# บทที่ 1
# บทนำ

## 1.1 ความเป็นมา

เทคโนโลยีการรู้จำเสียงพูดในปัจจุบันได้รับการพัฒนาอย่างกว้างขวาง (Furui,1985) ระบบรู้จำต่าง ๆ ได้เกิดขึ้นมากมายหลายวิธี แต่ละวิธีมีข้อดีข้อเสียแตกต่างกัน ระบบรู้จำโดยทั่วไปจำเป็นต้องใช้การคำนวณที่ซับซ้อน ดังนั้นตัวประมวลผลที่สามารถใช้ในงานรู้จำได้จำเป็นต้องมีความสามารถในการคำนวณสูง ในงานประยุกต์ที่ต้องการการรู้จำคำศัพท์จำนวนไม่มาก จึงไม่จำเป็นต้องใช้ระบบรู้จำที่มีความสามารถสูงแต่ต้องใช้การคำนวณสูงด้วย ยิ่งลดการคำนวณได้มากเท่าใด งานประยุกต์ขนาดเล็กยิ่งเป็นไปได้มากขึ้น

ในวิทยานิพนธ์นี้ เป้าหมายจึงมิใช่สร้างระบบรู้จำประสิทธิภาพสูง แต่เป้าหมายคือการสร้างระบบรู้จำที่มีการคำนวณต่ำ แต่สามารถนำไปใช้งานได้ดีในระดับหนึ่ง การคำนวณในระบบรู้จำแบ่งได้เป็น 2 ส่วน คือการคำนวณเพื่อดึงลักษณะสำคัญของเสียง (feature extraction) และการคำนวณในส่วนการรู้จำ ถ้าดึงลักษณะสำคัญของเสียงได้ดี จะช่วยลดภาระในการคำนวณในส่วนการรู้จำด้วย

ลักษณะสำคัญของเสียงที่เป็นที่นิยมได้แก่ สัมประสิทธิ์ของการประมาณพันธะเชิงเส้น Linear Prediction Coefficient (LPC), พลังงาน, คาบการสั่นของเสียง (Pitch Period) เป็นต้น (Rabiner and Juang,1989) การรู้จำโดยทั่วไปจึงต้องคำนวณลักษณะสำคัญดังกล่าวจากเสียงพูด แล้วผ่านระบบรู้จำ สำหรับมาตรฐานการเข้ารหัส G.729 เป็นมาตรฐานการเข้ารหัสเสียงพูดซึ่งสามารถใช้งานได้อย่างกว้างขวาง (GerHard,1997;Benyassine and Shlomot,1997) ดังนั้นฮาร์ดแวร์ราคาถูกที่รองรับมาตรฐานดังกล่าวควรจะมีใช้กันอย่างแพร่หลายในไม่ช้า และมาตรฐานการเข้ารหัสอื่น ๆ ที่ตามมาในอนาคตจะใช้หลักการการเข้ารหัสคล้ายคลึงกับการเข้ารหัส G.729 ซึ่งจำลองการสร้างเสียงของมนุษย์ ข้อมูลที่ได้รับจากการเข้ารหัสควรจะนำมาใช้ประโยชน์ในการรู้จำได้โดยตรงไม่จำเป็นต้องถอดรหัสเป็นสัญญาณเสียงก่อนแล้วทำการคำนวณเพื่อดึงลักษณะสำคัญอีก ทำให้การคำนวณในส่วนของการดึงลักษณะสำคัญลดลง ถ้าทำการรู้จำเสียงพูดจากข้อมูลที่ผ่านการเข้ารหัสได้โดยตรงจะช่วยลดการคำนวณได้อย่างมาก

## 1.2 ปัญหาของการรู้จำเสียงพูด

1.2.1 ขั้นตอนเทคนิคการหาจุดสิ้นสุดเสียงพูด (Endpoint Detection) เป็นขั้นตอนสำคัญขั้นตอนหนึ่งในการรู้จำเสียงพูด จุดสิ้นสุดเสียงพูดที่เหมาะสมจะช่วยให้อัตราการรู้จำเพิ่มสูงขึ้น ลักษณะสำคัญของเสียงที่ใช้หาจุดสิ้นสุดเสียงพูด ได้แก่ พลังงาน , คาบการสั่นของเสียง (pitch period) เป็นต้น (Evangelos and Nikos,1991) ในการเข้ารหัสของมาตรฐาน G.729 จะได้ คาบ

การสั่นของเสียง โดยตรง ไม่ต้องผ่านการคำนวณ แต่ในการเข้ารหัสดังกล่าว ไม่ให้พลังงานโดยตรง ทำให้ต้องคำนวณพลังงานเองจากรหัส

