



หลักการวิเคราะห์และสังเคราะห์เสียงพูด

ก่อนที่จะกล่าวถึง เรื่องการวิเคราะห์และการสังเคราะห์เสียงพูด เพื่อให้เข้าใจ ลักษณะของธรรมชาติในการกำเนิดเสียงของมนุษย์ เราควรจะมาเข้าใจกันอย่างคร่าวๆว่า เสียงพูดของคนเราสร้างออกมาได้อย่างไร

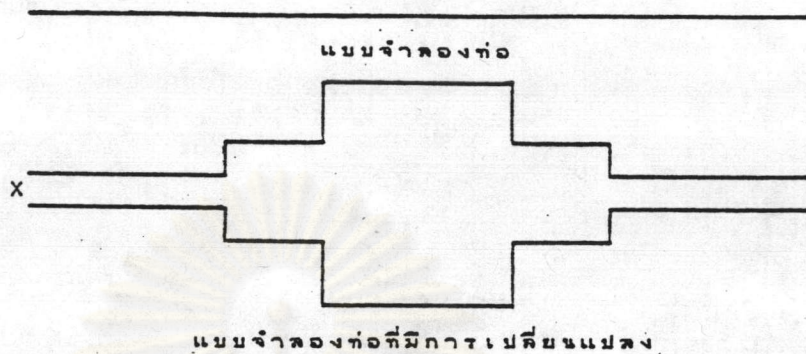
2.1 ทฤษฎีการสร้างเสียงพูด

มนุษย์สร้างเสียงพูดขึ้นมาได้โดยมนุษย์จะมีอวัยวะสำคัญที่ทำให้เกิดเสียงหรือที่เรียกว่า เส้นเสียง (vocal cords) ซึ่งอยู่ในกล่องเสียง (larynx) และจากการสั่นของเส้นเสียงจะทำให้เกิดคลื่นเสียงความถี่ต่างๆผ่านเข้าสู่ลำคอ ผ่านจากลำคอเข้าสู่ช่องปากหรือช่องจมูกออกไปภายนอก ซึ่งขนาดและรูปร่างของอวัยวะภายในช่องปากนี้จะ เป็นสิ่งกำหนดว่าคลื่นเสียงความถี่ ไหนจะยอมให้ปรากฏออกมาให้ได้ยิน (passed through) หรือคลื่นเสียงความถี่ไหนจะถูกดูดซับไว้ไม่ยอมให้ปรากฏออกมา (absorbed) ปรากฏการณ์ที่ทำให้เกิดเสียงพูดออกมาได้นี้เกิดขึ้นได้เนื่องจากรูปร่างของช่องคอและช่องปากของมนุษย์ที่สามารถเปลี่ยนแปลงไปมาได้ อย่างอิสระ ดังรูปที่ 2.1.1 ทำให้ผู้พูดมีความสามารถในการกำหนดคลื่นเสียงความถี่ต่างๆ ให้ปรากฏออกมา หรือดูดซับเอาไว้ได้ตามต้องการ การเรียงลำดับขั้นของการคัดเลือกความถี่ของคลื่นเสียงระดับต่างๆเหล่านี้มันเองที่ทำให้เกิดเสียงพูดออกมา

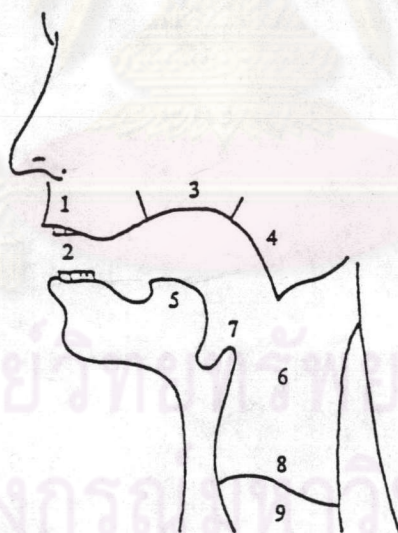
2.1.1 กระบวนการทำให้เกิดเสียง (articulation)

ธรรมชาติจะมีการกำหนดตำแหน่งของอวัยวะ ที่ใช้ในการทำให้เกิดเสียงของมนุษย์เอาไว้ นั่นคือ ริมฝีปาก ฟัน เพดาน และลิ้น ที่อยู่ตามลำดับตลอดช่วงของกล้ามเนื้อช่องคอ เพื่อให้สามารถเปลี่ยนแปลงขนาดและรูปร่างได้ตามต้องการ ดังรูปที่ 2.1.2 เราจะเปล่งคำพูดออกมาโดยการเปล่งออกมาเป็นลำดับขั้นของความถี่คลื่นเสียงที่ได้คัดเลือกไว้ออกมา นั่นคือ จะมีการเปลี่ยนแปลงช่องทางเดินเสียง (vocal tract) ของมนุษย์เป็นลำดับขั้นอย่างต่อเนื่อง

x



รูปที่ 2.1.1 รูปแบบอย่างง่าย ๆ ของช่องทางเดินเสียง(vocal tract) ของมนุษย์
 ที่แสดงถึงการเปลี่ยนแปลงรูปร่างอย่างต่อเนื่องของช่องเสียง
 โดย x แทนตำแหน่งของเส้นเสียง



1. ริมฝีปาก
2. ฟัน
3. เพดานแข็ง
4. เพดานอ่อน
5. ลิ้น
6. ช่องคอ
7. ลิ้นปิดหลอดลม
8. ตำแหน่งรอยเห็นเสียง
9. ช่องสายเสียง

รูปที่ 2.1.2 อวัยวะภายในของระบบการพูดของมนุษย์
 (Veltri, 1985)

อวัยวะที่ใช้ในการออกเสียงนั้นอาจจะแบ่งได้คร่าวๆได้เป็นสองพวกคือ
 (กาญจนา นาคสกุล, 2524)

1) อวัยวะที่เป็นส่วนกระทำอาการ (articulator) หมายถึงอวัยวะส่วนที่เคลื่อนไหวเพื่อผลักลมไปยังส่วนต่างๆ อวัยวะตัวกระทำอาการที่สำคัญ คือ ลิ้น ซึ่งเป็นส่วนที่เคลื่อนไหวได้มากที่สุด

2) อวัยวะที่เป็นฐานที่เกิดเสียงต่างๆ (point of articulator) หมายถึงตำแหน่ง หรือฐานที่เกิดของเสียงต่างๆ เช่น ริมฝีปาก ฟัน เพดานส่วนต่าง ๆ เป็นต้น

จากรูปที่ 2.1.2 อวัยวะส่วนที่มีหน้าที่โดยตรงในการออกเสียงพูดมี ดังนี้

1) ริมฝีปาก เป็นอวัยวะส่วนที่สามารถเคลื่อนไหวได้มาก และทำให้เสียงแตกต่างกันได้มาก เราอาจบังคับริมฝีปากให้อยู่ชิดกัน ห่างกัน ขึ้นออกมา หรือห่อกลม ฯลฯ ก็ได้ ลักษณะริมฝีปากต่างๆ ก็มีอิทธิพลต่อการออกเสียง และการทำให้เสียงแตกต่างกันไปทั้งสิ้น

2) ฟัน เป็นอวัยวะที่เกิดของเสียงหลายชนิด เช่น เมื่อฟันบนกดลงบนริมฝีปากล่างหรือกดกับฟันล่าง ลมที่ผ่านออกมาโดยแรงจะลอดช่องที่พอดฟัน ได้ออกมา ทำให้เกิดเป็นเสียงชนิดที่เรียกว่า เสียงเสียดแทรกที่เกิดที่ฟัน เป็นต้น

3) เพดานแข็ง หรือ เพดานปาก หมายถึงเฉพาะเพดานที่โค้งเป็นกระดูกแข็งเท่านั้น

4) เพดานอ่อน คือส่วนของเพดานที่อยู่ต่อเพดานแข็งไปข้างใน มีลักษณะเป็นกระดูกอ่อนที่ขยับขึ้นลงได้ เวลาหายใจเพดานอ่อนและลิ้นไก่ซึ่งอยู่ปลายเพดานอ่อนจะลดระดับลงมาเปิดช่องให้ลมออกไปทางจมูก เวลาพูดส่วนใหญ่ปลายเพดานอ่อนและลิ้นไก่อจะถูกยกขึ้นไปจรดกับผนังคอ นอกจากเวลาออกเสียงนาสิกเท่านั้นที่เพดานอ่อนจะลดระดับลงมา

5) ลิ้น เป็นส่วนที่เคลื่อนไหวมากที่สุดในการออกเสียง จึงต้องแบ่งออกเป็น 3 ส่วนด้วยกันตามหน้าที่ที่มีในการออกเสียง

ปลายลิ้น (blade of the tongue) หมายถึงส่วนปลายซึ่งสามารถจะยกขึ้นไปแตะอวัยวะส่วนต่าง ๆ ในปากตอนบนได้ง่าย

หน้าลิ้น (front of the tongue) หมายถึงลิ้นส่วนที่อยู่ตรงข้ามกับเพดานแข็ง ถ้าวางลิ้นราบกับปากอย่างในขณะที่ไม่ได้พูด

หลังลิ้น (back of the tongue) หมายถึงส่วนของลิ้น ซึ่งถ้าวางลิ้นราบกับปาก ตามปกติจะอยู่ตรงข้ามกับเพดานอ่อน

6) ช่องคอ (pharynx) หมายถึงโพรงคอซึ่งอยู่ถัดช่องปากลงไปจนถึงเส้นเสียง

7) ลิ้นปิดหลอดลม (epiglottis) เป็นก้อนเนื้อเล็กๆคล้ายลิ้นไก่ อยู่ต่อ

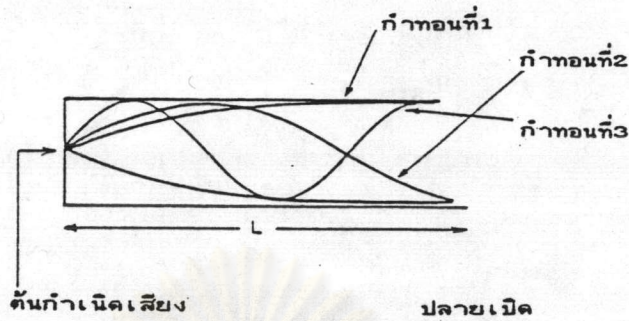
โคนสั้นลงไปในคอ มีหน้าที่ปิดช่องคอเมื่อรับประทานอาหาร และเปิดช่องคอเมื่อกิน

8),9) **เส้นเสียง (Vocal Cords)** เป็นอวัยวะสำคัญที่เกิดของเสียง เส้นเสียงมีลักษณะเป็นกล้ามเนื้อ 2 แผ่นอยู่ภายในกล่องเสียงและปิดขวางปากช่องหลอดลมจากด้านหลังมาด้านหน้า ระหว่างเส้นเสียงจะมีช่องว่างซึ่งเป็นทางให้ลมผ่านเข้าไปถึงปอดและออกมาจากปอดได้ ช่องนี้เรียกว่า ช่องว่างระหว่างเส้นเสียง (glottis) เส้นเสียงทั้งสองสามารถจะดึงออกให้ห่างจากกัน หรือดึงเข้ามาชิดกันก็ได้ เส้นเสียงเป็นส่วนทำให้เกิดเสียงพูดขึ้น ระบบของเสียงพูดสามารถที่จะพิจารณาได้ว่าประกอบด้วยลำดับของ ท่อ และ ช่องที่ต้องออกมาจากปอด ไปยังปากและจมูก ท่อและช่องนี้จะมีความยาวโดยประมาณ 7 นิ้ว เส้นเสียงจะอยู่ในตำแหน่งปลายที่ตรงข้ามกับซี่ปอด ทำหน้าที่ควบคุมการไหลของลมจากปอด ผ่านเข้าสู่ช่องเสียง ภายใต้การควบคุมของกล้ามเนื้อส่วนประกอบของช่องเสียงที่มีลักษณะเป็นท่อจะสามารถเปลี่ยนแปลงรูปร่างได้ในอัตราถึง 10 ครั้งต่อวินาที ส่วนเส้นเสียงนั้นจะสามารถเปิดปิดด้วยอัตราเร็วประมาณ 100-300 ครั้งต่อวินาที การเปลี่ยนแปลงรูปร่างของช่องเสียง และรูปร่างและตำแหน่งของสื่อกลางที่ทำให้เกิดเสียงดังกล่าวนี้รวมเรียกว่า กระบวนการทำให้เกิดเสียง (Veltri, 1985)

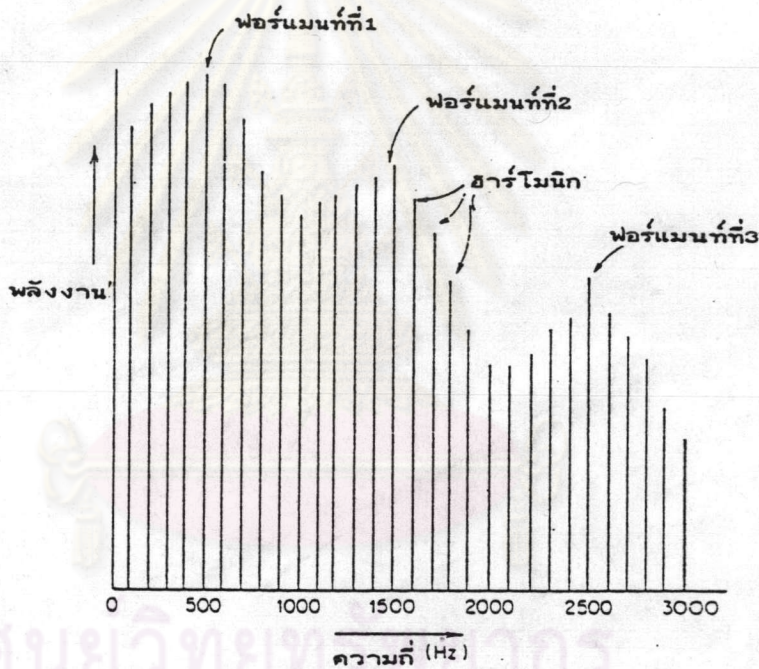
รูปแบบจำลองอย่างง่ายของช่องเสียง ก็อาจจะมองได้เป็นลักษณะของท่อทรงกลมที่มีต้นกำเนิดของเสียงอยู่ที่ปลายข้างหนึ่ง (ส่วนของกล่องเสียง) และปลายอีกข้างหนึ่งจะเปิด (ส่วนของปาก) ดังรูปที่ 2.1.3 ดังนั้นมันจะเกิดการก้ำกอนภายในท่อได้ที่มีความยาวคลื่นเท่ากับ $4L, 4L/3, 4L/5, \dots$, โดยที่ L ก็คือความยาวของท่อ ซึ่งถ้าคิดเป็นความถี่ที่เกิดการก้ำกอนก็จะได้ความถี่ที่ $c/4L, 3c/4L, 5c/4L, \dots$, Hz โดยที่ c คือค่าความเร็วของเสียงในอากาศ และถ้าจะคำนวณหาความถี่ในการก้ำกอนของช่องเสียงของคนซึ่งปกติช่องเสียงของคนเราจะมีความยาวประมาณ 7 นิ้ว หรือ 17 เซนติเมตรดังกล่าว และ c มีค่าเท่ากับ 340 เมตรต่อวินาที ดังนั้นก็จะมีการก้ำกอนที่ความถี่ประมาณ 500 Hz, 1500 Hz, 2500 Hz, ..., เป็นต้น

เมื่อกล่องเสียงกระตุ้นให้เกิดรูปคลื่นที่ประกอบไปด้วยฮาร์โมนิก (harmonic) ต่างๆมากมาย การก้ำกอนของช่องเสียงนี้จะสร้างรูปคลื่นเสียงที่มียอดสูงเด่นขึ้นมาเมื่อดูจากสเปกตรัมพลังงานของรูปคลื่นซึ่งเรียกกันว่า ฟอर्मานต์ (formants) ของเสียง ดังรูปที่ 2.1.4 ฟอर्मานต์ที่มีความถี่ต่ำสุดจะเรียกว่า ฟอर्मานต์ที่หนึ่ง ซึ่งจะอยู่ที่ความถี่ประมาณ 200 Hz ถึง 1000 Hz ทั้งนี้ขึ้นอยู่กับขนาดของช่องเสียงด้วย ส่วนฟอर्मานต์ที่สองที่อยู่ถัดไปก็จะอยู่ที่ความถี่ประมาณ 500 Hz ถึง 2500 Hz และฟอर्मานต์ที่สามที่ปรากฏในรูปจะอยู่ที่ความถี่ประมาณ 1500 Hz ถึง 3500 Hz เป็นต้น โดยที่ฟอर्मานต์ที่หนึ่งจะเป็นตัวที่แสดงคุณลักษณะที่สำคัญมาก

อันหนึ่งในการพิจารณาประยุกต์ใช้ในด้านการวิเคราะห์สัญญาณเสียงทางด้านสเปกตรัม



รูปที่ 2.1.3 แสดงการเกิดการก้ำทอนภายในแบบจำลองของช่องเสียง



รูปที่ 2.1.4 สเปกตรัมของพลังงานของเสียง

(Witten, 1982)

ในขณะที่เราทำเสียงกระซิบ เราจะสร้างเสียงพูดเหล่านี้ได้โดยไม่ต้องมีการสั่นของเส้นเสียงเลย ด้วยเหตุนี้จึงเป็นสิ่งที่ชี้ชัดว่าการเปลี่ยนแปลงรูปร่างของช่องทางเดินเสียงเป็นสิ่งที่สำคัญ ส่วนการสั่นของเส้นเสียงนั้นเป็นสิ่งที่รองลงไป ในความเป็นจริงแล้วระหว่างที่พูดตามปกติมนุษย์จะทำการสร้างเสียงขึ้นได้ทั้งโดยมีการสั่นและไม่มีการสั่นของเส้นเสียง เสียงทั้งสองรูปแบบนี้ เราสามารถจำแนกเป็นเสียงที่เกิดขึ้นขณะที่เส้นเสียงสั่นก็คือเสียงก้อง (voiced) และเสียงที่เกิดขึ้นขณะที่เส้นเสียงไม่สั่นก็คือเสียงไม่ก้อง (unvoiced)

เสียงก้อง เกิดจากการสั่นของเส้นเสียงซึ่งจะสั่นในลักษณะเป็นคาบ (periodic) กล่าวเนื้อหาของเส้นเสียงจะเป็นตัวควบคุมความเร็วในการสั่นของเส้นเสียงแบบเป็นคาบ ความเร็ว-ช้าของการสั่นนี้เป็นตัวกำหนดระดับเสียงสูงต่ำ (pitch) ในการพูด เสียงผู้ชายโดยปกติจะมีเสียงค่อนข้างต่ำคือมีอัตราการสั่นของเส้นเสียงอยู่ในช่วงความถี่ 50 Hz ถึง 250 Hz หรือค่าเฉลี่ยคือประมาณ 100 Hz สำหรับเสียงผู้ชาย ส่วนเสียงของผู้หญิงจะสูงกว่านี้ขึ้นไปอีกอาจจะถึงประมาณ 400 Hz โดยมีค่าเฉลี่ยประมาณ 200 Hz ส่วนความดังจะขึ้นอยู่กับแรงลมที่อัดออกมาจากปอด สำหรับเสียงก้องที่มีการเปล่งลมขึ้นไปทางช่องโพรงจมูกจะเรียกเสียงที่มีลักษณะนี้ว่า เสียงนาสิก (nasal sound) เช่นเสียงพยัญชนะ /ม/, /น/, /ง/ ตัวอย่างของเสียงก้องในภาษาไทยก็คือเสียงสระทุกเสียงในภาษาไทย (ประยูรฯ, 2527) ดังเช่นคำต่อไปนี้ กา ปลอดภัย ตระกร้า ฯลฯ และเสียงของพยัญชนะบางตัวดังเช่นในคำต่อไปนี้ มาก เป็นต้น เพื่อให้เป็นการแน่ใจถึงการสั่นของเส้นเสียงทดลองจับบริเวณลำคอขณะออกเสียงคำดังกล่าวจะสามารถรู้สึกถึงการสั่นของเส้นเสียงได้เป็นอย่างดี

เสียงไม่ก้อง เป็นเสียงที่ไม่ได้เกิดมาจากการสั่นของเส้นเสียง แต่เกิดขึ้นจากการที่ลมถูกกักอยู่เฉยๆหรือไหลผ่านช่องเปิดแคบๆ ในกรณีหลังจะมีลักษณะคล้ายกับสัญญาณรบกวน (noise) คือไม่มีลักษณะเป็นคาบ ตัวอย่างของเสียงไม่ก้องนั้นเช่นเสียงของพยัญชนะในคำว่า สาย และ ช่า ถ้าลองจับบริเวณลำคอแล้วออกเสียงเฉพาะตัว "ส" หรือ "ซ" จะไม่รู้สึถึงการสั่นของเส้นเสียงเลย เสียงพวกนี้จะถูกสร้างขึ้นที่บริเวณใกล้กับฟันและริมฝีปากโดยการบังคับให้ลมผ่านช่องเปิดเล็กๆระหว่างลิ้นกับฟันและริมฝีปาก ลักษณะดังนี้เป็นเช่นเดียวกับการเกิดเสียง ฟี่ ขึ้นขณะที่มีลมรั่วขึ้นที่ล้อรถยนต์ เสียงไม่ก้องนี้มีอยู่หลายลักษณะด้วยกันเช่น

เสียงเสียดแทรก (fricative) คือเสียงที่ลมต้องผ่านช่องแคบในปากออกมาอย่างไม่ค่อยสะดวก เช่น เสียง /ฟ/ /ส/ /ซ/ เป็นต้น

เสียงกัก (stop) คือเสียงที่มีการเปล่งลมออกมาจากการกักไว้ที่ใดที่หนึ่งภายในปากช่วงเวลาหนึ่ง เช่น เสียงตัวสะกดเสียงสั้นๆของคำว่า นับ นิด นึก เป็นต้น

เสียงระเบิด (plosive) คือเสียงลมที่ถูกปล่อยออกมาอย่างแรงหลังจากถูกกักเอาไว้ เช่น เสียงพยัญชนะ /พ/ /ค/ /ท/ /ก/ เป็นต้น

2.1.2 สรุปหลักสำคัญของการเกิดเสียงพูด

จากการอธิบายข้างต้นเราจะเห็นหลักสำคัญสามข้อ ที่ใช้ในช่องเสียงของมนุษย์ในการสร้างเสียงพูดดังนี้

- 1) ลักษณะการทำงานของเส้นเสียง (phonations) ซึ่งมีสองแบบคือ สั่น

ในกรณีของเสียงก้อง และ ไม่นั่น ในกรณีของเสียงไม่ก้อง

2) ลักษณะการเปล่งเสียง (manner of articulation) คือเป็นเสียงเปิด เช่น เสียงสระและอັสระ เสียงกัก เสียงเสียดแทรก เสียงร่ว เสียงข้างลิ้น เสียงนาสิก เป็นต้น

3) การเปลี่ยนแปลงขนาดและรูปร่างของช่องทางเดินเสียงเพื่อตัดแปลงหรือกักกันลมให้เกิดการกำทอนที่ความถี่ต่าง ๆ กัน

2.2 วิธีการเข้ารหัสสัญญาณเสียงแบบดิจิตอล

วิธีการเข้ารหัสสัญญาณเสียงแบบดิจิตอลก็คือ การแปลงสัญญาณอนาล็อกให้เป็นสัญญาณดิจิตอล (A/D) มีหลักการทำงานคือ แต่ละจุดของสัญญาณรูปคลื่นต้นแบบจะถูกแทนค่าโดยจำนวนเลขที่เป็นรหัสทางดิจิตอลซึ่งมีความสัมพันธ์กับค่าแอมพลิจูดของสัญญาณ โดยมีอัตราการตัดตัวอย่างสัญญาณ (Sampling rate) เป็นค่าที่บอกว่าจะมีการวัดค่าของสัญญาณต้นแบบแต่ละจุดบ่อยครั้งแค่ไหน ดังนั้นการเข้ารหัสสัญญาณเสียงแบบดิจิตอลก็คือลักษณะของการแทนค่ารูปคลื่นสัญญาณอนาล็อกด้วยลำดับของค่าจำนวนเลขในช่วงเวลาที่เท่าๆ กัน

การเข้ารหัสสัญญาณเสียงนั้น เริ่มแรกสัญญาณเสียงจะถูกนำไปผ่านตัวกรองสัญญาณอนาล็อกแบบความถี่ต่ำผ่าน (analog lowpass filter) เพื่อขจัดความถี่ที่สูงกว่าความถี่ที่ต้องการออกไปเสียก่อนและเป็นการป้องกันการเกิดการหลอ้มทับ (aliasing) ในการตัดตัวอย่างสัญญาณด้วย แล้วสัญญาณนี้จะถูกตัดเข้ามาด้วยอัตราการตัดตัวอย่างตามที่ได้ออกแบบไว้ อัตราการตัดตัวอย่างสัญญาณนี้อย่างน้อยต้องมีค่าเป็นสองเท่าของความถี่สูงสุดที่มีในสัญญาณอนาล็อก เพื่อที่จะได้สารสนเทศของความถี่ของสัญญาณที่ครบถ้วนตามทฤษฎีของการตัดตัวอย่างสัญญาณ (sampling theory) และอัตราการตัดตัวอย่างสัญญาณที่จุดวิกฤตตามทฤษฎีของการตัดตัวอย่างสัญญาณนี้เรียกว่า อัตราไนควิสต์ (Nyquist rate) (Carlson, 1975) ซึ่งนิยามเป็นคนแรกโดย Nyquist H. เมื่อปี ค.ศ. 1928 แต่มีการนิสุจน์และยอมรับกันอย่างแพร่หลายโดย Shannon เมื่อปี ค.ศ. 1949 (Parson, 1986) โดยคุณภาพของเสียงที่ได้จะขึ้นอยู่กับองค์ประกอบหลายอย่าง และความถี่ในการตัดตัวอย่างสัญญาณก็เป็นองค์ประกอบที่สำคัญที่สุดอันหนึ่ง อัตราการตัดตัวอย่างสัญญาณยิ่งมากเท่าไรก็จะยิ่งทำให้เสียงที่สร้างขึ้นมาได้ใกล้เคียงกับต้นแบบมากเท่านั้น

แถบความถี่สูงสุดในการสังเคราะห์สัญญาณเสียงของมนุษย์ควรมีค่าครอบคลุมประมาณ 5 kHz ถึงจะได้สัญญาณเสียงของมนุษย์ที่มีคุณภาพดี (Veltri, 1985) ดังนั้นในการเข้ารหัสเสียงพูดนี้ เราจึงควรกรองคลื่นเสียงเพื่อกำจัดความถี่ที่สูงเกิน 5 kHz ออกไปก่อนแล้วจึงทำการ

คัดตัวอย่างสัญญาณเป็นสัญญาณดิจิทัลที่ความถี่ 10 kHz ตามทฤษฎีการคัดตัวอย่างสัญญาณ นั่นคือ
คัดตัวอย่างสัญญาณเสียง 10,000 ครั้งใน 1 วินาทีแล้วบันทึกข้อมูลเหล่านี้ไว้

2.2.1 พืชีเอ็ม (PCM : Pulse Code Modulation)

พืชีเอ็ม เป็นแนวทางหนึ่งที่ได้เสนอการเข้ารหัสเสียงพูดที่มีคุณภาพสูงขึ้นมา แต่มีข้อเสียคือมันต้องการใช้หน่วยความจำในการเก็บข้อมูลจำนวนมากซึ่งในระดับของคอมพิวเตอร์
ที่ใช้ตามบ้านจะสามารถเก็บค่าศัพท์ได้เพียงไม่กี่ค่าเท่านั้น สำหรับเสียงพูดที่มีคุณภาพสูงจะต้อง
ใช้หน่วยความจำในการเก็บข้อมูลดิจิทัลประมาณตั้งแต่ 60,000 บิตขึ้นไปต่อเสียงพูดยาว 1
วินาที (KAZUO INOUE et al., 1983) ตามปกติการเพิ่มส่วนของฮาร์ดแวร์ เช่น ส่วนของการ
แปลงสัญญาณดิจิทัลเป็นสัญญาณอนาล็อก และการใช้วงจรประกอบทางตรรกศาสตร์ก็สามารถที่จะทำ
งานของพืชีเอ็มได้ ในปัจจุบันได้มีวงจรรวม (IC : Integrated Circuit) สำเร็จรูป
ที่ทำหน้าที่แปลงสัญญาณอนาล็อกให้เป็นสัญญาณดิจิทัลโดยใช้เทคนิคพืชีเอ็มออกวางขายให้เลือกใช้
หลายแบบตามวัตถุประสงค์ต่างๆกัน พืชีเอ็ม เป็นการแปลงสัญญาณอนาล็อกไปเป็นสัญญาณดิจิทัล
ที่มีการแปลงทั้งในด้านของเวลาและแอมพลิจูดของสัญญาณ การแปลงในด้านของเวลาก็คือเรื่อง
ของการคัดตัวอย่างสัญญาณ และการแปลงในด้านของแอมพลิจูดก็คือเรื่องของการควอนไทเซชัน
(quantization)

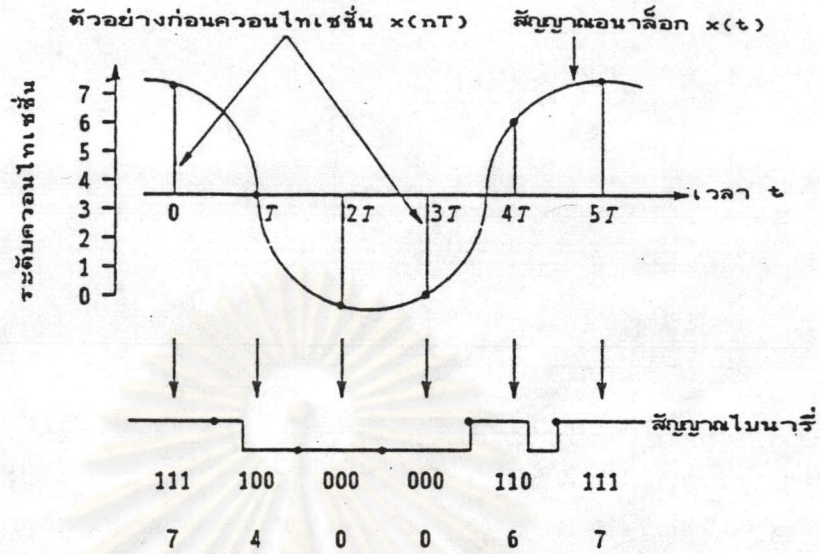
ควอนไทเซชัน เป็นการแปลงค่าแอมพลิจูดของสัญญาณแบบอนาล็อกให้เป็นค่า
แบบดิจิทัล ทำได้โดยการแทนค่าแอมพลิจูดที่จุดนั้นๆ ให้กลายเป็นค่าเลขฐานสองหรือเลข ไบนารี
ที่สัมพันธ์กัน วิธีการที่ง่ายที่สุดก็คือการแทนค่าแอมพลิจูดให้มีค่าเท่ากับค่า ไบนารีที่ระดับควอนไทเซ
ชัน(quantization level)ที่ใกล้เคียงที่สุด ความแตกต่างของระดับควอนไทเซชันแต่ละระดับ
ขึ้นอยู่กับว่าจะใช้จำนวนบิตในการแทนค่า ไบนารีมากน้อยเพียงใด ตัวอย่างเช่นการใช้เลข ไบนารี
ขนาด 12 บิตในการเก็บค่าจะทำให้ได้ระดับควอนไทเซชันที่แตกต่างกันไปถึง $2^{12} = 4096$
ระดับที่จะนำไปเปรียบเทียบกับค่าแอมพลิจูดของสัญญาณอนาล็อกที่อินพุตเข้ามา

ในการพิจารณาความคลาดเคลื่อนของการเข้ารหัสสัญญาณเสียงมักจะยึดเอา
ค่า SNR (Signal to Noise Ratio) ของพลังงานเฉลี่ยของสัญญาณเสียงกับพลังงานเฉลี่ย
ของสัญญาณผิดพลาดโดยคิดค่าเป็นหน่วย เดซิเบล (dB , $10\log_{10} \text{SNR}$) จากความสัมพันธ์
ของจำนวนบิตที่ใช้ในการแทนค่าต่อการคัดหนึ่งครั้ง เราสามารถประมาณค่า SNR ได้จากสมการ
ดังนี้ (O'Shaughnessy, 1987)

$$\text{SNR} = 6.02B - 7.27 \quad (\text{dB})$$

..... 2.2.1

โดย B คือจำนวนบิตที่ใช้ในการแทนค่าต่อการคัดหนึ่งครั้ง



รูปที่ 2.2.1 การเข้ารหัสสัญญาณเสียงด้วยเทคนิค พิชี่เอ็ม

จากสมการที่ 2.2.1 จะเห็นได้ว่าเมื่อมีการใช้จำนวนบิตเพิ่มขึ้นจากเดิมหนึ่งบิตจะมีผลทำให้ค่า SNR มีค่าเพิ่มขึ้นมาประมาณ 6 dB ดังนั้นในกรณีที่เรานำจำนวนบิตในการตัดตัวอย่างสัญญาณเป็นจำนวน 12 บิตก็จะทำให้ได้ค่า SNR ประมาณ 65 dB

การกำหนดค่าให้ทุกๆระดับควอนไทเซชันมีค่าเท่าๆกันหมด หรือในลักษณะที่มีการกระจายของระดับควอนไทเซชันเป็นแบบเชิงเส้น ดังที่ได้กล่าวมาแล้วอาจจะมีปัญหาขึ้นได้ในกรณีที่ของสัญญาณเสียงไม่ก้องซึ่งมีขนาดของแอมพลิจูดค่อนข้างต่ำเมื่อเทียบกับเสียงก้องแล้ว วิธีการแก้ไขปัญหายังหนึ่งก็คือการกำหนดค่าระดับควอนไทเซชันให้เป็นแบบสเกลล็อก คือระดับควอนไทเซชันจะค่อนข้างถี่มากในช่วงแอมพลิจูดต่ำๆ แล้วค่อยๆห่างออกไปที่ค่าแอมพลิจูดมากขึ้น การควอนไทเซชันแบบสเกลล็อกนี้ เราอาจจะสามารถใช้เลขไบนารีแทนค่าเพียง 7 บิตต่อการตัดตัวอย่างสัญญาณก็ได้สัญญาณที่มีคุณภาพพอใช้ได้แล้ววิธีการนี้คือ log PCM

นอกจากนี้ ยังอาจใช้เทคนิคในการสับเปลี่ยนระดับควอนไทเซชัน ให้สัมพันธ์กับขนาดแอมพลิจูดของสัญญาณที่เข้ามา ก็ได้ คือถ้าสัญญาณเสียงที่เข้ามา มีขนาดแอมพลิจูดต่ำ ก็จะใช้ระดับควอนไทเซชันที่มีช่วงค่อนข้างถี่ แต่ถ้าสัญญาณเสียงที่เข้ามา มีขนาดแอมพลิจูดสูง ก็จะไปเลือก ใช้ระดับควอนไทเซชันที่มีช่วงห่างๆ เป็นต้น วิธีนี้เรียกว่า APCM (Adaptively quantized Pulse Code Modulation)

2.3 หลักการวิเคราะห์เสียงพูด (speech analysis)

จากหัวข้อที่ 2.2 เราสามารถที่จะเก็บสัญญาณเสียงไว้ในรูปของสัญญาณดิจิทัลได้แล้ว แต่เนื่องจากข้อมูลที่ไต่ยังมีปริมาณมากเกินไป ไม่เหมาะสมที่จะนำไปประยุกต์ใช้ในด้านสังเคราะห์เสียงโดยตรง จึงควรที่จะลดความซับซ้อนของข้อมูลลงเก็บไว้เฉพาะเท่าที่จำเป็นเท่านั้น ในรูปแบบของตัวพารามิเตอร์ซึ่งจะสามารถลดปริมาณของหน่วยความจำที่ใช้เก็บข้อมูลลง โดยพารามิเตอร์เหล่านี้สามารถจะนำมาประมวลผลแทนค่ากลับออกมาเป็นสัญญาณเสียงได้ในลักษณะเดิม แม้ว่าคุณภาพของเสียงจะลดลงไปบ้างก็ตาม

ดังนั้นจึงจำเป็นที่จะต้องทำการวิเคราะห์สัญญาณเสียงพูดด้วยวิธีการต่างๆ อาจะทั้งในด้านของโดเมนเวลา (time domain) คือพิจารณาจากรูปคลื่นของสัญญาณเสียงตามแนวแกนเวลาโดยตรงหรือในด้านของโดเมนความถี่ (frequency domain) คือพิจารณาจากสเปกตรัมของสัญญาณเสียง โดยมีวัตถุประสงค์หลักของการวิเคราะห์สัญญาณเสียงคือ เพื่อที่จะเลือกเก็บตัวแทนของสัญญาณเสียงที่มีประโยชน์และจำเป็นที่สุดเท่านั้น ในรูปของตัวพารามิเตอร์ที่ประกอบไปด้วยสาระสนเทศที่มีความสัมพันธ์กัน ในรูปแบบที่มีประสิทธิภาพมากที่สุด

2.3.1 การวิเคราะห์ในโดเมนเวลา

การประมวลผลสัญญาณเสียงในโดเมนของเวลานี้ มีประโยชน์ในด้านทำการประมวลผลสัญญาณได้ไม่ยาก การคำนวณใช้เวลาไม่มากนักและง่ายในการวิเคราะห์พิจารณา

การวิเคราะห์ในโดเมนเวลา คือวิธีการประมวลผลสัญญาณเสียงพูด โดยยึดหลักพิจารณาจากรูปคลื่นสัญญาณที่แปรตามแกนเวลาซึ่งมีอยู่หลายวิธีด้วยกัน ถ้าเป็นการประมวลผลสัญญาณเสียงแบบดิจิทัลนั้นเราก็สามารถจะพิจารณาจากตัวอย่างแต่ละตัวอย่างที่คัดเข้ามา อย่างเช่นเอาที่พูดที่ได้จากเทคนิคการเข้ารหัสสัญญาณพีซีเอ็มดังที่ได้กล่าวมาแล้ว ผลของการวิเคราะห์สัญญาณเสียงในโดเมนเวลาก็มี เช่น ค่าพลังงานหรือแอมพลิจูดของเสียง การตรวจสอบว่าเสียงนั้นเป็นเสียงก้องหรือไม่ก้อง การหาคาบของพิชเป็นต้น จากวิธีการต่างๆเหล่านี้จะมีวิธีการที่เป็นพื้นฐานในการวิเคราะห์สัญญาณเสียงในโดเมนเวลาดังที่จะกล่าวต่อไป

2.3.1.1 การวัดพลังงานของสัญญาณเสียง

พลังงานของสัญญาณเป็นตัวแทนอันหนึ่งที่เรามักจะนำมาใช้ในการวิเคราะห์ลักษณะต่างๆของสัญญาณต่างๆไป โดยพลังงานของสัญญาณ $s(n)$ ใดๆที่แปรตามเวลาจะนิยามได้ว่า

$$E = \sum_{n=-\infty}^{\infty} s^2(n) \quad \dots\dots\dots 2.3.1$$

แต่สำหรับสัญญาณเสียงซึ่งเป็นสัญญาณที่แปรเปลี่ยนอยู่ตลอดเวลา ไม่มีเสถียรภาพตามเวลา เราจะต้องแบ่งสัญญาณออกมามีการพิจารณาเป็นช่วงเล็กๆตามแกนเวลาหรือเรียกว่าแบ่งออกเป็นเฟรม เช่น เฟรมละประมาณ 10-30 มิลลิวินาที ดังนั้นก็จะสามารถหาพลังงานของเสียงในแต่ละเฟรมได้เป็น

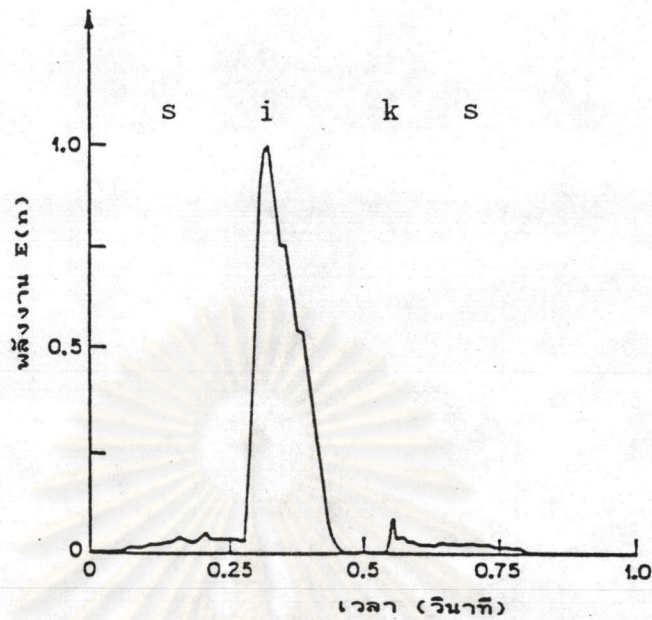
$$E(n) = \sum_{m=0}^{N-1} [w(m)s(n-m)]^2 \quad \dots\dots\dots 2.3.2$$

โดยที่ $w(m)$ คือวินโดว์ฟังก์ชัน (window function) ที่ใช้กำหนดช่วงเวลาในการศึกษาสัญญาณเสียง $s(n)$ ในหนึ่งเฟรม และ N คือจำนวนตัวอย่างของสัญญาณเสียงที่อยู่ภายในเฟรมหรือภายในกรอบของวินโดว์ฟังก์ชัน

การวัดค่าพลังงานดังในสมการที่ 2.3.2 นั้นมีข้อจำกัดตรงที่ว่า มันจะมีความไวต่อสัญญาณเสียงที่มีขนาดใหญ่ เนื่องจากเราใช้วิธียกกำลังสองค่าของสัญญาณอินพุต ดังนั้นการแก้ปัญหาอย่างหนึ่งก็คือวัดพลังงานของเสียงโดยใช้ฟังก์ชันดังนี้

$$E(n) = \sum_{m=0}^{N-1} |w(m)s(n-m)| \quad \dots\dots\dots 2.3.3$$

นั่นคือการใช้การรวมผลบวกของค่าสัมบูรณ์หรือเรียกว่าค่าแมกนิจูด (magnitude) แทนผลรวมของค่ายกกำลังสอง จากรูปที่ 2.3.1 แสดงถึงตัวอย่างรูปฟังก์ชันของพลังงานของคำว่า /six/ ภายใน ช่วงของช่องแคบแบบสี่เหลี่ยมขนาด 10 มิลลิวินาที จะสังเกตได้ว่าเราสามารถเห็นช่วงเริ่มต้นของคำก่อนจะเริ่มเสียง /s/ ซึ่งเป็นเสียงเสียดแทรก และช่วงหยุดก่อนจะเริ่มเสียง /k/ ที่มี ระดับของพลังงานต่ำจนเกือบเป็นศูนย์ได้อย่างชัดเจน



รูปที่ 2.3.1 รูปพลังงานของคำว่า /six/ [siks]
(Schafer and Rabiner, 1973)

2.3.1.2 การวิเคราะห์สัญญาณเสียงในช่วงเวลาสั้นๆ (short-time speech analysis)

เนื่องจากว่าสัญญาณเสียงนั้นเป็นสัญญาณที่แปรตามเวลาที่มีการแปรเปลี่ยนที่ไม่แน่นอน แต่ส่วนใหญ่แล้วผู้พูดจะสามารถทำการควบคุมได้ อย่างเช่นระหว่างเสียงพูดช่วงสั้นๆ รูปร่างของช่องทางเดินเสียงรวมทั้งลักษณะรูปแบบของการกระตุ้นอาจจะมีควมคงที่ได้นานที่สุดประมาณ 200 มิลลิวินาที แต่อย่างไรก็ตามในช่วงที่มีการเปลี่ยนแปลงอย่างรวดเร็ว เช่นสำหรับโน้ตสั้นๆก็อาจจะมีช่วงคงที่นี้ประมาณ 80 มิลลิวินาทีก็ได้ (O'shaughnessy, 1987)

ดังนั้นเทคนิคในการวิเคราะห์สัญญาณส่วนใหญ่แล้ว จะสมมติให้สัญญาณเสียงมีคุณสมบัติที่เปลี่ยนแปลงสัมพันธ์กับเวลาอย่างช้าๆ นั่นก็คือเราจะต้องแบ่งทำการพิจารณาหาพารามิเตอร์ของสัญญาณเสียงพูดในช่วงเวลาสั้นๆเหมือนมองผ่านช่องแคบๆที่เรียกว่า ซ็อต-โวม์ วินโดว์ (short-time window) เมื่อเทียบตามเวลาที่เสียงจะมีการเปลี่ยนแปลงได้เพื่อจะได้มองเห็นเหมือนกับว่าเราหาพารามิเตอร์นั้นๆได้มาจากสัญญาณเสียงที่อยู่ภายในช่องแคบๆ และมีความเสถียรภายในช่วงเวลานั้นๆ

เทคนิคส่วนใหญ่จะกำหนดให้พารามิเตอร์ได้มาจาก ค่าเฉลี่ยจากการพิจารณาภายในช่วงเวลาแคบๆนั้น สำหรับกรณีที่ต้องพิจารณาพารามิเตอร์ต่างๆที่มีการ

เปลี่ยนแปลง ก็จะทำให้การแบ่งสัญญาณเสียงออกเป็นช่องหลายๆช่องหรืออาจจะเรียกว่า กรอบการวิเคราะห์ (analysis frame) ดังนั้นพารามิเตอร์ต่างๆก็สามารถหาได้ทันทีซึ่งพอที่จะติดตามการเปลี่ยนแปลงที่เกิดขึ้นของสัญญาณเสียงอย่างเช่น สำหรับเสียงสระที่ยาวๆสัญญาณเสียงจะมีการเปลี่ยนแปลงในช่วงเวลาที่ช้าก็อาจจะกำหนดให้ช่องแคบมีขนาดใหญ่ประมาณ 100 มิลลิวินาทีก็ได้ แต่ในทางตรงกันข้ามถ้าในกรณีเสียงสั้นๆก็อาจจะต้องใช้ช่องแคบที่มีขนาดเล็กมากๆประมาณ 5-10 มิลลิวินาทีแทนเพื่อป้องกันการสูญหายในรายละเอียดของการต่อเนื่องกันของเสียงถัดไป

2.3.1.3 รูปแบบของช่องวิเคราะห์ (windowing)

รูปแบบของช่องวิเคราะห์ที่จะใช้ควรมีลักษณะดังนี้

1) ช่องวิเคราะห์จะต้องสั้นพอ ที่จะทำให้คุณสมบัติของเสียงที่กำลังพิจารณา ไม่มีการเปลี่ยนแปลงอย่างมีนัยสำคัญภายในช่องนั้น

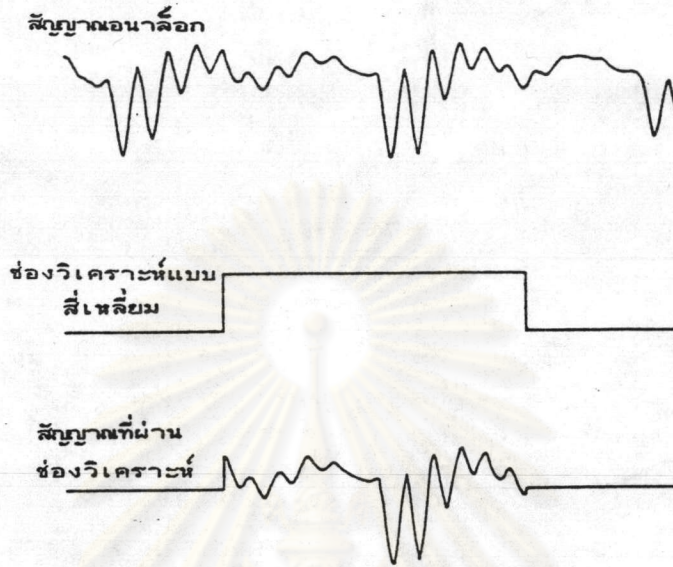
2) ช่องวิเคราะห์จะต้องยาวพอ ที่จะได้ตัวอย่างของเสียงเพื่อนำไปคำนวณหาพารามิเตอร์ตามที่ต้องการ อย่างเช่น ในกรณีที่สัญญาณรบกวนเข้ามาแทรกอยู่บางช่วงในสัญญาณเสียงด้วยถ้าเราเลือกใช้ช่องแคบที่มีขนาดใหญ่กว่าเมื่อทำการหาค่าพารามิเตอร์โดยเฉลี่ยแล้วก็จะทำให้ส่วนประกอบของสัญญาณรบกวนถูกตัดทิ้งหรือมองข้ามไป

3) ช่องวิเคราะห์จะต้องมีความยาวเหมาะสม คือ ไม่สั้นเกินไปกว่าช่วงหนึ่งคาบของสัญญาณเสียงในช่วงที่กำลังพิจารณา เองอน ไชนี้จะมีผลต่อค่าเฟรมเรท (frame rate) หรือจำนวนครั้งต่อวินาทีที่ทำการวิเคราะห์สัญญาณเสียงซึ่งเกิดขึ้นจากการขยับช่องวิเคราะห์ไปเป็นคาบๆตามแกนเวลา มากกว่าขนาดของช่องวิเคราะห์ ตามปกติเฟรมเรทจะมีค่าประมาณ 2 เท่าของส่วนกลับของขนาดช่องวิเคราะห์ นั่นก็คือช่องวิเคราะห์ที่ตัดๆกันไปจะมีการซ้อนทับกัน 50 เปอร์เซ็นต์

การเอาฟังก์ชันของช่องวิเคราะห์ที่มีช่วงขนาดจำกัด $w(n)$ มาคูณเข้ากับสัญญาณ $s(n)$ จะทำให้ได้กลุ่มตัวอย่างของเสียงพูดที่ถูกกำหนดน้ำหนักให้แปรไปตามรูปร่างของช่องวิเคราะห์ รูปแบบของช่องวิเคราะห์ที่ง่ายที่สุดก็คือ ช่องวิเคราะห์แบบสี่เหลี่ยม (rectangular shape window) ซึ่งมีนิยามดังนี้

$$w(n) = \begin{cases} 1 & \text{for } 0 \leq n \leq N-1 \\ 0 & \text{otherwise} \end{cases} \dots\dots\dots 2.3.4$$

ในสมการ 2.3.4 นี้คือการกำหนดช่วงของการวิเคราะห์ให้อยู่ภายในตัวอย่าง N ตัวอย่างที่ติดๆกันไป



รูปที่ 2.3.2 ช่องวิเคราะห์แบบเลื่อน

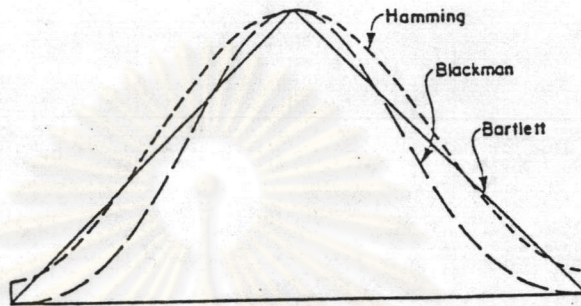
(Witten, 1982)

