

บทที่ 1

บทนำ

ที่มาของปัญหา

มีความพยายามที่จะทำให้เครื่องจักรได้ตอบกับมนุษย์โดยใช้เสียงพูดเกิดขึ้น เพราะเสียงพูดทำให้การสื่อสารระหว่างมนุษย์กับเครื่องจักรสะดวกและมีประสิทธิภาพมากขึ้น เนื่องจากปัจจุบันเทคโนโลยีทางด้านเสียงพูดได้มีการพัฒนาไปมากเช่นให้คอมพิวเตอร์รับคำสั่งจากเสียงพูดหรือพูดได้ตอบกับผู้ใช้ การที่จะทำให้คอมพิวเตอร์พูดได้นี้จะต้องมีระบบสังเคราะห์เสียง ระบบสังเคราะห์เสียงแบบแปลงข้อความเป็นเสียงพูด (Text to Speech System) เป็นระบบหนึ่งที่น่าไปประยุกต์ใช้ได้อย่างกว้างขวาง เช่นให้คอมพิวเตอร์อ่านหนังสือเป็นคำพูดโดยผู้ใช้ไม่ต้องกวาดสายตาอ่านเอง นำไปใช้ในดิกชันนารีแบบพูดได้ หรือนำไปใช้เพื่อช่วยเหลือคนตาบอดในการทำงานพิมพ์บนคอมพิวเตอร์ ที่ผ่านมามีการพัฒนาระบบสังเคราะห์เสียงขึ้นมาหลายแบบ ซึ่งแต่ละแบบก็มีข้อดีและข้อเสียแตกต่างกันไปแต่ทุกแบบก็มีจุดมุ่งหมายที่ว่าเสียงที่สังเคราะห์ขึ้นควรใกล้เคียงเสียงคนพูด หน่วยเสียงที่ใช้ควรมีขนาดเล็กและจำนวนไม่มากเพื่อให้ประหยัดเนื้อที่เก็บ อีกทั้งควรสังเคราะห์เสียงได้ครบทั้งภาษา

การวิจัยเรื่องการสังเคราะห์เสียงในประเทศไทยที่ผ่านมามีการใช้หน่วยหลายชนิดในระบบสังเคราะห์เสียง เช่น คำ (Words) พยางค์ (Syllables) หน่วยคู่เสียง (Diphones) และหน่วยอัครพยางค์ (Demisyllables) ในอดีตที่ผ่านมามีระบบสังเคราะห์เสียงที่ใช้คำในการสังเคราะห์เช่น ยืน ภู่วรรณ (2527), กฤษดา เรยส (2530), อาทร นันทียกุล (2533) งานของ ชัชวาลิต สรวารี (Saravari 1983) ได้ใช้หน่วยอัครพยางค์ การใช้คำเป็นหน่วยเสียงจะให้คุณภาพเสียงที่ดีแต่จะเปลืองเนื้อที่ในการเก็บมากหน่วยพยางค์ก็ให้คุณภาพเสียงที่ดีเช่นกันแต่ก็ยังใช้เนื้อที่เก็บหน่วยเสียงค่อนข้างมาก หน่วยคู่เสียงและหน่วยอัครพยางค์จะใช้เนื้อที่เก็บน้อยลงมากในขณะที่ยังให้คุณภาพเสียงที่ค่อนข้างดี งานที่กล่าวมาข้างต้นจะเน้นเกี่ยวกับเทคนิคของการสังเคราะห์เสียงและชนิดของหน่วยเสียง ยังไม่ได้มีการประยุกต์ความรู้ทางด้านภาษาศาสตร์เข้าไปใช้มากนัก ซึ่งถ้านำความรู้ทางด้านภาษาศาสตร์มาประยุกต์ใช้ในการสังเคราะห์เสียงจะสามารถทำให้ลดจำนวนหน่วยเสียงลงไปได้

วิทยานิพนธ์นี้ได้เสนอการใช้หน่วยเสียงอนุภาคในการสังเคราะห์เสียงสระ หน่วยเสียงอนุภาคคือ คลื่นเสียงที่ตัดออกมาจากคลื่นความถี่แบบกึ่งรายคาบ (Quasi-periodic) ของเสียงสระเพียงลูกคลื่นเดียว และได้นำเอาความรู้ทางด้านภาษาศาสตร์เกี่ยวกับเรื่องคุณลักษณะของความถี่หลักมูลกับเสียงวรรณยุกต์ในภาษาไทยเข้ามาใช้ในการสังเคราะห์เสียง หน่วยเสียงอนุภาคสามารถนำมาสังเคราะห์เสียงสระและวรรณยุกต์ขึ้นใหม่ได้ โดยหน่วยเสียงอนุภาคนี้จะมีขนาดเล็กมากเมื่อเทียบกับส่วนสระของพยางค์ สามารถที่จะนำหน่วยเสียงอนุภาคมาสังเคราะห์เสียงสระและวรรณยุกต์

ต่าง ๆ ได้โดยการเปลี่ยนค่าความถี่หลักมูลของมันให้มีค่าเป็นไปตามรูปแบบความถี่หลักมูลของ
วรรณยุกต์ที่ต้องการ

วัตถุประสงค์

1. เพื่อพัฒนาฐานข้อมูลหน่วยเสียงที่ใช้หน่วยเสียงอนุภาคเป็นหลักในการสังเคราะห์เสียง
2. เพื่อลดขนาดหน่วยเสียงให้เล็กลง
3. เพื่อสังเคราะห์เสียงวรรณยุกต์ขึ้นใหม่จากหน่วยเสียงอนุภาค
4. เพื่อสังเคราะห์เสียงสระขึ้นใหม่จากหน่วยเสียงอนุภาค

แนวความคิด

เสียงในพยางค์เปิดจะประกอบไปด้วย พยัญชนะต้นหรือพยัญชนะควบกล้ำ สระ และ
วรรณยุกต์ ซึ่งวรรณยุกต์นี้เป็นคุณลักษณะที่กลมกลืนอยู่ในเสียงสระ ไม่สามารถมองเห็นได้เด่นชัด
แต่สามารถวิเคราะห์ออกมาแสดงให้เห็นความแตกต่างได้ เสียงสระจะมีคุณสมบัติที่เป็นคลื่นเสียง
แบบเป็นรายคาบซึ่งมีรูปคลื่นคล้าย ๆ กัน แต่มีความถี่หลักมูล (Fundamental Frequency) ซึ่งเกิด
จากการสั่นของเส้นเสียงและรูปร่างของคลื่นเสียงที่ต่างกันไป ซึ่งความแตกต่างนี้ขึ้นอยู่กับผู้ออกเสียง
และเสียงของพยางค์นั้น ๆ ว่าเป็นสระ และวรรณยุกต์อะไร ในผู้พูดคนเดียวกันถ้าผู้พูดเปล่งเสียงออก
มาในสระเดียวกันแต่คนละวรรณยุกต์ เช่น อา อ่า อ้า เป็นคลื่นเสียงที่เป็นรายคาบมีรูปร่างคลื่นเสียง
คล้ายกันแต่มีค่าของความถี่หลักมูลไม่เท่ากัน แต่ในผู้พูดต่างคนกันจะทำให้ค่าของความถี่หลักมูล
ของแต่ละวรรณยุกต์คล้ายกับผู้พูดคนอื่นในขณะที่รูปคลื่นของสระอาจไม่เหมือนกัน จากคุณสมบัติ
ข้างต้นเราสามารถที่จะเก็บเฉพาะคลื่นเสียงที่ซ้ำกันของเสียงสระเอาไว้เพียงลูกเดียวได้โดยที่ไม่ต้อง
เก็บทั้งหมดแต่จะเก็บกรอบคลื่นแอมพลิจูด (Amplitude Envelope) ซึ่งเป็นตัวกำหนดความดังเบาของ
เสียงและค่าของความถี่หลักมูลซึ่งเป็นตัวกำหนดเสียงสูงต่ำของแต่ละวรรณยุกต์เอาไว้แทน ซึ่งจะทำ
ให้ประหยัดเนื้อที่เก็บลงได้มากเนื่องจากสระเป็นแกนของพยางค์ซึ่งมีความยาวมากกว่าส่วนอื่นใน
พยางค์แต่จะเก็บแทนด้วยหน่วยเสียงอนุภาคซึ่งเป็นลูกคลื่นเล็ก ๆ เพียงลูกเดียว

ในการทำฐานข้อมูลหน่วยเสียง จะนำพยางค์ที่เป็นต้นแบบมาตัดเอาพยัญชนะเก็บไว้ นำ
ส่วนสระมาวิเคราะห์หากกรอบคลื่นแอมพลิจูด หากค่าความถี่หลักมูล และตัดหน่วยเสียงอนุภาคเก็บไว้
เป็นฐานข้อมูล เมื่อจะสังเคราะห์กลับจะนำเอาหน่วยเสียงอนุภาคของสระที่ต้องการมาเปลี่ยนความถี่
หลักมูลและเปลี่ยนแปลงแอมพลิจูดให้ได้ตามรูปแบบที่ต้องการ เชื่อมต่อกันให้ได้ความยาวตามเสียง
พูดธรรมชาติและนำมาเชื่อมต่อกับหน่วยเสียงพยัญชนะ แล้วทำการเล่นกลับเป็นเสียงสังเคราะห์

ขอบเขตงานวิจัย

1. สังเคราะห์เสียงพยางค์เปิดในภาษาไทยเฉพาะสระเดี่ยวเสียงสั้น และสระเดี่ยวเสียงยาวได้แก่ อะ อา อี อี อี อือ อู อู เอะ เอ แอะ แอ โอะ โอ เอาะ ออ เออะ เออ รวม 18 เสียง
2. สังเคราะห์เสียงพยางค์เปิดตามเสียงวรรณยุกต์ทุกวรรณยุกต์ คือ สามัญ เอก โท ตรี และ จัตวา
3. ใช้เสียงที่บันทึกด้วยความละเอียด 16 บิต ความถี่ในการซึกตัวอย่าง 22 กิโลเฮิรตซ์ ด้วยการด์เสียง Soundblaster AWE64 Value ของบริษัท Creative Labs, Inc.
4. พัฒนาโปรแกรมวิเคราะห์และสังเคราะห์เสียงบนโปรแกรม Matlab

ขั้นตอนการวิจัย

1. ศึกษาทฤษฎีทางด้านภาษาศาสตร์ในด้านคุณสมบัติของเสียงในภาษาไทย
2. ศึกษากรรมวิธีทางคณิตศาสตร์ที่จะนำมาวิเคราะห์และสังเคราะห์เสียงพูด
3. สร้างโปรแกรมวิเคราะห์และสังเคราะห์เสียงพูด
4. บันทึกเสียงพูดเพื่อที่จะนำมาวิเคราะห์และเก็บผลการวิเคราะห์เป็นฐานข้อมูลเสียงพูดที่จะใช้ในการสังเคราะห์เสียง
5. กำหนดเสียงที่จะนำมาทดสอบการระบุเสียงวรรณยุกต์ (Identification Test) และการทดสอบทัศนคติต่อคุณภาพเสียง (Quality Judgement Test)
6. ทำการทดสอบการรับฟังกับกลุ่มตัวอย่างซึ่งเป็นนิสิตคณะอักษรศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย จำนวน 200 คน ซึ่งแบ่งเป็นการทดสอบการระบุเสียงวรรณยุกต์จำนวน 100 คน และ การทดสอบทัศนคติต่อคุณภาพเสียงจำนวน 100 คน ด้วยสระอา สระอี และสระอุ พยัญชนะ ซ เพื่อหาตัวแทนหน่วยเสียงที่จะนำมาเป็นฐานข้อมูล
7. สรุปผลการวิจัย ประเมินผล และเสนอแนะแนวทางในการวิจัยต่อไป
8. เสนอรายงานการวิจัยในรูปวิทยานิพนธ์

ประโยชน์ที่คาดว่าจะได้รับจากงานวิจัย

1. ได้วิธีการใหม่ในการสังเคราะห์เสียง
2. นำไปประยุกต์ใช้งานกับงานด้านภาษาศาสตร์ได้
3. สามารถสังเคราะห์เสียงสระและวรรณยุกต์ขึ้นใหม่ได้
4. สามารถควบคุมท่วงทำนองของเสียงสังเคราะห์ได้