โครงการ
การเรียนการสอนเพื่อเสริมประสบการณ์

ชื่อโครงการ การทดสอบความคงที่ของเสียงพูดสำหรับกลุ่มผู้ป่วยเด็กสมองพิการที่มีการพูดแบบดิสอาร์เทรีย

Speech Consistency Test for Cerebral Palsy Children with Dysarthric

ชื่อนิสิต นายอภิสิทธิ์ อังศุพันธุ์ **เลขประจำตัว** 5833447523

ภาควิชา ฟิสิกส์

ปีการศึกษา 2562

คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

โครงการ

การเรียนการสอนเพื่อเสริมประสบการณ์

การทดสอบความคงที่ของเสียงพูดสำหรับกลุ่มผู้ป่วยเด็กสมองพิการที่มีการพูดแบบดิสอาร์เทรีย

Speech Consistency Test for Cerebral Palsy Children with Dysarthric

ผู้จัดทำโครงงาน

นายอภิสิทธิ์ อังศุพันธุ์

อาจารย์ที่ปรึกษา

ผู้ช่วยศาสตราจารย์ ดร.ณัฐกร ทับทอง

ภาควิชาฟิสิกส์ คณะวิทยาศาสตร์

จุฬาลงกรณ์มหาวิทยาลัย

ชื่อโครงการ	การทดสอบความคงที่ของเสียงพูดสำหรับกลุ่มผู้ป่วยเด็กสมองพิการที่มีการพูดแบบดิสอาร์เทรีย
ผู้วิจัย	นายอภิสิทธิ์ อังศุพันธุ์ เลขประจำตัวนิสิต 5833447523
อาจารย์ที่ปรึกษา	ผู้ช่วยศาสตราจารย์ ดร.ณัฐกร ทับทอง
ภาควิชา/คณะ	ภาควิชาฟิสิกส์ คณะวิทยาศาสตร์
ปีการศึกษา	2562

โครงงานนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรบัณฑิต ภาควิชาฟิสิกส์ คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

คณะกรรมการได้ตรวจสอบและรับรองรายงานฉบับนี้แล้ว

..... ประธานกรรมการสอบ

(อาจารย์วิสิทธิ์ ลีลาศิริวงศ์)

..... กรรมการสอบ

(อาจารย์ ดร.ยุทธนา รุ่งธรรมสกุล)

..... อาจารย์ที่ปรึกษา

(ผู้ช่วยศาสตราจารย์ ดร.ณัฐกร ทับทอง)

ชื่อโครงการ การทดสอบความคงที่ของเสียงพูดสำหรับกลุ่มผู้ป่วยเด็กสมองพิการที่มีการพูดแบบดิสอาร์เทรีย

ผู้วิจัย นายอภิสิทธิ์ อังศุพันธุ์ เลขประจำตัวนิสิต 5833447523

อาจารย์ที่ปรึกษา ผู้ช่วยศาสตราจารย์ ดร.ณัฐกร ทับทอง

ภาควิชา/คณะ ภาควิชาฟิสิกส์ คณะวิทยาศาสตร์

ปีการศึกษา 2562

คำสำคัญ ความคงที่ของเสียงพูด, เด็กสมองพิการ, ดิสอาร์เทรีย

บทคัดย่อ

โครงงานนี้มุ่งพัฒนาการทดสอบความคงที่ของเสียงสำหรับกลุ่มผู้ป่วยเด็กสมองพิการที่มีการพูดแบบดิสอาร์เทรีย เพื่อนำไปช่วยทำนายอัตรารู้จำเสียงพูดของเด็กสมองพิการ

คลังข้อมูลเสียงพูดประกอบด้วยเสียงพูด 68 คำ บันทึกจากกลุ่มเด็กปกติ 4 คน และกลุ่มเด็กสมองพิการ 8 คน จำนวน 5 รอบ ถูกนำมาใช้ในการคำนวณหาค่าความคงที่ของเสียงพูด (SCS) สำหรับผู้รับการทดสอบแต่ละคน

เมื่อนำค่า SCS ไปเปรียบเทียบกับอัตราการรู้จำเสียงพูดที่ใช้แบบจำลองฮิดเดน มาร์คอฟ (HMMs) และข่ายงานประสาทเทียม (ANNs) พบว่าผู้รับการทดสอบที่มีค่า SCS สูง จะมีอัตราการรู้จำเสียงพูดสูง ส่วนผู้รับการทดสอบที่มีค่า SCS ต่ำก็จะมีอัตราการรู้จำเสียงพูดต่ำเช่นกัน

จากนั้น ผู้วิจัยได้นำค่า SCS มาหาความสัมพันธ์เชิงเส้นกับอัตราการรู้จำเสียงพูดที่ใช้ HMM และ ANN ซึ่งได้ค่าสัมประสิทธิ์การกำหนด (R^2) เท่ากับ 0.89 และ 0.86 ตามลำดับ ซึ่งอยู่ในเกณฑ์ดีมาก

ท้ายสุด ผู้วิจัยได้นำค่า SCS ไปทำนายอัตราการรู้จำเสียงพูด พบว่าผลการทำนายที่ได้มีคลาดเคลื่อนสัมบูรณ์เฉลี่ยจากอัตราการรู้จำจริงเท่ากับ 5% ซึ่งเป็นระดับที่น่าพอใจมาก

Topic Speech Consistency Test for Cerebral Palsy Children with Dysarthric

Name Mr.Apisit Angsuphan ID. 5833447523

Advisor Assistant Professor Dr.Nuttakorn Thubthong

Department/Faculty Department of Physics, Faculty of Science

Academic Year 2019

Keyword Speech Consistency, Cerebral Palsy, Dysarthric

Abstract

This project aims to develop a Speech Consistency Test for Cerebral Palsy Children with Dysarthric to predict the speech recognition rate.

Speech corpus contains 68 words were collected from 4 normal children and 8 cerebral palsy children for 5 times. The corpus was used to measure Speech Consistency Scores (SCSs) for every subject.

SCSs were compared with the speech recognition rate using hidden Markov models (HMM) and Artificial Neural Networks (ANNs). It was found that subjects with a high SCS will have a higher speech recognition rate while subjects with a low SCS also have a low speech recognition rate.

SCSs were then used to find the linear correlation with the speech recognition rate using HMM and ANN, which gave the linear trendlines with the coefficient of determinations (R^2) of 0.89 and 0.86, respectively.

Finally, the linear trendline using HMM was used to predict the speech recognition rate. It was found that the prediction results have an average absolute error of 5%, which is a very satisfactory level.

กิตติกรรมประกาศ

ข้าพเจ้าขอขอบคุณอาจารย์ที่ปรึกษาโครงงาน ผศ.ดร.ณัฐกร ทับทอง ที่ได้ให้คำแนะนำ คำปรึกษา แนวทางการใช้ชีวิต อีกทั้งยังเป็นธุระดูแลการทำโครงงานฉบับนี้ตลอดมา

ข้าพเจ้าขอขอบคุณ อาจารย์ ดร.อรพิน วรรณดิลก และ ผศ.พรเจริญ ผโลทัยดำเกิง ที่เป็นอาจารย์ที่ปรึกษาและให้คำปรึกษาทั้งแนวทางการใช้ชีวิต และทางด้านการเรียนตลอดมา

ข้าพเจ้าขอขอบคุณ รศ.ดร.อุดมศิลป์ ปิ่นสุข อาจารย์ ดร.สมฤทธิ์ วงศ์มณีโรจน์ และอาจารย์อีกหลายๆ ท่านที่คอยถามความเป็นไปของโครงงานและการดำเนินชีวิต

ข้าพเจ้าขอขอบคุณคณะกรรมการสอบโครงงานวิทยาศาสตร์ทุกท่านอันได้แก่ อาจารย์ ดร.ยุทธนา รุ่งธรรมสกุล และอาจารย์วิสิทธิ์ ลีลาศิริวงศ์ ที่ช่วยชี้แนะแนวในการทำโครงงาน และช่วยแก้ไขข้อบกพร่องของโครงงาน

ข้าพเจ้าขอขอบคุณผู้ที่อยู่ในห้องภาคของภาควิชาฟิสิกส์ทุกคนที่คอยให้กำลังใจ และคอยสนับสนุนให้ข้าพเจ้ามีสุขภาพจิตที่ดีขึ้นตลอดมา

ข้าพเจ้าขอขอบคุณมารดาที่คอยสนับสนุนให้คำแนะนำในการดำเนินชีวิต ค่าใช้จ่ายระหว่างเล่าเรียน และส่งเสียข้าพเจ้าให้เรียนสำเร็จการศึกษาถึงระดับปริญญาบัณฑิตตลอดมา

สารบัญ

บทที่ 1

บทนำ

1.1 ที่มาและความสำคัญของโครงงาน

การพูดเป็นวิธีการที่มนุษย์ใช้ในการสื่อสารมากที่สุดในโลก ผู้ที่มีปัญหาในการพูดอาจถูกจัดว่าเป็นผู้พิการ หากมีความบกพร่องของการพูดที่รุนแรง ทำให้เป็นผู้ด้อยโอกาส มีสถานภาพและคุณภาพชีวิตที่ต่ำกว่าคนปกติ มีความลำบากในการดำรงชีวิตอยู่ในสังคม และอาจทำให้ต้องอาศัยพึ่งพาผู้อื่นเป็นภาระแก่ครอบครัวและสังคม ลักษณะการพูดที่ผิดปกติ ได้แก่ พูดไม่ชัด พูดไม่มีเสียงหรือพูดติดอ่าง ความผิดปกติของเสียงพูดหลาย ๆ อย่าง สามารถรักษา ฝึก หรือแก้ไขให้เป็นปกติได้ เช่น เด็กที่พูดไม่ชัดจากการเรียนรู้ที่ผิด เป็นที่น่าเสียดายที่ปัญหาการพูดที่พบได้บ่อยบางครั้งไม่สามารถรักษาหรือแก้ไขต้นเหตุให้หายเป็นปกติได้ เช่น ความผิดปกติของการพูดแบบดิสอาร์เทรีย (Dysarthric speech disorders) การดูแลรักษาจึงมีความจำเป็นมากกว่าการฝึกฝนการออกเสียงใหม่ เนื่องจากการให้ความช่วยเหลือแก่ผู้ที่มีปัญหาการพูดสำหรับผู้พิการ ให้สามารถพูดสื่อสารกับผู้อื่นได้เป็นสิ่งสำคัญมาก เพราะเป็นวิธีการเพิ่มคุณภาพชีวิตโดยตรงต่อบุคคลนั้น ด้วยความก้าวหน้าทางเทคโนโลยี วิทยาศาสตร์ และการแพทย์สมัยใหม่ ตลอดจนการบูรณาการระหว่างความก้าวหน้าของสาขาต่างเข้าด้วยกัน จึงเป็นขบวนการทำงานสมัยใหม่ที่สามารถทำได้ และหากทำสำเร็จจะมีผลดีต่อทั้งผู้พิการ ครอบครัวและสังคมอย่างเด่นชัด เพราะทำให้การดูแลแก้ไขปัญหาความบกพร่องของการพูดทำได้อย่างมีประสิทธิภาพและประสิทธิผลมากที่สุด [1]

