

วิทยานิพนธ์นี้มีวัตถุประสงค์เพื่อศึกษาวิธีการสร้างระบบสำหรับช่วยถอดข้อความเสียงพูดภาษาไทยที่ทำงานได้อย่างอัตโนมัติสำหรับเสียงพูดเฉพาะบุคคล โดยใช้หลักการของระบบรู้จำเสียงพูดอย่างต่อเนื่องสำหรับคำศัพท์จำนวนมาก (Large Vocabulary Continuous Speech Recognition: LVCSR) ซึ่งใช้แบบจำลองฮิดเดนมาร์คอฟเป็นวิธีการพื้นฐานในการรู้จำ สำหรับเทคนิคที่นำเสนอได้มุ่งเน้นไปที่กระบวนการในส่วนของเตรียมข้อมูลสำหรับใช้สร้างแบบจำลองทางเสียงและแบบจำลองทางภาษาที่เหมาะสม ซึ่งได้นำเสนอวิธีการเรียงจับคู่ตำแหน่งของคำในข้อมูลเสียงและในข้อมูลถอดถอดความให้มีตำแหน่งตรงกันด้วยวิธีการไดนามิกทามวาร์ปปีง และยังสามารถเพิ่มส่วนของการวัดค่าความมั่นใจด้วยวิธีการค้นไม้การตัดสินใจ เพื่อการคัดกรองผลจากการเรียงจับคู่เฉพาะที่ถูกต้องสำหรับนำไปใช้ปรับปรุงแบบจำลองของระบบที่ถูกสร้างขึ้นจากชุดฐานข้อมูลไม่เฉพาะบุคคล จากการทดลองเรียงจับคู่ด้วยวิธีการที่นำเสนอกับชุดข้อมูลทดสอบให้ความถูกต้องของการรู้จำหน่วยเสียงสูงกว่าร้อยละ 50 และเมื่อใช้ผลของการจับคู่ด้วยเทคนิคการเรียงจับคู่ดังกล่าวในระบบการรู้จำเสียงพูดสำหรับคำศัพท์จำนวนมากสามารถให้ความถูกต้องในการถอดความสูงขึ้นจากแบบจำลองต้นแบบถึงร้อยละ 27 นับว่าสามารถเพิ่มความถูกต้องของการทำงานขึ้นได้อย่างมากสำหรับระบบที่มีทำงานอย่างอัตโนมัติ

Abstract

228186

The purpose of this thesis is to invent an automatic Thai speaker-dependent speech transcription for using in a Large Vocabulary Continuous Speech Recognition (LVCSR) system. The proposed technique is focused on the data preparation process, which arranges suitable data for training an acoustic model and a language model by using Hidden Markov Model (HMM). The Dynamic Time Warping (DTW) technique is used to match phone's position on recognized utterances and their transcription data in the data preparation process. The false matched results are rejected by thresholding with confidence score, which is obtained from Classification and Regression Tree (CART). After that, rejected results are used as a resource to adapt the original acoustic model, which is created from a general-domain corpus. The matching accuracy is higher than 50 percent on the test set. The transcribed result which uses the ultimate model is improved by 27 percent from the original model on the same test set.