

บทที่ 2 ทฤษฎีที่เกี่ยวข้อง

ทฤษฎีในบทนี้จะกล่าวถึง หลักการที่เกี่ยวข้องกับการกำเนิดเสียงพูด โดยเริ่มตั้งแต่เทคนิคการสังเคราะห์เสียงพูดในแต่ละแบบ และลักษณะทางกายภาพของการเปล่งเสียง ซึ่งจะกล่าวถึงหน้าที่อวัยวะที่ใช้ในการเปล่งเสียงเริ่มตั้งแต่ปอดจนถึงริมฝีปาก รวมถึงปรากฏการณ์การเกิดเสียง ซึ่งการก่อกำเนิดของหน่วยเสียงมีความสัมพันธ์กับลักษณะทางกายภาพโดยตรง รวมถึงการวิเคราะห์เสียงในงานด้านประมวลผลของสัญญาณเสียงเพื่อหาสัมประสิทธิ์ของ ML ของเสียง หลังจากนั้นจะเป็นทฤษฎีเกี่ยวกับเสียงเพื่อนำไปคัดแยกผู้ป่วยอาการ โรคซึมเศร้าโดยการเปรียบเทียบ ML ของคุณลักษณะ MFCC

2.1 เทคนิคการสังเคราะห์เสียงพูดในแต่ละแบบ

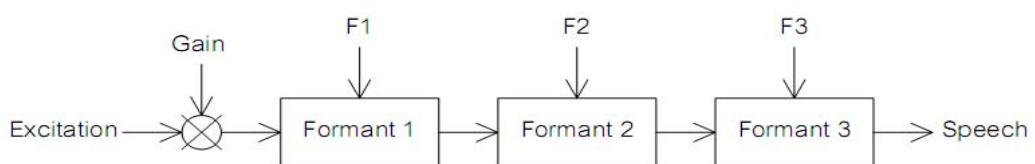
เทคนิคการสังเคราะห์เสียงพูดที่ใช้กันในปัจจุบันนี้มี 3 วิธีคือ

2.1.1 การนำหน่วยเสียงที่ทำการบันทึกเอาไว้มาเล่นใหม่

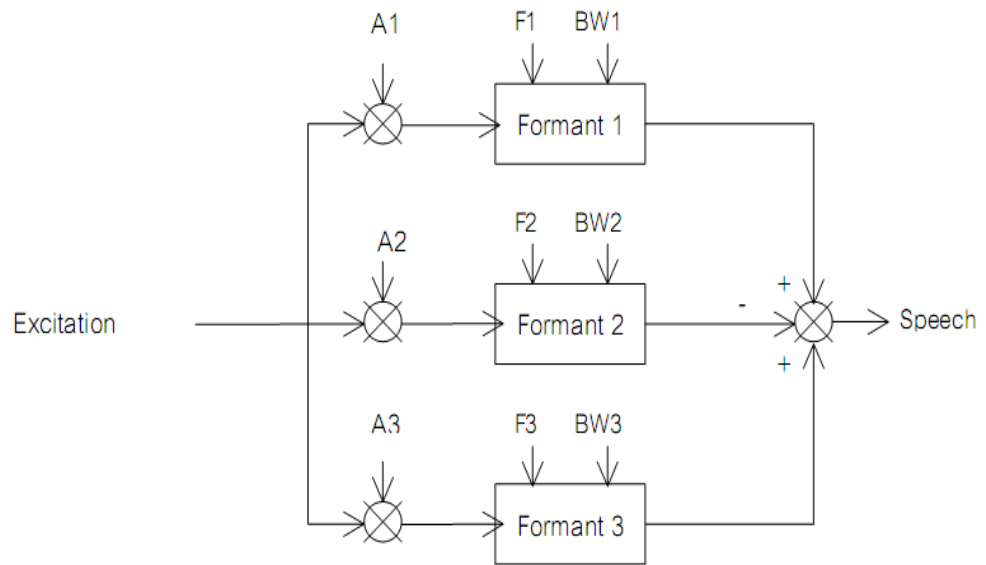
เป็นการนำหน่วยเสียงที่ทำการบันทึกเอาไว้มาเล่นใหม่ (Concatenative synthesis) โดยใช้วิธีการนำหน่วยเสียงที่บันทึกมาต่อกันตัวอย่างของหน่วยเสียงที่ใช้เช่นคำ (Words) พยางค์ (Syllables) หน่วยคู่เสียง (Diphones) และหน่วยเสียงอนุภาค (Micro phonemes) นอกจากนี้อาจจะเก็บ แต่ละหน่วยเสียงให้อยู่ในรูปของค่าพารามิเตอร์ด้วยวิธีการวิเคราะห์และสังเคราะห์ออกมา (Analysis-synthesis method) เมื่อต้องการสังเคราะห์เสียงออกมาก็นำค่าพารามิเตอร์เหล่านี้มาต่อกัน

2.1.2 การวิเคราะห์หาค่าฟอร์แมนต์เพื่อเป็นตัวแทนของหน่วยเสียง

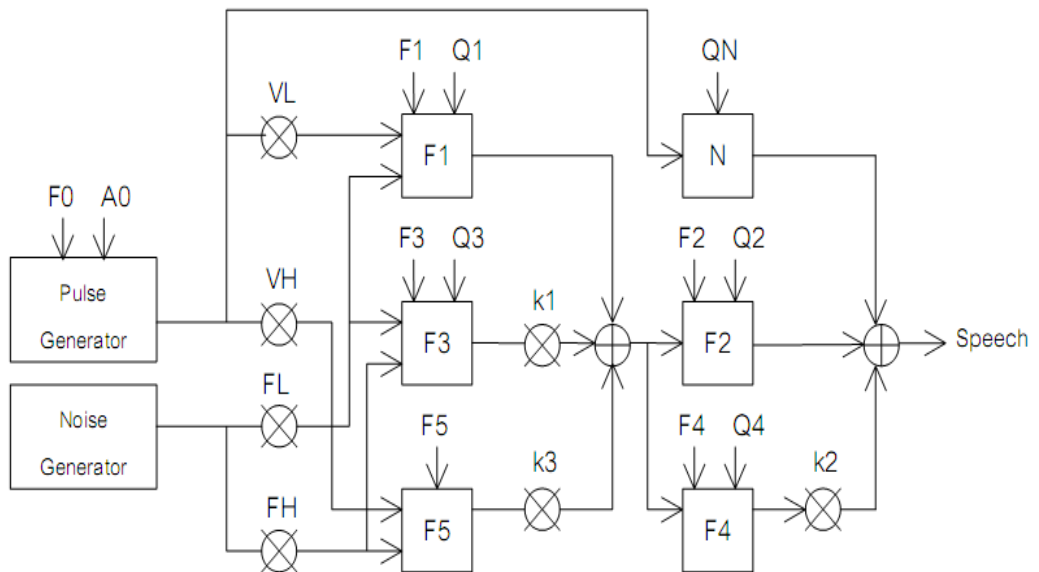
เป็นการวิเคราะห์หาค่าฟอร์แมนต์เพื่อเป็นตัวแทนของหน่วยเสียง (Formant synthesis) ในภาษาเพื่อนำค่าที่ได้ไปเป็นอินพุตให้กับทรานส์เฟอร์ฟังก์ชันของช่องทางเดินเสียง (Vocal tract) เพื่อสังเคราะห์เป็นเสียงออกมา สามารถแบ่งแบบจำลองได้เป็น 3 แบบคือแบบอนุกรมแบบขนาน และแบบผสมดังแสดงในภาพประกอบ



รูปที่ 2.1 ฟอร์แมนต์แบบอนุกรม



รูปที่ 2.2 ฟออร์แมนต์แบบขนาน



รูปที่ 2.3 ฟออร์แมนต์แบบผสม

2.1.3 วิธีจำลองระบบของการเกิดเสียงพูดของมนุษย์

เป็นวิธีจำลองระบบของการเกิดเสียงพูดของมนุษย์ (Articulatory synthesis) ประกอบด้วยแบบจำลองของ 3 ส่วนหลักคือแหล่งกำเนิดเสียง (Sound source production) รูปร่างช่องทางเดินเสียง (Vocal tract) และการแพร่เสียง (Radiation)

2.1.3.1 แหล่งกำเนิดเสียง (Sound source production)

เสียงที่มนุษย์ได้ยินนั้นมียู่มากมาย อาจเกิดจากมนุษย์เองหรือเกิดจากสิ่งแวดล้อม เสียงที่เกิดจากมนุษย์มีทั้งที่เป็นเสียงพูด ไม่เป็นเสียงพูด บ้างก็เป็นเสียงที่สื่อความหมาย บ้างก็ไม่ได้สื่อความหมายใดๆ เสียงพูดของมนุษย์เริ่มต้นจากการหดตัวของกล้ามเนื้อหน้าอก ซึ่งเป็นผลของอากาศจากปอดที่ดันขึ้นมา ผ่านไปยังคู่สายเสียง (Vocal cords) ที่อยู่ภายในกล่องเสียง (Larynx) ทำให้มีการสั่นสะเทือนที่บริเวณดังกล่าว เกิดเป็นเสียงที่ออกมาจากลำคอ แล้วใช้อวัยวะในช่องปากปรับให้กลายเป็นเสียงพูดได้แก่เสียงพูดทั้งพยัญชนะและสระรวมทั้งเสียงร้องเพลง ส่วนเสียงที่ไม่ใช่เสียงพูดอาจยกตัวอย่างได้ เช่น เสียงถอนหายใจ เสียงผิวปาก เป็นต้น

2.1.3.2 ช่องทางเดินเสียง (Vocal tract)

ทางเดินของเสียง (Vocal tract) คือ ส่วนที่ยาวประมาณ 7 นิ้ว จากกล่องเสียงไปจนถึงริมฝีปากนั้นทำงานคล้ายๆ กับท่อออร์แกนขนาดเล็ก ในขณะที่ลำอากาศจากปอดฝักลับเสียงออกไปนั้น เสียงที่ตั้งจะขึ้นอยู่กับความกว้างของช่องกลีบเสียง และแถบสั่นสะเทือนอันเป็นสายใยเหนียวๆ ตามขอบกลีบเสียงนั้นยืดออกไปมากน้อยเพียงใดในเวลาที่เราเปลี่ยนจากเสียงบ่นอุบอปปี้ไปเป็นเสียงร้องหวีดแหลม แถบสั่นสะเทือนดังกล่าวจะยืดออกประมาณ $1/4$ นิ้ว (แถบสั่นสะเทือนของพวกนักร้องโอเปร่าที่ฝึกหัดดีแล้วจะสามารถยืดออกได้ประมาณ $1/2$ นิ้ว) เสียงที่เกิดขึ้นในบริเวณนี้เมื่อผ่านกระบวนการขจัดกลาดกแต่งเสียงจากอวัยวะบางส่วน เช่น ริมฝีปาก ลิ้น ท่อทางเดินอากาศในจมูก และเพดานปาก ก็จะเกิดเป็นคำพูดที่เป็นภาษาของแต่ละชาติที่สามารถสื่อความหมายให้เข้าใจกันได้

2.1.3.3 การแพร่เสียง (Radiation)

เสียงจัดเป็นคลื่นตามยาวที่ต้องอาศัยตัวกลางในการเคลื่อนที่ สามารถแผ่ไปได้ทั้งในตัวกลางที่เป็นของแข็งของเหลว และก๊าซ การแผ่ของคลื่นเสียงเกิดจากการอัดตัวและขยายของโมเลกุลของตัวกลาง ดังนั้นจึงทำให้เสียงแผ่ในตัวกลางที่มีความหนาแน่นมากได้เร็วกว่าในตัวกลางที่มีความหนาแน่นน้อยหรืออาจกล่าวได้ว่าอัตราเร็วของเสียงในตัวกลางที่เป็นของแข็งมีค่ามากกว่าในตัวกลางที่เป็นของเหลว ในขณะที่เดียวกันอัตราเร็วของเสียงในของเหลวก็มีค่ามากกว่าในตัวกลางที่เป็นก๊าซถ้าแหล่งกำเนิดเสียงตั้งอยู่ในบริเวณที่เป็นสุญญากาศจะไม่มีเสียงเกิดขึ้นเพราะในบริเวณนั้นไม่มีโมเลกุลของตัวกลางที่จะอัดหรือขยายเสียงจึงแผ่ออกมาจากแหล่งกำเนิดไม่ได้

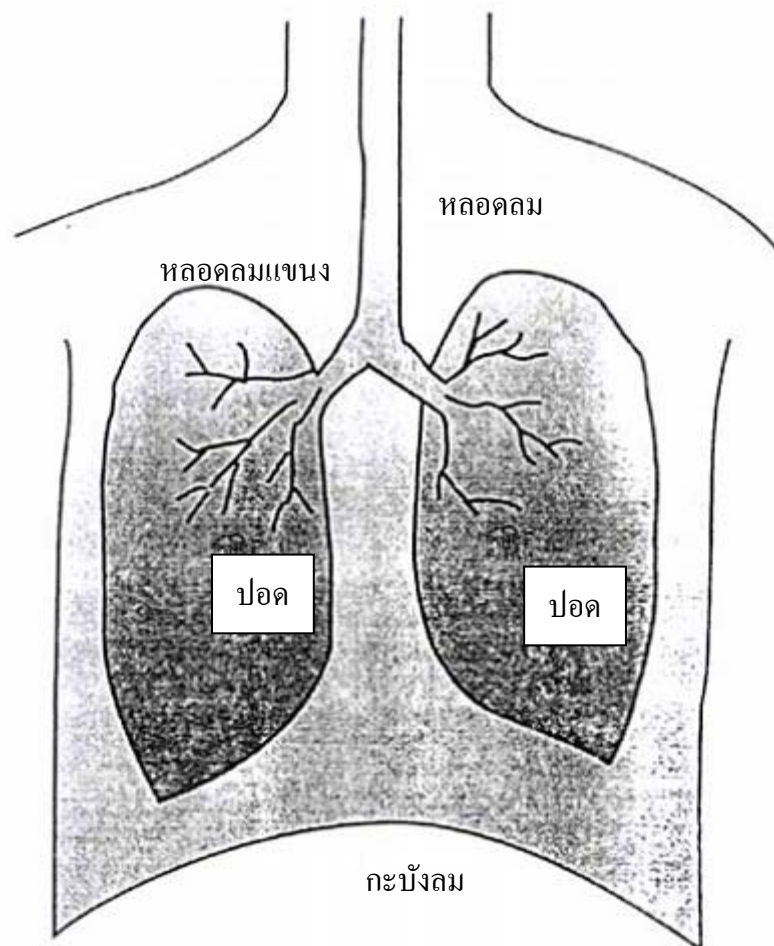
2.2 เสียงและอวัยวะที่ใช้ในการเปล่งเสียงและปรากฏการณ์การเกิดเสียง

ในการศึกษาเรื่องเสียงในภาษาเราจำเป็นต้องเรียนรู้เกี่ยวกับลักษณะหน้าที่และการทำงานของอวัยวะต่างๆที่ใช้ในการเปล่งเสียงเนื่องจากความรู้ดังกล่าวจะช่วยให้เราเข้าใจกระบวนการของการเปล่งเสียง

ซึ่งความเข้าใจนี้จะช่วยให้เราสามารถฟังและจำแนกเสียงในภาษาได้ดีลมนเป็นองค์ประกอบที่สำคัญในการเปล่งเสียงเสียงในภาษาต่างๆส่วนใหญ่จะใช้ลมที่เกิดจากปอดซึ่งเป็นลมหายใจออกอย่างไรก็ตามมีบางภาษาที่เสียงพยัญชนะบางเสียงใช้ลมที่เกิดจาก คอและเป็นลมเข้า (Implosives) เสียงพยัญชนะบางเสียงใช้ลมที่เกิดจากคอเป็นลมออก (Ejectives) พยัญชนะบางเสียงใช้ลมที่เกิดจากปากและเป็นลมเข้า (Clicks) อวัยวะที่ใช้ในการออกเสียงพูดแบ่งออกเป็น 3 กลุ่มได้แก่

2.2.1 ปอดและหลอดลม

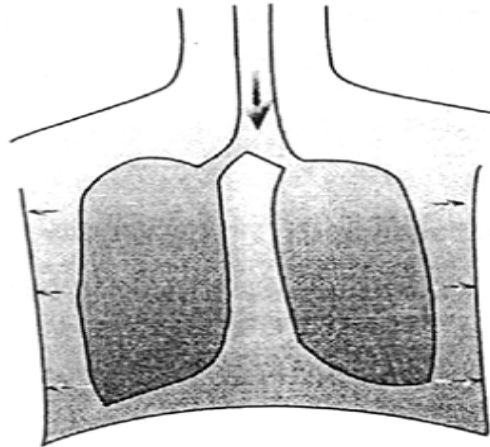
ปอดทั้งสองข้างประกอบด้วยถุงลมขนาดเล็กจำนวนมากปอดมีลักษณะเหมือน ฟองน้ำ ไม่มีกล้ามเนื้อประกอบอยู่ภายในปอดมีหลอดลมฝอยอยู่มากมายหลอดลมฝอยนี้จะมารวม กันเป็นหลอดลมแขนง และหลอดลมแขนงของปอดทั้งสองข้างจะมารวมกันเป็นหลอดลม



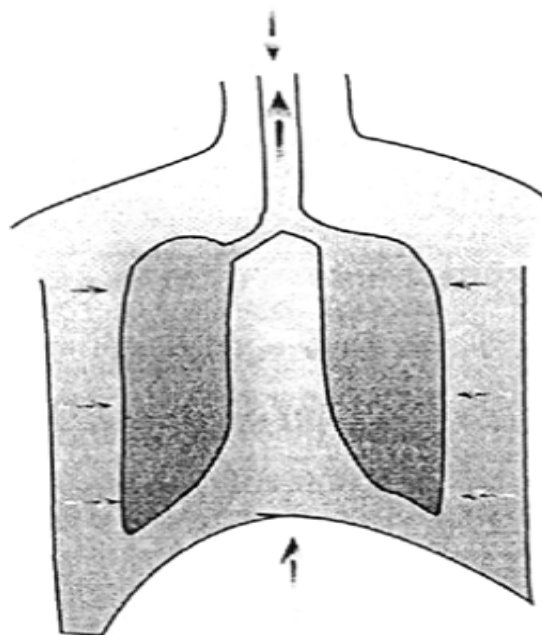
รูปที่ 2.4 ปอดและหลอดลม

ด้านล่างของปอดมีกะบังลมซึ่งเป็นแผ่นกล้ามเนื้อรูปโค้งแผ่นกะบังลมนี้จะเป็นตัวช่วย กดดันให้ช่องอกใหญ่ขึ้นหรือเล็กลงเวลาหายใจกล่าวคือถ้าแผ่นกะบังลมลดตัวลงกระดูกซี่โครงจะ ถูกกล้ามเนื้อดึง

ให้ขยายตัวออกขนาดของช่องอกจะใหญ่ขึ้นความดันของอากาศในช่องอกจะต่ำกว่า ความดันของอากาศภายนอกช่องอกอากาศก็จะไหลจากภายนอกร่างกายเข้าไปในปอดเป็นลมหายใจเข้าเมื่อสิ้นสุดลมหายใจเข้าแล้วช่องอกจะอยู่ในสภาพพักแผ่นกะบังลมจะดันตัวโค้งเล็กน้อย กระดูกซี่โครงจะถูกกล้ามเนื้อระหว่างซี่โครงดึงให้เข้ามาหากันขนาดของช่องอกจะเล็กลงความดันของอากาศในช่องอกจะสูงกว่าความดันภายนอกช่องอกอากาศก็จะไหลจากปอดไปนอกร่างกายเป็น ลมหายใจออก



รูปที่ 2.5 ลมเข้าสู่ปอด

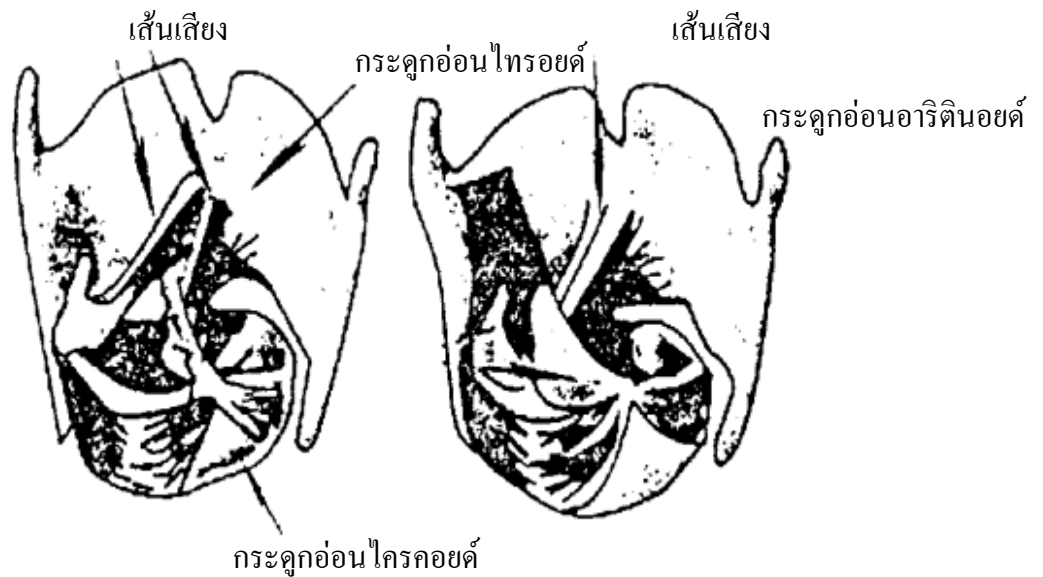


รูปที่ 2.6 ลมออกจากปอด

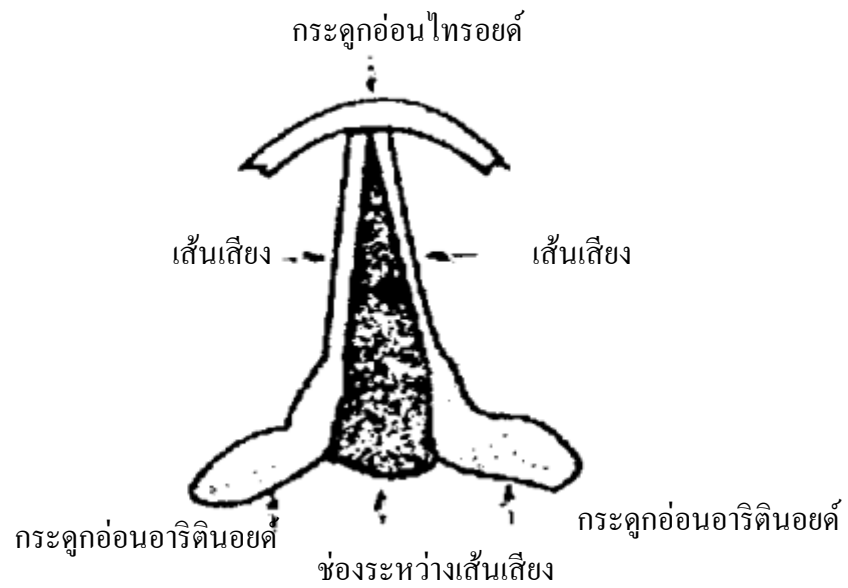
เราใช้ลมหายใจออกเป็นสิ่งสำคัญที่ช่วยให้เกิดเสียงพูด โดยที่ลมจะไหลผ่านหลอดลม เข้าสู่กล่องเสียง ผ่านสู่อวัยวะในช่องปากหรือช่องจมูกอวัยวะเหล่านี้จะดัดแปลงลมให้เป็นเสียงพูด ต่างๆกัน

2.2.2 กล่องเสียงและเส้นเสียง

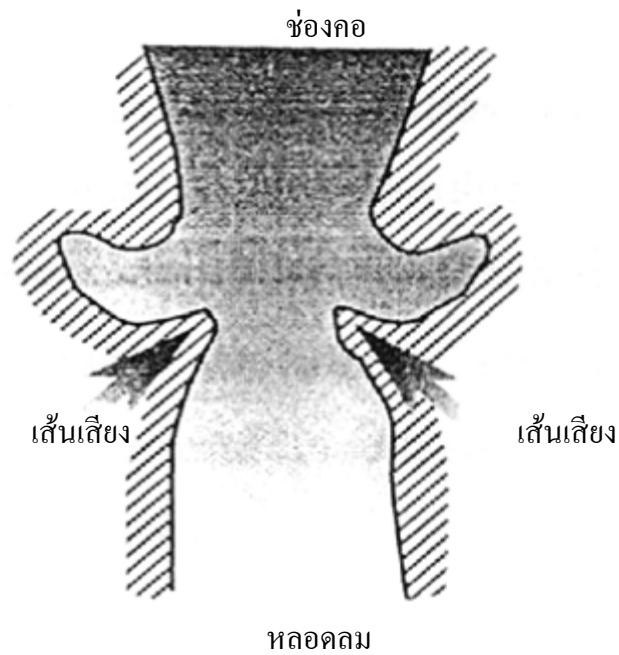
กล่องเสียงประกอบขึ้นด้วยกระดูกกระดูกอ่อนและกล้ามเนื้อหลายชิ้นส่วนหนึ่งของกล่องเสียงคือสิ่งที่เราเรียกกันว่าลูกกระเดือกกล่องเสียงอยู่เหนือหลอดลมด้านล่างของกล่องเสียง ประกอบด้วยกระดูกอ่อนชื่อไครคอยด์ (Cricoids) เป็นกระดูกอ่อนรูปคล้ายแหวนด้านกว้างหันมา ทางด้านหลังของร่างกาย ทางด้านหน้าของกล่องเสียงประกอบด้วยกระดูกอ่อนชื่อไทรอยด์ (Thyroid) เป็นกระดูกอ่อนแผ่นเดี่ยวรูปโป๊วครึ่งวงกลมกระดูกชิ้นนี้เป็นส่วนของกล่องเสียงที่เรา สัมผัสได้และเห็นได้ที่บริเวณคอนอกจากนี้กล่องเสียงยังประกอบด้วยกระดูกอ่อนรูปปิรามิดอีก 1 คู่ชื่อกระดูกอ่อนอาริตिनอยด์ (Arytenoids) กระดูกอ่อนคู่นี้อยู่ด้านหลังของกล่องเสียงบนกระดูกอ่อน รูปวงแหวนส่วนประกอบที่สำคัญที่สุดของกล่องเสียงได้แก่เส้นเสียงซึ่งมีอยู่ 2 เส้นมีลักษณะเป็น แผ่นประกอบด้วยเส้นเอ็นและกล้ามเนื้อเส้นเสียงทั้งคู่พาดตัวตามแนวนอนอยู่ตรงกลางของ กล่องเสียงปลายด้านหนึ่งของเส้นเสียงทั้งสองเส้นอยู่ติดกันและเชื่อมอยู่กับกระดูกอ่อนไทรอยด์ทาง ด้านหน้าของกล่องเสียงปลายอีกด้านหนึ่งของเส้นเสียงแต่ละเส้นเชื่อมอยู่กับกระดูกอ่อนอาริตินอยด์แต่ละอัน



รูปที่ 2.7 ส่วนประกอบของกล่องเสียง



รูปที่ 2.8 บริเวณเส้นเสียงขณะหายใจปกติ

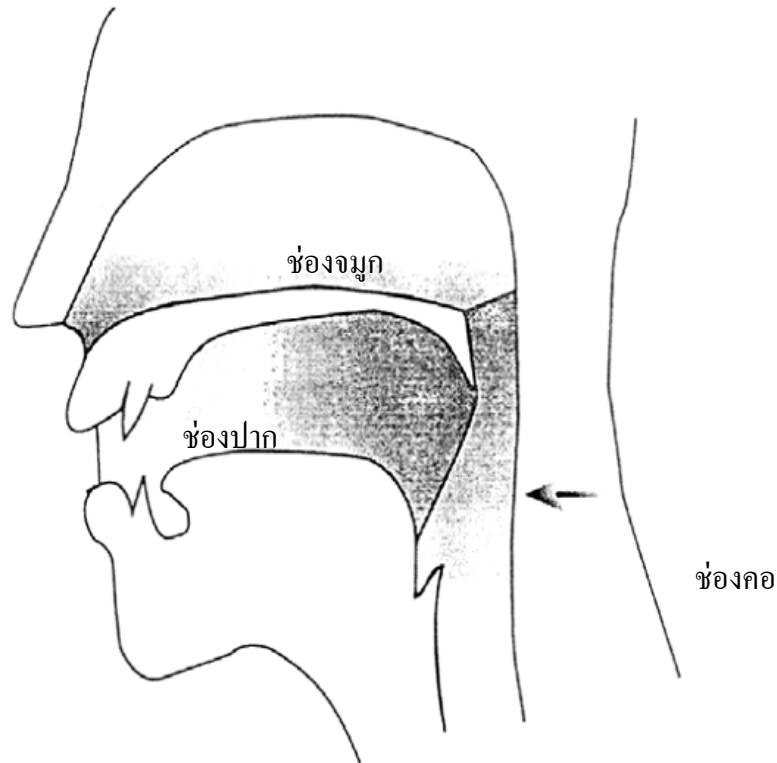


รูปที่ 2.9 ภาพตัดของเส้นเสียงตามแนวตั้งขณะหายใจ

เส้นเสียงมีหน้าที่หลักในการปิดกั้นไม่ให้อาหารตกลงไปในหลอดอาหารในการเปล่งเสียงหน้าที่สำคัญของเส้นเสียงคือเปลี่ยนลมจากปอดให้เป็นคลื่นเสียง

2.2.3 อวัยวะที่อยู่เหนือเส้นเสียง

ลมจากปอดเมื่อผ่านเส้นเสียงแล้วจะผ่านขึ้นมายังช่องคอและจะออกสู่ภายนอกทางช่องปากหรือช่องจมูกช่องใดช่องหนึ่งหรือทั้งสองช่องพร้อมๆกันอาณาเขตของช่องคอช่องปากและช่องจมูกมีดังนี้

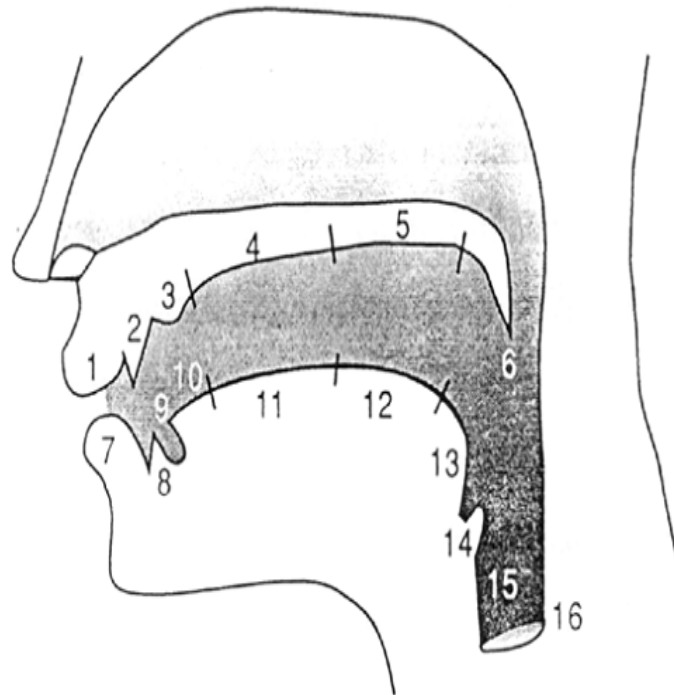


รูปที่ 2.10 อวัยวะที่อยู่เหนือเส้นเสียง

อวัยวะที่อยู่เหนือเส้นเสียงซึ่งมีองค์ประกอบของอวัยวะต่างๆเหล่านี้อาจจำแนกตามหน้าที่ในการช่วยให้เกิดเสียงเป็น 2 ประเภทคือ

- อวัยวะที่เป็นตำแหน่งที่เกิดเสียงต่างๆ (Passive articulator) เรียกอีกอย่างหนึ่งว่า ฐานที่เกิดเสียงเป็นอวัยวะที่ไม่เคลื่อนที่ในขณะที่ออกเสียงอยู่ทางด้านบนของช่องปากได้แก่ริมฝีปากบนฟันบนปุ่มเหงือกเพดานแข็งเพดานอ่อนและลิ้นไก่

- อวัยวะที่เป็นส่วนกระทำอาการ (Active articulator) เรียกอีกอย่างหนึ่งว่ากรณีเป็น อวัยวะที่เคลื่อนที่ขณะออกเสียงอยู่ด้านล่างของช่องปากได้แก่ริมฝีปากล่างฟันล่างปลายลิ้นลิ้นส่วนปลายลิ้นส่วนหน้าและลิ้นส่วนหลัง



รูปที่ 2.11 บริเวณต่างๆของอวัยวะในช่องปาก

- | | |
|--------------------------------|--|
| 1. ริมฝีปากบน (Upper lip) | 9. ปลายลิ้น (The tip of the tongue) |
| 2. ฟันบน (Upper teeth) | 10. ลิ้นส่วนปลาย (The blade of the tongue) |
| 3. ปุ่มเหงือก (Alveolar ridge) | 11. ลิ้นส่วนหน้า (The front of the tongue) |
| 4. เพดานแข็ง (Hard palate) | 12. ลิ้นส่วนหลัง (The back of the tongue) |
| 5. เพดานอ่อน (Soft palate) | 13. โคนลิ้น (The root of the tongue) |
| 6. ลิ้นไก่ (Uvula) | 14. แผ่นเนื้อปากหลอดลม (Epiglottis) |
| 7. ริมฝีปากล่าง (Lower lip) | 15. กรวยคอ (Pharynx) |
| 8. ฟันล่าง (Lower teeth) | 16. เส้นเสียง (Vocal cord) |

คำอธิบายภาพ

1 และ 7 คือ ริมฝีปากบนและริมฝีปากล่างเป็นอวัยวะที่เคลื่อนไหวได้มากริมฝีปากทั้งคู่นี้จะดัดแปลงลมให้เป็นเสียงต่าง ๆ กันโดยปิดริมฝีปาก เปิดเล็กน้อยเปิดกว้างรูปปากแผ่รูปปากห่อ ฯลฯ

2 และ 8 คือ ฟันช่วยกักลมทำให้เกิดเสียงหลายประเภทเช่นเมื่อฟันบนกดลงบนริมฝีปากล่างและ ปล่องยให้ลมเสียดแทรกช่องฟันออกมาเป็นเสียง ฟ [f] หรือใช้ลิ้นส่วนปลายแตะหลังฟันบนทำให้เกิดเสียงต่าง ๆ เช่น ด [d] ต [t] ฯลฯ

3 คือ ปุ่มเหงือกคือส่วนนูนออกมาจากแผ่นเพดานบนอยู่หลังฟันบนเป็นส่วนสำคัญในการเปล่งเสียงพยัญชนะอีกส่วนหนึ่งเช่น ร [r] ล [l] ฯลฯ

4 คือ เพดานแข็งคือส่วนโค้งที่เป็นกระดูกแข็งอยู่ถัดจากปุ่มเหงือกเข้าไปเป็นบริเวณที่เกิดเสียงพยัญชนะอีกส่วนหนึ่งเช่น จ [c] ช [ch] ฯลฯ

5 คือ เพดานอ่อนคือกระดูกที่อยู่ต่อกับเพดานแข็งสามารถเคลื่อนขยับขึ้นลงไปเล็กน้อย เป็นบริเวณที่เกิดเสียงพยัญชนะอีกส่วนหนึ่งเช่น ก [k] ค [kh] ฯลฯ

6 คือ ลิ้นไก่เป็นก้อนเนื้อเล็กๆอยู่ปลายเพดานอ่อนตรงกลางปากสามารถขยับขึ้นลงได้เป็นอวัยวะสำคัญที่ช่วยในการเปล่งเสียงทางจมูกและเสียงที่ไม่ใช่เสียงทางจมูก

9 ถึง 13 คือ ลิ้นเป็นอวัยวะที่เคลื่อนไหวได้มากในการเปล่งเสียงพูดแบ่งออกเป็น 5 ส่วนตามตำแหน่งที่ช่วยทำให้เกิดเสียงต่างๆ กัน

9 คือ ปลายลิ้นคือส่วนปลายสุดของลิ้นสามารถยกขึ้นไปแตะอวัยวะต่างๆ ในปากส่วนบน ได้ช่วยในการเกิดเสียงพยัญชนะ ร [r] ล [l] ฯลฯ

10 คือ ลิ้นส่วนปลายคือส่วนที่ติดกับปลายลิ้นสามารถยกขึ้นไปแตะอวัยวะในปากส่วนบน ได้ช่วยในการเกิดเสียงพยัญชนะ ต [t] ด [d] ฯลฯ

11 คือ ลิ้นส่วนหน้าคือส่วนที่อยู่ตรงกับเพดานแข็งเมื่อวางลิ้นราบช่วยในการเกิดเสียงพยัญชนะ จ [c] ช [ch] และสระหน้า ฯลฯ

12 คือ ลิ้นส่วนหลังคือส่วนของลิ้นที่อยู่ตรงข้ามกับเพดานอ่อนเมื่อวางลิ้นราบช่วยในการเกิดเสียงพยัญชนะ ก [k] ค [kh] และสระหลัง ฯลฯ

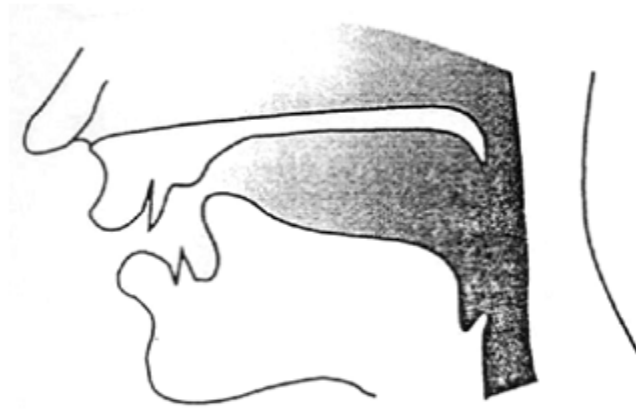
13 คือ โคนลิ้นคือส่วนที่อยู่ต่อจากลิ้นส่วนหลังลงไปในช่องคอ

14 คือ แผ่นเนื้อปากหลอดลมเป็นก้อนเนื้อเล็กๆคล้ายลิ้นไก่อยู่ต่อจากโคนลิ้นลงไปในช่องคอ มีหน้าที่ปิดช่องลมเมื่อกินอาหารและเปิดช่องลมเมื่อพูด

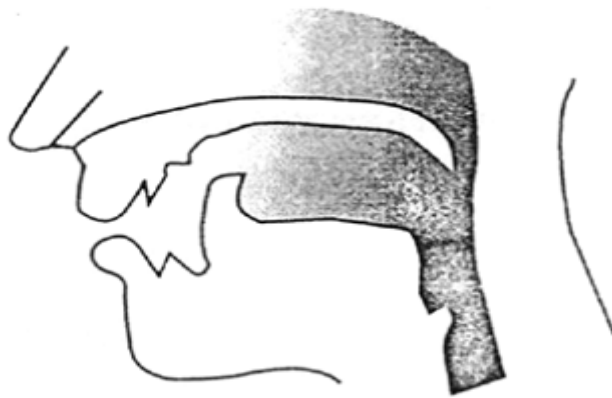
15 คือ กรวยคอคือโพรงคอที่อยู่ถัดจากช่องปากลงไปถึงเส้นเสียง

16 คือ เส้นเสียงเป็นส่วนที่สำคัญที่ช่วยทำให้เกิดเสียงพูด

ลิ้นเป็นอวัยวะที่เคลื่อนไหวได้อย่างคล่องแคล่วในหลายทิศทางทั้งกระดกขึ้น กระดกลงแลบไปข้างหน้าดึงไปข้างหลังนอกจากนี้ปลายลิ้นยังเคลื่อนตัวได้มากกว่าส่วนอื่นอาจ กระดกขึ้นไปในทิศทางของฟันบนและปุ่มเหงือกหรืออปลายลิ้นกลับเข้าไปด้านในและด้านล่างของลิ้นกระดกขึ้นไปในทิศทางของฟันบนปุ่มเหงือกเพดานแข็งทำให้เกิดเสียงต่าง ๆ กันได้มาก



รูปที่ 2.12 ปลายลิ้นกระดกขึ้น โดยไม่งอกลับ



รูปที่ 2.13 ปลายลิ้นกระดกขึ้นและงอกลับ

นอกจากนี้ลักษณะของลิ้นยังแบ่งออกเป็น 3 ลักษณะ

- แบบกลม
- แบบแบนราบ
- แบบมีร่องตรงกลาง



กลม



แบนราบ



มีร่องตรงกลาง

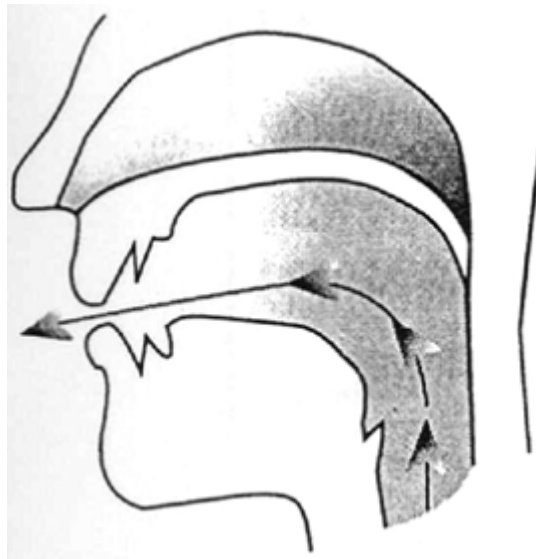
รูปที่ 2.14 ปลายลิ้นกระดกขึ้นและงอกลับ

ริมฝีปากมีส่วนในการคัดแปลงลมให้เป็นเสียงต่างๆกัน โดยริมฝีปากอาจปิดสนิท เปิดเล็กน้อยเปิดกว้างห่อกลมหรือทำเป็นรูปรีก็ได้

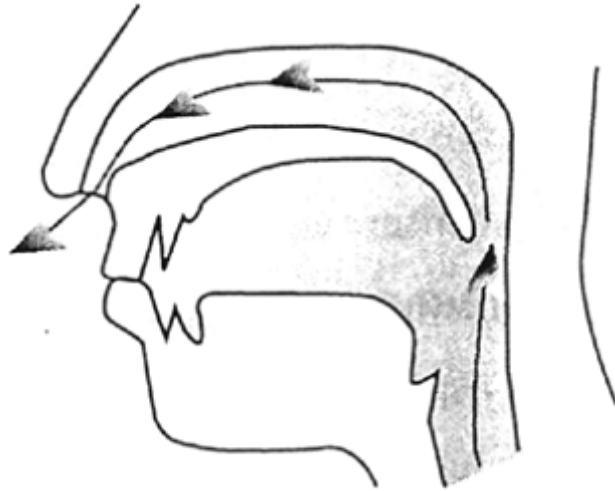


รูปที่ 2.15 ริมฝีปากลักษณะต่าง ๆ

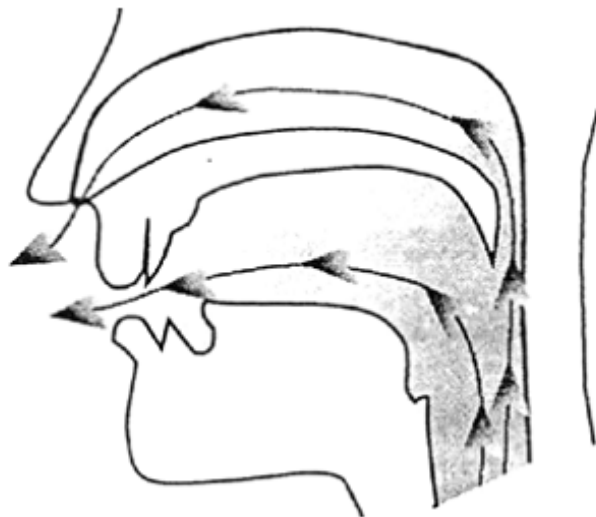
ลิ้นไก่และเพดานอ่อนเป็นส่วนที่เคลื่อนไหวได้เล็กน้อยโดยอาจยกขึ้นไปติดกับผนัง คอทำให้ลมจากปอดเคลื่อนไปสู่ช่องจมูกไม่ได้หรืออาจลดตัวลงปล่อยให้ลมจากปอดเข้าสู่ช่องจมูก



รูปที่ 2.16 ลมออกทางช่องปาก



รูปที่ 2.17 ลมออกทางช่องช่องจมูก



รูปที่ 2.18 ลมออกทางช่องปากและช่องจมูก

หากเพดานอ่อนและลิ้นไก่เคลื่อนไปติดกับผนังคอลมจากปอดจะผ่านไปทางช่องจมูกไม่ได้ต้องผ่านทางช่องปากเท่านั้น เสียงที่เกิดขึ้นก็เป็นเสียงที่เกิดจากช่องปากแต่ถ้าเพดานอ่อนและลิ้นไก่ลดลงอยู่ในลักษณะพัก ธรรมชาติลมจากปอดจะผ่านขึ้นไปทางช่องจมูกได้เสียงที่เกิดขึ้นจะมี 2 ประเภทคือ

- ถ้ามีการปิดกั้นลม ณ ที่ใดที่หนึ่งในช่องปากลมจะผ่านออกได้เฉพาะทางจมูกเสียงที่เกิดขึ้นเป็นเสียงทางจมูก (Nasal sounds) ซึ่งเกิดขึ้นได้ที่เสียงท้ายพยางค์ เช่นคำว่าลมเสียง "ม" จะปรากฏที่พยางค์ท้ายโดยลักษณะของเพดานอ่อนและลิ้นไก่จะลด ระดับลงมาปิดกั้นไม่ให้ลมผ่านออกทางปาก

- ถ้าไม่มีการปิดกั้นลมในช่องปากลมจะผ่านออกได้ทั้งทางปากทั้งทางจมูกเสียงที่เกิดขึ้น เป็นเสียงขึ้นจมูกซึ่งจะเกิดกับเสียงสระที่ติดกับเสียงทางจมูกและในบาง ภาษาจะมีสระทั้งที่เป็นเสียงสระขึ้นจมูกและเสียงสระที่ออกทางช่องปากโดยการเปล่งเสียงริมฝีปาก ทั้งคู่อยู่ติดกันก็กลมไว้ใน

ขณะเดียวกันเพดานอ่อนและลิ้นไก่ลดระดับลงอยู่ในลักษณะพับทำให้ลม ผ่านออกทางช่องจมูกเมื่อเปิดที่กักลมที่ริมฝีปากลมจะออกทางปากด้วย

2.3 ลักษณะผู้ป่วยในอาการซึมเศร้า

ความรู้สึกเศร้าและเสียใจเป็นความรู้สึกที่พบได้เป็นธรรมดาในคนปกติทั่วไปในเวลาใดเวลาหนึ่ง อาการซึมเศร้า (Depressive symptom) เป็นผลรวมของความรู้สึกเศร้าว่างเปล่า และเฉื่อยชาที่รุนแรงกว่าความรู้สึกเศร้าเสียใจธรรมดา อาการซึมเศร้าอาจเกิดจากสาเหตุในสิ่งแวดล้อมหรือเกิดร่วมกับโรคทางกายหรือโรคทางจิต

2.3.1 ภาวะซึมเศร้า

ภาวะซึมเศร้าสามารถเกิดขึ้นได้กับทุกเพศทุกวัย ที่อยู่ในภาวะดังกล่าวเริ่มตั้งแต่เด็กจนถึงผู้สูงวัยล้วนแต่เกิดภาวะซึมเศร้าได้ตลอดเวลา

- ภาวะซึมเศร้าในเด็กวัยเรียนแยกจากจากปัญหาอารมณ์อื่นๆ เช่น ความกังวลซึ่งมักพบร่วมกัน เด็กจะไม่ถูกพามาพบแพทย์เพราะเด็กบอกว่าเศร้าหรือผู้ปกครองคิดว่าซึมเศร้า แต่จะแสดงอาการอื่นๆ ที่แพทย์จะต้องนึกถึงภาวะซึมเศร้าเช่น อาการทางกาย แบบปวดศีรษะ ปวดท้อง การเรียนเลวลงเพราะไม่มีสมาธิ อารมณ์หงุดหงิดหรืออารมณ์เสียกับเรื่องเล็กน้อยๆ แยกตัว ไม่ร่าเริง และดูเหมือนหมดเรี่ยวแรงเด็กมักไม่มีปัญหาการกินและนอนเหมือนวัยรุ่นหรือผู้ใหญ่แต่บางรายอาจแสดงอาการของภาวะซึมเศร้าอย่างชัดเจน เช่น เฉื่อยชา พุดและทำอะไรช้า (Retardation) เบื่ออาหาร ขาดความมั่นใจ ความคิดและการพยายามฆ่าตัวตายพบน้อยในเด็กแต่เมื่อถามเด็กอาจบอกว่าอยากตาย

- เด็กวัยเรียนมักไม่บอกว่าเศร้า แต่อาจบ่นว่าเบื่อ ไม่มีอะไรน่าสนใจหรือบ่นว่าคนรอบข้าง เช่น พ่อแม่ ครู ไม่รักหรือไม่ยุติธรรมอาการเบื่อและไม่รู้สึกสนุกในเด็กที่ซึมเศร้าจะขึ้นตลอดเวลาต่างจากความเบื่อแบบธรรมดาเวลาที่เด็กไม่มีอะไรทำเป็นบางครั้งความรู้สึกว่าคนรอบข้างไม่รักและไม่ยุติธรรม แสดงถึงความไม่มั่นใจในตัวของเด็กซึ่งเมื่อถามเด็กมักตอบว่า ตัวเองรู้สึกที่ตัวเองไม่ดีหรือไม่เก่งเท่าเพื่อนหรือพ่อแม่รักพี่หรือน้องมากกว่าเด็ก

- ภาวะซึมเศร้าในวัยรุ่นจะมีอาการคล้ายคลึงกับภาวะซึมเศร้าในวัยผู้ใหญ่วัยรุ่นบอกได้ถึงความรู้สึกเศร้า เฉื่อยชา ไม่สนใจสิ่งแวดล้อมและความรู้สึกเหมือนหมดเรี่ยวแรงจะทำอะไร พบปัญหาการกินและการนอนได้บ่อยบางรายเบื่ออาหาร แต่บางรายจะกินจุมากขึ้น ปัญหาการนอนอาจเป็นได้ตั้งแต่นอนไม่หลับตื่นกลางคืน หรือตื่นเช้ากว่าปกติ ซึ่งทำให้วัยรุ่นรู้สึกเหนื่อยในตอนกลางวันวัยรุ่นที่ซึมเศร้าหลายรายจะหมกมุ่นและคิดมากเกี่ยวกับปัญหาสุขภาพและรูปร่างหน้าตาู้สึกว่าตัวเองไม่มีคุณค่าและมีความท้อแท้ในชีวิตในรายที่เป็นรุนแรงอาจมีความคิดหรือพยายามที่จะฆ่าตัวตาย

2.3.2 สาเหตุการเกิดภาวะซึมเศร้า

สาเหตุภาวะซึมเศร้าเกิดจากหลายสาเหตุแต่ละสาเหตุจะมีผลกระทบซึ่งกันและกันสาเหตุของภาวะซึมเศร้าที่ได้มีการศึกษาค้นพบประกอบด้วย

- สาเหตุทางพันธุกรรม พันธุกรรมมีบทบาทสำคัญมากในภาวะซึมเศร้าที่พบสลับกับอาการ Mania (Bipolar disorder) ใน Unipolar depression แม้จะพบบทบาทน้อยกว่าใน Bipolar แต่ก็มีส่วนจากหลายการศึกษาถึงบทบาทของพันธุกรรม โดยเฉพาะภาวะซึมเศร้าที่เกิดขึ้นครั้งแรกในวัยเด็กหรือวัยรุ่นการศึกษาเด็กที่พ่อแม่มีภาวะซึมเศร้าพบว่ามียีนการเกิดภาวะซึมเศร้าสูงกว่าเด็กทั่วไป ในทำนองเดียวกันการศึกษานิวคลีอไทป์ของเด็กที่ป่วยด้วยโรคซึมเศร้าพบว่ามียีนการป่วยด้วยโรคซึมเศร้าสูงกว่านิวคลีอไทป์ของเด็กทั่วไป

- สาเหตุทางจิตสังคมปัจจัยเสี่ยงทางจิตสังคมหลายอย่างเป็นสาเหตุให้เด็กเกิดโรคซึมเศร้าการที่เด็กมีแม่ที่มีปัญหาทุกข์ใจเลี้ยงดูลูกได้ไม่ดีหรือมีครอบครัวที่มีปัญหาทะเลาะเบาะแว้ง ปัญหาเศรษฐกิจและที่อยู่อาศัยทำให้เด็กมีความเสี่ยงที่จะเกิดโรคซึมเศร้า ปัจจัยเสี่ยงอื่นๆ ได้แก่เหตุการณ์ในชีวิตที่เกิดขึ้นกะทันหัน เช่น การเสียชีวิตหรือการหย่าร้างของพ่อแม่และเหตุการณ์เลวร้ายในชีวิตที่เรื้อรัง เช่น การถูกรังแกทางร่างกายและทางเพศการถูกทอดทิ้ง

- กลไกทางจิตใจ มีทฤษฎีหลายทฤษฎีที่พยายามอธิบายสาเหตุของภาวะซึมเศร้า Seligman กล่าวถึงการเรียนรู้ว่าไม่มีอะไรจะช่วยให้ (Learned helplessness) ซึ่งอธิบายว่าการที่เราพยายามหลีกเลี่ยงสิ่งรอบตัวที่ไม่ดีหลายครั้งแต่ไม่สำเร็จอาจทำให้ขาดแรงจูงใจ เฉยชา และซึมเศร้าได้ Beck กล่าวถึงทฤษฎีทางความคิดที่เป็นสาเหตุของภาวะซึมเศร้าว่าคนที่มีความคิดในทางลบเกี่ยวกับตัวเอง โลก และอนาคตมีความเสี่ยงที่จะเกิดภาวะซึมเศร้ากลไกทางจิตใจอื่นที่มีการศึกษาพบว่าเกี่ยวข้องกับภาวะซึมเศร้าในเด็ก ได้แก่ความรู้สึกว่าตนเองไม่มีคุณค่า ไร้ความสามารถ และไม่เป็นที่ยอมรับของเพื่อน

- กลไกทางชีววิทยาการสังเคราะห์ที่ทำให้เกิดภาวะซึมเศร้าและการค้นพบยาที่รักษาโรคซึมเศร้า นำไปสู่สมมติฐานของสารสื่อประสาท Monoamines ที่ว่าภาวะซึมเศร้าเกิดจากสารดังกล่าวลดลง หลายการศึกษาพบความผิดปกติของฮอร์โมนที่อาจเป็นสาเหตุของภาวะซึมเศร้า เช่น เด็กที่ซึมจะไม่ตอบสนองแบบปกติต่อ Dexamethasone suppression test, Thyroid releasing hormone test, การหลั่งของ Growth hormone ภายหลัง Hypoglycemia หรือให้ Clonidine นอกจากนี้มีการศึกษาพบความผิดปกติของคลื่นสมองขณะหลับในเด็กที่ซึมเศร้า

2.3.3 การช่วยเหลือ

การช่วยเหลือรักษาโรคซึมเศร้าประกอบด้วยการรักษาหลายแบบ เช่น การใช้ยา จิตบำบัดการเปลี่ยนแปลงสิ่งแวดล้อม การให้คำแนะนำ ในรายที่เป็นไม่รุนแรงการนัดเด็กหรือวัยรุ่นและครอบครัวมาพบแพทย์สม่ำเสมอ โดยแพทย์ให้คำแนะนำที่ถูกต้องเกี่ยวกับโรคและการจัดการเกี่ยวกับ

ความเครียดที่เป็นสาเหตุของโรคร่วมกับการให้กำลังใจและช่วยเหลือประคับประคองทางจิตใจก็อาจเพียงพอที่จะทำให้เด็กดีขึ้นในรายที่เป็นรุนแรงอาจต้องรักษาด้วยยาหรือเข้ารับการรักษาตัวในโรงพยาบาล

- การลดความเครียดผู้รักษาต้องพยายามลดความเครียดที่เป็นเหตุกระตุ้นให้ซึมเศร้า เช่นเด็กที่ซึมเศร้าจากการสอบตกซ้ำๆการให้ความช่วยเหลือเรื่องการเรียนรู้จะช่วยรักษาภาวะซึมเศร้าซึ่งควรทำควบคู่ไปกับการรักษาโดยวิธีอื่นๆ

- จิตบำบัดการให้คำปรึกษาแนะนำ (Counseling) แบบตัวต่อตัวจะช่วยให้เด็กรับรู้ถึงปัจจัยที่เกี่ยวข้องกับการเกิดภาวะซึมเศร้าจิตบำบัดแบบตัวต่อตัว (Individual psychotherapy) จะช่วยเด็กจัดการเกี่ยวกับความขัดแย้งภายในจิตใจ ครอบครัวบำบัด (Family therapy) จะช่วยแก้ไขการสื่อสารในครอบครัวซึ่งจะช่วยให้เด็กและสมาชิกในครอบครัวแบ่งปันและบอกเล่าความรู้สึกซึ่งทำให้ความรู้สึกรู้สึกของเด็กดีขึ้น

- วิธีปรับเปลี่ยนความคิดและพฤติกรรม (Cognitive behavioral therapy) ทำโดยเน้นให้เด็กหรือวัยรุ่นรับรู้วิถีคิดหรือพฤติกรรมที่ทำให้ซึมเศร้า เช่น ความคิดว่าตัวเองไม่มีคุณค่า ถูกคนอื่นดูถูกนำมาพูดคุยกับผู้รักษาและเปลี่ยนวิถีคิดให้เหมาะสมการรักษาวิธีนี้อาจทำได้ทั้งแบบเดี่ยวหรือตัวต่อตัวและแบบกลุ่มและอาจทำร่วมกับการฝึกทักษะทางสังคม (Social skill training) ในรายที่มีปัญหาการใช้ยาควบคู่กับจิตบำบัดในรายที่เป็นรุนแรง

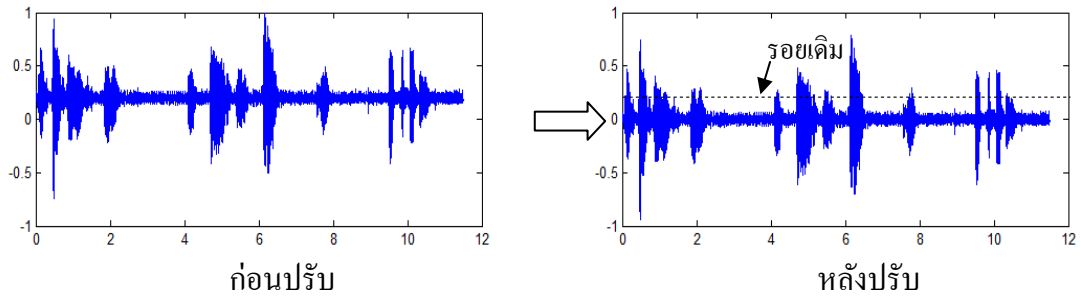
2.4 ทฤษฎีการวิเคราะห์เสียงพูด

การประมวลผลสัญญาณเบื้องต้น (Signal Preprocessing) เป็นขั้นตอนการประมวลผล สัญญาณเบื้องต้นเพื่อจัดเตรียมข้อมูลให้เหมาะต่อการประมวลผลในขั้นตอนต่อไปกระบวนการ ประมวลผลสัญญาณเบื้องต้นจะต้องอาศัยหลักการต่างๆดังนี้

2.4.1 การปรับสัญญาณสู่แกนศูนย์

ในการบันทึกเสียงแต่ละครั้งจะได้ระดับเสียงเฉลี่ย ต่างกันออกไปไม่เท่ากันจำเป็นต้องปรับระดับเสียงเฉลี่ยให้เท่ากันก่อนคือปรับให้เข้าสู่แกนศูนย์ได้ตามสมการ (2-1) และรูปที่ 2.19

$$Signal = Signal - mean(Signal) \quad (2-1)$$



รูปที่ 2.19 การปรับสัญญาณสู่แกนศูนย์

2.4.2 ขั้นตอนกรรมวิธีการเน้นล่วงหน้า

ขั้นตอนกรรมวิธีการเน้นล่วงหน้า (Pre-emphasis) เป็นการบีบอัดสัญญาณเสียงพูดในช่วงพิสัยพลวัต (Dynamic Range) มีผลทำให้อัตราส่วนสัญญาณต่อสัญญาณรบกวน (Signal to Noise Ratio) มีค่าสูงขึ้นโดยนำสัญญาณเสียงพูดผ่านวงจรกรองดิจิทัลอันดับที่หนึ่ง (First-Order Digital Filter) ที่มีฟังก์ชันถ่ายโอนดังแสดงในสมการที่ (2-2) และสมการที่ (2-3) และรูปที่ 2.20 แสดงการตอบสนองต่อความถี่ของสมการ (2-3) ในการกำหนดค่า α ต่างๆซึ่งโดยทั่วไปค่า α จะมีค่าอยู่ระหว่าง 0.9 ถึง 1

$$H(z) = 1 - \alpha Z^{-1} \tag{2-2}$$

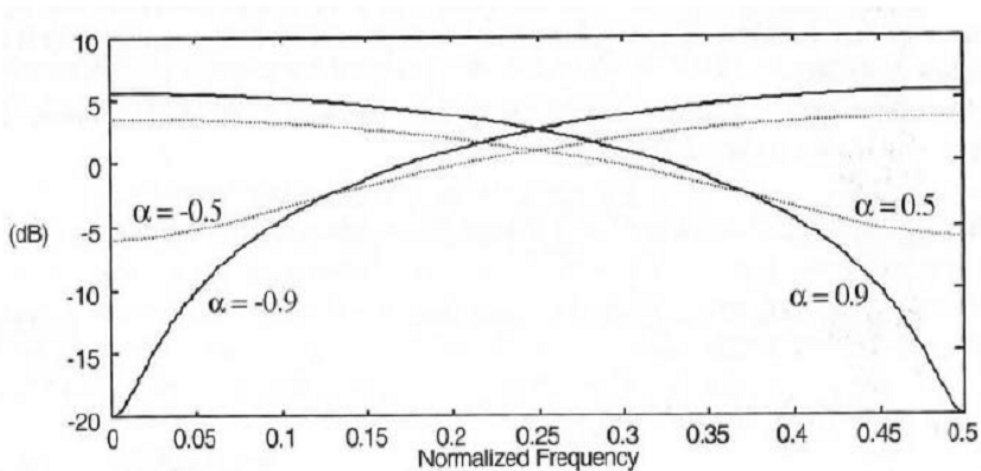
$$\tilde{S}(n) = S(n) - \alpha S(n-1) \tag{2-3}$$

เมื่อ α เป็นค่าสัมประสิทธิ์ของวงจรกรอง

เป็นค่าของสัญญาณเสียงที่พูดขาออกที่ผ่านกรรมวิธีการเน้นล่วงหน้า

เป็นค่าของสัญญาณเสียงพูดขาเข้าที่ n

และ $S(n-1)$ เป็นค่าของสัญญาณเสียงพูดขาเข้าที่ $n-1$



รูปที่ 2.20 การตอบสนองต่อความถี่เมื่อเลือกใช้ค่า α ต่างๆกัน [13]

2.4.3 การวางกรอบสัญญาณ

เนื่องจากเสียงพูดโดยรวมมีลักษณะไม่คงที่ และมีการเปลี่ยนแปลงอย่างช้าๆไปตามเวลาจึงจำเป็นต้องวิเคราะห์สัญญาณเสียงทีละช่วงสั้นๆเฉพาะที่อยู่ภายในกรอบสัญญาณ (Windowing) ที่กำหนดขึ้น ขนาดของกรอบสัญญาณที่นิยมใช้ในการวิเคราะห์สัญญาณเสียงมีค่าประมาณ 10-30 มิลลิวินาทีและควรวางกรอบสัญญาณถัดไปให้มีความเหลื่อมกันประมาณ $1/2$ ถึง $1/3$ ของขนาดกรอบสัญญาณ เพื่อให้ข้อมูลที่วิเคราะห์มีความ ต่อเนื่องกันกรอบสัญญาณหลักมี 3 ชนิด

- กรอบสัญญาณสี่เหลี่ยม (Rectangular Window) ดังสมการที่ (2-4)

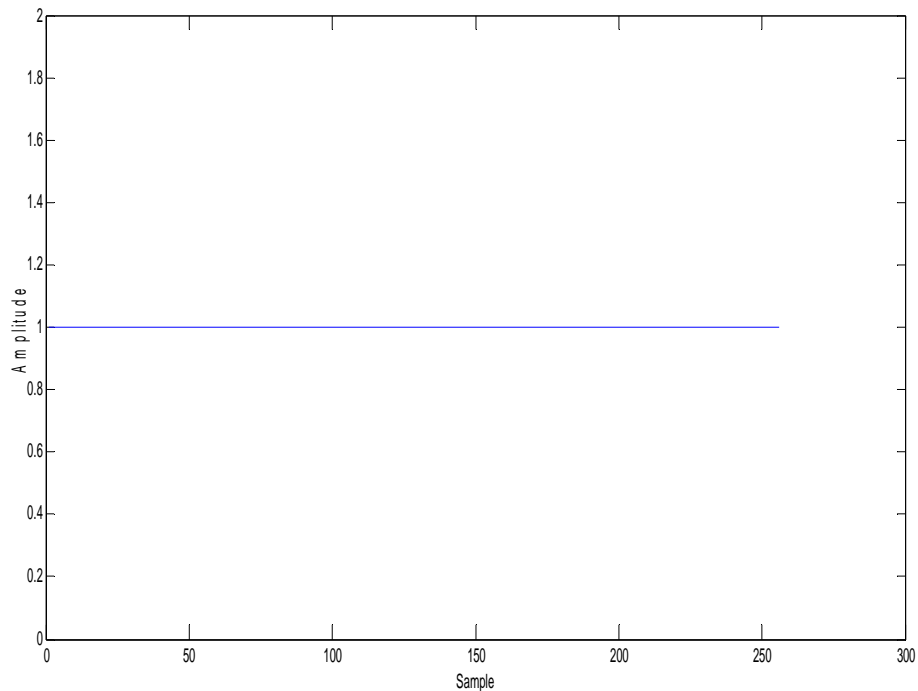
$$\omega(n) = \begin{cases} 1, & 0 \leq n \leq N-1 \\ 0, & \text{Otherwise} \end{cases} \quad (2-4)$$

เมื่อ $\omega(n)$ คือ ผลลัพธ์ของฟังก์ชันกรอบตำแหน่งที่ n

N คือ ความกว้างหน้าต่าง

n คือ ข้อมูลในหน้าต่างมีค่าตั้งแต่ 0 จนถึง $N-1$

ซึ่งมีลักษณะดังรูปที่ 2.21



รูปที่ 2.21 ฟังก์ชันกรอบสัญญาณสี่เหลี่ยม

- กรอบสัญญาณแฮมมิง (Hamming Window) ดังสมการที่ (2-5)

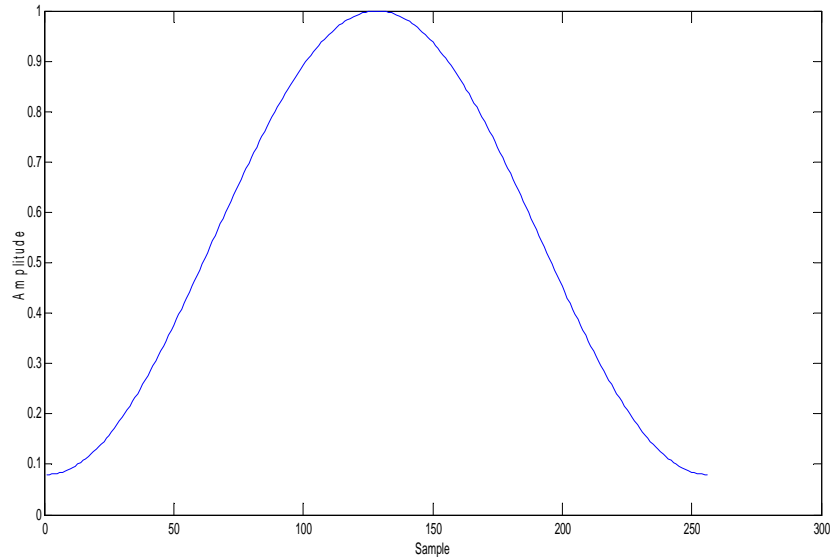
$$\omega(n) = \begin{cases} 0.54 - 0.46 \cos \frac{2\pi n}{N-1}, & 0 \leq n \leq N-1 \\ 0, & \text{Otherwise} \end{cases} \quad (2-5)$$

เมื่อ $\omega(n)$ คือ ผลลัพธ์ของฟังก์ชันกรอบตำแหน่งที่ n

N คือ ความกว้างหน้าต่าง

n คือ ข้อมูลในหน้าต่างมีค่าตั้งแต่ 0 จนถึง $N-1$

ซึ่งมีลักษณะดังรูปที่ 2.22



รูปที่ 2.22 ฟังก์ชันกรอบสัญญาณแฮมมิง

- กรอบสัญญาณแฮนนิ่ง (Hanning Window) ดังสมการที่ (2-6)

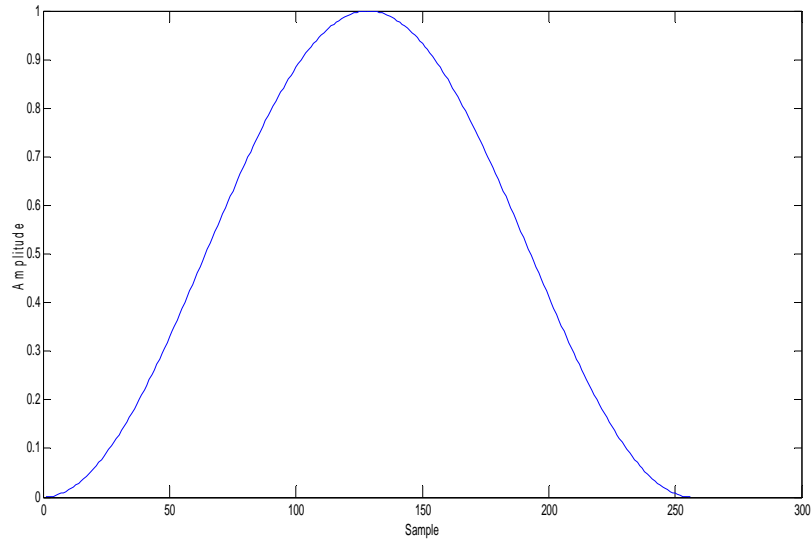
$$\omega(n) = \begin{cases} 0.5 \left[1 - \cos \frac{2\pi n}{N-1} \right], & 0 \leq n \leq N-1 \\ 0, & \text{Otherwise} \end{cases} \quad (2-6)$$

เมื่อ $\omega(n)$ คือ ผลลัพธ์ของฟังก์ชันกรอบตำแหน่งที่ n

N คือ ความกว้างหน้าต่าง

n คือ ข้อมูลในหน้าต่างมีค่าตั้งแต่ 0 จนถึง $N-1$

ซึ่งมีลักษณะดังรูปที่ 2.23



รูปที่ 2.23 ฟังก์ชันกรอบสัญญาณแฮนนิ่ง

กรอบสัญญาณในแบบแฮมมิงและแฮนนิ่งเหมาะสำหรับการประมวลผลสัญญาณเสียงเพราะสามารถเน้นสัญญาณเสียงในกรอบที่กำลังพิจารณาให้มีความสำคัญสูงสุดโดยลดความสำคัญของสัญญาณเสียงที่อยู่ในกรอบรอบข้างแต่ยังคงความต่อเนื่องของสัญญาณเสียงให้มีความต่อเนื่องกันในแต่ละกรอบสัญญาณทำให้เสียงที่ผ่านการวางกรอบสัญญาณยังคงข้อมูลครบถ้วนอยู่ [4]

2.4.4 การตัดหัวท้ายค่าโดยใช้ค่าพลังงาน

เป็นการวิเคราะห์หาพลังงานในช่วงเวลาสั้นๆ (Short-time energy) เนื่องจากช่วงเสียงพูดมีค่าพลังงานมากกว่าช่วงที่ไม่ใช่คำพูดการหาค่าพลังงานสามารถเขียนเป็นสมการได้ ดังสมการที่ (2-7)

$$E(n) = \sum_{m=-\alpha}^{\alpha} [x(m)w(n-m)]^2 \quad (2-7)$$

หรือเขียนเป็น

$$E(n) = \sum_{m=-\alpha}^{\alpha} x^2(m)h(n-m)$$

เมื่อ $h(n) = w^2(n)$ โดยที่ E คือค่าพลังงานที่เวลา n

การหาค่าพลังงานตามวิธีการนี้ จะมีความไวต่อสัญญาณที่มีแอมพลิจูดขนาดใหญ่ๆ และใช้เวลาคำนวณมาก เนื่องจากการยกกำลังสอง ดังนั้นจึงแก้ปัญหาโดยการเปลี่ยนมาใช้ค่าสัมบูรณ์ (Absolute) แทนค่ากำลังสองแทนตามสมการ (2-8) ผลการวิเคราะห์ที่ได้จะเรียกว่าพลังงานสัมบูรณ์แบบสี่เหลี่ยมและแบบสามเหลี่ยมสำหรับวิทยานิพนธ์นี้ เลือกใช้การวางกรอบแบบสามเหลี่ยม [8]

$$M(n) = \sum_{m=-\alpha}^{\alpha} |x(m)| w(n-m) \quad (2-8)$$

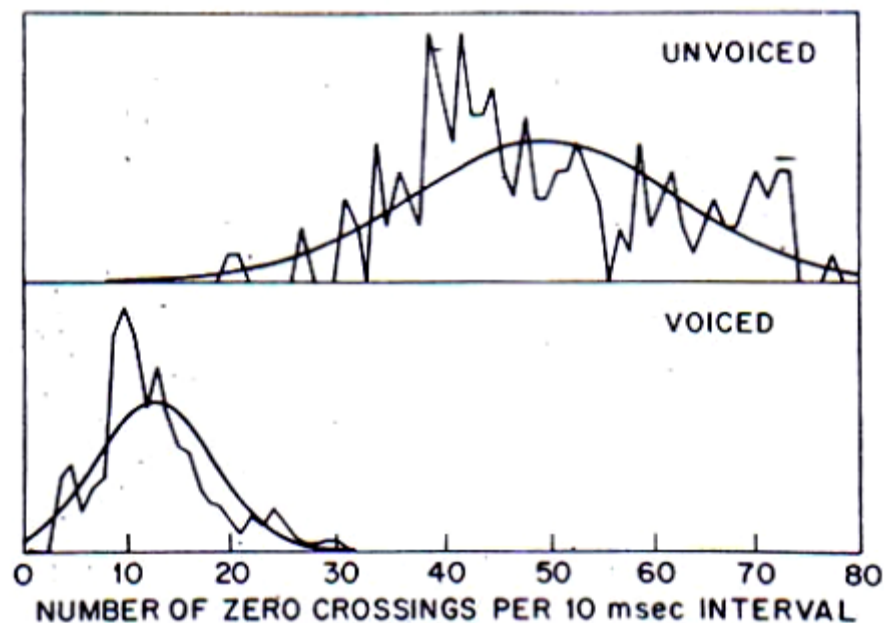
เมื่อ $M(n)$ คือ พลังงานสัมบูรณ์ที่เฟรม n

$x(m)$ คือ สัญญาณเสียงที่ผ่านการปรับคู่แกนศูนย์แล้ว

$w(n)$ คือ หน้าต่างกรอบสัญญาณโดยทั่วไปนิยมวางฟังก์ชันกรอบสัญญาณ

2.4.5 อัตราการตัดศูนย์

อัตราการตัดศูนย์ (Zero Crossing Rate) คือจำนวนครั้งของ สัญญาณเสียง ที่ตัดแกนศูนย์ในช่วงเวลาใดๆ ภายในหน้าต่างที่กำหนดปกติสัญญาณ เสียงที่มีค่า จุดตัดสูงจะเป็นเสียงไม่ก้องและสัญญาณเสียงที่มีค่าจุดตัดต่ำจะเป็นเสียงก้องแต่อย่างไรก็ตาม การกำหนดขนาดของค่าจุดตัดศูนย์ที่แน่นอนเพื่อจำแนกชนิดของเสียงนั้นจะต้องอาศัยผลจากการ ทดลองเป็นหลักในรูปที่ 2.24 จะแสดงให้เห็นถึงการกระจายของค่าจุดตัดศูนย์ของเสียงไม่ก้องและก้อง



รูปที่ 2.24 การกระจายของค่าจุดตัดศูนย์ของเสียงไม่ก้องและเสียงก้อง [13]

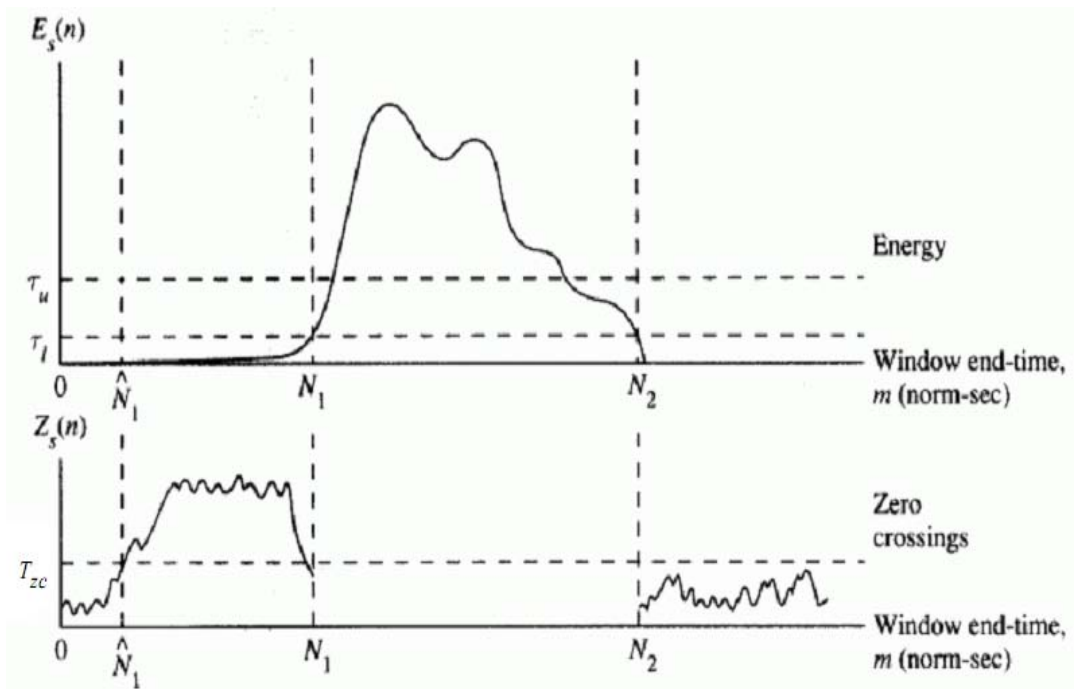
จุดตัดศูนย์สามารถนิยามได้ดังสมการที่ (2-9)

$$Z(n) = \sum_{m=-\alpha}^{\alpha} |\text{sgn}[x(m)] - \text{sgn}[x(m-1)]| w(n-m) \tag{2-9}$$

เมื่อ

$$\text{sgn}[x(n)] = \begin{cases} 1, & x(n) \geq 0 \\ -1, & x(n) < 0 \end{cases} \tag{2-10}$$

โดยที่ $w(n)$ คือ ฟังก์ชันกรอบสัญญาณ ที่เหลี่ยมที่ใช้กำหนดรูปร่างในการพิจารณาของ สัญญาณเสียง $x(m)$ ในหนึ่งเฟรมและ N คือจำนวนตัวอย่างทั้งหมดของสัญญาณเสียงที่อยู่ในเฟรม



รูปที่ 2.25 การตัดหัวท้ายคำโดยใช้ค่าพลังงานและอัตราการตัดศูนย์ร่วมกัน [13]

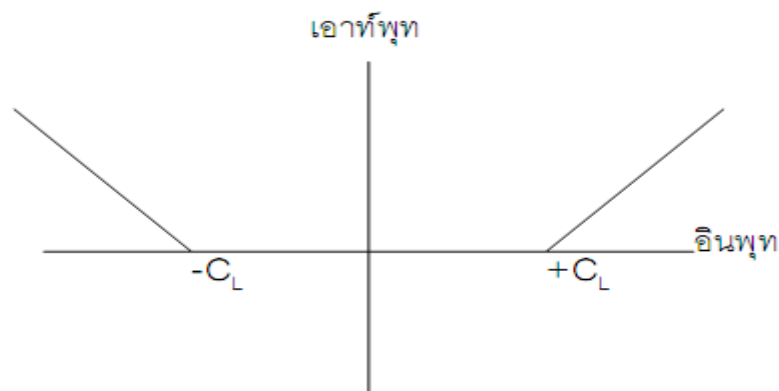
จากรูปที่ 2.25 แสดงการตัดหัวท้ายคำโดยใช้ค่าพลังงานและอัตราการตัดศูนย์ร่วมกันเพื่อหาจุดเริ่มต้นและสิ้นสุดของเสียงพูดการหาตำแหน่งเริ่มต้นและสิ้นสุดใช้พลังงานสัมบูรณ์ โดยใช้ระดับอ้างอิง T_u และ T_l เป็นตัวกำหนดตำแหน่งเริ่มต้นและสิ้นสุดของเสียงพูดตำแหน่งแรกของพลังงานที่เริ่มมากกว่า T_u ถือเป็นตำแหน่งเริ่มต้นของคำหลังจากนั้นถ้าตำแหน่งแรกของพลังงานที่ เริ่มต้นต่ำกว่าตำแหน่ง T_l ถือเป็นตำแหน่งสิ้นสุดของคำ (จากรูปที่ 2.25 คือจุด N_1 และ N_2) ขึ้น ต่อไปพิจารณาช่วงต้นของสัญญาณ (จากรูปที่ 2.25 คือตำแหน่ง N_1 ถึง N_2) โดยใช้อัตราการตัดศูนย์ของสัญญาณมาช่วยวิเคราะห์เพื่อให้ครอบคลุมสัญญาณเสียงที่พลังงานต่ำกว่าค่าที่กำหนดโดย T_u ด้วยการเผื่อสัญญาณไปทางซ้าย

มือและใช้การพิจารณาจากอัตราการตัดศูนย์ของสัญญาณตั้งแต่ช่วง N_1 ถึง N_2 หากอัตราการตัดศูนย์ของสัญญาณที่ตำแหน่งใดมีค่ามากกว่าหรือเท่ากับค่า Track zero crossing (T_{zc}) ก็จะกำหนดตำแหน่งนั้นเป็นตำแหน่งเริ่มต้นของเสียงพูดแทนตำแหน่งเดิมค่า T_{zc} สามารถคำนวณได้โดยสมมติให้ 10 กรอบหน้าต่างแรกเป็นช่วงเสียงเงียบและกำหนดให้มี อัตราการตัดศูนย์คงที่ในช่วงเวลาใดๆภายในหน้าต่าง [21] เช่นกำหนดอัตราการตัดศูนย์เป็น 25 ครั้งในช่วงเวลา 10 ms จากนั้นหาคำนวณค่าเฉลี่ย (I_{zc}') และส่วนเบี่ยงเบนมาตรฐาน (σ) ของ อัตราการตัดศูนย์ในช่วง 10 กรอบหน้าต่างแรกแล้วคำนวณ T_{zc} ได้จาก

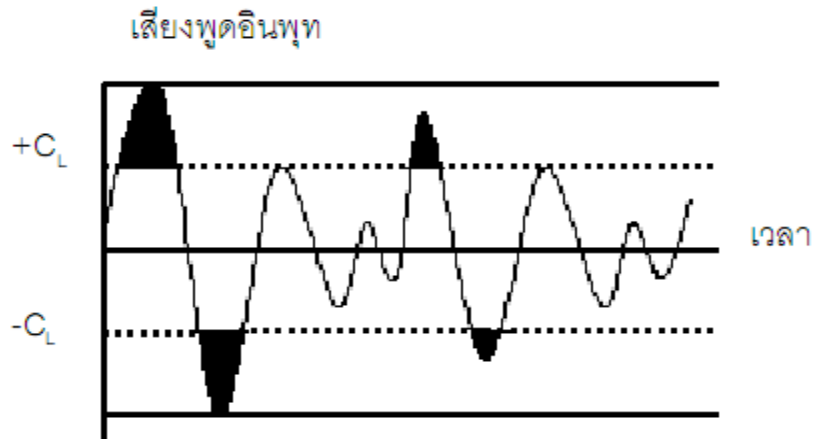
$$T_{zc} = \min\left(\frac{25}{10ms}, I_{zc}' + 2\sigma\right) \quad (2-11)$$

2.4.6 เทคนิคเซนเตอร์คลิปปีง

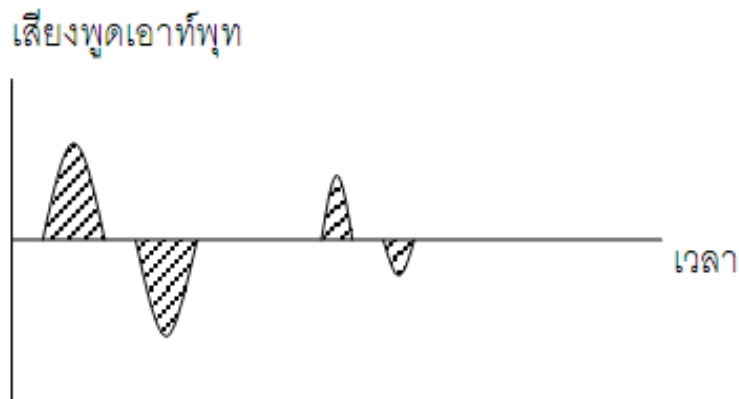
ปัญหาสำคัญประการหนึ่งในการหาความถูกต้องของพิชซ์ของเสียงในช่วงแถบความถี่ที่กว้างของสัญญาณก็คือผลของโครงสร้างฟอร์แมนท์ที่มีการวัดสัมพันธ์กับคาบของรูปคลื่นสัญญาณ ดังนั้นเพื่อเพิ่มความถูกต้องในการหาพิชซ์ของเสียงจะต้องทำการลดผลของฟอร์แมนท์นี้ลงให้มากที่สุดเท่าที่จะทำได้ซึ่งเทคนิคเซนเตอร์คลิปปีงเป็นวิธีที่นิยมเพราะนำมาประยุกต์ใช้งานได้สะดวก วิธีการนี้คือการตรวจจับระดับของสัญญาณว่าถ้ามีค่าต่ำกว่าระดับคลิปปีงที่กำหนดก็จะ กำหนดให้สัญญาณตรงนั้นมีค่าเป็นศูนย์ส่วนสัญญาณที่ยังมีค่าสูงกว่าระดับคลิปปีงก็จะเอามาหัก ลบด้วยค่าระดับคลิปปีงเหลือแต่ส่วนที่เกินไปจากระดับคลิปปีงเท่านั้นรูปที่ 2.26, 2.27 และ 2.28 แสดงถึงลักษณะสมบัติของสัญญาณอินพุต-เอาต์พุตที่นำมาทำกระบวนการเซนเตอร์คลิปปีงซึ่งจะเห็นได้ว่า ถ้ามีการกำหนดระดับคลิปปีงที่เหมาะสมแล้วโครงสร้างของรูปคลื่นที่เป็นส่วนของ ฟอร์แมนท์ต่างๆจะถูกกำจัดออกไปได้เกือบทั้งหมดซึ่งจะมีผลทำให้สัญญาณที่มีสเปคตรัมราบเรียบที่ได้ออกมามองเห็นคาบของมันได้ง่ายกว่าสัญญาณที่ยังไม่ได้ผ่านกระบวนการทำให้สเปคตรัมราบเรียบ



รูปที่ 2.26 ระดับคลิปปีง



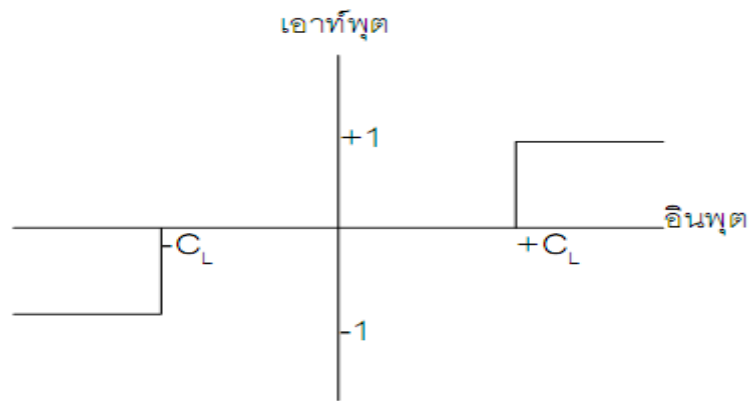
รูปที่ 2.27 เสียงพูดอินพุท



รูปที่ 2.28 เสียงพูดเอาต์พุท

ขั้นตอนแรกของการประมวลผลสำหรับแต่ละเฟรมนั้นก็จะทำการคำนวณระดับคลิปปิงของแต่ละเฟรมเนื่องจากเสียงเป็นสัญญาณที่มีช่วงการเปลี่ยนแปลงกว้างมากดังนั้นการเลือกระดับคลิปปิงจะต้องทำด้วยความระมัดระวังเพื่อป้องกันการสูญหายของข้อมูลที่สำคัญบางส่วนของรูปคลื่นในขณะที่รูปคลื่นมีแอมพลิจูดที่กำลังมีขนาดเพิ่มขึ้นหรือกำลังลดลงในระหว่างเฟรมนั้นๆกรณีแบบนี้จะเกิดขึ้นในช่วงที่เสียงกำลังจะเริ่มขึ้นหรือกำลังจะจบลงหรืออาจจะเป็นตรงรอยต่อของช่วง การเปลี่ยนแปลงของเสียงสมมติว่าเราเลือกให้ระดับคลิปปิงมีค่า CL สัญญาณเสียงแต่ละเฟรมมีขนาด 30 มิลลิวินาทีจะถูกแบ่งออกเป็น 3 ส่วนๆละ 10 มิลลิวินาทีในส่วนแรกและส่วนสุดท้าย อัลกอริทึมจะทำการหาค่าสมบรูณ์ที่จุดยอดสูงสุดเก็บเอาไว้สำหรับแต่ละส่วนค่าระดับคลิปปิงก็จะกำหนดเอาจากค่าเปอร์เซ็นต์ที่แน่นอนของค่าสมบรูณ์ที่จุดยอดสูงสุดที่น้อยที่สุดระหว่างสองค่านั้นตามวิธีการของ Sondhi จะกำหนดค่าระดับคลิปปิงไว้ที่ 30 เปอร์เซ็นต์ของค่าสมบรูณ์ที่จุดยอด ซึ่งจะใช้เป็นระดับอ้างอิงกับข้อมูลของเสียงแต่ละเฟรม

หลังจากการหาระดับคลิปปิงได้แล้วจะนำสัญญาณมาตัดช่วงที่อยู่ระหว่างระดับคลิปปิง นั้นทิ้งไปซึ่งลักษณะนี้จึงเรียกว่าเซนเตอร์คลิปปิง (Center clipping) ดังนั้นผลที่ได้ออกมากำหนด ให้เป็น 3 ค่าคือ +1, -1 สำหรับสัญญาณที่มีค่าเกินกว่าระดับคลิปปิงด้านบวกและต่ำกว่าด้านลบตาม ลำดับและค่า 0 สำหรับสัญญาณที่อยู่ในช่วงดังกล่าวจะเห็นได้ว่าไม่ว่าค่าของสัญญาณตัวอย่างจะมี ค่ามากหรือน้อยแค่ไหนก็จะถูกทำให้มีค่าเท่ากับ +1 และ -1 เหมือนกันหมดรูปประกอบ 2.29 แสดงให้เห็นถึงลักษณะของสัญญาณเสียงขาเข้าและขาออกเมื่อผ่านการเซนเตอร์คลิปปิงเพื่อนำผล มาทำการถอดคอคอริเลชันต่อไปผลของการเซนเตอร์คลิปปิงนี้จะเป็นการลดความยุ่งยากในการ คำนวณถอดคอคอริเลชันลงอย่างมากเหมาะสำหรับการนำไปปฏิบัติในขั้นตอนการเขียนโปรแกรม คอมพิวเตอร์



รูปที่ 2.29 ลักษณะของสัญญาณเสียงขาเข้าและขาออกเมื่อผ่านการเซนเตอร์คลิปปิง

2.5 การสกัดค่าคุณลักษณะสำคัญ

เป็นขั้นตอนซึ่งทำหน้าที่แปลงสัญญาณเสียงที่อยู่ในรูปดิจิทัลให้อยู่ในรูปตัวแทนที่มีความเหมาะสมในการวิเคราะห์คุณลักษณะของเสียง การสกัดค่าคุณลักษณะสำคัญที่เหมาะสม ของสัญญาณเสียงจึงนับเป็นขั้นตอนที่สำคัญมาก เพราะถ้าสามารถสกัดค่าคุณลักษณะสำคัญที่บ่งบอกความแตกต่างระหว่างสัญญาณเสียงแต่ละเสียงของผู้ป่วยได้ดีในขณะเดียวกันถ้ายังใช้จำนวนค่าลักษณะเด่นน้อยการวิเคราะห์ก็จะมีประสิทธิภาพก็จะลดลงในวิทยานิพนธ์นี้ เลือกใช้เทคนิคการสกัดค่าคุณลักษณะสำคัญ ทั้งที่วิเคราะห์จาก โดเมนความถี่

2.5.1 การคำนวณสเปคตรัม

นำสัญญาณเสียงที่ผ่านขั้นตอนการเน้นล่งหน้าแล้วมาวางกรอบ หน้าต่างแฮมมิงหรือแฮนนิ่งนำสัญญาณภายในหน้าต่างมาแปลงฟูริเยร์ไม่ต่อเนื่องแบบเร็วเพื่อเปลี่ยนเป็น โดเมนความถี่ตามสมการที่

(2-12)

$$\tilde{x}(k) = \sum_{n=0}^{N-1} x(n) \cdot w(n) e^{-j2\pi nk/N}; 0 \leq k < N \tag{2-12}$$

เมื่อ N คือ จำนวนข้อมูลในกรอบสัญญาณ

$x(n)$ คือ สัญญาณเสียงพูด

k คือ ลำดับของกรอบสัญญาณ

$w(n)$ คือ ฟังก์ชันกรอบสัญญาณ

แต่เนื่องจากข้อมูลเข้าเป็นจำนวนจริงทำให้ผลที่ได้สมมาตรจึงพิจารณาเพียงครึ่งเดียว

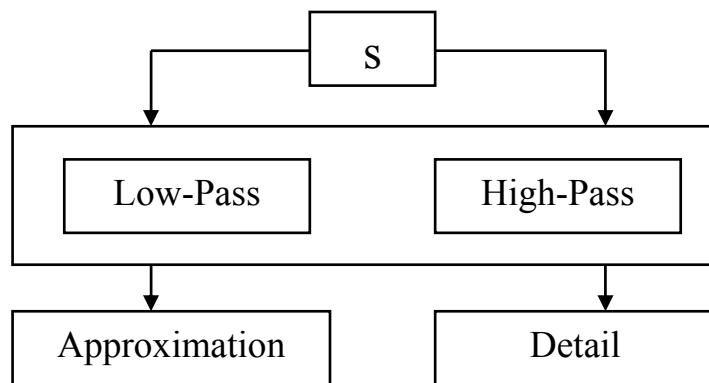
$$0 \leq k < \frac{N}{2} - 1 \tag{2-13}$$

2.5.2 การวิเคราะห์แบบตัวกรองสัญญาณ (Filter Bank Analysis)

การวิเคราะห์ตัวกรองสัญญาณนั้นจะทำการแยกสัญญาณต้นแบบเป็น 2 ส่วนคือส่วนที่มีความถี่ต่ำกับส่วนที่มีความถี่สูงโดยจะใช้ตัวกรองความถี่ต่ำผ่าน (Low Pass Filter) สำหรับแยกองค์ประกอบส่วนที่มีความถี่ต่ำและใช้ตัวกรองความถี่สูงผ่าน (High Pass Filter) สำหรับแยกองค์ประกอบความถี่สูงโดยสัญญาณที่ได้มานั้นจะผ่านอัตราสุ่มลง (Down Sampling) ด้วย 2 ซึ่งจะทำให้ผลลัพธ์จากการวิเคราะห์ลดลงครึ่งหนึ่งในการวิเคราะห์แต่ละขั้น โดยสรุปคือเมื่อนำสัญญาณต้นแบบมาผ่านการวิเคราะห์ด้วยการแปลงเวฟเล็ตแบบเต็มหน่วยแล้วจะทำให้สัญญาณต้นแบบถูกแยกองค์ประกอบด้วยตัวกรองความถี่สูงและตัวกรองความถี่ต่ำซึ่งจะแยกความถี่ได้ตามช่วงที่ต้องการและผลที่ได้จะถูกลดอัตราสุ่มครึ่งหนึ่งทำให้ได้สัญญาณเป็น 2 ส่วนคือ

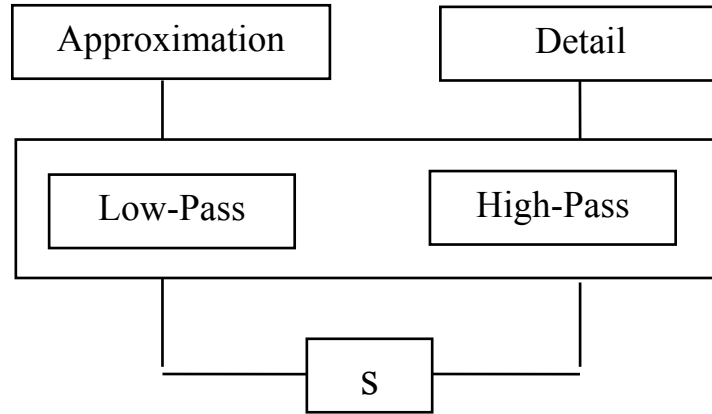
- สัญญาณองค์ประกอบความถี่สูงที่เรียกว่า “Detail”
- สัญญาณองค์ประกอบความถี่ต่ำที่เรียกว่า “Approximation”

ดังรูปที่ 2.30



รูปที่ 2.30 การแยกสัญญาณด้วยตัวกรองแบบ 2 ช่องสัญญาณ

นอกจากนี้ในโครงสร้างที่กลับกันการวิเคราะห์แบบสร้างกลับสองช่องสัญญาณ (Reconstruction 2 Channel Synthesis Filter Bank) จะนำส่วนของสัญญาณความถี่สูงและความถี่ต่ำมารวมกันเป็นสัญญาณต้นแบบได้ซึ่งกระบวนการนี้ว่าการสังเคราะห์สัญญาณ (Synthesis) โดยได้แสดงกระบวนการไว้ดังรูปที่ 2.31



รูปที่ 2.31 การแปลงกลับสัญญาณด้วยตัวกรองแบบ 2 ช่องสัญญาณ

และเมื่อเอาตัวกรองแบบสองช่องสัญญาณมาเรียงต่อกัน โดยใช้สัญญาณทางด้านความถี่ต่ำมาใช้ในการแยกองค์ประกอบอีกครั้งหนึ่ง

2.5.3 คำนวณหาพลังงานสเปกตรัมที่ผ่านตัวกรอง

ขั้นตอนนี้นำความถี่ที่ได้จากขั้นตอนที่ 1 มาหาขนาดกำลังสองได้ $\left| \tilde{x}(k) \right|^2$ ส่งผ่านชุดตัวกรองเพื่อเน้นความสำคัญของความถี่ที่อยู่ในช่วงกลาง ของชุดตัวกรองแต่ละตัวกรองตามสมการที่ (2-14)

$$E(j) = \sum_{k=0}^{n-1} \phi(j).(k) \left| \tilde{x}(k) \right|^2 ; 0 \leq j \leq J \tag{2-14}$$

2.5.4 การคำนวณค่าพลังงานเสียง

ความแตกต่างของพลังงานระหว่างหน่วยของเสียงจะแสดงว่าพลังงานนี้เป็นคุณสมบัติที่ดีในการแบ่งแยกระดับหน่วยของเสียงพลังงานจะถูกประมวลผลภายในหน้าต่างสี่เหลี่ยม $N=256$ ในลักษณะลอการิทึมของสัญญาณพลังงานเสียงซึ่งสามารถคำนวณได้จากสมการที่ (2-15)

$$E(v) = \log_{10} \sum_{n=1}^N S^2(n) \tag{2-15}$$

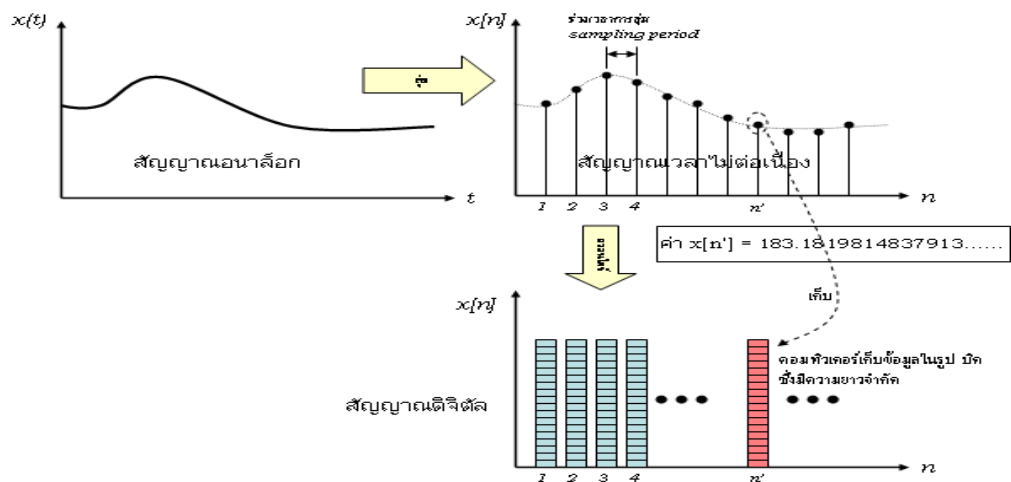
2.5.5 การประมวลผลสัญญาณดิจิทัล (DSP - Digital Signal Processing)

การประมวลผลสัญญาณดิจิทัล (DSP - digital signal processing) เป็นการศึกษาการประมวลผลสัญญาณที่อยู่ในรูปดิจิทัล (digital) โดยทั่วไป การประมวลผลสัญญาณ อาจแบ่งได้ตาม

1. รูปแบบของตัวแทนสัญญาณ: การประมวลผลสัญญาณดิจิทัล (Digital signal processing) และ การประมวลผลสัญญาณแอนะล็อก (Analog signal processing)
2. คุณสมบัติของสัญญาณ: การประมวลผลสัญญาณไม่สุ่ม (Deterministic signal processing) และ การประมวลผลสัญญาณสุ่ม (Stochastic/statistical signal processing)
3. ลักษณะการประมวลผลสัญญาณ: เชิงเส้น (Linear signal processing) และ ไม่เป็นเชิงเส้น (nonlinear signal processing)
4. แบ่งตามคุณลักษณะเฉพาะของสัญญาณ หรือ ลักษณะเฉพาะของการประมวลผล เช่น Adaptive signal processing, mutilate/multiresolution signal processing, chaotic signal processing ฯลฯ

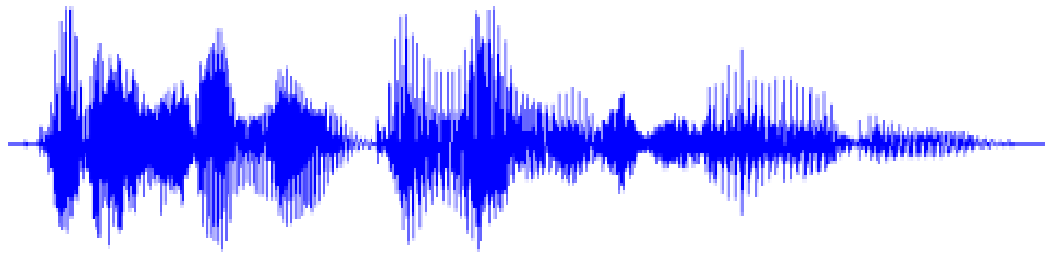
DSP นี้ อาจแบ่งออกได้เป็นในส่วนของซอฟต์แวร์และฮาร์ดแวร์หรือตามการประยุกต์เป็นการประมวลผลสัญญาณเสียง (audio signal processing) การประมวลผลภาพดิจิทัล (digital image processing) และการประมวลผลคำพูด (speech processing) ถึงแม้ว่าใน DSP นั้น สัญญาณที่เราพิจารณากันจะเป็นดิจิทัล แต่โดยทั่วไปสัญญาณเหล่านี้จากแหล่งกำเนิดจะอยู่ในรูปเดิมที่เป็นแอนะล็อก การได้มาซึ่งสัญญาณดิจิทัลซึ่งเป็นตัวแทนสัญญาณแอนะล็อกที่เราสนใจนี้ จะต้องผ่านกระบวนการแปลงสัญญาณแอนะล็อกเป็นดิจิทัล (Analog-to-Digital Conversion - ADC) หรือการดิจิทัล (digitization) ซึ่งประกอบด้วย การสุ่มตัวอย่าง (sampling) และการควอนไทซ์ (quantization) ให้อยู่ในรูปดิจิทัลก่อนที่จะทำการประมวลผลต่อไป

2.5.5.1 การสุ่มสัญญาณ (signal sampling) และการควอนไทซ์ (quantization)



รูปที่ 2.32 การสุ่มสัญญาณ (Signal sampling) และการควอนไทซ์ (quantization)

2.5.5.2 โดเมนของเวลาและสถานที่ (Temporal and Spatial domain)



รูปที่ 2.33 สัญญาณเสียงพูด

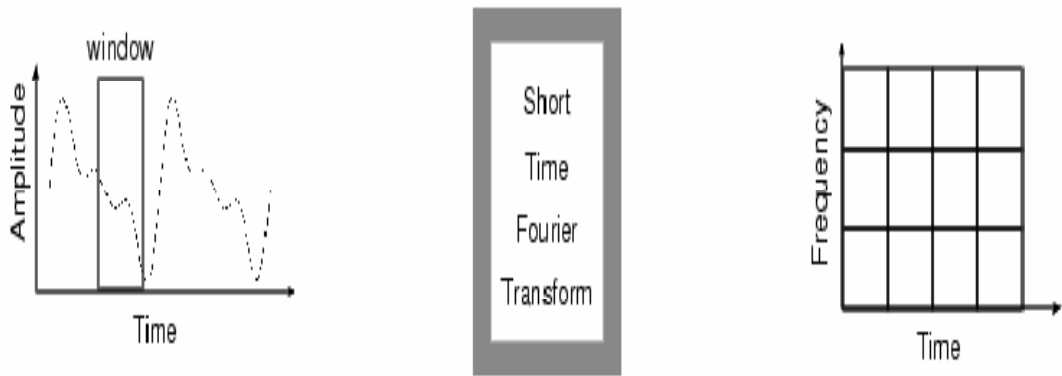
2.5.5.3 โดเมนของการแปลง (Transformed domain)

1. ความถี่ (frequency) หรือฟูรีเยร์ (Fourier)

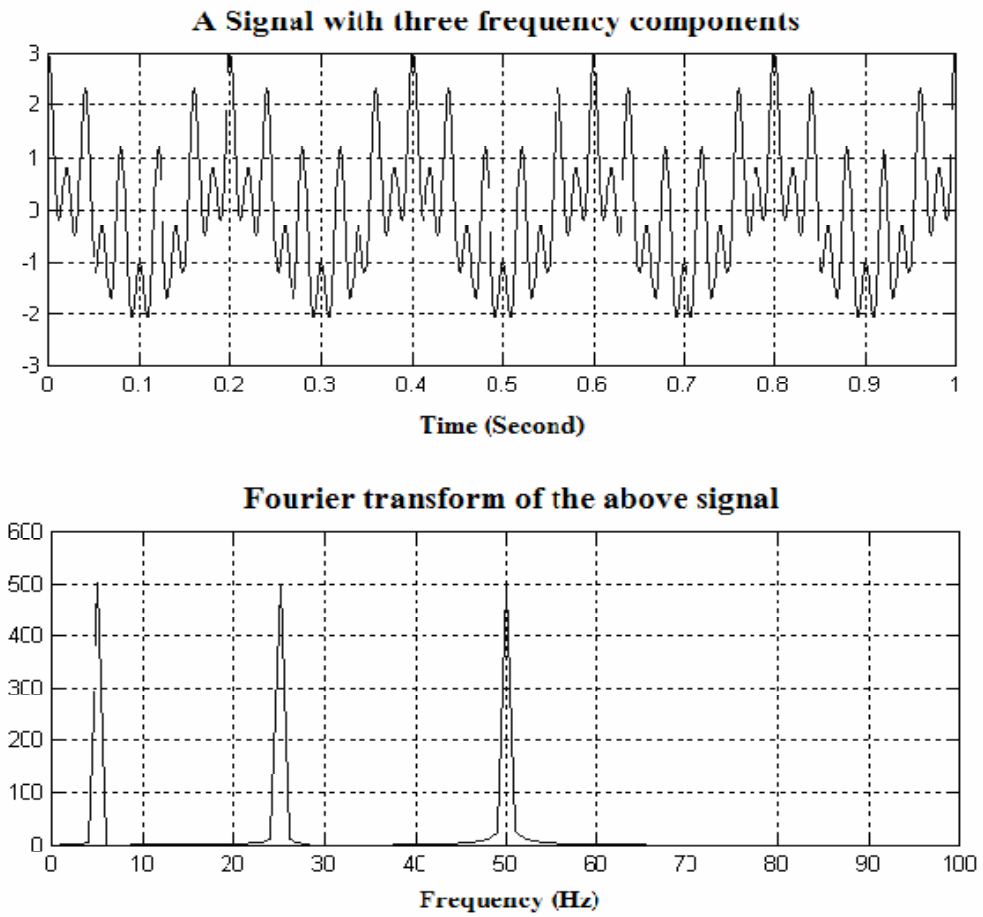
-การแปลงฟูรีเยร์ (Fourier transform) เป็นวิธีการทางคณิตศาสตร์ที่ใช้ในการวิเคราะห์สัญญาณมาเป็นเวลานาน โดยมีหลักการคือการแปลงสัญญาณจากโดเมนเวลา (Time Domain) มาเป็นโดเมนความถี่ (Frequency Domain) ซึ่งสมการหลักที่ใช้ในการวิเคราะห์สัญญาณด้วยการแปลงฟูรีเยร์คือ

$$X(f) = \int_{-\infty}^{\infty} x(t)e^{-j2\pi ft} dt \quad (2.16)$$

จากสมการหลักสังเกตได้ว่าการแปลงฟูรีเยร์เป็นการวิเคราะห์สัญญาณในรูปของสัญญาณเอกซ์โปเนนเชียล (Exponential Function) ผลการวิเคราะห์ที่ได้จะเป็นผลของการแตกองค์ประกอบของสัญญาณใดๆให้อยู่ในรูปของฟังก์ชันไซน์ (Sine Function) และโคไซน์ (Cosine Function) ที่มีขนาดและความถี่ที่แตกต่างกันในช่วงเวลาตั้งแต่ ∞ ถึง $-\infty$ โดยแสดงดังรูปที่ 2.34 ซึ่งการแปลงฟูรีเยร์นั้นจะมีความเหมาะสมกับสัญญาณที่เป็นรายคาบ (Stationary Signal) ที่มีลักษณะของสัญญาณคงที่ตลอดช่วงเวลาเพราะการวิเคราะห์ด้วยการแปลงฟูรีเยร์นั้นมีความแม่นยำทางด้านความถี่สูงแต่มีข้อจำกัดในสัญญาณที่ไม่เป็นรายคาบ (Non-Stationary Signal) เช่น สัญญาณที่มีสถานะชั่วคราวการวิเคราะห์สัญญาณดังกล่าวด้วยการแปลงฟูรีเยร์นั้นจะเกิดความผิดพลาดขึ้นรวมทั้งไม่สามารถบอกข้อมูลทางด้านเวลาของสัญญาณได้ซึ่งแสดงลักษณะดังกล่าวดังรูปที่ 2.35-2.36 ซึ่งถือว่าเป็นสิ่งสำคัญในการวิเคราะห์สัญญาณที่มีรูปแบบเป็นสัญญาณชั่วคราว

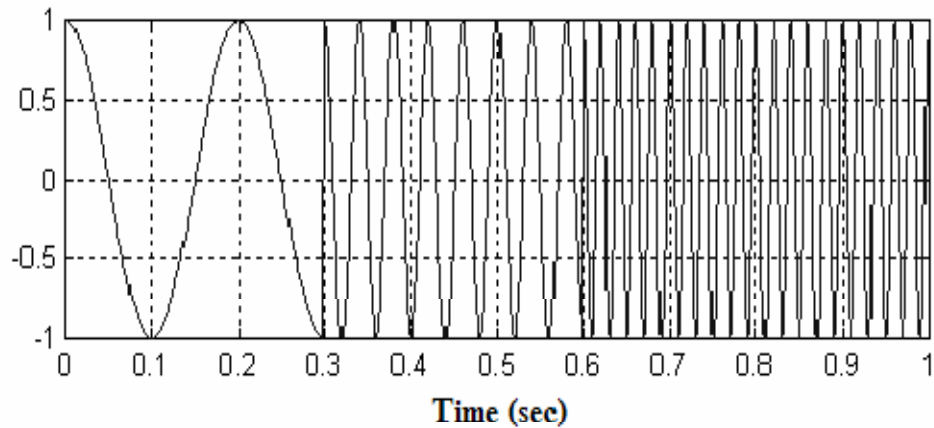


รูปที่ 2.34 การวิเคราะห์สัญญาณด้วยการแปลงฟูริเยร์

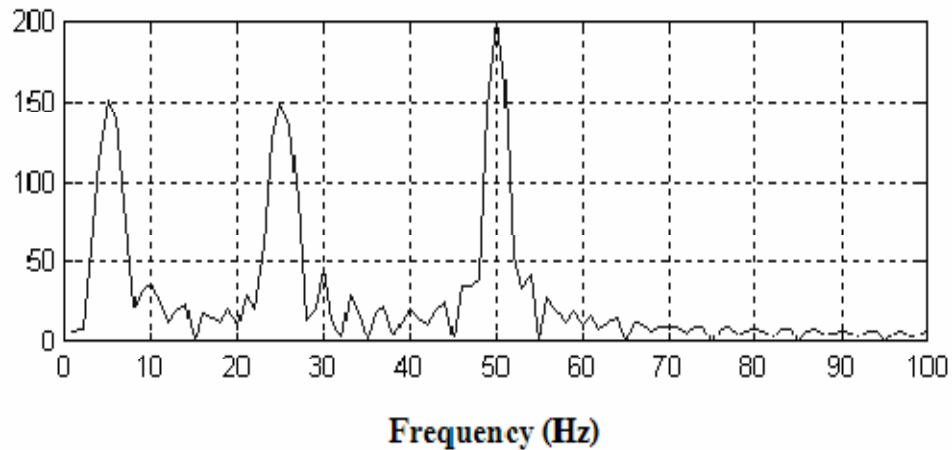


รูปที่ 2.35 การวิเคราะห์สัญญาณรายคาบด้วยการแปลงฟูริเยร์

A signal with three phase frequency components at varying times



Fourier transform of the above signal



รูปที่ 2.36 การวิเคราะห์สัญญาณที่ไม่เป็นรายคาบด้วยการแปลงฟูรีเยร์

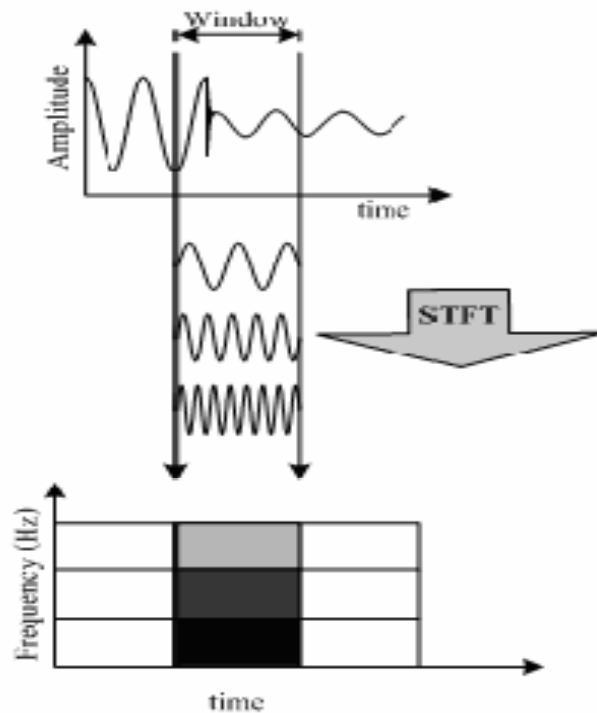
- การแปลงฟูรีเยร์เต็มหน่วย (Discrete Fourier Transform, DFT) นับเป็นวิธีการที่สำคัญในการวิเคราะห์การออกแบบและการสร้างแบบในการประมวลสัญญาณเต็มหน่วยมีการวิเคราะห์และการประมวลสัญญาณจำนวนมากที่กระทำภายใต้เวดล้อมทางความถี่หรือโดเมนฟูรีเยร์ (Frequency Domain or Fourier Domain) การแปลงฟูรีเยร์นี้สามารถที่จะคำนวณโดยตรงหรือใช้ขั้นตอนวิธีการคำนวณแบบเร็ว (Fast Fourier Transform, FFT) ก็ได้

2. เวลาและความถี่ (Time-Frequency)

-การแปลงฟูรีเยร์ในเวลาช่วงสั้น (Short Time Fourier Transform) เป็นวิธีการทางคณิตศาสตร์อีกวิธีการหนึ่งที่ใช้เพื่อแก้ไขข้อจำกัดของการแปลงฟูรีเยร์ที่ไม่สามารถวิเคราะห์ข้อมูลทางด้านเวลาได้โดยสมการการแปลงฟูรีเยร์ช่วงเวลาสั้นนั้นสามารถแสดงได้ดังสมการที่ 2.17

$$STFT(f, \tau) = \int_{-\infty}^{\infty} x(t)w(t - \tau)e^{-j2\pi ft} dt \quad (2.17)$$

เมื่อ $w(t - \tau)$ คือฟังก์ชันหน้าต่างที่ใช้ในการวิเคราะห์โดยที่ตำแหน่งเวลาที่ใช้ในการวิเคราะห์ถูกกำหนดโดยค่า τ และช่วงความถี่ในการวิเคราะห์กำหนดด้วยความถี่หรือความกว้างของฟังก์ชันหน้าต่างนั้นๆ ทำให้การแปลงฟูริเยร์ช่วงเวลานั้นสามารถวิเคราะห์ข้อมูลได้ทั้งด้านความถี่และทางเวลาดังรูปที่ 2.37 แต่การวิเคราะห์ด้วยการแปลงฟูริเยร์ช่วงเวลานั้นก็ยังมีข้อด้อยที่ลักษณะของความละเอียดของช่วงเวลาการวิเคราะห์ที่คงที่ (Fixed Resolution Transform) เพราะการใช้ช่วงการวิเคราะห์สัญญาณที่คงที่ในการวิเคราะห์สัญญาณทุกๆ ช่วงความถี่ไม่เหมาะสมในทางปฏิบัติเนื่องจากสัญญาณที่มีความถี่สูงจะมีการเปลี่ยนแปลงที่เร็วจึงควรใช้ช่วงเวลาที่แคบในการวิเคราะห์ ในขณะที่สัญญาณที่มีความถี่ต่ำกว่ามีการเปลี่ยนแปลงที่ช้าจึงควรใช้ช่วงเวลาที่กว้างกว่าในการวิเคราะห์ด้วยเหตุนี้การวิเคราะห์สัญญาณด้วยการแปลงฟูริเยร์ช่วงเวลานั้นจึงยังมีข้อด้อยในช่วงเวลาที่ใช้วิเคราะห์ซึ่งมีผลต่อขนาดของสัญญาณในแต่ละช่วงความถี่ในเวลาต่อมาจึงมีการพัฒนากระบวนการทางคณิตศาสตร์ในการวิเคราะห์สัญญาณเพิ่มเติมโดยพัฒนาช่วงเวลาการวิเคราะห์ให้เหมาะสมกับสัญญาณในแต่ละความถี่ซึ่งเรียกว่า “การแปลงเวฟเล็ต” นั่นเอง



รูปที่ 2.37 การวิเคราะห์สัญญาณด้วยการแปลงฟูริเยร์ช่วงเวลานั้น

3. เวลาและสเกล (time-scale) หรือเวฟเล็ต (wavelets)

-การแปลงเวฟเล็ต **Wavelet transform** การแปลงเวฟเล็ตของสัญญาณจะแสดงในรูปของผลรวมสัญญาณองค์ประกอบความถี่ต่ำ (Approximated Version) และองค์ประกอบความถี่สูง (Detailed Version) ในระดับต่างๆโดยจะมีค่าสัมประสิทธิ์เวฟเล็ตทำหน้าที่เป็นค่าน้ำหนักขององค์ประกอบความถี่แต่ละระดับ [21] ดังรูปที่ 3.6-3.7 ดังนั้นในการอธิบายสัญญาณใดๆด้วยการแปลงเวฟเล็ตจึงทำได้โดยนำกลุ่มของเวฟเล็ตที่มีโครงสร้างแบบฟังก์ชันเดียวกันเป็นตัวอธิบายซึ่งฟังก์ชันนี้จะเป็นฟังก์ชันต้นกำเนิดที่เรียกว่าเวฟเล็ตแม่ (Mother Wavelet) โดยที่คลื่นเวฟเล็ตแต่ละอันในกลุ่มจะเกิดจากการปรับสเกล (Scale ; a) หรือการแสดงการยืดหรือหดตัวของเวฟเล็ตแม่นั้นเองส่วนการเลื่อนตำแหน่ง (Translation or Shifting ; b) จะเป็นการแสดงตำแหน่งบนแกนเวลาโดยจะสามารถเขียนสมการของเวฟเล็ตได้ดังนี้

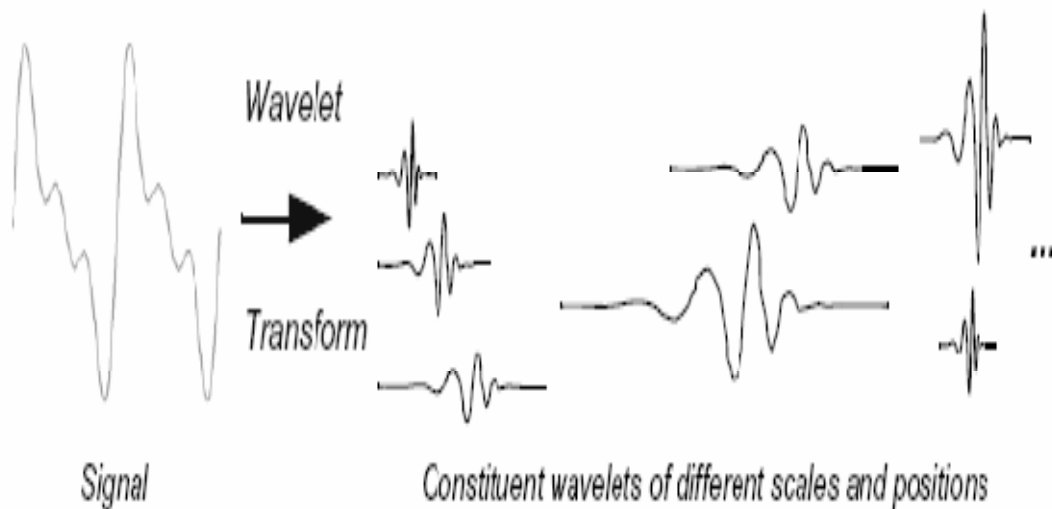
$$\Psi_{a,b}(t) = \frac{1}{\sqrt{a}} \Psi\left(\frac{t-b}{a}\right) \tag{2.18}$$

เมื่อ $\Psi(t)$ คือ ฟังก์ชันของเวฟเล็ตแม่ที่มีการปรับสเกล

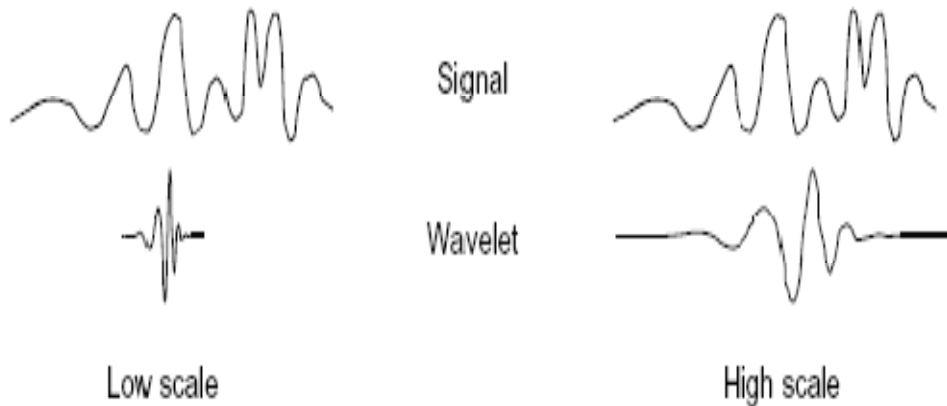
a คือ พารามิเตอร์ของการปรับสเกล (Scaling)

b คือ พารามิเตอร์การเลื่อนตำแหน่ง (Shifting)

ซึ่งในสมการจะมีการปรับค่าเพื่อให้สัญญาณที่ได้หลังจากปรับสเกลแล้วมีพลังงานเท่ากับเวฟเล็ตแม่เสมอด้วย $\frac{1}{\sqrt{a}}$



รูปที่ 2.38 การวิเคราะห์สัญญาณด้วยฟังก์ชันเวฟเล็ต



รูปที่ 2.39 คุณสมบัติการปรับสเกลของฟังก์ชันเวฟเล็ต

เมื่อนำสัญญาณใดๆมาผ่านกระบวนการแปลงเวฟเล็ตก็เปรียบเหมือนการแตกสัญญาณนั้นให้ออกมา
 ดังรูปของเวฟเล็ตแม่ที่การปรับสเกลและตำแหน่งที่แตกต่างกันไปซึ่งรูปแบบของการแปลงเวฟเล็ต
 โดยทั่วไปนั้นสามารถแบ่งออกได้เป็น 2 แบบคือการแปลงเวฟเล็ตแบบต่อเนื่อง (Continuous Wavelet
 Transform) และการแปลงเวฟเล็ตแบบเต็มหน่วย (Discrete Wavelet Transform)

- การแปลงเวฟเล็ตแบบต่อเนื่อง (Continuous Wavelet Transform) รูปแบบการแปลงเวฟ
 เล็ตแบบต่อเนื่องนั้นมีลักษณะการวิเคราะห์สัญญาณในทุกๆค่าของความถี่ซึ่งสามารถเขียนรูปแบบ
 แทนการแปลงเวฟเล็ตแบบต่อเนื่องได้ดังสมการที่ 2.19

$$CWT(a, b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} f(t) \Psi \left(\frac{t-a}{a} \right) dt \tag{2.19}$$

เมื่อ $f(t)$ คือ สัญญาณที่ทำการแปลงเวฟเล็ต

$\Psi(t)$ คือ เวฟเล็ตแม่ (Mother Wavelet)

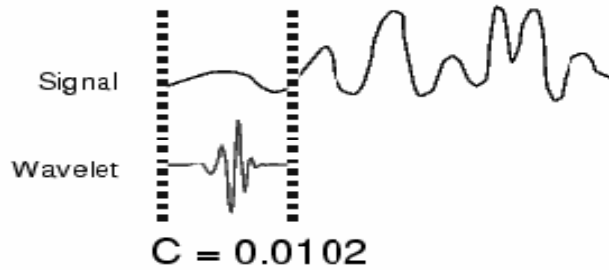
a คือ พารามิเตอร์ของการปรับสเกล (Scaling)

b คือ พารามิเตอร์การเลื่อนตำแหน่ง (Shifting)

ซึ่งพิจารณาเทอมของ $\Psi(t)$ ซึ่งเป็นเทอมของเวฟเล็ตแม่ที่เทียบได้กับเทอมของฟังก์ชันหน้าต่างในการ
 แปลงฟูริเยร์ช่วงเวลานั้นนั่นเองแต่ในการแปลงเวฟเล็ตจะมีการเปลี่ยนคุณสมบัติด้วยการเปลี่ยนค่าของ
 พารามิเตอร์ a และ b ทำให้ในช่วงที่มีความถี่สูงจะมีช่วงเวลาในการวิเคราะห์ที่แคบและในช่วงที่มี
 ความถี่ต่ำจะมีช่วงเวลาวิเคราะห์ที่กว้างซึ่งด้วยวิธีการดังกล่าวจะเป็นการลดข้อด้อยของการวิเคราะห์

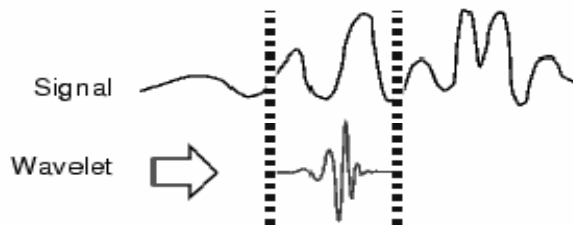
ด้วยวิธีการแปลงฟูรีเยร์ช่วงเวลาที่สั้นได้โดยกระบวนการของการแปลงเวฟเล็ตแบบต่อเนื่องนั้นได้แสดงขั้นตอนไว้ดังนี้

1. นำเวฟเล็ตแม่มาเปรียบเทียบกับส่วนแรกซึ่งเป็นจุดเริ่มต้นของสัญญาณในสเกลแรก
2. คำนวณสัมประสิทธิ์ซึ่งผลการวิเคราะห์จะแสดงความสัมพันธ์ระหว่างเวฟเล็ตแม่ในสเกลแรกซึ่งค่าสัมประสิทธิ์นี้ขึ้นอยู่กับการใช้รูปร่างของเวฟเล็ตแม่ด้วยโดยขั้นตอนที่ 1 และ 2 นั้นแสดงวิธีการได้ดังรูปที่ 2.40



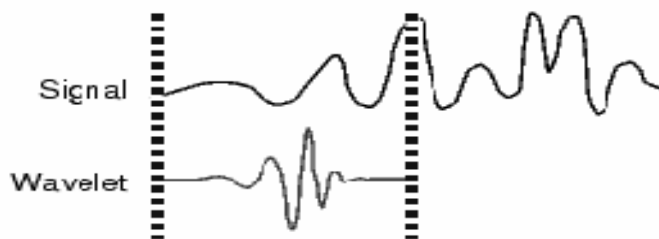
รูปที่ 2.40 กระบวนการแปลงเวฟเล็ตในขั้นตอนที่ 1 และ 2

3. เลื่อนตำแหน่งการวิเคราะห์ไปทางด้านขวาและทำขั้นตอนที่ 1 และ 2 ซ้ำจนกระทั่งครอบคลุมช่วงสัญญาณทั้งหมดในการปรับความละเอียดครั้งแรกดังรูปที่ 2.41



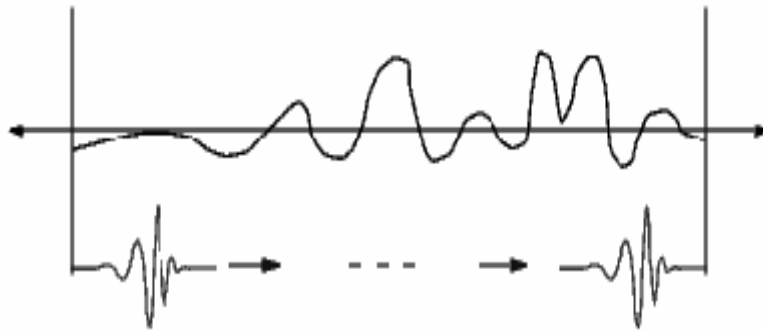
รูปที่ 2.41 กระบวนการแปลงเวฟเล็ตในขั้นตอนที่ 3

4. เปลี่ยนอัตราการปรับความละเอียดดังรูปที่ 2.42 โดยทำการขยายสเกลและทำตามขั้นตอนที่ 1-3 อีกครั้ง



รูปที่ 2.42 กระบวนการแปลงเวฟเล็ตในขั้นตอนที่ 4

5. ทำตามขั้นตอนที่ 1-4 อีกครั้งจนครบทุกอัตราการปรับความละเอียดดังรูปที่ 2.43



รูปที่ 2.43 กระบวนการแปลงเวฟเลตในขั้นตอนที่ 5

จากขั้นตอนทั้งหมดผลของการวิเคราะห์ที่ได้จะอยู่ในรูปของสัมประสิทธิ์ (Coefficient) ของแต่ละค่าการปรับสเกลและทุกตำแหน่งของการเคลื่อนสัญญาณอย่างไรก็ตามแม้ว่าการวิเคราะห์ด้วยวิธีนี้จะมี ความแม่นยำทางด้านเวลาและความถี่ก็ตามกระบวนการในการวิเคราะห์ต้องใช้เวลาในการวิเคราะห์ ค่อนข้างมากและข้อมูลของผลลัพธ์ที่ได้ก็มีความซับซ้อนเกินความจำเป็นซึ่งด้วยเหตุผลดังกล่าวการแปลงเวฟเลตแบบเต็มหน่วยจึงได้ถูกพัฒนาขึ้น

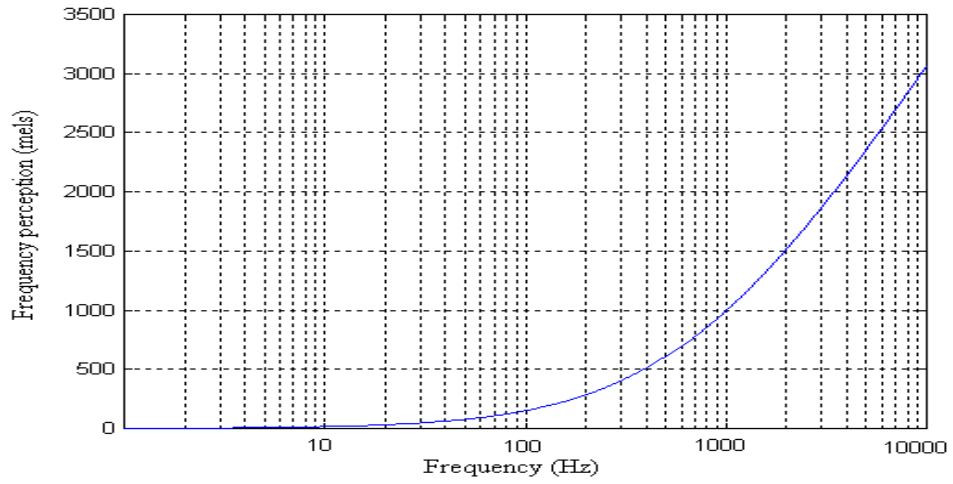
- การแปลงเวฟเลตแบบเต็มหน่วย (Discrete Wavelet Transform) การแปลงเวฟเลตแบบเต็มหน่วยเป็นการแปลงเวฟเลตรูปแบบหนึ่งที่มีลักษณะการวิเคราะห์โดยพัฒนารูปแบบการปรับสเกลและการเลื่อนตำแหน่งในลักษณะเป็นช่วงๆไม่ต่อเนื่องกันซึ่งในการวิเคราะห์ด้วยการเวฟเลตแบบเต็มหน่วยนั้นจำเป็นต้องกล่าวถึงทฤษฎีพื้นฐานที่เกี่ยวข้องเพื่อความเข้าใจในกระบวนการมากขึ้น

2.5.6 Mel-Scale

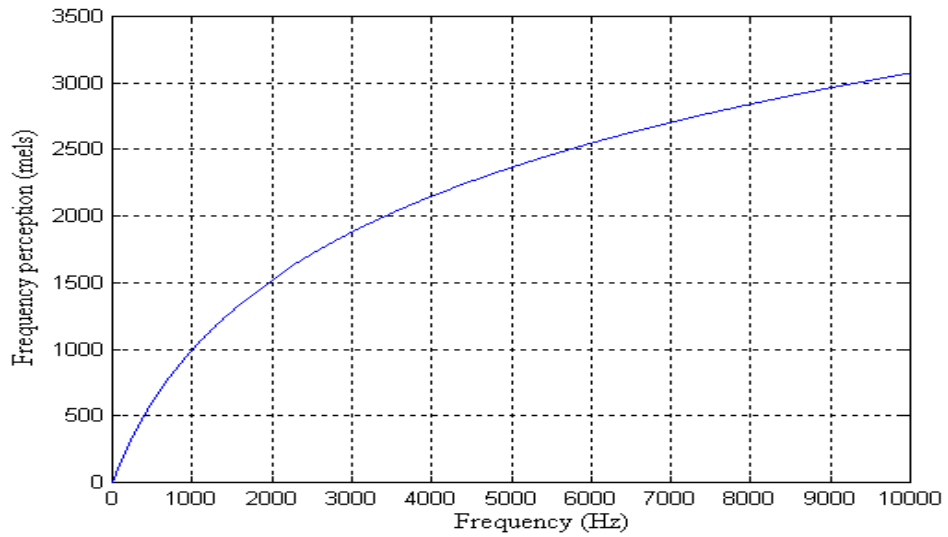
จะชี้ให้เห็นความเด่นชัดที่หูมนุษย์จะไวต่อการเปลี่ยนแปลงในช่วงความถี่ที่ต่ำของการตอบสนอง ความถี่สเปกตรัม Mel-Scale เป็นสูตรจริงใจสำหรับการเก็บตัวอย่างของคลื่นความถี่ที่อยู่บนพื้นฐานของการรับรู้ของหูมนุษย์ มันเป็นแผนที่ของคลื่นความถี่เชิงเส้นขึ้นอยู่กับการรับรู้ของหูมนุษย์ การทำแผนที่อยู่ที่ประมาณเชิงเส้นในช่วงความถี่ตั้งแต่ 0 Hz ถึง 1 kHz และลอการิทึมเกิน 1 kHz. Mel-Scale สามารถประมาณในรูปแบบใกล้ชิดของ:

$$F_{mel} = 2595 \log_{10} \left[1 + \frac{F_{HZ}}{700} \right] \quad (2-20)$$

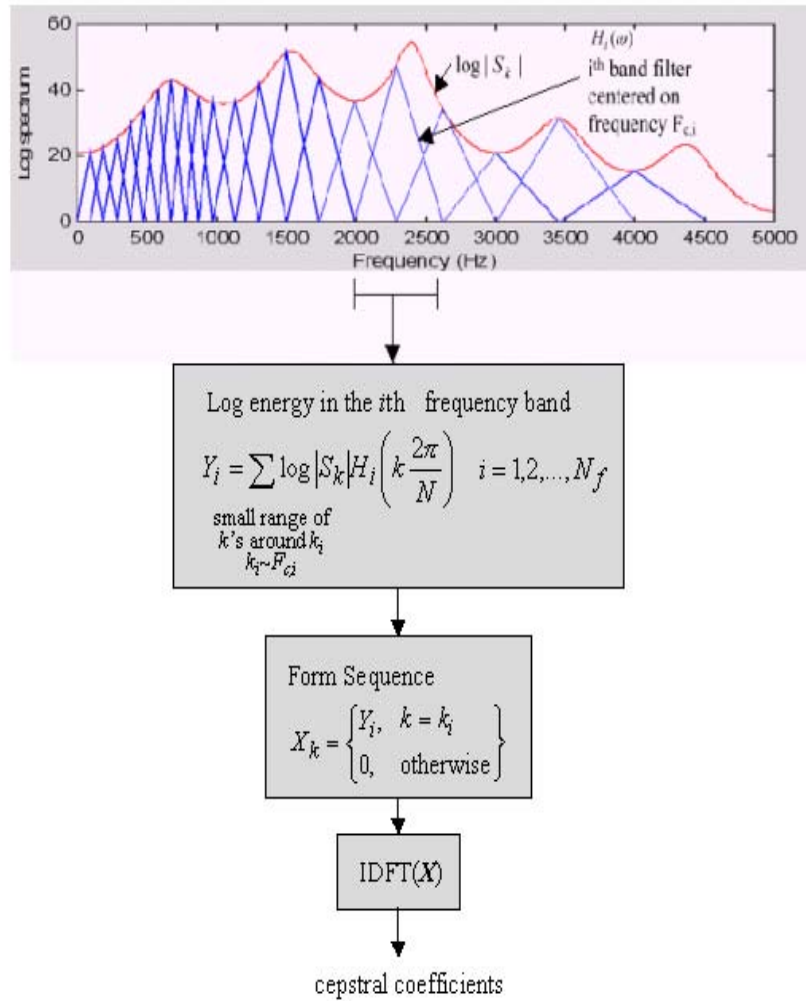
ซึ่งใน F_{mel} เป็นความถี่การรับรู้และ F_{Hz} เป็นความถี่ของสัญญาณจริง



รูปที่ 2.30 การแปลงลอการิทึมที่แสดงการแมประหว่างขนาดความถี่ของจริง (Hz) และระดับความถี่การรับรู้ (Mels)

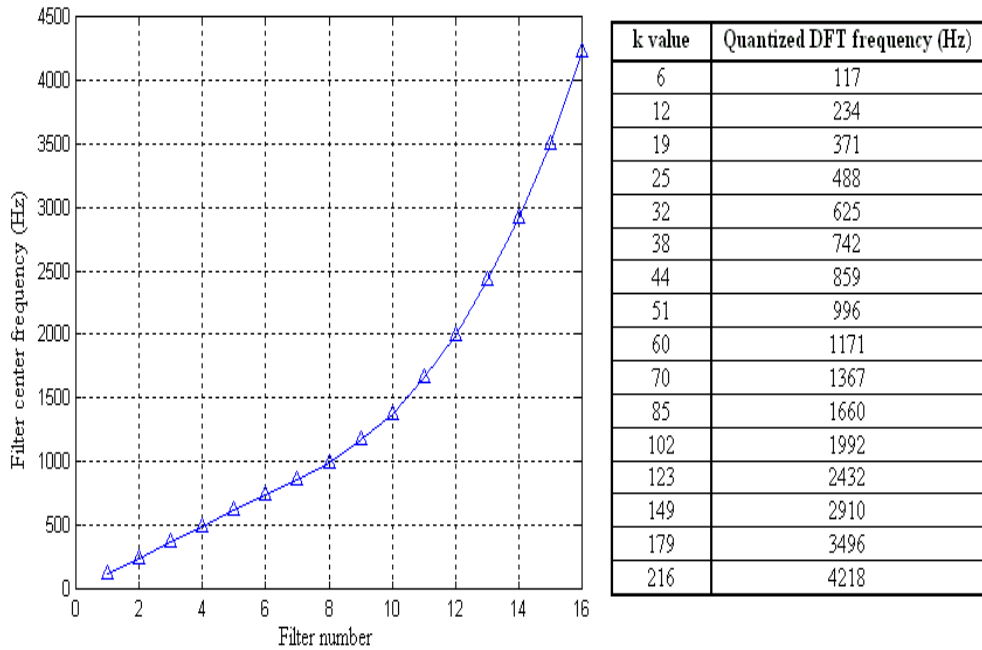


รูปที่ 2.31 การแปลงเชิงเส้นที่แสดงให้เห็นถึงการ Mapping ระหว่างขนาดความถี่ของจริง (Hz) และระดับความถี่การรับรู้ (Mels)



รูปที่ 2.32 Mel - Cepstral การสกัดสัมประสิทธิ์ filter bank

แผนภาพบล็อกของ Mel-Cepstral การสกัดสัมประสิทธิ์ filter bank จะมีภาพประกอบในรูปที่ 2.32 ขั้นตอนในการสกัด Mel-Cepstral สัมประสิทธิ์ประกอบด้วยขั้นตอนดังต่อไปนี้ ก่อนแปลงออกมาแต่ละส่วนของสัญญาณเสียงพูดเป็นประมาณสำหรับคลื่นเข้าสู่ระบบแมกนิจูด (ลอการิทึมของคลื่นขนาด) โดยการคำนวณผ่าน $N' = 512$ จุดวิเคราะห์ฟูริเย Discrete Transform (DFT) สมปทรีมเข้าสู่ระบบแมกนิจูดถูกรองแล้วโดยชุดของ $N_f = 16$ กรองผ่านแถบสามเหลี่ยมที่มีความถี่และศูนย์แบนด์วิดท์ที่มีการเลือกตาม Mel-Cepstral จะขยายแบนด์วิดธ์ของสัญญาณทั้งหมด $[0 - f_s / 2]$ รูปที่ 2.33 แสดงให้เห็นถึงความถี่กลางของตัวกรองเหล่านี้ ที่สามารถเห็นได้เป็นครั้งแรกแปดตัวกรองมีการกระจายอย่างสม่ำเสมอในช่วง 0 - 1kHz ในขณะที่อื่น ๆ แปดมีการกระจายลอการิทึมในช่วง 1-5 kHz วงจิดจำกัด บนของตัวกรองแต่ละตัวมีการตั้งค่าความถี่กลางของตัวกรองดังต่อไปนี้ในขณะที่วงจิดจำกัดล่างของตัวกรองแต่ละตัวมีการตั้งค่าความถี่กลางของตัวกรองก่อนหน้านั้นบนแกนความถี่



รูปที่ 2.33 ศูนย์กลางความถี่ของตัวกรอง 16 ที่เป็น Mel – Cepstral filter bank

2.5.6 ตัวประมาณค่าความควรจะเป็นสูงสุด (Maximum Likelihood Estimator)

ค่าพารามิเตอร์ที่ดีที่สุดของการแจกแจงความน่าจะเป็นคือค่าซึ่งจะทำให้ Likelihood หรือ Joint Probability ของเหตุการณ์ของ Observed Sample มีค่ามากที่สุด สมมติว่าแบ่ง Sample Space ออกเป็น ช่วงๆ แต่ละช่วงกว้าง dx และตัวอย่างซึ่งมีค่า x_1, x_2, \dots, x_n ที่ได้จาก Sample Space ดังกล่าวมีลักษณะ เป็นแบบ Independent and Identically Distributed ค่า Probability Density ของ $x = x_i$ คือ $f(x_i)$ และ ความน่าจะเป็นที่ตัวแปรแรนด้อมจะมีค่าใน Interval ซึ่งรวมถึง x_i คือ $f(x_i)dx$ เนื่องจาก ข้อมูลเป็นแบบอิสระต่อกัน คือ ผลคูณของ $f(x_1)dx, f(x_2)dx, \dots = [\prod_{i=1}^n f(x_i)dx^n]$ เมื่อ Interval Size dx มีค่าคงที่

เมื่อ L คือ ฟังก์ชัน Likelihood

$$L = [\prod_{i=1}^n f(x_i)] \tag{2.21}$$

เนื่องจาก Probability Density Function ส่วนใหญ่เป็น Exponential จึงสะดวกกว่าที่จะใช้ Log-Likelihood Function

$$\ln L = \sum_{i=1}^n \ln [f(x_i)] \tag{2.22}$$

ตัวประมาณแบบ Maximum Likelihood (ML) หมายถึงตัวประมาณค่าของ β ที่ให้ภาวะน่าจะเป็นที่ค่าประมาณจะเป็นค่าของพารามิเตอร์สูงสุด โดยการ maximize likelihood function (L) หรือ $Max\{\mathcal{L}\}$ ซึ่งพบว่าค่าประมาณแบบ ML จะประเมินได้มีขั้นตอนดังนี้

- สร้างฟังก์ชันภาวะน่าจะเป็นของพารามิเตอร์จากทฤษฎีทางสถิติ ฟังก์ชันภาวะน่าจะเป็น (Likelihood function) ของพารามิเตอร์จากค่าสังเกต (y) สามารถอธิบายได้จากฟังก์ชันของการแจกแจงร่วม (joint density หรือ products of density) ของค่าสังเกตเมื่อกำหนดพารามิเตอร์ (likelihood function of parameters given data is a product of density function of data given parameters) เนื่องจากในการสร้าง Likelihood function ต้องทราบการแจกแจงร่วม ดังนั้นในการสร้างตัวประมาณด้วยวิธีนี้จึงต้องทราบชนิดของการแจกแจง ซึ่งโดยทั่วไปนิยมสร้างข้อกำหนดให้มีการแจกแจงปกติ (normality assumption) ดังนั้น โมเดลเชิงเส้นจึงมีรูปแบบเป็น

$$y = X\beta + \varepsilon, \text{ โดย } \varepsilon \sim \text{MVN}(0, V) \quad (2.23)$$

เมื่อ MVN หมายถึง Multivariate normal distribution ดังนั้น likelihood function (L) จะมีรูปดังนี้

$$\begin{aligned} \mathcal{L}(\beta, V | y) &= \prod f(y | \beta, V) \\ &= (2\pi)^{-n/2} |V|^{-1/2} \exp\left\{-\frac{1}{2}(y - X\beta)'V^{-1}(y - X\beta)\right\} \end{aligned} \quad (2.24)$$

- ทำให้ค่า likelihood function สูงสุด ซึ่งในการหาค่าสูงสุดของฟังก์ชันที่อยู่ในรูปผลคูณหรือ exponential family นิยมแปลงฟังก์ชันนั้นให้อยู่ในรูปลอการิทึมก่อนแล้วจึงหาอนุพันธ์ ดังนั้น Log likelihood function (L) จะมีค่าดังนี้

$$\begin{aligned} \mathcal{L} &= (2\pi)^{-n/2} |V|^{-1/2} \exp\left\{-\frac{1}{2}(y - X\beta)'V^{-1}(y - X\beta)\right\} \\ \ln \mathcal{L} &= -\frac{n}{2} \ln(2\pi) - \frac{1}{2} \ln |V| - \frac{1}{2}(y - X\beta)'V^{-1}(y - X\beta) \\ L &= -\frac{n}{2} \ln(2\pi) - \frac{1}{2} \ln |V| - \frac{1}{2}(y' - \beta'X')V^{-1}(y - X\beta) \\ &= -\frac{n}{2} \ln(2\pi) - \frac{1}{2} \ln |V| - \frac{1}{2}(y'V^{-1}y - \beta'X'V^{-1}y - y'V^{-1}X\beta \\ &\quad + \beta'X'V^{-1}X\beta) \end{aligned} \quad (2.25)$$

ทำให้มีค่าสูงสุดได้โดยหาอนุพันธ์ตาม β แล้วเทียบให้เท่ากับศูนย์ซึ่งจะได้ว่า

$$\begin{aligned}
 \frac{\partial}{\partial \beta} L &= \frac{\partial}{\partial \beta} \left\{ -\frac{n}{2} \ln(2\pi) - \frac{1}{2} \ln |\mathbf{V}|^{-1/2} - \frac{1}{2} (\mathbf{y}'\mathbf{V}^{-1}\mathbf{y} - \beta'\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \right. \\
 &\quad \left. - \mathbf{y}'\mathbf{V}^{-1}\mathbf{X}\beta + \beta'\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\beta) \right\} \\
 &= -\frac{1}{2} (-\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} - \mathbf{X}'\mathbf{V}^{-1}\mathbf{y} + 2\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\beta) \\
 &= -\frac{1}{2} (-2\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} + 2\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\beta) \\
 &= \mathbf{X}'\mathbf{V}^{-1}\mathbf{y} - \mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\beta
 \end{aligned} \tag{2.26}$$

เมื่อเทียบอนุพันธ์ให้เท่ากับศูนย์จะได้ว่า

$$\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\beta = \mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \tag{2.27}$$

จาก Normal equation ที่ได้ ซึ่งจากสมการนี้ทำให้ประมาณ β ได้จาก

$$\hat{\beta} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \tag{2.28}$$