ความผิดปกติของการพูดแบบดิสอาร์เทรีย หมายถึง การพูดที่ผิดปกติซึ่งเป็นผลจากการทำงานของกล้ามเนื้อที่ใช้ในกลไกของการออกเสียงพูดบกพร่อง อันสืบเนื่องมาจากระบบประสาทส่วนกลางหรือสมอง (Central nervous system) หรือระบบประสาทส่วนปลาย (Peripheral nervous system) ทำงานผิดปกติ ทำให้กำลังของกล้ามเนื้อของอวัยวะที่ใช้ในการพูดทำงานไม่ได้ อาจแสดงออกในลักษณะอาการกล้ามเนื้ออ่อนแรงหรือแข็งเกร็ง หรือการทำงานประสานกันของกล้ามเนื้อทำงานได้ไม่ดี การพูดแบบดิสอาร์อาร์เทรียเป็นปัญหาที่พบได้บ่อยในผู้ป่วยเด็กสมองพิการ (Cerebral palsy children) ผู้ป่วยโรคหลอดเลือดสมอง (Stroke หรือ Cerebrovascular Accident หรือ CVA) และผู้ป่วยที่ได้รับอุบัติเหตุทางสมอง นอกจากระบบประสาทที่ใช้สั่งการการพูกบกพร่องแล้ว ผู้ป่วยกลุ่มนี้มักมีปัญหาทางด้านร่างกายหรือการทำงานของอวัยวะส่วนอื่นร่วมด้วย เช่น เป็นอัมพาตของกล้ามเนื้อแขนขา มีการปัญหาชักเกร็ง มีปัญหากลืนลำบาก มีขีดจำกัดของการเคลื่อนไหว การเคลื่อนที่ทำได้ช้า ผู้ป่วยเหล่านี้ถูกจัดว่าเป็นผู้พิการทางด้านร่างกายและการสื่อสารเพราะพูดผิดปกติ การเพิ่มความสามารถ

ในการสื่อสารโดยการพูดควบคู่ไปกับการทำงานด้านร่างกายเพื่อเพิ่มคุณภาพชีวิตเป็นสิ่งที่สำคัญมาก การดูแลแก้ไขปัญหาการพูดโดยหลักการทั่วไปคือเพิ่มประสิทธิภาพการทำงานของกลไกการพูดเป็นสำคัญ [1]

กลไกการออกเสียงพูดของคนปกติ เกิดจากการทำงานประสานงานกันของอวัยวะระบบต่าง ๆ ซึ่งเริ่มต้นตั้งแต่การทำงานของระบบการหายใจ ทำให้ได้ลมหายใจออกเป็นแหล่งพลังของเสียง ระบบการเปล่งเสียงพูด (Phonation) โดยการทำงานของเส้นเสียงและกล้ามเนื้อกล่องเสียงจะปรับกระแสลมหายใจที่ออกจากปอด เกิดการกักลมหรือการสั่นของเส้นเสียง ทำให้เกิดเป็นคลื่นเสียงที่มีคุณสมบัติด้านความถี่ของเสียง (ระดับเสียงสูง-ต่ำ) แตกต่างกันไป และโดยอาศัยการทำงานของกล้ามเนื้อของระบบแปรเสียงพูด (Articulation and resonation) เช่น กล้ามเนื้อกล่องเสียง ลิ้น ริมฝีปาก เพดานอ่อน ลิ้นไก่ จะทำให้เสียงที่มีความถี่แตกต่างกันนั้นเกิดเป็นเสียงต่างๆ ของระบบเสียงของแต่ละภาษา การได้ยินเสียงพูดของตนเองเป็นระบบการทำงานที่ช่วยทำให้ผู้พูดตะหนักและรับรู้ว่าการพูดของตนเองถูกต้องชัดเจนดีหรือไม่ การทำงานของทุกระบบต้องประสานงานกันอย่างเหมาะสมเพื่อให้ได้ผลลัพธ์ คือ คำพูดที่ชัดเจนถูกต้อง [1]

ผู้ที่มีปัญหาการพูดแบบดิสอาร์เทรียนั้น อาจมีความบกพร่องของกลไกการพูดทุกระบบหรือบางระบบเท่านั้น เช่น ระบบการหายใจที่ไม่ดี ทำให้พูดเสียงเบา พูดไม่มีเสียง เสียงพูดขาดหายในตอนท้าย พูดไม่ต่อเนื่องหรือพูดประโยคยาว ๆ ไม่ได้ บางรายอาจมีปัญหาพูดไม่ชัด ออกเสียงบางเสียงไม่ได้ หรือพูดช้า ทำให้ยากแก่การเข้าใจ ทำให้เกิดความล้มเหลวในการสื่อสารกับผู้อื่น การให้ความช่วยเหลือเกี่ยวกับความบกพร่องบางอย่าง เช่น การเพิ่มความดังของเสียงพูด สามารถแก้ไขได้ด้วยเทคโนโลยีวิทยาการหรือระบบคอมพิวเตอร์สมัยใหม่ แต่เนื่องจากผู้ป่วยมักมีอาการผิดปกติทางด้านร่างกายและการควบคุมการทำงานของกล้ามเนื้อในส่วนต่าง ๆ ร่วมด้วย โดยเฉพาะอย่างยิ่งการทำงานของกลุ่มกล้ามเนื้อที่ต้องการความแม่นยำในการทำงานสูง [6] ดังนั้นอุปกรณ์มาตรฐานต่าง ๆ เช่น อุปกรณ์ที่ช่วยในการควบคุมคอมพิวเตอร์ ได้แก่ เมาส์ หรือ แป้นพิมพ์ จะไม่สามารถใช้งานได้อย่างมีประสิทธิภาพ [1]

จากความผิดปกติด้านการพูดและร่างกายของผู้พิการ ระบบการรู้จำเสียงพูด (Speech recognition system) จะเป็นระบบหนึ่งที่จะช่วยให้ผู้พิการกลุ่มนี้มีการดำเนินชีวิตที่สะดวกและช่วยเหลือตัวเองได้มากขึ้น เพราะระบบนี้สามารถพัฒนาต่อไปเพื่อควบคุมอุปกรณ์อิเล็กทรอนิกส์ต่าง ๆ สำหรับอำนวยความสะดวก เช่น การใช้เสียงในการบังคับระบบรถเข็นอัตโนมัติ หรือ การใช้เสียงควบคุมการปิด-เปิดไฟ เป็นต้น [1]

ในการนำระบบการรู้จำเสียงพูดไปประยุกต์ใช้นั้น จำเป็นต้องมีการประเมินความผิดปกติของเสียงพูดในแบบต่าง ๆ ซึ่งในในปัจจุบันนิยมทำประเมินโดยการทดสอบความชัดเจนของเสียงพูด (Articulation Test) การ

วิเคราะห์ทางสวนศาสตร์ (Acoustic Analysis) และการทดสอบภาวะฟังความออก (Intelligibility Test) [2,3,4,5] ซึ่งการทดสอบทั้ง 3 แบบ ต้องใช้เวลาและกำลังคนในการดำเนินการทดสอบมาก โดยเฉพาะอย่างยิ่งการทดสอบความชัดเจนของเสียงและการวิเคราะห์ทางสวนศาสตร์จำเป็นต้องใช้ผู้เชี่ยวชาญเฉพาะด้านหรือผู้มีประสบการณ์การทำงานมานานพอสมควร จึงจะสามารถตรวจวิเคราะห์และให้การประเมินผลที่ถูกต้องได้ [1]

การวิเคราะห์ทางสวนศาสตร์สามารถให้ข้อมูลที่เป็นประโยชน์ในการวิเคราะห์เชิงคุณภาพของเสียงพูด หาข้อบกพร่อง และให้แนวทางการแก้ไขสัญญาณเสียงได้ แต่การวัดค่าตัวแปรทางสวนศาสตร์ต่าง ๆ ต้องใช้เวลาและความชำนาญ ผู้ทำการวัดแต่ละคนอาจมีรูปแบบการตัดสินใจที่ต่างกัน มาตรฐานการวัดจึงเป็นหลักสำคัญในการทดสอบแบบนี้ ขณะที่การทดสอบความชัดเจนของเสียงพูดและการทดสอบภาวะฟังความออกเป็นการประเมินผลที่ขึ้นอยู่กับบุคคลที่ทำการประเมิน หากผู้ประเมินไม่มีความชำนาญ มีภูมิหลังความรู้และประสบการณ์ต่างกัน ผลที่ได้อาจต่างกันไปบ้าง เช่น ผู้ฟังที่มีความใกล้ชิดหรือเคยมีประสบการณ์การทำงานกับผู้พิการทางเสียง อาจฟังและทำความเข้าใจกับภาษาของผู้พิการได้มากกว่าผู้ฟังที่ไม่คุ้นเคยหรือในกรณีที่ผู้ฟังได้รับฟังข้อมูลเสียงของเด็กสมองพิการอย่างต่อเนื่องนาน ๆ ก็จะคุ้นเคย และเกิดการเรียนรู้รูปแบบการพูดของเด็กสมองพิการ แต่หากทำการทดสอบโดยเปลี่ยนผู้ฟังไปเรื่อย ๆ ก็อาจเกิดความแตกต่างในด้านมาตรฐานการประเมินผล [1]

การวิจัยเพื่อพัฒนาระบบรู้จำสำหรับเด็กสมองพิการที่บกพร่องทางเสียงนั้น เราจำเป็นต้องมีเครื่องมือในการประเมินผลเบื้องต้น สำหรับใช้เป็นแนวทางในการเลือกลักษณะของระบบที่เหมาะสมกับความบกพร่องแต่ละแบบ เช่น หน่วยการรู้จำที่เหมาะสมหรือสวนะลักษณ์ (acoustic features) แบบต่าง ๆ เพื่อเพิ่มความถูกต้องในการรู้จำเสียงพูด นอกจากนี้เด็กสมองพิการบางคนอาจใช้ประโยชน์จากระบบรู้จำอย่างมีประสิทธิภาพ ในขณะที่บางคนมีข้อจำกัดที่ทำให้ไม่สามารถใช้ระบบได้ ดังนั้นการประเมินผลเบื้องต้นก่อนพัฒนาระบบจึงมีความจำเป็นมาก อย่างไรก็ตามการประเมินมาตรฐานที่กล่าวมา ยังไม่สามารถนำมาใช้ทำนายประสิทธิภาพของระบบการรู้จำเมื่อใช้กับผู้พิการแต่ละคนได้ [1]

จากความจำเป็นดังกล่าว โครงงานนี้จึงมุ่งศึกษาความเป็นไปได้ในการใช้ตัวประเมินผลแบบใหม่ คือ ค่าความคงที่ของเสียง (Speech Consistency Score, SCS) ซึ่งอาศัยหลักการของการวัดค่าความคล้ายและความต่างของสัญญาณเสียง เพื่อนำมาคำนวณค่าความคงที่ของเสียง ข้อดีของการทดสอบแบบนี้คือ

(1) พัฒนาง่าย ไม่ซับซ้อน

(2) ใช้ข้อมูลเสียงที่มีสำหรับทำการประเมินได้โดยตรง และใช้เวลาในกระบวนการประเมินน้อย

(3) สวนะลักษณ์ของเสียงที่ใช้ในการเปรียบเทียบสามารถปรับเปลี่ยนได้ตามความต้องการ ซึ่งสามารถเลือกให้เป็นสวนะลักษณ์ชุดเดียวกับที่ใช้ในระบบรู้จำเสียงพูด เพื่อนำผลการประเมินที่ได้ไปทำนายอัตราการรู้จำของผู้รับการทดสอบแต่ละคนได้

(4) สามารถพัฒนาต่อเนื่องเพื่อใช้เป็นระบบวิเคราะห์ความบกพร่องของเสียง และใช้เป็นตัวเลือกสวนะลักษณ์ที่เหมาะสมกับความบกพร่องของแต่ละคนได้

1.2 วัตถุประสงค์ของโครงงาน

1. เพื่อศึกษาและนำเสนอการประเมินความบกพร่องของเสียงพูด ในรูปแบบของการวัดค่าความคงที่ (ทั้งในด้านความเหมือนและความต่าง) ของเสียงในการออกเสียงแต่ละครั้ง

2. เพื่อประยุกต์ค่าความคงที่ของเสียงพูดในการทำนายอัตรารู้จำเสียงพูดกับเด็กสมองพิการได้

1.3 ขอบเขตของโครงงาน

ทดสอบกับผู้รับการทดสอบ 12 คน คือ กลุ่มเด็กปกติ 4 คน และกลุ่มเด็กสมองพิการ 8 คน

1.4 ประโยชน์ที่คาดว่าจะได้รับ

1. ได้การประเมินความบกพร่องของเสียงพูด ในรูปแบบของการวัดค่าความคงที่
2. ได้ค่าความคงที่ของเสียงพูดที่สามารถทำนายอัตรารู้จำเสียงพูดกับเด็กสมองพิการได้

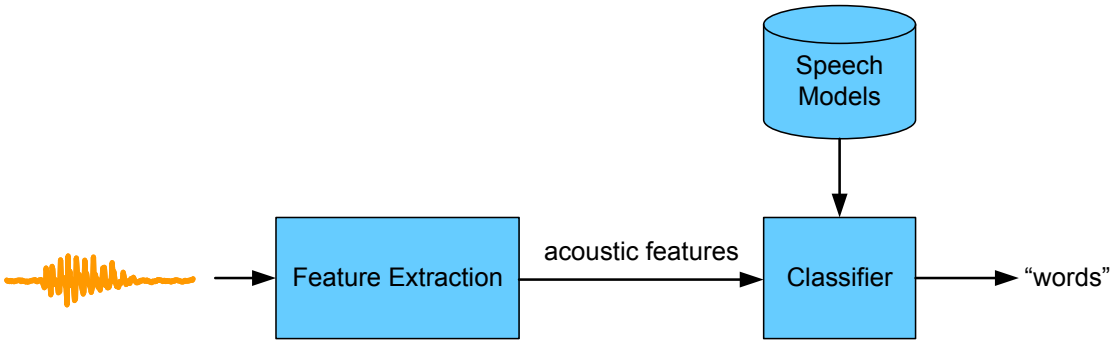
บทที่ 2

ทฤษฎีที่เกี่ยวข้อง

ในบทนี้จะกล่าวถึงความรู้พื้นฐานสำหรับโครงงานนี้ คือ การรู้จำเสียงพูด การสกัดสวนะลักษณ์ และตัวจำแนก โดยสังเขป

2.1 การรู้จำเสียงพูด (Speech Recognition)

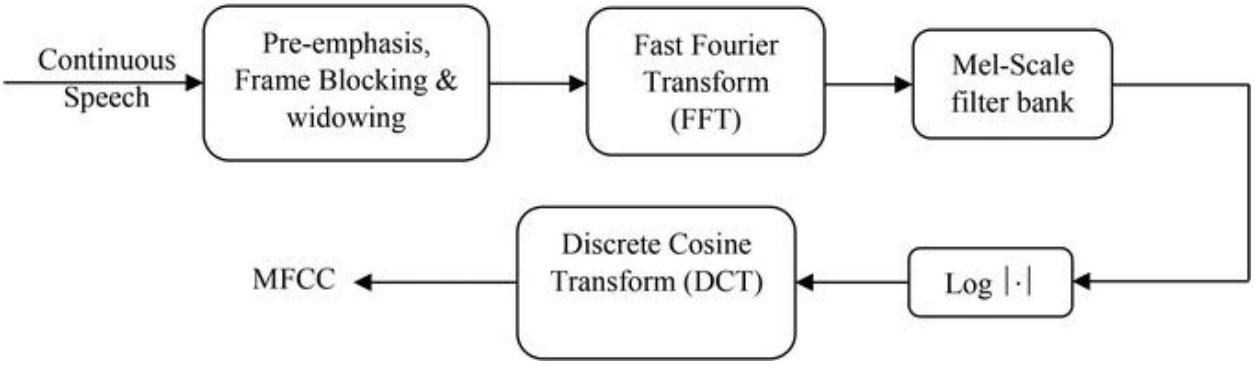
ระบบการรู้จำเสียงพูดคือระบบการแปลงเสียงพูดให้เป็นคำพูด ซึ่งประกอบด้วย 2 ส่วนหลัก คือ การสกัดสวนะลักษณ์ (Acoustic Feature extraction) และตัวจำแนก (Classifier) โดยส่วนแรกคือการสกัดลักษณะเด่นจากสัญญาณเสียงพูดซึ่งเรียกลักษณะเด่นนั้นว่าสวนะลักษณ์ (Acoustic Features) และส่วนที่สองคือการนำสวนะลักษณ์ ไปเทียบกับแบบจำลองเสียงพูด (Speech models) ที่สร้างจากตัวอย่างเสียงพูดจากกลุ่มคนจำนวนหนึ่ง และตอบว่าเสียงที่พูด คือคำพูดอะไร (ดูรูปที่ 2.)



รูปที่ 2.1 กระบวนการรู้จำเสียงพูด

2.2 การสกัดสวนะลักษณ์ (Acoustic Feature Extraction)

สวนะลักษณ์ คือ ลักษณะเด่นทางสวนะศาสตร์ของสัญญาณเสียงพูดซึ่งถูกใช้แทนสัญญาณเสียงพูดนั้นในขั้นตอนการรู้จำเสียงพูด สวนะลักษณ์ที่ใช้ในโครงงานนี้ คือ สัมประสิทธิ์เซบตรัมของความถี่เมล (Mel Frequency Cepstral Coefficient, MFCC) ซึ่งเป็นสวนะลักษณ์พื้นฐานที่เป็นที่นิยมที่สุดในปัจจุบัน ขั้นตอนการคำนวณหาค่า MFCC ทำได้ดังนี้ (ดูรูปที่ 2.2) [7]



รูปที่ 2.2 แผนผังการหาค่าสัมประสิทธิ์เซบตรัมของความถี่เมล (MFCC)

(ที่มา : https://www.intechopen.com/books/from-natural-to-artificial-intelligence-algorithms-and-applications/some-commonly-used-speech-feature-extraction-algorithms, 7 พฤษภาคม 2563)

1. การวิเคราะห์สเปกตรัม

การวิเคราะห์สเปกตรัม คือ การคำนวณหาสเปกตรัมกำลัง (Power Spectrum) ซึ่งคือสเปกตรัมที่แสดงความเข้ม (Intensity) ของสัญญาณที่ความถี่ต่าง ๆ มีขั้นตอนดังนี้

- การแบ่งเสียงออกเป็นกรอบย่อย โดยขนาดกรอบ (frame size) ขึ้นกับช่วงความถี่ (frequency bandwidth) ที่ต้องการพิจารณา สำหรับสัญญาณเสียงพูดนั้น นิยมใช้ขนาดของกรอบเท่ากับ 0.025

วินาที (เพื่อให้ความถี่ตอบสนองต่ำสุด เท่ากับ 40Hz) ซึ่งแต่ละกรอบมีการซ้อนเหลื่อมกัน (overlapping) เพื่อความต่อเนื่องของสัญญาณเสียง

- การปรับค่าน้ำหนักของสัญญาณเสียงแต่ละกรอบด้วยหน้าต่างแฮมมิง (Hamming window) ซึ่งเป็นการเพิ่มความชัดเจนของสเปกตรัมกำลังที่ต้องการ โดยใช้สมการ (2.1)

$$W(n) = 0.54 + 0.46\cos\left(\frac{2\pi n}{N-1}\right) \tag{2.1}$$

โดย N คือ จำนวนข้อมูลต่อ 1 กรอบ

- การแปลงแบบฟูเรียร์แบบไม่ต่อเนื่อง (Discrete Fourier transform) ซึ่งเป็นกระบวนการแปลงสัญญาณเสียงจากโดเมนเวลา (time domain) เป็นโดเมนความถี่ (frequency domain) ดังสมการ (2.2)

$$S(\omega) = \sum_{n=1}^{N} x(n)e^{-j\omega n} \tag{2.2}$$

เมื่อ $x(n)$ คือ สัญญาณจุดที่ n

- การคำนวณสเปกตรัมกำลัง (power spectrum) สัญญาณในโดเมนความถี่ที่ได้จะอยู่ในรูปตัวเลขเชิงซ้อน เมื่อคูณตัวมันด้วยสังยุคของมัน (conjugate) ดังสมการที่ (9.3) จะได้สเปกตรัมกำลัง

$$P(\omega) = \mathrm{Re}[S(\omega)]^2 + \mathrm{Im}[S(\omega)]^2 \tag{2.3}$$

โดยที่ $S(\omega)$ คือ สัญญาณในโดเมนความถี่ที่ผ่านการแปลงฟูเรียร์

ω คือ ความถี่เชิงมุม (rad/s)

2. การทำ Filter Bank

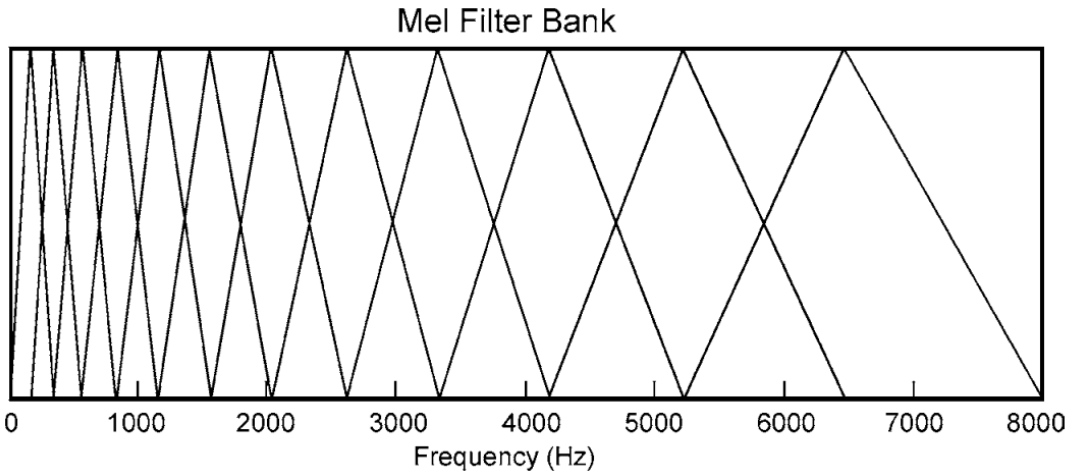
การวิเคราะห์สเปกตรัมมีความสัมพันธ์กับรูปร่างของช่องปาก สเปกตรัมจะถูกจัดเป็น N กลุ่ม โดยพิจารณาจากช่องความถี่ (filterbank channel) เนื่องจากต้องการลดความละเอียดของสเปกตรัมโดยยังคงรูปร่างของสเปกตรัมของสัญญาณเสียงพูดไว้ โดยจำนวนช่องความถี่ที่นิยมใช้ คือ 24 ช่อง ซึ่งสัมพันธ์กับกระบวนการประมวลผลของหูมนุษย์ เรียกวิธีการนี้ว่าการทำ filterbank

