บทที่ 1

บทนำ

1.1 ความสำคัญและที่มาของหัวข้อวิจัย

การสื่อสาร (communication) นับว่ามีความสำคัญและเป็นส่วนหนึ่งของมนุษย์ จึงมีความพยายามที่จะแปลงสาร (message) เป็นคำพูด (speech) เพื่อให้ความสะดวกแก่ผู้รับสารโดยที่ไม่ต้องกวาดสายตาอ่านข้อความตามวิธีเดิมอีกต่อไป วิธีการคือต้องอาศัยคอมพิวเตอร์ช่วยในการแปลงข้อความเป็นเสียงพูด (text to speech system) ตัวอย่างเช่นพจนานุกรมเสียงพูดซึ่งมีองค์ประกอบที่สำคัญของระบบนี้คือการสังเคราะห์เสียงพูด

จากระบบที่ได้ สามารถนำไปประยุกต์ใช้งานในหลาย ๆ ด้าน เช่นผู้พิการทางด้านการออกเสียงสามารถที่จะสื่อสารกับคนปกติที่ไม่รู้ภาษามือได้ แม้แต่คนหูหนวกหูตึงซึ่งจะมีปัญหาเรื่องการออกเสียงก็มีโอกาสที่จะสื่อสารได้เหมือนคนปกติ นอกจากนี้ยังนำไปใช้งานด้านการสอนเพื่อการออกเสียงที่ถูกต้องในภาษาต่าง ๆ ในอนาคตนั้นถ้าระบบการรู้จำเสียงและระบบการสังเคราะห์เสียงมีความสามารถที่เพียงพอก็จะสามารถนำไปเป็นเป็นตัวแปลภาษาเพื่อสื่อสารพูดคุยกันได้

ในปัจจุบันนี้งานที่เกี่ยวข้องกับเสียงพูดในภาษาไทยมักจะมุ่งเน้นไปที่การรู้จำเสียง ส่วนงานด้านการสังเคราะห์เสียงยังทำกันไม่มากนัก วิธีที่ง่ายที่สุดในการผลิตเสียงสังเคราะห์คือการเรียกใช้จากเสียงที่ได้ทำการบันทึกไว้แล้ว ซึ่งวิธีการนำคำพูดมาต่อกันนี้จะได้ความเป็นธรรมชาติของเสียงที่มีคุณภาพดี แต่มีข้อจำกัดในเรื่องของจำนวนคำศัพท์ (vocabulary database) เพราะมีเนื้อที่ในการเก็บข้อมูลจำกัดนั่นเอง ซึ่งวิธีนี้ยังถือว่าเป็นการสังเคราะห์เสียงได้ไม่เต็มที่นัก วิธีการสังเคราะห์เสียงที่ไม่มีขีดกำจัดในเรื่องของศัพท์คือการนำบางส่วนของสัญญาณเสียงพูดเช่น พยางค์ (syllables) หน่วยเสียง (phonemes) เสียงคู่ (diphones) ไปใช้งาน ก็สามารถแก้ปัญหาเรื่องของหน่วยความจำลงได้ แต่ก็ให้ความเป็นธรรมชาติของเสียงไม่ดีเพราะใช้เพียงบางส่วนของเสียงพูดเพื่อสร้างเสียงใหม่ ดังนั้นวิทยานิพนธ์นี้จึงได้ทำการศึกษาการสังเคราะห์เสียงพูดภาษาไทยที่ใช้เนื้อที่ในการเก็บข้อมูลน้อยและได้คุณภาพเสียงที่ดี ในเบื้องต้นจะทำการสังเคราะห์เฉพาะเสียงสระภาษาไทยตามวรรณยุกต์สามัญเท่านั้นเพื่อเป็นแนวทางในการสังเคราะห์เสียงพูดต่อไป

ในทางทฤษฎีวิธีที่มีความถูกต้องในการสังเคราะห์เสียงพูดที่สุดคือใช้แบบจำลองของอวัยวะที่ใช้ในการเปล่งเสียง (articulatory synthesis) โดยจำลองอวัยวะที่ใช้ในการเปล่งเสียง (articulators) ซึ่งวิธีนี้จะให้เสียงที่เป็นธรรมชาติและใช้เนื้อที่ในการเก็บข้อมูลน้อย แต่ก็มีปัญหาเรื่องความซับซ้อนของการจำลองอวัยวะที่ใช้ในการเปล่งเสียงทั้งหมด

วิธีการสังเคราะห์เสียงทั้งหมดมีทั้งข้อดีและข้อเสีย ยากที่จะบ่งบอกได้ว่าวิธีไหนดีที่สุด สำหรับวิธีการสังเคราะห์แบบบันทึกเสียงพูดประสบความสำเร็จสูงในปัจจุบัน ส่วนการสังเคราะห์เสียงโดยใช้แบบจำลองของอวัยวะที่ใช้ในการเปล่งเสียงจะเป็นวิธีที่มีประสิทธิภาพที่ดีในอนาคต

1.2 การตรวจเอกสาร

1.2.1 ต้นแบบเครื่องสังเคราะห์เสียงพูดด้วยวิธีเข้ารหัสแบบลิเนียร์พรีดิกทีฟ (กฤษดา เรเยส , 2530) งานวิจัยนี้ได้นำเสนอการสร้างระบบวิเคราะห์และสังเคราะห์เสียงพูดโดยอาศัยวิธีเข้ารหัสแบบลิเนียร์พรีดิกทีฟ (linear predictive coding) ซึ่งเป็นวิธีการที่แพร่หลายวิธีหนึ่ง ระบบวิเคราะห์และสังเคราะห์เสียงพูดที่สร้างขึ้นประกอบด้วยระบบไมโครคอมพิวเตอร์ภาคแปลงสัญญาณระหว่างสัญญาณอนาล็อกกับสัญญาณดิจิตอล และภาคประมวลผลสัญญาณ การทดลองในงานวิจัยชิ้นนี้จะชักตัวอย่างเสียงพูดด้วยความถี่ 10 kHz นำมาบันทึกลงหน่วยความจำภายในเครื่องไมโครคอมพิวเตอร์แล้วคำนวณหาค่าพารามิเตอร์ซึ่งเป็นตัวแทนของสัญญาณเสียง พารามิเตอร์เหล่านี้สามารถนำไปผลิตเป็นสัญญาณเสียงพูดได้ด้วยกรรมวิธีสังเคราะห์เสียง จากการประยุกต์ใช้วงจรรวมประมวลผลสัญญาณส่วนของการสังเคราะห์เสียงพูดสามารถทำงานได้ในลักษณะทำงานได้จริง (real time) การทำให้อัตราของข้อมูลเสียงน้อยที่สุดอาศัยการทดลองโดยเปลี่ยนแปลงจำนวนออร์เดอร์ของฟิลเตอร์ระหว่าง 2 ถึง 15 และเปลี่ยนแปลงขนาดของเฟรมระหว่าง 100 ถึง 300 แซมเปิล จากนั้นเสียงที่ได้จากการสังเคราะห์จะถูกประเมินคุณภาพด้วยค่าความผิดพลาดที่ถูกทำให้เป็นบรรทัดฐาน (normalized error) และจากการฟัง เพื่อหาองค์ประกอบในการสังเคราะห์เสียงที่ให้อัตราข้อมูลเสียงน้อยที่สุด ผลการทดลองของการสังเคราะห์เสียงโดยการกำหนดจำนวนออร์เดอร์ของฟิลเตอร์เท่ากับ 10 และจำนวนแซมเปิลในหนึ่งเฟรมเท่ากับ 200 เสียงที่ได้มีคุณภาพใช้งานได้เมื่อพิจารณาจากการฟัง การลดทอนข้อมูลสามารถลดทอนได้ประมาณ 15 เท่า อย่างไรก็ตามจำนวนแซมเปิลในหนึ่งเฟรมที่เหมาะสมจะขึ้นกับการเปลี่ยนแปลงคุณสมบัติตามเวลาของเสียงคำนั้น ๆ ซึ่งต้องอาศัยการทดลองเฉพาะเป็นคำ ๆ ไป ในเรื่องจำนวนออร์เดอร์ของฟิลเตอร์เช่นเดียวกัน แม้จะไม่ขึ้นกับการเปลี่ยนแปลงคุณสมบัติตามเวลามากนัก แต่ก็ขึ้นกับคุณสมบัติทางความถี่ของเสียงพูดคำนั้น ๆ โดยตรง จึงทำให้คุณภาพเสียงของแต่ละคำแตกต่างกันออกไป

1.2.2 การสังเคราะห์เสียงพูดจากข้อความภาษาไทย (อาทร นันทิยกุล, 2533) เป็นการสร้างเสียงพูดขึ้นมาจากข้อความภาษาไทยที่ถูกป้อนเป็นอินพุตเข้าสู่ระบบซึ่งได้ประยุกต์ลงบนไมโครคอมพิวเตอร์และมีภาคการแปลงสัญญาณระหว่างสัญญาณอนาล็อกกับสัญญาณดิจิตอลรวมอยู่ด้วย โดยใช้หลักการวิเคราะห์หน่วยย่อยของเสียงพูดคือ พยางค์ของเสียงต้นแบบมาทำการชักตัวอย่างด้วยความถี่ 10 kHz และนำมาทำการวิเคราะห์ด้วยวิธีเข้ารหัสแบบลิเนียร์พรีดิกทีฟ แบบ

ออร์เดอร์ 10 ทีละเฟรม (เฟรมละ 200 จุดหรือ 20 มิลลิวินาที) ได้เอาต์พุตออกมาเป็นชุดพารามิเตอร์ ประกอบด้วย 1) ค่าความผิดพลาดเฉลี่ย 2) ค่าคาบของพิตช์ และ 3) ค่าสัมประสิทธิ์ของการทำนาย (10 ค่า) ต่อหนึ่งเฟรม เพื่อนำเก็บเอาไว้ในหน่วยเก็บความจำสำรองในรูปแบบของพจนานุกรมข้อมูลซึ่งสามารถจะทำการแก้ไขพารามิเตอร์เพื่อปรับปรุงให้ได้เสียงสังเคราะห์ที่ดีขึ้น จากนั้นก็จะทำการสร้างเสียงพูดสังเคราะห์ขึ้นมา โดยนำเอาชุดพารามิเตอร์ของหน่วยย่อยของเสียงที่ได้มาจากการค้นหาในพจนานุกรมข้อมูลตามข้อความที่ป้อนเข้ามาเพื่อทำการสังเคราะห์ผ่านตัวกรองแลททิส (lattice filter) และภาคการแปลงสัญญาณดิจิตอลเป็นอนาล็อกออกลำโพงกลับคืนมาเป็นเสียงพูด โดยคุณภาพของเสียงที่สังเคราะห์ออกมาอยู่ในระดับปานกลางซึ่งวัดผลได้จากการรับฟังของกลุ่มตัวอย่าง และเสียงสังเคราะห์นี้จะมีคุณสมบัติของเสียงวรรณยุกต์ซึ่งเป็นลักษณะสำคัญของภาษาไทย

การป้อนข้อความของอินพุตจะป้อนในรูปของตัวสะกดตามคำอ่านเท่านั้น ซึ่งระบบการสังเคราะห์เสียงพูดจากข้อความภาษาไทยนี้ยังต้องมีการพัฒนาในส่วนของการแปลงข้อความที่เป็นภาษาเขียนให้เป็นคำอ่านอีก แต่อย่างไรก็ตามระบบการสังเคราะห์เสียงพูดนี้ก็ได้ใช้เทคนิคการสังเคราะห์เสียงพูดขึ้นมาจากชุดพารามิเตอร์โดยใช้เป็นซอฟแวร์ทั้งหมด ซึ่งทำให้สามารถนำไปประยุกต์ใช้ได้ง่ายและกว้างขวางสำหรับระบบการสังเคราะห์เสียงพูดแบบฐานความรู้ทั่ว ๆ ไป

1.2.3 Tone Transformation (Luksaneeyanawin Sudaporn, 1995) งานชิ้นนี้เป็นการสังเคราะห์เสียงวรรณยุกต์ขึ้นใหม่โดยใช้กรรมวิธี PSOLA (Pitch-Synchronous Overlap-Add) ซึ่งเป็นวิธีที่ใช้ในการเปลี่ยนท่วงทำนองของหน่วยเสียงโดยกรรมวิธีที่เกี่ยวกับการเปลี่ยนแปลงคาบพิตช์ (ในคลื่นเสียงตำแหน่งที่มียอดสูงสุดเรียกว่า ตำแหน่งพิตช์ ระยะระหว่างตำแหน่งพิตช์สองตำแหน่งเรียกว่าคาบพิตช์) เสียใหม่เพื่อให้เสียงสังเคราะห์มีท่วงทำนองใหม่โดยหาตำแหน่งพิตช์ของเสียงต้นแบบเตรียมเอาไว้ จากนั้นนำมาเลื่อนระยะห่างของพิตช์เหล่านั้นให้มีท่วงทำนองเป็นไปตามต้องการ ซึ่งงานชิ้นนี้ได้นำ กรรมวิธีนี้มาใช้ในการสังเคราะห์ซ้ำเสียงวรรณยุกต์ภาษาไทยโดยใช้หน่วยพยางค์ต้นแบบที่มี วรรณยุกต์เอก โดยการปรับคาบพิตช์ของเสียงโดยวิธี PSOLA ให้เป็นไปตามรูปแบบค่าความถี่หลักมูล (fundamental frequency เป็นความถี่หลักของเสียงซึ่งเสียงจะประกอบด้วยความถี่หลักมูลและฮาร์มอนิก) ของแต่ละวรรณยุกต์ ใช้หน่วยเสียง 1,200 หน่วยรูปเสียง (speech units) ในพื้นที่ 3 MB ผลที่ได้มีคุณภาพค่อนข้างดี จากผลการทดสอบการรับฟังพบว่าในการสังเคราะห์เสียงในสระเสียงยาวค่าคะแนนของการระบุวรรณยุกต์มีค่าเฉลี่ยสูงสุดและต่ำสุดเท่ากับ 100% และ30 % ตามลำดับ คะแนนของคุณภาพเสียงโดยค่าความพอใจมากเท่ากับ 5 และไม่พอใจเลยเท่ากับ 1 มีค่าเฉลี่ยสูงสุดและต่ำสุดเท่ากับ 4.2 และ 1.5 ตามลำดับ และในการสังเคราะห์เสียงสระเสียงสั้นคะแนนของการระบุวรรณยุกต์มีค่าเฉลี่ยสูงสุดและต่ำสุดเท่ากับ 98% และ 96 % ตามลำดับ คะแนนของคุณภาพเสียงมีค่าเฉลี่ยสูงสุดและต่ำสุดเท่ากับ 3.9 และ 2.0 ตามลำดับ

1.2.4 การสังเคราะห์เสียงวรรณยุกต์และสระในพยางค์เปิดภาษาไทยโดยใช้หน่วยเสียงอนุภาค (เอกพล อนุสุเรนทร์, 2541) งานวิจัยนี้เสนอ การสังเคราะห์เสียงพยางค์เปิดภาษาไทย โดยได้นำหน่วยเสียงอนุภาคและความรู้ด้านภาษาศาสตร์ที่เกี่ยวข้องกับค่าความถี่หลักมูลของแต่ละเฉพาะหน่วยเสียงอนุภาคและนำมาเชื่อมต่อกันเพื่อสร้างคลื่นเสียงของส่วนสระด้วยการปรับค่าแอมปลิจูดและค่าความถี่หลักมูลนี้สามารถที่จะสังเคราะห์เสียงวรรณยุกต์อื่น ๆ ขึ้นมาได้ วิธีการสังเคราะห์เสียงนี้ได้ทำการศึกษาเฉพาะการสังเคราะห์เสียงสระเดี่ยว และวรรณยุกต์ในพยางค์เปิดภาษาไทยเท่านั้น และได้ทำการทดสอบการรับฟังโดยกลุ่มตัวอย่างเพื่อทดสอบความถูกต้องในการระบุเสียงวรรณยุกต์ของเสียงสังเคราะห์ และเพื่อวัดทัศนะคติของกลุ่มตัวอย่างต่อคุณภาพของเสียงสังเคราะห์ จากการทดสอบการระบุเสียงวรรณยุกต์ของกลุ่มตัวอย่าง กลุ่มตัวอย่างสามารถจำแนกเสียงวรรณยุกต์ได้โดยเฉลี่ยถูกต้อง 95% (ค่าเบี่ยงเบนมาตรฐาน = 6.4) โดยมีพิสัยตั้งแต่ 58% ถึง 100% ค่าทัศนคติต่อคุณภาพเสียง (ไม่พอใจเลย = 1, พอใจมาก = 5) มีค่าโดยเฉลี่ยเท่ากับ 2.95 (ค่าเบี่ยงเบนมาตรฐาน = 0.69) โดยมีพิสัยตั้งแต่ 1.4 ถึง 4.2

จุดประสงค์ของงานวิจัยชิ้นนี้คือการลดจำนวนของหน่วยเสียงและขนาดของหน่วยเสียง ถ้าคิดขนาดของหน่วยเสียงที่ใช้หน่วยพยางค์จากพยางค์ที่สามารถเกิดขึ้นได้ในพยางค์เปิดสระเดี่ยวจะใช้หน่วยเสียงจำนวน 2,376 หน่วยและใช้หน่วยความจำ 77 MB แต่ถ้าเก็บโดยหน่วยเสียงอนุภาคจะทำให้หน่วยเสียงลดลงเหลือ 432 หน่วย และเนื้อที่เก็บจะเหลือ 2.6 MB

1.2.5 การอ่านออกเสียงคำหลายพยางค์ในภาษาไทยของคอมพิวเตอร์ (อภิญญา คำยา, ลลิตา นฤปิยะกุล, บุญเจริญ ศิริเนาวกุล, 2543) งานวิจัยชิ้นนี้ได้นำเสนอการแก้ปัญหาในส่วนของโปรแกรมการทำงานของคอมพิวเตอร์ที่ไม่สามารถแบ่งพยางค์ที่อยู่ติดกันให้ถูกต้องได้ซึ่งต้องอาศัยความรู้ด้านภาษาศาสตร์ในการวิเคราะห์ โปรแกรมการอ่านออกเสียงของเครื่องคอมพิวเตอร์ที่ใช้ในงานวิจัยชิ้นนี้เน้นการพัฒนาในส่วนของการถ่ายทอดภาษาเขียนให้เป็นสัทอักษร (phonetic symbol) ให้ถูกต้องมากที่สุดโดยใช้วิธีการทางภาษาศาสตร์เพื่อแก้ปัญหา ซึ่งสัทอักษรเป็นสัญลักษณ์ทางภาษาเพื่อแสดงคุณสมบัติของเสียงแต่ละเสียงกำหนดโดยนักภาษาศาสตร์ สัทอักษรที่ใช้เป็นสากลเรียกว่า IPA (International Pronunciation Alphabet) สัทอักษรแต่ละตัวทำหน้าที่แทนหน่วยเสียงซึ่งแสดงให้เห็นว่าแต่ละเสียงทำหน้าที่ต่างกัน

1.2.6 http://www.haskins.yale.edu/Haskins/MISC/ASY/asy.html เป็นเว็บไซต์ที่ให้ข้อมูลเกี่ยวกับการสังเคราะห์เสียงด้วยแบบจำลองการเปลี่ยนรูปร่างช่องทางเดินเสียง ซึ่งเป็นข้อมูลสังเขปเกี่ยวกับแบบจำลองที่ใช้ในการสังเคราะห์เสียงและพารามิเตอร์ที่ใช้ในแบบจำลองประกอบไปด้วยข้อมูลรูปร่างช่องทางเดินเสียงและพื้นที่หน้าตัดของสระภาษาอังกฤษจำนวน 6 หน่วยเสียงที่มีสัทอักษรตรงกันกับภาษาไทยคือ [ii] (อี), [ee] (เอ), [ae] (แอ), [uh] (เออ), [aa] (อา) และ [uu] (อู)

1.3 วัตถุประสงค์

สร้างแบบจำลองสำหรับการสังเคราะห์เสียงโดยวิธี articulatory synthesis และหาค่าพารามิเตอร์ที่เหมาะสมสำหรับใช้ในการสังเคราะห์เสียงพูดภาษาไทย เพื่อพัฒนาให้ใช้เนื้อที่จัดเก็บน้อยและได้คุณภาพเสียงที่ดี

1.4 ขอบเขตของงานวิจัย

1.4.1 สังเคราะห์เสียงสระเดี่ยวภาษาไทยได้แก่ อะ อา อิ อี อึ อือ อุ อู เอะ เอ แอะ แอ โอะ โอ เอาะ ออ เออะ เออ

1.4.2 สังเคราะห์เสียงสระประสมภาษาไทยได้แก่ เอียะ เอีย เอือะ เอือ อัวะ อัว

1.4.3 สังเคราะห์ตามเสียงวรรณยุกต์ภาษาไทย สามัญ เอก โท ตรี จัตวา

1.4.4 สังเคราะห์เป็นเสียงผู้ชายเท่านั้น

1.4.5 พัฒนาโปรแกรมและสังเคราะห์เสียงโดยวิธี articulatory synthesis

1.5 ขั้นตอนและวิธีการวิจัย

1.5.1 ศึกษาทฤษฎีด้านภาษาศาสตร์

1.5.2 ศึกษาการทำงานของอวัยวะที่ใช้ในการออกเสียง

1.5.3 ศึกษาแบบจำลองที่ใช้ในการสังเคราะห์เสียง

1.5.4 หาค่าพารามิเตอร์ที่ใช้ในการควบคุมแบบจำลอง

1.5.5 สร้างโปรแกรมที่ใช้สำหรับวิเคราะห์และสังเคราะห์เสียง

1.5.6 ทดสอบการทำงานของโปรแกรม

1.5.7 สรุปผลการวิจัย ประเมินผล และเสนอแนะแนวทางในการวิจัยต่อไป

1.6 ประโยชน์ที่คาดว่าจะได้รับ

1.6.1 ได้วิธีการใหม่ในการสังเคราะห์เสียงภาษาไทย ซึ่งกรรมวิธีนี้มีข้อดีตรงที่ทำงานได้เร็วและใช้เนื้อที่จัดเก็บน้อย และระบบสามารถสังเคราะห์ได้เสียงที่ต่อเนื่อง

1.6.2 กรรมวิธีวิเคราะห์และสังเคราะห์เสียง สามารถนำไปใช้ศึกษาวิจัยเกี่ยวกับด้านเสียงพูดได้ เช่น ระบบรู้จำเสียงพูด การเพิ่มคุณภาพเสียง เป็นต้น

1.6.3 สามารถใช้เป็นพื้นฐานสำหรับการศึกษาและพัฒนาในระดับที่สูงขึ้น

1.6.4 เป็นพื้นฐานด้านการศึกษาลักษณะการออกเสียงด้านภาษาศาสตร์