

การรู้จำเสียงพูดตัวเลขไทยโดยไม้อื่นต่อผู้พูดโดยการใช้ไดนามิกโทมวาร์บปีง



นาย ระพีพัฒน์ เพ็ญศิริ

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต

ภาควิชาวิศวกรรมไฟฟ้า

บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย

พ.ศ. 2538

ISBN 974-632-602-3

ลิขสิทธิ์ของบัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย

I 166546S7

SPEAKER-INDEPENDENT THAI NUMERICAL VOICE RECOGNITION
BY USING DYNAMIC TIME WARPING



Mr. Rapeepat Pensiri

A thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Engineering
Department of Electrical Engineering
Graduate School
Chulalongkorn University

1995

ISBN 974-632-602-3

หัวข้อวิทยานิพนธ์ การรู้จำเสียงพูดตัวเลขไทยโดยไม่ขึ้นต่อผู้พูดโดยการใช้ไดนามิก
โทมวาร์บปีง

โดย นาย ระพีพัฒน์ เพ็ญศิริ

ภาควิชา วิศวกรรมไฟฟ้า


อาจารย์ที่ปรึกษา รองศาสตราจารย์ ดร. สมชาย จิตะพันธ์กุล




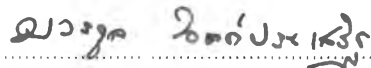
บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้บัณฑิตวิทยาลัยนี้เป็น
ส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาโทบัณฑิต

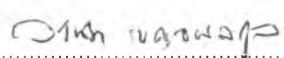

.....คนบดีบัณฑิตวิทยาลัย
(รองศาสตราจารย์ ดร. สันติ ฤงสูรธรรม)

คณะกรรมการสอบวิทยานิพนธ์


.....ประธานกรรมการ
(รองศาสตราจารย์ ดร. ณรงค์ อยู่ถนอม)


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


.....กรรมการ
(ดร. บวรกุล จิตต์ประเสริฐ)


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

พิมพ์ต้นฉบับบทคัดย่อวิทยานิพนธ์ภายในกรอบสี่เหลี่ยมนี้เพียงแผ่นเดียว



ระพีพัฒน์ เพ็ญศิริ : การรู้จำเสียงพูดตัวเลขไทยโดยไม่ขึ้นต่อผู้พูดโดยใช้ไดนามิกไทม์วาร์ปปีง
(SPEAKER-INDEPENDENT THAI NUMERICAL VOICE RECOGNITION BY USING DYNAMIC TIME
WARPING) อ.ที่ปรึกษา : รศ.ดร.สมชาย จิตะพันธ์กุล, 60 หน้า. ISBN 974-632-602-3

วิทยานิพนธ์ฉบับนี้มีจุดมุ่งหมายเพื่อศึกษาการรู้จำตัวเลขไทยแบบไม่ขึ้นต่อผู้พูดโดยใช้ไดนามิกไทม์วาร์ปปีง การวิเคราะห์ข้อมูลเสียงเพื่อหารูปแบบจะใช้เป็นค่าเดียว โดยการใช้สถิติสหสัมพันธ์ของเสียงในแต่ละเฟรม จากนั้นทำ พารามิเตอร์ของรูปแบบของค่านั้น ๆ จากนั้นทำการคำนวณหา distance ระหว่างแบบทดสอบ (test pattern) กับแบบอ้างอิง (reference pattern)

ผลการศึกษาด้วยวิธีการที่นำเสนอนี้ โดยทำการทดสอบบนเครื่อง IBM PC/AT compatible โดยอัตราการรู้จำของเสียงตัวเลขไทย 0 - 9 โดยไม่ขึ้นต่อผู้พูดจะมีค่าเป็น 79.25 % และอัตราความถูกต้องภายในกลุ่มที่ใช้สร้างแบบอ้างอิงจำนวน 20 คน จำนวน 600 คำ จะได้ 87.17 % และอัตราการรู้จำของเสียงตัวเลขไทย 0 - 9, "สิบ", "เอ็ด", "ยี่", "ร้อย", "พัน", "หมื่น", "แสน", และ "ล้าน" โดยไม่ขึ้นต่อผู้พูดภายในกลุ่มที่สร้างแบบอ้างอิงจำนวน 20 คน จำนวน 1080 คำ จะได้ 74.07 % ผลการรู้จำเสียงพูดที่ดีจะขึ้นกับการเลือกใช้พารามิเตอร์ในการแทนเสียงพูด และจะเห็นได้ว่าการนำเอาไดนามิกไทม์วาร์ปปีงมาใช้ร่วมกับเทคนิคนี้นั้นเหมาะกับการรู้จำเสียงที่ไม่มากแบบ