โดยก่อนการจัดกลุ่มจะทำการแปลงสเกลของจากความถี่สเกลเฮริตซ์ (Hertz) เป็นสเกลเมล (Mel) โดยใช้สมการ 2.4 ซึ่งสเกลเมลเป็นสเกลความถี่ที่มีลักษณะคล้ายคลึงกับการได้ยินเสียงด้วยหูของมนุษย์มากกว่า [8]

$$Mel(f) = 1125\log_{10}(1+ f / 700) \quad (2.4)$$

เมื่อ f คือ ความถี่ในหน่วยเฮริตซ์

ในการคำนวณค่าในแต่ละช่อง (channel) ของ filterbank ทั้ง N ค่าในสเกลเมล มีการถ่วงน้ำหนักด้วยตัวกรองรูปสามเหลี่ยม (Triangular Filter) ดังรูปที่ 2.2



รูปที่ 2.2 ตัวกรองรูปสามเหลี่ยมของความถี่เมล

(ที่มา: https://i.stack.imgur.com/q6a7X.png, 7 พฤษภาคม 2563)

3. การคำนวณเซบตรัมของความถี่เมล (Mel frequency cepstrum computation)

เซบตรัม คือ การแปลงฟูเรียร์ผกผันแบบไม่ต่อเนื่อง (Inverse Discrete Fourier Transform; DFT) ของฟังก์ชันล็อก (log) ของสเปกตรัมกำลัง [9] การใส่ล็อกจะช่วยลดความแปรปรวนเชิงพลศาสตร์ (dynamics) ของสัญญาณเสียง ในกรณีที่ล็อกของสเปกตรัมกำลังคือค่าจริง (real) และสมมาตร (symmetric) IDFT สามารถถูกลดรูปเป็นการแปลงโคไซน์ (Discrete Cosine Transform, DCT) ได้ [9,10]

การคำนวณหาสัมประสิทธิ์ของเซบตรัมของความถี่เมล ทำได้โดยการนำสเปกตรัมกำลังที่ผ่านการกรอง $\tilde{S}_k$ มาใส่ล็อกและส่งเข้าสู่การแปลงโคไซน์ ดังสมการที่ (2.5)

$$\tilde{c}_n = \frac{1}{N}\sum_{k=1}^{N}(\log \tilde{S}_k)\cos\left[n\left(k-\frac{1}{2}\right)\frac{\pi}{K}\right], \quad n=1,2,\cdots,L \tag{2.5}$$

เมื่อ L คือ อันดับของเซบตรัมที่ต้องการ

N คือ ขนาดของกรอบ

2.3 ตัวจำแนก (Classifier)

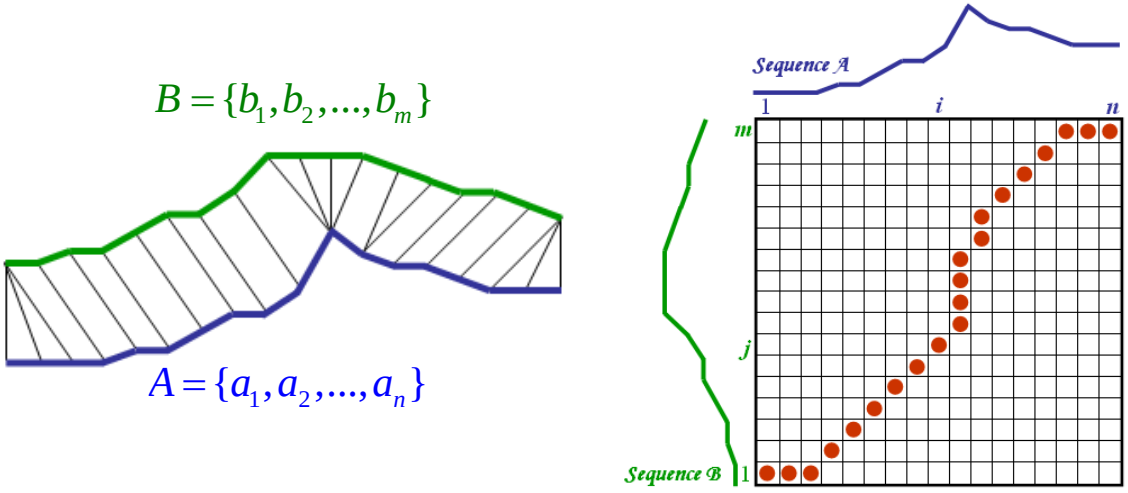
ตัวจำแนกมีหน้าที่ตอบว่าสวนะลักษณ์ที่ส่งเข้ามาคือคำพูดอะไร โดยจะทำการเทียบสวนะลักษณ์กับแบบจำลองเสียงพูดสร้างจากตัวอย่างเสียงพูดจำนวนหนึ่ง เช่น แบบจำลองฮิดเดน มาร์คอฟ (Hidden Markov Models; HMMs) และข่ายงานประสาทเทียม (Artificial Neural Networks; ANNs) ซึ่งเป็นเทคนิคการจำแนกที่นิยมใช้สำหรับงานด้านการรู้จำเสียงพูด อย่างไรก็ตาม โครงงานนี้ไม่ได้เน้นการสร้างระบบการรู้จำเสียงพูด เพียงแต่นำผลการรู้จำเสียงพูดโดยใช้แบบจำลองฮิดเดน มาร์คอฟ และข่ายงานประสาทเทียม จากงานวิจัยก่อนหน้าเพื่อมาใช้เปรียบเทียบเท่านั้น จึงไม่ขอกล่าวถึงรายละเอียดในที่นี้

โครงงานนี้จะเน้นการสร้างการทดสอบความคงที่ของเสียงพูดเท่านั้น จึงไม่จำเป็นต้องใช้ตัวจำแนกที่มีความซับซ้อน โดยต้องการเพียงเทคนิคสำหรับหาความห่างของสวนลักษณ์สองสาย (สกัดจากคำ 2 คำ) ที่ยาวไม่เท่ากัน ซึ่งได้เลือกใช้ไดนามิกไทม์วอร์ปปิง (Dynamic Time Wrapping) ซึ่งจะอธิบายโดยสังเขป ดังนี้

ไดนามิกไทม์วอร์ปปิง เป็นขั้นตอนวิธีสำหรับการเปรียบเทียบความคล้ายของลำดับที่มีความแตกต่างกันในด้านเวลาหรือความเร็ว เช่น รูปแบบการเดินของคน ๆ หนึ่งจะถูกนับว่ามีความคล้าย ไม่ว่าคน ๆ นั้นจะเดินอย่างรวดเร็ว เดินอย่างเชื่องช้า หรือแม้แต่เดินด้วยความเร่ง เมื่อพิจารณาจากผู้สังเกตเดียวกัน ซึ่งไดนามิกไทม์วอร์ปปิงสามารถนำไปประยุกต์ได้กับวิดีโอ เสียง และภาพ รวมไปถึงข้อมูลต่าง ๆ ที่สามารถแปลงให้อยู่ในรูปของข้อมูลเชิงเส้นได้ ตัวอย่างหนึ่งของการประยุกต์ขั้นตอนวิธีนี้ไปใช้คือ การรู้จำคำพูด โดยใช้ไดนามิกไทม์วอร์ปปิง เพื่อจัดการกับคำพูดที่มีความเร็วไม่เท่ากัน แม้จะสื่อความหมายเดียวกัน แสดงในรูปที่ 2.2

กำหนดให้ลำดับที่ขึ้นกับเวลา $A=\{a_1,a_2,...,a_n\}$ และ $B=\{b_1,b_2,...,b_m\}$ โดยลำดับทั้งสองมีขนาด n และ m ตามลำดับ ลำดับเหล่านี้อาจจะเป็นสัญญาณที่ไม่ต่อเนื่อง (อนุกรมเวลา) หรือ ลำดับของลักษณะเฉพาะ (feature) ที่ถูกสร้างขึ้นตามช่วงเวลา แสดงดังรูปที่ 2.2 (ด้านซ้าย)

จากนั้นจะหาเส้นทางไดนามิกไทม์วอร์ปปิงระหว่าง A และ B ซึ่งแสดงเป็นจุดสีแดงในรูปที่ 2.3 (ด้านขวา) และหาระยะทางของเส้นทางนั้น ซึ่งจะแทนความคล้ายของ A และ B ถ้าระยะทางมีค่าน้อย แสดงว่า A และ B มีความห่างกันน้อย หรือมีความคล้ายกันมาก นั่นเอง [11]



รูปที่ 2.3 แสดงวิธีการหาระยะห่างแบบ Dynamic Time Wrapping

(ที่มา: https://www.psb.ugent.be/cbd/papers/gentxwarper/DTWalgorithm.htm, 7 พฤษภาคม 2563)

บทที่ 3

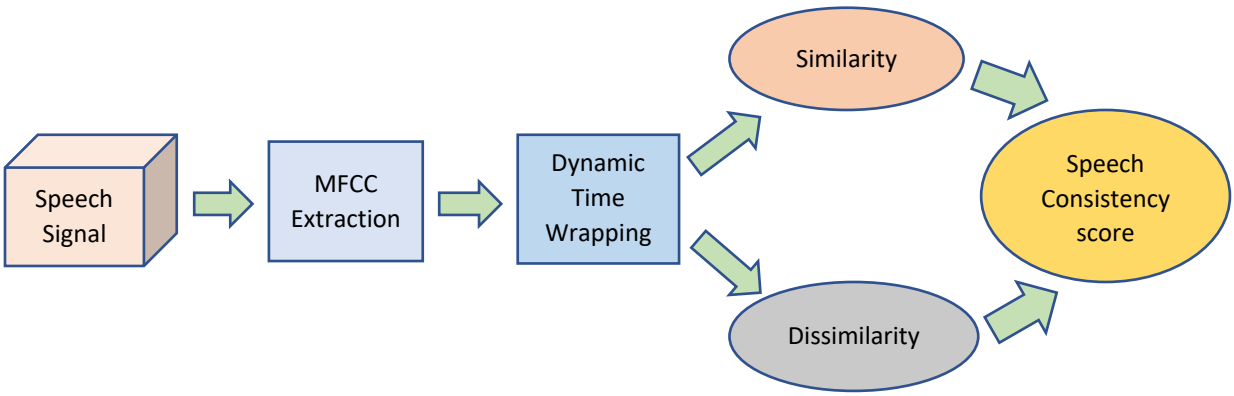
การทดสอบความคงที่ของเสียงพูด

การทดสอบความคงที่ของเสียงพูด คือ การทดสอบว่าผู้เข้ารับการทดสอบมีความคงที่ของเสียงพูดหรือไม่ ซึ่งความคล้ายของเสียงพูด (Similarity : SIM) คือ การที่ผู้เข้ารับการทดสอบพูดคำ ๆ เดิมซ้ำกันหลายครั้งแล้วเสียงออกมาเหมือนเดิม และค่าความต่างของเสียงพูด (Dissimilarity : DIS) คือ การที่ผู้รับการทดสอบพูดคำที่ต่างกันแล้วเสียงออกมาต่างกัน โดยขั้นตอนการทดสอบตวามคงที่ของเสียงพูดมีดังนี้ (ดูรูปที่ 3.1)

