วารสารวิศวกรรมและเทคโนโลยีอุตสาหกรรม มหาวิทยาลัยกาฬสินธุ์ ปีที่ 2 ฉบับที่ 2 มีนาคม – เมษายน 2567
Journal of Engineering and Industrial Technology, Kalasin University, Vol. 2 No. 2 March – April 2024

บทความวิจัย (Research Article)

การเปรียบเทียบเสียงภาษาไทยสำหรับผู้ต้องการเรียนรู้ภาษาไทย

ชญานิษฐ์ เชี่ยวชาญ[1], ทิวาภรณ์ ร่วมทรัพย์[1], ศรัณณ์ลักษณ์ เรียบเรียง[1] และ สุพาพร บรรดาศักดิ์[1*]

[1] สาขาเทคโนโลยีสารสนเทศ คณะวิทยาศาสตร์ ศรีราชา มหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตศรีราชา

*ผู้ประสานงานบทความต้นฉบับ: supaporn.band@ku.th
(รับบทความ: 1 มีนาคม 2567; แก้ไขบทความ: 17 เมษายน 2567; ตอบรับบทความ: 29 เมษายน 2567)

บทคัดย่อ

เนื่องจากช่วงหลังมานี้จะเห็นได้ว่ามีผู้คนสนใจเรียนภาษาไทยเพิ่มมากขึ้น และมาจากหลายประเทศทั่วโลก ทำให้ชาวต่างชาติหันมาสนใจเกี่ยวกับภาษาไทยเพิ่มมากยิ่งขึ้น กระแสความสนใจนี้มาพร้อมกับการเปิดอะไรหลาย ๆ อย่างทั้งในโลกดิจิทัล การเปิดพรมแดนในด้านของภาษา การเดินทางที่ง่ายขึ้น การส่งออกสินค้าไปจนถึงการส่งออกละคร ซีรี่ย์ และกระแส soft power ต่าง ๆ วัตถุประสงค์เพื่อให้ชาวต่างชาติที่เข้ามาเรียนรู้ภาษาไทยหรือเข้ามาทำงานในประเทศไทยได้รู้ว่าตัวเองนั้นพูดภาษาไทยได้ดีและคล้ายคลึงกับเจ้าของภาษามากแค่ไหน จึงทำให้ผู้วิจัยสร้างโมเดลพัฒนาเปรียบเทียบเสียงภาษาไทยสำหรับผู้ที่ต้องการเรียนรู้ภาษาไทย โดยเครื่องมือที่จะนำมาทำการเปรียบเทียบเสียงภาษาไทย ผู้วิจัยใช้โมเดลวิสเปอร์สมอล เป็นการจับคู่เสียงมาเปรียบเทียบเสียงพูดภาษาไทยด้วยความคล้ายของลำดับที่มีความแตกต่างกันในด้านเวลาหรือความเร็ว และมีการถอดข้อความจากเสียงของชาวต่างชาติต้องทำความเข้าใจลักษณะของสัญญาณเสียงในเชิงเวลาและความถี่ เพื่อที่จะใช้ในด้านการทำงานของการสื่อสารและต้องการที่จะฝึกภาษาไทย ซึ่งการออกเสียงภาษาไทยนั้นชาวต่างชาติอาจไม่แน่ใจว่าตนเองพูดได้ชัดเจนหรือคล้ายคลึงกับเจ้าของภาษามากเพียงใด และจากการทดลองที่ผู้วิจัยได้ทำการทดลองได้ผลการวิจัยว่าคำหรือประโยคที่ชาวต่างชาติพูดนั้นนำมาถอดเสียงในโมเดลวิสเปอร์สมอลผลของการถอดเสียงออกมามีความแม่นยำในทางภาษาค่อนข้างสูงมาก

คำสำคัญ: ภาษาไทย เปรียบเทียบเสียงภาษาไทย โมเดลวิสเปอร์สมอล ถอดข้อความจากเสียง

การอ้างอิงบทความ: ชญานิษฐ์ เชี่ยวชาญ และคณะ, "การเปรียบเทียบเสียงภาษาไทยสำหรับผู้ต้องการเรียนรู้ภาษาไทย," *วารสารวิศวกรรมและเทคโนโลยีอุตสาหกรรม มหาวิทยาลัยกาฬสินธุ์*, ปีที่ 2, ฉบับที่ 2, หน้า 36-47, 2567.

วารสารวิศวกรรมและเทคโนโลยีอุตสาหกรรม มหาวิทยาลัยกาฬสินธุ์ ปีที่ 2 ฉบับที่ 2 มีนาคม – เมษายน 2567
Journal of Engineering and Industrial Technology, Kalasin University, Vol. 2 No. 2 March – April 2024

บทความวิจัย (Research Article)

Comparison of Thai sounds for those who want to learn Thai

Chayanit Cheavchan[1], Tiwaporn Raumsub[1], Sarunluk Reabreang[1] and Supaporn Bundasuk[1*]

[1] Information Technology, Faculty of Science Sriracha, Kasetsart University Sriracha Campus

* Corresponding Author: supaporn.band@ku.th
(Received: March 1, 2024; Revised: April 17, 2024; Accepted: April 29, 2024)

Abstract

Because recently it can be seen that more and more people are interested in learning Thai. and come from many countries around the world Making foreigners increasingly interested in the Thai language. This trend of interest comes with the opening of many things in the digital world. Opening borders in the field of language Easier travel Exporting products to exporting dramas, series, and various soft power trends. The objective is for foreigners who come to learn Thai or come to work in Thailand. Knowing how well he spoke Thai and how similar he was to native speakers, the researchers created a model to develop a comparison of Thai sounds for those who want to learn Thai. The tool that will be used to compare the sounds of the Thai language, the researcher uses the Whisper Small model. It involves matching sounds to compare Thai spoken sounds with similarities in sequences that differ in time or speed. And there will be a transcript of the foreigner's voice. You must understand the characteristics of the sound signal in terms of time and frequency in order to use it in the work of communication and want to practice the Thai language. In Thai pronunciation, foreigners may not be sure whether they speak clearly or how similar they are to native speakers. And from the experiment that the researcher conducted, the results were that words or sentences spoken by foreigners were transcribed in the Whisper Small model. The results of the transliteration were quite accurate in terms of language a lot.

Keywords: Thai language, Compare Thai sounds, Model whisper small, Transcribe text from audio

Please cite this article as: C. Cheavchan, et al., “Comparison of Thai sounds for those who want to learn Thai,” *The Journal of Engineering and Industrial Technology, Kalasin University*, vol. 2, no. 2, pp. 36-47, 2024.

บทความวิจัย (Research Article)

1. บทนำ

ในปัจจุบันมีชาวต่างชาติจำนวนมากที่ให้ความสนใจอยากที่จะเรียนรู้ภาษาไทยเพื่อที่จะใช้ในด้านการทำงาน เพื่อที่อยากจะสามารถใช้สื่อสารและต้องการที่จะฝึกภาษาไทย แต่การลองออกเสียงภาษาไทยนั้นผู้ฝึกอาจไม่แน่ใจว่าตนเองพูดได้ดีชัดเจนและคล้ายกับเจ้าของภาษามากน้อย เพียงการฝึกการออกเสียงของนักเรียนที่พึ่งเริ่มเรียนฝึกออกเสียง ด้วยเหตุนี้จึงทำให้เกิดการเปรียบเทียบเสียงภาษาไทยโดยใช้โมเดลวิสเปอร์สมอล (Model whisper small) นี้ขึ้นมา Model whisper small คือ การเปรียบเทียบความคล้ายของลำดับที่มีความแตกต่างกันในด้านเวลาหรือความเร็วโดยแรกเริ่มจะนำชุดข้อมูลภาษาไทยมาใช้งานแต่ก็ยังยากต่อการนำมาวิเคราะห์ เพราะชุดข้อมูลเป็นประโยคภาษาไทยที่ยาวเกินไป จึงจะใช้วิธีการใช้ประโยคที่กระชับเพื่อให้กลายเป็นคำสั้น ๆ การที่จะถอดคำได้แต่ละประโยคนั้นจึงมีความจำเป็นอย่างยิ่งที่จะต้องใช้โมเดลที่มีความแม่นยำสูง หลังจากที่ถอดคำออกมาแล้วนั้นจะทำการวิเคราะห์เสียงของคำแต่ละคำจะทำการวิเคราะห์เสียงจากลักษณะต่าง ๆ ไม่ว่าจะเป็น รูปคลื่น ความเข้มข้นของเสียงและพิทช์ (Pitch) เพื่อมาวิเคราะห์ เมื่อวิเคราะห์เสร็จเรียบร้อยแล้วจะทำการเปรียบเทียบเสียงด้วย whisper small เพื่อหาความคล้ายของเสียง

2. เอกสารและทฤษฎีที่เกี่ยวข้อง

2.1 เอกสารและงานวิจัยที่เกี่ยวข้อง

2.2.1 ภาษาไทย

ภาษาไทยมีความเก่าแก่มาก ภาษาไทยมีที่มาจากที่ออสโต แต่มีความคล้ายกับภาษาจีน เพราะมีหลากหลายคำที่ยืมมาจากภาษาจีน ภาษาไทยได้ถูกประดิษฐ์ขึ้นมาเมื่อปี 1826 ถูกดัดแปลงมาจากภาษาบาลี และสันสกฤต มีพยัญชนะทั้งหมด 44 ตัว สระ 21 รูป และวรรณยุกต์ 5 เสียง ภาษาไทยแบ่งออกเป็น 2 สมัย ดังนี้

1. ภาษาไทยแบบดั้งเดิม คือ เป็นภาษาไทยที่อพยพมาอยู่ในสุวรรณภูมิ

2. ภาษาไทยประสม คือ ภาษาไทยที่ถูกนับตั้งแต่เข้ามาอยู่ในสุวรรณภูมิ [1]

2.2.2 การวิเคราะห์เสียง (Sound analysis)

ลักษณะของเสียงมี 2 รูปแบบดังนี้

1) รูปคลื่น (Waveform) คือ สัญญาณกราฟฟิคในรูปแบบของคลื่นเปลี่ยนแปลงระดับเมื่อเวลาผ่านไป waveform จะมีการเปลี่ยนแปลงหลายครั้งในระยะเวลาสั้น ๆ [2] ซึ่งมี 3 ลักษณะดังนี้

1.1) ความยาวคลื่น (Wavelength) ความยาวจากจุดเริ่มต้นไปยังจุดสิ้นสุด เป็นระยะห่างของลูกคลื่นสองลูกติดต่อกันแต่มีความสอดคล้องกัน ความยาวของคลื่นจะวัดจากยอดหนึ่งไปยังอีกยอดหนึ่ง [3]

1.2) แอมพลิจูด (Amplitude) คือ เป็นรูปแบบหนึ่งที่เกิดจากการสั่นสะเทือนจนเกิดเสียง เป็นการกระจัดสูงสุดจากเส้นศูนย์กลางถึงจุดสูงสุด ถ้าระยะห่างของกึ่งกลางหากจากเส้นมากขึ้น ความแปรผันของความดันจะยิ่งรุนแรงขึ้น

1.3) ความถี่ (Frequency) คือ จำนวนของคลื่นต่อวินาทีที่ผ่านจุดหนึ่งไป จะระบุอัตราการแปรผันของคลื่น [4]

2) สเปกตรัม (Spectrogram) คือ การแสดงภาพเพื่อแสดงความแรงของสัญญาณ ที่ปรากฏในรูปคลื่น สามารถดูระดับพลังงานตามเวลาได้ และยังสามารถใช้เพื่อแสดงความถี่ของคลื่นเสียง อีกทั้งยังแสดงแอมพลิจูดทั้งหมดบนกราฟเดียวได้อีกด้วย สเปกตรัมประกอบด้วย ความเข้มของเสียงและพิทช์ [5] ดังนี้

2.1) ความเข้มของเสียง (Intensity) คือ คลื่นตามความยาวที่ทำให้ได้ยินเสียงขึ้นมา ไม่สามารถวัดความดังได้อย่างเป็นกลางเนื่องจากทุกคนมีการรับรู้หรือได้ยินเสียงที่แตกต่างกันออกไป ความเข้มเสียงที่

สามารถรับรู้ได้มีค่ากว้างมาก ตั้งแต่ 10-12 Wm-2 ถึง 10 Wm-2 [6]

2.2) พิทซ์ (Pitch) คือ ระดับเสียงที่ขึ้นอยู่กับความถี่ของการสั่นของคลื่นถ้าความถี่จากการสั่นสะเทือนสูงขึ้น จะบอกได้ว่าเสียงนั้นมีความถี่และโหยหวน ความถี่มากกว่าก็จะทำให้ระดับเสียงสูงกว่า ในขณะที่ความถี่น้อยกว่าก็จะทำให้เกิดระดับเสียงที่ต่ำกว่า ซึ่งความถี่จะเป็นตัวกำหนดของระดับเสียง ได้มาจากการวัดความสั่นสะเทือนของวัตถุ ซึ่งพิทซ์ (pitch) จะมีความเกี่ยวข้องกับความถี่ (frequency) ความถี่นั้นจะเป็นตัวกำหนดระดับของเสียง มีลักษณะดังนี้

(1) คลื่นเสียงที่ความถี่ต่ำจะมีความยาวคลื่นที่ยาวและระดับเสียงที่ต่ำ

(2) คลื่นเสียงความถี่สูงจะมีความยาวที่สั้นและมีระดับเสียงที่สูง [7]

2.2.3 การสอนแบบโฟนิกส์

เป็นการศึกษาเสียงของการพูด การเปล่งเสียงและการออกเสียง การสอนภาษาแบบโฟนิกส์เป็นการสอนที่แสดงให้เห็นความสัมพันธ์ระหว่างเสียงและตัวอักษร วิธีการอ่านที่ถูกต้องควรเริ่มจากการฝึกการอ่าน การรู้พยัญชนะ สระ วรรณยุกต์ การผสมคำที่ทำให้เกิดความหมาย ขั้นตอนของวิธีการสอนแบบโฟนิกส์ มีอยู่ 4 ขั้นตอนการสอน ดังนี้

1) การออกเสียงคำแบบการวิเคราะห์เสียง (analytic phonics) ช่วยฝึกให้ผู้เรียน เรียนรู้และช่วยวิเคราะห์เสียงสามารถแยกแยะเสียงของแต่ละตัวอักษร นำไปสู่การพัฒนาเสียงของตัวอักษรในแต่ละคำได้อย่างมีระบบ

2) การเทียบเคียงในการออกเสียง (analogies phonics instruction) เป็นการสอนที่ต่อยอดมาจากการออกเสียงคำแบบการวิเคราะห์เสียง โดยผู้เรียนจะวิเคราะห์ส่วนประกอบของคำและเทียบเคียงการออกเสียงกับคำที่รู้

3) การสร้างคำด้วยการออกเสียง (Synthetic phonics) เป็นวิธีการเรียนรู้จากการสร้างคำของตัวอักษรผ่านการออกเสียงแล้วนำตัวอักษรแต่ละตัวมาผสมจนเกิดเป็นคำที่ฟังแล้วมีความหมายขึ้นมา

4) การอ่านออกเสียงแบบร่วมสมัย (contemporary phonics) เป็นการนำเอาวิธีการการออกเสียงคำแบบการวิเคราะห์เสียงและวิธีการสังเคราะห์การออกเสียงมารวมกัน เป็นการฝึกให้ผู้เรียนผสมคำใช้ตามความสามารถของผู้ที่เรียน กระบวนการการสอนในการเรียนรู้แบบโฟนิกส์ จะช่วยให้ผู้เรียนได้ฝึกฝนการอ่านอย่างเป็นขั้นตอนและมีความต่อเนื่องกัน ได้พัฒนาการฟังของผู้เรียน การพูดของผู้เรียน และการอ่านไปพร้อม ๆ กัน โดยผู้วิจัยได้จัดการการสอนภาษาแบบโฟนิกส์ให้เรียนรู้ตามขั้นตอนที่ได้ไว้ข้างต้นเพื่อไปพัฒนาการอ่านออกเสียงของชาวต่างชาติ [8]

2.2.4 การถอดข้อความจากเสียง

การถอดข้อความจากเสียง คือ การแปลงเสียงจากการฟังเสียง เขียนคำพูดใหม่ให้ออกมาเป็นไฟล์ข้อความ มีหลายประเภทด้วยกันยกตัวอย่าง เช่น

1) การสัมภาษณ์เกิดขึ้นบ่อยครั้งตามหน้าหนังสือพิมพ์ นักข่าวมักจะเขียนคำพูดของผู้ตอบหรือบุคคลที่ถูกสัมภาษณ์ได้อย่างมีคุณภาพ

2) การบันทึกของศาลการถอดข้อความเสียงมักจะมีความซับซ้อนมาก จะต้องระบุคำพูดของผู้พูดทั้งหมดอย่างชัดเจน ถอดรหัสคำ วลีทั้งหมด (รวมถึงเสียงกระซิบ อุทาน การทะเลาะวิวาท)

3) การบันทึกการประชุมงานนี้ค่อนข้างยาก เพราะมีผู้เข้าร่วมประชุมเป็นอย่างมาก อาจมีการขัดจังหวะกันในการพูด พูดไม่ดังพอ หรือเสียงสะท้อนจากห้องที่กว้างของที่ประชุม อาจต้องตั้งใจฟังเป็นพิเศษ เป็นต้น [9]

ดังนั้นเทคนิคที่ใช้ในการถอดเสียงพูดภาษาไทย คือ โมเดลวิสเปอร์สมอล (Model whisper small) เป็นโมเดลที่มีความสามารถในการทำความเข้าใจลักษณะของสัญญาณเสียงในเชิงเวลาและความถี่ได้ดีมาก โมเดลแต่เดิมสร้างขึ้นเพื่อใช้กับภาษาอังกฤษเป็น

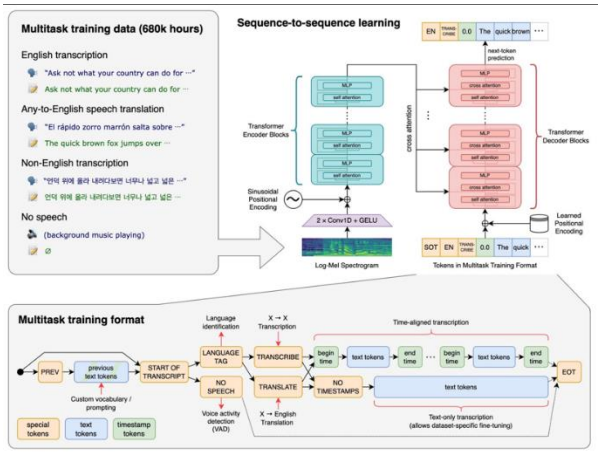
บทความวิจัย (Research Article)

ภาษาหลัก ดังนั้นการนำมาใช้กับภาษาไทยจึงจำเป็นต้องมีการเทรนเพิ่มให้เข้ากับชุดข้อมูลเสียงภาษาไทย จะได้ใช้กับการถอดข้อความเสียงได้อย่างแม่นยำยิ่งขึ้น ซึ่งผลการทำนายของคำได้ค่อนข้างที่จะแม่นยำแต่อาจจะมีความผิดพลาดอยู่เล็กน้อย [10]

2.2.5 Model Whisper

เป็นโมเดลย่อยตัวหนึ่งของ Whisper ที่ได้รับการฝึกล่วงหน้าสำหรับผู้ที่ต้องการรู้จักการจดจำเสียงคำพูดโดยอัตโนมัติ เป็นโมเดลการเรียนรู้ของเครื่องปัญญาประดิษฐ์ที่พัฒนาโดย OpenAI ประกาศปล่อยตัว Whisper ที่แปลงเสียงเป็นข้อความโมเดลนี้ถูกออกแบบมาเพื่อการจดจำเสียงพูดอัตโนมัติ (ASR) หรือแปลงคำพูดเป็นข้อความ Whisper ถูกสร้างขึ้นด้วยสถาปัตยกรรมแบบ end-to-end ที่มีการใช้ Transformer เป็น Encoder และ Decoder จะใช้การทำนายโมเดลพร้อมกับแปลข้อความเป็นภาษาที่เราสนใจจะแปลง โดยตัวโมเดลที่ปล่อยออกมามีจำนวนพารามิเตอร์อยู่ 4 ขนาด tiny, base, small, medium, large เท่ากับ 39M, 74M, 244M, 769M, 1550M ตามลำดับ

จุดเด่นของ Whisper Small คือ ขนาด ถึงแม้ว่าขนาดของโมเดลจะเล็กว่า Whisper รุ่นอื่น ๆ ขนาดเล็กใช้พลังงานในการประมวลผลน้อยกว่า Whisper small มีความแม่นยำสูง โมเดลสามารถจดจำเสียงพูดได้อย่างถูกต้องแม้ในสภาพแวดล้อมที่มีเสียงรบกวน เหมาะสำหรับการใช้งานที่ต้องการความแม่นยำสูง เช่น การถอดเสียงการประชุม หรือ การบันทึกเสียง Whisper Small รองรับการจดจำเสียงพูดได้หลายภาษา ถึงความแม่นยำในแต่ละภาษาจะมีความแตกต่างกันออกไป ภาษาที่มีความผิดพลาดต่ำสุด เช่น สเปน อิตาลี อังกฤษ และโปรตุเกส (อัตราการผิดพลาด WER ต่ำกว่า 5.0) แต่ในส่วนของภาษาไทยนั้นมี WER ที่ 13.2 และภาษาเกาหลีมี WER ที่ 15.2 ภาษาในอาเซียนอื่น ๆ ยังมีความผิดพลาดค่อนข้างสูง เช่น ลาวอยู่ที่ 101.6 เมียนมาร์อยู่ที่ 124.5 [11]



รูปที่ 1

ที่มา: https://www.blognone.com/node/130549

2.2.6 Find-tuning whisper

การใช้ Find-tune ตัวโมเดล Whisper จะช่วยให้การทำนายของตัวโมเดลนั้นเกิดความผิดพลาดได้น้อยลงและสร้างความแม่นยำให้กับตัวการทำนายได้เพิ่มมากขึ้น ซึ่งตัวโมเดล Whisper นี้ก็ได้ถูกพัฒนาขึ้นเพื่อที่จะให้มีความเหมาะสมกับภาษาไทยและทางบริษัทผู้พัฒนายังเปิดโมเดลเป็นสาธารณะให้กับนักพัฒนาและผู้ใช้งานทั่วไปได้ทดลองใช้กันเพื่อเพิ่มความหลากหลายของตัวเลือกโมเดล ASR ที่มีในปัจจุบัน [10]

3. วิธีดำเนินวิจัย

3.1 การวิเคราะห์การเปรียบเทียบเสียงภาษาไทย

แนวคิดในการวิเคราะห์คือจะนำชุดข้อมูลเสียงภาษาไทยมาทำการถอดความทั้งหมด 15 ประโยค และจากนั้นจะนำประโยค 5 จาก 15 ประโยคมาให้ชาวต่างชาติเพศหญิง 3 คนพูดประโยคเดียวกันหลังจากสกัดข้อมูลที่ได้มาแล้ว เพื่อนำมาเปรียบเทียบว่าเสียงที่ถอดความออกมามีความเปลี่ยนแปลงหรือคล้ายคลึงมากน้อยแค่ไหน

บทความวิจัย (Research Article)

3.2 เตรียมชุดข้อมูลเสียงภาษาไทย

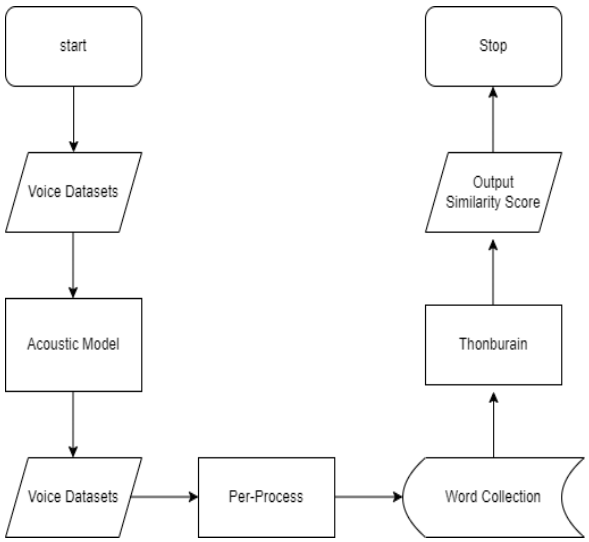
ตารางที่ 1 ชุดข้อมูลภาษาไทย

ชื่อไฟล์เสียง	ชุดประโยค
1.weba	ช่วยโทรหาฉันพรุ่งนี้ด้วย ถ้าคุณว่าง
2.weba	ฉันปวดหัวมาสักพักแล้ว
3.weba	เดือนพฤศจิกายนนี้จะมีคอนเสิร์ต
4.weba	ฉันคิดว่าคนไทยใจดีมากๆเลยนะคะ
5.weba	ขอโทษนะคะ ห้องน้ำอยู่ไหนหรอคะ

จากตารางที่ 1 ชุดข้อมูลนี้ประกอบไปด้วย ชื่อไฟล์เสียง และชุดประโยค ที่เป็นถูกต้องตามหลักภาษาไทย

3.3 Flowchart

ลำดับการพัฒนาระบบเสียงภาษาไทย



รูปที่ 2 Flowchart การพัฒนาระบบการเปรียบเทียบเสียงภาษาไทย

3.4 แสดงผลลัพธ์จากโมเดล

1) ทำการแปลงข้อมูลโดยใช้เทคนิค Whisper Small โดยจะทำการแปลงไฟล์เสียงให้ออกมาเป็นชุดข้อมูลเทสของ speaker_001

ตารางที่ 2 ชุดข้อมูลภาษาไทย

ประโยคที่พูด	ผลลัพธ์ที่ได้
ช่วยโทรหาฉันพรุ่งนี้ด้วย ถ้าคุณว่าง	ช่วยโทรหาฉันพรุ่งนี้ด้วย ถ้าคุณว่าง
ฉันปวดหัวมาสักพักแล้ว	ฉันปวดหัวมาสักพักแล้ว
เดือนพฤศจิกายนนี้จะมีคอนเสิร์ต	เดือนพฤษจิกายนต์นี้จะมีคอนเสิร์ต
ฉันคิดว่าคนไทยใจดีมาก ๆ เลยนะคะ	ฉันคิดว่าคนไทยใจดีมาก ๆ เลยนะคะ
ขอโทษนะคะ ห้องน้ำอยู่ไหนหรอคะ	ขอโทษนะคะ ห้องน้ำอยู่ไหนหรอคะ

จากตารางที่ 2 ผลลัพธ์ที่ได้มาจากโมเดล เห็นได้ว่าค่อนข้างมีความละเอียด แต่ก็มีบางจุดของแต่ละประโยคที่ผลของการทำนายผลลัพธ์ออกมาไม่ตรง คือ (ตัวอักษรสีแดง) เช่น ประโยคที่ 3 ทำให้การทำนายที่ออกมาเป็นความถูกต้องจาก 100% เป็น 99%

2) ทำการแปลงข้อมูลโดยใช้เทคนิค Whisper Small โดยจะทำการแปลงไฟล์เสียงให้ออกมาเป็นชุดข้อมูลเทสของ speaker_002

ตารางที่ 3 ชุดข้อมูลภาษาไทย

ประโยคที่พูด	ผลลัพธ์ที่ได้
ช่วยโทรหาฉันพรุ่งนี้ด้วย ถ้าคุณว่าง	ช่วยโทรหาฉันพรุ่งนี้ด้วย ถ้าคุณว่าง
ฉันปวดหัวมาสักพักแล้ว	ฉันโปรดหัวมาสักครักแล้ว
เดือนพฤศจิกายนนี้จะมีคอนเสิร์ต	เดือนพฤษจิกายนี้จะมีคอนเสิร์ต
ฉันคิดว่าคนไทยใจดีมาก ๆ เลยนะคะ	ฉันคิดว่าคนไทยใจดีมาก ๆ เลยนะคะ
ขอโทษนะคะ ห้องน้ำอยู่ไหนหรอคะ	ขอโทษนะคะ ห้องน้ำอยู่ไหนหรอคะ

จากตารางที่ 3 ผลลัพธ์ที่ได้มาจากโมเดล เห็นได้ว่าค่อนข้างมีความละเอียด แต่ก็มีบางจุดของแต่ละประโยคที่ผลของการทำนายออกมาไม่ตรง คือ (ตัวอักษรสีแดง) เช่น ประโยคที่ 2 และประโยคที่ 3 จากประโยคที่ 2 ทำให้การทำนายที่ออกมาเป็นความ

ถูกต้องจาก 100% เป็น 98% และจากประโยคที่ 3 ทำให้การทำนายที่ออกมาเป็นความถูกต้องจาก 100% เป็น 99%

3) ทำการแปลงข้อมูลโดยใช้เทคนิค Whisper Small โดยจะทำการแปลงไฟล์เสียงให้ออกมาเป็นชุดข้อมูลเทสของ speaker_003

ตารางที่ 4 ชุดข้อมูลภาษาไทย

ประโยคที่พูด	ผลลัพธ์ที่ได้
ช่วยโทรหาฉันพรุ่งนี้ด้วย ถ้าคุณว่าง	ช่วยโทรหาฉันพุ่นนี้ด้วย ถ้าคุณว่าง
ฉันปวดหัวมาสักพักแล้ว	ฉันปวดหัวมาจากพักแล้ว
เดือนพฤศจิกายนนี้จะมีคอนเสิร์ต	เดือนพฤศจิกายนี้จะมีคอนเสอร์
ฉันคิดว่าคนไทยใจดีมากๆ เลยนะคะ	ฉันคิดว่าคนไทยใจดีมักมากเลยนะคะ
ขอโทษนะคะ ห้องน้ำอยู่ไหนหรอคะ	ขอโทษนะคะ ห้องน้ำอยู่ไหนหรอคะ

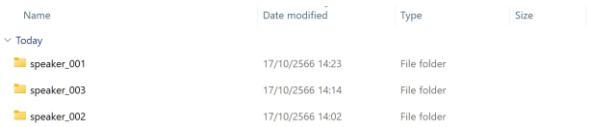
จากตารางที่ 4 ผลลัพธ์ที่ได้มาจากโมเดล เห็นได้ว่าค่อนข้างมีความละเอียด แต่ก็มีบางจุดของแต่ละประโยคที่ผลของการทำนายออกมาไม่ตรง คือ (ตัวอักษรสีแดง) เช่น ประโยคที่ 1 ประโยคที่ 2 ประโยคที่ 3 และประโยคที่ 4

จากประโยคที่ 1 ทำให้การทำนายที่ออกมาเป็นความถูกต้องจาก 100% เป็น 99% จากประโยคที่ 2 ทำให้การทำนายที่ออกมาเป็นความถูกต้องจาก 100% เป็น 99% จากประโยคที่ 3 ทำให้การทำนายที่ออกมาเป็นความถูกต้องจาก 100% เป็น 98% และจากประโยคที่ 4 ทำให้การทำนายที่ออกมาเป็นความถูกต้องจาก 100% เป็น 99%

3.5 คลังข้อมูล (Word Collection)

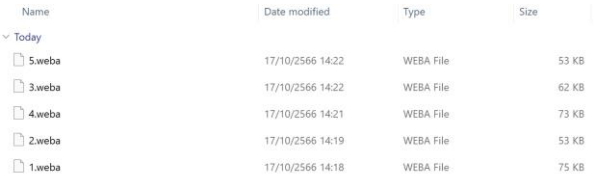
นำข้อมูลที่ทำการแปลงข้อมูลไปเก็บในคลัง ขั้นตอนการออกแบบโครงสร้างข้อมูล จะประกอบไปด้วย

1) เริ่มทำการแบ่งเสียงชาวต่างชาติเพศหญิงออกเป็น 3 คน คนละ 5 เสียง



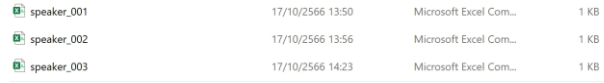
รูปที่ 3 การรวบรวมคำครั้งที่ 1

2) ข้อมูลประโยคทั้งหมด 5 ประโยค ไว้ใช้สำหรับเพื่อการถอดความ



รูปที่ 4 การรวบรวมคำครั้งที่ 2

3) แบ่งประโยคข้อมูลพากย์เสียงของแต่ละคน และนำข้อมูลทั้งหมดที่ได้ทำการถอดเสียงออกมา ได้แก่ ข้อมูลที่ได้มาจากการถอดคำ



รูปที่ 5 การรวบรวมคำครั้งที่ 3

3.6 การนำไปใช้

การศึกษาวิจัยในครั้งนี้ นำไปใช้เพื่อวิเคราะห์ว่าชาวต่างชาติที่ต้องการจะฝึกพูดภาษาไทยนั้น พูดได้ใกล้เคียงกับเจ้าของภาษามากน้อยแค่ไหน

4. ผลการดำเนินงาน

4.1 ทดสอบและประเมินผลการทดสอบ

การวิเคราะห์จากหัวข้อที่ผ่านมาในการถอดข้อความเสียงจากประโยคพูดภาษาไทยออกมาด้วย Model Whisper Small เพื่อนำข้อมูลของประโยคที่

บทความวิจัย (Research Article)

ได้มาวิเคราะห์ว่า สามารถถอดแบบเสียงภาษาไทยออกมาได้ถูกต้องมากน้อยเพียงใด โดยข้อมูลเสียงจะมีทั้งหมด 5 ประโยคทั้งประโยคที่ยาว ประโยคที่สั้น และความยากง่ายของรูปแบบการเขียนภาษาไทย ซึ่งให้ชาวต่างชาติเพศหญิง 3 คนมาพูดประโยคเดียวกันทั้งหมด เพื่อให้ได้ค่าความเหมือนของการถอดความนี้ซึ่งจะมีการจับคู่เสียงคู่เพื่อการทดสอบดังนี้

ตารางที่ 5 แสดงผลลัพธ์เสียงของ speaker_001

ประโยคที่พูด	ผลลัพธ์ที่ได้	คำที่ผิด	เปอร์เซ็นต์ความถูกต้อง	World error rate
ช่วยโทรหาฉันพรุ่งนี้ด้วยถ้าคุณว่าง	ช่วยโทรหาฉันพรุ่งนี้ด้วย ถ้าคุณว่าง	0	100%	295.899
ฉันปวดหัวมาสักพักแล้ว	ฉันปวดหัวมาสักพักแล้ว	0	100%	295.899
เดือนพฤศจิกายนนี้จะมีคอนเสิร์ต	เดือนพฤษจิกายนต์นี้จะมีคอนเสิร์ต	1	90%	295.899
ฉันคิดว่าคนไทยใจดีมากๆเลยนะคะ	ฉันคิดว่าคนไทยใจดีมาก ๆ เลยนะคะ	0	100%	295.899
ขอโทษนะคะ ห้องน้ำอยู่ไหนหรอคะ	ขอโทษนะคะ ห้องน้ำอยู่ไหนหรอคะ	0	100%	295.899

จากตารางที่ 5 ผลของการถอดเสียงออกมาได้ผลลัพธ์ที่ถูกต้อง 100% ทั้งหมด 4 ประโยค และมี 1 ประโยคที่ถอดเสียง ออกมาผิด 1 คำ

ตารางที่ 6 แสดงผลลัพธ์เสียงของ speaker_002

ประโยคที่พูด	ผลลัพธ์ที่ได้	คำที่ผิด	เปอร์เซ็นต์ความถูกต้อง	World error rate
ช่วยโทรหาฉันพรุ่งนี้ด้วย ถ้าคุณว่าง	ช่วยโทรหาฉันพรุ่งนี้ด้วย ถ้าคุณว่าง	0	100%	295.899
ฉันปวดหัวมาสักพักแล้ว	ฉันโปรดหัวมาสักครักแล้ว	2	98%	295.899
เดือนพฤศจิกายนนี้จะมีคอนเสิร์ต	เดือนพฤษจิกายนนี้จะมีคอนเสิร์ต	1	99%	295.899
ฉันคิดว่าคนไทยใจดีมากๆเลยนะคะ	ฉันคิดว่าคนไทยใจดีมาก ๆ เลยนะคะ	0	100%	295.899
ขอโทษนะคะ ห้องน้ำอยู่ไหนหรอคะ	ขอโทษนะคะ ห้องน้ำอยู่ไหนหรอคะ	0	100%	295.899

จากตารางที่ 6 ผลของการถอดเสียงออกมาได้ผลลัพธ์ที่ถูกต้อง 100% ทั้งหมด 3 ประโยค มี 1 ประโยคที่ถอดเสียง ออกมาผิด 1 คำ และมี 1 ประโยคที่ถอดเสียงออกมาผิด 2 คำ

ตารางที่ 7 แสดงผลลัพธ์เสียงของ speaker_003

ประโยคที่พูด	ผลลัพธ์ที่ได้	คำที่ผิด	เปอร์เซ็นต์ความถูกต้อง	World error rate
ช่วยโทรหาฉันพรุ่งนี้ด้วย ถ้าคุณว่าง	ช่วยโทรหาฉันพุ่นนี้ด้วย ถ้าคุณว่าง	1	99%	295.899
ฉันปวดหัวมาสักพักแล้ว	ฉันปวดหัวมาจากพักแล้ว	1	99%	295.899
เดือนพฤศจิกายนนี้จะมีคอนเสิร์ต	เดือนพฤศจิกายนี้จะมีคอนเสอร์	2	98%	295.899
ฉันคิดว่าคนไทยใจดีมากๆเลยนะคะ	ฉันคิดว่าคนไทยใจดีมักมากเลยนะคะ	1	99%	295.899
ขอโทษนะคะ ห้องน้ำอยู่ไหนหรอคะ	ขอโทษนะคะ ห้องน้ำอยู่ไหนหรอคะ	0	100%	295.899

จากตารางที่ 7 ผลของการถอดเสียงออกมาได้ผลลัพธ์ที่ถูกต้อง 100% ทั้งหมด 1 ประโยค มี 1 ประโยค

บทความวิจัย (Research Article)

ที่ถอดเสียง ออกมาผิด 1 คำ และมี 1 ประโยคที่ถอดเสียงออกมาผิด 2 คำ

นิยามของ WER (Word Error Rate) นั้นคือ การกำหนดข้อความที่ถูกต้องกับข้อความที่คอมพิวเตอร์ฟังได้นั้นยาว N คำ คอมพิวเตอร์ฟังผิดจากข้อความที่แทรกมาจากข้อความเดิม I คำ และเป็นคำที่หายไปจากเดิม D คำ และเป็นคำที่ถูกแทนที่คำเดิมไป S คำ ดังนั้น WER (Word Error Rate) คิดได้ดังนี้ WER = (I + D + S) / N ค่า WER จะวัดเป็น % ถ้ายิ่งค่า WER เยอะประสิทธิภาพก็จะยิ่งดี

4.2 โมเดลถอดเสียง

ใช้โมเดล Whisper Small เป็นโมเดลที่ถอดความจากเสียงพูดภาษาไทย ที่ถูกพัฒนามาจาก Open AI ซึ่ง Whisper นั้น เป็น OpenAI ที่ถูกพัฒนามาถูกเทรนด้วยเสียงของภาษาไทยจำนวนมาก จึงมีโมเดลที่สามารถเข้าใจภาษาไทยได้หลายหลายและจึงมีความแม่นยำค่อนข้างสูงมาก แต่จริง ๆ แล้ว Whisper นั้นถูกเทรนขึ้นมาเพื่อใช้กับภาษาอังกฤษเป็นหลักมากกว่า

4.3 การใช้ Fine – tune

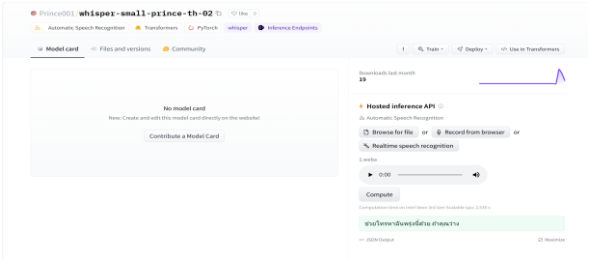
การที่เราจะนำมาใช้เป็นภาษาไทยนั้นจะต้องมีการเติม Fine – tune กับชุดข้อมูลของภาษาไทยนั้นก่อน เพื่อให้โมเดลสามารถที่จะรับรู้และเข้าใจในภาษาไทยได้อย่างแม่นยำ โมเดล Whisper นั้นจะสามารถตรวจจับภาษาได้หลายภาษานั้นแต่กับภาษาไทยยังมีข้อผิดพลาดอยู่หลายประการ เพื่อให้ใช้กับภาษาไทยได้อย่างมีประสิทธิภาพนั้นเราจะต้องนำโมเดลนี้มา Fine – tune ก่อนเพื่อให้เหมาะสมกับภาษาไทย

4.4 ผลของการทำนายและคุณภาพของ Model

โมเดลนั้นแสดงผลออกมาได้ค่อนข้างแม่นยำมาก แต่ก็ยังมีบางคำที่ยาก ๆ หรือสะกด ออกเสียงสอดคลองกับคำง่าย ๆ ที่ยังมีความผิดพลาด แต่โดยรวมส่วนใหญ่แล้ว Whisper small ทำได้ออกมาได้ใกล้เคียงและมีประสิทธิภาพสูง

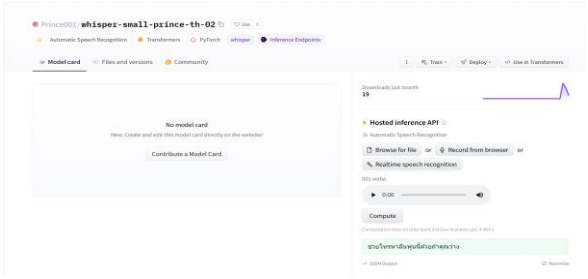
4.5 ผลของการการใช้ Model Whisper Small

การใช้โมเดล ในการทำนาย ทำดังนี้



รูปที่ 6 โมเดลที่ถอดเสียงออกมา

จากรูปรูปที่ 6 โมเดลสร้างขึ้นมาจาก Model Whisper Small ที่เขียนแบบ python เป็นโมเดลที่ถอดเสียงออกมาเป็นข้อความภาษาไทย โดยจากรูปรูปที่ 6 จะเห็นได้ว่าตัวโมเดลถอดข้อความเสียงออกมาได้อย่างถูกต้อง ครบถ้วน 100% ไม่มีข้อผิดพลาดจากผู้พูด speaker_001



รูปที่ 7 โมเดลที่ถอดเสียงออกมา

จากรูปรูปที่ 7 จากโมเดลถอดเสียงออกมาเป็นข้อความจะเห็นได้ว่ายังมีข้อผิดพลาดอยู่ ไม่ถูกต้อง เกิดข้อผิดพลาดจากผู้พูด speaker_003 เนื่องจากอาจจะเป็นคำที่มีการสะกดซับซ้อน โมเดลเลยไม่สามารถอ่านได้ถูกต้อง

5. สรุปและข้อเสนอแนะ

5.1 สรุปผลการทดลอง

จากการวิจัยครั้งนี้ผลลัพธ์ที่ได้ คือ

1) speaker_001 มีเปอร์เซ็นต์ความถูกต้องมากที่สุด ผลของการถอดเสียงถอดออกมาได้ผลลัพธ์ที่ถูกต้อง 100% ทั้งหมด 4ประโยค และมีเพียงประโยคเดียวที่ถอดความออกมาผิดเพียงแค่ 1 คำ

2) speaker_002 มีเปอร์เซ็นต์ความถูกต้องระดับกลาง ผลของการถอดเสียงออกมาได้ผลลัพธ์ที่ถูกต้อง 100% ทั้งหมด 3 ประโยค มี 1 ประโยคที่ถอดเสียง ออกมาผิด 1 คำ และมี 1 ประโยคที่ถอดเสียงออกมาผิด 2 คำ

3) speaker_003 มีเปอร์เซ็นต์ความถูกต้องน้อยที่สุด ผลของการถอดเสียงออกมาได้ผลลัพธ์ที่ถูกต้อง 100% ทั้งหมด 1 ประโยค มี 3 ประโยคที่ถอดเสียงออกมาผิด 1 คำ และมี 1 ประโยคที่ถอดเสียงออกมาผิด 2 คำ

จากข้อมูลทั้งหมดนั้นจะเห็นว่า เสียงของทั้ง 3 คน จะออกเสียงไม่ถูกต้องในคำ ๆ เดียวกัน ซึ่งหมายความว่าค่าที่ออกมานั้น มีความคล้ายคลึงกันมาก เปอร์เซ็นต์ความถูกต้องสูง

5.2 ผลสรุปจากการใช้ Model Whisper Small

Whisper นั้นมีความแตกต่างมาก ๆ กว่าระบบองค์กรพัฒนาเสียงตัวอื่น ๆ เพราะมันเป็น AI ที่เป็นแบบโอเพ่นเซอร์ที่มีการได้เรียนรู้ในภาษาต่าง ๆ มากมายสามารถเรียนรู้และจดจำทั้งสำเนียงของแต่ละบุคคล เสียงแบคกราวน์และคำศัพท์ยาก ๆ ที่เขียนยากมีตัวสะกดหลายตัวที่มีความซับซ้อนอย่างภาษาไทยได้ถูกต้องประมาณหนึ่งได้ดีและแม่นยำกว่า AI นั่นจึงเป็นเหตุผลหลักที่ผู้วิจัยเลือกโมเดลตัวนี้มาใช้กับงานวิจัย

ถึงแม้ว่า โมเดล Whisper ตัวนี้จะมีการเรียนรู้ที่ค่อนข้างเยอะและมีผลลัพธ์ที่แม่นยำพอสมควรแต่โมเดลก็ยังไม่สามารถที่จะถอดทุกภาษาที่มีอยู่ในโลกได้และยังมีข้อจำกัดอีกอย่างหนึ่งเลยก็ คือ โมเดลไม่สามารถถอดภาษาแบบแสดงผลทันที (Real time) ได้ จึงมีตัวเลือกในการใช้งานที่น้อยลงมาก

5.3 ประสิทธิภาพและข้อจำกัดของงานวิจัย

เนื่องจากตัวโมเดล ถูกเทรนขึ้นมาแบบไม่ได้มีประสิทธิภาพ 100% โดยที่มีการใช้สัญญาณข้อมูลขนาดใหญ่มาก อาจมีปัจจัยหลายปัจจัยที่มีผลต่อเสียง เช่น อุปกรณ์เสียงที่ใช้ การพูดของชาวต่างชาติและสภาพแวดล้อมซึ่งอาจทำให้มีความแตกต่างระหว่างเสียงที่ถูกสร้างขึ้นโดยโมเดลกับเสียงจริงได้บ้าง การออกเสียงออกไปอาจจะทำให้โมเดลรับเสียงบางเสียงได้อย่างไม่ชัดเจนและแม่นยำพอ ผู้วิจัยคิดว่าที่โมเดลถอดความออกมาได้ผิดพลาดบางคำนั้นเป็น เพราะโมเดลได้ยินหรือมีการเรียนรู้คำศัพท์ที่ไม่เพียงพอ โมเดลจึงมีการพยายามคาดเดาคำศัพท์ที่ได้ยินให้ใกล้เคียงกับคำที่พูดจริง ๆ ไปด้วยตัวมันเอง นั่นจึงเป็นเหตุผลที่โมเดล ยังไม่มีประสิทธิภาพแม่นยำที่เต็ม 100% และโมเดลนั้นยังมีความแม่นยำที่ต่ำมาก ๆ ในภาษาที่มีข้อมูลต่ำอีกด้วย เลยทำให้อัตราการทำงานมีความผิดพลาดที่ค่อนข้างสูงกว่าปกติ เพราะ มีการเรียนรู้ข้อมูลที่น้อย ส่วนในด้านของประสิทธิภาพในการทำงานด้านอื่น ๆ นั้น ก็ยังมีอีกหลายปัจจัยที่ส่งผลในเรื่องของความแม่นยำในการทำงานของตัวโมเดล เนื่องจากในขณะที่ใช้โมเดลในการบันทึกเสียงนั้นอาจจะมีเสียงสภาพแวดล้อมภายนอกเล็ดลอดเข้าไประหว่างการบันทึกผลจึงส่งผลให้การวัดผลของค่ามีความผิดพลาดและเกิดความไม่แม่นยำได้

5.4 ผลสรุปโดยรวม

สำหรับชาวต่างชาติทั้งหมดที่ผู้วิจัยได้รวบรวมข้อมูลมาและนำมาถอดเสียงข้อความใน Model Whisper ทั้งหมดนั้น ทุกคนมีสำเนียง เสียง และคำศัพท์ที่ค่อนข้างดีมาก ผลของการถอดความเสียงส่วนมากมีความแม่นยำในทางภาษาค่อนข้างสูงมาก

ถึงแม้จะไม่เป็นสำเนียงของคนไทยที่ชัดเจนนัก แต่สามารถเข้าใจได้อย่างมีประสิทธิภาพสูง และตัวโมเดลที่นำมาทดสอบเข้าใจได้เกือบถึง 100% โดยสรุป คือ คนต่างชาติที่ต้องการเรียนรู้ภาษาไทยนั้น มีเสียงสำเนียง ที่ชัดเจน เข้าใจได้ดีมากและมีความคล้ายคลึงกับคนไทยจริง ๆ มาก

5.5 ข้อเสนอแนะ

ข้อเสนอแนะในการที่จะไปพัฒนาระบบต่อไป

1) ทดลองใช้ Whisper Small ซึ่งเป็นเว็บถอดความที่ดีและเหมาะสม

2) ทำการเก็บข้อมูลเสียงใหม่ ๆ ให้เยอะและหลากหลาย

6. เอกสารอ้างอิง

[1] ทรูปลูกปัญญา, "ความเป็นมาของภาษาไทย," [ออนไลน์]. Available: https://www.dol.go.th/secretary/Pages/ความเป็นมาของภาษาไทย.aspx. [เข้าถึงเมื่อ: 30 สิงหาคม 2566].

[2] Newtech Insulation, "วิเคราะห์คลื่นความถี่เสียง," [ออนไลน์]. Available: https://noisecontrol365.com/service/Detail/0-6วิเคราะห์คลื่นความถี่เสียง. [เข้าถึงเมื่อ: 17 เมษายน 2567].

[3] The Editors of Encyclopaedia Britannica, "Wavelength," [online]. Available: https://www.britannica.com/science/wavelength. [Accessed: 30 August 2023].

[4] Karthik, "Amplitude of a Wave," [online]. Available: https://byjus.com/physics/amplitude-frequency-period-sound. [Accessed: 30 August 2023].

[5] "What is a Spectrogram?" [online]. Available: https://pnsn.org/spectrograms/what-is-a-spectrogram. [Accessed: 30 August 2023].

[6] Christian Auer and Antonia Korger, "Sound Intensity," [online]. Available: https://www.auersignal.com/en/technical-information/audible-signalling-equipment/sound-intensity. [Accessed: 30 August 2023].

[7] Wikipedia, "ระดับเสียง (pitch)," [ออนไลน์]. Available: https://th.m.wikipedia.org/wiki/ระดับเสียง_(ดนตรี). [เข้าถึงเมื่อ: 17 เมษายน 2567].

[8] ปุณยวัจน์ วรรณคาม, "ผลของการใช้วิธีสอนโฟนิกส์ในการพัฒนาความสามารถในการสื่อสารภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 4," ปริญญาการศึกษามหาบัณฑิต การค้นคว้าอิสระ, มหาวิทยาลัยนเรศวร, พิษณุโลก, 2564.

[9] Ratmir Belov, "การถอดความ – การแปลงคำพูดเป็นข้อความ," [ออนไลน์]. Available: https://pakhotin.org/th/career/transcription. [เข้าถึงเมื่อ: 31 สิงหาคม 2566].

[10] Looloo Technology, "Thonburian Whisper: โมเดลถอดความจากเสียงพูดภาษาไทย," [ออนไลน์]. Available: https://www.borntodev.com/2022/12/30/thonburian-whisper. [เข้าถึงเมื่อ: 31 สิงหาคม 2566].

[11] Lew, "OpenAI แจกโมเดลปัญญาประดิษฐ์แปลงเสียงเป็นข้อความพร้อมแปลเป็นอังกฤษ รองรับภาษาไทยและเกาหลีด้วย," [ออนไลน์]. Available:

https://www.blognone.com/node/13
0549. [เข้าถึงเมื่อ: 25 ธันวาคม 2566].