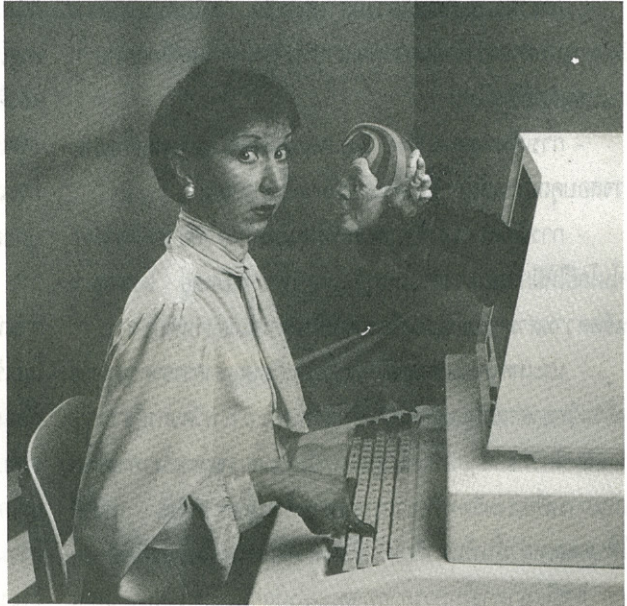




เพื่อ คอมพิวเตอร์ ทำงานตามเสียง



* พนารัตน์ ลิ้ม



“ท่านเคยคิดหรือไม่ว่า คนพิการที่ไม่มีแขนหรือคนตาบอดจะสามารถใช้เครื่องคอมพิวเตอร์ในการทำงานได้”

ท

งวันนี้อาจจะเห็นว่า การบันทึกข้อมูลลงในเครื่องคอมพิวเตอร์จะบันทึกโดยผ่านแป้นพิมพ์ ดังนั้นผู้ใช้คอมพิวเตอร์จำเป็นจะต้องนั่งประจำหน้าแป้นพิมพ์ติด ตามองจอภาพ

ในอนาคต ภาพเหล่านี้จะเปลี่ยนไป เมื่อมีผู้พยายามคิดค้นการบันทึกข้อมูลโดยใช้เสียงแทนการใช้แป้นพิมพ์ติด เนื่องจากการใช้เสียงในการถ่ายทอดข้อมูลเข้าไปในเครื่องคอมพิวเตอร์จะสามารถถ่ายทอดข้อมูลได้เร็วในอัตรา 200 คำต่อนาที เมื่อเทียบกับการพิมพ์ที่น้อยคนนักจะพิมพ์ได้ในอัตราที่มากกว่า 60 คำต่อนาที นอกจากนี้ การทำงานก็สามารถทำได้แม้ในที่มืด หรือในขณะที่มือไม่ว่างโดยที่ตาไม่ต้องจับจ้องอยู่เฉพาะจอภาพ หรือในสถานการณ์ที่ต้องการความเร่งด่วน การใช้พูดข้อมูลสู่คอมพิวเตอร์ จะใช้ได้รวดเร็วกว่าที่จะนั่งพิมพ์

เทคโนโลยีด้านการจำคำพูด (Speech recognition)

เป็นปัญญาประดิษฐ์ (Artificial intelligence) อย่างหนึ่งที่เกี่ยวข้องกับกระบวนการทางภาษา (Natural language processing) ที่คอมพิวเตอร์สามารถจำ รู้จักภาษามนุษย์ได้ ซึ่งจะทำให้การสื่อสารระหว่างคนกับเครื่องจักรเปลี่ยนไป คอมพิวเตอร์จะมีขีดความสามารถในการรองรับงานของทุก ๆ คนได้ และช่วยให้ผู้ใช้คอมพิวเตอร์อยู่เสมอสามารถทำงานได้อย่างมีประสิทธิภาพ

ในระบบการจำเสียง (Voice recognition system)

นั่น ท่านสามารถบอกคำสั่งและข้อมูลโดยตรงกับคอมพิวเตอร์ เช่น ปัจจุบัน ถ้าท่านต้องการเพิ่มข้อมูลหนึ่งจาก Drive B ท่านต้องพิมพ์คำสั่งลงไป แต่ถ้าเป็นคอมพิวเตอร์ระบบ Voice recognition ท่านจะได้ผลลัพธ์เช่นเดียวกัน เพียงแต่ท่านพูดผ่านไมโครโฟนของเครื่องคอมพิวเตอร์

* หัวหน้าภาควิชานิเทศศาสตร์ธุรกิจ คณะนิเทศศาสตร์ มหาวิทยาลัยธุรกิจบัณฑิตย: M.S. (MANAGEMENT INFORMATION SYSTEM) UNITED STATES INTERNATIONAL UNIVERSITY, U.S.A.