1. นำสัญญาณเสียงพูดมาจากคลังข้อมูล มาสกัดสวนลักษณ์แบบ MFCC เพื่อใช้เป็นตัวแทนสัญญาณเสียงเหล่านั้น

2. นำ MFCC ที่ได้ มาคำนวณหาค่าความคล้าย (SIM) และความต่าง (DIS) ตามสมการที่จะอธิบายในหัวข้อ 3.4 โดยใช้เทคนิคไดนามิกไทม์วอร์ปปิง

3. หาค่าความคงที่ของเสียงพูด (Speech Consistency Score, SCS) จากค่าความคล้าย และความต่าง



รูปที่ 3.1 แสดงขั้นตอนการวัดค่าความคงที่ของเสียงพูด

3.1 คลังข้อมูล (Speech Corpora)

คลังข้อมูลนี้ใช้โครงงานนี้ นำมาจากส่วนหนึ่งในงานวิจัยของณัฐกร ทับทอง และคณะ [1] มีรายละเอียดดังนี้

- ชุดคำสำหรับทดสอบความผิดปกติจำนวน 68 คำ ครอบคลุมทุกหน่วยเสียงหลักในภาษาไทย สำหรับทดสอบหน่วยเสียงสระ พยัญชนะต้น พยัญชนะท้าย พยัญชนะควบ และวรรณยุกต์จำนวน 22, 21, 8, 12 และ 5 คำ ตามลำดับ ดังแสดงในตารางที่ 3.1
- บันทึกเสียงจากผู้รับการทดสอบ 12 คน คือ กลุ่มเด็กปกติ 4 คน เพศละ 2 คน (NF01, NF02, NM01, NM02) และเด็กสมองพิการ 8 คน เพศละ 4 คน (DF01, DF02, DF03, DF04, DM01, DM02, DM03, DM04)
- บันทึกเสียงด้วยไมโครโฟนคาดศีรษะ AKG C440L (Condenser type) โดยใช้อัตราการชักข้อมูล (sampling rate) 16 kHz ความละเอียด (Resolution) 16 บิต แบบโมโน
- สถานที่บันทึกเสียง คือ ห้องฝึกพูด โรงเรียนศรีสังวาลย์ ปากเกร็ด นนทบุรี

ตารางที่ 3.1 รายการคำสำหรับทดสอบความผิดปกติของเสียงพูด [1]

(ก) เสียงสระ

คำ	เสียง	คำ	เสียง	คำ	เสียง	คำ	เสียง	คำ	เสียง	คำ	เสียง
พัด	a	ตา	aa	ลิ้น	i	หวี	ii	ผึ้ง	v	มือ	vv
กุ้ง	u	หมู	uu	เป็ด	e	เมฆ	ee	แกะ	x	แมว	xx
เงิน	q	เนย	qq	เงาะ	@	ล้อ	@@	โต๊ะ	o	โซ่	oo
เกี๊ยะ	ia	เตียง	iia			เสือ	vva			ขวด	uua

(ข) เสียงพยัญชนะต้น

คำ	เสียง	คำ	เสียง	คำ	เสียง	คำ	เสียง	คำ	เสียง	คำ	เสียง
ปู	p	ตู้	t	จาน	c	ไก่	k	อูฐ	#		
ผัก	ph	เทป	th	ช้าง	ch	ข้าว	kh	ลิง	l	เรือ	r
ม้า	m	หนู	n	งู	ng			แหวน	w	ยักษ์	j
ฟัน	f	เสื้อ	s	หู	h			บ้าน	b	ดาว	d

(ค) เสียงพยัญชนะท้าย

คำ	เสียง	คำ	เสียง	คำ	เสียง	คำ	เสียง	คำ	เสียง	คำ	เสียง
กบ	p	รถ	t	นก	k						
ผม	m	ค้อน	n	ยุง	ng	ว่าว	w	ด้าย	j		

(ง) เสียงพยัญชนะควบกล้ำ

คำ	เสียง	คำ	เสียง	คำ	เสียง	คำ	เสียง	คำ	เสียง	คำ	เสียง
แปรง	pr	ปลา	pl	พริก	phr	พลุ	phl	แตร	tr	ทรัมเป็ต	thr
กรง	kr	กล้วย	kl	กวาง	kw	ครู	khr	ขลุ่ย	khl	ควาย	khw

(จ) เสียงวรรณยุกต์

คำ	เสียง	คำ	เสียง	คำ	เสียง	คำ	เสียง	คำ	เสียง
งา	0	ถั่ว	1	หม้อ	2	ช้อน	3	หมี	4

3.2 การรู้จำเสียงพูดโดยใช้แบบจำลองฮิดเดน มาร์คอฟ (Hidden Markov Models, HMMs)

ในโครงงานนี้ใช้ผลการรู้จำเสียงพูดจากงานวิจัยของณัฐกร ทับทอง และคณะ [1,7] โดยมีรายละเอียดอย่างย่อ ดังนี้

- เป็นการรู้จำระดับคำแบบผู้พูดคนเดียว (Word based single speaker speech recognition)
- ใช้สวนะลักษณ์ MFCC อันดับ 12 อนุพันธ์อันดับที่หนึ่ง และที่สองของ MFCC
- ใช้จำนวนสถานะ (Numbers of states) 15 สถานะ และจำนวนมิกเจอร์ (Numbers of mixtures) 4 มิกเจอร์
- แบบจำลองฮิดเดน มาร์คอฟ ที่ใช้ในงานวิจัยนี้ คือ Hidden Markov Models Toolkit (HTK) ที่พัฒนาโดย Cambridge University Engineering Department (CUED) ประเทศอังกฤษ

3.3 การรู้จำเสียงพูดโดยใช้ข่ายงานประสาทเทียม (Artificial Neural Network, ANN)

ในโครงงานนี้ใช้ผลการรู้จำเสียงพูดจากงานวิจัยของณัฐกร ทับทอง และคณะ [1,7] โดยมีรายละเอียดอย่างย่อ ดังนี้

- เป็นการรู้จำระดับคำแบบผู้พูดคนเดียว (Word based single speaker speech recognition)
- ใช้สวนะลักษณ์ MFCC อันดับ 12 โดยในแต่ละคำจะถูกแทนด้วยสัมประสิทธิ์ของสวนะลักษณ์จำนวน 252 ค่า ซึ่งสกัดจากสัญญาณเสียงที่ตำแหน่งเวลาต่าง ๆ 21 ตำแหน่ง คือ 5, 10, 14, 19, 23, 28, 32, 37, 41, 46, 50, 55, 59, 64, 68, 73, 77, 82, 86, 91 และ 95 เปอร์เซ็นต์ ตามแกนเวลาของสัญญาณเสียง โดยใช้กรอบสัญญาณขนาด 25 มิลลิวินาที
- ใช้ข่ายงานประสาทเทียมแบบป้อนไปข้างหน้าสามชั้น (Three-layer feedforward network) ซึ่งประกอบด้วยชั้นข้อมูลเข้า (input layer) ชั้นซ้อนตัว (hidden layer) และชั้นข้อมูลออก (output layer) ในการฝึกจะใช้วิธีความผิดพลาดแบบแพร่กระจายย้อนกลับ (Error back-propagation) และกำหนดค่าโมเมนตัมเท่ากับ 0.9 และค่าอัตราการเรียนรู้ที่ใช้เท่ากับ 0.0001
- ข่ายงานประสาทเทียมที่ใช้ในงานวิจัยนี้ คือ NICO’s Toolkit ที่พัฒนาโดย Nikko Strom, Department of Speech,Music and Hearing,KTH, Sweden

3.4 การหาค่าคงที่ของเสียงพูด (Speech Consistency Score, SCS)

ในการออกเสียงคำเดียวกันหลาย ๆ ครั้ง จากผู้พูดคนเดียวกัน สัญญาณเสียงที่ออกมาในแต่ละครั้งจะมีความต่างกันทั้งในด้านความยาวของเสียงพูด จังหวะ และอัตราการพูด มีผลให้การหาระยะห่างระหว่างสัญญาณเสียง 2 สัญญาณโดยตรงไม่สามารถทำได้ ดังนั้น ไดนามิกไทม์วอร์ปปิง (Dynamic Time Wrapping, DTW) (ดูรายละเอียดในหัวข้อ 2.3) จึงถูกนำมาประยุกต์ใช้ ในการใช้งานจริงนั้น สัญญานเสียงถูกนำมาสกัดเป็นสวนลักษณ์ก่อน ซึ่งในโครงงานนี้ใช้ MFCC (ดูรายละเอียดในหัวข้อ 2.2) และจึงนำสวนลักษณ์ของสัญญาณเสียงทั้งสองมาหาระยะห่างของสัญญาณเสียงทั้งสอง โดยใช้ DTW ถ้าค่าระยะห่างมีค่าน้อยแสดงว่าสัญญาณที่ทดสอบมีความเหมือนกันมาก จากค่าระยะห่างดังกล่าวสามารถนำมาหาค่าความคงที่ของเสียง (SCS) สำหรับผู้รับการทดสอบแต่ละคนได้

ขั้นตอนการคำนวณ แบ่งออกเป็น 3 ส่วนคือ

1. การคำนวณค่าความคล้ายของเสียงพูด (Speech Similarity, *SIM*)

ค่าความคล้ายของเสียงพูด คือ ค่าที่แสดงความคล้ายของสัญญาณเสียงจากข้อมูลเสียงผู้พูดคนเดียวกันออกเสียงคำ ๆ เดิม M ครั้ง ซึ่งจะคำนวณหาค่าเฉลี่ยความต่างของสัญญาณเสียงทั้ง M ครั้ง เพื่อใช้เป็นตัวแทนของระยะห่างเฉลี่ยภายในกลุ่มคำเดียวกัน สมการที่ 3.1 แสดงการหาค่าระยะห่างเฉลี่ยของคำทดสอบที่ w จากการพูดทั้งหมด M ครั้งต่อคำ

$$\overline{X^w} = \frac{1}{{}_M C_2} \sum_{i=1}^{M-1} \sum_{j=i+1}^{M} |X_i^w - X_j^w| \tag{3.1}$$

เมื่อ X_i^w คือ ค่าสวนลักษณ์ (Acoustic Features) ของสัญญาณเสียงคำที่ w ครั้งที่ i (i คือ ลำดับการพูดครั้งที่ 1 ถึง M)

