สำนักหอสมุดกลาง พระจอมเกล้าลาดกระบัง

การสังเคราะห์เสียงภาษาไทยโดยใช้ HTK

THAI VOICE SYNTHESIS USING HTK

นายเสริมศักดิ์ กิตติธนนพร

นายอิทธิพล ดิลกรัตนพิจิตร

เลขหมู่.....
เลขทะเบียน 42801
วัน, เดือน, ปี 10 ส.ค. 2545

.b.....
.i.....

ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตร์บัณฑิต

ภาควิชาวิศวกรรมคอมพิวเตอร์

คณะวิศวกรรมศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ปีการศึกษา 2543

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การสังเคราะห์เสียงภาษาไทยโดยใช้ HTK

THAI VOICE SYNTHESIS USING HTK

โดย

นายเสริมศักดิ์ กิตติธนนพร

นายอิทธิพล ดิลกรัตนพิจิตร

อาจารย์ที่ปรึกษา

ผศ.ดร.บุญธีร์ เครือตราชู

ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตร์บัณฑิต

ภาควิชาวิศวกรรมคอมพิวเตอร์

คณะวิศวกรรมศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ปีการศึกษา 2543

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ปริญญานิพนธ์ปีการศึกษา 2543

ภาควิชา วิศวกรรมคอมพิวเตอร์

คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เรื่อง การสังเคราะห์เสียงภาษาไทยโดยใช้ HTK

THAI VOICE SYNTHESIS USING HTK

ผู้จัดทำ

1. นายเสริมศักดิ์ กิตติธนนพร รหัสประจำตัว 40010922
2. นายอิทธิพล ดิลกรัตนพิจิตร รหัสประจำตัว 40011010

อาจารย์ที่ปรึกษา

(ผศ.ดร.บุญธีร์ เครือตราชู)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การสังเคราะห์เสียงภาษาไทยโดยใช้ HTK

นายเสริมศักดิ์ กิตติธนนพร 40010922
นายอิทธิพล ดิลกรัตนพิจิตร 40011010
ผศ.ดร.บุญธีร์ เครือตราชู อาจารย์ที่ปรึกษา
ปีการศึกษา 2543

บทคัดย่อ

ปริญญานิพนธ์ฉบับนี้นำเสนอการสังเคราะห์เสียงภาษาไทยโดยในเบื้องต้นจะใช้โปรแกรม Hidden markov model Tool Kit ช่วยแบ่งไฟล์เสียงพูดออกเป็นไฟล์เสียงในระดับหน่วยเสียง(Phoneme) แต่ปรากฏว่าโปรแกรม HTK ไม่สามารถบอกช่วงเวลาของหน่วยเสียงต่าง ๆ ในไฟล์เสียงได้ถูกต้อง จึงต้องเปลี่ยนมาใช้การแบ่งหน่วยเสียงด้วยการ สังเกตด้วยตนเองแทน โดยทางผู้จัดทำได้กำหนดหน่วยเสียงย่อยขึ้นมาเอง

โปรแกรมสังเคราะห์เสียงภาษาไทยในปริญญานิพนธ์ฉบับนี้ จะรับอินพุตเป็นข้อความภาษาไทยและจะทำการตัดคำโดยใช้อัลกอริทึม Word Bucket แล้วนำคำที่ตัดได้ไปค้นหาคำอ่านจากพจนานุกรมคำศัพท์ คำอ่านจะอยู่ในรูปของสัญลักษณ์แทนเสียง โดยสัญลักษณ์แทนเสียงจะบอกถึงชื่อไฟล์หน่วยเสียงที่ต้องใช้ในการสร้างเสียงของคำนั้น ๆ จากนั้นจึงนำไฟล์หน่วยเสียงเหล่านั้นมาเรียงต่อกันจนครบหมดทั้งข้อความแล้วจึงนำไฟล์เสียงที่ได้ไปสร้างเสียงออกมาทางลำโพง เสียงที่ได้จากการสังเคราะห์เสียงในปริญญานิพนธ์ฉบับนี้อยู่ในระดับพอใช้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

THAI VOICE SYNTHESIS USING HTK

Sermsak Kittitanonporn

Itthipon Dilokrattanapijit

Assoc.Prof. Dr.Boontee Kruatrachue Adviser

ABSTRACT

This research presents Thai voice synthesis. At first, HTK (Hidden markov Tool Kit) was used to help separate phonemes in recorded wave files, but it couldn't show the correct time interval of the phonemes in wave files. So manual phoneme separating is used instead. The phonemes are defined to implement the voice synthesis.

Thai voice synthesis in this research has a Thai message as input. Then the message is separated into word level using Word Bucket algorithm. Each separated word is used to look up its phonetic symbols in the dictionary file. Each symbol is the name of the wave file for its sound. So each of those files is merged into one file to produce the sound of that word. This is repeated until every word in the message is done. The file that contains every word is played to generate the voices of the input message. The voice quality is acceptable.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กิตติกรรมประกาศ

ปริญญานิพนธ์ฉบับนี้คงไม่อาจสำเร็จลงได้ด้วยดีหากไม่ได้รับความช่วยเหลือและร่วมมือจากหลาย ๆ ฝ่ายด้วยกัน บุคคลแรกที่ต้องกล่าวถึงเพราะเป็นส่วนสำคัญที่ทำให้การทำงานในครั้งนี้สำเร็จลงได้ก็คือ อาจารย์ บุญธีร์ เครือตราชู อาจารย์ที่ปรึกษาที่ให้ความเอาใจใส่ ดูแล ให้คำแนะนำและช่วยเหลือแก่ข้าพเจ้าเป็นอย่างดีเสมอมา ซึ่งต้องขอขอบพระคุณเป็นอย่างมาก

บุคคลสำคัญที่สุดที่ทำให้ข้าพเจ้ามีวันนี้คือบิดา มารดาอันเป็นที่เคารพรักยิ่ง ซึ่งได้เลี้ยงดูข้าพเจ้ามาเป็นอย่างดี พร้อมทั้งให้โอกาสในการศึกษาอย่างเต็มที่ และยังให้กำลังใจ เอาใจใส่เสมอมาในทุก ๆ ด้านอันหาที่เปรียบมิได้ ข้าพเจ้าขอระลึกในพระคุณอันสุดประมาณและขอกราบขอบพระคุณมา ณ ที่นี้

บุคคลสุดท้ายที่อยากจะเอ่ยถึงคือรุ่นพี่และเพื่อน ๆ ทุกคนที่คอยอำนวยความสะดวกในการทำงานและให้คำแนะนำช่วยเหลือต่าง ๆ มาเป็นอย่างดี

เสริมศักดิ์ กิตติธนนพร

อิทธิพล ดิลกรัตนพิจิตร

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

	หน้า
บทคัดย่อภาษาไทย	I
บทคัดย่อภาษาอังกฤษ	II
กิตติกรรมประกาศ	III
สารบัญ	IV
สารบัญตาราง	VI
สารบัญภาพ	VII
บทที่ 1 บทนำ	1
1.1 ความสำคัญและที่มา	1
1.2 วัตถุประสงค์ของงานวิจัย	1
1.3 ขอบเขตของงานวิจัย	1
1.4 วิธีการดำเนินงาน	2
บทที่ 2 ระบบเสียงภาษาไทย	3
2.1 บทนำ	3
2.2 ความหมายของเสียง	3
2.3 การศึกษาเกี่ยวกับเสียง	3
2.4 อวัยวะที่ใช้ในการออกเสียงพูด	4
2.5 เสียงในภาษาไทย	6
2.6 เสียงของอักษรไทย	6
บทที่ 3 การตัดคำ	10
3.1 บทนำ	10
3.2 พจนานุกรมคำศัพท์	10
3.3 วิธีการตัดคำโดยใช้อัลกอริทึม Word Bucket	11
บทที่ 4 การกำหนดหน่วยเสียงและการสังเคราะห์เสียง	16
4.1 บทนำ	16
4.2 หน่วยเสียงพื้นฐาน	16
4.3 เทคนิคที่ใช้ในการสร้างเสียงพูด	23
4.4 การรวมหน่วยเสียง	23
บทที่ 5 การใช้งานโปรแกรม HTK	24
5.1 บทนำ	24
5.2 ขั้นตอนการใช้งานโปรแกรม HTK	24

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.3 สรุปการใช้งาน 36
บทที่ 6 การบันทึกเสียงและการแบ่งเสียงออกเป็นโฟนีม 38
6.1 บทนำ 38
6.2 ขั้นตอนในการทำงาน 38
6.3 ข้อสังเกตจากการเลือกโฟนีมที่ต้องการ 43
6.4 สรุปหลักในการบันทึกเสียง 45
6.5 สรุปหลักในการแบ่งไฟล์เสียง 45
บทที่ 7 การออกแบบ และการสร้างโปรแกรมสังเคราะห์เสียงภาษาไทย 47
7.1 บทนำ 47
7.2 แนวคิดในการออกแบบโปรแกรม 47
7.3 ส่วนประกอบหลักภายในโปรแกรม 48
7.4 การสร้างโปรแกรม 49
7.5 ภาพรวมการทำงานของโปรแกรม 56
บทที่ 8 ผลการทดลอง 57
8.1 ไฟล์เสียงที่ได้จากการทำงานของโปรแกรม 57
8.2 การเปรียบเทียบรูปคลื่นเสียงในเชิงเวลา 59
บทที่ 9 บทวิจารณ์และสรุป 61
9.1 บทวิจารณ์ 61
9.2 สรุป 63
9.3 แนวทางการพัฒนาต่อไป 63
ภาคผนวก 64
บรรณานุกรม 67

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

หน้า

ตารางที่

3-1 โครงสร้างของพจนานุกรมคำศัพท์ 11
4-1 เสียงของพยัญชนะต้นและสัญลักษณ์แทนเสียง 17
4-2 เสียงของสระที่ผันตามเสียงวรรณยุกต์และสัญลักษณ์แทนเสียง 18
4-3 เสียงของพยัญชนะสะกดที่ผันตามเสียงวรรณยุกต์และสัญลักษณ์แทนเสียง 20
6-1 คำที่เลือกมาใช้ในการแบ่งเสียงออกเป็นโฟนีมของสระที่ผันเสียงวรรณยุกต์แล้ว 40
6-2 คำที่เลือกมาใช้ในการแบ่งเสียงออกเป็นโฟนีมของพยัญชนะต้น 41
6-3 คำที่เลือกมาใช้ในการแบ่งเสียงออกเป็นโฟนีมของพยัญชนะสะกดที่ผันวรรณยุกต์แล้ว 41
8-1 การเปรียบเทียบรูปคลื่นในเชิงเวลา 59

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญภาพ

หน้า

รูปที่

2.1 ความสัมพันธ์ระหว่างส่วนย่อยทั้ง 3 ของวิชาสัทศาสตร์ 4
2.2 อวัยวะที่ใช้ในการออกเสียงพูด 4
3.1 โครงสร้างของประโยคภาษาไทย 10
3.2 วิธีการตัดคำโดยใช้ Word Bucket 12
3.3 แผนภูมิสายงานแสดงขั้นตอนการทำงานของอัลกอริทึม Word Bucket 14
5.1 Flow-Chart ของการใช้คำสั่ง HLEd 27
5.2 Flow-Chart ของการใช้ HCopy 29
5.3 รูปแบบของ Master Macro Files 32
5.4 Flow-Chart ของการใช้ HERest 32
5.5 รูปของ silence model และ short pause 33
5.6 Flow-chart ของการทำคำสั่ง HHEd 34
5.7 การใช้คำสั่ง HVite 35
5.8 ไฟล์ที่ทำการบันทึกคำว่า “ก้ม” 36
5.9 คลื่นเสียงของโฟนีม “ก” ที่โปรแกรม HTK ระบุมา 37
5.10 คลื่นเสียงของโฟนีม “ก” ที่ควรจะได้จริง 37
6.1 การบันทึกเสียงในลักษณะเป็นประโยค 38
6.2 การบันทึกเสียงในลักษณะเป็นคำเดี่ยว ๆ 39
6.3 ไฟล์เสียงของประโยค “สถาบันเทคโนโลยี” 42
6.4 ส่วนของโฟนีมที่ต้องการ (/ah4/) 42
6.5 เสียงของพยัญชนะต้น /ล/ จากคำว่า “ล้าน” 43
6.6 เสียงของสระ /อา/ ที่ผันวรรณยุกต์ /ตรี/ จากคำว่า “ล้าน” 43
6.7 เสียงของตัวสะกด /น/ ที่ผันวรรณยุกต์ /ตรี/ จากคำว่า “ล้าน” 44
6.8 เสียงของตัวสะกด /ว/ ที่ผันวรรณยุกต์ /สามัญ/ จากคำว่า “กาว” 44
6.9 การแบ่งไฟล์เสียงที่ถูกต้อง 45
6.10 การแบ่งไฟล์เสียงที่ผิด 46
6.11 การต่อไฟล์เสียงที่ถูกต้องเข้ากันได้พอดี 46
6.12 การต่อไฟล์เสียงที่ผิดเพราะการแบ่งไฟล์เสียงไม่ถูกต้อง 46
7.1 ภาพรวมของโปรแกรมสังเคราะห์เสียงภาษาไทย 47

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

7.2 ส่วนประกอบหลักภายในโปรแกรมสังเคราะห์เสียงภาษาไทย 48
7.3 แผนภูมิสายงานของการรับข้อความจากไฟล์ข้อความ 49
7.4 ตัวอย่างโปรแกรม ที่รับข้อความจากไฟล์ข้อความหรือจากผู้ใช้ผ่านแป้นพิมพ์ 50
7.5 การทำงานของเธรดตัดคำ 53
7.6 การทำงานของเธรดรวมไฟล์และเล่นไฟล์เสียง 55
8.1 รูปคลื่นของคำว่า "อ่านออกเสียงภาษาไทย" ที่ได้จากการรวมไฟล์เสียงที่เตรียมไว้ 58
8.2 รูปคลื่นของคำว่า "อ่านออกเสียงภาษาไทย" ที่ได้จากการบันทึกเสียงพูดจริง 58

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1

บทนำ

1.1 ความสำคัญและที่มา

ในปัจจุบัน เทคโนโลยีทางด้านคอมพิวเตอร์ได้มีการพัฒนาเพิ่มมากขึ้นเรื่อยๆอย่างรวดเร็วทั้งทางด้านฮาร์ดแวร์และซอร์ฟแวร์ โดยทางด้านฮาร์ดแวร์นั้นก็มีการพัฒนาให้คอมพิวเตอร์มีความเร็วในการทำงานเพิ่มมากขึ้น รวมถึงอุปกรณ์ที่ช่วยในการอำนวยความสะดวกในการทำงานออกมามากมาย ทั้ง พรินเตอร์ สแกนเนอร์และกล้องดิจิตอลเป็นต้น และเมื่อเรามาดูทางด้านซอร์ฟแวร์ก็มีการพัฒนาโปรแกรมออกมามากมาย เช่นโปรแกรมที่ช่วยผู้บริหารในการทำงานตั้งแต่ระดับล่างที่เป็นผู้ปฏิบัติงานประจำไปจนถึงระดับสูงสุดที่เป็นผู้กำหนดนโยบายของบริษัท และยังมีการพัฒนาทางด้านปัญญาประดิษฐ์(Artificial Intelligent) ซึ่งจุดประสงค์หลักก็คือทำให้มนุษย์สามารถใช้งานคอมพิวเตอร์ได้แบบเป็นธรรมชาติมากที่สุด ซึ่งหนึ่งในจำนวนโปรแกรมพวกนั้นก็คือโปรแกรมที่ใช้เสียงในการสั่งงานคอมพิวเตอร์ ซึ่งขณะนี้ก็ได้รับการพัฒนาไปมากพอสมควรแล้ว เพราะมีผู้พัฒนาอยู่มากมาย เนื่องจากสามารถนำไปใช้ได้โดยทั่วๆไป แต่ก็ยังมีโปรแกรมที่ต้องเกี่ยวข้องกับเสียงอีกลักษณะหนึ่งคือโปรแกรมที่ใช้ในการอ่านออกเสียง ซึ่งจะมีประโยชน์มากสำหรับผู้ที่พิการทางสายตาที่ไม่อาจมองเห็นตัวหนังสือได้ ทำให้บุคคลเหล่านี้ไม่สามารถที่จะเรียนรู้สิ่งใดๆจากหนังสือได้ ถึงแม้ว่าจะมีอักษรเบลล์ที่ช่วยให้เขาเหล่านั้นสามารถอ่านหนังสือได้ แต่ก็ยังไม่สะดวกเท่าที่ควร และยังต้องเสียเวลาในการฝึกฝนทักษะการอ่านอีกพอสมควรทีเดียว ดังนั้นเราจึงคิดว่าการพัฒนาโปรแกรมทางด้านนี้จะช่วยให้พวกเขาได้รับความสะดวกในการเรียนรู้สิ่งใหม่ๆ เพิ่มมากขึ้น ขอเพียงมีแฟ้มข้อมูลที่ต้องการอ่านและคอมพิวเตอร์ที่มีการ์ดเสียง(sound card) พร้อมลำโพงด้วยเท่านั้น

1.2 วัตถุประสงค์ของงานวิจัย

สร้างโปรแกรมสังเคราะห์เสียงภาษาไทยด้วยคอมพิวเตอร์ เพื่อทำการแปลงข้อความจากแฟ้มข้อความภาษาไทยให้เป็นเสียงพูดภาษาไทย โดยให้มีความสามารถในการรับข้อความได้ทั้งจากแฟ้มข้อความและแป้นพิมพ์ มีความสามารถในการตัดคำได้ถูกต้องตามหลักภาษาไทย และสามารถสังเคราะห์เสียงภาษาไทยได้ใกล้เคียงกับเสียงธรรมชาติมากที่สุดโดยใช้ทรัพยากรน้อยที่สุด

1.3 ขอบเขตของงานวิจัย

การออกแบบและสร้างโปรแกรมสังเคราะห์เสียงภาษาไทยในครั้งนี้จะมุ่งเน้นที่การทำงานหลักๆ คือสามารถตัดคำภาษาไทยที่ป้อนเข้ามาเป็นประโยคออกเป็นคำๆได้ถูกต้อง และสามารถสร้างเสียงให้ออกมาอยู่ในระดับที่น่าพอใจ สามารถนำไปใช้งานได้จริง แต่การที่จะทำให้การทำงานในส่วนต่างๆเช่น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การตัดคำหรือการสร้างเสียงให้ออกมาดีที่สุดนั้นจะต้องใช้เวลาและทรัพยากรเป็นจำนวนมาก ด้วยข้อจำกัดอันนี้เราจึงต้องเลือกวิธีที่คิดว่าเหมาะสมที่สุด แม้ว่าจะไม่ใช่วิธีที่ดีที่สุด แต่ผลงานที่ออกมาก็อยู่ในระดับที่ดีพอสมควร โดยในการตัดคำภาษาไทยเราจะตัดได้เฉพาะคำที่มีอยู่ในพจนานุกรมคำศัพท์เท่านั้น ส่วนการสร้างเสียงภาษาไทยในครั้งนี้ เราได้เลือกใช้วิธีการในการสังเคราะห์เสียงจากหน่วยเสียงที่เล็กกว่าหน่วยพยางค์ซึ่งเรียกว่าโฟนีม(phoneme) เพราะเป็นวิธีที่ประหยัดพื้นที่ในการจัดเก็บข้อมูลมากที่สุด และสามารถสังเคราะห์เสียงภาษาไทยได้ครอบคลุมทุกเสียง แม้ว่าจะทำให้ไม่ได้เสียงที่เป็นจริงตามธรรมชาติเหมือนกับวิธีการบันทึกเสียงพยางค์จริงก็ตาม โดยใน 1 พยางค์จะประกอบไปด้วย 3 ส่วนหลักๆคือ เสียงพยัญชนะต้น เสียงสระและเสียงวรรณยุกต์ โดยในบางพยางค์อาจจะมีเสียงตัวสะกดเพิ่มเข้ามาอีกด้วย

1.4 วิธีการดำเนินงาน

โครงงานนี้จะเริ่มด้วยการศึกษาถึงทฤษฎีพื้นฐานต่างๆที่เกี่ยวข้องกับโครงงาน โดยเริ่มจากระบบเสียงในภาษาไทย การแบ่งประโยคออกเป็นคำ การกำหนดหน่วยเสียงและการสังเคราะห์เสียง การใช้โปรแกรม HTK(Hidden markov model ToolKit) เพื่อช่วยในการแบ่งไฟล์เสียงที่ได้ทำการบันทึกไว้ออกเป็นโฟนีมย่อยๆ การบันทึกเสียงและการแบ่งเสียงออกเป็นโฟนีม ในบทที่ 2-6 ตามลำดับ ต่อจากนั้นก็ทำการออกแบบและพัฒนาโปรแกรมสังเคราะห์เสียงภาษาไทยในบทที่ 7 แล้วทดสอบการทำงานในบทที่ 8 และสุดท้ายคือสรุปและวิจารณ์ผลลัพธ์ของโปรแกรมในบทที่ 9

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2
ระบบเสียงภาษาไทย

2.1 บทนำ

ระบบเสียงภาษาไทยเป็นการศึกษาที่เกี่ยวข้องกับการสร้างเสียงภาษาไทย จะเริ่มจากการศึกษาอวัยวะที่ใช้ในการออกเสียงพูด การศึกษาเสียงในภาษาไทยและการศึกษาเสียงของอักษรไทย ซึ่งเสียงของอักษรไทยจะประกอบไปด้วยเสียงของพยัญชนะ เสียงสระและเสียงวรรณยุกต์

2.2 ความหมายของเสียง

พจนานุกรมฉบับราชบัณฑิตยสถานให้ความหมายของเสียงไว้ว่า "สิ่งที่ได้ยินด้วยหู คำที่เปล่งออกมา เช่น เสียงลือ ว่า คำเล่าลือ ความเห็น ความนิยม คะแนนเสียง" และให้ความหมายของคำว่าเสียงเมื่อนำไปประสมกับคำอื่นบางคำไว้ว่า "เสียงเขียว น.แสดงอาการโกรธจัด เสียงแข็ง น.คำพูดที่กล่าวออกมาโดยไม่ยำเกรง เสียงแตก น.เสียงพูดในตอนแตกเนื้อหนุ่ม เสียงทอง น.เสียงเพราะ ทำให้ค่าตัวในการร้องเพลงสูงขึ้น เสียงหลง น.เสียงที่แผดดังผิดปกติหรือผิดระดับเสียงดนตรี เสียงอ่อน น.คำพูดที่เพลาความแข็งแรงลง" จากความหมายดังกล่าวแล้วจะเห็นว่าความหมายของเสียงมีอยู่ 2 ส่วนสำคัญ คือส่วนแรกหมายถึงสิ่งที่หูอาจหรือสามารถได้ยิน จะเป็นเสียงอะไรก็ได้ เช่น เสียงฝนตก ฟ้าร้อง เสียงวัตถุกระทบกันเป็นต้น ส่วนที่ 2 หมายถึงเสียงพูดของมนุษย์ ส่วนหลังนี้มีความสำคัญที่เราจำเป็นต้องศึกษา เพราะเป็นเสียงในภาษาที่ใช้สื่อสารในชีวิตประจำวันทุกด้าน การศึกษาเสียงพูดคือการศึกษาภาษาที่แท้จริงของมนุษย์

2.3 การศึกษาเกี่ยวกับเสียง

กาญจนา นาคสกุล[1] ให้ความเห็นว่า "เสียงพูดนี้มีลักษณะที่จะอธิบายได้ด้วยหลักเกณฑ์ทางวิทยาศาสตร์ซึ่งเป็นสากล แม้ว่าภาษาหนึ่งๆจะมีเสียงแตกต่างกันออกไปมากบ้างน้อยบ้าง แต่ละเสียงก็สามารถที่จะนำมาพิจารณาและอธิบายให้รู้ลักษณะการออกเสียง และตำแหน่งที่เกิดของเสียงได้ คำอธิบายจะทำให้รู้และเข้าใจลักษณะของเสียงทุกเสียง ทำให้เห็นความแตกต่างและคล้ายคลึงกันของเสียงทุกเสียงได้ ซึ่งจะเป็นประโยชน์อย่างมากต่อการเรียนรู้"

เรื่องการศึกษาเกี่ยวกับเสียงนี้ ประสิทธิ์ กาพย์กลอน[2] ได้สรุปว่ามีวิชาที่ศึกษาเฉพาะอยู่ 2 แขนงใหญ่ๆ คือ

1. สัทศาสตร์(Phonetics) คือการศึกษาเสียงที่เราได้ยินว่าเกิดอย่างไร เป็นเสียงชนิดใดและมีธรรมชาติเป็นอย่างไร ในแขนงนี้จะศึกษาลักษณะของเสียงอย่างละเอียดทุกเสียงที่ได้ยิน แบ่งย่อยออกได้ 3 ส่วนคือ

---

[1] กาญจนา นาคสกุล **ระบบเสียงภาษาไทย**(กรุงเทพฯ:จุฬาลงกรณ์มหาวิทยาลัย,2520),หน้า6-7

[2] ประสิทธิ์ กาพย์กลอน **การศึกษาภาษาไทยแนวภาษาศาสตร์**(กรุงเทพฯ:ไทยวัฒนาพานิช,2516)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.1 สรีระสัทศาสตร์(Articulatory Phonetics) จะศึกษาอวัยวะต่างๆที่ใช้ในการออกเสียง เช่น ใช้อวัยวะส่วนใดทำเสียง เส้นเสียงสั่นสะเทือนหรือไม่ ลมที่พุ่งออกมาแรงหรือไม่ ลมออกมาทางปากหรือทางจมูก เป็นต้น

1.2 กลสัทศาสตร์(Acoustic Phonetics) ศึกษาว่าเสียงเปล่งออกไปในอากาศได้อย่างไร การศึกษาส่วนนี้ต้องอาศัยเครื่องมือวิทยาศาสตร์ประกอบด้วย

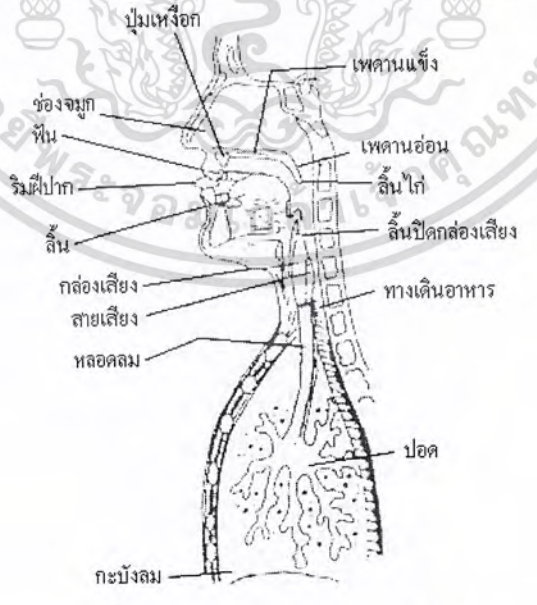
1.3 โสตสัทศาสตร์(Auditory Phonetics) ศึกษาเรื่องการรับฟังเสียงที่เปล่งออกไปว่าผู้รับฟังได้ยินเสียงได้อย่างไร

รูปที่ 2.1 ความสัมพันธ์ระหว่างส่วนย่อยทั้ง 3 ของวิชาสัทศาสตร์

2. วิชาว่าด้วยหน่วยเสียง(Phonemics) คือการศึกษาเรื่องหน่วยเสียงหรือเสียงสำคัญในภาษา เป็นการนำเอาเสียงที่วิเคราะห์แล้วตามวิชาสัทศาสตร์มาจัดเป็นหมวดหมู่ โดยอาศัยหลักการเฉพาะหน่วยเสียง

2.4 อวัยวะที่ใช้ในการออกเสียงพูด

อวัยวะที่ใช้ในการออกสียงพูดก็คืออวัยวะต่างๆที่มีไว้ทำหน้าที่อื่นๆนั่นเอง แต่มนุษย์ใช้อวัยวะเหล่านั้นมาใช้ในการออกเสียงต่างๆตามที่ต้องการได้ แม้ว่าจะมีความสลับซับซ้อนอย่างไรก็ตาม อวัยวะต่างๆมีดังในภาพ



รูปที่ 2.2 อวัยวะที่ใช้ในการออกเสียงพูด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปที่ 2.2 เราสามารถแบ่งอวัยวะที่ใช้ออกเสียงได้เป็น 3 พวกคือ

1. พวกเริ่มต้น คือพวกที่ก่อให้เกิดกระแสลม กำหนดลมและทำให้ลมเคลื่อนที่ ได้แก่ ปอด กะบังลม และกล้ามเนื้อระหว่างซี่โครง
2. พวกที่ใช้ในการออกเสียง ได้แก่ หลอดลม กล่องเสียงและเส้นเสียง
3. พวกที่ทำให้เสียงเปลี่ยนแปลงไปต่างๆ ได้แก่ ช่องปาก ฟัน ปุ่มเหงือก เพดานอ่อน เพดานแข็ง ลิ้น ริมฝีปากรวมทั้งช่องจมูกด้วย

หน้าที่ของอวัยวะต่างๆในการออกเสียงพูด

1. ริมฝีปาก เป็นอวัยวะส่วนที่เคลื่อนไหวได้มาก และทำให้เสียงแตกต่างกันได้มาก เราอาจจะบังคับริมฝีปากให้อยู่ชิดกัน ห่างกัน ยื่นออกหรือห่อกลม ลักษณะริมฝีปากต่างๆนี้ล้วนแต่มีอิทธิพลต่อการออกเสียงและทำให้เสียงแตกต่างกันไปทั้งสิ้น
2. ฟัน เป็นอวัยวะที่เกิดของเสียงหลายชนิด เช่นเมื่อฟันบนกดลงบนริมฝีปากล่าง หรือฟันล่าง ลมที่ผ่านออกมาโดยแรงจะลอดช่องที่พอผ่านได้ออกมา ทำให้เกิดเป็นเสียงชนิดที่เรียกว่าเสียงเสียดแทรก เป็นต้น
3. ปุ่มเหงือก เป็นส่วนนูนออกมาอยู่หลังฟันด้านบน ถ้าเอาลิ้นแตะดูจะรู้สึกว่ามีลักษณะเป็นคลื่น ปุ่มเหงือกเป็นบริเวณที่เกิดเสียงปุ่มเหงือก(alveolar sound)
4. เพดานแข็ง คือส่วนเฉพาะที่โค้งเป็นกระดูกแข็ง
5. เพดานอ่อน คือส่วนของเพดานที่อยู่ต่อจากเพดานแข็งไปข้างใน มีลักษณะเป็นกระดูกอ่อนที่ขยับขึ้น-ลงได้ เวลาหายใจเพดานอ่อนและลิ้นไก่ซึ่งอยู่ตอนปลายจะลดระดับลงมาเปิดช่องให้ลมออกไปทางจมูก เวลาพูดส่วนใหญ่ปลายเพดานอ่อนและลิ้นไก่จะถูกยกขึ้นไปจรดกับหลังคอ นอกจากเวลาออกเสียงนาสิกเท่านั้นที่เพดานอ่อนจะลดระดับลงมาเพื่อให้ลมออกทางช่องจมูก
6. ลิ้นไก่ เป็นก้อนเนื้อเล็กๆ อยู่ต่อปลายเพดานตรงกลางปาก อวัยวะส่วนนี้สั่นรัวได้
7. ลิ้น เป็นส่วนที่เคลื่อนไหวมากที่สุดในการออกเสียงพูด จึงต้องแบ่งออกเป็น 3 ส่วนตามหน้าที่ในการออกเสียงคือ
   7.1 ปลายลิ้น(Blade of the tongue) คือส่วนปลายลิ้นซึ่งสามารถยกขึ้นไปแตะอวัยวะส่วนต่างๆ ในปากตอนบนได้โดยง่าย
   7.2 หน้าลิ้น(Front of the tongue) คือลิ้นที่อยู่ตรงข้ามกับเพดานแข็ง
   7.3 หลังลิ้น(Back of the tongue) คือส่วนของลิ้นที่อยู่ตรงข้ามกับเพดานอ่อน
8. ลิ้นปิดหลอดลม เป็นก้อนเนื้อเล็กๆคล้ายลิ้นไก่ อยู่ต่อโคนลิ้นลงไปในคอ มีหน้าที่ปิดช่องลมเมื่อเวลารับประทานอาหาร และเปิดช่องลมเมื่อพูด
9. กรวยคอ คือโพรงคอที่อยู่ถัดจากปากลงไปจนถึงเส้นเสียง
10. ช่องจมูก หมายถึงโพรงในช่องจมูกซึ่งอยู่เหนือลิ้นไก่ขึ้นไป เป็นช่องที่ลมซึ่งผ่านเส้นเสียงขึ้นมาจะผ่านออกไปทางจมูกได้เมื่อเวลาหายใจและเวลาออกเสียงนาสิก ในเวลาที่พูดเสียงอื่นลิ้นไก่จะถูกยกขึ้นไปปิดช่องจมูก เพื่อให้ลมออกมาทางปาก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

11. เส้นเสียง เป็นอวัยวะสำคัญที่เกิดของเสียง มีลักษณะเป็นกล้ามเนื้อ 2 แผ่นปิดขวางอยู่บริเวณปากช่องหลอดลมจากด้านหลังมาด้านหน้า ระหว่างเส้นเสียงจะมีช่องว่างซึ่งเป็นทางผ่านให้ลมเข้าถึงปอดและออกมาจากปอดได้ ช่องว่างนี้เรียกว่าช่องระหว่างเส้นเสียง(Glottis) เส้นเสียงทั้งสองสามารถดึงออกให้ห่างจากกันหรือดึงเข้าหากันได้ ซึ่งเส้นเสียงนี้เป็นส่วนสำคัญที่ทำให้เกิดเสียงพูดขึ้นในภาษา
12. เส้นเสียงปลอม เป็นอวัยวะที่มีลักษณะเหมือนเส้นเสียงแต่อยู่เหนือเส้นเสียงขึ้นไป เส้นเสียงปลอมนี้เข้าใจกันว่าจะดึงเข้าหากันเมื่อเวลาพูดเสียงกระซิบ

2.5 เสียงในภาษาไทย

เสียงในภาษาไทยจำแนกออกได้เป็น 3 ชนิด คือเสียงแท้ เสียงแปรและเสียงดนตรี

1. เสียงแท้ คือเสียงที่เปล่งออกมาจากลำคอโดยตรง โดยไม่ถูกอวัยวะต่างๆในปากปิดหรือกั้นลมไว้ เป็นเสียงที่เปล่งออกมาได้โดยสะดวก เช่น เสียง อา อี เอ โอ ออ เป็นต้น สัญลักษณ์หรืออักษรที่ใช้แทนเสียงแท้เรียกว่า “สระ”
2. เสียงแปร คือเสียงที่เปล่งออกมาจากลำคอแล้วถูกอวัยวะต่างๆในปากปิดกักหรือกั้นลมไว้ ไม่ปล่อยให้ลมแล่นออกมาได้สะดวก ทำให้เสียงแปรไป เช่น เสียง /ป/ สัญลักษณ์หรือตัวอักษรที่ใช้แทนเสียงแปรเรียกว่า “พยัญชนะ”
3. เสียงดนตรีหรือเสียงผัน คือเสียงแท้หรือเสียงแปรนั่นเองที่เปล่งออกมาแล้วมีระดับเสียงสูงๆต่ำๆ คล้ายเสียงดนตรี สัญลักษณ์หรืออักษรที่ใช้แทนเสียงดนตรีเรียกว่า “วรรณยุกต์”

2.6 เสียงของอักษรไทย

เสียงกับตัวอักษรมีความแตกต่างกัน เพราะเสียงในภาษาก็คือเสียงพูดของมนุษย์ ส่วนตัวอักษรคือสัญลักษณ์ที่ใช้แทนเสียงพูดของมนุษย์ ไม่มีภาษาใดในโลกที่มีตัวอักษรถ่ายทอดเสียงพูดได้หมดทุกเสียง ตัวอักษรไทยได้ชื่อว่ามีลักษณะพิเศษ เพราะสามารถถ่ายทอดเสียงพูดในภาษาไทยได้ตรงเกือบหมดทุกเสียง นอกจากนี้ยังสามารถใช้ถ่ายทอดเสียงพูดในภาษาอื่นๆได้มากด้วย แม้ว่าภาษานั้นๆจะออกเสียงยากเพียงใดก็ตาม ดังที่เราถ่ายทอดเสียงภาษาอังกฤษเป็นอักษรไทย

2.6.1 เสียงพยัญชนะ

พยัญชนะไทยมี 44 รูป แต่มี 21 เสียงดังนี้

1. หน่วยเสียง ก
2. หน่วยเสียง ข ฃ ค ฅ ฆ
3. หน่วยเสียง ง
4. หน่วยเสียง จ
5. หน่วยเสียง ฉ ช ฌ
6. หน่วยเสียง ซ ส ศ ษ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

7. หน่วยเสียง ญ ย
8. หน่วยเสียง ด ฎ
9. หน่วยเสียง ต ฏ
10. หน่วยเสียง ถ ท ธ ฐ ฑ ฒ
11. หน่วยเสียง น ณ
12. หน่วยเสียง บ
13. หน่วยเสียง ป
14. หน่วยเสียง ผ พ ภ
15. หน่วยเสียง ฝ ฟ
16. หน่วยเสียง ม
17. หน่วยเสียง ร
18. หน่วยเสียง ล ฬ
19. หน่วยเสียง ว
20. หน่วยเสียง ห ฮ
21. หน่วยเสียง อ

ซึ่งในจำนวนทั้งหมดนี้จะมีเสียงพยัญชนะบางตัวที่สามารถควบกับเสียงอื่นได้รวมอยู่ด้วย ได้แก่ เสียง /ร/ /ล/ และ /ว/ ทำให้เกิดเสียงพยัญชนะควบกล้ำเพิ่มขึ้นอีกดังนี้

กร	เช่น	กรง	กรวด	เกรียบ
กล	เช่น	กลัว	เกลื่อน	กลับ
กว	เช่น	กว้าง	กวาง	กวาด
คร	เช่น	ครัว	ครอง	ครีม
คล	เช่น	คลอน	คลอก	คล้อง
คว	เช่น	ควาน	ควาย	ความ
ปร	เช่น	ปราบ	เปรียบ	ปรับ
ปล	เช่น	ปลด	ปลี	ปลอด
พร	เช่น	พร้อม	พราว	พราน
พล	เช่น	พลั่ว	พลอย	พลี
ตร	เช่น	ตรา	ตรม	ตรอง
ทร	เช่น	ทรา	ทรวง	ทรง
ดร*	เช่น	ดราฟต์	ดรัมเมเยอร์	
ฟร*	เช่น	ฟรี	ฟรักโทส	
ฟล*	เช่น	แฟลต	ฟลูออรีน	
บร*	เช่น	บริดจ์	บรอนซ์	
บล*	เช่น	บล็อก	บลูยีนส์	

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

คำที่มีเครื่องหมายดอกจัน(*) เป็นคำที่ยืมมาจากภาษาต่างประเทศ

2.6.2 เสียงสระ

สระในภาษาไทยมี 21 รูป แต่มี 32 เสียง ซึ่งแบ่งออกเป็นหมวดหมู่ได้ดังนี้

1. สระแท้ฐานเดียว มี 8 เสียงคือ อะ อา อิ อี อึ อื อุ อู
2. สระแท้สองฐาน มี 10 เสียงคือ เอะ เอ แอะ แอ โอะ โอ เออะ เออ เอาะ ออ
3. สระผสม มี 6 เสียงคือ เอียะ เอีย เอือะ เอือ อัวะ อัว
4. สระเกิน มี 8 เสียงคือ ฤ ฤา ฦ ฦา อำ ไอ ใอ เอา

เราจะสังเกตเห็นว่าเสียงสระบางเสียงต้องใช้รูปสระหลายๆรูปมาผสมกันให้เป็นเสียงหนึ่ง เช่น "เอะ" มีรูปไม้หน้า (เ -) กับ วิสรรชนีย์ (- ะ) ผสมกันเป็นต้น

หรือเราอาจจะแบ่งได้ในอีกลักษณะหนึ่ง คือเสียงสระเดี่ยว 18 หน่วยเสียง เสียงสระประสม 6 หน่วยเสียง และสระเกิน 8 เสียง ซึ่งประเภทหลังนี้จะมีเสียงพยัญชนะสะกดอยู่ด้วย จึงไม่นับเสียงสะกดเหล่านี้เป็นเสียงสระ

1. สระเดี่ยว(Monophthong) หมายถึงเสียงสระที่มีที่เกิดเพียงฐานเดียว คือเสียงที่เกิดจากการบังคับลิ้นให้อยู่ในระดับที่ต้องการเพียงจุดใดจุดหนึ่ง แล้วจึงเปล่งเสียงออกมา

แบ่งเป็น

1.1 สระเสียงสั้น มี 9 เสียงคือ อะ อิ อึ อุ เอะ แอะ โอะ เอาะ เออะ

1.2 สระเสียงยาว มี 9 เสียงคือ อา อี อือ อู เอ แอ โอ ออ เออ

2. สระประสม(Diphthong) หมายถึงเสียงสระที่มีที่เกิดตั้งแต่ 2 ฐานขึ้นไป กล่าวคือเสียงสระที่เกิดจากการใช้ลิ้นออกเสียงสระหนึ่งแล้วลิ้นยังไม่กลับเข้าที่เดิมและได้ออกเสียงอีกสระหนึ่ง เป็นลักษณะการเปลี่ยนระดับของลิ้นจากระดับหนึ่งไปอีกระดับหนึ่ง โดยลงเสียงหนักที่สระตัวแรก ซึ่งมี 6 เสียงคือ

เอียะ	เกิดจากการออกเสียงสระ	อิ+อะ
เอีย	เกิดจากการออกเสียงสระ	อี+อา
เอือะ	เกิดจากการออกเสียงสระ	อึ+อะ
เอือ	เกิดจากการออกเสียงสระ	อือ+อา
อัวะ	เกิดจากการออกเสียงสระ	อุ+อะ
อัว	เกิดจากการออกเสียงสระ	อู+อา

3. สระเกิน หมายถึงเสียงสระที่มีเสียงซ้ำกับเสียงสระเดี่ยวและมีเสียงพยัญชนะประสมอยู่ ซึ่งมี 8 เสียงคือ ฤ ฤา ฦ ฦา อำ ไอ ใอ เอา

เพื่อประโยชน์ในการศึกษารูปสระที่ใช้กับเสียงสระ ควรที่จะได้ทราบรูปของสระต่างๆทั้ง 21 รูปไว้ดังนี้

1. ะ วิสรรชนีย์หรือนมนางคู่
2. ั หันอากาศ ไม้ผัดหรือหางกังหัน
3. ็ ไม้ไต่คู้
4. า ลากข้าง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5. ิ พินทุหรือพินทุอิ
6. ์ ฝนทอง ฟันหนูหรือมูสิกทันต์
7. ์ หยาดน้ำค้าง นิคหิตหรือนฤคหิต
8. ุ ตีนเหยียด
9. ู ตีนคู้หรือลากตีน
10. เ ไม้หน้า
11. ใ ไม้ม้วน
12. ไ ไม้มลาย
13. โ ไม้โอ
14. ฤ ตัวรึ
15. ฤา ตัวรือ
16. ฦ ตัวลึ
17. ฦา ตัวลือ
18. ย ตัวยอ
19. รร ร หัน
20. ว ตัวววอ
21. อ ตัวออ

2.6.3 เสียงวรรณยุกต์

วรรณยุกต์มี 4 รูป 5 เสียง ซึ่งเราอาจแบ่งออกได้เป็น 2 ชนิดคือ

1. วรรณยุกต์มีรูป หมายถึงวรรณยุกต์ที่มีเครื่องหมายบอกระดับของเสียงให้เห็นชัดเจนอยู่เบื้องบนของตัวอักษร มี 4 รูปคือ ่ ้ ๊ ๋ เรียกชื่อว่าวรรณยุกต์ เอก โท ตรีและจัตวา ตามลำดับ และให้เขียนไว้เบื้องบนพยัญชนะต้น เช่น ก่ ก้ ก๊ และ ก๋ ถ้าเป็นอักษรควบหรืออักษรนำให้เขียนไว้เบื้องบนอักษรตัวที่ 2 เช่น ครุ่น เกล้า เป็นต้น
2. วรรณยุกต์ไม่มีรูป ได้แก่เสียงที่มีทำนองสูงต่ำตามหมู่ของอักษร โดยไม่ต้องมีรูปวรรณยุกต์กำกับก็อ่านออกเสียงได้ เหมือนมีรูปวรรณยุกต์กำกับอยู่ด้วย เช่น คุณ จะ พูด นัก เสียง จะมีเสียงวรรณยุกต์ สามัญ เอก โท ตรีและจัตวา ตามลำดับ

เมื่อใช้รูปวรรณยุกต์เอกกับอักษรสูงและอักษรกลางจะเป็นเสียงเอก เช่น ข่า กี่

เมื่อใช้รูปวรรณยุกต์เอกกับอักษรต่ำจะเป็นเสียงโท เช่น ล่า นี่

เมื่อใช้รูปวรรณยุกต์โทกับอักษรสูงและอักษรกลางจะเป็นเสียงโท เช่น ข้า โป้ง

เมื่อใช้รูปวรรณยุกต์โทกับอักษรต่ำจะเป็นเสียงตรี เช่น ค้า ม้า

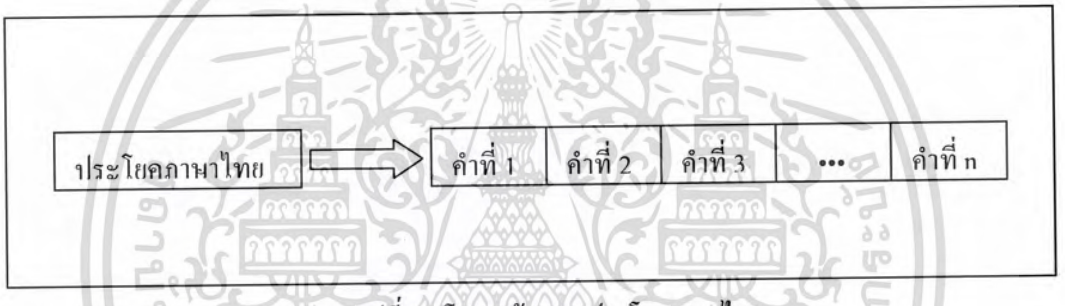
รูปวรรณยุกต์ตรีและจัตวาจะไม่ใช้กับอักษรต่ำและอักษรสูง จะใช้กับอักษรกลางเท่านั้น ซึ่งก็จะได้เสียงตรีและจัตวาเช่นเดียวกัน เช่น ก๊าด ป๋า ตุ๊กแก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3
การตัดคำ

3.1 บทนำ

ประโยคในภาษาไทยเกิดขึ้นจากการนำคำมาเรียงต่อกัน คำต่าง ๆ ภายในประโยคจะไม่มีการแบ่งแยกด้วยเครื่องหมายหรือช่องว่างใด ๆ ดังนั้นผู้อ่านจะต้องทำการแบ่งประโยคออกเป็นคำ ๆ ด้วยตนเอง โดยอาศัยความหมายของคำที่อยู่ด้านหน้าและคำที่จะตามมาด้านหลัง เมื่อต้องการให้คอมพิวเตอร์สามารถอ่านประโยคภาษาไทยได้จึงจำเป็นต้องอาศัยอัลกอริทึมที่ช่วยในการแบ่งคำหรือตัดคำภายในประโยคออกมาแล้วได้ความหมายของประโยคเหมือนความตั้งใจของผู้เขียนประโยคนั้น อัลกอริทึมที่ใช้ในการตัดคำในปริญญานิพนธ์ฉบับนี้คือ การตัดคำโดยใช้วิธี Word Bucket



รูปที่ 3.1 โครงสร้างของประโยคภาษาไทย

3.2 พจนานุกรมคำศัพท์

อัลกอริทึม Word Bucket ต้องอาศัยพจนานุกรมช่วยในการตัดคำ พจนานุกรมที่จะใช้ในปริญญานิพนธ์ฉบับนี้ มีโครงสร้าง ดังนี้คือ

1. คำศัพท์
2. คำอ่าน เป็นส่วนที่แสดงคำอ่านของคำศัพท์นั้น ๆ (ส่วนนี้เป็นส่วนที่เพิ่มเข้ามา ไม่มีความจำเป็นสำหรับการตัดคำ) คำอ่านจะกำหนดอยู่ในระดับโฟนีม (phoneme) โดยใช้สัญลักษณ์แทนหน่วยเสียงต่าง ๆ แต่ละสัญลักษณ์ของหน่วยเสียงจะคั่นด้วยช่องว่างหนึ่งช่อง

ดังแสดงในตารางที่ 3-1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

คำศัพท์	คำอ่าน
กรกฎาคม	g a1 r a3 g a1 d ah kh o m_
กรณี	g a1 r a3 n ih
กรรมการ	g a m_ m a3 g ah n_
ข้น	kh o2 n_2
ขนม	kh a1 n o4 m_4
ขนาด	kh a1 n ah1 d_1

ตารางที่ 3-1 โครงสร้างของพจนานุกรมคำศัพท์

3.3 วิธีการตัดคำโดยใช้อัลกอริทึม Word Bucket

อัลกอริทึม Word Bucket เป็นการตัดคำโดยวิธีการเก็บ "คำที่เป็นไปได้" ในการตัดคำแต่ละครั้งไว้ใน Word Bucket แล้วเลือกคำที่ยาวที่สุดใน Word Bucket เป็นคำที่ตัดได้ "คำที่เป็นไปได้" จะเป็นคำที่มีอยู่ในพจนานุกรมคำศัพท์เท่านั้น เช่น

Word Bucket [1] อา

Word Bucket [2] อาจ

Word Bucket [3] อาจารย์

คำที่เป็นไปได้ คือ อา, อาจ, อาจารย์

คำที่เป็นไปได้ที่ยาวที่สุด คือ อาจารย์

หมายเหตุ : ต้องมีคำทั้งสามนี้ใน พจนานุกรมคำศัพท์

ตัวอย่างการตัดคำจากประโยค

ประโยคที่จะตัดคำ คือ "สวัสดีครับผมชื่อโปรแกรมสังเคราะห์เสียงภาษาไทยครับ"

หลังจากการตัดคำ : สวัสดี ครับ ผม ชื่อ โปรแกรม สังเคราะห์ เสียง ภาษา ไทย ครับ

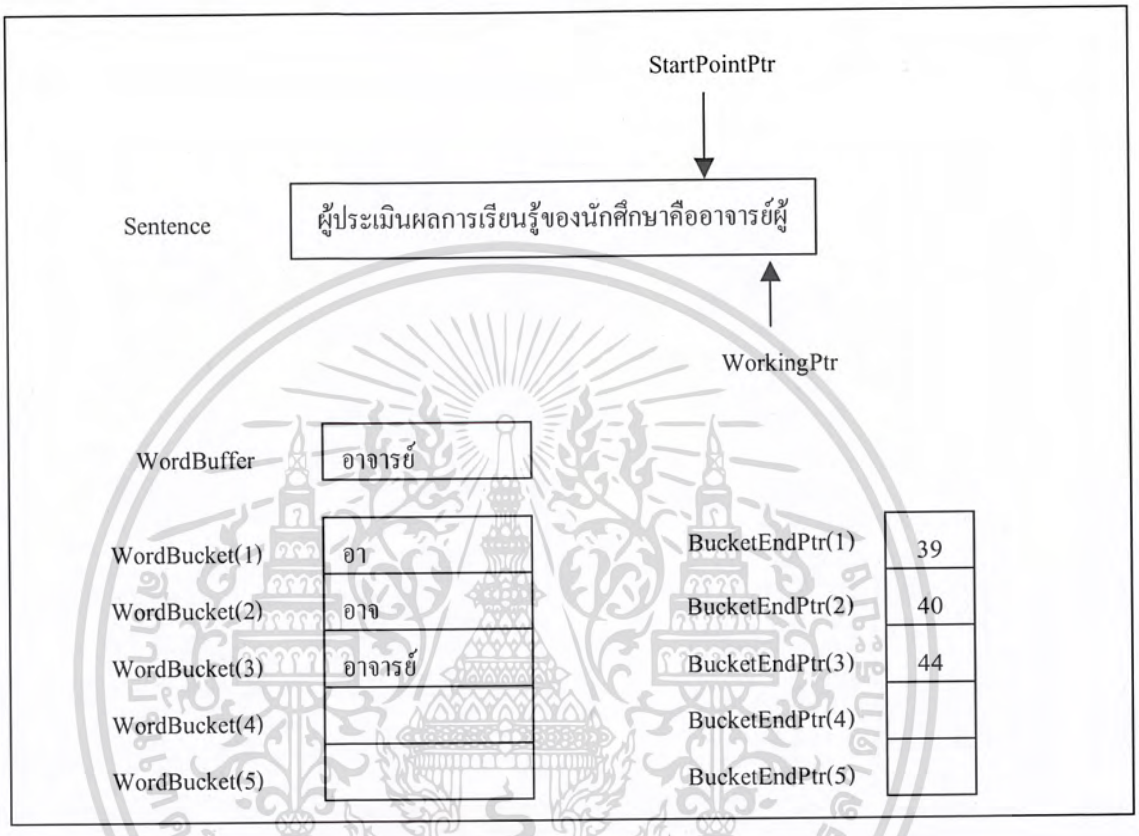
3.3.1 ตัวแปรที่ใช้ในการตัดคำ

ตัวแปรที่ใช้อัลกอริทึม Word Bucket มีดังนี้คือ

- ตัวแปรที่ใช้ชี้ข้อมูล
  1. StartPointPtr เป็นตัวแปรที่ใช้ชี้ตำแหน่งตัวอักษรเริ่มต้นของคำที่กำลังตัด
  2. WorkingPtr เป็นตัวแปรที่ใช้ชี้ไปยังตัวอักษรตัวต่อไปที่จะอ่านเข้ามายัง WordBuffer
- ตัวแปรที่ใช้ในการเก็บข้อมูล
  1. WordBuffer เป็นตัวแปรที่ใช้เก็บกลุ่มของตัวอักษรที่จะนำไปเปรียบเทียบกับพจนานุกรมคำศัพท์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2. WordBucket เป็นตัวแปรที่เก็บคำที่เป็นไปได้ในการตัดคำทั้งหมด เพื่อทำการเลือกคำที่ยาวที่สุดต่อไป
3. BucketEndPtr เป็นตัวเปรที่ใช้เก็บจำนวนตำแหน่งของตัวอักษรตัวสุดท้ายของคำต่าง ๆ ที่อยู่ใน WordBucket เพื่อใช้สำหรับเลื่อน StartPointPtr ไปยังตำแหน่งเริ่มต้นของคำถัดไป



รูปที่ 3.2 แสดงวิธีการตัดคำโดยใช้ Word Bucket

3.3.2 ขั้นตอนการทำงานของอัลกอริทึม

1. กำหนดค่าเริ่มต้นให้กับ StartPointPrt และ WorkingPtr ให้เท่ากับ 1 เพื่อชี้ไปยังตัวอักษรตัวแรกของข้อความที่ต้องการตัดคำ
2. นำตัวอักษรที่ชี้โดย WorkingPtr มาเก็บไว้ใน WordBuffer ซึ่งเป็นบัฟเฟอร์ที่ใช้เปรียบเทียบกับพจนานุกรมคำศัพท์และเพิ่มค่า WorkingPtr อีกหนึ่ง เพื่อชี้ไปยังตัวอักษรตัวต่อไป
3. เปรียบเทียบกลุ่มของตัวอักษรที่อยู่ใน WordBuffer กับพจนานุกรมคำศัพท์ว่ามีคำที่อยู่ในพจนานุกรมคำศัพท์ที่ขึ้นต้นด้วยกลุ่มของตัวอักษรที่อยู่ใน WordBuffer หรือไม่ ถ้ามีแสดงว่ายังคงมีคำที่เป็นไปได้ในการตัดที่ยาวกว่านี้อยู่ให้ไปขั้นตอนถัดไป ถ้าไม่มีแสดงว่าไม่มีคำที่เป็นไปได้ที่ยาวกว่านี้แล้วให้ไปขั้นตอนที่ 7

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4. เปรียบเทียบกลุ่มของตัวอักษรที่อยู่ใน WordBuffer กับพจนานุกรมคำศัพท์ ว่ามีคำที่อยู่ในพจนานุกรมคำศัพท์ที่ตรงกับคำที่อยู่ใน WordBuffer นี้หรือไม่ ถ้ามีแสดงว่าพบคำที่สามารถตัดได้ในพจนานุกรมคำศัพท์ให้ไปขั้นตอนต่อไป ถ้าไม่มีให้กลับไปขั้นตอนที่ 2 เพื่อหาคำที่สามารถตัดได้ต่อไป
5. นำคำที่ได้จากขั้นตอนที่ 4 เก็บลง WordBucket ซึ่งเป็นบัฟเฟอร์ขนาด 5 ช่อง สำหรับเก็บคำทั้งหมดที่สามารถตัดได้
6. ตรวจสอบว่า WordBucket เต็มหรือไม่ ถ้าเต็มไปขั้นตอนต่อไปถ้าไม่เต็มกลับไปขึ้นตอนที่ 2 เพื่อหาคำที่สามารถตัดได้ที่ยาวที่สุดต่อไป
7. เลือกคำใน WordBucket ที่ยาวที่สุด โดยปกติจะเป็นคำสุดท้ายที่ตัดได้แต่บางครั้งคำที่ยาวที่สุดที่ตัดได้อาจทำให้คำต่อไปอ่านไม่ได้ ดังนั้นเราจึงต้องกลับมาเลือกคำก่อนหน้าที่ยาวที่สุดที่ไม่ทำให้คำต่อไปอ่านไม่ได้

   ตัวอย่างเช่น คำว่า "หายาก"

