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%

อัมรินทร์ ดีมะการ — อัศนีย์ ก่อตระกูล — 20/07/2550

ลายมือชื่อนิสิต — ลายมือชื่อประธานกรรมการ