ภาควิชา วิศวกรรมไฟฟ้า
สาขาวิชา DSP
ปีการศึกษา ๒๕๓๕

ลายมือชื่อนิสิต
ลายมือชื่ออาจารย์ที่ปรึกษา
ลายมือชื่ออาจารย์ที่ปรึกษาร่วม



C515643 : MAJOR ELECTRICAL ENGINEERING
KEY WORD: SPEAKER-INDEPENDENT / THAI NUMERICAL / VOICE RECOGNITION / DYNAMIC TIME WARPING
RAPEEPAT PENSIRI : SPEAKER-INDEPENDENT THAI NUMERICAL VOICE RECOGNITION BY USING DYNAMIC TIME WARPING. THESIS ADVISOR : ASSO. PROF. SOMCHAI JITAPHUNKUL, Dr. 60 pp. ISBN 974-632-602-3

This thesis has the objective to study on speaker-independent Thai numerical voice recognition by using dynamic time warping. In analysis to find a pattern uses isolated word by discrete Hartley transform in each frame of voice. Then, to find parameters of pattern of each word, after that to calculate distance between a test pattern and a reference pattern.

This proposed method results the zero to nine independent voice recognition rate 79.25 % with 20 testing persons, 87.17 % with 20 training persons with 600 words and zero to nine, "sib", "ed", "yee", "roy", "pan", "hmuan", "san", "lan" independent voice recognition rate 74.07 % with 20 training persons with 1080 word, by testing on IBM PC/AT compatible. Good voice recognition result is depended on voice parameter selection and shows that using DTW for this technique is appropriate for no many voice recognized patterns.

ภาควิชา..... วิศวกรรมไฟฟ้า.....
สาขาวิชา..... DSP.....
ปีการศึกษา..... ๒๕๓๘.....

ลายมือชื่อนิสิต.....
ลายมือชื่ออาจารย์ที่ปรึกษา.....
ลายมือชื่ออาจารย์ที่ปรึกษาร่วม.....



กิตติกรรมประกาศ

ข้าพเจ้าขอกราบขอบพระคุณ รองศาสตราจารย์ ดร. สมชาย จิตะพันธ์กุล อาจารย์ที่ปรึกษาวิทยานิพนธ์ กรุณาสละเวลาให้คำปรึกษาและคำแนะนำต่าง ๆ ในการทำวิทยานิพนธ์ จนกระทั่งสามารถทำงานสำเร็จลุล่วงไปได้ด้วยดี ข้าพเจ้าขอกราบขอบพระคุณ รองศาสตราจารย์ ดร. ณรงค์ อยู่ถนอม, ดร. บวรกุล จิตต์ประเสริฐ, และอาจารย์ ดร. วาทิต เบญจพลกุล ที่ได้ให้คำแนะนำและวิจารณ์ที่เป็นประโยชน์เกี่ยวกับการทำวิทยานิพนธ์ฉบับนี้

นอกจากนี้ข้าพเจ้าขอขอบคุณ คุณสนธยา เมรินทร์ ที่ได้ช่วยเหลือในการพัฒนาโปรแกรมสำหรับจัดทำเมนู และขอขอบนิสิตที่หน่วยปฏิบัติการ DSP และที่หอพักศึกษิตินิเวศน์ ทุกคนที่สละเวลาในการจัดเก็บข้อมูลเสียง รวมทั้งขอขอบคุณผู้มีพระคุณทุกท่านที่ไม่ได้กล่าวถึง ณ ที่นี้ ที่มีส่วนช่วยในการให้ความช่วยเหลือและให้กำลังใจ จนกระทั่งวิทยานิพนธ์ฉบับนี้สำเร็จด้วยดี



สารบัญ

	หน้า
บทคัดย่อภาษาไทย	ง
บทคัดย่อภาษาอังกฤษ	จ
กิตติกรรมประกาศ	ฉ
สารบัญตาราง	ฉ
สารบัญภาพ	ฎ

บทที่

1. บทนำ	1
1.1 ความเบื้องต้น	2
1.2 วัตถุประสงค์ของงานวิจัย	2
1.3 ขอบเขตของงานวิจัย	2
1.4 ขั้นตอนและวิธีการ	2
1.5 เป้าหมายของงานวิจัย	3
2. การวิเคราะห์เสียงพูด	4
2.1 เสียงที่ใช้ในการวิเคราะห์	4
2.2 การวิเคราะห์สัญญาณ	5
2.2.1 การวิเคราะห์สัญญาณในเชิงเวลา	6
2.2.2 การวิเคราะห์สัญญาณในเชิงความถี่	8
2.3 ฟังก์ชันหน้าต่าง	10
2.4 การตัดคำ	12
3. การรู้จำเสียงพูด	16
3.1 โครงสร้างของระบบการรู้จำ	16
3.1.1 feature measurement	18
3.1.2 time registration of pattern	19
3.1.3 decision rule for recognition	20

3.2 ไดนามิกโทมัวร์ปิง	22
3.3 การกำหนดพารามิเตอร์และวิธีการวัดในการวิเคราะห์เสียง ..	33
3.4 การสร้างแบบอ้างอิง	35
4. การวิเคราะห์และผลการทดสอบ	36
4.1 ข้อมูลที่ใช้ในการทดสอบ	36
4.2 ผลการทดสอบ	37
5. บทสรุป	44
5.1 สรุปและวิจารณ์	44
5.2 ข้อเสนอแนะ	45
รายการอ้างอิง	46
ภาคผนวก ก	49
ภาคผนวก ข	53
ภาคผนวก ค	57
ประวัติผู้เขียน	60

