

# การศึกษาเทคนิคการสร้างเสียงภาษาไทยจากปัญญาประดิษฐ์เพื่อใช้ผลิตสื่อมัลติมีเดีย

## A Study of Techniques to Generate Thai Language from Artificial Intelligence (AI) Voice for Producing Multimedia.

กุลกานต์ สุทธิดารา<sup>1\*</sup> อลิษา แสงวิมาน<sup>1</sup> และอาทิตยา บินฮาซัน<sup>1</sup>  
Kullakan Suthidara<sup>1\*</sup>, Alisa Sangwiman<sup>1</sup> and Atitaya Binhasun<sup>1</sup>

### บทคัดย่อ

ปัจจุบันปัญญาประดิษฐ์หรือ AI (Artificial Intelligence) เข้ามามีบทบาทสำคัญในทุกภาคส่วน ซึ่งเทคโนโลยีการสร้างเสียงจากปัญญาประดิษฐ์ (Generate AI Voice) เป็นศาสตร์หนึ่งของ AI ที่ถูกนำมาใช้ประโยชน์ในการผลิตเสียงสังเคราะห์เลียนแบบเสียงพูดจริงของมนุษย์ เพื่อช่วยอำนวยความสะดวกในชีวิตประจำวันและการทำงาน การสร้างเสียงจากปัญญาประดิษฐ์นั้นอาศัยเทคโนโลยี 2 ด้านหลัก คือ (1) การประมวลผลภาษาธรรมชาติ (NLP: Natural language processing) ร่วมกับ (2) เทคโนโลยีสังเคราะห์เสียงจากข้อความ (TTS: Text-to-Speech Synthesis) เพื่อแปลงข้อความจากตัวหนังสือให้เป็นเสียงพูดในภาษาต่าง ๆ ได้ โดยงานเทคโนโลยีการเรียนรู้ คณะศิลปศาสตร์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรีได้ใช้เสียงภาษาอังกฤษจาก AI ในงานผลิตสื่อมัลติมีเดีย ซึ่งสามารถใช้แทนการอัดเสียงจากห้องสตูดิโอได้เป็นอย่างดี แต่สำหรับการสร้างเสียงภาษาไทยจาก AI นั้นยังมีปัญหาเรื่องการอ่านออกเสียงที่ไม่ถูกต้องและไม่เป็นธรรมชาติอยู่ เนื่องจากโครงสร้างทางภาษาไทยมีความซับซ้อนและไม่มีการเว้นช่องว่างระหว่างคำเหมือนภาษาอังกฤษ งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาเทคนิคการป้อนข้อความภาษาไทยในซอฟต์แวร์การสร้างเสียงจากปัญญาประดิษฐ์ (Generate AI Voice) ให้ได้เสียงพูดสังเคราะห์ที่แม่นยำและเป็นธรรมชาติ โดยมีกระบวนการ ดังนี้ 1) ศึกษารายละเอียดข้อมูลและวิเคราะห์ซอฟต์แวร์ที่ให้บริการแปลงข้อความให้เป็นเสียงพูดเพื่อเลือกซอฟต์แวร์ที่เหมาะสมกับการใช้งาน และเนื้อหาที่ต้องการผลิตสื่อ 2) ทดลองป้อนเนื้อหาในซอฟต์แวร์ที่เลือกเพื่อเก็บรวบรวมข้อมูลด้านความถูกต้องของการออกเสียง และด้านความเป็นธรรมชาติของเสียง (จังหวะ ความเร็ว และการเว้นวรรค) ผลการศึกษาพบว่า จากการทดลองใช้เทคนิค ความถูกต้องของการออกเสียงและความเป็นธรรมชาติของเสียงดีขึ้นตามลำดับ ในการทดลองครั้งที่ 1 ความถูกต้องของการออกเสียงคิดเป็นร้อยละ 92 จนถึงครั้งที่ 30 คิดเป็นร้อยละ 100 และความเป็นธรรมชาติของเสียงปรับจากเสียงที่อ่านเร็วเกินไป และขาดจังหวะหายใจ มาเป็นเสียงที่จังหวะพอดีเข้ากับสื่อที่ผลิต ดังนั้นขั้นตอนการเตรียมพร้อมของข้อความก่อนที่จะป้อนลงซอฟต์แวร์เป็นสิ่งสำคัญ เช่น การเว้นวรรคคำใหม่ การใช้สัญลักษณ์พิเศษ การถอดอักษรเป็นคาราโอเกะสำหรับคำที่มีความซับซ้อนในการอ่าน โดยสามารถสรุปเป็นเทคนิคที่สำคัญในการการสร้างเสียงจาก AI เพื่อใช้ผลิตสื่อมัลติมีเดียได้ 8 ข้อ

**คำสำคัญ:** การสร้างเสียงจากปัญญาประดิษฐ์ การผลิตสื่อมัลติมีเดีย ข้อความภาษาไทย

### Abstract

In the present, artificial intelligence technology or AI plays an important role in all sectors. The AI voice technology is one of the concepts of AI utilized to synthesize human voices to help facilitate daily life and work. Artificial intelligence speech generators are produced from two technologies: (1) NLP: Natural language processing and (2) TTS: Text-to-Speech Synthesis technology to convert from text to speech in different languages. Education Technology Unit at School of Liberal Arts, KMUTT uses AI-powered English voices in multimedia production which can also be used instead of recording from the studio. For creating Thai voices from AI, there

<sup>1</sup> คณะศิลปศาสตร์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี กรุงเทพมหานคร 10140

<sup>1</sup> School of Liberal Arts, King Mongkut's University of Technology Thonburi, Bangkok, 10140

\*Corresponding author: e-mail: kullakan.sut@kmutt.ac.th

Received: April 19, 2023, Accepted: October 23, 2023, Published: July 15, 2024



are still problems with correctness and unnatural pronunciation. This is because the structure of Thai language is complex, and there are no gaps like in English words. The objective of this research was to study techniques for entering Thai text in generating AI Voice software to obtain an accurate and a natural synthetic speech for multimedia production. The processes were as follows: 1) studying the data and analyze the software providing text to speech conversion for selecting the software suitable to use, and 2) experimenting with entering content in selected software to collect information on pronunciation accuracy and the naturalness of the sound (rhythm, speed, and spacing). The results of experimenting with techniques showed that the accuracy of pronunciation and the naturalness of the voice have improved accordingly. In the first trial, the pronunciation accuracy was found at 92 percent, and in the thirtieth trial, it was found at 100 percent. The naturalness of the tone was adjusted from too fast and being lack of breathing rhythm to fit the rhythm of the produced media. Therefore, it is important to prepare the texts before entering them in the software, such as new word spacing, special symbols, complicated karaoke transcription for wordings. It can be summarized as eight important techniques for creating sound from AI for using in multimedia production.

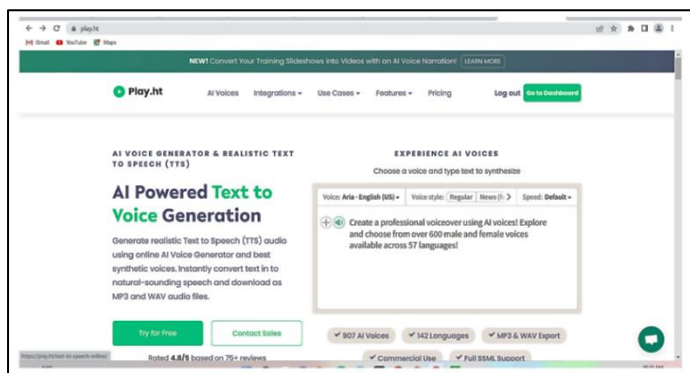
**Keywords:** generate artificial intelligence voice, media production, Thai text

## บทนำ

ปัจจุบันเทคโนโลยีปัญญาประดิษฐ์ (AI: Artificial Intelligence) เข้ามามีบทบาทสำคัญในทุกภาคส่วน โดยมีการใช้ประโยชน์จาก AI ในการสร้างเสียง (AI Voice) แทนเสียงจริงของมนุษย์ เพื่อช่วยอำนวยความสะดวกในชีวิตประจำวันและการทำงาน เช่น การตอบกลับการสนทนาแบบอัตโนมัติหรือแชทบอทฝ่ายบริการลูกค้า หนังสือเสียง การอ่านเอกสารหรือ e-book การทำเสียงพากย์ภาพยนตร์ รวมถึงการสร้างเสียงในสื่อมัลติมีเดียต่าง ๆ

AI Voice คือ เสียงสังเคราะห์ที่เลียนแบบเสียงพูดจริงของมนุษย์ ผ่านกระบวนการบ่อนข้อความตัวหนังสือลงในคอมพิวเตอร์ซอฟต์แวร์หรือแอปพลิเคชัน ข้อความที่บ่อนไปนั้นจะถูกประมวลผลและแปลงออกมาเป็นเสียงพูด โดยสามารถดาวน์โหลดไฟล์เสียงที่ได้นั้นไปประยุกต์ใช้กับงานต่าง ๆ ได้ และด้วยเทคโนโลยีที่ทันสมัยขึ้น จึงมีซอฟต์แวร์หรือแอปพลิเคชันที่ให้บริการสร้างเสียงสังเคราะห์จาก AI ออกมาเป็นจำนวนมากทั้งรูปแบบเว็บไซต์และรูปแบบที่ต้องติดตั้งลงเครื่อง มีการออกแบบให้ใช้งานได้ง่าย มีโทนเสียงให้เลือกหลากหลายตามความต้องการใช้งาน โดยสามารถเลือกเสียงตามเพศ อายุ เชื้อชาติ มีฟังก์ชันให้เพิ่มหรือลดเอฟเฟกต์ ปรับแต่งเสียงแหลมสูงหรือทุ้ม ปรับความเร็ว และระดับเสียงเพื่อให้ได้เสียงที่มีคุณภาพใกล้เคียงเสียงธรรมชาติที่สุด (Voicemod, 2023) ซึ่งหลักการทำงานพื้นฐานในการสร้างเสียงจาก AI ของแต่ละซอฟต์แวร์หรือแอปพลิเคชันสามารถแบ่งการทำงานได้เป็น 2 ส่วน ดังนี้

