# พูดกับคอมพิวเตอร์ใน พ.ศ. 2551: สิ่งที่ควรและไม่ควรคาดหวัง

โดย อ. ดร.อติวงศ์ สุชาโต
ภาควิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

## จากนิยายวิทยาศาสตร์สู่ความเป็นจริง

เรากำลังก้าวเข้าสู่ พ.ศ.2551 ภาพของคนพูดกับเครื่องคอมพิวเตอร์ไม่ว่าจะโดยตรงหรือผ่านทางเครือข่ายโทรศัพท์ไม่ใช่สิ่งที่แปลกตาสำหรับสังคมในปัจจุบันอีกต่อไป จากอดีตที่ภาพเหล่านี้ปรากฏแต่เพียงในนิยายวิทยาศาสตร์ ทุกวันนี้เราสามารถพบส่วนติดต่อผู้ใช้ (User Interface) ของคอมพิวเตอร์ที่สามารถรับรู้หรือแม้แต่เข้าใจความต้องการของผู้ใช้ผ่านภาษาพูดของมนุษย์ในเครื่องคอมพิวเตอร์ส่วนบุคคลที่ใช้ระบบปฏิบัติการ เช่น วินโดวส์วิสตา (Windows Vista) หรือ แมคโอเอส (Mac OS) รุ่นใหม่ๆ หรือแม้แต่การใช้เสียงพูดผ่านโทรศัพท์เพื่อใช้บริการต่างๆ ในเชิงพาณิชย์ โดยที่อีกปลายทางหนึ่งของการเชื่อมต่อไม่ได้มีผู้ให้บริการที่เป็นมนุษย์คอยรับฟังอยู่ ก็สามารถมีให้พบเห็นทั่วไปในประเทศสหรัฐอเมริกา และ ประเทศในแถบยุโรปตะวันตก โดยเฉพาะอย่างยิ่งในธุรกิจเกี่ยวกับสายการบิน การเงินธนาคาร และ โทรคมนาคม ซึ่งเป็นธุรกิจที่แข่งขันกันด้วยศักยภาพในการให้บริการลูกค้าและจำเป็นจะต้องให้บริการลูกค้าผ่านทางศูนย์บริการทางโทรศัพท์อย่างกว้างขวางและหลายแห่งมีปริมาณการใช้งานไม่ต่ำกว่าล้านธุรกรรม (Transaction) ต่อวัน การเพิ่มความสะดวกให้แก่ลูกค้ามากขึ้นในขณะที่ไม่ต้องเพิ่มต้นทุนในส่วนของการจ้างงานและอบรมบุคคลผู้ปฏิบัติหน้าที่เป็นเอเยนท์ (Agent) รับโทรศัพท์ เป็นความคุ้มค่าที่บริษัทเหล่านั้นเล็งเห็น

## ใช้งานจริงไม่ได้หรือ?

อย่างไรก็ตามความพึงพอใจของผู้ใช้ระบบที่มีส่วนติดต่อผู้ใช้ผ่านเสียงพูดนั้นขึ้นอยู่กับอัตราของธุรกรรมที่ถูกกระทำสำเร็จลุล่วงตรงตามที่ผู้ใช้ต้องการ หรือ กล่าวอีกนัยหนึ่งด้วยสำนวนที่เข้าใจง่ายคือ “เมื่อสั่งออกไปแล้ว ระบบจะต้องทำงานได้ตามคำพูดที่สั่งออกไป” หากไม่ได้ผลดังที่ต้องการความสะดวกที่คาดหวังว่าจะได้รับกลับจะกลายเป็นความรำคาญใจ ซึ่งส่งผลให้ผู้ใช้หันไปหาทางเลือกอื่นที่ไม่ใช่เสียงพูด ทำให้สูญเสียการได้รับประโยชน์จากการใช้เสียงพูดไปอย่างน่าเสียดาย ส่วนสำคัญที่ทำให้ธุรกรรมดังกล่าวไม่ถูกกระทำให้ลุล่วงเกิดจากความผิดพลาดในการตีความเสียงพูดที่เป็นคำสั่งจากผู้ใช้ หรือเรียกว่า “ผลการรู้จำที่ผิดพลาด” แน่นอนว่าความบกพร่องดังกล่าวเป็นหน้าที่ของวิศวกรผู้พัฒนาโปรแกรมและนักออกแบบผู้ออกแบบลำดับของการสนทนาและเลือกถ้อยคำที่ใช้ตอบโต้กับผู้ใช้ที่จะต้องดำเนินการแก้ไขปรับปรุง อย่างไรก็ตามในบางกรณีวิธีแก้ไขโดยพัฒนาเทคโนโลยีให้ดีขึ้นเป็นวิธีการที่อาจจะไม่สามารถทำได้หากประสิทธิภาพการรู้จำสูงถึงขีดจำกัดของเทคนิคในปัจจุบันแล้ว ถ้าเป็นเช่นนั้นหมายความว่าการได้รับอรรถประโยชน์จากระบบดังกล่าวไม่สามารถเกิดขึ้นได้จริงใช่หรือไม่

การได้รับอรรถประโยชน์จากระบบที่ประยุกต์ใช้งานส่วนติดต่อผู้ใช้ผ่านเสียงพูดไม่จำเป็นต้องพึ่งพาระบบที่ไร้ข้อบกพร่องเสมอไป หากแต่ผู้ใช้จะได้รับอรรถประโยชน์อย่างเต็มที่เมื่อความสามารถในการรู้จำเสียงพูดของระบบสอดคล้องกับความคาดหวังต่อระบบของผู้ใช้ ความคาดหวังที่ไม่สอดคล้องกับความเป็นจริงทำให้ผู้ใช้ใช้งานระบบในวิธีการที่ไม่ตรงกับวิธีการที่ระบบถูกออกแบบมาให้รองรับ

## พบกันครึ่งทาง

จุดที่ความคาดหวังของผู้ใช้มาบรรจบกับความสามารถของในการรู้จำเสียงพูดระบบจะเกิดขึ้นได้ต้องเริ่มมาจากความเข้าใจเบื้องต้นเกี่ยวกับระบบรู้จำเสียงพูด ประเด็นแรกคือเข้าใจสิ่งที่มันเป็น ผู้ใช้จะต้องคำนึงถึงเสมอว่าการสั่งการเครื่องคอมพิวเตอร์ด้วยเสียงพูดนั้น เป็นเพียงช่องทางในการอินพุทเช่นเดียวกับอุปกรณ์อินพุทอื่นๆ ที่มีการใช้งานแพร่หลายมาก่อน เช่น เมาส์ คีย์บอร์ด ทัชแพด และ จอยสติ๊ก เป็นต้น ถึงแม้ว่าผู้ใช้มักจะมีความคุ้นเคยกับอุปกรณ์เหล่านี้และสามารถควบคุมอุปกรณ์ให้ทำงานได้ดังใจอย่างเป็นธรรมชาติ แต่อย่าลืมว่าในขั้นต้นความสามารถ

ดังกล่าวเกิดจากการยอมรับ ทดลอง และ ฝึกฝนใช้งานอุปกรณ์แต่ละชิ้นในวิธีการถูกออกแบบเอาไว้สำหรับแต่ละอุปกรณ์ การทดลองใช้เพื่อสังเกตลักษณะการออกเสียงที่ส่งผลให้ผลการรู้จำที่ผิดพลาดลดลง และ การฝึกฝนเพื่อให้ออกเสียงในลักษณะดังกล่าวได้ จะทำให้ผู้ใช้ได้รับความสะดวกที่ไม่สามารถได้มาโดยผู้ใช้ที่ไม่เปิดใจรับความไม่สมบูรณ์พร้อมของระบบ แนวคิดเช่นนี้อาจกล่าวได้ว่าเป็นแนวคิดในลักษณะ “พบกันครึ่งทาง” นั่นคือผู้ใช้ไม่เพียงแต่รอคอยให้เทคโนโลยีก้าวหน้าจนสามารถปรับวิธีการในการโต้ตอบกับผู้ใช้ให้เข้ากับมนุษย์อย่างสมบูรณ์ แต่ก็ปรับตัวเองให้เข้ากับเครื่องในขณะเดียวกันเพื่อให้สามารถใช้ประโยชน์จากระบบในปัจจุบัน

กรณีศึกษาเกี่ยวกับการนำนวัตกรรมของส่วนติดต่อผู้ใช้ที่ถูกออกแบบมาให้ปรับตัวให้เข้ากับวิธีการทำงานของมนุษย์ และ ส่วนติดต่อผู้ใช้ที่ลดความซับซ้อนและยืดหยุ่นลงเพื่อลดข้อผิดพลาดในการประมวลผล มาใช้งานในเชิงพาณิชย์จริงๆ ที่น่าสนใจคือ กรณีของวิธีการรับอินพุทโดยการรู้จำลายมือของเครื่องคอมพิวเตอร์พกพา แอปเปิล นิวตัน (Apple Newton) เมื่อปี พ.ศ.2536 และ วิธีการรับอินพุทผ่านสัญกรณ์กราฟิตี (Graffiti®) ของเครื่องพีดีเอ จากบริษัท พาล์ม (Palm Inc.) ซึ่งถูกนำมาใช้หลังจากกรณีของเครื่องแอปเปิล นิวตัน (ดูกรอบ “กรณีศึกษาจากการรู้จำลายมือ)

## การใช้งานในแบบต่างๆ

การประยุกต์ใช้งานส่วนติดต่อผู้ใช้ผ่านเสียงพูดนั้นมีระดับความซับซ้อนของระบบแตกต่างกันออกไป ขึ้นอยู่กับวัตถุประสงค์ในการใช้งาน การแบ่งประเภทของการใช้งานนี้ไม่มีการกำหนดที่ตายตัว ในที่นี้จะขอยกตัวอย่างประเภทของการใช้งานที่พบเห็นกันในปัจจุบัน 3 ประเภท คือ การออกคำสั่งและควบคุม (Command-And-Control) การสอบถามในแวดวงจำกัด (Limited Domain Inquiry) และ การเขียนตามคำบอก (Dictation)

การออกคำสั่งและควบคุม ได้แก่ การใช้งานประเภทที่คำหนึ่งคำที่พูด หรือ ลำดับของคำที่พูดนั้น ถูกนำไปประมวลผลเป็นการทำงานอย่างใดอย่างหนึ่งของคอมพิวเตอร์โดยตรง เช่น การโทรออกด้วยเสียงพูด (Voice Dialing) การกรอกฟอร์มด้วยเสียงพูด การควบคุมเครื่องอิเลกทรอนิคส์ด้วยการพูดคำสั่ง หรือ การเลือกรายการจากเมนูด้วยการอ่านชื่อรายการที่ต้องการนั้น เป็นต้น

การสอบถามในแวดวงจำกัด เป็นการใช้งานที่ระบบถูกออกแบบมาให้รองรับลักษณะประโยค หรือ สำนวนการพูด ที่เป็นไปตามธรรมชาติของมนุษย์หรือรองรับรูปแบบประโยคที่หลากหลายกว่าการจับคู่คำต่อคำในการใช้งานประเภทการออกคำสั่งและควบคุม ระบบจะตีความสิ่งที่ผู้ใช้ต้องการจากคำพูดนั้น และ ตอบสนองกลับไปยังผู้ใช้ โดยที่แวดวง (Domain) ของการสนทนาหรือคำถามนั้นถูกกำหนดไว้อย่างตายตัว หากมีการพูดสิ่งที่ “ไม่เกี่ยวข้อง” กับแวดวงที่กำหนด ระบบมักจะทำงานผิดพลาด แต่หากระบบมีที่การตรวจสอบความเกี่ยวข้องกับแวดวงก็จะสามารถทำการปฏิเสธที่จะตอบสนองต่อคำพูดนั้นได้ ตัวอย่างการใช้งานประเภทนี้ ได้แก่ ระบบเอเยนต์อัจฉริยะที่ให้บริการข้อมูลเฉพาะทางต่างๆ เช่น พยากรณ์อากาศ แหล่งท่องเที่ยว หรือ ข้อมูลการเดินทาง เป็นต้น

การเขียนตามคำบอก เป็นความท้าทายที่ยิ่งใหญ่ของการรู้จำเสียงพูด เนื่องจากการเขียนตามคำบอกมักจะเกี่ยวข้องกับการเกิดขึ้นของคำและลำดับของคำที่เกือบจะไม่จำกัด ยิ่งกว่านั้นอรรถประโยชน์ของการใช้โปรแกรมเขียนตามคำบอกขึ้นกับความถูกต้องของคำที่ปรากฏในเอกสารอย่างชัดเจน ดังนั้นความผิดพลาดในการรู้จำจะส่งผลเสียต่ออรรถประโยชน์ของโปรแกรมเสมอ นั่นคือ คำที่รู้จำได้ผิดไม่ว่าจะเล็กน้อยแค่ไหนก็ถือว่าเป็นการทำให้ความต้องการของผู้ใช้ไม่บรรลุผล ความผิดพลาดในการรู้จำอย่างเล็กน้อยนั้นอาจจะไม่ส่งผลต่ออรรถประโยชน์ของการใช้งานประเภทการสอบถามในแวดวงจำกัด เนื่องจากผลตอบสนองที่ได้รับจากระบบนั้นอาจจะถูกต้องเหมาะสมแม้แต่ในกรณีที่มีการรู้จำบางคำผิดพลาด

การใช้งานในปัจจุบัน มีการจำกัดแวดวงให้แก่การเขียนตามคำบอกเพื่อให้ได้ระบบที่มีประสิทธิภาพในแง่ของความถูกต้องในการรู้จำมากขึ้น แวดวงที่เป็นที่นิยมได้แก่ การเขียนตามคำบอกในวงการกฎหมาย และ การเขียนตามคำบอกในวงการเวชศาสตร์ (Medical Dictation)

ในภาษาที่การรู้จำเสียงพูดมีความเจริญก้าวหน้ามากแล้ว ประเภทของการใช้งานทั้ง 3 จะสามารถพบเห็นได้โดยทั่วไปโดยระบบที่ซับซ้อนน้อยกว่าและเชื่อถือได้มากกว่าจะมีการนำไปใช้ในเชิงพาณิชย์อย่างแพร่หลาย เช่น การ

ออกคำสั่งและควบคุม ระบบที่ซับซ้อนมากกว่าจะพบในห้องวิจัยตามมหาวิทยาลัยและองค์กรทั้งของเอกชนและของรัฐ ส่วนในภาษาไทยนั้นมีการใช้งานประเภทการออกคำสั่งและควบคุมในเชิงพาณิชย์บ้าง อย่างไรก็ตามการเขียนตามคำบอกในภาษาไทยนั้นยังเป็นความท้าทายของนักวิจัยในปัจจุบัน

## คำศัพท์และไวยากรณ์

แม้ธรรมชาติจะสร้างให้มนุษย์สามารถควบคุมการเคลื่อนไหวของมือและนิ้วได้อย่างอิสระ การสื่อความหมายสู่เครื่องคอมพิวเตอร์ผ่านอุปกรณ์เช่นคีย์บอร์ด ก็ต้องทำโดยการเคลื่อนไหวของมือและนิ้วที่กดลงไปยังปุ่มที่ถูกกำหนดไว้ เช่นเดียวกับการสื่อความหมายสู่เครื่องคอมพิวเตอร์ผ่านเสียงพูด การออกเสียงก็จะต้องเป็นไปตามคำศัพท์และไวยากรณ์ (Grammar) ที่มีการกำหนดเอาไว้ ปัญญาประดิษฐ์ในปัจจุบันยังไม่สามารถทำให้เครื่องคอมพิวเตอร์มีวิจารณญาณในการคาดเดาความหมายหรือแม้แต่จะตระหนักถึงคำศัพท์ที่เกิดขึ้นใหม่หรือที่ไม่เคยถูกบันทึกเอาไว้ในคลังศัพท์ (Lexicon) ของระบบได้ การรู้จำเสียงพูดในปัจจุบันจึงเป็นการสกัดลักษณะสำคัญของสัญญาณเสียงเพื่อนำไปค้นหาว่าลักษณะสำคัญของสัญญาณเหล่านั้นตรงกับคำศัพท์ใดบ้างในคลังศัพท์ โดยไวยากรณ์ที่ถูกกำหนดมีหน้าที่สร้างความเป็นไปได้ทั้งหมดของลำดับของคำศัพท์เหล่านั้น ผลของการรู้จำสัญญาณเสียงพูดก็คือ ลำดับของคำศัพท์ในคลังศัพท์ที่ถูกต้องตามไวยากรณ์ของระบบ ซึ่งลักษณะสำคัญของสัญญาณเสียงที่สอดคล้องกับลำดับของคำศัพท์นั้นๆ มีความคล้ายกับสำคัญของสัญญาณเสียงที่ต้องการรู้จำมากที่สุดนั่นเอง