สารบัญตาราง

ตารางที่	หน้า
3.1.1 แสดงการเปรียบเทียบข้อดี ข้อเสียของการเลือกใช้ template	17
3.2.1 ตัวอย่างของชนิดของ local constraints	25
3.2.2 แสดงสมการไดนามิกโปรแกรมมิ่งต่าง ๆ	29
3.3.1 แสดงการกำหนดค่าของพารามิเตอร์เพื่อการรู้จำ	34
4.2.1 แสดงผลการรู้จำของเสียงกลุ่มที่ 1 (0-9) โดยใช้แบบอ้างอิง	37
ที่จำนวนผู้พูดต่าง ๆ กัน	
4.2.2 แสดงผลการรู้จำของกลุ่ม A1 ต่อแบบอ้างอิงที่สร้างจากผู้พูด	38
จำนวน 20 คน	
4.2.3 แสดงผลการรู้จำของกลุ่ม A2,A3 ต่อแบบอ้างอิงที่สร้างจากผู้พูด	39
จำนวน 20 คน	
4.2.4 แสดงผลการรู้จำของกลุ่ม B ต่อแบบอ้างอิงที่สร้างจากผู้พูด	40
จำนวน 20 คน	
4.2.5 แสดงผลการรู้จำของเสียงกลุ่มที่ 2 ต่อแบบอ้างอิงที่สร้างจากผู้พูด	40
จำนวน 20 คน	
4.2.6 แสดงผลการรู้จำของกลุ่ม A1 ต่อแบบอ้างอิงที่สร้างจากผู้พูด	41
จำนวน 20 คน โดยทดสอบเสียงกลุ่มที่ 2	
4.2.7 แสดงผลการรู้จำของกลุ่ม A2,A3 ต่อแบบอ้างอิงที่สร้างจากผู้พูด	42
จำนวน 20 คน โดยทดสอบเสียงกลุ่มที่ 2	
4.2.8 แสดงผลการรู้จำของกลุ่ม B ต่อแบบอ้างอิงที่สร้างจากผู้พูด	43
จำนวน 20 คน โดยทดสอบเสียงกลุ่มที่ 2	
ก.1 แสดงผลของพารามิเตอร์ a,b,และ c ของการตัดคำที่มีผลต่อการรู้จำ	49
โดยใช้แบบอ้างอิงจำนวน 20 คน แบบทดสอบจำนวน 20 คน	
ก.2 แสดงผลการทดสอบการเลือกใช้รูปแบบพารามิเตอร์	50
ก.3 แสดงผลการรู้จำของแบบทดสอบ B ต่อแบบอ้างอิง A1 ในช่วงความถี่ต่าง ๆ	51
ก.4 แสดงผลการรู้จำของการปรับขนาดความยาวเสียงพูดของแบบอ้างอิง	52

สารบัญภาพ

ภาพที่	หน้า
2.2.1 ตัวอย่างสัญญาณเสียงคำว่า “หนึ่ง”	5
2.2.2 แสดงรูปคลื่นในแต่ละเฟรมของคำว่า “หนึ่ง” ขนาดของเฟรมเท่ากับ 25 มิลลิวินาที	6
2.2.1.1 แสดงพลังงานของสัญญาณเสียงของคำว่า “ หนึ่ง” ตามสมการที่ 2.2.4 โดย N_1 เท่ากับ 100	7
2.2.2.1 ขั้นตอนการแปลงข้อมูลเสียง	9
2.3.1 (ก) แสดงรูปของ window function แบบต่าง ๆ	11
(ข) แสดงถึง spectrum ของสัญญาณ ที่ใช้ window function แบบ ต่าง ๆ ใน (ก)	11
2.4.1 แสดงจุดอ้างอิงเพื่อหาจุดเริ่มต้นและจุดสิ้นสุดของรูปคลื่นพลังงาน	13
2.4.2 แสดงขั้นตอนการตัดคำ	14-15
3.1.1 โครงสร้างของระบบการรู้จำ	16
3.1.2 โครงสร้างระบบการรู้จำแบบ isolated word recognition	18
3.1.2.1 ตัวอย่างของ time registration ของแบบทดสอบและแบบอ้างอิง	19
3.2.1 DTW ระหว่าง A และ B	22
3.2.2 แสดง local path constraints ที่ไปยังจุด (n,m)	24
3.2.3 ตัวอย่างของ weighting function of Type 2 constraints	26
3.2.4 ตัวอย่างการทำ smoothed weighting function ของ Type 1 constraints	27
3.2.5 ขั้นตอนการทำ DTW	31
3.2.6 แสดงการใช้ normalize/warp DTW algorithm	32
3.2.7 แสดงการเปรียบเทียบระหว่าง standard DTW และ normalize/warp DTW	33
ข.1 แสดงแบบอ้างอิงที่สร้างจากผู้พูดจำนวน 20 คน	53-56