บทที่ 1

บทนำ

1.1 แนวเหตุผลและสมมติฐาน

นับจากที่มนุษย์ได้เริ่มมองเห็นความสำคัญของเครื่องคอมพิวเตอร์ในการนำมาใช้ช่วยในการประมวลผลให้เป็นไปอย่างรวดเร็วและถูกต้องแม่นยำนั้น คอมพิวเตอร์นับวันก็ยิ่งจะแพร่หลายออกไปและได้กลายมาเป็นสิ่งจำเป็นสิ่งหนึ่งของสังคมมนุษย์ ปัจจุบันคอมพิวเตอร์เริ่มเข้ามามีบทบาทถึงภายในบ้านในรูปแบบของคอมพิวเตอร์ภายในบ้าน (home computer) หรือเป็นเครื่องคอมพิวเตอร์แบบส่วนบุคคล (personnal computer) ความสัมพันธ์และความใกล้ชิดกันระหว่างมนุษย์กับเครื่องคอมพิวเตอร์ก็ยิ่งเพิ่มขึ้นทุกที การติดต่อสื่อสารระหว่างมนุษย์กับเครื่อง (man-machine communication) ก็เริ่มจะกลายมาเป็นเรื่องสำคัญที่ต้องการการปรับปรุงและเปลี่ยนแปลงแก้ไข เพื่อให้การสื่อสารระหว่างมนุษย์กับเครื่องนั้นมีลักษณะของความเป็นกันเองให้มากที่สุด โดยมีแนวโน้มที่จะเลียนแบบให้ใกล้เคียงกับระบบประสาทสัมผัสของมนุษย์ในรูปแบบต่างๆกันไป ดังนั้นแนวทางอันหนึ่งที่จะเป็นการพัฒนาระบบการติดต่อสื่อสารระหว่างมนุษย์กับเครื่องคอมพิวเตอร์ให้มีความเป็นกันเองมากยิ่งขึ้นก็คือ การสื่อสารกันด้วยเสียงพูด เพราะเสียงพูดเป็นวิธีการติดต่อสื่อสารที่รวดเร็ว (สุพงศ์, 2525) จะเห็นได้ว่าถ้าในอนาคตเราสามารถที่จะทำการติดต่อสื่อสารกับเครื่องคอมพิวเตอร์โดยการใช้เสียงพูดโต้ตอบกันเหมือนกับเราพูดคุยกับมนุษย์ด้วยกันตามปกติแล้วจะช่วยลดช่องว่างระหว่างมนุษย์กับเครื่องจักรลง นอกจากนั้นแล้วแนวทางประยุกต์ที่จะให้ประโยชน์นอกเหนือออกไปก็คือการนำไปใช้ในการสอนสำหรับคนตาบอด การช่วยการศึกษาสำหรับนักศึกษาในด้านของภาษาต่างประเทศ และการนำไปใช้ประโยชน์ในการเสริมความสนุกสนานเพลิดเพลินของซอฟแวร์ต่างๆ (Sherwood, 1979) สำหรับในด้านอุตสาหกรรมก็อาจจะนำไปใช้ได้ในบริเวณที่มีเสียงดังรบกวนมากๆ และลำบากในการใช้เสียงตะโกนบ่อยๆครั้ง หรือในกรณีของงานควบคุมที่มือทั้งสองของผู้ควบคุมไม่ว่าง ซึ่งทั้งหมดนี้อาจจะนำไปสู่แนวทางของระบบโรงงานอัตโนมัติ (automated factory) และอาจจะรวมไปถึงการนำไปประยุกต์ในรูปแบบของสำนักงานอัตโนมัติด้วย (Bursky, 1985)

ด้วยเหตุนี้เพื่อเป็นการพัฒนาแนวทางการสื่อสารดังกล่าว งานวิจัยนี้จึงมุ่งพัฒนาให้เครื่องคอมพิวเตอร์ขนาดเล็กสามารถเปล่งเสียงออกมาเป็นภาษามนุษย์ให้ได้ เพื่อใช้ในการสื่อความหมายระหว่างกัน โดยเฉพาะอย่างยิ่งการพัฒนาให้เครื่องคอมพิวเตอร์ขนาดเล็กหรือคอมพิวเตอร์ที่ใช้ภายในบ้านสามารถพูดออกมาเป็นภาษาไทยเพื่อนำมาใช้กับคนไทย และตามเหตุผลแล้วภาษาไทยกับภาษาอื่นๆก็มีการเกิดขึ้นมาจากรากฐานของเสียงในลักษณะเดียวกัน ดังนั้นจึงเชื่อว่าการทำการสังเคราะห์เสียงพูดภาษาไทยโดยใช้เครื่องคอมพิวเตอร์ขนาดเล็กนั้นก็น่าจะทำได้ในรูปแบบเดียวกันกับการสังเคราะห์เสียงพูดของภาษาอื่นๆที่มีผู้ทำการวิจัยสำเร็จมาแล้วรวมทั้งที่ทำออกมาจำหน่ายในรูปของผลิตภัณฑ์ทางการค้าด้วย เพียงแต่ว่าภาษาไทยและภาษาของบางประเทศในแถบทวีปเอเชียนั้นเป็นภาษาที่มีเสียงวรรณยุกต์ (tone language) ที่ทำให้เป็นส่วนที่แตกต่างไปจากภาษาในแถบอื่นๆเท่านั้น (Chatchavalit Saravari and Satoshi Imai, 1983) ดังนั้นจากสมมติฐานอันนี้จึงทำให้เกิดหัวข้อการวิจัยชิ้นนี้ขึ้นมา

1.2 หลักการเบื้องต้น

การแปลงข้อความให้ออกมาเป็นเสียงพูดนั้น (Text-to-speech) คือลักษณะของการที่เราป้อนข้อความเป็นคำๆหรือชุดของหน่วยย่อยของคำเข้าไป แล้วเครื่องจะทำการแปลงชุดของข้อความนั้นให้ออกมาเป็นเสียงพูดตามชุดตัวอักขระนั้นได้ (Veltri, 1985) หลักการหรือขั้นตอนอย่างคร่าวๆของการแปลงข้อความให้ออกมาเป็นเสียงพูดจะสามารถทำได้โดยเริ่มจากระบบจะทำการรับข้อความเข้ามา แล้วก็เปลี่ยนข้อความนั้นให้เป็นชุดของโฟนีม (phoneme) หรือสัญญลักษณ์แทนเสียงของคำแต่ละคำแทน รวมทั้งพยายามที่จะกำหนดค่าต่างๆที่เกี่ยวข้องกับเสียง เช่น ความยาวของคำ พยางค์ที่จะเน้น ระดับความดังและระดับของพิทช์ (pitch) ที่ใช้ในการเน้น เป็นต้น ขั้นต่อไปก็คือการแปลโฟนีมและตัวแปรต่างๆข้างต้นให้กลายเป็นพารามิเตอร์ในโดเมนของความถี่ รวมทั้งการจัดการเกี่ยวกับการเปลี่ยนแปลงอย่างต่อเนื่องระหว่างโฟนีมหนึ่งกับโฟนีมตัวถัดไป แล้วพารามิเตอร์เหล่านี้ก็จะถูกทำการสังเคราะห์ให้ออกมาเป็นเสียงพูดอีกที

จากหลักการข้างต้นพอจะสรุปได้ว่า การแปลงตัวอักขระให้ออกมาเป็นเสียงพูดนั้นมี 2 ขั้นตอนใหญ่ๆคือ

1.2.1 ขั้นตอนที่ทำการรับข้อความเข้ามา เป็นส่วนที่ทำการรับข้อความเข้ามาแล้วแปลงเป็นสัญญลักษณ์แทนเสียงพร้อมทั้งพารามิเตอร์ที่ควบคุมลักษณะของเสียง เช่น ระดับความถี่ การเน้นเสียงที่พยางค์ เป็นต้น ในส่วนแรกนั้นเราจำเป็นจะต้องรู้ว่าข้อความที่ป้อนเข้ามานั้น

อ่านออกเสียงอย่างไรจึงจะถูกต้อง แล้วจึงแทนด้วยชุดสัญญลักษณ์ตามเสียงที่ถูกต้องนั้นซึ่งอาจจะทำได้โดยวิธีการดังนี้ (Bursky, 1985)

1) ใช้ชุดของกฎเกณฑ์ต่างๆตามหลักการแปลงตัวอักษรเป็นหน่วยเสียง (Letter to sound rule set) นั่นคือการรวบรวมกฎเกณฑ์ข้อบังคับต่างๆที่บัญญัติไว้ในหลักภาษาของการอ่าน เพื่อจะได้จัดเตรียมชุดสัญญลักษณ์ทางเสียงได้ถูกต้องดังนั้นจะมีกฎเกณฑ์เหล่านี้อยู่มากมาย แต่อย่างไรก็ตามย่อมจะต้องมีคำที่แปลกออกไปไม่ได้อ่านออกเสียงตามกฎเกณฑ์เหล่านั้น เป็นคำที่อยู่นอกเหนือกฎเกณฑ์ ทำให้การใช้วิธีการนี้เพียงอย่างเดียวไม่สามารถครอบคลุมการอ่านคำต่างๆได้อย่างถูกต้องทั้งหมด

2) การใช้ระบบพจนานุกรม (Dictionary) เป็นการแก้ปัญหาจากการใช้กฎเกณฑ์ต่างๆได้เป็นบางส่วน นั่นคือในกรณีที่เป็นคำที่อยู่นอกเหนือกฎเกณฑ์หรืออ่านออกเสียงผิดไปจากหลักการตามปกติก็จะจัดคำเหล่านั้นเป็นกลุ่มๆ พร้อมทั้งกำหนดคำอ่านที่ถูกต้องของแต่ละคำเอาไว้รวบรวมแยกไว้เป็นพจนานุกรมในแฟ้มข้อมูล ก็จะสามารถค้นหาสัญญลักษณ์แทนเสียงที่ถูกต้องของแต่ละคำได้ แต่อย่างไรก็ตามก็ยังมีข้อยกเว้นอยู่บ้าง เช่นในกรณีของคำพ้องรูป ที่สามารถอ่านออกเสียงได้หลายแบบโดยขึ้นกับความหมายของคำที่อยู่รอบๆคำนั้น ซึ่งจะไปเกี่ยวข้องเข้าสู่ปัญหาในแง่ของปัญญาประดิษฐ์ทำให้ยุ่งยากขึ้นไปอีกหลายระดับขั้น

1.2.2 ขั้นตอนที่ทำหน้าที่สังเคราะห์เสียงตามชุดสัญญลักษณ์แทนเสียงและพารามิเตอร์ เป็นส่วนที่ใช้พารามิเตอร์สังเคราะห์เสียงขึ้นมาตามชุดสัญญลักษณ์แทนเสียงให้ออกมาเป็นเสียงพูดที่เราได้ยินกัน สำหรับในส่วนที่สองนี้เป็นส่วนของการสังเคราะห์เสียงพูดโดยตรง จึงเป็นส่วนที่สำคัญมากสำหรับระบบการแปลงข้อความเป็นเสียงพูดนี้ เพราะเป็นส่วนที่จะทำให้เครื่องสามารถกำเนิดเสียงพูดออกมาได้ เทคนิคในการทำให้เครื่องสามารถสร้างเสียงพูดออกมาได้นั้นมีอยู่หลายวิธี แต่วิธีที่มีความยืดหยุ่นมากที่สุดวิธีหนึ่งก็คือ การใช้ลักษณะของการเก็บการแทนค่าของคำพูดต้นฉบับเอาไว้ (store a textual representation of the utterance) ซึ่งมีวิธีการที่แตกต่างกันออกไป 2 วิธีใหญ่ๆคือ (Witten, 1982)

1) การสร้างเสียงโดยการเก็บบันทึกเสียงพูด (speech storage) เป็นวิธีของการเก็บบันทึกเสียงจริงๆของมนุษย์เอาไว้อย่างตรงไปตรงมา โดยอาจจะแบ่งเก็บเป็นข้อมูลของหน่วยย่อยๆของคำหรือพยางค์ ซึ่งเมื่อมีการสร้างเสียงพูดขึ้นมาก็จะมีการอ้างถึงข้อมูลของเสียงหน่วยย่อยๆนี้นำมาเรียงต่อๆกันเกิดเป็นคำหรือประโยคขึ้นมา

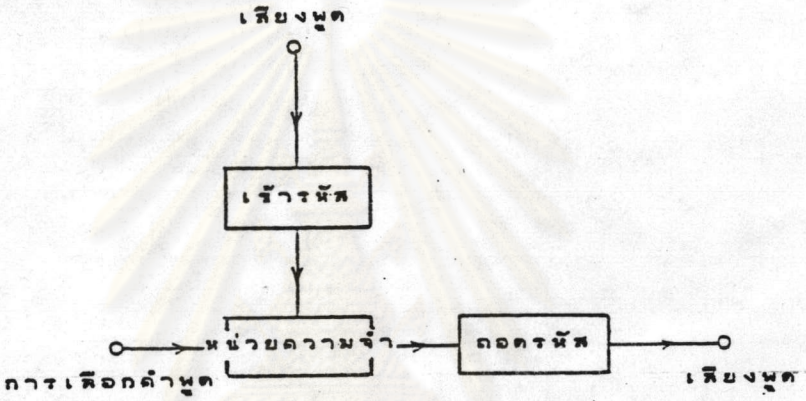
ข้อดี ก็คือสามารถจะสร้างเสียงพูดที่มีคุณภาพสูงมากๆหรือใกล้เคียงกับสำเนียงของมนุษย์ได้เป็นอย่างดี

ข้อเสีย ก็คือใช้หน่วยเก็บข้อมูลมากตามคุณภาพของเสียงที่ได้

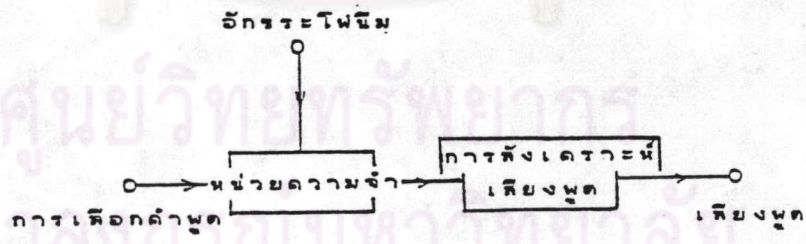
2) การสร้างเสียงโดยการสังเคราะห์เสียงพูด (speech synthesis) เป็นการสร้างเสียงพูดขึ้นมาโดยที่เครื่องจะเป็นตัวสร้างเสียงพูดขึ้นมาโดยตัวมันเอง โดยไม่จำเป็นต้องมีการบันทึกเสียงของมนุษย์สำหรับเสียงพูดที่ต้องการจะให้เครื่องพูดออกมาเอาไว้เลย

ข้อดี ใช้หน่วยความจำน้อยกว่าวิธีแรกมาก ทำให้สามารถมุ่งไปสู่การสร้างระบบสังเคราะห์เสียงที่ไม่จำกัดจำนวนคำศัพท์ และที่สำคัญที่สุดก็คือสามารถทำได้บนเครื่องคอมพิวเตอร์ขนาดเล็กและสิ้นเปลืองค่าใช้จ่ายน้อย

ข้อเสีย โปรแกรมในการสังเคราะห์เสียงพูดมีความยุ่งยากและเสียงที่ได้ออกมามีคุณภาพอยู่ในระดับปานกลาง



รูปที่ 1.1.1 การสร้างเสียงโดยการเก็บบันทึกเสียงพูด



รูปที่ 1.1.2 การสร้างเสียงโดยการสังเคราะห์เสียงพูด

แต่ทั้งสองวิธีมีสิ่งหนึ่งที่เหมือนกันคือ ต้องมีการเก็บอะไรบางสิ่งบางอย่างเอาไว้คือ

- การสร้างเสียงโดยการเก็บบันทึกเสียงพูด ต้องมีการเก็บข้อมูลซึ่งเป็นการแทนค่าเสียงพูดของมนุษย์โดยตรง ทำให้สำเนียงและการเน้นเสียงจะปรากฏอยู่ในตัวข้อมูล

ที่บันทึกเอาไว้ที่จะให้กลับออกมาเป็นเสียงพูด

- การสร้างเสียงโดยการสังเคราะห์เสียงพูด จะมีการเก็บชนิดของรูปแบบของเสียงพูดในรูปของเสียงหรือสัญญลักษณ์แทนเสียง ซึ่งจะนำมาใช้เป็นองค์ประกอบของเสียงต่างๆโดยการเน้นเสียงจะทำที่ตัวเครื่อง และสำเนียงของเสียงจะควบคุมโดยส่วนของโปรแกรมสังเคราะห์เสียงแทน

จากข้อดีข้อเสียของการสร้างเสียงพูดทั้งสองวิธี งานวิจัยนี้จึงมุ่งมาใช้วิธีทางด้าน การสังเคราะห์เสียงพูดในการสร้างเสียงพูดออกมา ในปัจจุบันนี้ระบบที่สามารถจะสร้างหรือสังเคราะห์เสียงพูดออกมาได้นั้นมีรากฐานอยู่ 2 รูปแบบคือ การสังเคราะห์เชิงวิเคราะห์ (analysis synthesis) กับ การสังเคราะห์เชิงก่อ (constructive synthesis) (Schalk, 1982)

การสังเคราะห์เชิงวิเคราะห์จะมีการสร้างคำศัพท์ขึ้นมาเป็นจำนวนที่จำกัดแต่มีคุณภาพในการฟังออกค่อนข้างสูง ส่วนการสังเคราะห์เชิงก่อจะเป็นแนวทางที่สร้างคำศัพท์ขึ้นมาได้ไม่จำกัดจำนวนแต่มีคุณภาพในการฟังออกต่ำกว่า เสียงพูดจะมีรากฐานอยู่ที่การเปลี่ยนแปลงของโฟนีมหรืออัลโลโฟน (allophone) ของเสียง

ในงานวิจัยอันนี้จะหันไปใช้เทคนิคการทำนายแบบเชิงเส้น หรือ LPC ซึ่งเป็นลักษณะของการสังเคราะห์เชิงวิเคราะห์ ในการวิเคราะห์สัญญาณเสียงออกมาเก็บไว้ในรูปของพารามิเตอร์ ซึ่งเป็นการลดจำนวนหน่วยความจำที่จำเป็นต้องใช้ในการเก็บสัญญาณเสียงลง และด้วยเทคนิคอันเดียวกันนี้ก็จะนำมาประยุกต์ใช้ในการสังเคราะห์เสียงโดยใช้พารามิเตอร์ที่เก็บเอาไว้นั้นแปลงกลับออกมาเป็นสัญญาณเสียงอย่างเดิมซึ่งคุณภาพของเสียงที่ได้จากเทคนิคอันนี้จัดอยู่ในระดับปานกลาง

1.3 วัตถุประสงค์

1) เพื่อพัฒนาระบบการสังเคราะห์เสียงพูด ให้สามารถรับข้อความภาษาไทยรวมทั้งสัญญลักษณ์พิเศษ แล้วเปล่งเสียงออกมาตามคำนั้นได้อย่างถูกต้อง
2) เป็นการพัฒนาระบบการสังเคราะห์เสียงพูดภาษาไทยเชิงวิเคราะห์ ที่ใช้เทคนิคแอลพีซี (LPC : Linear Prediction Coding) เพื่อเป็นแนวทางในการลดหน่วยความจำที่ใช้ และเพิ่มประสิทธิภาพของเสียงพูดภาษาไทย
3) ระบบสามารถที่จะทำการสังเคราะห์เสียงพูดภาษาไทย โดยรับข้อความภาษาไทยเข้ามาได้ทั้งจากแป้นพิมพ์โดยตรง หรือจากแฟ้มข้อมูลในแผ่นจานแม่เหล็ก

4) เสียงสังเคราะห์จะมีคุณสมบัติของภาษาไทยที่สำคัญอยู่คือเสียงวรรณยุกต์

5) สามารถเพิ่มเติมหน่วยของคำที่เก็บไว้ในพจนานุกรมข้อมูล รวมทั้งเปลี่ยนแปลงแก้ไขได้

1.4 ขั้นตอนการวิจัย

1) วิเคราะห์เสียงพูดภาษาไทยรวมทั้งหาพารามิเตอร์ที่จำเป็น โดยมีวัตถุประสงค์ที่จะนำพารามิเตอร์เหล่านั้น กลับมาทำให้เกิดเป็นเสียงพูดให้ได้อย่างเดิม

2) ศึกษาแนวทางในด้านของการประมวลผลสัญญาณดิจิตอล (digital signal proceesing)

3) แปลความหมายของพารามิเตอร์ และหาวิธีการสร้างเสียงพูดจากพารามิเตอร์

4) จำลองแบบลงบนเครื่องไมโครคอมพิวเตอร์

5) สร้างโปรแกรมเพื่อรับชุดอักขระรวมทั้งสัญญลักษณ์พิเศษในการเปลี่ยนแปลงลักษณะของเสียงที่สังเคราะห์ออกมา

6) สรุปผลการวิจัย

7) เสนอแนะแนวทางในการปรับปรุงแก้ไข

1.5 ขอบเขตการวิจัย

1) ใช้พยางค์เป็นหน่วยย่อยของคำ

2) รับข้อมูลเข้าเฉพาะข้อความภาษาไทยที่เป็นพยางค์ของคำอ่าน รวมทั้งสัญญลักษณ์พิเศษในการเปลี่ยนแปลงลักษณะของเสียงสังเคราะห์ด้วย

3) รับตัวอักขระภาษาไทยในรูปของอักขระภาษาไทยรหัสเกษตร

4) ในกรณีที่รับข้อความจากแป้นพิมพ์จะต้องพิมพ์เป็นชุดๆ ชุดละไม่เกิน 1 บรรทัดโดยมีรหัสของแป้น ENTER เป็นรหัสในการเริ่มต้นให้ระบบทำการอ่านในแต่ละข้อความหรือจะใช้สัญญลักษณ์พิเศษก็ได้

5) ระบบให้คุณภาพเสียงอยู่ในระดับปานกลาง

6) ประยุกต์ลงบนเครื่องไมโครคอมพิวเตอร์