หอสมุดกลาง สถาบันวิทยบริการ จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 3

การจำลองลงบนระบบไมโครคอมพิวเตอร์

ระบบการสังเคราะห์เสียงพูดจากข้อความภาษาไทยนี้ประกอบด้วยขั้นตอนใหญ่ๆ 2 ขั้นตอนที่จะต้องพิจารณาในการจำลองระบบคือ ขั้นตอนการวิเคราะห์เสียงพูด และขั้นตอนการสังเคราะห์เสียงพูด ซึ่งทั้งระบบส่วนใหญ่จะกำหนดให้มีการปฏิบัติงานอยู่บนไมโครคอมพิวเตอร์หรือใช้ตัวไมโครคอมพิวเตอร์เป็นหลักในการควบคุมการทำงาน เพราะสามารถทำได้ค่อนข้างสะดวก นอกจากบางส่วนที่จะต้องติดต่อควบคุมกับระบบการแปลงสัญญาณอนาล็อกเป็นดิจิตอลและระบบการแปลงสัญญาณดิจิตอลเป็นอนาล็อก รวมทั้งส่วนของหน่วยความจำเพิ่มเติม(extended memory) ซึ่งตามปกติส่วนเหล่านี้จะไม่ได้รวมอยู่ในตัวไมโครคอมพิวเตอร์ด้วย แต่จะไปอยู่ในส่วนของ TSB บอร์ด ซึ่งมีไมโครโปรเซสเซอร์ความเร็วสูงเป็นหัวใจในการควบคุมการทำงานเพื่อช่วยติดต่อและควบคุมระบบดังกล่าวแยกออกไปจากตัวไมโครคอมพิวเตอร์ และอาจจะนำคุณสมบัติเฉพาะตัวของมันมาช่วยในการ ทำการประมวลผลสัญญาณแบบดิจิตอลให้เสร็จสิ้นอย่างทันทีทันใดหรือการประมวลผลแบบเวลาจริง (real-time processing) ซึ่งรายละเอียดจะอยู่ในส่วนของแนวทางการวิจัยต่อในบทสุดท้ายต่อไป

ส่วนประกอบภายนอกทั้งหมดของระบบการสังเคราะห์เสียงพูดจากข้อความภาษาไทยที่ใช้ในการวิจัยนี้ประกอบด้วยส่วนประกอบหลักๆ 4 ส่วนด้วยกันดังนี้

1) ไมโครคอมพิวเตอร์รุ่น พีซี-เอ็กซ์ที (IBM PC-XT) ความเร็วของสัญญาณนาฬิกา 4.77 เมกกะเฮิร์ท มีหน่วยความจำ 640 กิโลไบท์ มีตัวประมวลผลช่วยทางคณิตศาสตร์ 8087 เพื่อช่วยในการคำนวณเลขจำนวนจริง (floating-point) ให้เร็วขึ้น ไมโครคอมพิวเตอร์นี้เอาไว้สำหรับทำการควบคุมการทำงานส่วนใหญ่ของระบบ รวมทั้งทำการติดต่อสื่อสารแลกเปลี่ยนข้อมูลกับส่วนประกอบอื่นๆด้วย

2) บอร์ด TSB (TMS32010 Single Board) คือบอร์ดสำหรับพัฒนาระบบของไมโครโปรเซสเซอร์สำหรับประมวลผลสัญญาณดิจิตอลความเร็วสูง TMS32010 ถูกออกแบบอยู่บนแผ่นพิมพ์สองหน้าอเนกประสงค์ที่ใช้เสียบลงบนสล็อตขนาด 8 บิทของไมโครคอมพิวเตอร์เพื่อให้ทำงานร่วมกันได้ นอกจากนั้นบอร์ด TSB สามารถจะทำงานได้โดยอิสระแยกออกจากไมโครคอมพิวเตอร์เหมือนระบบไมโครโปรเซสเซอร์แบบบอร์ดเดี่ยวทั่วๆไป บอร์ด TSB มีส่วนประกอบที่สำคัญๆคือ

- หน่วยความจำโปรแกรมขนาด 4 K x 16 บิท (program memory) เป็นหน่วย

เก็บความจำขนาด 4 กิโลเวิร์ด สำหรับเก็บโปรแกรมที่จะใช้ปฏิบัติการ

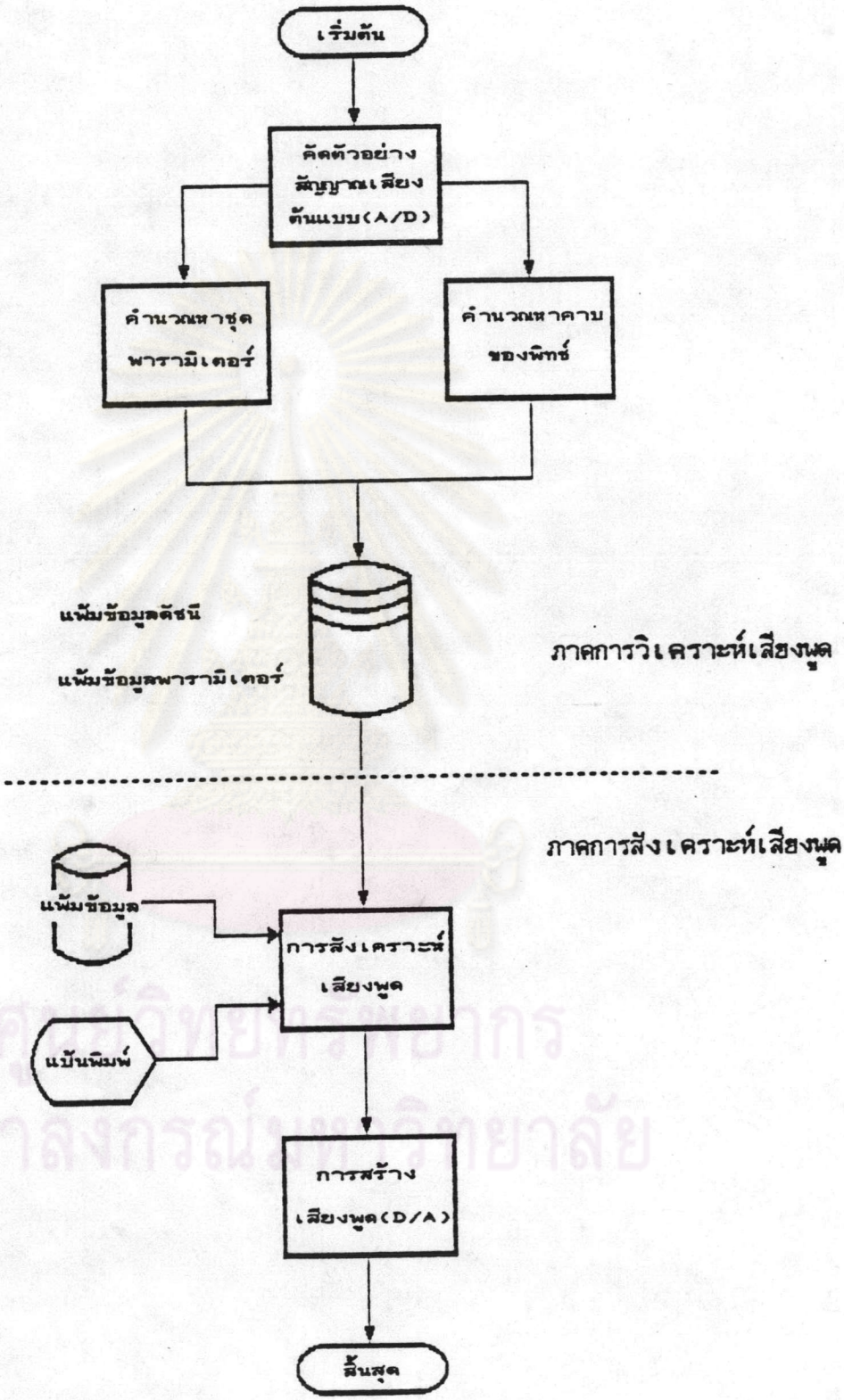
- หน่วยความจำข้อมูลขนาด 64 K x 16 บิท (data memory) เป็นหน่วยความจำเพิ่มเติม (extended memory) เพื่อไว้สำหรับเก็บข้อมูลของเสียงที่ถูกคัดตัวอย่างเข้ามา หรือข้อมูลที่จะทำการส่งถ่ายระหว่างบอร์ด TSB กับไมโครคอมพิวเตอร์ มีขนาดความจุ 64 กิโลเวิร์ด และความเร็วในการเข้าถึงข้อมูล 120 นาโนวินาที ซึ่งหน่วยความจำนี้สามารถถูกใช้โดยไมโครโปรเซสเซอร์ TMS32010 ได้โดยตรง
- ส่วนที่ทำหน้าที่ส่งผ่านข้อมูลระหว่างตัวโปรเซสเซอร์กับอุปกรณ์รอบตัว (I/O port function part) เช่น หน่วยความจำโปรแกรมกับหน่วยความจำข้อมูล เป็นต้น
- ส่วนที่ทำหน้าที่เกี่ยวกับสัญญาณอนาล็อก (analog part) เป็นส่วนที่ทำหน้าที่หลักเกี่ยวกับการแปลงสัญญาณระหว่างสัญญาณอนาล็อกกับสัญญาณดิจิตอล (Analog to Digital A/D , Digital to Analog D/A) โดยจะให้ความละเอียดของข้อมูลดิจิตอลขนาด 12 บิท และมีส่วนของภาคเชื่อมโยงเพื่อใช้ในการติดต่อสื่อสารข้อมูลกับอุปกรณ์ภายนอกเช่น ไมโครโฟน เพื่อใช้สำหรับรับเสียงพูดต้นแบบมาทำการสุ่มสัญญาณ และ ลำโพง สำหรับเปล่งเสียงจากข้อมูลที่ได้ออกมาเป็นต้น
- ส่วนเชื่อมโยงกับไมโครคอมพิวเตอร์ (interface to microcomputer) เป็นส่วนที่ทำให้ไมโครคอมพิวเตอร์ทำงานร่วมกับ TSB บอร์ดได้ ตั้งแต่การเข้าถึงหน่วยความจำโปรแกรมและหน่วยความจำข้อมูลบน TSB บอร์ด สามารถส่งถ่ายข้อมูลระหว่างกันได้ และการสร้างสัญญาณที่จะควบคุม TSB บอร์ด

3) ไมโครโฟนแบบทิศทางเดียว (uni-direction microphone) มีหน้าที่แปลงสัญญาณเสียงพูดให้เป็นสัญญาณไฟฟ้า มีความไวในการรับสัญญาณเสียงในทิศทางเพียงทิศทางเดียว เพื่อป้องกันสัญญาณเสียงรบกวนรอบๆข้างเข้ามา

4) ลำโพง มีหน้าที่แปลงสัญญาณไฟฟ้าให้เป็นสัญญาณเสียง เพื่อใช้ทดสอบฟังเสียงที่ต้องการ

จากอุปกรณ์ทั้ง 4 ส่วนนี้เราสามารถแยกพิจารณาลำดับการทำงานตามขั้นตอนของโปรแกรมได้ดังรูปที่ 3.1 ซึ่งประกอบด้วย 4 กระบวนการใหญ่ๆ คือ

- ภาคของการวิเคราะห์เสียงพูด (Speech analysis procedure)
- ภาคของการสังเคราะห์เสียงพูด (Speech synthesis procedure)
- ภาครับอักขระอินพุทภาษาไทย
- โปรแกรมอรรถประโยชน์



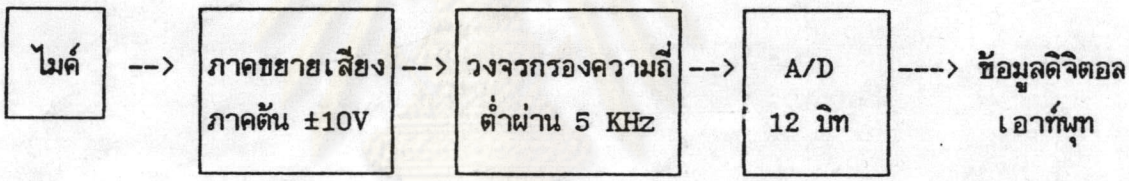
รูปที่ 3.1 ขั้นตอนการทำงานรวมของระบบ

3.1 ภาคของการวิเคราะห์เสียงพูด (Speech analysis procedure)

เป็นส่วนที่เริ่มตั้งแต่นำสัญญาณเสียงพูดต้นแบบมาทำการวิเคราะห์หาชุดของพารามิเตอร์ต่างๆ ซึ่งจะใช้เก็บไว้แทนข้อมูลของเสียงแต่ละคำทั้งหมดเอาไว้เพื่อใช้ในการสังเคราะห์เสียงพูดออกมาในขั้นตอนของกระบวนการสังเคราะห์เสียงต่อไป ภาคของการวิเคราะห์เสียงพูดนี้ถูกออกแบบให้ทำงานในลักษณะออฟไลน์ สามารถที่จะปฏิบัติการเมื่อไรก็ได้ที่จะทำการวิเคราะห์เสียงต้นแบบพยางค์ใหม่ๆเก็บเพิ่มเติมเข้าไว้ในพจนานุกรมข้อมูล ภาคนี้มีขั้นตอนและโปรแกรมย่อยๆ ที่ทำงานแยกกันดังนี้

3.1.1 ขั้นตอนสำหรับทำการคัดตัวอย่างสัญญาณเสียงต้นแบบ

เป็นขั้นตอนที่ทำหน้าที่แปลงสัญญาณเสียงพูดที่รับเข้ามาซึ่งเป็นสัญญาณอนาล็อกให้กลายเป็นสัญญาณดิจิตอล เพื่อนำมาใช้ในการวิเคราะห์และประมวลผลสัญญาณเสียงในรูปแบบของสัญญาณดิจิตอลต่อไป ส่วนประกอบของขั้นตอนนี้มีรายละเอียดคือ (รูปที่ 3.2)



คัดตัวอย่าง 10,000 จุด/วินาที

รูปที่ 3.2 ขั้นตอนสำหรับทำการคัดตัวอย่างสัญญาณเสียง

ไมโครโฟนจะทำหน้าที่รับสัญญาณเสียงพูดต้นแบบเข้ามาแปลงจากคลื่นเสียงให้เป็นสัญญาณไฟฟ้าแบบอนาล็อกป้อนเข้าสู่ภาคขยายเสียงภาคต้น (Pre-amp) เพื่อขยายให้สัญญาณมีขนาดของแอมปลิจูดมากขึ้นจนได้ระดับใกล้เคียงกับ ± 10 โวล์ทซึ่งเป็นระดับสัญญาณเข้าสูงสุดของวงจร A/D โดยภาคขยายเสียงภาคต้นนี้สามารถปรับอัตราการขยายได้ทั้งนี้เพื่อให้ขึ้นอยู่กับว่าเสียงพูดที่เข้ามาดังหรือเบา แต่สัญญาณที่จุดยอดคลื่นที่สูงที่สุดจะต้องไม่เกิน ± 10 โวล์ทเป็นอันขาดเพราะจะทำให้เกิดการตัดยอดคลื่น (cliping) ขึ้นซึ่งจะทำให้รูปคลื่นของสัญญาณเสียงผิดเพี้ยนไป จากนั้นสัญญาณเสียงที่ถูกขยายขึ้นมาแล้วนี้จะถูกป้อนผ่านวงจรกรองความถี่ต่ำผ่านที่จุดตัดประมาณ 5 กิโลเฮิร์ท ตามทฤษฎีของการคัดสัญญาณจากหัวข้อ 2.2 แล้วถึงป้อนเข้าสู่วงจรแปลงสัญญาณอนาล็อกเป็นสัญญาณดิจิตอลซึ่งมีไอซีเบอร์ ADC80-Z เป็นตัวจัดการ โดยจะให้ความละเอียดในการแปลงขนาด 12 บิทต่อการคัดสัญญาณหนึ่งครั้งและใช้เวลาในการแปลงประมาณ 25 ไมโครวินาที ความถี่ในการคัดสัญญาณสามารถกำหนดได้จากโปรแกรม จากนั้นก็จะเก็บข้อมูลที่ได้

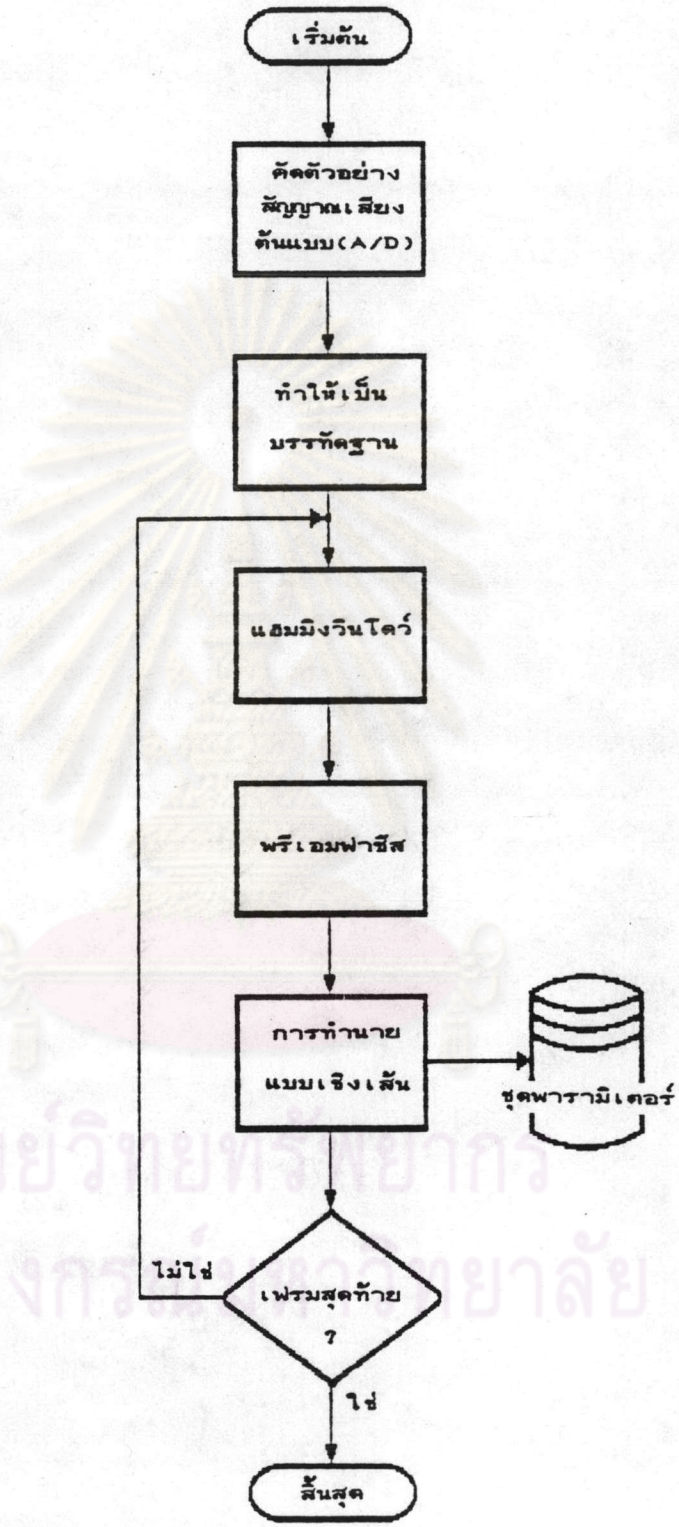
ออกมาในแต่ละครั้ง (2 ไบท์โดยที่ 4 บิทต่ำไม่ได้ใช้) เก็บเอาไว้ในหน่วยความจำเพิ่มเติมขนาด 64 กิโลเวิร์ดที่อยู่บนบอร์ด TSB ก่อนแล้วค่อยใช้ไมโครคอมพิวเตอร์ส่งถ่ายข้อมูลทั้งหมดเข้าไปเก็บไว้ในแผ่นจานแม่เหล็กของไมโครคอมพิวเตอร์เพื่อนำข้อมูลของเสียงต้นแบบไปใช้ในขั้นตอนต่อไป

