

ชื่อเรื่อง การซิงค์เสียงอัตโนมัติเพื่อความสะดวกสำหรับผู้สร้างคอนเทนต์

โดย นายเมธี บุญชู
นายโชคทวี พลเมือง

บทคัดย่อ

โครงการนี้พัฒนาระบบที่จัดเรียงข้อความให้ตรงกับเสียงในวิดีโอโดยอัตโนมัติ ซึ่งช่วยลดเวลาและความซับซ้อนในการสร้างและแก้ไขคำบรรยาย ระบบนี้ใช้เทคนิคการบิดเวลาแบบไดนามิก (Dynamic Time Warping หรือ DTW) เพื่อจัดเรียงข้อมูลอนุกรมเวลา แม้จะมีความแตกต่างในความเร็วหรือจังหวะของการพูด โดยแปลงเสียงเป็นค่าสัมประสิทธิ์เซปสตรัลความถี่เมล (Mel-Frequency Cepstral Coefficients หรือ MFCCs) แล้วใช้ อัลกอริทึม DTW ในการจับคู่ข้อความกับเสียงพูดอย่างแม่นยำ พร้อมสร้างเวลาแสดงผลที่ถูกต้องสำหรับแต่ละคำและประโยค

การทดสอบกับวิดีโอหลากหลายประเภทแสดงให้เห็นถึงความแม่นยำในการจัดเรียงสูงถึง 95% แม้ในสภาวะที่ท้าทาย เช่น อัตราการพูดที่แปรผันและเสียงรบกวนพื้นหลังระดับปานกลาง เมื่อเทียบกับวิธีการทำด้วยมือ วิธีการนี้สามารถลดเวลาในการผลิตคำบรรยายลงได้ประมาณ 70% ในขณะที่ยังคงรักษาความแม่นยำระดับสูงไว้ ผลลัพธ์เหล่านี้ยืนยันว่าการซิงโครไนซ์เสียงพูดกับข้อความโดยอัตโนมัติช่วยเพิ่มประสิทธิภาพในการผลิตเนื้อหาได้อย่างมีนัยสำคัญ โดยลดความจำเป็นในการปรับเวลาแสดงผลด้วยตนเอง ซึ่งเป็นงานที่น่าเบื่อและใช้เวลามาก

เทคโนโลยีนี้สามารถนำไปใช้งานได้ทันทีในสื่อการศึกษา เทคโนโลยีช่วยเหลือสำหรับผู้บกพร่องทางการได้ยิน และการสร้างคำบรรยายอัตโนมัติสำหรับแพลตฟอร์มเนื้อหาดิจิทัล สำหรับการพัฒนาในอนาคต จะมุ่งเน้นการปรับปรุงระบบให้เหมาะสมกับการประมวลผลแบบเรียลไทม์ และขยายการรองรับไปยังภาษาอื่น ๆ นอกเหนือจากภาษาอังกฤษ

Thesis Title: Automatic Audio Synchronization for Content Creators

By Mr. Metee Boonchu

Mr. Choktavee Ponlamuang

ABSTRACT

This project develops a system that automatically aligns text with audio in videos, significantly reducing the time and complexity of subtitle creation and editing. The system utilizes Dynamic Time Warping (DTW) to align time-series data despite variations in speech speed or rhythm. By converting audio into Mel-Frequency Cepstral Coefficients (MFCCs) and applying DTW algorithms, the system precisely matches text with speech, generating accurate timestamps for words and sentences.

Testing across diverse video content demonstrates 95% alignment accuracy, even under challenging conditions such as variable speaking rates and moderate background noise. Compared to manual methods, this approach reduces subtitle production time by approximately 70% while maintaining high accuracy. The results confirm that automated speech-text synchronization substantially improves content production efficiency by eliminating tedious manual timestamp adjustment.

The technology has immediate applications in educational media, assistive technology for individuals with hearing impairments, and automated subtitle generation for digital content platforms. Future work will focus on optimizing the system for real-time processing and expanding language support beyond English.