บทที่ 5.

เสียงพูด

5.1 อวัยวะที่ใช้ในการเปล่งเสียง(Articulation)

อวัยวะที่ใช้ในการเปล่งเสียง แบ่งได้เป็น 3 พวกใหญ่ ๆ คือ[27]

1. อวัยวะที่ใช้ในการสร้างลม คือส่วนที่ทำให้เกิดการเคลื่อนไหวของลม

2. อวัยวะส่วนที่เคลื่อนที่ได้(Active Articulators) หมายถึง อวัยวะที่ติดกับกระดูกคางส่วนล่าง ได้แก่ ริมฝีปากล่าง และลิ้น

3. อวัยวะส่วนที่เคลื่อนที่ไม่ได้(Passive Articulators) หมายถึงอวัยวุะที่ติดกับกระดูกครางส่วนบนได้แก่ ริมฝีปากบน ฟันบน ปุ่มเหงือก เพดานแข็ง เพดานอ่อน ลิ้นไก่ และอวัยวะที่เป็นช่องว่างได้แก่ ช่องคอ ปาก ช่องจมูก

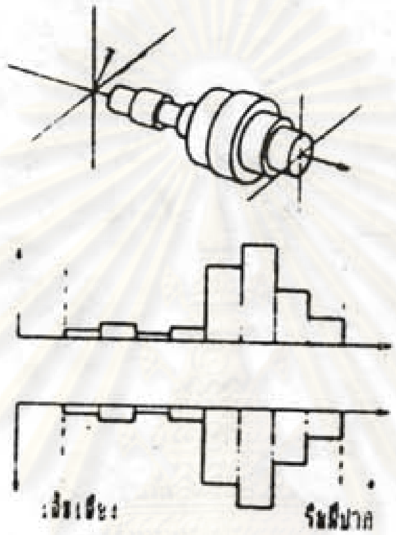
และถ้าจะมองถึงลำดับของการเกิดเสียงแล้ว จะแบ่งได้เป็น 3 ขั้นตอนใหญ่ ๆ คือ

ขั้นตอนที่1 จุดเริ่มต้น(Iniation) อวัยวะที่ใช้ในขั้นตอนนี้ คือ ปอด เป็นขั้นตอนที่ลมเริ่มถูกขับออกจากปอด ผ่านเข้าไปสู่ขั้นตอนที่2

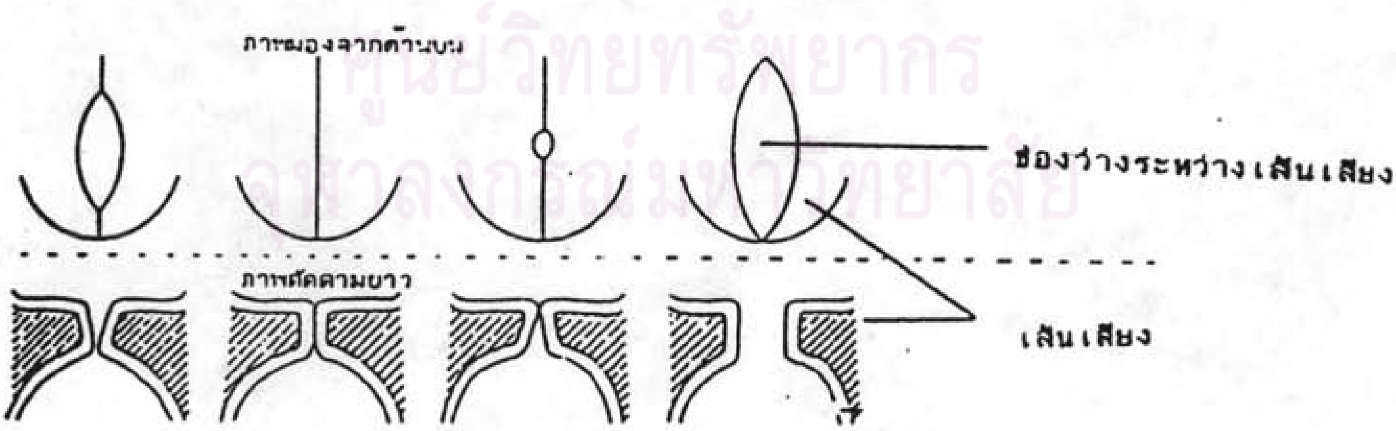
ขั้นตอนที่2 การดัดแปลงลมที่เส้นเสียง(Phonation) อวัยวะที่ใช้ในขั้นตอนนี้ คือส่วนที่ต่อจากปอดขึ้นมา จนถึงกล่องเสียง เป็นขั้นตอนที่ลมจากปอดจะผ่านมายังหลอดลมและกล่องเสียง ซึ่ง ณ ที่กล่องเสียงนี้ เส้นเสียงจะทำหน้าที่เป็นลิ้นเปิดปิด ทำให้เกิดเสียงได้ 2 ชนิดคือ ถ้าเส้นเสียงเปิดตลอดเวลาที่ลมผ่าน ลมจะผ่านออกมาได้โดยสะดวก ซึ่งจะทำให้เกิดเสียงชนิด ไม่ก้อง แต่ถ้าเส้นเสียงปิดกั้นลมไว้ ลมที่ผ่านออกมาจะเพิ่มแรงดันมากขึ้นจนเส้นเสียงเปิดและปิดสลับกันไป ทำให้เกิดเสียงชนิด ก้อง และเรียกความถี่ในการเปิดปิดของเส้นเสียงว่า ความถี่พื้นฐาน(FundamentalFrequency)

ขั้นตอนที่3 การเปลี่ยนแปลงลักษณะเสียง(Articulation) อวัยวะที่ใช้ในขั้นตอนนี้ คือ ส่วนที่ต่อจากกล่อเสียงจนถึงริมฝีปาก และช่องว่างใหญ่อีกสองส่วนคือ ช่องปากและช่องจมูก ขั้นตอนนี้ลมที่ผ่านออกมาจากกล่องเสียง จะถูกแปลงให้เกิดเสียงในลักษณะต่าง ๆ

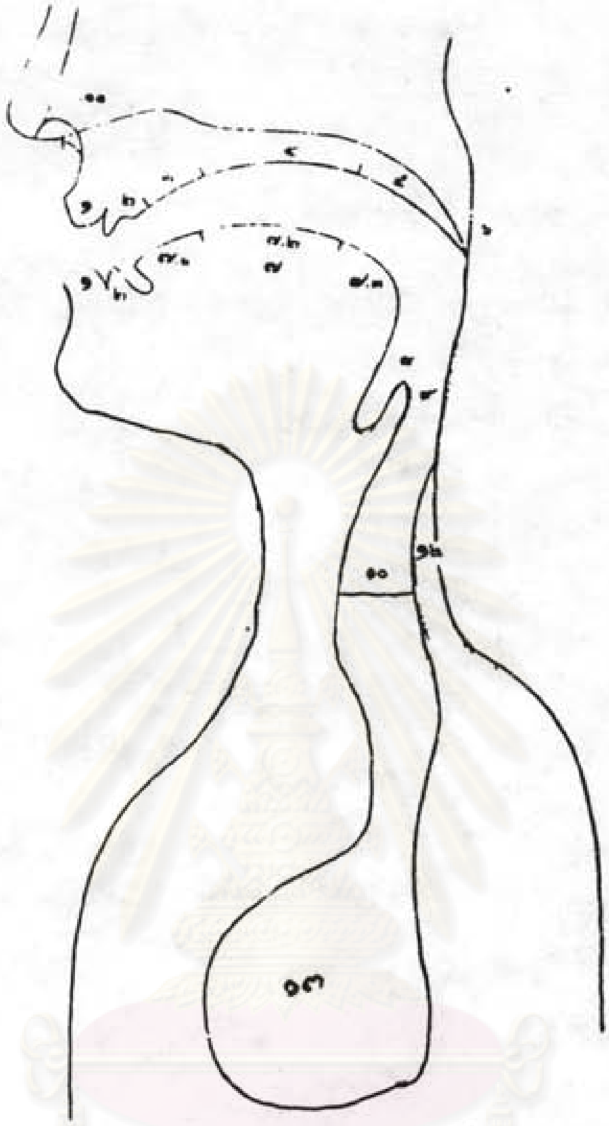
การเปลี่ยนแปลงของลักษณะเสียงต่าง ๆ นั้น จะเกิดจากอวัยวะภายในช่องปากและช่องจมูกเป็นส่วนใหญ่ ทั้งนี้เพราะว่าภายในช่องว่างภายในปากจะมีการเปลี่ยนแปลงขนาด ตลอดเวลาที่เปลี่ยนแปลงเสียง และช่องจมูกก็จะถูกควบคุมโดยเพดานอ่อน(Velume) ดังนั้นถ้าช่องว่างต่าง ๆ เหล่านี้เปลี่ยนขนาดไป ก็จะทำให้ความถี่กำทอน (Resonant Frequency) ของเสียงเปลี่ยนไปด้วย ซึ่งช่องว่างเหล่านี้เปรียบเสมือนกับ ท่อที่มีขนาดต่าง ๆ กันมาต่อกันอยู่ดังรูปที่ 9



รูปที่ 9 ช่องว่างภายในปากเปรียบเสมือนท่อขนาดต่าง ๆ มาต่อกัน



รูปที่ 10 แสดงการเคลื่อนไหวของเส้นเสียง



๑. ริมฝีปาก (Lips)

๒. ฟัน (Teeth)

๓. ปุ่มเหงือก (Alveola or Teeth-ridge)

๔. เพดานแข็ง (Hard Palate or Palate)

๕. เพดานอ่อน (Soft Palate or Velum)

๖. ลิ้นไก่ (Uvula)

๗. ลิ้น (Tongue)

๗.๑ ปลายลิ้น (Blade of the Tongue)

๗.๒ หน้าลิ้น (Front of the Tongue)

๗.๓ หลังลิ้น (Back of the Tongue)

๘. ลิ้นปิดหลอดลม (Epiglottis)

๙. ช่องคอ (Pharynx)

๑๐. เส้นเสียง (Vocal cords)

๑๑. ช่องจมูก (Nasal cavity)

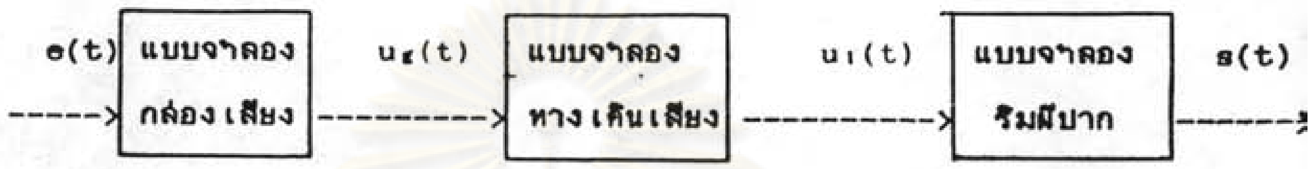
๑๒. เส้นเสียงปลอม (False Vocal cords)

๑๓. ปอด

รูปที่ 11 อวัยวะที่ใช้ในการออกเสียง[27]

5.2 แบบจำลองของเสียงพูด

แบบจำลองของเสียงพูด ได้พัฒนาครั้งแรกโดย Fant ในปี ค.ศ.1950 ซึ่งมีรูปแบบดังรูปที่ 12[16]



รูปที่ 12 แบบจำลองการเกิดเสียง

จากหลักการในการเกิดเสียงพูดในหัวข้อ 5.1 ดังกล่าว ทำให้แบบจำลองที่พัฒนาขึ้นมานี้มีลักษณะการทำงาน คล้าย ๆกับอวัยวะในการก่อให้เกิดเสียงของมนุษย์ โดยเริ่มจากสัญญาณที่เป็นจังหวะ (Periodic) หรือสัญญาณที่ไม่เป็นจังหวะ(Nonperiodic) โดยสัญญาณที่เป็นจังหวะนั้น จะมีช่วงเวลาห่างกัน P ซึ่งใช้แทนลักษณะการเกิดเสียงก้อง และสัญญาณที่ไม่เป็นจังหวะจะใช้เป็นตัวแทนของการเกิดเสียงไม่ก้อง สัญญาณในส่วนนี้ใช้สัญลักษณ์แทนเป็น e(t) ซึ่งถ้ากำหนดให้ช่วงจังหวะมีค่าเป็น I และค่า e(t) ถูกนอร์มอลไลซ์ให้มีค่าเป็น 1 แล้ว กลุ่มของสัญญาณนี้เมื่อแปลงให้อยู่ในรูปของ Z-Transform แล้วจะได้ว่า[16]

$$E(z) = \sigma \sum_{n=0}^{\infty} (z^{-I})^{n}$$

$$= \frac{\sigma}{1-z^{-I}}$$

ซึ่งเป็นส่วนที่จะเข้าไป(input)สู่แบบจำลองของกล่องเสียง(Glottal Model)ซึ่งมีรูปแบบ

$$G(z) = \frac{1}{(1-e^{-cT}z^{-1})^{2}} \quad \text{สำหรับ} \quad |z| > 1$$

$$= \frac{z^2}{(z-e^{-cT})^2}$$

หลังจากสัญญาณผ่านแบบจำลองของกล่องเสียงแล้วจะได้ผล(output) ออกมาเป็น ug(t) ซึ่งจะเข้าไปสู่แบบจำลองของทางเดินของเสียง(Vocal Tract Model) ซึ่งจะเป็นลักษณะของ ตัวกรองความถี่(Filter) แบบ all-pole มี 2-pole ต่อเนื่องกัน เพื่อเป็นตัวทำให้เกิดการกำทอนของเสียงคล้าย ๆ กับหน้าที่ช่องว่างภายในปากของคนเรา ซึ่งตัวกำทอนแต่ละตัว จะทำหน้าที่ก่อให้เกิดความถี่กำทอน(formant) ค่าหนึ่ง ๆ ในทางทฤษฎีแล้วตัวก่อให้เกิดการกำทอนนี้ จะต้องมีอยู่เป็นจำนวนมาก ซึ่งเราสามารถจะจำลองรูปแบบในส่วนช่องว่างภายในปากได้ว่าเป็น

$$V(z) = \frac{1}{\prod_{i=1}^{K}[1-2e^{-c_iT}\cos(b_iT)z^{-1}+e^{-2c_iT}z^{-2}]}$$

โดย K คือจำนวนความถี่กำทอนและแต่ละความถี่จะมีค่าเท่ากับ $b_i/2$ ซึ่งมีแถบความถี่ของสัญญาณ(Bandwidth)เท่ากับ $c_i/2$

หลังจากนั้นสัญญาณ $u_l(t)$ ที่ออกมาจากแบบจำลองของช่องปาก จะเข้าสู่แบบจำลองของริมฝีปาก ซึ่งแบบจำลองนี้จะทำหน้าที่แปลงสัญญาณ $u_l(t)$ ให้เป็นสัญญาณเสียงในแง่ของสัญญาณคลื่นในอากาศโดยแบบจำลอง ริมฝีปากจะมีลักษณะดังนี้

$$L(z) = 1-z^{-1}$$

ภายใต้ข้อสมมุติฐานที่ว่า แบบจำลองแต่ละส่วนมีความสัมพันธ์กันแบบ เชิงเส้น(linear) เราจะได้แบบจำลองของการสร้างเสียงในรูป Z-Transform ได้เป็น

$$S(z) = E(z)G(z)V(z)L(z)$$

$$\text{โดย } G(z)V(z)L(z) = \frac{(1-z^{-1})}{(1-e^{-cT}z^{-1})^2\{\prod_{i=1}^{K}[1-2e^{-c_iT}\cos(b_iT)z^{-1}+e^{-2c_iT}z^{-2}]\}}$$

โดยปกติ $-cT$ มักจะมีค่าน้อยกว่า 1 มากดังนั้น $[1-e^{-cT}z^{-1}]$ จึง

ประมาณได้ด้วย $(1-z^{-1})^2$ ซึ่งทำให้แบบจำลองของการสังเคราะห์ง่ายขึ้นเป็น

$$S(z) = E(z)\frac{1}{A(z)}$$

โดย $A(z) = \sum_{i=0}^{M} a_i z^{-i}$ $\qquad (a_i = 1, M > 2K+1)$

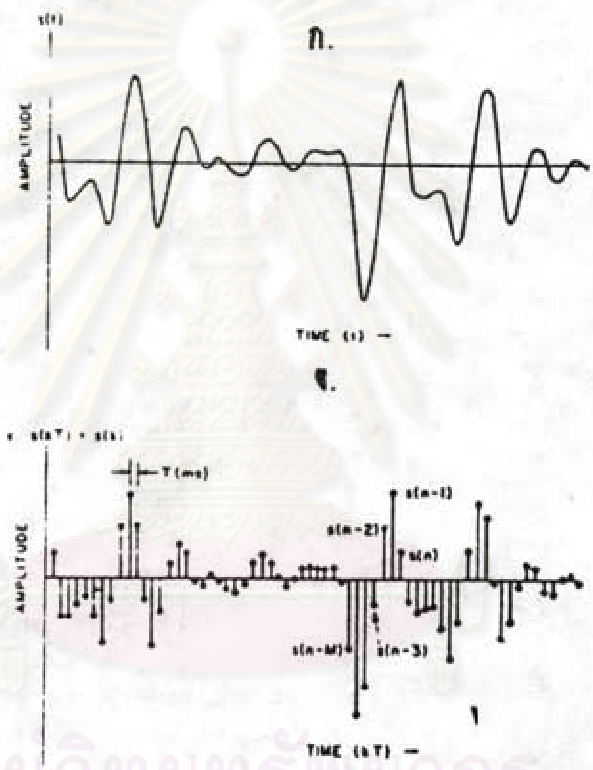
$$= \frac{1}{G(z)V(z)L(z)}$$

และจะได้แบบจำลองของการวิเคราะห์เสียงเป็น

$$E(z) = S(z)A(z)$$

5.3 แบบจำลองของเสียงพูดตามวิธีพยากรณ์เชิงเส้น(Linear Prediction

หลักการของการพยากรณ์เชิงเส้นคือ เราพยายามที่จะคาดการณ์เสียงพูดที่สุ่มขึ้นมาได้ โดยอาศัยความสัมพันธ์เชิงเส้นของข้อมูลเสียงพูดที่ผ่านมา ซึ่งถ้าเราพยายามที่จะหาให้ ผลรวมของผลต่างระหว่างเสียงพูดที่สุ่มขึ้นมากับเสียงพูดที่เราคาดการณ์ยกกำลังสอง ในช่วงเวลาหนึ่ง ๆ มีค่าน้อยที่สุดเราจะได้ค่าสัมประสิทธิ์ของการพยากรณ์อยู่ 1 ชุด (ซึ่งสัมประสิทธิ์ของการ พยากรณ์นี้ก็คือ ค่าที่จะเป็นตัวถ่วงน้ำหนักในความสัมพันธ์เชิงเส้นของเสียงพูดที่ผ่านมา)[16]



รูปที่ 13.1สัญญาณเสียงพูด ก. สัญญาณต่อเนื่อง ข. สัญญาณที่สุ่ม(sampling)ได้

ถ้าเรากำหนดว่า การพยากรณ์สัญญาณ ณ ตำแหน่งเวลาที่ n มีค่าผิดพลาดเป็น e(n) เราจะได้ว่า

$$e(n) = \sum_{i=0}^{m} a_i s(n-i)$$

$$= s(n) + \sum_{i=1}^{m} a_i s(n-i)$$

ซึ่งถ้าเราให้ s(n) เป็นค่าพยากรณ์และมีค่าเท่ากับ $-\sum_{i=1}^{m} a_i s(n-i)$

เราจะได้ว่า $e(n) = s(n)-s(n)$

ใน Z-Transform เราจะได้ว่า

$$E(z) = S(z)A(z)$$

โดย $$A(z) = 1+\sum_{i=1}^{M} a_i z^{-i}$$

หรือ $$S(z) = \frac{E(z)}{A(z)}$$

ซึ่งจะเห็นได้ว่า แบบจำลองของเสียงพูดตามวิธีการพยากรณ์เชิงเส้นมีความสอดคล้องกับแบบจำลองเสียงพูดในหัวข้อ 5.2

ถ้าเรากำหนดให้ เป็นผลรวมของกำลังสองของค่าผิดพลาดในการพยากรณ์ ในช่วงเวลาหนึ่ง ๆ จะได้ว่า

$$\alpha = \sum_{n=n_0}^{n_1} e^2(n)$$

$$= \sum_{n=n_0}^{n_1} [\sum_{i=0}^{M} a_i s(n-i)]^2$$

$$= \sum_{n=n_0}^{n_1} \sum_{i=0}^{M} \sum_{j=0}^{M} a_i s(n-i) s(n-j) a_j$$

โดย n_0 และ n_1 เป็นช่วงของการพยากรณ์

ให้ $$c_{ij} = \sum_{n=n_0}^{n_1} s(n-i)s(n-j)$$

จะได้ว่า $$\alpha = \sum_{i=0}^{M} \sum_{j=0}^{M} a_i c_{ij} a_j$$

เพื่อให้ค่าผิดพลาดในการพยากรณ์น้อยที่สุด เราจะต้องหาการ differentiate ค่า เทียบกับ a_k แล้วกำหนดให้เท่ากับ 0 ซึ่งจะได้ว่า

$$\frac{\partial \alpha}{\partial a_k} = 0 = 2\sum_{i=0}^{M} a_i c_{ik}$$

แต่ $a_0 = 1$ ดังนั้น เราจะได้สมการอยู่ชุดหนึ่งคือ

$$\sum_{i=1}^{M} a_i c_{ik} = -c_{0k} \qquad ;k=1,2,3,\dots\dots\dots M$$

ถ้าเราถอดสมการชุดนี้ออกมา จะได้ค่า สัมประสิทธิ์ของการพยากรณ์

(a_i ;i=1,2,3....M) ซึ่งมีวิธีการในการหาอยู่ 2 วิธี คือ วิธี Covarience และ วิธี AutoCorrelation

สมมุติว่าสัญญาณเสียงพูด ถูกสุ่มมา N จุด โดยมีค่าเป็น s(0),s(1),....s(N-1) สำหรับค่า n_0 และ n_1 ที่จะกำหนดนั้นมีอยู่ 2 แบบใหญ่ ๆ คือ

1. วิธี Covarience จะกำหนดให้ $n_0 = M$ และ $n_1 = N-1$ ดังนั้น ช่วงของการพยากรณ์โดยวิธีนี้ จะอยู่ในช่วง[M,N-1] ซึ่งจะได้ชุดของสมการเป็น

$$\sum_{i=1}^{n} a_i c_{ik} = -c_{0k} \quad ;k=1,2,.......M$$

และ $c_{ik} = \sum_{n=M}^{N-1} s(n-i)s(n-k)$ ซึ่งเขียนเป็นแมทริกซ์ได้ดังนี้

$$\begin{bmatrix} c_{11} & c_{12} & c_{13} \cdots\cdots c_{1M} \\ c_{21} & c_{22} & c_{23} \cdots\cdots c_{2M} \\ \cdot & \cdot & \cdots\cdots \\ \cdot & \cdot & \cdots\cdots \\ \cdot & \cdot & \cdots\cdots \\ c_{M1} & c_{M2} & c_{M3} \cdots\cdots c_{MM} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \cdot \\ \cdot \\ \cdot \\ a_M \end{bmatrix} = - \begin{bmatrix} c_{10} \\ c_{20} \\ \cdot \\ \cdot \\ \cdot \\ c_{M0} \end{bmatrix}$$

แต่ $c_{ij} = c_{ji}$ ดังนั้น Covarience Matrix จึงมีคุณสมบัติสมมาตร

2. วิธี AutoCorrelation จะกำหนดให้ $n_0 = -\infty$ และ $n_1 = \infty$ โดยให้ s(n) = 0 สำหรับ n < 0 และ n > N ซึ่งวิธีการนี้ทำให้การหาค่า c_{ij} ง่ายขึ้น โดย

$$c_{ij} = \sum_{n=-\infty}^{\infty} s(n-1)s(n-j)$$

$$= \sum_{n=-\infty}^{\infty} s(n)s(n+|i-j|)$$

$$= \sum_{n=0}^{N-1-|i-j|} s(n)s(n+|i-j|)$$

$$= r(|i-j|)$$

ทำให้เราได้ AutoCorrelation Matrix ดังนี้

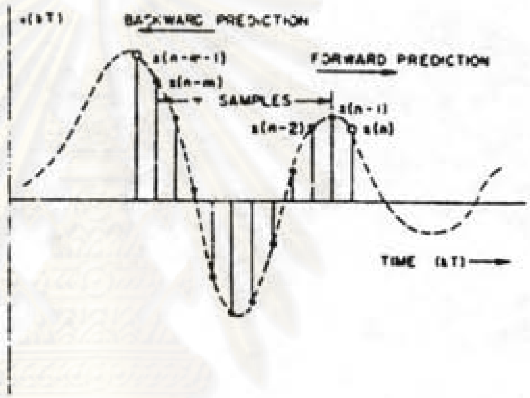
$$
\begin{bmatrix}
r(0) & r(1) & r(2) \ldots\ldots r(M-1) \\
r(1) & r(0) & r(1) \ldots\ldots r(M-2) \\
. & . & \ldots\ldots\ldots\ldots \\
. & . & \ldots\ldots\ldots\ldots \\
. & . & \ldots\ldots\ldots\ldots \\
r(M-1) & r(M-2) & r(M-3) \ldots r(0)
\end{bmatrix}
\begin{bmatrix}
a_1 \\ a_2 \\ . \\ . \\ . \\ a_M
\end{bmatrix}
= -
\begin{bmatrix}
r(1) \\ r(2) \\ . \\ . \\ . \\ r(M)
\end{bmatrix}
$$

ซึ่งแมทริกซ์นี้เป็น toeplitz แมทริกซ์ เพราะสมมาตรกัน และทุกตัวในแนวเส้นทะแยงมีค่าเท่ากัน

เทคนิคของ PARCOR[16]

PARCOR เป็นเทคนิค ที่ใช้หลักการของการพยากรณ์เชิงเส้น ทั้งด้านหน้า และด้านหลังของสัญญาณ ณ ตำแหน่งที่จะพยากรณ์ ซึ่งจะทำให้การคำนวณหาสัมประสิทธิ์ของการพยากรณ์(a_i) รวดเร็วขึ้น และสามารถใช้หลักการของการกระทำซ้ำ(Recursive) ซึ่งจะทำให้สะดวกในการเขียนโปรแกรม

หลักการของ PARCOR จะพิจารณาได้จากสัญญาณเสียง ที่สุ่มมาได้ดัง รูปที่ 13.2



รูปที่ 13.2 แสดงสัญญาณเสียงพูดที่ใช้เทคนิค PARCOR

ถ้าสัญญาณ x_m เป็นสัญญาณที่เราต้องการพยากรณ์ เราจะได้ค่าผิดพลาด ในการพยากรณ์ไปข้างหน้า($x_m^+(n)$) เป็น

$$x_m^+(n) = x(n) - \left[-\sum_{i=1}^{m} a_{mi}x(n-i)\right]$$

$$= \sum_{i=0}^{m} a_{mi}x(n-i) \quad \text{โดย } a_{m0} = 1$$

และจะได้ค่าผิดพลาดของการพยากรณ์ไปข้างหลัง($x_m^-(n)$) เป็น

$$x_m^-(n) = x(n-m-1) - \left[-\sum_{i=1}^{m} b_{mi}x(n-i)\right]$$

$$= \sum_{i=1}^{m+1} b_{mi}x(n-i) \quad \text{โดย } b_{m,m+1} = 1$$

ถ้ากำหนดให้ ผลรวมกำลังสองของค่าผิดพลาด ในการพยากรณ์ไปข้างหน้าเป็น α_m และผลรวมกำลังสองของค่าผิดพลาด ในการพยากรณ์ไปข้าง

หลังเป็น β_m เงื่อนไขที่จะทำให้ การพยากรณ์ผิดพลาดน้อยที่สุดคือ

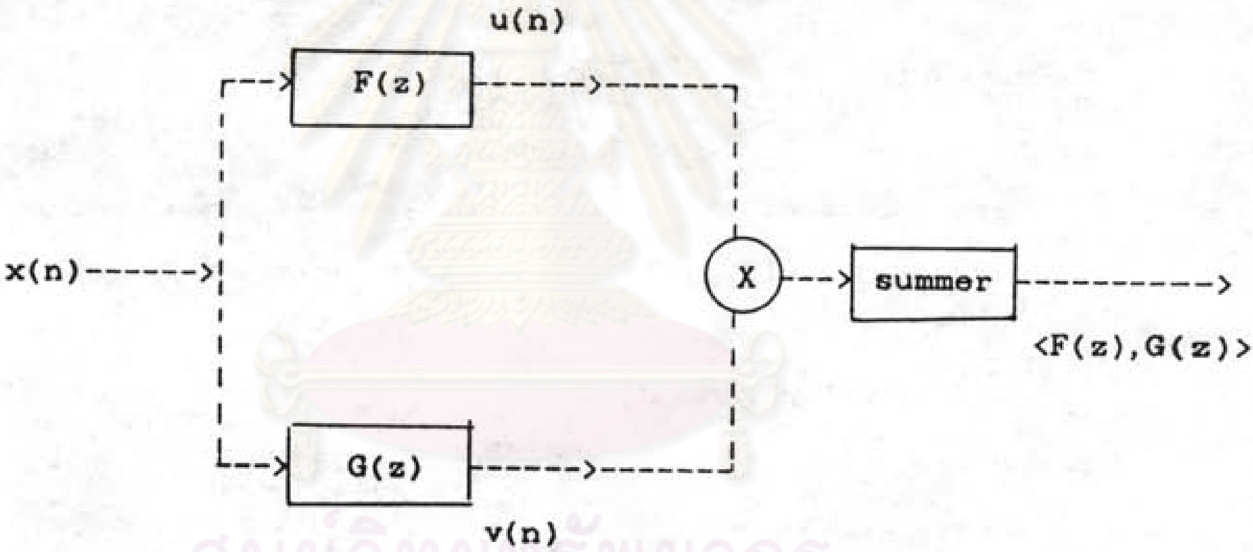
$$\frac{\partial \alpha_m}{\partial a_{mi}} = 0 \quad \text{และ} \quad \frac{\partial \beta_m}{\partial b_{mi}} = 0$$

โดย $$\alpha_m = \sum_{n=-\infty}^{\infty} [x_m^+(n)]^2$$

และ $$\beta_m = \sum_{n=-\infty}^{\infty} [x_m^-(n)]^2 \quad ;m=1,2.......M;i=1,2....m$$

<u>หลักการของ inner products และ orthogonality</u>

หลักการของ inner products ก็คือ การที่ filter 2 ตัว มีข้อมูล(input)เดียวกันในเวลาเดียวกัน เข้ามา และจะได้ผลลัพท์ออกมา 2 ค่า หลังจากนั้นผลลัพท์ทั้งสอง(ในเวลาเดียวกัน)จะคูณกัน และผลคูณนั้นจะถูกรวมกันในช่วงเวลาหนึ่ง ๆ ขบวนการทั้งหมดนี้เรียกว่า inner products



รูปที่ 14 แสดงถึงหลักการของ inner products

จากรูปที่ 14แสดงให้เห็นว่า ผลลัพท์ของ inner products ของ filter F(z) และ G(z) มีค่าเท่ากับ

$$\langle F(z),G(z)\rangle = \sum_{n=n_0}^{n_1} u(n)v(n)$$

และในทำนองเดียวกันก็จะได้ว่า

$$\langle Z^{-i},Z^{-j}\rangle = \sum_{n=n_0}^{n_1} x(n-i)x(n-j)$$

ซึ่งตามวิธีการแบบ Covarience จะได้ว่า

$$\langle Z^{-i},Z^{-j}\rangle = c_{ij} = c_{ji}$$

และตามวิธีการแบบ AutoCorrelation จะได้ว่า

$$\langle Z^{-i}, Z^{-j}\rangle = r(i-j) = r(j-i)$$

สมมุติว่า $x_m'(n)$ และ $x_m^-(n)$ เป็นผลลัพท์ของ filter

2 ตัวคือ $A_m(z)$ และ $B_m(z)$ โดย

$$A_m(z) = \sum_{i=0}^{m} a_{mi} z^{-i} \qquad ;a_{m0} = 1 \quad \ldots\ldots(n)$$

และ

$$B_m(z) = \sum_{i=1}^{m+1} b_{mi} z^{-i} \qquad ;b_{m,m+1} = 1 \quad \ldots$$

โดยอาศัยหลักการของ inner products เราจะได้ว่า

$$\alpha_m = \langle A_m(z), A_m(z)\rangle = |A(z)|^2$$

$$\beta_m = \langle B_m(z), B_m(z)\rangle = |B(z)|^2$$

และจากเงื่อนไข ของการพยากรณ์ที่ต้องการให้ ผลรวมของการผิดพลาดยกกำลังสองมีค่าน้อยที่สุด สามารถพิสูจน์ได้ว่า

$$\langle A_m(z), Z^{-j}\rangle = 0 \qquad ;j=1,2\ldots\ldots m$$

และ

$$\langle B_m(z), Z^{-j}\rangle = 0 \qquad ;j=1,2\ldots\ldots m$$

ผลที่ได้ ก็คือหลักการของ orthogonality

โดยใช้หลักการของ Inner Products และ Orthogonality ทำให้ได้

ผลสรุปว่า

$$\langle A_m(z), Z^{-j}\rangle = \sum_{n=n_0}^{n_1} x_m'(n)x(n-j) = 0$$

และ

$$\langle B_m(z), Z^{-j}\rangle = \sum_{n=n_0}^{n_1} x_m^-(n)x(n-j) = 0$$

จาก (ก) จะได้ว่า $A_{m-1}(z)$ เป็นโพลิโนเมียล(polinomial)ของ Z^{-i}

โดย $i=1,2\ldots(m-1)$ และจาก (ข) จะเห็นว่า $B_{m-1}(z)$ เป็นโพลิโนเมียลของ Z^{-i}

โดย $i=1,2\ldots\ldots m$ ดังนั้น $A_m(z)$ จึงสามารถหาได้จาก $A_{m-1}(z)$ กับ $B_{m-1}(z)$ โดยความสัมพันธ์ดังนี้

$$A_m(z) = A_{m-1}(z) + k_m B_{m-1}(z)$$

จากหลักการของ orthogonality เราจะได้ว่า

$$K_m = \frac{-1}{\beta_{m-1}} \langle A_{m-1}(z), B_{m-1}(z) \rangle$$

$$= \frac{-1}{\beta_{m-1}} \sum_{n=n_0}^{n_1} x_{m-1}^{+}(n)\, x_{m-1}^{-}(n)$$

และ

$$\alpha_{m+1} = \alpha_m - k_{m+1}^2 \beta_m$$

โดยตัวแปร k เรียกว่า partial correlation coefficient หรือสัมประสิทธิ์ของการสะท้อนกลับ (Reflection Coefficient) ซึ่งจะสอดคล้องกับ สัมประสิทธิ์ของการสะท้อนกลับ ที่หามาจากแบบจำลองของทางเดินของเสียง จากล่องเสียงถึงริมฝีปาก

5.4 วิธีในการหาค่าพารามิเตอร์

การหาค่าพารามิเตอร์ของเสียง โดยวิธีการพยากรณ์เชิงเส้นทั้งสองวิธีคือ แบบ AutoCorrelation และแบบ Covarience มีกรรมวิธีในการหาเหมือนกัน จะแตกต่างกันก็ตรงการกำหนดค่า n_0 และ n_1 ซึ่งได้กล่าวไว้ในหัวข้อ ค่าพารามิเตอร์ที่ได้ก็คือ สัมประสิทธิ์ของการพยากรณ์(a_i), สัมประสิทธิ์ของการสะท้อนกลับ(k_i)รวมทั้งค่าผิดพลาดในการพยากรณ์ยกกำลังสอง(α)

ขั้นที่1 คำนวณ Covarience หรือ AutoCorrelation แมทริกซ์

จากวิธี Covarience สัมประสิทธิ์ c_{ij} หาได้จาก

$$c_{ij} = \sum_{n=M}^{N-1} x(n-i)x(n-j)$$

$j = 0,1,2,\ldots,i;\ i = 0,1,2,\ldots,M$ เมื่อกำหนดเงื่อนไขในช่วง $n = M$ ถึง $n = N-1$ ค่า c_{ij} ที่ต้องคำนวณทั้งหมด จะมีจำนวนเท่ากับ $M(M+1)/2$ ด้วยวิธี Autocorrelation สัมประสิทธิ์ c_{ij} กลายเป็น $r(|i-j|)$

$$c_{ij} = r(|i-j|) = r(k) \qquad k = 0,1,2,\ldots,M$$

$$r(k) = \sum_{n=0}^{N-1-k} x(n)x(n+k)$$

ขั้นที่ 2 หาค่าเริ่มต้น

สำหรับค่าเริ่มต้นกรณีที่ $m=0$ จะได้ว่า

$a_{00} = 1$ และ $b_{01} = 1$

ซึ่งจากหลักของ inner product จะได้ว่า

$$\alpha_0 = \beta_0 = r(0)$$

และที่ $m=1$ จะได้ว่า

$$A_1(z) = A_0(z) + k_1\beta_0(z)$$
$$= 1 + k_1 Z^{-1}$$

เมื่อเทียบสัมประสิทธิ์แล้วจะได้ว่า

$a_{10} = 1, a_{11} = k_1$

และจะได้ $\alpha_1 = \alpha_0 - k_1^2\beta_0$

ขั้นที่ 3 หาค่าพารามิเตอร์เมื่อ m เพิ่มขึ้น

ตามวิธี AutoCorrelation เราจะได้ว่า

$$\beta_m = \alpha_m$$

$$b_{mi} = a_{m,m+1-i}$$

และ

$$K_{m+1} = \frac{-1}{\alpha_m} \sum_{i=0}^{m} r(m+1-i)a_{mi}$$

โดย

$$A_{m+1}(z) = A_m(z) + k_{m+1}\,\beta_m(z) \quad \text{............(ค)}$$

ดังนั้นแทนค่า k_{m+1} ใน (ค) และเทียบสัมประสิทธิ์จะได้ว่า

$$a_{m+1,0} = 1$$

$$a_{m+1,i} = a_{mi} + k_{m+1}b_{mi} \quad ;i=1,2........m$$

$$a_{m+1,m+1} = k_{m+1}$$

และสามารถคำนวณ α_{m+1} ใหม่ได้ว่า

$$\alpha_{m+1} = \alpha_m - k_{m+1}^2\,\beta_m$$

รายละเอียดของ ผังการทำงานในส่วนนี้อยู่ในภาคผนวก ก.

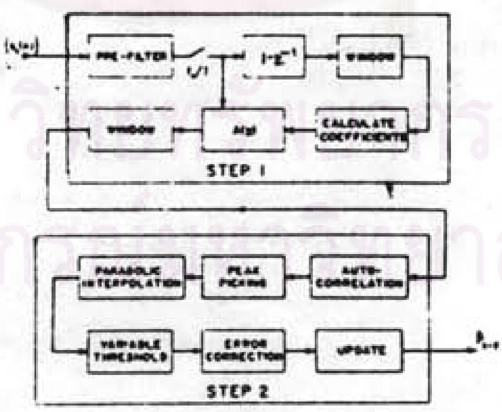
5.5 วิธีในการหาค่าความถี่มูลฐาน

โดยอาศัยคุณสมบัติของแบบจำลองเสียงพูด ตามแบบวิธี Linear Prediction เราสามารถที่จะหาค่า Pitch Period ของเส้นเสียงได้ โดยใช้คุณสมบัติของรูปแบบการวิเคราะห์ดังในหัวข้อที่ 5.3 ได้ว่า

$$E(z) = S(z)A(z)$$

สำหรับการวิจัยในครั้งนี้ได้ใช้วิธีการที่เรียกว่า SIFT (Simplified inverse filter tracking)

โดยหลักการก็คือ เราจะต้องหาค่า a_i ของ inverse filter เพื่อแปลงสัญญาณเสียงพูดให้เป็นสัญญาณที่เส้นเสียง $e(t)$ ซึ่งก็คือ error signal นั่นเอง วิธีการของ SIFT ได้แบ่งเป็นสองขั้นตอนคือ ขั้นตอนการหาค่า error signal และขั้นตอนการตัดสินใจ และประมาณค่า Pitch Period จาก error signal ที่ได้ ดังรูปที่ 15 สำหรับผังงานโดยละเอียดอยู่ในภาคผนวก ข.



รูปที่ 15 แสดงหลักการทำงานของ SIFT