หอสมุดกลาง สถาบันวิทยบริการ จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 4

วิเคราะห์รายละเอียดและผลการทดสอบ

วิเคราะห์รายละเอียดของการทดสอบ

จากการทดสอบทั้งหมดของการวิจัยนี้ สามารถจะแบ่งวิเคราะห์การพิจารณารายละเอียดของการทดสอบออกได้เป็นส่วนที่สำคัญๆ 3 ส่วนคือ ส่วนของการวิเคราะห์เสียงต้นแบบ ส่วนของการแก้ไขพารามิเตอร์ และส่วนของการสังเคราะห์เสียงพูด

4.1 ส่วนของการวิเคราะห์เสียงต้นแบบ

1) ขั้นตอนการเก็บข้อมูลอินพุทต้นแบบ

เริ่มแรกก่อนที่จะทำให้ระบบสามารถสังเคราะห์เสียงพูดออกมาได้นั้น จะต้องทำการเก็บเสียงพูดต้นแบบของหน่วยย่อยของคำแต่ละหน่วยเอาไว้ก่อน ในขอบเขตของการวิจัยนี้ก็ได้ใช้พยางค์เป็นหน่วยย่อยของคำในการเก็บ โดยได้กำหนดกลุ่มของพยางค์ตัวอย่างที่จะใช้เอาไว้ในด้านของ การแจ้งบอกราคาหรือจำนวนตัวเลข ซึ่งมีพยางค์ที่จำเป็นต้องใช้เลือกเก็บเอาไว้ในพจนานุกรมข้อมูลเป็นจำนวน 28 พยางค์ ดังตารางที่ 4.1

เสียงต้นฉบับนี้ใช้เสียงของผู้วิจัยเป็นเสียงต้นแบบ โดยพูดเน้นแยกทีละพยางค์ผ่านภาคการคัดตัวอย่างสัญญาณเสียงเก็บเอาไว้ในรูปของข้อมูลดิจิตอล ซึ่งสามารถนำมาพล็อตดูค่าแอมปลิจูดแปรผันตามเวลาได้โดยใช้ โปรแกรมแก้ไขสัญญาณ ได้รูปของสัญญาณเสียงออกมาดังรูปที่ 4.1

ค่าของสัญญาณเสียงต้นแบบเมื่อเทียบกับค่าสัญญาณรบกวนของระบบ ภายใต้สภาวะในห้องวิจัยที่แทรกเข้ามาขณะคัดตัวอย่างสัญญาณด้วยคิดได้เป็น 35 เดซิเบลซึ่งนับว่าค่อนข้างต่ำเมื่อเทียบกับค่า SNR ของสัญญาณที่วัดจากการคัดตัวอย่างสัญญาณด้วยความละเอียดในการคัดตัวอย่างขนาด 12 บิทซึ่งมีค่า SNR เท่ากับ 65 เดซิเบลตามหัวข้อที่ 2.2 ข้อมูลของสัญญาณเสียงต้นฉบับแต่ละพยางค์นี้สามารถที่จะนำมาส่งออกภาค D/A ให้กลับคืนมาเป็นเสียงได้โดยตรงเพื่อใช้ในการทดสอบดูว่าข้อมูลที่เก็บได้จากการสุ่มสัญญาณเสียงนั้นผิดเพี้ยนไปหรือไม่ ถ้าเสียงที่ได้ออกมาไม่ชัดพอหรือมีสัญญาณรบกวนแทรกอยู่มากก็จะต้องทำการเก็บเสียงพยางค์นั้นใหม่ หลังจากที่ได้ข้อมูลของสัญญาณเสียงแต่ละพยางค์ที่ชัดเจนพอใช้ได้แล้ว ก็จะนำมาตัดเอาเฉพาะส่วนที่

เป็นข้อมูลของพยางค์นั้นจริงๆ โดยพิจารณาดูจากค่าพลังงานของเสียงที่จุดเริ่มต้นและจุดสิ้นสุด ประกอบกับรูปของสัญญาณเสียงที่พล็อตขึ้นมาบนจอโดยโปรแกรมแก้ไขสัญญาณ แล้วลองส่งข้อมูลที่ตัดได้ส่งออก D/A เพื่อทดสอบฟังเสียงอีกครั้งว่าเสียงที่ได้ออกมานี้ยังคงสามารถฟังได้เป็นเสียงที่ถูกต้องของพยางค์นั้นๆอยู่ นั่นคือผลสุดท้ายของขั้นตอนนี้ก็จะได้ข้อมูลของพยางค์แต่ละพยางค์ล้วนๆที่จะมาใช้เป็นต้นแบบในการสังเคราะห์เสียงพูดต่อไป ซึ่งคุณภาพของเสียงต้นแบบนี้ก็จะเป็นส่วนสำคัญที่สุดในการกำหนดคุณภาพของเสียงที่ได้จากการสังเคราะห์ในขั้นตอนสุดท้าย

2) ขั้นตอนการวิเคราะห์เสียงต้นแบบให้ออกมาเป็นชุดพารามิเตอร์

เป็นขั้นตอนที่นำเอาข้อมูลของเสียงต้นแบบแต่ละพยางค์มาเป็นข้อมูลอินพุทผ่านอัลกอริทึม LPC เพื่อทำการวิเคราะห์เสียงออกมา โดยแบ่งข้อมูลออกสำหรับพิจารณาในช่วงเวลาสั้นๆเป็นเฟรมๆ เฟรมละ 20 มิลลิวินาที รวมทั้งการหาคาบของสัญญาณโดยใช้อัลกอริทึมเซนเตอร์คลิปปิง ด้วย ผลสุดท้ายของขั้นตอนนี้ก็จะได้พารามิเตอร์ 1 ชุดของแต่ละเฟรม โดยชุดพารามิเตอร์นี้ก็จะประกอบด้วย ค่าซิกม่า ค่าคาบของพิทช์ และค่าสัมประสิทธิ์การสะท้อน 10 ค่า เก็บรวมกันไว้เป็นข้อมูลรวมอยู่ในพจนานุกรมข้อมูลเพื่อเอาไว้ในการสังเคราะห์เสียงต่อไป

จากข้อมูลของเสียงต้นแบบขนาด 1 เฟรม จำนวน 200 จุดจะใช้เนื้อที่ในการเก็บ 400 ไบท์ (1 จุดใช้เนื้อที่ 2 ไบท์) เมื่อนำมาทำการวิเคราะห์แล้วจะได้ชุดของพารามิเตอร์ออกมาดังนี้

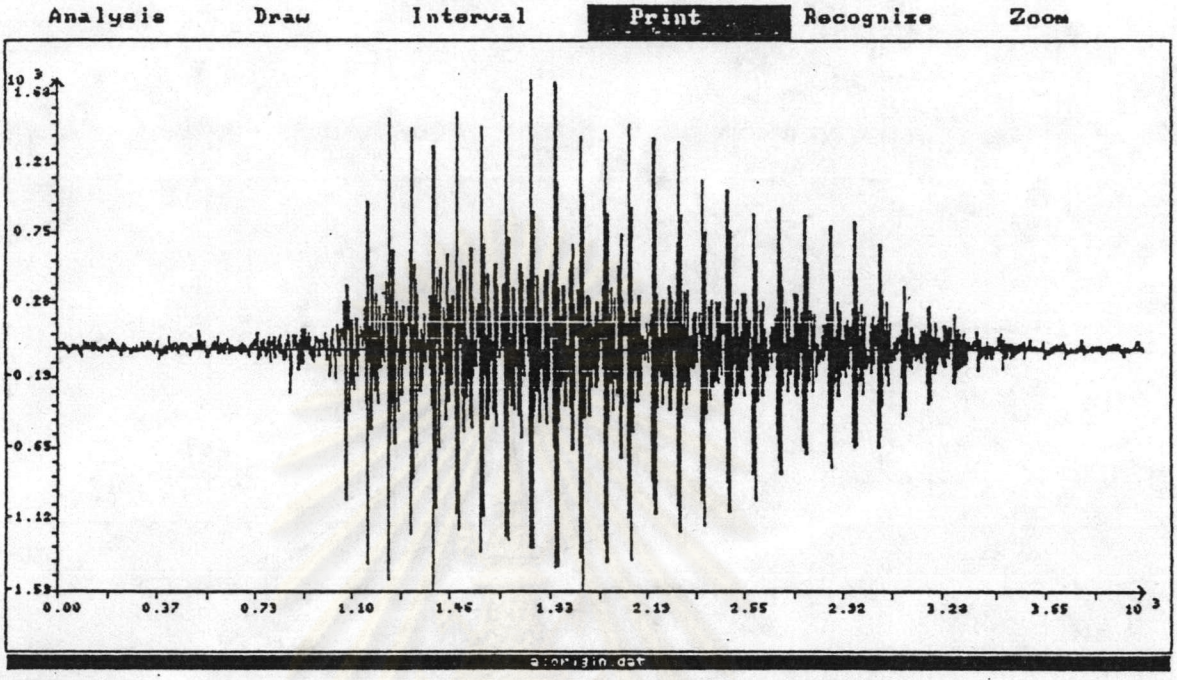
ค่าซิกม่า	ใช้เนื้อที่ 2 ไบท์	
ค่าคาบของพิทช์	ใช้เนื้อที่ 1 ไบท์	
ค่าสัมประสิทธิ์การสะท้อน 10 ตัว	ใช้เนื้อที่ 40 ไบท์	(ตัวละ 4 ไบท์)
รวมเป็นเนื้อที่ทั้งหมดหลังการวิเคราะห์	43 ไบท์	

นั่นคือใช้เนื้อที่ลดลงคิดเป็นร้อยละ 89.25 หรือลดลงจากเดิมประมาณ 10 เท่าเลยทีเดียว

	ชื่อพยางค์	จำนวนเฟรม	เวลา
1.	หนึ่ง	12	240 mS
2.	สอง	16	320 mS
3.	สาม	18	360 mS
4.	สี่	17	340 mS
5.	ห้า	17	340 mS
6.	หก	8	160 mS
7.	เจ็ด	9	180 mS
8.	แปด	13	260 mS
9.	เก้า	14	280 mS
10.	สิบ	8	160 mS
11.	เอ็ด	8	160 mS
12.	ยี่	18	360 mS
13.	ร้อย	17	340 mS
14.	พัน	13	260 mS
15.	หมื่น	19	380 mS
16.	แสน	18	360 mS
17.	ล้าน	14	280 mS
18.	จุด	7	140 mS
19.	บาด	15	300 mS
20.	สะ	13	260 mS
21.	ตาง	14	280 mS
22.	รา	16	320 mS
23.	คา	19	380 mS
24.	สุน	20	400 mS
25.	ขึ้น	13	260 mS
26.	ลง	17	340 mS
27.	เปิด	15	300 mS
28.	ปิด	7	140 mS

ตารางที่ 4.1 รายละเอียดพยางค์ตัวอย่างที่ใช้ในพจนานุกรมข้อมูล

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

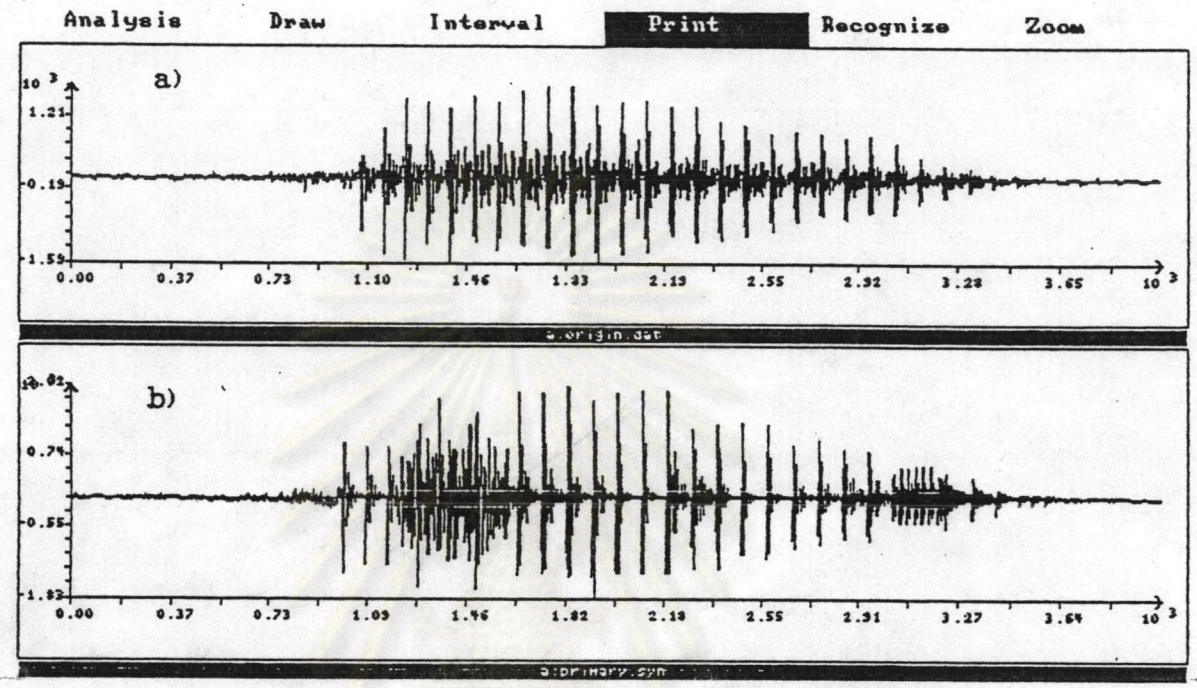


รูปที่ 4.1 รูปสัญญาณเสียงต้นแบบของพยางค์ "สอง"

4.2 ส่วนของการแก้ไขพารามิเตอร์

จากพารามิเตอร์ที่ได้มาจากขั้นตอนวิเคราะห์เสียงพูดนี้ ถ้านำมาผ่านกระบวนการสังเคราะห์เสียงให้กลับออกมาเป็นเสียงเลย จะได้เสียงพยางค์ที่ไม่ชัดหรือบางทีอาจจะผิดเพี้ยนไปมากจนฟังไม่ออกว่าเป็นพยางค์อะไร ทั้งนี้เนื่องมาจากอิทธิพลระหว่างเสียงของพยัญชนะกับเสียงสระภายในแต่ละพยางค์นั้นส่งผลรบกวนกัน หรืออาจจะเป็นเสียงของสระกับเสียงตัวสะกดกวนกันก็ได้ ซึ่งผลเหล่านี้จะทำให้ค่าพารามิเตอร์ที่ได้ออกมาในช่วงของการวิเคราะห์เสียงมีค่าที่ผิดพลาดไปบ้าง พารามิเตอร์ที่เห็นได้ชัดถึงผลอันนี้ก็คือ ค่าคาบของพิทช์ จากตัวอย่างในรูปที่ 4.2 ซึ่งเป็นรูปของสัญญาณเสียงของพยางค์ "สอง" ที่พล็อตตามแกนเวลา รูปบนเป็นรูปของสัญญาณเสียงต้นแบบ ส่วนรูปล่างเป็นสัญญาณเสียงที่ได้จากการสังเคราะห์เสียงจากพารามิเตอร์ที่ได้ออกมาจากการวิเคราะห์เสียงโดยตรง ไม่มีการแก้ไขใดๆทั้งสิ้น รูปที่ 4.3 จะแสดงค่าของพารามิเตอร์ที่พิจารณาคือ ค่าซิกม่า และค่าคาบของพิทช์ ในแต่ละเฟรมของพยางค์ "สอง" เมื่อพิจารณาควบคู่กับรูปที่ 4.3 ทางขวามือจะเห็นได้ว่าค่าคาบของพิทช์ในเฟรมที่ 7 , 8 และ 16 มีค่าที่ผิดไปจากที่ควรจะเป็น เช่น เฟรมที่ 7 , 8 มีค่าเป็นศูนย์หรือกลายเป็นเฟรมที่เป็นเสียงไม่ก้อง ทั้งๆที่ถูกแล้วเมื่อเทียบกับสัญญาณเสียงต้นแบบควรจะต้องเป็นเฟรมที่เป็นเสียง

ก้อง และมีค่าคาบของพิทช์ที่ใกล้เคียงกับเฟรมข้างๆด้วย ส่วนเฟรมที่ 16 ก็มีค่าคาบของพิทช์ที่ผิดไปมากจากต้นแบบ



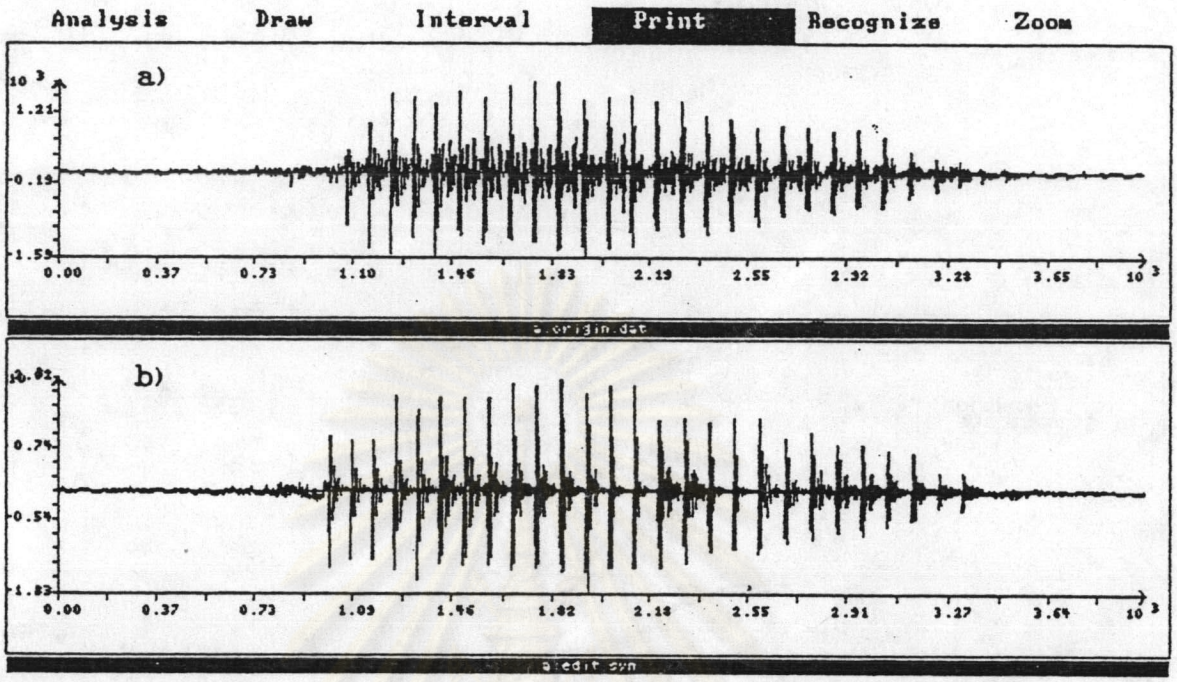
รูปที่ 4.2 รูปสัญญาณเสียงพยางค์ "สอง" ต้นแบบ (a) กับที่สังเคราะห์ได้ (b) จากพารามิเตอร์ก่อนแก้ไข

เมื่อทำการแก้ไขให้ค่าคาบของพิทช์ในเฟรมที่ 7 , 8 และ 16 ให้เป็นค่าที่เหมาะสมตามค่าพารามิเตอร์ด้านขวามือในรูปที่ 4.3 แล้ว นำพารามิเตอร์ชุดใหม่มาผ่านการสังเคราะห์เสียงใหม่ สัญญาณเสียงที่ได้ออกมาจะเป็นดังรูปที่ 4.4 ซึ่งจะเห็นได้ว่าใกล้เคียงกับสัญญาณเสียงต้นแบบมากขึ้น การตรวจสอบความใกล้เคียงของเสียงสัญญาณต้นแบบกับสัญญาณเสียงที่สังเคราะห์ขึ้นมาจากชุดพารามิเตอร์ที่แก้ไขแล้วนั้น นอกจากจะดูจากรูปสัญญาณในแกนเวลาดังรูปที่ 4.4 แล้วยังสามารถพิจารณาในด้านของแกนความถี่ก็ได้ โดยพิจารณาจากสเปคตรัมของสัญญาณเสียงทั้งสองซึ่งจะได้ดังรูป 4.5 รูปบนเป็นรูปสเปคตรัมของสัญญาณเสียงต้นแบบ รูปล่างเป็นรูปสเปคตรัมของสัญญาณเสียงที่สังเคราะห์ขึ้นมาจากพารามิเตอร์ที่ได้ทำการแก้ไขแล้ว นอกจากนั้นวิธีในการตรวจสอบความใกล้เคียงของสัญญาณเสียงที่ได้จากการสังเคราะห์ได้ดีที่สุด นั่นก็คือใช้การนำสัญญาณเสียงทั้งสองมาผ่าน D/A ออกลำโพงแล้วทดลองฟังเสียงเปรียบเทียบกันว่าเหมือนหรือแตกต่างกันอย่างไร

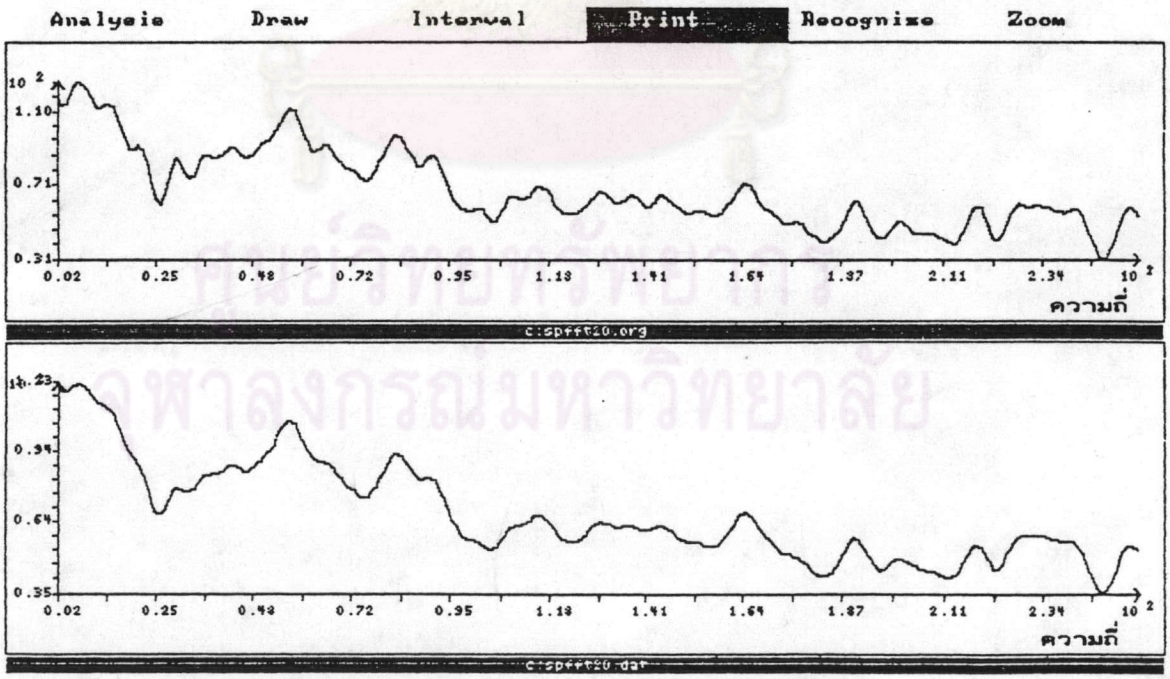
```
**************************
indxtmp.label ="  lek2samp"
indxtmp.offset=0
indxtmp.length=863
==========================
 head.K (number of frames) = 20
 head.N (number of points) = 200
 head.M (number of orders) = 10
 -------------------------
  perframe.sigma = 16
  perframe.pitch=0
  perframe.sigma = 20
  perframe.pitch=0
  perframe.sigma = 25
  perframe.pitch=0
  perframe.sigma = 30
  perframe.pitch=0
  perframe.sigma = 43
  perframe.pitch=0
  perframe.sigma = 85
  perframe.pitch=80
  perframe.sigma = 138
  perframe.pitch=0
  perframe.sigma = 85
  perframe.pitch=0
  perframe.sigma = 120
  perframe.pitch=88
  perframe.sigma = 123
  perframe.pitch=93
  perframe.sigma = 168
  perframe.pitch=92
  perframe.sigma = 70
  perframe.pitch=92
  perframe.sigma = 100
  perframe.pitch=95
  perframe.sigma = 42
  perframe.pitch=95
  perframe.sigma = 31
  perframe.pitch=92
  perframe.sigma = 37
  perframe.pitch=28
  perframe.sigma = 18
  perframe.pitch=87
  perframe.sigma = 14
  perframe.pitch=47
  perframe.sigma = 11
  perframe.pitch=0
  perframe.sigma = 10
  perframe.pitch=0
```

```
**************************
indxtmp.label ="  lek2edit"
indxtmp.offset=0
indxtmp.length=863
==========================
 head.K (number of frames) = 20
 head.N (number of points) = 200
 head.M (number of orders) = 10
 -------------------------
  perframe.sigma = 16
  perframe.pitch=0
  perframe.sigma = 20
  perframe.pitch=0
  perframe.sigma = 25
  perframe.pitch=0
  perframe.sigma = 30
  perframe.pitch=0
  perframe.sigma = 43
  perframe.pitch=0
  perframe.sigma = 85
  perframe.pitch=80
  perframe.sigma = 138
  perframe.pitch=83
  perframe.sigma = 85
  perframe.pitch=86
  perframe.sigma = 120
  perframe.pitch=88
  perframe.sigma = 123
  perframe.pitch=93
  perframe.sigma = 168
  perframe.pitch=92
  perframe.sigma = 70
  perframe.pitch=92
  perframe.sigma = 100
  perframe.pitch=95
  perframe.sigma = 42
  perframe.pitch=95
  perframe.sigma = 31
  perframe.pitch=92
  perframe.sigma = 37
  perframe.pitch=89
  perframe.sigma = 18
  perframe.pitch=87
  perframe.sigma = 14
  perframe.pitch=0
  perframe.sigma = 11
  perframe.pitch=0
  perframe.sigma = 10
  perframe.pitch=0
```

รูปที่ 4.3 แสดงค่าพารามิเตอร์ของพยางค์ "สอง" ที่ได้มาจากการวิเคราะห์เสียง



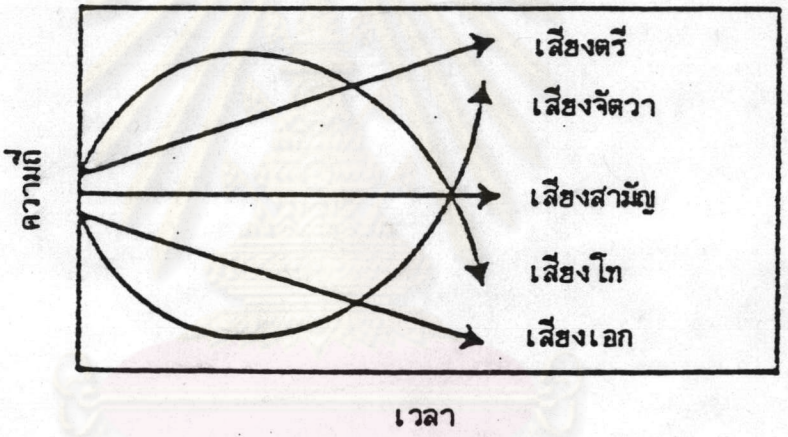
รูปที่ 4.4 รูปสัญญาณเสียงพยางค์ "สอง" ต้นแบบ (a) กับที่สังเคราะห์ได้ (b) จากพารามิเตอร์หลังแก้ไข



รูปที่ 4.5 รูปแสดงสเปคตรัมของเฟรมที่ 20 ของสัญญาณเสียงต้นแบบ (a) และสัญญาณเสียงสังเคราะห์ (b)

หลักเกณฑ์ในการแก้ไขพารามิเตอร์ค่าคาบของพิทช์นั้นนอกจากจะใช้การพิจารณาดูจากค่าที่ควรจะเป็นจากค่าคาบของพิทช์ของเฟรมก่อนหน้ากับเฟรมที่ถัดไปแล้ว ยังสามารถจะพิจารณาดูจากแนวโน้มการเปลี่ยนแปลงค่าคาบของพิทช์ตามเสียงวรรณยุกต์ของพยางค์นั้นได้อีกด้วย จากเอกสารอ้างอิงของ Chatchavalit Saravari and Satoshi Imai ได้กล่าวถึงผลการทดลองของ Abramson เมื่อปี ค.ศ. 1975 ในเรื่องของเสียงวรรณยุกต์ในภาษาไทยว่าเสียงวรรณยุกต์จะแบ่งออกเป็นสองกลุ่มใหญ่ๆคือ

- กลุ่มของเสียงวรรณยุกต์ที่มีการแปรเปลี่ยน (dynamic tones)
  ได้แก่เสียงวรรยุกต์ โท (dropped tone) และ จัตวา (rising tone)
- กลุ่มของเสียงวรรณยุกต์ที่คงค่า (static tones)
  ได้แก่เสียงวรรณยุกต์ สามัญ (level tone) เอก (grave tone) และ ตรี (acute tone)



รูปที่ 4.6 แนวทางการเปลี่ยนแปรความถี่ของเสียงวรรณยุกต์ไทย

จากรูปที่ 4.6 จะแสดงให้เห็นถึงแนวทางการเปลี่ยนแปรความถี่ตามแกนเวลาของเสียงวรรณยุกต์ทั้ง 5 เสียง ที่พอจะใช้เป็นแนวทางในการพิจารณาค่าคาบของพิทช์(ซึ่งเป็นสัดส่วนผกผันกับค่าความถี่)ในเฟรมต่างๆได้ด้วย อย่างเช่น ในกรณีที่พยางค์มีเสียงวรรณยุกต์เป็นเสียงสามัญ ค่าคาบของพิทช์ก็ควรจะมีค่าค่อนข้างจะคงที่ในทุกๆเฟรม ส่วนพยางค์ที่มีเสียงวรรณยุกต์เป็นเสียงเอก ค่าคาบของพิทช์ก็ควรจะมีค่าเพิ่มขึ้นอย่างสม่ำเสมอจากเฟรมแรกจนถึงเฟรมสุดท้าย หรือถ้าพยางค์เป็นเสียงวรรณยุกต์จัตวา ก็ควรจะมีค่าคาบของพิทช์ค่อยๆเพิ่มขึ้นไปในเฟรมแรกๆแล้วลดลงอย่างรวดเร็วในเฟรมช่วงหลัง เป็นต้น

4.3 ส่วนของการสังเคราะห์เสียงพูด

หลังจากขั้นตอนการแก้ไขพารามิเตอร์แล้วก็จะได้ พจนานุกรมข้อมูลที่พร้อมจะใช้ในการสังเคราะห์เสียงพูด โดยชุดพารามิเตอร์ของพยางค์แต่ละพยางค์จะมี ตัวชี้ ชี้อยู่พร้อมกับมีป้ายชื่อบอกด้วยว่าชุดพารามิเตอร์ไหนตรงกับพยางค์อะไร

เมื่อเริ่มให้โปรแกรมดำเนินการ โปรแกรมจะรอรับอินพุทจากแป้นพิมพ์ (หรืออาจจะเป็นแฟ้มข้อมูลก็ได้) ซึ่งจะพิมพ์ข้อความที่เป็นคำอ่านภาษาไทยทีละพยางค์ที่มีอยู่ในพจนานุกรมข้อมูลลงไป โดยอาจจะรวมทั้งอักขระที่ใช้ในการเปลี่ยนแปลงลักษณะของเสียงที่จะสังเคราะห์ออกมาด้วย โปรแกรมจะไปค้นหาชุดพารามิเตอร์ของพยางค์นั้น ๆ ในพจนานุกรมข้อมูลขึ้นมาทำการสังเคราะห์เสียงได้เป็นข้อมูลของเสียงจากการสังเคราะห์เก็บเอาไว้ใน หน่วยความจำเพิ่มเติมของ TSB บอร์ดแล้วส่งออกไป D/A ออกสู่ลำโพงอีกที ช่วงเวลาในการสังเคราะห์เสียงนี้จะแปรตามประสิทธิภาพของเครื่องไมโครคอมพิวเตอร์ที่ใช้ และเนื่องจากตัวโปรแกรมมีการเรียกใช้และคำนวณตัวเลขแบบจำนวนจริง (floating-point) เป็นหลัก ดังนั้นการที่เครื่องไมโครคอมพิวเตอร์มีตัวประมวลผลช่วยทางด้านคณิตศาสตร์ (math co-processor) ด้วยก็จะทำให้ความเร็วในการสังเคราะห์เพิ่มขึ้นมากกว่า 10 เท่าเลยทีเดียว จากการทดลองเครื่องไมโครคอมพิวเตอร์ที่มีอยู่สามารถสรุปเวลาที่ใช้ในการสังเคราะห์ต่อข้อมูล 1 เฟรมได้ดังตารางที่ 4.2 นี้

เครื่อง IBM XT (8088,4.77MHz)	ไม่มี 8087	11.45 sec/frame
เครื่อง IBM XT (8088,4.77MHz)	มี 8087	1.10 sec/frame
เครื่อง IBM AT (80286,12MHz)	ไม่มี 80287	2.64 sec/frame
เครื่อง IBM AT (80286,12MHz)	emulation	0.28 sec/frame
เครื่อง IBM 386 , 25 MHz	ไม่มี 80387	1.24 sec/frame
เครื่อง IBM 386 , 25 MHz	emulation	0.065 sec/frame

ตารางที่ 4.2 เวลาที่ใช้ในการคำนวณสังเคราะห์เสียงกับเครื่องไมโครคอมพิวเตอร์ชนิดต่างๆ

จากข้อจำกัดของเครื่องไมโครคอมพิวเตอร์ที่ใช้ในการทดลองนี้ ซึ่งเป็นเครื่อง IBM PC-XT 4.77 MHz มีตัวประมวลผลช่วยทางด้านคณิตศาสตร์เบอร์ 8087 จะใช้เวลาในการสังเคราะห์เสียงประมาณ 1.10 วินาทีต่อเฟรม จึงไม่สามารถทำให้เป็นในลักษณะประมวลผลเวลาจริงได้ การสังเคราะห์เสียงพูดจึงจำเป็นต้องมีการหยุดรอการคำนวณ หลังจากที่ข้อมูลบน

หน่วยความจำเพิ่มเติมเต็ม 40 กิโลเวิร์ด (หรือเสียงพูดยาวประมาณ 4 วินาที) และข้อมูลถูกส่งออก D/A เป็นเสียงออกมาแล้วเพื่อทำการสังเคราะห์เสียงพยางค์ต่อไป

