

## ระบบ Speech Re-Synthesis สำหรับผู้บกพร่องทางการออกเสียง

### 1. ที่มาและความสำคัญของโครงการ

ในปัจจุบัน เทคโนโลยีการรู้จำเสียงพูด (Speech-to-Text: STT) ได้รับการพัฒนาอย่างกว้างขวางและถูกนำไปใช้งานในหลายบริบท เช่น ระบบสั่งงานด้วยเสียง การถอดเสียงอัตโนมัติ และผู้ช่วยอัจฉริยะ อย่างไรก็ตาม ระบบ STT มีข้อจำกัดสำคัญเมื่อใช้งานกับผู้บกพร่องทางการออกเสียง เช่น ผู้ป่วยโรคทางระบบประสาท ผู้ป่วยโรคหลอดเลือดสมอง หรือเด็กที่มีความผิดปกติด้านการออกเสียง เนื่องจากเสียงพูดมีความไม่ชัดเจนและแตกต่างจากเสียงปกติ ส่งผลให้การรู้จำเสียงมีความผิดพลาดสูง

นอกจากนี้ แม้ STT จะสามารถแปลงเสียงเป็นข้อความได้ แต่ยังไม่สามารถแก้ปัญหาหลักของผู้ใช้กลุ่มนี้ได้อย่างแท้จริง กล่าวคือ ผู้ใช้ยังไม่สามารถ “สื่อสารด้วยเสียงพูด” กับผู้อื่นได้โดยตรง

ดังนั้น โครงการนี้จึงมุ่งเน้นการพัฒนา ระบบ Speech Re-Synthesis ซึ่งเป็นระบบที่รับเสียงพูดที่ฟังยากหรือไม่ชัดเจน และทำการสังเคราะห์เสียงใหม่ให้เป็นเสียงพูดที่ชัดเจน เป็นธรรมชาติ และสามารถนำไปใช้สื่อสารได้ในทันที โดยมีเป้าหมายเพื่อคืนความสามารถในการสื่อสารด้วยเสียงให้แก่ผู้บกพร่องทางการออกเสียง

### 2. แนวคิดหลักของ Speech Re-Synthesis

Speech Re-Synthesis แตกต่างจาก Speech-to-Text อย่างชัดเจน โดยไม่ได้มีเป้าหมายเพียงการแปลงเสียงเป็นข้อความ แต่เน้นการสร้าง **สัญญาณเสียงพูดใหม่** ที่มนุษย์สามารถฟังเข้าใจได้ง่าย และมีคุณภาพใกล้เคียงเสียงพูดปกติ

แนวคิดสำคัญของโครงการ คือ

- การแยก “เนื้อหาเสียงพูด (content)” ออกจาก “ลักษณะการออกเสียงที่ผิดปกติ”
- การใช้แบบจำลองปัญญาประดิษฐ์เพื่อเรียนรู้โครงสร้างของเสียงพูดที่ชัดเจน
- การสังเคราะห์เสียงใหม่จากเนื้อหาเดียวกัน แต่มีการออกเสียงที่ถูกต้องและเป็นธรรมชาติ

เมื่อเปรียบเทียบกับ STT ซึ่งให้ผลลัพธ์เป็นข้อความ ระบบ Speech Re-Synthesis ต้องสร้างผลลัพธ์เป็นคลื่นเสียง (waveform) ที่มีความละเอียดสูง และต้องคำนึงถึงจังหวะ น้ำเสียง และลักษณะการพูดของมนุษย์ จึงถือเป็นปัญหาที่มีความท้าทายสูงกว่า แต่ให้ผลกระทบเชิงบวกต่อผู้ใช้งานมากกว่า

### 3. ความเป็นไปได้ในการพัฒนาโดยไม่ใช้เสียงผู้บกพร่องจริง

หนึ่งในประเด็นสำคัญของโครงการนี้ คือ การหลีกเลี่ยงการเก็บข้อมูลเสียงของผู้บกพร่องทางการออกเสียงโดยตรง ซึ่งมีข้อจำกัดด้านจริยธรรม ความเป็นส่วนตัว และความยากในการรวบรวมข้อมูล

โครงการนี้เสนอแนวทางการพัฒนาระบบโดยไม่ต้องใช้เสียงผู้ป่วยจริง ดังนี้

- ใช้เสียงพูดปกติเป็นฐานข้อมูล
- สร้างเสียงพูดที่ “ฟังยาก” ด้วยเทคนิคการจำลองความผิดปกติของการออกเสียง (Simulated Speech Distortion) เช่น การยืดเวลาเสียง การลดความชัดของหน่วยเสียง หรือการเปลี่ยนลักษณะของ formant
- ฝึกแบบจำลองให้เรียนรู้การแปลงเสียงที่ถูกบิดเบือนให้กลับมาเป็นเสียงพูดที่ชัดเจน

แนวทางนี้ช่วยให้สามารถสร้างข้อมูลแบบ paired data ได้อย่างมีประสิทธิภาพ ลดความเสี่ยงด้านจริยธรรม และยังเหมาะสมกับการทำโครงการเชิงวิชาการในระดับมหาวิทยาลัย

### 4. โครงสร้างระบบโดยรวม (System Overview)

ระบบ Speech Re-Synthesis สำหรับผู้บกพร่องทางการออกเสียงที่นำเสนอในโครงการนี้ ถูกออกแบบเป็นระบบประมวลผลเสียงพูดแบบเป็นลำดับขั้น (pipeline-based system) โดยมีเป้าหมายในการรับสัญญาณเสียงพูดที่มีความไม่ชัดเจนหรือถูกบิดเบือน และทำการสังเคราะห์เสียงพูดใหม่ให้มีความชัดเจน เป็นธรรมชาติ และสามารถนำไปใช้ในการสื่อสารได้อย่างมีประสิทธิภาพ

โครงสร้างระบบโดยรวมประกอบด้วยขั้นตอนหลักดังต่อไปนี้

ขั้นตอนแรก คือ การรับสัญญาณเสียงเข้า (Input Speech Signal) ระบบจะรับสัญญาณเสียงพูดที่มีความไม่ชัดเจน ซึ่งอาจเกิดจากความผิดปกติของการออกเสียง หรือเสียงที่ถูกบิดเบือนจากกระบวนการจำลองความผิดปกติ โดยสัญญาณเสียงดังกล่าวจะถูกปรับให้อยู่ในรูปแบบมาตรฐานเดียวกัน เช่น อัตราการสุ่มตัวอย่าง (sampling rate) และระดับความดัง เพื่อให้เหมาะสมกับการประมวลผลในขั้นถัดไป

ขั้นตอนที่สอง คือ การแปลงเสียงเป็นคุณลักษณะเชิงลึก (Speech Representation Extraction) ระบบจะใช้แบบจำลองการเรียนรู้แบบไม่ใช้ป้ายกำกับ (self-supervised learning) เพื่อแปลงสัญญาณเสียงดิบให้เป็นคุณลักษณะเชิงลึกที่สะท้อนข้อมูลสำคัญของเสียงพูด เช่น โครงสร้างของหน่วยเสียงและลักษณะทางภาษาศาสตร์ โดยไม่ขึ้นกับลักษณะการออกเสียงที่ผิดปกติ ขั้นตอนนี้มีบทบาทสำคัญในการลดผลกระทบจากความเพี้ยนของเสียง

ขั้นตอนที่สาม คือ การสกัดเนื้อหาเสียงพูด (Content Representation) จากคุณลักษณะเชิงลึกที่ได้ ระบบจะมุ่งเน้นการแยกข้อมูลด้านเนื้อหาเสียงพูด (linguistic content) ออกจากลักษณะเฉพาะของเสียงที่ไม่ชัดเจน เพื่อให้ได้ตัวแทนของเนื้อหาที่สามารถนำไปใช้ในการสร้างเสียงพูดใหม่ได้อย่างมีประสิทธิภาพ

ขั้นตอนที่สี่ คือ การสังเคราะห์เสียงพูดใหม่ (Speech Re-Synthesis) ระบบจะนำตัวแทนเนื้อหาเสียงพูดที่ได้ไปผ่านแบบจำลองสังเคราะห์เสียง เพื่อสร้างสัญญาณเสียงพูดใหม่ที่มีความชัดเจนและเป็นธรรมชาติ โดยกระบวนการนี้อาจกำหนดให้เสียงพูดที่สร้างขึ้นมีลักษณะเป็นเสียงกลาง (neutral voice) หรือเสียงที่เหมาะสมกับบริบทการใช้งาน ทั้งนี้เป้าหมายหลักคือการทำให้ผู้ฟังสามารถเข้าใจเนื้อหาเสียงพูดได้ง่ายขึ้น

ขั้นตอนสุดท้าย คือ การส่งออกผลลัพธ์ (Output Speech Signal) ระบบจะส่งออกสัญญาณเสียงพูดที่ผ่านการสังเคราะห์แล้ว ซึ่งสามารถนำไปใช้ในการสื่อสารหรือประเมินผลคุณภาพเสียงต่อไปได้ โดยโครงสร้างระบบดังกล่าวสามารถปรับขยายเพื่อรองรับการทำงานแบบใกล้เคียงเวลาจริง (near real-time) ในอนาคต

โดยสรุป โครงสร้างของระบบถูกออกแบบให้มีความยืดหยุ่น สามารถพัฒนาและประเมินผลได้ในแต่ละขั้นตอนอย่างเป็นระบบ และเหมาะสมสำหรับการนำไปประยุกต์ใช้ในงานด้านเทคโนโลยีช่วยเหลือผู้บกพร่องทางการสื่อสาร

## 5. การประเมินผลระบบ

การประเมินผลของระบบไม่ได้พิจารณาเพียงคุณภาพเสียงในเชิงเทคนิคเท่านั้น แต่ยังคำนึงถึงความเข้าใจของมนุษย์ โดยใช้เกณฑ์ เช่น

- ความสามารถในการทำให้เสียงพูดเข้าใจง่ายขึ้น (Intelligibility)
- อัตราความผิดพลาดของระบบรู้จำเสียง (WER) ก่อนและหลังการประมวลผล
- การประเมินคุณภาพเสียงจากผู้ฟัง (Subjective Listening Test)

การประเมินในลักษณะนี้ช่วยสะท้อนประสิทธิภาพของระบบได้ใกล้เคียงกับการใช้งานจริง

## 6. สรุปภาพรวมของโครงการ

โครงการ Speech Re-Synthesis สำหรับผู้บกพร่องทางการออกเสียง เป็นโครงการที่ผสมผสานองค์ความรู้ด้านปัญญาประดิษฐ์ การประมวลผลสัญญาณเสียง และเทคโนโลยีเพื่อช่วยเหลือมนุษย์ (Assistive Technology) โดยมีความท้าทายเชิงเทคนิคสูงกว่าระบบ Speech-to-Text ทั่วไป แต่สามารถดำเนินการได้จริงโดยอาศัยเทคนิค self-supervised learning และการจำลองข้อมูลเสียง

โครงการนี้มีศักยภาพในการต่อยอดทั้งในเชิงวิจัยและเชิงการใช้งานจริง และสามารถเป็นรากฐานสำคัญสำหรับการพัฒนาเทคโนโลยีช่วยเหลือผู้บกพร่องทางการสื่อสารในอนาคต