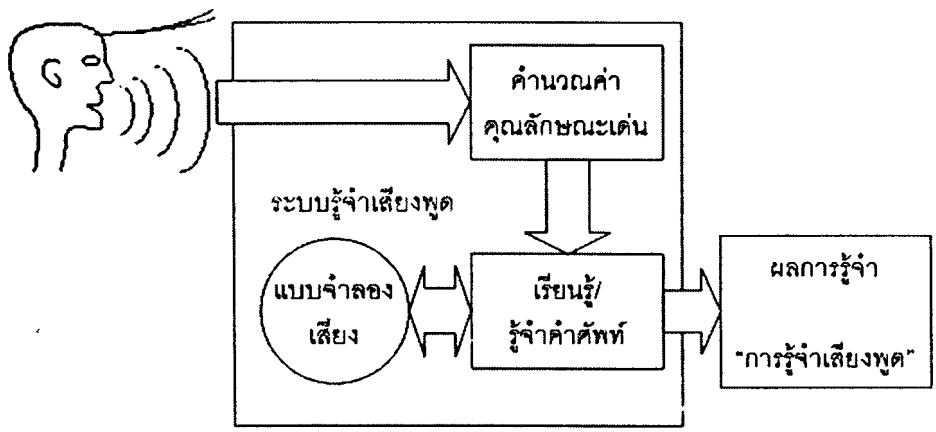
บทที่ 2

ทฤษฎีและหลักการ

2.1 การรู้จำเสียงพูด

ในการใช้งานระบบรู้จำเสียงพูดประกอบด้วย 2 ขั้นตอนที่สำคัญได้แก่ การฝึกฝน (training phase) และการรู้จำ (recognition phase) ขั้นตอนการฝึกฝนระบบรู้จำเป็นการสั่งสอนระบบรู้จำให้จดจำเสียงพูดตามสัญลักษณ์ที่กำหนดได้ โดยการป้อนข้อมูลเสียงและข้อมูลกำกับเสียง (คือข้อมูลที่ระบุว่าคำแต่ละคำอยู่ในช่วงใดของเสียงทั้งประโยค) ผ่านกระบวนการเรียนรู้เพื่อสร้างแบบจำลองเสียงที่เก็บค่าลักษณะเด่นของแต่ละคำ ขั้นตอนต่อมา คือขั้นตอนการรู้จำเป็นการเปรียบเทียบเสียงไม่ทราบค่ากับแบบจำลองเสียงว่าใกล้เคียงคำใดที่สุด ก็จะให้ผลการรู้จำเป็นคำนั้นออกมา ดังภาพประกอบ 2.1

หลักการทำงานภายในของระบบรู้จำเสียงพูดยังมีการประมวลผลที่ซับซ้อน ซึ่งได้อธิบายในหัวข้อถัดไป

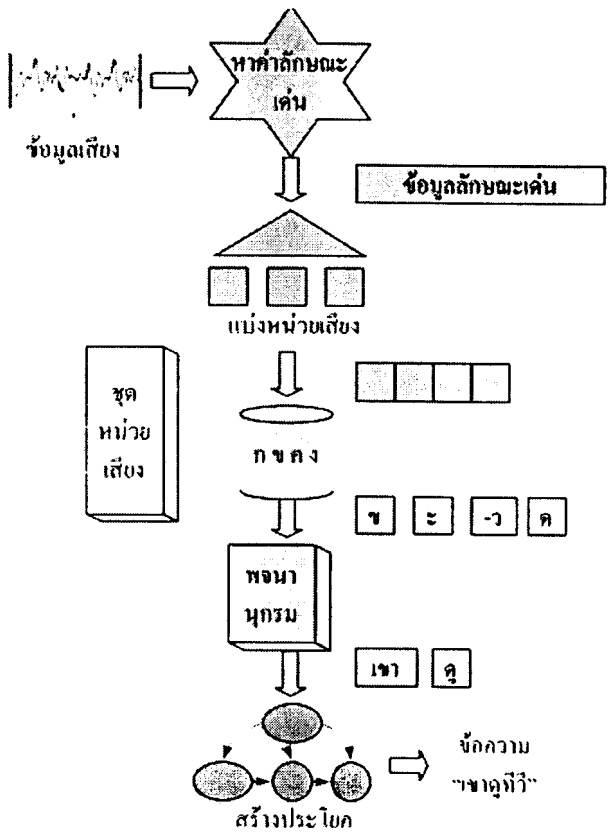


ภาพประกอบ 2.1 หลักการทำงานของระบบรู้จำเสียงพูด

ในการรู้จำเสียงพูดต้องมีการกำหนดค่าพารามิเตอร์สำหรับการรู้จำให้กับระบบเสียงก่อน ซึ่งค่าพารามิเตอร์เหล่านี้คือ แบบจำลองเสียง พจนานุกรม ไวยากรณ์ภาษา เสียงทดสอบ (ข้อมูลเสียงที่ใช้สำหรับการทดสอบระบบรู้จำ) จากนั้นจึงนำข้อมูลเสียงเข้าขั้นตอนการรู้จำ

ขั้นตอนการรู้จำเสียงพูด เริ่มจากการนำข้อมูลเสียงมาหาค่าลักษณะเด่น แบบเดียวกับที่ใช้ในการสร้างแบบจำลองเสียง จากนั้นแบ่งค่าข้อมูลออกเป็นส่วนๆ (segmentation) ตามช่วงเวลาและลักษณะของเสียงที่เปลี่ยนไป แล้วจึงนำข้อมูลแต่ละส่วนเข้าสู่การจำแนกหน่วยเสียง (classification) ได้ผลการจำแนกเป็นหน่วยเสียงออกมา จากนั้นนำกลุ่มหน่วยเสียงที่ได้มาตรวจสอบกับพจนานุกรม ได้เป็นชุดคำศัพท์ แล้วนำมาจัดเรียงเป็นประโยคตามไวยากรณ์ที่กำหนด ดังภาพประกอบ 2.2

ข้อมูลเสียงจะผ่านการแบ่งข้อมูลออกเป็นส่วนๆ จนถึงการจัดเรียงเป็นประโยค แล้วย้อนกลับมาแบ่งข้อมูลใหม่เช่นนี้หลายรอบ เพื่อคัดเลือกคำตอบที่ให้ค่าความเป็นไปได้สูงสุดให้เป็นผลการรู้จำออกมา

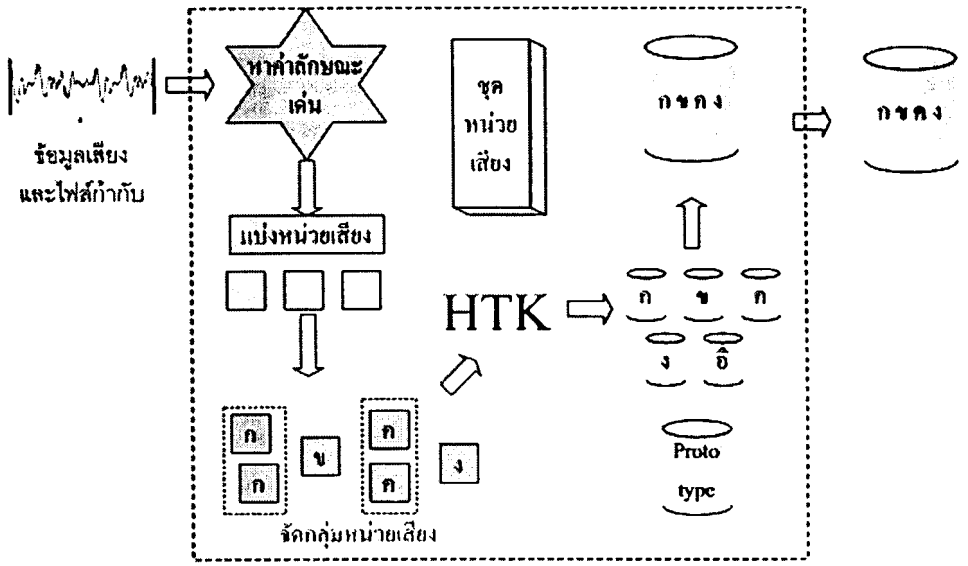


ภาพประกอบ 2.2 กระบวนการรู้จำเสียงพูด

2.2 การสร้างแบบจำลองเสียงด้วยวิธีการของฮิดเดนมาร์คอฟ

แบบจำลองเสียงคือฐานข้อมูลที่เก็บค่าลักษณะเด่นมูลฐานของหน่วยเสียงทั้งหมด การสร้างแบบจำลองเสียงต้องใช้ตัวอย่างเสียงฝึกฝนที่มีจำนวนมากพอ เพื่อให้ได้ค่าลักษณะเด่นที่เหมาะสมสำหรับบันทึกลงในแบบจำลองเสียง

ในการสร้างแบบจำลองเสียงหรือการฝึกฝนระบบรู้จำ จำเป็นต้องมีการกำหนด ชุดหน่วยเสียง วิธีการคำนวณค่าลักษณะเด่น โครงสร้าง HMM และข้อมูลเสียงสำหรับการฝึกฝน (เป็นเสียงที่ใช้สำหรับการสอนระบบรู้จำ)



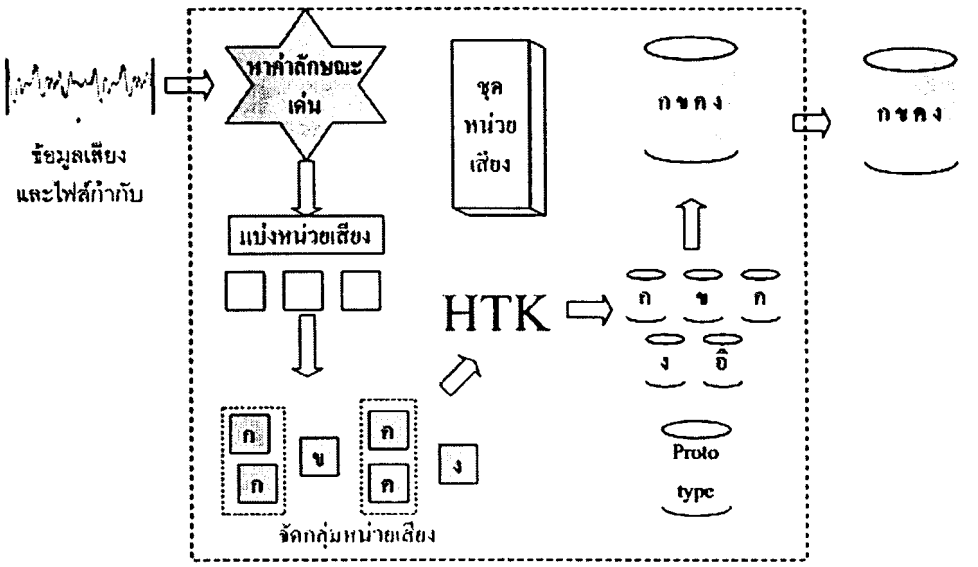
ภาพประกอบ 2.3 ขั้นตอนการสร้างแบบจำลองเสียง

จากภาพประกอบ 2.3 แสดงถึงขั้นตอนการสร้างแบบจำลองเสียง โดยเริ่มจากการนำข้อมูลเสียงมาหาค่าลักษณะเด่น และทำการแยกค่าลักษณะเด่นออกเป็นช่วงหน่วยเสียงตามที่ระบุในข้อมูลกำกับเสียง (phonetic transcription) แล้วจัดเป็นกลุ่มของหน่วยเสียงแต่ละหน่วย จากนั้นจึงหาค่าเฉลี่ย (mean) และค่าความแปรปรวน (covariance) แบบเกาส์เซียน (Gaussian model) ในแต่ละหน่วยเสียงเพื่อบันทึกลงในแบบจำลองเสียง โดยการคำนวณของชุดโปรแกรม HTK แล้วจึงรวมแบบจำลองหน่วยเสียงทั้งหมดลงในแบบจำลองเดียวกันเพื่อนำไปใช้ในการรู้จำเสียงพูดต่อไป

2.2 การสร้างแบบจำลองเสียงด้วยวิธีการของฮิดเดนมาร์คอฟ

แบบจำลองเสียงคือฐานข้อมูลที่เก็บค่าลักษณะเด่นมูลฐานของหน่วยเสียงทั้งหมด การสร้างแบบจำลองเสียงต้องใช้ตัวอย่างเสียงฝึกฝนที่มีจำนวนมากพอ เพื่อให้ได้ค่าลักษณะเด่นที่เหมาะสมสำหรับบันทึกลงในแบบจำลองเสียง

ในการสร้างแบบจำลองเสียงหรือการฝึกฝนระบบรู้จำ จำเป็นต้องมีการกำหนด ชุดหน่วยเสียง วิธีการคำนวณค่าลักษณะเด่น โครงสร้าง HMM และข้อมูลเสียงสำหรับการฝึกฝน (เป็นเสียงที่ใช้สำหรับการสอนระบบรู้จำ)



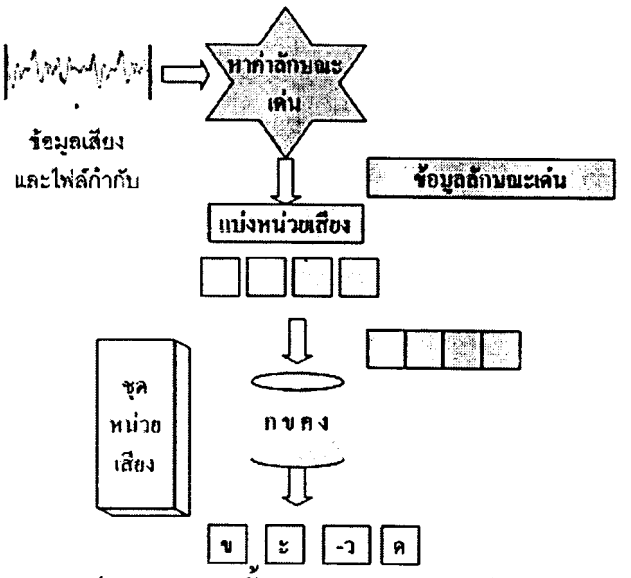
ภาพประกอบ 2.3 ขั้นตอนการสร้างแบบจำลองเสียง

จากภาพประกอบ 2.3 แสดงถึงขั้นตอนการสร้างแบบจำลองเสียง โดยเริ่มจากการนำข้อมูลเสียงมาหาค่าลักษณะเด่น และทำการแยกค่าลักษณะเด่นออกเป็นช่วงหน่วยเสียงตามที่ระบุในข้อมูลกำกับเสียง (phonetic transcription) แล้วจัดเป็นกลุ่มของหน่วยเสียงแต่ละหน่วย จากนั้นจึงหาค่าเฉลี่ย (mean) และค่าความแปรปรวน (covariance) แบบเกาส์เซียน (Gaussian model) ในแต่ละหน่วยเสียงเพื่อบันทึกลงในแบบจำลองเสียง โดยการคำนวณของชุดโปรแกรม HTK แล้วจึงรวมแบบจำลองหน่วยเสียงทั้งหมดลงในแบบจำลองเดียวกันเพื่อนำไปใช้ในการรู้จำเสียงพูดต่อไป

2.3 การจำแนกหน่วยเสียง

เป็นขั้นตอนหนึ่งของการรู้จำเสียงพูดคือการวิเคราะห์เสียงที่ไม่ทราบค่าว่าเป็นหน่วยเสียงใด ขั้นตอนนี้เป็นการหาค่าความแม่นยำของแบบจำลองเสียงทำให้ทราบถึงจุดอ่อน/จุดแข็งในการรู้จำหน่วยเสียงแต่ละหน่วย และความแม่นยำสูงสุดที่เป็นไปได้

สิ่งที่จำเป็นต้องใช้สำหรับการจำแนกหน่วยเสียงคือแบบจำลองเสียง วิธีการคำนวณค่าลักษณะเด่นที่ใช้สร้างแบบจำลอง ชุดหน่วยเสียงที่มีในแบบจำลอง และ เสียงทดสอบ



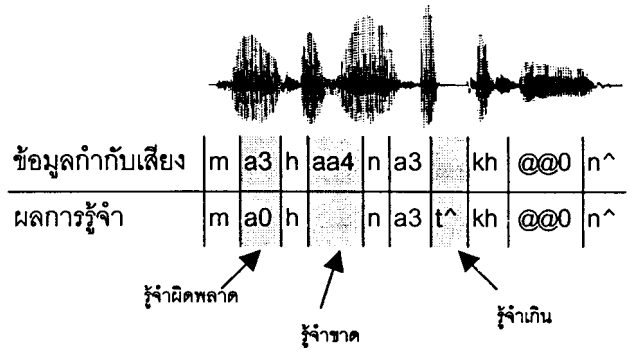
ภาพประกอบ 2.4 ขั้นตอนการจำแนกหน่วยเสียง

วิธีการจำแนกหน่วยเสียง เริ่มจากการนำข้อมูลเสียงมาคำนวณหาค่าคุณลักษณะเด่น และแบ่งเป็นหน่วยเสียงตามข้อมูลกำกับเสียง แล้วนำเข้าสู่กระบวนการจำแนกหน่วยเสียงด้วยแบบจำลองเสียง นำผลการจำแนกมาเปรียบเทียบกับข้อมูลกำกับเสียงเพื่อหาค่าความแม่นยำ ดังภาพประกอบ 2.4

2.4 การเปรียบเทียบผลการรู้จำ

การเปรียบเทียบผลการรู้จำเป็นเปรียบเทียบผลลัพท์ที่ได้จากการรู้จำหรือการจำแนกเทียบกับข้อมูลกำกับเสียง มี 4 แบบด้วยกัน

1. ข้อมูลในผลการรู้จำตรงกันกับข้อมูลกำกับเสียง เรียกว่า การรู้จำถูกต้อง (correction)
2. ข้อมูลในผลการรู้จำไม่ตรงกันกับข้อมูลกำกับเสียงเรียกว่า รู้จำผิดพลาด (substitution error)
3. ผลการรู้จำไม่มีแต่ข้อมูลกำกับเสียงมี เรียกว่าเรียกว่า การรู้จำขาด (deletion error)
4. ในผลการรู้จำมีแต่ข้อมูลกำกับเสียงไม่มี เรียกว่า การรู้จำเกิน (insertion error)



ภาพประกอบ 2.5 การเปรียบเทียบผลการรู้จำเสียงพูดและข้อมูลกำกับเสียง

การเปรียบเทียบผลการรู้จำมี 2 วิธีคือ การหาค่าความแม่นยำและ การหาค่าความถูกต้อง ซึ่งแต่ละวิธีจะมีการคำนวณที่แตกต่างกัน

1. **ค่าความแม่นยำ** เป็นการคำนวณผลการรู้จำเทียบกับข้อมูลกำกับเสียงโดยอาศัยปัจจัยความผิดพลาดทุกอย่างที่มี ตั้งแต่ การรู้จำทั้งหมด การรู้จำผิดพลาด การรู้จำขาด การรู้จำเกิน ดังสมการที่ 2.1

$$Accuracy = \frac{N - D - S - I}{N} \times 100\% \qquad (2.1)$$

N จำนวนข้อมูลในข้อมูลกำกับเสียง

D จำนวนที่รู้จำขาด

S จำนวนที่รู้จำผิดพลาด

I จำนวนที่รู้จำเกิน

2. **ค่าความถูกต้อง** เป็นการคำนวณผลการรู้จำเทียบกับข้อมูลกำกับเสียงโดยอาศัยปัจจัยความผิดพลาดบางอย่างได้แก่ การรู้จำทั้งหมด การรู้จำผิดพลาด การรู้จำขาด ดังสมการที่ 2.2 เนื่องจากในบางกรณีการรู้จำเกินไม่จำเป็นต้องนำมาคำนวณ เช่น การรู้จำเสียงพูดในบริเวณที่มีสัญญาณรบกวนมาก

$$Correct = \frac{N - D - S}{N} \times 100\% \tag{2.2}$$

N จำนวนข้อมูลในข้อมูลกำกับเสียง

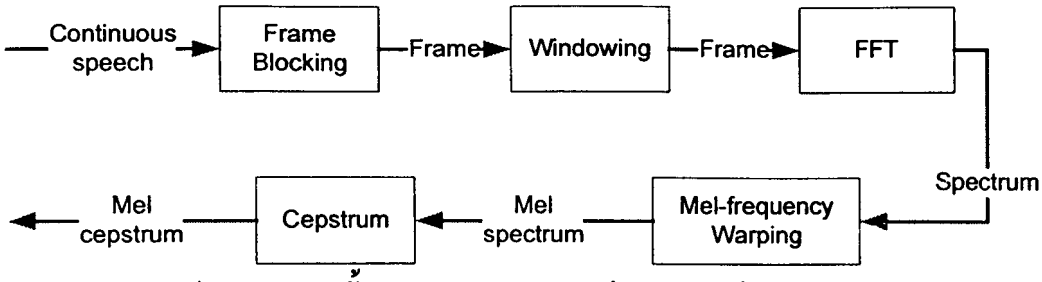
D จำนวนที่รู้จำขาด

S จำนวนที่รู้จำผิดพลาด

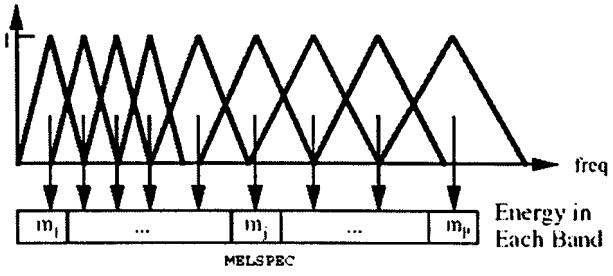
2.5 การคำนวณค่าลักษณะเด่น

เป็นการค้นหาจุดเด่นของเสียงเพื่อให้ระบบรู้จำสามารถจดจำได้ การคำนวณค่าลักษณะเด่นของเสียงมีหลายวิธีด้วยกันคือ MFCC และ LPC ในงานวิจัยที่ใช้แบบจำลองฮิดเดนมาร์คอฟส่วนใหญ่เลือกใช้ MFCC (Gowdy, J.N. and Tufekci, Z., 2000) เพราะมีความแม่นยำสูงกว่า LPC อย่างชัดเจน ในงานวิจัยนี้จึงได้นำเสนอวิธีการคำนวนค่าลักษณะเด่นแบบ MFCC

ในการคำนวณค่าคุณลักษณะเด่นแบบ MFCC จากภาพประกอบ 2.6 เริ่มจากการนำข้อมูลเสียงมาแบ่งเป็นเฟรม (frame blocking) แล้วนำข้อมูลแต่ละเฟรมมาคูณด้วยแฮมมิงวินโดวส์ (hamming window) เพื่อลดความแตกต่างระหว่างข้อมูลแต่ละเฟรมจากนั้นจึงเข้ากระบวนการแปลงข้อมูลจากแกนเวลาเป็นแกนความถี่โดยการแปลงรูปข้อมูลแบบฟาสฟูริเยร์ (fast Fourier transform: FFT) แล้วจึงนำมากรองช่วงความถี่ด้วยฟิลเตอร์ของเมล (Mel-frequency warping) เป็นการกำหนดขอบเขตช่วงของความถี่แต่ละค่า มีลักษณะเป็นรูปสามเหลี่ยมหลายรูปต่อกัน ดังภาพประกอบ 2.7 จากนั้นจึงนำมาคำนวณหาค่าเคปสตรัม (Cepstrum) ในขั้นสุดท้าย ได้เป็นค่าลักษณะเด่นออกมา ซึ่งเป็นชุดข้อมูลที่มีจำนวนเท่ากับจำนวนเฟรมของเสียงนั้นๆ ภายในข้อมูลแต่ละชุดจะมีข้อมูลย่อยที่มีจำนวนเท่ากับจำนวนวินโดวส์ของเมล



ภาพประกอบ 2.6 ขั้นตอนการคำนวณหาค่าลักษณะเด่นแบบ MFCC

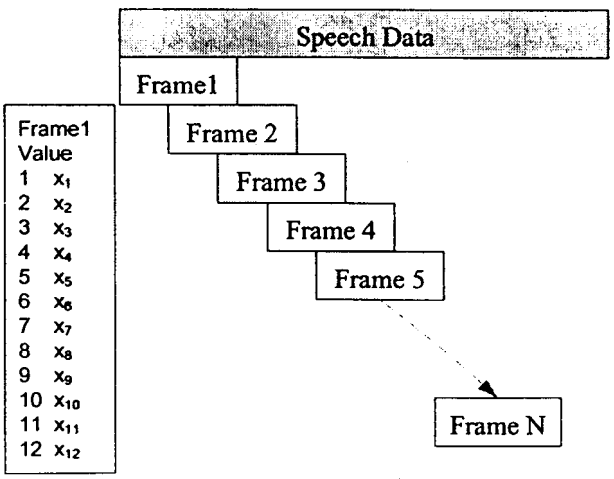


ภาพประกอบ 2.7 ขอบเขตช่วงของความถี่ของเมล

ในขั้นตอนนี้จะผู้ใช้สามารถกำหนดพารามิเตอร์สำหรับการคำนวณค่าลักษณะเด่นได้

- ความยาวของเฟรม (frame size) ค่าที่นิยมใช้อยู่ในช่วง 20-30 มิลลิวินาที
- ระยะห่างระหว่างเฟรม (frame rate) ค่าที่นิยมใช้อยู่ในช่วง 10-15 มิลลิวินาที
- จำนวนสัมประสิทธิ์เคปสตรัม ค่าที่นิยมใช้ประมาณ 12

ข้อมูลเสียง เมื่อนำมาหาค่าลักษณะเด่นจะได้ข้อมูล N ชุดขึ้นอยู่กับความยาวของเสียง ความยาวของเฟรม และระยะห่างระหว่างฟรม โดยที่ข้อมูลแต่ละชุดมี 12 ค่า ดังภาพประกอบ 2.8



ภาพประกอบ 2.8 ตัวอย่างค่าลักษณะเด่นของเสียงในช่วงหนึ่ง

เทคนิคที่ใช้ในกับการปรับปรุงค่าลักษณะเด่น

1. การปรับปรุงค่าลักษณะเด่น MFCC

ค่าลักษณะเด่นที่มีอยู่อาจไม่เพียงพอต่อการนำมารู้จำ จึงมีการเพิ่มลักษณะเด่นบางอย่างเข้าไป เช่น ค่าอนุพันธ์อันดับที่หนึ่ง (MFCC+D) เป็นการนำค่าความแตกต่างของคุณลักษณะเด่นที่อยู่ติดกัน ดังสมการที่ 2.3 ซึ่งการเพิ่มค่าค่าอนุพันธ์อันดับที่หนึ่งทำให้ค่าลักษณะเด่นมีจำนวนเพิ่มขึ้นเป็น 2 เท่า

$$d_t = \frac{\sum_{\theta=1}^{\Theta} \theta(c_{t+\theta} - c_{t-\theta})}{2\sum_{\theta=1}^{\Theta} \theta^2} \tag{2.3}$$

d_t คือ ค่าอนุพันธ์อันดับที่หนึ่งของค่าลักษณะเด่นที่เฟรม t

Θ คือ ครึ่งหนึ่งของขนาดหน้าต่าง ที่ใช้ในการกำหนดค่าอนุพันธ์ ค่าปริยายคือ 2 เฟรม

c_t คือค่าคุณลักษณะเด่นที่เฟรม t

ค่าอนุพันธ์อันดับที่สอง (MFCC+DA) มาจากการหาค่าอนุพันธ์ของ MFCC+D ดังสมการที่ 2.3 การเพิ่มค่าค่าอนุพันธ์อันดับที่สองทำให้ค่าลักษณะเด่นมีจำนวนเพิ่มขึ้นเป็น 3 เท่า ค่าดังกล่าวนี้ทำให้จำนวนข้อมูลในประมวลผลมากขึ้นและช่วยเพิ่มความแม่นยำในการรู้จำ

$$a_t = \frac{\sum_{\theta=1}^{\Theta} \theta(d_{t+\theta} - d_{t-\theta})}{2\sum_{\theta=1}^{\Theta} \theta^2} \tag{2.4}$$

a_t คือค่าอนุพันธ์อันดับที่สองของค่าลักษณะเด่นที่เฟรม t

Θ คือ ครึ่งหนึ่งของขนาดหน้าต่าง ที่ใช้ในการกำหนดค่าอนุพันธ์ ค่าปริยายคือ 2 เฟรม

d_t คือค่าอนุพันธ์อันดับที่ 1 ของค่าลักษณะเด่นที่เฟรม t

ตัวอย่างผลการคำนวณค่าลักษณะเด่นเป็นดังนี้

MFCC $= [c_1, c_2, c_3, c_4, c_5, c_6, c_7, c_8, c_9, c_{10}, c_{11}, c_{12}]^T$

ค่าอนุพันธ์อันดับที่ 1 $= [d_1, d_2, d_3, d_4, d_5, d_6, d_7, d_8, d_9, d_{10}, d_{11}, d_{12}]^T$

ค่าอนุพันธ์อันดับที่ 2 $= [a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8, a_9, a_{10}, a_{11}, a_{12}]^T$

จากนั้นจึงนำค่าอนุพันธ์ที่ได้มาจัดเรียงให้อยู่ในรูปแบบของเวกเตอร์ดังนี้

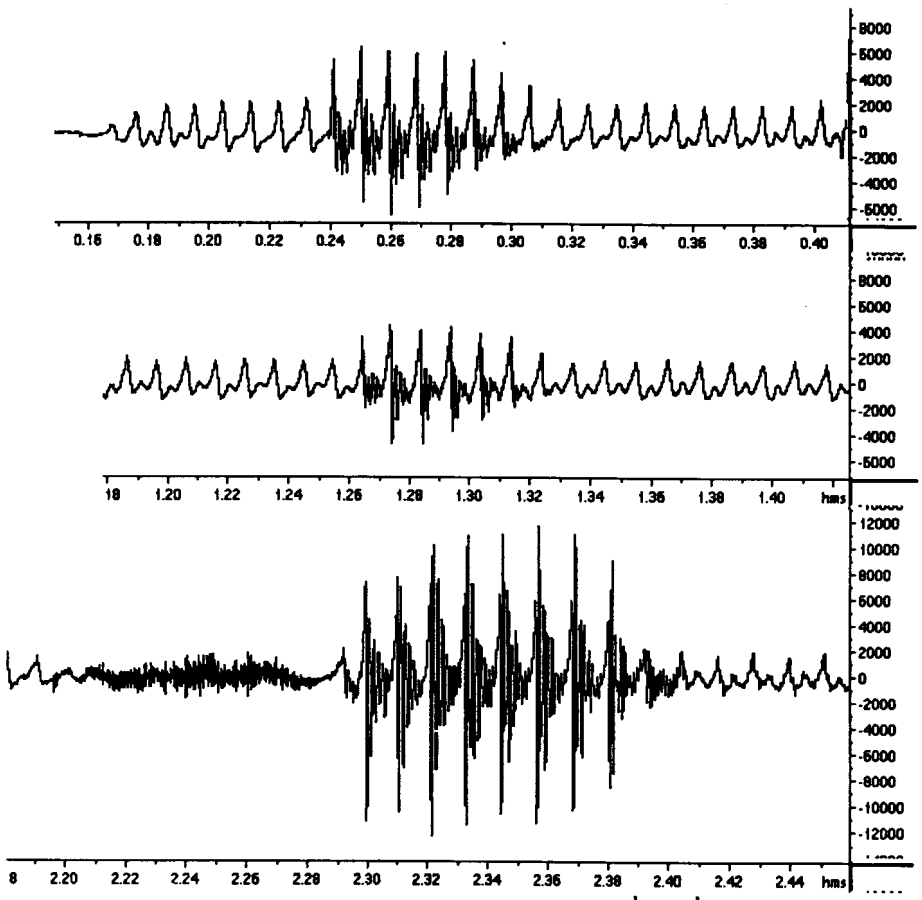
เมื่อเพิ่มค่าอนุพันธ์อันดับที่ 1 ทำให้มีค่าลักษณะเด่นจำนวน 24 ค่า

$$\text{MFCC+D} = [c_1,\ldots c_{12}, d_1,\ldots d_{12}]^T$$

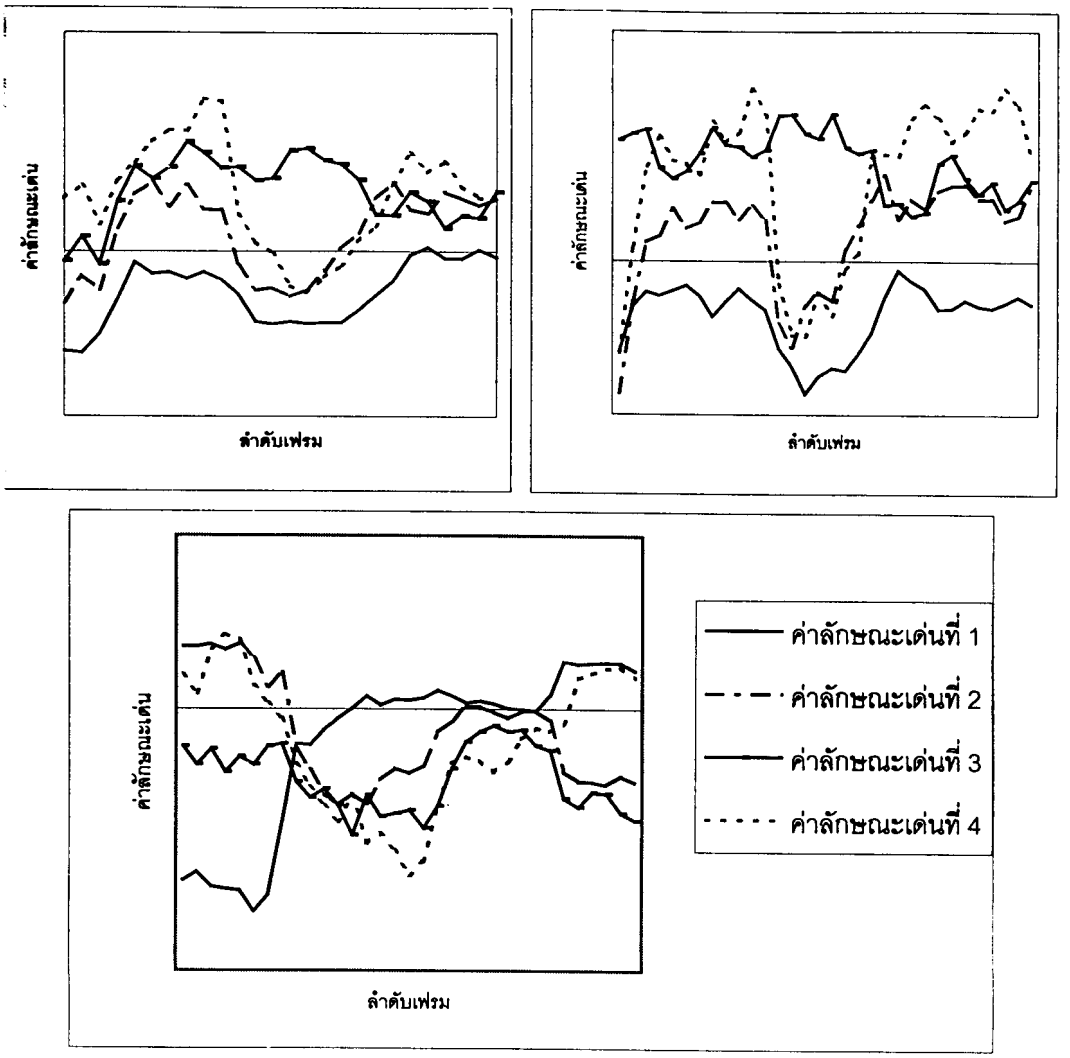
เมื่อเพิ่มค่าอนุพันธ์อันดับที่ 1 และ 2 ทำให้มีค่าลักษณะเด่นจำนวน 36 ค่า

$$\text{MFCC+DA} = [c_1,\ldots c_{12}, d_1,\ldots d_{12}, a_1,\ldots a_{12}]^T$$

เมื่อพิจารณาค่าลักษณะเด่นที่ดี เสียงพูดคำเดียวกันจะมีความเหมือนกันอย่างเห็นได้ชัด ตัวอย่างเช่น เสียงคำว่า หนึ่ง หนึ่ง และ สอง เมื่อเปรียบเทียบลักษณะของสัญญาณเสียง จากภาพประกอบ 2.9 พบว่าความแตกต่างของ หนึ่ง และ สอง มีน้อยมากในขณะที่ค่าลักษณะเด่นของเสียงดังกล่าวมีความแตกต่างที่เห็นได้อย่างชัดเจนดังภาพประกอบ 2.10



ภาพประกอบ 2.9 ความแตกต่างของสัญญาณเสียง หนึ่ง หนึ่ง สอง ตามลำดับ



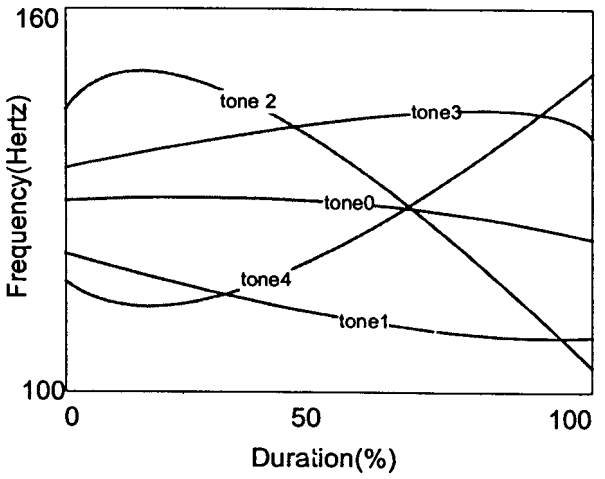
ภาพประกอบ 2.10 ความแตกต่างของค่าลักษณะเด่นของเสียง หนึ่ง หนึ่ง สอง ตามลำดับ

2. การเพิ่มค่าพิตช์

การรู้จำวรรณยุกต์ถือว่าเป็นสิ่งสำคัญสำหรับระบบรู้จำเสียงพูดภาษาไทย เนื่องจากค่าลักษณะเด่น MFCC ไม่สามารถแยกระดับวรรณยุกต์ได้อย่างชัดเจน จากงานวิจัยของหลี่ ถาน (Tan, L. and Karnjanadecha, M., 2003) พบว่าค่าพิตช์คือค่าของความถี่ที่มีค่าแอมปลิจูดสูงสุดในช่วงเสียงหนึ่ง ซึ่งมีความสอดคล้องกับระดับวรรณยุกต์ของภาษาไทยดังภาพประกอบ 2.11 โดยที่ tone0 ถึง tone4 คือ เสียงไม่มีวรรณยุกต์หรือเสียงสามัญ จนถึง วรรณยุกต์จัตวา

การเพิ่มค่าพิตช์ สามารถทำได้ในขั้นตอนของการคำนวณค่าลักษณะเด่น โดยเพิ่มค่าพิตช์ลงไปในค่าลักษณะเด่น การแก้ปัญหาวิธีนี้จะสร้างความซับซ้อนในการคำนวณค่าลักษณะเด่นทั้ง MFCC และคำนวณค่าพิตช์ และจำเป็นต้องสร้างแบบจำลองเสียงตามวิธีการคำนวณค่าลักษณะเด่น

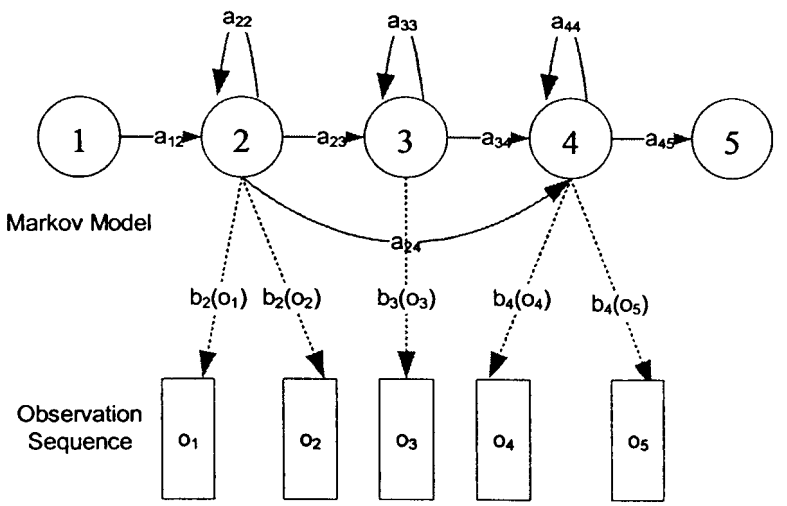
แบบดังกล่าว ซึ่งภายในงานวิจัยนี้จึงไม่ได้นำเสนอวิธีการเพิ่มค่าพิตช์ลงในค่าลักษณะเด่น ตลอดจนค่าความแม่นยำหลังจากเพิ่มค่าพิตช์



ภาพประกอบ 2.11 ค่าพิตช์ของการออกเสียงคำในภาษาไทยที่ระดับวรรณยุกต์ต่างๆ กัน 5 ระดับ (ที่มา: Potisuk, S., Harper, M. P. and Gandour, J., 1999)

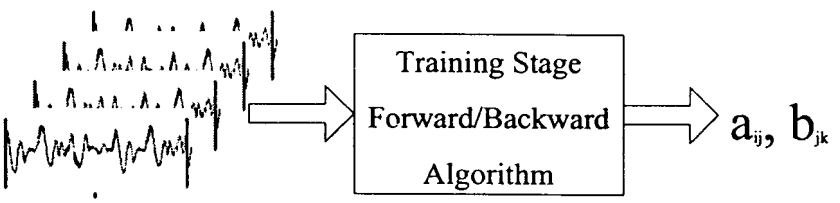
2.6 แบบจำลองฮิดเดนมาร์คอฟ

แบบจำลองฮิดเดนมาร์คอฟ คือแบบจำลองการเปลี่ยนแปลงสัญญาณในรูปแบบต่างๆ เพื่อการจดจำและเปรียบเทียบกับสัญญาณอื่นๆ แบบจำลองนี้จะจดจำการเปลี่ยนสเตทของสัญญาณ (สเตทคือ ลักษณะของสัญญาณในแต่ละช่วงเวลาที่มีรูปแบบเฉพาะตัวแตกต่างกันไป) ภายในประกอบด้วยตัวแปร 2 ชุดคือ a_{ij} และ b_{jk} ชุดตัวแปร a_{ij} จะควบคุมการเปลี่ยนสเตทภายในแบบจำลอง ส่วน b_{jk} หรือ $b_j(o_k)$ จะควบคุมผลการเปรียบเทียบกับสัญญาณที่เข้ามา (o_k) เมื่อมีสัญญาณใดๆ เข้ามาในแต่ละสเตทจะมีการคำนวณความน่าจะเป็นที่แบบจำลองนี้จะให้สัญญาณแบบเดียวกันในสเตทดังกล่าวได้ โดยการคำนวณจาก 2 ชุดตัวแปรและ สัญญาณที่เข้ามา ดังภาพประกอบ 2.12 ซึ่งการใช้งานแบบจำลองนี้จะต้องมี 2 ส่วนคือ การสร้างแบบจำลอง และการเปรียบเทียบสัญญาณ



ภาพประกอบ 2.12 หลักการทำงานของแบบจำลองเสียงแบบฮิดเดนมาร์คอฟ

การสร้างแบบจำลอง ต้องนำกลุ่มสัญญาณ จำนวนหนึ่งมาเป็นข้อมูลฝึกฝนเพื่อกำหนดค่าตัวแปรภายในแบบจำลองเสียงซึ่งก็คือ a_{ij} และ b_{jk} โดยวิธีการ Forward/Backward Algorithm (Duda, R.O., Hart, P.E. and Stork, D.G., 2000: 138) ดังภาพประกอบ 2.13 ซึ่งจะได้ค่าของชุดตัวแปรทั้งสองออกมา ดังตัวอย่างในภาพประกอบ 2.14 ขั้นตอนนี้จะต้องกำหนดจำนวนสเตทและมิกซเจอร์เพื่อควบคุมความละเอียดของสัญญาณที่แบบจำลองจดจำได้



ภาพประกอบ 2.13 ขั้นตอนภายในของการสร้างแบบจำลองเสียง

$$a_{ij} = \begin{bmatrix} 1.0,0.0,0.0,0.0 \\ 0.2,0.3,0.1,0.4 \\ 0.2,0.5,0.2,0.1 \\ 0.8,0.1,0.0,0.1 \end{bmatrix}$$

$$b_{jk} = \begin{bmatrix} 1.0,0.0,0.0,0.0,0.0 \\ 0.0,0.3,0.4,0.1,0.2 \\ 0.0,0.1,0.1,0.7,0.1 \\ 0.0,0.5,0.2,0.1,0.2 \end{bmatrix}$$

ภาพประกอบ 2.14 ตัวอย่างข้อมูลของชุดตัวแปร a_{ij} และ b_{jk}

การเปรียบเทียบสัญญาณ เมื่อมีสัญญาณข้อมูลใดๆ ผ่านการคำนวณค่าลักษณะเด่นในแต่ละเฟรม และค่าลักษณะเด่นในแต่ละเฟรมผ่านเข้ามาในแต่ละสเตท ก็จะมีการคำนวณค่าความน่าจะเป็นของสเตทนั้นดังสมการที่ 2.5 (Duda, R.O., Hart, P.E. and Stork, D.G., 2000: 131) เพื่อแสดงให้เห็นว่าสัญญาณที่เข้ามามีความคล้ายคลึงกับแบบจำลองเสียงเพียงใด ซึ่งค่าที่ได้จาก สเตทสุดท้ายจะเป็นค่าสุทธิที่บอกว่าสัญญาณดังกล่าวเหมือนหรือต่างจากแบบจำลองเสียง ดังภาพประกอบ 2.15 สเตทที่ 0 คือสเตทสุดท้ายของแบบจำลอง ค่าที่ได้จากสเตทนี้คือความน่าจะเป็นที่สัญญาณ $o_1o_3o_2o_0$ จะเป็นสัญญาณเดียวกันกับแบบจำลองที่มีชุดตัวแปรในภาพประกอบ 2.13 ซึ่งมีค่าเป็น 0.001 หากแบบจำลองนี้คือ ก ก็จะหมายความว่าความน่าจะเป็นที่สัญญาณ $o_1o_3o_2o_0$ เป็น ก มีค่า 0.001

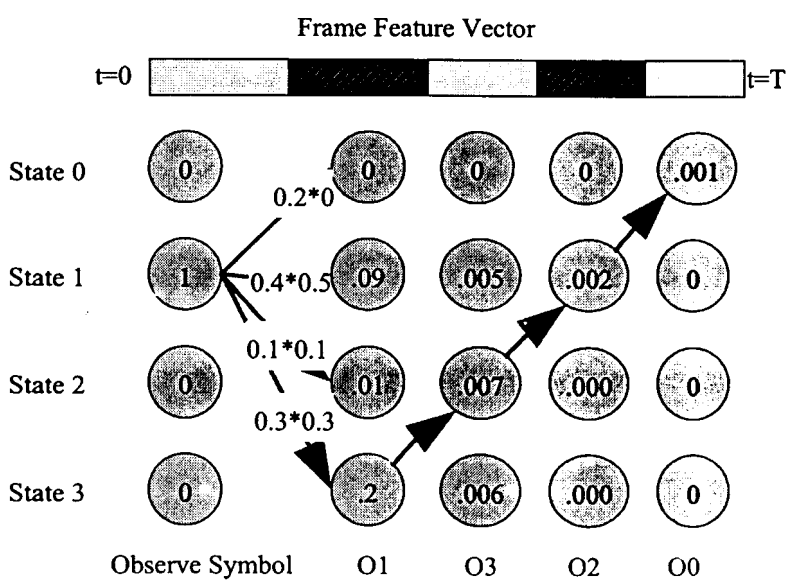
ดังนั้นสัญญาณชุดหนึ่งๆ จะถูกนำไปเปรียบเทียบกับแบบจำลองเสียงทั้งหมดที่มีอยู่ แล้วคัดเลือกแบบจำลองที่ให้ค่าความน่าจะเป็นที่สูงที่สุดเป็นผลการรู้จำสัญญาณนั้นๆ

$$P(o^T) = \sum_{r=1}^{r_{max}} \prod_{t=1}^{T} P(o(t) \mid s(t)) P(s(t) \mid s(t-1)) \qquad (2.5)$$

$P(o^T)$ ความน่าจะเป็นที่จะเกิดสัญญาณ o^T

$P(o(t) \mid s(t))$ ความน่าจะเป็นที่จะเกิดสัญญาณ o ที่สเตท s ที่เวลา t $= b_{s(t)o(t)}$

$P(s(t) \mid s(t-1))$ ความน่าจะเป็นที่จะมีการเปลี่ยนสเตท s(t-1) เป็น s(t) $= a_{s(t-1)s(t)}$



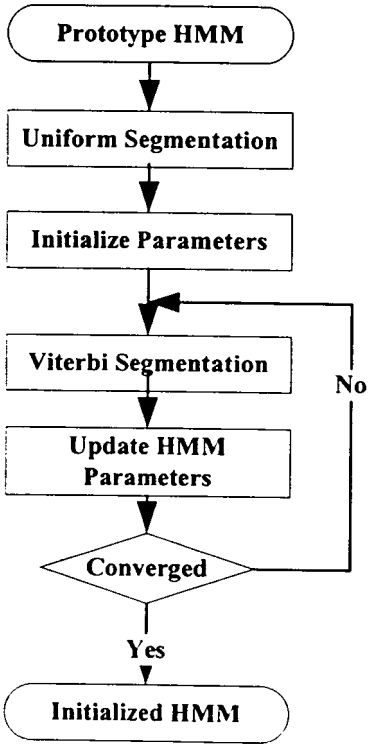
ภาพประกอบ 2.15 การทำงานของแบบจำลองเสียงเพื่อรู้จำเสียงไม่ทราบค่า

2.7 ขั้นตอนของ HTK ในการสร้างแบบจำลองเสียง

ภายในขั้นตอนการสร้างแบบจำลองเสียงของ HTK มีขั้นตอนการประมวลผลด้วยกัน 3 ขั้นตอนคือ การกำหนดค่าเริ่มต้น (model initialization) ขั้นตอน Single Re-estimation และขั้นตอน Embedded Re-estimation

2.7.1 กำหนดค่าเริ่มต้น (model initialization)

เป็นการกำหนดค่าเริ่มต้นให้กับแบบจำลองเสียง ด้วยโปรแกรม Hinit จากชุดโปรแกรม HTK จากภาพประกอบ 2.16 ขั้นตอนนี้เริ่มจากการนำเสียงมาแบ่งเป็นส่วนย่อยที่มีความยาวเท่าๆกัน (uniform segmentation) โดยจำนวนส่วนย่อยนี้จะเท่ากับจำนวน State จากนั้นจะใช้ข้อมูลในส่วนย่อยนี้เพื่อคำนวณหาค่าการกระจายแบบเกาส์เซียน (Gaussian distribution function) ให้กับชุดตัวแปรในทุก State จากนั้นจึงทำการคำนวณใหม่อีกครั้งหนึ่งโดยใช้วิเทอร์บีอัลกอริทึม แบ่งข้อมูลและปรับปรุงค่าตัวแปรภายในแบบจำลองอีก ทำเช่นนี้ไปเรื่อยๆ ประมาณ 20 รอบ จำนวนรอบในการปรับปรุงค่านี้ผู้ใช้สามารถกำหนดเองได้



ภาพประกอบ 2.16 ขั้นตอนการกำหนดค่าเริ่มต้นให้กับแบบจำลองเสียง

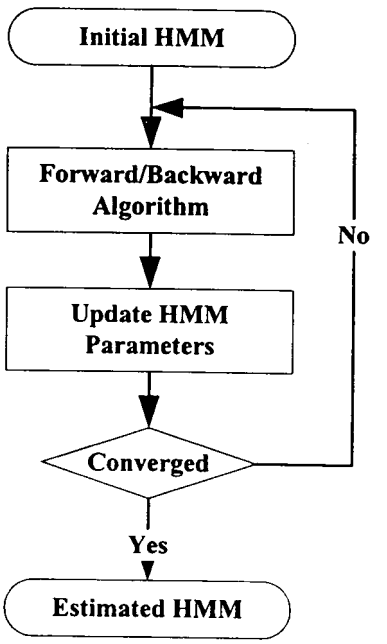
(ที่มา: Young, S., et al., 2000: 139)

2.7.2 ขั้นตอน Single Re-estimate และ Embedded Re-estimate

ขั้นตอนนี้เป็นการปรับคุณสมบัติบางอย่างของแบบจำลองเสียงให้เหมาะสมกับเสียงพูด คำนวณโดยโปรแกรม HRest ของชุดโปรแกรม HTK

เริ่มจากการนำข้อมูลเสียงผ่านกระบวนการคำนวณแบบ Forward/Backward Algorithm ดังสมการที่ 2.6 (Young, S., et al., 2000: 151) แล้วไปปรับปรุงค่าของชุดตัวแปรภายในแบบจำลองเสียงจากขั้นตอนแรก และตรวจสอบการลู่เข้าของข้อมูล (จากสมการดังกล่าวจะให้ค่าที่ลู่เข้าแน่นอน) จากนั้นก็ทำขั้นตอนนี้ซ้ำไปเรื่อยๆ ตามจำนวนรอบที่กำหนด ดังภาพประกอบ 2.17

$$a_{ij} = \frac{\sum_{r=1}^{R} \frac{1}{P_r} \sum_{t=1}^{T_r - 1} \alpha_i^r(t) a_{ij} b_j(o_{t+1}^r) \beta_j^r(t+1)}{\sum_{r=1}^{R} \frac{1}{P_r} \sum_{t=1}^{T_r} \alpha_i^r(t) \beta_j^r(t)} \tag{2.6}$$



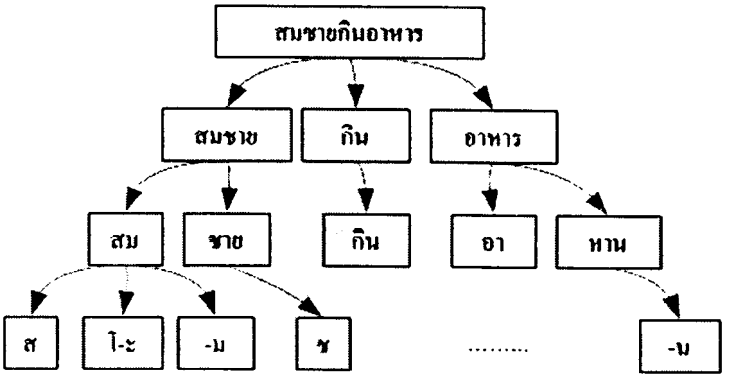
ภาพประกอบ 2.17 การสร้างแบบจำลองเสียงในขั้นตอน Single Re-estimation (ที่มา: Young, S., et al., 2000: 143)

ฝ่ายหอสมุด
คุณหญิงหลง อรรถกระวีสุนทร

2.8 ชุดหน่วยเสียง

ในระบบรู้จำเสียงพูดชุดหน่วยเสียงคือรายการเสียงย่อยสุดที่มีอยู่ในแบบจำลองเสียง โดยหลักไวยากรณ์ของภาษาไทยนั้นได้กำหนดองค์ประกอบของข้อความ คำ และ พยางค์ ดังภาพประกอบ 2.8

- ข้อความ มาจากคำหลายคำประกอบกัน
- คำ มาจากพยางค์หลายพยางค์ประกอบกัน
- พยางค์ มาจากพยัญชนะ สระ ตัวสะกด วรรณยุกต์ประกอบกัน



ภาพประกอบ 2.18 ส่วนประกอบของข้อความในภาษาไทย

จึงกล่าวได้ว่าพยัญชนะ สระ ตัวสะกด คือหน่วยเสียงที่ย่อยที่สุดในภาษาไทย เช่นเดียวกับงานวิจัยหน่วยเสียงภาษาไทย (Thongprasert, R., 2002) ได้กล่าวว่าเสียงย่อยสุดในภาษาไทยมี 3 ชนิดคือ เสียงพยัญชนะต้น เสียงสระ และ เสียงตัวสะกด จึงมีการนิยามหน่วยเสียงเดี่ยว (mono-phone) สำหรับใช้ในระบบรู้จำ โดยกำหนดให้เสียงพยางค์ภาษาไทยอยู่ในรูปแบบ C_i+V+C_f มีจำนวนหน่วยเสียงเดี่ยว 68 หน่วยเสียง ดังข้อมูลในตาราง 2.1 ดังนั้นระบบรู้จำเสียงพูดภาษาไทยจะต้องจดจำหน่วยเสียงเหล่านี้ได้ทั้งหมด

พยัญชนะต้น (C_i)			
เดี่ยว	ตัวอย่าง	ผสม	ตัวอย่าง
P	ปาก	pr	ประสาน
t	เต้น กุฏิ	phr	พราน
c	จะ	tr	เตรียม
k	ก่อน	kr	กราบ
z	อาน	khr	คร่า
ph	พบ ภัย ผ่าน	pl	ปลา
th	ทิ้ง ธง เฒ่า ฐาน มณโฑ	phl	พลาด
ch	ชอบ เฌอ	thr	จันทรา
kh	คน เขิน ฆ่า	kl	เกลอ
b	บอก	khl	เคลื่อน
d	ด้าน ชฎา	kw	กวาง
m	ไม่	khw	ขวา
n	นาน เณร	เสียงทับศัพท์	
ng	เงิน	br	เบรน
l	เล่น กีฬา	bl	บลู
r	รอ ฤทัย	fr	ฟราย
f	ฝน ฟัน	fl	เฟลม
s	สาย ศิลา	dr	ดรากอน
h	โหน เฮฮา	17 หน่วย	
w	ว่า		
j	ย้อน หญิง		
21 หน่วย			

สระ (V)			
เดี่ยว	ตัวอย่าง	ผสม	ตัวอย่าง
a	อะ	ia	เอียะ
aa	อา	iia	เอีย
i	อิ	va	เอือะ
ii	อี	vva	เอือ
v	อึ	ua	อัวะ
vv	อื	uua	อัว
u	อุ	6 หน่วย	
uu	อู		
e	เอะ		
ee	เอ		
x	แอะ		
xx	แอ		
o	โอะ		
oo	โอ		
@	เอาะ		
@@	ออ		
q	เออะ		
qq	เออ		
18 หน่วย			

ตัวสะกด (C_f)	
เดี่ยว	ตัวอย่าง
p^	พบ
t^	เกร็ด
k^	ปาก
n^	หาร
m^	ลม
ng^	ฟาง
j^	ยาย
w^	กาว
เสียงทับศัพท์	
f^	กราฟ
l^	แอล
s^	เอส
ch^	คลัช
12 หน่วย	

ตาราง 2.1 สัญลักษณ์หน่วยเสียงทั้งหมดในภาษาไทย (ที่มา: Thongprasert, R., 2002)

งานวิจัยนี้ได้ปรับปรุงหน่วยเสียงกลุ่มสระให้มีระดับวรรณยุกต์เพิ่มขึ้นอีก 5 ระดับเพื่อให้ผลการรู้จำครอบคลุมเสียงวรรณยุกต์ ทำให้หน่วยเสียงชุดใหม่ $C_i+V_t+C_n$ มีจำนวนหน่วยเสียง 170 หน่วยเสียง ดังแสดงในตาราง 2.2 หน่วยเสียงสระชุดเดิมที่กำหนดสัญลักษณ์เป็นตัวอักษรภาษาอังกฤษในตารางที่ 2.1 ต่างจากเสียงสระชุดใหม่ที่มีวรรณยุกต์กำกับ คือการเพิ่มตัวเลขบอกระดับวรรณยุกต์ต่อท้ายหน่วยเสียงสระ ดังแสดงในตารางที่ 2.3

พยัญชนะ	ก ค ง จ ช ด ต ท น บ ป พ ฟ ม ย ร ล ว ส อ ฮ กร กล คร คล คว ตร ทร ปร ปล พร พล บร บล ฟร ฟล คร
สระเดี่ยว 5 ระดับวรรณยุกต์	-ะ -า อิ อี อึ อือ อุ อู เ-ะ เ- แ-ะ แ- โ-ะ โ- เ-าะ -อ เ-อะ เ-อ เอียะ เอีย เออะ เออ อัวะ อัว (เพิ่ม 0-4 บอกระดับวรรณยุกต์)
ตัวสะกด	-บ -ด -ก -น -ม -ง -ย -ว -ฟ -ล -ส -ช

ตาราง 2.2 ชุดหน่วยเสียง $C_i+V_t+C_f$

รูปสระ	สระชุดเดิม	สระชุดใหม่
สระ อะ เสียงสามัญ	a	a0
สระ อะ เสียงเอก		a1
สระ อะ เสียงโท		a2
สระ อะ เสียงตรี		a3
สระ อะ เสียงจัตวา		a4

ตาราง 2.3 ตัวอย่างรูปแบบของหน่วยเสียงสระวรรณยุกต์ V_t

การปรับปรุงชุดหน่วยเสียง

1. ชุดหน่วยเสียง $C_i+V_t+C_n$

เนื่องจากตัวสะกดมีส่วนสำคัญในการระบุระดับวรรณยุกต์ ภายในงานวิจัยนี้จึงได้พัฒนาชุดหน่วยเสียงอีกชุดหนึ่งจากชุดหน่วยเสียง $C_i+V_t+C_f$ เพิ่มหน่วยเสียงตัวสะกดให้มีระดับวรรณยุกต์เพิ่มขึ้นอีก 5 ระดับเป็น 218 หน่วยเสียง ดังตาราง 2.4 ชุดหน่วยเสียงนี้จึงเป็นตัวเลือกหนึ่งสำหรับระบบรู้จำเสียงพูด

พยัญชนะ	ก ค ง จ ช ด ต ท น บ ป พ ฟ ม ย ร ล ว ส อ ฮ กร กล คร คล คว ตร ทร ปร ปล พร พล บร บล ฟร ฟล ดร
สระเดี่ยว 5 ระดับวรรณยุกต์	-ะ -า อิ อี อึ อือ อุ อู เ-ะ เ- แ-ะ แ- โ-ะ โ- เ-าะ -อ เ-อะ เ-อ เอียะ เอีย เอือะ เอือ อัวะ อัว (เพิ่ม _0-4 บอกระดับวรรณยุกต์)
ตัวสะกด 5 ระดับวรรณยุกต์	-บ -ด -ก -น -ม -ง -ย -ว -ฟ -ล -ส -ช (เพิ่ม _0-4 บอกระดับวรรณยุกต์)

ตาราง 2.4 หน่วยเสียงชุดตัวสะกดวรรณยุกต์ $C_i+V_t+C_{fi}$

2. ชุดหน่วยเสียงลดรูป (Reduced Phonetic Set) $C_{ir}+V_r+C_{fr}$

ปัญหาจากข้อมูลเสียง การสนทนาทั่วไป คำศัพท์บางคำอาจมีการออกเสียงไม่ตรงตามหลักภาษา แต่ผู้พูดและผู้ฟังเข้าใจตรงกัน ส่วนหนึ่งเป็นเพราะการออกเสียงคลาดเคลื่อนเพียงเล็กน้อย หรือเป็นความเคยชินของผู้พูด เช่น คำว่า มหาวิทยาลัย อ่านว่า มะ หา วิด ทะ ยา ลัย สังเกตคำว่า มะ ที่ต้องออกเสียงวรรณยุกต์ตรี แต่คนส่วนใหญ่ออกเสียงวรรณยุกต์สามัญ นอกจากนี้ยังมีเสียง ร และ ล ต้องออกเสียงชัดเจนระบบจึงจะรู้จำได้ถูกต้อง ความคลาดเคลื่อนในจุดนี้จะทำให้ชุดหน่วยเสียง $C_i+V_t+C_f$ ไม่สามารถแก้ปัญหาได้

จากปัญหาดังกล่าวผู้วิจัยจึงได้สร้างหน่วยเสียงอีกชุดหนึ่ง $C_{ir}+V_r+C_{fr}$ โดยรวมหน่วยเสียงที่มีความคล้ายคลึงกันเข้าด้วยกันและแยกหน่วยเสียงที่มีการออกเสียงต่างกันออกจากกันดังนี้

- แยกเสียงพยัญชนะควบออกจากกัน
- รวมสระเสียงสั้น สระเสียงยาว ทุกระดับวรรณยุกต์ เป็นหน่วยเสียงเดียว
- ตัดหน่วยเสียงตัวสะกดที่เป็นคำทับศัพท์ออก

เหลือหน่วยเสียงทั้งหมด 41 หน่วยเสียง ดังตาราง 2.5

พยัญชนะ	ก ค ง จ ช ด ต ท น บ ป พ ฟ ม ย ล ว ส อ ฮ
สระเดี่ยว	อะ อิ อึ อุ เอะ แอะ โอะ เอาะ เออะ เอียะ เอือะ อัวะ
ตัวสะกด	-บ -ด -ก -น -ม -ง -ย -ว

ตาราง 2.5 ชุดหน่วยเสียงภาษาไทยแบบ 41 หน่วยเสียง

หน่วยเสียงชุดนี้มีหน่วยเสียงไม่ครบถ้วนตามหลักภาษาไทย ทำให้ผลการรู้จำจากหน่วยเสียงชุดนี้ไม่สมบูรณ์พอที่จะระบุว่าเป็นคำใดในภาษาไทย จำเป็นต้องมีระบบจำแนกหน่วยเสียงในส่วนที่ขาดไป ซึ่งได้แก่ ส่วนวิเคราะห์ว่าเป็นเสียงสระสั้นหรือยาว เป็นวรรณยุกต์ใด เป็นเสียงตัวสะกดทับศัพท์หรือไม่ เป็นเสียงพยัญชนะเป็น ร หรือ ล และอื่นๆ

เป้าหมายอีกอย่างหนึ่งของหน่วยเสียงชุดนี้ คือการตรวจสอบความสามารถในการรู้จำวรรณยุกต์ของชุดหน่วยเสียงแรก โดยเปรียบเทียบความแม่นยำในการรู้จำสระของชุดหน่วยเสียงทั้งสอง ความแตกต่างที่ได้จะเป็นตัวชี้วัดว่าระบบรู้จำวรรณยุกต์ได้ดีเพียงใด

3. การจดจำเสียงรอบข้าง (context dependency)

ลักษณะของเสียงในภาษาไทยในช่วงการเปลี่ยนจากเสียงหนึ่งไปอีกเสียงหนึ่งนั้นจะมีความสัมพันธ์กัน เช่น เสียง "กิน" และ "การ" เสียง "ก" เหมือนกัน แต่การออกเสียงต่างกันเพราะอยู่ในหน่วยเสียงแวดล้อมที่ต่างกัน การจดจำเสียงรอบข้างจึงเป็นเทคนิคที่พิจารณาหน่วยเสียงปัจจุบันควบคู่กับหน่วยเสียงก่อนหน้าและหน่วยเสียงที่ตามหลัง เช่น เสียง "กิน" และ "การ" หากพิจารณาที่หน่วยเสียงแรก "ก" ระบบจะพิจารณาว่าตรงกับหน่วยเสียง "ก" ที่อยู่กับสระ "อิ" หรือ "ก" ที่อยู่กับสระ "า" ดังนั้นหากว่าเสียงปัจจุบันเป็นเสียงหนึ่งในสองนี้ผลการรู้จำจะเป็น "ก" ดังตารางที่ 2.6

เสียงที่เข้าสู่ระบบ	หน่วยเสียงระบบรู้จำได้	ผลการรู้จำ
เสียง ก ในคำว่า กิน	ก ที่ตามด้วยสระ อิ	ก
เสียง ก ในคำว่า การ	ก ที่ตามด้วยสระ อา	ก

ตาราง 2.6 ผลการรู้จำของหน่วยเสียงแบบจดจำเสียงรอบข้าง

วิธีนี้มีข้อเสียคือต้องเก็บรูปแบบการออกเสียงหลากหลายรูปแบบทำให้ต้องใช้ตัวอย่างเสียงจำนวนมากในการสร้างแบบจำลองเสียงขึ้นมา ซึ่งการพิจารณาเสียงปัจจุบันกับเสียงรอบข้างมีด้วยกัน 3 แบบคือ

1. พิจารณาเสียงปัจจุบันกับเสียงที่อยู่ก่อนหน้า (left context bi-phone)
2. พิจารณาเสียงปัจจุบันกับเสียงเสียงที่อยู่ตามหลัง (right context bi-Phone)
3. พิจารณาเสียงปัจจุบันกับเสียงกับทั้งเสียงที่อยู่ก่อนหน้า เสียงที่ตามหลัง (tri-phone)

การนำเทคนิคนี้ไปใช้จำเป็นต้องมีตัวอย่างเสียงจำนวนมากซึ่งข้อมูลเสียงที่ใช้ในงานวิจัยนี้มีจำนวนตัวอย่างต่อหน่วยเสียงไม่เพียงพอที่จะสร้างแบบจำลองเสียงที่ใช้ชุดหน่วยเสียงแบบจดจำเสียงรอบข้าง

2.9 พจนานุกรม (dictionary)

ในกระบวนการรู้จำเสียงพูด หลังจากขั้นตอนจำแนกหน่วยเสียง จะเป็นส่วนพจนานุกรมมีหน้าที่พิจารณากลุ่มหน่วยเสียงเทียบกับรายการคำศัพท์ที่มีอยู่ในพจนานุกรม โดยพิจารณาจากลำดับหน่วยเสียงและให้ผลเป็นคำศัพท์ ตัวอย่างเช่น พยัญชนะ "ว" สระ "อะ เสียงตรี" ตัวสะกด "-ด" เมื่อผ่านพจนานุกรมแล้ว ระบบจะให้คำตอบเป็นคำว่า "วัด" คำเดียว

ขั้นตอนนี้เป็นการกำหนดคำศัพท์ที่ระบบรู้จำได้ จากการกำหนดค่าในพจนานุกรมโดยระบุข้อความด้านขวาเป็นรายการคำศัพท์ที่รู้จำได้ ส่วนข้อความด้านซ้ายเป็นลำดับหน่วยเสียงของคำศัพท์เหล่านั้น ดังตาราง 2.7

รูปเสียง	ลำดับหน่วยเสียง
มหานคร	m a3 h aa4 n a3 kh @@0 n^
อาราม	z aa0 r aa0 m^
วัด	w a3 t^

ตาราง 2.7 ตัวอย่างพจนานุกรมคำศัพท์

เทคนิคในการปรับปรุงพจนานุกรม

หากปรับปรุงพจนานุกรมสามารถทำได้หลายรูปแบบ เช่น ทำให้ระบบรู้จำสามารถรู้จำเสียงในระดับต่างๆได้ เช่น การรู้จำหน่วยเสียงสามารถทำได้โดยการกำหนดพจนานุกรมให้ รูปเสียงและลำดับหน่วยเสียง เป็นรายการหน่วยเสียงทั้งคู่ จะได้ผลการรู้จำเป็นหน่วยเสียงออกมา ทำให้ทราบถึงความถูกต้องของระบบในการจดจำหน่วยเสียงได้ ดังนั้นขั้นตอนนี้จึงเป็นการกำหนดชุดข้อมูลที่ระบบรู้จำได้ นอกจากนี้การรู้จำพยางค์ สามารถทำได้โดยให้รูปเสียงเป็นชื่อพยางค์ ส่วนลำดับหน่วยเสียงเป็นลำดับหน่วยเสียงของพยางค์นั้นๆ ดังตาราง 2.9

การกำหนดพจนานุกรมที่เหมาะสมจะช่วยในการรู้จำเสียงพูดในระดับที่ต่างกันเพื่อวิเคราะห์หาสาเหตุของความผิดพลาดในการรู้จำเสียงพูดได้

รูปเสียง	ลำดับหน่วยเสียง
ม	m
อ๊ะ	a3
ห	h
อ๋ำ	aa4

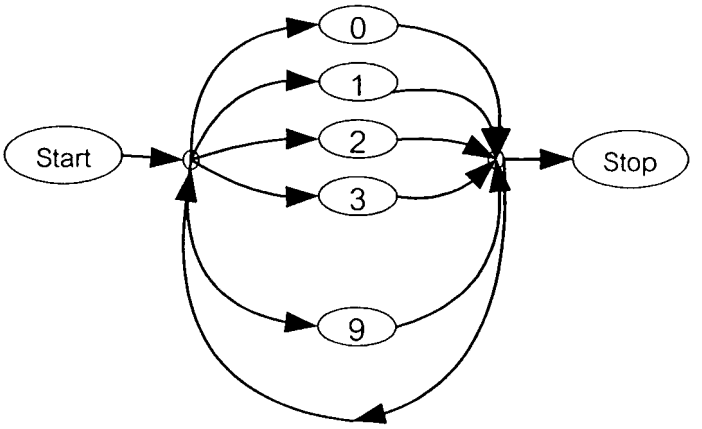
ตาราง 2.8 ตัวอย่างพจนานุกรมหน่วยเสียง

รูปเสียง	ลำดับหน่วยเสียง
มะ	m a3
หา	h aa4
นะ	n a3
คอน	kh @@0 n^

ตาราง 2.9 ตัวอย่างพจนานุกรมพยางค์

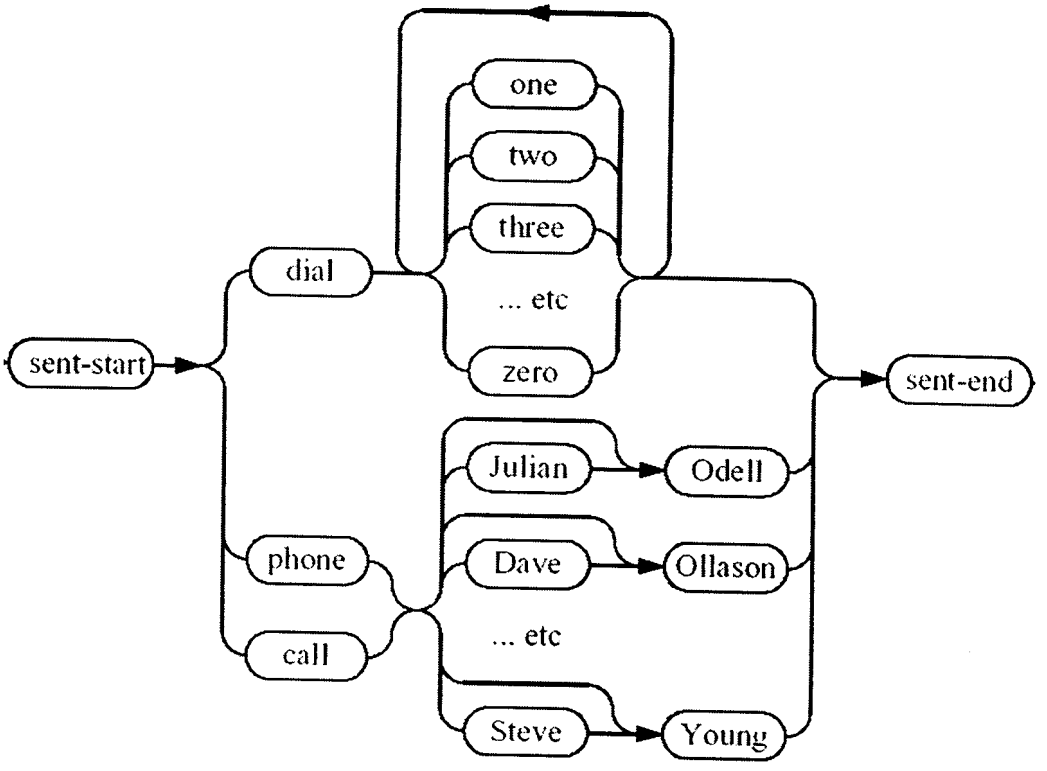
2.10 ไวยากรณ์ (grammar model)

ส่วนไวยากรณ์หรือส่วนสร้างประโยค มีหน้าที่ตรวจสอบลำดับของคำศัพท์ที่เกิดขึ้นและควบคุมให้ลำดับของคำศัพท์ที่เกิดขึ้นนั้นตรงตามกฎที่ผู้ใช้กำหนด หรืออีกนัยหนึ่งคือผู้ใช้สามารถกำหนดลำดับของคำศัพท์ให้เหมาะสมกับการรู้จำได้



ภาพประกอบ 2.19 ตัวอย่างไวยากรณ์สำหรับการรู้จำตัวเลข

จากภาพประกอบ 2.19 เป็นตัวอย่างการกำหนดไวยากรณ์สำหรับการรู้จำเสียงพูดที่เป็นตัวเลขทั่วไป ผลการรู้จำเสียงตัวเลขตัวแรกจากตัวเลขทั้งหมด และหลังจากเลขตัวแรกก็ยังมีโอกาสเกิดตัวเลขถัดไปอย่างไม่จำกัด ตัวอย่างเช่น 1 2 4 7 9 5 0 หรือ 1 3 หรือ 5 เป็นต้น สังเกตลูกศรย้อนกลับหมายถึงการเรียกซ้ำอย่างไม่รู้จบ หากไม่มีลูกศรย้อนกลับก็หมายความว่า เป็นการรู้จำเป็นตัวเลขเพียงตัวเดียวเท่านั้น



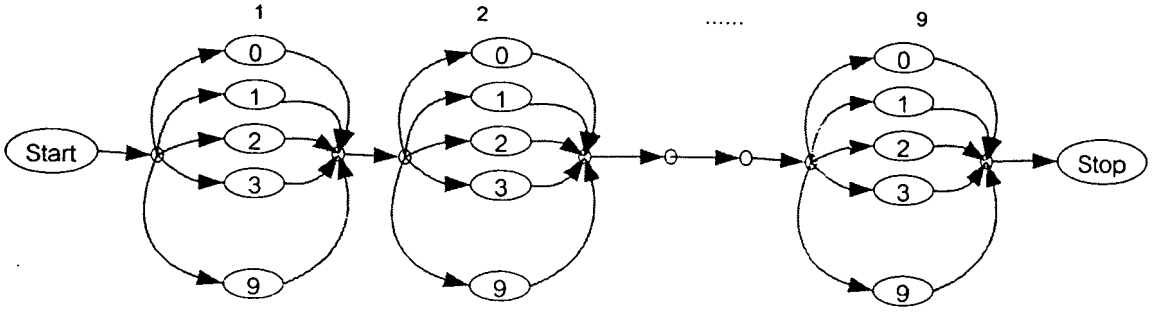
ภาพประกอบ 2.20 ตัวอย่างไวยากรณ์สำหรับการต่อโทรศัพท์

อีกตัวอย่างหนึ่งคือการติดต่อโทรศัพท์ โดยการกำหนดลำดับของคำศัพท์ตั้งแต่ การโทรออก หมายเลขโทรศัพท์ และชื่อผู้รับ ดังภาพประกอบ 2.20 ตัวอย่างผลการรู้จำที่เกิดขึ้นได้คือ dial nine one one หรือ call Julian Odel หรือ call Odel สำหรับเส้นที่กระโดดข้ามวงคำพูดใดๆ หมายถึงคำพูดในวงวงนั้นจะปรากฏในผลการรู้จำหรือไม่ก็ได้

จะเห็นได้ว่าส่วนไวยากรณ์ภาษาเป็นส่วนที่มีความซับซ้อนสูงสุดเนื่องจากระบบต้องเลือกคำศัพท์ที่อยู่ในกฎมาจัดเรียงให้อยู่ในรูปแบบไวยากรณ์ที่กำหนด หากไวยากรณ์มีความซับซ้อนมากก็จะทำให้ระบบใช้เวลาในการประมวลผลนาน และอาจผิดพลาดได้ง่ายหากแบบจำลองเสียงไม่มีความแม่นยำสูงพอ

เทคนิคการปรับปรุงไวยากรณ์

ความแม่นยำในการรู้จำนอกจากจะขึ้นอยู่กับแบบจำลองเสียง แล้วยังขึ้นอยู่กับวิธีการในการให้คำตอบของระบบรู้จำเช่นกัน ซึ่งวิธีการให้คำตอบนี้จะขึ้นอยู่กับพจนานุกรม และ ไวยากรณ์ ตัวอย่างเช่นการรู้จำหมายเลขโทรศัพท์ โดยปกติพจนานุกรมมีคำศัพท์ที่เป็นตัวเลข ศูนย์ถึงเก้า ไวยากรณ์จะให้คำตอบเป็นตัวเลขทุกตัวไม่จำกัดจำนวน แต่ความจริงแล้วเบอร์โทรศัพท์จะมีจำนวนตัวเลขที่คงที่คือ 9 ตัว ซึ่งสามารถกำหนดไวยากรณ์ให้คำตอบเป็นตัวเลขทุกตัวแบบจำกัดจำนวนเป็น 9 ตัวได้ ซึ่งจะช่วยแก้ปัญหาการรู้จำตัวเลขเกิน 9 หลักได้ดังภาพประกอบ 2.21 จะเห็นได้ว่าการปรับปรุงไวยากรณ์และพจนานุกรมมีส่วนช่วยให้การรู้จำมีความแม่นยำสูงขึ้น



ภาพประกอบ 2.21 ตัวอย่างไวยากรณ์สำหรับการรู้จำหมายเลขโทรศัพท์แบบตัวเลข 9 ตัว

ดังนั้นในการปรับปรุงไวยากรณ์ต้องดูความเหมาะสมกับแบบจำลองเสียง หากไวยากรณ์มีความซับซ้อนมากแล้วแบบจำลองมีความแม่นยำต่ำก็จะให้ผลการรู้จำที่ผิดพลาดยิ่งกว่าการใช้ไวยากรณ์ที่ซับซ้อนน้อยกว่า