เนื่องจากเสียงต้นแบบที่ใช้นั้นตัดมาเฉพาะส่วนที่มีข้อมูลจริงๆ ดังนั้นเมื่อเอามาต่อกันให้เกิดเป็นคำโดยตรงอาจจะได้คำที่ฟังไม่ออกตามที่ควรจะเป็น ทั้งนี้เพราะในคำที่มีหลายพยางค์จะต้องมีช่วงว่างระหว่างพยางค์หรือช่วงจังหวะระหว่างพยางค์ด้วย โดยช่วงว่างนี้จะสั้นหรือยาวขึ้นอยู่กับคำแต่ละคำ ถ้าทิ้งช่วงว่างระหว่างพยางค์นานเกินไปก็จะฟังเหมือนกับพูดพยางค์เดี่ยวๆแยกจากกันคนละพยางค์ แต่ถ้าทิ้งช่วงว่างระหว่างพยางค์สั้นเกินไปเสียงทั้งสองพยางค์จะกลมกลืนกันจนฟังไม่ได้ใจความ ดังนั้นในส่วนนี้จึงอาจจะมีการใช้สัญญลักษณ์สำหรับการแก้ไขลักษณะของเสียงที่สังเคราะห์ออกมาได้ เพื่อทำการตกแต่งให้เสียงแต่ละพยางค์เมื่อนำมาต่อเรียงกันให้เกิดเป็นคำหรือวลีสั้นๆ หรือเป็นประโยคก็ตามมีความใกล้เคียงกันกับเสียงธรรมชาติหรือเพื่อให้เสียงสังเคราะห์ที่ได้ออกมาฟังได้เหมือนจริงมากยิ่งขึ้น เช่นการใช้สัญญลักษณ์ '/' เพื่อทิ้งช่วงว่างระหว่างพยางค์หรือคำเป็นเวลา 20 มิลลิวินาที และการเปลี่ยนแปลงค่าคาบของพิทช์เป็นต้น ทั้งนี้ขึ้นกับเสียงลงท้ายของพยางค์หน้าและเสียงขึ้นต้นของพยางค์ถัดไป และลักษณะการพูดของสัญญาณเสียงต้นแบบเช่นพูดเสียงสั้นหรือลากเสียงยาวด้วย โดยต้องอาศัยการฟังเสียงที่สังเคราะห์ออกมาได้และทำการปรับแต่งจนกว่าจะได้เสียงที่ฟังแล้วเหมือนจริงเป็นธรรมชาติที่สุดเท่าที่จะทำได้

หลักเกณฑ์ของการใช้สัญญลักษณ์พิเศษสำหรับกลุ่มพยางค์ตัวอย่างที่เก็บอยู่ในพจนานุกรมต้นแบบนี้พอจะยกขึ้นมาเป็นบางกรณีได้ดังเช่น

- พยางค์ที่ขึ้นต้นคือ "สิบ" ต้องเว้นช่องว่างก่อนที่จะให้เปล่งเสียงพยางค์ต่อไป โดยต้องคั่นด้วย '/' ประมาณ 4 ตัวถึงจะตามด้วยพยางค์ถัดไป ("เอ็ด ..... เก้า") ถ้าเว้นช่วงว่างระหว่างพยางค์ "สิบ" กับพยางค์ถัดไปน้อยกว่านี้จะทำให้การรับรู้โดยการฟังทำได้ยากขึ้น ตัวอย่างของข้อความที่ใช้เป็นอินพุทเช่น สิบ //// สาม , สิบ //// เอ็ด

- พยางค์ที่ตามหลังด้วยพยางค์ "สิบ" หรือ "ร้อย" หรือ "หมื่น" เว้นช่วงว่างโดยคั่นด้วย '/' 1 ตัว เช่น ยี่ / สิบ , หนึ่ง / ร้อย , ห้า / ร้อย

- พยางค์ที่ตามหลังด้วยพยางค์ "พัน" หรือ "แสน" เว้นช่วงว่างระหว่างพยางค์โดยคั่นด้วย '/' 2 ตัว สำหรับพยางค์ "แสน" ควรจะเปลี่ยนค่าคาบของพิทช์ให้เพิ่มขึ้นอีก 10 จะได้เสียงที่ฟังชัดเจนขึ้น เช่น หก // พัน , แปด // แสน+10

- พยางค์ที่ตามหลังด้วยพยางค์ "ล้าน" เว้นช่วงว่างระหว่างพยางค์ข้างหน้าโดยคั่นด้วย '/' 3 ตัว และควรจะเปลี่ยนค่าคาบของพิทช์ของพยางค์ "ล้าน" ให้เพิ่มขึ้นอีก 5 ด้วย เช่น สี่ /// ล้าน+5

- ช่วงว่างระหว่างคำหรือระหว่างวลี ควรจะเว้นช่วงว่างโดยการคั่นด้วย '/' ตั้งแต่ 5 ตัวขึ้นไป

ตัวอย่างของข้อความภาษาไทยพร้อมทั้งสัญญลักษณ์พิเศษเพื่อให้ได้คำหรือวลีที่ฟังออกได้ชัดเจนที่ใช้ในการทดสอบ เช่น

