บทที่ 6.

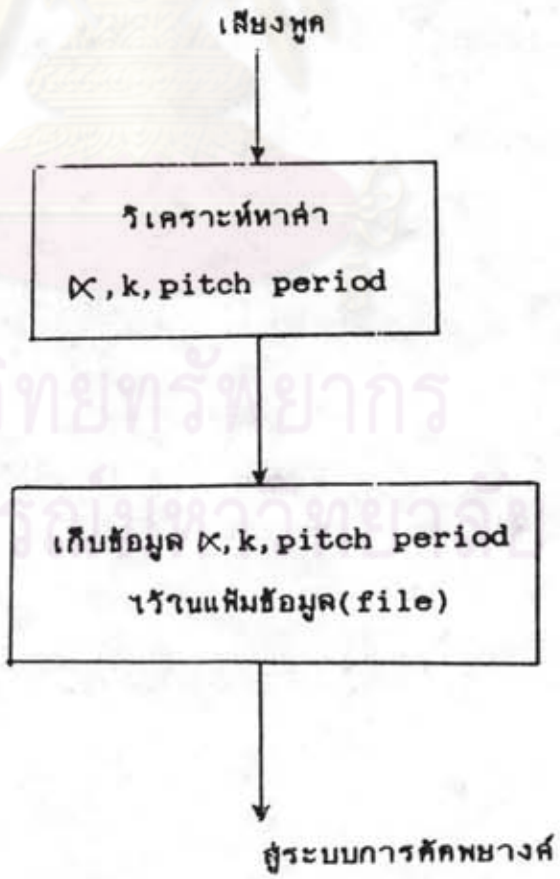
ระบบในการคัดพยางค์

6.1 ผังงานของระบบ

ซอฟท์แวร์(software) ของระบบที่จะใช้ในการคัดพยางค์นี้ได้แบ่งเป็น 3 ส่วนใหญ่ ๆ ส่วนที่ 1 เป็นส่วนที่เกี่ยวกับการวิเคราะห์ (analysis) เพื่อหาค่าพารามิเตอร์(parameter)ของเสียงเก็บไว้ซึ่งจะเป็นส่วนสำคัญในการตรวจรู้เสียงพูด และส่วนที่ 2 จะเป็นส่วนของการนำค่าพารามิเตอร์บางค่ามาใช้ในการคัดพยางค์ซึ่งพารามิเตอร์ที่จะใช้ในการคัดพยางค์นี้ มีอยู่ 2 อย่าง คือ ค่าพลังงานและ pitch ส่วนที่ 3 จะเป็นส่วนของการให้เครื่องตรวจรู้คำพูด

6.1.1 ผังงานของระบบการวิเคราะห์

ผังงานของระบบในส่วนนี้มีโครงสร้างดังนี้

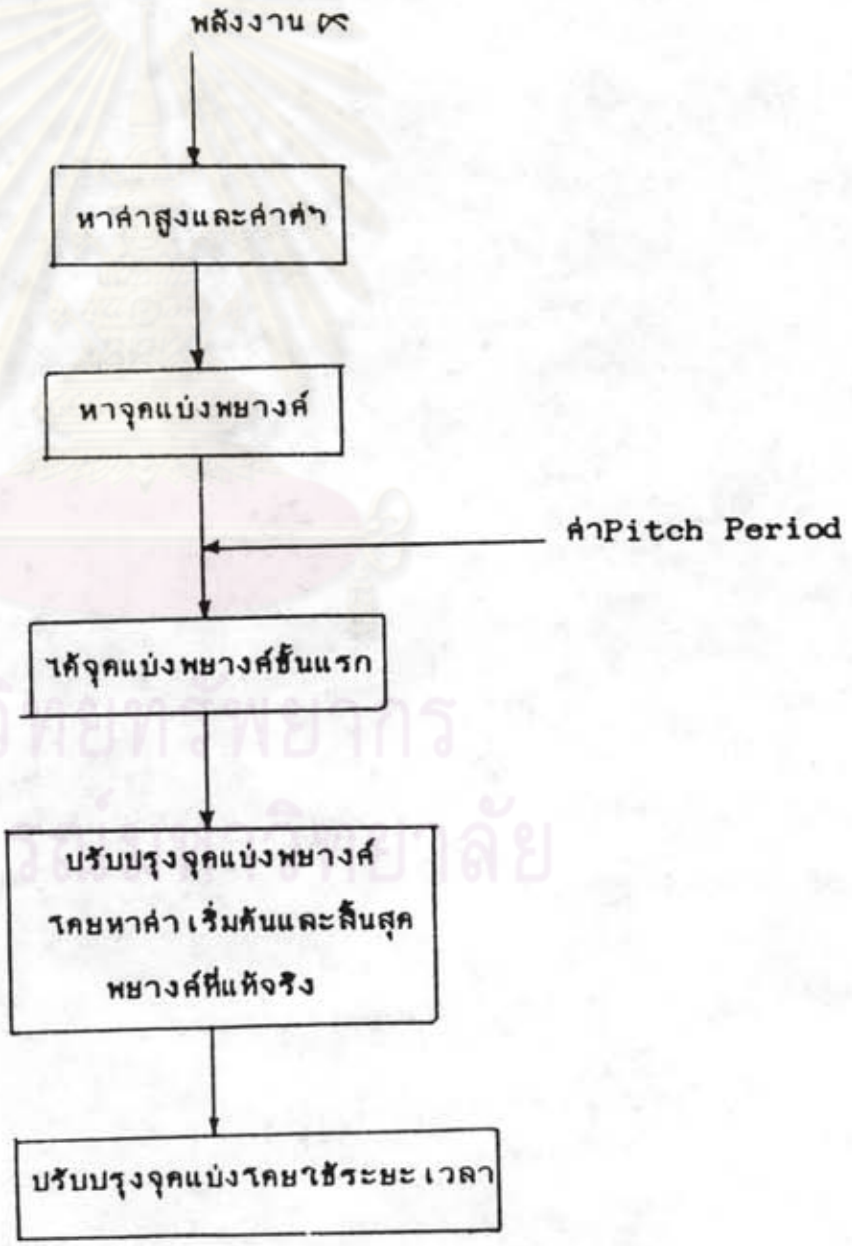


ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ในขั้นตอนของการวิเคราะห์นี้ ส่วนที่รับเข้ามา (input) คือเสียงพูด และใช้ช่วงกว้างของการวิเคราะห์ (window size) ครั้งละ 25 msec โดยการวิเคราะห์จะเคลื่อนไป (framing rate) ครั้งละ 5 msec ส่วนลำดับ (order) ที่ใช้ในการวิเคราะห์ใช้จำนวน 15 ค่า ผลลัพธ์ (output) ของขั้นตอนนี้คือค่าพารามิเตอร์ พลังงาน α, reflection coefficient, pitch period ซึ่งจะเก็บเป็นแฟ้มข้อมูลและส่งผ่านไปยังขั้นตอนที่ 2

6.1.2 <u>ผังงานของระบบการสังเคราะห์</u>

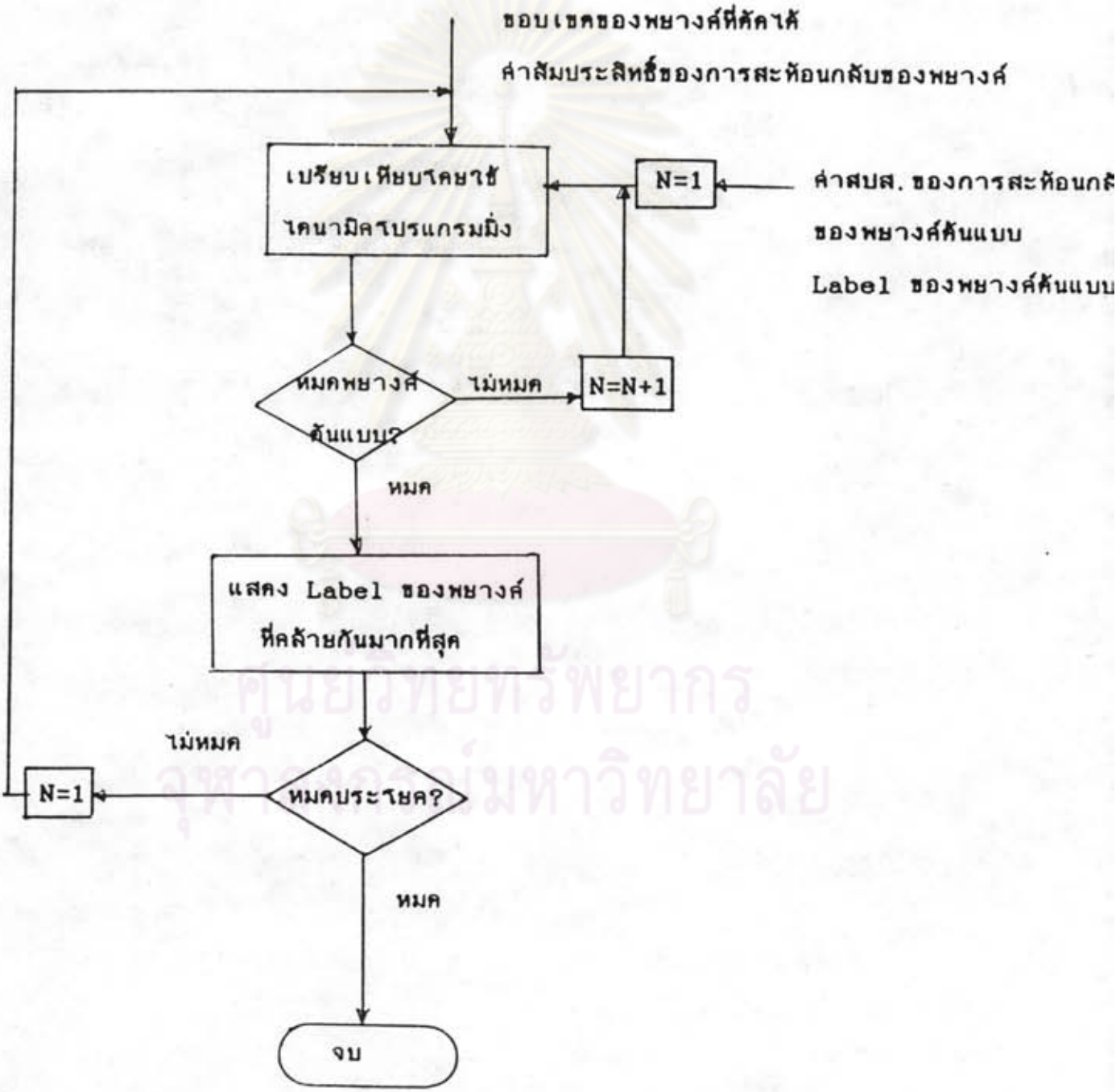
ผังงานของระบบในส่วนนี้ มีโครงสร้างดังนี้



ในขั้นตอนของการตัดพยางค์นี้ ส่วนที่รับเข้ามา (input) คือ ค่าพลังงาน α และpitch period โดยจะใช้ค่าพลังงานเป็นหลักในการตัดพยางค์ และจะใช้ pitch period กับระยะเวลา (duration) เพื่อช่วยให้การตัดพยางค์ได้ผลถูกต้องมากขึ้น ส่วนผลลัพธ์ของขั้นตอนนี้ ก็คือขอบเขตของพยางค์ที่ได้ตัดออกมา

6.1.3 <u>ผังงานของระบบการตรวจรู้</u>

ผังงานของระบบในส่วนนี้มีโครงสร้างดังนี้



ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ส่วนของขั้นตอนการตรวจรู้นี้ เป็นการใช้เทคนิคของไดนามิคโปรแกรมมิ่ง เปรียบเทียบระหว่างค่าสัมประสิทธิ์ของการสะท้อนกลับ ของพยางค์ที่ตัดได้กับพยางค์ที่เป็นต้นแบบ โดยถ้าพยางค์ใดมีค่าความแตกต่างน้อยที่สุด ก็จะนำ Label ของพยางค์ต้นแบบนั้นมาแสดงผล

6.2 วิธีการในการตัดพยางค์

วิธีการในการตัดพยางค์นั้น มีรายละเอียดดังนี้

6.2.1 การหาค่าสูงและค่าต่ำของค่าพลังงาน

วิธีการที่ใช้สำหรับขั้นตอนนี้ แบ่งเป็น 2 ส่วน คือส่วนต้นและส่วนท้ายของค่าพลังงาน กับส่วนอื่น ๆ

