บทคัดย่อ

ประสิทธิภาพของ ระบบรู้จำเสียงพูดอัติโนมัติและ ระบบสังเคราะห์เสียงพูด ขึ้นอยู่กับความครอบคลุมของหน่วยเสียงจากคลังข้อความที่เหมาะสม งานวิจัยนี้เสนอการสร้างคลังข้อความอัตโนมัติ จากการกระจายตัวของหน่วยเสียงตามที่กำหนดการกระจายตัวของหน่วยตามที่กำหนดนั้น สามารถกำหนดได้จากชนิดของหน่วยเสียง ขนาดของคลังข้อความ เกณฑ์ขั้นต่ำของจำนวนหน่วยเสียง และรูปแบบของการกระจายตัวเป้าหมาย ได้คัดเลือกข้อความมาจากข้อมูลจากอินเตอร์เน็ต โดยข้อความนั้นจะถูก จัดเก็บมาอย่างต่อเนื่องโดย กระบวนการดึงบทความจากหน้าเว็บบนอินเตอร์เน็ต จนกระทั่งได้คลังข้อความที่เหมาะสม ในวิทยานิพนธ์นี้ยังได้ประยุกต์ใช้วิธีการเชิงละโมบ เพื่อเลือกประโยคที่เหมาะสมที่จะทำให้เกิดการกระจายตัวของหน่วยเสียงตามเป้าหมาย ในการทดลองได้ใช้ข้อความจากฐานข้อมูล Large Vocabulary Continuous Speech Recognition (LVCSR) corpus for Thai language ในการสร้างเป้าหมายของการกระจายตัวหน่วยเสียง ผลการทดลองที่ได้คือ จำนวนของข้อมูลข้อความที่ดึงมาจากอินเตอร์เน็ตที่เพิ่มขึ้นสามารถทำให้การกระจายตัวของหน่วยเสียงเป็นไปตามเป้าหมายได้ และเกิดความครอบคลุมทางหน่วยเสียงคู่ ถึง 99.13% คลังข้อความที่ถูกสร้างขึ้นนี้ จึงสามารถนำไปใช้ในการสร้างคลังเสียงพูดได้อย่างมีประสิทธิภาพ

ABSTRACT

Performance of Automatic Speech Recognition (ASR) and Text-to-Speech (TTS) systems depend on appropriate text corpus. This article explains about the automated text corpus generating method using custom phonetic distribution. This distribution is defined by phonemes type, corpus size, minimum criterion number of phonemes, and target phonetic distribution. Generally, the system selects text data from the internet by continuously downloading them using web crawler. The greedy algorithm is applied to extract the proper sentences, in order to fit with the target phonetic distribution until the appropriate text corpus is established. The experiment is done by using the text from Large Vocabulary Continuous Speech Recognition (LVCSR) corpus for Thai language to generate target phonetic distribution. The result shown that, the increased number of data drawn from the internet is able to accomplish target phonetic distribution and generate diphone coverage for 99.13%. This text corpus then generate speech corpus efficiently.