Speech recognition

จะรับข้อมูลที่เป็นคลื่นเสียง

กลั่นกรองและเปลี่ยน

เป็นสัญญาณที่ใช้ในคอมพิวเตอร์

เพื่อนำไปเปรียบเทียบกับ

ตัวแม่แบบคำศัพท์ไวยากรณ์

ซึ่งได้เก็บไว้ในหน่วยความจำ



แม้ว่าปัจจุบันระบบการจำเสียงจะมีขีดจำกัดในการทำงาน แต่ยังเป็นประโยชน์มากสำหรับงานบางอย่าง ในโรงงานที่ทำงานด้วยเครื่องจักร พนักงานสามารถที่จะควบคุมเครื่องจักรได้โดยการพูดบอกคำสั่ง โดยไม่ต้องกังวลที่จะต้องคอยจับบังคับเครื่องตลอดเวลา

สายการบินบางแห่งใช้เสียงในการควบคุมระบบการหากระเป๋าดำเนินทางของผู้โดยสารเมื่อถึงปลายทาง โดยการพูดบอกชื่อสถานีปลายทางผ่านไมโครโฟน คอมพิวเตอร์จะทำการตรวจเช็ค และจัดลำเลียงกระเป๋าออกมา

ในการควบคุมสินค้าคงเหลือก็เช่นเดียวกัน พนักงานจะต้องใช้มือและตาในการตรวจนับ แต่ถ้าใช้การพูดคำสั่งจะทำให้การทำงานเร็วขึ้น นอกจากนี้ในระบบการจำคำพูดสามารถนำมาใช้เป็นประโยชน์กับการจองตั๋วเครื่องบิน หอพักโรงแรม หรือแม้กระทั่งการสอบถามกำหนดการต่าง ๆ ได้โดยตรง

จวบจนปัจจุบันจะเห็นได้ว่า ระบบการจำคำพูดยังไม่แพร่หลายมากนักกับเครื่องคอมพิวเตอร์ เพราะยังมีข้อจำกัดบางอย่าง เช่น ความเชื่อถือได้ยังไม่ดีนัก และราคาแพง

หน่วยจำคำพูด (Speech recognizers) แบ่งออกได้หลายประเภท

1. หน่วยจำคำพูดเป็นคำ ๆ (Word recognizers)
2. หน่วยจำคำพูดที่เป็นประโยคต่อเนื่อง (Continuous speech recognizers)
3. หน่วยจำเสียงผู้พูดเฉพาะราย (Speaker dependent)
4. หน่วยจำเสียงผู้พูดทั่วไป (Speaker independent)

หน่วยจำคำพูดเป็นคำ ๆ (Word recognizers)

หน่วยจำคำพูดเป็นระบบที่จำคำพูดเป็นคำ ๆ แยกจากกัน ระบบนี้จะสามารถทำได้เฉพาะคำศัพท์สั้น ๆ หรือวลีง่าย ๆ ในการให้คำสั่งจะต้องเป็นข้อมูลที่ชัดเจน มีการเว้นระยะของคำ

การสร้าง hardware และ software ที่สามารถจำคำพูดของมนุษย์ได้นั้นเป็นเรื่องยาก ปัจจุบันในทางปฏิบัติแล้ว หน่วยจำคำพูดมักจะเป็นแบบจำคำพูดเป็นคำ ๆ

หน่วยจำคำพูดที่เป็นประโยคต่อเนื่อง (Continuous speech recognizers)

หน่วยจำคำพูดประเภทนี้จะจำคำพูดที่มีลักษณะต่อเนื่อง ท่าน

สามารถพูดเป็นประโยค โดยหน่วยคำพูดสามารถที่จะจำและเข้าใจ ข้อมูลที่ใส่เข้าไป แต่อย่างไรก็ตามเป็นการยากที่จะสร้างหน่วยคำพูดแบบต่อเนื่องนี้ เพราะความยากในการแบ่งคำในประโยค หรือวลีที่ต่อเนื่องออกจากกัน ในขณะที่ผู้พูดพูดเป็นประโยคติดต่อกันเป็นการยากที่ระบบจะรู้ว่าแต่ละคำสิ้นสุดที่ตรงไหน และเริ่มคำใหม่ตรงไหน ซึ่งจะต้องใช้เทคโนโลยีที่ซับซ้อนในการคิดค้นทำหน่วยคำพูดต่อเนื่องนี้