ถ้ากำหนดให้ x1, x2, x3 เป็นค่าพลังงาน 3 ค่าที่ต่อเนื่องกัน ตามลำดับ จะมีวิธีการที่จะพิจารณาว่าเป็นค่าสูงหรือต่ำได้ดังนี้

6.2.1.1 สำหรับส่วนต้นและท้ายของค่าพลังงาน มีวิธีการกำหนดโดยอาศัยค่าพลังงาน เพียง 2 ค่าคือ ในส่วนต้นจะอาศัยค่า x1 และ x2 โดยกำหนดว่า

1. ถ้า x1<x2 ตำแหน่งของ x1 จะเป็นค่าต่ำ
2. ถ้า x1>x2 ตำแหน่งของ x1 จะเป็นค่าสูง

และในส่วนท้ายจะอาศัยค่า x2 และ x3 โดยกำหนดว่า

1. ถ้า x2<x3 ตำแหน่งของ x3 จะเป็นค่าสูง
2. ถ้า x2>x3 ตำแหน่งของ x3 จะเป็นค่าต่ำ

6.2.1.2 สำหรับส่วนอื่น ๆ ของค่าพลังงาน จะหาค่าสูงหรือค่าต่ำได้โดยอาศัยค่าพลังงาน 3 ค่า ซึ่งกำหนดว่า

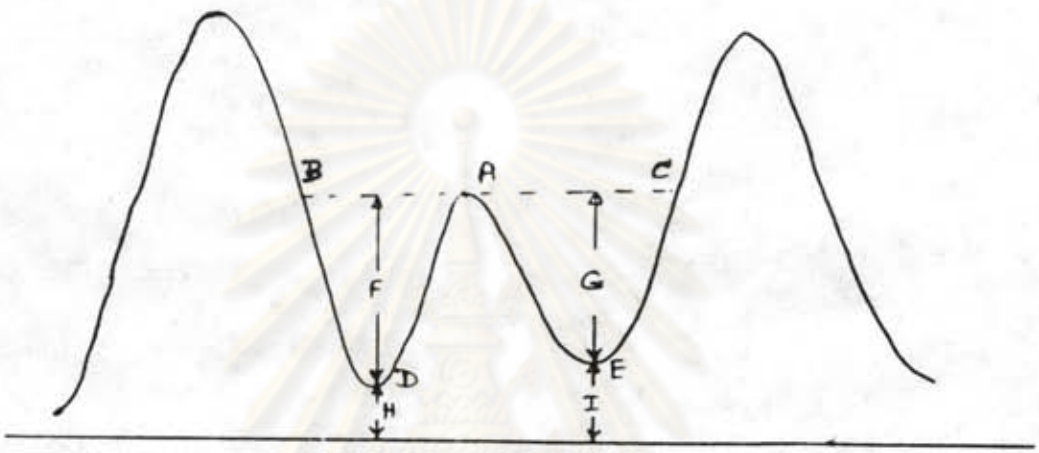
1. ถ้า x2>x1 และ x2>x3 ตำแหน่งของ x2 จะเป็นค่าสูง
2. ถ้า x2=x1 และ x3<x2 ตำแหน่งของ x2 จะเป็นค่าสูง
3. ถ้า x2>x1 และ x2=x3 ตำแหน่งของ x2 จะเป็นค่าสูง
4. ถ้า x2<x1 และ x2<x3 ตำแหน่งของ x2 จะเป็นค่าต่ำ

5. ถ้า x2<x1 และ x2=x3 ค่าแหน่งของ x2 จะเป็นค่าต่ำ

6. ถ้า x2=x1 และ x2<x3 ค่าแหน่งของ x2 จะเป็นค่าต่ำ

6.2.2 การหาจุดแบ่งของพยางค์

ในขั้นตอนจะหาจุดแบ่งโดยใช้วิธีการดังนี้[2]



รูปที่ 16 แสดงการตัดพยางค์

ณ จุดสูง (Maximum) ของทุก ๆ ค่า ซึ่งในตัวอย่างนี้ ก็คือจุด A เราจะมองไปทั้งทางด้านซ้ายและด้านขวา เพื่อหาตำแหน่งของค่าพลังงาน α ที่มีค่ามากกว่าค่าในตำแหน่ง A ซึ่งในที่นี้ด้านซ้ายมือคือตำแหน่ง B และด้านขวามือคือตำแหน่ง C หลังจากนั้นจะพิจารณาภายในช่วง A-B และในช่วง A-C เพื่อหาจุดต่ำสุด (local minimum.) ในช่วงนั้น ๆ ซึ่งในที่นี้คือ จุด D และ จุด E และหาผลต่างระหว่างค่า ณ ตำแหน่งสูง คือจุด A กับค่า ณ ตำแหน่งต่ำสุดในช่วงนั้น ซึ่งในที่นี้ด้านซ้ายคือ ตำแน่ง D ผลต่างคือระยะ F ด้านขวามือคือตำแน่ง E ผลต่างคือระยะ G และ ณ ตำแหน่งต่ำสุดในช่วงนั้น ๆ จะหาผลต่างกับค่าที่ตั้งไว้ ซึ่งผลต่าง ณ ตำแหน่ง D คือระยะ H และผลต่าง ณ ตำแหน่ง E คือระยะ I หลังจากนั้นจะพิจารณาว่าระยะ F มากกว่าค่าระดับบนที่ตั้งไว้หรือไม่ซึ่งในการทดลองนี้ได้ตั้งไว้ ณ ระดับ และพิจารณาว่าระยะ H น้อยกว่าค่าระดับล่างที่ตั้งไว้หรือไม่ ซึ่งในการทดลองนี้ตั้งไว้ที่ระดับ 10% ถ้าเป็นไปตามเงื่อนไขทั้ง 2 ข้อ ก็จะถือว่าตำแหน่งจุดต่ำ

สุดนั้น เป็นจุดแบ่งของพยางค์ ในขณะเดียวกัน ณ ตำแหน่ง E ก็ใช้หลักเกณฑ์นี้ในการพิจารณา เช่นเดียวกัน ซึ่งถ้าเขียนออกมาเป็นกฎตรรกศาสตร์ก็คือ

ถ้าระยะ G > ระดับบน และระยะ E < ระดับล่าง เราจะถือว่า ณ ตำแหน่งนั้นเป็นจุดแบ่งของพยางค์

6.2.3 การปรับปรุงโดยใช้ค่าความถี่มูลฐานมาช่วย

ในขั้นตอนนี้ เราจะพิจารณาจุดแบ่งของพยางค์เพิ่มจากเดิมโดยอาศัย เป็นตัวช่วยในการพิจารณา ทั้งนี้เพราะว่าลักษณะของค่าพลังงานที่ใช้ในการตัดพยางค์ นั้น เป็นค่า นอมอลไลซ์ (normalize) ด้วยค่าสูงสุดของประโยค ดังนั้น จึงมีค่าระยะ F บางค่าที่อยู่ต่ำกว่าค่าระดับบน ที่กำหนดไว้แต่เป็นขอบเขตของพยางค์ เราจึงต้องมาพิจารณาดูว่า ณ ตำแหน่งที่ค่าต่ำสุดในช่วงใด ๆ มีค่าน้อยกว่า ระดับล่าง ที่กำหนดไว้ แต่ระยะช่วงบนน้อยกว่าระดับบน ที่กำหนดไว้ เราก็จะพิจารณาดูค่าความถี่พื้นฐาน ณ เวลานั้นว่ามี ค่าความถี่พื้นฐาน หรือไม่ ถ้ามีก็จะไม่เป็นจุดแบ่งของพยางค์ แต่ถ้าไม่มีก็จะถือว่าตำแหน่งนั้น เป็นจุดแบ่งของพยางค์ด้วยเช่นกัน ซึ่งถ้าเขียนเป็นกฎเกณฑ์จะได้ว่า

ถ้าระยะบน(G)< ระดับบน และระยะล่าง(I) < ระดับล่าง และไม่มีค่าความถี่พื้นฐาน จะถือว่าเป็นจุดแบ่งของพยางค์

6.2.4 การปรับปรุงจุดแบ่งเพื่อหาจุดเริ่มต้นและจุดสิ้นสุดพยางค์ที่แท้จริง

หลังจากเราหาตำแหน่งของจุดแบ่งพยางค์ โดยวิธีดังกล่าวมาแล้ว บางตำแหน่งเป็นค่า ซึ่งยังไม่ใช่ขอบเขตของต้นพยางค์หรือท้ายพยางค์ที่แท้จริง จึงต้องปรับปรุงเพื่อหาตำแหน่งต้นพยางค์หรือท้ายพยางค์ที่แท้จริงให้ได้ โดยอาศัยหลักการที่ว่า ณ จุดแบ่งพยางค์ตำแหน่งใด ๆ ก็ตาม ถ้าค่าช้างเคียงด้านซ้ายมีค่าไม่มากกว่า 5 เท่าจะไม่ถือว่าเป็นขอบเขตของพยางค์ที่แท้จริงจะต้องเลื่อนจุดแบ่งนั้นมาด้านซ้าย แล้วพิจารณาจุดข้างเคียงซ้ายใหม่ ซึ่งถ้ามีค่ามากกว่า ก็จะถือว่าตำแหน่งนั้นเป็นจุดแบ่งพยางค์ที่แท้จริง และในขณะเดียวกันด้านขวามือของตำแหน่งนั้นก็ใช้กฎเกณฑ์ เช่นเดียวกัน

6.2.5 การปรับปรุงจุดแบ่งโดยอาศัยระยะเวลา

ในขั้นตอนนี้ เราจะพิจารณาจุดแบ่งพยางค์ที่ได้รับมา เพื่อดูว่า พยางค์แต่ละพยางค์ที่ได้แบ่งไว้นั้นมีระยะเวลา น้อยกว่าค่าที่กำหนดไว้หรือไม่ สำหรับการทดลองนี้ได้ตั้งค่าไว้ เป็นระยะเวลา 92.5 msec. (ตัวเลขนี้ได้มาจากผลการทดลอง) ทั้งนี้เพื่อป้องกันการแบ่งหน่วยเสียง(phoneme) เป็นหน่วยพยางค์ ถ้าหากว่าระยะเวลาของพยางค์ที่แบ่งไว้น้อยกว่าระยะเวลาที่กำหนด เราจะไม่พิจารณาว่าจุดแบ่งท้ายพยางค์ เป็นขอบเขตท้ายของพยางค์นั้น และจะพิจารณาขอบเขตต่อไป เพื่อให้ได้ขอบเขตท้ายพยางค์ที่แท้จริง

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

6.3 วิธีการในการตรวจรู้

เนื่องจากเสียงพูดของคนเรา แม้จะเป็นคนเดียวกันพูดคำ ๆ เดียวกันในแต่ละครั้งจะมีอัตราความเร็วเปลี่ยนแปลงไปได้ง่าย ดังนั้นจึงเป็นสาเหตุให้เกิดความไม่เป็นเชิงเส้น(nonlinear)ขึ้น ซึ่งเป็นปัญหาหลักอย่างหนึ่ง ในด้านการเปรียบเทียบเสียงพูดที่เป็นต้นแบบ กับเสียงพูดที่ต้องการจะให้เครื่องตรวจรู้

ได้มีความพยายามที่จะแก้ไขปัญหาในด้านนี้ โดยใช้เทคนิคที่เรียกว่า การแปลงเชิงเส้น(linear transformation) ซึ่งจากรายงานแสดงให้เห็นว่าเทคนิคชนิดนี้ใช้แก้ปัญหาดังกล่าวได้ไม่ดีนัก เนื่องจากเสียงมีการเปลี่ยนแปลง ที่ค่อนข้างมากและไม่มีความเป็นเชิงเส้น

แต่เทคนิคของไดนามิคโปรแกรมมิ่งนั้น จะกำจัดความแตกต่างทางด้านเวลาของรูปแบบ(pattern) 2 ชุด ได้โดยพยายามหาความคล้ายคลึงของรูปแบบทั้งสองให้มากที่สุด ซึ่งก็คือพยายามเปรียบเทียบรูปแบบโดยพิจารณาถึงความแตกต่างที่น้อยที่สุด(Minimize Residual Distance) ซึ่งจากบทความของ Sakoe และ Chiba ได้รายงานว่า วิธีการไดนามิคโปรแกรมมิ่งนี้ ได้นำไปใช้ในการตรวจรู้เสียงพูดตัวเลขของภาษาญี่ปุ่นได้ผลถูกต้องถึง 99.8%

หลักการของไดนามิคโปรแกรมมิ่ง

ก. การนอร์มอลไลซ์ระยะเวลา(Time-Normalize Distance)

เสียงพูดของคนเราสามารถแทนได้ด้วย ชุดของค่าพารามิเตอร์ ซึ่งค่าพารามิเตอร์นี้ จะมีลักษณะเป็นเวคเตอร์ สมมุติให้มีค่าพารามิเตอร์ของเสียงอยู่ 2 ชุด คือ

$$A = a_1, a_2, a_3, \ldots\ldots\ldots\ldots a_I$$

$$B = b_1, b_2, b_3, \ldots\ldots\ldots\ldots b_J$$

ซึ่งค่าพารามิเตอร์ แต่ละตัวจะเป็นเวคเตอร์ โดยในการวิจัยนี้ได้ใช้ค่า สัมประสิทธิ์ของการสะท้อนกลับ(Reflection Coefficient) เป็นพารามิเตอร์ที่จะใช้เปรียบเทียบโดยมี 15 อันดับ(order)

การที่เราจะหาความแตกต่างที่น้อยที่สุด ของเสียงทั้งสองชุดนี้ เรา

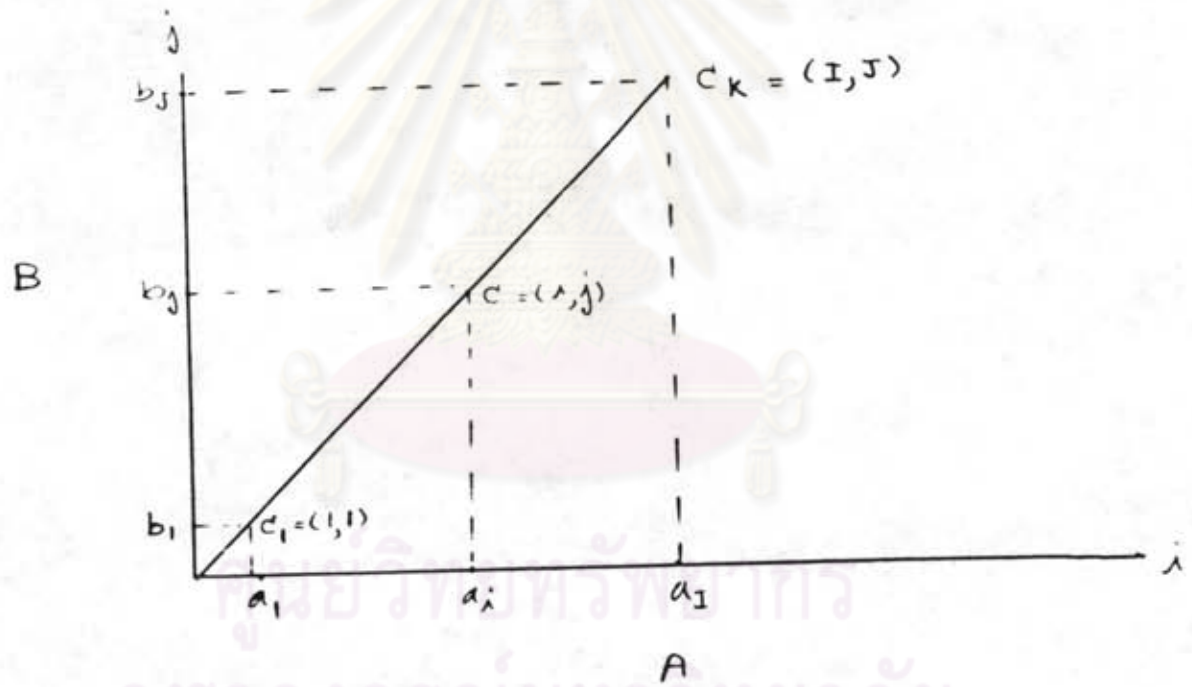
สามารถที่จะกำหนดเป็นฟังก์ชั่น ความแตกต่างด้านเวลา(Timing Difference) ระหว่างเสียงสองชุดได้โดย ให้เป็น c=(i,j) ซึ่งความแตกต่างของพารามิเตอร์ทั้งสองชุดจะเป็นชุดของ c โดย

$$F = c(1), c(2), c(3) \ldots\ldots c(k) \ldots\ldots\ldots c(K)$$

โดย $c(k) = (i(k), j(k))$

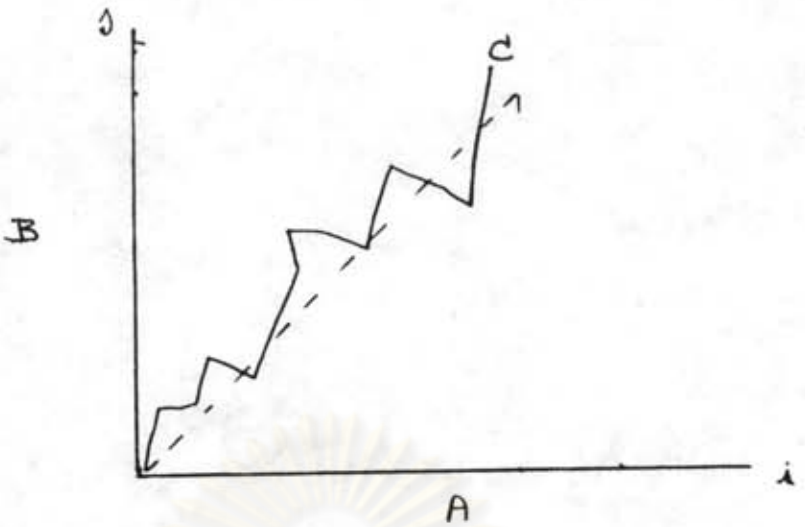
ชุดของ c ใน F ก็จะเป็นฟังก์ชั่น ซึ่งเป็นการแปลงจากรูปแบบ A ไปสู่รูปแบบ B หรือในทำนองเดียวกัน ก็อาจกล่าวได้ว่าเป็น Warping function

ถ้าไม่มีความแตกต่างด้านเวลา(Time-Difference) ของเสียงทั้งสองชุดแล้ว Warping function ที่ได้ก็จะเป็นแนวทะแยงตามเส้น j=i ดังรูปที่ 17



รูปที่ 17 แสดงทางเดินของ c ในกรณีที่พารามิเตอร์ทั้งสองชุดไม่มีความแตกต่างด้านเวล

แต่ถ้าพารามิเตอร์ของเสียงทั้งสองชุดนั้น มีความแตกต่างกันมากทางด้านเวลา Warping function ที่ได้ก็จะเบี่ยงเบนไปจากแนวทะแยงมากขึ้น ดังรูปที่ 18



รูปที่ 18 แสดงทางเดินของ c ในกรณีที่พารามิเตอร์ทั้งสองชุดมีความแตกต่างด้านเวลา

ถ้าเรากำหนดให้ ความแตกต่างระหว่างพารามิเตอร์ทั้งสองชุดเป็น d(c) โดย

$$d(c) = d(i,j) = \sum_{m=1}^{\text{order}=M} |a_{im} - b_{jm}|$$

เราจะได้ว่า ค่าความแตกต่างที่น้อยที่สุด ระหว่างเสียงชุด A และเสียงชุด B คือ

$$D(A,B) = \min_{F} \left[\frac{\sum_{k=1}^{K} d(c(k)).w(k)}{\sum_{k=1}^{K} w(k)} \right]$$

โดย w(k) คือ ค่าถ่วงน้ำหนักที่จะให้กับทางย่อยแต่ละ c(k) และ $\sum_{k=1}^{K} (w(k))$ ก็คือ ตัวที่จะนอร์มอลไลซ์ความแตกต่างในเส้นทางนั้น ๆ

สำหรับค่าถ่วงน้ำหนัก w(k) นี้ ถ้าไม่ขึ้นอยู่กับ Warping function F เราสามารถแยกออกมาอยู่นอกวงเล็บเป็น

$$D(A,B) = \frac{1}{N} \min_{F} \left[\sum_{k=1}^{K} d(c(k)).w(k) \right]$$

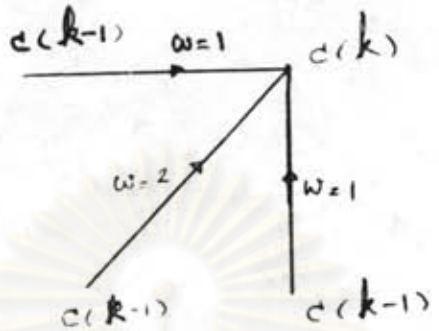
ซึ่ง N หรือสัมประสิทธิ์ของการนอร์มอลไลซ์(Normalize Coefficient) สามารถแยกพิจารณาได้ 2 แบบคือ แบบสมมาตร(Symetric form) และ แบบไม่สมมาตร(Asymetric form)

แบบสมมาตร จะได้ว่า

$$N = w(k) = (i(k)-i(k-1))+(j(k)-j(k-1))$$

ซึ่งจะได้ว่า N = I+J เป็นการให้ความสำคัญทั้งชุด A และชุด B โดยเสมือนกับว่าทั้ง A และ B ได้ถูกแปลงไปอยู่ในแกนขวางราว $l = i+j$ ซึ่งการ

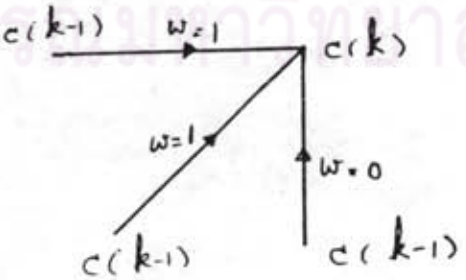
ถ่วงน้ำหนักในแบบนี้ เขียนเป็นไดอะแกรมได้ดังรูปที่ 19



รูปที่ 19 แสดงการถ่วงน้ำหนักแบบสมมาตร

แบบไม่สมมาตร จะได้ว่า

N = w(k) = (i(k)-i(k-1)) ซึ่งจะได้ว่า N = I เป็นการให้ความสำคัญกับรูปแบบของชุด A โดยเสมือนกับว่า B ถูกแปลงไปสู่ A หรือ N = w(k) = (j(k)-j(k-1)) ซึ่งจะได้ว่า N = J และเป็นการให้ความสำคัญกับรูปแบบของชุด B โดยเสมือนกับ A ถูกแปลงไปสู่ B ซึ่งการถ่วงน้ำหนักในแบบนี้เขียนเป็นไดอะแกรมได้ดังรูปที่ 20



รูปที่ 20 แสดงการถ่วงน้ำหนักแบบไม่สมมาตร

แต่จากการทดลองโดย Sakoe และ Chiba ได้ผลว่า การนอร์มอลไลซ์ โดยกำหนดให้เป็นแบบสมมาตรจะได้ผลถูกต้องมากกว่า

ข. เงื่อนไขของ Warping function

เพื่อให้การแปลงเสียงของชุด A และชุด B เป็นไปตามลักษณะของเสียงที่แท้จริง จึงจำเป็นต้องมีเงื่อนไขดังนี้

1. เงื่อนไขการเพิ่มขึ้น(Monotonic)

ในแง่ของฟิสิกส์แล้ว เสียงสองชุดที่จะเหมือนกันได้นั้น i(k) และ j(k) จะต้องมีค่าเพิ่มขึ้นเสมอ คือ

$$i(k-1) \leqslant i(k) \text{ และ } j(k-1) \leqslant j(k)$$

2. เงื่อนไขของการต่อเนื่อง(Continuity)

เงื่อนไขข้อนี้เพื่อกำหนดให้ c(k) มีค่าต่อเนื่องโดย

$$i(k)-i(k-1) \leqslant 1 \text{ และ } j(k)-j(k-1) \leqslant 1$$

3. เงื่อนไขของขอบเขต(Boundary Condition)

ก็คือขอบเขตเริ่มต้นและขอบเขตสุดท้ายของเสียงทั้งสองชุด ที่มาเปรียบเทียบกันนั่นเอง คือ

$$i(1) = 1 \ , \ j(1) = 1$$

$$i(K) = I \ , \ j(K) = J$$

การหาค่านอร์มอลไลซ์ระยะเวลาที่น้อยที่สุดโดยเทคนิค ไดนามิคโปรแกรมมิ่ง

หลักการของไดนามิคโปรแกรมมิ่ง คือ พยายามค้นหาทุก ๆ เส้นทางจาก i=1 ถึง I และ j=1 ถึง J เพื่อหาค่าความแตกต่างของเสียงทั้งสองชุดที่น้อยที่สุด โดยเริ่มจาก i=1 และ j=1 ซึ่งค่า i และ j จะเพิ่มขึ้นครั้งละ 1 เพื่อหาความแตกต่างที่น้อยที่สุดในแต่ละขั้น จน i = I และ j = J ซึ่งเขียนเป็นสมการได้ดังนี้

เงื่อนไข เริ่มต้น

$$g_1(c(1)) = d(c(1)).w(1)$$

สมการไดนามิคโปรแกรมมิ่ง เพื่อหาค่า g ในแต่ละขั้นของ i และ j

คือ

$$g_k(c(k)) = \text{Min}[g_{k-1}(c(k-1)+d(c(k)).w(k)]$$

และเมื่อ i = I และ j = J แล้ว จะได้

$D(A,B) = \frac{1}{N} g\ (c(k))$ ซึ่งเป็นความแตกต่างที่น้อยที่สุด ที่นอร์มอลไลซ์เวลาแล้ว

สำหรับการวิจัยในครั้งนี้ ใช้วิธีการหา D(A,B) ดังนี้

$$g(1,1) = 2d(1,1)$$

$$g(i,j) = \min \begin{bmatrix} g(i,j-1)+d(i,j) \\ g(i-1,j-1)+2d(i,j) \\ g(i-1,j)+d(i,j) \end{bmatrix}$$

และ $D(A,B) = \frac{1}{N} g(I,J)$ โดย $N = I+J$ ซึ่งเขียนออกมาเป็นโครงสร้างของโปรแกรมได้ดังนี้

```
infini = 100000.0;
g(0,0) = 0;
for j=1 to J do g(0,j)=infini;
    for i=1 to I do{ g(i,0) = infini
                     for j=1 to J do{ d=distance(i,j)
                                      g(i,j)=min[g(i-1,j]+d or
                                      g(i-1,j-1)+2d
                                      g(i,j-1)+d]}}
D = g(i,j)/(I+J);
```

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

การทำงาน

การทำงานของโปรแกรมนี้จะแบ่งเป็น 2 ส่วนใหญ่ คือ ส่วนที่ทำหน้าที่วิเคราะห์และเก็บค่าพารามิเตอร์ของหน่วยเสียงพยางค์ที่จะใช้เป็นต้นแบบ และส่วนที่ทำหน้าที่วิเคราะห์, ตัดพยางค์และตรวจรู้ประโยคต่อเนื่อง

ในส่วนแรกมีการทำงานดังนี้

ขั้นที่ 1. จะต้องทำการพูดเสียงในแต่ละพยางค์ผ่านวงจร A/D เพื่อแปลงสัญญาณเสียงให้เป็นสัญญาณไม่ต่อเนื่อง หลังจากนั้นจะนำสัญญาณที่ได้มาลบด้วยค่าคงที่จำนวนหนึ่ง เพื่อที่จะยกระดับสัญญาณลงมา (เนื่องจากวงจร A/D ได้ยกระดับเสียงพูดขึ้นไปเพื่อใช้ในการสุ่ม (Sampling)) และเก็บไว้ในแฟ้มข้อมูล

ขั้นที่ 2. จะนำแฟ้มข้อมูลของแต่ละพยางค์มาหาค่าพารามิเตอร์ด้วยวิธีการพยากรณ์เชิงเส้น ซึ่งใช้สัมประสิทธิ์ของการพยากรณ์ทั้งหมด 15 ค่า (15 order) โดยผ่านโปรแกรมหาค่าพลังงาน α และค่าสัมประสิทธิ์ของการสะท้อนกลับ ซึ่งในการวิจัยนี้ได้ใช้ช่วงในการวิเคราะห์ (window size) ขนาด 250 ค่า หรือ 25 msec. และเลื่อนการวิเคราะห์ออกครั้งละ 50 ค่า หรือ 5 msec. ค่าพลังงาน ที่ได้จะเก็บไว้ในแฟ้มข้อมูลหนึ่งส่วนค่าสัมประสิทธิ์ของการสะท้อนกลับซึ่งเป็นอะเรย์ 1 มิติ ขนาด 15 ค่า จะเก็บไว้ในอีกแฟ้มข้อมูลหนึ่ง

ขั้นที่ 3. เป็นการนำค่าพารามิเตอร์สัมประสิทธิ์ของการสะท้อนกลับของแต่ละพยางค์มารวมอยู่ในแฟ้มข้อมูลที่จะใช้เป็นต้นแบบในการจำ ซึ่งในขั้นตอนนี้จะมีแฟ้มข้อมูลอีกแฟ้มหนึ่งเพื่อใช้เป็นที่เก็บสัญญลักษณ์(Label) ของพารามิเตอร์แต่ละชุด

ในส่วนของการตรวจรู้นั้นมีขั้นตอนดังนี้

ขั้นที่ 1 และ ขั้นที่ 2 จะเหมือนกับส่วนของการสร้างต้นแบบ

ขั้นที่ 3 จะนำค่าพารามิเตอร์พลังงาน α ที่วิเคราะห์ได้ในขั้นตอนที่ 2

มาผ่านโปรแกรมในการตัดพยางค์ ซึ่งจะได้ขอบเขตของพยางค์ หลังจากที่ได้ขอบเขตของแต่ละพยางค์ในประโยคแล้ว เราจะนำค่าพารามิเตอร์สัมประสิทธิ์ของการสะท้อนกลับในแต่ละพยางค์ที่วิเคราะห์ได้ในขั้นตอนที่ 2 มาเปรียบเทียบกับสัมประสิทธิ์ของการสะท้อนกลับของต้นแบบที่เตรียมไว้ในส่วนแรก โดยการเปรียบเทียบในแต่ละพยางค์จะเริ่มตั้งแต่พยางค์แรกจนถึงพยางค์สุดท้ายของแฟ้มข้อมูลที่เป็นต้นแบบ โดยถ้าค่าเปรียบเทียบกับต้นแบบพยางค์ใดมีค่าต่ำสุด ก็จะนำสัญญลักษณ์ของพยางค์นั้นมาแสดงผลและจะเริ่มเปรียบเทียบพยางค์ที่เหลือต่อ ๆ ไป จนครบทุกพยางค์ในประโยคที่พูด

ขั้นตอนการทดลอง

การทดลองได้แบ่งเป็น 2 ส่วน คือ ส่วนของการตัดพยางค์และส่วนของการจดจำ ส่วนของการตัดพยางค์ได้ใช้ตัวอย่างเสียงผู้พูดเป็นชาย 1 ท่าน หญิง 2 ท่าน ซึ่งในที่นี้ขอกำหนดให้ ช. หมายถึงชายซึ่งมีอายุ 28 ปี กำเนิดที่จังหวัดสิงห์บุรี ญ1. หมายถึงหญิงซึ่งมีอายุ 18 ปี กำเนิดที่จังหวัดขอนแก่น ญ2. หมายถึงหญิงซึ่งมีอายุ 25 ปี กำเนิดที่กรุงเทพมหานคร โดยผู้พูดทั้ง 3 ท่าน ได้พูดโดยใช้สำเนียงกรุงเทพมหานคร

ประโยคที่ใช้เป็นตัวอย่างใช้วิธีสุ่ม(random)แบ่งเป็น 2 ส่วน คือ ประโยคที่เกี่ยวกับชื่อและนามสกุลมีทั้งหมด 15 ประโยค และประโยคคำพูดทั่ว ๆ ไป อีก 15 ประโยค ดังปรากฏในภาคผนวก ค.

จากการทดลองได้ผลในการตัดพยางค์ดังตารางที่ 7.1-7.4 ซึ่งจะสังเกตุเห็นได้ว่า ในการพูดของตัวอย่างทั้ง 3 ท่านนั้น คำที่ตัดเกินส่วนใหญ่เป็นเสียงเสียดแทรกและเสียงนาสิก เช่นคำว่า "ชั่ว" ในประโยคที่ 21 ซึ่งพูดโดย ช.และ ญ1. หรือคำว่า "น้อย" ในประโยคที่ 27 ที่พูดโดย ช. ซึ่งแสดงให้เห็นว่าหน่วยเสียงประเภทเหล่านี้ มีลักษณะของพลังงานคล้ายหน่วยพยางค์มาก เนื่องจากหน่วยเสียงเหล่านี้มีระยะเวลาในการออกเสียงค่อนข้างนาน

สำหรับพยางค์ต่อเนื่องกันที่ระบบไม่สามารถตัดได้นั้น จากตารางที่ 7.1-7.4 จะเห็นว่าเป็นพยางค์ทั่ว ๆ ไป แต่ถ้าพิจารณาถึงขนาดของพลังงานแล้ว

จะเห็นว่า เหตุที่ระบบไม่สามารถจะคัดพยางค์เหล่านี้ได้ เป็นเพราะ

1. เป็นพยางค์ที่มีเสียงสั้นมาก เช่นในคำว่า "หวิ"

2. เนื่องจากขนาดของพลังงานในพยางค์ที่ใกล้เคียงกัน มีขนาดที่แตกต่างกันมาก

ส่วนของการตรวจรู้มีรายละเอียดของการทดลองดังนี้

ตัวอย่างที่ใช้ในการทดลองใช้วิธีสุ่มเลขประโยคต่อเนื่องเกี่ยวกับตัวเลข 0-9 จำนวน 8 ประโยค ดังปรากฏในภาคผนวก จ.

จากการทดลองได้ผลในการคัดพยางค์และตรวจรู้ดังตารางที่ 8. และ 9. จะเห็นว่าระบบสามารถที่จะคัดพยางค์ได้ถูกต้องทั้งหมด และตรวจรู้ได้ผิดพลาดเพียงพยางค์เดียว คือ 0 ตรวจรู้เป็น 7 ซึ่งแสดงให้เห็นว่า พยางค์ที่พูดต่อเนื่องกันนั้น มีผลกระทบต่อกันอยู่บ้าง

	จำนวนพยางค์ที่พูด	คัดถูกต้อง		คัดเกิน		คัดขาด	
			%		%		%
ช	234	223	95.29	5	2.13	6	2.58
ญ1	124	110	88.70	3	2.41	11	8.89
ญ2	110	104	94.54	-	0	6	5.46
รวม	468	437	93.37	8	1.70	23	4.93

ตารางที่ 7.1 แสดงผลในการคัดพยางค์

ประโยคที่	คัดขาด		คัดเกิน	
	จำนวน	คำ	จำนวน	คำ
11	-	-	1	สาว
12	-	-	1	สาว
21	-	-	2	ชั่ว, ชั่ว
27	-	-	1	น้อย
17	1	ทวิ	-	-
18	1	คุณไม่	-	-
20	1	ชีวิต	-	-
23	2	ไม่ได้	-	-
29	1	ฟาลง	-	-

ตารางที่ 7.2 พยางค์ที่คัดผิดพลาด(ผู้ใหญ่ ช.)

ประโยคที่	คัดขาด		คัดเกิน	
	จำนวน	คำ	จำนวน	คำ
1	1	ชาลี	-	-
3	2	ธีระ, โรธิน	-	-
6	1	ฒนา	-	-
11	2	สริน, รินจะ	-	-
18	2	คุณแล้ว, คุณไม่	-	-
20	1	ทั้งหลาย	-	-
26	1	ตำแหน่ง	-	-
30	1	โลยี	-	-
7	-	-	1	ผู้ร
20	-	-	1	ชี
21	-	-	1	ชั่ว

ตารางที่ 7.3 พยางค์ที่คัดผิดพลาด(ผู้ใหญ่ ฌ1.)

ประโยคที่	คัดขาด		คัดเกิน	
	จำนวน	คำ	จำนวน	คำ
2	1	มั่นคง	-	-
13	1	รงศักดิ์	-	-
22	1	ผู้เชี่ยว	-	-
23	2	เนโน, ไม่ได้	-	-
24	1	คำยัง	-	-

ตารางที่ 7.4 พยางค์ที่คัดผิดพลาด(พูดโดย ญ2.)

จำนวนพยางค์ที่พูด	คัดถูกต้อง	
		%
32	32	100

ตารางที่ 8 แสดงผลในการคัดพยางค์เพื่อใช้ในการตรวจรู้

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

คำที่พูด \ ผล	0	1	2	3	4	5	6	7	8	9
0	3							1		
1		5								
2			3							
3				3						
4					1					
5						4				
6							4			
7								7		
8									2	
9										2

ตารางที่ 9 แสดงผลในการตรวจรู้

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย