ชื่อวิทยานิพนธ์	ระบบรู้จำเสียงพูดภาษาไทยอัตโนมัติแบบคำต่อเนื่องโดยใช้แบบจำลองฮิดเดนมาร์คอฟ
ผู้เขียน	นายพงศ์ธร ตันธนกิจ
สาขาวิชา	วิศวกรรมคอมพิวเตอร์
ปีการศึกษา	2547

บทคัดย่อ

งานวิจัยนี้กล่าวถึงการสร้างระบบรู้จำเสียงพูดภาษาไทย โดยใช้แบบจำลองฮิดเดนมาร์คอฟเพื่อหาส่วนประกอบที่เหมาะสมสำหรับการรู้จำเสียงพูดภาษาไทยแบบคำต่อเนื่องและเพื่อสร้างแบบจำลองเสียงเริ่มต้นสำหรับการพัฒนาต่อในอนาคต ผลจากการวิจัยทำให้ได้แบบจำลองเสียง ที่ใช้ค่าลักษณะเด่นแบบ MFCC+DA โครงสร้าง HMM แบบ 3 สถานะ 8 มิกซเจอร์ ใช้ชุดหน่วยเสียง C_i+V_t+C_f 170 หน่วยเสียง โดยการฝึกฝนด้วยข้อมูลเสียงพูดทั่วไป 900 ประโยคจากผู้พูด 45 คนที่บันทึกภายในห้องบันทึกเสียง แล้วทำการทดสอบการจำแนกหน่วยเสียงโดยใช้แบบจำลองเสียงดังกล่าว ด้วยข้อมูลเสียง 200 ประโยคจากผู้พูด 10 คนบันทึกเสียงภายในห้องบันทึกเสียง ได้ค่าความแม่นยำ 58.3% เมื่อแยกตามกลุ่มพยัญชนะ สระ และตัวสะกด เป็น 75.3% 39.6% และ 77.6% ตามลำดับ สรุปได้ว่าการจำแนกหน่วยเสียงกลุ่มสระยังขาดความแม่นยำ เมื่อตรวจสอบหาสาเหตุพบว่าเป็นความผิดพลาดในส่วนของการจำแนกวรรณยุกต์ซึ่งอยู่ในกลุ่มสระ และเมื่อทดสอบด้วยการรู้จำเสียงพูด ความแม่นยำลดลงเหลือเพียง 27.6% แสดงถึงประสิทธิภาพโดยรวมที่ต้องปรับปรุง สาเหตุที่สำคัญของปัญหาดังกล่าวเนื่องมาจากฐานข้อมูลเสียงที่มีจำนวนตัวอย่างเสียงน้อยไม่เพียงพอต่อการฝึกฝนระบบรู้จำ สุดท้ายนี้ค่าความแม่นยำที่ได้จากงานวิจัยนี้สามารถนำไปเป็นค่าอ้างอิง สำหรับการพัฒนาระบบรู้จำเสียงพูดต่อไปในอนาคตได้

คำสำคัญ: Speech Recognition, Thai Language Speech Processing, Phonetic Classification

Thesis Title	Automatic Continuous Speech Recognition of Thai Language Using Hidden Markov Model
Author	Mr.Pongtorn Tantanakit
Major Program	Computer Engineering
Academic Year	2004

Abstract

This paper describes the development of a Thai speech recognition system using Hidden Markov Model. The objective is to determine speech modeling techniques and to build phonetic models for Thai speech recognition. Each phone was modeled with a 3-state 8-mixture HMM. There were a total of 170 $C_i+V_t+C_f$ phone units. Speech feature consisted of 12 MFCC coefficients, 12 delta coefficients and 12 delta-delta coefficients. Training speech consisted of 900 sentences spoken by 45 speakers and the testing speech data consisted of 200 sentences spoken by 10 speakers. Classification accuracy of 58.3% was obtained from test data using 170 phonetic units. However when the phones were divided into 3 groups according to their functions, classification accuracy was 75.3% for first consonants, 39.6% for vowels and 77.6% for final consonants. In recognition phase, where segmentation errors, deletion errors, and insertion errors were accounted for, the accuracy was dropped down to 27.6%. Most of errors were resulted from misclassification of tones. Since tones are difficult to classify and sometime are not clearly announced, it is suggested that more training data is needed and better technique for tone classification should be applied. This work can be served as a baseline for other experiments in the future.