ขั้นตอนการทำงานต่างๆ ที่กล่าวมานี้จะควบคุมการทำงานโดยโปรแกรมสองส่วนที่ทำงานร่วมกันคือ โปรแกรมส่วนที่ทำงานอยู่บนบอร์ด TSB และโปรแกรมส่วนที่ทำงานอยู่บนไมโครคอมพิวเตอร์ โปรแกรมที่ทำงานอยู่บนบอร์ด TSB เป็นโปรแกรมภาษาแอสเซมบลี้ของไมโครโปรเซสเซอร์ TMS32010 ที่ผ่านการแปลโปรแกรม (compile) และเชื่อมโยงโปรแกรม (link) จนได้เป็นโปรแกรมภาษาเครื่องที่พร้อมที่จะปฏิบัติการเรียบร้อยแล้วนำมาโหลดไว้ในส่วนของหน่วยความจำโปรแกรมบนบอร์ด TSB ซึ่งตัวไมโครโปรเซสเซอร์ TMS32010 จะทำการปฏิบัติการตามขั้นตอนของโปรแกรมส่วนนี้เกี่ยวกับการควบคุมการคัดตัวอย่างสัญญาณเสียงทั้งหมดจนกระทั่งได้ข้อมูลในรูปแบบดิจิตอลออกมา จำนวนของข้อมูลที่ได้จากการคัดตัวอย่างนี้จะขึ้นอยู่กับว่าจะให้ทำการคัดสัญญาณเสียงต้นแบบเป็นเวลานานแค่ไหน แต่ทั้งนี้ก็ถูกจำกัดโดยขนาดความจุสูงสุดของหน่วยความจำเพิ่มเติมคือ 64 กิโลเวิร์ด นั่นคือที่อัตราความถี่ในการคัดสัญญาณ 10,000 ครั้งต่อวินาทีนี้ก็จะสามารถเก็บสัญญาณเสียงพูดได้นานที่สุด 6.4 วินาที (ซึ่งตามปกติหน่วยย่อยของเสียงเช่นพยางค์หรือคำจะใช้เวลาประมาณ 0.5 ถึง 1 วินาทีเท่านั้น) หลังจากคัดสัญญาณเก็บไว้ในหน่วยความจำเพิ่มเติมจนครบจำนวนแล้ว โปรแกรมก็จะรอการส่งถ่ายข้อมูลไปสู่ไมโครคอมพิวเตอร์ซึ่งก็จะใช้โปรแกรมติดต่อรับข้อมูลไปเก็บไว้ในจานแม่เหล็กต่อไป

3.1.2 โปรแกรมสำหรับคำนวณหาชุดพารามิเตอร์ของสัญญาณเสียง

เป็นโปรแกรมที่นำเอาข้อมูลดิจิตอลของเสียงต้นแบบมาผ่านฟังก์ชั่นต่างๆ เพื่อคำนวณหาค่าพารามิเตอร์ โดยแบ่งสัญญาณออกมาพิจารณาเป็นเฟรมๆไปผ่านขั้นตอนของการคำนวณตามวิธีของการทำนายแบบเชิงเส้น แล้วเก็บค่าพารามิเตอร์เหล่านี้เอาไว้ใช้แทนข้อมูลของเสียงทั้งหมดต่อไป รายละเอียดของโปรแกรมเป็นดังรูปที่ 3.3

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย



รูปที่ 3.3 ผังงานของโปรแกรมคำนวณหาชุดพารามิเตอร์

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

1) รับข้อมูลดิจิตอลของเสียงต้นแบบเข้ามาทำให้เป็นข้อมูลขนาด 16 บิท แบบ 2 คอมพลีเมนท์
2) ทำข้อมูลให้เป็นบรรทัดฐาน (normalize) โดยการเก็บค่าสัมบูรณ์สูงสุดของข้อมูลทั้งหมดเอามาเปรียบเทียบกับค่าคงที่ ค่าของข้อมูลค่าอื่นๆ ก็จะเป็นสัดส่วนโดยตรงกับค่าคงนี้
3) แบ่งพิจารณาข้อมูลออกเป็นเฟรมๆ เฟรมละ 200 จุด (20 มิลลิวินาที) แล้วพิจารณาแต่ละเฟรมผ่านช่องแคบแบบแฮมมิงขนาด 300 จุด (แต่ละเฟรมจะเหลื่อมทับกัน 100 จุด)
4) ผ่านฟังก์ชั่น พรี-เอมฟาซีส (pre-emphasis)
5) ส่งข้อมูลแต่ละเฟรมเข้าสู่ฟังก์ชั่นการทำนายแบบเชิงเส้น โดยใช้เทคนิคของ Durbin และ Levinson หาค่า ซิกม่า กับ ค่าสัมประสิทธิ์การทำนายจำนวน 10 ค่า ต่อเฟรม
6) ส่งข้อมูลแต่ละเฟรมเข้าสู่ฟังก์ชั่นการหาคาบของพิทช์ โดยใช้เทคนิคเซนเตอร์คลิปปิง ได้ค่าคาบของพิทช์ออกมา (มีค่าเป็นศูนย์กรณีที่เป็นเฟรมของเสียงไม่ก้อง หรือ เฟรมของสัญญาณรบกวน)
7) รวมพารามิเตอร์ทั้งหมดเข้าด้วยกันแล้วจัดเข้าเป็นระบบของพจนานุกรมข้อมูล ซึ่งประกอบด้วยแฟ้มข้อมูล 2 แฟ้มข้อมูลด้วยกัน แฟ้มข้อมูลดัชนีจะเก็บป้ายชื่อของพยางค์แต่ละพยางค์พร้อมกับตัวชี้ที่จะบอกตำแหน่งของชุดพารามิเตอร์ของมันที่เก็บอยู่ในแฟ้มข้อมูลพารามิเตอร์ โดยที่แฟ้มข้อมูลทั้งสองมีโครงสร้างดังนี้

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

แฟ้มข้อมูลดัชนี :

ป้ายชื่อพยางค์	1	ไบท์
ตำแหน่งที่อยู่	1	ไบท์
ความยาวที่ใช้	1	ไบท์

แฟ้มข้อมูลพารามิเตอร์ :

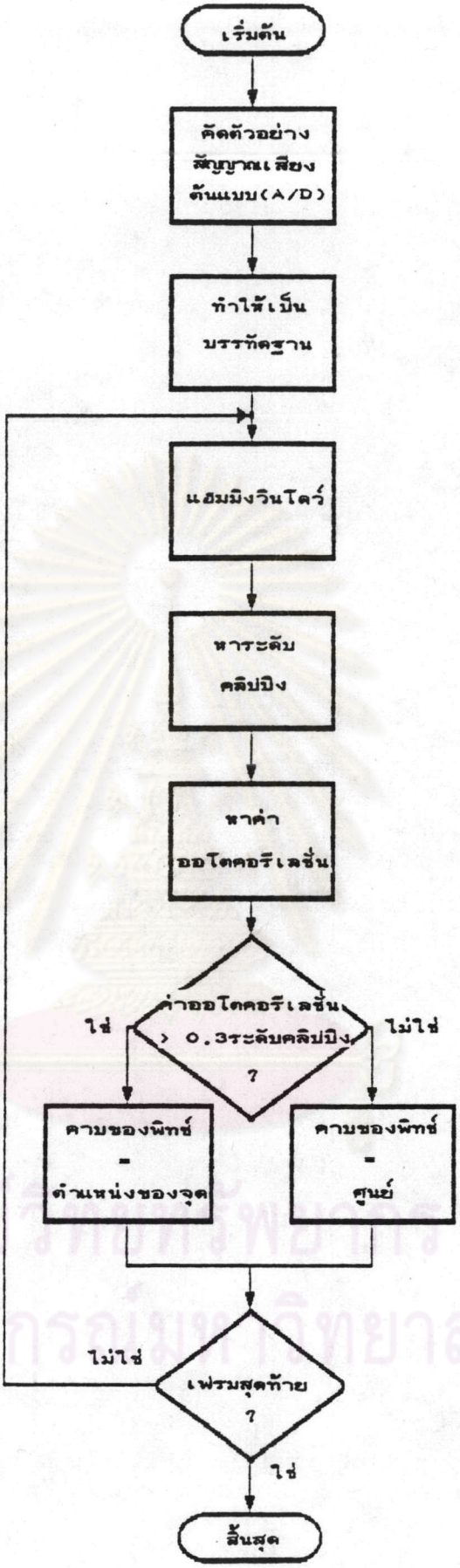
****

จำนวนเฟรมของพยางค์	1	ไบท์
จำนวนจุดต่อเฟรม	1	ไบท์
อันดับของการทำนายแบบเชิงเส้น	1	ไบท์
===== ข้อมูลของเฟรมที่ 1 =====		
ค่าซิกม่า	2	ไบท์
ค่าคาบของพิทช์	1	ไบท์
ค่าสัมประสิทธิ์การสะท้อน 10 ค่า	10 x 4	ไบท์
===== ข้อมูลของเฟรมที่ 2 =====		
ค่าซิกม่า	2	ไบท์
ค่าคาบของพิทช์	1	ไบท์
ค่าสัมประสิทธิ์การสะท้อน 10 ค่า	10 x 4	ไบท์
===== ข้อมูลของเฟรมที่ 3 =====		
ค่าซิกม่า	1	ไบท์
ค่าคาบของพิทช์	2	ไบท์
ค่าสัมประสิทธิ์การสะท้อน 10 ค่า	10 x 4	ไบท์

รูปที่ 3.4 ภาพแสดงโครงสร้างของระบบพจนานุกรมข้อมูล

3.1.3 โปรแกรมสำหรับคำนวณหาคาบของพิทช์

เป็นโปรแกรมที่นำเอาข้อมูลดิจิตอลของเสียงต้นแบบเช่นกัน มาคำนวณหาคาบของพิทช์ในแต่ละเฟรมโดยวิธีเซนเตอร์คลิปปิง ในกรณีที่เสียงนั้นเป็นเสียงไม่ก้องก็จะได้ค่าของคาบของพิทช์นั้นเป็นศูนย์ ซึ่งทำให้สามารถพิจารณาได้ว่าเสียงในเฟรมนั้นเป็นเสียงก้องหรือไม่จากค่าคาบของพิทช์นี้ด้วย รายละเอียดดังรูปที่ 3.5



รูปที่ 3.5 ผังงานของโปรแกรมคำนวณหาคาบของพิทช์

1) รับข้อมูลจากขั้นตอนที่ 4) ของโปรแกรมการคำนวณหาชุดพารามิเตอร์
2) แบ่งข้อมูลออกเป็นเฟรมๆ เฟรมละ 300 จุด (30 มิลลิวินาที) โดยที่เฟรมถัดไปจะเขยิบถัดจากจุดเริ่มต้นของเฟรมแรก 100 จุด (เหลื่อมทับกับเฟรมแรก 200 จุด)
3) หาค่าระดับคลิปปิง โดยคิดจาก 30 เปอร์เซนต์ของค่าที่ต่ำที่สุดระหว่างค่าสัมบูรณ์สูงสุดของ 100 จุดแรกกับ 100 จุดท้ายของเฟรม
4) ใช้ค่าระดับคลิปปิงมาเปรียบเทียบกับข้อมูลทุกๆ จุดในเฟรม โดยกำหนดเงื่อนไขเซนเตอร์คลิปปิงแบบสามระดับ และหาค่าอัตสหสัมพันธ์ออกมา
5) จากค่าที่ได้หลังจากคำนวณอัตสหสัมพันธ์ จากค่าที่มากที่สุดค่าแรกถ้ามีค่ามากเกินกว่า 30 เปอร์เซนต์ของค่าอัตสหสัมพันธ์ที่จุดที่ศูนย์แล้วเฟรมนั้นจะเป็นเฟรมของเสียงก้อง และมีค่าคาบของพิทช์เท่ากับตำแหน่งที่จุดนั้น แต่ถ้าไม่มากกว่าก็จะเป็นเฟรมของเสียงไม่ก้อง

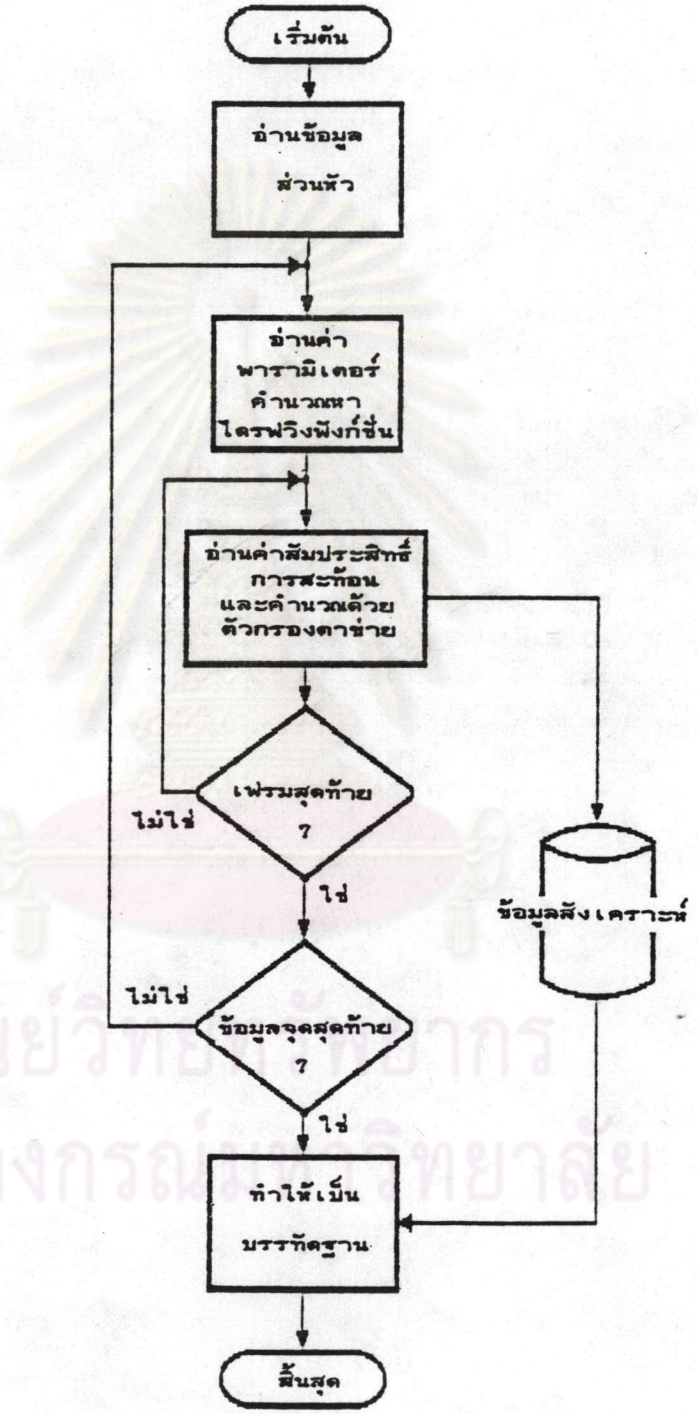
3.2 ภาคของการสังเคราะห์เสียงพูด (Speech synthesis procedure)

เป็นส่วนที่ทำการสังเคราะห์เสียงพูดออกมา โดยใช้ชุดพารามิเตอร์ที่ได้มาจากภาคการวิเคราะห์เสียงพูดมาทำการคำนวณเพื่อให้ได้ข้อมูลดิจิตอลกลับคืนมา ภาคของการสังเคราะห์เสียงพูดนี้มีโปรแกรมและขั้นตอนย่อยๆ คือ

3.2.1 โปรแกรมสังเคราะห์เสียงพูด

เป็นโปรแกรมที่นำเอาชุดพารามิเตอร์ มาทำการคำนวณตามวิธีการของตัวกรองตาข่าย เพื่อให้ได้ข้อมูลของเสียงสังเคราะห์แบบดิจิตอลกลับออกมาแล้วเก็บไฟล์ข้อมูลของเสียงสังเคราะห์นี้เอาไว้เพื่อทำให้เกิดเป็นเสียงในขั้นตอนต่อไป รายละเอียดมีดังนี้

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

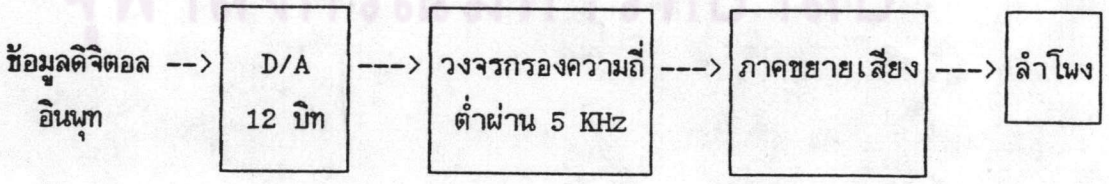


รูปที่ 3.6 ผังงานของโปรแกรมสังเคราะห์เสียงพูด

1) อ่านข้อมูลส่วนหัวของพยางค์ คือ จำนวนเฟรมของพยางค์นั้น จำนวนจุดต่อเฟรม และอันดับของการทำนายแบบเชิงเส้น
2) วนซ้ำเท่ากับจำนวนเฟรม
3) อ่านค่าพารามิเตอร์ต่างๆ และคำนวณหาไดรฟวิงฟังก์ชั่นของเสียงก้องและเสียงไม่ก้อง
4) วนซ้ำเท่ากับจำนวนจุดในเฟรม
5) คำนวณตัวกรองตาข่าย แบบตัวคูณสองตัว โดยใช้ค่าสัมประสิทธิ์การสะท้อนและค่าไดรฟวิงฟังก์ชั่นเป็นอินพุท แล้วได้ค่าข้อมูลของเสียงสังเคราะห์เป็นเอาท์พุทออกมาหนึ่งจุด
6) ทำข้อมูลเอาท์พุททั้งหมดเป็นบรรทัดฐาน โดยหาค่าสัมบูรณ์ที่มากที่สุดมากำหนดให้เท่ากับค่าคงที่ ข้อมูลค่าอื่นๆก็จะเป็นสัดส่วนโดยตรงกับค่าคงที่นี้ เก็บข้อมูลทั้งหมดไว้ในบัฟเฟอร์

3.2.2 ขั้นตอนการสร้างเสียงพูด

เป็นขั้นตอนที่นำเอาข้อมูลของเสียงสังเคราะห์ที่เป็นข้อมูลแบบดิจิตอลมาแปลงให้กลายเป็นสัญญาณแบบอนาล็อก แล้วส่งออกภาคขยายเสียงเพื่อเปล่งเป็นเสียงออกมา เริ่มจากไมโครคอมพิวเตอร์ส่งถ่ายข้อมูลทั้งหมดที่สังเคราะห์ออกมาได้ไปยังหน่วยความจำเพิ่มเติมบน TSB บอร์ด โปรแกรมบน TSB บอร์ดก็จะทะยอยส่งข้อมูลทีละจุดออกไปยังวงจรแปลงสัญญาณดิจิตอลเป็นอนาล็อก โดยที่จะต้องส่งให้มีระยะเวลาที่ถูกต้องคือ 10,000 จุดในเวลาหนึ่งวินาทีเท่ากับตอนที่คัดสัญญาณเข้ามาเพื่อจะได้เสียงที่มีจังหวะถูกต้องไม่ผิดเพี้ยน วงจรแปลงสัญญาณดิจิตอลเป็นอนาล็อกนั้นมีไอซีเบอร์ AD565A เป็นหลัก สัญญาณอนาล็อกที่ได้ออกมาจะต้องนำไปผ่านวงจรกรองต่ำผ่านที่จุดตัด 5 กิโลเฮิร์ทก่อนเพื่อให้สัญญาณมีความราบเรียบขึ้น แล้วถึงเข้าสู่ภาคขยายเสียงออกสู่ลำโพงเปล่งเป็นเสียงพูดออกมา



รูปที่ 3.7 ขั้นตอนการสร้างเสียงพูด

3.3 ภาครับอักขระอินพุทภาษาไทย

เป็นโปรแกรมที่ทำการรับข้อความภาษาไทยที่เป็นรูปแบบของคำอ่านเข้ามา รวมทั้งอักขระพิเศษที่ใช้ในการเปลี่ยนแปลงลักษณะของเสียงที่จะสังเคราะห์ออกมา โปรแกรมจะกวาดหาอักขระภาษาไทยทีละพยางค์แล้วเอาไปค้นหาในแฟ้มข้อมูลดัชนี เพื่อดึงเอาชุดของพารามิเตอร์ของพยางค์นั้นออกมาจากไฟล์ข้อมูลพารามิเตอร์อีกทีหนึ่งเพื่อส่งไปยังภาคของการสังเคราะห์เสียง ความสามารถของโปรแกรมรับอักขระอินพุทภาษาไทยมีดังนี้

1) รับอักขระภาษาไทยรหัสเกษตรและสัญญลักษณ์พิเศษที่กำหนดได้จากทาง แป้นพิมพ์ หรือจากแฟ้มข้อมูลก็ได้
2) ใช้สัญญลักษณ์ เว้นวรรค,ขีด(-),<ENTER> เป็นตัวแบ่งระหว่างพยางค์
3) ใช้สัญญลักษณ์ จุลภาค(,) เป็นตัวบอกให้โปรแกรมแปล่งเสียงในบัฟเฟอร์ออกมาก่อนจบบรรทัด (ตามปกติ โปรแกรมจะเปล่งเสียงออกมาเมื่อพบรหัสขึ้นบันทัดใหม่)
4) ใช้สัญญลักษณ์ เครื่องหมายหาร (/) เป็นตัวเว้นช่วงว่างระหว่างพยางค์ โดยเครื่องหมายหารหนึ่งตัวจะเว้นว่างเป็นเวลานาน 20 มิลลิวินาที (สามารถใช้เครื่องหมายหารซ้อนๆกันหลายๆตัวได้ในกรณีที่ต้องการเว้นช่วงนานๆ เช่น เว้นช่วงว่างระหว่างคำ เป็นต้น)
5) ใช้สัญญลักษณ์ เครื่องหมายบวก (+) หรือเครื่องหมายลบ (-) ตามด้วยตัวเลขต่อท้ายพยางค์ เพื่อให้โปรแกรมแปลี่ยนแปลงค่าคาบของพิทช์ไปจากเดิมก่อนทำการสังเคราะห์เป็นเสียงออกมา
6) ใช้ตัวอักษร p หรือ P เพื่อบอกให้โปรแกรมเปล่งเสียงของข้อมูลในบัฟเฟอร์ที่ได้มาจากการสังเคราะห์เสียงครั้งล่าสุดออกมาใหม่อีกครั้ง
7) ใช้ตัวอักษร f หรือ F เพื่อบอกให้โปรแกรมรับอินพุทจากแฟ้มข้อมูลแทนแป้นพิมพ์
8) เนื่องจากความจำกัดในความสามารถทางด้านเวลาในการคำนวณ ของเครื่องไมโครคอมพิวเตอร์ที่ใช้ จึงทำให้โปรแกรมสามารถเปล่งเสียงอย่างต่อเนื่องได้ไม่เกิน 4 วินาทีต่อครั้งเท่านั้น

