

ชื่อเรื่อง ระบบซิงค์เสียงพูดกับข้อความ

โดย Mr. Sakponhrik Vy

บทคัดย่อ

โครงการนี้พัฒนาอัลกอริทึมที่ซิงโครไนซ์ข้อความที่กำหนดกับคลิปวิดีโอที่คู่กัน โดยขั้นแรกเสียงพูดจะถูกสังเคราะห์จากข้อความที่กำหนดโดยระบบสังเคราะห์เสียงพูดจากข้อความ ที่ประโยคหรือทีละส่วนที่กำหนด จากนั้นระบบจะจับคู่เสียงที่สังเคราะห์กับเสียงพูดในวิดีโอโดยใช้เทคนิค Dynamic-Time warping (DTW) เนื่องจากเรารู้ความยาวระยะเวลาของเสียงที่สังเคราะห์ที่ได้แต่ละส่วน เราจึงสามารถใช้ข้อมูลระยะเวลานี้ เชื่อมโยงกลับไปยังเวลาที่สอดคล้องกับเสียง ณ เวลาต่างๆ ในวิดีโอ แล้วจะได้ผลลัพธ์เป็น time stamp ของแต่ละประโยคหรือคำที่เชื่อมโยงถึงเวลาในเสียงของวิดีโอ จากการทดสอบในหลายกรณีจะเห็นว่าโครงการนี้ประสบความสำเร็จโดยระบบทำงานได้ดี และมีความแม่นยำเท่ากับซอฟต์แวร์และการถอดข้อความโดยคนจริง

Thesis Title : Speech-Text Alignment System

By Mr. Sakponhrik Vy

ABSTRACT

This project develops an algorithm that synchronizes a given text with the corresponding video clips. First, a speech voice is synthesized from the given text by a text-to-speech system, sentence by sentence. Then, the synthesized voice is mapped with the speech component from the video using Dynamic-Time warping (DTW). Since the sentence durations of the synthesized voice can be obtained easily, we can use them to map back to the time that corresponds to the words in the speech component in the video. This will result in a time stamp of each sentence or word associated with the audio timeline of the video. After many tests, the project was successful. The system performing well and being accurate as the software and human transcription.

Faculty Of Engineering