1.2.2 ขนาดของชุดรหัสที่เหมาะสม จำนวนชุดรหัสที่เหมาะสมกับจำนวนคำศัพท์จะช่วยในการจดจำรูปแบบของเสียงพูดแต่ละเสียงได้อย่างมีประสิทธิภาพ ในการรู้จำเสียงพูดภาษาไทย โดยกำหนดคำศัพท์จำนวนไม่มาก (ไม่เกิน 100 คำ) จำนวนชุดรหัสที่เหมาะสมคือ 128 – 256 ชุดรหัส(เสาวลักษณ์ อารีย์พงศา,2538) ซึ่งจำนวนชุดรหัสของมาตรฐาน G.729 ที่ใช้ในส่วนของลักษณะสำคัญเชิงความถี่คือ 128 ชุดรหัส

1.2.3 จำนวนสถานะของแบบจำลองฮิดเดน มาร์คอฟที่เหมาะสม จำนวนสถานะที่เหมาะสมกับเสียงพูดแต่ละชุดมีผลโดยตรงกับอัตรารู้จำ ดังนั้นต้องเลือกจำนวนสถานะให้เหมาะสม

1.2.4 จำนวนเสียงพูดต้นแบบ (Training Set) ถ้าจำนวนเสียงพูดต้นแบบมีไม่มากพอ จะไม่ครอบคลุมการแปรเปลี่ยนทั้งหมดของเสียงพูดทำให้อัตรารู้จำต่ำ ถ้าจำนวนเสียงพูดต้นแบบมีมากเกิน จะไม่มีผลต่ออัตรารู้จำในส่วนที่เกินนั้นมากนัก อีกทั้งเพิ่มภาระในการคำนวณขณะทำการฝึกฝนด้วย

ในวิทยานิพนธ์นี้ แบ่งออกเป็น 5 บทดังนี้

บทที่ 1 บทนำ กล่าวถึง ความเป็นมา และความสำคัญของปัญหา

บทที่ 2 แนวคิดและทฤษฎีที่เกี่ยวข้อง กล่าวถึง ทฤษฎีเกี่ยวกับการรู้จำเสียงพูด แบบจำลองการสร้างเสียง การดึงลักษณะสำคัญของเสียงที่ใช้ในการรู้จำ และมาตรฐานการเข้ารหัส G.729

บทที่ 3 วิธีดำเนินการวิจัย กล่าวถึง ขั้นตอนการดึงลักษณะสำคัญของเสียงจากการเข้ารหัส G.729 ในด้านความถี่ พลังงาน และ คาบการสั่นของเสียง (pitch period) และนำลักษณะสำคัญดังกล่าวใช้ในการรู้จำในหลายวิธี

บทที่ 4 ผลการวิจัย กล่าวถึง ผลที่ได้จากการนำลักษณะสำคัญของเสียง ผ่านระบบรู้จำ เปรียบเทียบผลที่ได้ในแง่มุมต่าง ๆ เช่น วิธีการใช้ลักษณะสำคัญของเสียง จำนวนชุดรหัสที่ใช้ จำนวนสถานะของแบบจำลองฮิดเดน มาร์คอฟ เป็นต้น

บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ กล่าวถึง ผลสรุปของการวิจัยทั้งหมดและข้อเสนอแนะเพื่อใช้ในงานวิจัยในอนาคต

## 1.3 วัตถุประสงค์ของการวิจัย

สร้างระบบรู้จำเสียงพูดไทยที่ผ่านการเข้ารหัส G.729 โดยตรง ที่ลดการคำนวณในส่วนของการวัดค่าลักษณะสำคัญของเสียงโดยมีอัตราการรู้จำมากกกว่าร้อยละ 85

## 1.4 ขอบเขตของการวิจัย

1. ดึงลักษณะสำคัญของเสียงพูดจากการเข้ารหัส G.729 ในด้านต่าง ๆ โดยมีเงื่อนไขว่าการคำนวณซับซ้อนน้อยที่สุด ได้แก่ คุณสมบัติสำคัญด้านความถี่, คาบการสั่นของเสียง (Pitch Period), พลังงาน เป็นต้น

2. สร้างระบบรู้จำเสียงพูดไทยโดยตรงจากการเข้ารหัส G.729 จากคำพยางค์เดียว 20 คำ และคำพูดตัวเลข 0 ถึง 9

## 1.5 ประโยชน์ที่คาดว่าจะได้รับ

1. ได้ระบบรู้จำที่มีความซับซ้อนต่ำเหมาะกับงานประยุกต์ที่ใช้ตัวประมวลผลประสิทธิภาพต่ำ

2. เมื่อมาตรฐาน G.729 ได้รับความนิยมแพร่หลาย ทำให้ได้คุณลักษณะเด่นเกี่ยวกับเสียงพูดทันที ไม่จำเป็นต้องถอดรหัสเป็นเสียงพูดก่อน แล้วคำนวณหาลักษณะสำคัญของเสียงภายหลัง ช่วยลดการคำนวณได้มาก