จำนวนคำศัพท์ทั้งหมดในคลังศัพท์ของระบบที่ใช้ในงานต่างๆ กันจะแตกต่างกันไป ตั้งแต่ระดับไม่กี่คำ จนถึงระดับที่มีคำศัพท์ทั้งหมดมากกว่า 50,000 คำ ยกตัวอย่างเช่น ระบบบริการลูกค้าทางโทรศัพท์ที่สามารถรับหมายเลขบัตรเครดิตจากเสียงผู้ใช้ ซึ่งเป็นการใช้งานประเภทการออกคำสั่งและควบคุมนั้น โดยทั่วไปจะมีขนาดของคลังศัพท์อยู่ที่ 10 คำศัพท์ นั่นคือ หมายเลขศูนย์ถึงหมายเลขเก้านั่นเอง ส่วนระบบเขียนตามคำบอกอาจจะมีคำศัพท์เป็นหลักหมื่นอันเป็นคำศัพท์จากพจนานุกรมใหญ่ๆ ในภาษานั้นๆ เป็นต้น

ระบบที่มีจำนวนคำศัพท์เท่ากันอาจจะมีความซับซ้อนที่แตกต่างกันเพราะไวยากรณ์ที่แตกต่างกัน เช่น ระบบที่มีการใช้งานการออกคำสั่งและควบคุมแบบ “พูดหรือกด (Press-Or-Say)” ซึ่งผู้ใช้เลือกรายการในเมนูโดยการกดปุ่มบนแป้นโทรศัพท์หรือพูดหมายเลขปุ่มแทนนั้น หากรายการที่มีให้เลือกมีมากที่สุด 10 รายการ คำศัพท์ในคลังศัพท์ก็อาจจะมี 10 คำศัพท์ช่างก็เป็นหมายเลขเช่นเดียวกับการรับหมายเลขบัตรเครดิต หากแต่ไวยากรณ์ของการใช้งานแบบพูดหรือกดยอมให้มีความเป็นไปได้แค่หมายเลขใดหมายเลขหนึ่ง แต่หมายเลขบัตรเครดิตเกิดจากการประสมกันของตัวเลขทั้งสิบเป็นจำนวน 16 ตัว จึงมีความเป็นไปได้ที่มากกว่า ความซับซ้อนจึงมากกว่า

## ความยืดหยุ่นต่อเอกลักษณ์ของเสียงผู้ใช้ที่แตกต่างกัน

เสียงของมนุษย์มีความเป็นเอกลักษณ์ที่สามารถบ่งบอกตัวตนของผู้เปล่งเสียงได้ เอกลักษณ์ต่างๆ นี้ อาจจะอยู่ในรูปของคุณภาพของเสียง (Voice Quality) ที่แตกต่างกัน คำว่าคุณภาพของเสียงนี้มิได้หมายความถึงเสียงที่ดีหรือไม่ดี เพราะหรือไม่เพราะ หากแต่หมายถึงลักษณะทางอคูสติก (Acoustic) เช่น ความทุ้มความแหลม หรือ ลักษณะของเสียงที่ขึ้นจมูก เป็นต้น นอกจากคุณภาพของเสียงแล้วสไตล์การพูดเฉพาะตัวของแต่ละบุคคลก็มีส่วนทำให้ลักษณะสำคัญที่สกัดมาจากสัญญาณเสียงเพื่อทำการเปรียบเทียบกับลักษณะสำคัญที่สอดคล้องกับคำศัพท์ในคลังศัพท์แตกต่างกันออกไป แม้จะสกัดมาจากสัญญาณของเสียงพูดคำๆ เดียวกัน ความเป็นเอกลักษณ์เฉพาะตัวของเสียงของบุคคลหนึ่งๆ นี้มีคุณค่าในการนำไปประยุกต์ใช้งานประเภทการบ่งชี้ผู้พูด (Speaker Identification) และ การทวนสอบผู้พูด (Speaker Verification) ในขณะที่เอกลักษณ์ดังกล่าวก่อให้เกิดปัญหาในการสร้างระบบที่สามารถรู้จำเสียงพูดของโดยไม่ทราบผู้พูดมาก่อน

ระบบที่ถูกออกแบบมาเพื่อให้ผลการรู้จำมีประสิทธิภาพดีเมื่อผู้ใช้เป็นผู้ใดก็ได้เรียกว่า ระบบแบบไม่ขึ้นกับผู้พูด (Speaker-independent) เนื่องจากระบบรู้จำเสียงพูดจะต้องมีการสร้างแบบจำลองการเกิดลักษณะสำคัญของเสียงที่สอดคล้องกับคำศัพท์เหล่านั้นก่อนที่จะทำการรู้จำเสียงของผู้ใช้ เพื่อให้ระบบรู้จำเสียงพูดแบบไม่ขึ้นกับผู้พูดมีแบบจำลองที่ครอบคลุมลักษณะเสียงของผู้ใช้ทั่วๆ ไป แบบจำลองจำเป็นจะต้องสร้างขึ้นมาจากตัวอย่างเสียงพูดของ

ผู้พูดจำนวนมาก แบบจำลองที่ดีมักจะต้องถูกสร้างขึ้นมาจากตัวอย่างเสียงของผู้พูดจำนวนหลายพันคน โดยการกระจายของจำนวนผู้พูดซึ่งถูกแบ่งตามปัจจัยที่มีผลต่อคุณภาพเสียงและสไตล์การพูดอย่างเด่นชัดควรจะสะท้อนการกระจายของจำนวนผู้ใช้ที่แท้จริงเมื่อระบบถูกนำไปใช้ โดยทั่วไปปัจจัยเหล่านี้รวมไปถึง เพศ ช่วงอายุ และ ถิ่นที่อยู่อาศัย การประยุกต์ใช้ระบบรู้จำเสียงในการให้บริการลูกค้าทางโทรศัพท์มักจะต้องเป็นระบบรู้จำเสียงพูดแบบไม่ขึ้นกับผู้พูด เนื่องจากระบบจะต้องรองรับผู้ใช้ที่มีคุณภาพเสียงและสไตล์การพูดที่แตกต่างกันเป็นจำนวนมาก

ในทางทฤษฎีเทคนิคการรู้จำเสียงพูดในปัจจุบันสามารถใช้ในการสร้างระบบรู้จำเสียงพูดแบบไม่ขึ้นกับผู้พูดที่มีประสิทธิภาพสูงได้ หากมีการจัดเก็บตัวอย่างเสียงที่มีจำนวนมากและมีความแตกต่างที่ครอบคลุมกลุ่มผู้ใช้งานเพียงพอ หากแต่ในทางปฏิบัตินั้นการกระทำดังกล่าวจะสำเร็จได้จำเป็นจะต้องใช้การลงทุนที่เพียงพอ โดยเฉพาะอย่างยิ่งค่าใช้จ่ายในกระบวนการที่ทำเพื่อให้ได้มาซึ่งตัวอย่างเสียงจากผู้พูดหลายพันคนภายใต้ข้อกำหนดของปัจจัยที่เหมาะสม

อย่างไรก็ตามสำหรับระบบที่ถูกออกแบบและสร้างขึ้นมาเพื่อผู้ใช้ที่รู้ตัวตนแน่นอน การสร้างแบบจำลองการเกิดลักษณะสำคัญของเสียงสามารถกระทำได้ง่ายกว่ามาก เนื่องจากต้องการตัวอย่างเสียงเฉพาะของบุคคลนั้น ส่งผลให้ความแปรปรวนของคุณภาพเสียงและสไตล์การพูดไม่สูงเท่ากับในกรณีของระบบแบบไม่ขึ้นกับผู้พูด สำหรับระบบแบบขึ้นกับผู้พูด (Speaker-dependent) เหล่านี้ ในความเป็นจริงผู้ใช้สามารถคาดหวังผลการรู้จำที่มีอัตราความผิดพลาดต่ำได้ ตัวอย่างการใช้งานระบบประเภทนี้เช่น ระบบควบคุมเครื่องมือเครื่องใช้ภายในบ้านด้วยเสียงพูดสำหรับผู้พิการแขนขา เนื่องจากระบบดังกล่าวสามารถออกแบบและสร้างขึ้นมาเพื่อผู้ใช้โดยเฉพาะได้

## ความเหลื่อมล้ำของประสิทธิภาพในภาษาไทยและภาษาตะวันตก

คำถามที่คงจะอยู่ในใจหลายคนคือ ทำไมประสิทธิภาพการรู้จำเสียงพูดภาษาไทยถึงยังไม่ทัดเทียมกับประสิทธิภาพการรู้จำเสียงพูดในภาษาตะวันตก เช่น อังกฤษ เยอรมัน สเปน และ ภาษาเอเชียบางภาษาเช่น ญี่ปุ่น ความก้าวหน้าทางเทคโนโลยีด้านวิทยาศาสตร์คอมพิวเตอร์และการประมวลผลสัญญาณที่สูงกว่าอาจจะเป็นคำตอบหลักของความเหลื่อมล้ำในวิทยาการสาขาอื่นๆ แต่สำหรับปัญหาด้านการรู้จำเสียงพูดปัจจัยที่สำคัญกว่าคือ ประสบการณ์และเงินทุนในการสร้างฐานข้อมูลเสียงที่ยังมีจำกัด นอกจากนั้น ถึงแม้ว่าเทคนิคที่ใช้ในการสร้างแบบจำลองลักษณะสำคัญของสัญญาณเสียงสามารถนำมาประยุกต์ใช้กับภาษาใดก็ได้อย่างไม่ยากลำบาก แต่การเลือกค่าตัวแปรในการสร้างแบบจำลองที่เหมาะสมนั้นแตกต่างกันไปตามธรรมชาติของการพูดภาษาหนึ่งๆ ตัวแปรเหล่านี้มิใช่เป็นเพียงค่าเชิงตัวเลขที่สามารถทำการทดลองทางวิทยาศาสตร์เพื่อหาค่าเหมาะสมที่สุดได้ ค่าที่เหมาะสมดังกล่าวจะต้องได้มาจากผลของการศึกษาทางภาษาศาสตร์ของภาษานั้นๆ ตัวอย่างของตัวแปรเหล่านี้ ได้แก่ ชุดของหน่วยเสียงที่ครอบคลุมเสียงทั้งหมดในภาษา การจัดกลุ่มหน่วยเสียงตามสมบัติทางสวนสัทศาสตร์ (Acoustic-phonetics) เพื่อสร้างแบบจำลอง เป็นต้น

## หลีกเลี่ยงเสียงรบกวน

ผู้ใช้ไม่ควรคาดหวังผลการรู้จำเสียงพูดที่ดีหากเสียงพูดนั้นยังไม่ชัดพอสำหรับผู้ฟังที่เป็นมนุษย์ การใช้ส่วนติดต่อผู้ใช้ผ่านเสียงพูดควรจะทำในสภาวะที่ “เงียบ” ซึ่งสภาวะดังกล่าวอาจจะไม่จำเป็นต้องเป็นสภาวะที่ไม่มีเสียงใดๆ นอกจากเสียงพูดที่ต้องการรู้จำ เทคนิคการรู้จำเสียงพูดในปัจจุบันนั้นสามารถรองรับการรู้จำเสียงพูดที่ปะปนอยู่ในเสียงรบกวนที่มีอัตราส่วนสัญญาณต่อเสียงรบกวนต่ำได้ในระดับหนึ่ง เมื่อเสียงรบกวนเหล่านี้เป็นเสียงที่มีสมบัติคงที่ทางสถิติ ดังเช่น เสียงรบกวนโดยรอบ (Ambience Noise) จากเครื่องปรับอากาศ เป็นต้น อย่างไรก็ตามการรู้จำเสียงพูดให้มีประสิทธิภาพที่ดีในสภาวะที่เสียงรบกวนมากยังคงเป็นปัญหาที่ท้าทายนักวิจัยและพัฒนาในปัจจุบันและน่าจะยังคงเป็นปัญหาหลักในอนาคตอันใกล้ด้วยเช่นกัน

## สู่ความคาดหวังที่สอดคล้องกับความเป็นจริง

ประโยชน์ของส่วนติดต่อผู้ใช้ผ่านเสียงพูดนั้นมีอยู่มากมาย ผู้ใช้ไม่ควรปฏิเสธเทคโนโลยีเพียงเพราะความสามารถของมันไม่ตรงกับความคาดหวังที่ไม่ใกล้เคียงโลกแห่งความเป็นจริง หากเปิดใจให้กับเทคโนโลยีก็จะพบว่าแม้จะมีข้อบกพร่องเทคโนโลยีการรู้จำเสียงพูดสามารถก่อให้เกิดวิธีการทำงานใหม่ๆ อันจะนำมาสู่การทำงานที่มีประสิทธิภาพมากขึ้น เพลิดเพลินขึ้น หรือ แม้แต่ให้กำเนิดบางสิ่งที่ไม่สามารถกระทำได้มาก่อน เวลาและความพยายามที่ใช้ในการเรียนรู้ฝึกฝนให้เคยชินกับระบบจะนำไปสู่ความสะดวกและประสิทธิผลในการมำงานที่เพิ่มขึ้นอย่างคุ้มค่า ความเข้าใจในระดับความสามารถ ข้อจำกัดต่างๆ และ ระดับของความซับซ้อนที่แตกต่างกันของการประยุกต์ใช้งานการรู้จำเสียงพูดในปัจจุบัน เหล่านี้เป็นสิ่งที่ผู้ใช้ระบบควรคำนึงถึง และที่สำคัญกว่านั้นคือ เป็นสิ่งที่ผู้เลือกใช้เทคโนโลยีควรคำนึงถึงหากต้องการสร้างธุรกิจที่ประสบความสำเร็จจากการรู้จำเสียงพูด

### กรณีศึกษาจากการรู้จำลายมือ

เมื่อปี พ.ศ. 2536 บริษัทแอปเปิล คอมพิวเตอร์ (Apple Computer) ได้นำเครื่องพีดีเอ (Personal Digital Assistant) รุ่นแอปเปิล นิวตัน ออกวางจำหน่าย โดยเครื่องแอปเปิลนิวตันนี้เป็นพีดีเอรุ่นแรกที่มีความสามารถรู้จำลายมือของผู้ใช้ ความสามารถนี้นับเป็นนวัตกรรมที่สำคัญของการป้อนข้อมูล อย่างไรก็ตามความถูกต้องของการรู้จำลายมือไม่สูงเท่าที่คนคาดหวัง หลายคนวิจารณ์ระบบรู้จำลายมือนี้ในแง่ลบ สิ่งนี้ทำให้ชื่อเสียงของแอปเปิล นิวตันไม่ดีนัก จนทำให้แอปเปิลเลิกผลิตเครื่องรุ่นนี้ไป

หลังจากนั้นไม่นานบริษัทพาล์ม (Palm Inc.) ได้เปิดตัวพีดีเอที่เรียกว่า พาล์ม ไพลอต (Palm Pilot) ซึ่งมีความสามารถในการรับข้อมูลป้อนเข้าจากผู้ใช้ทางการเขียนบนหน้าจอเช่นเดียวกับแอปเปิล นิวตัน หากแต่การเขียนนี้จะต้องอยู่ในรูปแบบการเขียนที่เรียกว่า กราฟฟิตี (Graffiti®) ซึ่งผู้ใช้จะต้องทำความคุ้นเคยกับการเขียนแบบนี้ก่อนที่จะใช้ได้อย่างคล่องแคล่ว เนื่องจากการตรวจสอบการเขียนจากกราฟฟิตี ซึ่งมีการกำหนดลำดับทิศทางการเขียนแต่ละเส้นของตัวอักขระอย่างชัดเจนนั้น สามารถทำได้ง่ายกว่าระบบรู้จำลายมือของผู้เขียน ความถูกต้องของการตรวจจับตัวอักขระย่อมสูงกว่า ผลปรากฏว่าพาล์ม ไพลอตเป็นที่นิยม

กรณีของกราฟิตีและการรู้จำลายมือนี้ แสดงให้เห็นว่า ในบางครั้งมนุษย์และคอมพิวเตอร์ก็ควรปรับตัวมาพบกันครึ่งทาง