การรู้จำเสียงพูดตัวเลขไทยโดยไม่ขึ้นต่อผู้พูดโดยการใช้ไดนามิกไทม์วาร์ปปิง

นาย ระพีพัฒน์ เพ็ญศิริ

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต
ภาควิชาวิศวกรรมไฟฟ้า
บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย
พ.ศ. 2538

ISBN 974-632-602-3

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	การรู้จำเสียงพูดตัวเลขไทยโดยไม่ขึ้นต่อผู้พูดโดยการใช้ไดนามิกไทม์วาร์ปปิง
โดย	นาย ระพีพัฒน์ เพ็ญศิริ
ภาควิชา	วิศวกรรมไฟฟ้า
อาจารย์ที่ปรึกษา	รองศาสตราจารย์ ดร. สมชาย จิตะพันธ์กุล

บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

.....คณบดีบัณฑิตวิทยาลัย

(รองศาสตราจารย์ ดร. สันติ ถุงสุวรรณ)

คณะกรรมการสอบวิทยานิพนธ์

.....ประธานกรรมการ

(รองศาสตราจารย์ ดร. ณรงค์ อยู่ถนอม)

.....อาจารย์ที่ปรึกษา

(รองศาสตราจารย์ ดร. สมชาย จิตะพันธ์กุล)

.....กรรมการ

(ดร. บวรกูล จิตต์ประเสริฐ)

.....กรรมการ

(อาจารย์ ดร. วาทิต เบญจพลกุล)

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

วิทยานิพนธ์ฉบับนี้มีจุดมุ่งหมายเพื่อศึกษาการรู้จำตัวเลขไทยแบบไม่ขึ้นต่อผู้พูดโดยใช้ไดนามิกไทม์วาร์ปปิง การวิเคราะห์ข้อมูลเสียงเพื่อหารูปแบบจะใช้เป็นคำเดี่ยว โดยการใช้ดีสครีตทรานส์ฟอร์มของเสียงในแต่ละเฟรม จากนั้นทำหาพารามิเตอร์ของรูปแบบของคำนั้น 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, "ศูนย์", "เอ็ด", "ยี่", "ร้อย", "พัน", "หมื่น", "แสน", และ "ล้าน" โดยไม่ขึ้นต่อผู้พูดภายในกลุ่มที่สร้างแบบอ้างอิงจำนวน 20 คน จำนวน 1080 คำ จะได้ 74.07 % ผลการรู้จำเสียงพูดที่ดีจะขึ้นกับการเลือกใช้พารามิเตอร์ในการแทนเสียงพูด และจะเห็นได้ว่าการนำเอาไดนามิกไทม์วาร์ปปิงมาใช้ร่วมกับเทคนิคนี้นั้นเหมาะกับการรู้จำเสียงที่ไม่มากแบบ

ภาควิชา วิศวกรรมไฟฟ้า ลายมือชื่อนิสิต .....

สาขาวิชา DSP ลายมือชื่ออาจารย์ที่ปรึกษา .....

ปีการศึกษา 2538 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 ลายมือชื่อนิสิต .....

สาขาวิชา DSP ลายมือชื่ออาจารย์ที่ปรึกษา .....

ปีการศึกษา 2538 ลายมือชื่ออาจารย์ที่ปรึกษาร่วม .....

กิตติกรรมประกาศ

ข้าพเจ้าขอกราบขอบพระคุณ รองศาสตราจารย์ ดร. สมชาย จิตะพันธ์กุล อาจารย์ที่ปรึกษาวิทยานิพนธ์ กรุณาสละเวลาให้คำปรึกษาและคำแนะนำต่าง ๆ ในการรรทำวิทยานิพนธ์จนกระทั่งสามารถทำงานสำเร็จลุล่วงไปได้ด้วยดี ข้าพเจ้าขอกราบขอบพระคุณ รองศาสตราจารย์ ดร. ณรงค์ อยู่ถนอม, ดร. บวรกูล จิตต์ประเสริฐ, และอาจารย์ ดร. วาทิต เบญจพลกุล ที่ได้ให้คำแนะนำและวิจารณ์ที่เป็นประโยชน์เกี่ยวกับการทำวิทยานิพนธ์ฉบับนี้

นอกจากนี้ข้าพเจ้าขอขอบคุณ คุณสนธยา เมรินทร์ ที่ได้ช่วยเหลือในการพัฒนาโปรแกรมสำหรับจัดทำเมนู และขอขอบนิสิตที่หน่วยปฏิบัติการ DSP และที่หอพักศึกษิตนิเวศน์ทุกคนที่สละเวลาในการจัดเก็บข้อมูลเสียง รวมทั้งขอขอบคุณผู้มีพระคุณทุกท่านที่ไม่ได้กล่าวถึง ณ ที่นี้ ที่มีส่วนช่วยในการให้ความช่วยเหลือและให้กำลังใจ จนกระทั่งวิทยานิพนธ์ฉบับนี้สำเร็จด้วยดี

สารบัญ

สารบัญตาราง

สารบัญภาพ