ในการประยุกต์ส่วนมากจะใช้ช่องวิเคราะห์ที่มีช่วงกว้างกว่าช่วงของข้อมูลที่คงที่ และเน้นเฉพาะช่วงตรงกลางของช่องวิเคราะห์ให้เป็นส่วนของข้อมูลที่จะทำการพิจารณา อย่างเช่นข้อมูลสัญญาณเสียงมีการคงที่ในช่วงเวลา 10 มิลลิวินาทีที่อาจจะใช้ช่องวิเคราะห์ขนาด 20 มิลลิวินาที โดยในช่วงกึ่งกลางช่องวิเคราะห์ขนาด 10 มิลลิวินาทีจะมีการเน้นเพิ่มน้ำหนักให้มากกว่าช่วง 5 มิลลิวินาทีที่ส่วนปลายทั้งสองข้างของช่อง

รูปแบบของช่องวิเคราะห์นี้ มีฟังก์ชันหลายลักษณะด้วยกันตัวอย่างเช่น Blackman , Barlett , Hamming เป็นต้น โดยฟังก์ชันของช่องวิเคราะห์ที่นิยมใช้กันมากในการวิเคราะห์สัญญาณเสียงก็คือ ฟังก์ชันของ Hamming ซึ่งมีรูปร่างตามลักษณะของโคไซน์พลัส (cosine pulse) มีนิยามดังนี้

$$w(n) = \begin{cases} 0.54 - 0.46 \cos (2\pi n)/(N-1) & \text{for } 0 \leq n \leq N-1 \\ 0 & \text{otherwise} \end{cases} \dots\dots\dots 2.3.5$$

ส่วนที่เรียวลงไปตามขอบของช่องวิเคราะห์ มีผลทำให้เกิดการเลื้อนออกไปแบบเป็นคาบของเฟรมวิเคราะห์ตามสัญญาณอินพุท แต่ไม่มีผลกระทบต่อพารามิเตอร์ของเสียงที่เนื่องจากขอบเขตของคาบของพิทช์ (pitch periodic boundaries) หรือการเปลี่ยนแปลงอื่นๆที่มีลักษณะเดียวกันนี้ ในสัญญาณเสียง



รูปที่ 2.3.3 ลักษณะของช่องวิเคราะห์บางชนิด (Witten, 1982)

2.3.1.4 การวิเคราะห์ออโตคอร์รีเลชันในช่วงเวลาสั้นๆ

(short-time autocorrelation analysis)

ออโตคอร์รีเลชันของสัญญาณเสียงคือค่าของ สเปกตรอลแมกนิจูด (spectral magnitude) ยกกำลังสอง ซึ่งมีความสัมพันธ์กับส่วนกลับของการแปลงฟูเรียร์ (Fourier Transform) ของพลังงานในรูปของสเปกตรัม ออโตคอร์รีเลชันจะให้ข้อมูลเกี่ยวกับฮาร์โมนิคของสัญญาณเสียง แอมพลิจูดของฟอร์แมนท์ ตลอดจนแสดงถึงความเป็นคาบของสัญญาณโดยไม่เกี่ยวข้องกับเรื่องของเฟส ดังนั้นจึงสามารถจะนำไปประยุกต์ใช้ในด้านการศึกษาของพิทช์ การดูว่าสัญญาณเสียงช่วงนั้นเป็นเสียงก้องหรือไม่ และการประยุกต์ใช้ในเทคนิคแอลพีซี ดังที่จะกล่าวต่อไป ออโตคอร์รีเลชันฟังก์ชันนิยาม ได้ดังนี้

$$\Phi(m) = \frac{1}{N} \sum_{n=0}^{N-m} s(n)s(n+m) \quad 0 \leq m \leq M_0-1 \quad \dots\dots\dots 2.3.6$$

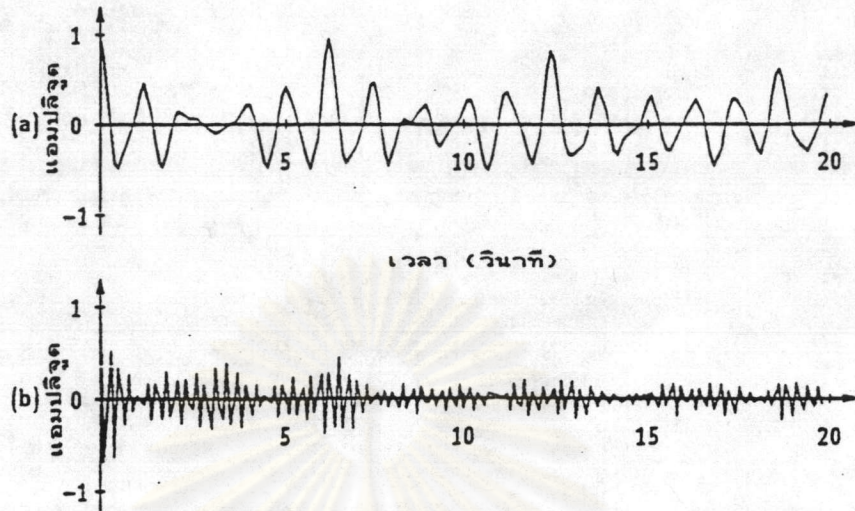
นั่นคือการรวมผลของการคูณสัญญาณตัวอย่างเข้ากับสัญญาณตัวอย่าง
ตัวก่อนหน้า ภายในจำนวนตัวอย่างที่จำกัดจาก 0 ถึง $N-m$ และแปลงให้เป็นค่าบรรทัดฐาน
(normalized) โดยการหารด้วยจำนวนตัวอย่าง N โดยที่ M_0 เป็นค่าของการ lag ที่มากที่สุด
ที่สุดที่จะให้มิได้สำหรับตัวอย่างที่กำลังพิจารณา ตัวอย่างเช่นถ้าต้องการใช้กับสัญญาณที่เป็นคาบ P
ค่า M_0 ก็ควรที่จะมากกว่าค่า P แต่ทั้งนี้ค่า m จะมีค่าอยู่ในช่วงเท่าไรนั้นก็ขึ้นอยู่กับนำไป
ประยุกต์ใช้ในด้านต่างๆด้วย เช่นในกรณีของเทคนิคแอลพีซี ค่า m จะอยู่ในช่วง 0 ถึง 16 หรือ
ตามจำนวนอันดับ (order) ของสมการตั้งที่จะกล่าวในเรื่องของแอลพีซี ส่วนในการประยุกต์
ทางด้านความถี่พื้นฐานนั้นก็อาจจะใช้ค่า m เป็นค่าประมาณเท่ากับค่าคาบของพิชท์ที่คิดเป็นจำนวน
ครั้งของการตัดแต่ละตัวอย่าง แต่ถ้าในกรณีสัญญาณไม่เป็นคาบก็อาจจะเริ่มจากคาบที่สั้นที่สุดที่
เป็นไปได้คือ 3 มิลลิวินาที (ในกรณีเสียงของผู้หญิง) จนถึงคาบที่ยาวที่สุดก็คือ 20 มิลลิวินาที
(ในกรณีเสียงของผู้ชาย) นั่นก็คือในกรณีที่คิดด้วยความถี่ 10 kHz หรือคิดทุกๆ 0.1 มิลลิวินาที
จะต้องคำนวณค่าออคโตคอร์เรเลชันประมาณ 170 ครั้งขึ้นไปในแต่ละเฟรมดังนี้ เป็นต้น
(O'shaughnessy, 1987) จากสมการ 2.3.6 นั้นเราจะสามารถพิสูจน์ได้ว่า

$$\Phi(m) = \Phi(-m) \quad \dots\dots\dots 2.3.7$$

และฟังก์ชันออคโตคอร์เรเลชันจะมีค่าสูงสุดเมื่อ $m=0$ และค่า $\Phi(0)$
นี้จะมีค่าเท่ากับค่าพลังงานของสัญญาณ (หรือค่าพลังงานเฉลี่ยสำหรับสัญญาณแบบสุ่มหรือสัญญาณที่เป็น
คาบ) ในกรณีที่สัญญาณมีลักษณะเป็นคาบ (periodic signal) ที่มีคาบเท่ากับจำนวน P ตัว
อย่างแล้วจะได้ว่าฟังก์ชันออคโตคอร์เรเลชันของสัญญาณนั้นก็จะมีคาบเท่ากับ P ด้วย

$$\Phi(m) = \Phi(m+P) \quad \dots\dots\dots 2.3.8$$

หรืออาจจะพูดอีกอย่างได้ว่าค่าสูงสุดของ $\Phi(m)$ จะเกิดขึ้นเมื่อ
 $m = 0, \pm P, \pm 2P, \pm 3P, \dots\dots\dots$



รูปที่ 2.3.4 ตัวอย่างแสดงฟังก์ชันออโตคอร์เรเลชันของ

a) เสียงก้อง b) เสียงไม่ก้อง (O'shaughnessy, 1987)

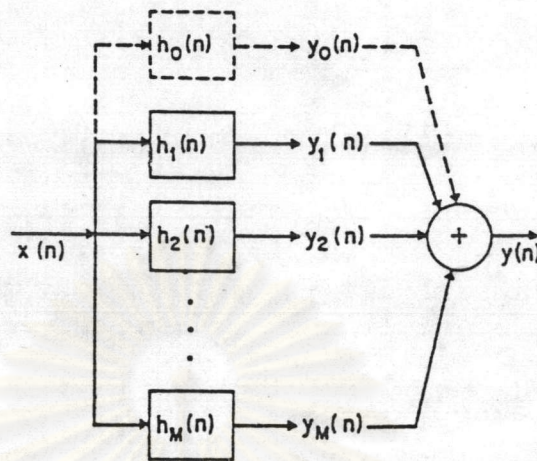
2.3.2 การวิเคราะห์ทางสเปกตรัม

การวิเคราะห์ทางสเปกตรัมของสัญญาณเสียง ก็คือการวิเคราะห์หรือพิจารณา สัญญาณเสียงในด้านของ โดเมนความถี่นั่นเอง เหตุที่ต้องมีการวิเคราะห์ทางด้านสเปกตรัมนั้นก็ เนื่องจากพารามิเตอร์บางตัวในการประมวลผลสัญญาณเสียงนั้น ส่วนใหญ่จะสามารถหาได้จาก โดเมนความถี่ และสัญญาณเสียงที่สร้างออกมาจากการเปลี่ยนแปลงรูปร่างของช่องทางเดินเสียง ของคนนั้นก็จะสามารถพิจารณาเป็นรูปแบบในโดเมนของความถี่จะง่ายและคงที่กว่าการพิจารณาใน โดเมนเวลาเป็นต้น

การวิเคราะห์ทางสเปกตรัมในช่วงเวลาสั้นๆ ก็เป็นวิธีการดั้งเดิมอันหนึ่งของการประมวลผลสัญญาณเสียงที่สำคัญ ด้วยข้อสมมติฐานที่ว่าสัญญาณเสียงในช่วงเวลายาวๆจะมีการเปลี่ยนแปลงอยู่ตลอดเวลาไม่มีสภาพของการคงที่ แต่ถ้าพิจารณาในช่วงเวลาสั้นๆก็อาจจะ สมมติว่าเป็นสัญญาณที่คงที่ได้ ดังนั้นการทำการแปลงฟูเรียร์ของสัญญาณเสียงช่วงสั้นๆช่วงหนึ่งจะ สามารถแทนเป็นค่าสเปกตรัมของสัญญาณเสียงในช่วงเวลาดังกล่าวได้เป็นอย่างดี การพิจารณา ดูสเปกตรัมของสัญญาณเสียงในช่วงเวลาอันสั้นนั้น จะเป็นพื้นฐานของเทคนิคการวิเคราะห์เสียง ต่างๆเช่น เทคนิค channel vocoder , phase vocoder , spectrogram เป็นต้น

เทคนิคทางด้านสเปกตรัมที่จะนำมาอ้างอิงในที่นี้จะมีดังนี้

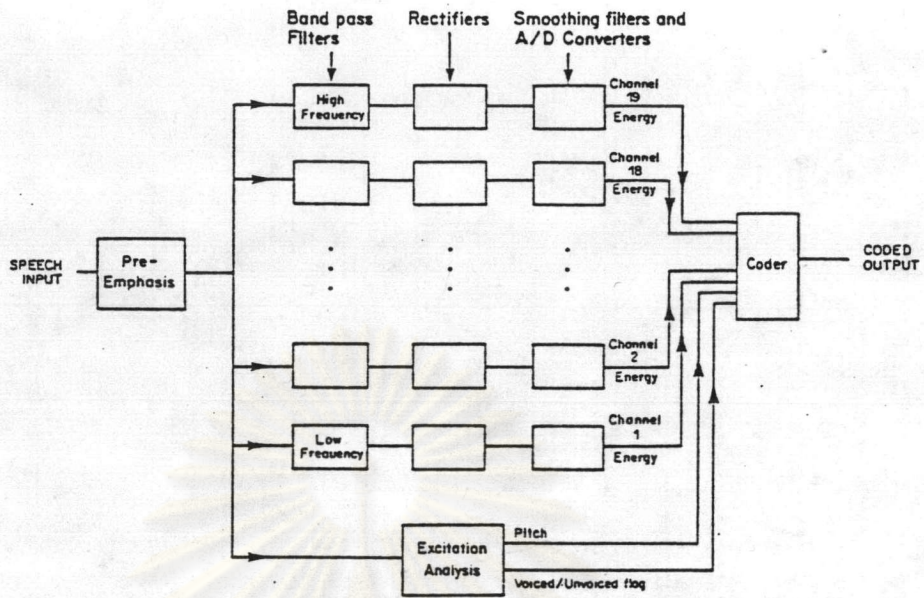
2.3.2.1 การวิเคราะห์ฟิลเตอร์แบงก์ (Filter bank analysis)



รูปที่ 2.3.5 การวิเคราะห์ฟิลเตอร์แบงก์
(Schafer and Rabiner, 1973)

เป็นเทคนิคทางด้าน การวิเคราะห์สัญญาณเสียง โดยใช้สเปกตรัมที่นิยมกันมากอันหนึ่ง เนื่องจากสามารถทำได้ง่ายแล้ว ได้ผลออกมาทันทีที่ทันใจและไม่สิ้นเปลืองมากนัก โดยการนำใช้กลุ่มของตัวกรองแบบเฉพาะย่านความถี่ผ่าน (bandpass filter) ซึ่งอาจจะเป็นได้ทั้งแบบอนาล็อกหรือแบบดิจิทัลก็ได้ ตัวกรองแต่ละอันจะทำการกรองเอาความถี่เฉพาะที่ต่างช่วงกันมาทำการวิเคราะห์ ทั้งนี้ความถี่ทั้งหมดตลอดย่านจะต้องครอบคลุมแถบความถี่ที่สำคัญทั้งหมดของสัญญาณเสียงอินพุท เทคนิคอันนี้เป็นเทคนิคพื้นฐานของระบบการติดต่อสื่อสารข้อมูล เช่น เทคนิคของแชนเนลโวกโคดเดอร์ (channel vocoder) หรือ เฟสโวกโคดเดอร์ (phase vocoder) เป็นต้น (Gold and Rader, 1967)

จุฬาลงกรณ์มหาวิทยาลัย



รูปที่ 2.3.6 แชลแนลโวโคเดอร์
(Holmes, 1980)

2.3.2.2 การวิเคราะห์ฟูเรียร์ในช่วงเวลาสั้นๆ

(Short-time Fourier Analysis)

เป็นวิธีการดั้งเดิมของการวิเคราะห์สัญญาณทางด้านสเปกตรัม

การวิเคราะห์แบบฟูเรียร์จะสามารถแทนค่าสัญญาณเสียงออกมาได้ในเทอมของ แอมพลิจูดและเฟส ซึ่งก็เป็นฟังก์ชันอันดับหนึ่งของความถี่ ถ้าเรามองช่องทางเดินเสียงของคนเป็นระบบเชิงเส้นระบบหนึ่ง การแปลงแบบฟูเรียร์ของสัญญาณเสียงก็คือผลลัพธ์ที่เกิดจากการแปลงของการกระตุ้นจากการสั่นของเส้นเสียงกับการตอบสนองภายในช่องทางเดินเสียงนั่นเอง นิยามของการแปลงฟูเรียร์ของสัญญาณ $g(t)$ ใดๆเป็นดังนี้

$$g(t) = \int_{-\infty}^{\infty} G(f)e^{+j2\pi ft} df \quad \dots\dots\dots 2.3.9$$

$$G(f) = \int_{-\infty}^{\infty} g(t)e^{-j2\pi ft} dt \quad \dots\dots\dots 2.3.10$$

เป็นการแปลงระหว่างโดเมนเวลาที่ต่อเนื่อง กับโดเมนความถี่

ที่ต่อเนื่อง จะสังเกตได้ว่าในบางครั้งจะมีการทำให้เป็นบรรทัดฐานกับสมการทั้งสองโดยการคูณค่า $1/2\pi$ เข้าไปซึ่งจำเป็นในกรณีที่ตัวแปรความถี่มีหน่วยเป็น radians/s แต่ในรูปแบบในที่นี้เราจะใช้หน่วยเป็น Hz แทน

ฟูเรียร์ซีรีส์ (Fourier Series) คือการปฏิบัติการกับรูปคลื่นที่มีลักษณะเป็นคาบตามแกนเวลา ถ้ารูปคลื่นมีคาบอยู่ในช่วง $[0, b)$ จะได้ฟังก์ชันการแปลงที่มีนิยามดังนี้

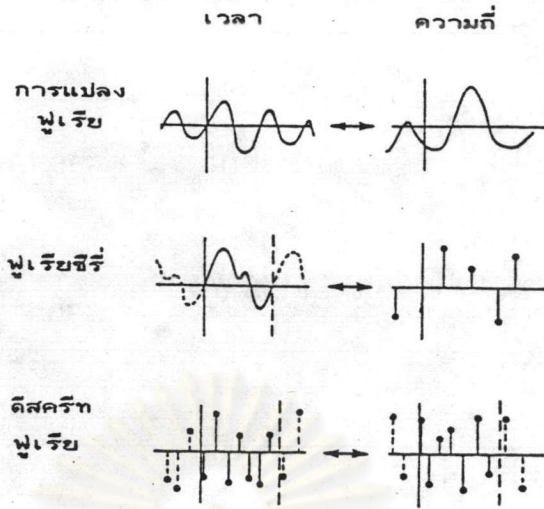
$$g(t) = \sum_{r=-\infty}^{\infty} G(r)e^{+j2\pi r t/b} \dots\dots\dots 2.3.11$$

$$G(r) = \frac{1}{b} \int_0^b g(t)e^{-j2\pi r t/b} dt \dots\dots\dots 2.3.12$$

ฟูเรียร์ซีรีส์เป็นการแปลงระหว่างฟังก์ชันที่เป็นคาบในโดเมนเวลากับฟังก์ชันแบบไม่ต่อเนื่องในโดเมนความถี่ ถ้าเราทำการแปลงรูปคลื่นในโดเมนเวลาที่มีลักษณะเป็นฟังก์ชันแบบไม่ต่อเนื่องและมีช่วงที่จำกัด เราจะได้รูปคลื่นในโดเมนของความถี่ซึ่งมีช่วงที่จำกัดมีลักษณะเป็นคาบและเป็นฟังก์ชันที่ไม่ต่อเนื่อง ลักษณะการแปลงแบบนี้ก็คือ การแปลงแบบดิสครีตฟูเรียร์ (Discrete Fourier Transform) ซึ่งก็คือการทำฟังก์ชันที่เป็นคาบแบบไม่ต่อเนื่องในโดเมนเวลาไปเป็นฟังก์ชันที่เป็นคาบและไม่ต่อเนื่องในโดเมนความถี่นั่นเอง โดยมีนิยามกำหนดไว้ดังนี้

$$g(n) = \frac{1}{N} \sum_{r=0}^{N-1} G(r)e^{+j2\pi r n/N} \dots\dots\dots 2.3.13$$

$$G(r) = \sum_{n=0}^{N-1} g(n)e^{-j2\pi r n/N} \dots\dots\dots 2.3.14$$



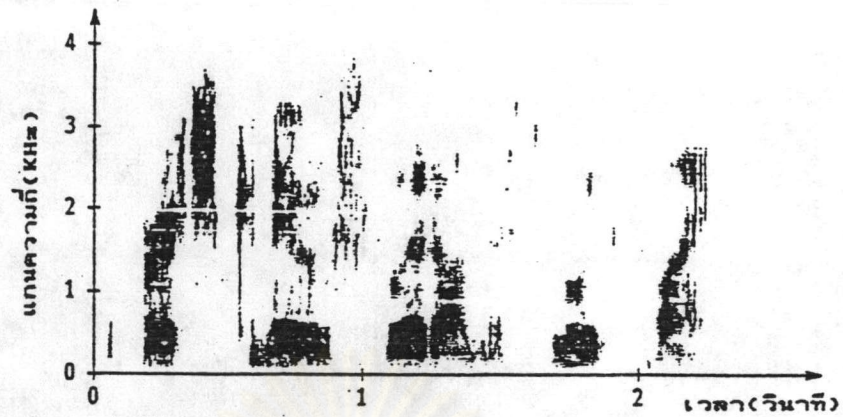
รูปที่ 2.3.7 รูปแสดง การแปลงฟูรีเย ฟูรีเยซีรี และการแปลงดีสครีทฟูรีเย

2.3.2.3 การแสดงสเปคตรัมของเสียง

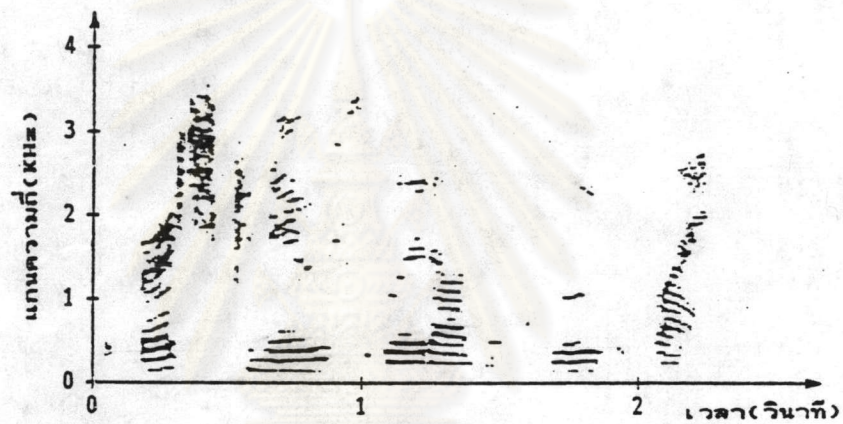
อุปกรณ์ที่สำคัญในการวิเคราะห์สัญญาณเสียงซึ่งมีขึ้นเมื่อประมาณ 40 ปีที่แล้วก็คือ สเปคโตรแกรม (spectrogram) หรือ ชาร์นสเปคโตรกราฟ (sound spectrograph) ซึ่งจะแสดงให้เห็นถึงลักษณะเฉพาะของสัญญาณเสียงในรูปสามมิติภายในช่วงเวลาสั้นๆประมาณ 2-3 วินาที โดยการใช้การแปลงฟูรีเยในช่วงเวลาสั้นๆของสัญญาณเสียงมาพล็อตรูปภาพโดยมีแกนอนเป็นแกนของเวลา และแกนตั้งเป็นแกนของความถี่ (f จาก 0 ถึง T) และแสดงค่าของแมกนิจูดด้วยความเข้มของการพิมพ์ออกมาดังรูปที่ 2.3.8

สเปคโตรแกรมแบบช่วงการกรองกว้าง (wideband spectrogram) จะแสดงให้เห็นถึงคาบของพิทช์ออกมาในแนวตั้ง ซึ่งจะสัมพันธ์กับช่วงที่แอมพลิจูดของเสียงมีค่ามากที่สุดขณะที่เส้นเสียงกำลังจะปิด ดังนั้นในกรณีของเสียงก้อง เราสามารถที่จะตรวจพบได้โดยง่ายจากการค้นหาเส้นในแนวตั้งที่มีลักษณะเป็นคาบนี้ การวิเคราะห์ความถี่ฟอร์แมนท์และแอมพลิจูดของฮาร์โมนิกภายใต้ฟอร์แมนท์แต่ละฟอร์แมนท์จะอยู่ในช่วงแถบความถี่ประมาณ 300 Hz แสดงให้เห็นเป็นแถบสีดำครอบคลุมในแต่ละฟอร์แมนท์ทางแนวตั้ง ความถี่ตรงกึ่งกลางของแต่ละแถบสีดำก็คือความถี่ฟอร์แมนท์ โดยประมาณนั่นเอง

สเปคโตรแกรมแบบช่วงการกรองแคบ (narrowband spectrogram) จะแสดงให้เห็นถึงฮาร์โมนิกแต่ละอันแทนคาบของพิทช์ และมีประโยชน์ค่อนข้างน้อยในการวิเคราะห์สัญญาณเสียงออกเป็นส่วนๆเนื่องจากให้รายละเอียดในด้านความถี่น้อยมาก แบบนี้จะใช้วิเคราะห์ความถี่ในช่วง 45 Hz



The birch canoe slid on the smooth planks.



รูปที่ 2.3.8 ภาพของสเปกโตรแกรม a) ไวร์แบนด์ b) นาโรแบนด์
(O'shaughnessy, 1987)

เนื่องจากแอมพลิจูดของเสียงจะมีค่าลดลงประมาณ -6 dB ต่อออกเตฟ (octave) เมื่อสัญญาณเสียงออกมาจากริมฝีปาก และรายละเอียดด้านความถี่สูงจะปรากฏขึ้นมารบกวนเพิ่มขึ้นตาม ดังนั้นจึงมีการทำพรี-เอมฟาไซซิง (pre-emphasizing) กับสัญญาณเสียงโดยการทำดิฟเฟอเรนเชียลสัญญาณเสียงที่เป็นอนาล็อก $S_u(t)$ ก่อนผ่านเข้าไป A/D เพื่อแปลงเป็นสัญญาณดิจิทัลอีกทีหนึ่ง หรือ โดยการดิฟเฟอเรนเชียลสัญญาณที่ไม่ต่อเนื่องทางด้านเวลา $S(n) = S_u(nT)$ เพื่อเป็นการชดเชยลดสัญญาณความถี่สูงลงไป ถ้าสัญญาณเสียงที่มีการสังเคราะห์เสียงขึ้นมาจากข้อมูลที่ผ่านการพรี-เอมฟาไซซิงมาแล้วนั้น ในขั้นตอนสุดท้ายของการสังเคราะห์เสียงจะต้องทำการในทำนองกลับกันคือทำดี-เอมฟาไซซิง (de-emphasizing) โดยทำอินทิเกรชันกลับคืนด้วยเพื่อให้ได้คุณสมบัติทางไดนามิกเรจกลับคืนมา ในการประมวลผลทางดิจิทัลสามารถทำการพรี-เอมฟาไซซิงด้วยสมการ

$$y(n) = s(n) - as(n-1) \quad \dots\dots\dots 2.3.15$$

โดยค่า a จะต้องมีค่าระหว่าง 0.9 ถึง 1 ตามปกติแล้วการทำ
 พรี-เอมฟาไซซิงนี้จะใช้เฉพาะกับเสียงก้องเท่านั้น ทั้งนี้เนื่องจากเสียงไม่ก้องจะมีค่าสเปคตรัม
 ราบเรียบที่ความถี่สูงๆอยู่แล้วดังนั้นค่า a จะมีค่าเข้าใกล้ 0 ในกรณีเสียงไม่ก้องซึ่งเราสามารถ
 จะทำให้ค่า a มีผลลัพธ์ออกมาใกล้เคียงตามที่กำหนดนี้ได้โดยการกำหนดให้ค่า a มีค่าเกี่ยวข้องกับ
 กับค่าออคโตคอร์เรเลชันของสัญญาณเสียงดังนี้

$$a = R(1) / R(0) \quad \dots\dots\dots 2.3.16$$

โดยค่า $R(1)$ คือค่าคอร์เรเลชันของสัญญาณเสียงซึ่งมีการหน่วง
 เวลาไปหนึ่งครั้งของตัวอย่าง ค่า $R(0)$ คือค่าคอร์เรเลชันของสัญญาณซึ่งไม่มีการหน่วงเวลา
 ซึ่งในกรณีของเสียงก้องแล้วการหาค่าคอร์เรเลชันของสัญญาณจะทำให้ได้ค่า $R(1)$ มีค่าใกล้เคียงกับ
 ค่า $R(0)$ มากดังนั้นก็จะได้ค่า a เข้าใกล้กับ 1 ส่วนในกรณีของเสียงไม่ก้องนั้นค่าคอร์เรเลชัน
 ของสัญญาณตัวอย่างหนึ่งกับตัวอย่างถัดไปจะมีค่าน้อยหรือไม่มีเลย ดังนั้นอัตราส่วนของ $R(1)$ กับ
 $R(0)$ จะมีค่าเข้าใกล้ศูนย์

2.3.3 การหาคาบของพิทช์ (Pitch detection)

อัลกอริทึมในการหาคาบของสัญญาณเสียงมีอยู่หลายวิธีด้วยกัน เริ่มตั้งแต่
 ประมาณปี ค.ศ 1976 ที่มีชื่อเสียงก็เช่นวิธีการหาพิทช์ของ Sondhi และอัลกอริทึม SIFT
 (Simplified Inverse Filter Tracking) ของ Markel ในปี ค.ศ. 1972 เป็นต้น
 แต่ละวิธีการก็จะออกแบบเฉพาะตามความสามารถของฮาร์ดแวร์ของตนเพื่อทำให้เกิดการประมวล
 ผลได้แบบทันทีทันใด ทั้งนี้เนื่องจากว่าอัลกอริทึมในการหาพิทช์นี้จะต้องใช้หลักการทางตรรกหลาย
 ขั้นตอน ทำให้ต้องเสียเวลาในการคำนวณมากถึงแม้ว่าการประยุกต์ในบางรูปแบบของอัลกอริทึม
 นี้จะสามารถประยุกต์โดยใช้โปรแกรมทั้งหมด แต่ก็จะต้องมีการใช้ตัวประมวลผลพิเศษมาช่วย
 เพื่อผลในการตอบสนองให้เป็นที่พอใจอย่างทันทีทันใด

ปัญหาที่สำคัญประการหนึ่งในการหาความถูกต้องของพิทช์ของเสียงในช่วงแถบ
 ความถี่ที่กว้างของสัญญาณก็คือ ผลของโครงสร้างฟอร์แมนท์ (formant structure) ที่มีการ
 วัดสัมพันธ์กับคาบของรูปคลื่นสัญญาณ ดังนั้นเพื่อเพิ่มความถูกต้องในการหาพิทช์ของเสียงจะต้องทำ
 การลดผลของฟอร์แมนท์นี้ลงให้มากที่สุดเท่าที่จะทำได้ เทคนิคในการจัดรูปร่างของสเปคตรัม

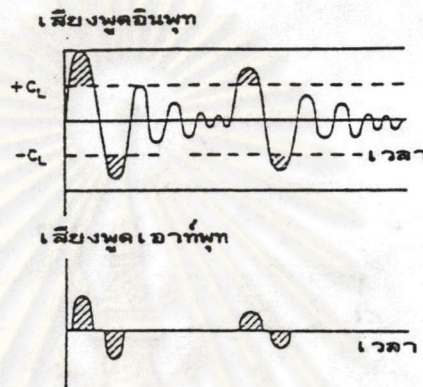
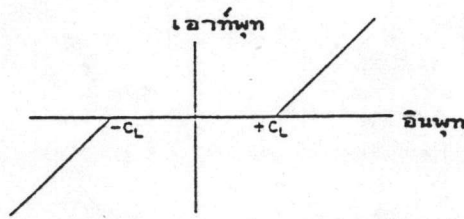
ในรูปคลื่นที่เกี่ยวข้องกับฟอร์แมนที่นี้ เรียกว่า การทำสเปกตรัมให้ราบเรียบ (spectral flattening) Sondhi ได้เสนอวิธีการทำสเปกตรัมให้ราบเรียบอยู่สองวิธีคือ

- วิธีฟิลเตอร์แบงค์ (filter bank method) วิธีการนี้คือการนำสัญญาณเสียงไปผ่านกลุ่มของตัวกรองสัญญาณแบบแถบความถี่ผ่าน (band pass filter) ที่มีแถบของความถี่แตกต่างกันไปตลอดย่านความถี่ของสัญญาณเสียง สัญญาณสุดท้ายที่ผ่านออกมาจากตัวกรองแต่ละตัวจะถูกทำให้เป็นบรรทัดฐานกับค่าแอมพลิจูดให้เป็น 1 หรือเป็นการทำสเปกตรัมให้ราบเรียบโดยการนำมารวมด้วยค่ากำลังงานในช่วงเวลาสั้นๆ สัญญาณรวมทั้งหมดที่มีการทำให้สเปกตรัมราบเรียบจะ ได้มาจาก นำสัญญาณสุดท้ายจากตัวกรองแต่ละตัวมารวมเข้าด้วยการหน่วงเวลาที่เหมาะสม ถึงแม้ว่าวิธีการนี้จะสามารถใช้ได้ดีในหลายๆกรณีก็ตามแต่ก็ยังมีข้อเสียอยู่หลายอย่าง เช่น ต้องใช้ฮาร์ดแวร์ในส่วนที่เป็นวงจรตัวกรองและวงจรปรับระดับสัญญาณแต่ละความถี่ต่างๆเป็นจำนวนมาก และในกรณีที่เกิดมีฮาร์โมนิกที่ไม่มีพิทซ์อยู่ในตัวกรองในแถบความถี่ใดความถี่หนึ่งขึ้นมาจะมีผลทำให้สัญญาณที่ออกมาจากตัวกรองนั้นมีระดับสัญญาณที่ต่ำมาก แต่อย่างไรก็ตามการปรับระดับสัญญาณที่ความถี่นั้นก็ยังเป็นสิ่งที่จำเป็นอยู่ แต่จะมีผลในด้านการจัดสัญญาณรบกวนซึ่งมีขนาดใหญ่เมื่อเทียบกับขนาดของสัญญาณจริงมากกว่าที่จะมีผลในด้านการตรวจจับพิทซ์ของสัญญาณ

- เทคนิคเซ็นเตอร์คลิปปิง (center clipping technique) วิธีการนี้คือการตรวจจับระดับของสัญญาณว่า ถ้ามีค่าต่ำกว่าระดับคลิปปิงที่กำหนดไว้เมื่อไรก็จะกำหนดให้สัญญาณตรงนั้นมีค่าเป็นศูนย์ ส่วนสัญญาณที่ยังมีค่าสูงกว่าระดับคลิปปิงก็จะเอาค่ามาหักลบด้วยค่าระดับคลิปปิงเหลือแต่ส่วนที่เกินไปจากระดับคลิปปิงเท่านั้น รูปที่ 2.3.9 แสดงถึงลักษณะสมบัติของสัญญาณอินพุท-เอาต์พุทที่นำมาทำกระบวนการเช่นเซ็นเตอร์คลิปปิง ซึ่งจะเห็นได้ว่าถ้ามีการกำหนดระดับคลิปปิงที่เหมาะสมแล้ว โครงสร้างของรูปคลื่นที่เป็นส่วนของฟอร์แมนต่างๆจะถูกกำจัดออกไปได้เกือบทั้งหมด ซึ่งจะมีผลทำให้สัญญาณที่มีสเปกตรัมราบเรียบที่ได้ออกมามองเห็นคาบของมันได้ง่ายกว่าสัญญาณที่ยังไม่ได้ผ่านกระบวนการทำให้สเปกตรัมราบเรียบ

วิธีการทำให้สเปกตรัมของสัญญาณราบเรียบทั้งสองวิธีข้างต้นนั้น วิธีการเซ็นเตอร์คลิปปิงจะเป็นวิธีที่มีการนำมาประยุกต์ใช้ได้ค่อนข้างจะสะดวกกว่า โดยการนำเอาฮาร์ดแวร์มาประยุกต์ใช้แทนอัลกอริธึมบางส่วนด้วย เพื่อให้ผลของการตอบสนองเป็นไปอย่างทันทีทันใด (Dubnowski, 1976)

อัลกอริธึมของการหาพิทซ์ของสัญญาณเสียงนี้ จะเริ่มจากการนำสัญญาณเสียงผ่านเข้าตัวกรองความถี่แบบต่ำผ่านที่มีช่วงกว้างของแถบความถี่อยู่ในช่วง 900 เฮิรท์ จากนั้นจะนำมาแปลงให้เป็นสัญญาณแบบดิจิตอลแล้วตัดแบ่งออกเป็นเฟรมๆ เฟรมละ 30 มิลลิวินาทีเพื่อนำไปวิเคราะห์และประมวลผลต่อไป

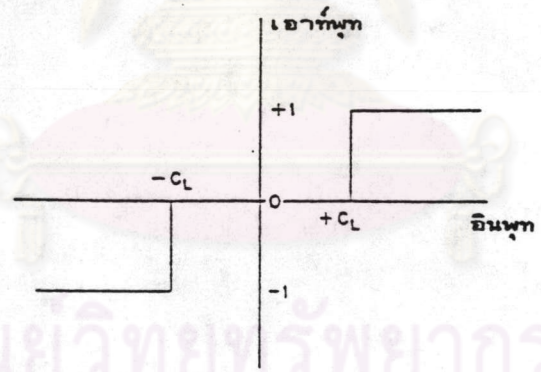


รูปที่ 2.3.9 ลักษณะสมบัติของสัญญาณเมื่อผ่านกระบวนการเช่นเตอร์คลิปปิง (Sondhi, 1968)

ขั้นแรกของการประมวลผลสำหรับแต่ละเฟรมนั้นก็จะทำการคำนวณระดับคลิปปิงของแต่ละเฟรม เนื่องจากเสียงเป็นสัญญาณที่มีช่วงของการเปลี่ยนแปลงกว้างมากดังนั้นการเลือกระดับคลิปปิงจะต้องทำด้วยความระมัดระวัง เพื่อป้องกันการสูญหายของข้อมูลที่สำคัญบางส่วนของรูปคลื่น ในขณะที่รูปคลื่นมีแอมพลิจูดที่กำลังมีขนาดเพิ่มขึ้น หรือกำลังลดลงในระหว่างแต่ละเฟรมนั้นๆ กรณีแบบนี้จะเกิดขึ้นในช่วงที่เสียงกำลังจะเริ่มตันหรือกำลังจะจบลง หรืออาจจะเป็นตรงรอยต่อของช่วงการเปลี่ยนแปลงของเสียง สมมติว่าเราเลือกให้ระดับคลิปปิงมีค่าเป็น c_L สัญญาณเสียงแต่ละเฟรมที่มีขนาด 30 มิลลิวินาทีจะถูกแบ่งออกเป็น 3 ส่วนๆ ละ 10 มิลลิวินาที ในส่วนของ 10 มิลลิวินาทีแรกกับส่วนของ 10 มิลลิวินาทีสุดท้ายอัลกอริทึมจะทำการหาค่าสมบรูณ์ที่จุดยอดสูงสุดเก็บเอาไว้สำหรับของแต่ละส่วน ค่าระดับคลิปปิงก็จะกำหนดเอาจากค่าเปอร์เซ็นต์ที่แน่นอนของค่าสมบรูณ์ที่จุดยอดสูงสุดที่น้อยที่สุดระหว่างสองค่านั้น ตามวิธีการของ Sondhi นั้นก็จะกำหนดค่าระดับคลิปปิงไว้ที่ 30 เปอร์เซ็นต์ของค่าสมบรูณ์ที่จุดยอด ซึ่งจะใช้เป็นระดับอ้างอิงกับข้อมูลของเสียงตลอดช่วง 30 มิลลิวินาทีเลย แต่ทั้งนี้เพื่อเป็นการหลีกเลี่ยงการสูญหายของข้อมูลของเสียงที่มีระดับสัญญาณต่ำๆ ก็จำเป็นต้องใช้ระดับคลิปปิงให้ต่ำตามไปด้วย นั่นหมายถึง

ว่า การที่จะเลือกระดับคลิปปิงให้เหมาะสมกับระดับของสัญญาณเสียงเพื่อหลีกเลี่ยงปัญหาเหล่านี้ เป็นเรื่องที่ไม่ง่ายนัก ไม่สามารถกำหนดตายตัวเป็นกฎเกณฑ์ที่ลงตัวได้

หลังจากการหาระดับคลิปปิงได้แล้ว จะนำสัญญาณมาตัดช่วงที่อยู่ตรงกลาง ระหว่างระดับคลิปปิงด้านบนและด้านล่างทิ้งไป ซึ่งลักษณะนี้จึงเรียกว่า เซ็นเตอร์คลิปปิง (center clipping) ดังนั้นผลที่ได้ออกมา ก็จะเป็นไปได้ 3 ค่าคือค่า +1 , -1 สำหรับตัวอย่าง สัญญาณที่มีค่าเกินกว่าระดับคลิปปิงด้านบนและต่ำกว่าด้านล่างตามลำดับ และค่า 0 สำหรับสัญญาณ ที่มีค่าอยู่ภายในช่วงระดับคลิปปิงระหว่างด้านบนและด้านล่าง จะเห็นได้ว่าไม่ว่าค่าของสัญญาณ ตัวอย่างจะมีค่ามากหรือน้อยแค่ไหน ก็จะถูกทำให้มีค่าเท่ากับหนึ่งเหมือนกันหมด (ในแต่ละด้านบนหรือล่าง) ถ้าเกินค่าของระดับคลิปปิงแล้ว นั่นคือเป็นการทำการตัดยอดคลื่นที่มีค่ามากๆ (in-finite peak clipped) ด้วย รูปที่ 2.3.10 จะแสดงให้เห็นถึงลักษณะสมบัติของ สัญญาณเสียงขาเข้าและขาออกเมื่อผ่านการเซ็นเตอร์คลิปปิง และการตัดยอดคลื่นที่มีค่ามากๆด้วย จากนั้นก็จะเอาผลมาทำการออโตคอร์เรเลชันต่อไป ผลของการเซ็นเตอร์คลิปปิงและการตัดยอด คลื่นนั้นจะเป็นการลดความยุ่งยากในการคำนวณของชั้นออโตคอร์เรเลชันลงอย่างมาก เพราะไม่ต้อง ทำการคูณหรือบวกในการคำนวณฟังก์ชันออโตคอร์เรเลชันของสัญญาณที่ผ่านการตัดยอดคลื่นมานี้เลย



รูปที่ 2.3.10 ลักษณะสมบัติของเทคนิคเซ็นเตอร์คลิปปิง

ขั้นตอนของการประมวลผลต่อไป โดยการคำนวณออโตคอร์เรเลชันสำหรับสัญญาณ ใน 1 เฟรมหรือ 30 มิลลิวินาทีที่ผ่านการคลิปปิงแล้วนี้ จะสามารถกำหนดได้เป็น

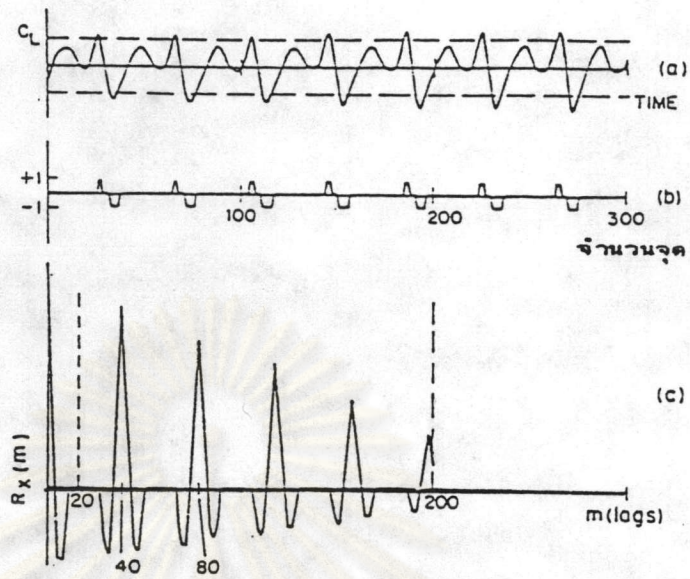
$$R_x(m) = \sum_{n=0}^{299-m} x(n)x(n+m) \quad m=M_1, M_{1+1}, \dots, M_F \quad \dots \dots \dots 2.3.17$$

โดยที่ M_1 คือค่าเริ่มต้นและ M_2 คือค่าสุดท้ายสำหรับการคำนวณฟังก์ชัน
 ออโตคอร์เรเลชัน ซึ่งตามปกติแล้วจะกำหนดให้ $M_1 = 25$ และ $M_2 = 200$ สำหรับเสียงที่มีช่วง
 คาบของเสียงอยู่ภายในช่วง 400 เฮิรตซ์ จนถึงต่ำลงมาถึง 50 เฮิรตซ์ จากสมการ 2.3.17 เทอม
 แต่ละเทอมที่อยู่ในรูปของ $x(n)x(n+m)$ ซึ่งค่าของ $x(n)$ จะเห็นได้เพียงค่า $+1, -1, 0$ เท่า
 นั้น ดังนั้น เราสามารถเขียนเทอมนี้ใหม่ในรูปแบบดังนี้

$$\begin{aligned} x(n)x(m) &= 0 \quad \text{ถ้า } x(n) = 0 \quad \text{หรือถ้า } x(n+m) = 0 \quad \dots\dots\dots 2.3.18 \\ &= 1 \quad \text{ถ้า } x(n) = x(n+m) = \pm 1 \\ &= -1 \quad \text{ถ้า } x(n) = -x(n+m) = \pm 1 \end{aligned}$$

นั่นก็คือเราสามารถให้เหตุผลทางตรรกะมาใช้คำนวณเทอมแต่ละเทอมในฟังก์ชัน
 ออโตคอร์เรเลชันได้ ซึ่งถ้าทำการประยุกต์โดยใช้ฮาร์ดแวร์แล้วก็สามารถจะใช้วงจรนับธรรมชาติ
 แบบง่ายๆ มาใช้ในการหาค่าออโตคอร์เรเลชันก็ได้แล้ว รูปที่ 2.3.11 เป็นตัวอย่างของการ
 ประมวลผลของสัญญาณใน 1 เฟรม รูปบนจะเป็นสัญญาณที่ผ่านวงจรกรองแบบต่ำผ่านมาแล้ว
 และจะทำการเช่นเตอร์คลิปปิงด้วยระดับสัญญาณคลิปปิงเท่ากับ $\pm C_L$ รูปกลางหลังจากการทำ
 เช่นเตอร์คลิปปิงแล้ว และรูปล่างแสดงถึงฟังก์ชันออโตคอร์เรเลชันของสัญญาณเสียงที่ผ่านการคลิปปิง
 แล้ว ช่วงของคาบหรือพิชของเสียงนี้ถ้าดูจากรูปล่างก็จะพบว่ามีความถี่อยู่ในช่วงของแนวเส้นประ
 คือช่วงที่ $m=20$ จนถึง $m=200$ ซึ่งถ้าหาคาบของเสียงในเฟรมนี้ก็จะได้ค่าเท่ากับ 40 ตัวอย่าง
 สัญญาณหรือ 4 มิลลิวินาที หรืออาจจะพูดได้ว่ามีความถี่ของพิชเท่ากับ 250 เฮิรตซ์

เราควรจะระลึกไว้เสมอว่าในการคำนวณฟังก์ชันออโตคอร์เรเลชันจากสมการที่
 2.3.17 นั้นเราสมมติว่า ตัวอย่างของสัญญาณตัวอย่างที่อยู่นอกเหนือไปจาก 30 มิลลิวินาทีที่เรา
 กำลังพิจารณาอยู่นี้จะถือว่าเป็นศูนย์หมด ผลกระทบอันนี้จะเป็นการถ่วงน้ำหนักฟังก์ชันออโตคอร์
 เรเลชันด้วยการลาดเอียงแบบเชิงเส้น (linear taper) โดยเริ่มจากค่าเป็น 1 ที่จุด $m=0$
 ไปจนถึงค่าเป็น 0 ที่จุด $m=300$ ผลกระทบอันนี้จะเห็นได้จากรูปที่ 2.3.11 ล่าง ซึ่งจะเห็น
 ว่าจุดยอดของฟังก์ชันออโตคอร์เรเลชันจะมีการลาดเอียงลดหลั่นกันลงมาเป็นแบบเชิงเส้นจนถึงศูนย์
 ประโยชน์ของการลาดเอียงแบบเชิงเส้นบนฟังก์ชันออโตคอร์เรเลชันนี้จะมีผลดีต่อจุดยอดแต่ละจุดที่เป็น
 จุดคาบของพิช เพราะมันจะลดผลของจุดยอดอื่นๆที่ไม่ใช่จุดยอดตรงคาบของพิชซึ่งอาจจะมีผล
 ของฟังก์ชันออโตคอร์เรเลชันเด่นชัดมากกว่าจุดยอดตรงคาบของพิชจริงๆซึ่งจะทำให้การหาคาบของ
 พิชผิดพลาดได้



รูปที่ 2.3.11 ตัวอย่างการประมวลผลของสัญญาณใน 1 เพรม

(Dubnowski, 1976)

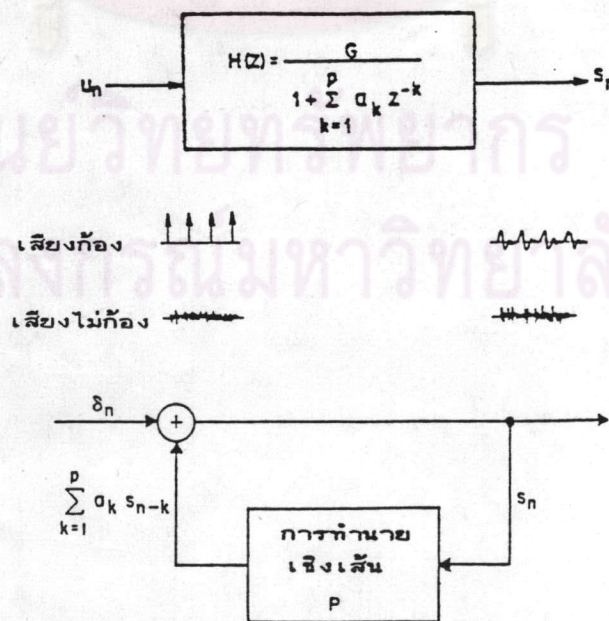
นอกจากนั้นในเรื่องระดับจุดยอดของสัญญาณ จะมีการกำหนดระดับอ้างอิงของสัญญาณในภาวะสงบ (silence level threshold) ซึ่งหามาจากการวัดระดับจุดยอดของสัญญาณในสภาวะสงบเป็นเวลา 50 มิลลิวินาที ระดับอ้างอิงต่ำสุดของสัญญาณในสภาวะสงบนี้จะถูกนำมาเทียบกับระดับสัญญาณที่จุดยอดของสัญญาณในเฟรมนี้เพื่อทำการแยกแยะว่าสัญญาณในเฟรมที่กำลังทำการประมวลผลอยู่นี้เป็นข้อมูลของสัญญาณเสียงหรือไม่ ถ้าเป็นข้อมูลของสภาวะแวดล้อมในสภาวะสงบก็จะไม่ทำการคำนวณหาคาบของพิทช์ ดังนั้นระดับอ้างอิงต่ำสุดของสัญญาณแวดล้อมในสภาวะสงบนี้จะต้องทำการเปลี่ยนแปลงค่าให้เหมาะสมตามสภาวะแวดล้อม เพื่อจะได้ทำการแยกแยะสัญญาณเสียงออกจากสัญญาณจากสภาวะแวดล้อมได้อย่างถูกต้อง ถ้าเราสามารถแยกได้แล้วว่าข้อมูลในเฟรมนี้เป็นข้อมูลของเสียงจริงก็จะนำไปประมวลผลออกโตคอร์รีเลชัน โดยจะทำการค้นหาค่าออกโตคอร์รีเลชันที่มากที่สุดภายในช่วง $m=M_1$ จนถึง $m=M_2$ โดยจะทำการเก็บเอาไว้ทั้งตำแหน่งและค่าของจุดที่มีค่ามากที่สุดนี้ไว้ ถ้าค่าที่จุดนี้มีค่ามากกว่าระดับอ้างอิงต่ำสุดของการจำแนกว่าเป็นเสียงก้องหรือไม่ก้อง ก็จะถือว่าข้อมูลส่วนนี้เป็นข้อมูลของสัญญาณเสียงก้องและคาบของพิทช์ก็จะเริ่มที่ตำแหน่งของจุดนี้ แต่ถ้าน้อยกว่าระดับอ้างอิงต่ำสุดก็จะถือว่าข้อมูลในช่วงนี้เป็นข้อมูลของสัญญาณเสียงไม่ก้อง

2.4 การวิเคราะห์โดยใช้เทคนิคการทำนายแบบเชิงเส้นหรือแอลพีซี

หลักการของการทำนายแบบเชิงเส้นปรากฏขึ้นมาเป็นครั้งแรกในหนังสือ Wiener's 1984 บทที่ 2 "The Linear Prediction for a Single Time Series" (Wiener, 1966) จากนั้นมาก็ได้มีการนำมาประยุกต์ใช้กันอย่างแพร่หลายในแขนงต่างๆกันไป นักวิจัยกลุ่มแรกๆที่นำเอาเทคนิคการทำนายแบบเชิงเส้นนี้มาประยุกต์ใช้โดยตรงในด้านของการวิเคราะห์และการสังเคราะห์เสียงพูดก็คือ Saito กับ Itakura (ปี ค.ศ. 1966) และ Atal กับ Schroeder (ปี ค.ศ. 1967) จากนั้นเทคนิคการทำนายแบบเชิงเส้นหรือแอลพีซีนี้ก็กลายเป็นเทคนิคที่นิยมใช้และแพร่หลายเป็นต้นมา

เทคนิคการทำนายแบบเชิงเส้น เป็นเทคนิคอันหนึ่งในการเข้ารหัสสัญญาณเสียงพูดโดยใช้อัตราการเก็บหน่วยความจำน้อยและเป็นที่ยอมรับกันแพร่หลายกันมากในปัจจุบัน จนกระทั่งกลายเป็นเครื่องมือที่สำคัญอันหนึ่งทั้งในการวิเคราะห์และสังเคราะห์เสียงพูด หลักการพื้นฐานของแอลพีซีก็คือเราสามารถทำการประมาณค่าตัวอย่างของสัญญาณเสียง ได้ด้วยตัวอย่างของสัญญาณเสียงที่คัดตัวอย่างมาในอดีตนำมารวมกันแบบเชิงเส้น และด้วยการทำให้ผลต่างกำลังสองของสัญญาณเสียงตัวอย่างที่ตัดมาจริงกับสัญญาณเสียงที่ประมาณค่าขึ้นมามีค่าน้อยที่สุด ก็จะทำให้เราสามารถหาค่าสัมประสิทธิ์ของการทำนาย(prediction coefficient)ที่จะนำมาใช้ในสมการรวมแบบเชิงเส้นได้

2.4.1 รูปแบบจำลองของการสร้างสัญญาณเสียงโดยเทคนิคการทำนายแบบเชิงเส้น



รูปที่ 2.4.1 รูปแบบจำลองการกำเนิดเสียงโดยใช้เทคนิคการทำนายแบบเชิงเส้น

จากรูปที่ 2.4.1 แสดงถึงรูปแบบการจำลองของการกำเนิดสัญญาณเสียงโดยใช้เทคนิคแอลพีซีทั้งในโดเมนของเวลา (บน) (Sambur, 1976) และในโดเมนความถี่ (ล่าง) (Atal and Hanauer, 1971) ซึ่งในส่วนของตัวกรองดิจิทัลที่แปรตามสัญญาณ (time-varying digital filter) นั้นจะแทนครอบคลุมถึงผลกระทบต่างๆทางด้านสเปคตรัม เช่น การแผ่กระจายของคลื่นเสียง (radiation) ลักษณะของช่องทางเดินเสียง (vocal tract) และการสั่นของเส้นเสียง (glottal excitation) ดังนั้นจากรูปแบบจำลองดังกล่าวเราสามารถเขียนรูปแบบของการสังเคราะห์เสียงให้เป็นแบบออลโพล (all pole) ในรูปของการแปลงแบบแซด ได้ดังนี้

$$S(z) = E(z) 1/A(z) \quad (\text{รูปแบบของการสังเคราะห์เสียง}) \dots\dots\dots 2.4.1$$

โดยมีการนิยามดังนี้

$$A(z) = \sum_{i=0}^M a_i z^{-i} \quad (a_0 = 1) \dots\dots\dots 2.4.2$$

ค่า M คืออันดับของการทำนายแบบเชิงเส้นหรือเทคนิคแอลพีซีนั่นเอง (Markel and Gray, 1977) และ A(z) นี้เรียกว่าตัวกรองแบบออลซีโร่ (all zero filter) และจากสมการที่ 2.4.1 จะเรียกส่วนกลับของตัวกรองแบบออลซีโร่ (1/A(z)) นี้ว่าตัวกรองแบบออลโพล (all pole filter) ก็ได้ ความหมายของสมการที่ 2.4.1 นี้ก็คือถ้าเราทำการป้อนสัญญาณ E(z) เข้าไปสู่ตัวกรองแบบออลโพล ก็จะได้เอาท์พุทคือ S(z) ออกมาซึ่งก็คือการแปลงแบบแซดของรูปแบบจำลองของสัญญาณเสียงนั่นเอง

ถ้าเราคูณสมการ 2.4.1 ทั้งสองข้างด้วย A(z) ผลจะออกมาเป็น

$$E(z) = S(z)A(z) \quad (\text{รูปแบบของการวิเคราะห์เสียง}) \dots\dots\dots 2.4.3$$

สมการนี้จะหมายถึงรูปแบบจำลองของการวิเคราะห์สัญญาณเสียง โดยถ้าเราให้ S(z) เป็นสัญญาณอินพุทเข้าไปยังส่วนกลับของตัวกรองแบบออลโพลหรือก็คือ A(z) (ซึ่งเราจะหาค่าสัมประสิทธิ์ของมันได้จากการวิเคราะห์รูปแบบคลื่นของสัญญาณเสียง) ก็จะได้เอาท์พุทออกมาเป็น E(z) หรือเป็นไดรฟวิงฟังก์ชัน (driving function) ของรูปแบบการสังเคราะห์เสียงนั่นเอง

2.4.2 รูปแบบของการทำนายแบบเชิงเส้น (Linear Prediction Model)

จากหลักการพื้นฐานของแอลพีดังที่ได้กล่าวในตอนต้นแล้วนั้น ถ้าสมมติให้สัญญาณเสียงตัวอย่างของเราที่ทำการคัดตัวอย่างที่เวลา n เป็น x(n) ต่อไปก็จะนำไปเปรียบเทียบกับสัญญาณเสียงตัวอย่างที่ทำการสุ่มมาก่อนหน้าคือ x(n-1) โดยคาดว่า x(n) ควรจะมีค่าใกล้เคียงกับ x(n-1) ที่สุด โดยการคูณสัญญาณตัวก่อนหน้านั้นด้วยค่าคงที่ a₁ ค่าผิดพลาดที่เกิดขึ้นจากการเปรียบเทียบสัญญาณจริงกับสัญญาณที่ได้มาจากการทำนายนี้จะเรียกว่า ค่าผิดพลาดของสัญญาณ (error signal) หรือค่าผิดพลาดจากการทำนาย (prediction error) มีสัญลักษณ์เป็น e(n) ซึ่งสามารถเขียนออกมาเป็นสมการได้ดังนี้

$$e(n) = x(n) - a_1x(n-1) \quad \dots\dots\dots 2.4.4$$

แต่เพื่อให้การทำนายมีความแม่นยำยิ่งขึ้น ก็จะต้องนำสัญญาณตัวอย่างนั้นไปทำการเปรียบเทียบกับสัญญาณเสียงตัวอย่างก่อนๆ หลายๆ ค่า ในที่นี้สมมติให้ทำการเปรียบเทียบกับค่าสัญญาณตัวอย่างก่อนหน้าเป็นจำนวน M ค่าก็จะได้สมการเป็น

$$e(n) = x(n) - a_1x(n-1) - a_2x(n-2) - \dots\dots\dots - a_Mx(n-M)$$

$$= x(n) - \sum_{k=1}^M a_kx(n-k) \quad \dots\dots\dots 2.4.5$$

จะเห็นได้ว่าค่าตัวคูณคงที่ a_k จะเป็นตัวปรับให้เกิดค่าผิดพลาดจากการทำนายลดน้อยลงได้ เราจะเรียกค่า a_k นี้ว่า สัมประสิทธิ์ของการทำนาย (prediction coefficients) (Markel and Gray, 1976)

ถ้ากลับมาพิจารณาสมการที่ 2.4.3 โดยการแทนค่า A(z) จากสมการ 2.4.2 จะได้ว่า

$$E(z) = \sum_{i=0}^M a_iS(z)z^{-i} \quad \dots\dots\dots 2.4.6$$

หรือเขียนในรูปฟังก์ชันของเวลาจะได้เป็น

$$e(n) = \sum_{i=0}^M a_i s(n-i) \quad a_0 = 1 \quad \dots\dots\dots 2.4.7$$

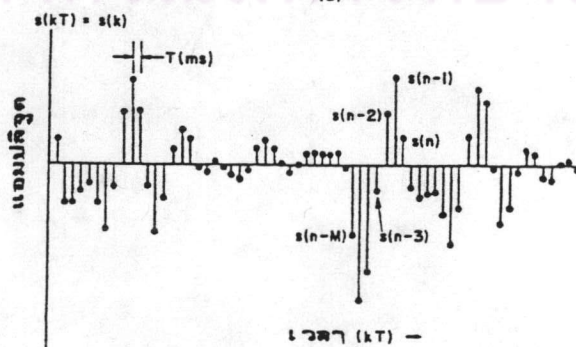
$$= s(n) + \sum_{i=1}^M a_i s(n-i) \quad \dots\dots\dots 2.4.8$$

ก็จะเห็นได้ว่าการทำนายที่มอดับเท่ากับ M ก็จะต้องใช้การรวมตัวอย่างแบบเชิงเส้นจากตัวอย่าง M ตัวอย่างก่อนหน้านั้น ถ้าเราให้ $\hat{s}(n)$ เป็นตัวอย่างที่เราทำนายขึ้นมาจะได้

$$e(n) = s(n) - \hat{s}(n) \quad \dots\dots\dots 2.4.9$$

โดยที่
$$\hat{s}(n) = - \sum_{i=1}^M a_i s(n-i) \quad \dots\dots\dots 2.4.10$$

นั่นก็คือ โดรฟวิงฟังก์ชัน $e(n)$ สามารถที่จะมองได้เป็นค่าผิดพลาดจากการทำนายระหว่างสัญญาณจริงกับสัญญาณที่เราทำนายขึ้นมา ค่า $-a_i, i=1,2,\dots,M$ ก็คือค่าสัมประสิทธิ์ของการทำนายที่จะต้องทำการหา ส่วนเครื่องหมายลบทำให้ค่า $e(n)$ เป็นค่าของความแตกต่างของสัญญาณทั้งสองตัว



รูปที่ 2.4.2 การตัดตัวอย่างของสัญญาณเสียงทุกๆช่วงเวลา T

จากรูปแสดงถึงการตัดตัวอย่างของสัญญาณเสียงพูด $s(t)$ ทุกๆช่วงเวลา T มิลลิวินาทีทุกตัวอย่างที่ n $s(n)$ ก็จะได้มาจากการทำนายโดยใช้การรวมกันแบบเชิงเส้นของสัญญาณก่อนหน้านั้น M ตัวอย่าง คือ $s(n-1), s(n-2), \dots, s(n-M)$ สมการที่ 2.4.10 เขียนเป็นการแปลงแบบแซดได้เป็น

$$\hat{S}(z) = F(z)S(z) \quad \dots\dots\dots 2.4.11$$

โดยที่
$$F(z) = - \sum_{i=1}^M a_i z^{-i} \quad \dots\dots\dots 2.4.12$$

$F(z)$ คือฟังก์ชันตัวกรองของการทำนายแบบเชิงเส้น ส่วน $\hat{S}(z) \leftrightarrow \hat{s}(n)$ และ $S(z) \leftrightarrow s(n)$ นั้นคือการนิยามตามความสัมพันธ์ระหว่างข้อมูลตัวอย่างกับคู่ของการแปลงแบบแซด จากสมการที่ 2.4.9, 2.4.10 รูปแบบของการทำนายเชิงเส้นในโดเมนของการแปลงแบบแซดจะมีลักษณะดังนี้

$$\begin{aligned} E(z) &= S(z)[1-F(z)] \\ \text{หรือ} \quad &= S(z)A(z) \quad \dots\dots\dots 2.4.13 \end{aligned}$$

$$\begin{aligned} \text{โดยที่} \quad A(z) &= 1 - F(z) \\ &= 1 + \sum_{i=1}^M a_i z^{-i} \quad \dots\dots\dots 2.4.14 \end{aligned}$$

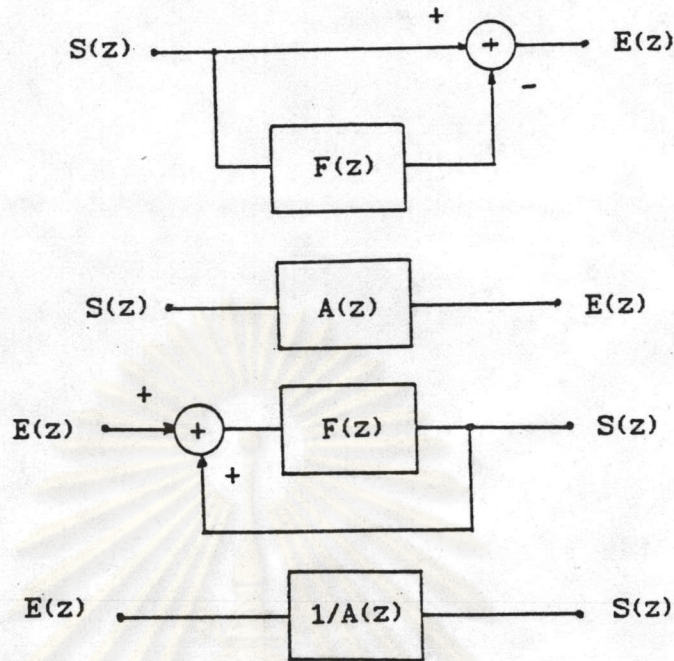
สำหรับรูปแบบในด้านการสังเคราะห์สัญญาณเสียงก็จะได้เป็น

$$\begin{aligned} S(z) &= E(z)/1-F(z) \\ \text{หรือ} \quad &= E(z)/A(z) \quad \dots\dots\dots 2.4.15 \end{aligned}$$

เขียนในโดเมนของเวลาได้เป็น

$$s(n) = e(n) - \sum_{i=1}^M a_i s(n-i) \quad \dots\dots\dots 2.4.16$$

ซึ่งทั้งสองลักษณะแสดงเป็นบล็อกไดอะแกรมได้ดังนี้



รูปที่ 2.4.3 บล็อกไดอะแกรมของแบบจำลองของการวิเคราะห์ (บน) และการสังเคราะห์(ล่าง) ของการทำนายแบบเชิงเส้น

สมการในเทอมของตัวกรองการทำนายแบบเชิงเส้นเสนอขึ้นมาโดย Atal เมื่อปี ค.ศ. 1970 และในเทอมของส่วนกลับของตัวกรองการทำนายแบบเชิงเส้นโดย Markel เมื่อปี ค.ศ. 1971 โดยสรุปใจความสำคัญของการทำนายแบบเชิงเส้นว่า พารามิเตอร์ของฟังก์ชัน $A(z)$ สามารถหาได้มาจากรูปคลื่นของสัญญาณเสียงโดยตรงจากการประยุกต์ใช้หลักการของค่าผิดพลาดกำลังสองที่น้อยที่สุด (least mean square error) จากสมการที่ 2.4.9 เนื่องจาก $e(n)$ คือค่าผิดพลาดระหว่างสัญญาณจริงกับสัญญาณที่ทำขึ้นมา ดังนั้นจากหลักการดังกล่าวเราก็จะต้องหาสัมประสิทธิ์ $a_i, i=1,2,\dots,M$ ที่จะทำให้ค่า $e(n)$ มีค่าน้อยที่สุด แต่เนื่องจากเราจะทำการแบ่งสัญญาณเสียงมาพิจารณาเป็นเฟรมๆไป ดังนั้นจึงกำหนดค่าผิดพลาดยกกำลังสองรวมขึ้นมาให้เป็น α ซึ่งมีนิยามดังนี้

$$\alpha = \sum_{n=n_0}^{n_1} e^2(n) \quad \dots\dots\dots 2.4.17ก$$

$$= \sum_{n=n_0}^{n_1} \left[\sum_{i=0}^M a_i s(n-i) \right]^2 \quad \dots\dots 2.4.17\text{ข}$$

$$= \sum_{n=n_0}^{n_1} \sum_{i=0}^M \sum_{j=0}^M a_i s(n-i) s(n-j) a_j \quad \dots\dots 2.4.17\text{ค}$$

ค่า n_0 กับ n_1 เป็นตัวกำหนดขอบเขตที่จะทำการพิจารณาซึ่งจะเกิดค่าผิดพลาดน้อยที่สุด ถ้าเรานิยามตัวแปรขึ้นมาตัวหนึ่งให้มค่าดังนี้

$$c_{ij} = \sum_{n=n_0}^{n_1} s(n-i) s(n-j) \quad \dots\dots 2.4.18$$

ก็จะทำให้เขียนสมการที่ 2.4.17 ได้ใหม่ว่า

$$\alpha = \sum_{i=0}^M \sum_{j=0}^M a_i c_{ij} a_j \quad \dots\dots 2.4.19$$

ทำให้ค่าผิดพลาดยกกำลังสองมีค่าน้อยที่สุด โดยการดิฟเฟอเรนเชียลสมการ 2.4.19 ทั้งสองข้างเทียบกับ a_k โดยที่ $k=1, 2, \dots, M$ แล้วจัดให้เท่ากับศูนย์ก็จะได้

$$\frac{\partial \alpha}{\partial a_k} = 0 = 2 \sum_{i=0}^M a_i c_{ik} \quad \dots\dots 2.4.20$$

แต่ $a_0 = 1$ เพราะฉะนั้น

$$\sum_{i=1}^M a_i c_{ik} = -c_{0k} \quad k=1, 2, \dots, M \quad \dots\dots 2.4.21$$

สมการที่ 2.4.21 นี้เป็นสมการเชิงเส้นที่มีตัวไม่รู้ค่าอยู่ M ตัว $\{a_i\}$ ซึ่งจะต้องหาออกมาจากสมการเชิงเส้น M สมการ ค่า c_{ik} $i=0, 1, \dots, M$ นั้นเรารู้ค่าได้จากสมการที่ 2.4.18 อยู่แล้ว ถ้าเราหาค่า a_i $i=0, 1, \dots, M$ ออกมาได้จากสมการที่ 2.4.21 นี้ก็สามารถนำไปหาค่าผิดพลาด $e(n)$ ในการทำการวิเคราะห์เสียงได้จากสมการที่ 2.4.8 และอาจนำไปสังเคราะห์เป็นสัญญาณเสียงกลับคืนมาได้ด้วยสมการที่ 2.4.16 สำหรับวิธีการแก้สมการที่ 2.4.21 เพื่อหาค่า a_i ออกมานั้นมีสามวิธีที่น่าสนใจและสำคัญคือ

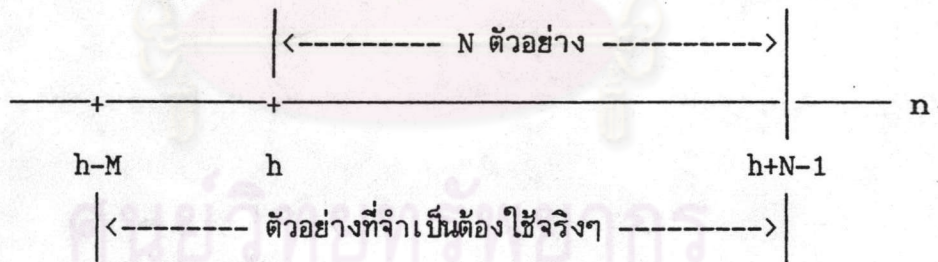
- 1) วิธีโควาเรียนซ์ (Covariance Method)
- 2) วิธีออโตคอร์เรเลชัน (Autocorrelation Method)
- 3) วิธีพาร์คอร์ (PARCOR Method)

2.4.3 วิธีการหาค่าสัมประสิทธิ์ของการทำนาย

2.4.3.1 วิธีโควาเรียนซ์

เป็นแนวทางที่ใช้การนิยามสัญญาณเสียง $s(n)$ ออกเป็นส่วนๆ และมีการจำกัดขอบเขตในการรวม เพื่อให้พิจารณาอยู่เฉพาะภายในช่วงที่มีการหาผลรวมของค่าผิดพลาดยกกำลังสองเท่านั้น เช่นสมมติว่าถ้าเรามีตัวอย่างของสัญญาณเสียงอยู่ N ตัวอย่างคือ $s(n) = s(0), s(1), s(2), \dots, s(N)$ ถ้าเราเริ่มต้นคำนวณที่จุด $n=h$ ใดๆ เพื่อทำการหาค่าสัมประสิทธิ์ของการทำนายระหว่างตัวอย่างที่ h ถึงตัวอย่างที่ $h+N$ ดังนั้นการหาค่าผิดพลาดยกกำลังสองรวมทั้งน้อยที่สุดก็จะอยู่ภายในกรอบ $[h, h+N-1]$ นั่นคือ

$$\alpha = \sum_{n=h}^{h+N-1} e^2(n) \quad \dots\dots\dots 2.4.22$$



รูปที่ 2.4.4 ขอบเขตของตัวอย่างที่ใช้ในวิธีการโควาเรียนซ์

และนำเอาตัวอย่างมาใช้ในการคำนวณหาสัมประสิทธิ์ของโควาเรียนซ์เมทริกซ์ c_{1j} ดังนี้

$$\sum_{i=1}^M a_i c_{1j} = -c_{0j} \quad j=1, 2, \dots, M \quad \dots\dots\dots 2.4.23$$

โดยที่
$$c_{i,j} = \sum_{n=h}^{h+N-1} s(n-i)s(n-j) \quad 0 \leq i < M, \quad 1 \leq j < M \quad \dots\dots\dots 2.4.24$$

จากสมการที่ 2.4.23 เขียนในรูปเมทริกซ์ได้เป็น

$$\begin{bmatrix} c(1,1) & c(1,2) & c(1,3) & \dots\dots & c(1,M) \\ c(2,1) & c(2,2) & c(2,3) & \dots\dots & c(2,M) \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ c(M,1) & c(M,2) & c(M,3) & \dots\dots & c(M,M) \end{bmatrix} \begin{bmatrix} a(1) \\ a(2) \\ \dots \\ \dots \\ a(M) \end{bmatrix} = \begin{bmatrix} c(0,1) \\ c(0,2) \\ \dots \\ \dots \\ c(0,M) \end{bmatrix}$$

..... 2.4.25

จากสมการที่ 2.4.24 เราจะได้ว่า $c_{i,j} = c_{j,i}$ ดังนั้นเมทริกซ์ในสมการที่ 2.4.25 ก็จะเป็นเมทริกซ์ที่สมมาตรตามแนวเส้นทแยงมุม เพราะฉะนั้นในการหาผลลัพท์ของสมการหรือเมทริกซ์นี้ก็จะทำได้อย่างมีประสิทธิภาพสูงขึ้น เช่นการใช้ทฤษฎีของ Cholesky ในการหาผลลัพท์ออกมา วิธีโควาเรียนซ์นี้ไม่ต้องมีการใช้ฟังก์ชันของช่องวิเคราะห์เข้ามากำหนดรูปร่างของสัญญาณเสียง และมันจะสามารถให้ค่าสัมประสิทธิ์ของการทำนายที่ค่อนข้างจะแม่นยำ ในขณะที่ใช้เฟรมวิเคราะห์เล็กกว่าวิธีออโตคอร์เรเลชันด้วย ซึ่งตามปกติก็จะใช้ขนาดของเฟรมประมาณ 50-100 ตัวอย่างสัญญาณเสียง และทำการคำนวณครั้งต่อไปทุกๆ 100 ถึง 250 ตัวอย่างสัญญาณ (Parson, 1986)

2.4.3.2 วิธีออโตคอร์เรเลชัน

จากสมการ 2.4.17 ถ้าเรากำหนดให้ $n_0 = -\infty$ และ $n = \infty$ และนิยามให้ $s(n)$ มีค่าเท่ากับศูนย์ในกรณีที่ $n < 0$ และ $n \geq N$ นั่นคือการกำหนดรูปแบบของแถบแบบสี่เหลี่ยมขนาด $(0, N]$ จะทำให้สมการที่ 2.4.18 กลายเป็น

$$c_{i,j} = \sum_{n=-\infty}^{\infty} s(n-i)s(n-j)$$

$$\begin{aligned}
 &= \sum_{n=-\infty}^{\infty} s(n)s(n+|i-j|) \\
 &= \sum_{n=0}^{N-1-|i-j|} s(n)s(n+|i-j|) \\
 &= r(|i-j|) \qquad \dots\dots\dots 2.4.26
 \end{aligned}$$

ในกรณีนี้แม้ว่าค่าผิดพลาด $e(n)$ จะเป็นการหาค่าต่ำสุดภายในช่วงที่ไม่มีการจำกัดขอบเขตก็ตาม แต่ผลก็จะได้เช่นเดียวกับการหาค่าผิดพลาดต่ำสุดภายในช่วง $[0, N+M-1]$ เหมือนกัน เนื่องจากว่าค่า $s(n)$ เป็นศูนย์สำหรับ $n \geq N$ และ $n < 0$ นั้นเอง

$$\sum_{i=1}^M a_i r(|i-j|) = -r(j) \qquad j=1,2,\dots,M \qquad \dots\dots\dots 2.4.27$$

โดยที่
$$r(l) = \sum_{n=0}^{N-1-l} s(n)s(n+l) \qquad (l \geq 0) \qquad \dots\dots\dots 2.4.28$$

และ
$$e(n) = \sum_{i=0}^M a_i s(n-i) \qquad a_0 = 1 \text{ และ } n=0,1,\dots,N+M-1 \qquad \dots\dots\dots 2.4.29$$

จากสมการที่ 2.4.27 เขียนเป็นเมทริกซ์ได้เป็น

$$\begin{bmatrix} R(0) & R(1) & R(2) & \dots\dots & R(M-1) \\ R(1) & R(0) & R(1) & \dots\dots & R(M-2) \\ \dots & \dots & \dots & \dots\dots & \dots \\ \dots & \dots & \dots & \dots\dots & \dots \\ R(M-1) & R(M-2) & R(M-3) & \dots\dots & R(0) \end{bmatrix} \begin{bmatrix} a(1) \\ a(2) \\ \dots \\ \dots \\ a(M) \end{bmatrix} = \begin{bmatrix} R(1) \\ R(2) \\ \dots \\ \dots \\ R(M) \end{bmatrix} \qquad \dots\dots\dots 2.4.30$$

ก็จะได้เมทริกซ์ขนาด $M * M$ สามารถแก้สมการหาผลลัพท์ออกมาได้หลายวิธีด้วยกัน แต่วิธีอันหนึ่งที่นิยมใช้กันก็คือวิธีของ Durbin และ Levinson ซึ่งจะใช้ขั้นตอนหรือจำนวนครั้งในการคำนวณน้อยกว่าที่จะต้องใช้ในการแก้สมการตามวิธีพื้นฐานปกติ สำหรับวิธีของ โดคอรี่ เลชันนี้จำเป็นที่จะต้องมีการใช้ฟังก์ชันของช่องแคบมากำหนดรูปแบบของสัญญาณ โดยสัญญาณที่อยู่นอกช่องแคบจะมีค่าเป็นศูนย์หมด ในการคำนวณนั้นตามปกติจะใช้ช่องแคบที่มีขนาดประมาณ 100-250 ตัวอย่างสัญญาณเสียงเพื่อใช้ในการวิเคราะห์ในหนึ่งเฟรม

2.4.3.3 การเปรียบเทียบระหว่างวิธี โควา เรียนซ์กับวิธี โดคอรี่ เลชัน

วิธีการของวิธี โดคอรี่ เลชัน จำเป็นจะต้องมีการใช้การกำหนดขนาดของช่องแคบ หรือนำเอาวินโดว์ฟังก์ชันเข้ามาคูณกับสัญญาณเสียงด้วย ซึ่งหมายถึงว่าในทางปฏิบัติ เราจะต้องพิจารณาจากตัวอย่างสัญญาณเป็นช่วงเวลาที่ยาวนานกว่าวิธี โควา เรียนซ์ โดยจะเห็นได้จากวิธีของ โดคอรี่ เลชันจะกำหนดจำนวนตัวอย่างของสัญญาณภายในช่องแคบที่จะนำมาพิจารณามีค่า N ประมาณ 256 ตัวอย่างและสำหรับวิธี โควา เรียนซ์จะกำหนดค่า N ประมาณ 100 ตัวอย่าง สำหรับในด้านของการคำนวณนั้น ได้มีการเปรียบเทียบทั้งสองเทคนิคนี้โดย Witten ปี ค.ศ. 1980 ได้ว่าวิธีการหาผลลัพท์จากการอินเวอร์สเมทริกซ์ของ Durbin และ Levinson สำหรับวิธีของ โดคอรี่ เลชันจะมีประสิทธิภาพดีกว่าการหาผลลัพท์โดยการแยกองค์ประกอบ (decomposition) ของ Cholesky สำหรับวิธี โควา เรียนซ์ที่จำนวนครั้งเท่าๆกัน (Witten, 1982)

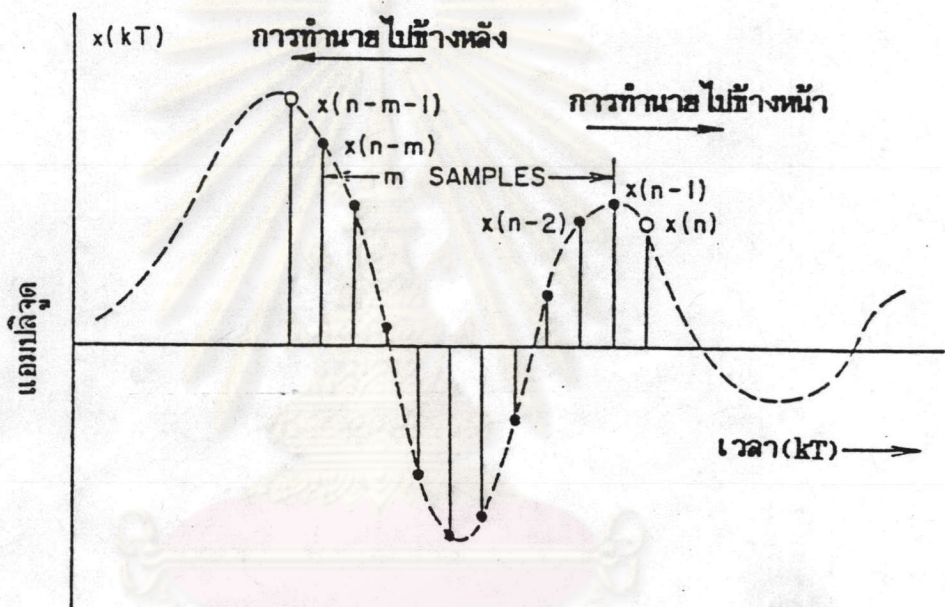
องค์ประกอบซึ่งใช้ในการให้นำหนักสำหรับวิธี โควา เรียนซ์นั้นยากในการที่จะแบ่งปริมาณเป็นระดับต่างๆในทันทีทันใดได้ แต่สำหรับวิธีของ โดคอรี่ เลชันนั้น เราสามารถใช้การคำนวณทางคณิตศาสตร์เป็นแบบกำหนดความละเอียด (fix point) ได้เลย ทำให้สามารถนำไปตัดแปลงเป็นวงจรการคำนวณทางด้านฮาร์ดแวร์ได้โดยง่าย นอกจากนี้สิ่งที่สำคัญสิ่งหนึ่งก็คือวิธี โควา เรียนซ์ในบางครั้งอาจจะทำให้ได้ผลลัพท์ออกมาที่ไม่เสถียรได้ ในขณะที่วิธีของ โดคอรี่ เลชันนั้นสามารถจะแสดงให้เห็นได้ว่าผลลัพท์ที่ออกมาจะเสถียรอย่างแน่นอน แต่กระนั้นก็ตามผลจากการประมาณคุณสมบัติของสัญญาณโดยการใช้ช่องแคบ และผลกระทบของการกระทำกับตัวอย่างของสัญญาณจำนวนมากของวิธีของ โดคอรี่ เลชัน จะทำให้ค่าสัมประสิทธิ์ที่ใช้แทนสัญญาณเสียงที่ได้จากวิธี โควา เรียนซ์จะมีความแม่นยำค่อนข้างจะสูงกว่า ซึ่งเป็นข้อดีที่สำคัญของวิธี โควา เรียนซ์

2.4.3.4 พาร์คอร์ (PARCOR : Partial Correlation)

เป็นการประยุกต์เทคนิคการทำนายแบบเชิงเส้นกับสัญญาณเสียง

แนวทางหนึ่งซึ่งเสนอขึ้นมาโดย Itakura และ Saito (1969-1972) โดยใช้ค่าสัมประสิทธิ์พาร์เชียลคอรีเลชัน (partial correlation coefficients) และแนวทางนี้ก็ได้นำมาใช้ในการกำหนดนิยามรายละเอียดภายในของส่วนกลับตัวกรอง $A(z)$ (Wakita, 1982)

วิธีพาร์คอร์นี้จะทำการทำนายแบบเชิงเส้นที่มีอันดับเท่ากับ m ทั้งการทำนายไปข้างหน้าและทำนายย้อนกลับหลังดังรูปข้างล่าง โดยกำหนดให้ค่าผิดพลาดในการทำนายไปข้างหน้า m อันดับที่เวลา n โดยมีค่าเป็น $x_m^+(n)$ และค่าผิดพลาดในการทำนายย้อนกลับ m อันดับที่เวลา n โดยมีค่าเป็น $x_m^-(n)$ โดยนิยามค่าทั้งสองได้ดังนี้



รูปที่ 2.4.5 หลักการของวิธีพาร์คอร์

$$\begin{aligned}
 x_m^+(n) &= x(n) - [-\sum_{i=1}^m a_{m,i} x(n-i)] \\
 &= \sum_{i=0}^m a_{m,i} x(n-i) \quad a_{m,0} = 1 \quad \dots\dots\dots 2.4.31
 \end{aligned}$$

$x_m^+(n)$ ได้มาจากผลต่างของสัญญาณที่เวลา n กับสัญญาณที่ได้จากการทำนายที่เกิดจากสมการการรวมแบบเชิงเส้นของสัญญาณที่อยู่ก่อนหน้าไป m หน่วยของเวลา ส่วน $x_m^-(n)$ ก็ได้มาจากผลต่างของสัญญาณที่เวลา n กับสัญญาณที่ได้จากการทำนายที่เกิดจากการรวมแบบเชิงเส้นของสัญญาณที่อยู่ถัดไปตามแกนเวลา โดยเริ่มคิดที่เวลา $n-m-1$

$$\begin{aligned}
 x_m^-(n) &= x(n-m-1) - [-\sum_{i=1}^m b_{m,i} x(n-i)] \\
 &= \sum_{i=1}^{m+1} b_{m,i} x(n-i) \quad b_{m,m+1} = 1 \quad \dots\dots\dots 2.4.32
 \end{aligned}$$

ตามเทคนิคการทำนายแบบเชิงเส้นจะต้องหาผลรวมของค่าผิดพลาดกำลังสองของทั้งสองด้านได้เป็น

$$\alpha_m = \sum_{n=n_0}^{n_1} [x_m^+(n)]^2 \quad m=1,2,\dots,M \quad \dots\dots\dots 2.4.33$$

$$\beta_m = \sum_{n=n_0}^{n_1} [x_m^-(n)]^2 \quad m=1,2,\dots,M \quad \dots\dots\dots 2.4.34$$

และหาค่าต่ำสุดของผลรวมของค่าผิดพลาดกำลังสองโดยการทำการหาค่าต่ำสุดของฟังก์ชันเชิงเส้นกับพารามิเตอร์ที่สนใจแล้วจึงให้เท่ากับศูนย์ได้

$$\frac{\partial \alpha_m}{\partial a_{m1}} = 0 \quad \text{และ} \quad \frac{\partial \beta_m}{\partial b_{m1}} = 0 \quad i=1,2,\dots,m \quad \dots\dots\dots 2.4.35$$

แต่จากเงื่อนไขนี้ ยังไม่สามารถนำไปใช้แก้สมการเพื่อหาค่าสัมประสิทธิ์ของการทำนายได้โดยตรง Markel และ Gray ได้เสนอแนวทางในการหาค่าตอบออกมาเป็นค่าสัมประสิทธิ์ของพหุนามด้วยสัญลักษณ์ k_m โดยใช้หลักการของอินเนอร์โปรดักต์และคุณสมบัติของออร์โธโกนอลที่จะกล่าวต่อไปนี้

2.4.3.5 หลักการของอินเนอร์โปรดักต์และคุณสมบัติออร์โธโกนอล

(Inner Products and Orthogonality Principle)

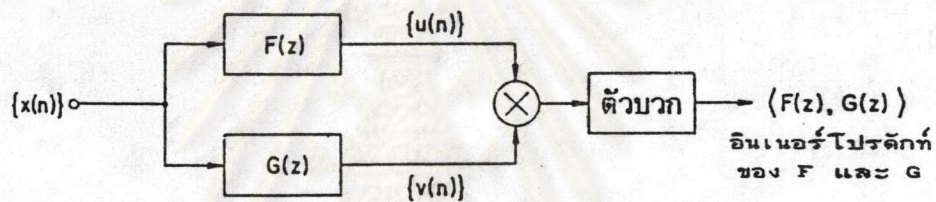
ค่าผิดพลาดจากการทำนายของสัญญาณทั้งสองด้านคือ $x_m^+(n)$ และ $x_m^-(n)$ สามารถจะพิจารณาให้เป็นเอาต์พุตที่ออกมาจากตัวกรอง $A_m(z)$ และ $B_m(z)$ ตามลำดับ โดยมีสัญญาณอินพุตเป็น $x(n)$ โดยที่

$$A_m(z) = \sum_{i=0}^m a_{m,i} z^{-i} \quad \text{โดย } a_{m,0}=1 \quad \dots\dots\dots 2.4.36ก$$

และ $B_m(z) = \sum_{i=1}^{m+1} b_{m,i} z^{-i} \quad \text{โดย } b_{m,m+1}=1 \quad \dots\dots\dots 2.4.36ข$

ผลรวมของค่าผิดพลาด α_m และ β_m แทนถึงพลังงานที่ออกจากตัวกรองภายในช่วงเวลาจาก $n=n_0$ ถึง $n=n_1$

หลักการของอินเนอร์โปรดักต์เริ่มจากการสมมติให้สัญญาณเสียง $x(n)$ ป้อนเข้าผ่านตัวกรองสองตัวคือ $F(z)$ กับ $G(z)$ ซึ่งมีค่าสัมประสิทธิ์เป็นจำนวนจริงดังรูป



รูปที่ 2.4.6 แสดงถึงอินเนอร์โปรดักต์ของตัวกรอง $F(z)$ กับ $G(z)$

(Markel and Gray, 1976)

เอาท์พุทของตัวกรองแต่ละตัวคือ $u(n)$ กับ $v(n)$ ตามลำดับ จะถูกนำมาคูณและบวกกันตั้งแต่ตัวอย่างที่ $n=n_0$ ถึง $n=n_1$ ผลที่ได้ออกมาจะเรียกว่าอินเนอร์โปรดักต์ของ $F(z)$ กับ $G(z)$ โดยมีสัญลักษณ์เป็น $\langle F(z), G(z) \rangle$ และนิยามได้ดังนี้

$$\langle F(z), G(z) \rangle = \sum_{n=n_0}^{n_1} u(n)v(n) \quad \dots\dots\dots 2.4.37$$

ถ้าตัวกรอง $F(z)$ และ $G(z)$ อยู่ในรูปของ

$$F(z) = \sum_{i=0}^{\infty} f_i z^{-i} \quad \text{และ} \quad G(z) = \sum_{i=0}^{\infty} g_i z^{-i} \quad \dots\dots\dots 2.4.38$$

อินเนอร์โปรดักต์ของ $F(z)$ กับ $G(z)$ จะเท่ากับ

$$\langle F(z), G(z) \rangle = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} f_i \left[\sum_{n=n_0}^{n_1} x(n-i)x(n-j) \right] g_j \quad \dots\dots\dots 2.4.39$$

โดยที่ $\langle z^{-i}, z^{-j} \rangle = \sum_{n=n_0}^{n_1} x(n-i)x(n-j) \quad \dots\dots\dots 2.4.40$

ด้วยวิธีโควาเรียนซ์ ซึ่งกำหนดให้ $n_0 = M$ และ $n_1 = N-1$ จะได้

$$\langle z^{-i}, z^{-j} \rangle = C_{i,j} = C_{j,i} \quad \dots\dots\dots 2.4.41$$

ส่วนวิธีออกโคคอร์รีเลชันซึ่งมีขอบเขต $n_0 = -\infty, n_1 = \infty$ และค่าของข้อมูลจะไม่เป็นศูนย์เมื่ออยู่ในช่วง $[0, N-1]$ เท่านั้นจะได้ว่า

$$\langle z^{-i}, z^{-j} \rangle = r(i-j) = r(j-i) \quad \dots\dots\dots 2.4.42$$

จากการนิยามอินเนอร์โปรดักต์ข้างต้นค่าผิดพลาดจากการทำนายทั้งสองด้านคือ α_m กับ β_m ถ้าให้มันเป็นเอาท์พุทของตัวกรอง $A_m(z)$ กับ $B_m(z)$ ตามลำดับจะสามารถเขียนให้อยู่ในรูปแบบตามนิยามของอินเนอร์โปรดักต์ได้เป็น

$$\alpha_m = \langle A_m(z), A_m(z) \rangle \quad \dots\dots\dots 2.4.43ก$$

$$\beta_m = \langle B_m(z), B_m(z) \rangle \quad \dots\dots\dots 2.4.43ข$$

ซึ่งในกรณีนี้คืออินเนอร์โปรดักต์ของตัวกรองตัวมันเอง จะเรียกใหม่ว่า นอร์มสแควร์ (norm square) และมีนิยามดังนี้

$$\|F(z)\|^2 = \langle F(z), F(z) \rangle \quad \dots\dots\dots 2.4.44$$

จุดวิกฤตของการหาค่าต่ำสุดของค่าผิดพลาดจากการทำนายสามารถอธิบายได้ด้วยการหาค่าต่ำสุดของนอร์มสแควร์ของตัวกรอง $A_m(z)$ และ $B_m(z)$ ซึ่งเป็นรูปแบบ

ของสมการโพลิโนเมียลจาก $m=1,2,\dots,M$ ต่อไปจะนำหลักการของอินเนอร์โปรดักท์เข้ามาประยุกต์ใช้เพื่อแสดงถึงคุณสมบัติออร์โธโกนอลดังนี้

ถ้าตัวกรอง $A_m(z)$ และ $B_m(z)$ เป็นตัวทำให้ได้ α_m และ β_m ที่น้อยที่สุดจริงๆแล้ว ถ้าเราบวกค่า cz^{-j} โดยที่ $j=1,2,\dots,m$ และ c เป็นค่าคงที่เข้าไปในโพลิโนเมียลแล้ว นอร์มสแควร์หลังการบวกจะต้องใหญ่กว่านอร์มสแควร์ที่ยังไม่ได้บวกเพราะฉะนั้นจะได้

$$\|A_m(z) + cz^{-j}\|^2 \geq \|A_m(z)\|^2 \quad \text{โดย } j=1,2,\dots,m \quad \dots 2.4.45$$

ถ้าทำการกระจายอสมการ 2.4.45 ออกมาในรูปของอินเนอร์โปรดักท์จะได้

$$2c\langle A_m(z), z^{-j} \rangle + c^2\langle z^{-j}, z^{-j} \rangle \geq 0 \quad \dots 2.4.46$$

ถ้า $\langle z^{-j}, z^{-j} \rangle$ ไม่ได้มีค่าเป็นศูนย์แล้ว ก็สามารถจะเลือกค่า c ให้มีค่าเป็น

$$c = -\langle A_m(z), z^{-j} \rangle / \langle z^{-j}, z^{-j} \rangle \quad \dots 2.4.47$$

จะทำให้สมการที่ 2.4.47 กลายเป็น

$$-[\langle A_m(z), z^{-j} \rangle]^2 \geq 0 \quad \dots 2.4.48$$

แต่ถ้าเกิด $\langle z^{-j}, z^{-j} \rangle$ มีค่าเป็นศูนย์แล้ว ค่า c ก็สามารถที่จะเลือกให้มีค่าเป็น $c = -\langle A_m(z), z^{-j} \rangle$ เพื่อให้ได้ผลลัพธ์ออกมาเท่ากับอสมการที่ 2.4.48 โดยที่ค่า $j=1,2,\dots,m$ จากสมการที่ 2.4.48 นั้นเนื่องจากตัวกรอง $A_m(z)$ เป็นตัวกรองจำนวนจริง ดังนั้นสมการที่ 2.4.48 จะเป็นจริงได้ก็ต่อเมื่อ

$$\langle A_m(z), z^{-j} \rangle = 0 \quad \text{โดย } j=1,2,\dots,m \quad \dots 2.4.49$$

และสำหรับตัวกรอง $B_m(z)$ ก็จะมีสัญญาณได้ผลในทำนองเดียวกันคือ

$$\langle B_m(z), z^{-j} \rangle = 0 \quad \text{โดย } j=1, 2, \dots, m \quad \dots\dots\dots 2.4.50$$

ความสัมพันธ์ของทั้งสองสมการคือ 2.4.49 กับ 2.4.50 นี้จะแสดงให้เห็นถึงคุณสมบัติของ โข โจนอล ซึ่งจะนำไปประยุกต์ใช้กับการหาค่าน้อยที่สุดของค่าผิดพลาดจากการทำนายทั้งสองด้านคือ α_m และ β_m ต่อไป

ความสัมพันธ์กันแบบออร์โธโกนอลนี้สามารถแสดงในเทอมของสัญญาณผิดพลาดกับอินพุตที่มีการหน่วงเวลา ได้จากนิยามของอินเนอร์โปรดักต์และสัญญาณผิดพลาดในสมการที่ 2.4.23 กับ สมการที่ 2.4.24 ได้ดังนี้

$$\langle A_m^+(z), z^{-j} \rangle = \sum_{n=n_0}^{n_1} x_m^+(n)x(n-j) = 0 \quad \dots\dots\dots 2.4.51ก$$

$$\langle B_m^-(z), z^{-j} \rangle = \sum_{n=n_0}^{n_1} x_m^-(n)x(n-j) = 0 \quad \dots\dots\dots 2.4.51ข$$

2.4.3.6 การหาสัมประสิทธิ์พาร์คอร์และโครงสร้างแบบแลททิซ (PARCOR lattice structure)

การหาผลลัพท์โพลีโนเมียลของตัวกรอง $A_m(z)$ และ $B_m(z)$ โดยใช้คุณสมบัติความสัมพันธ์ออร์โธโกนอลในลักษณะของการเรียกใช้ตัวเองตามวิธีของ Levinson Recursion แสดง ได้ดังต่อไปนี้

จากสมการที่ 2.4.49 , 2.4.50 จะได้ค่าเริ่มต้นของความสัมพันธ์เมื่อเริ่มประมวลผลตั้งแต่ $m=1, 2, \dots, M$ ดังนี้

$$A_0(z) = 1 \quad \text{และ} \quad B_0(z) = z^{-1} \quad \dots\dots\dots 2.4.52$$

เนื่องจากทั้ง $A_m(z)$ และ $B_m(z)$ เป็นฟังก์ชันโพลีโนเมียลของ z^{-j} เมื่อ $j=0, 1, \dots, m-1$ ดังนั้นการรวมกันแบบเชิงเส้นในรูปของ $A_{m-1}(z) + k_m B_{m-1}(z)$ ก็จะเป็นฟังก์ชันโพลีโนเมียลที่ออร์โธโกนอลกับ z^{-j} โดยที่ $j=1, 2, \dots, m-1$ ด้วย ถ้า K_m

เป็นค่าที่เลือกมาเพื่อทำให้การรวมกันแบบเชิงเส้นออโธโกนอลกับ z^{-m} แล้ว $A_m(z)$ สามารถเขียนได้ในรูปดังนี้

$$A_m(z) = A_{m-1}(z) + k_m B_{m-1}(z) \quad \dots\dots\dots 2.4.53$$

และถ้าค่า k_m ทำให้ $A_m(z)$ ออโธโกนอลกับ z^{-m} แล้วจะทำให้

$$\langle A_m(z), z^{-m} \rangle = \langle A_{m-1}(z), z^{-m} \rangle + k_m \langle B_{m-1}(z), z^{-m} \rangle = 0 \quad \dots\dots\dots 2.4.54$$

และในการทำงานเดียวกันทั้ง $A_{m-1}(z)$ และ $B_{m-1}(z)$ ก็ออโธโกนอลกับ z^{-m} เมื่อ $j=1, 2, \dots, m-1$ ด้วย ดังนั้นเทอมในสมการที่ 2.4.54 สามารถเขียนกระจายได้เป็น

$$\langle A_{m-1}(z), z^{-m} \rangle = \langle A_{m-1}(z), B_{m-1}(z) \rangle = \langle 1, B_{m-1}(z) \rangle \quad \dots\dots\dots 2.4.55$$

$$\text{และ} \quad \langle B_{m-1}(z), z^{-m} \rangle = \langle B_{m-1}(z), B_{m-1}(z) \rangle = \|B_{m-1}(z)\|^2 = \beta_{m-1} \quad \dots\dots\dots 2.4.56$$

จากสมการที่ 2.4.54 แทนค่าด้วยสมการที่ 2.4.56 กับ 2.4.55 ได้ว่า

$$k_m = \frac{-\langle A_{m-1}(z), z^{-m} \rangle}{\langle B_{m-1}(z), z^{-m} \rangle} \quad \dots\dots\dots 2.4.57$$

$$k_m = - \frac{1}{\beta_{m-1}} \langle A_{m-1}(z), B_{m-1}(z) \rangle \quad \dots\dots\dots 2.4.58$$

$$= - \frac{1}{\beta_{m-1}} \sum_{n=n_0}^{n_1} x_{m-1}^+(n) x_{m-1}^-(n) \quad \dots\dots\dots 2.4.59$$

จากสมการ 2.4.59 นี้จะหาค่าสัมประสิทธิ์ของพหุนาม k_m ออกมาได้ นอกจากนี้เรายังอาจพิจารณาลงไปในรูปแบบของโครงสร้างแลททิซได้อีกด้วยจากสมการที่ 2.4.48 นำเอาสมการที่ 2.4.41 มาใช้จะได้เป็น

$$\begin{aligned} \langle A_m(z), z^{-j} \rangle &= \sum_{i=0}^m a_{m,i} \langle z^{-i}, z^{-j} \rangle \\ &= \sum_{i=0}^m a_{m,i} r(i-j) = 0 \quad \text{เมื่อ } j=1, 2, \dots, m \quad \dots\dots\dots 2.4.60 \end{aligned}$$

ถ้าทำการกลับอันดับของดัชนี โดยให้ $l=m+1-j$ และ $i=m+1-k$ สมการที่ 2.4.60 เขียนได้ใหม่เป็น

$$\sum_{k=1}^{m+1} a_{m, m+1-k} r(l-k) = 0 \quad l=1, 2, \dots, m \quad \dots\dots\dots 2.4.61$$

$$\text{แต่ } b_{mk} = a_{m, m+1-k} \quad k=1, 2, \dots, m+1 \quad \dots\dots\dots 2.4.62$$

การแปลงแบบแซดของสมการที่ 2.4.62 จะเป็น

$$B_m(z) = z^{-(m+1)} A_m(1/z) \quad \dots\dots\dots 2.4.63$$

ซึ่งจะทำให้ได้ตามหลักของออร์โธโกนอลดังนี้

$$\langle B_m(z), z^{-l} \rangle = 0 \quad l=1, 2, \dots, m \quad \dots\dots\dots 2.4.64$$

ดังนั้น $B_m(z)$ ก็คือโพลิโนเมียลที่มีค่าสัมประสิทธิ์เหมือนกับ $A_m(z)$ เพียงแต่ว่ามีอันดับที่กลับกัน เพราะฉะนั้นสมการที่ 2.4.53 มารวมกับสมการที่ 2.4.63 จะได้

$$B_m(z) = z^{-1} [k_m A_{m-1}(z) + B_{m-1}(z)] \quad \dots\dots\dots 2.4.65$$

ต่อจากนั้นถ้ามองโดยใช้การแปลงแบบเซตมาอธิบายลักษณะของ
ตัวกรองโดยให้ $X(z)$ แทนการแปลงแบบเซตของข้อมูลอินพุต $x(n)$ จะได้

$$X_m^+(z) = A_m(z)X(z) \quad \text{และ} \quad X_m^-(z) = B_m(z)X(z) \quad \dots\dots\dots 2.4.66$$

เพราะฉะนั้นจากสมการที่ 2.4.53 กับ 2.4.65 ก็จะได้

$$X_m^+(z) = X_{m-1}^+(z) + k_m X_{m-1}^-(z) \quad \dots\dots\dots 2.4.67$$

$$X_m^-(z) = z^{-1} [k_m X_{m-1}^+(z) + X_{m-1}^-(z)] \quad \dots\dots\dots 2.4.68$$

โดย $X_m^+(z)$ และ $X_m^-(z)$ แทนการแปลงแบบเซตของ $x_m^+(n)$
และ $x_m^-(n)$ ตามลำดับ ดังที่สมการที่ 2.4.67 และ 2.4.68 นี้เขียนใหม่ในรูปของโดเมน
เวลาได้เป็น

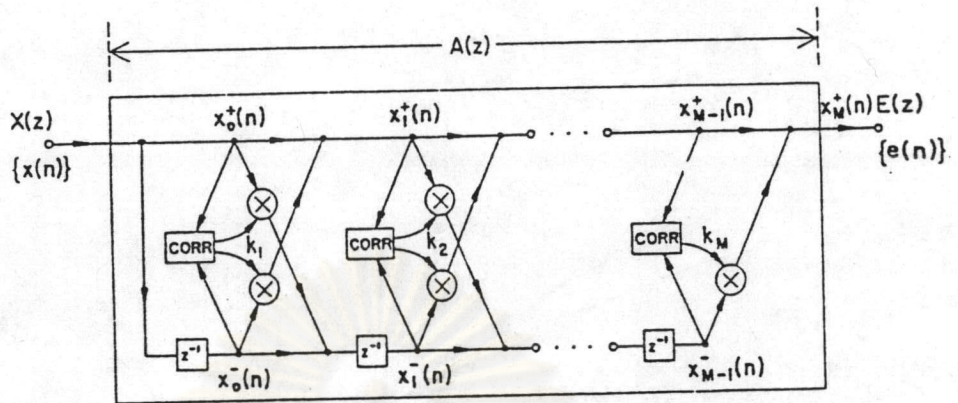
$$x_m^+(n) = x_{m-1}^+(n) + k_m x_{m-1}^-(n) \quad \dots\dots\dots 2.4.69$$

$$x_m^-(n) = k_m x_{m-1}^+(n-1) + x_{m-1}^-(n-1) \quad \dots\dots\dots 2.4.70$$

จากสมการที่ 2.4.23 และ 2.4.24 จะกำหนดขอบเขตของ
สองสมการข้างบนได้ดังนี้

$$x_0^+(n) = x(n) \quad \text{และ} \quad x_0^-(n) = x(n-1) \quad \dots\dots\dots 2.4.71$$

จากสมการที่ 2.4.69 , 2.4.70 และ 2.4.71 นี้สามารถ
แสดงให้เห็นเป็นรูปของตัวกรองการวิเคราะห์ได้ดังรูป



รูปที่ 2.4.7 โครงสร้างของส่วนกลับของตัวกรอง $\{A(z)\}$ ในรูปของโครงสร้างแลททิซ

ในกรอบสี่เหลี่ยมที่เขียนว่า CORR คือขั้นตอนในการหาค่าสัมประสิทธิ์ k_m โดยการใช้สมการที่ 2.4.59

พารามิเตอร์ k_m ที่นิยามโดย Itakura และ Saito ในชื่อของค่าสัมประสิทธิ์พาร์คอร์มีค่าเท่ากันทั้งด้านไปและด้านย้อนกลับ เพียงแต่ว่าเครื่องหมายต่างกันเท่านั้นเมื่อนำไปเปรียบเทียบกับวิธีออคอร์เลชัน สำหรับวิธีออคอร์เลชันนั้นจะได้ค่านอร์มสแควร์ของ $A_m(z)$ กับ $B_m(z)$ มีค่าเท่ากันหรือเป็นตัวเดียวกันนั่นเอง ดังนั้นสำหรับในวิธีออคอร์เลชันแล้วจะได้ว่า

$$\alpha_m = \beta_m = \sum_{n=-\infty}^{\infty} [x_m^+(n)]^2 = \sum_{n=-\infty}^{\infty} [x_m^-(n)]^2 \quad \dots\dots\dots 2.4.72$$

ทำให้ได้

$$k_m = - \frac{\sum_{n=-\infty}^{\infty} x_{m-1}^+(n)x_{m-1}^-(n)}{\left(\sum_{n=-\infty}^{\infty} [x_{m-1}^+(n)]^2 \sum_{n=-\infty}^{\infty} [x_{m-1}^-(n)]^2 \right)^{1/2}}$$

..... 2.4.73

จากสมการนี้จะเห็นว่าค่าสัมประสิทธิ์ k_m จะมีค่าอยู่ระหว่าง ± 1 เสมอ นอกจากนี้ค่าสัมประสิทธิ์ของพาร์คอร์ยังมีส่วนเกี่ยวข้องกับค่าสัมประสิทธิ์การสะท้อนกลับ (reflection coefficients) ในแง่ของวิชา Acoustics ซึ่งกล่าวถึงการทำงานของช่องเสียงของมนุษย์ว่าทำการกรองสัญญาณเสียงได้อย่างไร พุดถึงรูปแบบของช่องทางเดินเสียง และแสดงให้เห็นว่าการแปรเปลี่ยนรูปร่างของช่องทางเดินเสียงนั้นสามารถจำลองให้เห็นเป็นลักษณะของท่อยาวที่มีเส้นผ่าศูนย์กลางแตกต่างกันออกไปขึ้นมา เมื่อรูปคลื่นผ่านจากท่ออันหนึ่งไปยังอีกอันหนึ่งนั้นมันจะมีการสร้างรูปคลื่นขึ้นมาใหม่ในทิศทางตรงกันข้ามซึ่งเราเรียกว่า รูปคลื่นสะท้อนกลับ (reflected waveforms) รูปคลื่นสะท้อนกลับนี้จะมีค่าของพารามิเตอร์ตัวหนึ่งที่เกี่ยวข้องอยู่ เรียกว่า สัมประสิทธิ์การสะท้อนกลับ (reflection coefficients) ซึ่งจะมีส่วนที่สัมพันธ์กันกับฟอร์แมนท์ (formant) หรือค่าสูงสุด (peak) ในสัญญาณเสียง ดังนั้นเราอาจจะนำค่าสัมประสิทธิ์การสะท้อนกลับนี้มาวิเคราะห์และใช้ในการเปลี่ยนแปลงแบบจำลองของช่องทางเดินเสียงได้ และจากการที่ค่าสัมประสิทธิ์พาร์คอร์สามารถเทียบเท่ากับค่าสัมประสิทธิ์การสะท้อนกลับดังกล่าวนี้เอง บางทีจึงมีการเรียกค่า k_m นี้ว่า ค่าสัมประสิทธิ์ของการสะท้อนกลับ อย่างไรก็ตามคุณลักษณะของตัวกรองนั้นเราอาจจะใช้พารามิเตอร์หรือค่าสัมประสิทธิ์ใดก็ได้ ในระหว่างสัมประสิทธิ์การทำนายกับสัมประสิทธิ์การสะท้อนกลับมาใช้เป็นตัวกำหนดการแปรค่าของตัวกรอง แต่สำหรับค่าสัมประสิทธิ์การทำนายหรือ a_n (จากสมการที่ 2.4.21) นั้นอาจจะทำให้สถานะของตัวกรองเกิดไม่เสถียรขึ้นมาได้ เนื่องจากค่าของสัมประสิทธิ์การทำนายมีค่าที่ไม่แน่นอน ไม่มีขอบเขตที่แน่ชัด ด้วยเหตุนี้สัมประสิทธิ์การทำนายจึงไม่เหมาะสมที่จะนำมาใช้ เพราะว่าการเปลี่ยนแปลงเพียงเล็กน้อยของมันสามารถที่จะมีผลทำให้เกิดการเปลี่ยนแปลงอย่างมากมาเกี่ยวกับคุณลักษณะสมบัติของตัวกรองที่ใช้ในการสังเคราะห์ ส่วนค่าสัมประสิทธิ์การสะท้อนจะมีค่าอยู่ในช่วง ± 1 หรือทำให้ตัวกรองมีสถานะที่เสถียรอยู่ตลอดเวลา

2.4.3.7 ขั้นตอนการคำนวณค่าสัมประสิทธิ์พาร์คอร์

จากข้างต้นในเรื่องของพาร์คอร์นั้น เราสามารถนำมาประยุกต์ใช้ได้โดยตรงทั้งวิธีออคโตคอร์รีเลชันและวิธีโควาเรเนียน ขั้นตอนนั้นจะเป็นการเข้าสู่รายละเอียดในการแก้สมการเพื่อนำไปสู่การหาสัมประสิทธิ์ของการทำนายเชิงเส้นและสัมประสิทธิ์ของพาร์คอร์ k_m ตามลำดับโดยใช้วิธีการออคโตคอร์รีเลชัน จากสมการที่ 2.4.53

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

เราจะทำการหาค่า $A_m(z)$ จาก $A_{m-1}(z)$ และ $B_{m-1}(z)$ และสามารถหาค่า $A_m(z)$ ได้โดยการใช่วิธีคำนวณซ้ำ (recursive) โดยเริ่มต้นจากสมการที่ 2.4.52 คือ $A_0(z) = 1$ ดังตัวอย่างเช่น

$$\text{ที่ } m=1 \quad A_1(z) = 1 + k_1 B_0(z)$$

$$\text{ที่ } m=2 \quad A_2(z) = 1 + k_1 B_0(z) + k_2 B_1(z)$$

.. .. =

หรือสรุปเป็นสมการทั่วไปได้ว่า

$$A_m(z) = 1 + \sum_{i=1}^m k_i B_{i-1}(z) \quad \text{เมื่อ } m > 0 \quad \dots\dots\dots 2.4.74$$

จากสมการที่ 2.4.55 เขียนใหม่ได้เป็น

$$\langle A_m(z), B_m(z) \rangle = \langle 1, B_m(z) \rangle = \langle A_m(z), z^{-(m+1)} \rangle \quad \dots\dots\dots 2.4.75$$

จากสมการที่ 2.4.43 โดยอาศัยสมการที่ 2.4.75 จะได้ว่า

$$\alpha_m = \|A_m(z)\|^2 = \langle A_m(z), A_m(z) \rangle = \langle 1, A_m(z) \rangle \quad \dots\dots\dots 2.4.76ก$$

$$\beta_m = \|B_m(z)\|^2 = \langle B_m(z), B_m(z) \rangle = \langle z^{-(m+1)}, B_m(z) \rangle \quad \dots\dots\dots 2.4.76ข$$

นอร์มสแควร์ของสมการที่ 2.4.74 มีค่าเป็น

$$\|A_m(z) - 1\|^2 = \sum_{i=1}^m k_i^2 \beta_{i-1} \quad \dots\dots\dots 2.4.77$$

แตกค่าทางซ้ายมือของสมการที่ 2.4.77 ให้อยู่ในรูปแบบตามสม

การที่ 2.4.44 ได้เป็น

$$\begin{aligned} \|A_m(z) - 1\|^2 &= \|A_m(z)\|^2 - 2\langle A_m(z), 1 \rangle + \|1\|^2 \\ &= \|1\|^2 - \alpha_m \end{aligned} \quad \dots\dots\dots 2.4.78$$

เทียบสมการที่ 2.4.77 กับ 2.4.78 จะได้ว่า

$$\alpha_m = \|1\|^2 - \sum_{i=1}^m k_i^2 \beta_{i-1} \quad \dots\dots\dots 2.4.79$$

แทนค่า m ด้วย m+1 แล้วลบ α_{m+1} ออกด้วย α_m จะได้เป็น

$$\begin{aligned} \alpha_{m+1} &= \alpha_m - k_{m+1}^2 \beta_m \\ &= (1 - k_{m+1}^2) \alpha_m \quad \text{เนื่องจาก } \alpha_m = \beta_m \end{aligned} \quad \dots\dots\dots 2.4.80$$

จากสมการที่ 2.4.53 เพิ่มค่าของดัชนีไปอีกหนึ่งเขียนใหม่ได้เป็น

$$A_{m+1}(z) = A_m(z) + k_{m+1} B_m(z) \quad \dots\dots\dots 2.4.81$$

และจากสมการที่ 2.4.54 เพิ่มดัชนีไปอีกหนึ่งก็จะได้เป็น

$$\langle A_m(z), z^{-(m+1)} \rangle + k_{m+1} \beta_m = 0 \quad \dots\dots\dots 2.4.82$$

จากสมการที่ 2.4.60 และถ้า β_m ไม่เท่ากับศูนย์แล้วจะได้ว่า

$$\begin{aligned} k_{m+1} &= - (1/\beta_m) \langle A_m(z), z^{-(m+1)} \rangle \\ &= - (1/\alpha_m) \sum_{i=0}^m r(m+1-i) a_{m,i} \end{aligned} \quad \dots\dots\dots 2.4.83$$

เพราะฉะนั้นจะสามารถหาค่าสัมประสิทธิ์ของตัวกรองได้โดยการแทนค่า k_{m+1} ลงในสมการที่ 2.4.81 ได้เป็น

$$a_{m+1,0} = 1 \quad \dots\dots 2.4.84ก$$

$$a_{m+1,i} = a_{mi} + k_{m+1}b_{mi} \quad \text{โดย } i=1,2,\dots,m \quad \dots\dots 2.4.84ข$$

$$a_{m+1,m+1} = k_{m+1} \quad \dots\dots 2.4.84ค$$

ค่า b_{mi} ในสมการ 2.4.84ข อาจจะแทนได้ด้วยสมการที่ 2.4.61 ซึ่งเขียนใหม่ได้เป็น

$$b_{mi} = a_{m,m+1-i} \quad \text{โดย } i=1,2,\dots,m+1 \quad \dots\dots 2.4.85$$

2.4.3.8 สรุปขั้นตอนการคำนวณหาค่าสัมประสิทธิ์

ในการแก้สมการ $A_m(z)$ ที่มีอันดับเท่ากับ m โดยนั้นเราจะต้องได้ผลลัพธ์ออกมาคือ ค่าสัมประสิทธิ์ k_m และผลรวมของค่าผิดพลาดยกกำลังสอง α_m ในแต่ละขั้นตอน โดยทำทีละขั้นตอนตั้งแต่ m มีค่า = 1 จนถึงค่าอันดับของตัวกรอง $A_m(z)$ หรือค่า M นั้นเอง กระบวนการในการคำนวณแบ่งเป็นส่วนที่สำคัญ 3 ส่วนด้วยกันดังนี้

1) การคำนวณหาค่าออโตคอร์เรลชัน

สำหรับวิธีของออโตคอร์เรลชันนั้นค่า $x(n) = 0$ เมื่อ $n < 0$ และ $n > N-1$ จะได้ค่า $c_{i,j}$ ที่นิยามตามที่กล่าวเอาไว้ในตอนต้นแล้วว่า

$$c_{i,j} = c_{0,i-j} = c_{i-j,0} = r(i-j) \quad \dots\dots 2.4.86$$

สำหรับสมการที่มีอันดับเท่ากับ M จะได้ค่าสัมประสิทธิ์ออกมา $M+1$ ตัวจากสมการ

$$c_{0,k} = r(k) = \sum_{n=-\infty}^{\infty} x(n)x(n-k)$$

$$= \sum_{n=0}^{N-1-k} x(n)x(n+k) \quad \text{โดย } k=0,1,\dots,M$$

..... 2.4.87

โดย $x(n)$ เมื่อ $n = 0,1,\dots,N-1$ คือสัญญาณเสียงที่
 คัดตัวอย่างมาในหนึ่งเฟรม จากสมการที่ 2.4.87 นี้ก็จะสามารถหาค่าอโตคอรีเลชั่น $r(k)$,
 $k=0,1,\dots,M$ ออกมาได้

2) การกำหนดค่าเริ่มต้นสำหรับการแก้สมการแบบ Levinson recursion

จากสมการที่ 2.4.36 จะได้การกำหนดค่าเริ่มต้นออกมา
 เป็นสมการที่ 2.4.52 ดังนี้

$$\begin{array}{ll} A_0(z) = 1 & \text{และ} & B_0(z) = z^{-1} \\ \text{หรือ} & a_{00} = 1 & \text{และ} & b_{01} = 1 \end{array} \quad \text{..... 2.4.88}$$

จากสมการที่ 2.4.43 แทนค่าในสมการที่ 2.4.76 โดย
 $m=0$ จะได้

$$\alpha_0 = \beta_0 = \langle A_0(z), A_0(z) \rangle = \langle B_0(z), B_0(z) \rangle = r(0) \quad \text{..... 2.4.89}$$

จากสมการที่ 2.4.58 ใช้สมการที่ 2.4.55 กับ 2.4.60
 เมื่อ $m=1$ จะได้ว่า

$$\begin{aligned} k_1 &= - (1/\beta_0) \langle A_0(z), B_0(z) \rangle \\ &= - r(1)/r(0) \end{aligned} \quad \text{..... 2.4.90}$$

จากสมการ 2.4.84ก และ 2.4.84ข แทนค่า $m = 0$ ได้

$$a_{10} = 1 \quad \text{และ} \quad a_{11} = k_1 \quad \dots\dots\dots 2.4.91$$

จากสมการที่ 2.4.80 เมื่อ $m = 0$ จะได้

$$\alpha_1 = (1 - k_1^2)\alpha_0 = (1 - k_1^2)r(0) \quad \dots\dots\dots 2.4.92$$

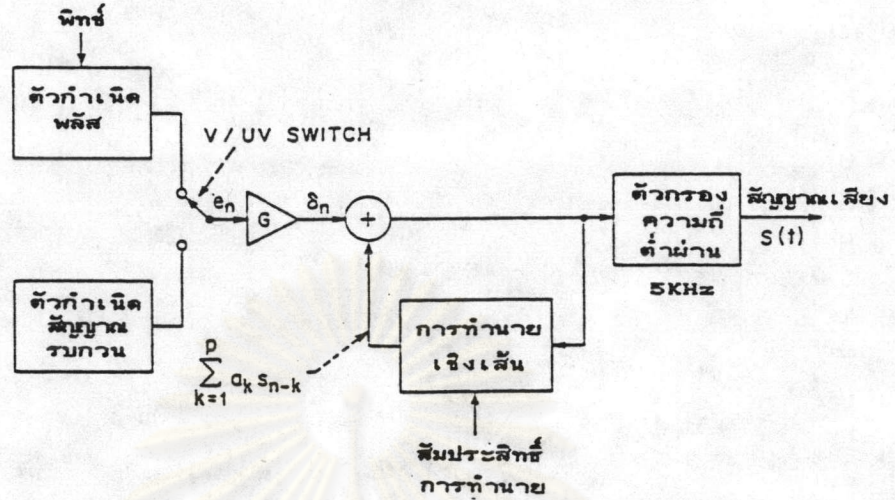
3) การคำนวณค่าแบบรีเคอร์ซีฟ

โดยใช้สมการที่ 2.4.80 , 2.4.83 , 2.4.84 , 2.4.85 ทำการคำนวณหาค่าสัมประสิทธิ์พาร์คอร์ k_m , $m=1,2,\dots,M$ ตลอดจนสัมประสิทธิ์ของการทำนาย a_i เมื่อ $i=-1,\dots,M$ โดยใช้การคำนวณวนรอบทั้งหมด $M-1$ รอบ โดยเริ่มจาก $m=1,2,\dots,M-1$ และได้ค่าผิดพลาดยกกำลังสองรวม α_m ออกมาด้วย ซึ่งจะนำไปใช้ในการสังเคราะห์เสียงต่อไป

จากสรุปขั้นตอนรวมทั้งสามขั้นตอนนำไปเขียนเป็นตัวโปรแกรมภาษาซีใช้ในการคำนวณเพื่อทำการวิเคราะห์สัญญาณเสียงด้วยวิธีอโตคอร์รีเลชันกับสัญญาณเสียงพูดที่ทำการสุ่มเข้ามาได้ดังรายละเอียดในบทที่ 3

2.5 การสังเคราะห์เสียงพูดโดยใช้เทคนิคการทำนายแบบเชิงเส้น

การสังเคราะห์เสียงพูดโดยใช้เทคนิคการทำนายแบบเชิงเส้นนี้ ก็คือการนำเอาพารามิเตอร์ที่ได้มาจากขั้นตอนการวิเคราะห์ในหัวข้อ 2.4 มาทำการสังเคราะห์กลับคืนออกมาเป็นเสียงพูดเช่นเดิม จากรูปแบบจำลองในการสร้างเสียงโดยใช้เทคนิคการทำนายแบบเชิงเส้นในหัวข้อที่แล้วสามารถเขียนให้เห็นเป็นแผนภาพได้ดังนี้



รูปที่ 2.5.1 แผนภาพของการสังเคราะห์เสียงโดยใช้เทคนิคการทำนายเชิงเส้น (Atal and Hanauer, 1971)

จากภาพจะเห็นได้ว่าพารามิเตอร์ที่ใช้ในการควบคุมการสังเคราะห์เสียงโดยใช้เทคนิคการทำนายเชิงเส้นก็มี คาบของพิทช์ (ในกรณีของเสียงก้อง) ตัวชี้ที่จะบอกว่าเป็นเสียงก้องหรือไม่ก้อง ค่าแอมพลิจูดของตัวอย่างสัญญาณ และชุดของสัมประสิทธิ์การทำนาย ตัวกำเนิดพัลส์ (pulse generator) จะทำหน้าที่สร้างพัลส์ที่มีแอมพลิจูดขนาดหนึ่งหน่วยขึ้นมาหนึ่งลูกทุกๆ จุดเริ่มต้นตามคาบของพิทช์ ส่วนตัวกำเนิดสัญญาณรบกวน (white-noise generator) ก็จะช่วยสร้างสัญญาณสุ่มที่กระจัดกระจายไม่มีรูปแบบที่แน่นอนขึ้นมา โดยมีส่วนเบี่ยงเบนมาตรฐานในทุกๆ ตัวอย่างเท่ากันหนึ่ง ตัวเลือกที่จะใช้ต้นกำเนิดเสียง (excitation source) แบบไหนนั้นก็ขึ้นอยู่กับตัวชี้ว่าเป็นเสียงก้องหรือไม่ ถ้าเป็นเสียงก้องก็จะเลือกต้นกำเนิดเสียงเป็นตัวกำเนิดพัลส์โดยมีคาบของพิทช์เป็นตัวควบคุม ส่วนถ้าเป็นเสียงไม่ก้องก็จะเลือกต้นกำเนิดเสียงเป็นตัวกำเนิดสัญญาณรบกวนแทนตามลักษณะของการเกิดเสียงตามธรรมชาติ ค่าแอมพลิจูดของต้นกำเนิดเสียงจะมีการปรับโดยการกำหนดค่า G จากนั้นก็就会有การนำเอาสัญญาณที่ได้จากการทำนายคือ

$$s(n) = \sum_{k=1}^M a_k s_{n-k}$$

ซึ่งมีการแปรตามค่าสัมประสิทธิ์ของการทำนาย หรือคือการจำลองแบบการแปรเปลี่ยนของช่องเสียงมารวมเข้ากับสัญญาณกระตุ้น δ_n เพื่อจัดเป็นส่วนของการสังเคราะห์สัญญาณเสียงตามอันดับของตัวกรอง แล้วสุดท้ายสัญญาณเสียงที่ได้ออกมา ก็จะนำไปผ่านตัวกรองแบบต่ำผ่านที่มีความถี่ประมาณ 5 kHz เพื่อจัดเรียงให้สัญญาณเสียงที่สังเคราะห์ออกมาเป็นรูปคลื่นของสัญญาณอนาล็อกอย่างราบเรียบและต่อเนื่อง การสังเคราะห์เสียงนี้จะมีการควบคุมเปลี่ยนแปลงค่าพารามิเตอร์ต่างๆเป็นค่าใหม่ทุกๆครั้งเมื่อครบคาบของนิพจน์ในกรณีเสียงก้อง หรือทุกๆ 10 มิลลิวินาทีสำหรับเสียงไม่ก้อง

จากรูปแบบข้างต้น เราสามารถแบ่งพิจารณาออกเป็นสองส่วนใหญ่ว่าได้คือ ส่วนของต้นกำเนิดกระตุ้นทั้งเสียงก้องและเสียงไม่ก้องมาคูณกับอัตราขยาย G กับส่วนของแบบจำลองของช่องเสียง (vocal tract) ที่ทำหน้าที่กรองสัญญาณกระตุ้นที่เป็นสัญญาณอินพุตเข้ามาให้ออกไปเป็นสัญญาณเสียงโดยใช้สัมประสิทธิ์การทำนายที่หามาได้จากขั้นตอนการวิจัยใช้ในการจำลองรูปแบบของช่องเสียงในช่วงเวลาหนึ่งๆ หรืออาจจะใช้สัมประสิทธิ์การสะท้อนแทนเมื่อมองช่องเสียงเป็นแบบจำลองของท่อสั้นๆที่มีขนาดพื้นที่ต่างๆมาเรียงต่อกันก็ได้ ซึ่งความสัมพันธ์ระหว่างพื้นที่หน้าตัดของท่อสั้นๆตามแบบจำลองทางด้านวิชากลศาสตร์ (Acoustic tube model) กับสัมประสิทธิ์การสะท้อนกลับหรือสัมประสิทธิ์พาร์คอร์เป็นดังนี้

$$k_m = \frac{A_m - A_{m-1}}{A_m + A_{m-1}} \quad \dots \dots \dots 2.5.1$$

... เมื่อ A_m คือพื้นที่หน้าตัดของท่อลำดับที่ m ...

2.5.1 ต้นกำเนิดเสียงและอัตราขยายในการสังเคราะห์

จากหัวข้อ 2.1 มนุษย์สามารถกำเนิดเสียงที่เป็นเสียงก้องออกมาได้โดยการอาศัยการสั่นของเส้นเสียงเป็นตัวต้นกำเนิดเสียงทำให้เกิดสัญญาณที่เป็นคาบตามจังหวะการสั่นของเส้นเสียงผ่านช่องเสียงที่แปรเปลี่ยนรูปร่างตลอดเวลาเพื่อทำหน้าที่เป็นตัวกรองเอาเฉพาะสัญญาณความถี่ที่ต้องการให้ผ่านออกมาที่ช่องปากเกิดเป็นเสียงตามที่ต้องการ แต่สำหรับเสียงไม่ก้องแล้วต้นกำเนิดกระตุ้นไม่ได้เกิดมาจากเส้นเสียงแต่เกิดมาจากการบังคับให้เกิดแรงดันภายในช่องปากเป็นผลลงลมเคลื่อนที่ผ่านอวัยวะต่างๆภายในช่องปากที่ควบคุมให้เกิดเป็นเสียงแบบไม่ก้องขึ้นมาดังที่ได้กล่าวไปแล้ว

จากเรื่องแบบจำลองการวิเคราะห์และการสังเคราะห์สัญญาณเสียงโดยใช้เทคนิคการทำนายแบบเชิงเส้นนั้นจะได้ว่าเอาทั้งหมดที่ออกมาจากส่วนกลับของตัวกรองในขั้นของการ

วิเคราะห์สัญญาณเสียงนั้นถ้าเรามองในด้านย้อนกลับ ให้สัญญาณเอาท์พุทไปเป็นสัญญาณอินพุทหรือ ไดรฟ์ฟังก์ชัน (driving function) ป้อนเข้าไปยังตัวกรองสังเคราะห์ก็จะได้สัญญาณเสียงออกมา จากหลักการอันนี้เราสามารถนำมาใช้คำนวณหาไดรฟ์ฟังก์ชันด้วยวิธีอโตคอร์รีเลชัน โดยยึดค่าของอัตราขยาย (gain) เป็นหลัก วิธีการอันนี้เสนอขึ้นมาโดย Markel และ Gray เมื่อประมาณปี ค.ศ.1974

สมมติให้ข้อมูลของเสียงมีจำนวน N ตัวอย่างต่อหนึ่งเฟรมวิเคราะห์ และค่าพิณลาดยกกำลังสองรวม (α) มีค่าเป็นอัตราขยายยกกำลังสอง (σ^2) แล้วจะทำให้

1) ในกรณีของเสียงไม่ก้อง ให้มีต้นกำเนิดเสียงเป็นต้นกำเนิดแบบสุ่มเทียม (psudo-random number generator) หรือต้นกำเนิดสัญญาณรบกวนที่สุ่มมาด้วยฟังก์ชัน $g(n)$ และมีอัตราขยายแปรตามต้นกำเนิดเสียงนี้เป็น σ_g^2 แล้ว จากหลักการสมมูลของพลังงานในตัวอย่าง N ตัวอย่างซึ่งมีไดรฟ์ฟังก์ชันเป็น $e(n)$ จะหาได้ดังนี้

$$Ne^2(n) = g^2(n)\sigma_g^2/\sigma^2$$

หรือ $e(n) = g(n)\sigma_g/(\sigma N^{1/2})$ 2.5.2

ถ้าหากค่าจากการสุ่มมีการกระจายอย่างสม่ำเสมอในช่วง $[-b$ ถึง $b]$ โดยแล้วจะได้ค่า $\sigma_g = b/3^{1/2}$ ดังนั้น

$$e(n) = g(n) \frac{\sigma_g}{b} (3/N)^{1/2} \quad \text{สำหรับเสียงไม่ก้อง} \quad \dots\dots\dots 2.5.3$$

2) ในกรณีของเสียงก้อง เอาท์พุทจากส่วนกลับของตัวกรองจะเป็นลำดับของขบวนพัลส์ที่มีขนาดหนึ่งหน่วย (unit pulse train) และมีคาบระหว่างพัลส์โดยคิดเทียบเป็นจำนวนครั้งของการสุ่มเท่ากับ I ครั้ง จะได้ $e(n)$ เป็น

$$e^2(n) \delta_{n,mI} N/I = \sigma^2 \quad \text{โดย } m=0, 1, 2, \dots, \quad \dots\dots\dots 2.5.4$$

หรือ

$$e(n) = \begin{cases} \sigma(I/N)^{1/2} & n = 0, I, 2I, 3I, \dots, \\ 0 & \text{else where} \end{cases} \quad \dots\dots\dots 2.5.5$$

ซึ่งจะเห็นได้ว่าในกรณีที่มีค่าเฉลี่ยของสัญญาณกระตุ่มมีค่าไม่เป็นศูนย์ซึ่งจะมีผลทำให้มีการเพิ่มการไปอัสในขั้นตอนที่จะเข้าสู่ตัวกรองสุดท้ายก่อนที่จะออกมาเป็นสัญญาณเสียงได้ มันจะทำให้สัญญาณเสียงสังเคราะห์ที่ออกมา มีการเพิ่มขึ้นของความถี่มูลฐาน ผลก็คือจะมีเสียงความถี่ต่ำรบกวนออกมา ดังนั้นสามารถแก้ไขได้โดยการจัดให้ค่าเฉลี่ยของสัญญาณกระตุ่มมีค่าเป็นศูนย์ ดังนี้

$$e(n) = \begin{cases} \sigma(I/N)^{1/2} & n = 0, I, 2I, 3I, \dots, \\ -\sigma(I/N)^{1/2} / (I-1) & n \neq 0, I, 2I, 3I, \dots, \end{cases}$$

สำหรับเสียงก้อง 2.5.6

2.5.2 โครงสร้างของการสังเคราะห์เสียงพูด

โครงสร้างของการสังเคราะห์เสียงพูดในที่นี้จะกล่าวในแง่ของค่าสัมประสิทธิ์การสะท้อนกลับของรูปแบบท่อนทางวิชาการสัทศาสตร์ ซึ่งก็คือค่าของสัมประสิทธิ์พาร์คอร์ที่เราได้กล่าวมาแล้วในบทที่ 2.4.3 นั้นเอง ลักษณะของโครงสร้างนี้นับว่าเป็นแนวทางที่ยังใหม่อยู่สำหรับด้านการประมวลผลสัญญาณดิจิทัล (Digital Signal Processing) ซึ่งเป็นผลที่ได้มาจากนักวิจัยทางด้านเสียงพูดชื่อดังคือ Itakura และ Saito กับ Gray และ Markel ในช่วงปี ค.ศ. 1971-1975 โดยเฉพาะ โครงสร้างนี้สามารถแสดงให้เห็นในรูปแบบทั่วไปของตัวกรองแบบดิจิทัล (digital filter) ในรูปแบบดังนี้

$$G(z) = P(z)/A(z) \quad \dots\dots\dots 2.5.7$$

โดยที่ตัวเศษคือ P(z) และตัวส่วนคือ A(z) เป็นชุดของโพลีโนเมียลทั้งคู่มี

รูปแบบเป็น

$$P(z) = P_M(z) = \sum_{m=0}^M p_{mM} z^{-m} \quad \dots\dots\dots 2.5.8$$

$$\text{และ } A(z) = A_M(z) = \sum_{m=0}^M a_{Mm} z^{-m} \quad \dots\dots\dots 2.5.9$$

โดยที่ $a_{M0}=1$

ถ้า $P(z) = 1$ แล้ว ตัวกรอง $G(z)$ ก็จะอยู่ในรูปของ $1/A(z)$ ซึ่งก็คือรูปแบบของตัวกรองสังเคราะห์แบบออลโพล ซึ่งจะพบเห็นได้ว่าการใช้กันอย่างแพร่หลายในด้านของการสังเคราะห์เสียงพูด

จุดสำคัญที่สุดของ โครงสร้างตัวกรองในการสังเคราะห์เสียงพูดก็คือความเสถียรของตัวกรอง (stability) ความเสถียรของตัวกรองใดๆก็ตามที่มีรูปแบบตามสมการที่ 2.5.7 หรือ $G(z) = 1/A(z)$ สามารถดูได้จากราก (root) ของโพลีโนเมียลของตัวส่วนคือ $A(z)$ เมื่อค่าสัมประสิทธิ์ของตัวกรองเป็นค่าคงที่ ถ้ารากของมันเมื่อพล็อตในระนาบซัด (z-plane) แบบวงกลมขนาดหนึ่งหน่วย (คือ $|z| = 1$) อยู่บนหรือหลุดออกไปจากระนาบวงกลมซัดนี้แล้วจะถือว่าตัวกรองนั้นไม่เสถียร นั่นคือมันจะเกิดการแกว่ง (oscillation) อย่างคงที่ตลอดไปหรืออาจจะแกว่งเพิ่มขึ้นแบบทวีคูณที่อัตราหนึ่งหน่วย (Jury, 1964) สำหรับทางด้าน การสังเคราะห์เสียงแล้ว ตามปกติค่าสัมประสิทธิ์ของตัวกรองจะต้องมีการเปลี่ยนแปลงใหม่ทุกๆช่วงเวลาในหน่วยของมิลลิวินาที ทำให้การติดตามพิจารณาค่าความเสถียรของตัวกรองที่แปรตามเวลาทำได้ยากมาก แต่ก็ยังนับว่าโชคดีที่ค่าสัมประสิทธิ์ที่แปรเปลี่ยนในการสังเคราะห์เสียงนั้นตามปกติจะไม่ค่อยเกิดปัญหาการไม่เสถียรขึ้นเมื่อ เราสมมติให้ค่าสัมประสิทธิ์นั้นเป็นค่าคงที่

2.5.3 โครงสร้างของการสังเคราะห์เสียงพูดแบบแลททิส

วรรณกรรมทางด้าน การประมวลผลสัญญาณทางดิจิทัลต่างๆ เช่นของ Rabiner และ Rader เมื่อปี ค.ศ.1972 , Oppenheim และ Schaffer เมื่อปี ค.ศ.1975 , Rabiner และ Gold เมื่อปี ค.ศ. 1975 เป็นต้น ได้อ้างอิงถึงโครงสร้างของตัวกรองในรูปแบบของตัวกรองเชิงเส้นธรรมดาแบบต่างๆไปในรูปของ $P(z)/A(z)$ แต่สำหรับ Itakura และ Saito กับ Gray และ Markel ในช่วงปี ค.ศ. 1973-1975 ได้เสนอโครงสร้างแบบใหม่ขึ้นมาในการนิยาม $P(z)/A(z)$ โครงสร้างแนวใหม่ที่มีความสำคัญในการทำการสังเคราะห์เสียงดังที่จะกล่าวต่อไปเนื่องจากเหตุผลสองประการ คือ ประการแรก มันสามารถจะนำเอาสัมประสิทธิ์การสะท้อน k_n มาใช้เป็นสัมประสิทธิ์ของตัวกรองได้โดยตรง ประการที่สองในการนำเอาไปประยุกต์ลงในระบบคอมพิวเตอร์จะสามารถสนับสนุนในด้านความแม่นยำ จำนวนครั้งของ

การคูณและการบวกในแต่ละส่วน และความยุ่งยากในการคำนวณให้หมดสิ้นไป โครงสร้างตัวกรองสังเคราะห์เสียงพูดนั้น ได้ถูกพัฒนาขึ้นมาหลายรูปแบบหลายลักษณะด้วยกัน เช่น โครงสร้างแบบตรง (direct form) โครงสร้างแลททิสแบบตัวคูณตัวเดียว (one-multiplier) โครงสร้างแลททิสแบบตัวคูณสองตัว (two-multiplier) โครงสร้างแบบตัวคูณสี่ตัว (four-multiplier) เป็นต้น แต่ทั้งนี้พื้นฐานของโครงสร้างแต่ละรูปแบบก็มาจากนิยามอันเดียวกันเพียงแต่จะกำหนดตัวพารามิเตอร์ของแต่ละภาคของโครงสร้างแตกต่างกันไป ในงานวิจัยขั้นนี้ จะเลือกเอาโครงสร้างแลททิสแบบตัวคูณสองตัวมาใช้ในการสังเคราะห์เสียงพูด โดยใช้สัมประสิทธิ์การสะท้อนโดยตรง

จากเรื่องของการหาสัมประสิทธิ์พาร์คอร์และโครงสร้างแบบแลททิส สมการที่ 2.4.53 และสมการที่ 2.4.65 สามารถแสดงได้ว่าส่วนกลับของตัวกรองคือ $A(z)$ จากวิธีออกโตคอร์รีเลชันสามารถหาได้โดยการใช้อนุกรมแบร์เคอร์ที่ดังนี้

$$A_m(z) = A_{m-1}(z) + k_m B_{m-1}(z) \quad \dots\dots\dots 2.5.10$$

$$zB_m(z) = k_m A_{m-1}(z) + B_{m-1}(z) \quad \dots\dots\dots 2.5.11$$

เมื่อ $m=1, 2, \dots, M$ และมีค่าเริ่มต้นเป็น $A_0(z) = 1$ และ $zB_0 = 1$ ถ้า $E(z)$ คือสัญญาณอินพุตเข้าสู่ตัวกรองการสังเคราะห์ และ $X(z)$ คือเอาต์พุตที่ออกจากตัวกรองการสังเคราะห์นั้นคือ

$$X(z) = G(z)E(z) \quad \dots\dots\dots 2.5.12$$

ถ้าคูณสมการ 2.5.10 และ 2.5.11 ด้วย $E(z)/A_m(z) = E(z)/A(z)$ จะได้

$$\frac{A_m(z)E(z)}{A(z)} = \frac{A_{m-1}(z)E(z)}{A(z)} + \frac{k_m B_{m-1}(z)E(z)}{A(z)} \quad \dots\dots\dots 2.5.13$$

$$\frac{zB_m(z)E(z)}{A(z)} = \frac{k_m A_{m-1}(z)E(z)}{A(z)} + \frac{B_{m-1}(z)E(z)}{A(z)} \quad \dots\dots\dots 2.5.14$$

และ
$$\frac{zB_0(z)E(z)}{A(z)} = \frac{A_0(z)E(z)}{A(z)} = \frac{E(z)}{A(z)} \quad \dots\dots\dots 2.5.15$$

จุดประสงค์ของการคูณนี้ก็นำค่า $A_0(z) = 1$ มาใช้ให้เกิดความสัมพันธ์กับเอาท์พุท $X(z) = E(z)/A(z)$ ตามสมการที่ 2.5.15 ถ้าให้ค่า $m = M$ ก็จะได้สัญญาณอินพุทนิยามดังนี้

$$\frac{A_M(z)E(z)}{A(z)} = E(z) \quad \dots\dots\dots 2.5.16$$

โดยที่ $E(z)$ เป็นสัญญาณอินพุทที่ $m=M$ จนกระทั่ง $m=0$ บล็อกไดอะแกรมแสดงสมการตั้งแต่ที่ 2.5.13 ถึง 2.5.16 โดยแบ่งเป็นภาคๆ ตั้งแต่ภาคต้นที่อินพุทเมื่อ $m=M$ ภาคกลางๆที่จุด m ใดๆ และภาคท้ายเมื่อ $m=0$ ได้ดังรูปที่ 2.5.2 ซึ่งจะเรียกเป็น ภาค M , ภาค m , และภาค 1 ตามลำดับ โดยที่ m นั้นหมายถึงดัชนีของโพลีโนเมียล $A_m(z)$ หรือ $B_m(z)$ ที่ปลายทางด้านซ้ายของมัน สำหรับส่วนเริ่มต้นคือส่วนที่ $m=M$ จะสามารถเขียนสมการได้ใหม่จากสมการที่ 2.5.13 เป็น

$$\frac{A_{m-1}(z)E(z)}{A(z)} = \frac{A_m(z)E(z)}{A(z)} - \frac{k_{m-m-1} B_{m-m-1}(z)E(z)}{A(z)} \quad \dots\dots\dots 2.5.17$$

โครงสร้างส่วนต่างๆตั้งแต่ $m=M$ จนถึง $m=1$ จะเป็นภาคนิยามภาคต่างๆของตัวกรองแบบออลโพล $1/A(z)$

โพลีโนเมียลตัวเศษคือ $P(z)$ สามารถเขียนในเทอมของออลโพลีโนเมียลคือ $zB_m(z)$ และ $1/A(z)$ ได้ดังนี้

$$P(z) = P_M(z) = \sum_{m=0}^M v_m zB_m(z) \quad \dots\dots\dots 2.5.18$$

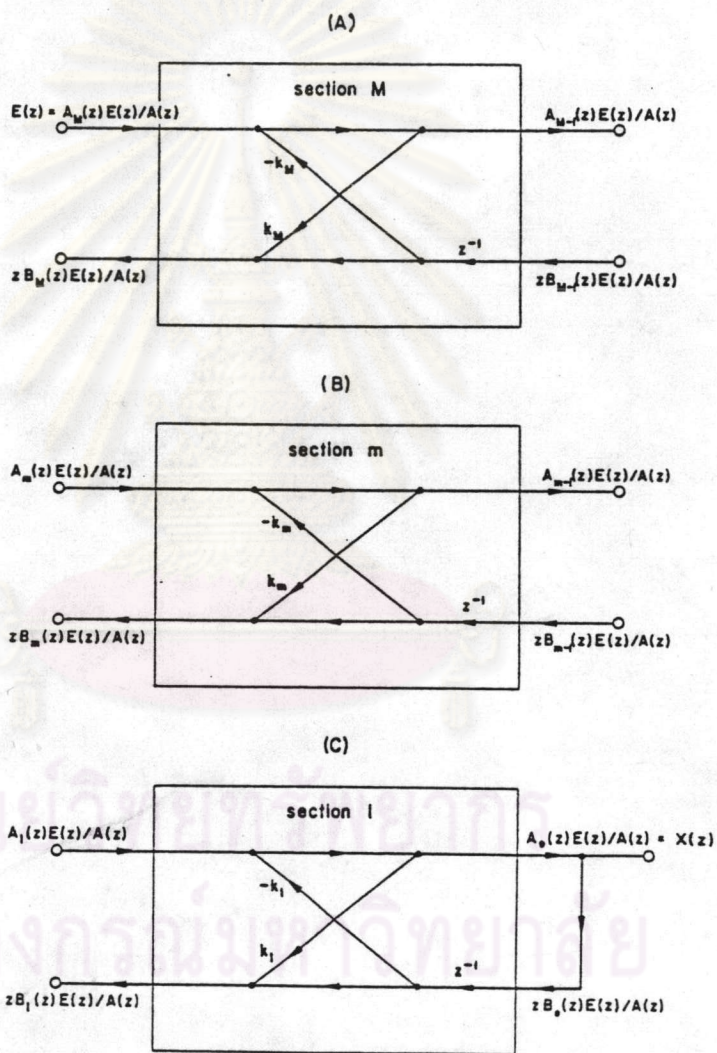
ดังนั้น

$$G(z) = P(z)/A(z) = \sum_{m=0}^M v_m [zB_m(z)/A(z)] \quad \dots\dots\dots 2.5.19$$

โดยที่ v_m คือแทพพารามิเตอร์ของแต่ละภาค m ใดๆ ดังนั้นผลลัพธ์ $X(z)$ ก็จะขึ้นตรงกับค่าของสัมประสิทธิ์การสะท้อนและแทพพารามิเตอร์ดังนี้

$$X(z) = E(z)P(z)/A(z)$$

$$= \sum_{m=0}^M v_m [zB_m(z)E(z)/A(z)] \dots\dots\dots 2.5.20$$



รูปที่ 2.5.2 ภาพแสดงการไหลของสัญญาณของตัวกรอง
 (A) ภาคที่ M (B) ภาคที่ m ไฉๆ (C) ภาคที่ 1

ดังนั้นจากชุดสมการที่ 2.5.17, 2.5.14, 2.5.15 จะนิยามสัญลักษณ์ใหม่เพื่อให้สมการระดับชั้นดังนี้

$$E_m^+(z) = \frac{A_m(z)E(z)}{A(z)} \quad \langle \text{-----} \rangle \quad e_m^+(n) \quad \dots\dots\dots 2.5.21$$

$$E_m^-(z) = \frac{B_m(z)E(z)}{A(z)} \quad \langle \text{-----} \rangle \quad e_m^-(n) \quad \dots\dots\dots 2.5.22$$

เพราะฉะนั้นภาคต่างๆของตัวกรองซึ่งมีรูปแบบเป็นโครงสร้างแลททิสมแบบตัวคูณสองตัวสามารถเขียนแทนได้ด้วยสมการดังนี้

$$E_{m-1}^+(z) = E_m^+(z) - k_m E_{m-1}^-(z) \quad \dots\dots\dots 2.5.23$$

$$zE_m^-(z) = k_m E_{m-1}^+(z) + E_{m-1}^-(z) \quad \dots\dots\dots 2.5.24$$

เมื่อ $m=M, M-1, \dots, 1$

โดยมีสัญญาณอินพุตที่ภาคแรก $m=M$ เป็น

$$E(z) = E_M^+(z) \quad \dots\dots\dots 2.5.25$$

ส่วนสัญญาณเอาต์พุตที่ออกจากตัวกรองแบบออล โพลคือ $1/A(z)$ ที่ $m=0$ จะเป็น

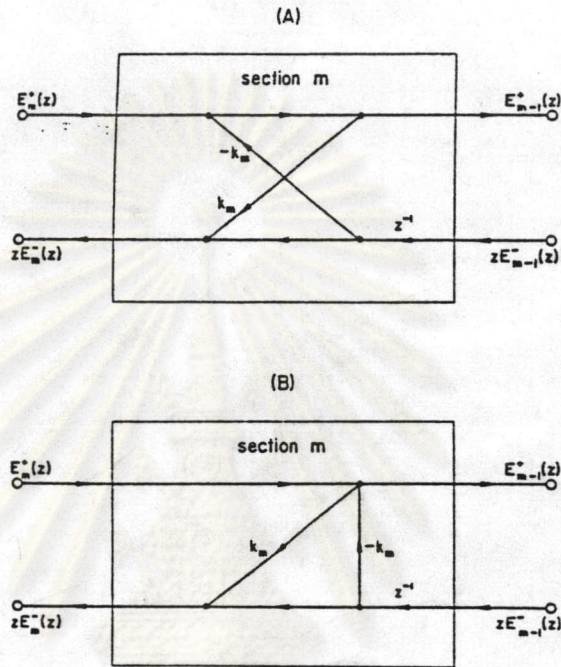
$$zE_0^-(z) = E_0^+(z) \quad \dots\dots\dots 2.5.26$$

หรือถ้าเขียนในรูปของ $X(z)$ จากสมการที่ 2.5.20 ก็จะได้เป็น

$$\begin{aligned} X(z) &= G(z)E(z) \\ &= \sum_{m=0}^M v_m z \frac{B_m(z)E(z)}{A(z)} \\ &= \sum_{m=0}^M v_m z E_m^-(z) \quad \dots\dots\dots 2.5.27 \end{aligned}$$

นั่นคือเอาที่พิกัดที่ออกจากตัวกรองแบบออล โพล ในภาคท้ายสุดก็คือ

$$X(z) = E^+_o(z) \quad \text{เมื่อ } G(z) = 1/A(z) \quad \dots\dots\dots 2.5.28$$



รูปที่ 2.5.3 แสดงกราฟการไหลของสัญญาณของ โครงสร้างแลททิสแบบตัวคูณสองตัว

จากรูปข้างบนแสดงกราฟการไหลของสัญญาณของภาคที่ m โดยของ โครงสร้างแลททิสแบบตัวคูณสองตัว โดยรูปบนแสดงในลักษณะที่มีการพาดผ่านของสัญญาณในลักษณะไขว้กันเป็นตาราง ส่วนรูปล่างแสดงกราฟซึ่งเขียนขึ้นมาใหม่โดยไม่ให้มีลักษณะของการพาดผ่านอยู่ จะเห็นได้ว่าในภาคที่ m โดยนี้จะมีจุดที่มีการคูณกันของสัญญาณอยู่สองจุด จุดที่มีการบวกกันอยู่สองจุดและส่วนของการหน่วงสัญญาณอีกหนึ่งส่วน