เมื่อนำคำที่ใช้ทดสอบความบกพร่องของเสียงทั้งหมด n คำ มาค่าระยะห่างเฉลี่ย และนำค่าระยะห่างเฉลี่ยเหล่านั้นมาหาค่าเฉลี่ยระยะห่างรวมของคำทั้งหมด ดังสมการที่ 3.2 ค่าที่ได้จะแสดงความคล้ายของเสียงพูด (*SIM*)

$$SIM = \frac{1}{n} \sum_{w=1}^{n} \overline{X^w} \tag{3.2}$$

2. การคำนวณค่าความต่างของเสียงพูด(Speech Dissimilarity, *DIS*)

ค่าความต่างของเสียงพูด คือ ค่าที่แสดงระยะห่างระหว่างคำที่ต่างกันจากเสียงของผู้พูดคนเดียวกัน ซึ่งคือคำระยะห่างเฉลี่ยระหว่างคำนั้นเอง

ในการคำนวณเริ่มต้นจากการหาครั้งที่เสียงพูดที่มีระยะห่างจากครั้งอื่นน้อยที่สุด (พูดทั้งหมด M ครั้ง) เพื่อใช้เป็นตัวแทนของคำที่ w ดังสมการที่ 3.3

$$k_w = \arg\min_{i}^{M} \left(\sum_{j=1,\, j\neq i}^{M} \left| X_i^w - X_j^w \right| \right) \tag{3.3}$$

เมื่อ k_w คือ ครั้งที่มีระยะห่างจากครั้งอื่นน้อยที่สุด ของคำที่ w

จากนั้นคำนวณหาระยะห่างเฉลี่ยของตัวแทนของคำทั้งหมด ซึ่งคือค่าความต่างของเสียงพูด (*DIS*) นั้นเอง ดังสมการที่ 3.4

$$DIS = \frac{1}{{}^{N}C_2} \sum_{i=1}^{N} \sum_{j=i+1}^{N} |X_{k_i}^i - X_{k_j}^j| \tag{3.4}$$

เมื่อ N คือ จำนวนคำที่ใช้ในการทดสอบ

3. การคำนวณค่าความคงที่ของเสียงพูด (Speech Consistency Score, SCS)

SIM และ *DIS* เป็นค่าที่ขึ้นกับผู้รับการทดสอบแต่ละคน ไม่สามารถใช้เปรียบเทียบระหว่างบุคคลได้ ดังนั้นในการทำนายผลการรู้จำเพื่อประเมินและเปรียบเทียบความบกพร่องของสัญญาณเสียงของผู้พูดแต่ละคนจึงต้องใช้อัตราส่วนระหว่าง *SIM* และ *DIS* ซึ่งเรียกว่าค่าความคงที่ของเสียงพูด (*SCS*) ดังสมการที่ 3.6

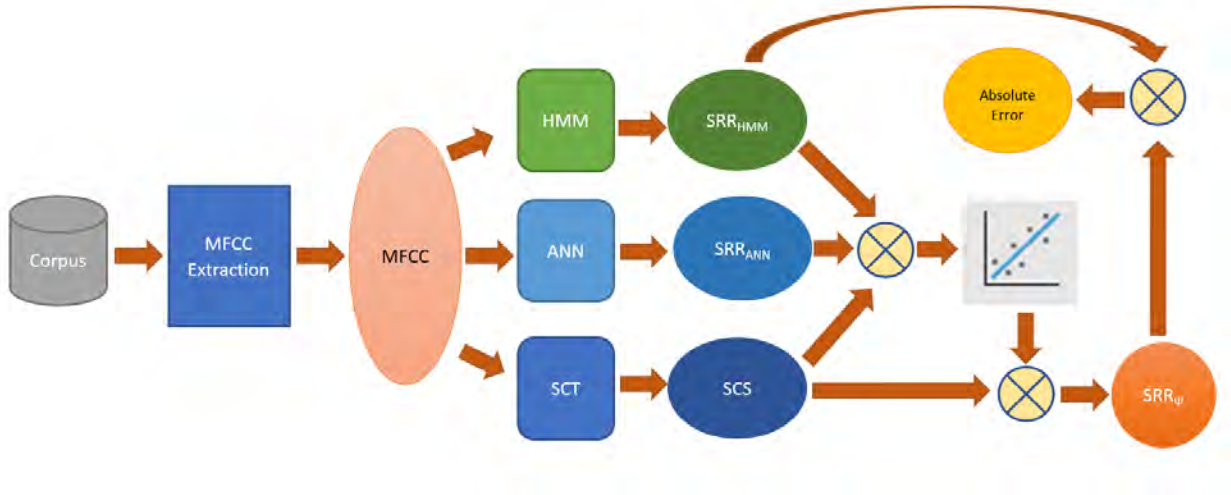
$$SCS = \frac{DIS}{SIM} \tag{3.5}$$

3.5 การทดลองการทำนายอัตราการรู้จำโดยใช้ความคงที่ของเสียงพูด

เนื่องจากค่าอัตราการรู้จำเสียงพูด (Speech Recognition Rate, SRR) เป็นค่าที่แสดงถึงประสิทธิภาพการประยุกต์ระบบการรู้จำเสียงพูดกับการช่วยเหลือเด็กสมองพิการจริง แต่การสร้างระบบการรู้จำเสียงพูดนั้นมีกระบวนการซับซ้อน และต้องใช้ข้อมูลมาก ในขณะที่ค่า SCS หาได้ง่ายกว่า ดังนั้น ถ้าสามารถทำนาย SRR จาก SCS ได้อย่างถูกต้อง จะเป็นการดีมาก เพราะจะช่วยในการตัดสินใจว่า ควรสร้างระบบการรู้จำเสียงพูดเพื่อเด็กคนนั้นหรือไม่ ซึ่งจะช่วยลดความสูญเสียจากการสร้างระบบการรู้จำเสียงพูดแล้วเด็กสมองพิการคนนั้นใช้ไม่ได้

ผังการทดลองการทำนายอัตราการรู้จำโดยใช้ความคงที่ของเสียงพูด แสดงดังรูปที่ 3.2

1. ใช้คลังข้อมูลเสียงพูดสำหรับทดสอบความผิดปกติของเสียงพูด จากผู้รับการทดสอบ 12 คน คือ เด็กปกติ 4 คน และเด็กสมองพิการ 8 คน โดยแต่ละคนพูดทดสอบ 68 คำ ๆ ละ 5 ครั้ง
2. นำเสียงพูดทั้งหมดมาสกัดหา MFCC
3. นำ MFCC ที่ได้ไปใช้กับระบบการรู้จำเสียงพูดแบบ HMM และ ANN เพื่อหาอัตราการรู้จำเสียงพูด (SRR)
4. นำ MFCC เข้าสู่การทดสอบความคงที่ของเสียงพูด (Speech Consistency Test, SCT) เพื่อหาค่า DIS, SIM และ SCS ตามลำดับ
5. นำค่า SCS กับ SRR_{HMM} และ SCS กับ SRR_{ANN} มาหาเส้นแนวโน้มเชิงเส้น (Linear Trendline) และสัมประสิทธิ์การกำหนด (Coefficient of Determination, R^2)
6. ใช้เส้นแนวโน้มเชิงเส้นในการทำนายอัตราการรู้เสียงพูด (SRR_{ψ}) จากกำหนดค่า SCS
7. เปรียบเทียบ SRR_{ψ} กับ SRR_{HMM} และ SRR_{ANN} เพื่อหาความคลาดเคลื่อนสัมบูรณ์



รูปที่ 3.2 ผังการทดลองการทำนายอัตราการรู้จำโดยใช้ความคงที่ของเสียงพูด

บทที่ 4

ผลการทดลองและการวิเคราะห์ผลการทดลอง

ในบทนี้จะกล่าวถึงผลการทดลองผลการทดลองและการวิเคราะห์ผลการทดลอง 2 ส่วน คือ การเปรียบเทียบค่าความคงที่ของเสียงพูดและอัตราการรู้จำเสียงพูด และการทำนายความสามารถในการใช้ระบบรู้จำเสียงพูด

4.1 การเปรียบเทียบค่าความคงที่ของเสียงพูดและอัตราการรู้จำเสียงพูด

ตารางที่ 4.1 แสดงการเปรียบเทียบอัตราการรู้จำเสียงพูดโดยของระบบการรู้จำเสียงพูดโดยใช้ HMM และ ANN (SRR_{HMM} และ SRR_{ANN}) กับ SCS สำหรับกลุ่มเด็กปกติ พบว่าผู้รับการทดสอบที่มีค่า SCS สูงจะมีอัตราการรู้จำเสียงพูดสูง ส่วนผู้รับการทดสอบที่มีค่า SCS ต่ำก็จะมีอัตราการรู้จำเสียงพูดต่ำเช่นกัน โดยที่ค่า SCS เฉลี่ยสำหรับกลุ่มเด็กปกติคือ 2.44 (S.D = 0.150)

ตารางที่ 4.2 แสดงการเปรียบเทียบอัตราการรู้จำเสียงพูดโดยของระบบการรู้จำเสียงพูดโดยใช้ HMM และ ANN กับ SCS สำหรับกลุ่มเด็กสมองพิการ พบว่ามีความสอดคล้องกับกลุ่มเด็กปกติ คือ ผู้รับการทดสอบที่มีค่า SCS สูงจะมีอัตราการรู้จำเสียงพูดสูง ส่วนผู้รับการทดสอบที่มีค่า SCS ต่ำก็จะมีอัตราการรู้จำเสียงพูดต่ำเช่นกัน โดยค่าเฉลี่ยสำหรับกลุ่มเด็กสมองพิการอยู่ที่ 1.76 (S.D = 0.169)

รูปที่ 4.1 แสดงความสัมพันธ์ระหว่างค่า SIM กับค่า DIS โดยรูปที่ 4.1(ก) เป็นผลการทดสอบของกลุ่มผู้รับการทดสอบปกติแต่ละคน จะเห็นได้ว่าค่า SIM จะถูกแสดงอยู่ด้านในและมีค่าน้อยกว่าค่า DIS เสมอ อัตราส่วนระยะห่างระหว่างกราฟทั้งสองเส้นจะมีความหมายแสดงค่า SCS ซึ่งคือถ้ามีค่าสูงแสดงว่ามีค่าความคงที่ของเสียงพูดมาก

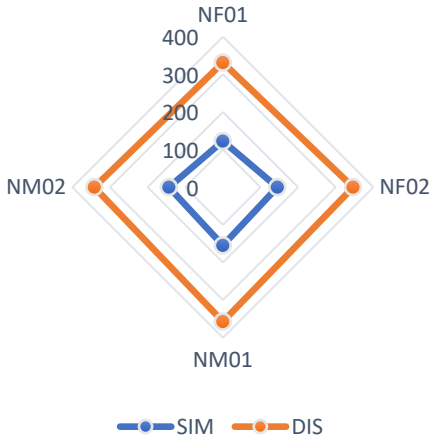
รูปที่ 4.1 (ข) แสดงผลการทดสอบกับเด็กสมองพิการแต่ละคน พบว่าระยะห่างของเส้นกราฟทั้งสองมีค่าน้อยมากกว่าในกรณีของเด็กปกติ ซึ่งหมายความว่า เด็กปกติมีความคงที่ของเสียงพูดสูงกว่าเด็กสมองพิการ

