

หลักการและทฤษฎีสำคัญ

รายละเอียดในบทที่ 2 นี้จะกล่าวถึงทฤษฎีต่างๆที่เกี่ยวข้องและอ้างอิงถึงในงานวิจัยนี้ โดยเริ่มจากทฤษฎีที่เกี่ยวข้องกับการประมวลผลสัญญาณเบื้องต้น การวิเคราะห์และวัดค่าลักษณะสำคัญและการจำแนกรูปแบบโดยใช้โครงข่ายประสาทเทียม กลศาสตร์และภาษาศาสตร์ และการรู้จำแบบรูปโดยวิธีเชิงสถิติ

2.1 การประมวลผลสัญญาณเบื้องต้น (Signal Preprocessing)

2.1.1 การมวธีเน้นล่วงหน้า (Preemphasis)

การมวธีเน้นล่วงหน้าเป็นขั้นตอนแรกที่กระทำกับสัญญาณเสียง $s(n)$ โดยการผ่านวงจรกรอง FIR อันดับหนึ่งที่มีฟังก์ชันถ่ายโอนดังสมการที่ (2.1) และ (2.2) เพื่อปรับสเปกตรัมให้มีความเรียบ เมื่อ α เป็นค่าสัมประสิทธิ์ของวงจรกรอง $\tilde{s}(n)$ เป็นค่าของสัญญาณเสียงพูดขาออกที่ผ่านกรวมวธีเน้นล่วงหน้าที่ n $s(n)$ เป็นค่าของสัญญาณเสียงพูดขาเข้าที่ n และ $s(n-1)$ เป็นค่าของสัญญาณเสียงพูดขาเข้าที่ $n-1$

$$H(z) = 1 - \alpha z^{-1} \tag{2.1}$$

$$\tilde{s}(n) = s(n) - \alpha s(n-1) \tag{2.2}$$

ค่าสัมประสิทธิ์ของวงจรกรอง α จะถูกนิยามกำหนดให้เท่ากับ 0.95 (Rabiner and Juang, 1993)

2.1.2 กรวมวธีวางกรอบสัญญาณ (Smoothing Window)

สัญญาณที่ผ่านกรวมวธีเน้นล่วงหน้าจะถูกแบ่งออกเป็นกรอบข้อมูลเสียงพูดโดยการคูณสัญญาณเสียงพูดกับฟังก์ชันกรอบ (Window Function) ผลของการใช้ฟังก์ชันกรอบคือการลดทอนแอมพลิจูดอย่างช้าๆที่บริเวณปลายแต่ละข้างของกรอบข้อมูลเสียงพูดเพื่อหลีกเลี่ยงความไม่ต่อเนื่องที่จุดปลาย กำหนดให้ฟังก์ชันกรอบคือ $w(n)$ มีค่าอยู่ในช่วง $0 \leq n \leq N-1$ โดยที่ N คือ ความกว้างของฟังก์ชันกรอบ สัญญาณที่ผ่านฟังก์ชันกรอบคือ

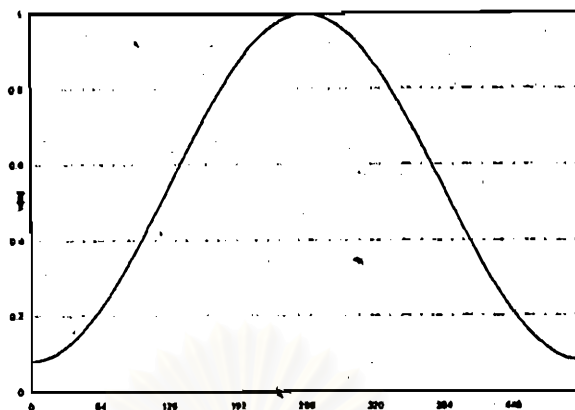
$$\tilde{x}_i(n) = x_i(n)w(n) \quad 0 \leq n \leq N-1 \tag{2.3}$$

ฟังก์ชันกรอบที่นิยมใช้สำหรับการวิเคราะห์เสียงพูดคือฟังก์ชันกรอบชนิด Hamming Window ซึ่งแสดงในรูปที่ 2.1 และมีรูปแบบตามสมการที่ (2.4)

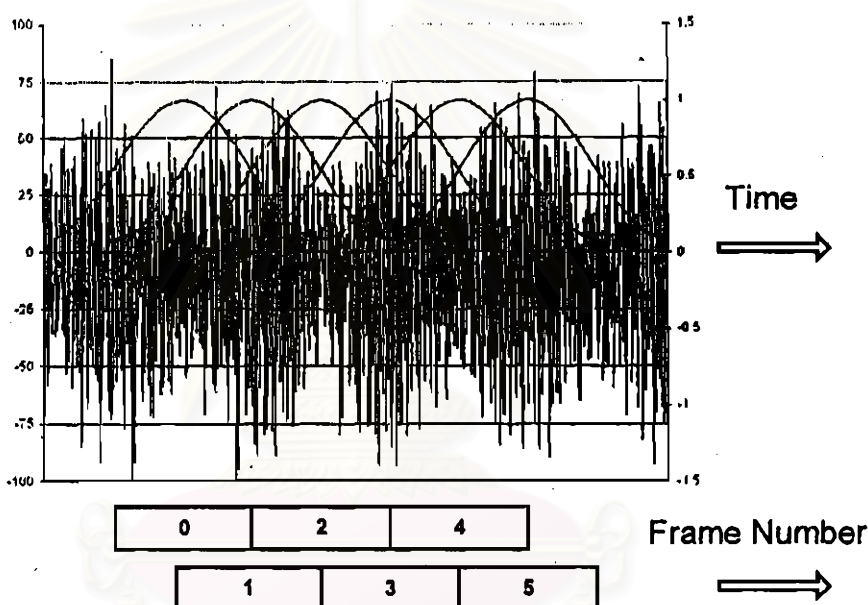
$$w(n) = 0.54 - 0.46 \cos\left(\frac{2\pi n}{N-1}\right) \quad 0 \leq n \leq N-1 \tag{2.4}$$

สัญญาณที่ได้จากฟังก์ชันกรอบจะมีช่วงประสิทธิผลอยู่ค่าหนึ่ง ดังนั้นสัญญาณในแต่ละกรอบข้อมูลเสียงพูดจะมีจำนวนตัวอย่างเท่ากับช่วงประสิทธิผลของความยาวฟังก์ชันกรอบ ฟังก์ชันกรอบจะถูกเลื่อนครั้งละ M

ตัวอย่าง โดยที่ M มีค่าน้อยกว่าช่วงประสิทธิผลของฟังก์ชันกรอบเพื่อให้การเปลี่ยนแปลงของสัญญาณเป็นไปอย่างช้าๆ ดังแสดงในรูปที่ 2.2



รูปที่ 2.1 ฟังก์ชันกรอบชนิด Hamming Window



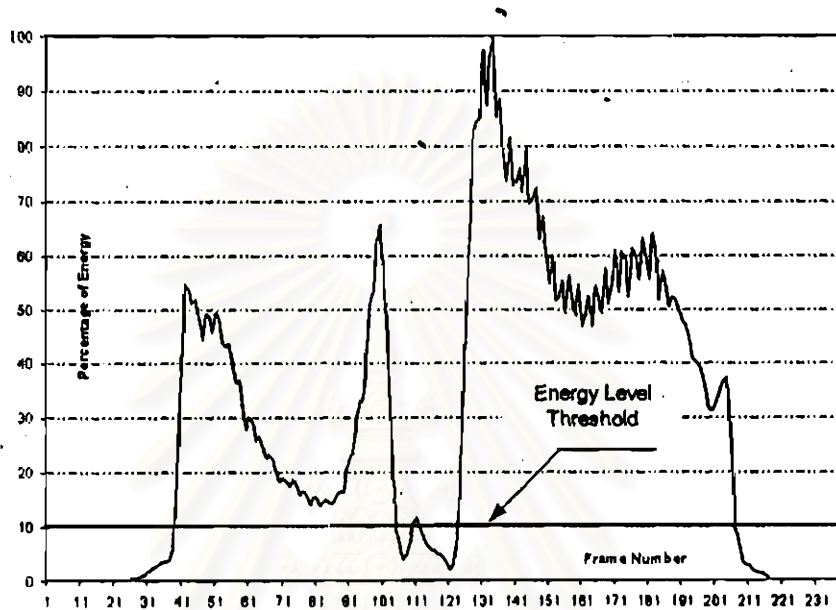
รูปที่ 2.2 การแบ่งกรอบข้อมูลเสียงพูดด้วยฟังก์ชันกรอบชนิด Hamming Window

2.1.3 การหาจุดสิ้นสุดเสียงพูด (Endpoint Detection)

ขั้นตอนการหาจุดสิ้นสุดเสียงพูดเป็นการแยกเสียงพูดออกจากสัญญาณรบกวนต่างๆ เพื่อนำส่วนที่เป็นสัญญาณเสียงพูดมาวิเคราะห์และทำการรู้จำ การหาจุดสิ้นสุดเสียงพูดโดยอาศัยแผนภูมิระดับเส้นพลังงานเป็นวิธีการที่นิยมใช้ (วิศรุต อาชุนทร, 2539; วุฒิพงษ์ พรสุขจันทร์, 2539; ชัย วุฒิวีวัฒน์ชัย; 2540) ดังนั้นในงานวิจัยนี้จะอาศัยแผนภูมิระดับเส้นพลังงาน (Energy Level Contour) ของเสียงพูดแต่ละกรอบข้อมูลเสียงพูดดังสมการที่ (2.5) โดย $E_x(m)$ เป็นค่าระดับพลังงานของกรอบข้อมูลเสียงพูดที่ m $s(n)$ เป็นค่าสัญญาณที่ n และ N เป็นจำนวนตัวอย่างเสียงพูดของกรอบข้อมูลเสียงพูดที่ m (Deller, Proakis, and Hansen, 1993)

$$E_x(m) = \sum_{n=0}^{N-1} s^2(n) \quad (2.5)$$

ขั้นตอนการมีวิธีหาขอบเขตพยางค์จะทำการกำหนดจุดเริ่มเปลี่ยนระดับพลังงาน (Energy Level Threshold) โดยกำหนดเป็นร้อยละของระดับพลังงานสูงสุด การกำหนดจุดเริ่มต้นกรอบข้อมูลเสียงพูดส่วนจะนับจากจุดที่มีการเปลี่ยนระดับพลังงานมากกว่าจุดเริ่มเปลี่ยนระดับพลังงาน การกำหนดจุดสิ้นสุดกรอบข้อมูลเสียงพูดจะนับจากจุดที่มีการเปลี่ยนระดับพลังงานต่ำกว่าจุดเริ่มเปลี่ยนระดับพลังงานดังแสดงในรูปที่ 2.3



รูปที่ 2.3 การหาจุดเริ่มต้นและสิ้นสุดของพยางค์โดยวิธีพลังงานของเสียงพูดคำว่า "มีอ้าย"

2.2 การวิเคราะห์และวัดค่าลักษณะสำคัญ (Feature Measurement)

การวิเคราะห์และวัดค่าลักษณะสำคัญเป็นวิธีการหาตัวแทนของสัญญาณเสียงพูดเพื่อวิเคราะห์และจำแนกเสียงพูดที่มีความแตกต่างกันสำหรับระบบรู้จำเสียงพูด ค่าลักษณะสำคัญนอกจากจะช่วยในการวิเคราะห์และจำแนกเสียงพูดแล้วยังเป็นตัวแทนของเสียงพูดได้โดย ไม่จำเป็นต้องนำเสียงสัญญาณพูดเข้าระบบรู้จำโดยตรง เป็นการประหยัดหน่วยความจำและเวลาในการรู้จำ วิธีการหาคุณลักษณะสำคัญของสัญญาณเสียงพูดมีหลายวิธีเช่น แบบจำลองการประมาณพหุเชิงเส้น ความถี่มูลฐาน ความถี่ฟอร์แมนท์ เป็นต้น

2.2.1 สัมประสิทธิ์การประมาณพหุเชิงเส้น (Linear Prediction Coefficient)

ข้อดีของการใช้แบบจำลองการประมาณพหุเชิงเส้นสำหรับการรู้จำเสียงพูดมีดังนี้ (Rabiner and Juang, 1993)

1. การประมาณพหุเชิงเส้นเป็นแบบจำลองที่ดีสำหรับสัญญาณเสียงพูดในส่วนที่เป็น quasi steady state voiced ซึ่งแบบจำลอง all-pole ของการประมาณพหุเชิงเส้นจะให้การประมาณสเปกตรัมของช่องทางเดินเสียงที่ดี แต่บริเวณเสียงโอดยะและช่วงที่มีการเปลี่ยนแปลงแบบชั่วคราวของ

สัญญาณเสียงพูดจะให้ผลที่ต่ำกว่าบริเวณเสียงโฆระ แต่แบบจำลองการประมาณพหุเชิงเส้นยังคงเป็นแบบจำลองที่สามารถใช้งานได้ดีสำหรับงานรู้จำเสียงพูด

2. การใช้แบบจำลองการประมาณพหุเชิงเส้นสามารถวิเคราะห์ช่องทางเดินเสียงและสามารถเป็นตัวแทนคุณลักษณะของช่องทางเดินเสียงได้เป็นอย่างดี
3. การใช้แบบจำลองการประมาณพหุเชิงเส้นเป็นวิธีการทางคณิตศาสตร์ที่มีความแม่นยำและง่ายสำหรับการประยุกต์ใช้งานจริง
4. แบบจำลองการประมาณพหุเชิงเส้นให้ผลดีสำหรับการรู้จำเสียงพูด ระบบรู้จำเสียงพูดจำนวนมากใช้แบบจำลองการประมาณพหุเชิงเส้นสำหรับการประมวลผลส่วนหน้า (Front-end Processing)

แบบจำลองการประมาณพหุเชิงเส้น (Linear Prediction Model)

สัญญาณเสียง $s(n)$ ที่เวลา n สามารถประมาณได้ด้วยผลรวมเชิงเส้นของสัญญาณเสียงในอดีต p ตัวอย่างดังสมการ

$$s(n) \approx a_1 s(n-1) + a_2 s(n-2) + \dots + a_p s(n-p) \quad (2.7)$$

โดยค่าสัมประสิทธิ์ a_1, a_2, \dots, a_p มีค่าคงที่ตลอดการรอบข้อมูลเสียงพูดที่วิเคราะห์ เปลี่ยนสมการที่ (2.7) โดยมีพจน์กระตุ้น (Excitation Term) $Gu(n)$

$$s(n) = \sum_{i=1}^p a_i s(n-i) + Gu(n) \quad (2.8)$$

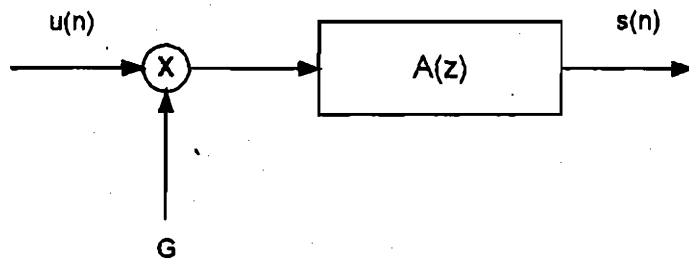
โดยที่ $u(n)$ เป็นการกระตุ้นที่ถูกทำให้เป็นบรรทัดฐานเดียวกัน (Normalized Excitation) และ G เป็นอัตราขยายของการกระตุ้น วิเคราะห์สมการที่ (2.8) ในโดเมน z จะได้ความสัมพันธ์

$$S(z) = \sum_{i=1}^p a_i z^{-i} S(z) + GU(z) \quad (2.9)$$

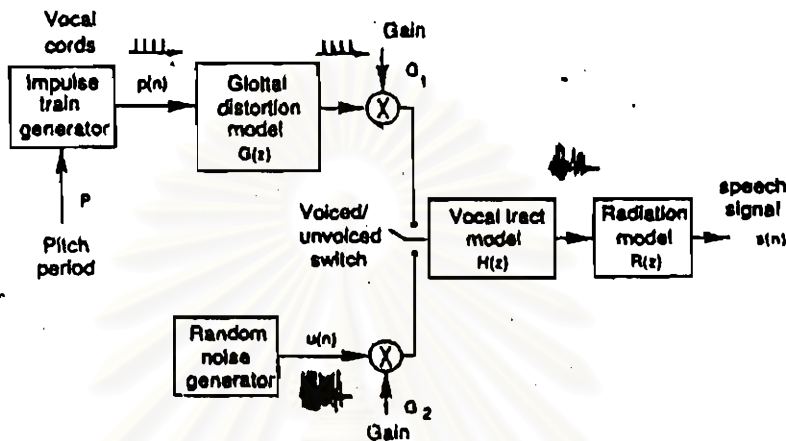
และฟังก์ชันการถ่ายโอน

$$H(z) = \frac{S(z)}{GU(z)} = \frac{1}{1 - \sum_{i=1}^p a_i z^{-i}} = \frac{1}{A(z)} \quad (2.10)$$

ในรูปที่ 2.4 อธิบายความหมายของสมการที่ (2.10) โดยที่ $u(n)$ ถูกคูณด้วย G เป็นสัญญาณเข้าของระบบ all-pole $H(z) = \frac{1}{A(z)}$ เกิดเป็นสัญญาณเสียงพูด $s(n)$ ฟังก์ชันการกระตุ้นของเสียงพูดเป็นชววนคลื่นกึ่งรายคาบ (Quasiperiodic) สำหรับเสียงโฆระและฟังก์ชันการกระตุ้นของเสียงพูดเป็นแหล่งสัญญาณรบกวนแบบสุ่มสำหรับเสียงอโฆระ แบบจำลองที่เหมาะสมสำหรับการสังเคราะห์เสียงพูดที่มีความสอดคล้องกับการประมาณพหุเชิงเส้นแสดงในรูปที่ 2.5



รูปที่ 2.4 แบบจำลองการประมาณพันธะเชิงเส้นสำหรับเสียงพูด (Rabiner and Juang, 1993)



รูปที่ 2.5 แบบจำลองสังเคราะห์เสียงพูดจากแบบจำลองการประมาณพันธะเชิงเส้น (Vuuren, 1998)

สมการวิเคราะห์การประมาณพันธะเชิงเส้น

จากรูปที่ 2.4 ความสัมพันธ์ระหว่าง $s(n)$ และ $u(n)$ เป็นดังนี้

$$s(n) = \sum_{k=1}^p a_k s(n-k) + Gu(n) \tag{2.11}$$

พิจารณาลรวมเชิงเส้นของตัวอย่างสัญญาณเสียงพูดในอดีตเพื่อที่จะประมาณ $\tilde{s}(n)$

$$\tilde{s}(n) = \sum_{k=1}^p a_k s(n-k) \tag{2.12}$$

ค่าผิดพลาดจากการประมาณ $e(n)$

$$e(n) = s(n) - \tilde{s}(n) = s(n) - \sum_{k=1}^p a_k s(n-k) \tag{2.13}$$

ฟังก์ชันการถ่ายโอนของค่าผิดพลาดจากการประมาณ

$$A(z) = \frac{E(z)}{S(z)} = 1 - \sum_{k=1}^p a_k z^{-k} \tag{2.14}$$

เมื่อสัญญาณเสียงพูด $s(n)$ เกิดจากระบบเชิงเส้นดังแสดงในรูปที่ 2.4 ค่าผิดพลาดจากการประมาณ $e(n)$ จะเท่ากับ $Gu(n)$

ปัญหาพื้นฐานของการวิเคราะห์การประมาณพันธะเชิงเส้นคือการคำนวณค่าสัมประสิทธิ์การทำนาย $\{a_k\}$ โดยตรงจากสัญญาณเสียงพูดที่ทำให้คุณสมบัติของสเปกตรัมของวงจรวงจรองเชิงเลขที่แสดงในรูปที่ 2.5

สอดคล้องกับรูปคลื่นสัญญาณเสียงภายในกรอบข้อมูลเสียงพูดที่วิเคราะห์ คุณลักษณะสเปกตรัมของสัญญาณเสียงพูดจะแปรเปลี่ยนไปตามเวลา ค่าสัมประสิทธิ์การทำนายที่เวลา n จะเกิดจากการประมาณส่วนย่อยของสัญญาณเสียงพูดในเวลาสั้นๆ ที่เวลา n การคำนวณค่าสัมประสิทธิ์การทำนายคือการทำหาค่าผิดพลาดกำลังสองเฉลี่ยของค่าจากการทำนายมีค่าต่ำที่สุดที่ส่วนย่อยของสัญญาณเสียงพูดนั้น

กำหนดสัญญาณเสียงพูดช่วงเวลาสั้นๆ (short-term speech) และค่าผิดพลาดในส่วนย่อย (error segment) ที่เวลา n เป็น

$$s_n(m) = s(n+m) \quad (2.15)$$

$$e_n(m) = e(n+m) \quad (2.16)$$

ต้องการหาค่าผิดพลาดกำลังสองเฉลี่ยของค่าของสัญญาณที่เวลา n มีค่าต่ำที่สุด

$$E_n = \sum_m e_n^2(m) \quad (2.17)$$

แทนค่า $e_n(m)$ ในพจน์ของ $s_n(m)$ จะได้

$$E_n = \sum_m \left[s_n(m) - \sum_{k=1}^p a_k s_n(m-k) \right]^2 \quad (2.18)$$

ในการแก้สมการที่ (2.18) สำหรับค่าสัมประสิทธิ์การทำนายจะต้องหาค่าอนุพันธ์ E_n เทียบกับค่าสัมประสิทธิ์ a_k และให้เท่ากับศูนย์

$$\frac{\partial E_n}{\partial a_k} = 0, \quad k = 1, 2, \dots, p \quad (2.19)$$

จะได้

$$\sum_m s_n(m-i)s_n(m) = \sum_{k=1}^p \hat{a}_k \sum_m s_n(m-i)s_n(m-k) \quad (2.20)$$

สมการทางขวามือ $\sum_m s_n(m-i)s_n(m-k)$ คือความแปรปรวนร่วมของ $s_n(m)$

$$\phi_n(i, k) = \sum_m s_n(m-i)s_n(m-k) \quad (2.21)$$

สมการที่ (2.20) เขียนในรูปแบบใหม่คือ

$$\phi_n(i, 0) = \sum_{k=1}^p \hat{a}_k \phi_n(i, k) \quad (2.22)$$

ซึ่งเป็นชุดของ p สมการ p ตัวแปร ค่าผิดพลาดกำลังสองเฉลี่ย E_n ที่มีค่าต่ำที่สุดเขียนได้เป็น

$$E_n = \sum_m s_n^2(m) - \sum_{k=1}^p \hat{a}_k \sum_m s_n(m)s_n(m-k) \quad (2.23)$$

$$= \phi_n(0, 0) - \sum_{k=1}^p \hat{a}_k \phi_n(0, k) \quad (2.24)$$

ค่าเฉลี่ยของกำลังสองของค่าผิดพลาดที่มีค่าต่ำที่สุดประกอบด้วยพจน์ที่มีค่าคงที่ $\phi_n(0, 0)$ และพจน์ที่ขึ้นอยู่กับค่าสัมประสิทธิ์การทำนาย

ในการแก้สมการที่ (2.22) เพื่อให้ได้ค่าสัมประสิทธิ์การทำนายที่เหมาะสมที่สุดจะต้องคำนวณ $\phi_n(i, k)$ สำหรับ $1 \leq i \leq p$ และ $0 \leq k \leq p$ และทำการแก้สมการจำนวน p สมการ ในทางปฏิบัติวิธี

การแก้สมการมี 2 วิธีคือ วิธีอัตโนมัติสัมพันธ์และวิธีความแปรปรวนร่วม ในงานวิจัยนี้จะกล่าวถึงวิธีอัตโนมัติสัมพันธ์เท่านั้น

วิธีอัตโนมัติสัมพันธ์ (Autocorrelation Method)

วิธีการที่ง่ายและตรงไปตรงมาในการกำหนดค่า m คือการสมมติให้ส่วนย่อยของสัญญาณเสียง $s_n(m)$ จะมีค่าเท่ากับศูนย์นอกช่วง $0 \leq m \leq N-1$ โดยที่ N เป็นความกว้างของฟังก์ชันกรอบ ซึ่งสอดคล้องกับการคูณสัญญาณเสียงพูด $s(m+n)$ ด้วยฟังก์ชันกรอบขนาดจำกัด $w(m)$ ซึ่งจะให้ค่าศูนย์นอกช่วง $0 \leq m \leq N-1$ ดังนั้นสัญญาณเสียงที่นำมาวิเคราะห์จะเขียนในรูปแบบสมการคือ

$$s_n(m) = \begin{cases} s(m+n)w(m), & 0 \leq m \leq N-1 \\ 0, & \text{otherwise.} \end{cases} \quad (2.25)$$

จากสมการที่ (2.25) ค่าผิดพลาดกำลังสองเฉลี่ยคือ

$$E_n = \sum_{m=0}^{N-1+p} e_n^2(m) \quad (2.26)$$

และ

$$\phi_n(i, k) = \sum_{m=0}^{N-1+p} s_n(m-i)s_n(m-k) \quad \begin{matrix} 1 \leq i \leq p \\ 0 \leq k \leq p \end{matrix} \quad (2.27)$$

หรือ

$$\phi_n(i, k) = \sum_{m=0}^{N-1-(i-k)} s_n(m)s_n(m+i-k) \quad \begin{matrix} 1 \leq i \leq p \\ 0 \leq k \leq p \end{matrix} \quad (2.28)$$

ในสมการที่ (2.28) เป็นฟังก์ชันของ $i-k$ ดังนั้นฟังก์ชันความแปรปรวนร่วม $\phi_n(i, k)$ จะลดรูปเป็นฟังก์ชันอัตโนมัติสัมพันธ์อย่างง่าย

$$\phi_n(i, k) = r_n(i-k) = \sum_{m=0}^{N-1-(i-k)} s_n(m)s_n(m+i-k) \quad (2.29)$$

เนื่องจากฟังก์ชันอัตโนมัติสัมพันธ์มีความสมมาตรคือ $r_n(-k) = r_n(k)$ สมการการประมาณพหุคูณเชิงเส้นเขียนได้เป็น

$$\sum_{k=1}^p r_n(i-k)\hat{a}_k = r_n(i), \quad 1 \leq i \leq p \quad (2.30)$$

และเขียนเป็นรูปแบบเมตริกซ์คือ

$$\begin{bmatrix} r_n(0) & r_n(1) & r_n(2) & \cdot & \cdot & \cdot & r_n(p-1) \\ r_n(1) & r_n(0) & r_n(1) & \cdot & \cdot & \cdot & r_n(p-2) \\ r_n(2) & r_n(1) & r_n(0) & \cdot & \cdot & \cdot & r_n(p-3) \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ r_n(p-1) & r_n(p-2) & r_n(p-3) & \cdot & \cdot & \cdot & r_n(0) \end{bmatrix} \begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \\ \cdot \\ \cdot \\ \hat{a}_p \end{bmatrix} = \begin{bmatrix} r_n(1) \\ r_n(2) \\ r_n(3) \\ \cdot \\ \cdot \\ r_n(p) \end{bmatrix} \quad (2.31)$$

เมตริกซ์ของค่าอัตโนมัติสัมพันธ์ อยู่ในรูปของเมตริกซ์ Toeplitz ซึ่งมีลักษณะสมมาตร และทุกๆ สมาชิกในแนวทแยงมุมมีค่าเท่ากัน ทำให้การแก้สมการเพื่อหาค่าสัมประสิทธิ์ a_k โดยวิธีนี้ง่ายกว่าการใช้วิธีความแปรปรวนร่วม (Covariance) วิธีหนึ่งที่เหมาะสมในการแก้สมการที่อยู่ในรูปเมตริกซ์ Toeplitz คือ ขั้นตอนวิธีการ Levinson-Durbin

ขั้นตอนวิธีการ Levinson-Durbin

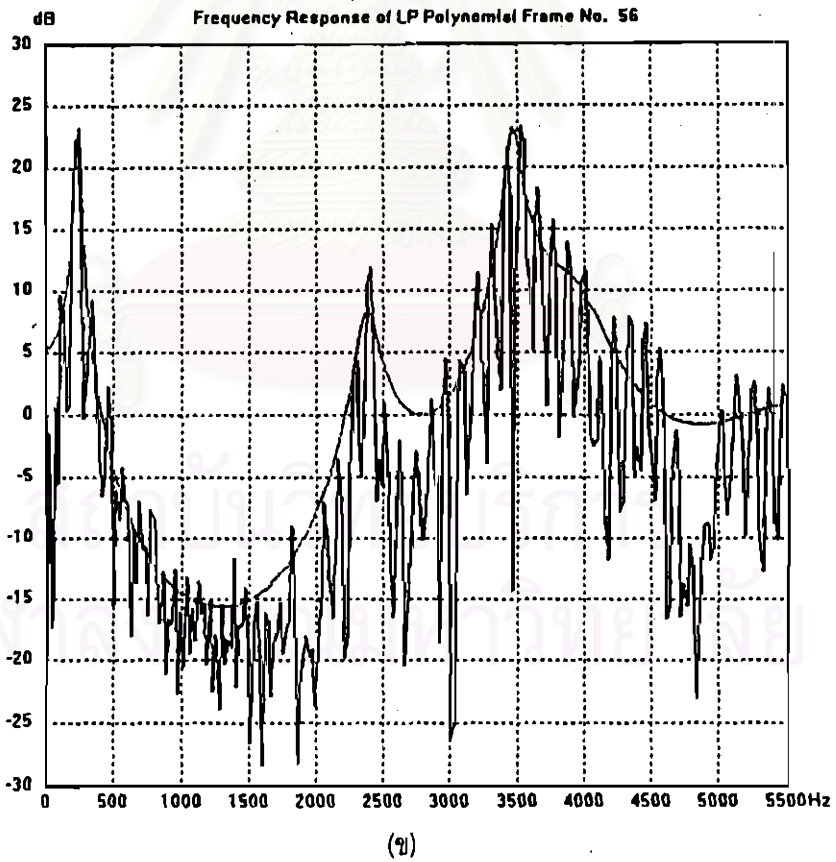
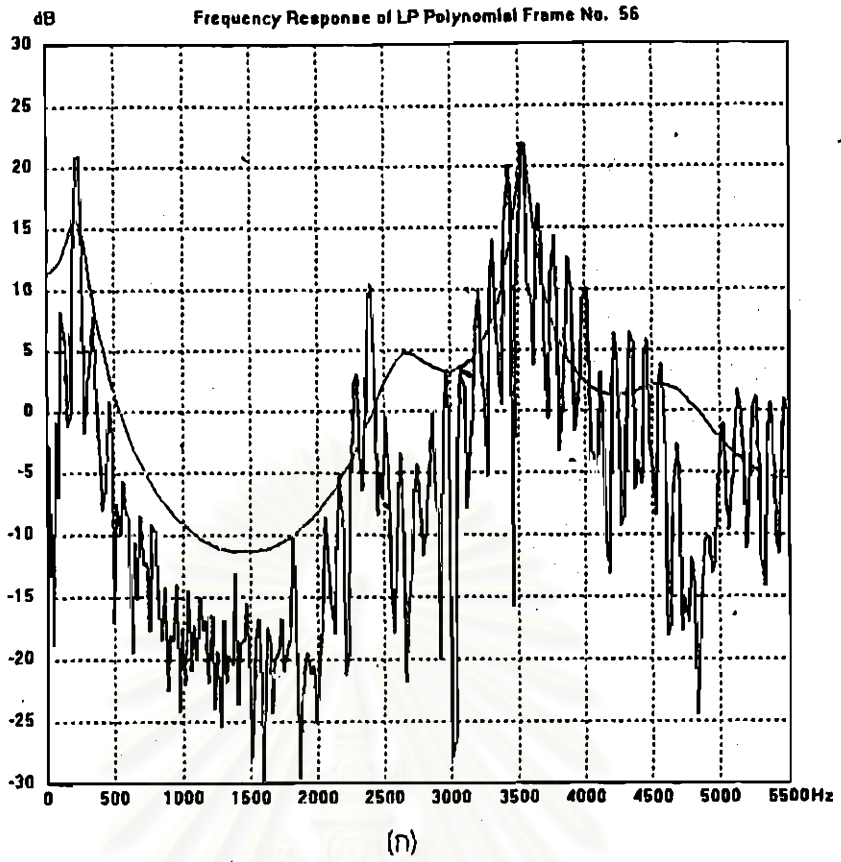
กระบวนการขั้นตอนจากการวิเคราะห์การประมาณพหุระเชิงเส้นคือการเปลี่ยนจากอัตสหสัมพันธ์ $p+1$ ค่าในแต่ละกรอบข้อมูลเสียงพูดเป็นชุดค่าสัมประสิทธิ์การประมาณพหุระเชิงเส้นซึ่งประกอบด้วยสัมประสิทธิ์การประมาณพหุระเชิงเส้น สัมประสิทธิ์การสะท้อน (Reflection Coefficient) หรือค่าอื่นๆ วิธีการเปลี่ยนจากสัมประสิทธิ์อัตสหสัมพันธ์เป็นชุดค่าสัมประสิทธิ์การประมาณพหุระเชิงเส้นคือขั้นตอนวิธีการ Levinson-Durbin ซึ่งมีขั้นตอนดังตารางที่ 2.1 (Deller, Proakis, and Hansen, 1993)

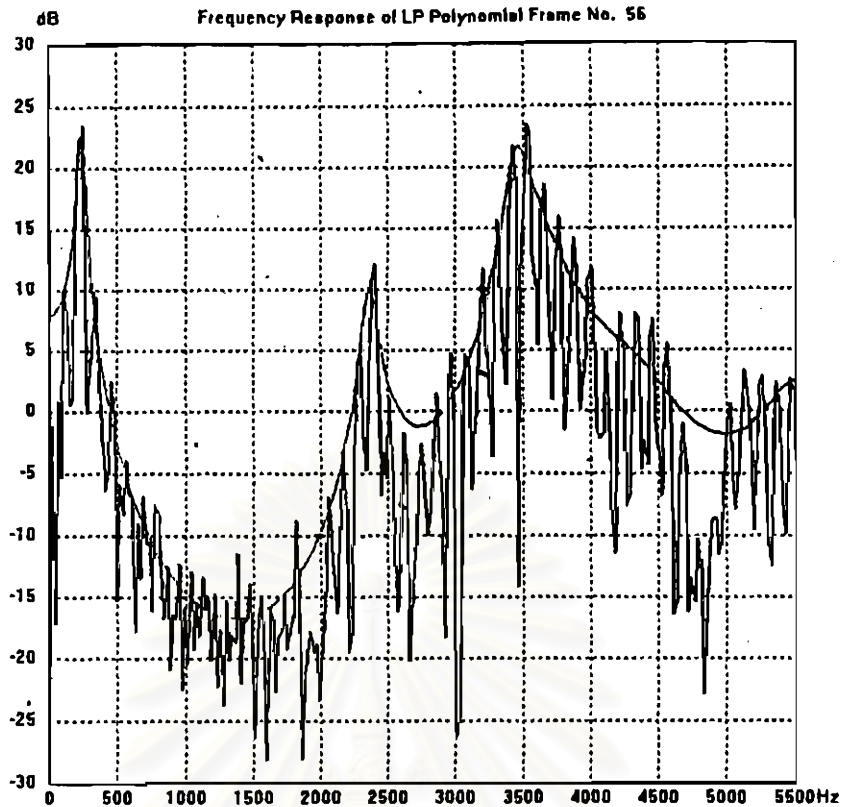
ตารางที่ 2.1 ขั้นตอนวิธีการ Levinson-Durbin

เงื่อนไขเริ่มต้น	
$\xi^0 = r(0)$	
การวนซ้ำ $l = 1, 2, 3, \dots, p$	
1. คำนวณสัมประสิทธิ์การสะท้อนที่ l	$\kappa(l) = \frac{1}{\xi^{l-1}} \left[r(l) - \sum_{i=1}^{l-1} a^{l-1}(i)r(l-i) \right]$
2. สร้างชุดพารามิเตอร์ของการประมาณพหุระเชิงเส้น	$a'(l) = \kappa(l)$ $a'(i) = a^{l-1}(i) - \kappa(l)a^{l-1}(l-i) \quad i = 1, \dots, l-1$
3. คำนวณค่าผิดพลาดที่ l	$\xi^l = \xi^{l-1}(1 - \kappa(l)^2)$
4. กลับไปขั้นตอนที่ 1 แทนค่า l ด้วยค่า $l+1$ ถ้า $l < p$	

โดยที่ p คือ อันดับของพหุนามการประมาณพหุระเชิงเส้น

แบบจำลองการประมาณพหุระเชิงเส้นสามารถแทนลักษณะของสัญญาณเสียงพูดในโดเมนความถี่ได้ดีเพียงใดขึ้นอยู่กับอันดับของพหุนามการประมาณพหุระเชิงเส้นดังแสดงในรูปที่ 2.6





(ค)

รูปที่ 2.6 สเปกตรัมของเสียงสระ "อิ" ที่คำนวณพหุนามการประมาณพัหะเชิงเส้นที่มีอันดับต่างกัน

- (ก) พหุนามการประมาณพัหะเชิงเส้น 8 อันดับ
- (ข) พหุนามการประมาณพัหะเชิงเส้น 10 อันดับ
- (ค) พหุนามการประมาณพัหะเชิงเส้น 12 อันดับ

2.2.2 สัมประสิทธิ์ Cepstral (Cepstral Coefficient)

คุณลักษณะสำคัญที่คำนวณได้จากแบบจำลองการประมาณพัหะเชิงเส้นคือค่าสัมประสิทธิ์ Cepstral ซึ่งมีวิธีการคำนวณดังแสดงในสมการที่ (2.32) และ (2.33) (Rabiner and Juang, 1993)

$$c(m) = a_m + \sum_{k=1}^{m-1} \left(\frac{k}{m}\right) c_k a_{m-k} \quad 1 \leq m \leq p \quad (2.32)$$

$$c(m) = \sum_{k=1}^{m-1} \left(\frac{k}{m}\right) c_k a_{m-k} \quad m > p \quad (2.33)$$

โดยที่ a_m เป็นค่าสัมประสิทธิ์การประมาณพัหะเชิงเส้น

$c(m)$ เป็นค่าสัมประสิทธิ์ cepstral

p เป็นอันดับของสัมประสิทธิ์การประมาณพัหะเชิงเส้น

ค่าสัมประสิทธิ์ Cepstral คำนวณได้ 2 วิธีคือคำนวณจากการแปลงฟูเรียร์อย่างรวดเร็วและคำนวณจากแบบจำลองการประมาณพัหะเชิงเส้น ซึ่งการคำนวณค่าสัมประสิทธิ์ Cepstral จากแบบจำลองการประมาณพัหะเชิงเส้นจะมีข้อดีคือใช้เวลาในการคำนวณที่น้อยกว่า

2.2.3 ความถี่มูลฐาน (Fundamental Frequency)

คุณสมบัติพื้นฐานของเสียงที่เกิดจากเส้นเสียง (vocal cord) คือมีความเป็นรายคาบที่มีช่วงเวลา T_0 ซึ่งส่วนกลับของช่วงเวลาดังกล่าวเรียกว่าความถี่มูลฐาน (Fant, 1970) แสดงในสมการที่ (2.34)

$$F_0 = \frac{1}{T_0} \quad (2.34)$$

ความถี่มูลฐาน F_0 เป็นตัวแสดงระดับเสียงสูงต่ำทำให้สามารถบอกรวมของแต่ละพยางค์ได้ นอกจากความถี่มูลฐานจะสามารถบอกรวมของพยางค์ได้แล้วยังสามารถแบ่งเสียงออกเป็นเสียงโหระ (voiced sound) และเสียงอโหระ (unvoiced sound) เสียงโหระเท่านั้นที่มีความถี่มูลฐาน กรรมวิธีการหาความถี่มูลฐานมี 2 วิธีคือ ฟังก์ชันผลต่างขนาดเฉลี่ย (Average Magnitude Difference Function, AMDF) และ Real Cepstrum ในงานวิจัยนี้จะหาความถี่มูลฐานโดยใช้วิธี Real Cepstrum เพราะใช้เวลาในการคำนวณน้อยกว่าและให้ความแม่นยำที่มากกว่าวิธีฟังก์ชันผลต่างขนาดเฉลี่ย การหาความถี่มูลฐานโดยใช้วิธี Real Cepstrum มีขั้นตอนดังแสดงในรูปที่ 2.7

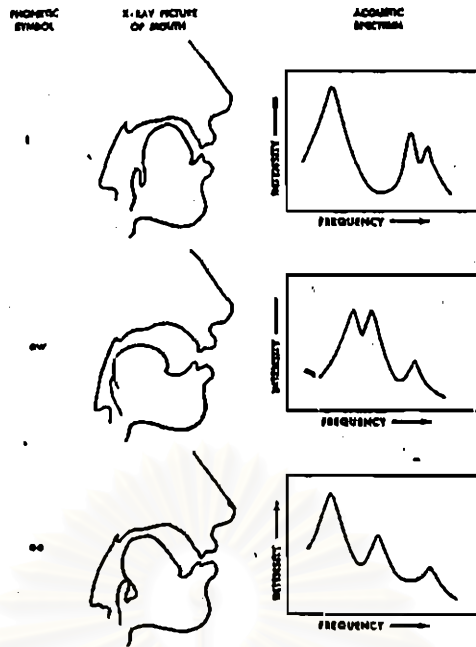


รูปที่ 2.7 กรรมวิธีการหา Real Cepstrum (Deller, Proakis, and Hansen, 1993)

2.2.4 ความถี่ฟอร์แมนท์ (Formant Frequency)

ช่องทางเดินเสียง (vocal tract) มีลักษณะเหมือนกับท่ออากาศและทำหน้าที่เป็นตัวเรโซแนนซ์ ช่องทางเดินเสียงจะมีความถี่ธรรมชาติ ค่าหนึ่งและจะตอบสนองกับคลื่นเสียงที่มีความถี่ตรงกับความถี่ของช่องทางเดินเสียง สเปกตรัมของคลื่นเสียงจะมีจุดยอดอยู่ที่ความถี่ธรรมชาติของช่องทางเดินเสียง ถ้าช่องทางเดินเสียงมีรูปร่างที่แตกต่างกันออกไปความถี่ธรรมชาติของช่องทางเดินเสียงจะมีค่าเปลี่ยนไปด้วยดังแสดงในรูปที่ 2.8

การเรโซแนนซ์ของช่องทางเดินเสียงเรียกว่าฟอร์แมนท์ (Formant) และความถี่เรโซแนนซ์ของช่องทางเดินเสียงเรียกว่าความถี่ฟอร์แมนท์ (Formant Frequency) ความถี่มูลฐานและฮาร์โมนิกของความถี่มูลฐานเกิดจากการสั่นของเส้นเสียง (vocal cord) แต่ความถี่ฟอร์แมนท์เกิดจากรูปร่างของช่องทางเดินเสียง ความถี่อันเกิดจากการสั่นของเส้นเสียงและความถี่อันเกิดจากลักษณะของช่องทางเดินเสียงจะเป็นอิสระต่อกัน ในรูปที่ 2.9 แสดงรูปคลื่นสัญญาณเสียงที่ออกจากปากและรูปคลื่นสัญญาณเสียงที่เส้นเสียง รูปคลื่นสัญญาณเสียงที่เส้นเสียงจะถูกปรับเปลี่ยนจากช่องทางเดินเสียงเป็นรูปคลื่นที่มีความซับซ้อน (Borden and Harris, 1980) ความถี่ฟอร์แมนท์ที่ต่ำที่สุดเรียกว่าความถี่ฟอร์แมนท์ที่หนึ่งและความถี่ฟอร์แมนท์ถัดไปเรียกว่าความถี่ฟอร์แมนท์ที่สองตามลำดับ



รูปที่ 2.8 รูปร่างของช่องทางเดินเสียงและสเปกตรัม (Denes and Pinson, 1967)

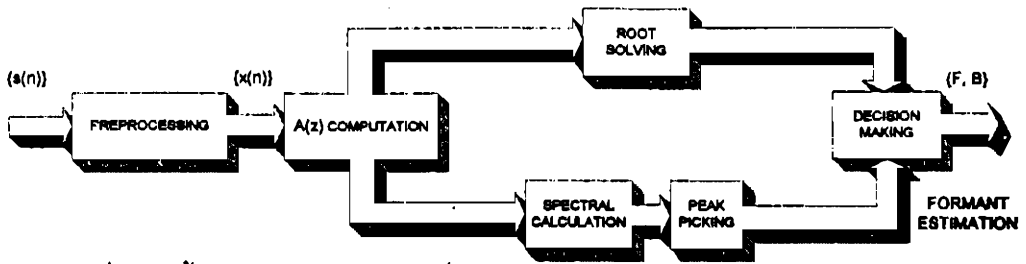


รูปที่ 2.9 รูปคลื่นสัญญาณเสียงที่ออกจากปากและที่เส้นเสียง (Borden and Harris, 1980)

วิธีการคำนวณความถี่ฟอร์แมนท์จะอยู่บนพื้นฐานของแบบจำลองการประมาณพหุเชิงเส้น (Markel and Gray Jr., 1980; Furui, 1989; Deller, Proakis, and Hansen, 1993) ที่มีฟังก์ชันถ่ายโอน ดังแสดงในสมการที่ (2.35)

$$H(z) = \frac{1}{1 - \sum_{l=1}^p a_l z^{-l}} = \frac{1}{A(z)} \quad (2.35)$$

การคำนวณความถี่ฟอร์แมนท์จากแบบจำลองการประมาณพหุเชิงเส้นมี 2 วิธีคือคำนวณจากรากของพหุนามสัมประสิทธิ์การประมาณพหุเชิงเส้น $A(z)$ และคำนวณจากวิธีการตรวจจับค่ายอด (Peak Picking) ของสเปกตรัมแสดงในรูปที่ 2.10



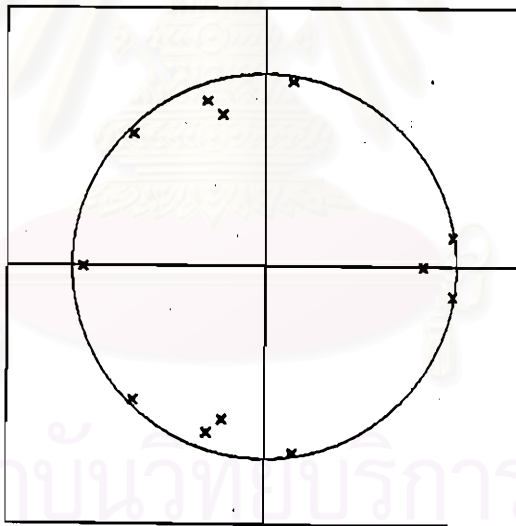
รูปที่ 2.10 ขั้นตอนการคำนวณความถี่ฟอร์แมนท์ (Markel and Gray Jr., 1980)

การคำนวณความถี่ฟอร์แมนท์จากรากของพหุนามสัมประสิทธิ์การประมาณพันระเชิงเส้น

เนื่องจากแบบจำลองการประมาณพันระเชิงเส้นเป็นแบบจำลองที่มีเสถียรภาพค่าโพลของฟังก์ชันถ่ายโอนจะอยู่ในวงกลมหนึ่งหน่วยดังแสดงในรูปที่ 2.11 กำหนดให้ $z = \text{Re}(z) + j\text{Im}(z)$ เป็นรากของพหุนามสัมประสิทธิ์การประมาณพันระเชิงเส้น ความถี่ฟอร์แมนท์จะคำนวณได้จาก

$$F = \left(\frac{f_s}{2\pi} \right) \arctan \left(\frac{\text{Im}(z)}{\text{Re}(z)} \right) \quad (2.36)$$

โดยที่ f_s เป็นค่าความถี่ซิกตัวอย่าง (Sampling Frequency) ในการคำนวณความถี่ฟอร์แมนท์จะใช้รากของพหุนามสัมประสิทธิ์การประมาณพันระเชิงเส้นที่อยู่ในซีกบนของวงกลมหนึ่งหน่วยเท่านั้น



รูปที่ 2.11 โพลของฟังก์ชันถ่ายโอนของแบบจำลองการประมาณพันระเชิงเส้น

การคำนวณความถี่ฟอร์แมนท์จากวิธีการตรวจจับค่ายอดของสเปกตรัม

การคำนวณความถี่ฟอร์แมนท์จากวิธีการตรวจจับค่ายอดของสเปกตรัมจะมีขั้นตอนดังต่อไปนี้คือ การคำนวณการแปลงฟูเรียร์แบบเร็ว N จุดของค่าสัมประสิทธิ์การประมาณพันระเชิงเส้น ตรวจจับค่ายอดของสเปกตรัมที่คำนวณได้จากการแปลงฟูเรียร์แบบเร็วและคำนวณความถี่ฟอร์แมนท์จากสมการที่ (2.37)

$$F = k \frac{f_s}{N} \quad (2.37)$$

โดยที่ k เป็นค่ายอดของสเปกตรัม

2.3 โครงข่ายประสาทเทียม (Artificial Neural Networks)

การศึกษาเรื่องโครงข่ายประสาทสมัยใหม่เริ่มต้นในศตวรรษที่ 19 ทำให้มีความตื่นตัวในการศึกษาการทำงานของสมองมนุษย์และพบว่าสมองมนุษย์ประกอบด้วยปมประสาท (neuron) จำนวนมากแต่ลูปประสาทมีการเชื่อมต่อกันจำนวนมากทำให้การทำงานของสมองมนุษย์ซับซ้อนอย่างรวดเร็วเนื่องจากมีความไม่เป็นเชิงเส้นและมีการทำงานแบบขนาน จึงทำให้มีการสร้างแบบจำลองการทำงานของสมองมนุษย์แบบง่าย ๆ ขึ้น โครงข่ายประสาทเทียมจึงเป็นแบบจำลองการทำงานของสมองมนุษย์ที่ถูกนำมาประยุกต์สร้างเป็นวงจรอิเล็กทรอนิกส์หรือเป็นซอฟต์แวร์เพื่อการประมวลผลในงานต่างๆ

องค์ประกอบพื้นฐานของโครงข่ายประสาทเทียม

โครงข่ายประสาทเทียมประเภทต่างๆจะมีองค์ประกอบที่เหมือนกัน 4 อย่างคือ (Tebelskis, 1995)

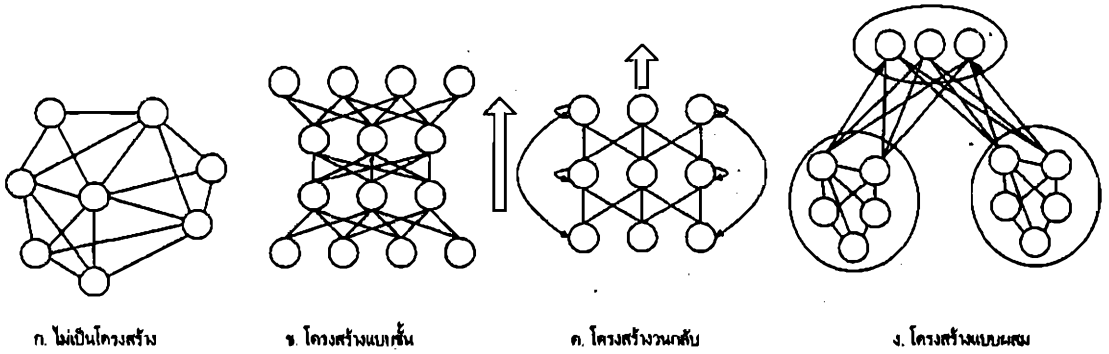
1. หน่วยประมวลผล (Processing Units)
2. การเชื่อมต่อ (Connections)
3. กระบวนการคำนวณ (Computing Procedure)
4. กระบวนการฝึกฝน (Training Procedure)

หน่วยประมวลผล (Processing Units)

หน่วยประมวลผลในโครงข่ายประสาทเทียมจะแบ่งออกเป็น 3 ส่วนคือ หน่วยข้อมูลเข้าทำหน้าที่รับข้อมูลจากภายนอก หน่วยซ่อนตัวทำหน้าที่แปลงข้อมูลภายในและหน่วยข้อมูลออกทำหน้าที่ตัดสินใจหรือควบคุมสัญญาณ

การเชื่อมต่อ (Connections)

หน่วยประมวลผลในโครงข่ายประสาทเทียมจะถูกจัดเรียงเป็นโครงสร้างต่างๆโดยการเชื่อมต่อซึ่งมีค่ากำกับไว้ ค่าที่กำกับการเชื่อมต่อเรียกว่าน้ำหนักการเชื่อมต่อ โครงข่ายประสาทเทียมสามารถมีการเชื่อมต่อเป็นลักษณะต่างๆได้ 4 แบบดังแสดงในรูปที่ 2.12 คือไม่เป็นโครงสร้าง (Unstructured) โครงสร้างแบบชั้น (Layered) โครงสร้างวนกลับ (Recurrent) และ โครงสร้างแบบผสม (Modular)



รูปที่ 2.12 ลักษณะของโครงข่ายประสาทเทียม (Tebelskis, 1995)

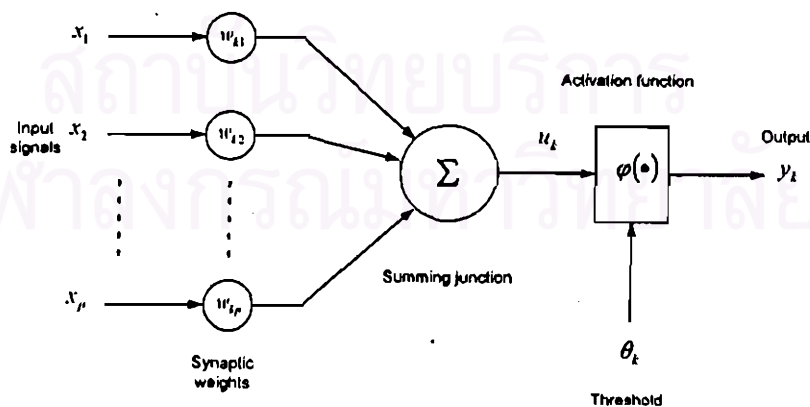
กระบวนการคำนวณ (Computing Procedure)

โหนดเป็นหน่วยประมวลผลข้อมูลซึ่งเป็นพื้นฐานของโครงข่ายประสาทเทียม ในรูปที่ 2.13 แสดงแบบจำลองของโหนดที่มีองค์ประกอบดังนี้ (Haykin, 1994)

1. การเชื่อมต่อซึ่งแต่ละการเชื่อมต่อจะมีคุณลักษณะคือน้ำหนักการเชื่อมต่อ สัญญาณเข้า x_j ของการเชื่อมต่อ j กับโหนดที่ k จะถูกคูณด้วยน้ำหนักการเชื่อมต่อ w_{kj}
2. การบวกสำหรับการรวมสัญญาณเข้าที่ถูกคูณด้วยน้ำหนักการเชื่อมต่อ
3. ฟังก์ชันการกระตุ้นสำหรับจำกัดแอมพลิจูดของสัญญาณที่ออกจากโหนด

กระบวนการฝึกฝน (Training Procedure)

การฝึกฝนโครงข่ายประสาทเทียมคือการปรับเปลี่ยนค่าน้ำหนักการเชื่อมต่อหรือบางกรณีเป็นการปรับเปลี่ยนโครงสร้างของโครงข่ายประสาทเทียมซึ่งเป็นการเพิ่มหรือลบการเชื่อมต่อของโหนด การปรับเปลี่ยนค่าน้ำหนักการเชื่อมต่อจะมีลักษณะทั่วไปมากกว่าการปรับเปลี่ยนโครงสร้างเพราะว่าการที่ค่าน้ำหนักการเชื่อมต่อเท่ากับศูนย์คือการลบการเชื่อมต่อนั้นออกจากโครงข่ายประสาทเทียม อย่างไรก็ตามการเปลี่ยนแปลงโครงสร้างของโครงข่ายประสาทเทียมจะเป็นการเพิ่มความเร็วในการเรียนรู้และเพิ่มความสามารถในการรู้จำรูปแบบทั่วไป



รูปที่ 2.13 แบบจำลองการทำงานที่ไม่เป็นเชิงเส้นของโหนด (Haykin, 1994)

โครงข่ายประสาทเทียมมีโครงสร้างเป็นชั้นและมีความไม่เป็นเชิงเส้น การปรับเปลี่ยนค่าน้ำหนักการเชื่อมต่อทำโดยวิธีการวนซ้ำเช่น Gradient Descent การปรับเปลี่ยนค่าน้ำหนักการเชื่อมต่อแต่ละครั้งจะต้องไม่ทำให้การเรียนรู้ที่ผ่านมาสถูยเสียไป ค่าคงที่ที่ใช้ควบคุมขนาดของการปรับเปลี่ยนน้ำหนักการเชื่อมต่อเรียกว่า อัตราการเรียนรู้ (Learning Rate) การกำหนดค่าอัตราการเรียนรู้มีความสำคัญมาก ถ้ากำหนดค่าอัตราการเรียนรู้น้อยเกินไปจะทำให้การเรียนรู้นานมากและถ้ากำหนดค่าอัตราการเรียนรู้มากจะทำให้สูญเสียการเรียนรู้ที่ผ่านมา

โครงข่ายประสาทเทียมแบบ Multilayer Perceptron (MLP)

โครงข่ายประสาทเทียมแบบ MLP ประกอบด้วยชั้นข้อมูลเข้าสำหรับรับข้อมูล ชั้นซ่อนตัวและชั้นข้อมูลออกสำหรับการคำนวณ สัญญาณเข้าจะผ่านเข้าไปในโครงข่ายประสาทเทียมในทิศทางเดียวจากชั้นหนึ่งไปสู่อีกชั้นหนึ่ง โครงข่ายประสาทเทียมแบบ MLP ถูกประยุกต์ใช้สำหรับงานที่มีความซับซ้อนได้เป็นอย่างดี โดยมีกระบวนการฝึกฝนเป็นแบบ Supervised และใช้ขั้นตอนวิธีการส่งค่าย้อนกลับ (Backpropagation) สำหรับการฝึกฝน กระบวนการฝึกฝนแบบส่งค่าย้อนกลับ ประกอบด้วย 2 ส่วนย่อยคือ การส่งผ่านไปข้างหน้า (forward pass) และการส่งผ่านย้อนกลับ (backward pass) สำหรับการส่งผ่านไปข้างหน้าข้อมูลจะผ่านเข้าโครงข่ายประสาทเทียมที่ชั้นข้อมูลเข้า และจะส่งผ่านจากชั้นหนึ่งไปสู่อีกชั้นหนึ่งจนกระทั่งถึงชั้นข้อมูลออก ส่วนการส่งผ่านย้อนกลับค่าน้ำหนักการเชื่อมต่อจะถูกปรับเปลี่ยนให้สอดคล้องกับกฎการแก้ข้อผิดพลาด (error-correction rule) คือผลต่างของผลตอบแท้จริง (actual response) กับผลตอบเป้าหมาย (target response) เกิดเป็นสัญญาณผิดพลาด (error signal) ซึ่งสัญญาณผิดพลาดนี้จะถูกส่งย้อนกลับเข้าสู่โครงข่ายประสาทเทียมในทิศทางตรงข้ามกับการเชื่อมต่อ ค่าน้ำหนักการเชื่อมต่อจะถูกปรับจนกระทั่งผลตอบแท้จริงเข้าใกล้ผลตอบเป้าหมาย

โครงข่ายประสาทเทียมแบบ MLP มีลักษณะเด่น 3 ประการคือ

- 1. แบบจำลองของแต่ละโหนดจะมีความไม่เป็นเชิงเส้นและความไม่เป็นเชิงเส้นดังกล่าวจะต้องมีความราบเรียบคือสามารถหาอนุพันธ์ได้ทุกจุด ความไม่เป็นเชิงเส้นดังกล่าวถูกกำหนดจากฟังก์ชันซิกมอยด์

$$y_j = \frac{1}{1 + \exp(-v_j)} \tag{2.38}$$

โดยที่ v_j เป็นผลรวมภายในของโหนดที่ j และ y_j คือสัญญาณออกของโหนดที่ j

- 2. โครงข่ายประสาทเทียมจะมีจำนวนชั้นซ่อนตัวที่มากกว่าหนึ่งชั้นได้ซึ่งไม่ใช่ชั้นข้อมูลเข้าและชั้นข้อมูลออก โหนดในชั้นซ่อนตัวนี้จะทำให้โครงข่ายประสาทเทียมสามารถเรียนรู้งานที่มีความซับซ้อนได้ดีขึ้น
- 3. โครงข่ายประสาทเทียมแบบ MLP จะมีการเชื่อมต่อกันมาก

รูปที่ 2.14 แสดงตัวอย่างโครงสร้างของโครงข่ายประสาทเทียมแบบ MLP ที่มีจำนวนชั้นซ่อนตัว 2 ชั้น สัญญาณที่ในโครงข่ายประสาทเทียมแบบ MLP มี 2 ประเภทคือ Function Signal และ Error Signal ดังแสดงในรูปที่ 2.15 และมีรายละเอียดดังนี้

1. Function Signal เป็นสัญญาณเข้าที่มาจากโหนดในชั้นก่อนหน้าและส่งผ่านไปข้างหน้าจากโหนดหนึ่งไปสู่อีกโหนดหนึ่ง
2. Error Signal เป็นสัญญาณที่เกิดขึ้นที่โหนดในชั้นข้อมูลออกของในโครงข่ายประสาทเทียมและถูกส่งผ่านย้อนกลับจากชั้นหนึ่งไปสู่อีกชั้นหนึ่ง

ขั้นตอนวิธีการส่งค่าย้อนกลับ (Backpropagation Algorithm)

สัญญาณผิดพลาดที่โหนดที่ j ของชั้นข้อมูลออกสำหรับชุดฝึกที่ n กำหนดในสมการที่ (2.39)

$$e_j(n) = d_j(n) - y_j(n) \quad (2.39)$$

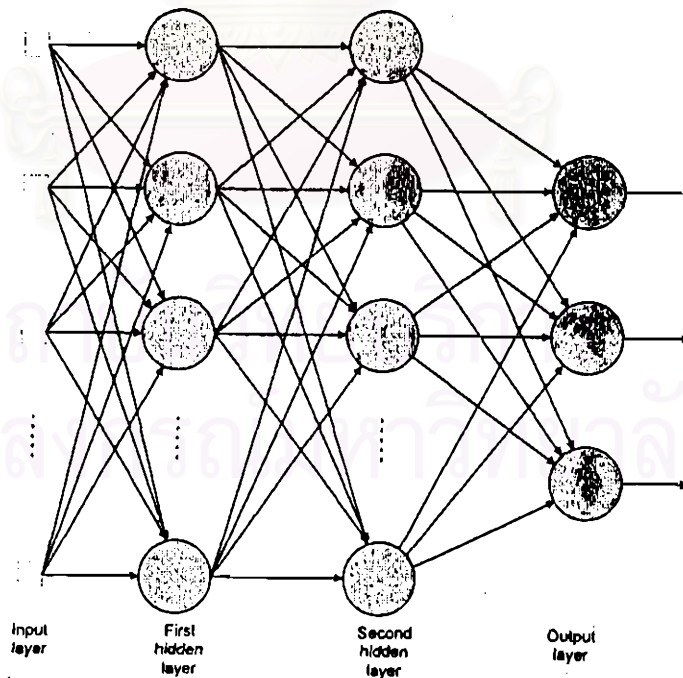
โดยที่ $d_j(n)$ คือข้อมูลออกเป้าหมาย (target output) y_j คือสัญญาณออกแท้จริง (actual output)

กำหนดค่าผิดพลาดของโหนดที่ j ในรูปแบบค่าผิดพลาดกำลังสอง (square error) คือ $\frac{1}{2}e_j^2(n)$ และผลรวมของค่าผิดพลาดกำลังสองดังกล่าวคือ $E(n)$ ซึ่งเขียนในรูปสมการคือ

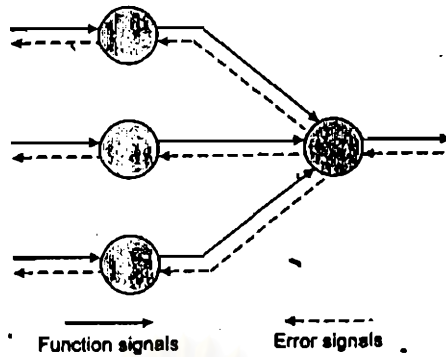
$$E(n) = \frac{1}{2} \sum_{j \in C} e_j^2(n) \quad (2.40)$$

โดยที่ C เป็นจำนวนโหนดในชั้นข้อมูลออกทั้งหมด กำหนดให้ N เป็นจำนวนข้อมูลฝึกฝน ค่าผิดพลาดกำลังสองเฉลี่ย (average sum squared error) คือผลรวม $E(n)$ ของทุกข้อมูลฝึกฝนจำนวน N และถูกปรับให้เป็นบรรทัดฐานเดียวกันดังสมการที่ (2.41)

$$E_{av} = \frac{1}{N} \sum_{n=1}^N E(n) \quad (2.41)$$



รูปที่ 2.14 ตัวอย่างโครงสร้างของโครงข่ายประสาทเทียมแบบ MLP ที่มีจำนวนชั้นซ่อนตัว 2 ชั้น (Haykin, 1994)



รูปที่ 2.15 ทิศทางการไหลของสัญญาณในโครงข่ายประสาทเทียมแบบ MLP (Haykin, 1994)

รูปที่ 2.16 แสดงการไหลของสัญญาณของโหนดที่ j จากชั้นทางซ้ายมือไปสู่ชั้นขวามือรวมกันเป็นสัญญาณ $v_j(n)$ ดังสมการ

$$v_j(n) = \sum_{i=0}^p w_{ji}(n) y_i(n) \quad (2.42)$$

โดยที่ p เป็นจำนวนข้อมูลเข้าของโหนด j คำนวณน้ำหนักการเชื่อมต่อ w_{j0} เท่ากับค่าจุดเริ่มเปลี่ยน θ_j สัญญาณที่ออกจากโหนดที่ j คือ

$$y_j(n) = \varphi_j(v_j(n)) \quad (2.43)$$

ขั้นตอนวิธีการส่งค่าย้อนกลับจะมีค่าแก้ไขน้ำหนักการเชื่อมต่อ $\Delta w_{ji}(n)$ กับน้ำหนักการเชื่อมต่อ $w_{ji}(n)$ ซึ่งเป็นสัดส่วนกับค่าลดหลั่น (Gradient) $\partial E(n)/\partial w_{ji}(n)$ จากกฎลูกโซ่ขยาย $\partial E(n)/\partial w_{ji}(n)$ จะได้

$$\frac{\partial E(n)}{\partial w_{ji}(n)} = \frac{\partial E(n)}{\partial e_j(n)} \frac{\partial e_j(n)}{\partial y_j(n)} \frac{\partial y_j(n)}{\partial v_j(n)} \frac{\partial v_j(n)}{\partial w_{ji}(n)} \quad (2.44)$$

พิจารณาอนุพันธ์ทั้งสองข้างของสมการที่ (2.40) เทียบกับ $e_j(n)$ จะได้

$$\frac{\partial E(n)}{\partial e_j(n)} = e_j(n) \quad (2.45)$$

อนุพันธ์ของสมการที่ (2.39) เทียบกับ $y_j(n)$ จะได้

$$\frac{\partial e_j(n)}{\partial y_j(n)} = -1 \quad (2.46)$$

อนุพันธ์ของสมการที่ (2.43) เทียบกับ $v_j(n)$ จะได้

$$\frac{\partial y_j(n)}{\partial v_j(n)} = \varphi'_j(v_j(n)) \quad (2.47)$$

อนุพันธ์ของสมการที่ (2.42) เทียบกับ $w_{ji}(n)$ จะได้

$$\frac{\partial v_j(n)}{\partial w_{ji}(n)} = y_i(n) \quad (2.48)$$

จากสมการที่ (2.45) ถึง สมการที่ (2.48) แทนในสมการที่ (2.44) จะได้

$$\frac{\partial E(n)}{\partial w_{j\mu}(n)} = -e_j(n)\phi'_j(v_j(n))y_\mu(n) \quad (2.49)$$

$\Delta w_{j\mu}(n)$ ที่นำมาปรับเปลี่ยนค่าน้ำหนักการเชื่อมต่อ $w_{j\mu}(n)$ ถูกกำหนดจากกฎเดลตา (delta rule)

$$\Delta w_{j\mu}(n) = -\eta \frac{\partial E(n)}{\partial w_{j\mu}(n)} \quad (2.50)$$

โดยที่ η คืออัตราการเรียนรู้ จากสมการที่ (2.49) และ สมการที่ (2.50) จะได้

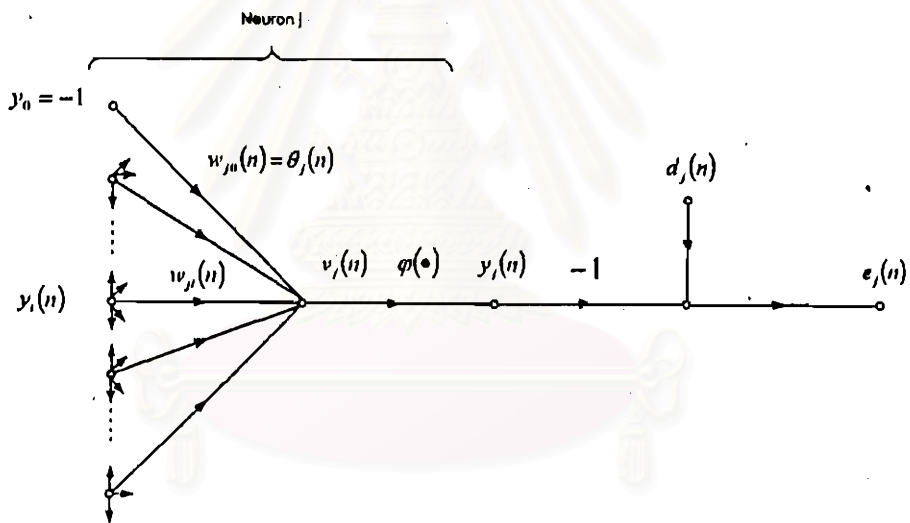
$$\Delta w_{j\mu}(n) = \eta \delta_j(n) y_\mu(n) \quad (2.51)$$

โดยที่ $\delta_j(n)$ คือค่าลดหลั่นเฉพาะที่ (local gradient) เป็นไปตามสมการ

$$\delta_j(n) = -\frac{\partial E(n)}{\partial e_j(n)} \frac{\partial e_j(n)}{\partial y_j(n)} \frac{\partial y_j(n)}{\partial v_j(n)} \quad (2.52)$$

$$= e_j(n)\phi'_j(v_j(n)) \quad (2.53)$$

จากสมการที่ (2.52) และสมการที่ (2.53) ปัจจัยที่สำคัญในการคำนวณการปรับเปลี่ยนน้ำหนักการเชื่อมต่อคือ สัญญาณผิดพลาด $e_j(n)$ ที่ด้านออกของโนด j และขึ้นอยู่กับตำแหน่งของโนด j ซึ่งแบ่งเป็น 2 ตำแหน่งคือ ตำแหน่งของโนด j อยู่ที่ชั้นข้อมูลออกและตำแหน่งของโนด j อยู่ที่ชั้นซ่อนตัว



รูปที่ 2.16 แสดงการไหลของสัญญาณของโนดที่ j (Haykin, 1994)

ตำแหน่งของโนด j อยู่ที่ชั้นข้อมูลออก

ถ้าตำแหน่งของโนดอยู่ที่ชั้นข้อมูลออกจะใช้สมการที่ (2.39) ในการคำนวณสัญญาณผิดพลาด $e_j(n)$ และคำนวณ local gradient $\delta_j(n)$ โดยสมการที่ (2.53)

ตำแหน่งของโนด j อยู่ที่ชั้นซ่อนตัว

ถ้าตำแหน่งของโนดอยู่ที่ชั้นซ่อนตัว สัญญาณผิดพลาด $e_j(n)$ สำหรับโนดในชั้นนี้จะคำนวณแบบวนซ้ำจากโนดในชั้นก่อนหน้าซึ่งทำให้ขั้นตอนวิธีการส่งค่าย้อนกลับมีความซับซ้อนมากกว่าในกรณีตำแหน่งของโนดอยู่ที่ชั้นข้อมูลออก จากสมการที่ (2.53) จะคำนวณค่าลดทอนเฉพาะที่ของโนดในชั้นซ่อนตัวคือ

$$\delta_j(n) = -\frac{\partial E(n)}{\partial y_j(n)} \frac{\partial y_j(n)}{\partial v_j(n)} \quad (2.54)$$

$$= -\frac{\partial E(n)}{\partial y_j(n)} \varphi'_j(v_j(n)) \quad (2.55)$$

จากรูปที่ 2.17 แสดงการคำนวณ $\partial E(n)/\partial y_j(n)$ ได้ดังนี้

$$E(n) = \frac{1}{2} \sum_{k \in C} e_k^2(n) \quad \text{โนด } k \text{ คือโนดในชั้นข้อมูลออก} \quad (2.56)$$

อนุพันธ์ของสมการที่ (2.56) เทียบกับ $y_j(n)$ จะได้

$$\frac{\partial E(n)}{\partial y_j(n)} = \sum_k e_k \frac{\partial e_k(n)}{\partial y_j(n)} \quad (2.57)$$

จากนั้นใช้กฎลูกโซ่สำหรับ $\partial e_k(n)/\partial y_j(n)$

$$\frac{\partial E(n)}{\partial y_j(n)} = \sum_k e_k(n) \frac{\partial e_k(n)}{\partial v_k(n)} \frac{\partial v_k(n)}{\partial y_j(n)} \quad (2.58)$$

แต่

$$e_k(n) = d_k(n) - y_k(n) \quad (2.59)$$

$$= d_k(n) - \varphi_k(v_k(n)) \quad (2.60)$$

และ

$$\frac{\partial e_k(n)}{\partial v_k(n)} = -\varphi'_k(v_k(n)) \quad (2.61)$$

$$v_k(n) = \sum_{j=0}^q w_{kj}(n) y_j(n) \quad (2.62)$$

โดยที่ q เป็นจำนวนข้อมูลเข้าของโนด k อนุพันธ์ของสมการที่ (2.62) เทียบกับ $y_j(n)$ จะได้

$$\frac{\partial v_k(n)}{\partial y_j(n)} = w_{kj}(n) \quad (2.63)$$

แทนสมการที่ (2.61) และสมการที่ (2.63) ในสมการที่ (2.59) จะได้

$$\frac{\partial E(n)}{\partial y_j(n)} = -\sum_k e_k(n) \varphi'_k(v_k(n)) w_{kj}(n) \quad (2.64)$$

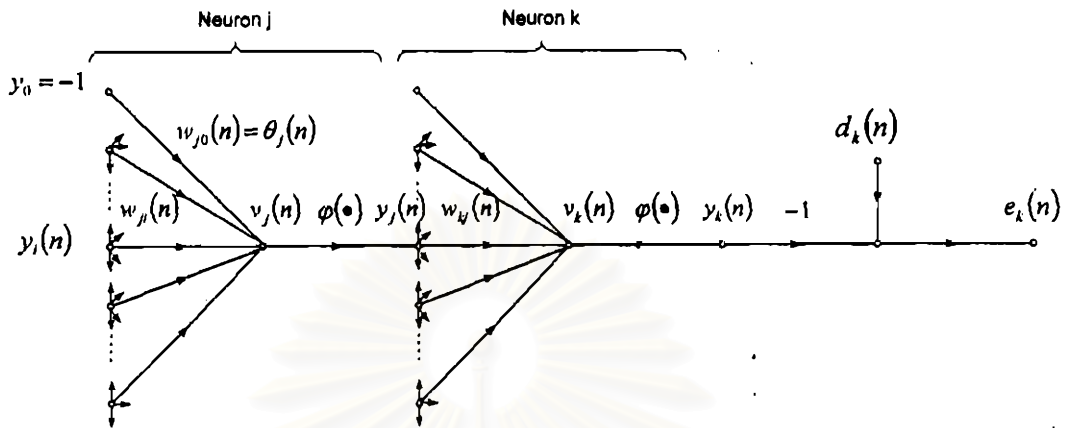
$$= -\sum_k \delta_k(n) w_{kj}(n) \quad (2.65)$$

จากสมการที่ (2.65) และสมการที่ (2.57) จะได้ค่าลดทอนเฉพาะที่ $\delta_j(n)$ สำหรับโนด j คือ

$$\delta_j(n) = \varphi'_j(v_j(n)) \sum_k \delta_k(n) w_{kj}(n) \quad (2.66)$$

ค่าแก้ไข $\Delta w_{ij}(n)$ ที่ใช้ปรับเปลี่ยนค่าน้ำหนักการเชื่อมต่อระหว่างโนด i และโนด j ถูกกำหนดโดย

$$\begin{pmatrix} \text{Weight} \\ \text{correction} \\ \Delta w_{ji}(n) \end{pmatrix} = \begin{pmatrix} \text{learning} \\ \text{rate parameter} \\ \eta \end{pmatrix} \begin{pmatrix} \text{local} \\ \text{gradient} \\ \delta_j(n) \end{pmatrix} \begin{pmatrix} \text{input signal} \\ \text{of neuron } j \\ y_j(n) \end{pmatrix} \quad (2.67)$$



รูปที่ 2.17 การไหลของสัญญาณสำหรับโหนด k ในชั้นข้อมูลออก
ที่เชื่อมต่อกับโหนด j ในชั้นซ่อนตัว (Haykin, 1994)

ค่าลดหลั่นเฉพาะที่ $\delta_j(n)$ ขึ้นอยู่กับตำแหน่งของโหนด j

1. ถ้าโหนด j เป็นโหนดในชั้นข้อมูลออก $\delta_j(n)$ จะเท่ากับผลคูณของ $\varphi'(v_j(n))$ กับสัญญาณผิดพลาด $e_j(n)$ ดังสมการที่ (2.53)
2. ถ้าโหนด j เป็นโหนดในชั้นซ่อนตัว $\delta_j(n)$ จะเท่ากับผลคูณของ $\varphi'(v_j(n))$ กับผลรวมของผลคูณระหว่างค่าน้ำหนักการเชื่อมต่อกับ $\delta_j(n)$ ของโหนดในชั้นที่สูงกว่าชั้นซ่อนตัวปัจจุบันที่เชื่อมต่อกับโหนด j ดังสมการที่ (2.66)

ฟังก์ชันซิกมอยด์ (Sigmoid Function)

การคำนวณ δ สำหรับแต่ละโหนดของโครงข่ายประสาทเทียมแบบ MLP ฟังก์ชันกระตุ้น $\varphi(\bullet)$ จะต้องมี
ความต่อเนื่องและหาอนุพันธ์ได้ ฟังก์ชันที่นิยมใช้คือฟังก์ชันซิกมอยด์

$$\begin{aligned} y_j(n) &= \varphi_j'(v_j(n)) \\ &= \frac{1}{1 + \exp(-v_j(n))} \quad -\infty < v_j(n) < \infty \end{aligned} \quad (2.68)$$

$y_j(n)$ จะมีค่าอยู่ในช่วง $0 \leq y_j(n) \leq 1$

หาอนุพันธ์ทั้งสองข้างของสมการที่ (2.68) เทียบกับ $v_j(n)$

$$\begin{aligned} \frac{\partial y_j(n)}{\partial v_j(n)} &= \varphi_j'(v_j(n)) \\ &= \frac{\exp(-v_j(n))}{[1 + \exp(-v_j(n))]^2} \end{aligned} \quad (2.69)$$

จากสมการที่ (2.68) กำจัดพจน์เอกซ์โพเนนเชียล $\exp(-v_j(n))$ จากสมการที่ (2.69)

$$\varphi'_j(v_j(n)) = y_j(n)(1 - y_j(n)) \quad (2.70)$$

สำหรับโนด j ที่อยู่ในชั้นข้อมูลออก

$$\begin{aligned} \delta_j(n) &= e_j(n)\varphi'_j(v_j(n)) \\ &= (d_j(n) - o_j(n))b_j(n)(1 - o_j(n)) \end{aligned} \quad (2.71)$$

$o_j(n)$ เป็นค่าออกของโนดในชั้นข้อมูลออก $d_j(n)$ เป็นผลตอบเป้าหมาย

สำหรับโนด j ที่อยู่ในชั้นซ่อนตัว

$$\begin{aligned} \delta_j(n) &= \varphi'_j(v_j(n)) \sum_k \delta_k(n) w_{kj}(n) \\ &= y_j(n)(1 - y_j(n)) \sum_k \delta_k(n) w_{kj}(n) \end{aligned} \quad (2.72)$$

ในรูปที่ 2.18 แสดงขั้นตอนวิธีการส่งค่าย้อนกลับของโครงข่ายประสาทเทียมแบบ MLP

ค่าคงที่โมเมนตัม (momentum constant)

วิธีการเพิ่มอัตราการเรียนรู้โดยที่หลีกเลี่ยงการทำให้เกิดความไม่มีเสถียรภาพ ทำได้โดยการดัดแปลงกฎเดลตาในสมการที่ (2.51) โดยเพิ่มพจน์โมเมนตัม (Haykin, 1994) ดังแสดงในสมการที่ (2.73)

$$\Delta w_{jk}(n) = \alpha \Delta w_{jk}(n-1) + \eta \delta_j(n) y_k(n) \quad (2.73)$$

โดยที่ α เป็นจำนวนจริงบวกเรียกว่าค่าคงที่โมเมนตัม การมีพจน์โมเมนตัมจะเป็นการควบคุมการป้อนกลับรอบๆ $\Delta w_{jk}(n)$ ดังแสดงในรูปที่ 2.19

กระบวนการฝึกฝนโครงข่ายประสาทเทียมแบบ MLP แบ่งออกเป็น 5 ขั้นตอนดังนี้ (Haykin, 1994)

1. กำหนดค่าเริ่มต้น กำหนดให้ค่าน้ำหนักการเชื่อมต้อมีการแจกแจงแบบสมมาตร
2. กำหนดชุดข้อมูลฝึกฝน กำหนดลำดับของข้อมูลฝึกฝนและผลตอบเป้าหมายของข้อมูลฝึกฝนนั้น
3. การป้อนไปข้างหน้า คำนวณค่า $v_j(n)$ และ $y_j(n)$ จากสมการต่อไปนี้

$$v_j^{(l)}(n) = \sum_{i=0}^l w_{ji}^{(l)}(n) y_i^{(l-1)}(n)$$

$$y_j^{(l)}(n) = \frac{1}{1 + \exp(-v_j^{(l)}(n))}$$

ถ้าโนด j อยู่ในชั้นซ่อนตัว ($l = 1$)

$$y_j^{(0)}(n) = x_j(n)$$

ถ้าโนด j อยู่ในชั้นข้อมูลออก ($l = L$)

$$y_j^{(L)}(n) = o_j(n)$$

คำนวณสัญญาณผิดพลาด

$$e_j(n) = d_j(n) - o_j(n)$$

- 4. การย้อนกลับ คำนวณค่าล้นเฉพาะที่ δ แบบย้อนกลับจากชั้นหนึ่งไปสู่อีกชั้นหนึ่ง ถ้าโนด j อยู่ในชั้นข้อมูลออก

$$\delta_j^{(l)}(n) = e_j(n) o_j(n) (1 - o_j(n))$$

ถ้าโนด j อยู่ในชั้นซ่อนตัว

$$\delta_j^{(l)}(n) = y_j^{(l)}(n) (1 - y_j^{(l)}(n)) \sum_k \delta_k^{(l+1)}(n) w_{jk}^{(l+1)}(n)$$

ปรับเปลี่ยนค่าน้ำหนักการเชื่อมต่อในชั้น l

$$w_{\mu\nu}^{(l)}(n+1) = w_{\mu\nu}^{(l)}(n) + \alpha [w_{\mu\nu}^{(l)}(n) - w_{\mu\nu}^{(l)}(n-1)] + \eta \delta_j^{(l)}(n) y_j^{(l-1)}(n)$$

โดยที่ η เป็นอัตราการเรียนรู้และ α เป็นค่าคงที่โมเมนตัม

- 5. การวนซ้ำ ทำการวนซ้ำจนกระทั่งค่าพารามิเตอร์อิสระของโครงข่ายประสาทเทียมจะมีเสถียรภาพและค่า E_{av} น้อยกว่าระดับที่กำหนด

กฎเกณฑ์การตัดสินใจ

ในการจำแนกรูปแบบ m กลุ่มจากปริภูมิข้อมูลทั้งหมด จำนวนโนดในชั้นข้อมูลออกของโครงข่ายประสาทเทียมแบบ MLP จะเท่ากับ m เพื่อแทนการจำแนกทั้งหมดที่เป็นไปได้ดังแสดงในรูปที่ 2.20 เวกเตอร์ x_j มี p มิติและกลุ่มที่ X ถูกจำแนกคือกลุ่มที่ k จากทั้งหมด m กลุ่มแทนด้วยสัญลักษณ์ C_k กำหนดให้ $y_{k,j}$ เป็นค่าสัญญาณออกของโนดที่ k ของโครงข่ายประสาทเทียมแบบ MLP และเป็นผลตอบของ x_j

$$y_{k,j} = F_k(x_j) \quad k = 1, 2, 3, \dots, m \tag{2.74}$$

โดยที่ฟังก์ชัน $F_k(\bullet)$ กำหนดการแปลงจากข้อมูลเข้าไปเป็นข้อมูลออกที่ k

ถ้าใช้ค่า 0 และ 1 ซึ่งเป็นค่าขอบเขตล่างและขอบเขตบนของฟังก์ชันซิกมอยด์ ตามลำดับแทนผลตอบเป้าหมาย จะเขียนเงื่อนไขในการกำหนดค่าแทนผลตอบเป้าหมายได้เป็น

$$d_{k,j} = \begin{cases} 1 & x_j \in C_k \\ 0 & x_j \notin C_k \end{cases} \tag{2.75}$$

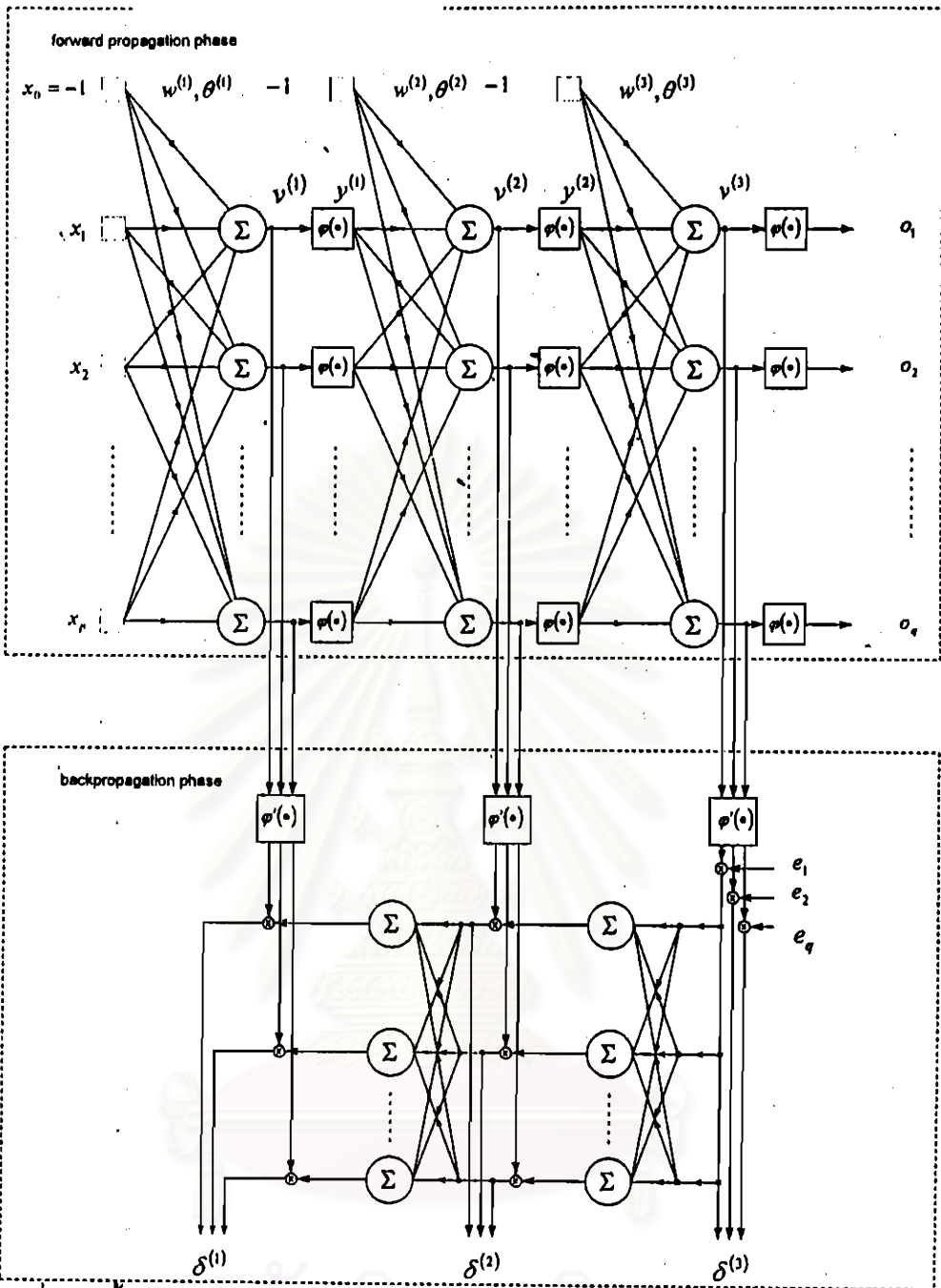
กฎเกณฑ์การตัดสินใจที่เหมาะสมสำหรับการจำแนกรูปแบบโดยใช้โครงข่ายประสาทเทียมแบบ MLP ที่ใช้ฟังก์ชันซิกมอยด์ คือ

จำแนก X ให้อยู่ในกลุ่ม C_k ถ้า

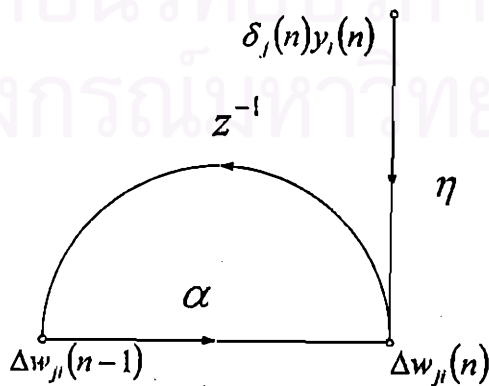
$$F_k(X) > F_j(X) \quad \text{สำหรับ } j \neq k \tag{2.76}$$

โดยที่ $F_k(X)$ และ $F_j(X)$ เป็นองค์ประกอบของฟังก์ชันการแปลง

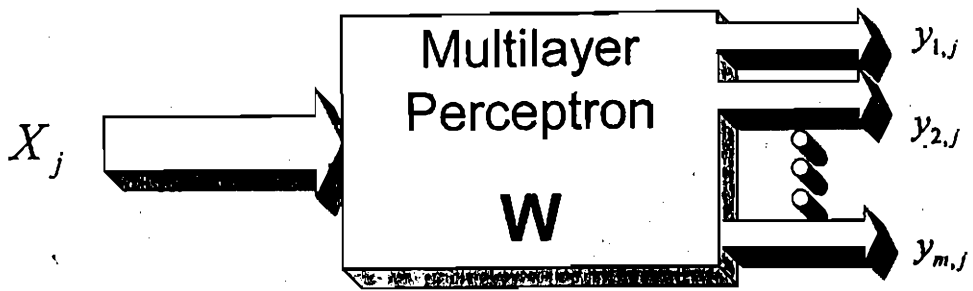
$$F(X) = \begin{bmatrix} F_1(X) \\ F_2(X) \\ \vdots \\ F_m(X) \end{bmatrix}$$



รูปที่ 2.18 ขั้นตอนวิธีการส่งค่าย้อนกลับของโครงข่ายประสาทเทียมแบบ MLP (Haykin, 1994)



รูปที่ 2.19 กราฟการไหลของสัญญาณที่แสดงผลของค่าคงที่โมเมนต์ (Haykin, 1994)



รูปที่ 2.20 แผนผังการจำแนกรูปแบบ (Haykin, 1994)

2.4 กลศาสตร์และภาษาศาสตร์ (Acoustic Phonetic and Linguistic)

พยางค์ในภาษาไทยเกิดจากหน่วยเสียง 3 แบบคือ (Luksaneeyanawin, 1993)

1. หน่วยเสียงพยัญชนะ (Consonantal Phonemes) จำนวน 33 หน่วยเสียง แบ่งเป็นพยัญชนะเดี่ยว (Consonant) จำนวน 21 หน่วยเสียงและพยัญชนะคู่ (Consonant Cluster) จำนวน 12 หน่วยเสียง
2. หน่วยเสียงสระ (Vowel Phonemes) จำนวน 24 หน่วยเสียง แบ่งเป็นสระเดี่ยว (Monophthong) จำนวน 18 หน่วยเสียงและสระผสม (Diphthong) จำนวน 6 หน่วยเสียง
3. หน่วยเสียงวรรณยุกต์ (Tonal Phonemes) จำนวน 5 หน่วยเสียง

เสียงพูดหรือพยางค์ในภาษาไทยเกิดจากการประกอบกันของสระ 1 หน่วยหรือสระคู่ 1 หน่วย พยัญชนะ อาจมีมากกว่า 1 พยัญชนะได้แต่ไม่มากกว่า 3 พยัญชนะและวรรณยุกต์ พยางค์ในภาษาไทยสามารถแทนได้ด้วยโครงสร้างดังนี้ (Luksaneeyanawin, 1993)

$$S = C_1(C_2)V(V)C_3T$$

C_1 คือพยัญชนะต้น

C_2 คือตัวสะกด

V คือสระ

T คือวรรณยุกต์

เสียงสระ

เสียงสระจะเป็นเสียงส่วนกลางของพยางค์ส่วนเสียงพยัญชนะจะเป็นเสียงส่วนนอกของพยางค์ (อมร ทวีศักดิ์, 2536) เสียงสระเป็นเสียงที่อากาศผ่านออกจากปากโดยไม่มีฐานกรณ์กักกั้นทางเดินของอากาศ การที่สระมีเสียงต่างๆ กันเพราะช่องคอ ช่องปากมีรูปร่างต่างกันในขณะที่อากาศกำลังผ่านออกจากปากไป (วิไลวรรณ ชินฐานันท์, 2533) สระเดี่ยวเป็นสระที่ตำแหน่งของลิ้นตั้งแต่เริ่มต้นเสียงสระจนจบเสียงสระไม่เปลี่ยนแปลง ส่วนสระผสมเป็นสระที่เสียงเมื่อเริ่มต้นกับเสียงตอนจบไม่เหมือนกัน ตำแหน่งของลิ้นเปลี่ยนไป เพื่อความสะดวกในการอธิบายเสียงสระจะแบ่งตำแหน่งของลิ้นออกเป็น 3 ส่วนคือส่วนหน้า ส่วนกลาง และส่วนหลัง นอกจากจะแบ่งตำแหน่งของลิ้นที่ใช้ในการออกเสียงแล้วยังสามารถแบ่งระดับของลิ้นเป็นสูง กลางและต่ำตามลำดับ สำหรับเสียงสระเดี่ยวในภาษาไทยแบ่งตามตำแหน่งของลิ้นและระดับของลิ้นแสดงใน

ตารางที่ 2.2 ส่วนสระผสมในภาษาไทยเกิดจากเสียงสระเดี่ยว 2 เสียงมาผสมกันได้แก่ สระเอียะ /ia/ เอีย /iia/ เอือะ /va/ เอือ /vva/ อัวะ /ua/ อิว /uua/

คลื่นเสียงของสระทุกตัวมีลักษณะเป็นรายคาบซึ่งเกิดจากการกำทอน ความถี่กำทอนสามารถบ่งชี้การทำงานของอวัยวะในการออกเสียงสระซึ่งมีอยู่ 3 ความถี่ที่สำคัญคือ ความถี่ฟอร์แมนท์ที่ 1 ฟอร์แมนท์ที่ 2 และฟอร์แมนท์ที่ 3 ตามลำดับ ซึ่งความถี่ทั้งสามนี้สามารถแยกความแตกต่างของสระออกได้ อธิบายจากความสัมพันธ์ระหว่างความถี่ฟอร์แมนท์ทั้งสามกับการทำงานของอวัยวะในการออกเสียงสระได้ดังนี้ (สุดาพร ลักษณ์นิยานวิน, 2529)

กฎเกี่ยวกับความยาวของช่องทางเดินเสียง

ความถี่กำทอนของเสียงสระจะมีค่าตรงข้ามกับความยาวของทางช่องทางเดินเสียง ถ้าช่องออกเสียงยาว ค่าของความถี่ฟอร์แมนท์จะต่ำ ถ้าเราแบ่งช่องปากออกเป็น 3 ตอนใหญ่คือ

- ก. บริเวณระหว่างลิ้นส่วนกลางกับเพดานแข็ง
- ข. บริเวณระหว่างลิ้นส่วนหลังกับเพดานอ่อน
- ค. บริเวณระหว่างโคนลิ้นกับผนังคอ

จะเขียนกฎความสัมพันธ์ของค่าความถี่ฟอร์แมนท์กับการออกเสียงสระได้ดังนี้

1. กฎเกี่ยวกับค่าของความถี่ฟอร์แมนท์ที่ 1

ค่าของความถี่ฟอร์แมนท์ที่ 1 จะลดลงถ้าบริเวณระหว่างลิ้นส่วนกลางและเพดานแข็ง และบริเวณระหว่างลิ้นส่วนหลังกับเพดานอ่อนแคบลง

2. กฎเกี่ยวกับค่าของความถี่ฟอร์แมนท์ที่ 2

ค่าของความถี่ฟอร์แมนท์ที่ 2 จะเพิ่มขึ้นถ้าช่องระหว่างลิ้นส่วนกลางกับเพดานแข็งแคบลง

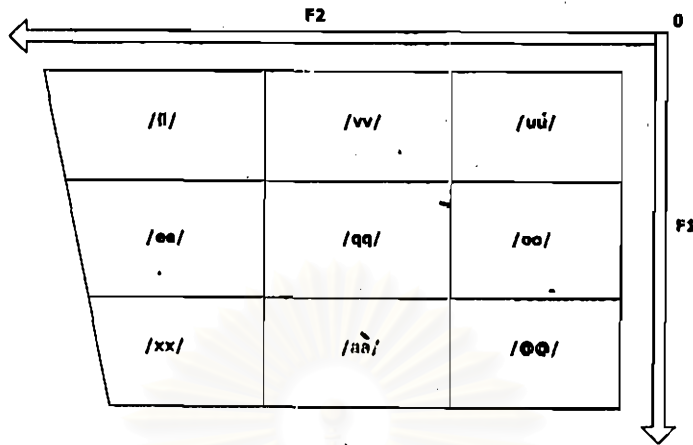
3. กฎเกี่ยวกับการห่อริมฝีปาก

ค่าของความถี่ฟอร์แมนท์ทุกตัวจะลดลงถ้ามีการห่อริมฝีปาก ช่องริมฝีปากเวลาห่อกลมยิ่งแคบเท่าใดค่าของความถี่ฟอร์แมนท์ก็จะลดต่ำลงเท่านั้น

ตารางที่ 2.2 หน่วยเสียงสระเดี่ยวภาษาไทย

ความสูงของลิ้น	ตำแหน่งการเคลื่อนไหวของลิ้น		
	ด้านหน้า	กึ่งกลาง	ด้านหลัง
สูง	อี, อี /i/, /ii/	อึ, อือ /v/, /vv/	อุ, อุ /u/, /uu/
กลาง	เอะ, เอ /e/, /ee/	เออะ, เออ /q/, /qq/	โอะ, โอ /o/, /oo/
ต่ำ	แอะ, แอ /x/, /xx/	อะ, อา /a/, /aa/	เอาะ, ออ /@/, /@@/

ถ้านำกฎเกี่ยวกับความยาวของช่องทางเดินเสียงมาสร้างเป็นตารางเทียบเสียงสระภาษาไทยจะมีลักษณะดังแสดงในรูปที่ 2.21



รูปที่ 2.21 ความสัมพันธ์ของค่าความถี่ฟอร์แมนท์กับความยาวของทางช่องทางเดินเสียงสำหรับเสียงสระภาษาไทย (ดัดแปลงจาก สุดาพร ลักษณะียนาวิน, 2529)

2.5 การรู้จำแบบรูปโดยวิธีเชิงสถิติ (Statistical Pattern Recognition)

ขั้นตอนที่สำคัญสำหรับการรู้จำแบบรูปโดยวิธีเชิงสถิติคือการพัฒนากลยุทธ์ในการตัดสินใจ (decision strategy) การเลือกกฎเกณฑ์การตัดสินใจมีหลายวิธีเช่น (Schalkoff, 1992)

1. การแปลง a priori class probability $P(w_i)$ ไปเป็น a posteriori probability $P(w_i|x)$ โดยที่ w_i เป็นกลุ่มข้อมูลที่ i และ x เป็นข้อมูลที่นำมาตัดสินใจ
2. การสร้างตัววัดความผิดพลาดและเลือกกฎเกณฑ์การตัดสินใจที่จะทำให้ความผิดพลาดที่คำนวณจากตัววัดความผิดพลาดที่สร้างขึ้นมีค่าน้อยที่สุด

ในงานวิจัยนี้จะกล่าวถึงกฎเกณฑ์การตัดสินใจในแบบแรกเท่านั้นและจะสนใจเฉพาะกรณีที่มีข้อมูลมีการแจกแจงแบบเกาส์ (Gaussian Distribution) มีจำนวนกลุ่มข้อมูล 2 กลุ่ม และข้อมูลมี 1 มิติ การอธิบายการแจกแจงแบบเกาส์เป็นไปตามสมการที่ (2.77)

$$p(x|w_i) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left\{-\frac{1}{2}\left(\frac{x - \mu_i}{\sigma_i}\right)^2\right\} \quad (2.77)$$

โดยที่ μ_i เป็นค่าเฉลี่ยของกลุ่มที่ i และ σ_i เป็นความแปรปรวนของกลุ่มที่ i จากกฎของ Bayes

$$P(w_i|x) = \left[\frac{p(x|w_i)P(w_i)}{p(x)} \right] \quad (2.78)$$

โดยที่ $p(x)$ คือ unconditional density function คำนวณจากสมการที่ (2.79)

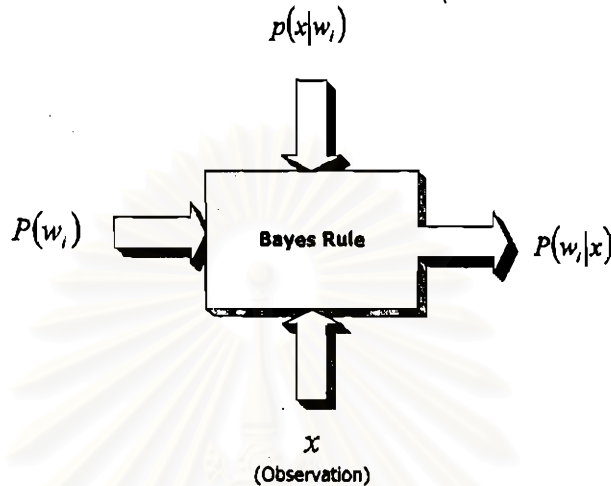
$$p(x) = \sum_i p(x|w_i)P(w_i) \quad (2.79)$$

กฎเกณฑ์การตัดสินใจเป็นดังนี้

ถ้า $p(x|w_1)P(w_1) > p(x|w_2)P(w_2)$ เลือกให้ x อยู่ในกลุ่ม w_1

ถ้า $p(x|w_1)P(w_1) < p(x|w_2)P(w_2)$ เลือกให้ x อยู่ในกลุ่ม w_2

ในรูปที่ 2.22 แสดงการแปลง a priori class probability $P(w_i)$ ไปเป็น a posteriori probability $P(w_i|x)$ โดยใช้กฎของ Bayes



รูปที่ 2.22 การแปลง a priori class probability $P(w_i)$ ไปเป็น a posteriori probability $P(w_i|x)$ (Schalkoff, 1992)

การคำนวณความผิดพลาดจากแบบจำลองทางสถิติที่มีการแจกแจงแบบเกาส์

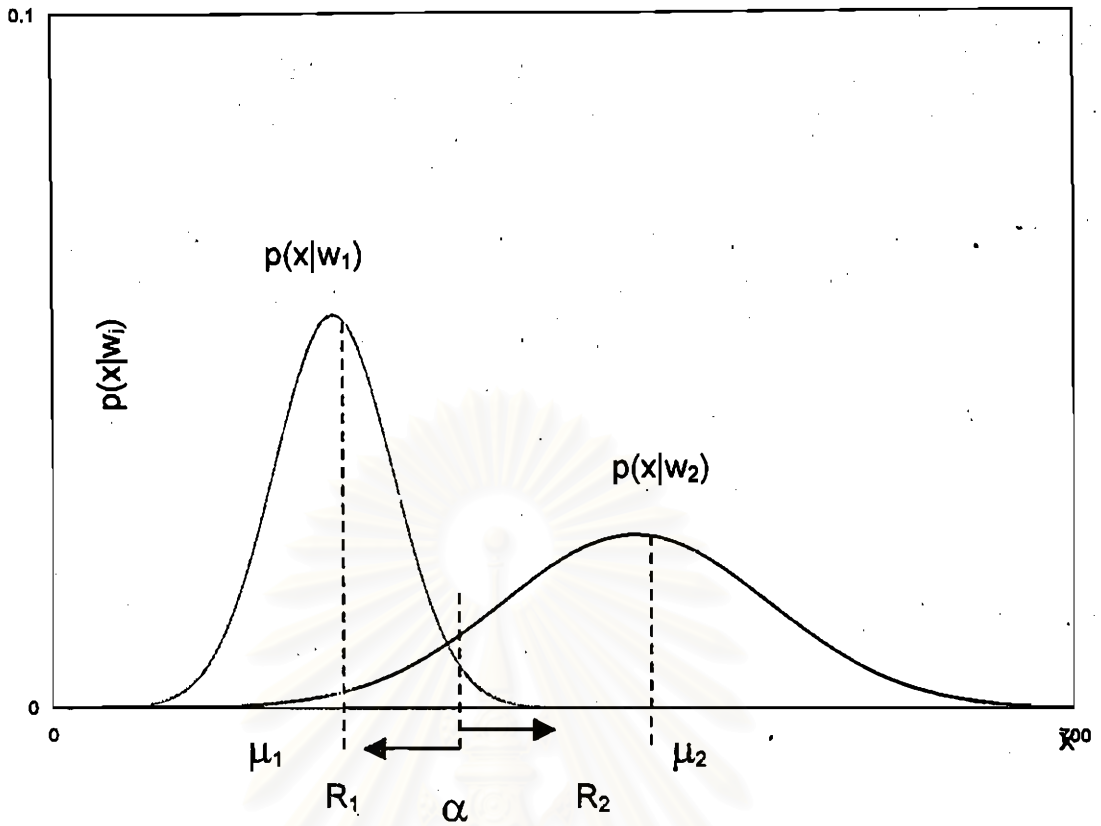
การตัดสินใจในกรณีที่ใช้แบบจำลองทางสถิติที่มีการแจกแจงแบบเกาส์มีความผิดพลาดเกิดขึ้นคือ $P(error)$ ที่ถูกกำหนดโดย (Schalkoff, 1992)

$$P(error) = P(x \text{ is assigned to the wrong class})$$

ในกรณีที่จำนวนกลุ่มข้อมูลมี 2 กลุ่มจะได้

$$\begin{aligned} P(error) &= P(\text{choose } w_1 \text{ and } x \text{ actually from } w_2) \\ &\quad + P(\text{choose } w_2 \text{ and } x \text{ actually from } w_1) \\ &= P(error|w_1)P(w_1) + P(error|w_2)P(w_2) \\ &= P(x \in R_2|w_1)P(w_1) + P(x \in R_1|w_2)P(w_2) \\ &= P(x < \alpha|w_2)P(w_2) + P(x > \alpha|w_1)P(w_1) \\ &= P(w_2) \int_{-\infty}^{\alpha} p(\zeta|w_2) d\zeta + P(w_1) \int_{\alpha}^{\infty} p(\zeta|w_1) d\zeta \end{aligned} \quad (2.80)$$

การเลือกค่าจุดแบ่งกลุ่มข้อมูล α จะมีผลต่อการตัดสินใจในการที่จะแบ่งข้อมูล 1 มิติเป็นสองกลุ่มคือ R_1 และ R_2 ตามลำดับซึ่งรายละเอียดของบริเวณที่มีความผิดพลาดเกิดขึ้นแสดงในรูปที่ 2.23



รูปที่ 2.23 บริเวณที่มีความผิดพลาดในการตัดสินใจของการใช้แบบจำลองทางสถิติที่มีการแจกแจงแบบเกาส์ (Schalkoff, 1992)

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย