

การรู้จำเสียงตัวเลขภาษาไทยแบบไม่ขึ้นกับผู้พูด โดยใช้แอลพีซี  
และนิรอรเน็ตเวิร์กแบบแบ็กพรอพากะชัน

นายวุฒิพงษ์ พรสุขจันทร์



วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต

ภาควิชาวิศวกรรมไฟฟ้า

บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2539

ISBN 974-635-547-3

ลิขสิทธิ์ของบัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย

I 1738 0546

**SPEAKER-INDEPENDENT THAI NUMERAL SPEECH RECOGNITION USING  
LPC AND THE BACK PROPAGATION NEURAL NETWORK**

**Mr. Wuthipong Pornsukjantra**

**A Thesis Submitted in Partial Fulfillment of the Requirements**

**for the Degree of Master of Engineering**

**Department of Electrical Engineering**

**Graduate School**

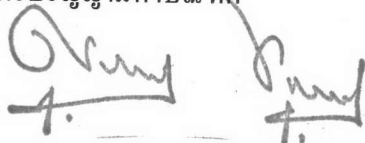
**Chulalongkorn University**

**Academic Year 1996**

**ISBN 974-635-547-3**

หัวข้อวิทยานิพนธ์ การรู้จำเสียงตัวเลขภาษาไทยแบบไม่ขึ้นกับผู้พูดโดยใช้แอลพีซีและ  
นิเวรอลเน็ตเวิร์กแบบแบ็กพรอพากะชัน  
โดย นาย วุฒิพงษ์ พรสุขจันทร์  
ภาควิชา ภาควิชาวิศวกรรมไฟฟ้า  
อาจารย์ที่ปรึกษา รองศาสตราจารย์ ดร. สมชาย จิตะพันธ์กุล

บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้บัณฑิตวิทยาลัยรับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรบัณฑิต

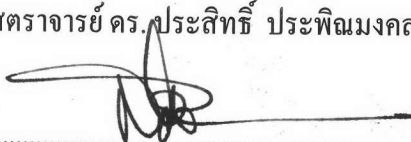


..... รักษาการแทนคณบดีบัณฑิตวิทยาลัย  
(ศาสตราจารย์ นายแพทย์ สุภวัฒน์ ชุตินวงศ์)

คณะกรรมการสอบวิทยานิพนธ์



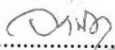
..... ประธานกรรมการ  
(ศาสตราจารย์ ดร. ประสิทธิ์ ประพัฒมงคลการ)



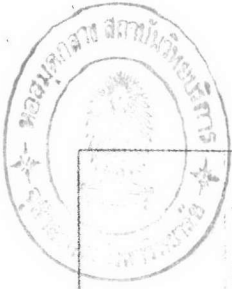
..... อาจารย์ที่ปรึกษา  
(รองศาสตราจารย์ ดร. สมชาย จิตะพันธ์กุล)



..... กรรมการ  
(ดร. จุฬารัตน์ ตันประเสริฐ)



..... กรรมการ  
(อาจารย์ ดร. วาทีต เบญจพลกุล)



## พิมพ์ต้นฉบับบทความวิทยานิพนธ์ภายในกรอบสี่เหลี่ยมนี้เพียงแผ่นเดียว

วุฒิพงษ์ พรสุขจันทรา : การรู้จำเสียงตัวเลขภาษาไทยแบบไม่ขึ้นกับผู้พูดโดยใช้แอลพีซี  
และนิวรอลเน็ตเวิร์กแบบแบ็กพรอพาเกชัน (SPEAKER-INDEPENDENT THAI NUMERAL  
SPEECH RECOGNITION USING LPC AND THE BACK PROPAGATION NEURAL  
NETWORK)


อาจารย์ที่ปรึกษา : รองศาสตราจารย์ ดร.สมชาย จิตะพันธ์กุล, 63 หน้า. ISBN 974-635-547-3

การวิจัยครั้งนี้มีจุดมุ่งหมายเพื่อพัฒนาวิธีการรู้จำเสียงตัวเลขภาษาไทยโดยใช้นิวรอลเน็ตเวิร์กแบบแบ็ก  
พรอพาเกชัน ลักษณะเด่นของเสียงพูดทำได้โดยใช้การประมาณพันธะเชิงเส้น(แอลพีซี) ชุดของสัมประสิทธิ์การ  
ประมาณพันธะเชิงเส้นใช้เป็นข้อมูลอินพุตสำหรับนิวรอลเน็ตเวิร์ก

กลุ่มข้อมูลที่ใช้เป็นเสียงตัวเลขพยางค์เดียวคือศูนย์ถึงเก้า และตัวเลขสองและสามพยางค์จำนวน 12 คำ  
กลุ่มคนในชุดฝึกมีจำนวน 30 คน กลุ่มคนในชุดทดสอบมีจำนวน 12 คน

ผลการวิจัยพบว่าผลการรู้จำแบบไม่ขึ้นกับผู้พูดเท่ากับ 89.4 % สำหรับคำพยางค์เดียว และเท่ากับ 84.7 %  
สำหรับคำ 2 และ 3 พยางค์

ภาควิชา ..... วิศวกรรมไฟฟ้า .....  
สาขาวิชา ..... วิศวกรรมไฟฟ้า .....  
ปีการศึกษา ..... 2539 .....

ลายมือชื่อนิสิต ..... วุฒิพงษ์ พรสุขจันทรา .....  
ลายมือชื่ออาจารย์ที่ปรึกษา .....  .....  
ลายมือชื่ออาจารย์ที่ปรึกษาร่วม ..... — .....

พิมพ์ต้นฉบับบทความวิทยานิพนธ์ภายในกรอบสี่เหลี่ยมนี้เพียงแผ่นเดียว

## C716075 : MAJOR ELECTRICAL ENGINEERING

KEY WORD: SPEECH RECOGNITION / NEURAL NETWORK / LPC

WUTHIPONG PORNUSUKJANTRA : SPEAKER-INDEPENDENT THAI NUMERAL SPEECH RECOGNITION USING LPC AND THE BACK PROPAGATION NEURAL NETWORK.

THESIS ADVISOR : ASSOC. PROF. SOMCHAI JITAPUNKUL, Ph.D.

63 pp. ISBN 974-635-547-3

This research has the objective to develop speaker-independent Thai numeral speech recognition using back propagation neural network. Feature of speech is extracted by linear predictive coding(LPC). Set of LPC coefficients is used as input data for neural network.

The data set is divided into 2 groups: 1. zero to nine Thai digits 2. 12 words of two and three syllables numeral speech. Number of persons in the training set is 30 persons. Other 12 persons are deployed in the test set.

The result of this research show that speaker-independent recognition rate is 89.4 % for one syllable speech and 84.7 % for two and three syllables speech.

ภาควิชา.....วิศวกรรมไฟฟ้า

สาขาวิชา.....วิศวกรรมไฟฟ้า

ปีการศึกษา..... 2539

ลายมือชื่อนิสิต..... วิทยานิพนธ์ วุฒิพงษ์

ลายมือชื่ออาจารย์ที่ปรึกษา.....

ลายมือชื่ออาจารย์ที่ปรึกษาร่วม..... -

## กิตติกรรมประกาศ

ข้าพเจ้าขอกราบขอบพระคุณ รองศาสตราจารย์ ดร. สมชาย จิตะพันธ์กุล อาจารย์ที่ปรึกษาวิทยานิพนธ์ ที่กรุณาสละเวลาให้คำปรึกษาและคำแนะนำต่าง ๆ ในการทำวิทยานิพนธ์ฉบับนี้ ข้าพเจ้าขอกราบขอบพระคุณ ศาสตราจารย์ ดร. ประสิทธิ์ ประพัฒน์มงคล, ดร. จุฬารัตน์ ต้นประเสริฐ, และอาจารย์ ดร. วาทิต เบญจพลกุล ที่ได้ให้คำแนะนำและคำวิจารณ์ที่เป็นประโยชน์เกี่ยวกับการทำวิทยานิพนธ์ฉบับนี้ ข้าพเจ้าขอกราบขอบพระคุณ อาจารย์ปราโมทย์ ศรีสุขสันต์ หัวหน้าห้องปฏิบัติการโทรคมนาคม ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ ที่ได้ให้โอกาส, คำแนะนำและให้ความรู้แก่ข้าพเจ้า

ข้าพเจ้าขอขอบพระคุณ นายธีระ ภัทรพรนันท์ ,นางสาวเสาวลักษณ์ อารีย์พงศา และ นายกิตติพงษ์ เจนวิถีสุข ที่ได้ให้คำแนะนำ,ข้อคิดเห็น และให้ข้อมูลที่เป็นประโยชน์ ข้าพเจ้าขอขอบพระคุณนายภาคภูมิ บุญญานันต์ สำหรับคำแนะนำที่เป็นประโยชน์ ข้าพเจ้าขอขอบพระคุณทุก ๆ ท่าน ที่ได้สละเวลาอันมีค่า มาช่วยบันทึกเสียงเพื่อใช้เป็นข้อมูลในการทดลอง โดยเฉพาะอย่างยิ่งเพื่อนทุกคนที่มีส่วนช่วยให้คำแนะนำและกำลังใจ