   WordBucket (1) = หา BucketEndPtr (1) = 2

   WordBucket (2) = หาย BucketEndPtr (2) = 3

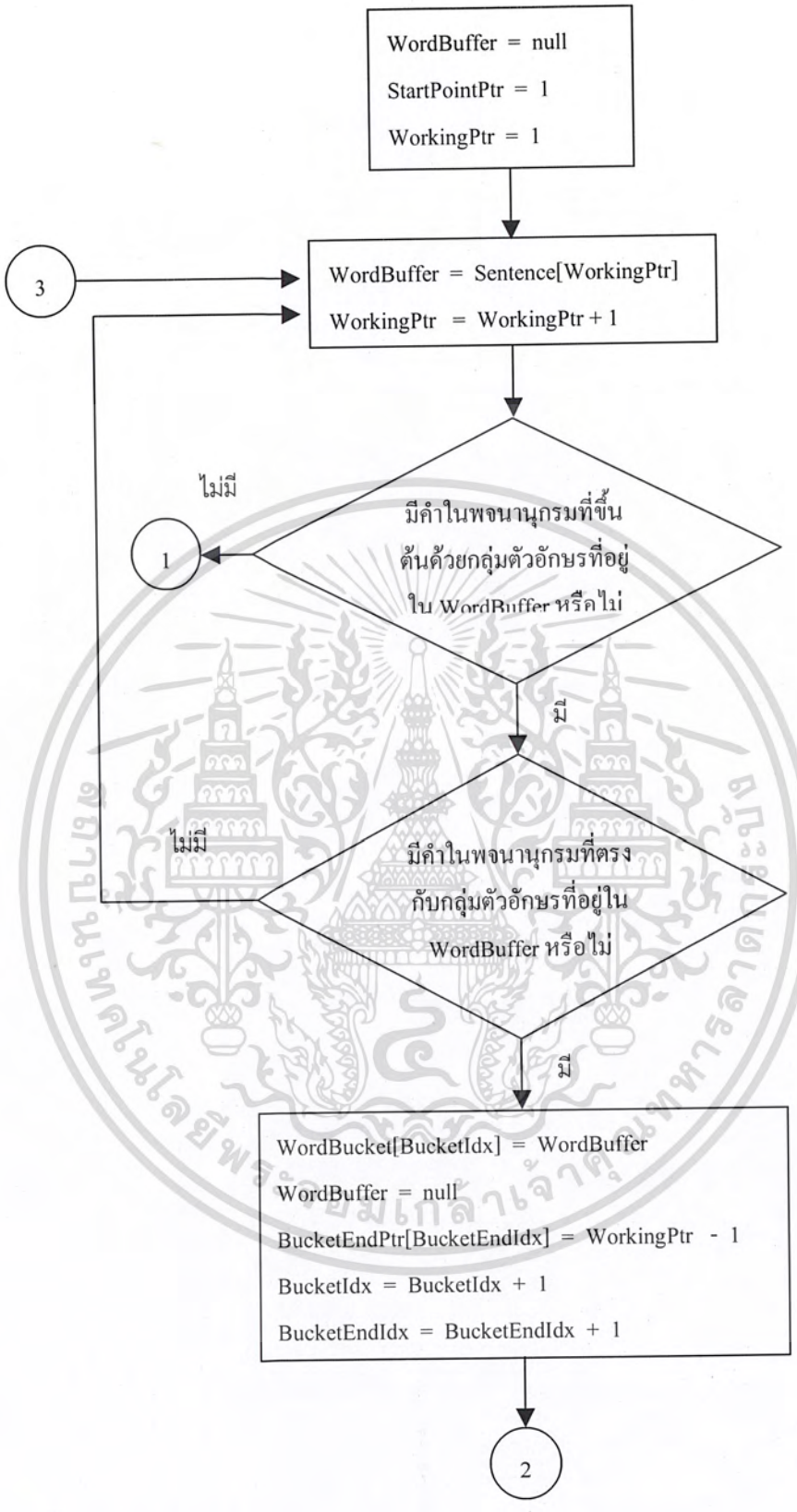
   ถ้าเราเลือกคำที่ยาวที่สุด คือ คำว่า "หาย" จะทำให้คำต่อไปเป็น "าก" ซึ่งเป็นคำที่ไม่สามารถอ่านได้ เราจึงต้องถอยกลับมาเลือกคำว่า "หา" แทน ซึ่งวิธีตรวจสอบของเราจะใช้การตรวจสอบว่าอักษรตัวแรกของคำต่อไปขึ้นต้นด้วยสระหรือไม่ ถ้าขึ้นต้นด้วยสระให้ถอยกลับไปเลือกคำที่ยาวที่สุดคำก่อนหน้า โดยเขียนโปรแกรมในลักษณะของ Recursive
8. ตรวจสอบว่าคำที่เลือกเป็นคำที่เข้ามาใน WordBucket เป็นคำสุดท้ายหรือไม่ ถ้าใช่ก็ไม่ต้องทำการเปลี่ยนแปลงค่าใน WorkingPtr เพราะจะชี้ไปยังตัวอักษรตัวแรกของคำต่อไปอยู่แล้วให้ไปขั้นตอนที่ 10 ได้เลย แต่ถ้าไม่ใช่ให้ไปขั้นตอนต่อไป
9. ถ้าคำที่เลือกไม่ได้เป็นคำที่เข้ามาใน WordBucket เป็นคำสุดท้าย แสดงว่าหลังจากเลือกคำนี้ WorkingPtr ไม่ได้ชี้ไปยังตัวอักษรตัวแรกของคำต่อไป ต้องทำการย้าย WorkingPtr ไปยังตำแหน่งที่ถูกต้อง ดังนี้ WorkingPtr = BucketEndPtr + 1

   โดย BucketEndPtr เป็นตัวแปรที่เก็บค่าตำแหน่งตัวอักษรสุดท้ายของคำคำนั้น
10. ให้ StartPointPtr = WorkingPtr เพื่อให้ StartPointPtr ชี้ไปยังตัวอักษรตัวแรกของคำถัดไป แล้วกลับไปทำขั้นที่ 2 เพื่อตัดคำถัดไป

สรุป

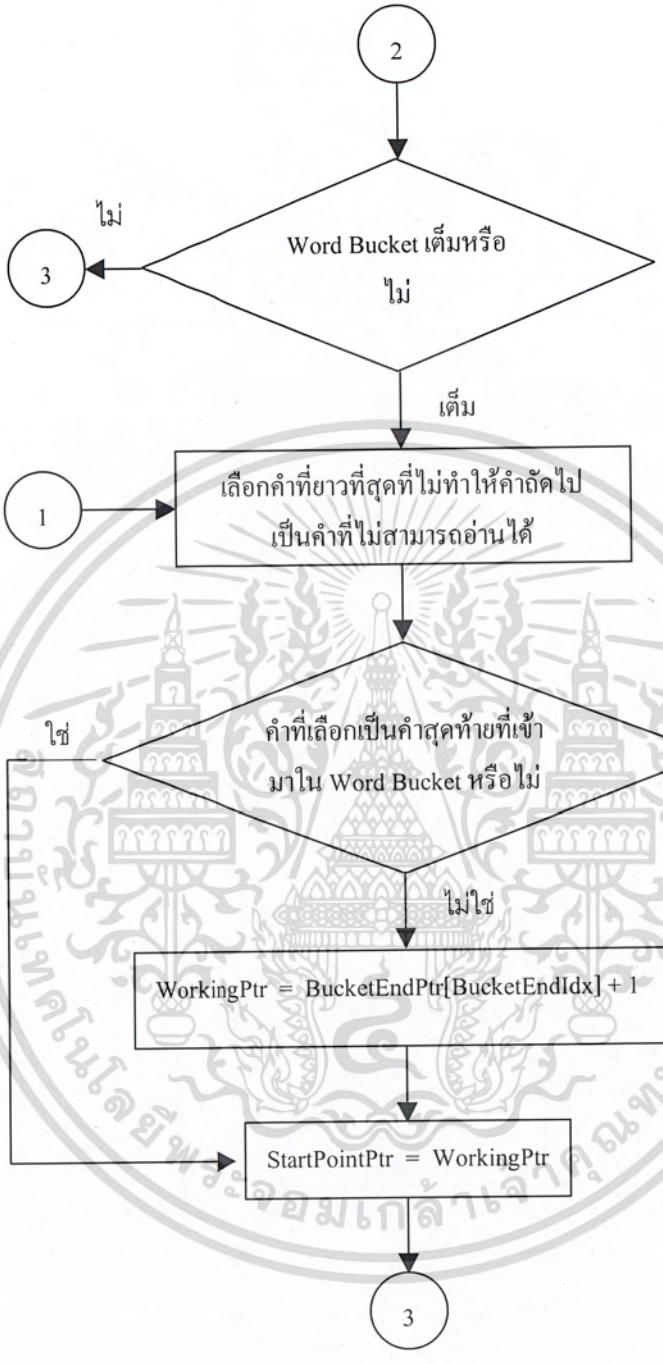
อัลกอริทึม Word Bucket เป็นการตัดคำที่เลือกคำที่ยาวที่สุดใน Word Bucket เป็นคำที่ตัดได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.3 แผนภูมิสายงานแสดงขั้นตอนการทำงานของอัลกอริทึม Word Bucket

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 3.3 แผนภูมิสายงานแสดงขั้นตอนการทำงานของอัลกอริทึม Word Bucket (ต่อ)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4
การกำหนดหน่วยเสียงและการสังเคราะห์เสียง

4.1 บทนำ

ในบทนี้จะกล่าวถึงรูปแบบการเก็บบันทึกหน่วยเสียงเพื่อนำไปใช้ในการสร้างเสียงพูด เทคนิคที่นำมาใช้ในการสร้างเสียงพูด และในส่วนสุดท้ายของบทจะกล่าวถึงการรวมหน่วยเสียงในระบบเสียงภาษาไทย

4.2 หน่วยเสียงพื้นฐาน(Basic Phoneme)

หน่วยเสียง(phoneme) เป็นลักษณะเฉพาะแทนเสียงแต่ละเสียงที่ไม่ซ้ำกัน หน่วยเสียงพื้นฐานในภาษาไทยได้แก่ หน่วยเสียงพยัญชนะ หน่วยเสียงสระและหน่วยเสียงวรรณยุกต์ การเก็บบันทึกสามารถทำได้หลายวิธี แต่ประเด็นสำคัญจะต้องนำหน่วยเสียงเหล่านี้ไปใช้ให้เกิดความยืดหยุ่น และเมื่อรวมกันแล้วต้องสามารถให้เสียงที่ถูกต้องเหมือนเสียงพูดตามธรรมชาติมากที่สุด

รูปแบบการเก็บบันทึกหน่วยเสียงมีดังนี้

แบบที่ 1

1. บันทึกเสียงพยัญชนะต้นเดี่ยว 21 หน่วยเสียง และพยัญชนะควบกล้ำ 17 หน่วยเสียง รวมเป็น 38 หน่วยเสียง
2. บันทึกเสียงสระ 24 หน่วยเสียง ผันเสียงวรรณยุกต์ 5 เสียง ได้ 120 หน่วยเสียง
3. บันทึกเสียงพยัญชนะสะกดแม่ /กง/ /กน/ /กม/ /เกย/ /เกอว/ 5 หน่วยเสียง ผันเสียงวรรณยุกต์ 5 หน่วยเสียงได้ 25 หน่วยเสียง และพยัญชนะสะกดแม่ /กก/ /กด/ /กบ/ 3 หน่วยเสียง ผันเสียงวรรณยุกต์ 3 หน่วยเสียงได้ 9 หน่วยเสียง รวมเป็น 34 หน่วยเสียง

หน่วยเสียงรวมทั้งสิ้น 38+120+34 = 192 หน่วยเสียง

ตัวอย่างการผสมเสียง

บ้าน จะได้ บ+อ้า+น้

รัก จะได้ ร+อ๊ะ+ก๊

เหนียว จะได้ น+เอี๋ย+ว๋

สวย จะได้ ส+อั๊ว+ว๋

ปลอด จะได้ ปล+อ่อ+ด่

เป็นต้น

การบันทึกเสียงในแบบที่ 1 นี้จะประหยัดเนื้อที่ในหน่วยความจำมากที่สุด และเสียงที่ผสมได้ก็สามารถฟังได้เข้าใจดีพอสมควร ซึ่งในการทำงานครั้งนี้เราจะทำการบันทึกในลักษณะนี้ ดังมีรายละเอียดดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สำนักหอสมุดกลาง พระจอมเกล้าเจ้าคุณทหารลาดกระบัง

1. เสียงของพยัญชนะต้น มีทั้งหมด 38 เสียง ดังตารางที่ 4-1

เสียงพยัญชนะต้น			
รูปพยัญชนะ	สัญลักษณ์แทนเสียง	รูปพยัญชนะ	สัญลักษณ์แทนเสียง
ก	g	ห ฮ	h
ข ค ฆ	kh	อ	oo
ง หง-(หงาย)	ng	กร	gr
จ จร-(จริง)	j	กล	gl
ฉ ช ฌ	ch	กว	gw
ซ ส ศ ษ ทร-(ทราย)	s	คร ขร	kr
ญ ย หย-(หย่า) อย-(อย่า)	y	คล ขล	kl
ฎ ด ฑ-(มณฑป)	d	ควฃ ขว	kw
ฏ ต	dt	พร ผร	phr
ฐ ฑ ฒ ถ ท ธ	th	พล ผล	phl
ณ น หน-(หนู)	n	ปร	pr
บ	b	ปล	pl
ป	p	ดร	dr
ผ พ ภ	ph	ทร	thr
ฝ ฟ	f	ตร	tr
ม หม-(หมี)	m	ฟร	fr
ร หร-(เหรียญ)	R	ฟล	fl
ล ฬ หล-(หลาน)	L	บร	br
ว หว-(หวี)	W	บล	bl

ตารางที่ 4-1 เสียงของพยัญชนะต้นและสัญลักษณ์แทนเสียง

2. เสียงของสระที่ผันตามเสียงวรรณยุกต์ มีทั้งหมด 21x5 = 105 เสียง ดังตารางที่ 4-2 ซึ่งเราจะตัดเสียง เอียะ เอือะ และ อัวะ ออก เพราะในภาษาไทยไม่ค่อยได้มีการนำมาใช้เท่าใดนัก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

42801

เสียงสระที่ผันตามเสียงวรรณยุกต์							
เสียงสระ	เสียงวรรณยุกต์	สัญลักษณ์	ตัวอย่าง	เสียงสระ	เสียงวรรณยุกต์	สัญลักษณ์	ตัวอย่าง
อะ	สามัญ	a	บัน	อา	สามัญ	ah	ภา
	เอก	a1	กระ		เอก	ah1	อ่าน
	โท	a2	ลั่น		โท	ah2	ลาด
	ตรี	a3	พระ		ตรี	ah3	ช้าง
	จัตวา	a4	หวัง		จัตวา	ah4	ถา
อิ	สามัญ	I	ริม	อี	สามัญ	ih	ยี
	เอก	i1	จิก		เอก	ih1	ปี่
	โท	i2	สิ้น		โท	ih2	ที่
	ตรี	i3	ริ้ว		ตรี	ih3	ชี้
	จัตวา	i4	สิน		จัตวา	ih4	ผี
อึ	สามัญ	v	จึง	อือ	สามัญ	vh	ดือ
	เอก	v1	หนึ่ง		เอก	vh1	ดื่น
	โท	v2	ซึ่ง		โท	vh2	ชื่อ
	ตรี	v3	นึก		ตรี	vh3	ชื้อ
	จัตวา	v4	ถึง		จัตวา	vh4	ถือ
อุ	สามัญ	u	คุณ	อู	สามัญ	uh	ง
	เอก	u1	สุด		เอก	uh1	อยู่
	โท	u2	นุ่ม		โท	uh2	พู่
	ตรี	u3	ทุก		ตรี	uh3	รู้
	จัตวา	u4	ถุง		จัตวา	uh4	หรู
เอะ	สามัญ	e	เป็น	เอ	สามัญ	eh	เกง
	เอก	e1	เด็ก		เอก	eh1	เหล่
	โท	e2	เล่น		โท	eh2	เป้
	ตรี	e3	เล็ก		ตรี	eh3	เคว้ง
	จัตวา	e4	เห็น		จัตวา	eh4	เฉ
แอะ	สามัญ	x	แกรม	แอ	สามัญ	xh	แตน
	เอก	x1	แห่ง		เอก	xh1	แต่
	โท	x2	แท่ง		โท	xh2	แย่
	ตรี	x3	และ		ตรี	xh3	แล้ว
	จัตวา	x4	แข็ง		จัตวา	xh4	แสง

ตารางที่ 4-2 เสียงของสระที่ผันตามเสียงวรรณยุกต์และสัญลักษณ์แทนเสียง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เสียงสระที่ผันตามเสียงวรรณยุกต์							
เสียงสระ	เสียงวรรณยุกต์	สัญลักษณ์	ตัวอย่าง	เสียงสระ	เสียงวรรณยุกต์	สัญลักษณ์	ตัวอย่าง
โอะ	สามัญ	o	บน	โอ	สามัญ	oh	โน
	เอก	o1	สด		เอก	oh1	โอ่ง
	โท	o2	ก้ม		โท	oh2	โลก
	ตรี	o3	รถ		ตรี	oh3	โจ๊ก
	จัตวา	o4	ขน		จัตวา	oh4	โขง
เอาะ	สามัญ	or	คอม	ออ	สามัญ	orh	จอม
	เอก	or1	เกาะ		เอก	orh1	ออก
	โท	or2	ต้อง		โท	orh2	ชอบ
	ตรี	or3	เงาะ		ตรี	orh3	ร้อน
	จัตวา	or4	หอย		จัตวา	orh4	ของ
เออะ	สามัญ	er	เงิน	เออ	สามัญ	erh	เธอ
	เอก	er1	เปอะ		เอก	erh1	เหม่อ
	โท	er2	เปิ้ล		โท	erh2	เก้อ
	ตรี	er3	เทอะ		ตรี	erh3	เช้ิต
	จัตวา	er4	เหยิน		จัตวา	erh4	เสริม
อัว	สามัญ	uah	กลัว	เอือ	สามัญ	vah	เตือน
	เอก	uah1	จั่ว		เอก	vah1	เสื่อ
	โท	uah2	ชั่ว		โท	vah2	เสื้อ
	ตรี	uah3	ย้วย		ตรี	vah3	เยื้อง
	จัตวา	uah4	สวย		จัตวา	vah4	เหลือง
เอีย	สามัญ	iah	เดียว				
	เอก	iah1	เฉี่ยว				
	โท	iah2	เที่ยว				
	ตรี	iah3	เคี้ยว				
	จัตวา	iah4	เสียง				

ตารางที่ 4-2(ต่อ) เสียงของสระที่ผันตามเสียงวรรณยุกต์และสัญลักษณ์แทนเสียง

3. เสียงของพยัญชนะสะกดที่ผันวรรณยุกต์ ซึ่งจะแบ่งออกเป็น

- แม่ /กง/ /กน/ /กม/ /เกย/ /เกอว/ ซึ่งผันวรรณยุกต์ได้ครบทั้ง 5 เสียงคือ เสียงสามัญ เอก โท ตรีและจัตวา ก็จะได้ 5x5 = 25 เสียง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

- แม่ /กก/ /กด/ / กบ/ ซึ่งผันวรรณยุกต์ได้เพียง 3 เสียงคือ เสียงเอก โทและตรี ก็จะได้ 3x3 = 9 เสียง

รวมแล้วก็จะได้ 25+9 = 34 เสียง ดังตารางที่ 4-3

เสียงพยัญชนะสะกดที่ผันตามเสียงวรรณยุกต์							
พยัญชนะสะกด	เสียงวรรณยุกต์	สัญลักษณ์	ตัวอย่าง	พยัญชนะสะกด	เสียงวรรณยุกต์	สัญลักษณ์	ตัวอย่าง
แม่ /กง/	สามัญ	ng_	ยาง	แม่ /เกอว/	สามัญ	w_	กาว
	เอก	ng_1	อย่าง		เอก	w_1	ข่าว
	โท	ng_2	ย่าง		โท	w_2	ข้าว
	ตรี	ng_3	ล้าง		ตรี	w_3	ร้าว
	จัตวา	ng_4	ฉาง		จัตวา	w_4	ขาว
แม่ /กน/	สามัญ	n_	บัน	แม่ /กก/	เอก	g_1	กาก
	เอก	n_1	ป่าน		โท	g_2	มาก
	โท	n_2	บ้าน		ตรี	g_3	พัก
	ตรี	n_3	ล้าน	แม่ /กด/	เอก	d_1	บาด
	จัตวา	n_4	หลาน		โท	d_2	ลาด
แม่ /กม/	สามัญ	m_	จาม		ตรี	d_3	วัด
	เอก	m_1	ห่าม	แม่ /กบ/	เอก	b_1	ดาบ
	โท	m_2	ย่าม		โท	b_2	ราบ
	ตรี	m_3	ม้าม		ตรี	b_3	ซับ
	จัตวา	m_4	หลาม				
แม่ /เกย/	สามัญ	y_	กาย				
	เอก	y_1	จ่าย				
	โท	y_2	ร่าย				
	ตรี	y_3	ท้าย				
	จัตวา	y_4	หลาย				

ตารางที่ 4-3 เสียงของพยัญชนะสะกดที่ผันตามเสียงวรรณยุกต์และสัญลักษณ์แทนเสียง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จะเห็นได้ว่าฐานข้อมูลเสียงของเรานั้นมีอยู่น้อยมากคือมีเพียง 38+105+34 = 177 ไฟล์ ซึ่งจะทำให้ประหยัดพื้นที่ในการเก็บได้มากทีเดียว และยังช่วยเพิ่มความเร็วในการหาไฟล์เสียงที่ต้องการนำมารวมกันเป็นคำอีกด้วย เพราะไม่ต้องเสียเวลาในการค้นหาข้อมูลมากนักก็จะเจอไฟล์ที่ต้องการได้ ถึงแม้ว่าคุณภาพเสียงที่ได้ออกมาอาจจะไม่เป็นธรรมชาติเท่าใดนัก แต่ก็เรียกได้ว่าดีพอสมควร สามารถฟังออกเป็นประโยคเป็นคำได้เข้าใจ

แบบที่ 2

1. บันทึกเสียงพยัญชนะต้น(เดี่ยว+ควบกล้ำ) 38 หน่วยเสียง ผสมเสียงสระ อะ ได้ทั้งหมด 38 หน่วยเสียง
2. บันทึกสระเสียงสั้น 12 หน่วยเสียง ผันเสียงวรรณยุกต์ 3 หน่วยเสียงคือเสียงเอก โทและตรี ได้ 36 หน่วยเสียง
3. บันทึกสระเสียงสั้นที่มีตัวสะกดแม่ /กก/ /กด/ /กบ/ ผันเสียงวรรณยุกต์ 3 หน่วยเสียงคือเสียงเอก โทและตรี ได้ 12x3x3 = 108 หน่วยเสียง
4. บันทึกสระเสียงสั้นที่มีตัวสะกดแม่ /กง/ /กน/ /กม/ /เกย/ /เกอว/ ผันเสียงวรรณยุกต์ 5 หน่วยเสียงคือสามัญ เอก โท ตรีและจัตวา ได้ 12x5x5 = 300 หน่วยเสียง
5. บันทึกสระเสียงยาว 12 หน่วยเสียง ผันเสียงวรรณยุกต์ 5 หน่วยเสียงคือเสียงสามัญ เอก โท ตรีและจัตวา ได้ 60 หน่วยเสียง
6. บันทึกสระเสียงยาวที่มีตัวสะกดแม่ /กก/ /กด/ /กบ/ ผันเสียงวรรณยุกต์ 3 หน่วยเสียงคือเสียงเอก โทและตรี ได้ 12x3x3 = 108 หน่วยเสียง
7. บันทึกสระเสียงยาวที่มีตัวสะกดแม่ /กง/ /กน/ /กม/ /เกย/ /เกอว/ ผันเสียงวรรณยุกต์ 5 หน่วยเสียงคือสามัญ เอก โท ตรีและจัตวา ได้ 12x5x5 = 300 หน่วยเสียง

หน่วยเสียงรวมทั้งสิ้น 38+36+108+300+60+108+300 = 950 หน่วยเสียง

ตัวอย่างการผสมเสียง

บ้าน จะได้ บะ+อ้าน
รัก จะได้ ระ+อั๊ก
เหนียว จะได้ นะ+เอี๋ยว
ปลอด จะได้ ปละ+อ่อด
สวย จะได้ สะ+อ๋วย

เป็นต้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

แบบที่3

จะคล้ายกับแบบที่ 2 มาก แต่ต่างกันที่ข้อแรกเท่านั้นคือจะบันทึกเสียงพยัญชนะต้น(เดี่ยว+ควบกล้ำ) ผสมเสียงสระเสียงสั้นทุกเสียง ซึ่งมีอยู่ 12 หน่วยเสียงก็จะได้ 38x12 = 456 เสียง ส่วนที่เหลือ(ข้อ2-ข้อ7) จะเหมือนกันทุกประการ

เมื่อรวมกันทั้งหมดแล้วก็จะได้หน่วยเสียงรวมทั้งสิ้น 456+36+108+300+60+108+300 = 1368 หน่วยเสียง

ตัวอย่างการผสมเสียง

บ้าน จะได้ บะ+อ้าน

รัก จะได้ ระ+อั๊ก

เหนียว จะได้ เนียะ+เอี๋ยว

ปลอด จะได้ เปลาะ+อ่อด

สวย จะได้ สัวะ+อ๋วย

เป็นต้น

แบบที่ 4

1. บันทึกพยัญชนะต้น(เดี่ยว+ควบกล้ำ) 38 หน่วยเสียง ผสมสระ(สั้น+ยาว) 24 หน่วยเสียง ผันวรรณยุกต์ 5 เสียง ได้ทั้งหมด 4560 เสียง
2. บันทึกพยัญชนะต้น(เดี่ยว+ควบกล้ำ) 38 หน่วยเสียง ผสมสระ(สั้น+ยาว) 24 หน่วยเสียง ผันวรรณยุกต์ 5 เสียง และตัวสะกด 8 เสียง ได้ทั้งหมด 36480

หน่วยเสียงรวมทั้งสิ้น 4560+36480 = 41040 หน่วยเสียง

การบันทึกในแบบนี้จะเป็นลักษณะของคำอ่านในภาษาไทย และเป็นคำที่ครอบคลุมทุกเสียงในภาษาไทยทั้งหมด

ตัวอย่างการออกเสียง

ประชากร จะได้ ประ+ชา+กร

บุรุษ จะได้ บุ+หรุด

ของขวัญ จะได้ ของ+ขวัญ

เรียบร้อย จะได้ เรียบ+ร้อย

ไปรษณีย์ จะได้ ไปร+สะ+นี

เป็นต้น

แบบที่ 5

เป็นการบันทึกเสียงเป็นคำๆเลย ซึ่งวิธีนี้จะได้เสียงที่เป็นธรรมชาติมากที่สุด แต่ก็จะเปลืองหน่วยความจำมากที่สุดเช่นเดียวกัน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.3 เทคนิคที่ใช้ในการสร้างเสียงพูด มี 2 วิธีคือ

1. การสร้างโดยการเก็บบันทึกเสียงพูด(speech storage)

เป็นวิธีของการเก็บบันทึกเสียงจริงๆของมนุษย์เอาไว้อย่างตรงไปตรงมา โดยอาจจะแบ่งเก็บเป็นข้อมูลของหน่วยย่อยๆของคำหรือพยางค์ ซึ่งเมื่อมีการสร้างเสียงพูดขึ้นมา ก็จะมีการอ้างถึงข้อมูลของหน่วยเสียงย่อยๆนี้นำมาเรียงต่อกันเกิดเป็นคำหรือประโยคขึ้นมา

การสร้างเสียงโดยการบันทึกเสียงพูด จะใช้การเก็บหน่วยเสียงตามแบบที่ 4 และ 5

ข้อดี คือสามารถสร้างเสียงพูดที่มีคุณภาพเสียงสูงมากๆ ใกล้เคียงกับสำเนียงของมนุษย์ได้เป็นอย่างดี

ข้อเสีย คือใช้หน่วยเก็บข้อมูลมากตามคุณภาพของเสียงที่ได้และคำศัพท์ที่มีอยู่ในพจนานุกรมเสียง

2. การสร้างเสียงโดยการสังเคราะห์เสียงพูด(speech synthesis)

การสร้างเสียงจากการสังเคราะห์ จะสร้างเสียงโดยการนำค่าพารามิเตอร์ของแต่ละเสียงมาใช้ในการหาเสียงจริง ดังนั้นการเก็บข้อมูลเสียงจะเก็บเฉพาะค่าพารามิเตอร์ของแต่ละเสียงเท่านั้น

การสร้างเสียงโดยการสังเคราะห์เสียงพูด จะใช้การเก็บหน่วยเสียงตามแบบที่ 1 2 และ 3

ข้อดี คือใช้หน่วยความจำน้อยกว่าวิธีแรกมาก ทำให้สามารถมุ่งไปสู่การสร้างระบบสังเคราะห์เสียงที่ไม่จำกัดจำนวนคำศัพท์ และที่สำคัญที่สุดก็คือสามารถทำได้บนเครื่องคอมพิวเตอร์ขนาดเล็กและสิ้นเปลืองค่าใช้จ่ายน้อย

ข้อเสีย คือโปรแกรมในการสังเคราะห์เสียงพูดมีความยุ่งยากและเสียงที่ได้ออกมามีคุณภาพอยู่ในระดับปานกลาง

4.4 การรวมหน่วยเสียง

จากสัญลักษณ์แทนเสียงที่ได้จากการแปลงหน่วยคำภายในประโยค เมื่อได้กำหนดหน่วยเสียงตามสัญลักษณ์แทนเสียงแต่ละตัวแล้ว เราสามารถนำแต่ละหน่วยเสียงที่เก็บบันทึกในรูปแฟ้มเสียง(.WAV) มาทำการเชื่อมต่อกันเป็นเสียงของแต่ละพยางค์ในหน่วยคำ ซึ่งในที่สุดเราก็จะได้แฟ้มเสียงรวมทั้งหมดของแฟ้มข้อมูล

ขั้นตอนการรวมหน่วยเสียง

1. ทำการวิเคราะห์สัญลักษณ์แทนเสียงแต่ละพยางค์ แล้วจัดรูปแบบในการรวมแฟ้มเสียงตามวิธีบันทึก/รวมแฟ้มในแบบที่ 1 ซึ่งจะอยู่ในรูปของเสียง

[พยัญชนะต้น] + [สระ+วรรณยุกต์] + [พยัญชนะสะกด+วรรณยุกต์]

2. เตรียมแฟ้มที่ใช้บันทึกผลลัพธ์ของการต่อแฟ้มเสียง

3. ทำการต่อแฟ้มเสียง โดยเรียกแฟ้มเสียงของหน่วยเสียงตามสัญลักษณ์แทนเสียงเข้ามา และนำแฟ้มเสียงของตัวถัดไปมาต่อท้ายเรียงกันตามลำดับจนหมดแฟ้มเสียง

4. เมื่อทำการต่อแฟ้มเสียงจนครบแล้วก็จะสามารถเล่นเสียงตามวิธีการเล่นเสียงของ Audio wave ทั่วๆไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5
การใช้งานโปรแกรม HTK

5.1 บทนำ

ในการทำงานครั้งนี้เราจะใช้โปรแกรม Cool Edit Pro ในการบันทึกเสียง โดยเสียงที่บันทึกนั้นจะบันทึกในลักษณะเป็นประโยค คำหรือพยางค์ซึ่งก็จะต้องครอบคลุมทุกเสียงที่จะนำมาใช้เป็นฐานข้อมูลเสียง และใช้โปรแกรม HTK(Hidden markov model ToolKit) ช่วยในการแบ่งไฟล์เสียงที่เราได้ทำการบันทึกเอาไว้ออกเป็นโฟนีม

โดยแท้จริงแล้วโปรแกรม HTK นั้นเป็นโปรแกรมที่ช่วยในการทำการรู้จำเสียง แต่เราก็สามารถนำส่วนหนึ่งของโปรแกรมมาช่วยในการแบ่งไฟล์เสียงของเราออกเป็นโฟนีมย่อยๆได้ ดังจะกล่าวต่อไป

5.2 ขั้นตอนการใช้งานโปรแกรม HTK

ขั้นตอนที่ 1 –การทำ Dictionary

ขั้นตอนของการสร้าง Dictionary ก็คือการสร้างรายการของคำที่ต้องการที่เรียงแล้ว(Sorted list) ซึ่งสามารถใช้ Editor ทั่วไปได้ โดยคำที่มีอยู่ใน Dictionary นั้นจะต้องเป็นคำศัพท์ที่ครอบคลุมคำทุกคำที่จะนำมาใช้ในการรู้จำ โดย Dictionary นั้นจะเป็นการสะกดคำเพื่ออ่านออกเสียง หรือก็คือในคำแต่ละคำที่จะทำการรู้จำนั้นจะประกอบด้วยเสียง Phoneme อะไรบ้าง โดยรูปแบบของ Dictionary คือ

WORD [outsym] p1 p2 p3 . . .

หมายความว่าคำว่า WORD นั้นจะมีการออกเสียงในลำดับของเสียง p1 p2 p3 . . .สำหรับตัวอักษรที่อยู่ใน Square Brackets จะหมายถึง Output ของคำนั้นเมื่อคำนั้นได้ถูกทำการรู้จำเรียบร้อยแล้ว ถ้าหากไม่มีการใส่ไว้เลยจะหมายถึงว่าคำๆนั้นจะเป็น Output (ในที่นี้ถ้าไม่มี Square Brackets คำที่เป็น Output ก็คือคำว่า WORD) แต่ถ้ามี Square Brackets แล้วแต่ข้างในไม่มีอะไรเลยจะหมายถึงไม่มีอะไรเป็น Output

ต่อไปนี้เป็นตัวอย่างของ Dictionary ที่ใช้ในการทำงาน

[] sil
silence [] sil
sil [] sil
SENT-START []
SENT-END []
หนึ่ง n vh1 ng sp
สอง s orh4 ng sp
สาม s ah4 m sp
สี่ s ih1 sp
ห้า h ah2 sp
หก h o1 g sp

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เจ็ด j e1 d sp

แปด p xh1 d sp

เก้า g a2 w sp

ศูนย์ s uh4 n sp (enter)

จะเห็นได้ว่า SENT-START และ SENT-END จะมี Silence Model sil เป็นการสะกดการออกเสียงและจะมี Output เป็น null หรือไม่มีอะไรเลย

ขั้นตอนที่ 2 -การบันทึกข้อมูลเสียง

ในการทำการรู้จำและทำการทดสอบผลของการรู้จำนั้นจำเป็นต้องใช้ข้อมูลเสียงทั้งสิ้น โดยการบันทึกข้อมูลเสียงสามารถทำได้จากโปรแกรมที่มีอยู่ทั่วไปที่สามารถทำให้ได้ไฟล์ที่มี Format เป็นนามสกุล .wav โดยในที่นี้เราจะใช้โปรแกรม Cool Edit Pro ในการบันทึกเสียง โดยในการบันทึกเสียงจะต้องเป็นไฟล์ที่เป็นแบบ 16 บิท mono และมี Sampling Rate เท่ากับ 44 KHz

ขั้นตอนที่ 3 -การสร้าง Transcription Files

ในการทำการรู้จำนั้น การทำการ train เซ็ทของ HMMs ทุกๆไฟล์เสียงที่นำมาใช้เป็นข้อมูลในการทำการรู้จำนั้นจะต้องมีรายการหรือ Transcription ในรูปแบบของ Phone Level หรือการออกเสียงที่เกี่ยวข้องกับเสียงที่ทำการรู้จำ ในการใช้งาน HTK ในที่นี้จะใช้วิธีการแบบ Flat-Start ซึ่งจะทำให้การทำ Transcription Files นั้นมีความง่ายขึ้น(ถ้าทำแบบ Bootstrap จะต้องมีการบอกช่วงเวลาของการออกเสียงของแต่ละเสียงในแต่ละคำเป็น Labeled ให้กับแต่ละคำในแต่ละไฟล์ด้วย) ในการสร้าง Transcription Files ตามวิธี Flat-Start นั้นจะมี Phone Transcription Files อยู่ 2 เซ็ทที่มีความจำเป็นในการใช้งาน เซ็ทแรกนั้นจะเป็นแบบที่ไม่มี Short-Pause(sp) Model ระหว่างคำ และเมื่อ Phone Model ได้ถูกสร้างขึ้นมาแล้ว sp model จะถูกแทรกลงไปในระหว่างคำเพื่อคอยดูแลในเรื่องของการหยุดในระหว่างคำที่เกิดขึ้นโดยผู้พูด

จุดเริ่มต้นของทั้งสองเซ็ทของ Phone Transcription คือ Orthographic Transcription ในรูปแบบของ HTK Label Format ซึ่งสามารถสร้างได้โดยใช้ Text Editor ซึ่ง Transcription File ที่จำเป็นต้องสร้างก็จะอยู่ในรูปแบบดังต่อไปนี้

#!MLF!#

"*/0001.lab"

หนึ่ง

.

"*/0002.lab"

สอง

.

"*/0003.lab"

สาม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

.

"*/0004.lab"

สี่

. (enter)

เป็นต้น

จาก Transcription ข้างต้น จะเห็นได้ว่าเมื่อเป็นรายการของคำพูดซึ่งจะเป็นเหมือนกับประโยคที่ทำการบันทึกเสียงไว้แต่ละคำจะถูกแยกอยู่คนละบรรทัด ซึ่งในแต่ละบรรทัดจะต้องประกอบด้วยคำที่มีอยู่ใน Dictionary ในบรรทัดแรกของรายการคำพูดจะเป็นตัวบ่งชี้ว่าไฟล์จะเป็น Master Label File(MLF) ซึ่งในไฟล์หนึ่งไฟล์นี้จะประกอบด้วยรายการคำพูดทั้งหมดที่เป็นคำพูดที่ใช้ในการทำการ trainในครั้งนั้นๆ

รูปแบบที่ใช้เป็น Path Name ใน MLF นั้นควรจะทำความเข้าใจเสียก่อนว่าเป็น pattern ไม่ใช่ชื่อ เมื่อ HTK ได้ทำการประมวลผลไฟล์ข้อมูลเสียง มันก็จะหา Transcription File ที่มีชื่อเหมือนไฟล์เสียง แต่ต่างกันที่ส่วนท้ายของ Path เช่น เมื่อไฟล์ /hwork1.1.2/sound/0001.wav ได้ถูกทำการประมวลผล HTK ก็จะทำการหา Label File ที่ชื่อว่า /hwork1.1.2/sound/0001.lab อย่างไรก็ตาม ในที่นี้เราใช้เครื่องหมาย ‘ * ‘ ทำให้ HTK ค้นหาจากคนละ path กันได้(หรือที่ใดๆก็ได้)

เมื่อไฟล์ MLF ในแบบ Word Level ได้ถูกสร้างขึ้น จะสามารถสร้างไฟล์ MLF ในแบบ Phone Level ได้โดยใช้คำสั่ง HLEd ตัวอย่างเช่น สมมติว่าไฟล์ MLF ในระดับ Word Level ที่แสดงให้ดูข้างต้นนั้นมีชื่อว่า words.mlf เราจะใช้คำสั่ง

*HLEd –l * –d dict –i phones0.mlf mkphones0.led words.mlf*

เพื่อทำการสร้าง Phone Level Transcription โดย –d เป็น option หมายถึงการอ้างอิง Dictionary ชื่ออะไร -i เป็นการบอกว่า output ที่ได้จะทำการเขียนลง MLF File ที่ชื่ออะไร –l มีความจำเป็นในการสร้าง Path ‘*’ ตามรูปแบบของ output ที่ได้ ซึ่งลักษณะของไฟล์ที่อยู่ในระดับ Phoneme(phones0.mlf) นั้นจะมีลักษณะดังต่อไปนี้

"*/0021.lab"

sil

n

vh1

ng

s

orh4

ng

s

ah4

m

s

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

uh4

n

h

o1

g

g

a2

w

h

ah2

s

uh4

n

sil

.

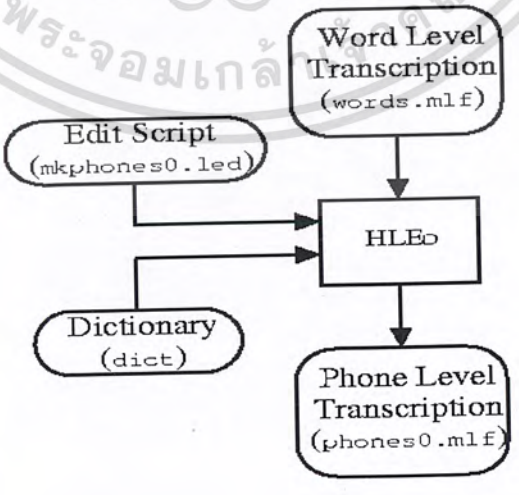
ส่วนไฟล์ที่ชื่อ mkphones0.led นั้นจะประกอบไปด้วยคำสั่งดังต่อไปนี้

EX

IS sil sil

DE sp

คำสั่ง EX เป็นการสั่งให้ทำการแทนที่แต่ละคำในไฟล์ words.mlf ด้วยการสะกดของแต่ละคำที่ได้จากไฟล์ Dictionary(dict) คำสั่ง IS เป็นการสั่งให้ทำการแทรก silence model sil ลงไปทุกๆครั้งที่มีการเริ่มต้นและจบประโยคหรือไฟล์ และคำสั่ง DE เป็นการทำการลบทุกๆ short pause sp labels ซึ่งไม่เป็นที่ต้องการในตอนนี้ออกให้หมด ซึ่งในการประมวลผลนี้ได้แสดงให้ดูในรูปที่ 5.1



รูปที่ 5.1 Flow-Chart ของการใช้คำสั่ง HLEd

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ขั้นตอนที่4 – การทำการ CODING DATA

ในขั้นตอนสุดท้ายของการทำการเตรียมข้อมูลนั้น คือการทำการแปลงค่าข้อมูลสัญญาณเสียงพูดจากสัญญาณเสียงพูดธรรมดาไปเป็น Feature vector ของเสียงพูดโดยเป็นการเอาสัญญาณเสียงไปผ่านกระบวนการที่ทำให้ได้ลักษณะเฉพาะของเสียงออกมา โดย HTK สามารถทำการวิเคราะห์ทั้งในแบบ FFT-based และ LPC-based ซึ่งในที่นี้ได้ใช้ Mel Frequency Cepstral Coefficients(MFCCs) ซึ่งได้มาจาก log spectra ของ FFT-based

การทำการแปลงข้อมูลจะทำได้โดยการใช้เครื่องมือของ HTK ที่ชื่อ HCopy และทำการกำหนด Configure เพื่อให้สามารถทำการแปลงข้อมูล Input ไปเป็น MFCC Vector ได้อย่างอัตโนมัติ ในการทำจะต้องมี Configuration File ในที่นี้ได้ใช้ชื่อว่า config_mfcc ซึ่งเป็นไฟล์ที่ระบุถึงค่า Parameter ที่มีความสำคัญที่ใช้ในการแปลง ซึ่งในที่นี้จะสามารถเขียนขึ้นมาได้ดังนี้

```
#Coding parameters
SOURCEKIND=WAVEFORM
SOURCEFORMAT=NOHEAD
SOURCERATE=625
TARGETKIND=MFCC_D_A_0
TARGETRATE=175000.0
SAVECOMPRESSED=T
SAVEWITHCRC=T
USEHAMMING=T
ENORMALISE=T
PREEMCOEF=0.97
NUMCHANS=24
NUMCEPS=12
WINDOWSIZE=250000.0
```

การตั้งค่าบางอย่างในนี้จะเป็นค่า default อยู่แล้ว แต่ที่แสดงให้ดูเพื่อความสมบูรณ์แบบ จากตัวอย่างจะระบุว่าค่า target parameter จะเป็น MFCC โดยใช้ค่า C_0 เป็น energy component ซึ่งมีค่า frame period 17.5 ms (HTK จะใช้หน่วย 100 ms) และ Output ที่ได้จะถูกบันทึกในแบบ Compressed Format นอกจากนั้นยังมีการใช้ CRC Checksum อีกด้วย FFT จะใช้ Hamming Window และสัญญาณจะมี First Order Preemphasis โดยใช้ค่าสัมประสิทธิ์ 0.97 Filterbank จะมี 24 channel และจะให้ค่าสัมประสิทธิ์12 MFCC เป็น Output และค่าตัวแปร ENORMALISE โดยปกติจะเป็นค่า true ซึ่งหมายความว่าจะมีการทำ Energy Normalisation บนไฟล์เสียงที่ถูกบันทึกมา ซึ่งการทำงานนี้ไม่สามารถทำได้กับ Live Audio ดังนั้นถ้าหากมีการใช้งานกับ Live Audio ค่านี้จะต้องมีการตั้งค่าให้เป็น false

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ในการทำคำสั่ง HCopy รายชื่อไฟล์ที่เป็น Source File กับ รายชื่อไฟล์ที่จะเป็น Output นั้นจะต้องถูกสร้างขึ้นดังตัวอย่าง

f:\hwork1.1.2\sound\0001.wav f:\hwork1.1.2\train\0001.mfc

f:\hwork1.1.2\sound\0002.wav f:\hwork1.1.2\train\0002.mfc

f:\hwork1.1.2\sound\0003.wav f:\hwork1.1.2\train\0003.mfc

f:\hwork1.1.2\sound\0004.wav f:\hwork1.1.2\train\0004.mfc

f:\hwork1.1.2\sound\0005.wav f:\hwork1.1.2\train\0005.mfc

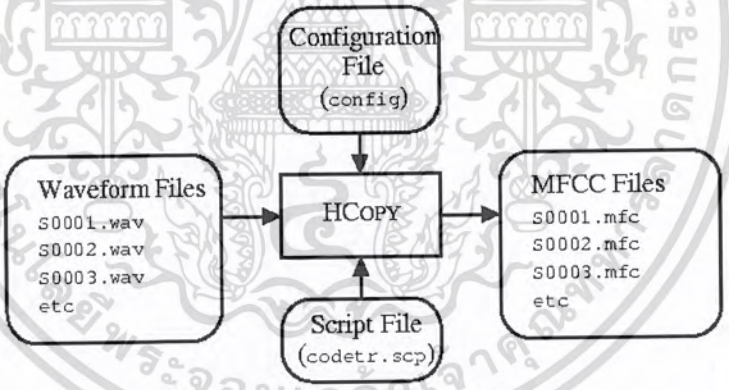
ไฟล์ที่ประกอบด้วยรายชื่อของไฟล์ดังข้างต้นนั้นจะถูกเรียกว่า Script Files และจะใช้นามสกุลเป็น scp สำหรับ Script Files นั้นจะสามารถระบุในคำสั่งได้โดยใช้ –S ซึ่งเป็น Option ซึ่งทำให้ในการทำงานกับคำสั่งนั้นสามารถประมวลผลไฟล์ทีละมากๆบนคำสั่งเดียว

สมมติว่า Script File ที่ประกอบด้วยชื่อไฟล์ตามที่แสดงไว้ด้านบนนี้อยู่ในไฟล์ชื่อ codetr.scp ข้อมูลที่จะนำมา train นั้นจะถูกทำการแปลงโดยใช้คำสั่งดังต่อไปนี้

HCopy –T 1 –C config_mfcc –S codetr.scp

การทำงานคำสั่งนี้แสดงดังรูปที่ 5.2

โดยจากตัวอย่างคำสั่งด้านบน –T 1 เป็นการทำให้เกิด Trace เพื่อติดตามผลการทำงาน –C หมายถึงการมีการใช้ Config File และ –S หมายถึงมีการใช้ Script File



รูปที่ 5.2 Flow-Chart ของการใช้ HCopy

ขั้นตอนที่ 5 –การสร้าง Flat Start Monophones

ในขั้นตอนแรกในการทำการ train จะต้องทำการระบุ Prototype Model ก่อน ซึ่งค่า parameters ต่างๆของ model นี้จะไม่มีความสำคัญ เพราะจุดประสงค์หลักก็คือการให้ค่าเริ่มต้นแก่แบบจำลอง ในระบบที่เป็น phone-based แบบจำลองที่ดีจะใช้ 3-state left-right ที่ไม่มีการกระโดดข้าม state ซึ่งสามารถให้ค่าได้ดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

```
~o <VecSize> 39 <MFCC_O_D_A> <StreamInfo> 1 39
~h "proto"
<BeginHMM>
 <NumStates> 5
 <State> 2 <NumMixes> 1
 <Stream> 1
 <Mixture> 1 1.0000
  <Mean> 39
   0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
  <Variance> 39
   1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
 <State> 3 <NumMixes> 1
 <Stream> 1
 <Mixture> 1 1.0000
  <Mean> 39
   0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
  <Variance> 39
   1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
 <State> 4 <NumMixes> 1
 <Stream> 1
 <Mixture> 1 1.0000
  <Mean> 39
   0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0 0.0
  <Variance> 39
   1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
 <TransP> 5
 0.000e+0  1.000e+0  0.000e+0  0.000e+0  0.000e+0
 0.000e+0  6.000e-1  4.000e-1  0.000e+0  0.000e+0
```

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

0.000e+0 0.000e+0 6.000e-1 4.000e-1 0.000e+0

0.000e+0 0.000e+0 0.000e+0 6.000e-1 4.000e-1

0.000e+0 0.000e+0 0.000e+0 0.000e+0 0.000e+0

<EndHMM>

ซึ่งจะเห็นได้ว่าแต่ละ ellipsed vector จะมีความยาว 39 ซึ่งค่า 39 นี้สามารถคำนวณได้จากความยาวของ Parameterised static vector(MFCC_0 =13) บวกกับจำนวน delta coefficient(+13) บวกกับจำนวน acceleration coefficient(+13)

เครื่องมือของ HTK ชื่อ HCompV จะทำการค้นหาเซ็ทของไฟล์ข้อมูลซึ่งจะทำการคำนวณค่า global mean และ variance และเซ็ททั้งหมดของ Gaussian ใน HMM ที่ให้มาเพื่อให้มีค่า mean และค่า variance ที่เหมือนกัน สมมติว่ารายการของไฟล์ที่ต้องการนำมา train อยู่ในไฟล์ที่ชื่อ train.scp ดังตัวอย่าง

f:\hwork1.1.2\train\0001.mfc

f:\hwork1.1.2\train\0002.mfc

f:\hwork1.1.2\train\0003.mfc

f:\hwork1.1.2\train\0004.mfc

f:\hwork1.1.2\train\0005.mfc

และไฟล์ config_mfcc ได้ถูกเปลี่ยนชื่อเป็น config โดยมีรายละเอียดดังนี้

#Coding parameters

SOURCEFORMAT=HTK

TARGETKIND=MFCC_0_D_A

TARGETRATE=175000.0

การใช้คำสั่ง *HCompV –C config –f 0.01 –m –S train.scp –M hmm0 proto*

จะเป็นการสร้าง proto ในรูปแบบใหม่ใน Directory ชื่อ hmm0 ซึ่งค่า zero means และ unit variances ตามข้างต้นจะถูกแทนที่ด้วย global speech mean และ variance จากตัวอย่างของ prototype HMM ที่ระบุไว้ด้านบนนั้น ได้มีการระบุค่า parameter เป็น MFCC_0_D_A ซึ่งมีความหมายว่าค่าสัมประสิทธิ์ delta และ acceleration จะถูกคำนวณและถูกแนบด้วยค่าสัมประสิทธิ์ statistic MFCC ที่ถูกคำนวณแล้ว และถูกเก็บระหว่างทำการแปลงข้อมูล(Coding Process)

จากตัวอย่างคำสั่งจะเห็นได้ว่า มีการใช้ Option ซึ่งก็คือ –f เป็นการทำให้ค่า Variance Floor Macro(vFloors) ถูกสร้างขึ้นซึ่งมีค่าเท่ากับ 0.01 คูณด้วยค่า Global Variance ซึ่งจะนำค่านี้ไปใช้ต่อไป -m เป็นการเรียกให้มีการคำนวณค่า means เหมือนกับ variances. Prototype ที่ได้ใหม่นี้จะถูกเก็บไว้ใน directory ที่ชื่อ hmm0 เราจะสามารถสร้าง hmmdefs ซึ่งเป็นไฟล์ที่ประกอบไปด้วย Monophones HMMs ที่ต้องการซึ่งสามารถสร้างโดยการทำการคัดลอกจาก prototye ที่ได้และทำการเปลี่ยนชื่อเป็นชื่อ monophone ที่ต้องการโดย Monophone HMMs นั้น 1 ชื่อจะมี 1 model และจะต้องมีให้ครบตามชื่อ phone ที่ได้จาก phone level mlf file(phones0.mlf) รูปที่ 5.3 นี้จะแสดงถึงรูปแบบของ Master Macro Files ซึ่งจะมีไฟล์ที่ชื่อ macros และ hmmdefs

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

macros

```
~o
<STREAMINFO> 1 39
<VECSIZE> 39<NULLD><MFCC_D_A_0>
~v "varFloor1"
<VARIANCE> 39
 2.386773e-001 2.781330e-001 ...
```

hmmdefs

```
~h "a"
<BEGINHMM>...
<ENDHMM>
~h "d"
<BEGINHMM>...
```

รูปที่ 5.3 รูปแบบของ Master Macro Files

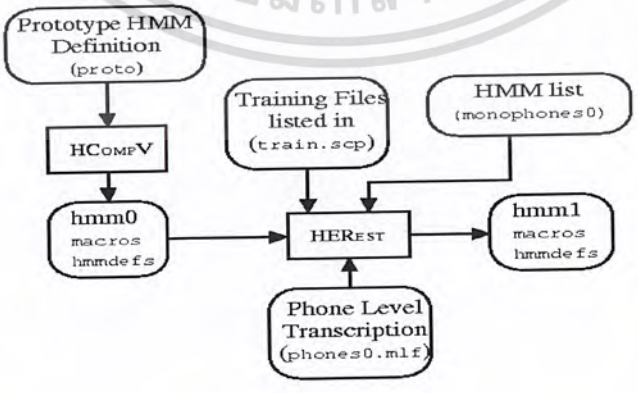
Flat Start monophones ที่ถูกเก็บค่าไว้ใน hmm0 จะถูกทำการคำนวณซ้ำ(Re-Estimated) โดยใช้เครื่องมือสำหรับการทำ re-estimation เรียกว่า HERest โดยจะสามารถใช้ได้ดังนี้

HERest –C config –I phones0.mlf –t 250.0 150.0 1000.0 –S train.scp –H hmm1\macros –H hmm1\hmmdefs –M hmm2 monophones0

จะทำให้เกิดผลคือระบบจะทำการนำค่าของแบบจำลองจาก hmm0 ตามที่มีในรายการแบบจำลอง monophones0(ไฟล์นี้ประกอบไปด้วยรายการของ phoneme ที่จะใช้ทั้งหมด ยกเว้น sp ซึ่งจะถูกเพิ่มเข้าไปในไฟล์ monophones1) ซึ่งจะทำการ re-estimate โดยใช้ข้อมูลที่มีใน train.scp และจะทำให้ได้แบบจำลองใหม่ซึ่งจะเก็บไว้ใน directory ที่ชื่อ hmm1 ไฟล์ส่วนมากที่ใช้ในคำสั่ง HERest ได้เคยอธิบายมาแล้ว ยกเว้นไฟล์ macros ซึ่งจะมี global options macro และ variance floor macro vFloors ซึ่งถูกสร้างมาก่อนหน้านี้ โดย global options macro จะอธิบายถึงชนิดของ parameter ของ HMM และ ขนาดของ vector ซึ่งจะมีลักษณะดังนี้ ~o

<STREAMINFO> 1 39

<VECSIZE> 39<NULLD><MFCC_D_A_0>



รูปที่ 5.4 Flow-Chart ของการใช้ HERest

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โดยการใช้ option –t จะเป็นการตั้งค่า Pruning Thresholds เพื่อใช้ในระหว่างทำการ train ซึ่งจะสามารถลดการคำนวณลงได้ ไฟล์ที่ใช้ในการ train ส่วนใหญ่นั้นสามารถกำหนดค่า pruning threshold ให้น้อยได้ แต่ในบางไฟล์ไม่สามารถทำได้เพราะอาจทำให้ประสิทธิภาพนั้นต่ำลง จึงต้องการค่า pruning ที่มากขึ้น HERest นั้นสามารถเพิ่มค่า pruning threshold ได้เองโดยอัตโนมัติ จากตัวอย่างข้างต้น pruning นั้นโดยทั่วไปจะอยู่ที่ 250.0 ถ้าหากการทำการ re-estimation นั้นไม่ประสบผลสำเร็จค่า threshold จะถูกเพิ่มขึ้นทีละ 150.0 และไฟล์นั้นจะถูกประมวลผลใหม่ ซึ่งจะมีการทำซ้ำจนกระทั่งได้ผลตามต้องการ หรือเมื่อค่า pruning นั้นเกิน 1000.0 ณ. จุดนี้จะเป็นการปลอดภัยถ้าสมมติว่ามีปัญหาเกิดขึ้นกับไฟล์ที่ใช้ในการ train และจำเป็นจะต้องแก้ไขโดยด่วน(ซึ่งส่วนใหญ่นั้นจะเกิดจาก transcriptionที่ไม่ถูกต้อง) หรือไม่ก็ทำการ train โดยใช้ไฟล์ใหม่ รูป 5.4 เป็นการแสดงให้เห็นถึงการประมวลผลที่เกิดขึ้นในขั้นตอนนี้

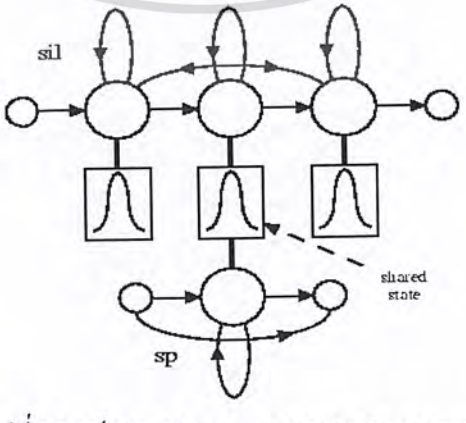
options ที่ใช้ในที่นี้ –H หมายถึงจะนำไฟล์ input จากที่ใด(Path)มาใช้ในการทำ re-estimation และ -M นั้นบอกว่าไฟล์ output ที่ได้จะถูกเก็บไว้ที่ใด สำหรับ –I นั้นเป็นการบ่งบอกถึงการทำการเรียกข้อมูลไฟล์แบบ MLF

ในแต่ละครั้งที่ใช้คำสั่ง HERest นั้นมันจะทำการ re-estimation เพียงครั้งเดียว ซึ่งในความเป็นจริงนั้นควรที่จะทำ HERest นี้ 2 ครั้งโดยเปลี่ยนชื่อ input และ output directory ใน option –H และ –M ตามลำดับ

ขั้นตอนที่ 6 – Fixing the Silence Models

ในขั้นตอนก่อนหน้านี้เป็นการสร้าง 3 state left-to-right HMM สำหรับแต่ละ phone และรวมไปถึง HMM สำหรับ silence model sil ขั้นตอนต่อไปนี้จะเป็นการเพิ่มเติมการย้าย state แบบพิเศษจาก state2 ไป state4 และจาก state4 กลับไปยัง state2 ใน silence model ซึ่งแนวคิดนี้ทำให้แบบจำลองนั้นมีประสิทธิภาพดีขึ้น โดยยอมให้แต่ละ state มีการดูดซับ ลดทอนสัญญาณรบกวนต่างๆในข้อมูลที่ใช้ในการ train โดยการที่มีการย้ายกลับหลังได้นั้นทำให้ไม่ต้องเกิดการย้ายไปสู่คำถัดไปในทันที

และในขั้นตอนนี้ 1 state short pause sp model จะถูกสร้างขึ้น และถูกเรียกว่า tee-model ซึ่งมีการเปลี่ยน state โดยตรงจาก entry node ไปยัง exit node และ sp ยังมี state ที่เป็น emitting sate ที่ผูกกับ state กลางของ silence model



รูปที่ 5.5 รูปของ silence model และ short pause

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

วิธีในการสร้าง silence model นั้นประกอบด้วย 2 ขั้นตอนต่อไปนี้คือ

1.ใช้ text editor เพื่อทำการคัดลอก state ตรงกลางของ sil model จาก hmm3\hmmdefs เพื่อทำการสร้าง sp model ใหม่และทำการเก็บไฟล์ MMF hmmdefs ซึ่งมี sp model อันใหม่ไว้ใน directory ที่ชื่อ hmm4(ดูตัวอย่างได้ที่ภาคผนวก ก. และให้สังเกตค่าตัวเลขให้ดี)

2.ใช้ HMM editor ที่ชื่อ HHEd เพื่อทำการเพิ่มการย้าย state แบบพิเศษที่ต้องการ และทำการเชื่อม(tied) state ของ sp กับ state กลางของ sil

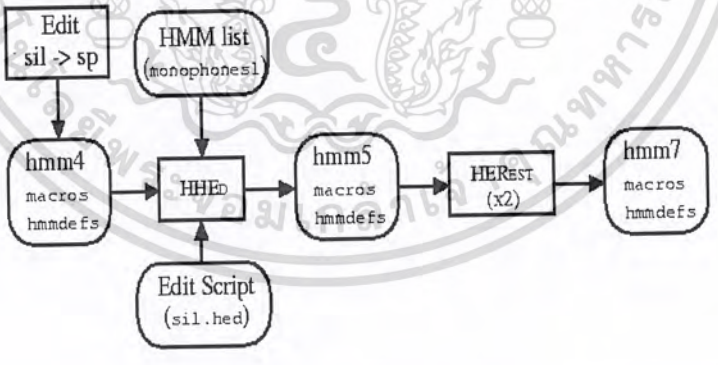
HHEd นั้นทำงานในลักษณะเดียวกันกับ HLEd คือมีการใช้เซ็ทของคำสั่งที่มีอยู่ใน script file ช่วยในการปรับปรุงเซ็ทของ HMM โดยในการใช้ HHEd นั้นจะใช้คำสั่งดังนี้

HHEd -H hmm4\macros –H hmm4\hmmdefs –M hmm5 sil.hed monophones1

โดยที่ไฟล์ sil.hed จะประกอบด้วยคำสั่งดังต่อไปนี้

```
AT 2 4 0.2 {sil.transP}
AT 4 2 0.2 {sil.transP}
AT 1 3 0.3 {sp.transP}
TI silst {sil.state[3],sp.state[2]}
```

คำสั่ง AT นั้นเป็นการเพิ่มการย้าย state ให้แก่ trasition matrices และคำสั่ง TI จะเป็นการสร้าง tied-state ที่เรียกว่า silst ซึ่งค่า parameters ต่างๆของ tied-state นี้จะถูกเก็บไว้ในไฟล์ hmmdefs และในแต่ละ silence model ค่า parameters ของ stateเดิมที่เคยมีอยู่จะถูกแทนที่ด้วยชื่อของ macro เมื่อมาถึง ณ จุดนี้สามารถบอกได้ว่ารายการของ phone ที่มีตอนนี้ได้เปลี่ยนไปแล้ว เพราะว่า monophones0 นั้นถูกเปลี่ยนแปลงไปโดยการเพิ่ม sp model เข้ามา ซึ่งไฟล์ใหม่ที่ได้นี้ถูกตั้งชื่อว่า monophones1 และได้ถูกใช้โดยคำสั่ง HHEd ที่ได้แสดงให้ดูตามข้างต้น



รูปที่ 5.6 Flow-chart ของการทำคำสั่ง HHEd

ในตอนสุดท้ายของขั้นตอนนี้จะมีการใช้ HERest 2 ครั้ง โดยใช้ phone transcription ที่มี sp model ในระหว่างคำดังนี้

HERest –C config –I phones0.mlf –t 250.0 150.0 1000.0 –S train.scp –H hmm5\macros –H hmm5\hmmdefs –M hmm6 monophones1

ซึ่งเมื่อทำถึงขั้นนี้แล้วนั้นจะได้ hmm7 ขึ้นมา โดยในขั้นตอนที่ 7 นี้สามารถแสดงให้เห็นได้โดยรูปที่ 5.6

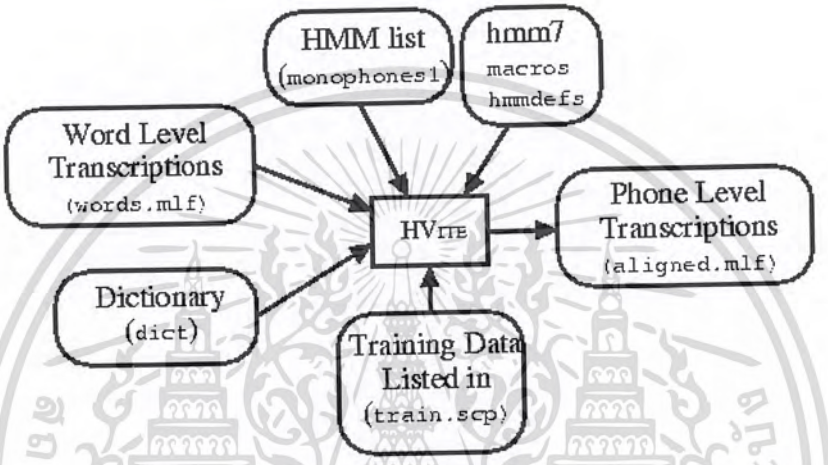
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ขั้นตอนที่ 7 –Realigning the Training Data

ดังที่ได้กล่าวไว้ก่อนจะเข้าสู่ขั้นตอนที่ 6 แล้วว่าคำบางคำนั้นสามารถออกเสียงได้มากกว่า 1 แบบ ในขั้นตอนนี้เราจึงจะทำการสร้าง phone model อันใหม่ขึ้นมาเพื่อให้แบบจำลองนั้นสามารถแยกแยะเสียงเหล่านั้นได้ โดยการใช้คำสั่ง

*Hvite –l * –b silence –C config –a –H hmm7\macros –H hmm7\hmmdefs –I aligned.mlf -m –t 250.0 –y lab –I words.mlf –S train.scp dict monophones1*

คำสั่งนี้จะใช้ HMMs ใน hmm7 เพื่อทำการแปลงจาก word level transcription (words.mlf) ให้กลายเป็น phone level transcription (aligned.mlf) โดยใช้การออกเสียงจากไฟล์ dict ดังรูปที่ 5.7



รูปที่ 5.7 การใช้คำสั่ง HVite

สิ่งสำคัญที่คำสั่งนี้ต่างไปจากการแปลงจาก word เป็น phone โดยคำสั่ง HLEd ในขั้นตอนที่ 3 คือ recogniser จะพิจารณาการออกเสียงของแต่ละคำทั้งหมดและ output ที่ได้จะตรงกับข้อมูลเสียงมากที่สุด

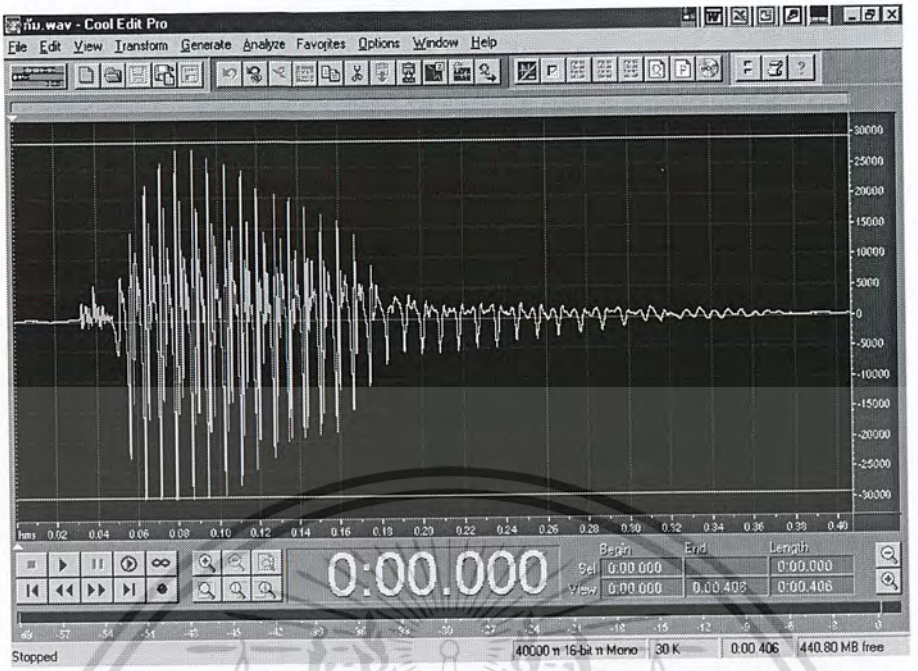
คำสั่งนี้ option –b จะใช้เพื่อแทรก silence model ลงไปที่ต้นและท้ายของแต่ละ utterance ส่วน –t ใช้ในการตั้งค่า pruning คือ 250.0 ตอนนี้ phone ใหม่ได้ถูกสร้างขึ้นมาแล้ว เราจึงทำการ re-estimate อีก 2 ครั้งโดยใช้คำสั่ง HERest ดังนี้

HERest –C config –I aligned.mlf –t 250.0 150.0 1000.0 –S train.scp –H hmm7\macros –H hmm7\hmmdefs –M hmm8 monophones1

หลังจากได้ทำคำสั่งนี้แล้ว เราก็จะได้ไฟล์ที่ชื่อ aligned.mlf ซึ่งรายละเอียดในไฟล์นี้จะบ่งบอกถึงช่วงเวลาของแต่ละโฟนีมของแต่ละคำที่เราได้ทำการบันทึกเสียงเอาไว้ ช่วงเวลาที่ระบุมาจะมีหน่วยเป็น 100 นาโนวินาที

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตัวอย่างไฟล์ที่ทำการบันทึก



รูปที่ 5.8 ไฟล์ที่ทำการบันทึกคำว่า "ก้ม"

ตัวอย่างส่วนหนึ่งของไฟล์ aligned.mlf ที่ระบุช่วงเวลาของคำว่า "ก้ม"

#!MLF!#

"*/0001.lab"

0 350000 sil -366.190216 silence

350000 1050000 g -619.151123 \241\351\301

1050000 1750000 o2 -379.381317

1750000 3675000 m -739.867249

3675000 3675000 sp -1.203973

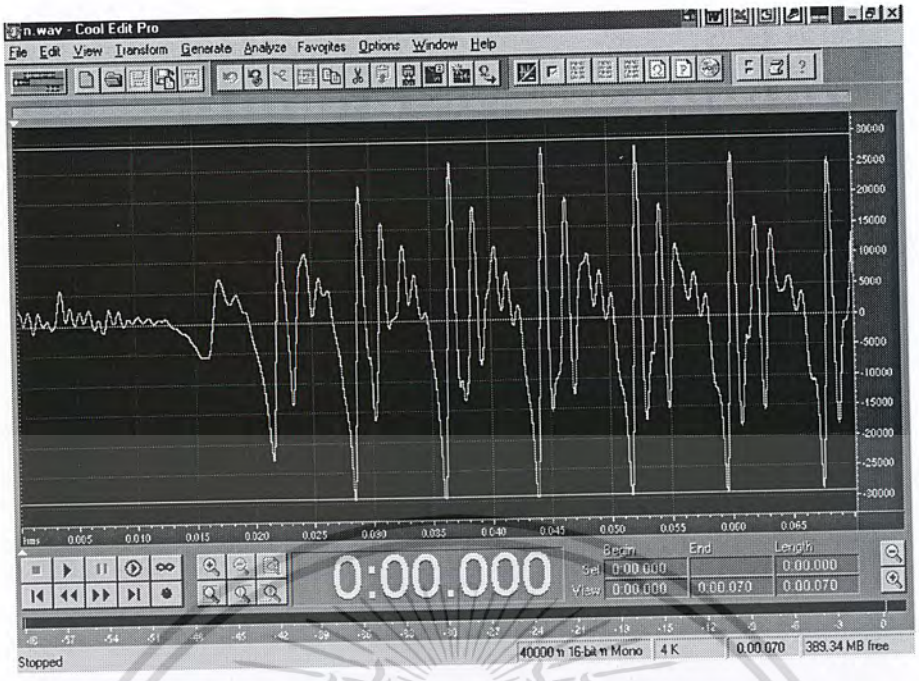
3675000 4025000 sil -128.435760 silence

.

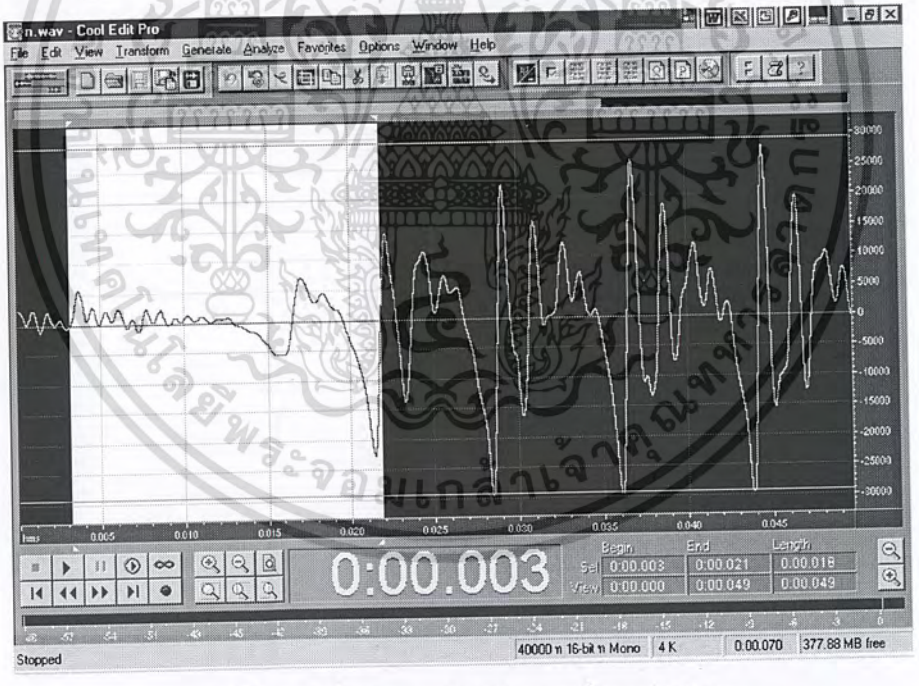
5.3 สรุปการใช้งาน

จากช่วงเวลาที่โปรแกรม HTK ได้ระบุมานั้น เมื่อได้นำไปทดสอบฟังดูปรากฏว่าเสียงที่ได้ออกมานั้นไม่ตรงกับความเป็นจริง เช่นช่วงเวลาของเสียงโฟนีม "ก" (g) ที่ HTK ระบุมาคือ 0.035-0.105 วินาที เมื่อเราได้ทำการแบ่งไฟล์ที่ได้บันทึกไว้ตามช่วงเวลาที่ระบุมาดังรูปที่ 5.9 แล้วลองฟังดู นอกจากจะได้ยินเสียงโฟนีมของ "ก" แล้วยังมีเสียงโฟนีมของสระ โอะ ที่ผันวรรณยุกต์โทติดมาด้วย และเมื่อเราได้ลองกับโฟนีมอื่นๆของคำอื่นด้วยก็จะได้ผลลัพธ์ในลักษณะเดียวกันนี้ นอกจากนั้นช่วงเวลาที่ได้ยังไม่ตรงกับจุดที่คลื่นเสียงจะครบลูกคลื่นพอดี ทำให้เมื่อต้องการนำแต่ละโฟนีมมารวมกันเป็นพยางค์เพื่อทำการสังเคราะห์เสียงแล้วเฟสของคลื่นจะไม่ตรงกัน ทำให้ผลลัพธ์คือเสียงที่ได้ออกมาฟังไม่ค่อยรู้เรื่องหรือไม่ดีนัก ซึ่งเสียงของโฟนีม "ก" ที่ควรจะได้ควรจะใกล้เคียงกับส่วนที่เป็นแถบสีขาวดังในรูปที่ 5.10

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 5.9 คลื่นเสียงของโฟนีม "ก" ที่โปรแกรม HTK ระบุมา



รูปที่ 5.10 คลื่นเสียงของโฟนีม "ก" ที่ควรจะได้จริง

จากผลลัพธ์ที่ได้ออกมาทำให้เราสามารถสรุปได้ว่าโปรแกรม HTK นั้นไม่สามารถระบุช่วงเวลาของแต่ละโฟนีมได้ดีเท่าใดนัก จึงไม่สามารถใช้โปรแกรมนี้เพื่อช่วยในการแบ่งไฟล์เสียงออกเป็นโฟนีมได้ตามที่เข้าใจในตอนแรก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 6
การบันทึกเสียงและการแบ่งเสียงออกเป็นโฟนีม

6.1 บทนำ

เนื่องจากการใช้งานโปรแกรม HTK เพื่อช่วยในการแบ่งไฟล์เสียงที่ทำการบันทึกออกเป็นโฟนีม ได้เกิดปัญหาขึ้นดังที่ได้กล่าวมาแล้วในบทที่ 5 เราจึงได้ตัดสินใจที่จะใช้การสังเกตและทดลองแบ่งไฟล์ออกเป็นโฟนีมเอง ดังจะได้กล่าวต่อไป

6.2 ขั้นตอนในการทำงาน

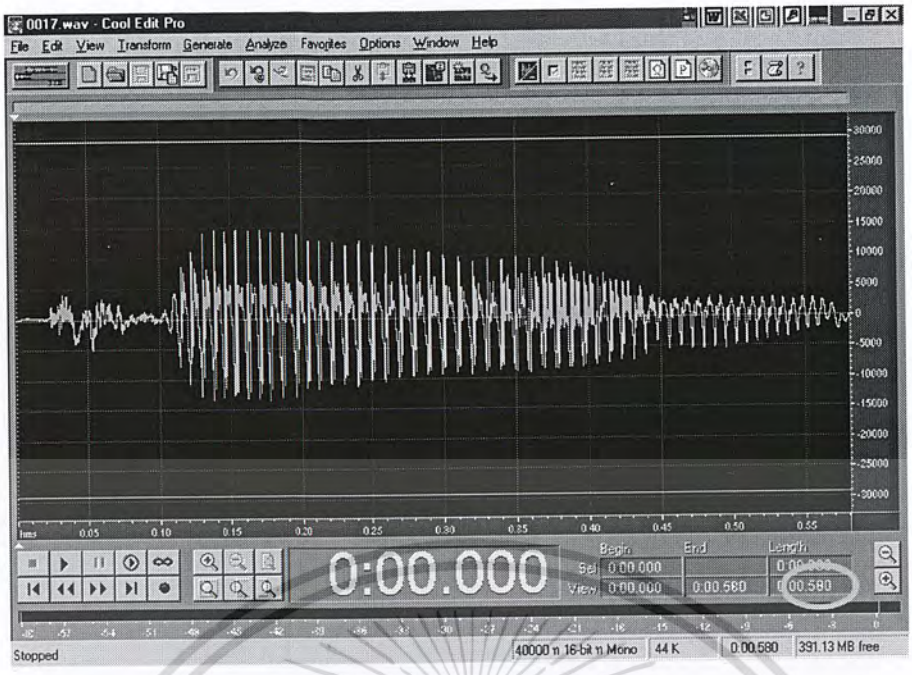
ขั้นตอนที่ 1 –บันทึกเสียง

ในขั้นตอนแรกนี้เราจะทำการบันทึกเสียงที่จำเป็นต้องใช้งานก่อน ซึ่งก็จะต้องครอบคลุมทุกโฟนีมที่จะนำมาใช้เป็นฐานข้อมูลเสียงเพื่อนำมารวมกันให้เป็นพยางค์และคำต่อไป โดยเราจะทำการบันทึกเสียงในแบบ 16 bit mono ที่ sampling rate เท่ากับ 44 KHz โดยใช้โปรแกรม Cool Edit Pro

เสียงที่เราจะทำการบันทึกนี้อาจจะอยู่ในลักษณะเป็นคำๆเดียวหรือเป็นประโยคก็ได้ แต่จากการทดลองได้พบว่าการบันทึกเป็นประโยคจะให้ผลลัพธ์ออกมาดีกว่าการบันทึกเป็นคำๆเดียว เพราะว่าการบันทึกเป็นคำเดี่ยวๆ โดยธรรมชาติของมนุษย์จะออกเสียงลากยาวออกไปมากกว่าปกติ จะมีลักษณะไม่เป็นธรรมชาติเหมือนเวลาที่เราพูดคุยกันตามปกติ ซึ่งเสียงที่ออกมาแต่ละคำจะสั้นกว่าการออกเสียงเดียวกันเพียงคำเดียว เช่น จากรูปที่ 6.1 ในแถบสีขาวเป็นคำว่า “โขง” ที่บันทึกในลักษณะเป็นประโยค ซึ่งใช้เวลาเพียง 0.426 วินาที ดังเวลาที่ปรากฏในวงกลมมุมล่างขวามือ แต่ถ้าเป็นคำว่า “โขง” ที่บันทึกแบบเป็นคำเดี่ยวๆแล้ว เวลาที่ใช้คือ 0.580 วินาที ดังเวลาที่ปรากฏในวงกลมมุมล่างขวามือของรูปที่ 6.2

รูปที่ 6.1 การบันทึกเสียงในลักษณะเป็นประโยค

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 6.2 การบันทึกเสียงในลักษณะเป็นคำเดี่ยวๆ

อย่างไรก็ตามการบันทึกในลักษณะเป็นประโยคก็ยังมีข้อเสียเหมือนกัน เนื่องจากการบันทึกในลักษณะนี้ถ้าไม่ระวังให้ดี เสียงที่ได้อาจจะเพี้ยนได้ง่ายๆ โดยที่เราไม่รู้ตัว เพราะในการฟังเสียงที่บันทึกในแบบประโยคจะทำให้เรารับรู้ได้เองว่าแต่ละคำควรจะได้ยินว่าอะไร เนื่องจากความหมายของประโยคที่บันทึกจะบ่งบอกเอง แต่เมื่อเราเลือกฟังเฉพาะบางคำที่อยู่ในประโยค เราอาจจะพบว่าคำๆนั้นเราได้ออกเสียงเพี้ยนไป โดยมักจะเพี้ยนในส่วนของเสียงวรรณยุกต์ ซึ่งเป็นส่วนสำคัญของภาษาไทยที่ทำให้เกิดคำที่มีความหมายแตกต่างไปจากเดิมได้มาก ต้องระวังให้ดี

ตัวอย่างประโยคที่เราได้ทำการบันทึกไว้เพื่อใช้ในการแบ่งเสียงออกเป็นโฟนีมมีดังนี้

รักดีหามจั่ว	รักชั่วหามเสา	ขี่ช้างจับตั๊กแตน
ผีฝากถุงข้าวสารให้ฉัน	ไก่จิกเด็กตายบนปากโอ่ง	งูใหญ่นอนอยู่ ณ ริมวัดโมพีโลก
น้ำร้อนปลาเป็น	โปรแกรมคอมพิวเตอร์	อ่านออกเสียงภาษาไทย
สถาบันเทคโนโลยี	พระจอมเกล้า	ลาดกระบัง
เธอจะไปเที่ยวไหน	รถไฟมาแล้ว	คุณเตือนด้วยความหวังดี
เปิ้ลใส่เสื้อสีเหลือง	ถือเงินไปซื้อของ	ถึงตลาดสดแห่งหนึ่ง
ซึ่งหรูหรามาก	ไม่เปอะเลอะเทอะ	เธอก้มดูเงาะและชมพู่
เห็นว่ามันสวยดี	แต่พึ่งนึกขึ้นได้	เธอต้องการหอยฝาเดียว
เอาไปเคี้ยวเล่น	จึงเยื้องกายต่อไป	จนเจอเพื่อนซี้ชื่อนุ่ม
ซึ่งรู้สิ้นทุกอย่าง	แต่ชอบเหม่อลอย	ที่แย่กว่านั้น
เธอกลัวเสื่อมาก	ในที่สุด	ก็ขนโจ๊กใส่เป้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

แล้วตื่นขึ้นมา	เพราะแสงแดดส่อง	กับดินสอแข็งแข็งหนึ่งแท่ง
เป่าปี่ดังลั่น	เสื้อเชิ้ตเกาะอก	เหยินขับรถเฉี่ยวแม่น้ำโขง
เสริมใส่กางเกงขาก๊วย	สินเหล่าดูสะดือ	เคว้งคว้างเฉไฉ

คำที่เราได้ทำการเลือกมาใช้ในการแบ่งเสียงออกเป็นโฟนีมของสระที่ผันเสียงวรรณยุกต์แล้ว ของพยัญชนะต้น และของตัวสะกดที่ผันเสียงวรรณยุกต์แสดงดังตารางที่ 6-1, 6-2 และ 6-3 ตามลำดับ

โฟนีม	คำ	โฟนีม	คำ	โฟนีม	คำ	โฟนีม	คำ	โฟนีม	คำ
a	บัน	v	จึง	e	เป็น	o	บน	er	เงิน
a1	กระ	v1	หนึ่ง	e1	เด็ก	o1	สด	er1	เปอะ
a2	ลั่น	v2	ซึ่ง	e2	เล่น	o2	ก้ม	er2	เปิ้ล
a3	พระ	v3	นึก	e3	เทศ	o3	รถ	er3	เทอะ
a4	หวัง	v4	ถึง	e4	เห็น	o4	ขน	er4	เหยิน
ah	ภา	vh	ดือ	eh	เก่ง	oh	โน	erh	เธอ
ah1	อ่าน	vh1	ตื่น	eh1	เหล่	oh1	โอ่ง	erh1	เหม่อ
ah2	ลาด	vh2	ชื่อ	eh2	เป้	oh2	โลก	erh2	เตอร์
ah3	ช้าง	vh3	ซื้อ	eh3	เคว้ง	oh3	โจ๊ก	erh3	เชิ้ต
ah4	ถา	vh4	ถือ	eh4	เฉ	oh4	โขง	erh4	เสริม
I	ริม	u	คุณ	x	แกรม	or	คอม	uah	กลัว
i1	จิก	u1	สุด	x1	แห่ง	or1	เกาะ	uah1	จั่ว
i2	สิ้น	u2	นุ่ม	x2	แท่ง	or2	ต้อง	uah2	ชั่ว
i3	พิว	u3	ทุก	x3	และ	or3	เงาะ	uah3	ก๊วย
i4	สิน	u4	ถุง	x4	แข็ง	or4	หอย	uah4	สวย
ih	ยี	uh	ง	xh	แตน	orh	จอม	iah	เดียว
ih1	ปี่	uh1	อยู่	xh1	แต่	orh1	ออก	iah1	เฉี่ยว
ih2	ที่	uh2	พู่	xh2	แย่	orh2	ชอบ	iah2	เที่ยว
ih3	ซี้	uh3	รู้	xh3	แล้ว	orh3	ร้อน	iah3	เคี้ยว
ih4	ผี	uh4	หรู	xh4	แสง	orh4	ของ	iah4	เสียง
vah	เตือน	vah1	เสื่อ	vah2	เสื้อ	vah3	เยื้อง	vah4	เหลือง

ตารางที่ 6-1 คำที่เลือกมาใช้ในการแบ่งเสียงออกเป็นโฟนีมของสระที่ผันเสียงวรรณยุกต์แล้ว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

โฟนีม	คำ	โฟนีม	คำ	โฟนีม	คำ	โฟนีม	คำ	โฟนีม	คำ
g	การ	th	ทาง	l	ลาม	kr	คราว	tr	ทรา
kh	คาง	n	นาน	w	วาง	kl	คลาย	dr	ดรา
ng	งาน	b	บาง	s	ซาว	kw	ควาย	fr	ฟราน
j	จาม	p	ปาน	h	ฮา	pr	ปราม	fl	ฟลาย
ch	ชาม	ph	พาย	oo	อาน	pl	ปลาย	br	บรา
y	ยาว	f	ฟาง	gr	กราม	phr	พราน	bl	บลา
d	ดาว	m	มา	gl	กลาย	phl	พลาง		
dt	ตาย	r	ราว	gw	กวาง	tr	ตรา		

ตารางที่ 6-2 คำที่เลือกมาใช้ในการแบ่งเสียงออกเป็นโฟนีมของพยัญชนะต้น

โฟนีม	คำ	โฟนีม	คำ	โฟนีม	คำ	โฟนีม	คำ	โฟนีม	คำ
n_	บัน	ng_	ยาง	m_	จาม	y_	กาย	w_	กาว
n_1	ป่าน	ng_1	อย่าง	m_1	ห่าม	y_1	จ่าย	w_1	ข่าว
n_2	บ้าน	ng_2	ย่าง	m_2	ย่าม	y_2	ร่าย	w_2	ข้าว
n_3	ล้าน	ng_3	ล้าง	m_3	ม้าม	y_3	ท้าย	w_3	ร้าว
n_4	หลาน	ng_4	ฉาง	m_4	หลาม	y_4	หลาย	w_4	ขาว
g_1	กาก	d_1	บาด	b_1	ดาบ				
g_2	มาก	d_2	ลาด	b_2	ราบ				
g_3	พัก	d_3	วัด	b_3	ซับ				

ตารางที่ 6-3 คำที่เลือกมาใช้ในการแบ่งเสียงออกเป็นโฟนีมของพยัญชนะสะกดที่ผันวรรณยุกต์แล้ว

ขั้นตอนที่ 2 –แบ่งเสียงออกเป็นโฟนีม

เมื่อได้ทำการบันทึกเสียงเรียบร้อยแล้ว ในขั้นตอนนี้เราจะนำไฟล์เสียงที่บันทึกไว้แล้วนั้นมาทำการแบ่งออกเป็นโฟนีมตามที่เราต้องการ ดังมีรายละเอียดดังนี้

1. เปิดไฟล์เสียงที่มีคำที่มีโฟนีมที่ต้องการ เช่น ถ้าเราต้องการโฟนีม /ah4/ เราก็จะเปิดไฟล์เสียงที่บันทึกไว้ว่า “สถาบันเทคโนโลยี” ดังรูปที่ 6.3

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

รูปที่ 6.3 ไฟล์เสียงของประโยค "สถาบันเทคโนโลยี"

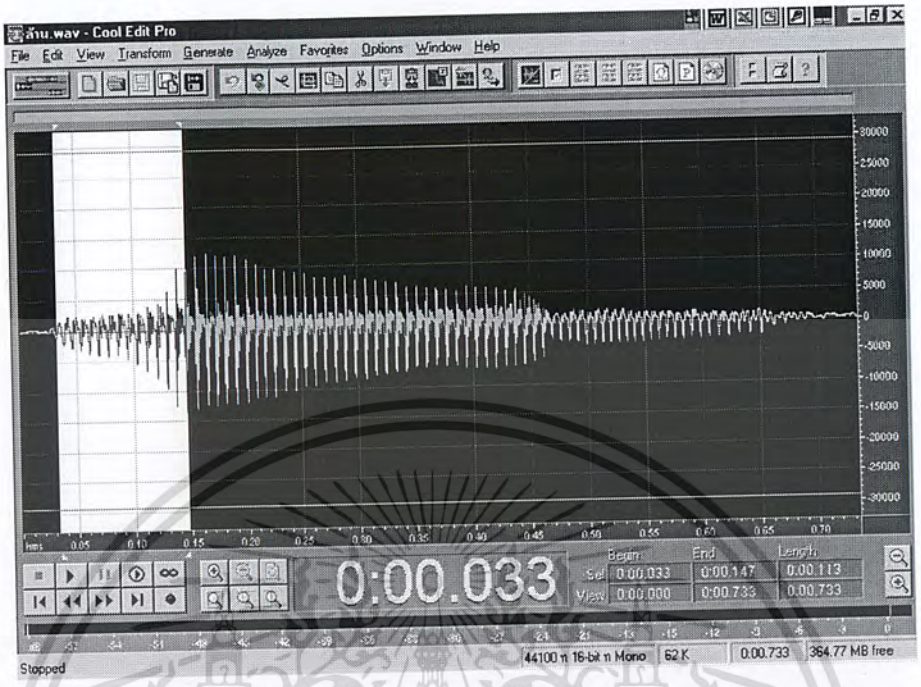
2. เลือกช่วงที่คาดว่าน่าจะเป็นส่วนของโฟนีมที่ต้องการ แล้วลองฟังดูว่าใช่ที่ต้องการจริงหรือไม่ ถ้าไม่ใช่ก็ให้เลื่อนไปเลือกช่วงอื่นๆจนกว่าจะได้ที่ต้องการจริงๆ ดังรูปที่ 6.4

รูปที่ 6.4 ส่วนของโฟนีมที่ต้องการ (/ah4/)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

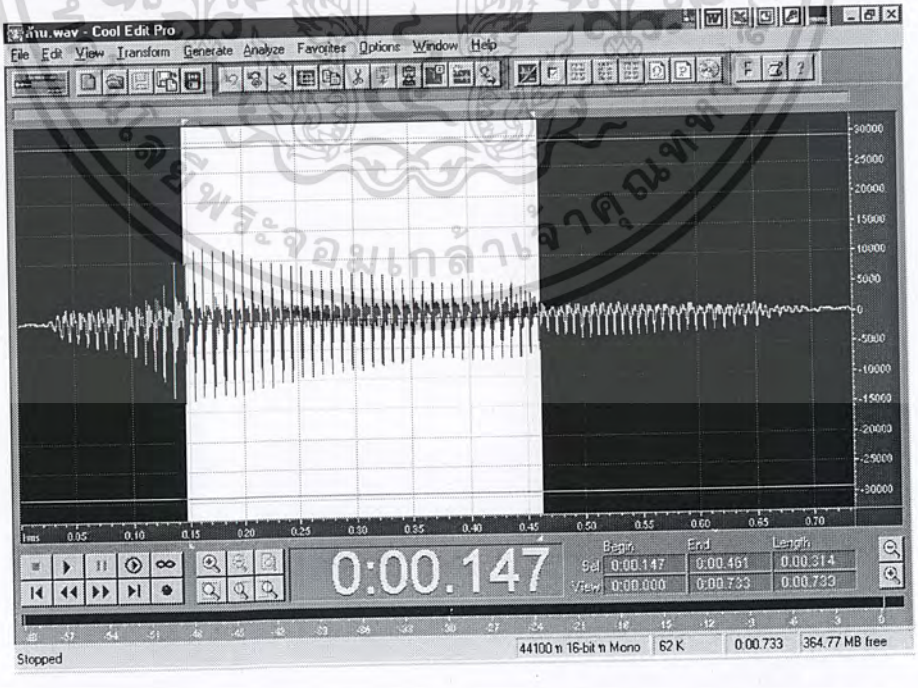
6.3 ข้อสังเกตจากการเลือกโฟนีมที่ต้องการ

1. เสียงของพยัญชนะต้นจะสั้นมาก อยู่ในช่วงต้นของคำ ดังตัวอย่างในรูปที่ 6.5



รูปที่ 6.5 เสียงของพยัญชนะต้น /ล/ จากคำว่า "ล้าน"

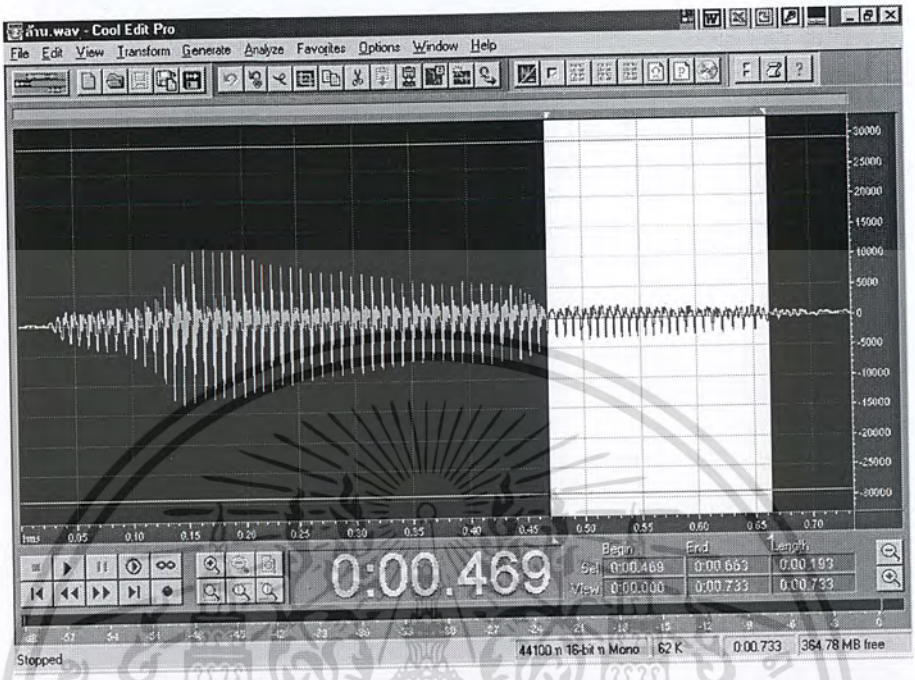
2. เสียงของสระที่ผันวรรณยุกต์จะค่อนข้างยาวเมื่อเทียบกับเสียงของพยัญชนะต้นและเสียงของตัวสะกดที่ผันวรรณยุกต์ โดยเกือบจะเป็นทั้งหมดของคำ ดังตัวอย่างในรูปที่ 6.6



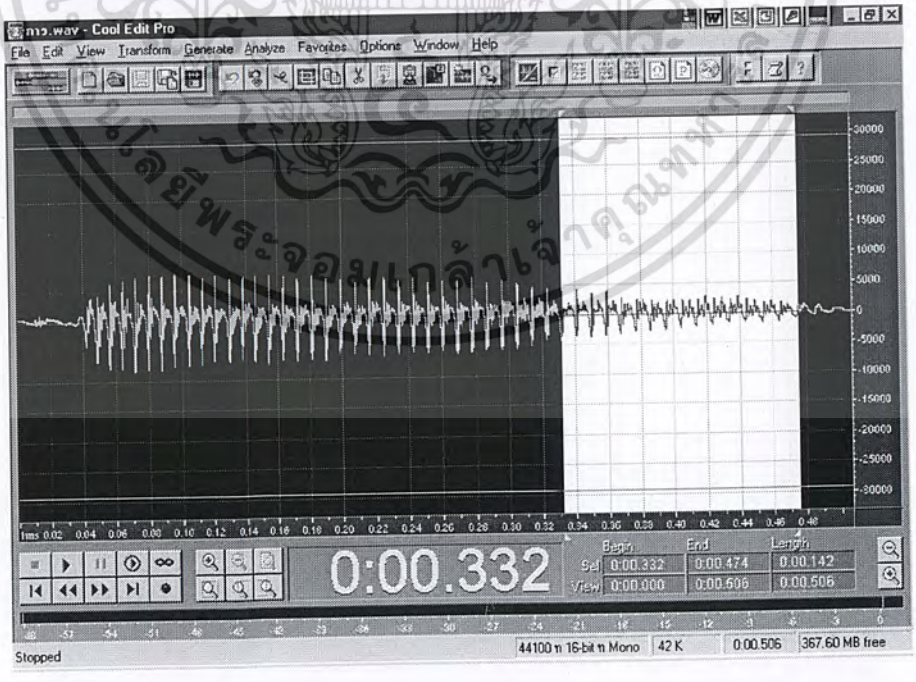
รูปที่ 6.6 เสียงของสระ /อา/ ที่ผันวรรณยุกต์ /ตรี/ จากคำว่า "ล้าน"

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3. เสียงของตัวสะกดที่ผันวรรณยุกต์จะอยู่ส่วนท้ายของคำ แบ่งเป็น
   - ตัวสะกดที่สังเกตเห็นได้ชัด คือแม่ /กง/ /กน/ และ /กม/ ดังตัวอย่างในรูปที่ 6.7
   - ตัวสะกดที่สังเกตได้ยาก คือแม่ /เกย/ /เกอว/ /กก/ /กด/ และ /กบ/ ดังตัวอย่างในรูปที่ 6.8



รูปที่ 6.7 เสียงของตัวสะกด /น/ ที่ผันวรรณยุกต์ /ตรี/ จากคำว่า "ล้าน"



รูปที่ 6.8 เสียงของตัวสะกด /ว/ ที่ผันวรรณยุกต์ /สามัญ/ จากคำว่า "กาว"

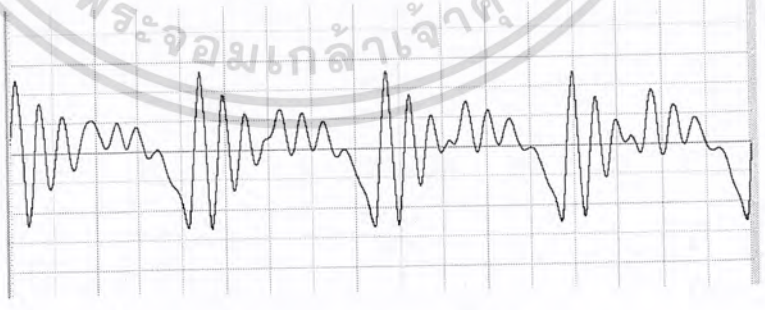
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

6.4 สรุปหลักในการบันทึกเสียง

1. เลือกอุปกรณ์(ไมโครโฟน) ที่มีคุณภาพดีที่สุด เพราะไมโครโฟนทุกตัวจะให้คุณภาพเสียงที่ดีได้ไม่เท่ากัน และเมื่อเสียบเข้าพอร์ตพร้อมที่จะทำงานแล้ว จะมีเสียงรบกวนในตัวของมันเองในระดับหนึ่ง เราจะต้องเลือกให้มีเสียงรบกวนน้อยที่สุด
2. ทำการบันทึกเสียงในห้องที่เงียบที่สุด อย่าให้มีเสียงรบกวนจากภายนอก
3. การบันทึกเสียงแต่ละครั้งต้องพยายามออกเสียงให้ชัดถ้อยชัดคำมากที่สุด อย่าให้เพี้ยน ทั้งเสียงพยัญชนะต้นโดยเฉพาะพยัญชนะควบกล้ำหรือเสียงสระที่ผันวรรณยุกต์แล้ว และโดยเฉพาะเสียงพยัญชนะสะกดที่ผันวรรณยุกต์มักจะมีการเพี้ยนได้ง่ายโดยไม่รู้ตัว เพราะเมื่อฟังเป็นคำดั้งเดิม(ที่บันทึกไว้) ก็จะฟังออกดี แต่พอแบ่งเป็นส่วนๆแล้วนำไปผสมกับส่วนอื่นที่มาจากคำอื่น จึงจะรู้ว่าเสียงที่ได้นั้นเพี้ยนออกไป เพราะภาษาไทยถ้าเสียงวรรณยุกต์เปลี่ยนไป ก็จะทำให้ความหมายของคำนั้นเปลี่ยนไปด้วย เช่น เรามีเสียงพยัญชนะต้น /ล/ มีเสียงสระ /า/ ที่ผันวรรณยุกต์ตรีและเราตัดเสียงพยัญชนะสะกด /น/ ที่ผันเสียงวรรณยุกต์ตรีมาจากคำว่า “ลิ้น” แล้วนำมาผสมกัน เสียงที่ควรจะได้คือ “ล้าน” แต่บางทีอาจจะออกมาเป็น “หลาน” ได้ ต้องระวังให้ดี
4. การบันทึกเสียงแต่ละครั้งต้องพยายามให้ระดับเสียงหรือความดังของเสียง(Amplitude) ใกล้เคียงกันมากที่สุด เพราะถ้าต่างกันมากเกินไปอาจทำให้เพี้ยนได้ ซึ่งตรงนี้อาจแก้ได้โดยการใช้คำสั่งในโปรแกรมที่ใช้บันทึกเสียงในการลดหรือเพิ่มแอมปลิจูดให้อยู่ในระดับที่ต้องการได้

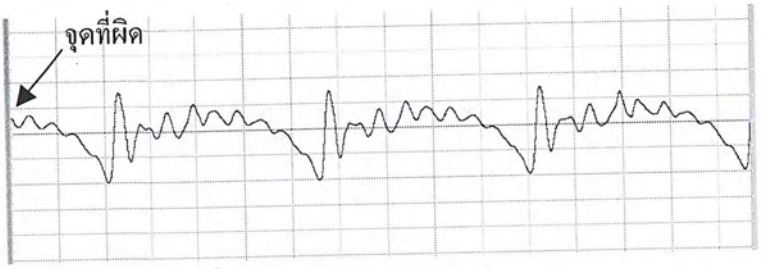
6.5 สรุปหลักในการแบ่งไฟล์เสียง

1. จะต้องแบ่งให้ตรงตามเสียงที่ต้องการ ไม่มีส่วนที่ไม่ต้องการติดมาด้วย เช่นถ้าต้องการเสียงพยัญชนะต้น /ล/ ก็อย่าให้มีเสียงสระหรือเสียงวรรณยุกต์ติดมาด้วย
2. เนื่องจากไฟล์เสียงที่เราบันทึกไว้จะมีลักษณะเป็นคลื่นรูปไซน์(sine wave) เวลาแบ่งจะต้องแบ่งตรงจุดที่ครบลูกคลื่นพอดี(crossing zero) ในทุกไฟล์ ดังรูปที่ 6.9 อย่าแบ่งระหว่างกลางลูกคลื่น ดังรูปที่ 6.10



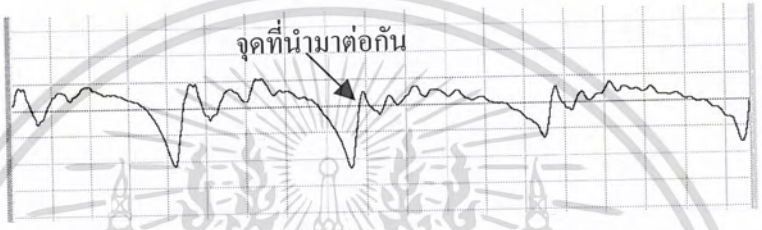
รูปที่ 6.9 การแบ่งไฟล์เสียงที่ถูกต้อง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

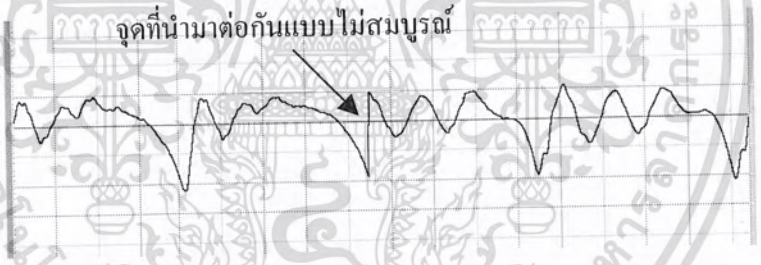


รูปที่ 6.10 การแบ่งไฟล์เสียงที่ผิด

เพราะว่าเวลาเราเรียกโปรแกรมให้ทำงาน โปรแกรมจะทำการนำไฟล์เหล่านี้มาต่อกัน ดังรูปที่ 6.11 ถ้าต่อกันไม่พอดี ดังรูปที่ 6.12 เสียงที่ได้ออกมาจะไม่เรียบ จนบางทีก็อาจทำให้ฟังไม่รู้เรื่องไปเลยก็ได้



รูปที่ 6.11 การต่อไฟล์เสียงที่ถูกต้องเข้ากันได้พอดี



รูปที่ 6.12 การต่อไฟล์เสียงที่ผิดเพราะการแบ่งไฟล์เสียงไม่ถูกต้อง

3. ในบางครั้งแม้ว่าเราจะแบ่งตรงจุดที่ครบลูกคลื่นพอดีแล้วก็ตาม เสียงที่ได้ก็อาจจะยังไม่ดีพอ เราจะต้องสังเกตให้ดี แล้วลองเลื่อนไปหรือเลื่อนกลับไปยังลูกคลื่นถัดไปแทน

มาถึงตอนนี้ เราก็ได้ไฟล์เสียงของโฟนีมต่างๆที่จะนำมาใช้ในการสังเคราะห์เสียงภาษาไทยได้ครอบคลุมทุกเสียงแล้ว ไฟล์เสียงเหล่านี้จะถูกเก็บไว้เป็นฐานข้อมูลในการให้โปรแกรมสังเคราะห์เสียงภาษาไทยนำมารวมกันให้เป็นพยางค์ คำและประโยคต่อไป ซึ่งในบทต่อไปเราจะอธิบายถึงการทำงานของตัวโปรแกรมว่ามีการทำงานอย่างไรบ้าง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 7
การออกแบบ และการสร้าง
โปรแกรมสังเคราะห์เสียงภาษาไทย

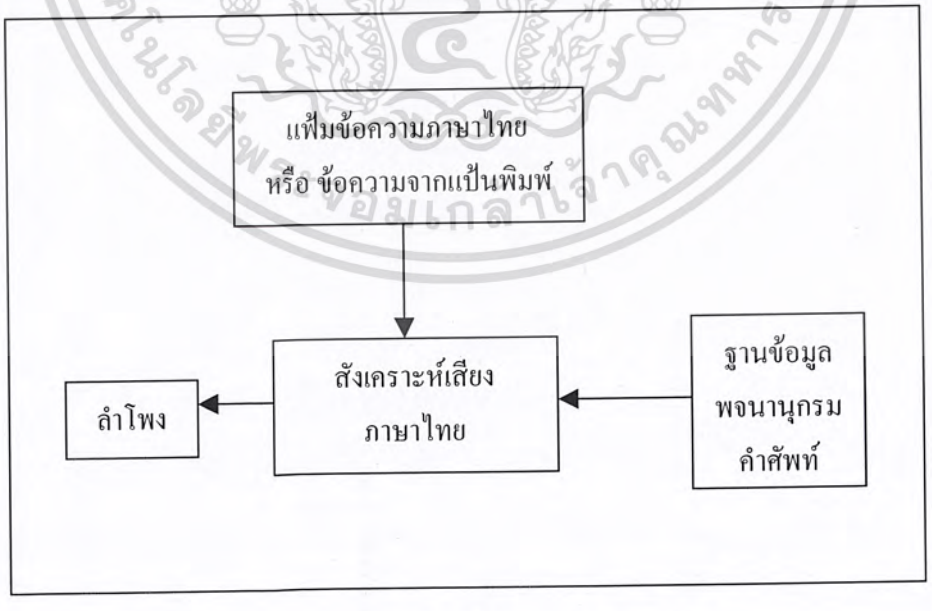
7.1 บทนำ

ในบทนี้ จะกล่าวถึงการออกแบบโปรแกรมสังเคราะห์เสียงภาษาไทย อันประกอบด้วยแนวคิดในการออกแบบ ส่วนประกอบต่าง ๆ ภายในโปรแกรม และ วิธีการทำงานของโปรแกรม

7.2 แนวคิดในการออกแบบโปรแกรม

โปรแกรมสังเคราะห์เสียงภาษาไทยที่จะสร้างขึ้น เป็นโปรแกรมที่รับไฟล์ข้อความภาษาไทย(Text File) หรือข้อความจากแป้นพิมพ์ จากนั้นโปรแกรมจะนำประโยคที่อยู่ในไฟล์มาตัดแบ่งออกเป็นคำ ๆ แล้วนำคำที่ตัดได้ไปวิเคราะห์ว่าคำ ๆ นั้นประกอบด้วยหน่วยเสียงย่อย ๆ อะไรบ้างโดยอ้างอิงจากไฟล์พจนานุกรมคำศัพท์ ต่อจากนั้นจึงนำไฟล์หน่วยเสียงที่บันทึกไว้มาประกอบกันให้ได้เป็นไฟล์เสียงของคำที่ตัดได้จากประโยคจนครบทั้งไฟล์ข้อความ แล้วจึงนำไฟล์เสียงที่สร้างขึ้นไปสร้างเสียงออกทางลำโพง

เนื่องจากในระหว่างที่มีการนำไฟล์หน่วยเสียงมาประกอบกันนั้น เราสามารถตัดคำไปในระว่างนั้นได้ ดังนั้นเพื่อการทำงานที่เร็วขึ้นของโปรแกรมจึงควรมีการแยกขั้นตอนการตัดคำกับการประกอบไฟล์หน่วยเสียงออกจากกันในรูปแบบของเธรด



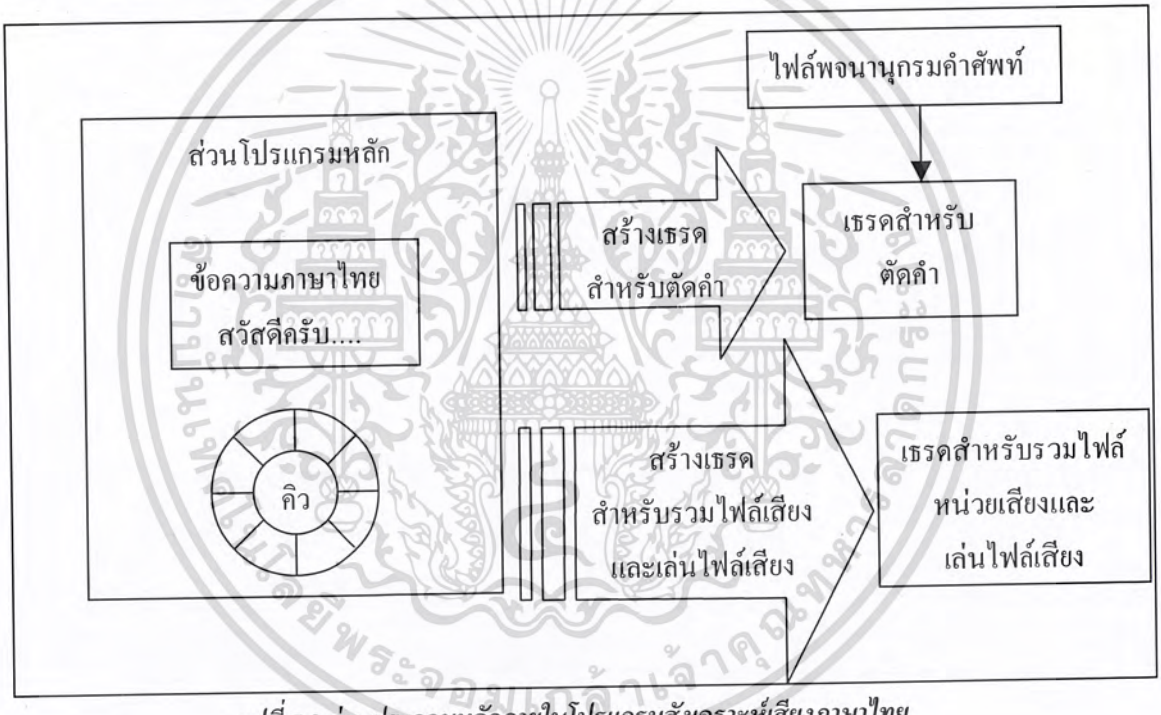
รูปที่ 7.1 ภาพรวมของโปรแกรมสังเคราะห์เสียงภาษาไทย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

7.3 ส่วนประกอบหลัก ภายในโปรแกรม

โปรแกรมสังเคราะห์เสียงภาษาไทยในปริญญานิพนธ์ฉบับนี้ประกอบด้วยส่วนต่าง ๆ ดังนี้คือ

- ส่วนโปรแกรมหลัก ส่วนนี้ทำหน้าที่ในการสร้างส่วนติดต่อผู้ใช้ซึ่งจะรับข้อความภาษาไทยจากผู้ใช้ รอรับการกดปุ่มคำสั่งและจัดการกับเหตุการณ์ (Event) ที่เกิดขึ้น
- ส่วนของการตัดคำ ส่วนนี้ทำหน้าที่ในการตัดคำในประโยคออกเป็นคำเดี่ยว ๆ โดยอาศัยอัลกอริทึม Word Bucket เข้ามาช่วยในการตัดคำ และนำคำที่ได้ไปหาหน่วยเสียงที่ประกอบเป็นคำ ๆ นั้นจากไฟล์พจนานุกรม แล้วส่งรายชื่อหน่วยเสียงที่ได้เข้าไปในคิว
- ส่วนของการรวมไฟล์หน่วยเสียงและเล่นไฟล์เสียง ส่วนนี้ทำหน้าที่รวมไฟล์หน่วยเสียงโดยนำรายชื่อไฟล์ของหน่วยเสียงมาจากคิว เมื่อรวมไฟล์หน่วยเสียงของคำได้จนครบทุกคำที่ตัดได้จากไฟล์ข้อความแล้วก็จะนำไฟล์เสียงที่ได้ไปสร้างเสียงออกทางลำโพง



รูปที่ 7.2 ส่วนประกอบหลักภายในโปรแกรมสังเคราะห์เสียงภาษาไทย

7.4 การสร้างโปรแกรม

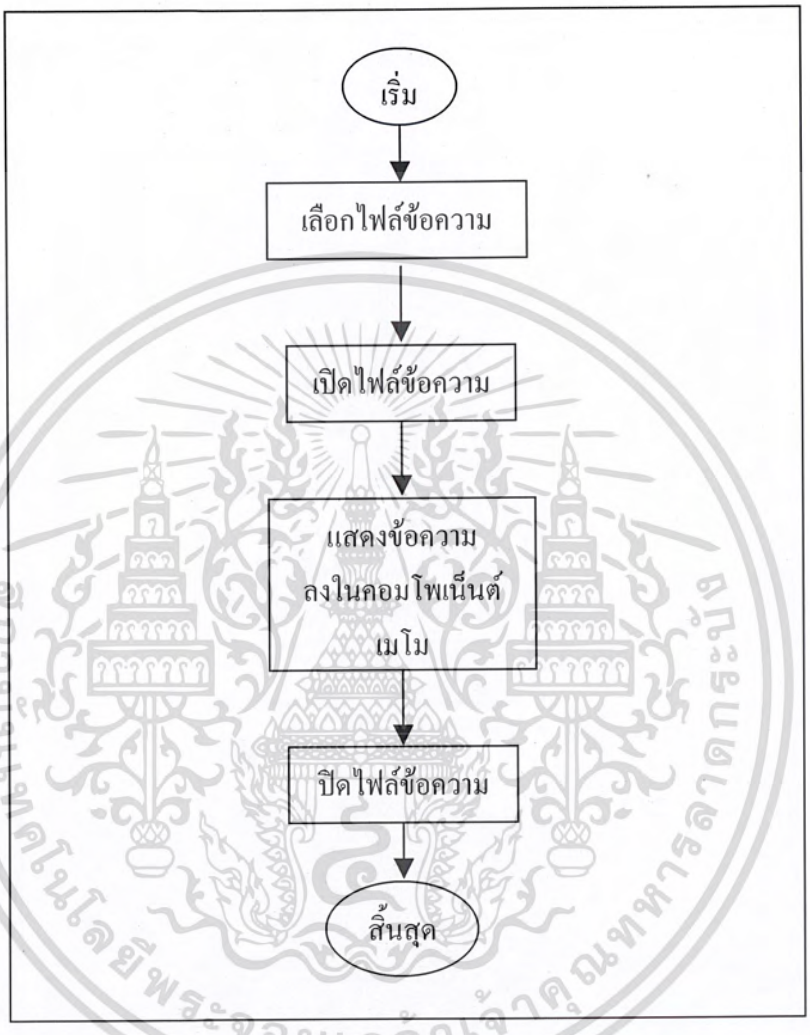
ในส่วนของการสร้างโปรแกรม จะขออธิบายตามส่วนประกอบของโปรแกรมดังนี้

1. ส่วนรับข้อความ
2. ส่วนตัดคำ
3. ส่วนรวมไฟล์และเล่นไฟล์เสียง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

7.4.1 ส่วนรับข้อความ

ส่วนรับข้อความเป็นส่วนที่รับข้อความภาษาไทยที่จะนำมาสังเคราะห์เสียง โดยสามารถรับได้สองทางคือ รับจากไฟล์ข้อความที่มีอยู่แล้วหรือผู้ใช้ป้อนเข้าทางแป้นพิมพ์โดยตรง



รูปที่ 7.3 แผนภูมิสายงานของ การรับข้อความจากไฟล์ข้อความ

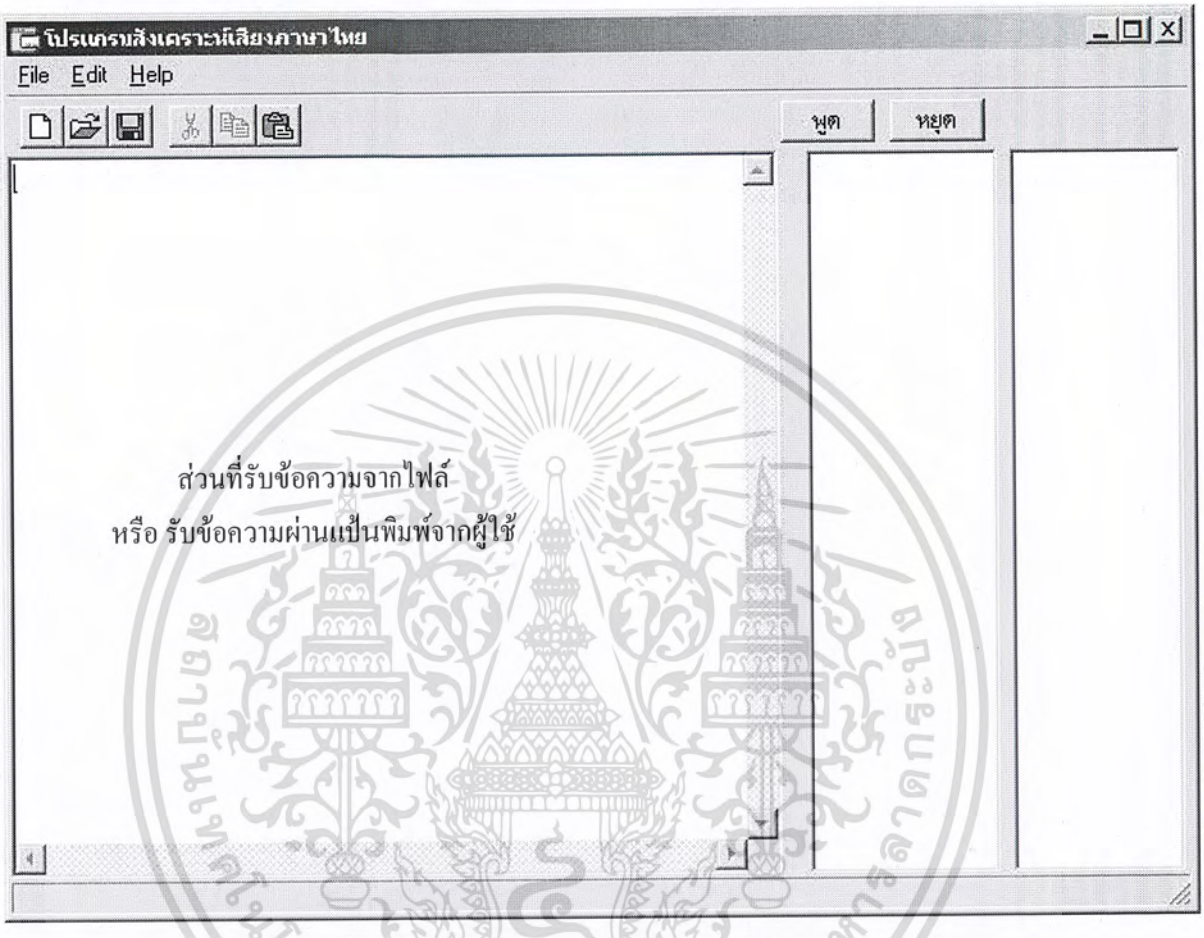
อธิบายการทำงานของการรับข้อความจากไฟล์ข้อความ

1. ให้ผู้ใช้เลือกไฟล์ข้อความ
2. เปิดไฟล์ข้อความที่ผู้ใช้เลือก
3. นำไฟล์ข้อความที่ผู้ใช้เลือกมาแสดงผลลงในคอมโพเน็นต์เมโม
4. ปิดไฟล์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

อธิบายการรับข้อความจากแป้นพิมพ์

ขั้นตอนนี้ผู้ใช้สามารถพิมพ์ข้อความภาษาไทยที่ต้องการลงบนคอมโพเน็นต์เมโมได้โดยตรง



รูปที่ 7.4 ตัวอย่างโปรแกรม ที่รับข้อความจากไฟล์ข้อความหรือจากผู้ใช้ผ่านแป้นพิมพ์

7.4.2 ส่วนตัดคำ

โพรเซสของการตัดคำจะมีลักษณะการทำงานเป็นเธรดซึ่งแยกออกจากส่วนของโปรแกรมหลักโดยจะถูกสร้างขึ้นเมื่อผู้ใช้สั่งให้ทำการสังเคราะห์เสียง และจะทำลายตัวเองเมื่อเสร็จสิ้นการประมวลผล

การทำงานของเธรดการตัดคำมีขั้นตอนดังนี้ (ดูรูป ที่ 7.5 ประกอบ)

1. อ่านข้อความบรรทัดแรกจากคอมโพเน็นต์เมโมมาเก็บไว้ในบัฟเฟอร์ที่ชื่อ tmpString กำหนดค่าให้กับตัวแปร workingPosition และ startPosition เท่ากับ 1 เพื่อชี้ไปยังตัวอักษรตัวแรกของข้อความในบัฟเฟอร์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2. นำตัวอักษรที่ชี้โดย workingPosition มาเก็บไว้ในตัวแปร wordBuff ซึ่งเป็นบัฟเฟอร์ที่ใช้เปรียบเทียบกับพจนานุกรมคำศัพท์ และเพิ่มค่า workingPosition อีกหนึ่ง เพื่อชี้ไปยังตัวอักษรตัวถัดไป
3. เปรียบเทียบกลุ่มของตัวอักษรที่อยู่ใน wordBuff กับพจนานุกรมคำศัพท์ ว่ามีคำที่อยู่ในพจนานุกรมที่ขึ้นต้นด้วยกลุ่มตัวอักษรที่อยู่ใน wordBuff หรือไม่ ถ้ามี แสดงว่ายังมีคำที่เป็นไปได้ในการตัดที่ยาวกว่านี้อยู่ให้ทำขั้นตอนต่อไป แต่ถ้าไม่มี แสดงว่าไม่มีคำที่เป็นไปได้ที่ยาวกว่านี้แล้วให้ข้ามไปขั้นตอนที่ 6
4. เปรียบเทียบกลุ่มของตัวอักษรที่อยู่ใน wordBuff กับพจนานุกรมคำศัพท์ว่ามีคำที่อยู่ในพจนานุกรมที่ตรงกับคำที่อยู่ใน wordBuff หรือไม่ ถ้ามี แสดงว่าพบคำที่สามารถตัดได้ในพจนานุกรมคำศัพท์ให้ทำขั้นตอนต่อไป ถ้าไม่มีกลับไปขั้นตอนที่ 2 เพื่อหาคำที่สามารถตัดได้ต่อไป
5. นำคำที่ตัดได้เก็บลงใน wordBucket ซึ่งเป็นบัฟเฟอร์ขนาด 5 ช่อง ซึ่งสามารถเก็บคำที่ตัดทั้งหมดไว้ได้ 5 คำ
6. เลือกคำใน wordBucket ที่ยาวที่สุด โดยปกติจะเป็นจะเป็นคำสุดท้ายที่ตัดได้ แต่บางครั้งคำที่ยาวที่สุดที่ตัดได้อาจทำให้คำต่อไปอ่านไม่ได้ ดังนั้นเราจึงต้องกลับมาเลือกคำที่ยาวที่สุดที่อยู่ก่อนหน้าที่ไม่ทำให้คำต่อไปอ่านไม่ได้
   ตัวอย่างเช่น คำว่า "หายาก"

   wordBucket (1) = หา bucketEndPtr (1) = 2

   wordBucket (2) = หาย bucketEndPtr (2) = 3

   ถ้าเราเลือกคำที่ยาวที่สุด คือ คำว่า "หาย" จะทำให้คำต่อไปเป็น "าก" ซึ่งไม่สามารถอ่านได้ เราจึงต้องย้อนกลับไป มาเลือกคำว่า "หา" แทน ซึ่งวิธีการตรวจสอบของเราจะใช้การตรวจสอบว่าอักษรตัวแรกของคำต่อไปขึ้นต้นด้วย "า" หรือ " ั " หรือไม่ ถ้าขึ้นต้นสระให้ย้อนกลับไปเลือกคำที่ยาวที่สุดคำก่อนหน้า โดยเขียนโปรแกรมในลักษณะของรีเคอร์ซีฟ
7. ตรวจสอบว่าคำที่เลือกเป็นคำที่เข้ามาใน wordBucket เป็นคำสุดท้ายหรือไม่ ถ้าใช่ ก็ไม่ต้องทำการเปลี่ยนแปลงค่าใน workingPosition เพราะจะชี้ไปยังตัวอักษรตัวแรกของคำถัดไปอยู่แล้วให้ไปขั้นตอนที่ 9 ได้เลย แต่ถ้าไม่ใช่ ให้ไปขั้นตอนต่อไป
8. ถ้าคำที่เลือกไม่ได้เป็นคำที่เข้ามาใน wordBucket เป็นคำสุดท้าย แสดงว่าหลังจากเลือกคำนี้ workingPosition ไม่ได้ชี้ไปยังตัวอักษรตัวแรกของคำถัดไปต้องย้าย workingPosition ไปยังตำแหน่งที่ถูกต้อง คือ

   workingPosition = bucketEndPtr (n) + 1

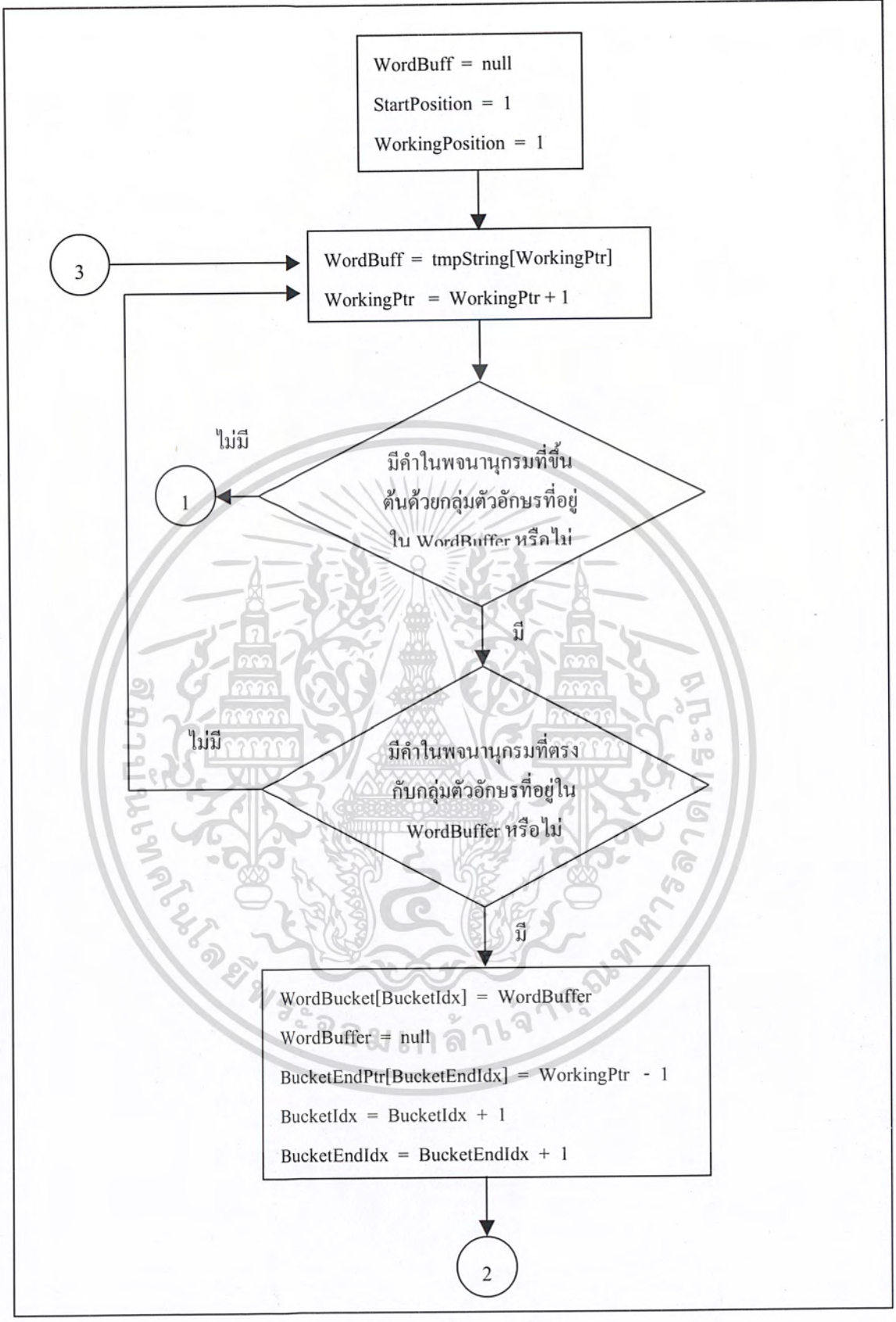
   โดย bucketEndPtr(n) เป็นตัวแปรที่เก็บตำแหน่งของตัวอักษรตัวสุดท้ายของคำที่ n ที่อยู่ใน wordBucket(n)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

9. ให้ StartPosition = workingPtr เพื่อให้ StartPosition ชี้ไปยังตัวอักษรตัวแรกของคำถัดไป
10. เมื่อได้คำที่ตัดได้แล้วก็นำไปเปรียบเทียบกับพจนานุกรมคำศัพท์ เพื่อหาว่าคำนั้นประกอบด้วยไฟล์หน่วยเสียงชื่ออะไรบ้าง เมื่อได้แล้วก็จะนำชื่อไฟล์เหล่านั้นใส่ลงในคิวเพื่อส่งให้เธรดรวมไฟล์และเล่นไฟล์เสียงประมวลผลต่อไป
11. ย้อนกลับไปทำขั้นตอนที่ 2 จนกว่าจะตัดคำได้ครบทุกบรรทัดของคอมโพเน็นต์เมโม เมื่อตัดคำครบทุกบรรทัดของเมโมแล้ว ก็ให้ใส่คำว่า "Stop" เข้าไปในคิวเพื่อเป็นการบอกเธรดรวมไฟล์และเล่นไฟล์เสียงว่า คำก่อนหน้านี้เป็นคำสุดท้ายที่จะนำไปสร้างเสียง เมื่อทำงานเสร็จเธรดนี้ก็จะทำลายตัวเอง

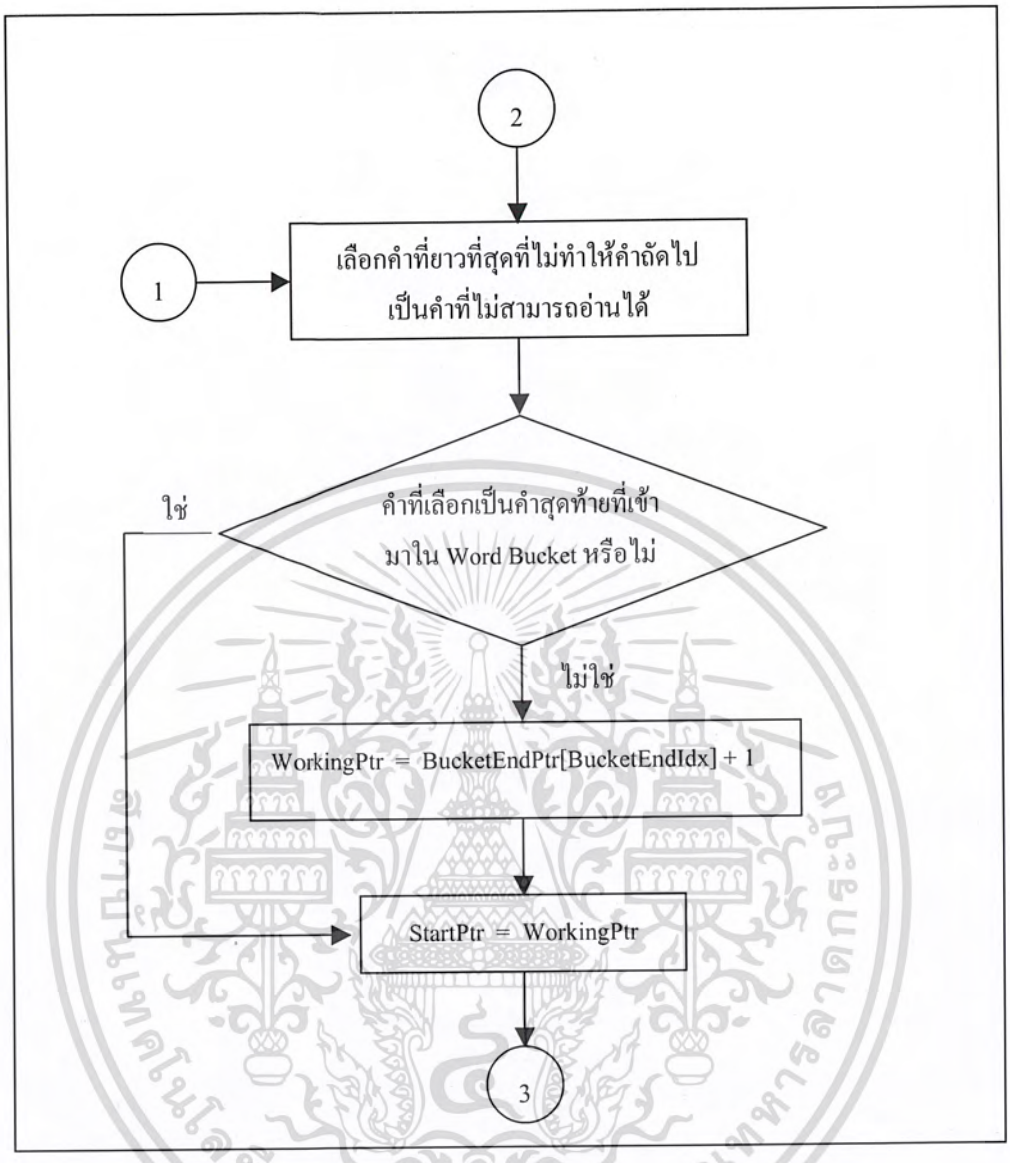
ข้อสังเกต จะเห็นได้ว่าการตัดคำที่ใช้ในปริญญานิพนธ์ฉบับนี้ ถึงแม้จะใช้อัลกอริทึม Word Bucket แต่ก็ไม่มีการตรวจสอบว่า WordBucket เต็มหรือไม่ ที่เป็นเช่นนี้เพราะคำศัพท์ในไฟล์พจนานุกรมคำศัพท์ได้ถูกกำหนดขึ้นด้วยคำที่มีขนาดเล็กที่สุด(มีตัวอักษรน้อยที่สุด) คำศัพท์ที่มีขนาดใหญ่(มีจำนวนตัวอักษรมาก) มักประกอบขึ้นจากคำศัพท์เล็ก ๆ หลายคำจึงมีโอกาสเป็นไปได้น้อยมากที่ Word Bucket จะเต็ม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 7.5 การทำงานของเธรดตัดคำ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 7.5 การทำงานของเธรดตัดคำ (ต่อ)

7.4.3 ส่วนรวมไฟล์และเล่นไฟล์เสียง

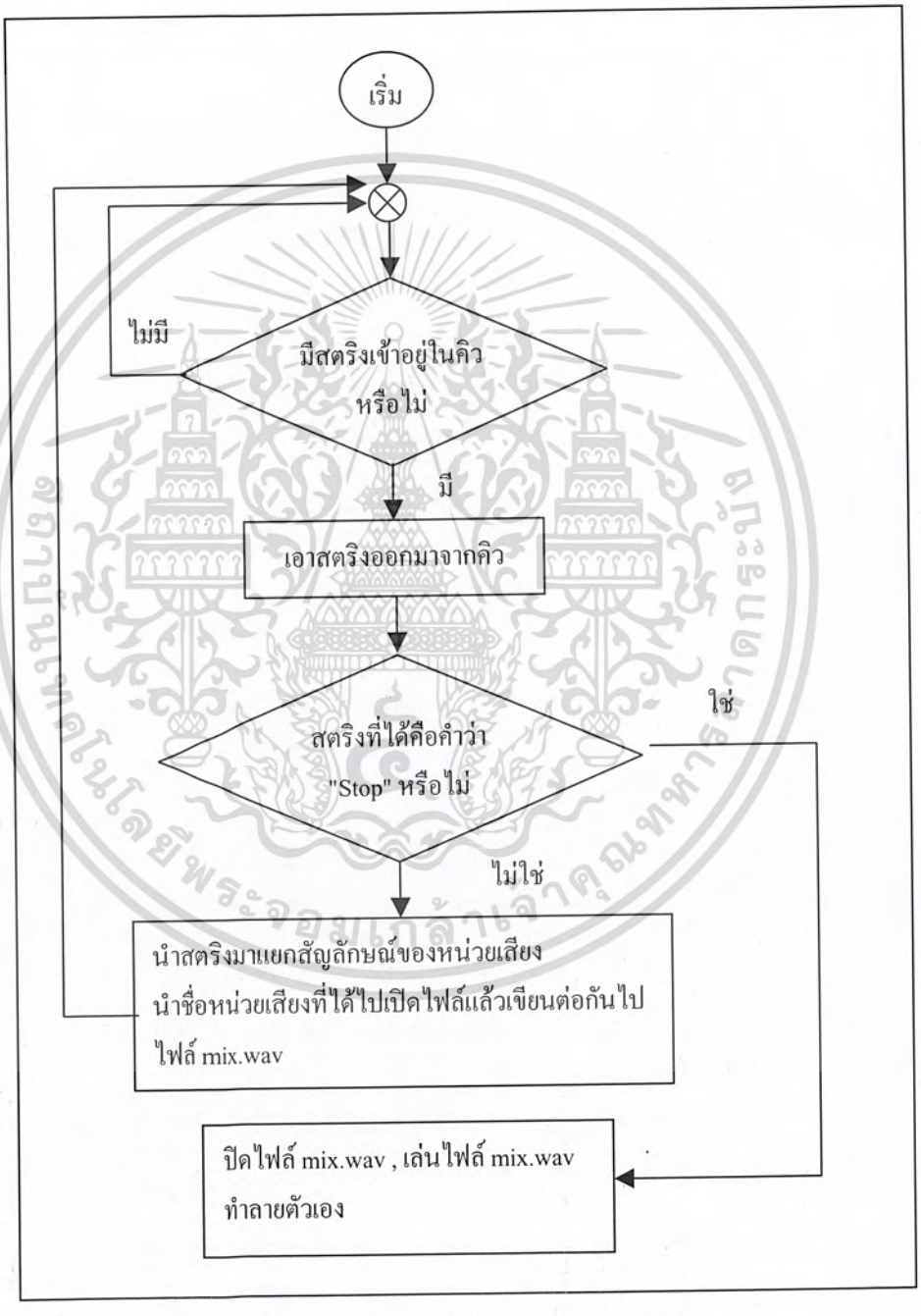
โพรเซสการรวมไฟล์เสียงและเล่นไฟล์เสียงจะมีลักษณะเหมือนกับโพรเซสของการตัดคำ คือมีลักษณะการทำงานเป็นเธรด ซึ่งจะถูกสร้างขึ้นเมื่อผู้ใช้สั่งให้สังเคราะห์เสียง และจะทำลายตัวเองเมื่อเสร็จสิ้นการประมวลผล

การทำงานของเธรดการรวมไฟล์และเล่นไฟล์เสียงมีดังนี้ (ดูรูปที่ 7.6 ประกอบ)

1. คอยสำรวจคิวจนกว่าจะมีสตริงชื่อไฟล์หน่วยเสียงเข้ามา แล้วจะนำค่านั้นเข้ามาตรวจสอบว่าเป็นคำว่า "Stop" หรือไม่ ถ้าใช่ก็จะข้ามไปขั้นตอนที่ 3 ถ้าไม่ใช่ก็ให้ทำขั้นตอน 2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2. นำสตริงที่ได้ไปแยกสัญลักษณ์หน่วยเสียงออกเป็นแต่ละหน่วยเสียงโดยแต่ละสัญลักษณ์จะเป็นชื่อของไฟล์เสียงที่เก็บไว้ เมื่อแยกได้สัญลักษณ์หน่วยเสียงหนึ่งก็จะนำไฟล์ที่มีชื่อตามสัญลักษณ์นั้นมาเขียนลงในไฟล์รวมที่ชื่อ mix.wav แล้ว กลับไปทำข้อ 1
3. ปิดไฟล์เสียง mix.wav แล้วเล่นไฟล์นี้โดยใช้คอมโพเน็นต์ มีเดียเพลเยอร์เพื่อให้สร้างเสียงออกมาทางลำโพง เมื่อทำงานเสร็จเธรดนี้ก็จะทำลายตัวเอง



รูปที่ 7.6 การทำงานของเธรดรวมไฟล์เสียงและเล่นไฟล์เสียง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

7.5 ภาพรวมการทำงานของโปรแกรม

เมื่อผู้ใช้รันโปรแกรมสังเคราะห์เสียงภาษาไทยก็จะพบกับหน้าจอดังรูปที่ 7.3 จากนั้นผู้ใช้ก็สามารถเปิดไฟล์ข้อความภาษาไทย หรือ ป้อนข้อความภาษาไทยลงในช่องว่างทางซ้ายมือ เมื่อข้อความพร้อมแล้ว ก็สามารถกดปุ่ม "พูด" เพื่อให้โปรแกรมทำงาน โดยมีรายละเอียดดังนี้

เมื่อปุ่ม "พูด" ถูกกด

- ส่วนโปรแกรมหลักจะสร้างเธรดสำหรับตัดคำและเธรดสำหรับรวมไฟล์หน่วยเสียงและเล่นไฟล์เสียงที่ได้รวมไว้
- เธรดที่ทำหน้าที่ตัดคำเริ่มตัดคำโดยอ่านข้อมูลมาจากคอมโพเน็นต์เมโมครั้งละหนึ่งบรรทัด จากนั้นก็ตัดคำโดยใช้อัลกอริทึม Word Bucket ตามที่ได้กล่าวไปแล้วในบทที่ 3 แต่จะมีการดัดแปลงบางส่วนคือเมื่อไม่สามารถหาคำที่ขึ้นต้นด้วยตัวอักษรตัวใดตัวหนึ่งได้ก็จะข้ามตัวอักษรตัวนั้น แล้วเริ่มตัดคำที่ตัวอักษรถัดไปและทำเช่นเดียวกันนี้ไปเรื่อย ๆ จนกว่าจะหมดทั้งบรรทัดแล้วจึงนำข้อความในบรรทัดถัดไปมาตัดคำจนครบทุกบรรทัดของเมโม เมื่อตัดคำได้หนึ่งคำก็จะแสดงผลคำคำนั้นในคอมโพเน็นต์เมโมที่อยู่ทางขวาแล้วจะนำคำนั้นไปเปรียบเทียบกับพจนานุกรมคำศัพท์อีกครั้ง เพื่อหาว่าคำนั้นประกอบด้วยไฟล์หน่วยเสียงชื่ออะไรบ้าง เมื่อได้แล้วก็จะนำชื่อไฟล์เหล่านั้นใส่ลงในคิวเพื่อส่งให้เธรดที่ทำหน้าที่รวมไฟล์หน่วยเสียงและเล่นไฟล์เสียง ถ้าหากคิวเต็ม เธรดสำหรับตัดคำก็จะต้องรอจนกว่าจะมีที่ว่างให้ใส่ชื่อไฟล์หน่วยเสียง(ซึ่งเก็บในรูปแบบสตริงโดยแต่ละชื่อไฟล์หน่วยเสียงจะมีช่องว่างหนึ่งช่องคั่นไว้) แล้วจึงจะไปตัดคำต่อไปได้ เมื่อตัดคำครบทุกบรรทัดของเมโมแล้ว เธรดสำหรับตัดคำก็จะใส่คำว่า "Stop" ลงในคิวเพื่อเป็นการบอกเธรดที่รวมไฟล์เสียงให้ทราบว่าได้ตัดคำครบทุกคำแล้ว การทำงานของเธรดก็จะสิ้นสุดลง
- เธรดที่ทำหน้าที่รวมไฟล์หน่วยเสียงและเล่นไฟล์เสียงจะคอยตรวจดูว่ามีสตริงชื่อไฟล์หน่วยเสียงของคำอยู่ในคิวบ้างหรือไม่ ถ้ามีก็จะนำสตริงชื่อไฟล์หน่วยเสียงมาแยกชื่อไฟล์ออกจากกันแล้วนำไฟล์หน่วยเสียงที่มีชื่อตามนั้นมาเขียนต่อกันไปในไฟล์ที่ชื่อ mix.wav เธรดนี้จะทำเช่นนี้เรื่อยไปจนกว่าจะเจอคำว่า "Stop" ในคิวก็จะปิดไฟล์เสียงที่เขียนไว้ แล้วนำไฟล์นั้นมาเล่นเพื่อให้สร้างเสียงออกมาทางลำโพง แล้วจึงสิ้นสุดการทำงานของเธรด

สรุป

โปรแกรมสังเคราะห์เสียงที่ได้สร้างขึ้นนั้นมีส่วนการทำงานหลัก 3 ส่วน ซึ่งทำงานในลักษณะเป็นเธรด โดยประกอบด้วยเธรดการทำงานหลัก เธรดสำหรับตัดคำ และ เธรดสำหรับรวมไฟล์เสียงและเล่นไฟล์เสียง โดยการตัดคำจะใช้อัลกอริทึม Word Bucket และใช้ไฟล์พจนานุกรมคำศัพท์เพื่อช่วยในการตัดคำและเก็บข้อมูลหน่วยเสียงของคำ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 8

ผลการทดลอง

ในบทนี้จะนำเสนอผลการทำงานของโปรแกรมและคุณภาพเสียงที่ได้

8.1 ไฟล์เสียงที่ได้จากการทำงานของโปรแกรม

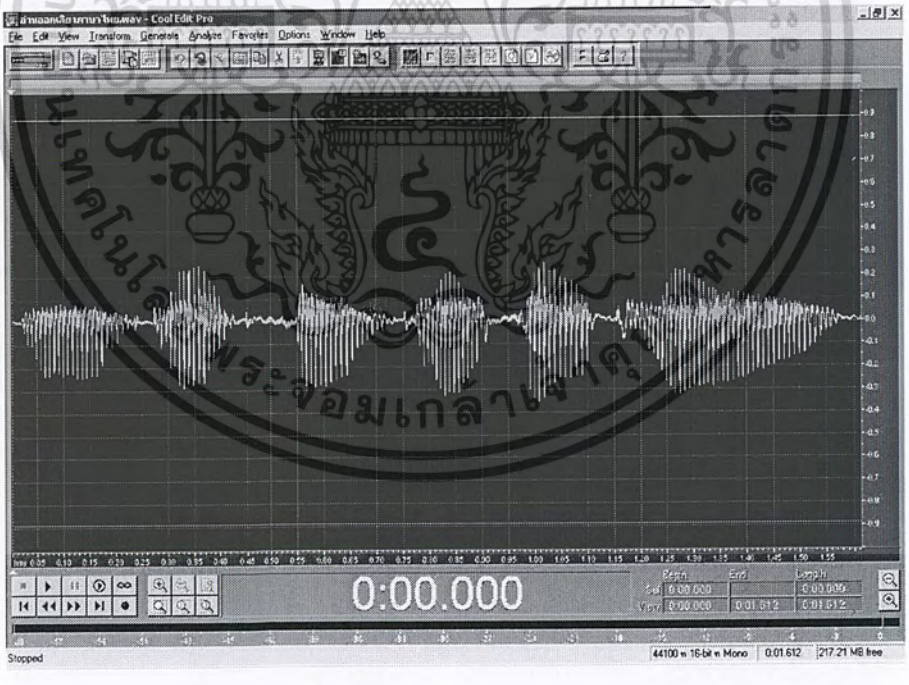
1. เสียงที่ได้จากโปรแกรมสังเคราะห์เสียงมีคุณภาพใกล้เคียงกับเสียงที่ได้จากการบันทึกเสียงพูดจริง โดยอาจจะมีความเพี้ยนบ้างในบางคำ
2. เมื่อดูลักษณะรูปคลื่นในเชิงเวลา(Time Domain) ปรากฏว่ารูปคลื่นที่เกิดจากการรวมไฟล์หน่วยเสียงมีลักษณะคล้ายกับลักษณะรูปคลื่นของการบันทึกเสียงพูดจริง
3. เสียงที่เกิดจากการรวมไฟล์หน่วยเสียงเมื่อรวมกันเป็นประโยคจะไม่เป็นธรรมชาติ เนื่องจากขาดความต่อเนื่องระหว่างคำ เพราะในการพูดจริงช่องว่างระหว่างคำจะไม่แน่นอน แต่เสียงจากโปรแกรมสังเคราะห์เสียงที่ได้จะมีช่องว่างระหว่างหน่วยเสียงเป็นเสียงเงียบที่มีความยาวแน่นอน คือ 0.15 วินาที อีกประการหนึ่งคือเสียงที่ได้จากการสังเคราะห์ขาดการทอดเสียงท้ายประโยค ทำให้ฟังดูไม่เป็นธรรมชาติ

เมื่อรันโปรแกรมและป้อนอินพุตตัวอย่างให้กับโปรแกรม

1. ข้อความภาษาไทยว่า "อ่านออกเสียงภาษาไทย" จะได้ผลลัพธ์ดังรูปที่ 8.1 โดยเสียงที่บันทึกจริงจะเป็นดังรูปที่ 8.2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

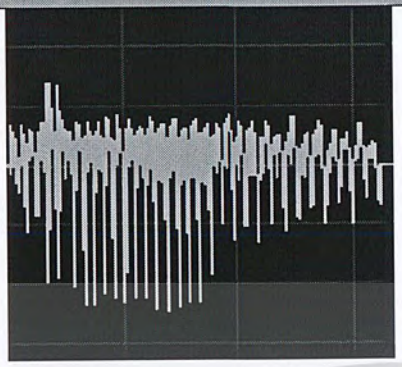
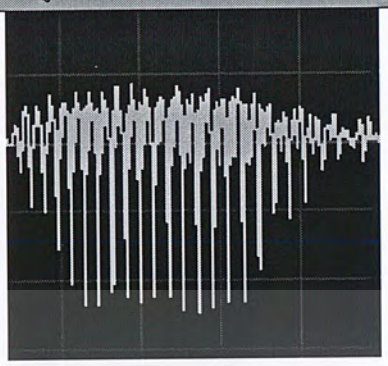
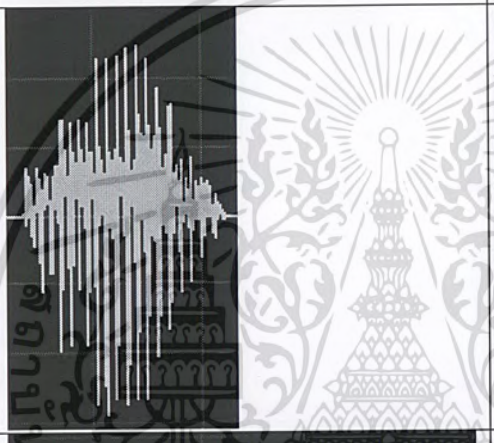
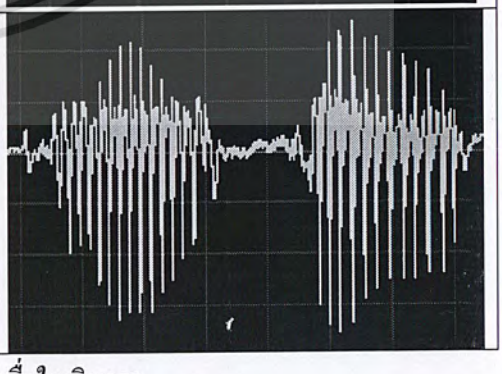
รูปที่ 8.1 รูปคลื่นของคำว่า "อ่านออกเสียงภาษาไทย" ที่ได้จากการรวมไฟล์เสียงที่เตรียมไว้



รูปที่ 8.2 รูปคลื่นของคำว่า "อ่านออกเสียงภาษาไทย" ที่ได้จากการบันทึกเสียงพูดจริง

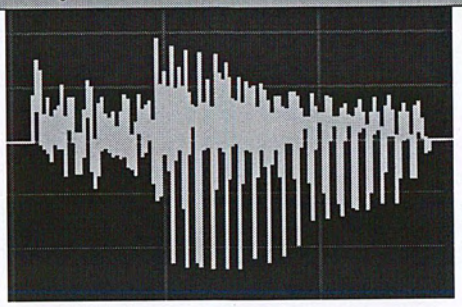
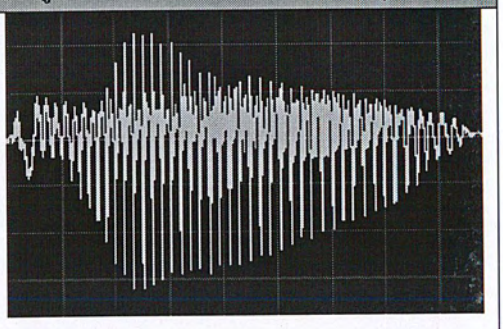
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

8.2 การเปรียบเทียบรูปคลื่นเสียงในเชิงเวลา

คำ	รูปคลื่นเสียงที่ได้จากการสังเคราะห์	รูปคลื่นที่ได้จากการบันทึกเสียงพูดจริง
อ่าน		
ออก		
เสียง		
ภาษา		

ตารางที่ 8-1 การเปรียบเทียบรูปคลื่นในเชิงเวลา

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

คำ	รูปคลื่นเสียงที่ได้จากการสังเคราะห์	รูปคลื่นที่ได้จากการบันทึกเสียงพูดจริง
ไทย		

ตารางที่ 8-1 การเปรียบเทียบรูปคลื่นในเชิงเวลา (ต่อ)

จากตารางการเปรียบเทียบรูปคลื่นในเชิงเวลาของคำจะสังเกตได้ว่าลักษณะของรูปคลื่นที่ได้จากการสังเคราะห์มีความคล้ายกันกับลักษณะของรูปคลื่นที่ได้จากการบันทึกเสียงพูดจริง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 9
บทวิจารณ์และสรุป

9.1 บทวิจารณ์

จากการทดลองใช้งานทำให้เราได้พบข้อจำกัดต่างๆของโปรแกรมดังต่อไปนี้

1. แฟ้มที่จะนำมาทำการอ่านจะต้องมีแต่ตัวหนังสือเท่านั้น ห้ามมีรูปปนอยู่ด้วย เพราะโปรแกรมไม่สามารถวิเคราะห์ได้
2. ไม่สามารถกลับไปอ่านคำก่อนหน้าได้ถ้าฟังไม่ทัน นอกจากจะสั่งให้หยุดโปรแกรมแล้วเริ่มต้นอ่านใหม่ตั้งแต่แรก
3. ไม่สามารถอ่านคำที่ไม่มีอยู่ในพจนานุกรมได้ เพราะการตัดคำของเราต้องทำการเปรียบเทียบกับคำศัพท์ที่มีอยู่ในพจนานุกรมคำศัพท์ แต่สามารถแก้ไขได้โดยการเพิ่มคำศัพท์นั้นลงไปในพจนานุกรมคำศัพท์
4. ไม่สามารถอ่านตัวเลขได้แบบเป็นธรรมชาติ เช่น "ฉันมีเงิน 234 บาท" ควรจะอ่านว่า "ฉันมีเงิน*สองร้อยสามสิบสี่*บาท" แต่โปรแกรมจะอ่านว่า "ฉันมีเงิน*สองสามสี่*บาท"
5. โปรแกรมจะอ่านสัญลักษณ์พิเศษนอกเหนือจากตัวอักษรปกติได้เพียงบางตัวคือ "(" ")" และ "." นอกนั้นจะอ่านไม่ได้
6. จากการทดลองฟังเสียงที่ได้ทำการสังเคราะห์ขึ้นมา ได้พบว่าเสียงที่ได้จากการผสมหน่วยเสียงที่มีเสียงพยัญชนะสะกดในแม่ /กก/ /กด/ และ /กบ/ จะไม่ค่อยดีเท่าใดนัก เนื่องจากตัวสะกดในมาตราเหล่านี้ จะเป็นการกักเสียง(หยุดการออกเสียง) ที่เปล่งออกมา ณ ฐานเสียงที่ต่างกัน จึงไม่สามารถที่จะจำลองเสียงให้เหมือนจริงได้ นั่นคือเสียงตัวสะกดเหล่านี้น่าจะมีส่วนขึ้นกับเสียงสระที่มาก่อนหน้าด้วย ดังนั้นการที่เราตัดเสียงตัวสะกดเหล่านี้มาโดยไม่คำนึงถึงเสียงสระที่มาก่อนหน้า ทำให้ได้เสียงที่ไม่มีคุณภาพเพียงพอ
7. จากการทดลองฟังเสียงที่ได้ทำการสังเคราะห์ขึ้นมา ได้พบว่าบางคำจะออกเสียงเพี้ยนไปในส่วนของเสียงตัวสะกด โดยเฉพาะตัวสะกดในแม่ /กง/ /กน/ และ /กม/ จะมีเสียงที่คล้ายกันมาก เช่นคำว่า "ผม" อาจจะได้ยินว่า "ผล" เนื่องจากในการตัดเอาตัวสะกดมานั้น เราจะตัดมาจากคำที่ผสมกับสระ "อา" และได้นำมาทดสอบกับเพียงบางสระเท่านั้น เมื่อนำไปผสมกับสระอื่นที่ไม่ได้นำมาทดสอบจึงอาจจะทำให้ได้เสียงที่เพี้ยนไป
8. ในการนำไฟล์เสียงของแต่ละคำมาต่อกันนั้น เราจะทำการแทรกเสียงเงียบ(silence) เข้าไปเป็นช่วงเวลาที่แน่นอน และจากการทดลองได้พบว่าเราควรจะทำการแทรกเสียงเงียบเข้าไประหว่างประโยคด้วยช่วงเวลาที่มากกว่าที่แทรกระหว่างคำ เช่น "สวัสดีครับ ผมคือโปรแกรมสังเคราะห์เสียงภาษาไทยครับ" ระหว่างคำว่า "สวัสดี" กับ "ครับ" จะมีช่วงเงียบ 0.15 วินาที และก่อนจะขึ้นคำว่า "ผม" ก็อาจจะมีช่วงเงียบ 0.3 วินาที นอกจากนั้นควรจะแทรกเสียงเงียบต่อจากพยางค์ที่มีตัว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สะกดในแม่ /กก/ /กด/ /กบ/ ด้วย เพราะพยางค์เหล่านี้จะเป็นการหยุดเสียงทันที ก่อนที่จะอ่านพยางค์ต่อไปจึงควรจะมีช่วงเงียบก่อน เพื่อให้มีความเป็นธรรมชาติมากขึ้น

9. การแบ่งคำที่ต้องการสระเสียงยาว(อา อี อือ อู เอ แอ โอ ออ เออ เอีย เอือ อัว) ที่ผันตามวรรณยุกต์จัตวาควรจะตัดมาจากคำที่ไม่มีตัวสะกด เช่นคำว่า “เสีย” หรือ “เหลือ” เนื่องจากถ้าตัดมาจากคำที่มีตัวสะกด เช่นคำว่า “เสียง” หรือ “เหลือง” เสียงของวรรณยุกต์จัตวาจะไปอยู่ที่ตัวสะกดแทน ทำให้เสียงที่ตัดออกมาได้จะค่อนข้างออกไปทางเสียงวรรณยุกต์เอกแทน โดยในลักษณะนี้จะไม่เกิดขึ้นกับสระเสียงสั้น เนื่องจากสระเสียงสั้น(อะ อิ อึ อุ เอะ แอะ โอะ เอาะ เออะ เอียะ เอือะ อัวะ) จะไม่สามารถผันวรรณยุกต์จัตวาได้โดยไม่มีตัวสะกด เช่นคำว่า “หลุม” “เห็น” หรือ “ขิง” ดังนั้นจึงต้องตัดมาจากคำที่มีตัวสะกดแน่นอน
10. ไม่สามารถอ่านคำกำกวมให้ถูกต้องตามสถานการณ์ได้ เช่น “ตากลม” อ่านได้ 2 วิธีคือ “ตา+กลม” หรือ “ตาก+ลม” ในปริญญานิพนธ์นี้จะสามารถอ่านได้เพียงวิธีเดียวคือ “ตาก+ลม” เพราะว่าการตัดคำของเราจะใช้วิธีเลือกคำที่ยาวที่สุดที่มีอยู่ในพจนานุกรมคำศัพท์ คำว่า “ตา” และ “ตาก” ต่างก็เป็นคำที่มีอยู่ในพจนานุกรมคำศัพท์ แต่คำว่า “ตาก” เป็นคำที่ยาวที่สุด โปรแกรมจึงอ่านได้ว่า “ตาก+ลม” แต่เราอาจสามารถแก้ไขได้โดยใช้วิธีการวิเคราะห์ประโยคเข้ามาช่วย เพื่อดูว่าควรจะเลือกตัดคำที่ตำแหน่งใด ซึ่งเป็นวิธีทางด้านปัญญาประดิษฐ์
11. ความต่อเนื่องในการอ่านข้อความของโปรแกรมยังไม่ดีเท่าที่ควร เนื่องจากโปรแกรมต้องมีการเปรียบเทียบคำศัพท์ในเนื้อหาที่ต้องการอ่านกับพจนานุกรมคำศัพท์ ซึ่งถูกเก็บอยู่ในหน่วยจัดเก็บข้อมูลซึ่งทำงานได้ช้า(หน่วยความจำสำรอง : Secondary Storage) บ่อยครั้ง เพื่อใช้ในการตัดคำ และยังต้องเสียเวลาในการรวมไฟล์ของหน่วยเสียงย่อยๆ ให้เป็นหน่วยพยางค์เพื่อที่จะให้อ่านออกเสียงมาอีก แต่ปัญหาตรงนี้เราสามารถแก้ได้โดยการให้โปรแกรมทำการตัดคำและรวมไฟล์ของหน่วยเสียงย่อยๆจนได้ครบทั้งแฟ้มข้อมูลที่ต้องการอ่านก่อน แล้วจึงสั่งให้โปรแกรมทำการอ่านออกเสียงมาทีเดียวทั้งแฟ้มข้อมูลเลย การแก้ปัญหาด้วยวิธีนี้อาจจะทำให้เกิดปัญหาในข้อต่อไปตามมา
12. จะต้องมีพื้นที่ในหน่วยจัดเก็บข้อมูลเพียงพอที่จะเก็บแฟ้มเสียงที่ได้สร้างขึ้นมาจากแฟ้มข้อมูล ซึ่งก็ขึ้นอยู่กับความยาวของแฟ้มข้อมูลที่นำมาอ่าน แต่เนื่องจากในปัจจุบันนี้หน่วยจัดเก็บข้อมูลได้ถูกพัฒนาจนมีความจุที่สูงมาก ปัญหานี้จึงไม่น่าเป็นห่วงเท่าใดนัก
13. เมื่อได้มีการอ่านแฟ้มข้อมูลหนึ่งไปแล้ว แล้วนำแฟ้มข้อมูลใหม่มาอ่าน ไฟล์เสียงที่สร้างขึ้นจากแฟ้มข้อมูลเก่าจะถูกเขียนทับด้วยไฟล์เสียงที่สร้างขึ้นจากแฟ้มข้อมูลใหม่ ทำให้ไม่สามารถอ่านแฟ้มข้อมูลเก่าได้ในทันทีเมื่อต้องการนำกลับมาอ่านใหม่ ต้องเสียเวลาในการตัดคำและรวมไฟล์เสียงขึ้นใหม่ ทางแก้ทางหนึ่งคือ เมื่อใดก็ตามที่ต้องการอ่านแฟ้มข้อมูลอันใหม่ ก็ให้นำไฟล์เสียงอันเก่าย้ายไปเก็บในโฟลเดอร์(folder) อื่นที่ไม่ใช่ที่โปรแกรมกำหนดเอาไว้ก่อน หรืออีกวิธีหนึ่งก็คือทุกครั้งที่มีการอ่านแฟ้มข้อมูลอันใหม่ ก็ให้โปรแกรมสร้างไฟล์เสียงในชื่ออื่นเปลี่ยนไปทุกครั้ง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

9.2 สรุป

โปรแกรมสังเคราะห์เสียงภาษาไทยด้วยคอมพิวเตอร์เป็นโปรแกรมที่มีประโยชน์ต่อผู้ที่พิการทางสายตา ทำให้สามารถรับฟังเสียงภาษาไทยที่ได้จากการสังเคราะห์เสียงจากข้อความภาษาไทยหรือสำหรับผู้พิการทางการพูดให้สามารถใช้คอมพิวเตอร์พูดแทนตนเอง ซึ่งมีการทำงานดังต่อไปนี้

1. รับข้อความที่ต้องการสังเคราะห์เสียงจากแฟ้มข้อความภาษาไทย(Text file) เพื่อนำข้อความเหล่านั้นมาสังเคราะห์เสียงออกทางลำโพง นอกจากสามารถรับข้อความจากแฟ้มข้อความแล้ว ยังสามารถรับข้อความจากทางแป้นพิมพ์ได้อีกด้วย
2. เมื่อได้ข้อความมาแล้ว เราก็นำข้อความนั้นมาทำการแยกออกเป็นคำๆ โดยเลือกคำที่ยาวที่สุดเท่าที่จะเป็นไปได้ โดยการเปรียบเทียบกับพจนานุกรมคำศัพท์
3. เมื่อตัดคำเสร็จเรียบร้อยแล้ว ก็จะนำคำนั้นมาทำการหาคำอ่านโดยการค้นหาคำอ่านของแต่ละคำได้จากพจนานุกรมคำศัพท์
4. เมื่อได้คำอ่านแล้ว ก็จะนำคำอ่านนั้นมาหาสัญลักษณ์แทนเสียง
5. หลังจากได้สัญลักษณ์แทนเสียงแล้ว ก็นำแฟ้มเสียงของแต่ละเสียงที่แทนด้วยสัญลักษณ์แทนเสียงนั้นมารวมกันเป็นแฟ้มเสียงทั้งหมด
6. ขั้นตอนสุดท้ายคือการนำแฟ้มเสียงไปแปลงเป็นเสียงส่งออกทางลำโพงให้เราได้ยินต่อไป

9.3 แนวทางการพัฒนาต่อไป

1. การแบ่งไฟล์เสียงอาจจะทำในอีกลักษณะหนึ่ง ซึ่งจะแตกต่างไปจากที่ได้ใช้ในการทำงานครั้งนี้ คืออาจจะทำการแบ่งไฟล์เสียงในแบบที่ 2 หรือ 3 ที่ได้อธิบายไว้ในบทที่ 4 หรือในแบบอื่นๆก็ได้
2. ทำให้โปรแกรมสามารถเพิ่มคำศัพท์ที่ยังไม่มีในพจนานุกรมได้ด้วย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก

การสร้าง sp จาก sil

~h "sil"

<BEGINHMM>

<NUMSTATES> 5

<STATE> 2

<MEAN> 39

-4.481729e+000 -4.988771e+000 4.725426e-001 8.716468e-001 3.114777e+000 -1.628793e+000 2.656590e+000 -4.726635e+000 1.540224e+000 -2.574618e+000 1.870777e+000 -4.013587e+000 6.441782e+001 1.131465e+000 -2.802686e-001 -1.006600e-002 -1.584653e-001 9.889801e-001 -7.050928e-001 2.172060e-001 -7.270316e-001 5.650804e-001 -1.428351e-001 -2.753033e-001 1.431123e-001 -9.932485e-001 -3.991371e-001 8.120268e-002 -3.916197e-002 3.268724e-002 -3.687306e-001 2.379665e-001 -1.104178e-001 3.084177e-001 -1.667813e-001 1.642951e-001 1.767361e-001 4.958390e-003 2.520020e-001

<VARIANCE> 39

2.026278e+001 3.895103e+000 4.962233e+000 5.778740e+000 2.056620e+001 2.135576e+001 1.347709e+001 2.386639e+001 1.798274e+001 1.349549e+001 1.389248e+001 1.268451e+001 9.963229e+000 3.113118e+000 3.333883e-001 5.131708e-001 5.807106e-001 2.678430e+000 2.913157e+000 1.790531e+000 2.778957e+000 1.907376e+000 1.685013e+000 1.523898e+000 1.437356e+000 1.576283e+000 3.035933e-001 6.045505e-002 9.720118e-002 1.170082e-001 3.514265e-001 4.318790e-001 3.348437e-001 3.445941e-001 2.920358e-001 2.837451e-001 2.704780e-001 2.684844e-001 4.456758e-001

<GCONST> 9.067432e+001

<STATE> 3

<MEAN> 39

-1.882606e+000 -3.554463e+000 1.659426e-001 5.695080e-001 2.726086e+000 -1.685616e+000 1.684329e+000 -3.286708e+000 1.424847e+000 -2.085776e+000 1.097108e+000 -2.247970e+000 4.162634e+001 2.410250e-002 6.801444e-002 -9.889011e-003 -3.758664e-003 -3.466492e-002 3.746291e-002 -3.793253e-002 5.766057e-002 -3.662661e-002 1.433644e-002 3.746801e-004 5.485542e-002 -6.098068e-001 -8.335336e-003 9.690197e-003 2.238820e-003 -1.132704e-003 -1.335554e-002 -8.233030e-004 -4.228565e-003 -8.020276e-003 -1.142018e-002 -9.332317e-005 -5.762242e-003 -4.305856e-004 -9.150078e-003

<VARIANCE> 39

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.892350e+000 8.365601e+000 3.846982e+000 4.211987e+000 1.076755e+001 9.036107e+000

9.848803e+000 1.420859e+001 1.078936e+001 1.098874e+001 8.808033e+000 1.003700e+001

8.887128e+002 1.797630e-001 2.355286e-001 3.505588e-001 3.882942e-001 6.302609e-001

7.545438e-001 8.286784e-001 8.659801e-001 1.018844e+000 8.986608e-001 9.457800e-001

8.353303e-001 8.880465e+000 4.273483e-002 4.430741e-002 7.269581e-002 7.994711e-002

1.251430e-001 1.585774e-001 1.765345e-001 1.827718e-001 2.121781e-001 1.873781e-001

1.926342e-001 1.740429e-001 1.413381e+000

<GCONST> 7.409518e+001

<STATE> 4

<MEAN> 39

-2.438808e+000 -1.485552e+000 6.098307e-001 2.719999e-001 1.198622e+000 -5.363513e-001

7.653816e-001 -2.435678e+000 3.708174e-001 -5.364203e-001 6.019283e-001 -5.061205e-001

2.570110e+001 -7.064192e-001 2.877237e-001 6.407951e-001 8.524866e-003 -1.423467e-001

8.344807e-002 1.514168e-001 -5.517167e-001 -2.022321e-001 4.125653e-001 4.485451e-002

5.476460e-001 1.279666e+000 -3.027843e-001 -1.428554e-002 2.536712e-001 3.412914e-002

7.495335e-002 -6.131095e-002 1.585990e-001 -3.469583e-001 -1.537429e-002 -4.782392e-003

4.836437e-002 6.261525e-002 1.918460e+000

<VARIANCE> 39

1.702749e+001 6.295977e+000 1.049955e+001 3.153865e+000 8.278751e+000 5.522054e+000

6.572440e+000 2.097840e+001 4.353487e+000 6.020226e+000 4.650790e+000 6.604639e+000

9.803580e+002 3.245868e+000 4.304452e-001 1.813011e+000 5.282696e-001 9.266866e-001

8.967347e-001 1.060456e+000 3.945073e+000 5.686834e-001 7.490652e-001 6.957111e-001

8.726550e-001 1.351110e+002 3.563516e-001 7.883416e-002 1.750682e-001 8.108416e-002

1.578618e-001 1.473610e-001 1.295619e-001 3.300810e-001 9.694073e-002 1.083466e-001

9.238505e-002 1.593818e-001 1.180044e+001

<GCONST> 8.622438e+001

<TRANSP> 5

0.000000e+000 1.000000e+000 0.000000e+000 0.000000e+000 0.000000e+000

0.000000e+000 7.665885e-001 2.334115e-001 0.000000e+000 0.000000e+000

0.000000e+000 0.000000e+000 9.617694e-001 3.823058e-002 0.000000e+000

0.000000e+000 0.000000e+000 0.000000e+000 7.006519e-001 2.993480e-001

0.000000e+000 0.000000e+000 0.000000e+000 0.000000e+000 0.000000e+000

<ENDHMM>

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

~h "sp"

<BEGINHMM>

<NUMSTATES> 3

<STATE> 2

<MEAN> 39

-1.882606e+000 -3.554463e+000 1.659426e-001 5.695080e-001 2.726086e+000 -1.685616e+000

1.684329e+000 -3.286708e+000 1.424847e+000 -2.085776e+000 1.097108e+000 -2.247970e+000

4.162634e+001 2.410250e-002 6.801444e-002 -9.889011e-003 -3.758664e-003 -3.466492e-002

3.746291e-002 -3.793253e-002 5.766057e-002 -3.662661e-002 1.433644e-002 3.746801e-004

5.485542e-002 -6.098068e-001 -8.335336e-003 9.690197e-003 2.238820e-003 -1.132704e-003 -

1.335554e-002 -8.233030e-004 -4.228565e-003 -8.020276e-003 -1.142018e-002 -9.332317e-005 -

5.762242e-003 -4.305856e-004 -9.150078e-003

<VARIANCE> 39

3.892350e+000 8.365601e+000 3.846982e+000 4.211987e+000 1.076755e+001 9.036107e+000

9.848803e+000 1.420859e+001 1.078936e+001 1.098874e+001 8.808033e+000 1.003700e+001

8.887128e+002 1.797630e-001 2.355286e-001 3.505588e-001 3.882942e-001 6.302609e-001

7.545438e-001 8.286784e-001 8.659801e-001 1.018844e+000 8.986608e-001 9.457800e-001

8.353303e-001 8.880465e+000 4.273483e-002 4.430741e-002 7.269581e-002 7.994711e-002

1.251430e-001 1.585774e-001 1.765345e-001 1.827718e-001 2.121781e-001 1.873781e-001

1.926342e-001 1.740429e-001 1.413381e+000

<GCONST> 7.409518e+001

<TRANSP> 3

0.000000e+000 1.000000e+000 0.000000e+000

0.000000e+000 9.617694e-001 3.823058e-002

0.000000e+000 0.000000e+000 0.000000e+000

<ENDHMM>

sp นั้นเชื่อมกับ sil โดยมี state กลางเหมือนกัน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บรรณานุกรม

[1] Steve Young และคณะ "A Tutorial Example of Using HTK" *the htk book version 2.2* Entropic , January 1999

[2] รัฐวรนนท์ รอดสัตรู *HTK Programmer Manual* KMITL,1999

[3] รัฐติ์พงษ์ พุทธเจริญและอัศวิน อุดมศิริ *การสังเคราะห์เสียงภาษาไทยด้วยคอมพิวเตอร์*,ปริญญานิพนธ์วิศวกรรมศาสตร์บัณฑิต,กรุงเทพฯ:ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

[4] สมจิตร อาจอินทร์,ผศ. *หลักการเขียนโปรแกรมภาษาปาสคาล* ศูนย์หนังสือมหาวิทยาลัยขอนแก่น, พฤษภาคม 2540

[5] ประพนธ์ อัศวภาณุวัฒน์ *เทคนิคและการพัฒนาโปรแกรมด้วยเดลไฟ* กรุงเทพฯ:ซีเอ็ดยูเคชั่น, 2543

[6] กมลมาศ กำจรกิจการ *คู่มือ Borland Delphi 5 ฉบับสมบูรณ์* กรุงเทพฯ:โปรวิชั่น, 2543

[7] Frank-Engo *How to program Delphi 3* California:Ziff-Davis, 1997

[8] Marco Cantù *Delphi 5* San Francisco:Sybex, 1999

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้