สิบ //// เจ็ด

เจ็ด / สิบ //// สะ / ตาง

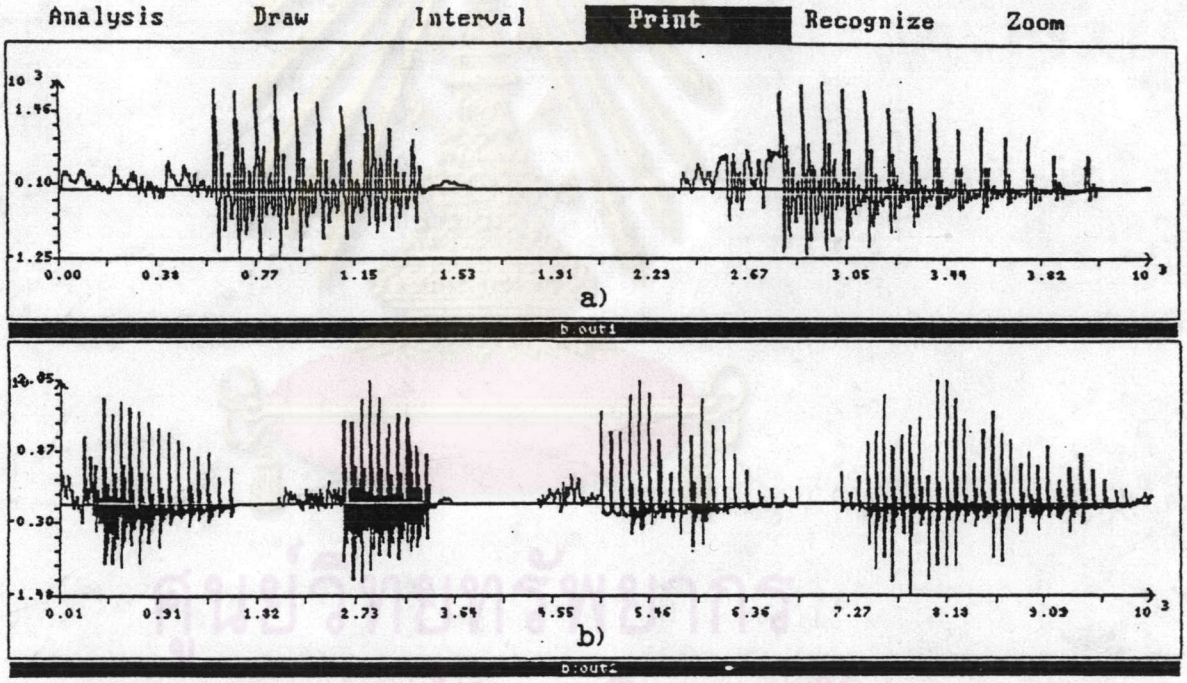
ยี่ / สิบ //// เอ็ด+5

ห้า /// ล้าน+5 ///// บาด

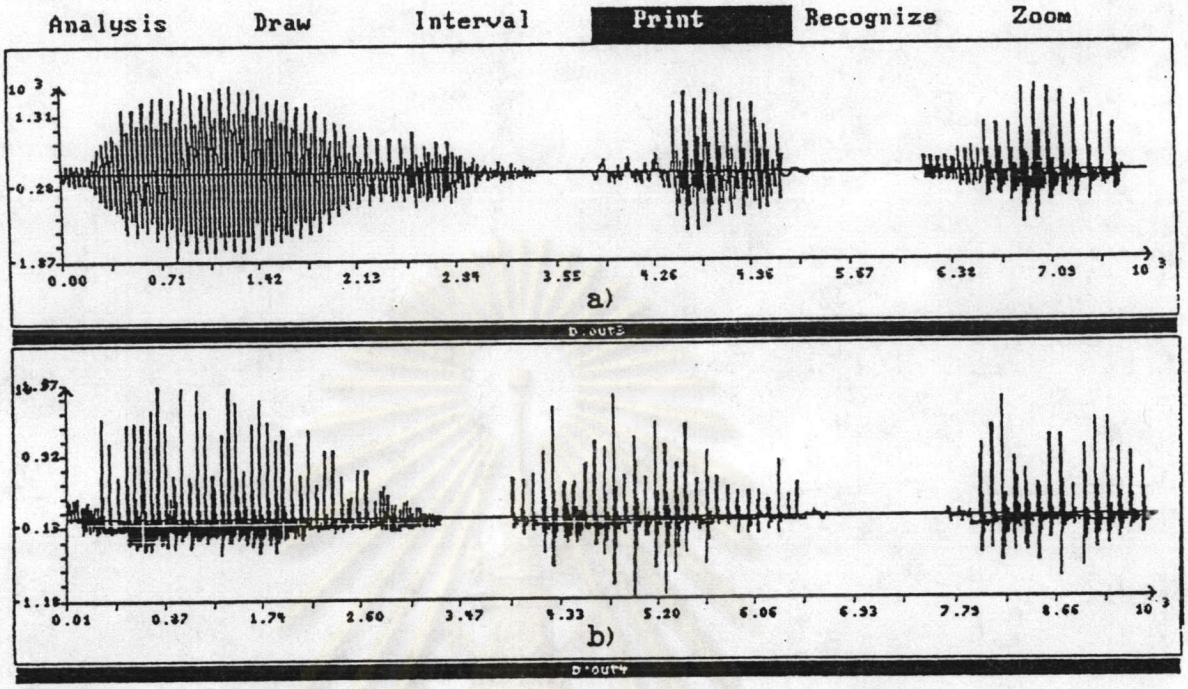
รา-5 คา-5 / ขึ้น+5

รา-5 คา-5 / เปิด+5

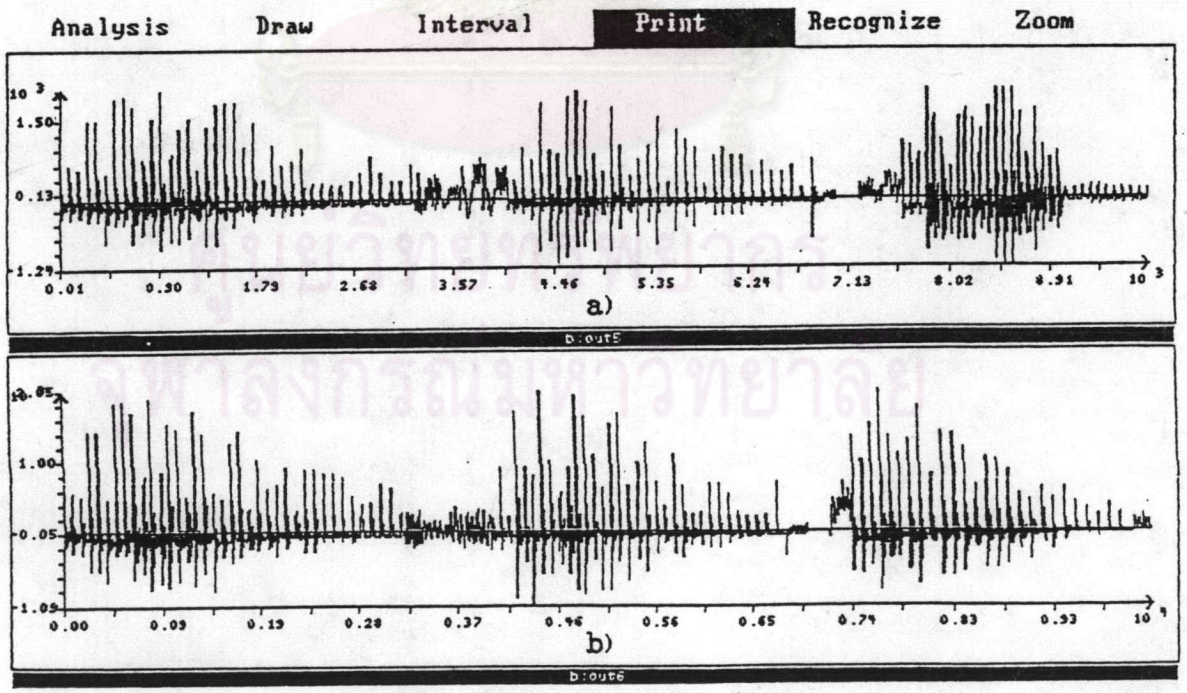
ซึ่งมีรูปสัญญาณตามรูปที่ 4.7 , 4.8 และ 4.9 ตามลำดับ



รูปที่ 4.7 รูปสัญญาณของวลี "สิบเจ็ด" (a) และ "เจ็ดสิบสตางค์" (b)



รูปที่ 4.8 รูปสัญญาณของวลี "ยี่สิบเอ็ด" (a) และ "ห้าล้านบาท" (b)



รูปที่ 4.9 รูปสัญญาณของวลี "ราคาขึ้น" (a) และ "ราคาเปิด" (b)

ผลการทดสอบ

จากข้อมูลเสียงพยางค์ต่างๆที่นำมาใช้ทดสอบเป็นจำนวนทั้งสิ้น 28 พยางค์ เมื่อนำมาทำการทดสอบการรับรู้ทางด้านการฟัง โดยเชิญผู้ทดสอบจำนวน 20 คนมาทดลองฟังเสียงที่สังเคราะห์ขึ้นมาในสองรูปแบบคือ

- ใช้พยางค์เดี่ยว จำนวน 22 พยางค์ (พยางค์จากตารางที่ 4.1 ที่มีความหมาย)
- ใช้พยางค์หลายๆพยางค์ต่อกันเป็นวลี (ยาวไม่เกิน 4 วินาที) จำนวน 22 วลี ดังนี้

1. สิบเจ็ด
2. เจ็ดสิบสะตาง
3. ยี่สิบเอ็ด
4. ห้าล้านบาด
5. สองบาดห้าสิบสะตาง
6. ขึ้นราคา
7. ลงราคา
8. ราคาขึ้น
9. ราคาลง
10. ราคาเปิด
11. ราคาปิด
12. สูนบาดสูนสะตาง
13. สูนจุดสูนสูน
14. ขึ้นลง
15. เปิดปิด
16. สองพันห้าร้อยสามสิบสาม
17. หนึ่งพันเก้าร้อยเก้าสิบ
18. สี่ร้อยหกสิบเก้าจุดสูนสอง
19. ราคาเปิดสองร้อยเจ็ดสิบเอ็ดบาดสิบสะตาง
20. ราคาปิดสองร้อยห้าบาดเจ็ดสะตาง
21. ราคาลงหกสิบหกบาดสามสะตาง
22. เจ็ดล้านหนึ่งแสนห้าหมื่นแปดพันยี่สิบเอ็ด

โดยที่แต่ละวลีมีการใส่สัญลักษณ์สำหรับแก้ไขลักษณะของเสียงที่สังเคราะห์ออกมาเพื่อให้เสียงที่ได้ออกมาใกล้เคียงเสียงจริงแล้ว

ผลการทดสอบของแต่ละรูปแบบได้ออกมาดังนี้

– แบบคำพยางค์เดี่ยว

	ชื่อพยางค์	จำนวนผู้ที่รับรู้ได้ถูกต้อง	คิดเป็นร้อยละ
1.	หนึ่ง	20	100
2.	สอง	20	100
3.	สาม	20	100
4.	สี่	20	100
5.	ห้า	20	100
6.	หก	20	100
7.	เจ็ด	20	100
8.	แปด	20	100
9.	เก้า	20	100
10.	สิบ	20	100
11.	ร้อย	20	100
12.	พัน	20	100
13.	หมื่น	20	100
14.	แสน	20	100
15.	ล้าน	20	100
16.	จุด	18	90
17.	บาด	19	95
18.	สูน	20	100
19.	ขึ้น	19	95
20.	ลง	20	100
21.	เปิด	20	100
22.	ปิด	20	100

ตารางที่ 4.3 ผลการรับรู้แบบคำพยางค์เดี่ยว

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

- แบบคำพูดหลายพยางค์

	คำพูด	จำนวนผู้ที่รับรู้ได้ถูกต้อง	คิดเป็นร้อยละ
1.	สิบเจ็ด	20	100
2.	เจ็ดสิบสะตาง	20	100
3.	ยี่สิบเอ็ด	20	100
4.	ห้าล้านบาด	20	100
5.	สองบาดห้าสิบสะตาง	20	100
6.	ขึ้นราคา	20	100
7.	ลงราคา	20	100
8.	ราคาขึ้น	20	100
9.	ราคาลง	20	100
10.	ราคาเปิด	20	100
11.	ราคาปิด	20	100
12.	สูนบาดสูนสะตาง	20	100
13.	สูนจุดสูนสูน	19	95
14.	ขึ้นลง	20	100
15.	เปิดปิด	20	100
16.	สองพันห้าร้อยสามสิบสาม	20	100
17.	หนึ่งพันเก้าร้อยเก้าสิบ	20	100
18.	สี่ร้อยหกสิบเก้าจุดสูนสอง	20	100
19.	ราคาเปิดสองร้อยเจ็ดสิบเอ็ดบาดสิบสะตาง	18	90
20.	ราคาปิดสองร้อยห้าบาดเจ็ดสะตาง	20	100
21.	ราคาลงหกสิบหกบาดสามสะตาง	20	100
22.	เจ็ดล้านหนึ่งแสนห้าหมื่นแปดพันยี่สิบเอ็ด	18	90

ตารางที่ 4.4 ผลการรับรู้แบบคำพูดหลายพยางค์

จากผลการทดสอบที่ได้จะเห็นได้ว่า ในกรณีคำพูดหลายพยางค์การรับรู้จากการฟังจะได้ผลความถูกต้องมากกว่าแบบคำพยางค์เดี่ยว ทั้งนี้เนื่องจากแบบคำพูดหลายพยางค์จะเป็นวลีที่พอ

จะมีความหมายอยู่บ้าง และความหมายจะอยู่ภายในวงจำกัดคือเป็นการบอกจำนวนตัวเลขเป็นหลัก ทำให้ผู้ฟังมุ่งเน้นความสนใจมาพิจารณาตีความหมายเฉพาะที่เกี่ยวข้องกับจำนวนหรือตัวเลขเพียงอย่างเดียวเท่านั้น และอาจจะอาศัยพยางค์ข้างๆที่รับรู้ได้เป็นสิ่งช่วยในการตีความหมายและจับใจความคำพูดทั้งวลีได้ง่ายขึ้น จึงสามารถที่จะคาดคะเนวลีที่ได้ยินได้ถูกต้องมากกว่าคำพยางค์เดี่ยวๆ จากตารางที่ 4.4 คำพูดวลีที่ 13, 19 และ 22 จะมีผลการรับรู้ได้ถูกต้องน้อยกว่าคำอื่นๆ อาจจะพอสรุปได้ว่า วลีที่ 13 มีแต่พยางค์ที่มีเสียงสั้นและเสียง 'ส' ของคำว่า "สูน" ในส่วนที่เป็นเสียงไม่ก้องไม่ค่อยมีความชัดเจนนักอาจจะถูกเสียงสระข้างๆรบกวนถ้าไม่ตั้งใจฟังอาจจะตีความหมายผิดพลาดไป ส่วนวลีที่ 22 นั้นเป็นวลีที่มีจำนวนพยางค์ค่อนข้างมากถึง 11 พยางค์ต่อเนื่องกัน แต่เนื่องจากความสามารถในการจดจำของคนในช่วงระยะเวลาสั้นๆ (short term memory) มีขีดจำกัด ตัวอย่างเช่นการจดจำหมายเลขโทรศัพท์ถ้าเราบอกเป็นช่วงๆเช่น 3 ตัวแรกตามด้วย 4 ตัวหลังผู้ฟังจะรู้สึกจดจำได้ง่ายกว่าการที่เราบอกรวดเดียว 7 ตัวเลย ซึ่งก็เช่นเดียวกันกับวลีที่ยาวๆผู้ฟังอาจจะจดจำคำพูดทั้งวลีได้ยากหรืออาจจะเป็นในลักษณะได้หน้าลืมหลังเป็นต้น จึงทำให้ผลการรับรู้ของวลีที่ 19 และ 22 นี้น้อยกว่าวลีอื่นๆ ส่วนในกรณีคำพยางค์เดี่ยวผู้ฟังจะรับรู้ได้ยากกว่านอกจากจะพอรู้ขอบเขตของคำบ้างว่าเกี่ยวข้องกับอะไร โดยเฉพาะพยางค์ที่มีเสียงสั้นจะเห็นว่าผลการรับรู้มีโอกาสที่จะผิดพลาดหรือฟังไม่ออกได้มากกว่าพยางค์ที่มีเสียงยาว ในการฟังครั้งแรกผู้ฟังอาจจะยังตีความไม่ออก ต้องฟังซ้ำครั้งที่สอง หรือสามถึงจะพอตีความได้ว่าเป็นคำว่าอะไร จากผลอันนี้อาจจะนำไปใช้แก้ไขการเก็บสัญญาณเสียงต้นแบบให้มีช่วงเวลาที่ยาวนานขึ้นหรืออาจจะใส่สัญญลักษณ์เพื่อเพิ่มคาบของพิทช์ให้มากขึ้นสำหรับพยางค์ที่มีเสียงสั้นๆ ในขั้นตอนที่จะสังเคราะห์เสียงเพื่อจะได้เสียงที่สังเคราะห์ออกมาฟังได้ง่ายขึ้น

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย