



การวิเคราะห์เสียงพูด

ในการวิเคราะห์เสียงพูด สิ่งที่สำคัญ คือการศึกษาวิเคราะห์เสียงพูดจริง เพื่อหาคุณสมบัติ หรือตัวแทน (Signal Representation) ของเสียง ตามหลักการต่าง ๆ ซึ่งมีรายละเอียด ดังจะกล่าวต่อไป

2.1 เสียงพูด

คนเราเปล่งเสียงพูดด้วยอวัยวะที่ใช้ในการออกเสียง (Organs of Speech) ทำเสียงตามที่มีอยู่ในระบบภาษาของตน แม้ว่าคนที่อยู่ในสังคมเดียวกัน จะใช้ภาษาเดียวกัน แต่ถ้าพิจารณาเสียงที่เปล่งออกมาจริง ๆ แต่ละครั้งก็อาจจะสังเกตเห็นลักษณะที่แตกต่างกันได้ เราจึงสามารถจำเสียง จำวิธีพูดของคนที่เราคุ้นเคยได้ เสียงพูดนี้มีลักษณะที่จะอธิบายได้ด้วยหลักเกณฑ์ทางวิทยาศาสตร์ ซึ่งเป็นสากล แม้ว่าภาษาหนึ่ง ๆ จะมีเสียงแตกต่างกันไปมากบ้างน้อยบ้าง แต่แต่ละเสียงก็สามารถจะนำมาพิจารณา และอธิบายให้รู้ลักษณะการออกเสียง และตำแหน่งที่เกิดของเสียงได้ คำอธิบายนี้ จะทำให้เข้าใจลักษณะเสียงทุกเสียง วิชาที่ว่าด้วยเสียงพูด เรียกว่า วิชาสัทศาสตร์ (Phonetics) (อดม วโรตม์สิริจิตต์ธัม, 2527)

ในการศึกษาเรื่องเสียงพูด แบ่งได้เป็น 2 ลักษณะ คือ (กาญจนา นาคสกุล, 2524)

ก. สรีรศาสตร์ (Articulatory Phonetics) เป็นการศึกษาเสียงพูดจากอวัยวะ และการเคลื่อนไหวของอวัยวะที่ทำให้เกิดเสียงพูด การอธิบายก็จะอธิบายโดยอาศัยลักษณะและอาการเคลื่อนไหวของอวัยวะที่เกี่ยวข้องในการเปล่งเสียงพูดนั้น

ข. กลศาสตร์ (Acoustic Phonetics) เป็นการศึกษาเสียงพูดจากลักษณะคลื่นเสียงที่แผ่เปล่งออกมาแล้ว และผู้ฟังได้ยิน ว่ามีลักษณะทางกลศาสตร์เป็นอย่างไร การศึกษาตามแนวนี้ ต้องอาศัยความรู้ทางฟิสิกส์ และคณิตศาสตร์ เข้าช่วยอธิบายลักษณะของคลื่นเสียง

2.1.1 อวัยวะที่ใช้ในการออกเสียง (Organs of Speech)

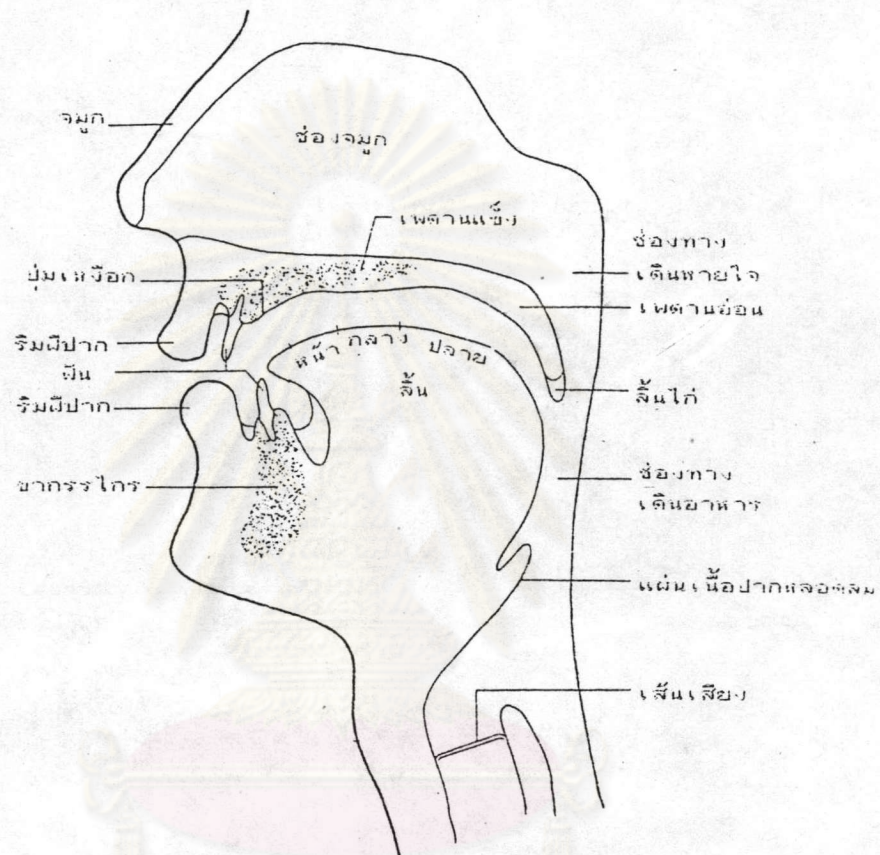
(กาญจนา นาคสกุล, 2524)

อวัยวะที่ใช้ในการออกเสียงพูดมีอยู่หลายส่วน แต่ละส่วนสามารถ ทำให้เสียงพูดแตกต่างกันไปได้ อวัยวะเหล่านี้มีปาก และส่วนต่าง ๆ ในปาก ช่องคอ กล้องเสียง และช่องปาก และช่องจมูก ดังรายละเอียดในรูปที่ 2.1.1 อวัยวะที่ใช้ในการออกเสียงแบ่งได้เป็น 2 พวก คือ



ก. อวัยวะที่ใช้ในการกระทำอาการ (Articulator) หมายถึง อวัยวะส่วนที่เคลื่อนไหว เพื่อผลิตลม ไปยังส่วนต่าง ๆ อวัยวะตัวกระทำอาการที่สำคัญ คือ ลิ้น ซึ่งเป็นส่วนที่เคลื่อนไหวได้มากที่สุด

ข. อวัยวะที่เป็นตำแหน่งที่เกิดเสียงต่าง ๆ (Point of Articulator) หมายถึงตำแหน่ง หรือฐานกรณ์ที่เกิดของเสียงต่าง ๆ เช่น ริมฝีปาก ฟัน เพดานส่วนต่าง ๆ เป็นต้น



รูปที่ 2.1.1 แสดงอวัยวะที่ใช้ในการออกเสียงพูด  
(Parson T.W., 1986)

อวัยวะส่วนที่มีหน้าที่โดยตรงในการออกเสียงพูดมี ดังนี้

ก. ริมฝีปาก เป็นอวัยวะส่วนที่สามารถเคลื่อนไหวได้มาก และทำให้เสียงแตกต่างกันได้มาก เราอาจบังคับริมฝีปากให้อยู่ชิดกัน ห่างกัน ชื่นออกมา หรือห่อกลม ฯลฯ ก็ได้ ลักษณะริมฝีปากต่าง ๆ นี้ล้วนแต่มีอิทธิพลต่อการออกเสียง และการทำให้เสียงแตกต่างกันไปทั้งสิ้น

ข. ฟัน เป็นอวัยวะที่เกิดของเสียงหลายชนิด เช่น เมื่อฟันบนกดลงบนริมฝีปากล่าง หรือกดกับฟันล่าง ลมที่ผ่านออกมาโดยแรง จะลอดช่องที่ฟันผ่านได้ออกมา ทำให้เกิดเป็นเสียงชนิดที่เรียกว่า เสียงเสียดแทรกที่เกิดที่ฟัน เป็นต้น



ค. ปุ่มเหงือก เป็นส่วนนูนออกมาอยู่หลังฟันด้านบน ถ้าเอาลิ้นแตะดูจะรู้สึกว่ามีลักษณะเป็นคลื่น ปุ่มเหงือกเป็นบริเวณที่เกิดเสียงปุ่มเหงือก (Alveolar Sound) ปุ่มเหงือกนั้นเป็นตำแหน่งสำคัญในการอธิบายเรื่องเสียงตำแหน่งหนึ่ง

ง. เพดานแข็ง หรือ เพดานปาก คือส่วนเฉพาะเพดานที่โค้งเป็นกระดูกแข็ง

จ. เพดานอ่อน คือส่วนของเพดานที่อยู่ต่อเพดานแข็งไปข้างใน มีลักษณะเป็นกระดูกอ่อนที่ขยับขึ้นลงได้ เวลาหายใจเพดานอ่อน และลิ้นไก่ซึ่งอยู่ปลายเพดานอ่อนจะลดระดับลงมาเปิดช่องให้ลมออกไปทางจมูก เวลาพูดส่วนใหญ่ปลายเพดานอ่อน และลิ้นไก่อจะยกยกขึ้นไปจรดกับหลังคอก นอกจากเวลาออกเสียงนาสิกเท่านั้น ที่เพดานอ่อนจะลดระดับลงมา

ฉ. ลิ้นไก่ เป็นก้อนเนื้อเล็ก ๆ อยู่ต่อปลายเพดานอ่อนตรงกลางปาก ลิ้นรัวได้

ช. ลิ้น ลิ้นเป็นส่วนที่เคลื่อนไหวมากที่สุดในการออกเสียงพูด จึงต้องแบ่งออกเป็น 3 ส่วนด้วยกัน ตามหน้าที่ที่มิในการออกเสียง

ปลายลิ้น (Blade of the Tongue) คือลิ้นส่วนปลายซึ่งสามารถจะยกขึ้นไปแตะอวัยวะส่วนต่าง ๆ ในปากตอนบนได้ง่าย

หน้าลิ้น (Front of the Tongue) คือลิ้นส่วนที่อยู่ตรงข้ามกับเพดานแข็ง ถ้าวางลิ้นราบกับปากอย่างในขณะที่ไม่ได้พูด

หลังลิ้น (Back of the Tongue) คือส่วนของลิ้นซึ่งถ้าวางลิ้นราบกับปาก ตามปกติจะอยู่ตรงข้ามกับเพดานอ่อน

ช. แผ่นเนื้อปากหลอดลม (Epiglottis) เป็นก้อนเนื้อเล็กๆ คล้ายลิ้นไก่ อยู่ต่อโคนลิ้นลงไปในคอ มีหน้าที่ปิดช่องลมเมื่อรับประทานอาหาร และเปิดช่องลมเมื่อพูด

ฉ. ช่องคอ (Pharynx) หมายถึงช่องคอซึ่งอยู่ถัดช่องปากลงไปจนถึงเส้นเสียง

ญ. เส้นเสียง (Vocals Cords) เป็นอวัยวะสำคัญที่เกิดของเสียง เส้นเสียงมีลักษณะเป็นกล้ามเนื้อ 2 แผ่นภายในกล่องเสียง ปิดขวางอยู่ปากช่องหลอดลมจากด้านหลังมาด้านหน้าระหว่างเส้นเสียงจะมีช่องว่าง ซึ่งเป็นทางให้ลมผ่านเข้าไปถึงปอด และออกมาจากปอดได้ ช่องนี้เรียกว่า ช่องว่างระหว่างเส้นเสียง (Glottis) เส้นเสียงทั้งสองสามารถจะดึงออกให้ห่างจากกัน หรือดึงเข้ามาชิดกันก็ได้ เส้นเสียงเป็นส่วนสำคัญที่ทำให้เกิดเสียงพูดขึ้น

ฎ. ช่องจมูก (Nasal Cavity) หมายถึงโพรงในช่องจมูก ซึ่งอยู่เหนือลิ้นไก่ขึ้นไป เป็นช่องที่ลมซึ่งผ่านเส้นเสียงขึ้นมาจะผ่านออกไปทางจมูกได้เมื่อเวลาหายใจ และเวลาออกเสียงนาสิก ในเวลาพูดเสียงอื่น ๆ ลิ้นไก่อจะยกยกขึ้นไปปิดช่องจมูกเพื่อให้ลมออกมาทางปาก

เสียงที่เกิดขึ้น ไม่ว่าจะเป็นเสียงประเภทใด จะผ่านไปตามทางเดินของเสียง (Vocal Tract) โดยจะเริ่มตั้งแต่ช่องว่างระหว่างเส้นเสียง ถึงริมฝีปาก ในทางวิทยาศาสตร์สรุปได้ว่า ทางเดินเสียง คือท่อนำเสียงที่มีรูปร่างไม่แน่นอน (Non Uniform Acoustic Tube)



มีความยาวประมาณ 17 เซนติเมตร มีพื้นที่หน้าตัดที่เปลี่ยนแปลงได้ระหว่าง 0 ถึง 20 ตารางเซนติเมตร (Flanagan J.L., 1972)

### 2.1.2 การเกิดของเสียง (Speech Production) (อดัม วโรตม์ลิกคิตต์ม, 2527)

ขั้นตอนการเกิดของเสียงแบ่งได้เป็น 3 ขั้นตอนใหญ่ ๆ คือ

ก. ขั้นเริ่มต้น (Initiation) ขั้นตอนนี้เป็นขั้นตอนที่ลมเริ่มถูกขับออกจากปอด เพื่อเข้าสู่ขั้นตอนที่สองต่อไป

ข. ขั้นตอนดัดแปลงลมที่เส้นเสียง (Phonation) เป็นขั้นตอนที่ลมจากปอดจะผ่านมายังหลอดลม และกล่องเสียง ซึ่งที่กล่องเสียงนี้ เส้นเสียงจะทำหน้าที่เป็นลิ้นเปิดปิด ทำให้เกิดเสียงได้ 2 ชนิด คือ เสียงก้อง (Voiced Sounds) และ เสียงไม่ก้อง (Unvoiced Sounds) อวัยวะที่ใช้ในขั้นตอนนี้ คือ ส่วนที่ต่อจากปอดขึ้นมาจนถึงกล่องเสียง

ค. ขั้นตอนเปลี่ยนแปลงลักษณะเสียง (Articulation) ในขั้นตอนนี้ลมที่ผ่านออกมาจากกล่องเสียง จะถูกแปลงให้เกิดเสียงในลักษณะต่าง ๆ อวัยวะที่ใช้ในขั้นตอนนี้ คือ ส่วนที่ต่อจากกล่องเสียงจนถึงริมฝีปาก

เสียงที่เกิดขึ้นจากขั้นตอนดังกล่าว จะถูกแบ่งออกเป็น 2 ประเภทใหญ่ ๆ คือ

ก. เสียงก้อง (Voiced Sounds) เกิดเนื่องจากการออกเสียงในขณะที่เส้นเสียงถูกดึงเข้ามาใกล้กันจนเกือบปิดช่องทางลมเสียงสนิท ลมที่ดันขึ้นมาจากปอดจะทำให้เส้นเสียงทั้งสองต้องสั่นสะบัด คือ ลมจะพยายามหาทางออก ทำให้เส้นเสียงส่วนที่อยู่ด้านล่างเปิดออก แล้วส่วนกลางก็เปิดออกตาม แล้วเปิดเรื่อยไปจนถึงส่วนบน พอส่วนบนเปิดส่วนล่างก็ปิด สลับกันอยู่ดังนี้ แต่อาการที่เกิดขึ้นนี้เร็วมาก จนเป็นอาการสั่น ลมที่ออกมาไม่สะดวกเพราะต้องบีบตัวผ่านช่องแคบ เป็นจังหวะ ๆ จึงทำให้เกิดเป็นเสียงขึ้น เรียกว่า เสียงก้อง ตัวอย่างของเสียงก้อง ได้แก่ บ น เป็นต้น

ข. เสียงไม่ก้อง (Unvoiced Sounds) เป็นการออกเสียงในขณะที่เส้นเสียงยังเปิดกว้าง โดยเปิดช่องระหว่างเส้นเสียงหรือช่องคอหอย (Glottis) ให้ลมหายใจผ่านเข้าออกสะดวก เราเรียกเสียงเหล่านี้ว่าเสียงไม่ก้อง ตัวอย่างของเสียงไม่ก้อง ได้แก่ พ ท ป ส เป็นต้น

### 2.2 หลักการในการวิเคราะห์เสียงพูด (Speech Analysis)

ในการวิเคราะห์เสียงพูดสามารถแบ่งออกได้เป็น 3 ลักษณะใหญ่ ๆ คือ ในลักษณะแรกจะกล่าวถึง วิธีการเปลี่ยนสัญญาณเสียงพูด หรือสัญญาณอนาล็อกไปเป็นสัญญาณดิจิทัล ซึ่งจะนำไปใช้ในการวิเคราะห์ต่อไป ในลักษณะที่สองจะกล่าวถึงการวิเคราะห์ในโดเมนเวลา ซึ่งเป็นการ



วิเคราะห์โดยตรงกับสัญญาณ และลักษณะสุดท้ายจะเป็นการวิเคราะห์ในโดเมนความถี่

2.2.1 การเข้ารหัสสัญญาณดิจิทัล (Digital Waveform Coding)

การแปลงสัญญาณเสียงพูดให้เป็นสัญญาณดิจิทัล แบ่งได้เป็น 2 วิธี คือ

ก. พัลส์โคดโมดูเลชัน (Pulse Code Modulation) หรือ พีซีเอ็ม (PCM) เป็นการสุ่มตัวอย่างสัญญาณอนาลอกในช่วงเวลาที่คงที่ ย่านความถี่ (Bandwidth) ที่ได้จะเท่ากับครึ่งหนึ่งของความถี่ในการสุ่ม (Sampling Frequency) จากนั้นจะนำสัญญาณที่สุ่มมาเปลี่ยนเป็นข้อมูลเลขฐานสอง ด้วยกระบวนการควอนไทเซชัน (Quantization) ความถี่ในการสุ่มสัญญาณอยู่ระหว่าง 6 ถึง 20 กิโลเฮิรตซ์ ในการพิจารณาความคลาดเคลื่อนของการเข้ารหัสสัญญาณ มักจะใช้ค่า Signal to Noise Ratio (SNR) ของกำลังงานเฉลี่ยของสัญญาณเสียง กับ กำลังงานเฉลี่ยของสัญญาณรบกวนอันเกิดจากความคลาดเคลื่อน โดยคิดค่าเป็นหน่วย เดซิเบล (dB,  $10\log_{10} SNR$ ) (Parson T.W., 1986)

$$SNR = 6.02B - 7.27 \quad (dB) \quad \dots\dots(2.2.1)$$

โดย B คือ จำนวนบิตที่ใช้ในการแทนค่าต่อการสุ่มหนึ่งครั้ง ตัวอย่างเช่น การควอนไทซ์แบบเชิงเส้น ความละเอียด 12 บิต จะได้ค่า SNR เท่ากับ 60 dB และการเปลี่ยนแปลงความละเอียด 1 บิต จะทำให้ค่า SNR เปลี่ยนไป 6 dB การควอนไทซ์ที่ได้ผลดีที่สุด คือ ระบบสเกลล็อก ซึ่งสามารถลดจำนวนบิตลงเหลือ 7-8 บิต ที่คุณภาพใช้งานได้

ข. ดิฟเฟอเรนเชียลควอนไทเซชัน (Differential Quantization)

วิธีการนี้เกิดจากความจริงที่ว่า ในการสุ่มตัวอย่างสัญญาณเสียงพูดตาม Nyquist Rate ค่าสหสัมพันธ์ (Correlation) ของสัญญาณที่ทำการสุ่มต่อเนื่องกันยังคงมีค่าสูงอยู่ นั่นคือเรามีโอกาสทำนายค่าของสัญญาณให้ได้ผลดีโดยไม่ยากนัก เนื่องจากความสามารถในการทำนายได้ของสัญญาณ จากรูปที่ 2.2.1  $x(n)$  คือค่าทำนายของสัญญาณ  $x(n) = x(nT)$  ในที่นี้ T คือเวลาระหว่างการสุ่มตัวอย่าง (Sampling Period) ถ้าการทำนายของตัวทำนาย  $P(z)$  มีความถูกต้องพอสมควร ค่าความแปรปรวน (Variance) ของค่าที่แตกต่างระหว่างสัญญาณจริง กับสัญญาณทำนาย คือ  $d(n) = x(n) - x(n)$  จะมีค่าน้อยกว่าค่าความแปรปรวนของสัญญาณ  $x(n)$  ดังนั้น ค่าความแปรปรวนของค่าผิดพลาดจากการควอนไทซ์ของสัญญาณ  $d(n)$  ย่อมมีค่าน้อยกว่าค่าความแปรปรวนของค่าผิดพลาดจากการควอนไทซ์ของตัวสัญญาณ  $x(n)$

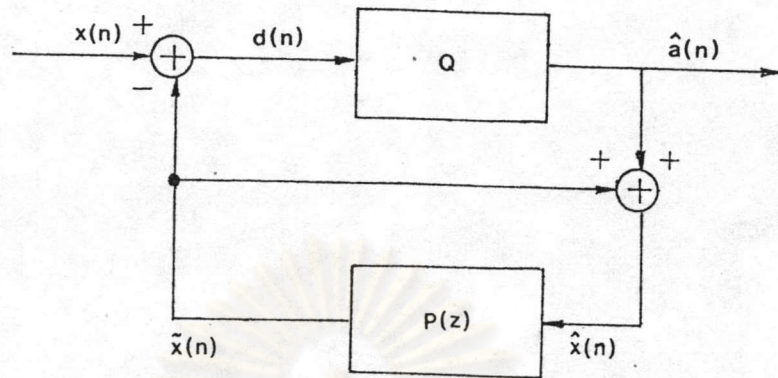
ให้ค่าผิดพลาดจากการควอนไทซ์ของ  $d(n)$  เท่ากับ  $e(n)$  หรือ

$$e(n) = d(n) - d(n) \quad \dots\dots(2.2.1)$$

เนื่องจาก  $d(n) = x(n) - x(n)$  และ  $d(n) = x(n) - x(n)$

ดังนั้น 
$$e(n) = x(n) - x(n) \quad \dots\dots(2.2.2)$$





รูปที่ 2.2.1 แสดงแผนภาพหลักการทำงานของดิฟเฟอเรนเชียลควอนไทเซชัน

นั่นคือ ค่าผิดพลาดจากการควอนไทซ์ของสัญญาณที่ได้  $\{x(n)\}$  จะมีค่าเท่ากับ ค่าผิดพลาดจากการควอนไทซ์ค่าที่แตกต่างกันระหว่างสัญญาณจริงกับสัญญาณทำนาย ผลก็คือในจำนวนบิตเท่ากัน ดิฟเฟอเรนเชียลควอนไทเซชันจะได้ SNR สูงกว่าพีซีเอ็ม ตัวอย่างการใช้งานด้วยหลักการนี้

1. การทำควอนไทเซชันแบบเชิงเส้นกับสัญญาณ  $d(n)$  เรียกว่า Linear Delta Modulation (LDM)
2. ดิฟเฟอเรนเชียลพีซีเอ็ม (DPCM) ซึ่งใช้การควอนไทซ์เพียง 1 บิต
3. อแดปทีฟดิฟเฟอเรนเชียลพีซีเอ็ม (ADPCM) เป็นพีซีเอ็มที่มีการเปลี่ยนแปลง Step Size เพื่อจะทำให้คุณภาพของเสียงดีขึ้น

### 2.2.2 การวิเคราะห์ในโดเมนเวลา (Time Domain Analysis)

หลักการนี้ เป็นการวิเคราะห์หาคคุณสมบัติของเสียงพูดในโดเมนเวลา เนื่องจากเสียงพูดมีคุณสมบัติเปลี่ยนแปลงตามเวลา ในการวิเคราะห์จึงต้องแบ่งเสียงพูดออกเป็นช่วง ๆ (Frame) โดยมีช่วงเวลายู่ระหว่าง 10-30 มิลลิวินาที ในช่วงเวลาดังกล่าว เสียงจะมีการเปลี่ยนแปลงคุณสมบัติน้อยมาก ดังนั้นในแต่ละเฟรมจึงสมมติให้ เสียงไม่เปลี่ยนแปลงตามเวลา ทำให้การวิเคราะห์ทำได้ง่ายยิ่งขึ้น

ตัวอย่างการวิเคราะห์เสียงพูดในโดเมนเวลา มีดังนี้



ก. การวัดพลังงานของสัญญาณ (Energy Measurement)

(Rabiner L.R., 1975)

พลังงานของเสียงใน 1 เฟรม ซึ่งมีจำนวน  $N$  ตัวอย่าง จะมีค่าเท่ากับ

$$E(n) = \sum_{m=0}^{N-1} [w(m)x(n-m)]^2 \quad \dots\dots(2.2.3)$$

โดยที่  $w(m)$  คือ ฟังก์ชันหน้าต่าง (Window Function) แต่อย่างไรก็ตาม สมการ (2.2.3) มีข้อเสีย คือ จะมีความไวต่อสัญญาณที่มีขนาดใหญ่ เนื่องจากการยกกำลังสอง ดังนั้นจึงมีการดัดแปลงใหม่โดยใช้ค่าสัมบูรณ์แทน ดังนี้

$$E(n) = \sum_{m=0}^{N-1} |w(m)x(n-m)| \quad \dots\dots(2.2.4)$$

ข. การวิเคราะห์หัตถสัมพันธ์ในช่วงเวลาสั้น

(Short-Time Autocorrelation Analysis)

(O'Shaughnessy D., 1987)

เป็นวิธีการสำคัญในการวิเคราะห์ความเป็นคาบ (period) ของสัญญาณเสียงพูดสมการของฟังก์ชันหัตถสัมพันธ์ มีดังนี้

$$\phi_1(m) = (1/N) \sum_{n=0}^{N-m} x_1(n)x_1(n+m) \quad \text{โดย } 0 \leq m \leq M_0-1 \quad \dots\dots(2.2.5)$$

โดยที่ 1 เป็นตัวบอกถึงเฟรมใด ๆ และ  $N$  คือ จำนวนตัวอย่างสุ่มใน 1 เฟรม

$$x_1(n) = x_1(n+1) \quad \text{โดย } 0 \leq n \leq N-1 \quad \dots\dots(2.2.6)$$

ถ้าสัญญาณ  $x(n)$  เป็นสัญญาณที่มีคาบเท่ากับ  $P$  นั่นคือ  $x(n+p) = x(n)$  ที่  $n$  ใด ๆ ถ้าให้  $M_0 \gg P$  เราจะได้

$$\phi(m) = \phi(m+P) \quad \dots\dots(2.2.7)$$

และนี่คือ ฟังก์ชันหัตถสัมพันธ์ ที่จะได้คาบด้วย นอกจากนั้นการลดลงของค่าหัตถสัมพันธ์ จากจุดยอดที่  $m = 0$  กับค่าที่  $m$  เพิ่มขึ้นเรื่อย ๆ จะบ่งบอกถึงความสามารถในการทำนายของสัญญาณ



2.2.3 การวิเคราะห์ทางโดเมนความถี่ (Frequency Domain Analysis)

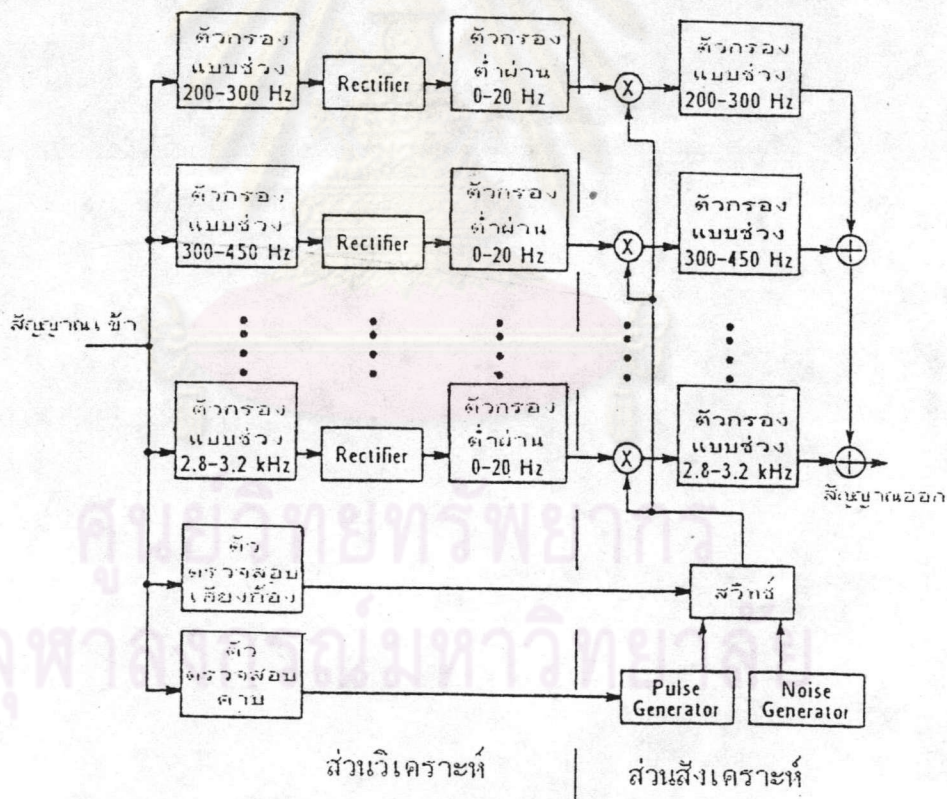
หลักการนี้ จะอาศัยการเปลี่ยนรูปข้อมูลระหว่างโดเมนเวลา กับ โดเมนความถี่ เทคนิคนี้จะช่วยในการระบุถึงส่วนประกอบทางความถี่ (Spectrum) ของคลื่นเสียง ตัวอย่างของการวิเคราะห์ทางสเปกตรัม ได้แก่

ก. เทคนิคของแชนเนลโคเดอ (Channel Vocoder)

(Schroeder M.R., 1966)

เทคนิคนี้จะประกอบด้วยกลุ่มของตัวกรองแบบช่วงความถี่ (Bandpass Filter) คลุมย่านความถี่ระหว่าง 200-3000 เฮิรตซ์ จำนวน 14 ช่อง ในแต่ละช่องจะมีวงจร Rectifier เพื่อดึงเอาเฉพาะ Envelope ของสัญญาณที่ผ่านตัวกรอง

ตัวตรวจสอบเสียงก้อง (Voiced-Unvoiced Detector) จะทำหน้าที่ตัดสินว่าเสียงพูดนั้นเป็นเสียงก้อง หรือเสียงไม่ก้อง ตัวตรวจสอบเสียงสูงต่ำ (Pitch detector) จะทำหน้าที่หาความถี่มูลฐาน (Fundamental Frequency) ของเสียงขณะนั้น



รูปที่ 2.2.2 แสดงส่วนประกอบของแชนเนลโคเดอ

(O'Shaughnessy D., 1987)



## ข. การวิเคราะห์ฟูรีเยร์ในช่วงเวลาสั้น

(Short-Time Fourier Analysis)

(Witten I.H., 1982)

เป็นเทคนิคที่ใช้ในการหาความสัมพันธ์ดังกล่าววิธีหนึ่ง โดยเริ่มจากการหาการแปลงของฟูรีเยร์สำหรับสัญญาณต่อเนื่อง ในรูปของ

$$X(f) = \int_{-\infty}^{\infty} x(t)e^{-i2\pi ft} dt \quad \dots (2.2.8ก)$$

$$x(t) = \int_{-\infty}^{\infty} X(f)e^{i2\pi ft} df \quad \dots (2.2.8ข)$$

เมื่อ  $-\infty < f < \infty$ ,  $-\infty < t < \infty$  และ  $i = -1$  โดยที่  $X(f)$  จะหมายถึงฟังก์ชันโดเมนความถี่ และ  $x(t)$  หมายถึงฟังก์ชันของโดเมนเวลา ในลักษณะเดียวกัน ถ้าเราทำการแปลงรูปคลื่นในโดเมนเวลาที่มีลักษณะเป็นฟังก์ชันแบบไม่ต่อเนื่อง และมีช่วงที่จำกัด ลักษณะการแปลงแบบนี้ เรียกว่า การแปลงแบบดิสครีตฟูรีเยร์ (Discrete Fourier Transform) สามารถเขียนได้ในรูปของ

$$X(j) = (1/N) \sum_{k=0}^{N-1} x(k)e^{-i2\pi jk/N} \quad \dots (2.2.9ก)$$

$$x(k) = \sum_{j=0}^{N-1} X(j)e^{i2\pi jk/N} \quad \dots (2.2.9ข)$$

เมื่อ  $j=0,1,\dots,N-1$  และ  $k=0,1,\dots,N-1$  โดยที่ทั้ง  $X(j)$  และ  $x(k)$  เป็นอนุกรมเชิงซ้อน เราสามารถเขียนสมการข้างต้นใหม่ โดยการให้  $W_N = e^{2\pi i/N}$  คือ

$$X(j) = (1/N) \sum_{k=0}^{N-1} x(k)W_N^{-jk} \quad \dots (2.2.10ก)$$

$$x(k) = \sum_{j=0}^{N-1} X(j)W_N^{+jk} \quad \dots (2.2.10ข)$$



ซึ่งสามารถแสดงตัวอย่างได้ คือ ถ้าเราจะแปลงรูปของข้อมูล 256 จุด(N) ของสัญญาณที่มีอัตรา  
การสุ่มที่ 8 กิโลเฮิร์ตซ์ เราจะได้ 256 ช่วงความถี่ระหว่าง 0 ถึง 8 กิโลเฮิร์ตซ์ ดังแสดงใน  
ตารางที่ 2.2.1 จากตารางจะเห็นว่า ทุก ๆ 31 เฮิร์ตซ์ในโดเมนความถี่ หรือทุก ๆ 32  
มิลลิวินาทีของเสียงที่จะถูกแปลงรูป โดยอาศัยความสัมพันธ์ข้างต้น

โดเมนเวลา		โดเมนความถี่	
ตัวอย่างที่	เวลา	ตัวอย่างที่	ความถี่
0	0 usec	0	0 Hz
1	125	1	31
2	250	2	62
...	...	...	...
254	31,750	254	7938
255	31,875 usec	255	7969 Hz

ตารางที่ 2.2.1 แสดงถึงตัวอย่างความสัมพันธ์ของโดเมนความถี่กับโดเมนเวลา

### 2.3 การหาขอบเขตของสัญญาณ (Endpoint Detection)

พื้นฐานของสัญญาณในการรับรู้เสียงพูดกลุ่มคำโดด คือ สัญญาณจะประกอบด้วยส่วนของ  
เสียงพูด อาจมีสัญญาณรบกวน (Noise Signal) นำหน้าและต่อท้ายด้วย ดังนั้น เมื่อต้องการ  
เฉพาะส่วนของสัญญาณ เราจึงจำเป็นต้องมีวิธีการแบ่งแยกส่วนของสัญญาณให้ถูกต้อง กระบวนการ  
ที่แยกส่วนของสัญญาณเสียงพูด ออกจากส่วนที่เป็นสัญญาณรบกวน เรียกว่าการหาขอบเขตของสัญญาณ  
(Endpoint Detection) ความจำเป็นของการหาขอบเขตของสัญญาณ ในการรับรู้เสียงพูดกลุ่ม  
คำโดด เนื่องจาก

- ก. ความถูกต้องของการรับรู้เสียงพูด จะขึ้นโดยตรงกับความถูกต้องของการหาขอบ  
เขตของสัญญาณ (Martin, 1974)
- ข. การหาขอบเขตสัญญาณที่ถูกต้อง จะทำให้เวลาที่ใช้ในการคำนวณน้อยที่สุด



พารามิเตอร์พื้นฐานที่ใช้ในการหาขอบเขตของสัญญาณ คือ พลังงานของสัญญาณ (Energy of Signal) โดยการเปรียบเทียบพลังงานของแต่ละเฟรม กับค่ากำหนด จุดเริ่มต้นของสัญญาณเสียงพูดจะได้จากเฟรมที่มีค่าพลังงานเกินค่ากำหนดเป็นเฟรมแรก และจุดสิ้นสุด จะได้จากเฟรมที่มีค่าพลังงานต่ำกว่าค่ากำหนดเป็นจุดแรก เช่นกัน

อย่างไรก็ตาม ได้มีการปรับปรุงการหาขอบเขตของสัญญาณ โดยอาศัยหลักการพื้นฐานดังกล่าว ได้เป็นดังนี้ (Lamel L.F., 1981)

ให้อินพุท คือ อะเรย์ของพลังงาน  $R_1(0)$ ,  $l=1,2,\dots,L$  โดย  $L$  เป็นจำนวนของเฟรมทั้งหมด จากนั้นนำอะเรย์ของพลังงานมาผ่านขั้นตอนดังนี้

### 2.3.1 การปรับปรุกระดับของพลังงาน (Adaptive Level Equalizer)

เป็นขั้นแรกของกระบวนการ โดยจะทำการปรับปรุกระดับของข้อมูล เทียบกับค่าลือกของระดับสัญญาณเบื้องหลัง (Background Signal) ได้ผลลัพธ์ คือ

$$R_1(0) = \log[R_1(0)] - Q \quad \text{โดย } l = 1, 2, \dots, L \dots (2.3.1)$$

โดย  $Q$  เป็นค่าเฉลี่ยของระดับสัญญาณเบื้องหลัง ซึ่งได้จากการหาค่า  $E_{min}$  ดังนี้

$$E_{min} = \min_{1 \leq l \leq L} \{ \log[R_1(0)] \}$$

จากนั้นจะทำการสร้างแผนภูมิแท่ง (Histogram) ระหว่าง  $\log[R_1(0)]$  กับ ค่า  $l$  โดยเลือกส่วนที่มีค่าระดับพลังงานสเกลลือก ต่ำกว่า 10 เดซิเบล แล้วจะทำการเลือกจุดเฉลี่ย 3 จุดจากแผนภูมิแท่งนี้ จะได้จุดยอดของแผนภูมิ ซึ่งจะใช้เป็นค่าของ  $Q$  นั้นเอง

### 2.3.2 การตรวจหารูปคลื่นพลังงาน (Energy Pulse Detection)

จากเอาต์พุทของการปรับปรุกระดับของพลังงาน  $R_1(0)$  เราจะทำการกำหนดระดับพลังงานอ้างอิง (Energy Thresholds) 4 ค่า คือ  $k_1$ ,  $k_2$ ,  $k_3$  และ  $k_4$  ดังรูปที่

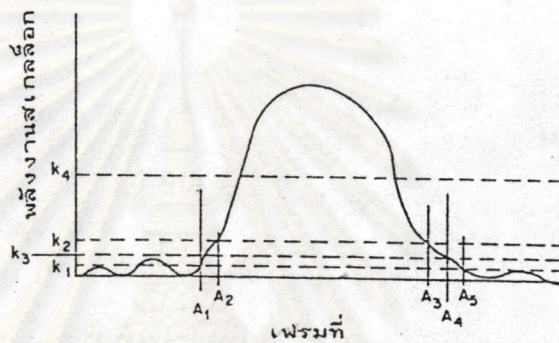
2.3.1 เพื่อใช้ในการหาขอบเขตของรูปคลื่นพลังงาน สมมุติฐานในการตรวจหารูปคลื่นพลังงาน คือ ในสัญญาณอินพุทที่เข้ามา จะมีรูปคลื่นพลังงานอย่างน้อย 1 รูป

ในการตรวจหารูปคลื่นพลังงาน จะทำจากซ้ายไปขวาโดยการเพิ่มขึ้นของค่า  $l$  เมื่อค่า  $R_1(0)$  มีค่าเกินจุดอ้างอิงแรก  $k_1$  เฟรมที่  $A_1$  จะถูกบันทึกไว้ ถ้าค่า  $R_1(0)$  มีค่าเกินจุดอ้างอิง  $k_2$  ก่อนที่จะตกกลับมาที่จุดอ้างอิง  $k_1$  จุดเริ่มต้นของรูปคลื่นจะเป็นเฟรมที่  $A_1$  แต่



ถ้าระยะเวลาจากเฟรม  $A_1$  ถึง  $A_2$  มีค่ายาวเกินไป จุดเริ่มต้นของรูปคลื่นจะถูกกำหนดเป็นเฟรมที่  $A_2$  แทน ส่วนจุดสิ้นสุดก็จะได้ในลักษณะเดียวกัน โดยใช้จุดอ้างอิงคือ จุด  $k_2$  และ  $k_3$  อย่างไรก็ตามถ้าระยะเวลาระหว่างเฟรมที่  $A_3$  และ  $A_4$  มีค่ายาวเกินไป ซึ่งเกิดจากมีลมหายใจที่ปลายคำเฟรมที่  $A_3$  จะเป็นเฟรมสุดท้ายของรูปคลื่นพลังงาน

การวิเคราะห์ทั้งสองขั้นตอนนี้ จะทำจนกว่าไม่พบรูปคลื่นพลังงานแล้ว โดยมีข้อยกเว้น คือ ถ้ารูปคลื่นมีค่าน้อยกว่าจุดอ้างอิง  $k_4$  รูปคลื่นนี้จะไม่ถูกนำมาคิด เนื่องจากถือว่าเป็นส่วนหนึ่งของคำ และถ้าระยะเวลาของรูปคลื่นมีค่าน้อยกว่า 5 เฟรม รูปคลื่นนี้จะไม่นำมาคิดเช่นกัน



รูปที่ 2.3.1 แสดงถึงการกำหนดจุดอ้างอิงเพื่อหาจุดเริ่มต้นและสิ้นสุดของรูปคลื่นพลังงาน (Lamel L.F., 1981)

#### 2.4 การหาฟิลเตอร์พารามิเตอร์โดยวิธีแอลพีซี (Markel J.D., 1978)

หลักการเบื้องต้นของการหาพารามิเตอร์โดยวิธีแอลพีซีคือ สัญญาณปัจจุบันสามารถประมาณได้ด้วย ผลรวมเชิงเส้นของสัญญาณในอดีต เราเรียกกระบวนการดังกล่าวว่า การทำนายเชิงเส้น (Linear Prediction) ให้  $\tilde{S}(n)$  คือสัญญาณเสียงลุ่มที่  $n$  ไต ๆ  $S(n)$  คือสัญญาณที่  $n$  ที่ได้จากการทำนาย  $a_1$  คือสัมประสิทธิ์ของการทำนาย (Predictor Coefficient) โดยใช้สัญญาณในอดีตทั้งหมด  $M$  ตัว หรือการทำนายมีออร์เดอร์ (Order) เท่ากับ  $M$  สมการของการทำนายเชิงเส้น คือ

$$S(n) = - \sum_{i=1}^M a_i S(n-i) \quad \dots (2.4.1)$$



การแปลงแบบแซท (Z-Transform) ของสมการ (2.4.1) คือ

$$\tilde{S}(z) = F(z)S(z) \quad \dots\dots(2.4.2)$$

โดย

$$F(z) = - \sum_{i=1}^M a_i z^{-i} \quad \dots\dots(2.4.3)$$

ซึ่ง  $F(z)$  มีชื่อว่า ฟิลเตอร์ทำนายเชิงเส้น (Linear Predictor Filter) ถ้าเอาสัญญาณจริงลบด้วยสัญญาณที่ได้จากการทำนาย จะได้ค่าผิดพลาด  $e(n)$  คือ

$$e(n) = S(n) - \tilde{S}(n) \quad \dots\dots(2.4.4)$$

$$= S(n) + \sum_{i=1}^M a_i S(n-i)$$

$$= \sum_{i=0}^M a_i S(n-i) \quad \text{โดย } a_0 = 1 \quad \dots\dots(2.4.5)$$

การแปลงแบบแซทของสมการ (2.4.5) คือ

$$E(z) = S(z)[1 - F(z)] \quad \dots\dots(2.4.6)$$

$[1 - F(z)]$  ในสมการ (2.4.6) มีชื่อเรียกว่าอินเวอร์สฟิลเตอร์ (Inverse Filter) และให้ใช้สัญลักษณ์  $A(z)$

จะได้

$$E(z) = S(z)A(z) \quad \dots\dots(2.4.7)$$

โดย

$$A(z) = 1 - F(z) = 1 + \sum_{i=1}^M a_i z^{-i} \quad \dots\dots(2.4.8)$$

สมการ (2.4.7) คือแบบจำลองของการวิเคราะห์ (Analysis Model) เมื่อเราทราบค่าชุดของ  $a_i$  ก็สามารถหาค่าผิดพลาดของการทำนายได้ ในแบบจำลองของการสังเคราะห์ (Synthesis Model) เราสามารถหาค่า  $S(z)$  ได้ถ้าทราบชุดของ  $a_i$  และ  $E(z)$  จากความสัมพันธ์

$$S(z) = \frac{E(z)}{1 - F(z)} = \frac{E(z)}{A(z)} \quad \dots\dots(2.4.9)$$

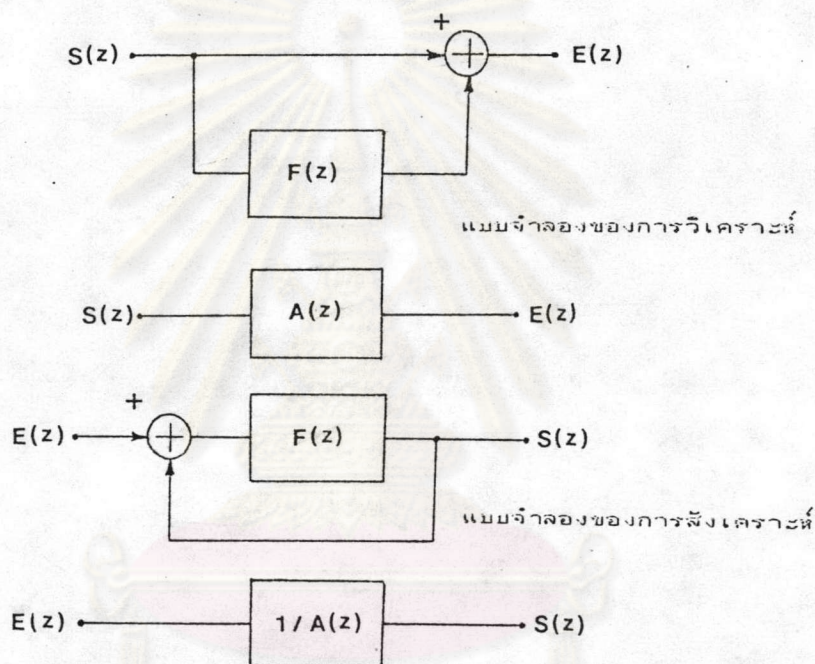
หรือ

$$S(n) = e(n) - \sum_{i=1}^M a_i S(n-i) \quad \dots\dots(2.4.10)$$



รูปที่ 2.4.1 จะแสดงถึงแบบจำลองของการวิเคราะห์ และสังเคราะห์ ตามสมการ (2.4.7) และ (2.4.9) ในการวิเคราะห์เราจะแบ่งเสียงพูดออกเป็นส่วนย่อย ๆ เรียกว่า เฟรม (Frame) แต่ละเฟรมกินเวลาประมาณ 10-20 มิลลิวินาที ซึ่งช่วงเวลาดังกล่าว เราสมมุติให้คุณสมบัติของเสียงไม่เปลี่ยนแปลงตามเวลา  $a_1$  จะเป็นค่าคงที่ในการหาพารามิเตอร์ ของแบบจำลองนี้ อาศัยเงื่อนไข ให้ค่าผิดพลาดยกกำลังสองรวม (Total Square Error) ใน หนึ่งเฟรมมีค่าน้อยที่สุด สมมุติให้เฟรมที่ทำการวิเคราะห์อยู่ระหว่าง  $n=n_0$  ถึง  $n=n_1$  ค่าผิดพลาด ยกกำลังสองรวมมีค่า เท่ากับ

$$\alpha = \sum_{n=n_0}^{n_1} e^2(n) \quad \dots\dots (2.4.11)$$



รูปที่ 2.4.1 แสดงถึงแบบจำลองของการวิเคราะห์และสังเคราะห์เสียงพูดด้วยวิธีแอลพีซี (Markel J.D., 1978)

แทนค่า  $e(n)$  จากสมการ (2.4.5) จะได้

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

$$= \sum_{n=n_0}^{n_1} \sum_{i=0}^M \sum_{j=0}^M a_i S(n-i) S(n-j) a_j \quad \dots\dots (2.4.12)$$



$$\text{ถ้ากำหนดให้ } C_{i,j} = \sum_{n=0}^{n_1} S(n-i)S(n-j) \quad \dots (2.4.13)$$

จะได้สมการ (2.4.12) เป็น

$$\alpha = \sum_{i=0}^M \sum_{j=0}^M a_j C_{i,j} a_i \quad \dots (2.4.14)$$

กำหนดเงื่อนไขค่าผิดพลาดยกกำลังสองรวมในหนึ่งเฟรมมีค่าน้อยที่สุด โดยหาค่าอนุพันธ์เทียบกับ  $a_k$  ให้เท่ากับ 0

$$\frac{\partial \alpha}{\partial a_k} = 0 = 2 \sum_{i=0}^M a_i C_{i,k} \quad \text{โดย } k = 1, 2, \dots, M \quad \dots (2.4.15)$$

เนื่องจาก  $a_0 = 1$  ดังนั้น จะได้ชุดของสมการเชิงเส้นจำนวน  $M$  สมการ คือ

$$\sum_{i=1}^M a_i C_{i,k} = -C_{0,k} \quad \text{โดย } k = 1, 2, \dots, M \quad \dots (2.4.16)$$

ค่าของ  $a_1$  ที่ได้จากการแก้สมการ (2.4.16) สามารถนำไปหาค่าผิดพลาด  $e(n)$  ได้ด้วยสมการ (2.4.5)

การแก้สมการเพื่อหาค่า  $a_1$  แบ่งได้เป็น 3 วิธีสำคัญ คือ

1. วิธีโควาเรียนซ์ (Covariance Method)
2. วิธีออโตโครีเลชัน (Autocorrelation Method)
3. วิธีพาร์คอร์ (PARCOR)

#### 2.4.1 วิธีโควาเรียนซ์ (Parson T.W., 1986)

จากสมการ (2.4.11) ถ้ากำหนดเงื่อนไขค่าผิดพลาดยกกำลังสองในช่วง  $n_0=M$  ถึง  $n_1=N-1$  จากสมการ (2.4.13) เราต้องการค่าของสัญญาณจริงทั้งหมด  $N$  ตัว คือ  $\{S(n)\} = \{S(0), S(1), \dots, S(N-1)\}$  จึงจะคำนวณค่า  $C_{i,j}$  โดย  $i=0, 1, \dots, M$  และ  $j = 1, 2, \dots, M$  ชุดของสมการจะได้เป็น

$$\sum_{i=1}^M a_i C_{i,j} = -C_{0,j} \quad \text{โดย } j = 1, 2, \dots, M \quad \dots (2.4.17)$$



ซึ่ง  $C_{ij} = \sum_{n=M}^{N-1} S(n-i)S(n-j) \dots\dots (2.4.18)$

หรือเขียนอยู่ในรูปเมตริกซ์ ได้เป็น

$$\begin{bmatrix} C_{11} & C_{12} & \dots & C_{1M} \\ C_{21} & C_{22} & \dots & C_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ C_{M1} & C_{M2} & \dots & C_{MM} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_M \end{bmatrix} = - \begin{bmatrix} C_{01} \\ C_{02} \\ \vdots \\ C_{0M} \end{bmatrix}$$



จากสมการ (2.4.18) จะพบว่า  $C_{ij} = C_{ji}$  ดังนั้นเมตริกซ์จึงมีคุณสมบัติสมมาตร และค่าในเส้นทแยงมีความสัมพันธ์ คือ

$$C_{i+1,j+1} = C_{ij} + S(-i-1)S(-j-1) - S(n-1-i)S(n-1-j) \dots\dots (2.4.20)$$

การหาคำตอบ  $a_1$  โดยสมการ (2.4.19) เรียกว่าวิธีโควาเรียนซ์ (Covariance Method) ซึ่งสามารถทำได้โดยมีประสิทธิภาพด้วยทฤษฎีของโชเลสกี (Cholesky's Theorem)

2.4.2 วิธีอัตราสัมพันธ์ (Parson T.W., 1986)

จากสมการ (2.4.13) ถ้ากำหนดให้  $n_0 = -\infty$  และ  $n_1 = \infty$  และให้  $S(n) = 0$  เมื่อ  $n < 0$  และ  $n \geq N$  ค่าของ  $C_{ij}$  จะได้เป็น

$$\begin{aligned} C_{ij} &= \sum_{n=-\infty}^{\infty} S(n-i)S(n-j) = \sum_{n=-\infty}^{\infty} S(n)S(n+|i-j|) \\ &= \sum_{n=0}^{N-1-|i-j|} S(n)S(n+|i-j|) \\ &= r(|i-j|) \quad \text{เมื่อ } i = 0, 1, \dots, M \dots\dots (2.4.21) \\ & \quad \quad \quad j = 1, 2, \dots, M \end{aligned}$$



ซึ่งการสังเกตนี้หมายความว่า เราจะกำหนดเงื่อนไขค่าผิดพลาดยกกำลังสองรวมกันน้อยที่สุด ในช่วง  $n_0 = -\infty$  ถึง  $n_1 = \infty$  แต่จะให้ผลเหมือนกับ  $n_0 = 0$  ถึง  $n_1 = N+M-1$  เนื่องจาก  $S(n)$  จะมีค่าเป็นศูนย์นอกช่วงดังกล่าว จึงทำให้ได้ชุดของสมการเชิงเส้นจำนวน  $M$  สมการ คือ

$$\sum_{i=1}^M a_i r(|i - j|) = -r(j) \quad \text{โดย } j = 1, 2, \dots, M \dots (2.4.22)$$

เมื่อเขียนอยู่ในรูปเมตริกซ์จะได้

$$\begin{bmatrix} r(0) & r(1) & r(2) & \dots & r(M-1) \\ r(1) & r(0) & r(1) & \dots & r(M-2) \\ r(2) & r(1) & r(0) & \dots & r(M-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r(M-1) & r(M-2) & r(M-3) & \dots & r(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_M \end{bmatrix} = - \begin{bmatrix} r(1) \\ r(2) \\ r(3) \\ \vdots \\ r(M) \end{bmatrix} \dots (2.4.23)$$

ในที่นี้  $[r(j)]$  คือเมตริกซ์ของค่าอัตสหสัมพันธ์ มีลักษณะเป็น Toeplitz Matrix คือมีคุณสมบัติสมมาตร และ ค่าตามแนวเส้นทแยงมีค่าเท่ากัน วิธีหาคำตอบอย่างมีประสิทธิภาพสามารถทำได้ด้วยวิธีของ Dubin/Levinson ข้อดีของวิธีอัตสหสัมพันธ์ คือ การคำนวณเพื่อแก้สมการมีน้อยกว่าวิธีโควาเรียนซ์ และมีความแน่นอนด้านเสถียรภาพ แต่สัญญาณจริง  $S(n)$  ที่จะใช้คำนวณ  $r(j)$  จะต้องผ่านหน้าต่างเพื่อลดความผิดพลาดทางองค์ประกอบความถี่ ซึ่งที่ใช้กันโดยทั่วไป คือ หน้าต่างแฮมมิง (Hamming Window) โดยสัญญาณที่ใช้ในการคำนวณ  $r(j)$  คือ  $S_1(n) = S_1(n)W(n)$  ซึ่ง  $S_1(n)$  เป็นสัญญาณสุ่มของเสียง และ  $W(n)$  คือฟังก์ชันของหน้าต่างมีค่าเป็น

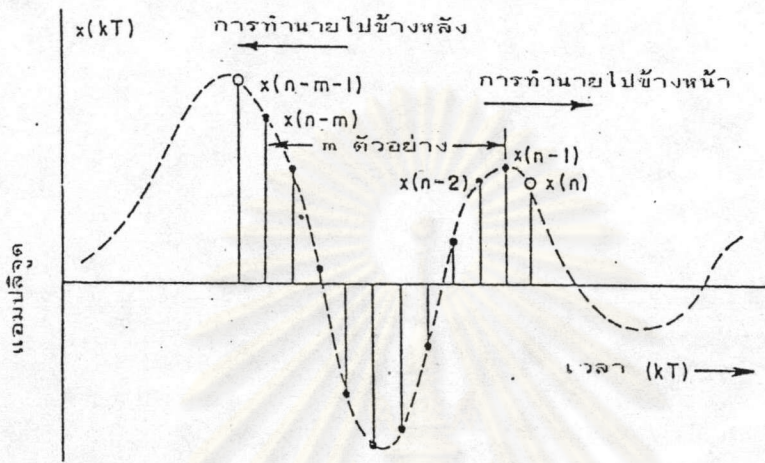
$$W(n) = 0.54 - 0.46 \cos (2\pi n / (N-1)) \quad \text{โดย } 0 \leq n \leq N-1 \dots (2.4.24)$$

### 2.4.3 วิธีพาร์คอร์ (PARCOR) (Markel J.D., 1978)

หลักการเบื้องต้นของวิธีพาร์คอร์ (PARCOR) ซึ่งย่อมาจาก สหสัมพันธ์บางส่วน (Partial Correlation) คือ การทำนายสัญญาณด้วยตัวทำนายเชิงเส้น (Linear Predictor) ที่มีออร์เดอร์เท่ากับ  $m$  ทั้งสองทาง คือทำนายไปทั้งข้างหน้า และย้อนหลัง จากรูปที่ 2.4.2 แสดงชุดของสัญญาณสุ่ม  $x(kT)$  จำนวน  $m+2$  ค่า การทำนายไปข้างหน้า หมายถึงการทำนายค่า  $x(n)$  จากค่า  $x(n-1), x(n-2), \dots, x(n-m)$  ค่าผิดพลาดของการทำนายที่ออร์เดอร์  $m$  ณ เวลา  $n$  แทนด้วย  $x_m^+(n)$  เท่ากับ



$$\begin{aligned}
 x_m^+(n) &= x(n) - \left[ -\sum_{i=1}^m a_{m1} x(n-i) \right] \\
 &= \sum_{i=0}^m a_{m1} x(n-i) \quad \text{โดย } a_{m0} = 1 \quad \dots\dots (2.4.25)
 \end{aligned}$$



รูปที่ 2.4.2 แสดงถึงชุดของสัญญาณสลับและการทำนายทั้งสองทาง (Markel J.D., 1978)

การทำนายย้อนหลังคือ การทำนายค่า  $x(n-m-1)$  จากค่า  $x(n-m), x(n-m+1), \dots, x(n-1)$  ค่าผิดพลาดที่ออร์เดอร์  $m$  ณ เวลา  $n$  แทนด้วย  $x_m^-(n)$  เท่ากับ

$$\begin{aligned}
 x_m^-(n) &= x(n-m-1) - \left[ -\sum_{i=1}^m b_{m1} x(n-i) \right] \\
 &= \sum_{i=1}^{m+1} b_{m1} x(n-i) \quad \text{โดย } b_{m,m+1} = 1 \quad \dots\dots (2.4.26)
 \end{aligned}$$

จากนั้นใช้เงื่อนไขค่าผิดพลาดยกกำลังสองรวมกันน้อยที่สุด กับการทำนายทั้งสองทิศทาง ให้  $\alpha_m$  และ  $\beta_m$  คือ ค่าผิดพลาดยกกำลังสองรวมของการทำนายไปข้างหน้า และ ข้างหลัง ตามลำดับ

$$\alpha_m = \sum_{n=n_0}^{n1} [x_m^+(n)]^2 \quad \text{และ} \quad \beta_m = \sum_{n=n_0}^{n1} [x_m^-(n)]^2$$

โดย  $m = 1, 2, \dots, M \quad \dots\dots (2.4.27)$

ใช้เงื่อนไขค่าผิดพลาดยกกำลังสองรวมกันน้อยที่สุด ดังนี้



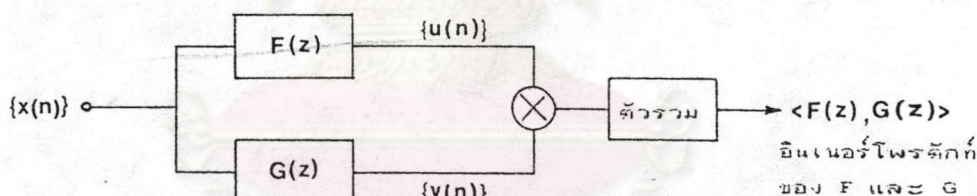
$$(\partial \alpha_m / \partial a_{m1}) = 0 \quad \text{และ} \quad (\partial \beta_m / \partial b_{m1}) = 0$$

โดย  $m = 1, 2, \dots, m$   
 $i = 1, 2, \dots, m$

อย่างไรก็ตาม เงื่อนไขนี้ยังไม่สามารถนำไปใช้ตั้งสมการ เพื่อหาค่าของสัมประสิทธิ์ตัวทำนายได้โดยตรง จำเป็นต้องมีเงื่อนไขเพิ่มเติม ซึ่งนาย J.D. Markel และ A.H. Gray ได้อาศัยหลักการของอินเนอร์โปรดักต์ และคุณสมบัติออร์โธโกนัลของตัวทำนาย เพื่อใช้ในการหาค่าตอบ โดยคำตอบที่ได้อยู่ในรูปของ สัมประสิทธิ์พาร์คอร์ (PARCOR Coefficient) แทนด้วยสัญลักษณ์  $k_m$

#### 2.4.4 หลักการอินเนอร์โปรดักต์และคุณสมบัติออร์โธโกนัลของตัวทำนาย

อาศัยนิยามและทฤษฎีเกี่ยวกับอินเนอร์โปรดักต์ในวิชาพีชคณิตเชิงเส้นประยุกต์ มาใช้ประยุกต์กับสัญญาณเสียง สมมติให้  $\{x(n)\}$  คือชุดของสัญญาณสุ่มที่เป็นอินพุตร่วมของฟิลเตอร์  $F(z)$  และ  $G(z)$  เอาท์พุตของฟิลเตอร์  $F(z)$  คือ  $\{u(n)\}$  และเอาท์พุตของฟิลเตอร์  $G(z)$  คือ  $\{v(n)\}$  ดังในรูปที่ 2.4.3 อินเนอร์โปรดักต์ของฟิลเตอร์  $F(z)$  และ  $G(z)$  ในช่วง  $n=n_0$  ถึง  $n=n_1$  แทนด้วยสัญลักษณ์  $\langle F(z), G(z) \rangle$  จะเท่ากับ ผลรวมของผลคูณระหว่าง  $u(n)$  กับ  $v(n)$  คือ



รูปที่ 2.4.3 แสดงถึงอินเนอร์โปรดักต์ของฟิลเตอร์  $F(z)$  กับ  $G(z)$

(Markel J.D., 1978)

$$\langle F(z), G(z) \rangle = \sum_{n=n_0}^{n_1} u(n)v(n) \quad \dots (2.4.28)$$

ถ้าให้  $F(z)$  และ  $G(z)$  เป็นรีลฟิลเตอร์ (Real Filter) จะได้อยู่ในรูป

$$F(z) = \sum_{i=0}^{\infty} f_i z^{-i} \quad \text{และ} \quad G(z) = \sum_{i=0}^{\infty} g_i z^{-i} \quad \dots (2.4.29)$$



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

$$\langle F(z), G(z) \rangle = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} f_i \left[ \sum_{n=n_0}^{n_1} x(n-i)x(n-j) \right] g_j \quad \dots (2.4.30)$$

ถ้าให้  $f$  และ  $g$  มีค่าเป็น 1 ตลอด จะได้

$$\langle z^{-i}, z^{-j} \rangle = \sum_{n=n_0}^{n_1} x(n-i)x(n-j) \quad \dots (2.4.31)$$

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

$$\langle z^{-i}, z^{-j} \rangle = C_{ij} = C_{ji} \quad \dots (2.4.32)$$

ด้วยวิธีอัตโนมัติกำหนดให้  $n_0 = -\infty$  และ  $n_1 = \infty$  โดยให้  $x(n)$  มีค่าเท่ากับ 0 เมื่อ  $n < 0$  และ  $n > N-1$  จะได้

$$\langle z^{-i}, z^{-j} \rangle = r(i-j) = r(j-i) \quad \dots (2.4.33)$$

ให้ค่าผิดพลาดจากการทำนายทั้งสองทาง คือ  $x_m^+(n)$  และ  $x_m^-(n)$  เป็นเอาต์พุตของฟิลเตอร์  $A_m(z)$  และ  $B_m(z)$  ซึ่งสอดคล้องกับสมการ (2.4.25) และ (2.4.26) โดย

$$A_m(z) = \sum_{i=0}^m a_{m1} z^{-i} \quad \text{โดย } a_{m0} = 1 \quad \dots (2.4.34ก)$$

$$B_m(z) = \sum_{i=1}^{m+1} b_{m1} z^{-i} \quad \text{โดย } b_{m,m+1} = 1 \quad \dots (2.4.34ข)$$

ด้วยนิยามของอินเนอร์โปรดักต์ ค่าผิดพลาดยกกำลังรวมสามารถแสดงได้ในรูปของ

$$\alpha_m = \langle A_m(z), A_m(z) \rangle = \|A_m(z)\|^2 \quad \dots (2.4.35ก)$$

$$\beta_m = \langle B_m(z), B_m(z) \rangle = \|B_m(z)\|^2 \quad \dots (2.4.35ข)$$

สมมติให้เงื่อนไขค่าผิดพลาดยกกำลังรวมน้อยที่สุด กับการทำนายไปข้างหน้า ดังนั้น  $\alpha_m$  ในสมการ (2.4.35ก) จะมีค่าน้อยที่สุด และเนื่องจาก  $A_m(z)$  เป็นโพลีโนเมียลของ  $z^{-1}$  โดย



$i = 0, 1, \dots, m$  เมื่อนำค่าใด ๆ ซึ่งเป็นโพลีโนเมียลของ  $z^{-j}$  โดย  $j = 1, 2, \dots, m$  บวกเข้ากับ  $A_m(z)$  นอร์มสแควร์ (Norm Square) ที่ได้จะต้องมีค่ามากกว่า หรือเท่ากับ  $\alpha_m$  คือ

$$\|A_m(z) + Cz^{-j}\|^2 \geq \|A_m(z)\|^2 \quad \text{โดย } j = 1, 2, \dots, m \quad \dots (2.4.36)$$

โดย  $C$  คือค่าคงที่ใด ๆ เมื่อแยกอสมการ (2.4.36) และย้ายข้าง จะได้

$$2C\langle A_m(z), z^{-j} \rangle + C^2\langle z^{-j}, z^{-j} \rangle \geq 0$$

เราเลือก  $C = -\langle A_m(z), z^{-j} \rangle / \langle z^{-j}, z^{-j} \rangle$  ถ้า  $\langle z^{-j}, z^{-j} \rangle \neq 0$

ดังนั้นจะได้  $-\langle A_m(z), z^{-j} \rangle^2 \geq 0$

เนื่องจาก  $A_m(z)$  เป็นรีลฟิลเตอร์ ดังนั้นอสมการจะเป็นจริงต่อเมื่อ

$$\langle A_m(z), z^{-j} \rangle = 0 \quad \text{โดย } j = 1, 2, \dots, m \quad \dots (2.4.37)$$

ในลักษณะเดียวกัน เมื่อเราเลือกค่า  $C = -\langle A_m(z), z^{-j} \rangle$  จะได้ว่า

$$\langle B_m(z), z^{-j} \rangle = 0 \quad \text{โดย } j = 1, 2, \dots, m \quad \dots (2.4.38)$$

สรุปคือว่า ถ้าใช้เงื่อนไขค่าผิดพลาดยกกำลังสองรวมน้อยที่สุด กับตัวทำนาย ผลที่ได้คือ  $A_m(z)$  และ  $B_m(z)$  จะออร์โธโกนัล กับ  $z^{-j}$  โดย  $j = 1, 2, \dots, m$

#### 2.4.5 การหาสัมประสิทธิ์พาร์ตอร์

เนื่องจาก  $A_{m-1}(z)$  เป็นโพลีโนเมียลของ  $z^{-j}$  เมื่อ  $j=0, 1, \dots, m-1$  และ  $B_{m-1}(z)$  เป็นโพลีโนเมียลของ  $z^{-j}$  เมื่อ  $j = 1, 2, \dots, m$  ดังนั้น ผลรวมเชิงเส้นของ  $A_{m-1}(z)$  กับ  $B_{m-1}(z)$  จะเป็นโพลีโนเมียลของ  $z^{-j}$  เมื่อ  $j=0, 1, \dots, m$  เพราะฉะนั้น  $A_m(z)$  จึงสามารถหาได้จากผลรวมเชิงเส้นของ  $A_{m-1}(z)$  กับ  $B_{m-1}(z)$  คือ

$$A_m(z) = A_{m-1}(z) + k_m B_{m-1}(z) \quad \dots (2.4.39)$$

โดย  $k_m$  เป็นค่าที่ทำให้  $A_m(z)$  อยู่ในเงื่อนไขค่าผิดพลาดยกกำลังสองรวมน้อยที่สุด คือ



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

ดังนั้นจะได้

$$\langle A_m(z), z^{-m} \rangle = \langle A_{m-1}(z), z^{-m} \rangle + k_m \langle B_{m-1}(z), z^{-m} \rangle \quad \dots (2.4.40)$$

จากการกำหนดค่าฟิลเตอร์  $A_m(z)$  และ  $B_m(z)$  ตามสมการที่ (2.4.34ก) และ (2.4.34ข) จะได้ว่าอินเนอร์โพรดักต์ระหว่าง  $A_{m-1}(z)$  กับ  $B_{m-1}(z)$  สามารถเขียนโดยแตก  $B_{m-1}(z)$  เป็นเทอมย่อย ๆ ได้

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

จากหลักการของอินเนอร์โพรดักต์ เราสามารถพิสูจน์ได้ว่า  $A_{m-1}(z)$  และ  $B_{m-1}(z)$  ออโธโกนัลกับ  $z^{-j}$  โดย  $j = 1, 2, \dots, m-1$  ดังนั้น เทอมทางขวามือจะเหลือเพียง

$$\langle A_{m-1}(z), B_{m-1}(z) \rangle = \langle A_{m-1}(z), z^{-m} \rangle \quad \dots (2.4.41)$$

ในทำนองเดียวกัน เมื่อแตกโพลีโนเมียลของ  $B_{m-1}(z)$  และใช้คุณสมบัติออโธโกนัล สามารถแสดงให้เห็นได้ว่า

$$\langle B_{m-1}(z), B_{m-1}(z) \rangle = \langle B_{m-1}(z), z^{-m} \rangle = \|B_{m-1}(z)\|^2 = \beta_{m-1} \quad \dots (2.4.42)$$

ด้วยวิธีการเดียวกันสามารถแสดงได้ว่า นอร์มสแควร์ของ  $A_{m-1}(z)$  จะเท่ากับ

$$\|A_{m-1}(z)\|^2 = \alpha_{m-1} = \langle 1, A_{m-1}(z) \rangle \quad \dots (2.4.43)$$

จากสมการ (2.4.39) จะได้

$$k_m = \frac{-\langle A_{m-1}(z), z^{-m} \rangle}{\langle B_{m-1}(z), z^{-m} \rangle} \quad \dots (2.4.44)$$



แทนค่าสมการ (2.4.44) จากสมการ (2.4.41) และ (2.4.42) จะได้

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

$$= - \frac{1}{\beta_{m-1}} \sum_{n=n_0}^{n_1} x_{m-1}^+(n) x_{m-1}^-(n) \quad \dots (2.4.45)$$



$k_m$  ที่ได้คือ ค่าสัมประสิทธิ์พาร์เตอร์ ซึ่งมีความหมายเป็นค่านอร์มัลไลซ์ของสหสัมพันธ์ระหว่างค่าผิดพลาดในการทำนาย ไปข้างหน้า และย้อนหลัง นอกจากนี้ค่าพารามิเตอร์  $k_m$  ยังสอดคล้องกับค่าสัมประสิทธิ์การสะท้อน (Reflection Coefficient) ในท่อนำเสียง ในบางครั้งเราจึงเรียกค่า  $k_m$  ว่าค่าสัมประสิทธิ์การสะท้อน จากสมการ (2.4.45) จะเห็นว่า  $k_m$  ที่ได้มีค่าใกล้เคียงกับที่ Saito และ Itakura ตั้งขึ้นมาในรูปของ

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

ซึ่งจะเห็นว่า  $k_m$  จะมีค่าอยู่ระหว่าง -1 ถึง 1 ซึ่งเป็นข้อดีในการนำไปใช้งาน การคำนวณสามารถทำได้ด้วย Fixed Point Arithmetic และมีความแน่นอนทางด้านเสถียรภาพในการคำนวณอีกด้วย

#### 2.4.6 กรรมวิธีการคำนวณค่าสัมประสิทธิ์พาร์เตอร์

การคำนวณสมการสามารถทำได้ทั้งวิธีโควาเรียนซ์ และวิธีอัตโนมัติ ในที่นี้เราจะใช้วิธีอัตโนมัติ จากความสัมพันธ์ของอินเวอร์สพิลเตอร์  $A_m(z)$  ดังในสมการ (2.4.39) จะเห็นว่า  $A_m(z)$  สามารถหาได้จาก  $A_{m-1}(z)$  และ  $B_{m-1}(z)$  จากนิยาม  $A_0(z) = 1$  และ  $B_0(z) = z^{-1}$  ดังนั้น เราจึงสามารถคำนวณ  $A_m(z)$  ที่  $m = 1$  ได้ คือ



$$A_1(z) = A_0(z) + k_1 B_0(z)$$

$$= 1 + k_1 B_0(z)$$

หา  $A_m(z)$  ที่  $m = 2$  จะได้

$$A_2(z) = A_1(z) + k_2 B_1(z)$$

$$= 1 + k_1 B_0(z) + k_2 B_1(z)$$

ดังนั้น  $A_m(z)$  สามารถหาได้จากการคำนวณต่อ ๆ กันแบบรีเคอร์ซีฟ (Recursive) โดยมีเทอมทั่วไป คือ

$$A_m(z) = 1 + \sum_{i=1}^m k_i B_{i-1}(z) \quad \dots\dots (2.4.46)$$

นอร์มสแควร์ของสมการ (2.4.46) เมื่อย้าย 1 มาข้างซ้ายจะได้

$$\|A_m(z) - 1\|^2 = \sum_{i=1}^m k_i^2 B_{i-1}^2(z)$$

เทอมทางซ้ายมือสามารถแตกได้เป็น

$$\|A_m(z) - 1\|^2 = \|A_m(z)\|^2 - 2\langle A_m(z), 1 \rangle + \|1\|^2$$

เช่นเดียวกับสมการ (2.4.43) สามารถแสดงให้เห็นว่า  $\|A_m(z)\|^2 = \langle 1, A_m(z) \rangle = \alpha_m$   
ดังนั้น

$$\alpha_m = \|1\|^2 - \sum_{i=1}^m k_i^2 B_{i-1}^2 \quad \dots\dots (2.4.47)$$

แทนค่า  $m$  ด้วย  $m+1$  จะได้  $\alpha_{m+1}$  แล้วลบออกด้วย  $\alpha_m$  จะได้

$$\alpha_{m+1} = \alpha_m - k_{m+1}^2 B_m^2 \quad \dots\dots (2.4.48)$$

ใช้เงื่อนไขค่าผิดพลาดยกกำลังสองรวมน้อยที่สุด กับการทำนายไปข้างหน้าและย้อนหลัง ความสัมพันธ์ที่ได้คือ

$$\min \alpha_m = \min \beta_m$$



จะได้

$$\alpha_{m+1} = (1 - k_{m+1}^2) \alpha_m \quad \dots (2.4.49)$$

ในการหาค่า  $k$  จากสมการ (2.4.45) สามารถเขียนใหม่ได้เป็น

$$k_{m+1} = - \frac{1}{\beta_m} \langle A_m(z), B_m(z) \rangle$$

แทน  $\beta_m$  ด้วย  $\alpha_m$  และจาก (2.4.41) แทน  $\langle A_m(z), B_m(z) \rangle$  ด้วย  $\langle A_m(z), z^{-(m+1)} \rangle$  ได้

$$\begin{aligned} k_{m+1} &= -(1/\alpha_m) \langle A_m(z), z^{-(m+1)} \rangle \\ &= -(1/\alpha_m) \sum_{i=0}^m r(m+1-i) a_{m1} \quad \dots (2.4.50) \end{aligned}$$

จากสมการ (2.4.39) สามารถเขียนใหม่ได้เป็น

$$A_{m+1}(z) = A_m(z) + k_{m+1} B_m(z) \quad \dots (2.4.51)$$

แทนค่า  $k_{m+1}$  จากสมการ (2.4.50) ลงไปในสมการ (2.4.51) แล้วเทียบสัมประสิทธิ์ทั้งสองข้างจะได้

$$a_{m+1,0} = 1 \quad \dots (2.4.52ก)$$

$$a_{m+1,i} = a_{m1} + k_{m+1} b_{m1} \quad \text{โดย } i = 1, 2, \dots, m \quad \dots (2.4.52ข)$$

$$a_{m+1,m+1} = k_{m+1} \quad \dots (2.4.52ค)$$

และ  $b_{m1}$  ในสมการ (2.4.52ข) สามารถเขียนใหม่ได้เป็น

$$b_{m1} = a_{m,m+1-1} \quad \text{โดย } i = 1, 2, \dots, m+1 \quad \dots (2.4.53)$$

สรุปสมการที่ใช้ในการคำนวณหาค่า  $k_m$  ประกอบด้วยสมการ (2.4.49) (2.4.50) (2.4.52) และ (2.4.53) ซึ่งสมการเหล่านี้อยู่ในรูปการคำนวณแบบรีเคอร์ซีฟทั้งสิ้น ในการนำสมการต่าง ๆ มาคำนวณด้วยเครื่องคอมพิวเตอร์ สามารถสรุปเป็นขั้นตอนต่าง ๆ ได้ดังนี้



### ขั้นที่ 1 คำนวณค่าอัตสหสัมพันธ์

กำหนดให้ในหนึ่งเฟรมของการวิเคราะห์ มีจำนวนสัญญาณสุ่ม เท่ากับ  $N$  ตัวอย่าง และฟิลเตอร์มีจำนวนออร์เตอร์เท่ากับ  $M$  ตามวิธีอัตสหสัมพันธ์ ค่าอัตสหสัมพันธ์ที่ใช้ในการคำนวณขั้นต่อ ๆ ไป จะมีจำนวนเท่ากับ  $M+1$  ค่า คือ

$$r(k) = \sum_{n=0}^{N-1-k} x(n)x(n+k) \quad \text{โดย } k = 0, 1, \dots, M \dots (2.4.54)$$

โดย  $x(n)$  เมื่อ  $n = 0, 1, \dots, N-1$  คือสัญญาณเสียงที่สุ่มมาในหนึ่งเฟรม

### ขั้นที่ 2 คำนวณค่าเริ่มต้น

จากสมการ (2.4.50) แทน  $m = 0$  และ สมการ (2.4.52ค) จะได้

$$\alpha_0 = \langle A_0(z), A_0(z) \rangle = r(0)$$

ซึ่งจากเงื่อนไขเริ่มต้นตามสมการ (2.4.34ก)  $a_{00} = 1$  และ  $\alpha_0 = r(0)$  ดังนั้น

$$k_1 = -r(1)/r(0) \dots (2.4.55)$$

จากสมการ (2.4.52ก) และ (2.4.52ข) แทนค่า  $m = 0$  ได้

$$a_{10} = 1 \quad \text{และ} \quad a_{11} = k_1 \dots (2.4.56)$$

จากสมการ (2.4.49) แทน  $m = 0$  จะได้

$$\alpha_1 = (1 - k_1^2)\alpha_0 = (1 - k_1^2)r(0) \dots (2.4.57)$$

### ขั้นที่ 3 ทำการคำนวณแบบรีเคอร์ซีฟ

ขั้นตอนนี้จะนำค่าเริ่มต้นจากขั้นตอนที่แล้ว มาทำการคำนวณต่อ ๆ กัน แบบรีเคอร์ซีฟ โดยอาศัยสมการ (2.4.49) (2.4.50) (2.4.52) และ (2.4.53) ในการหาค่าสัมประสิทธิ์พาร์คอร์จนครบจำนวนออร์เตอร์ คือ  $k_i$  โดย  $i = 1, 2, \dots, M$  และพร้อมกันนั้นก็จะได้ค่าสัมประสิทธิ์ของตัวทำนาย  $a_i$  โดย  $i = 0, 1, \dots, M$  จะต้องมีค่าการคำนวณต่อ ๆ กันไปทั้งหมด  $M-1$  รอบ คือแทนค่า  $m = 1, 2, \dots, M-1$  เมื่อคำนวณเสร็จขั้นที่  $m = M-1$  จะได้คำตอบครบตามที่ต้องการ รวมทั้งค่า  $\alpha_m$  หรือค่าผิดพลาดยกกำลังสองรวมในเฟรม



ขั้นตอนทั้งสามสามารถนำมาเขียนเป็นโปรแกรมย่อยชื่อ AUTO เพื่อใช้ในการวิเคราะห์เสียงพูด ทหารามิเตอร์ของฟิลเตอร์ในแต่ละเฟรม จำนวนสัญญาณสุ่มที่ใช้ในการวิเคราะห์ในหนึ่งเฟรมกำหนดที่ตัวแปร  $N$  จำนวนออร์เตอร์ของฟิลเตอร์กำหนดที่ตัวแปร  $M$  ตารางที่ 2.4.4 แสดงชื่อตัวแปรในโปรแกรม เทียบกับสัญลักษณ์ในสมการ และรูปที่ 2.4.5 แสดงถึงผังโปรแกรมย่อยของขั้นตอนดังกล่าวมา

สัญลักษณ์ในสมการ	ตัวแปรในโปรแกรมย่อย
$N$	$N$
$x(n)$ โดย $n=0, 1, \dots, N-1$	$x(n+1)$
$M$	$M$
$a_{m1}$ โดย $0 \leq i \leq m \leq M$	$A(i+1)$
$\alpha_m$ โดย $m=0, 1, \dots, M$	ALPHA
$k_m$ โดย $m=1, 2, \dots, M$	RC(m)
$r(i)$ โดย $i=0, 1, \dots, M$	$R(i+1)$
$\beta_m$ โดย $i=0, 1, \dots, M-1$	ALPHA
$b_{m1}$ โดย $1 \leq i \leq m+1 \leq M$	$B(i)$

ตารางที่ 2.4.2 แสดงถึงชื่อตัวแปรในโปรแกรม AUTO เทียบกับสัญลักษณ์

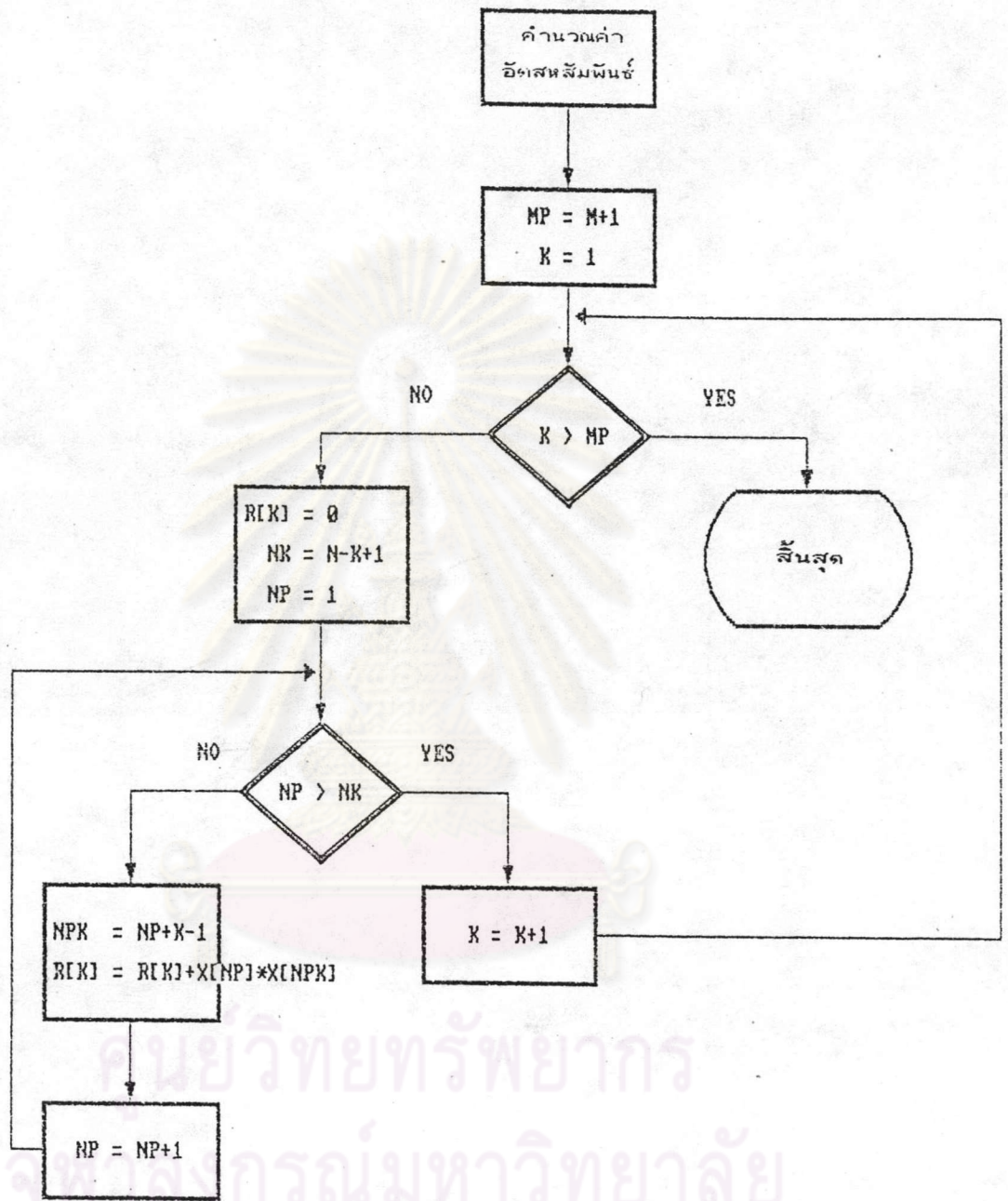






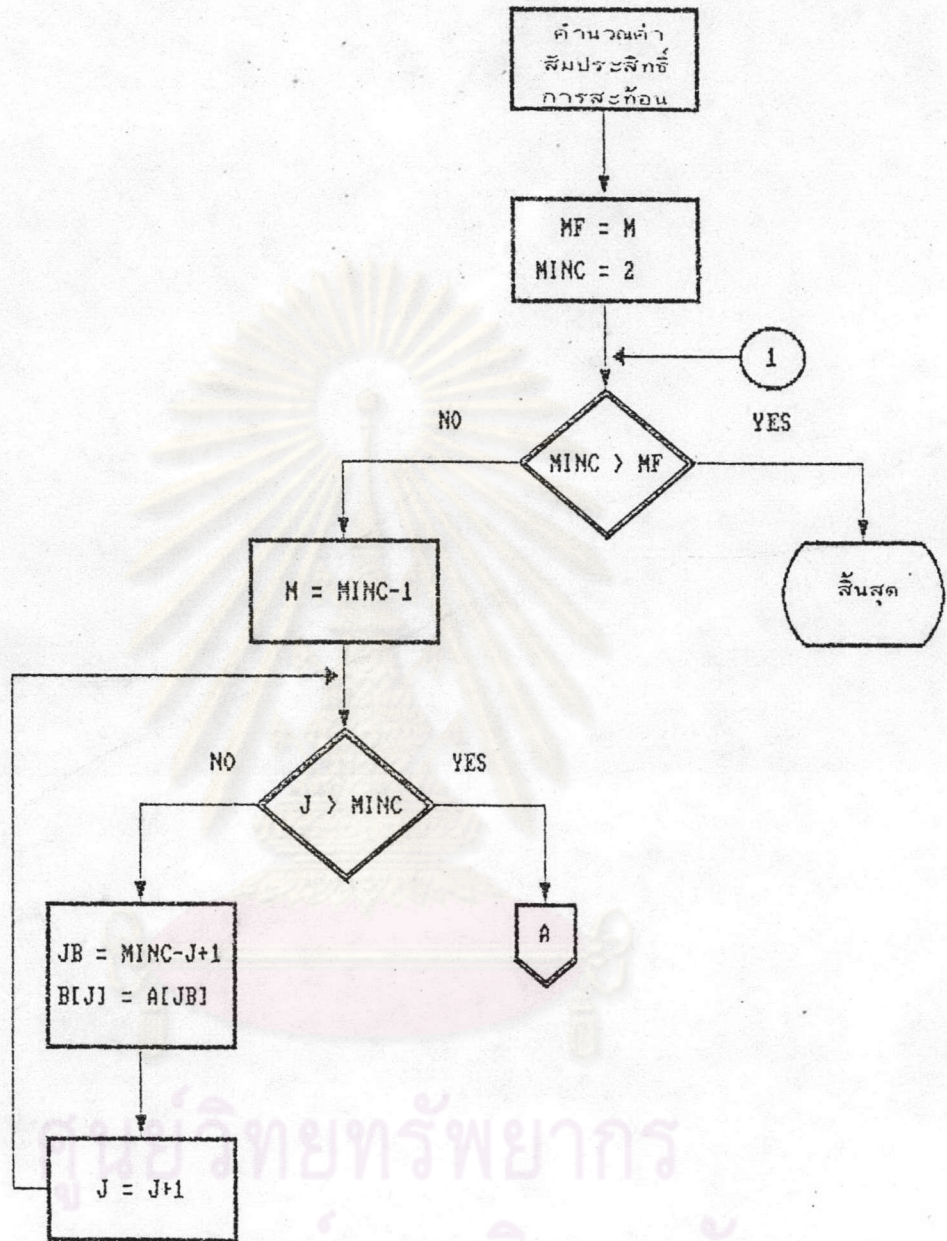
รูปที่ 2.4.5 แสดงถึงผังโปรแกรมย่อย AUTO





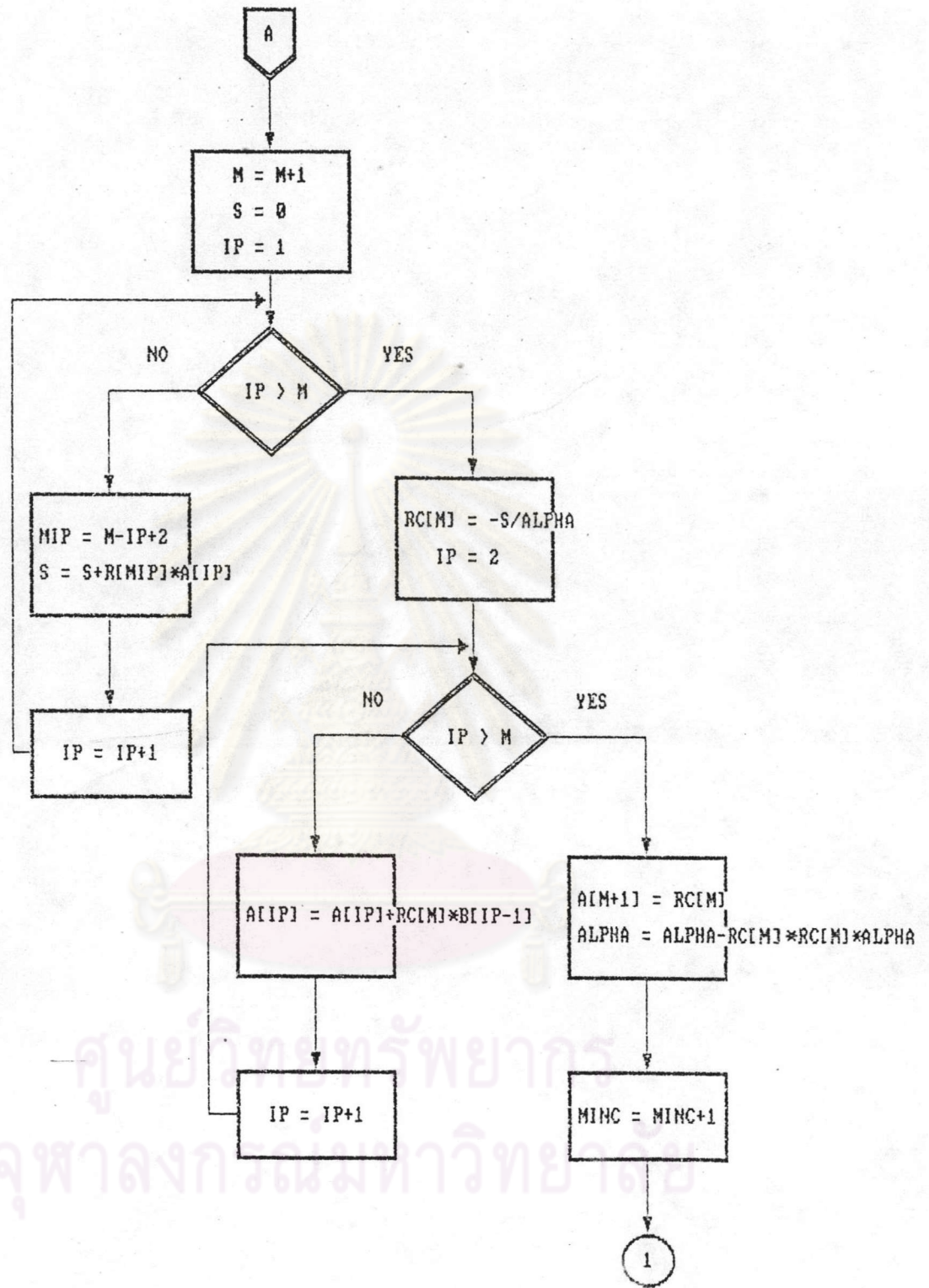
รูปที่ 2.4.5 แสดงถึงผังโปรแกรมย่อย AUTO (ต่อ)





รูปที่ 2.4.5 แสดงถึงผังโปรแกรมย่อย AUTO (ต่อ)





รูปที่ 2.4.5 แสดงถึงผังโปรแกรมย่อย AUTO (ต่อ)



## 2.5 การคำนวณหาคาบของสัญญาณ

การคำนวณหาคาบของสัญญาณ สามารถทำได้หลายวิธี เช่น วิธีของ Sondhi (Sondhi M.M., 1968) วิธี SIFT (Markel J.D., 1972) วิธีการใช้ Cepstrum (Noll A.M., 1967) เป็นต้น แต่อย่างไรก็ตาม วิธีส่วนใหญ่จะมีการคำนวณมาก และมีขั้นตอนที่ยุ่งยาก ซับซ้อน ทำให้ไม่เหมาะสมในการนำไปใช้งาน กับความต้องการการตอบสนองแบบทันทีทันใด

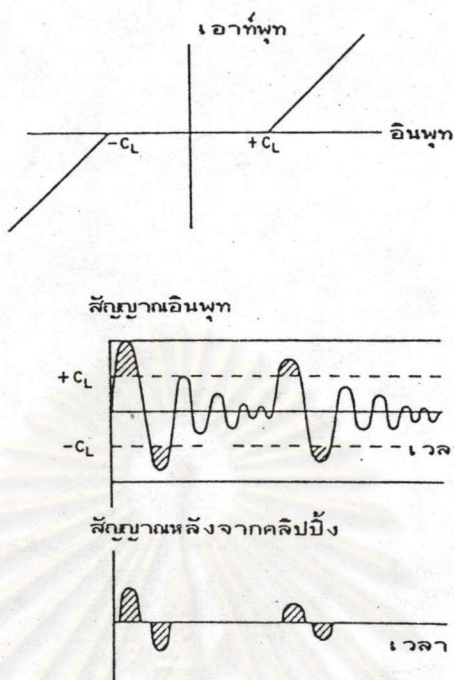
ปัญหาที่สำคัญประการหนึ่งในการหาคาบของสัญญาณ ให้ความถูกต้องในช่วงแถบความถี่ของสัญญาณ คือ อิทธิพลของโครงสร้างฟอร์แมนต์ (Formant Structure) ซึ่งจะมีความสัมพันธ์กับคาบของรูปคลื่น ดังนั้นในการคำนวณหาคาบของสัญญาณให้ถูกต้อง เราจำเป็นต้องลดอิทธิพลของโครงสร้างฟอร์แมนต์ลง หรือกำจัดออกไปเลย ซึ่งวิธีการนี้ เรียกว่า การทำสเปกตรัมราบเรียบ (Spectral Flattening) ซึ่งสามารถทำได้ 2 วิธี คือ (Sondhi M.M., 1968)

ก. วิธีฟิลเตอร์แบงก์ (Filter Bank Method) โดยจะนำสัญญาณมาผ่านตัวกรองแบบช่วง ซึ่งกระจายอยู่เต็มแถบความถี่ของสัญญาณ สัญญาณที่ออกมาจะถูกรวมอีควัลไลซ์ เป็นหน่วยย่อย ๆ ของแอมพลิจูด (Spectrally Flattened) โดยจะแบ่งตามพลังงานในช่วงเวลาสั้น ผลรวมของหน่วยย่อยนี้ ได้จากการรวมหน่วยย่อย ๆ กับเวลาหน่วงที่เหมาะสม แม้ว่าวิธีนี้จะสามารถใช้ได้เป็นอย่างดี แต่ก็มีปัญหาบางประการในทางปฏิบัติ คือ ประการแรก วิธีนี้จำเป็นต้องใช้ฮาร์ดแวร์จำนวนมาก เพื่อทำการกรอง และอีควัลไลซ์ (Equalized) ประการที่สอง ในบางกรณีที่สัญญาณไม่มีฮาร์โมนิกของคาบในช่วงของตัวกรองแบบช่วง ผลที่ได้เมื่อผ่านการทำอีควัลไลซ์ จะมีระดับของสัญญาณรบกวนสูง ซึ่งจะขัดขวางในการคำนวณหาคาบของสัญญาณ

ข. วิธีเซนเตอร์คลิปปิง (Center Clipping) วิธีการนี้อาศัยหลักการ คือ ค่าของสัญญาณที่มีระดับได้ระดับคลิปปิง (Clipping Level) จะถูกกำหนดให้มีค่าเป็น 0 ส่วนค่าที่เหนือกว่าระดับคลิปปิงจะถูกลบด้วยค่าระดับคลิปปิง จากรูปที่ 2.5.1 จะแสดงให้เห็นถึงลักษณะของผลลัพธ์ที่ได้จากการทำเซนเตอร์คลิปปิง และยังแสดงให้เห็นถึงการทำสเปกตรัมให้ราบเรียบด้วยวิธีนี้ จากรูปจะเห็นว่า ถ้ามีการกำหนดระดับคลิปปิงที่เหมาะสม ส่วนของสัญญาณที่มีอิทธิพลของโครงสร้างฟอร์แมนต์จะถูกกำจัดลงได้

จากวิธีดังกล่าวจะเห็นว่า วิธีเซนเตอร์คลิปปิงเป็นวิธีที่สะดวกกว่าในการทำสเปกตรัมให้ราบเรียบ จึงได้มีการพัฒนาวิธีการนี้บนฮาร์ดแวร์ (Dubnowski J.J., 1976) เพื่อหาคาบของสัญญาณซึ่งมีขั้นตอนดังแสดงในรูปที่ 2.5.2 คือ สัญญาณอนาล็อกจะผ่านตัวกรองต่ำผ่าน เพื่อให้ได้แถบความถี่ประมาณ 900 เฮิรตซ์ แล้วจึงนำไปผ่านตัวแปลงสัญญาณอนาล็อกเป็นสัญญาณดิจิทัล สัญญาณที่ได้จะถูกแบ่งเป็นเฟรม มีขนาด 30 มิลลิวินาที เพื่อใช้ในกระบวนการต่อไป





รูปที่ 2.5.1 แสดงถึงลักษณะอินพุต เอาต์พุต และกระบวนการของการทำเช่นเตอร์คลิปปิง (Sondhi M.M., 1968)

ในขั้นแรก จะทำการคำนวณระดับคลิปปิงในแต่ละส่วน เนื่องจากเสียงมีลักษณะการแปรเปลี่ยนกว้าง ทำให้การเลือกระดับคลิปปิงจึงต้องทำด้วยความระมัดระวัง เพื่อป้องกันสารสนเทศของรูปคลื่นสูญหาย ในขณะที่แอมพลิจูดมีการเปลี่ยนแปลงไป วิธีการเลือก  $C_L$  ทำโดย แบ่งข้อมูลในแต่ละเฟรมเป็นส่วน ๆ ละ 10 มิลลิวินาที ในส่วนแรก และส่วนที่สาม จะถูกนำไปหาขอดที่มีค่าสูงสุดในช่วงนั้น ระดับคลิปปิงจะเลือกจากขอดที่มีค่าน้อยระหว่างขอดทั้งสองคูณกับเปอร์เซ็นต์ที่กำหนดจากการทดลอง พบว่าค่าเปอร์เซ็นต์ที่ใช้กำหนดระดับคลิปปิงจะมีค่าประมาณร้อยละ 30 (Sondhi M.M., 1968)

หลังจากกำหนดระดับคลิปปิงแล้ว ค่าของสัญญาณในช่วงนี้จะถูกกำหนดค่าใหม่ โดยค่าของสัญญาณที่มีค่ามากกว่าระดับคลิปปิงบวก จะให้ค่าเป็น +1 ค่าของสัญญาณที่มีค่าน้อยกว่าระดับคลิปปิงลบ จะให้ค่าเป็น -1 ส่วนค่าที่อยู่ในช่วงบวก ลบ ระดับคลิปปิงจะให้ค่าเป็น 0 ในรูปที่ 2.5.3 จะแสดงให้เห็นถึงลักษณะอินพุต เอาต์พุตของการใช้วิธีเช่นเตอร์คลิปปิง จากการกำหนดค่าในลักษณะนี้ จะช่วยลดความซับซ้อนในการคำนวณฟังก์ชันออโตลอง เนื่องจากไม่จำเป็นต้องมีการคูณ หรือบวก ในการคำนวณหาฟังก์ชันออโตลองสำหรับสัญญาณนี้



ในขั้นต่อไป จะทำการคำนวณค่าฟังก์ชันออโต สำหรับสัญญาณ 30 มิลลิวินาที คือ

$$R_x(m) = \sum_{n=0}^{299-m} x(n)x(n+m) \quad \text{โดย } m = M_1, M_1+1, \dots, M_f \quad \dots (2.5.1)$$

โดย  $M_1$  เป็นค่าเริ่มต้น และ  $M_f$  เป็นค่าสุดท้าย สำหรับการคำนวณฟังก์ชันออโต เนื่องจากแต่ละนิพจน์ในสมการที่ (2.5.1) อยู่ในรูปของ  $x(n)x(n+m)$  และค่า  $x(n)$  จะมีค่าได้เพียง +1, 0 หรือ -1 เท่านั้น ดังนั้นผลคูณในสมการที่ (2.5.1) จึงสามารถมีค่าได้เป็น

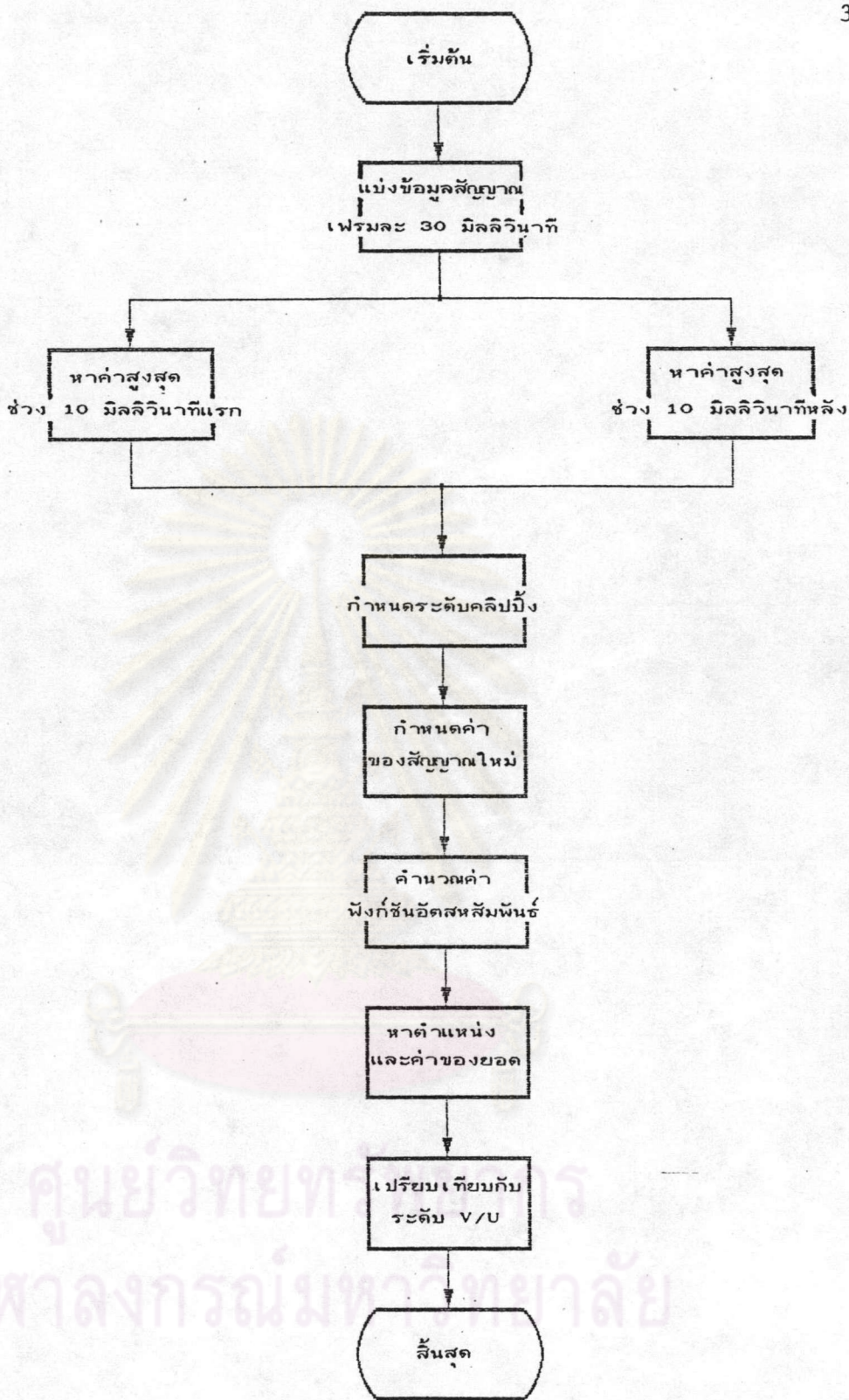
$$\begin{aligned} x(n)x(n+m) &= 0 \text{ ถ้า } x(n) = 0 \text{ หรือถ้า } x(n+m) = 0 \\ &= 1 \text{ ถ้า } x(n) = x(n+m) = \pm 1 \\ &= -1 \text{ ถ้า } x(n) = -x(n+m) = \pm 1 \end{aligned} \quad \dots (2.5.2)$$

ในรูปที่ 2.5.4 จะแสดงให้เห็นถึง ตัวอย่างของสัญญาณที่ทำการเซนเตอร์คลิปปิง และทำการหาฟังก์ชันออโต รูปบนแสดงถึงสัญญาณที่ผ่านตัวกรองต่ำผ่าน และทำการหารระดับคลิปปิงแล้ว รูปกลางแสดงถึงสัญญาณหลังจากผ่านระดับคลิปปิงแล้ว และรูปล่างแสดงถึงฟังก์ชันออโตของสัญญาณ ปกติคาบของสัญญาณจะอยู่ในช่วง  $m=20$  ถึง  $200$  สำหรับตัวอย่างนี้ คาบของสัญญาณจะอยู่ประมาณ 40 ตัวอย่าง หรือ ประมาณ 250 เฮิร์ตซ์

จะสังเกตเห็นว่า ในการคำนวณค่าฟังก์ชันออโตในสมการที่ (2.5.1) ตัวอย่างที่อยู่ นอกเหนือส่วนที่พิจารณา จะมีค่าเป็นศูนย์ ซึ่งจะมีผลต่อฟังก์ชันออโต โดยอาศัยลิเนียร์เทปเปอร์ (Linear Taper) ซึ่งเริ่มต้นจากค่า 1 ที่  $m=0$  และมีค่า 0 ที่  $m=300$  ซึ่งจะเห็นผลได้ชัดเจน ดังรูปที่ 2.5.4 การใช้ลิเนียร์เทปเปอร์สำหรับฟังก์ชันออโต จะทำให้การหายอดที่คาบของสัญญาณ ชัดเจนขึ้น ในกรณีที่ยอดของคาบสัญญาณมีมากกว่า 1 โดยจะลดความเป็นไปได้ของยอดอื่น ๆ ลง เนื่องจากอัตราการเปลี่ยนแปลงต่อลิเนียร์เทปเปอร์ของยอดหลัง ๆ มีมากกว่า

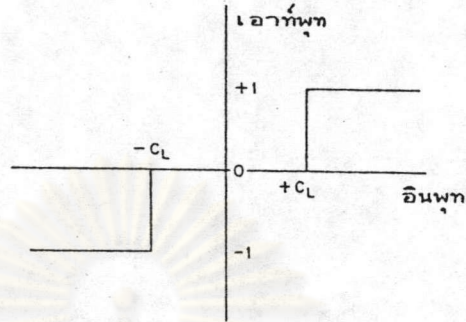
จากค่าของฟังก์ชันออโต จะทำการหาค่าสูงสุดในช่วง  $m=M_1$  ถึง  $m=M_f$  โดยจะเก็บทั้ง ตำแหน่ง และค่าไว้ ถ้ายอดนี้มีค่าเกิน voiced-unvoiced threshold ส่วนนี้จะถือเป็นส่วน ก้อง (Voiced Section) และคาบของสัญญาณจะมีค่าเท่ากับ ตำแหน่งของยอดสูงสุดนั้น แต่ถ้า ยอดนี้มีค่าต่ำกว่า ส่วนนี้จะถือว่าเป็นส่วนไม่ก้อง (Unvoiced Section)



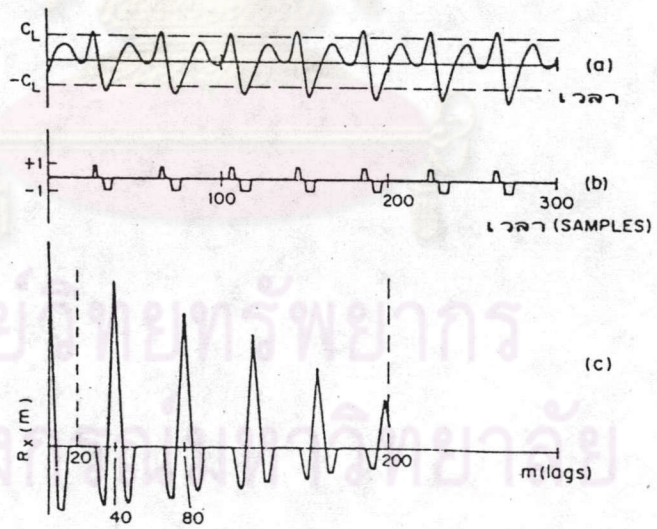


รูปที่ 2.5.2 แสดงขั้นตอนทั้งหมดของการหาค่าของสัญญาที่ทำโดยอาศัยฮาร์ดแวร์  
(Dubnowski J.J., 1976)





รูปที่ 2.5.3 แสดงถึงลักษณะอินพุต เอาต์พุต ของการทำเช่นเตอร์คลิบปี้ง และการกำหนดค่าสัญญาณใหม่ (Sondhi M.M., 1968)



รูปที่ 2.5.4 แสดงถึงตัวอย่างสัญญาณที่ผ่านการทำเช่นเตอร์คลิบปี้ง และทำการหาฟังก์ชันออโต (Sondhi M.M., 1968)



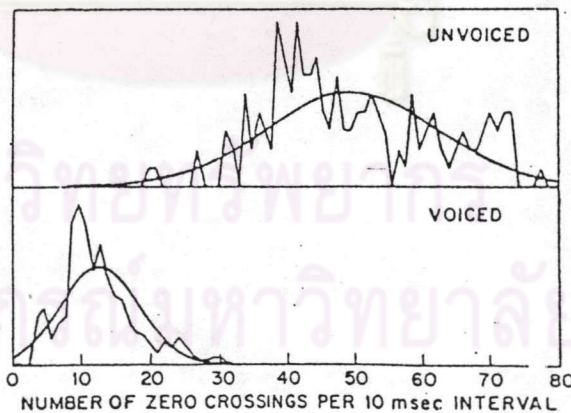
การเกิดซีโรคลอสซิงจะเกิดขึ้นจากการที่รูปคลื่นของสัญญาณมีการตัดกับแกนเวลา นั่นคือค่าของสัญญาณจะมีการเปลี่ยนสัญลักษณ์ทางคณิตศาสตร์ (Algebraic Sign) นั้นเอง อัตราการเกิดซีโรคลอสซิงจะเป็นเครื่องมืออย่างง่าย ที่ใช้ในการวัดความถี่ในการเปลี่ยนแปลงข้อมูลของสัญญาณ ค่าซีโรคลอสซิงสามารถนำมาใช้ในการตัดสินว่าสัญญาณเสียงนั้นเป็นเสียงก้อง หรือเสียงไม่ก้อง เนื่องจากเสียงก้องส่วนใหญ่จะมีค่าพลังงานอยู่ในช่วงความถี่ต่ำ ส่วนเสียงไม่ก้องจะมีพลังงานอยู่ในช่วงความถี่สูง และเนื่องจากค่าซีโรคลอสซิงมีความสัมพันธ์โดยตรงกับค่าความถี่ของสัญญาณ ดังนั้นจะได้ว่าสัญญาณเสียงที่มีค่าซีโรคลอสซิงสูงจะเป็นเสียงไม่ก้อง และสัญญาณเสียงที่มีค่าซีโรคลอสซิงต่ำจะเป็นเสียงก้อง แต่อย่างไรก็ตามการกำหนดขนาดของค่าซีโรคลอสซิงที่แน่นอนเพื่อจำแนกชนิดของเสียงนั้นจะอาศัยผลของการทดลองเป็นหลัก ในรูปที่ 2.6.1 จะแสดงให้เห็นถึงฮิสโตแกรม (Histogram) ของอัตราการเกิดซีโรคลอสซิงของสัญญาณเสียงก้องและเสียงไม่ก้อง

ในการหาค่าซีโรคลอสซิงสามารถทำได้โดยอาศัยสมการที่ (2.6.1) คือ

$$T[s(n)] = 0.5|sign(s(n)) - sign(s(n-1))| \quad \dots (2.6.1)$$

โดยที่ค่า  $sign(s(n))$  จะได้จาก

$$sign(s(n)) = \begin{cases} 1 & \text{สำหรับ } s(n) \geq 0 \\ -1 & \text{สำหรับค่าอื่น ๆ} \end{cases}$$



รูปที่ 2.6.1 แสดงถึงฮิสโตแกรมของอัตราการเกิดซีโรคลอสซิงของสัญญาณเสียงก้องและเสียงไม่ก้อง