1. ส่วน font-end (ส่วนที่ทุกคนสามารถมองเห็นได้ของเว็บไซต์หรือแอปพลิเคชัน)



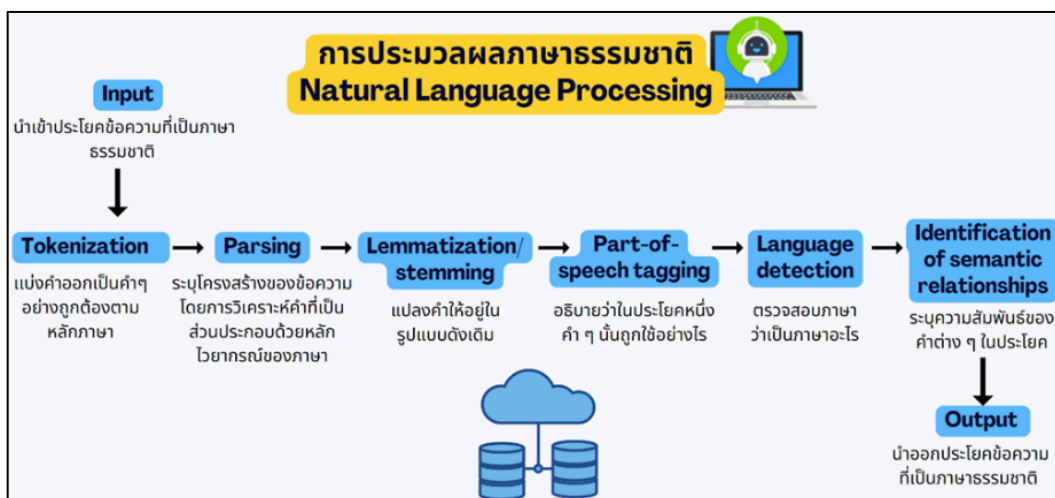
ภาพที่ 1 ตัวอย่าง Website interface

จากภาพที่ 1 ตัวอย่าง ส่วน font-end หรือหน้า website interface เช่น Play.ht สามารถใช้งานแบบออนไลน์ได้ โดยลำดับแรกต้องลงทะเบียนเพื่อสร้างบัญชีผู้ใช้งานก่อน จากนั้นเลือกโทนเสียงที่ต้องการ และป้อนข้อความเป็นประโยคลงในกล่องข้อความ สามารถปรับแต่งคุณภาพเสียงตามต้องการจากเมนูฟังก์ชันต่าง ๆ ในส่วนของ font-end นี้ จากนั้นซอฟต์แวร์จะประมวลผลข้อความด้วยระบบ AI ออกมาเป็นเสียงพูด ขั้นตอนสุดท้าย คือ การดาวน์โหลดไฟล์เสียงไปใช้งาน

## 2. ส่วน back-end (ส่วนของการทำงานเบื้องหลัง)

ส่วน back- end หรือการประมวลผลนั้น อาศัยเทคโนโลยี 2 ด้านหลัก คือ (2.1) การประมวลผลภาษาธรรมชาติ (NLP: Natural language processing) ร่วมกับ (2.2) เทคโนโลยีสังเคราะห์เสียงจากข้อความ (TTS: Text-to-Speech Synthesis) เพื่อแปลงข้อความจากตัวหนังสือให้เป็นเสียงพูดในภาษาต่าง ๆ ได้ (<https://th.wikipedia.org>) โดยมีรายละเอียดดังนี้

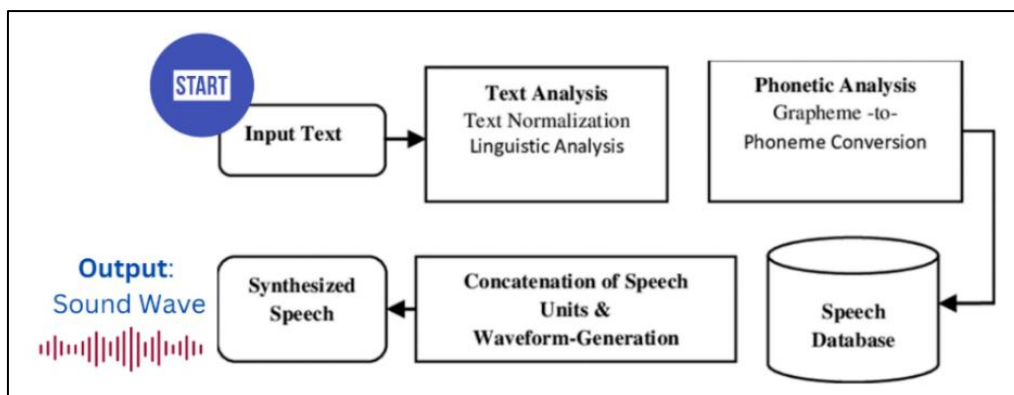
(2.1) เทคโนโลยีด้านการประมวลผลภาษาธรรมชาติ (NLP: Natural language processing) คือ การประมวลผลภาษาธรรมชาติหรือภาษามนุษย์ ที่ทำให้คอมพิวเตอร์สามารถเข้าใจภาษามนุษย์ด้วย เป็นวิธีการวิเคราะห์ข้อความด้วยการรวบรวมความรู้ว่ามนุษย์เข้าใจและใช้ภาษาอย่างไร (ชาภิยะห์, 2565) เพื่อพัฒนาเครื่องมือและเทคนิคที่เหมาะสม ซึ่งจะทำให้ระบบคอมพิวเตอร์เข้าใจและจัดการภาษาธรรมชาติเพื่อทำงานต่าง ๆ ที่ต้องการได้ (Joseph, et al., 2016) โดยกระบวนการทำงานต้องใช้ความรู้หลายศาสตร์บูรณาการร่วมกัน เช่น คณิตศาสตร์ ภาษาศาสตร์ จิตวิทยา และประกอบด้วยหลากหลายวิธีการประมวลผลและแปลความหมายของภาษาปกติของมนุษย์ หลักการทำงานขั้นพื้นฐานของ NLP มีดังภาพที่ 2



ภาพที่ 2 การประมวลผลภาษาธรรมชาติ (NLP) (Perkins, 2010; SAS Global Forum, 2020)

จากภาพประกอบที่ 2 หลักการทำงานขั้นพื้นฐานของ NLP เริ่มจากการนำเข้าประโยคข้อความและแบ่งประโยคออกเป็นคำหรือหน่วยย่อยอย่างถูกต้องตามหลักภาษา (เช่น การแยกประธาน กริยา กรรม ของประโยค) อีกนัยหนึ่งคือ NLP ทำหน้าที่ย่อข้อความในภาษาที่มนุษย์ใช้ในการสื่อสารลงเป็นส่วนย่อย ๆ เพื่อง่ายต่อการทำความเข้าใจของเครื่อง จากนั้นทำการวิเคราะห์ และประมวลผลความสัมพันธ์ (กลุ่มภารกิจบริการและพัฒนาสื่อการเรียนการสอน สำนักคอมพิวเตอร์, 2564) ในการวิเคราะห์นั้นจะมีการวิเคราะห์ในเชิงโครงสร้าง (Syntactic Analysis) การวิเคราะห์ในเชิงความหมาย (Semantic Analysis) การวิเคราะห์ในเชิงตีความ (Pragmatic Analysis) ซึ่งมีการทำงานร่วมกับฐานข้อมูลคำศัพท์ที่มีอยู่ด้วย (Dilmegani, 2022) จากนั้นจะประเมินค่าเพื่อหาคำตอบ นำออกเป็นประโยคข้อความที่เป็นภาษาธรรมชาติของมนุษย์ต่อไป

(2.2) เทคโนโลยีสังเคราะห์เสียงจากข้อความ (Text-to-Speech Synthesis: TTS) มีระบบประมวลผลข้อความเป็นคำพูด (ภาพที่ 3)



ภาพที่ 3 ระบบประมวลผลข้อความเป็นคำพูด TTS System (Mache and Mahender, 2016)

จากภาพที่ 3 ระบบการสร้างเสียงพูดนั้นเริ่มต้นจากการป้อนข้อความตัวหนังสือบนหน้าเว็บไซต์ จากนั้นซอฟต์แวร์จะจัดระเบียบโครงสร้างและวิเคราะห์ทางภาษาโดยมีการแยกประโยคให้เป็นหน่วยย่อย (unit) อักษรและอักขระจะถูกประมวลผลและวิเคราะห์ ซึ่งซอฟต์แวร์จะวิเคราะห์การออกเสียงของแต่ละหน่วยเสียง แล้วไปค้นหารูปแบบคลื่นเสียงของคำนั้นที่มีอยู่ในฐานข้อมูล (หนึ่งหน่วยเสียงจะมีรูปแบบคลื่นของแต่ละหน่วย) และได้รูปแบบคลื่นเสียงของหน่วยนั้นมา จากนั้นนำรูปแบบคลื่นเสียงของแต่ละหน่วยมาเชื่อมต่อกันเพื่อสร้างคำและประโยค และสร้างรูปแบบคลื่นทั้งหมดจนได้เป็นเสียงคำพูดสังเคราะห์

งานเทคโนโลยีการเรียนรู้ (EdTech) คณะศิลปศาสตร์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรีได้นำเทคโนโลยีการสร้างเสียงจากปัญญาประดิษฐ์ (Generate AI Voice) มาใช้ในขั้นตอนการผลิตสื่อการเรียนการสอน (ในขั้นตอนที่ 4-5 ของภาพประกอบที่ 4) แทนการใช้ห้องอัดเสียงในห้องสตูดิโอสำหรับบางชิ้นงาน



ภาพที่ 4 ขั้นตอนการผลิตสื่อการเรียนการสอน

เนื่องจากการปิดสถานที่ตามมาตรการป้องกันการแพร่ระบาดของโรคติดเชื้อไวรัสโคโรนา 2019 (COVID-19) รวมถึงการเรียนการสอนในยุคความปกติใหม่ (New Normal) นั้น รูปแบบการเรียนการสอนได้เปลี่ยนไป อาจารย์ผู้สอนมีความจำเป็นต้องใช้สื่อจำนวนมากเพื่อจัดการเรียนการสอนออนไลน์ ซึ่งงานเทคโนโลยีการเรียนรู้พบว่า เสียงจากการสังเคราะห์จาก AI นั้นใช้งานได้ดี ช่วยลดปัญหาการแทรกเข้ามาของเสียงรบกวน (Noise) ระหว่างการอัด ความดังของเสียงมีความสม่ำเสมอได้เสียงคมชัด ลดขั้นตอนการตัดเสียงรบกวนและขั้นตอนการเพิ่มลดระดับเสียงได้ดี และช่วยพากย์เสียงแทนเสียงจริงของอาจารย์ผู้สอนได้ ได้ไฟล์เสียง MP3 ที่มีขนาด

เล็ก โดยขนาดไฟล์ที่ได้จากการอัดพากย์เสียง มีขนาด 1013 KB และไฟล์ที่ได้จากการสังเคราะห์เสียง AI มีขนาด 210 KB โดยเปรียบเทียบจากการใช้เนื้อหาเดียวกัน ทั้งยังมีคุณภาพสูง เหมาะสำหรับนำไปผลิตสื่อมัลติมีเดีย ทำให้ประหยัดพื้นที่ (Storage) การจัดเก็บ เพราะพื้นที่จัดเก็บเป็นปัญหาสำคัญของงานมัลติมีเดีย

ซึ่งงานเทคโนโลยีการเรียนรู้มีการผลิตสื่อด้วยเสียงจากการสังเคราะห์จาก AI แล้ว 127 ชิ้นงาน ตามลิงก์ QR Code ตัวอย่างงานสื่อการเรียนการสอนออนไลน์ (ภาพประกอบที่ 5)



ภาพที่ 5 ตัวอย่างงานสื่อการเรียนการสอนออนไลน์

แต่อย่างไรก็ตามการใช้เสียงจาก AI ขณะนี้ยังมีข้อจำกัดอยู่มากในการใช้งาน ซึ่งซอฟต์แวร์ที่มีอยู่หลายแหล่งหรือหลายเว็บไซต์นั้นรองรับการสังเคราะห์เสียงภาษาอังกฤษได้ดีกว่าภาษาไทย เนื่องจากปัจจัยหลายประการ เช่น ภาษาอังกฤษมีหน่วยเสียงของคำที่อยู่ในฐานข้อมูลที่สามารถเทียบเคียงรูปแบบคลื่นของแต่ละหน่วยได้มากกว่า และด้วยปัจจัยของโครงสร้างทางภาษาไทยที่ซับซ้อนกว่าภาษาอังกฤษ ทั้งสระ วรรณยุกต์ คำเดียวอ่านได้หลายรูปแบบ รวมถึงมีรูปแบบประโยคที่เขียนติดกันหมด ทำให้ความหมายสามารถเปลี่ยนแปลงได้ เช่น ประโยคที่ว่า “คนขับรถไปอุดร” มีความหมายทั้ง “คน-ขับรถ-ไปอุดร” หรือ “คนขับรถ-ไปอุดร” เป็นต้น (Jazz.Ordinaryday, 2020) ภาษาไทยเป็นภาษาที่ไม่ได้แบ่งกลุ่มโดยธรรมชาติ ซึ่งไม่มีช่องว่างระหว่าง 2 คำ ในขณะที่ภาษาอังกฤษมีช่องว่าง และมหัพภาคหรือจุด (.) ช่วยในการเว้นวรรคประโยค อีกทั้งประโยคภาษาอังกฤษมีเครื่องหมายเป็นสัญลักษณ์ เช่น เครื่องหมายปรัศนี (?) ใช้เมื่อสิ้นสุดประโยคคำถาม ทำให้ซอฟต์แวร์มีการปรับคลื่นเสียงตามสัญลักษณ์ด้วย จึงเพิ่มความเป็นธรรมชาติของเสียง ดังนั้นการสังเคราะห์เสียงพูดภาษาไทยให้แม่นยำและเป็นธรรมชาติ ยังเป็นหนึ่งในประเด็นหลักที่สำคัญในการใช้เทคโนโลยีการสร้างเสียงจากปัญญาประดิษฐ์

งานเทคโนโลยีการเรียนรู้ จึงมีความจำเป็นต้องศึกษาหาเทคนิคการป้อนข้อความภาษาไทย ให้ซอฟต์แวร์ได้ประมวลผลและสังเคราะห์เสียงพูดได้แม่นยำและเป็นธรรมชาติที่สุด โดยทดลองป้อนข้อความที่ต้องการผลิตสื่อการเรียนการสอนในซอฟต์แวร์ ซึ่งมีหลายเว็บไซต์ให้บริการแปลงข้อความทั้งภาษาไทยและภาษาอังกฤษให้เป็นเสียงพูด มีทั้งแบบใช้ฟรีและแบบเสียค่าใช้จ่าย ซึ่งทั้ง 2 แบบจะมีข้อแตกต่างในเรื่องของการจำกัดจำนวนอักขระ ประเภทของโทนเสียง การรองรับภาษาที่หลากหลาย การเข้าใช้งานทั้งแบบออนไลน์และติดตั้งลงบนคอมพิวเตอร์ นามสกุลไฟล์ และการดาวน์โหลดไฟล์เสียงเพื่อนำไปใช้งาน

### วัตถุประสงค์การวิจัย

เพื่อศึกษาเทคนิคการป้อนข้อความภาษาไทยในซอฟต์แวร์การสร้างเสียงจากปัญญาประดิษฐ์ให้ได้เสียงพูดสังเคราะห์ที่แม่นยำและเป็นธรรมชาติสำหรับผลิตสื่อมัลติมีเดีย



## ระเบียบวิธีวิจัย

การวิจัยครั้งนี้เป็นการวิจัยเชิงทดลองเพื่อนำไปสู่บทสรุปเชิงเทคนิคของการสร้างเสียงจากปัญญาประดิษฐ์สำหรับใช้ผลิตสื่อมัลติมีเดีย และลดขีดจำกัดและอุปสรรคของเสียงจากปัญญาประดิษฐ์ โดยได้ออกแบบขั้นตอนของการศึกษาวิจัยไว้ดังนี้

### 1. การวิเคราะห์และคัดเลือกซอฟต์แวร์

ในการศึกษารายละเอียดข้อมูลของซอฟต์แวร์ต่าง ๆ ที่ให้บริการแปลงข้อความให้เป็นเสียงพูดนั้น มีการวิเคราะห์เปรียบเทียบลักษณะเฉพาะ (Characteristics) ของแต่ละซอฟต์แวร์ในมิติต่าง ๆ เพื่อใช้เป็นแนวทางในการเลือกซอฟต์แวร์ที่เหมาะสมกับการใช้งานและเนื้อหาที่ต้องการผลิตสื่อ ได้แก่ จุดเด่นของซอฟต์แวร์ ความหลากหลายของโทนเสียง (เพศ อายุ สำเนียงภาษา) ความเป็นธรรมชาติของเสียง ระดับความเร็วของเสียง การรองรับเสียงภาษาไทย ความสะดวกของการติดตั้ง ฟังก์ชันการใช้งาน และการออกแบบให้ใช้งานได้ง่าย (user-friendly) เป็นต้น

### 2. การทดลองป้อนเนื้อหาในซอฟต์แวร์ที่เลือก

ในขั้นตอนการทดลองใช้งานซอฟต์แวร์ที่เลือกได้แล้วนั้น ทีมผู้วิจัยได้ศึกษาเทคนิคการป้อนข้อความภาษาไทย โดยใช้ประโยชน์จากเนื้อหาที่ใช้ผลิตสื่อการเรียนการสอน ช่วงเดือนมกราคม – เดือนพฤษภาคม 2565 ซึ่งคัดเลือกสื่อภาษาไทยแบบสุ่มจำนวน 10 สื่อ และคัดเลือกคำของสื่อแบบจำเพาะ เน้นเลือกคำที่มีความซับซ้อนอ่านออกเสียงได้ เช่น คำที่มี รห้น (รร) คำที่มีหลายพยางค์ อักษรย่อ เพื่อหาเทคนิคให้ซอฟต์แวร์สามารถอ่านได้อย่างถูกต้องแม่นยำ และมีความเป็นธรรมชาติ (สามารถปรับความเร็วและความถี่ของเสียงรวมถึงจังหวะการพูดจากการตั้งค่าในซอฟต์แวร์ โดยในครั้งที่ 1 ได้ใส่ข้อความทั้งหมด 100 คำลงในซอฟต์แวร์ (ข้อจำกัดของซอฟต์แวร์ในการป้อนจำนวนคำ คือ หากใส่คำเกิน 100 คำจะพบปัญหาระยะเวลาการประมวลผลและการดาวน์โหลดไฟล์เป็นเวลานาน) และไม่มีตั้งค่าใด ๆ ใช้ค่า default ที่มากับ ซอฟต์แวร์ และมีการปรับเทคนิคการป้อนคำรวมถึงการตั้งค่าในแต่ละครั้งเพื่อให้ได้คำที่ดีที่สุด โดยมีเนื้อหาในการทดลอง (ตามภาพที่ 6)

นักศึกษสามารถประยุกต์ใช้ผลการเปลี่ยนแปลงเจตจำนงของรูปแบบในสิ่งต่าง ๆ สภาพบรรยากาศเปลี่ยนแปลงจากสภาพเดิม คริสต์ศตวรรษอย่างมีนัยสำคัญ โดยไม่มีกฎเพราะการมีอยู่ของสิ่งมีชีวิตในธรรมชาติ หลักฐานการเรียนรู้ได้และนักศึกษสามารถส่งผลงานเพื่อรอบอกจากอาจารย์ที่ปรึกษาได้ เอกลักษณ์ หรือ ลักษณะที่เป็นหนึ่งเดียวของ สิ่งใด ๆ ในกลุ่มประเภทเดียวกัน โดยระบุได้จากกุศโลบาย การเปรียบเทียบ องค์ประกอบโครงสร้างภายนอก ระบบภายใน เรื่องนอกกรอบ ก็หมายถึงว่าเราเห็น เอกลักษณ์ของสิ่งนั้นจากการสร้างลักษณะเด่นหรือเอกลักษณ์

### ภาพที่ 6 ข้อความเพื่อใช้ในการทดลอง

ทั้งนี้ ทีมผู้วิจัยทำการบันทึกผลการทดลองการออกเสียงของซอฟต์แวร์ โดยมีเกณฑ์การเก็บข้อมูล 2 ด้านคือ

#### 2.1 ด้านความถูกต้องของการออกเสียง

ทีมผู้วิจัยป้อนข้อความที่จัดเตรียมไว้ในซอฟต์แวร์ที่เลือกไว้ให้อ่านแล้วประเมินความถูกต้องของเสียงที่ออกพร้อมบันทึกคำและจำนวนคำที่ออกเสียงถูกต้องโดยวิเคราะห์เป็นคำร้อยละของจำนวนคำทั้งหมดเป็นรอบแรก จากนั้นทีมผู้วิจัยได้ทดลองในรอบถัดไปโดยปรับการป้อนคำด้วยเทคนิคต่าง ๆ เช่น เทคนิคสะกดคำ การใส่สัญลักษณ์พิเศษ แล้วประเมินความถูกต้องของการออกเสียงในแต่ละรอบการปรับ เพื่อให้ได้เทคนิคการป้อนคำที่ออกเสียงถูกต้องมากที่สุด

#### 2.2 ด้านความเป็นธรรมชาติของเสียง (จังหวะ ความเร็ว และการเว้นวรรค)

ทีมผู้วิจัยได้ทดลองความเป็นธรรมชาติของเสียงโดยปรับค่าต่าง ๆ ของซอฟต์แวร์และประเมินความเป็นธรรมชาติของเสียงในแต่ละรอบของการปรับค่า เพื่อหาตัวเลขที่เป็นค่ามาตรฐาน (standard) ในการตั้งค่าซอฟต์แวร์ (ตามภาพประกอบที่ 7) เพื่อสรุปเป็นเทคนิคการใช้งาน ซึ่งประกอบด้วย

- การ Pause หรือ ค่าจังหวะการหยุด คือ การเว้นจังหวะช่วงที่ต้องการจะเว้นวรรคในการหยุดเสียง

- ความถี่เสียง (KHz) คือ เป็นการเน้นเสียงและปรับแต่งของเสียง (เสียงทุ้ม เสียงกลางหรือเสียงแหลม)
- Speed ความเร็ว คือ ความเร็วของการอ่านเสียง และ Tone ความดังของเสียง คือ การตั้งค่ารายละเอียดความถี่ (สูง-ต่ำ) ของคลื่นเสียง



ภาพที่ 7 ตัวอย่างการทดลองป้อนเนื้อหาในซอฟต์แวร์ Play.ht

### 3. การสรุปผลการทดลอง

หลังจากเก็บผลการทดลองแล้ว จะรวบรวมข้อมูลสรุปเป็นเทคนิคที่สำคัญด้านต่าง ๆ เพื่อนำไปใช้ประโยชน์ในการสร้างเสียงจากปัญญาประดิษฐ์เพื่อใช้ผลิตสื่อมัลติมีเดีย

## ผลการวิจัย

### 1. ผลการวิเคราะห์และคัดเลือกซอฟต์แวร์

จากการศึกษาข้อมูลรายละเอียดและวิเคราะห์ซอฟต์แวร์จำนวน 5 ซอฟต์แวร์สามารถสรุปเป็นตารางคุณสมบัติเฉพาะ(Characteristics) และจุดเด่นของซอฟต์แวร์ ตามตารางที่ 1

ซึ่งทีมผู้วิจัยได้เลือกใช้ซอฟต์แวร์ Play.ht ซึ่งสามารถใช้งานแบบออนไลน์ผ่านเว็บไซต์ โดยไม่ต้องทำการติดตั้งซอฟต์แวร์ลงเครื่องคอมพิวเตอร์ การดาวน์โหลดเสียงเพื่อไปใช้งานไม่ซับซ้อน (ได้เป็นไฟล์เสียง MP3 และ WAV) ทำให้มีความสะดวกในการนำไปผลิตสื่อมัลติมีเดีย เสียงภาษาไทยยังมีความเป็นธรรมชาติมากกว่าซอฟต์แวร์อื่น แต่ยังอ่านผิด และอ่านติดกัน ซอฟต์แวร์นี้มีฟังก์ชันให้เพิ่มหรือลดเอฟเฟกต์ ปรับแต่งเสียงแหลมสูงหรือทุ้ม ปรับความเร็ว และระดับเสียงเพื่อให้ได้เสียงที่มีคุณภาพใกล้เคียงเสียงธรรมชาติมากกว่าซอฟต์แวร์อื่น จึงเหมาะกับการทดลองปรับและตั้งค่าทางด้านเทคนิคต่าง ๆ ให้ได้เสียงที่เป็นธรรมชาติใกล้เคียงเสียงจริงของมนุษย์ที่สุด ซึ่งการทดลองครั้งนี้ทีมผู้วิจัยได้เลือกเสียงชายไทย โทนเสียงวัยกลางคนชื่อเสียง Niwat เนื่องจากเสียงเหมาะสมกับเนื้อหาของสื่อที่ผู้วิจัยต้องการผลิต

ทีมผู้วิจัยได้เลือกใช้ซอฟต์แวร์ Play.ht (ตารางที่ 1) ซึ่งสามารถใช้งานแบบออนไลน์ผ่านเว็บไซต์ โดยไม่ต้องทำการติดตั้งซอฟต์แวร์ลงเครื่องคอมพิวเตอร์ การดาวน์โหลดเสียงเพื่อไปใช้งานไม่ซับซ้อน (ได้เป็นไฟล์เสียง MP3 และ WAV) ทำให้มีความสะดวกในการนำไปผลิตสื่อมัลติมีเดีย เสียงภาษาไทยยังมีความเป็นธรรมชาติมากกว่าซอฟต์แวร์อื่น แต่ยังอ่านผิดและอ่านติดกัน ซอฟต์แวร์นี้มีฟังก์ชันให้เพิ่มหรือลดเอฟเฟกต์ ปรับแต่งเสียงแหลมสูงหรือทุ้ม ปรับความเร็ว และระดับเสียงเพื่อให้ได้เสียงที่มีคุณภาพใกล้เคียงเสียงธรรมชาติมากกว่าซอฟต์แวร์อื่น จึงเหมาะกับการทดลองปรับและตั้งค่าทางด้านเทคนิคต่าง ๆ ให้ได้เสียงที่เป็นธรรมชาติใกล้เคียงเสียงจริงของมนุษย์ที่สุด ซึ่งการทดลองครั้งนี้ทีมผู้วิจัยได้เลือกเสียงชายไทย โทนเสียงวัยกลางคนชื่อเสียง Niwat เนื่องจากเสียงเหมาะสมกับเนื้อหาของสื่อที่ผู้วิจัยต้องการผลิต

**ตารางที่ 1** คุณลักษณะเฉพาะ (Characteristics) และจุดเด่นของซอฟต์แวร์

ความหลากหลายของโทนเสียง (เพศ อายุ สำเนียงภาษา)	ความเป็น ธรรมชาติ ของเสียง	การเลือก ระดับความเร็ว ของเสียง	การรองรับ เสียง ภาษาไทย	ความสะดวก ของการติดตั้ง	ฟังก์ชันการใช้งาน และการ ออกแบบให้ใช้งานได้ง่าย (user-friendly)
<b>ซอฟต์แวร์: TTS</b> <a href="http://www.fromtexttospeech.com/">http://www.fromtexttospeech.com/</a>					
- รองรับได้ 8 ภาษา - ในแต่ละภาษารองรับเสียงได้ทั้ง เพศ ชาย และ หญิง ซึ่งมี ตัวเลือกเสียงให้เลือก 17 เสียง	/	/	x	ใช้งานแบบ ออนไลน์	- ใช้งานง่ายและฟรี - download เป็นไฟล์ mp3 ได้เลย
<b>ซอฟต์แวร์: Microsoft Edge</b> (Microsoft Edge Browser)					
- รองรับได้ 20 ภาษา - รองรับได้ 73 เสียง ทั้งชาย และหญิง - เสียงผู้ชายในภาษาไทยมี 1 เสียง เสียงผู้หญิงในภาษาไทย มี 1 เสียง	/	/	/	ใช้งานแบบ ออนไลน์บน Microsoft Edge Browser	- เสียงมีความเป็นธรรมชาติ ทั้งภาษาไทยและอังกฤษ - ไม่มีค่าใช้จ่าย สามารถสร้าง เสียงได้ไม่จำกัด - มีตัวเลือกเสียงให้เลือกน้อย - ต้องใช้โปรแกรมอื่นช่วยใน การดาวน์โหลดไฟล์เสียง ทำให้มีขั้นตอนที่ซับซ้อนใน การนำมาใช้งาน
<b>ซอฟต์แวร์: Play.ht</b> <a href="https://play.ht/">https://play.ht/</a>					
- รองรับ 142 ภาษา - รองรับเสียงได้ทั้งเพศ ชาย และ หญิง 907 เสียง - เสียงผู้ชายในภาษาไทยมี 1 เสียง เสียงผู้หญิงในภาษาไทย มี 2 เสียง - เสียงภาษาไทยยังมีความเป็น ธรรมชาติ แต่มีการอ่านผิด และติดกัน	/	/	/	ใช้งานแบบ ออนไลน์	- มีค่าใช้จ่ายเป็นรายปี - download เป็นไฟล์ mp3 และ wav ได้เลย - มีฟังก์ชันให้เพิ่มหรือลด เอฟเฟกต์ ปรับแต่งเสียง แหลมสูงหรือทุ้ม ปรับ ความเร็ว และระดับเสียง เพื่อให้ได้เสียงที่มีคุณภาพ ใกล้เคียงเสียงธรรมชาติที่ หลากหลาย
<b>ซอฟต์แวร์: Descript</b> <a href="https://www.descript.com/">https://www.descript.com/</a>					
- สามารถถอดเสียง ที่อัดไว้ได้ - รองรับการถอดเสียงได้ 23 ภาษา	/	/	x	ติดตั้งโปรแกรม ลงเครื่อง คอมพิวเตอร์	- สามารถสร้างเสียงด้วย ตัวเองโดยการอัดเสียง (Overdub) ลงในโปรแกรม - เสียงมีความเป็นธรรมชาติ - สามารถเลือกใช้เสียงอื่นได้ นอกจากเสียงตัวเอง ทำให้ มีเสียงให้เลือกหลากหลาย มากยิ่งขึ้น
<b>ซอฟต์แวร์: Naturalreader</b> <a href="https://www.naturalreaders.com/">https://www.naturalreaders.com/</a>					
- รองรับ 29 ภาษา - ในแต่ละภาษารองรับเสียงได้ ทั้งเพศ ชาย และ หญิง - เสียงผู้หญิงในภาษาไทยมี 1 เสียง - เสียงภาษาไทยเสียงยังขาด ความเป็นธรรมชาติ แต่มี การอ่านผิด และติดกัน	/	/	/	ใช้งานแบบ ออนไลน์	- ใช้งานง่าย - ไม่ต้องติดตั้งโปรแกรม - เลือกใช้เสียงแบบฟรี สามารถdownload เป็น ไฟล์ mp3 ได้เลย



## 2. ผลการทดลองป้อนเนื้อหาในซอฟต์แวร์ที่เลือก เพื่อเก็บรวบรวมข้อมูล ประเมินผล และสรุปเป็นเทคนิคการใช้งาน

หลังจากเลือกซอฟต์แวร์ และกำหนดเสียงที่ใช้ได้แล้ว ทีมผู้วิจัยได้ทดลองป้อนเนื้อหา โดยในครั้งที่ 1 ได้ใส่ข้อความทั้งหมด 100 คำลงในซอฟต์แวร์ และไม่มีตั้งค่าใด ๆ (ใช้ค่า default ที่มากับ ซอฟต์แวร์) มีการปรับเทคนิคการป้อนค่าและการตั้งค่าในแต่ละครั้ง เพื่อให้ได้ค่าที่ดีที่สุด หลักการเบื้องต้นในการปรับเทคนิค คือ การเว้นวรรค ใส่สัญลักษณ์พิเศษตามแบบภาษาอังกฤษ เพื่อให้มีการแบ่งกลุ่มคำ และปรับค่าต่าง ๆ ของซอฟต์แวร์ตามตารางที่ 2

ตารางที่ 2 ทดลองเทคนิคการใช้งานซอฟต์แวร์

จำนวนครั้งที่ทดลอง	การทดลองใช้เทคนิค	ผลการทดลองด้านความถูกต้องของการออกเสียง	ผลการทดลองด้านความเป็นธรรมชาติของเสียง (ความเร็วและจังหวะ)
ครั้งที่ 1 (ยังไม่ได้ปรับแต่งเทคนิคใดเลย)	<ul style="list-style-type: none"> <li>- จัดเตรียมข้อความ และใส่ข้อความทั้งหมดลงในซอฟต์แวร์ (ตัวอย่างข้อความ 100 คำ)</li> <li>- ระดับความเร็ว Speed เสียง 100</li> <li>- ใช้อุปกรณ์ลำโพงในการฟัง</li> </ul>	<ul style="list-style-type: none"> <li>- ความถูกต้อง คิดเป็นร้อยละ 92 ของข้อความทั้งหมด</li> <li>- มีคำอ่านผิด</li> <li>- มีคำที่อ่านออกเสียงไม่ถูกต้อง เช่น เจตจำนง, ธรรมชาติ, ไต ๆ, รอบอก</li> <li>- ข้อความมีตัวสะกดผิดอยู่</li> </ul>	<ul style="list-style-type: none"> <li>- เสียงยังเป็นพรืด ขาดจังหวะหายใจ ยังไม่ธรรมชาติอยู่</li> <li>- เสียงอ่านเร็วเกินไป</li> <li>- ภาษาไทยบางคำอ่านไม่ตรง (เสียงวรรณยุกต์ และเสียงควบกล้ำ)</li> <li>- ได้ยินเสียงยังไม่ชัดเจน</li> </ul>
ครั้งที่ 2	<ul style="list-style-type: none"> <li>- จัดระเบียบข้อความและป้อนคำลงในซอฟต์แวร์ 100 คำ</li> <li>- ปรับระดับความเร็ว Speed เสียงอยู่ที่ 70</li> <li>- ตรวจสอบความถูกต้องตัวสะกดของข้อความ</li> <li>- ใช้อุปกรณ์หูฟัง (Headphone) ในการฟัง</li> </ul>	<ul style="list-style-type: none"> <li>- ความถูกต้อง คิดเป็นร้อยละ 92 ของข้อความทั้งหมด</li> <li>- มีคำที่อ่านออกเสียงไม่ถูกต้อง เช่น เจตจำนง, ธรรมชาติ, ไต ๆ, รอบอก</li> </ul>	<ul style="list-style-type: none"> <li>- เสียงยังเป็นพรืด ขาดจังหวะหายใจ ยังไม่ธรรมชาติอยู่</li> <li>- เสียงอ่านช้าเกินไป</li> <li>- ภาษาไทยบางคำอ่านไม่ตรง (เสียงวรรณยุกต์ และเสียงควบกล้ำ)</li> <li>- ได้ยินเสียงชัดเจน</li> </ul>
ครั้งที่ 3	<ul style="list-style-type: none"> <li>- ใส่ข้อความแบ่งย่อยเป็นประโยค 3 ประโยค</li> </ul>	<ul style="list-style-type: none"> <li>- ความถูกต้อง คิดเป็นร้อยละ 92 ของข้อความทั้งหมด</li> <li>- คำที่อ่านออกเสียงไม่ถูกต้อง เช่น เจตจำนง, ธรรมชาติ, ไต ๆ, รอบอก</li> </ul>	<ul style="list-style-type: none"> <li>- เสียงยังเร็วและขาดจังหวะการหายใจ แต่เริ่มมีความเป็นธรรมชาติ</li> <li>- มีคำอ่านผิด</li> <li>- ภาษาไทยบางคำอ่านไม่ตรง (เสียงวรรณยุกต์ และเสียงควบกล้ำ)</li> </ul>
ครั้งที่ 4	<ul style="list-style-type: none"> <li>- ใส่ข้อความแบ่งย่อยเป็นประโยคสั้น ๆ</li> <li>- ปรับระดับความเร็ว Speed เสียงอยู่ที่ 75</li> </ul>	<ul style="list-style-type: none"> <li>- ความถูกต้อง คิดเป็นร้อยละ 92 ของข้อความทั้งหมด</li> <li>- คำที่อ่านออกเสียงไม่ถูกต้อง เช่น เจตจำนง, ธรรมชาติ, ไต ๆ, รอบอก</li> </ul>	<ul style="list-style-type: none"> <li>- เสียงเริ่มจับจังหวะเว้นการหายใจได้ แต่เสียงยังไม่เป็นธรรมชาติตอนจบประโยค</li> <li>- ภาษาไทยบางคำอ่านไม่ตรง (เสียงวรรณยุกต์ และเสียงควบกล้ำ)</li> <li>- เสียงอ่านช้าเกินไป</li> </ul>
ครั้งที่ 5	<ul style="list-style-type: none"> <li>- ปรับระดับความเร็ว Speed เสียงอยู่ที่ 80</li> <li>- ใส่สัญลักษณ์จุลภาค (,) เมื่อจบประโยค</li> <li>- แก่คำที่อ่านออกเสียงไม่ถูกต้อง คำว่า “ธรรมชาติ” เปลี่ยนเป็น ทำ-มะ-ชาด</li> </ul>	<ul style="list-style-type: none"> <li>- ความถูกต้อง คิดเป็นร้อยละ 92 ของข้อความทั้งหมด</li> <li>- คำว่า “ธรรมชาติ” ยังอ่านออกเสียงไม่ถูกต้อง</li> </ul>	<ul style="list-style-type: none"> <li>- เสียงอ่านช้าเกินไป</li> <li>- เสียงเริ่มจับจังหวะเว้นการหายใจได้ แต่เสียงยังไม่ธรรมชาติตอนจบประโยค</li> <li>- ภาษาไทยคำว่า “ธรรมชาติ” อ่านไม่ตรง (เสียงวรรณยุกต์ และเสียงควบกล้ำ)</li> </ul>

หมายเหตุ: เมื่อใช้งานซอฟต์แวร์ในระยะเวลาอันยาวนานอาจมีบางครั้งที่ซอฟต์แวร์มีปัญหาขัดข้องเนื่องจากการโหลดงานจำนวนมากในการใช้งานซอฟต์แวร์แบบออนไลน์

ตารางที่ 2 (ต่อ)

จำนวนครั้งที่ ทดลอง	การทดลองใช้เทคนิค	ผลการทดลองด้านความ ถูกต้องของการออกเสียง	ผลการทดลองด้านความเป็นธรรมชาติ ของเสียง (ความเร็วและจังหวะ)
ครั้งที่ 6	- แก่คำว่า “ธรรมชาติ” เป็น ภาษาคาราโอเกะใส่สัญลักษณ์ ยัติภังค์ (-) แต่ละคำ เปลี่ยนเป็น ทำ-มะ-ชาด	- แก่คำว่า “ธรรมชาติ” เป็น ภาษาคาราโอเกะใส่ สัญลักษณ์ ยัติภังค์ (-) แต่ละ คำ เปลี่ยนเป็น ทำ-มะ-ชาด	- แก่คำว่า “ธรรมชาติ” เป็นภาษาคารา โอเกะใส่สัญลักษณ์ ยัติภังค์ (-) แต่ละคำ เปลี่ยนเป็น ทำ-มะ-ชาด
ครั้งที่ 7	- ปรับระดับความเร็ว Speed เสียงอยู่ที่ 85 - แก่คำที่อ่านออกเสียงไม่ถูกต้อง คำว่า “ธรรมชาติ” เปลี่ยนเป็น ทำมะชาด	- ความถูกต้อง คิดเป็นร้อยละ 93 ของข้อความทั้งหมด - คำว่า “ธรรมชาติ” อ่านออก เสียงถูกต้อง	- เสียงอ่านช้าเกินไป - ภาษาไทยคำว่า “ธรรมชาติ” อ่านได้ ตรง (เสียงวรรณยุกต์ และเสียง ควบกล้ำ)
ครั้งที่ 8	- แก่คำที่อ่านออกเสียงไม่ถูกต้อง คำว่า “ต่าง ๆ” “ใด ๆ” เปลี่ยนเป็น “ต่าง ๆ” “ใด ๆ” - ปรับระดับความเร็ว Speed เสียงอยู่ที่ 90	- ความถูกต้อง คิดเป็นร้อยละ 95 ของข้อความทั้งหมด - คำว่า “ต่าง ๆ” “ใด ๆ” อ่านออกเสียงถูกต้อง	- ภาษาไทยคำว่า “ต่าง ๆ” “ใด ๆ” อ่าน ได้ตรง (เสียงวรรณยุกต์) - เสียงอ่านพอดี ไม่เร็วเกินไป ไม่ช้าเกินไป
ครั้งที่ 9	- แก่คำที่อ่านออกเสียงไม่ถูกต้อง คำว่า “คริสต์ศตวรรษ” เปลี่ยนเป็น “คริตตะสะตะวัต” - ปรับระดับความเร็ว Speed เสียงอยู่ที่ 95	- ความถูกต้อง คิดเป็นร้อยละ 95 ของข้อความทั้งหมด - คำว่า “คริสต์ศตวรรษ” ยัง อ่านออกเสียงไม่ถูกต้อง	- ภาษาไทยคำว่า “คริสต์ศตวรรษ” อ่าน ไม่ตรง (เสียงวรรณยุกต์ และเสียงควบ กล้ำ) - เสียงอ่านพอดี ไม่เร็วเกินไป ไม่ช้าเกินไป
ครั้งที่ 10	- แก่คำว่า “คริสต์ศตวรรษ” เป็น ภาษาคาราโอเกะใส่สัญลักษณ์ จุลภาค (,) เปลี่ยนเป็น คริตตะ สะตะวัต	- ความถูกต้อง คิดเป็นร้อยละ 96 ของข้อความทั้งหมด - คำว่า “คริสต์ศตวรรษ” อ่าน ออกเสียงถูกต้อง	- ภาษาไทยคำว่า “คริสต์ศตวรรษ” อ่าน ได้ตรง (เสียงวรรณยุกต์ และเสียง ควบกล้ำ)
ครั้งที่ 11	- แก่คำว่า “เจตจำนง” เป็นภาษา คาราโอเกะใส่สัญลักษณ์ จุลภาค (,) แต่ละคำ เปลี่ยนเป็น เจต,จำนง	- ความถูกต้อง คิดเป็นร้อยละ 96 ของข้อความทั้งหมด - คำว่า “เจตจำนง” ยังอ่าน ออกเสียงไม่ถูกต้อง	- ภาษาไทยคำว่า “เจตจำนง” อ่านไม่ตรง เสียงวรรณยุกต์
ครั้งที่ 12	- แก่คำว่า “เจตจำนง” เป็นภาษา คาราโอเกะใส่สัญลักษณ์ ยัติภังค์ (-) แต่ละคำ เปลี่ยนเป็น เจต-จำ-นง	- ความถูกต้อง คิดเป็นร้อยละ 97 ของข้อความทั้งหมด - คำว่า “เจตจำนง” อ่านออก เสียงถูกต้อง	- ภาษาไทยคำว่า “เจตจำนง” อ่านได้ตรง เสียงวรรณยุกต์
ครั้งที่ 13	- แก่คำว่า “ทิว” เป็นภาษาคารา โอเกะใส่สัญลักษณ์ ยัติภังค์ (-) แต่ละคำ เปลี่ยนเป็น ทิ-อิ	- ความถูกต้อง คิดเป็นร้อยละ 97 ของข้อความทั้งหมด - คำว่า “ทิว” อ่านออกเสียง ไม่ถูกต้อง	- ภาษาไทยคำว่า “ทิว” อ่านไม่ตรงเสียง วรรณยุกต์
ครั้งที่ 14	- แก่คำที่อ่านออกเสียงไม่ถูกต้อง คำว่า “ทิว” เปลี่ยนเป็น “ทิดิ”	- ความถูกต้อง คิดเป็นร้อยละ 98 ของข้อความทั้งหมด - คำว่า “ทิว” อ่านออกเสียง ถูกต้อง	- ภาษาไทยคำว่า “ทิว” อ่านได้ตรงเสียง วรรณยุกต์
ครั้งที่ 15	- แก่คำว่า “กุสโลบาย” เป็นภาษา คาราโอเกะใส่สัญลักษณ์ ยัติภังค์ (-) แต่ละคำ เปลี่ยนเป็น กุ-สะ-โล-บาย	- ความถูกต้อง คิดเป็นร้อยละ 98 ของข้อความทั้งหมด - คำว่า “กุสโลบาย” อ่านออก เสียงไม่ถูกต้อง	- ภาษาไทยคำว่า “กุสโลบาย” อ่านไม่ ตรงเสียงวรรณยุกต์

หมายเหตุ: เมื่อใช้งานซอฟต์แวร์ในระยะเวลาอันยาวนานอาจมีบางครั้งที่ซอฟต์แวร์มีปัญหาขัดข้องเนื่องจากการโหลดงานจำนวนมากในการใช้งานซอฟต์แวร์แบบออนไลน์

## ตารางที่ 2 (ต่อ)

จำนวนครั้งที่ทดลอง	การทดลองใช้เทคนิค	ผลการทดลองด้านความถูกต้องของการออกเสียง	ผลการทดลองด้านความเป็นธรรมชาติของเสียง (ความเร็วและจังหวะ)
ครั้งที่ 16	- แก่คำที่อ่านออกเสียงไม่ถูกต้อง คำว่า “กุศโลบาย” เปลี่ยนเป็น “กุ สะ โล บาย”	- ความถูกต้อง คิดเป็นร้อยละ 98 ของข้อความทั้งหมด - คำว่า “กุศโลบาย” อ่านออกเสียงไม่ถูกต้อง	- ภาษาไทยคำว่า “กุศโลบาย” อ่านไม่ตรงเสียงวรรณยุกต์
ครั้งที่ 17	- แก่คำที่อ่านออกเสียงไม่ถูกต้อง คำว่า “กุศโลบาย” เปลี่ยนเป็น “กุสะโลบาย”	- ความถูกต้อง คิดเป็นร้อยละ 99 ของข้อความทั้งหมด - คำว่า “กุศโลบาย” อ่านออกเสียงถูกต้อง	- ภาษาไทยคำว่า “กุศโลบาย” อ่านได้ตรงเสียงวรรณยุกต์
ครั้งที่ 18	- แก่คำที่อ่านออกเสียงไม่ถูกต้อง คำว่า “นอกรอบ” เปลี่ยนเป็น “นอก รอบ”	- ความถูกต้อง คิดเป็นร้อยละ 99 ของข้อความทั้งหมด - คำว่า “นอกรอบ” อ่านออกเสียงไม่ถูกต้อง	- ภาษาไทยคำว่า “นอกรอบ” อ่านไม่ตรงเสียงวรรณยุกต์
ครั้งที่ 19	- แก่คำว่า “นอกรอบ” เป็นภาษาคาราโอเกะใส่สัญลักษณ์จุลภาค (.) แต่ละคำ เปลี่ยนเป็น นอก,รอบ	- ความถูกต้อง คิดเป็น 100% ของข้อความทั้งหมด - คำว่า “นอกรอบ” อ่านออกเสียงถูกต้อง	- ภาษาไทยคำว่า “นอกรอบ” อ่านตรงเสียงวรรณยุกต์
ครั้งที่ 20	- ใส่สัญลักษณ์ จุลภาค (,) และเคาะ spacebar 1 ครั้ง และใส่สัญลักษณ์ มหัพภาค (.) เมื่อจบประโยค	- ความถูกต้อง คิดเป็น 100% ของข้อความทั้งหมด	- เสียงเริ่มเป็นธรรมชาติตอนจบประโยค
ครั้งที่ 21	- ปรับเสียงตรง Volume ในระดับ 120 dB	- ความถูกต้อง คิดเป็น 100% ของข้อความทั้งหมด	- ไฟล์เสียงที่ดาวน์โหลดออกมาดังเกินไป
ครั้งที่ 22	- ปรับเสียงตรง Volume ในระดับ Medium 0 dB	- ความถูกต้อง คิดเป็น 100% ของข้อความทั้งหมด	- ได้เสียงที่ดังเป็นมาตรฐาน - ไม่ดังเกินไปและไม่เบาเกินไป
ครั้งที่ 23	- ปรับประเมินคำเสียง Rate +50%	- ความถูกต้อง คิดเป็น 100% ของข้อความทั้งหมด	- การอ่านเร็วมากเกินไป
ครั้งที่ 24	- ปรับประเมินคำเสียง Rate -50%	- ความถูกต้อง คิดเป็น 100% ของข้อความทั้งหมด	- การอ่านช้ามากเกินไป
ครั้งที่ 25	- ปรับประเมินคำเสียง Rate เป็น Medium 0 %	- ความถูกต้อง คิดเป็น 100% ของข้อความทั้งหมด	- ได้คำเสียงอ่านที่เป็นมาตรฐาน - ความเร็วคงที่ไม่ช้าไม่เร็วเกินไป
ครั้งที่ 26	- ปรับการตั้งค่าระดับเสียง Pitch -50%	- ความถูกต้อง คิดเป็น 100% ของข้อความทั้งหมด	- เสียงจะทุ้มต่ำมากเกินไป
ครั้งที่ 27	- ปรับการตั้งค่าระดับเสียง Pitch +50%	- ความถูกต้อง คิดเป็น 100% ของข้อความทั้งหมด	- เสียงจะแหลมสูงมากเกินไป
ครั้งที่ 28	- ปรับประเมินคำเสียง Pitch เป็น Medium 0%	- ความถูกต้อง คิดเป็น 100% ของข้อความทั้งหมด	- เสียงอยู่ในเกณฑ์ที่ดีเหมาะกับการผลิตสื่อ
ครั้งที่ 29	- ใส่ Add a pause ระหว่างประโยค - ปรับตั้งค่า KHz ตามDefault ของซอฟต์แวร์	- ความถูกต้อง คิดเป็น 100% ของข้อความทั้งหมด	- ใส่ Add a pause เสียงหยุดแบบกระทันหันเกินไปไม่เป็นธรรมชาติ - การตั้งค่า KHz ไม่มีผลอะไรกับเสียง
ครั้งที่ 30	- ใส่สัญลักษณ์ จุลภาค (,) เคาะ spacebar 2 ครั้ง และใส่สัญลักษณ์ มหัพภาค (.) เมื่อจบประโยค เพื่อให้จังหวะการพูดเป็นธรรมชาติมากที่สุด	- ความถูกต้อง คิดเป็น 100% ของข้อความทั้งหมด	- เสียงเป็นธรรมชาติ - สำเนียงและภาษาออกเสียงได้อย่างถูกต้อง - เสียงและจังหวะพอดีเข้ากับสื่อที่ผลิต

หมายเหตุ: เมื่อใช้งานซอฟต์แวร์ในระยะเวลาอันยาวนานอาจมีบางครั้งที่ซอฟต์แวร์มีปัญหาขัดข้องเนื่องจากการโหลดงานจำนวนมากในการใช้งานซอฟต์แวร์แบบออนไลน์

## สรุปผลการวิจัย

จากผลการวิจัยพบว่า เมื่อลองปรับเทคนิคในแต่ละครั้ง ความถูกต้องของการออกเสียงและความเป็นธรรมชาติของเสียงดีขึ้นตามลำดับ ในการทดลองครั้งที่ 1 ความถูกต้องของการออกเสียงคิดเป็นร้อยละ 92 จนถึงครั้งที่ 30 คิดเป็นร้อยละ 100 และความเป็นธรรมชาติของเสียงปรับจากเสียงที่อ่านเร็วเกินไป และขาดจังหวะหายใจ มาเป็นเสียงที่จังหวะพอดีเข้ากับสื่อที่ผลิต ดังนั้นขั้นตอนการเตรียมพร้อมของข้อความก่อนที่จะป้อนลงซอฟต์แวร์เป็นสิ่งสำคัญ เช่น การเว้นวรรคคำใหม่ การใช้สัญลักษณ์พิเศษ การถอดอักษรเป็นคาราโอเกะสำหรับคำที่มีความซับซ้อนในการอ่าน (ตามตัวอย่างการป้อนข้อความในภาพที่ 8) รวมถึงการแบ่งข้อความออกเป็นประโยคย่อย ๆ เพื่อไม่ให้ซอฟต์แวร์ทำงานหนักมากเกินไป จะช่วยให้เครื่องประมวลผลแม่นยำขึ้น

**ข้อความก่อนใช้เทคนิค**

นักศึกษาสามารถประยุกต์ใช้ผลการเปลี่ยนแปลง**เจด-จ้ง**ของรูปแบบในสิ่งต่าง ๆ สภาพบรรยากาศเปลี่ยนแปลงจากสภาพเดิม **คริสต์ศตวรรษ**อย่างมีนัยสำคัญ โดยไม่มี**กัฏ** เพราะการมีอยู่ของสิ่งมีชีวิตใน**ธรรมชาติ** หลักฐานการเรียนรู้ได้และนักศึกษาสามารถส่งผลงานเพื่อ**รอบอก**จากอาจารย์ที่ปรึกษาได้ เอกลักษณ์ หรือ ลักษณะที่เป็นหนึ่งเดียวของสิ่งใด ๆ ในกลุ่มประเภทเดียวกัน โดยระบุได้จาก**กฏ**ไลบรารีการเปรียบเทียบ องค์ประกอบโครงสร้างภายนอก ระบบภายใน เรื่องนอกกรอบ ก็หมายถึงว่าเราเห็นเอกลักษณ์ของสิ่งนั้นจากการสร้างลักษณะเด่นหรือเอกลักษณ์

**การเตรียมข้อความแบบมีเทคนิคร่วมด้วย**

นักศึกษาสามารถประยุกต์ใช้ผลการเปลี่ยนแปลงของ**เจด-จ้ง**ในสิ่งต่าง ๆ สภาพบรรยากาศเปลี่ยนแปลงไปจากสภาพเดิม **คริสต์ศตวรรษ**อย่างมีนัยสำคัญ โดยไม่มี**กัฏ** เพราะการมีอยู่ของสิ่งมีชีวิตใน**กามะชาติ** หลักฐานการเรียนรู้ได้และนักศึกษาสามารถส่งผลงานเพื่อ**รอบอก**จากอาจารย์ที่ปรึกษาได้ เอกลักษณ์ หรือ ลักษณะที่เป็นหนึ่งเดียวของสิ่งใด ๆ ในกลุ่มประเภทเดียวกัน โดยระบุได้จาก**กฏ**ไลบรารีการเปรียบเทียบ องค์ประกอบโครงสร้างภายนอก ระบบภายใน **นอก, กรอบ** ก็หมายถึงว่าเราเห็นเอกลักษณ์ของสิ่งนั้นจากการสร้างลักษณะเด่น (เอกลักษณ์) .

ภาพที่ 8 การเตรียมข้อความ ก่อน-หลัง ใช้เทคนิค

## อภิปรายผลการวิจัยและข้อเสนอแนะ

การใช้เทคนิคสามารถช่วยป้อนข้อความภาษาไทยในซอฟต์แวร์การสร้างเสียงจากปัญญาประดิษฐ์ (Generate AI Voice) ให้เสียงพูดสังเคราะห์ที่แม่นยำและเป็นธรรมชาติมากขึ้นนั้น สามารถสรุปเป็นเทคนิคที่สำคัญในการการสร้างเสียงจากปัญญาประดิษฐ์เพื่อใช้ผลิตสื่อมัลติมีเดียได้ดังภาพที่ 9

**8 TIPS TO USE AI VOICE**

ดาวน์โหลดเทคนิคการสร้างเสียงจาก AI VOICE

<div style="background-color: #FFD700; padding: 5px; margin-bottom: 5px;"><b>1 เทคนิคการเตรียมข้อความ</b></div> <ul style="list-style-type: none"> <li>จัดระเบียบข้อความ เช่น การเว้นวรรคคำ หรือ ประโยค (ควรมีการตรวจความถูกต้องของข้อความก่อน)</li> </ul>	<div style="background-color: #FFD700; padding: 5px; margin-bottom: 5px;"><b>5 เทคนิคการแก้ไขคำอ่านออกเสียงไม่ถูกต้อง</b></div> <ul style="list-style-type: none"> <li>เปลี่ยนรูปคำหรือเปลี่ยนสระวรรณยุกต์ เช่น ธรรมชาติ เปลี่ยนเป็น ธรรมชาติ</li> <li>ช่วงเวลาที่อ่านออกเสียงของคำ อาจเป็นปัจจัยหนึ่งที่ทำให้เกิดข้อผิดพลาดในการอ่าน</li> </ul>
<div style="background-color: #FFD700; padding: 5px; margin-bottom: 5px;"><b>2 เทคนิคการป้องกันคำลงในซอฟต์แวร์</b></div> <ul style="list-style-type: none"> <li>หากประโยคยาวให้แบ่งออกเป็นประโยคสั้น ๆ</li> <li>ใส่ข้อความที่เป็นประโยคไม่ควรยาวเกิน 2-3 บรรทัด</li> </ul>	<div style="background-color: #FFD700; padding: 5px; margin-bottom: 5px;"><b>6 เทคนิคการใช้ SPEED ความเร็วในการอ่านออกเสียง</b></div> <ul style="list-style-type: none"> <li>ตั้งค่าโหนดเสียงเป็นค่า default ของซอฟต์แวร์ เหมาะกับการผลิตสื่อ</li> <li>เลือกปรับระดับความเร็วให้เหมาะสมกับสื่อ</li> <li>ปรับระดับความเร็ว Speed อยู่ที่ 90-95</li> </ul>
<div style="background-color: #FFD700; padding: 5px; margin-bottom: 5px;"><b>3 เทคนิคการเว้นจังหวะอ่านออกเสียงให้เป็นธรรมชาติ</b></div> <ul style="list-style-type: none"> <li>ควรเว้นวรรคแต่ละประโยคโดยการใส่สัญลักษณ์จุลภาค (.)</li> <li>ใส่สัญลักษณ์ จุลภาค (.) และ spacebar 2 ครั้ง และใส่สัญลักษณ์ บทบาท (.) เมื่อจบประโยค</li> </ul>	<div style="background-color: #FFD700; padding: 5px; margin-bottom: 5px;"><b>7 เทคนิคการแก้ไขปัญหาวอฟต์แวร์ขัดข้อง</b></div> <ul style="list-style-type: none"> <li>logout ออกจากซอฟต์แวร์ และเปิด Tab</li> <li>ลบข้อความออกจากซอฟต์แวร์ แล้วเปิด Tab</li> <li>เป็นระยะของการเข้าซอฟต์แวร์สักพักแล้วลองเข้าไปอีกครั้ง</li> </ul>
<div style="background-color: #FFD700; padding: 5px; margin-bottom: 5px;"><b>4 เทคนิคการการแยกเสียงภาษาไทยกับภาษาอังกฤษ</b></div> <ul style="list-style-type: none"> <li>ปรับระดับความยากหรือการใส่ภาษาการอ่าน</li> <li>ใช้ภาษาการอ่านเป็นคำอ่านภาษาไทย เช่น Module แก้ไขใน Google</li> <li>ภาษาการอ่านอ่านได้คำสั้นๆ</li> </ul>	<div style="background-color: #FFD700; padding: 5px; margin-bottom: 5px;"><b>8 เทคนิคการเลือกใช้อุปกรณ์ที่ใช้ในการฟัง</b></div> <ul style="list-style-type: none"> <li>การใช้หูฟังจะช่วยให้การรับฟังเสียงได้ดี</li> <li>ไม่ควรปรับความดังเสียงบ่อยๆ อาจทำให้ระบบอัตโนมัติของเครื่องรวบได้</li> </ul>

ภาพที่ 9 เทคนิคในการสร้างเสียงจากปัญญาประดิษฐ์

จากข้อสรุปเทคนิคที่สำคัญในการการสร้างเสียงจากปัญญาประดิษฐ์เพื่อใช้ผลิตสื่อมัลติมีเดีย งานเทคโนโลยีการเรียนรู้ คณะศิลปศาสตร์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี สามารถนำไปประยุกต์ใช้งานได้ดังนี้

1. นำเทคนิคไปปรับใช้กับการผลิตสื่อการเรียนการสอนในรายวิชาต่าง ๆ ของคณะฯ ซึ่งหน่วยงานได้ผลิตสื่อการเรียนการสอนในรูปแบบวิดีโอเป็นจำนวนมากให้กับผู้สอน โดยได้นำซอฟต์แวร์การสร้างเสียงจากปัญญาประดิษฐ์ (Generate AI Voice) มาช่วยในการสร้างเสียงและใช้เทคนิคในการใส่ข้อความลงในซอฟต์แวร์ช่วยอำนวยความสะดวกให้กับผู้สอนมากขึ้น มีทางเลือกให้ผู้สอนมากขึ้น เช่นในกรณีที่ผู้สอนไม่สะดวกเข้ามาอัดเสียงที่ห้องสตูดิโอของคณะฯ และเสียงที่ส่งเคราะห์ออกมาไม่น่ามีเสียงรบกวน (Noise) จึงช่วยลดขั้นตอนการตัดต่อเสียงของทีมงานผลิตสื่อด้วย

2. นำไปพัฒนางานในการผลิตสื่ออื่น ๆ ของคณะฯ และมหาวิทยาลัย ได้อย่างรวดเร็ว เช่น เมื่อมีความต้องการใช้วิดีโออย่างเร่งด่วนซึ่งมีเนื้อที่ต้องการสื่อสารอยู่แล้วแต่ขาดคนพากย์เสียง ทีมงานผลิตสื่อสามารถนำข้อความมาออกแบบ storyboard และทำเสียงประกอบได้ทันทีและทันเวลาต่อการใช้งาน เช่น การผลิตกราฟิกเคลื่อนไหว (motion graphic) สำหรับคู่มือการใช้งานต่าง ๆ ได้อย่างรวดเร็ว

3. นำเทคนิคการสร้างเสียงสังเคราะห์ไปเป็นส่วนหนึ่งในการจัดกิจกรรม Sharing Session ให้กับอาจารย์ผู้สอนและบุคลากรที่มีความสนใจในการการผลิตสื่อมัลติมีเดีย

4. นำไปต่อยอดงานวิจัยเพื่อพัฒนางานด้านผลิตสื่อ โดยใช้เทคนิคที่ได้ไปเป็นแนวทางในการทดลองกับซอฟต์แวร์อื่น ๆ และเปลี่ยนชนิดของเสียง เพื่อสร้างความหลากหลายของเสียงในงานผลิตสื่อและงานอื่น ๆ มากขึ้น

นอกจากเทคนิคต่าง ๆ ที่ทีมผู้วิจัยได้ทดลองเพื่อใช้แก้ปัญหาความแม่นยำและความเป็นธรรมชาติของเสียงภาษาไทยจากการสร้างเสียงจากปัญญาประดิษฐ์ (Generate AI Voice) ในเบื้องต้นแล้วนั้นปัจจัยสำคัญของการเพิ่มคุณภาพและประสิทธิภาพโดยรวมของซอฟต์แวร์การสร้างเสียงจากปัญญาประดิษฐ์ (Generate AI Voice) คือ การรวบรวมชุดข้อมูล (datasets) ที่มีรูปแบบโทนเสียงภาษาไทยมากขึ้น เพื่อนำมาสอน (training dataset) ให้กับเครื่อง และมีชุดข้อมูลที่นำมาทดสอบ (testing dataset) ความแม่นยำความถูกต้องของรูปแบบที่เครื่องได้เรียนรู้ (<https://airesearch.in.th/releases/speech-emotion-dataset/>) ซึ่งชุดข้อมูลจะเกี่ยวกับการวิเคราะห์ความสัมพันธ์ระหว่างไวยากรณ์ (syntax) ความหมาย (semantics) การตีความ (pragmatics) ให้มีความเหมาะสมที่สุด (Dutoit, 1997) รวมถึงเรียนรู้การสร้างคำพูดที่มีความซับซ้อนของระบบสนทนา และระบบอื่น ๆ ที่ใช้การสร้างภาษา (Taylor, 2009) เพื่อให้เครื่องสามารถเก็บสถิติของชุดข้อมูล มีรูปแบบที่ทำงานได้ดีกับข้อมูลที่ไม่เคยเรียนรู้มาก่อน และวิเคราะห์ประโยคใหม่ ๆ ได้

หากมีการพัฒนาปัญญาประดิษฐ์ (AI) ให้เรียนรู้และเข้าใจภาษาไทยเพิ่มมากขึ้น จะสามารถเพิ่มโอกาสให้หลายภาคส่วนได้นำเสียงสังเคราะห์มาปรับใช้เพื่ออำนวยความสะดวกในชีวิตประจำวัน การทำงาน รวมถึงประโยชน์สาธารณะ เช่น การใช้เสียงด้วยปัญญาประดิษฐ์ช่วยผู้มีความบกพร่องด้านการพูดได้ ซึ่งการสังเคราะห์เสียงภาษาไทยให้ถูกต้องแม่นยำและได้เสียงที่เป็นธรรมชาตินั้นจะมีการพัฒนาต่อไปในอนาคต

### กิตติกรรมประกาศ

การวิจัยเรื่อง การศึกษาเทคนิคการสร้างเสียงภาษาไทยจากปัญญาประดิษฐ์เพื่อใช้ผลิตสื่อมัลติมีเดีย สามารถดำเนินการจนประสบความสำเร็จลุล่วงไปด้วยดี เนื่องจากได้รับความอนุเคราะห์และสนับสนุนเป็นอย่างดีจากคณะศิลปศาสตร์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรีที่สนับสนุนการทำวิจัย การเช่าซอฟต์แวร์แบบมีลิขสิทธิ์ให้ทีมผู้วิจัยได้ใช้และทดลอง ขอขอบพระคุณ ดร.พรเลิศ อาภาทนต์ รองคณบดีฝ่ายบริหาร ของคณะศิลปศาสตร์ ที่ได้ให้คำปรึกษาและขอแนะนำ

ขอขอบคุณเพื่อนร่วมงานทุกคนที่แลกเปลี่ยนเรียนรู้ร่วมกัน สุดท้ายนี้ผู้วิจัยหวังว่างานวิจัยฉบับนี้จะเป็นประโยชน์สำหรับหน่วยงานที่เกี่ยวข้อง และผู้ที่สนใจศึกษาต่อไป



## เอกสารอ้างอิง

- กลุ่มภารกิจบริการและพัฒนาสื่อการเรียนการสอน สำนักคอมพิวเตอร์ (Taksin University). 2564. เทคโนโลยีคอมพิวเตอร์ใน  
อนาคต สาขาของปัญญาประดิษฐ์. [online]. Available: [https://cst.tsu.ac.th/courseonline/course/  
computer\\_it/AI/nature.html](https://cst.tsu.ac.th/courseonline/course/computer_it/AI/nature.html). (สืบค้นเมื่อ สิงหาคม 2565).
- ชาภิยะห์ ตาเยะ. 2565. NLP คืออะไร. [online]. Available: <https://www.mindphp.com>. (สืบค้นเมื่อ สิงหาคม 2565).
- Dilmegani, C. 2022. A Complete Guide to NLP: What it is, How it Works & Use Cases. [Online]. Available:  
<https://research.aimultiple.com/nlp/>. (Retrieved August 2022).
- Dutoit, T. 1997. Conclusions and Perspectives. pp. 271-274. In: Media, B.V. (Ed.). An introduction to text-to-  
speech synthesis. Springer Science+Business. Netherland.
- Jazz.ordinaryday. 2020. รู้จัก “Thai NLP” ตัวช่วยสุดเจ๋งที่ทำให้คอมพิวเตอร์เข้าใจภาษาไทยได้มากขึ้น. [online]. Available:  
<https://www.mangozero.com/what-is-thai-nlp/>. (สืบค้นเมื่อ สิงหาคม 2565).
- Joseph, S.J., Hlomani, H., Letsholo, K., Kaniwa, F. and K. Sedimo. 2016. Natural Language Processing: A Review.  
International Journal of Research in Engineering and Applied Sciences. 6(3): 207-210.
- Mache, S., and C.N. Mahender. 2016. Development of Text-to-Speech Synthesizer for Pali Language. Journal of  
Computer Engineering. 18(3): 35-42.
- Perkins, J. 2010. Python Text Processing with NLTK 2.0 Cookbook, United Kingdom: Packt Publishing. 265 pages.
- SAS Global Forum. 2020. Natural Language Processing with SAS Special Collection. Cary, NC: SAS Institute Inc.  
USA. 74 pages.
- Taylor, P. 2009. Conclusion. in Cambridge University Press, Text-to-Speech Synthesis, (p. 550). United States of  
America: Cambridge University Press, New York. 626 pages.
- Voicemod. 2023. What is an AI voice?. [Online]. Available: <https://www.voicemod.net/ai-voice-technology>.  
(Retrieved March, 2023).