ระบบนี้มีพื้นฐานการทำงานเหมือนกับ word recognizer ต่างกันตรงที่มีความซับซ้อนมากกว่า ปัญหาหลักของระบบนี้คือความยากที่จะแยกความแตกต่างของคำแต่ละคำ แต่ถ้าหากมีการใส่ข้อมูลเป็นคำ ๆ โดยมีการหยุดเป็นช่วง ๆ ระบบจะสามารถทำงานได้สบาย และมีความน่าเชื่อถือมากขึ้น การหยุดพูดเป็นช่วง ๆ ระยะห่างกัน 50-300 มิลลิเซคเกิล ก็เพียงพอสำหรับระบบในการแยกคำระหว่างคำขึ้นต้นและคำสุดท้าย การทำงานของระบบ continuous speech systems นี้จะใช้วิธีการออกเสียงของตัวอักษรหรือกลุ่มตัวอักษรเพื่อลดความแตกต่างของเสียง

หน่วยจำเสียงผู้พูดเฉพาะราย (Speaker dependent)

ระบบนี้จะทำงานกับเสียงคน ๆ เดียวเท่านั้น ถ้าเสียงพูดแปลกไปจะไม่ยอมรับข้อมูล เนื่องจากคุณภาพเสียงของแต่ละคนต่างกัน จึงเป็นการยากที่จะสร้างวงจรไฟฟ้าและระบบคอมพิวเตอร์เพื่อจำเสียงของคนทุก ๆ คน ระบบนี้จะทำงานง่ายขึ้นและมีความน่าเชื่อถือมากขึ้นถ้ามีการกำหนดเสียงผู้พูดเฉพาะรายลงไปในการบอ้อนข้อมูล

ในระบบ speaker dependent word สิ่งแรกที่คอมพิวเตอร์ถูกกำหนดให้จำก็คือ การจำคำที่พูดโดยผู้พูดที่ถูกระบุไว้ ขั้นตอนนี้เปรียบเสมือนการฝึกหรือการบันทึกคำพูด ศัพท์ และไวยากรณ์ คำเหล่านี้จะถูกเก็บไว้ในระบบ เพื่อเป็นตัวแม่แบบไว้ล่วงหน้า เมื่อคอมพิวเตอร์นำข้อมูลที่บอ้อนเข้ามาเปรียบเทียบกับตัวแม่แบบแล้วจะมีการจับคู่คำพูดที่เหมือนหรือดีที่สุด แล้วแสดงผลลัพธ์ แต่อย่างไรก็ตาม การพูดที่ระบุตัวผู้พูดก็ยังมีปัญหา เพราะเสียงที่พูดบางครั้ยังมีความแตกต่างกันเล็กน้อย การออกเสียงแต่ละคำอาจจะใช้เวลาเล็กน้อยต่างกัน คำแต่ละคำอาจจะพูดในระดับเสียงต่างกัน ทำให้ความหมายของคำแตกต่างกันไป ผู้ที่ใช้ระบบนี้จะพูดคำทั้งหมดที่เขาจะต้องใช้ ดังนั้นคำเหล่านี้จะถูกบันทึกเก็บไว้ในระบบ



คนตาบอด

หรือ

ผู้พิการทางกายภาพ

สามารถ

ถามคำถาม

หรือบอกคำสั่ง

ไปยัง

คอมพิวเตอร์

ได้สะดวกกว่าการใช้แป้นพิมพ์ติด



หน่วยจำเสียงผู้พูดทั่ว ๆ ไป (speaker independent)

