อัมรินทร์ ดีมะการ 2550: การรู้จำเสียงตัวเลขต่อเนื่องภาษาไทยแบบไม่ขึ้นกับบุคคลโดยใช้ฮิดเดนมาคอร์ฟโมเดล ปริญญาวิศวกรรมศาสตรมหาบัณฑิต (วิศวกรรมคอมพิวเตอร์) สาขาวิชาวิศวกรรมคอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์ ประธานกรรมการที่ปรึกษา: รองศาสตราจารย์อัศนีย์ ก่อตระกูล, D.Eng. 65 หน้า

การรู้จำเสียงพูดตัวเลขแบบต่อเนื่องเป็นงานวิจัยที่สามารถประยุกต์ใช้ได้หลากหลาย กับงานที่เกี่ยวข้องกับการกรอกข้อมูลตัวเลข เช่น การหมุนเลขโทรศัพท์ด้วยเสียงพูด ปัญหาที่สำคัญของเสียงพูดแบบต่อเนื่องคือขอบเขตของคำพูดในแต่ละเสียงพูด นอกจากนี้การรู้จำเสียงพูดตัวเลขแบบต่อเนื่องยังไม่สามารถใช้หลักไวยากรณ์ทางภาษาในการช่วยเพิ่มความถูกต้องของระบบเหมือนในระบบอื่น เช่น ระบบรู้จำเสียงพูดบทสนทนา ระบบรู้จำเสียงที่สามารถรู้จำได้หลายบุคคลเป็นระบบที่สะดวกต่อการใช้งานมากกว่าแบบขึ้นกับบุคคล งานวิจัยนี้จึงเสนอวิธีการรู้จำเสียงตัวเลขต่อเนื่องภาษาไทยแบบไม่ขึ้นกับบุคคล

ระบบรู้จำเสียงพูดตัวเลขต่อเนื่องภาษาไทยที่ไม่ขึ้นกับบุคคลนี้ประกอบด้วย 5 ส่วน คือ 1) การประมวลผลสัญญาณเสียงพูดเบื้องต้นซึ่งทำหน้าที่ในการวางกรอบสัญญาณเพื่อเตรียมข้อมูลเสียงให้เหมาะสม 2) การดึงค่าลักษณะสำคัญของเสียง ซึ่งทำหน้าที่หาลักษณะสำคัญที่สามารถใช้เป็นตัวแทนของเสียงพูด 3) การฝึกฝนรูปแบบที่ทำหน้าที่ในการสอนระบบให้เรียนรู้ลักษณะสำคัญของเสียงพูดแต่ละตัวเลข 4) การจำแนกรูปแบบสำหรับวิเคราะห์เปรียบเทียบค่าคุณลักษณะสำคัญว่าเป็นตัวเลขใด และ 5) การหาคำตอบ โดยหาลำดับที่ดีที่สุดเพื่อเป็นคำตอบ ในงานวิจัยนี้จะเน้น 3 ขั้นตอนคือ ขั้นตอนที่ 2, 3 และ 4 ในส่วนของขั้นตอนที่ 2 การดึงค่าคุณลักษณะสำคัญของเสียงที่นำมาใช้ในระบบประกอบด้วย 6 คุณลักษณะได้แก่ ค่าสัมประสิทธิ์เซปสตรัมบนสเกลเมล ผลต่างทางเวลาของสัมประสิทธิ์เซปสตรัมบนสเกลเมลในลำดับที่หนึ่ง ผลต่างทางเวลาของสัมประสิทธิ์เซปสตรัมบนสเกลเมลในลำดับที่สอง ค่าพลังงานเสียง ค่าผลต่างพลังงาน และ ค่าผลต่างพลังงานลำดับที่สอง โดยจะทดลองใช้เริ่มจาก 2 คุณลักษณะไปจนถึง 6 คุณลักษณะเพื่อหาจำนวนค่าคุณลักษณะสำคัญที่เหมาะสมต่อการรู้จำเสียง ในขั้นตอนที่ 3 และ 4 การฝึกฝนรูปแบบและจำแนกรูปแบบ ใช้ทฤษฎีแบบจำลองฮิดเดนมาร์คอฟแบบต่อเนื่อง พารามิเตอร์ที่สำคัญของแบบจำลองนี้คือจำนวนสถานะของแบบจำลองและจำนวนของกลุ่มเกาส์เซียนที่เลือกใช้ในแบบจำลอง โดยในส่วนนี้มีการเปลี่ยนค่าจำนวนสถานะและกลุ่มเกาส์เซียน ด้วยการปรับค่าพารามิเตอร์เพื่อหาคำตอบที่เหมาะสมใช้ค่าสถานะเริ่มต้นที่ 5 และ ค่าเกาส์เซียนเริ่มที่ 2

ผลการทดลองประสิทธิภาพของระบบ ใช้ชุดข้อมูลฝึกฝนเสียงพูดจากผู้พูดจำนวน 100 คน แบ่งเป็นชาย 50 คนและหญิง 50 คน รวม 2000 เสียง แต่ละเสียงยาว 7 คำเรียงติดกัน ข้อมูลทดสอบประกอบด้วย 40 คน แบ่งเป็น ชาย 20 และหญิง 20 คน พบว่าจำนวนคุณลักษณะที่ทำให้อัตรารู้จำสูงสุดในการทดลองนี้คือใช้ 5 คุณลักษณะ(จำนวนรวม 41 ค่า) และ แบบจำลองเสียงพูดที่ให้ประสิทธิภาพดีที่สุดคือ แบบจำลองที่มีจำนวน 9 สถานะ 8 กลุ่มเกาส์เซียน อัตราการรู้จำเสียงพูดเฉลี่ยเท่ากับ 74.25%

Amarin Deemagarn 2007: Speaker-Independent Thai Connected Digit Speech Recognition System using Hidden Markov Model. Master of Engineering (Computer Engineering), Major Field: Computer Engineering, Department of Computer Engineering. Thesis Advisor: Associate Professor Asanee Kawtrakul, D.Eng. 65 pages.

Connected digit speech recognition has an important role to apply in many speech applications such as voice- telephone dialing. The major problem in connected digit speech recognition system is to detect the boundary of each word in digit string. Moreover, the language model can not be used to enhance the system precision, like the other speech recognition systems, such as dialogue speech recognition system. Nevertheless the speaker-independent speech recognition system is more practical than the speaker-dependent one. Thus this research presents the Speaker-Independent Thai Connected Digit Speech Recognition using Hidden Markov Model.

The system consists of five steps, which are speech signal preprocessing for preparing speech signal into speech frame, speech feature extraction for distilling the necessary information of speech and representing as the speech feature parameters, pattern training for learning speech feature in each digit, pattern classification for comparing the similarity between reference models and speech input, and decoding process for finding the best sequence result. This research focuses on speech feature extraction, pattern training and pattern classification. Speech feature extraction consists of 6 features, which are the Mel Frequency Cepstral Coefficients (MFCC), delta MFCC, delta-delta MFCC, energy, delta energy and delta-delta energy.
To find appropriate speech features, we varied feature parameters from 2 to 6 features. In pattern training and pattern classification, the Continuous Density Hidden Markov Model (CDHMM) is used to train the speech patterns and to recognize the speech input. Finally the Viterbi search algorithm is applied for decoding process.
The important parameters of CDHMM are the number of states and the number of Gaussian mixtures.
In this research we adjusted these two parameters to find an appropriate value. The experimental number of states for CDHMM and the number of Gaussian mixtures are starting from 5 and 2 respectively.

For the experiment, there are two sets of samples; training and testing sets. In training set, we use data from 100 speakers (50 females, 50 males) for 2000 utterances. Each utterance is seven digit lengths. In testing set, 40 speakers (20 females, 20 males) are used. The best model is the model that uses 5 mains of speech feature (41 values) with 9 states and 8 Gaussian mixtures. The average word accuracy is 74.25%.