ข้าพเจ้าขอขอบพระคุณครอบครัวของข้าพเจ้าที่ให้การสนับสนุนและให้กำลังใจแก่ข้าพเจ้า โดยเฉพาะอย่างยิ่งคุณพ่อคุณแม่ที่ให้ทุกสิ่งทุกอย่างแก่ข้าพเจ้า ทั้ง ๆ ที่ข้าพเจ้าไม่ค่อยมีเวลาปฏิบัติตนเป็นลูกที่น่ารัก

## สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ง
บทคัดย่อภาษาอังกฤษ.....	จ
กิตติกรรมประกาศ.....	ฉ
สารบัญตาราง.....	ฉ
สารบัญรูปประกอบ.....	ญ
คำศัพท์.....	ฐ

### บทที่

1. บทนำ.....	1
2. หลักการรู้จำเสียงพูด.....	3
2.1 การตัดหัวท้ายคำ(End Point Detection).....	3
2.2 การนอร์มัลไลซ์(Normalization).....	4
2.3 การวัดค่าลักษณะสำคัญ(Feature measurement).....	5
2.4 การหาความคล้ายคลึงกันของรูปแบบ(Pattern Similarity Determination).....	9
2.3 กฎเกณฑ์การตัดสินใจ(Decision Rule).....	14
3. กระบวนการรู้จำเสียงพูดตัวเลขไทย.....	15
3.1 โปรแกรมตัดหัวท้ายคำและหาจำนวนพยางค์.....	15
3.2 การนอร์มัลไลซ์(Normalization).....	19
3.3 การประมาณพื้นที่เชิงเส้น.....	22
3.4 นิเวรอลเน็ตเวิร์ก.....	25
3.5 กฎเกณฑ์การตัดสินใจ.....	29
4. การทดลองและการวิเคราะห์ผลการทดลอง.....	30
4.1 ข้อมูลที่ใช้ในการทดสอบ.....	30
4.2 การวิเคราะห์หาค่าพารามิเตอร์ที่เหมาะสม.....	30
4.3 ผลการทดลอง.....	38
5. สรุปและวิจารณ์.....	49
5.1 สรุปผลการวิจัย.....	49

5.2 ปัญหา.....	49
5.3 ข้อเสนอแนะ .....	50
เอกสารอ้างอิง .....	51
ภาคผนวก ก.....	53
ภาคผนวก ข.....	59
ประวัติผู้เขียน.....	63



## สารบัญตาราง

ตารางที่	หน้า
4.1 ผลการทดลองหาค่า $b$ ที่เหมาะสม .....	32
4.2 ผลการทดลองหาค่า $a$ ที่เหมาะสม.....	33
4.3 ผลการทดลองหาค่า $m$ ที่เหมาะสม .....	33
4.4 ผลการทดลองหาค่า $n$ ที่เหมาะสม .....	34
4.5 ความถูกต้องของโปรแกรมนับจำนวนพยางค์.....	34
4.6 ค่า momentum และเวลาในการฝึก.....	36
4.7 ความสัมพันธ์ของจำนวนโหนดในระดับซ่อนตัวและอัตราการรู้จำ .....	38
4.8 อัตราการรู้จำของเสียงพูดกลุ่มที่ 1 โดยทดสอบตามจำนวนผู้พูดที่ใช้ในการฝึก .....	39
4.9 ผลการรู้จำเสียงพูดกลุ่ม A1 จำนวน 30 คนพูดคำละ 2 ครั้ง เมื่อฝึกนิรอลเน็ตเวิร์ก โดยใช้ผู้พูดกลุ่ม A1 จำนวน 30 คน .....	41
4.10 ผลการรู้จำเสียงพูดกลุ่ม A2 จำนวน 30 คนพูดคำละ 1 ครั้ง เมื่อฝึกนิรอลเน็ตเวิร์ก โดยใช้ผู้พูดกลุ่ม A1 จำนวน 30 คน .....	41
4.11 ผลการรู้จำเสียงพูดกลุ่ม B จำนวน 12 คนพูดคำละ 3 ครั้ง เมื่อฝึกนิรอลเน็ตเวิร์ก โดยใช้ผู้พูดกลุ่ม A1 จำนวน 30 คน .....	42
4.12 ผลการรู้จำเสียงพูดของแต่ละบุคคลในกลุ่ม A1,A2.....	43
4.13 ผลการรู้จำเสียงพูดของแต่ละบุคคลในกลุ่ม B โดยพูด 10 คำ คำละ 3 ครั้ง.....	44
4.14 อัตราการรู้จำของเสียงพูดกลุ่มที่ 2 โดยทดสอบตามจำนวนผู้พูดที่ใช้ในการฝึก .....	44
4.15 ผลการรู้จำเสียงพูดกลุ่ม A1 จำนวน 30 คนพูดคำละ 2 ครั้ง เมื่อฝึกนิรอลเน็ตเวิร์ก โดยใช้ผู้พูดกลุ่ม A1 จำนวน 30 คน .....	46
4.16 ผลการรู้จำเสียงพูดกลุ่ม A2 จำนวน 30 คนพูดคำละ 1 ครั้ง.....	46
4.17 ผลการรู้จำเสียงพูดกลุ่ม B จำนวน 12 คนพูดคำละ 3 ครั้ง.....	47

## สารบัญรูปประกอบ

รูปที่	หน้า
2.1 โมเดลของการรู้จำรูปแบบ.....	3
2.2 การนอร์แมลไลซ์โดยใช้การเปลี่ยนอัตราการจัดตัวอย่าง.....	5
2.3 โครงสร้างของการฝึก.....	10
2.4 โครงสร้างของ multi-layer perceptron neural network .....	11
2.5 รายละเอียดของโหนดในนิวรอลเน็ตเวิร์ก .....	12
3.1 รูปคลื่นของคำสองพยางค์ที่มีสัญญาณรบกวน.....	16
3.2 รูปคลื่นและพลังงานของคำพยางค์เดียว.....	17
3.3 ขั้นตอนของการตัดหัวท้ายเสียงพูด.....	18
3.4 จำนวนเสียงพูดที่ความยาวต่าง ๆ ของเสียง 1,2 และ 3 พยางค์.....	19
3.5 ขั้นตอนของการนอร์แมลไลซ์สัญญาณเสียงพูด.....	19
3.6 ตัวอย่างของการนอร์แมลไลซ์สัญญาณเสียงพูด.....	21
3.7 ขั้นตอนการคำนวณค่าสัมประสิทธิ์ LPC.....	22
3.8 ขั้นตอนการหาค่าสัมประสิทธิ์ LPC โดยใช้วิธีของ Levinson-Durbin.....	24
3.9 ขั้นตอนกระบวนการการเรียนรู้แบบ backpropagation.....	27
3.10 โครงสร้างของแฟ้มข้อมูลที่ใช้เก็บตัวอย่างคู่ข้อมูลอินพุตเอาต์พุต.....	29
4.1 ค่าระดับพลังงานของเสียง “เจ็ดสิบแปด”.....	31
4.2 ค่าระดับพลังงานของเสียง “สิบสอง” .....	31
4.3 อัตราการเรียนรู้ในแต่ละรอบของการฝึก .....	37
4.4 อัตราการรู้จำเทียบกับจำนวนคนในชุดฝึก .....	39
4.5 อัตราการรู้จำของเสียงพูดกลุ่มที่ 2 เทียบกับจำนวนคนในชุดฝึก.....	45
4.6 เสียงเจ็ดสิบแปดที่ผิดเป็นเสียงยี่สิบเอ็ด.....	47
4.7 เสียงยี่สิบสามที่ผิดเป็นเสียงเจ็ดสิบแปด.....	48

## คำอธิบายคำศัพท์

การตัดหัวท้ายคำ	end point detection
การนอร์มัลไลซ์	normalization
การวัดค่าลักษณะสำคัญ	feature measurement
การหาความคล้ายคลึงกันของรูปแบบ	pattern similarity determination
การประมาณค่าในช่วงเชิงเส้น	linear interpolation
การประมาณพหุระเชิงเส้น	Linear Predictive Coding(LPC)
วิธีอัตโนมัติสัมพันธ์	autocorrelation method
วิธีความแปรปรวนร่วม	covariance method
ผลรวมเชิงเส้น	linear combination
นิเวรอลเน็ตเวิร์ก	neural network
การฝึก	training
ชุดฝึก	training set
น้ำหนักการเชื่อมต่อ	connection weight
ตัวอย่างคู่ข้อมูลอินพุตเอาต์พุต	input output pattern