

ระบบการรับรู้เสียงพูดแบบต่างบุคคล



นาย ไพศาล ธรรมไพฑิถอง

ศูนย์วิทยทรัพยากร

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

ภาควิชาวิศวกรรมคอมพิวเตอร์

บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย

พ.ศ. 2533

ISBN 974-577-834-6

ลิขสิทธิ์ของบัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย

MULTISPEAKER RECOGNITION SYSTEM

Mr. Paisan Thumpothong

คณะวิทยาศาสตร์
จุฬาลงกรณ์มหาวิทยาลัย

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science
Graduate School

Chulalongkorn University

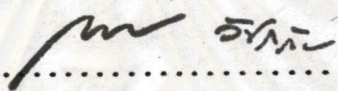
1990

ISBN 974-577-834-6

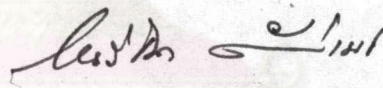


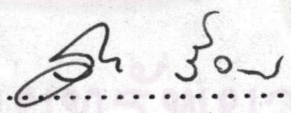
หัวข้อวิทยานิพนธ์ ระบบการรับรู้เสียงพูดแบบต่างบุคคล
โดย นาย ไพศาล ธรรมโพธิทอง
ภาควิชา วิศวกรรมคอมพิวเตอร์
อาจารย์ที่ปรึกษา ผู้ช่วยศาสตราจารย์ ดร. วีระ ธีรวิทักษ์

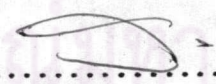
บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้มหาวิทยาลัยแห่งนี้เป็นส่วนหนึ่งของ
ของการศึกษาตามหลักสูตรปริญญาโทบัณฑิต

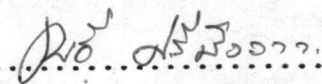

..... คณบดีบัณฑิตวิทยาลัย
(ศาสตราจารย์ ดร. ถาวร วัชรวิทย์)

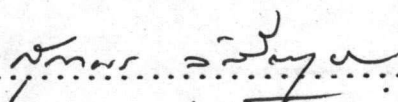
คณะกรรมการสอบวิทยานิพนธ์


..... ประธานกรรมการ
(รองศาสตราจารย์ ไกรวิชิต ตันติเมธ)


..... อาจารย์ที่ปรึกษา
(ผู้ช่วยศาสตราจารย์ ดร. วีระ ธีรวิทักษ์)


..... กรรมการ
(รองศาสตราจารย์ สมชาย ทยานง)


..... กรรมการ
(ผู้ช่วยศาสตราจารย์ เมธิ ศรีสิงวาล)


..... กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร. สูดานพร ลิขณยานาวิน)



ไพศาล ธรรมโพธิทอง : ระบบการรับรู้เสียงพูดแบบต่างบุคคล
(MULTISPEAKER RECOGNITION SYSTEM) อ.ที่ปรึกษา : ผศ.ดร. วีระ ธีรวิทักษ์,
109 หน้า, ISBN 974-577-834-6

ระบบการรับรู้เสียงพูดที่ทำการวิจัยนี้ เป็นแบบคำเดี่ยวต่างบุคคล ซึ่งขั้นตอนในการวิจัยแบ่งเป็น 3 ส่วน คือขั้นตอนแรกจะทำการหาค่าสัมประสิทธิ์ของสัญญาณ และค่าคาบเสียงสูงต่ำของกลุ่มคำ 2 กลุ่ม เพื่อนำมาสร้างเป็นแบบอ้างอิงจากผู้พูด 30 คน ขั้นตอนที่สองกลุ่มของแบบอ้างอิงทั้ง 2 กลุ่ม จะถูกนำมาผ่านการคัดเลือกให้ได้ค่าละ 10 แบบ โดยอาศัยเทคนิคการจัดกลุ่ม และทำการให้เลเบล ขั้นที่สามจะทำการทดสอบโดยอาศัยแบบทดสอบจากผู้พูด 5 คน มาทำการเปรียบเทียบกับแบบอ้างอิง การคำนวณหา ค่าความแตกต่างระหว่างแบบทดสอบกับแบบอ้างอิง ได้ใช้เทคนิคของไดนามิกโปรแกรมมิ่ง และกฎการตัดสินใจได้อาศัยเทคนิคของ K-Nearest Neighbour (KNN) การกำหนดค่าความแตกต่างที่ยอมรับได้ ถูกนำมาใช้เพื่อปรับปรุงอัตราความถูกต้องในการรับรู้ การทำ Post Processing โดยใช้ค่า Zero Crossing เพื่อปรับปรุงอัตราความถูกต้องของการรับรู้ระหว่างคำว่า "ปาก" และ "มาก"

ผลการวิจัยจากคำทดสอบกลุ่มแรก ได้อัตราความถูกต้องร้อยละ 85.8 และคำทดสอบกลุ่มที่สอง ได้อัตราความถูกต้องร้อยละ 94.3

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ภาควิชา
สาขาวิชา
ปีการศึกษา 2532

ลายมือชื่อนิติต
ลายมือชื่ออาจารย์ที่ปรึกษา
ลายมือชื่ออาจารย์ที่ปรึกษาร่วม



PAISAN THUMPOTHONG : "MULTISPEAKER RECOGNITION SYSTEM. THESIS
ADVISOR : ASST. PROF. VEERA RIEWPITAK, 109 PP.

This multispeaker voice recognition system is an isolated word : speaker independent system. The procedure in developing this system is of 3 steps 1) The analysis of 2 sets of reference words from 30 speakers in terms of the reflection coefficient and pitch period values, 2) the reference patterns of these 2 sets of words are created and 10 desired patterns are created by Clustering Technique and then labelled, 3) the test patterns from 5 subjects are compared with the reference patterns. The distance between the test and the reference patterns is formed by the Dynamic Programming Technique. The decision rules are created by using the K-nearest neighbour (KNN) technique. A range of acceptable distance value has been set to improve recognition rate. Post processing using Zero crossing was also added to improve the recognition rate in the word pair /pa:k/ and /ma:k/.

85.8 % of accurate recognition is formed in the first set of words and 94.3 % is formed in the second set of words.

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ภาควิชา
สาขาวิชา
ปีการศึกษา 2532

ลายมือชื่อนิสิต
ลายมือชื่ออาจารย์ที่ปรึกษา
ลายมือชื่ออาจารย์ที่ปรึกษาร่วม



กิตติกรรมประกาศ

ผู้วิจัยขอขอบพระคุณ ผศ. ดร. วีระ วีระพิทักษ์ ที่ได้กรุณาเสียสละเวลาให้คำปรึกษา
ชี้แนะ รวมทั้งแนวคิดต่าง ๆ และกำลังใจช่วยให้การทำวิทยานิพนธ์ฉบับนี้สำเร็จไปได้ด้วยดี

ขอขอบคุณหัวหน้า และเพื่อนร่วมงานที่บริษัท ที่เห็นความสำคัญของการศึกษา และสนับสนุน
ด้วยดีตลอดมา

ขอขอบคุณ คุณอาทร นันทิกุล คุณวัลณพ ตันถิติ คุณชลธิชา กุลกิตติมาศ ที่ช่วยให้
คำปรึกษา และกำลังใจในการทำวิทยานิพนธ์ด้วยดี

ขอขอบคุณ คุณชัยยง ภักดีกิจเจริญ ที่ช่วยพัฒนาบอร์ด CU-TSB 32010 จนสามารถ
ใช้ในการวิจัยได้เป็นอย่างดี

ขอขอบคุณ เพื่อนนักศึกษาทุกท่านที่ให้ความร่วมมือในการเก็บข้อมูลเพื่อทำการวิจัยเป็น
อย่างดี การวิจัยในครั้งนี้สำเร็จได้ก็มาจากข้อมูลเหล่านี้

สุดท้ายผู้วิจัยขอกราบขอบพระคุณ บิดา มารดา ผู้ให้กำเนิด และมีอุปการะคุณอันหาที่
เปรียบมิได้มาตลอด

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย



สารบัญ

	หน้า
บทคัดย่อภาษาไทย	ง
บทคัดย่อภาษาอังกฤษ	จ
กิตติกรรมประกาศ	ฉ
สารบัญตาราง	ญ
สารบัญภาพ	ฎ
 บทที่	
1. บทนำ	1
1.1 ความเป็นมาของปัญหา	1
1.2 วัตถุประสงค์ของวิทยานิพนธ์	4
1.3 ขอบเขตของวิทยานิพนธ์	4
1.4 ขั้นตอนและวิธีดำเนินการ	4
1.5 ประโยชน์ที่คาดว่าจะได้รับจากการทำวิทยานิพนธ์	4
2. การวิเคราะห์เสียงพูด	5
2.1 เสียงพูด	5
2.1.1 อวัยวะที่ใช้ในการออกเสียง	5
2.1.2 การเกิดของเสียง	8
2.2 หลักการในการวิเคราะห์เสียงพูด	8
2.2.1 การเข้ารหัสสัญญาณดิจิทัล	9
2.2.2 การวิเคราะห์ในโดเมนเวลา	10
2.2.3 การวิเคราะห์ในโดเมนความถี่	12
2.3 การวิเคราะห์หาขอบเขตของคำ	14

2.4	การหาพารามิเตอร์ฟิลเตอร์โดยวิธีแอลพีซี	16
2.4.1	วิธีโควาเรียนซ์	19
2.4.2	วิธีอัลทล์สมพันธ์	20
2.4.3	วิธีพาร์คอร์	21
2.4.4	หลักการอินเนอร์โปรดักท์และคุณสมบัติออร์โธโกนัล ของตัวทำนาย	23
2.4.5	การหาสัมประสิทธิ์พาร์คอร์	25
2.4.6	กรรมวิธีการคำนวณค่าสัมประสิทธิ์พาร์คอร์	27
2.5	การหาคาบของสัญญาณ	36
3.	การรับรู้เสียงพูด	42
3.1	การเลือกแบบอ้างอิงเพื่อใช้ในการรับรู้	42
3.1.1	การหาจุดศูนย์กลางของกลุ่ม	43
3.1.2	การจัดกลุ่มใหม่	43
3.2	การเปรียบเทียบแบบไดนามิกโปรแกรมมิ่ง	43
3.2.1	ฟังก์ชันแวร์บิง	44
3.2.2	สมการไดนามิกโปรแกรมมิ่ง	48
3.3	กฎการตัดสินใจ	49
3.3.1	กฎ Nearest Neighbour	52
3.3.2	กฎ K-Nearest Neighbour	52
4.	การพัฒนาระบบการรับรู้เสียงพูดบนไมโครคอมพิวเตอร์	53
4.1	โปรแกรมรับข้อมูล	56
4.2	โปรแกรมการวิเคราะห์	58
4.2.1	โปรแกรมหาขอบเขตของสัญญาณ	58
4.2.2	โปรแกรมหาค่าสัมประสิทธิ์พาร์คอร์	60
4.2.3	โปรแกรมหาคาบของสัญญาณ	60
4.3	โปรแกรมการรับรู้	63
4.3.1	โปรแกรมสร้างแฟ้มข้อมูลพารามิเตอร์	63
4.3.2	โปรแกรมคัดเลือกแบบอ้างอิง	66
4.3.3	โปรแกรมการรับรู้	66

5. การทดสอบและการประเมินผล	69
5.1 ข้อมูลทดสอบ	69
5.2 ผลการทดสอบ	70
6. สรุปและข้อเสนอแนะ	76
6.1 สรุป	76
6.2 ข้อเสนอแนะ	77
6.3 ข้อเสนอแนะในการประยุกต์ใช้งาน	78
เอกสารอ้างอิง	79
ภาคผนวก	91
ภาคผนวก ก หลักเกณฑ์คัดเลือกคำที่ใช้ในการทดสอบ	92
ภาคผนวก ข รายละเอียดและการใช้งานของบอร์ด TSB	95
ภาคผนวก ค คุณสมบัติของไมโครโปรเซสเซอร์ TMS32010	100
ภาคผนวก ง แสดงตัวอย่างของภาพสัญญาณเสียงที่ใช้ในการทดสอบ ..	105
ประวัติผู้เขียน	109

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญตาราง

ตารางที่		หน้า
1.1	แสดงการเปรียบเทียบข้อดี ข้อเสีย ของการเลือกหน่วยเสียง	3
1.2	แสดงถึงวิธีการจำแนกวิธีการรับรู้เสียงพูดต่าง ๆ	3
2.2.1	แสดงถึงตัวอย่างความสัมพันธ์ของโดเมนความถี่กับโดเมนเวลา	14
2.4.2	แสดงถึงข้อดีข้อเสียในโปรแกรม AUTO เกี่ยวกับสัญลักษณ์	31
3.2.1	แสดงสมการไดนามิคโปรแกรมมิ่งต่าง ๆ เมื่อเงื่อนไขการเปลี่ยนแปลง ความชันเปลี่ยนไป	50
5.1	แสดงการกระจายความถูกต้องในการรับรู้ของคำในกลุ่มที่ 1	72
5.2	แสดงการกระจายความถูกต้องในการรับรู้ของคำในกลุ่มที่ 2	73
5.3	แสดงการกระจายความถูกต้องในการรับรู้ของคำในกลุ่มที่ 1 หลังจากกำหนดค่าความแตกต่างที่ยอมรับได้	74
5.4	แสดงการกระจายความถูกต้องในการรับรู้ของคำในกลุ่มที่ 2 หลังจากกำหนดค่าความแตกต่างที่ยอมรับได้	75
ค.1	แสดงเงื่อนไขการทดสอบตัวสะสม	92

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญภาพ

ภาพที่		หน้า
1.1	แสดงถึงรูปคลื่นเสียงของคำว่า "เอ" ที่พูดจากบุคคลเดียวกัน 3 ครั้ง ..	2
2.1.1	แสดงอวัยวะที่ใช้ในการออกเสียงพูด	6
2.2.1	แสดงแผนภาพหลักการทำงานของดีฟเฟอเรนเชียลควอนไทเซชัน	10
2.2.2	แสดงส่วนประกอบของแซมเปิลแอนด์โฮลดิ้ง	12
2.3.1	แสดงถึงการกำหนดจุดอ้างอิงเพื่อหาจุดเริ่มต้น และจุดสิ้นสุดของรูปคลื่นพลังงาน	16
2.4.1	แสดงถึงแบบจำลองของการวิเคราะห์ และการสังเคราะห์เสียงพูดด้วยวิธีแอลพีซี	18
2.4.2	แสดงถึงชุดของสัญญาณส่งและการทำนายทั้งสองทาง	22
2.4.3	แสดงถึงอินเนอร์โปรดักต์ของฟิลเตอร์ $F(z)$ กับ $G(z)$	23
2.4.4	แสดงถึงโปรแกรมย่อย AUTO	32
2.5.1	แสดงถึงลักษณะอินพุท เอาท์พุทและกระบวนการของการทำเซนเตอร์คลิปปิง	37
2.5.2	แสดงขั้นตอนทั้งหมดของการหาคาบของสัญญาณที่ทำโดยอาศัยฮาร์ดแวร์ ..	39
2.5.3	แสดงถึงลักษณะอินพุท เอาท์พุทของการทำเซนเตอร์คลิปปิง และการกำหนดค่าสัญญาณใหม่	40
2.5.4	แสดงถึงตัวอย่างสัญญาณที่ผ่านการทำเซนเตอร์คลิปปิง และการหาฟังก์ชันออโต	40
3.2.1	แสดงให้เห็นลักษณะของฟังก์ชันแอมพลิจูด	45
3.2.2	แสดงถึงเงื่อนไขการเปลี่ยนความชันของฟังก์ชันแอมพลิจูด	47
3.2.3	แสดงถึงโปรแกรมย่อย DYNAMIC	51
4.1	แสดงถึงความสัมพันธ์ของโปรแกรมต่าง ๆ ในระบบการรับรู้เสียงพูดด้วยไมโครคอมพิวเตอร์	55
4.2	แสดงขั้นตอนการทำงานของโปรแกรมรับข้อมูล	57
4.3	แสดงขั้นตอนการทำงานของโปรแกรมหาขอบเขตของสัญญาณ	59
4.4	แสดงขั้นตอนการทำงานของโปรแกรมหาค่าสัมประสิทธิ์พาร์คอร์	61
4.5	แสดงขั้นตอนการทำงานของโปรแกรมหาค่าคาบของสัญญาณ	62
4.6	แสดงถึงโครงสร้างของแฟ้มข้อมูลพารามิเตอร์ และแฟ้มข้อมูลลาเบล ...	64

4.7	แสดงขั้นตอนของโปรแกรมสร้างแฟ้มข้อมูลพารามิเตอร์	65
4.8	แสดงขั้นตอนของโปรแกรมคัดเลือกแบบอ้างอิง	67
4.9	แสดงขั้นตอนของโปรแกรมการรู้	68
ข.1	แสดงขั้นตอนของวงจรการขยายสัญญาณอนาล็อกเข้า	87
ข.2	แสดงขั้นตอนของวงจรการขยายสัญญาณอนาล็อกเข้า	88
ค.1	แสดงการทำงานของสถาปัตยกรรมฮาร์ดแวร์	90
ค.2	แสดงตัวนับของเรจิสเตอร์เสริม	94
ค.3.1	แสดงถึงการเพิ่มของตัวนับโดยอัตโนมัติ	95
ค.3.2	แสดงถึงการลดของตัวนับโดยอัตโนมัติ	95
ค.4	แสดงถึงองค์ประกอบของเรจิสเตอร์สถานะ	96
ค.5	แสดงถึงเรจิสเตอร์สถานะที่บันทึกโดยคำสั่ง SST	97
ค.6	แสดงถึงวิธีการติดต่อกับอุปกรณ์ภายนอก	98
ค.7.1	แสดงถึงช่วงเวลาของคำสั่ง IN	99
ค.7.2	แสดงถึงช่วงเวลาของคำสั่ง OUT	99
ค.8.1	แสดงถึงช่วงเวลาของคำสั่ง TBLR	100
ค.8.2	แสดงถึงช่วงเวลาของคำสั่ง TBLW	100
ค.9	แสดงถึงแผนผังลอจิกในการขัดจังหวะอย่างง่าย	102
ค.10	แสดงถึงช่วงเวลาในขณะที่เกิดการขัดจังหวะ	103
ค.11	แสดงถึงช่วงเวลาของสัญญาณรีเซ็ต	103
ค.12	แสดงถึงการต่อหน้าพินภายใน	104
ค.13	แสดงถึงการต่อแหล่งกำเนิดความถี่ภายนอก	104
ง.1	แสดงภาพสัญญาณเสียงของคำว่า "ปาก"	105
ง.2	แสดงภาพสัญญาณเสียงของคำว่า "ปิด"	106
ง.3	แสดงภาพสัญญาณเสียงของคำว่า "ป็น"	107
ง.4	แสดงภาพสัญญาณเสียงของคำว่า "เป็น"	108
ง.1	แสดงภาพสัญญาณเสียงของคำว่า "ปาก"	105