

**THAI SYLLABLE SPEECH RECOGNITION USING HIDDEN
MARKOV MODELS**

PORNCHAI PO-ARAMSRI

**A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR
THE DEGREE OF MASTER OF SCIENCE
(COMPUTER SCIENCE)
FACULTY OF GRADUATE STUDIES
MAHIDOL UNIVERSITY
2003**

**ISBN 974-04-3339-1
COPYRIGHT OF MAHIDOL UNIVERSITY**

การรู้จำเสียงพยางค์ภาษาไทยโดยใช้ Hidden Markov Models (THAI SYLLABLE SPEECH RECOGNITION USING HIDDEN MARKOV MODELS)

พรชัย โพธิ์อร่ามศรี 4137587 SCCS/M

วท.ม. (วิทยาการคอมพิวเตอร์)

คณะกรรมการควบคุมวิทยานิพนธ์ : ศุภชัย ตั้งวงศ์สานต์, Ph.D., ชมทิพ พรพนมชัย, Ph.D.

บทคัดย่อ

ภาษาพูดถือได้ว่าเป็นการสื่อสารของมนุษย์ที่เป็นธรรมชาติที่สุดวิธีการหนึ่ง การติดต่อระหว่างมนุษย์กับเครื่องจักรโดยอาศัยเสียง หรือภาษาพูดจึงนับเป็นก้าวที่สำคัญที่นำไปสู่ยุคใหม่แห่งการปฏิสัมพันธ์ ระหว่างมนุษย์กับเครื่องจักร Speech Recognition หรือการรู้จำเสียงพูดเป็นการเปิดโอกาสให้คอมพิวเตอร์สามารถเข้าใจคำพูดของมนุษย์ได้ การรู้จำเสียงพูดสำหรับภาษาอังกฤษนั้นถือว่าประสบความสำเร็จเป็นอย่างดี ทำให้เราสามารถพบเห็นการประยุกต์ใช้งานได้โดยทั่วไปในปัจจุบัน อย่างไรก็ตาม สำหรับการรู้จำเสียงพูดภาษาไทยแล้วยังห่างไกลจากความเป็นจริง และยังไม่เป็นที่ยอมรับโดยทั่วไป

ในงานของเรามีเป้าหมายในการพัฒนาระบบรู้จำเสียงพูดภาษาไทย โดยกำหนดขอบเขตของงานไว้เป็นระบบรู้จำเสียงพยางค์โดด (Isolated syllable) และขึ้นกับผู้พูด (Speaker dependence) เพื่อให้บรรลุวัตถุประสงค์ดังกล่าว เราจึงทำการศึกษา และพัฒนาระบบรู้จำเสียงพูด โดยอาศัยแบบจำลอง Hidden Markov Model (HMM) โดยนำหลักการของ Generalized Mixture Tying HMM (GENONE HMM) มาใช้ นอกเหนือไปจาก Continuous-Density HMM (CDHMM) และ Semi-Continuous HMM (SCHMM) ที่มีใช้อยู่แล้ว เราได้ใช้ Viterbi training procedure ในการปรับพารามิเตอร์ภายใน HMM ซึ่งวิธีการดังกล่าวได้รับการปรับปรุงให้สามารถใช้งานกับ GENONE HMM ได้ นอกจากนี้ Beam pruning technique ยังได้ถูกนำมาปรับใช้เพื่อลด Computational time for recognition ด้วย สุดท้ายเราได้ทำการทดสอบความสามารถของระบบนี้ในการรู้จำเสียงวรรณยุกต์ไทยด้วย สำหรับ Acoustic feature ที่ใช้ประกอบด้วย Mel-Frequency Cepstrum Coefficients (MFCC), ค่าพลังงาน (Energy) และค่า Pitch Frequency ระบบทั้งหมดพัฒนาขึ้นโดยใช้ภาษา C/C++ โดยใช้ Microsoft Visual C++ 6.0 เพื่อใช้งานบนเครื่อง Intel-based ที่ใช้ MS-Windows 2000

ผลการทดสอบความแม่นยำในการรู้จำเสียงพยางค์โดด สามารถทำได้สูงกว่า 98% เวลาในการทดสอบการรู้จำสามารถลดลงได้ถึง 65.4% โดยไม่ลดความแม่นยำ เมื่อใช้ Beam pruning technique ด้วยค่า Beam width ที่เหมาะสม ผลการทดสอบความแม่นยำในการรู้จำเสียงวรรณยุกต์ไทย สามารถทำได้ถึง 97.88%, 97.36%, 98.81%, 90.67% และ 100.0% สำหรับวรรณยุกต์เสียง สามัญ, เอก, โท, ตรี และจัตวา ตามลำดับ สรุปแล้วระบบดังกล่าวนี้เป็นที่น่าพอใจตามวัตถุประสงค์ของงานวิจัย โดยเฉพาะอย่างยิ่งในประเด็นความแม่นยำและความเร็วในการทดสอบรู้จำ

91 หน้า. ISBN 974-04-3339-1

THAI SYLLABLE SPEECH RECOGNITION USING HIDDEN MARKOV MODELS

PORNCHAI PO-ARAMSRI 4137587 SCCS/M

M.Sc. (COMPUTER SCIENCE)

THESIS ADVISORS: SUPACHAI TANGWONGSAN, Ph.D., CHOMTIP PORNANOMCHAI, Ph.D.

ABSTRACT

Spoken language is one of the most intuitive communication methods for humans. Interfacing between human and machine by using speech or spoken language is an important part of the new era of human-computer interfacing. Speech recognition gives human speech understanding ability to a computer. Speech recognition of English language is very successful and could be commonly found in many daily applications, in the present day. However, Thai speech recognition is still far from reality or common acceptance.

Our work is to develop a Thai speech recognition system but limit the problem scope to the speaker dependence and isolated syllable problem. To reach that goal, the Hidden Markov Model was studied and the HMM-based speech recognition system was developed. Moreover, the Generalized Mixture Tying HMM (GENONE HMM), which is a special scheme of HMM, was adopted in addition to the Continuous-Density HMM (CDHMM) and Semi-Continuous HMM (SCHMM). The Viterbi training procedure, which is used for training the HMMs, was also modified for the GENONE HMM. In addition, the Beam pruning technique was adapted and integrated into the system to reduce computational time for recognition. Moreover, the experiment on the Thai tone recognition was also arranged by using the HMM-based speech recognition system. The acoustic features used in our system were the Mel-Frequency Cepstrum Coefficient (MFCC), Energy and Pitch frequency. The system was developed and tested for running on the Intel-based machine running MS Windows 2000. All programs were coded in the C/C++ programming language by using the Microsoft Visual C++ 6.0 as a development tool.

The recognition accuracy of the speaker-dependent Thai isolated syllable speech recognition using HMM was above 98%. By using Beam pruning technique with an appropriate beam width, recognition time was reduced by 65.4% with no additional error. Thai tone recognition using HMM yielded recognition accuracy of 97.88%, 97.36%, 98.81%, 90.67% and 100.0% for MID, LOW, FALLING, HIGH and RISING tones respectively. In conclusion, the system is quite satisfactory to meet the research objectives, especially in terms of recognition accuracy and speed.

KEY WORDS: THAI SPEECH RECOGNITION / HIDDEN MARKOV MODEL / GENERALIZED MIXTURE TYING / GENONE / VITERBI TRAINING / BEAM PRUNING