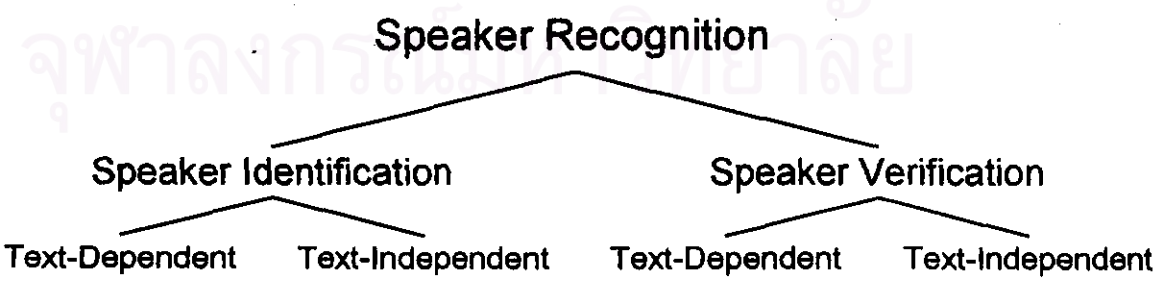
บทที่ 1
บทนำ

ในปัจจุบันการติดต่อสื่อสารเป็นปัจจัยพื้นฐานของการดำรงชีพของมนุษย์ การสื่อสารที่เป็นธรรมชาติที่สุดของมนุษย์ที่ได้ใช้ตลอดมาตั้งแต่ในอดีตจนถึงปัจจุบันก็คือเสียงที่เปล่งออกมานั่นเอง จึงได้มีแนวความคิดว่าเสียงพูดสามารถนำมาใช้บ่งบอกเอกลักษณ์ของแต่ละบุคคล เนื่องมาจากการสังเกตว่ามนุษย์สามารถรู้จำบุคคลที่ใกล้ชิดได้ แม้จะได้ยินแค่เสียงพูดของคนๆ นั้นก็ตาม ทำให้มีการพัฒนาและประยุกต์เทคโนโลยีของกรรมวิธีสัญญาณดิจิตอลมาออกแบบระบบการเข้าถึงข้อมูล หรือระบบรักษาความปลอดภัยเพื่อป้องกันการเข้าสู่ระบบจากผู้บุกรุก ระบบการรู้จำผู้พูด (Speaker Recognition System) ที่ใช้อยู่ในปัจจุบันนั้นสามารถแบ่งออกได้ตามการใช้งานได้เป็นสองระบบใหญ่ๆ (Campbell, 1997) คือ

1. ระบบการบ่งชี้ผู้พูด (Speaker Identification System) คือระบบที่สามารถจำแนกผู้พูดได้ว่าเป็นใคร เมื่อมีเสียงพูดของผู้พูดคนนั้นมากเพียงพอ และผู้พูดคนนั้นต้องเป็นบุคคลที่อยู่ในกลุ่มของบุคคลที่สามารถเข้าสู่ระบบได้เท่านั้น
2. ระบบการตรวจสอบผู้พูด (Speaker Verification System) คือระบบที่ตรวจสอบเสียงของผู้พูดที่เข้าสู่ระบบได้ว่าเป็นเสียงที่อยู่ในกลุ่มที่เราต้องการให้เข้าสู่ระบบได้หรือไม่ แล้วยอมรับ (Accept) หรือปฏิเสธ (Reject) การเข้าสู่ระบบ

นอกจากนี้ระบบการบ่งชี้ผู้พูดอาจแบ่งออกได้ตามลักษณะของประโยคหรือวลีที่ใช้ในการเข้าสู่ระบบ คือเป็นระบบที่กำหนดบทคำพูดอันได้แก่ประโยคหรือวลีที่ใช้ในการเข้าสู่ระบบ (Text-Dependent or Fixed-Text) กับระบบที่ไม่กำหนดบทคำพูดที่ใช้ในการเข้าสู่ระบบ (Text-Independent or Free-Text) ดังแสดงในรูปที่ 1.1 ในระบบ Free-Text นั้นต้องใช้ค่าสถิติของสัญญาณเสียงที่มากเพียงพอในการสกัดลักษณะสำคัญเฉพาะของผู้พูดนั้นๆ จึงต้องใช้เสียงตัวอย่างที่มีความยาว 10-30 วินาทีสำหรับการฝึกฝนและ 5-10 วินาทีในการจำแนกเสียง ส่วนระบบ Fixed-Text ต้องการความยาวของเสียงตัวอย่าง 2-3 วินาทีในการฝึกฝนและแยกแยะเสียง สรุปได้ว่าระบบ Fixed-Text มีสมรรถนะสูงกว่าระบบ Free-Text (Naik, 1990)



รูปที่ 1.1 การแบ่งระบบการรู้จำผู้พูดตามลักษณะการใช้งาน

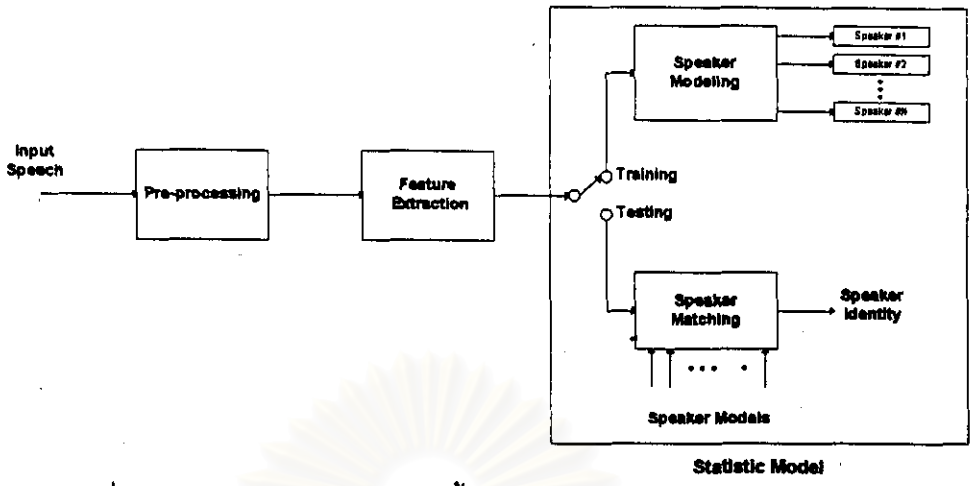
สำหรับการค้นคว้าและวิจัยการรู้จำผู้พูดนั้นในต่างประเทศได้วิจัยมานานแล้ว โดยเริ่มการวิจัยไปพร้อมๆ กับการรู้จำเสียงพูด ซึ่งเห็นได้จากบทความวิจัยที่ตีพิมพ์ในวารสารเชิงวิชาการและอุปกรณ์อิเล็กทรอนิกส์ที่ทันสมัยต่างๆ ที่รวมประยุกต์การใช้งานของการรู้จำเสียงกับการรู้จำผู้พูดได้อย่างดี เช่น โทรศัพท์มือถือที่มีการรู้จำชื่อบุคคลและจดจำเสียงของเจ้าของเครื่องได้ ระบบป้องกันการเข้าออกอาคารโดยใช้เสียงพูด ระบบธนาคารอิเล็กทรอนิกส์ เป็นต้น ส่วนการค้นคว้าและวิจัยการรู้จำผู้พูดในประเทศไทยนั้นเพิ่งเริ่มต้นได้ไม่นานนัก โดยในช่วงเริ่มแรกเป็นการวิจัยทางด้านภาษาศาสตร์ที่มุ่งค้นคว้าและวิเคราะห์ลักษณะการออกเสียงสระภาษาไทยบางสระของผู้พูดแต่ละคน โดยใช้เครื่องวิเคราะห์สเปกตรัม (Spectrogram) ในการวิเคราะห์เสียงพูดและใช้การรับรู้ทางการฟังของมนุษย์เป็นเครื่องมือในการบ่งชี้ผู้พูด (วิสิทธิ์ ลีลาศิริวงศ์, 2535) ต่อมาได้มีการวิเคราะห์การออกเสียงพูดของผู้พูดแต่ละคนแต่ไม่ได้บ่งชี้ผู้พูด และงานวิจัยล่าสุดเป็นการบ่งชี้ผู้พูดจากเสียงตัวเลขโดดภาษาไทย โดยใช้การเปรียบเทียบทางเวลาแบบพลวัต (Dynamic Time Warping, DTW) (คณะนักวิจัยและพัฒนาระบบระบุผู้พูดสำหรับภาษาไทย, 2541) ในงานวิจัยนี้ได้วิจัยเกี่ยวกับการบ่งชี้ผู้พูดแบบขึ้นกับบทคำพูด (Text-Dependent Speaker Identification) โดยวิเคราะห์สัญญาณเสียงพูดและลักษณะการพูดของผู้พูดแต่ละบุคคลจากบทคำพูดที่เป็นกำหนดไว้ ซึ่งบทคำพูดที่ใช้ในงานวิจัยนี้เป็นคำพูดภาษาไทยแบบต่อเนื่อง (Thai Continuous Speech)

1.1 วัตถุประสงค์

1. เพื่อศึกษาโครงสร้างและการออกเสียงภาษาไทยของผู้พูดแต่ละคน
2. เพื่อศึกษาและหาลักษณะสำคัญที่บ่งชี้ผู้พูดแต่ละคนจากบทคำพูดภาษาไทยที่กำหนดไว้ตามแนวทางภาษาศาสตร์
3. เพื่อพัฒนากรรมวิธีการบ่งชี้ผู้พูดแบบขึ้นกับบทคำพูดภาษาไทยที่ใช้พูด

1.2 หลักการและเหตุผล

ระบบการบ่งชี้ผู้พูดสามารถแบ่งออกเป็น 3 ส่วนใหญ่ๆ คือการประมวลผลเบื้องต้น การสกัดลักษณะสำคัญ และแบบจำลองทางสถิติ ดังแสดงในรูปที่ 1.2 ซึ่งการประมวลผลเบื้องต้น (Pre-processing) คือการเตรียมข้อมูลสัญญาณเสียงพูดเบื้องต้นก่อนที่จะนำไปสกัดหาค่าลักษณะสำคัญที่ต้องการในขั้นตอนการสกัดลักษณะสำคัญ (Feature Extraction) และนำไปสร้างเป็นแบบจำลองเสียงพูดของผู้พูดแต่ละคนโดยใช้แบบจำลองทางสถิติ (Statistic Model) ส่วนรายละเอียดในส่วนต่างๆ ที่กล่าวมาจะกล่าวถึงโดยละเอียดในบทที่ 2 และ 3 ต่อไป



รูปที่ 1.2 แบบจำลองของระบบการบ่งชี้ผู้พูด (Speaker Identification System)

(Assaleh K.T. and Mammone R.J., 1994)

ในการแยกแยะผู้พูดแต่ละคนของระบบการรู้จำเสียงพูดของผู้พูดจำเป็นที่ต้องเลือกใช้ลักษณะสำคัญที่สามารถแสดงความแตกต่างของผู้พูดได้อย่างชัดเจน ลักษณะสำคัญสามารถแบ่งออกได้หลายชนิดและหลายวิธีขึ้นอยู่กับลักษณะในการพิจารณา ซึ่งในที่นี้ยกตัวอย่างสองวิธี (Necioglu, Clements, and Barnwell, 1996; Naik, 1990) คือ

แบ่งตามชนิดของลักษณะสำคัญ

1. ลักษณะสำคัญระดับสูง (High level) หรือเสียงซ้อน (Suprasegmental) คือลักษณะที่เกิดร่วมกับเสียงเรียกที่บ่งชี้ ได้แก่ ภาษาท้องถิ่น ลักษณะการพูด อารมณ์ในขณะที่พูด และสภาพแวดล้อม เป็นต้น
2. ลักษณะสำคัญระดับต่ำ (Low level) หรือเสียงเรียก (Segmental) คือเสียงที่ประกอบกันเข้าเป็นคำพูด ได้แก่ ความถี่มูลฐาน (Pitch), ขนาดของสเปกตรัม (Spectral magnitude), ความถี่ฟอร์แมนท์ (Formant frequency), พลังงาน (Energy profile) และอื่นๆ

แบ่งตามการสกัดของลักษณะสำคัญ

1. ลักษณะสำคัญของเส้นเสียง (Glottal Feature) คือลักษณะสำคัญที่สกัดได้จากกระบวนการเกิดของสัญญาณเสียงโดยไม่คำนึงถึงผลกระทบของช่องทางเดินเสียงส่วนบน
2. ลักษณะสำคัญของช่องทางเดินเสียง (Vocal Tract Feature) คือลักษณะสำคัญที่สกัดเอาความถี่เรโซแนนซ์ที่เกิดจากช่องทางเดินเสียง (ความถี่ฟอร์แมนท์)
3. ลักษณะสำคัญของรูปแบบการพูด (Prosodic Feature) คือลักษณะสำคัญที่แสดงถึงแบบรูปการพูดของผู้พูดได้แก่ ความเร็วในการพูด สำเนียงสูงต่ำ Pitch เป็นต้น

การเลือกลักษณะสำคัญที่จะนำมาใช้ในการรู้จำผู้พูดนั้นควรมีคุณสมบัติดังต่อไปนี้

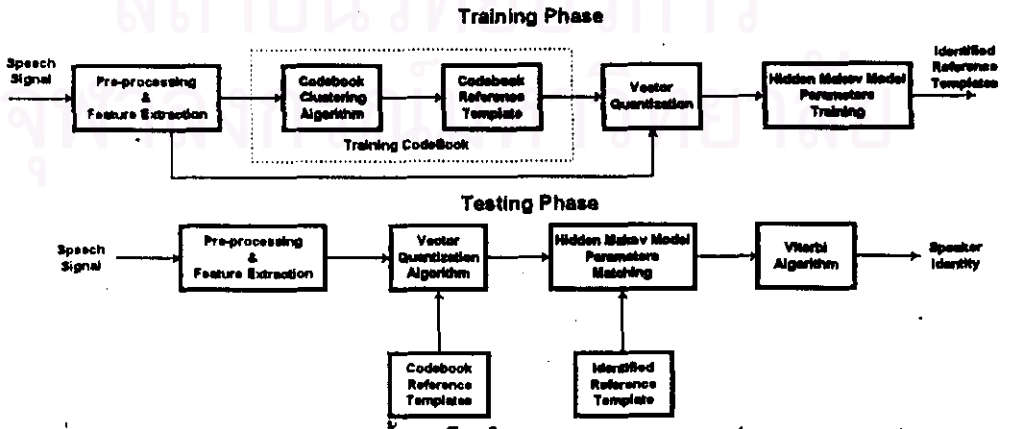
- ความสามารถในการแยกแยะผู้พูดแต่ละคนเมื่อคำนึงถึงความผิดพลาดที่เกิดจากการแปรปรวนของผู้พูดในกลุ่มที่กำลังพิจารณา
- ง่ายต่อการสกัดออกจากสัญญาณเสียง
- มีเสถียรภาพทางเวลา
- ไม่จำลองเสียงของผู้พูดที่ไม่ต้องการ (Impostor)

งานวิจัยนี้ได้นำเสนอลักษณะวิธีการต่างๆ ที่เลือกใช้ในการวิเคราะห์ระบบการบ่งชี้ผู้พูดสามวิธีดังนี้

1. สัมประสิทธิ์การประมาณพันธะเชิงเส้น (Linear Predictive Coefficients)
2. สัมประสิทธิ์เซปสตรอล (Cepstral Coefficients)
3. สัมประสิทธิ์เซปสตรอลบนความถี่เมล (Mel Frequency Cepstral Coefficients)

สำหรับขั้นตอนวิธีการทางสถิติที่นำมาจำลองระบบนั้นได้มีงานวิจัยที่นำเสนออยู่หลายวิธี เช่น การเปรียบเทียบทางเวลาแบบพลวัต (Dynamic Time Warpping, DTW) แบบจำลองฮิดเดนมาร์คอฟ (Hidden Markov Models, HMM) โครงข่ายประสาทเทียม (Neural Network, NN) เป็นต้น ในงานวิจัยนี้จะใช้แบบจำลองฮิดเดนมาร์คอฟแบบดิสครีต (Discrete Hidden Markov Models, DHMM) เนื่องจากเป็นแบบจำลองที่นิยมมาประยุกต์ใช้การรู้จำเสียงพูดภาษาไทย (วิศรุต อาชุบุตร, 2539; เสาวลักษณ์ อารีย์พงศา, 2538) และเป็นแบบจำลองที่เหมาะสมกับการรู้จำเสียงพูดต่อเนื่องเมื่อเปรียบเทียบกับแบบจำลองอื่น (Ahkuputa V., 1998) เนื่องจากแบบจำลองฮิดเดนมาร์คอฟแบบดิสครีตมีลักษณะเฉพาะของข้อมูลขาเข้าเป็นแบบลำดับ (Sequence) ทำให้สามารถรับข้อมูลสัญญาณเสียงพูดแบบต่อเนื่อง (Continuous Speech) ได้ดีกว่าแบบจำลองทางสถิติอื่นที่มีการทำนอร์มอลไลซ์ทางเวลา (Time Normalization) กับข้อมูลก่อนนำเข้าสู่แบบจำลองเพราะทำให้มีการสูญเสียของข้อมูลเกิดขึ้นระหว่างการทำนอร์มอลไลซ์ทางเวลา

สำหรับรายละเอียดและขั้นตอนวิธีการของแบบจำลองฮิดเดนมาร์คอฟแบบดิสครีตกล่าวในบทที่ 2 ส่วนขั้นตอนและวิธีการในการฝึกฝนและทดสอบระบบการบ่งชี้ผู้พูดดังแสดงในรูปที่ 1.3 จะกล่าวในบทที่ 3



รูปที่ 1.3 แผนภาพของระบบการบ่งชี้ผู้พูดโดยใช้แบบจำลองฮิดเดนมาร์คอฟแบบดิสครีต

1.3 ปัญหาของการบ่งชี้ผู้พูด

1. ความไม่คงที่ของลักษณะการพูดของผู้พูดแต่ละคน
2. จำนวนของข้อมูลเสียงของผู้พูดไม่เพียงพอ
3. ไม่สามารถหาลักษณะสำคัญที่บ่งชี้ผู้พูดแต่ละคนได้อย่างชัดเจน

1.4 เป้าหมายและขอบเขตของงานวิจัย

1. พัฒนากรรมวิธีการบ่งชี้ผู้พูดโดยใช้เสียงพูดจากประโยคภาษาไทยที่กำหนดไว้
2. บ่งชี้ผู้พูดจำนวน 12 คน เป็นเพศชาย 6 คนและเพศหญิง 6 คน โดยที่ผู้พูดคนนั้นต้องเป็นบุคคลที่มีอยู่ในฐานข้อมูลของระบบ และมีอัตราการบ่งชี้ผู้พูดประมาณร้อยละ 90

1.5 ขั้นตอนและวิธีการดำเนินการ

1. ศึกษาระบบการบ่งชี้ผู้พูด ขั้นตอนวิธีการและผลงานการวิจัยที่เกี่ยวข้องที่ได้ทำมาแล้วในอดีต
2. ศึกษาและค้นคว้าหาลักษณะสำคัญของเสียงจากผู้พูดแต่ละคน
3. เก็บรวบรวมข้อมูลเสียงของผู้พูดแต่ละคนเพื่อนำมาวิเคราะห์
4. วิเคราะห์ผลลักษณะสำคัญของเสียงผู้พูดแต่ละคนเพื่อหาสิ่งที่บ่งชี้ผู้พูด
5. รวบรวมลักษณะสำคัญที่จะใช้ในการรู้จำมาประยุกต์กับระบบที่ใช้
6. สร้างระบบการบ่งชี้ผู้พูด
7. วิเคราะห์ผลที่ได้จากระบบการบ่งชี้ผู้พูดและปรับปรุงแก้ไข
8. สรุปและรวบรวมผลงานวิจัยทั้งหมดพร้อมจัดทำวิทยานิพนธ์

1.6 ประโยชน์ที่คาดว่าจะได้รับ

1. เป็นพื้นฐานในการทำวิจัยระบบการรู้จำผู้พูดแบบอัตโนมัติที่ใช้กับภาษาไทย
2. สามารถรู้จำเสียงของผู้พูดที่อยู่ในระบบที่ฝึกฝนแล้ว
3. สามารถนำไปประยุกต์ใช้กับระบบรักษาความปลอดภัย