ตารางที่ 4.1 อัตราการรู้จำเสียงพูดโดยใช้ HMM และ ANN และค่าความคงที่เสียงของกลุ่มเด็กปกติ

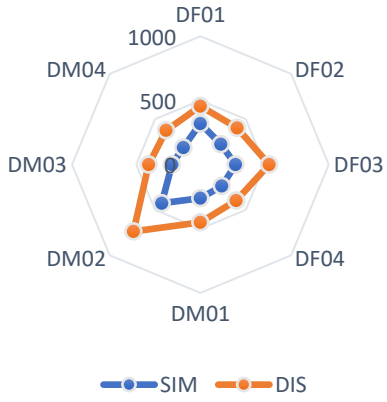
ผู้รับการทดสอบ	SRR_{HMM}	SRR_{ANN}	SIM	DIS	SCS
NF01	0.98	0.95	122.61	331.42	2.70
NF02	0.92	0.84	144.79	346.16	2.39
NM01	0.95	0.84	154.46	357.14	2.31
NM02	0.94	0.91	143.34	341.60	2.38

ตารางที่ 4.2 อัตราการรู้จำเสียงพูดโดยใช้ HMM และ ANN และค่าความคงที่เสียงของกลุ่มเด็กสมองพิการ

ผู้รับการทดสอบ	SRR_{HMM}	SRR_{ANN}	SIM	DIS	SCS
DF01	0.38	0.38	320.11	455.08	1.42
DF02	0.49	0.54	224.76	404.518	1.80
DF03	0.77	0.65	273.69	536.01	1.96
DF04	0.51	0.47	236.31	396.22	1.67
DM01	0.55	0.53	261.44	449.82	1.72
DM02	0.49	0.37	425.59	736.76	1.73
DM03	0.72	0.60	221.86	403.31	1.88
DM04	0.75	0.77	187.97	376.99	2.00



(ก) เด็กปกติ



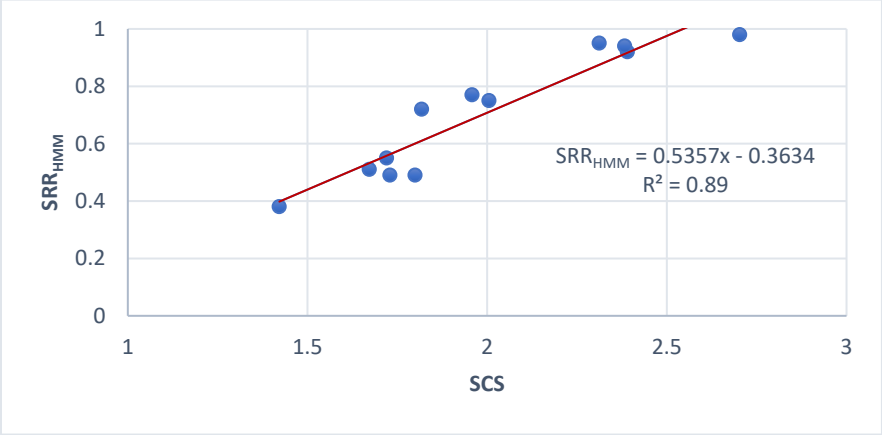
(ข) เด็กสมองพิการ

รูปที่ 4.1 ความต่างระหว่างค่า SIM และค่า DIS ของกลุ่มเด็กปกติ และกลุ่มเด็กสมองพิการ

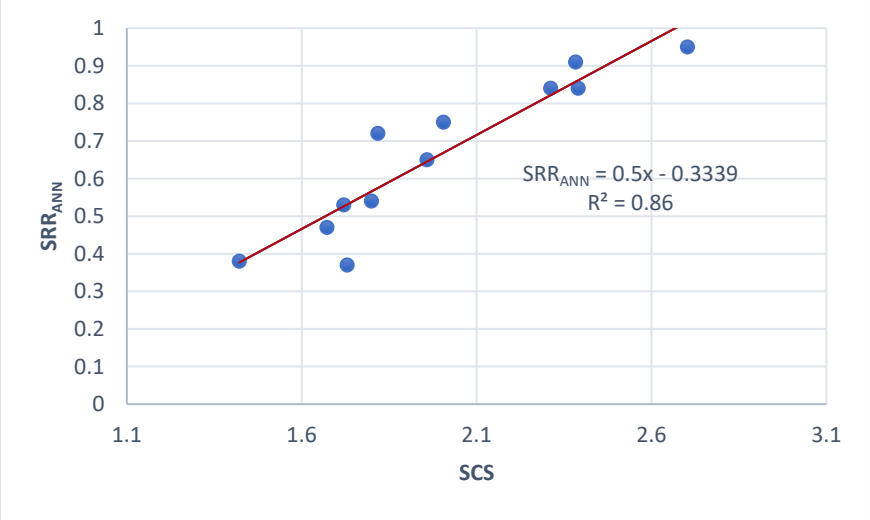
4.2 การทำนายความสามารถในการใช้ระบบรู้จำเสียงพูด

การทำนายความสามารถในการใช้ระบบรู้จำเสียงพูดนั้น จะใช้การสร้างเส้นแนวโน้มเชิงเส้น (Linear Trendline) ระหว่างค่า SCS กับ SRR_{HMM} และค่า SCS กับ SRR_{ANN} จากตารางที่ 4.1 – 4.2 และทำนายเส้นแนวโน้มเชิงเส้นนั้น

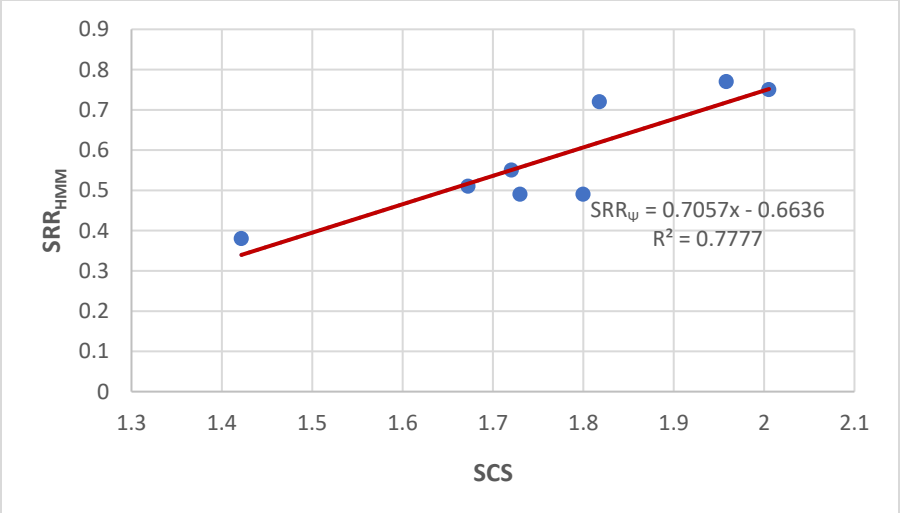
รูปที่ 4.2 – 4.3 แสดงการวิเคราะห์เส้นแนวโน้มเชิงเส้นระหว่างค่า SCS กับ SRR_{HMM} และ SRR_{ANN} โดยมีค่า R^2 เท่ากับ 0.89 และ 0.86 ตามลำดับ ซึ่งแสดงว่าค่า SSC มีความสัมพันธ์กับ SRR สูง โดย SCS จะมีความสัมพันธ์กับ SRR_{HMM} มากกว่าเล็กน้อย จึงเลือกใช้ HMM ในการวิเคราะห์ถัดไป



รูปที่ 4.2 ความสัมพันธ์ระหว่างค่า SCS กับอัตราการรู้จำเสียงพูดเมื่อใช้ HMM สำหรับกลุ่มเด็กปกติ และกลุ่มเด็กสมองพิการ



รูปที่ 4.3 ความสัมพันธ์ระหว่างค่า SCS กับอัตราการรู้จำเสียงพูดเมื่อใช้ ANN สำหรับกลุ่มเด็กปกติ และกลุ่มเด็กสมองพิการ



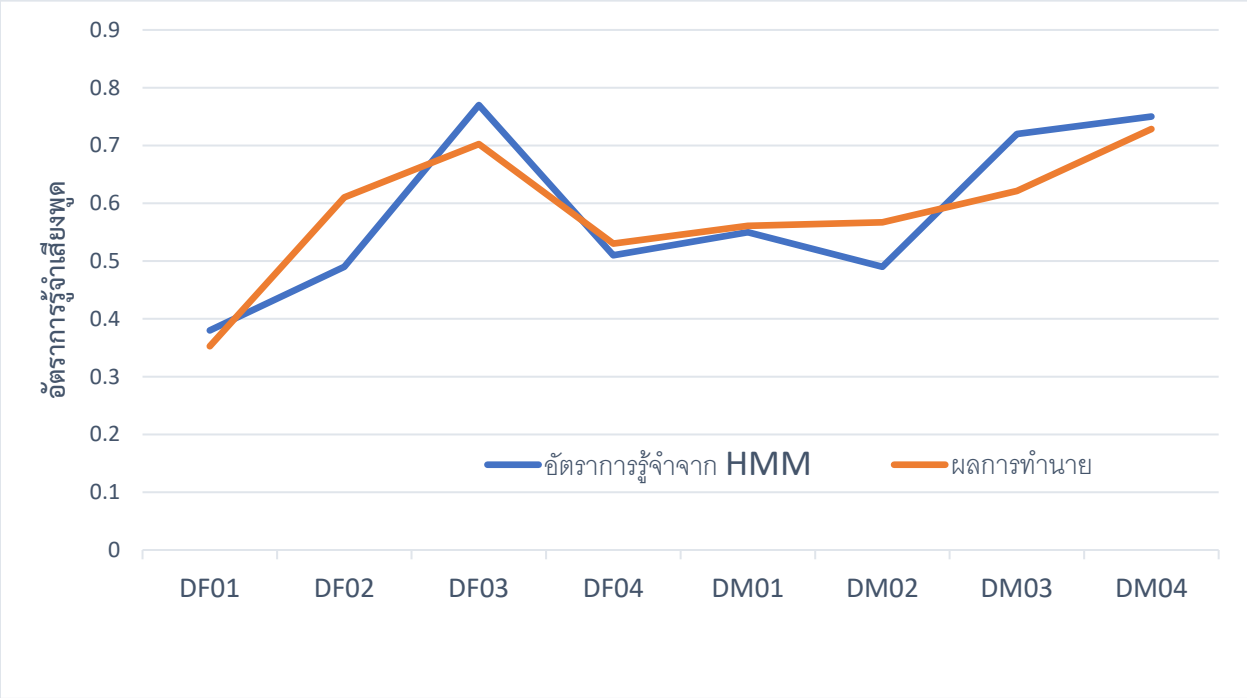
รูปที่ 4.4 ความสัมพันธ์ระหว่างค่า SCS กับอัตราการรู้จำเสียงพูด เมื่อใช้ HMM สำหรับกลุ่มเด็กสมองพิการ

เมื่อวิเคราะห์เฉพาะกลุ่มเด็กสมองพิการ เส้นแนวโน้มเชิงเส้นระหว่างค่า SCS กับ SRR_{HMM} จะให้ค่า R^2 เท่ากับ 0.77 ดังรูปที่ 4.4 โดยเส้นแนวโน้มนี้จะให้ในการทำนายอัตราการรู้จำเสียงพูด (SRR**ψ**) ต่อไป

ตารางที่ 4.3 และรูปที่ 4.5 แสดงการเปรียบเทียบอัตราการรู้จำเมื่อใช้ HMM (SRR_{HMM}) และผลการทำนายอัตรารู้จำเสียงพูด (SRR**ψ**) พบว่าความคลาดเคลื่อนสัมบูรณ์ (Absolute Error) เฉลี่ยเท่ากับ 0.05 (5%) โดยมีค่าต่ำสุด คือ 0.01 (1%) และสูงสุด คือ 0.12 (12%) ซึ่งแสดงว่า SCS มีความสามารถในการทำนาย SRR**ψ** ในระดับที่น่าพอใจมาก

ตารางที่ 4.3 การเปรียบเทียบอัตราการรู้จำเมื่อใช้ HMM และผลการทำนายอัตราการรู้จำเสียงพูด (*SRRψ*)

ผู้บอกภาษา	**SRR_{HMM}**	**SRR**ψ	**ความคลาดเคลื่อนสัมบูรณ์**
DF01	0.38	0.35	0.03
DF02	0.49	0.61	0.12
DF03	0.77	0.70	0.06
DF04	0.51	0.53	0.02
DM01	0.55	0.57	0.01
DM02	0.49	0.62	0.07
DM03	0.72	0.73	0.10
DM04	0.75	0.73	0.02
		เฉลี่ย	**0.05**



รูปที่ 4.5 ความสัมพันธ์ระหว่างอัตราการรู้จำที่ทำนายได้จากค่า SCS กับอัตราการรู้จำเสียงพูดเมื่อใช้ HMM สำหรับกลุ่มเด็กสมองพิการ

บทที่ 5

สรุป และข้อเสนอแนะ

5.1 สรุป

โครงงานนี้จัดทำขึ้นเพื่อพัฒนาการทดสอบความคงที่ของเสียงพูด(Speech Consistency Test) ในผู้ป่วยเด็กสมองพิการแบบดิสอาร์เทรีย และนำค่าความคงที่ของเสียงพูด (SCS) ที่ได้จากการทดสอบ ไปใช้ทำนายอัตราการรู้จำเสียงพูด โดยได้ทำการทดลองกับคลังข้อมูลสำหรับทดสอบความปกติของเสียงพูดเสียงจากกลุ่มเด็กปกติ 4 คน และกลุ่มเด็กสมองพิการ 8 คน [1]

จากการทดลองพบว่า เมื่อเปรียบเทียบค่า SCS กับอัตราการรู้จำเมื่อใช้แบบจำลองฮิดเดน มาร์คอฟ (HMMM) และข่ายงานประสาทเทียม (ANN) (SRR_{HMM} และ SRR_{ANN}) พบว่าผู้รับการทดสอบที่มีค่า SCS สูงจะมีอัตราการรู้จำเสียงพูดสูง ส่วนผู้รับการทดสอบที่มีค่า SCS ต่ำก็จะมีอัตราการรู้จำเสียงพูดต่ำเช่นกัน

เมื่อพิจารณากราฟความสัมพันธ์ระหว่างค่าความเหมือน (SIM) และค่าความต่าง (DIS) พบว่ากราฟของกลุ่มเด็กปกติจะมีความห่างกันของค่า SIM และ DIS มากกว่ากราฟของกลุ่มเด็กสมองพิการ แสดงว่าเด็กปกติมีความคงที่ของเสียงพูดสูงกว่าเด็กสมองพิการ

เมื่อนำค่า SCS มาหาความสัมพันธ์เชิงเส้นกับ SRR_{HMM} และ SRR_{ANN} โดยใช้ผู้บอกภาษาทั้งสองกลุ่ม และสร้างเป็นเส้นแนวโน้มเชิงเส้น (Linear Trendline) พบว่าได้ค่าสัมประสิทธิ์การกำหนด (R^2) เท่ากับ 0.89 และ 0.86 ตามลำดับ ซึ่งมีค่าสูง แสดงว่า SCS มีความสัมพันธ์กับ SRR_{HMM} และ SRR_{ANN} อย่างมาก โดย SCS มีความสัมพันธ์กับ SRR_{HMM} มากกว่ากับ SRR_{ANN} เล็กน้อย

เมื่อวิเคราะห์ความสัมพันธ์เฉพาะกลุ่มเด็กสมองพิการ พบว่าเส้นแนวโน้มเชิงเส้นระหว่างค่า SCS กับ SRR_{HMM} ให้ค่า R^2 เท่ากับ 0.77 ซึ่งยังมีความสัมพันธ์ในระดับสูง

เมื่อนำเส้นแนวโน้มเชิงเส้นที่ได้ไปใช้ในการทำนายอัตรารู้จำเสียงพูด (SRR_{ψ}) พบว่ามีความคลาดเคลื่อนสัมบูรณ์ (Absolute Error) จาก SRR_{HMM} เฉลี่ยเท่ากับ 0.05 (5%) โดยมีค่าต่ำสุด คือ 0.01 (1%) และสูงสุด คือ 0.12 (12%) ซึ่งแสดงว่า SCS มีความสามารถในการทำนาย SRR_{ψ} ในระดับที่น่าพอใจมาก

5.2 ปัญหา และข้อจำกัด

1. ปัญหาที่เกิดจากสัมประสิทธิ์ MFCC

เนื่องจากสัมประสิทธิ์ MFCC บางตัวมีค่าเป็นอนันต์และลบอนันต์ ผู้วิจัยจึงต้องเขียนคำสั่งของโปรแกรมในแทนค่าอนันต์นั้นให้มีค่าเท่ากับศูนย์ ซึ่งจากปัญหาทางด้านเวลาผู้วิจัยไม่สามารถเข้าไปดูและแก้ไขข้อมูลได้ทุกตัว และไม่สามารถทำให้ค่าอนันต์เหล่านั้นมีค่าใกล้เคียงกับค่าสัมประสิทธิ์ที่ใกล้เคียงกันได้เพราะค่าสัมประสิทธิ์ที่สกัดได้นั้นมีเป็นจำนวนมาก

2. เรื่องความไม่เท่าเทียมกันของกลุ่มตัวอย่าง

กลุ่มตัวอย่างที่นำมาทดลองนั้นมีจำนวนของกลุ่มเด็กปกติ 4 คนและกลุ่มเด็กสมองพิการ 8 คน ซึ่งทำให้ข้อมูลที่นำมาทำการทดลองมีการเอนเอียงไปทางกลุ่มเด็กสมองพิการมากกว่ากลุ่มเด็กปกติ

5.3 ข้อเสนอแนะ

1. การเลือกใช้กลุ่มตัวอย่าง

ควรเพิ่มกลุ่มตัวอย่างให้มากขึ้นและมีความเท่ากันในด้านจำนวนคน เพื่อให้ผลการทดลองน่าเชื่อถือมากขึ้น

2. การเลือกใช้ไดนามิกไทม์วอร์ปปิง

ในโครงงานนี้ผู้ทำการทดลองได้ใช้ไดนามิกไทม์วอร์ปปิงแบบดั้งเดิม ซึ่งในปัจจุบันมีงานวิจัยใหม่ ๆ ในการหาระยะห่างด้วยไดนามิกไทม์วอร์ปปิงแบบใหม่ ๆ เพื่อเพิ่มประสิทธิภาพในการหาค่าระยะห่างและอาจทำให้ผลการทำนายอัตราการรู้จำของเสียงพูดดีขึ้นได้

เอกสารอ้างอิง

[1] ณัฐกร ทับทอง, วิสิทธิ์ ลีลาศิริวงศ์, ศรีวิมล มโนเชี่ยวพินิจ และประกาศิต กายะสิทธิ์, “ระบบการรู้จำเสียงพูดภาษาไทยแบบอัตโนมัติสำหรับกลุ่มเด็กสมองพิการที่มีการพูดแบบดิสอาร์เทรีย เล่มที่ 1”, *รายงานการวิจัยพัฒนาและวิศวกรรมฉบับสมบูรณ์ ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ*, 2547.

[2] Bernthal, J. E. and Bankson, N. W., *Articulation and Phonological Disorders* (3rd ed.), Boston: Prentice Hall, 1993.

[3] Blaney B. and Wilson, J., “Acoustic variability in dysarthria computer speech recognition”, *Clinical Linguistics & Phonetics*, 14 (4): 307-327, 2000.

[4] De Bodt, M. S., Hernandez-Daz Huici, M. E., and Van De Heyning, P. H., “Intelligibility as a Linear Combination of Dimensions in Dysarthric Speech”, *Journal of Communication Disorders, 35*, 283 – 292, 2002.

[5] Kent, Ray D., Weismer, G., Kent, J. F., Vorperian, H. K. and Duffy, J. R., “Acoustic Studies of Dysarthric Speech: Methods, Progress, and Potential”, *Journal of Communication Disorders, 32*, 141-186, 1999.

[6] Talbot, N., *Improving the Speech Recognition in the ENABL project*, TMH-QPSR 1/2000.

[7] ณัฐกร ทับทอง, วิสิทธิ์ ลีลาศิริวงศ์, ศรีวิมล มโนเชี่ยวพินิจ และประกาศิต กายะสิทธิ์, “ระบบการรู้จำเสียงพูดภาษาไทยแบบอัตโนมัติสำหรับกลุ่มเด็กสมองพิการที่มีการพูดแบบดิสอาร์เทรีย เล่มที่ 2”, *รายงานการวิจัยพัฒนาและวิศวกรรมฉบับสมบูรณ์ ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ*, 2547.

[8] Stevens, S. S. and Volkman, J., “The relation of pitch to frequency: A revised scale”, *American Journal of Psychology*, 53,329-353, 1940.

[9] Davis, S. and Mermelstein, P., “Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentnces”, *IEEE Transactions on Acoustics, Speech and Signal Processing*, 28(4), 357–366, 1980.

[10] Furui, S., “Speaker independent isolated word recognizer using dynamic featues of speech spectrum”, *IEEE Transactions on Acoustics, Speech and Signal Processing*, 34(1), 52–59, 1986.

[11] นางสาวพิขามญชุ์ อนันตเศรษฐ์, “ไดนามิกไทม์วอร์ปปิงถ่วงน้ำหนักแบบเพิ่มสมรรถนะสำหรับการจำแนกประเภทอนุกรมเวลา”, *วิทยานิพนธ์หลักสูตรปริญญาวิทยาศาสตร์บัณฑิต*, จุฬาลงกรณ์มหาวิทยาลัย, 2554.

[12] ศรีวิมล มโนเชี่ยวพินิ, นันทนา ประชาฤทธิ์ภักดี และ สิริกัญญา เลิศศรัณญพงศ์ “ความสามารถในการเปล่งเสียงภาษาไทยระดับคำของเด็กไทยปกรติวัย 3 – 10 ปี”, *สารศิริราช*, 49: 752-59, 2539.

[13] Shriberg, L., Austin, D., Lewis, B. A., McSweeny, J.L. and Wilson, D. L., “The percentage of consonants correct (PCC) metric: extensions and reliability data”, *Journal of Speech, Language, and Hearing Research*, 40(4): 708 – 722, 1997.