3.4 โปรแกรมอรรถประโยชน์

เป็นชุดของโปรแกรมที่มีไว้สำหรับ ใช้เป็นเครื่องมือช่วยให้การทำงานต่างๆ ง่ายขึ้น สะดวกขึ้นและอำนวยความสะดวกแก่การวิจัย โปรแกรมที่สำคัญในชุดของโปรแกรมอรรถประโยชน์มีดังนี้

3.4.1 โปรแกรมสำหรับแก้ไขสัญญาณข้อมูล (signal editor)

เป็นโปรแกรมที่ใช้เป็นเครื่องมือช่วยในการแก้ไขสัญญาณข้อมูลที่ได้มาจากการคัดสัญญาณเสียงต้นแบบ โดยการนำเอาข้อมูลของสัญญาณเสียงที่เป็นสัญญาณแบบดิจิตอลมาพล็อตขึ้นบนจอภาพให้เห็นรูปร่างของสัญญาณและสามารถทำการตกแต่งแก้ไขข้อมูลได้โดยตรง โปรแกรมจะเริ่มจากการกำหนดชื่อแฟ้มข้อมูลดิจิตอลที่ต้องการนำมาแก้ไข กำหนดจุดที่จะให้เริ่มต้นทำการพล็อต ถ้าไม่มีการกำหนดโปรแกรมจะเริ่มจากจุดแรกของข้อมูล กำหนดว่าจะให้พล็อตเป็นจำนวนกี่จุดต่อจากจุดแรกถ้าไม่กำหนดจะทำการพล็อตทุกๆจุดข้อมูล จะต้องกำหนดว่าจะให้พล็อตที่วินโดว์ไหน โดยจอภาพจะมีการแบ่งออกเป็น 3 วินโดว์ วินโดว์ที่ 1 หมายถึงพล็อตเต็มทั้งจอ วินโดว์ที่ 2 และ 3 หมายถึงครึ่งจอด้านบนและล่างตามลำดับ เมื่อกำหนดครบแล้วโปรแกรมจะทำการอ่านข้อมูลขึ้นมาเก็บไว้บนหน่วยความจำ แล้วนำเอาข้อมูลแต่ละค่ามาพล็อตบนจอตามที่กำหนด โปรแกรมจะนำข้อมูลทั้งหมดมาพล็อตอย่างเหมาะสม เช่นอาจจะพล็อตจุดเว้นจุด จุดเว้นสองจุด ขึ้นอยู่กับปริมาณข้อมูลทั้งหมดที่เรากำหนด กับจำนวนบัฟเฟอร์บนจอภาพตามแกนนอน ที่สามารถจะแสดงผลได้ เมื่อพล็อตข้อมูลขึ้นบนจอเรียบร้อยแล้วเราสามารถจะทำการซูมข้อมูลบางส่วนบนจอได้ โดยอาศัยการเลือกฟังก์ชั่นซูม แล้วกำหนดจุดเริ่มต้นกับจุดท้ายสุด โปรแกรมจะทำการดึงข้อมูลเฉพาะส่วนนั้นขึ้นมาแสดงผลบนจอแทน จากฟังก์ชั่นซูมนี้ทำให้เราสามารถเลือกข้อมูลเอาเฉพาะส่วนที่ต้องการและตัดส่วนที่เป็นสัญญาณรบกวนทิ้งได้

3.4.2 โปรแกรมบำรุงรักษาพจนานุกรมข้อมูล

คือโปรแกรมที่ใช้เป็นเครื่องมือช่วยในการเปลี่ยนแปลงแก้ไขระบบพจนานุกรมข้อมูลที่มีอยู่โดยไม่กระทบกระเทือนต่อโครงสร้างเดิม ฟังก์ชั่นที่เตรียมเอาไว้ให้ใช้งานมีดังนี้

1) สามารถเรียกดูรายละเอียดของพยางค์ต่างๆ ที่มีอยู่ในระบบพจนานุกรมข้อมูลได้ เช่น ตำแหน่งที่อยู่ ความยาวที่ใช้ จำนวนเฟรม จำนวนจุดต่อเฟรม ค่าอันดับที่ใช้ในการทำนายแบบเชิงเส้น และค่าพารามิเตอร์ต่างๆ (ค่าซิกม่า ค่าคาบของพิทช์ ค่าสัมประสิทธิ์การสะท้อน) ฯลฯ ถ้ามีการระบุชื่อพยางค์ด้วยก็จะแสดงรายละเอียดเฉพาะพยางค์นั้น
2) สามารถเปลี่ยนแปลงค่าพารามิเตอร์ของพยางค์ที่ระบุให้ทันกาลได้
3) สามารถเปลี่ยนแปลงป้ายชื่อของพยางค์ที่ระบุได้
4) สามารถทำการลบข้อมูลทั้งหมดของพยางค์ที่ระบุ ออกไปจากระบบพจนานุกรมข้อมูล โดยที่ข้อมูลของพยางค์อื่นไม่ถูกกระทบกระเทือนด้วย
5) สามารถทำการเพิ่มเติมเฉพาะป้ายชื่อของพยางค์ที่ระบุได้ ในกรณีที่ต้องการให้พยางค์นั้นสามารถอ้างถึงได้มากกว่าหนึ่งชื่อ