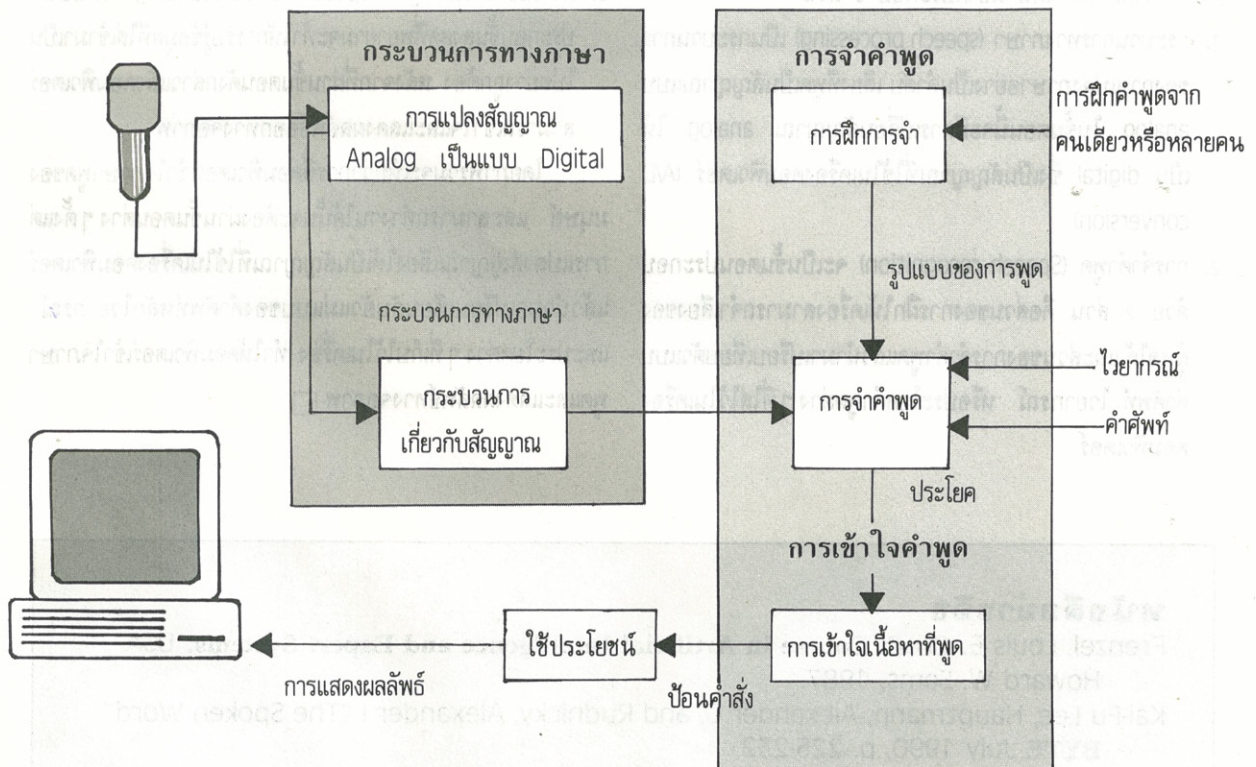
ทุกคนสามารถใช้ระบบนี้ได้ หน่วยจำคำพูดถูกออกแบบเพื่อใช้กับเสียงของทุกคน แม้จะมีความแตกต่างของคุณภาพเสียง ระบบก็สามารถทำงานได้ ระบบนี้มีความซับซ้อน และมีราคาแพงมาก รวมทั้งมีข้อจำกัดในด้านคำศัพท์อีกด้วย

ระบบการจำคำพูดทำงานอย่างไร

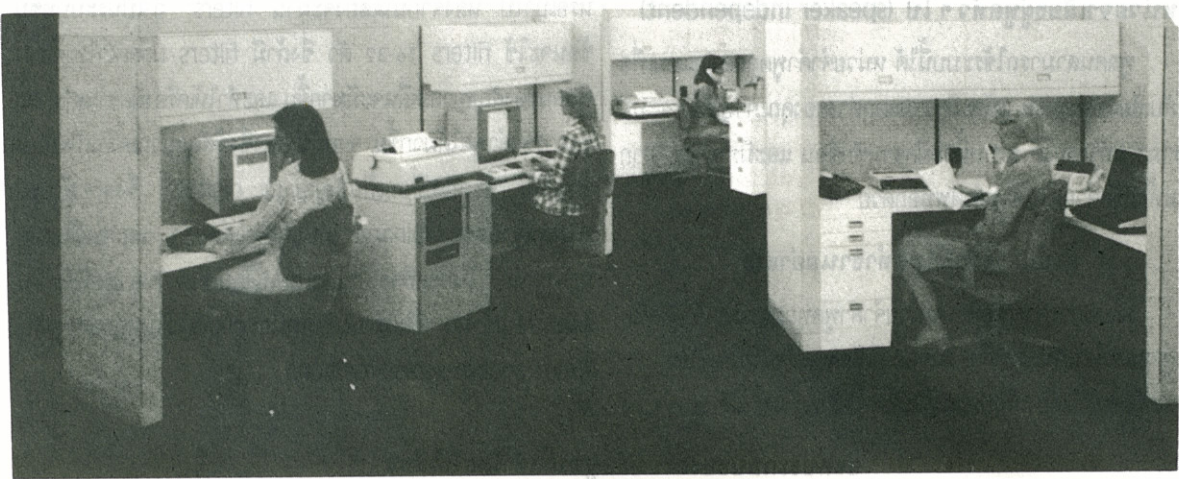
ระบบการจำคำพูดของหน่วยจำคำพูดทุกประเภท ใช้เทคนิคพื้นฐานเหมือนกันกล่าวคือ ในขณะที่เสียงพูดผ่านไมโครโฟน และผ่านเครื่องขยายเสียงเพื่อปรับระดับเสียง เพราะบางครั้งเสียงอาจจะดัง

หรือเบาไป หลังจากนั้นเสียงจะผ่าน filters¹ ถ้าเป็นระบบที่ซับซ้อนจะใช้ filters 16-32 ตัว ซึ่งถ้ามี filters มากเท่าไร ความชัดเจนของสัญญาณคลื่นจะมีมากขึ้น และทำให้เครื่องมีความสามารถในการจำคำพูดได้มากขึ้น จากนั้นเสียงซึ่งมีสัญญาณคลื่นเป็นแบบ analog² จะถูกเปลี่ยนเป็นสัญญาณ digital³ ซึ่งจะอยู่ในรูปรหัสเลขฐานสอง (binary) แทนคำพูดแต่ละคำ และจะเป็นแบบมาตรฐาน ในการที่จะเป็นแม่แบบเปรียบเทียบกับคำพูดที่จะเป็นข้อมูลป้อนเข้ามา ตัวแม่แบบคำพูดนี้จะถูกเก็บไว้ในหน่วยความจำ (RAM-Random access memory)

องค์ประกอบของระบบข้อมูลนำเข้าโดยใช้เสียง (A voice-input system)



1. **filter** วงจรไฟฟ้าที่ควบคุมความถี่คลื่นเสียงให้อยู่ในระดับที่แน่นอน มีความถี่ 3 ระดับคือ สูง กลาง ต่ำ
2. **analog** การทำงานเป็นแบบคลื่นที่มีลักษณะต่อเนื่องเหมือนเส้นกราฟขึ้นลง
3. **digital** การทำงานเป็นแบบตัวเลข ในคอมพิวเตอร์จะอยู่ในรูปของรหัสเลขฐานสอง ไม่ใช่เป็นคลื่นแบบ analog



(มกราคม ๒๕๓๖ ภาพจาก A)

จากภาพข้างต้น มีองค์ประกอบ 3 ส่วน

- 1. กระบวนการทางภาษา (speech processing) เป็นกระบวนการของการแปลงภาษาอย่างเป็นลำดับ เสียงที่พูดเป็นสัญญาณแบบ analog ในขั้นตอนนี้จะมีการเปลี่ยนสัญญาณ analog ให้เป็น digital ซึ่งเป็นสัญญาณที่ใช้ในเครื่องคอมพิวเตอร์ (A/D conversion)
- 2. การจำคำพูด (Speech recognition) จะเป็นขั้นตอนประกอบด้วย 2 ส่วน คือส่วนของการฝึกให้เครื่องสามารถจำเสียงของผู้พูดได้ และส่วนของการจำคำพูดแล้วนำมาเปรียบเทียบกับแบบคำศัพท์ ไวยากรณ์ หรือประโยคคำพูดต่างๆ ที่ใส่ไว้ในเครื่องคอมพิวเตอร์

3. ความเข้าใจในคำพูด (Speech understanding) เป็นองค์ประกอบขั้นสูงสุดที่พยายามจะทำให้การรับรู้ข้อมูลที่ใส่เข้ามาเป็นไปอย่างถูกต้อง หลังจากที่ผ่านมาขั้นตอนดังกล่าวแล้วคอมพิวเตอร์สามารถเข้าใจและแสดงผลลัพธ์ออกทางจอภาพ

โดยภาพรวมจะเห็นว่าการที่คอมพิวเตอร์เข้าใจภาษาพูดของมนุษย์ และสามารถทำงานได้นั้นจะต้องผ่านขั้นตอนต่าง ๆ ตั้งแต่การแปลงสัญญาณเสียงให้เป็นสัญญาณที่ใช้ในเครื่องคอมพิวเตอร์ แล้วนำมาเปรียบเทียบกับตัวแม่แบบของคำศัพท์หลักไวยากรณ์ และประโยคต่างๆ ที่เก็บไว้ในเครื่อง ทำให้คอมพิวเตอร์เข้าใจภาษาพูดและแสดงผลลัพธ์ทางจอภาพ □

หนังสืออ้างอิง

Frenzel, Louis E. **Crash Course in Artificial Intelligence and Expert Systems.** USA. : Howard W. Sams, 1987.
Kai-Fu Lee, Hauptmann, Alexander G. and Rudnicky, Alexander I "The Spoken Word" **BYTE** July 1990, p. 225-232.