บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ในบทนี้จะกล่าวถึงทฤษฎีต่าง ๆ ที่จำเป็นและเกี่ยวกับงานวิจัยนี้ โดยเริ่มจากทฤษฎีทางภาษาศาสตร์ ซึ่งจะกล่าวถึงเรื่องราวของทฤษฎีการสร้างเสียงพูด พยางค์ หน่วยเสียงพยัญชนะ หน่วยเสียงสระ และหน่วยเสียงวรรณยุกต์ ทั้งในแง่ของสวนสัทศาสตร์ (Acoustic phonetics) และสรีรสัทศาสตร์ (Articulatory phonetics) พร้อมทั้งได้เสนองานวิจัยที่ผ่านมาแทรกในเนื้อหาด้วย จากนั้นจะกล่าวถึงทฤษฎีการวิเคราะห์เสียง โดยเริ่มอธิบายถึงความจำเป็นที่ต้องใช้วิธีการวิเคราะห์สัญญาณในช่วงเวลาสั้น ๆ สำหรับวิเคราะห์เสียงพูด ตามด้วยการวิเคราะห์เสียงในโดเมนเวลาและความถี่ โดยนำเสนอวิธีการต่าง ๆ เช่น การหาค่าพลังงาน การหาค่าจุดตัดศูนย์ การวิเคราะห์ฟูเรียร์ เป็นต้น ในตอนท้ายจะกล่าวถึงทฤษฎีทางสถิติพื้นฐานที่ใช้ในงานวิจัยนี้ และหลักการเข้ากลุ่ม

ทฤษฎีทางภาษาศาสตร์

1. ทฤษฎีการสร้างเสียงพูด

มนุษย์สร้างเสียงพูดขึ้นมาได้โดยมนุษย์จะมีเส้นเอ็นในลำคอที่ทำให้เกิดเสียงหรือที่เรียกว่า เส้นเสียง (vocal cords) ซึ่งอยู่ในช่องหลอดลม และจากการสั่นของเส้นเสียงจะทำให้เกิดคลื่นเสียงความถี่ต่าง ๆ ผ่านเข้าสู่ลำคอ ผ่านจากลำคอเข้าสู่ปากหรือช่องจมูกออกไปภายนอก ซึ่งขนาดและรูปร่างของอวัยวะภายในช่องปากนี้จะเป็นสิ่งกำหนดว่าคลื่นเสียงความถี่ไหนจะยอมให้ปรากฏออกมาให้ได้ยิน (passed through) หรือคลื่นเสียงความถี่ไหนที่จะถูกดูดซับไว้ไม่ยอมให้ปรากฏออกมา (absorbed) ปรากฏการณ์ที่ทำให้เกิดเสียงพูดออกมาได้นี้ เนื่องจากรูปร่างของช่องคอและช่องปากของมนุษย์ที่สามารถเปลี่ยนแปลงไปมาได้อย่างอิสระ ทำให้ผู้พูดมีความสามารถในการกำหนดคลื่นเสียงความถี่ต่าง ๆ ให้ปรากฏออกมาหรือดูดซับเอาไว้ได้ตามต้องการ การเรียงลำดับชั้นของการคัดเลือกความถี่ของคลื่นเสียงระดับต่าง ๆ เหล่านี้นั่นเองที่ทำให้เกิดการแปลงเป็นเสียงพูดออกมา

1.1 อวัยวะออกเสียง (The Organs of Speech)

ธรรมชาติได้กำหนดตำแหน่งของอวัยวะที่ใช้ในการทำให้เกิดเสียงของมนุษย์เอาไว้ นั่นคือ ริมฝีปาก ฟัน เพดาน และลิ้น ที่อยู่ตามลำดับตลอดช่วงของกล้ามเนื้อช่องคอ เพื่อทำให้สามารถเปลี่ยนขนาดและรูปร่างได้ตามต้องการ ดังรูปที่ 2.1 เราจะเปล่งคำพูดออกมาโดยการเปล่งออกมาเป็นลำดับขั้นของความถี่คลื่นเสียงที่ได้คัดเลือกไว้ออกมา นั่นคือ จะมีการเปลี่ยนแปลงตำแหน่งของช่องทางเดินเสียง (vocal tract) ของมนุษย์เป็นลำดับขั้นอย่างต่อเนื่อง

อวัยวะที่ใช้ในการออกเสียงนั้น อาจจะแบ่งคร่าว ๆ ได้เป็น 2 พวก คือ (กาญจนา นาคสกุล, 2524)

1) อวัยวะที่เป็นส่วนกระทำอาการ (Articulator) หมายถึงอวัยวะส่วนที่เคลื่อนไหวเพื่อผลักลมไปยังส่วนต่าง ๆ อวัยวะที่เป็นส่วนกระทำอาการที่สำคัญ คือ ลิ้น ซึ่งเป็นส่วนที่เคลื่อนไหวได้มากที่สุด

2) อวัยวะที่เป็นฐานที่เกิดเสียงต่าง ๆ (Point of articulator) หมายถึงตำแหน่ง หรือฐานที่เกิดของเสียงต่าง ๆ เช่น ริมฝีปาก ฟัน เพดานส่วนต่าง ๆ เป็นต้น

จากรูปที่ 2.1 อวัยวะส่วนที่มีหน้าที่ในการออกเสียงพูดมีดังนี้

ริมฝีปาก (lips) เป็นอวัยวะส่วนที่สามารถเคลื่อนไหวได้มาก และทำให้เสียงแตกต่างกันได้มาก เราอาจบังคับริมฝีปากให้อยู่ชิดกัน ห่างกัน ยื่นออกมา หรือห่อกลม ฯลฯ ก็ได้ ลักษณะริมฝีปากต่าง ๆ นี้ ล้วนแต่มีอิทธิพลต่อการออกเสียง และการทำให้เสียงแตกต่างกันไปทั้งสิ้น

ฟัน (teeth) เป็นอวัยวะที่เกิดของเสียงหลายชนิด เช่น เมื่อฟันบนกดลงบนริมฝีปากล่างหรือกดกับฟันล่าง ลมที่ผ่านออกมาโดยแรงจะลอดช่องที่พอผ่านได้ออกมา ทำให้เกิดเป็นเสียงชนิดที่เรียกว่า เสียงเสียดแทรกที่เกิดที่ฟัน เป็นต้น

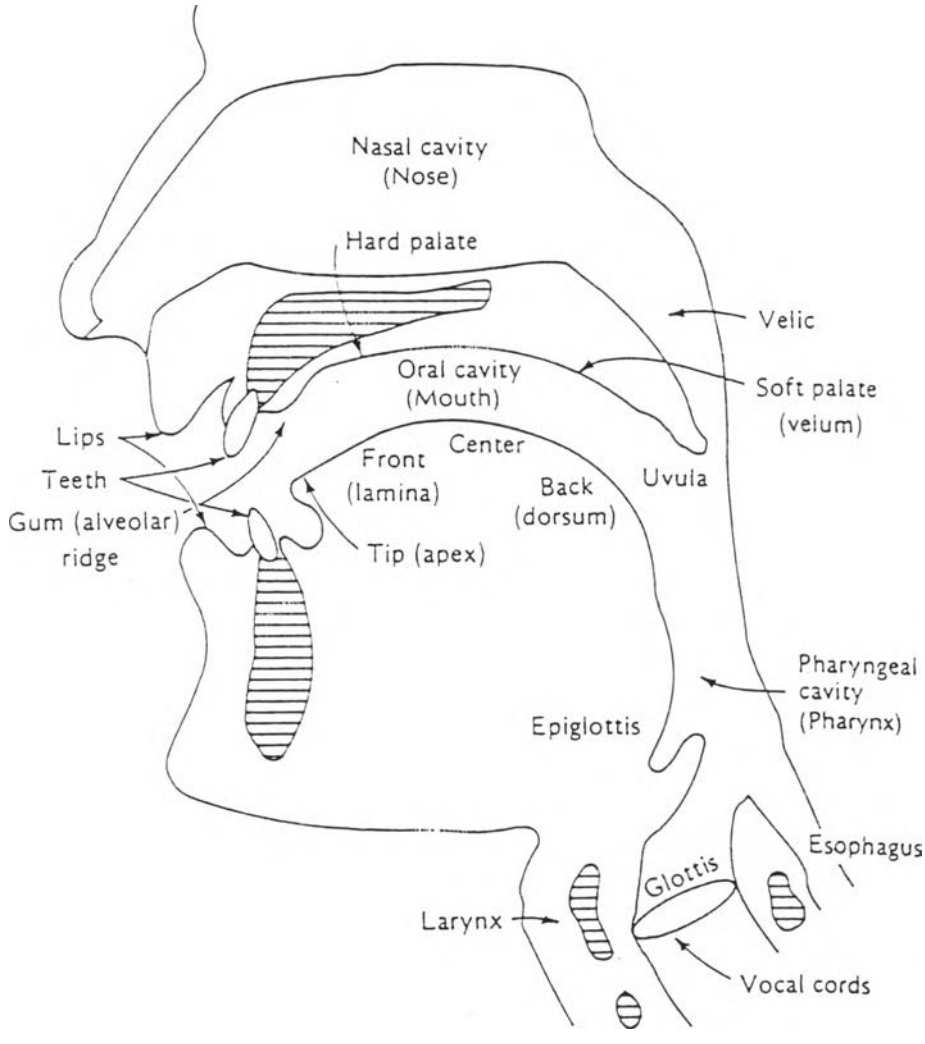
ปุ่มเหงือก (gum ridge) คือส่วนที่อยู่ต่อจากฟันบน

เพดานแข็ง หรือเพดานปาก (hard palate) หมายถึงเฉพาะเพดานที่โค้งเป็นกระดูกแข็งเท่านั้น

เพดานอ่อน (soft palate) คือส่วนของเพดานที่อยู่ต่อเพดานแข็งไปข้างใน มีลักษณะเป็นกระดูกอ่อนที่ขยับขึ้นลงได้ เวลาหายใจเพดานอ่อนและลิ้นไก่ซึ่งอยู่ปลายเพดานอ่อนจะลดระดับลงมาเปิดช่องให้ลมออกไปทางจมูก เวลาพูดส่วนใหญ่ปลายเพดานอ่อนและลิ้นไก่จะถูกยกขึ้นไปจรดกับหลังคอ นอกจากเวลาออกเสียงนาสิกเท่านั้นที่เพดานอ่อนจะลดระดับลงมา

ลิ้น (tongue) เป็นส่วนที่เคลื่อนไหวมากที่สุดในการออกเสียงพูด จึงต้องแบ่งเป็น 3 ส่วนด้วยกันตามหน้าที่ในการออกเสียง

รูปที่ 2.1 อวัยวะภายในของระบบการพูดของมนุษย์ (Ladefoged, 1975)



ปลายลิ้น (blade of the tongue) หมายถึงส่วนปลายซึ่งสามารถจะยกขึ้นไปแตะอวัยวะส่วนต่าง ๆ ในปากตอนบนได้ง่าย

หน้าลิ้น (front of the tongue) หมายถึงลิ้นส่วนที่อยู่ตรงข้ามกับเพดานแข็ง ขณะที่วางลิ้นราบกับปากตอนที่ไม่ได้พูด

หลังลิ้น (back of the tongue) หมายถึงส่วนของลิ้นที่อยู่ตรงข้ามกับเพดานอ่อน ขณะที่ว่างลิ้นราบกับปาก

ช่องปาก (oral cavity) ช่องปากนี้เป็นอวัยวะอีกส่วนหนึ่งที่มีความสำคัญในการพูด กล่าวคือ ช่องปากจะทำหน้าที่เป็นช่องกำทอน (resonant chamber) ซึ่งมีบทบาทในการกำทอนเสียงที่เดินทางผ่านมาถึงบริเวณนี้ ทั้งนี้เพราะช่องปากสามารถจะเปลี่ยนแปลงเป็นช่อง

กำทอนรูปร่างต่าง ๆ กัน เนื่องจากรูปร่างของมันแปรผันไปตามการจัดท่าทางของลิ้น ริมฝีปาก และ ขากรรไกร

โพรงจมูก (nasal cavity) คุณสมบัติหรือลักษณะของเสียงพูดที่เกิดขึ้นจะแปรผันไปตามการปิด-เปิดของช่องทางออกของอากาศที่จะออกสู่โพรงจมูก ซึ่งเป็นผลจากการยกขึ้นหรือเลื่อนลงของเพดานอ่อน

ช่องคอ (pharynx) หมายถึงโพรงคอซึ่งอยู่ถัดจากปากลงไปจนถึงเส้นเสียง

ลิ้นปิดหลอดลม (epiglottis) เป็นก้อนเนื้อเล็ก ๆ คล้ายลิ้นไก่ อยู่ต่อจากโคนลิ้นลงไปในลำคอ มีหน้าที่ปิดช่องลมขณะรับประทานอาหาร และเปิดช่องลมขณะพูด

เส้นเสียง (vocal cords) เป็นอวัยวะสำคัญที่เกิดของเสียง เส้นเสียงมีลักษณะเป็นกล้ามเนื้อ 2 แผ่นปิดขวางอยู่ปากช่องหลอดลมจากด้านหลังมาด้านหน้า ระหว่างเส้นเสียงจะมีช่องว่างซึ่งเป็นทางให้ลมผ่านเข้าไปถึงปอด และออกมาจากปอดได้ ช่องนี้เรียกว่า ช่องว่างระหว่างเส้นเสียง (glottis) เส้นเสียงทั้งสองสามารถจะดึงออกให้ห่างจากกัน หรือดึงเข้ามาประชิดกันก็ได้ เส้นเสียงเป็นส่วนทำให้เกิดเสียงพูดขึ้น

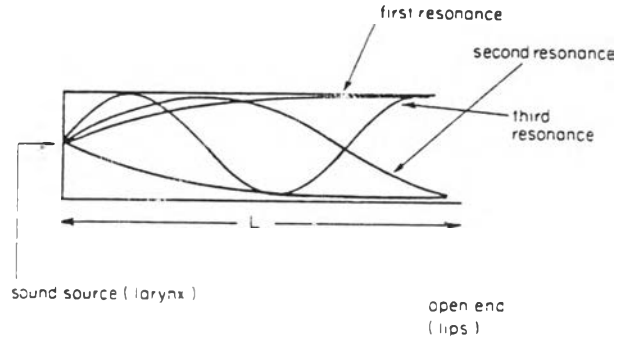
1.2 กระบวนการทำให้เกิดเสียง (Speech Production)

ระบบของเสียงพูดสามารถที่จะพิจารณาได้ว่า ประกอบด้วยลำดับของท่อ และช่องที่ต่อออกมาจากปอดไปยังปากและจมูก ท่อและช่องนี้จะมีความยาวโดยประมาณ 7 นิ้ว เส้นเสียงจะอยู่ในตำแหน่งปลายที่ตรงข้ามกับขั้วปอด ทำหน้าที่ควบคุมการไหลของลมให้ผ่านปอดเข้าสู่ช่องทางเดินเสียง ภายใต้การควบคุมของกล้ามเนื้อส่วนประกอบของช่องทางเดินเสียงที่มีลักษณะเป็นท่อ จะสามารถเปลี่ยนแปลงรูปร่างได้ในอัตราถึง 10 ครั้งต่อวินาที ส่วนเส้นเสียงนั้นจะสามารถเปิด-ปิดด้วยอัตราเร็วประมาณ 100-300 ครั้งต่อวินาที การเปลี่ยนแปลงรูปร่างของช่องทางเดินเสียงและรูปร่างและตำแหน่งของสื่อกลางที่ทำให้เกิดเสียงดังกล่าวนี้รวมเรียกว่า กระบวนการทำให้เกิดเสียง (Veltri, 1985)

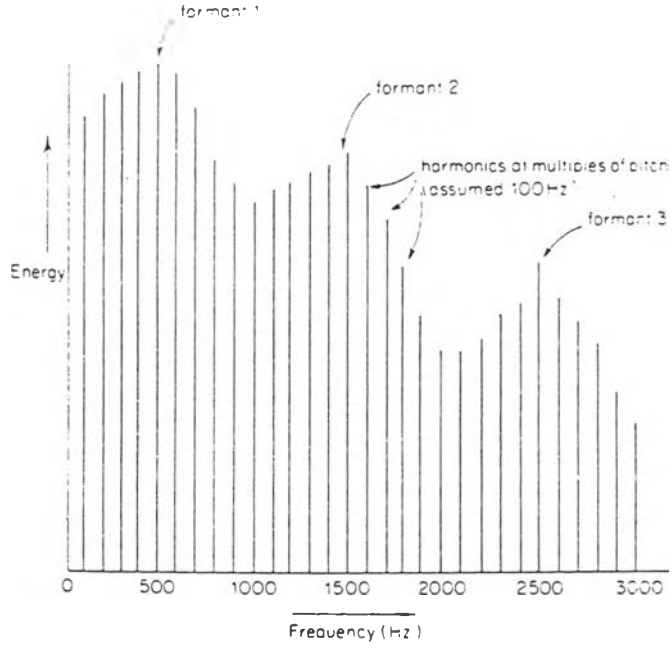
รูปแบบจำลองอย่างง่ายของช่องทางเดินเสียง ก็อาจจะมองได้เป็นลักษณะของท่อทรงกลมดันที่มีต้นกำเนิดของเสียงอยู่ที่ปลายข้างหนึ่ง (ส่วนของกล่องเสียง) และปลายอีกข้างหนึ่งจะเปิด (ส่วนของปาก) ดังรูปที่ 2.2 ดังนั้นมันจะเกิดการกำทอนภายในท่อได้ที่ความยาวคลื่นเท่ากับ 4L, 4L/3, 4L/5,.... โดยที่ L คือความยาวของท่อ ถ้าคิดเป็นความถี่ที่เกิดการกำทอนจะได้ความถี่ที่ c/4L, 3c/4L, 5c/4L,... Hz โดยที่ c คือค่าความเร็วของเสียงในอากาศ และถ้าจะคำนวณหาความถี่ในการกำทอนของช่องทางเดินเสียงของคน ซึ่งปกติช่องทางเดินเสียงของคนเราจะมีความยาวประมาณ 7 นิ้ว หรือ 17 เซนติเมตร และ c มีค่าเท่ากับ 340 เมตรต่อวินาที ดังนั้นจึงมีการกำทอนที่ความถี่ประมาณ 500 Hz, 1500 Hz, 2500 Hz,... เป็นต้น

เมื่อกล่องเสียงกระตุ้นให้เกิดคลื่นที่ประกอบไปด้วยฮาร์มอนิก (harmonic) ต่าง ๆ มากมาย การกำทอนของช่องทางเดินเสียงนี้จะสร้างรูปคลื่นเสียงที่มียอดสูงเด่นขึ้นมาเมื่อดูจากสเปกตรัมพลังงานของรูปคลื่นซึ่งเรียกว่า ฟอร์แมนท์ (formants) ของเสียง ดังรูปที่ 2.3 ฟอร์แมนท์ที่มีความถี่ต่ำที่สุดจะเรียกว่า ฟอร์แมนท์ที่หนึ่ง ซึ่งจะมีค่าประมาณ 200 - 1000 Hz ทั้งนี้ขึ้นอยู่กับขนาดของช่องทางเดินเสียงด้วย ส่วนฟอร์แมนท์ที่สองที่อยู่ถัดไปก็จะมีค่าประมาณ 500 - 2500 Hz และฟอร์แมนท์ที่สามมีค่าประมาณ 1500 - 3500 Hz เป็นต้น โดยที่ฟอร์แมนท์ที่หนึ่งและสองนี้จะเป็นตัวที่แสดงคุณลักษณะที่สำคัญมากอันหนึ่งในการพิจารณาประยุกต์ใช้ในด้านการวิเคราะห์สัญญาณเสียงทางด้านสเปกตรัม

รูปที่ 2.2 การเกิดการกำทอนภายในแบบจำลองของช่องทางเดินเสียง (Witten, 1982)



รูปที่ 2.3 สเปกตรัมของพลังงานของเสียง (Witten, 1982)



2. พยางค์

2.1 ทฤษฎีทั่วไปเกี่ยวกับพยางค์

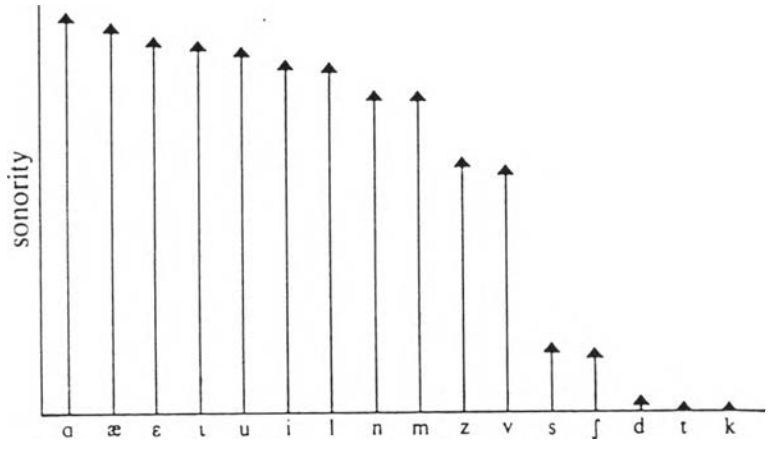
ได้มีทฤษฎีทางภาษาศาสตร์อยู่หลายทฤษฎีที่พยายามจะให้คำจำกัดความของคำว่า "พยางค์" ซึ่งทฤษฎีเหล่านี้สามารถแยกออกเป็น 3 ลักษณะใหญ่ ๆ คือ การให้คำจำกัดความในแง่คุณสมบัติของเสียง (Properties of Sounds) การให้คำจำกัดความในแง่กิจกรรมผู้พูด (Activities of Speaker) และการให้คำจำกัดความในแง่ของหน้าที่ (Functional)

1) คำจำกัดความในแง่คุณสมบัติของเสียง (Properties of Sounds)

Otto Jesperson นักสัทศาสตร์ชาวเดนมาร์ค ได้ให้คำจำกัดความของพยางค์โดยคำนึงถึง "sonority of sound" sonority ของเสียง คือ ความดังของเสียง (loudness) ที่มีความสัมพันธ์กับความดังของเสียงอื่น ซึ่งมีความยาวของเสียง (length) ความหนักเบาของเสียง (stress) และระดับสูง-ต่ำ (pitch) เท่ากัน

ความดังของเสียงส่วนใหญ่จะขึ้นอยู่กับความเข้มของเสียง (acoustic intensity) ซึ่งหมายถึงจำนวนพลังงานของเสียง (acoustic energy) ที่ใช้ เมื่อนำเสียงที่มีระดับความสูง-ต่ำ ความหนักเบาของเสียง และความยาวของเสียงเท่ากัน มาจัดกลุ่มเปรียบเทียบ เราจะสามารถจะวัดความเข้มของเสียงได้ ตัวอย่าง sonority ของหน่วยเสียงต่าง ๆ แสดงดังรูปที่ 2.4

รูปที่ 2.4 ความสัมพันธ์ของ Sonority ของเสียงต่าง ๆ ในภาษาอังกฤษ
(Ladefoged, 1975)



แต่อย่างไรก็ตามความเข้มของเสียงแต่ละเสียงจะแปรเปลี่ยนไปตามตัวผู้พูดแต่ละคนด้วย บางคนอาจจะออกเสียง /i/ ดังกว่าเสียง /l/ ก็ได้

Jesperson ได้อธิบายพยางค์โดยกล่าวว่า พยางค์จะเป็นกลุ่มของหน่วยเสียงที่มีหน่วยเสียงหนึ่งดังกว่าหน่วยเสียงข้างเคียง เช่นคำว่า plain จะเห็นว่ามีหน่วยเสียง /a/ ที่มีระดับความดังมากกว่าหน่วยเสียงข้างเคียง ดังนั้นจึงมี 1 พยางค์ หรือคำว่า visit จะมีหน่วยเสียง /i/ 2 ตัวที่มีระดับความดังมากกว่าเสียงข้างเคียง ดังนั้นจึงมี 2 พยางค์

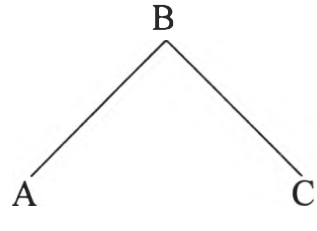
แต่ตามทฤษฎีระดับความดัง (Sonority) ของ Jesperson นี้ จะมีปัญหาสำหรับคำบางคำเช่น spa จะเห็นว่ามีหน่วยเสียงที่มี sonority มากกว่าข้างเคียงอยู่ 2 หน่วยเสียง คือ /s/ และ /a/ จึงควรเป็นคำ 2 พยางค์ แต่ความจริงแล้วคำนี้เป็นคำพยางค์เดียว

2) คำจำกัดความในแง่กิจกรรมผู้พูด (Activities of Speaker)

การอธิบายพยางค์ในแง่ของผู้พูดนั้น ได้มีนักสัทศาสตร์หลายคนได้ให้ความเห็นไว้ โดยเริ่มจาก Ferdinand de Saussure นักภาษาศาสตร์ชาวสวิตเซอร์แลนด์ ได้พยายามอธิบายพยางค์โดยอาศัยปริมาณความกว้างของอวัยวะในการเปล่งเสียง ระดับความมากน้อยในการออกเสียง (degree of opening of the sounds) เช่น เสียงสระมีการเปิดมากกว่าพยัญชนะ และพยัญชนะระเบิดมีระดับการออกเสียงมากกว่าพยัญชนะเสียดแทรก เป็นต้น Saussure ได้อธิบายว่าขอบเขตของพยางค์นั้นจะอยู่ที่ช่วงต่อระหว่างเสียงที่มีระดับในการออกเสียงน้อยกับเสียงที่มีระดับในการออกเสียงมาก นอกจากนั้น Saussure ยังใช้สัญลักษณ์ 2 ตัว ที่ใช้ในการอธิบายพยางค์โดยเขาเรียกเสียงที่มีระดับในการออกมากที่อยู่ตอนต้นพยางค์ว่าเสียงระเบิด (Explosion,<) และเสียงที่มีระดับในการออกน้อยที่อยู่ตอนท้ายพยางค์ว่าเสียงอุบ (Implosive,>) ดังนั้นพยางค์จะเขียนเป็นสัญลักษณ์ได้ดังนี้ (<>) และขอบเขตของพยางค์ก็จะเขียนได้เป็น (><)

นอกจากนั้น Maurice Grammont และ Pierre Fouchi นักสัทศาสตร์ชาวฝรั่งเศส ยังได้อธิบายพยางค์ในแง่ที่ว่า อวัยวะที่ใช้ในการออกเสียงจะมีการเกร็งในตอนต้นพยางค์ โดยแสดงในรูปที่ 2.5

รูปที่ 2.5 เส้น A-B แสดงการเกร็งที่เพิ่มขึ้นตอนต้นพยางค์ ส่วนเส้น B-C แสดงการเกร็งที่ลดลงตอนท้ายพยางค์



ต่อมา R.H. Stetson นักสัทศาสตร์ชาวอเมริกัน ได้อธิบายพยางค์โดยพิจารณาจากจังหวะการเกร็งของกล้ามเนื้อช่องอก โดยเขาได้อธิบายว่า ในตอนต้นพยางค์ทุกพยางค์จะมีการเกร็งของกล้ามเนื้อช่องอก (chest pulse) แต่ในภายหลังได้มีการพิสูจน์แล้วทฤษฎีดังกล่าวก็ไม่เป็นจริงดังกล่าว

3) คำจำกัดความในแง่ของหน้าที่ (Functional)

ทฤษฎีนี้ได้พยายามอธิบายพยางค์โดยพิจารณาจากโครงสร้าง (Structure) ดังนี้

[ส่วนขอบพยางค์] + ส่วนแกน + [ส่วนขอบพยางค์]

[Marginal sound] + Nucleus + [Marginal sound]

<u>หมายเหตุ</u> [ ] = จะมีหรือไม่มีก็ได้

การอธิบายพยางค์โดยทฤษฎีนี้จะต้องระบุให้แน่ชัดว่าจะใช้อธิบายพยางค์ของภาษาใด เพราะว่าหน่วยเสียงสระและหน่วยเสียงพยัญชนะที่ประกอบกันเป็นพยางค์ในแต่ละภาษานั้นแตกต่างกันออกไป

2.2 พยางค์ในภาษาไทย

พยางค์ในภาษาไทย คือ เสียงที่เปล่งออกมาครั้งหนึ่ง ๆ มีเสียงดังเด่น 1 เสียง และเสียงที่อยู่ข้างเคียงอย่างน้อย 2 เสียง พยางค์อาจจะเป็นคำได้ในกรณีที่พยางค์นั้น ๆ มีความหมาย (พัธนี โชติกเสถียร อ้างจากภาษาไทย 3, 2533)

เราถือว่าหน่วยเสียงที่ดังเด่นกว่าหน่วยเสียงอื่น ๆ ซึ่งปรากฏในเสียงพูดเป็นแกนกลางพยางค์ ส่วนหน่วยเสียงอื่น ๆ ที่ไม่ดังเด่นก็ทำหน้าที่เป็นส่วนประกอบ หรือเสริมพยางค์นั้น ๆ หน่วยเสียงจะดังเด่นหรือไม่ ขึ้นอยู่กับลักษณะประจำตัวของหน่วยเสียงนั้น ๆ เช่น หน่วยเสียงสระมีลักษณะประจำตัวเป็นเสียงก้อง จะดังเด่นกว่าหน่วยเสียงอื่น ๆ เวลาออกเสียงพูดในแต่ละครั้ง โดยปกติหน่วยเสียงสระจึงเป็นส่วนสำคัญที่ทำให้เกิดพยางค์ หรืออาจพูดได้ว่าพยางค์ย่อมมีหน่วยเสียงสระเป็นใจกลาง

พยางค์ในภาษาไทยสามารถเขียนในรูปของโครงสร้างได้ดังนี้

$$S = C(C)\overset{T}{V}(V)(C)$$

โดยที่

S = พยางค์

C = พยัญชนะ

V = สระ

T = วรรณยุกต์ ซึ่งมี 5 เสียง

โดย 0 = เสียงสามัญ

1 = เสียงเอก

2 = เสียงโท

3 = เสียงตรี

4 = เสียงจัตวา

(V) = สระเสียงยาว

(C) = พยัญชนะควบกล่ำ

ลักษณะโครงสร้างของพยางค์ต่างกันมีผลต่อการผันของเสียงวรรณยุกต์ซึ่งลักษณะต่าง ๆ นั้น ได้สรุปและแสดงในตารางที่ 2.1 โครงสร้างพยางค์ที่สามารถเกิดเสียงวรรณยุกต์ใดได้ จะแทนด้วยสัญลักษณ์บวก (+) ในหลักของวรรณยุกต์นั้น แต่ถ้าไม่สามารถเกิดเสียงวรรณยุกต์นั้นได้ ก็จะแสดงด้วยสัญลักษณ์ลบ (-)

ตารางที่ 2.1 เสียงวรรณยุกต์ที่สามารถเกิดขึ้นได้ (+) และไม่สามารถเกิด (-) กับพยางค์ในภาษาไทย ในโครงสร้างพยางค์ต่าง ๆ (ดัดแปลงจาก กาญจนา นาคสกุล, 2524)

เสียงวรรณยุกต์ / โครงสร้าง	สามัญ	เอก	โท	ตรี	จัตวา
1. C(C)VV	+	+	+	+	+
2. C(C)VVN	+	+	+	+	+
3. C(C)VN	+	+	+	+	+
4. C(C)VS	-	+	-	+	-
5. C(C)VVS	-	+	+	-	-

หมายเหตุ N แทนหน่วยเสียงพยัญชนะนาสิก และพยัญชนะครึ่งสระ

S แทนหน่วยเสียงพยัญชนะกัก

ในภาษาไทยยังมีคำที่ผันเสียงวรรณยุกต์ต่างไปจากตารางที่ 2.1 อีก เช่น คำที่ใช้แสดงอารมณ์ คำอุทาน และ คำยืมจากต่างประเทศ เป็นต้น

3. หน่วยเสียงพยัญชนะ

พยัญชนะในภาษาไทย มี 21 หน่วยเสียง แบ่งเป็น 2 กลุ่ม คือ พยัญชนะกัก (Stop Consonants) 11 หน่วยเสียง และพยัญชนะไม่กัก (Non-Stop Consonants) 10 หน่วยเสียง แสดงในตารางที่ 2.2

หน่วยเสียงพยัญชนะทั้ง 21 หน่วยเสียง สามารถที่จะอยู่ในตำแหน่งต้นพยางค์ได้ทุกหน่วยเสียง แต่จะมีหน่วยเสียงพยัญชนะที่ปรากฏท้ายพยางค์ได้เพียง 9 หน่วยเสียงเท่านั้น คือ เสียงกักไม่พ่นลม 4 หน่วยเสียง (p,t,k,ʔ) เสียงนาสิก 3 หน่วยเสียง (m,n,ŋ) และเสียงกึ่งสระ 2 หน่วยเสียง (w,j)

ตารางที่ 2.2 เสียงพยัญชนะในภาษาไทย

ที่เกิดเสียง / ลักษณะของเสียง		ริมฝีปาก	ปุ่มเหงือก	เพดานแข็ง	เพดานอ่อน	เส้นเสียง
พยัญชนะกัก	ไม่พ่นลม	p*	t*	c	k*	ʔ*
	พ่นลม	ph	th	ch	kh	
	ก้อง	b	d			
พยัญชนะไม่กัก	นาสิก	m*	n*		ŋ*	
	เสียดแทรก	f	s			h
	กระทบ		r			
	ข้างลิ้น		l			
	กึ่งสระ	w*		j*		

หมายเหตุ หน่วยเสียงพยัญชนะที่มีเครื่องหมายดอกจันทน์ (*) คือ หน่วยเสียงที่สามารถปรากฏที่ตำแหน่งท้ายพยางค์ได้

ลักษณะทางสวนสัทศาสตร์ของหน่วยเสียงพยัญชนะแต่ละตัว มีความแตกต่างกันขึ้นอยู่กับตำแหน่งและลักษณะการเปล่งเสียง ต่อไปนี้จะกล่าวถึงลักษณะการเปล่งเสียงของหน่วยเสียงพยัญชนะที่ใช้ในงานวิจัยนี้ ทั้ง 10 หน่วยเสียง คือ /p/, /ph/, /d/, /c/, /ch/, /n/, /f/, /s/, /r/ และ /w/ สามารถดูรูปประกอบได้จากรูปที่ 2.6 (ดุษฎีพร ชำนิโรคศานต์ อ้างจากภาษาไทย 3, 2533)

1) เสียงกักเกิดที่ฐานริมฝีปาก ไม่ก้องไม่พ่นลม /p/

ในการเปล่งเสียง ริมฝีปากอยู่ชิดกันสนิท ลมจะถูกกักในช่องปาก เส้นเสียงอยู่ห่างกัน มีช่องระหว่างเส้นเสียงทำให้ลมจากปอดผ่านช่องระหว่างเส้นเสียงได้สะดวก เมื่อจะเปล่งเสียงต้องเปิดที่กักที่ริมฝีปากให้ลมผ่านออก ปริมาณลมเป็นไปตามปกติหรืออาจกล่าวได้ว่าไม่มีกลุ่มลมตามออกมา

2) เสียงกักเกิดที่ฐานริมฝีปาก ไม่ก้องพ่นลม /ph/

ในการเปล่งเสียง ริมฝีปากอยู่ชิดสนิทกัน ลมจะถูกกักในช่องปาก เส้นเสียงอยู่ห่างกัน มีช่องระหว่างเส้นเสียง ขณะเปล่งเสียงลมจากปอดจะผ่านช่องระหว่างเส้นเสียงได้สะดวก ในขณะเดียวกันก็เปิดที่กักที่ริมฝีปากให้ลมผ่านออกไป ปริมาณลมมากกว่าปกติหรืออาจกล่าวได้ว่ามีกลุ่มลมตามออกมา

3) เสียงกักเกิดที่ฐานปุ่มเหงือก ก้อง /d/

ในการเปล่งเสียง ลิ้นส่วนปลายจะแตะปุ่มเหงือก ลมจะถูกกักในบริเวณดังกล่าว เส้นเสียงอยู่ชิดกันแต่ไม่ชิดกันสนิท ขณะเปล่งเสียง ลมจากปอดจะผ่านเส้นเสียงขึ้นมาทำให้เสียงเกิดการสั่นสะเทือน ปริมาณลมเป็นไปตามปกติ

4) เสียงกักเสียดแทรกเกิดที่ฐานปุ่มเหงือก-เพดานแข็ง ไม่ก้อง ไม่พ่นลม /c/

เมื่อเริ่มต้นจะเปล่งเสียง ลิ้นส่วนปลายจะขึ้นไปแตะที่ปุ่มเหงือก-เพดานแข็ง ลมจะถูกกักในบริเวณดังกล่าว เส้นเสียงอยู่ห่างกัน มีช่องระหว่างเส้นเสียงทำให้ลมจากปอดผ่านช่องระหว่างเส้นเสียงได้สะดวก เมื่อจะเปล่งเสียง ลิ้นส่วนปลายจะลดระดับลงเล็กน้อยทำให้มีช่องแคบ ๆ ระหว่างลิ้นส่วนปลายและปุ่มเหงือก-เพดานแข็ง ลมที่ถูกกักจะผ่านช่องระหว่างลิ้นส่วนปลายและปุ่มเหงือก-เพดานแข็งออกมาในลักษณะของการเสียดแทรก ปริมาณลมเป็นไปตามปกติ หรืออาจกล่าวได้ว่าไม่มีกลุ่มลมตามออกมา

5) เสียงกักเสียดแทรกเกิดที่ฐานปุ่มเหงือก-เพดานแข็ง ไม่ก้อง พ่นลม /ch/

เมื่อเริ่มต้นจะเปล่งเสียง ลิ้นส่วนปลายจะขึ้นไปแตะที่ปุ่มเหงือก-เพดานแข็ง ลมจะถูกกักในบริเวณดังกล่าว เส้นเสียงอยู่ห่างกัน ลมจากปอดจะผ่านช่องระหว่างเส้นเสียงได้สะดวก เมื่อจะเปล่งเสียง ลิ้นส่วนปลายจะลดระดับลงเล็กน้อย ทำให้มีช่องแคบ ๆ ระหว่างลิ้นส่วนปลายและปุ่มเหงือก-เพดานแข็งออกมาในลักษณะของการเสียดแทรก ปริมาณลมมากกว่าปกติ หรืออาจกล่าวได้ว่ามีกลุ่มลมตามออกมา

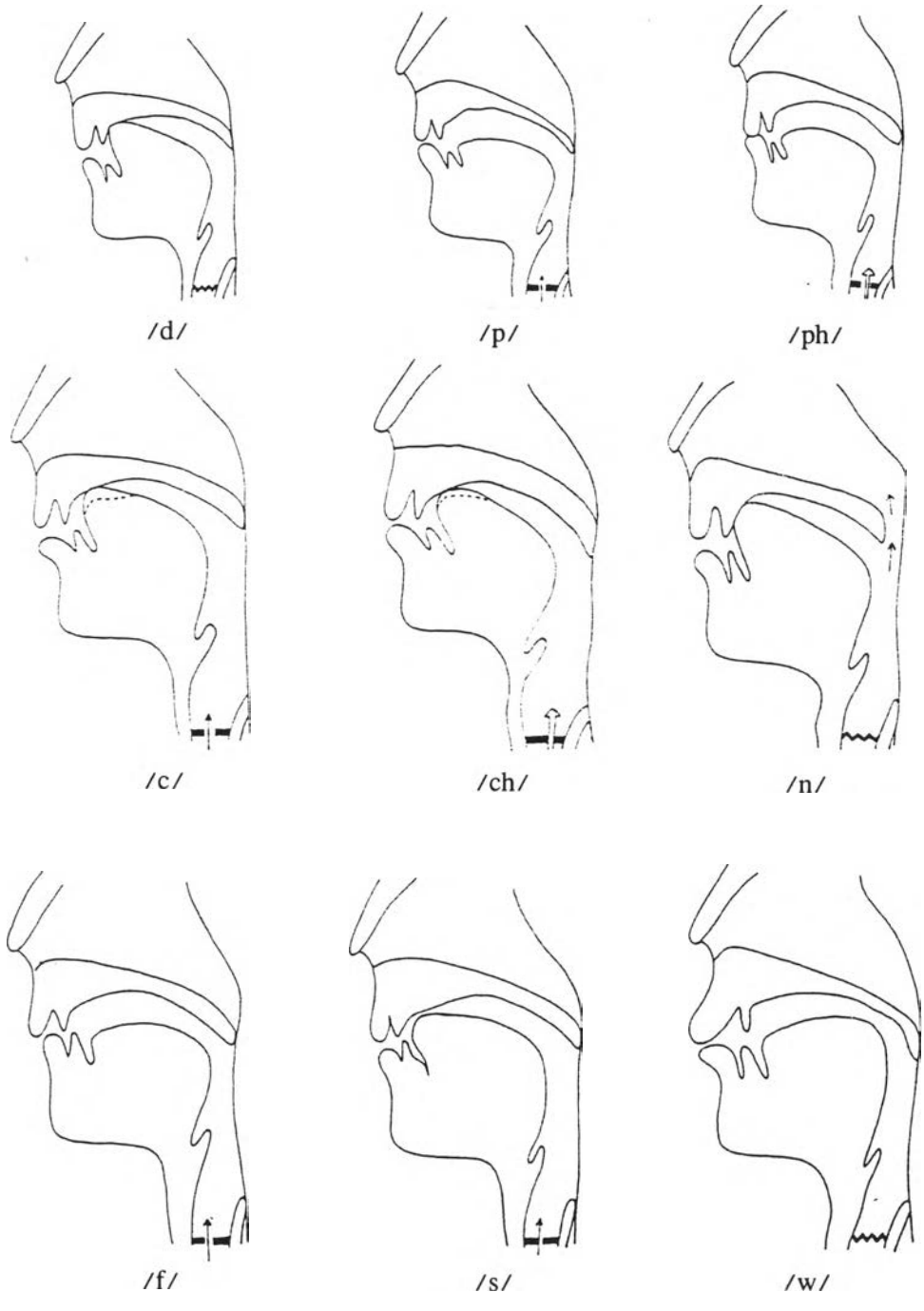
6) เสียงนาสิกเกิดที่ฐานปุ่มเหงือก /n/

เมื่อเริ่มต้นจะเปล่งเสียง ลิ้นส่วนปลายจะแตะปุ่มเหงือก ลมจะถูกกักในบริเวณดังกล่าว ในขณะเดียวกันเพดานอ่อนและลิ้นไก่ลดระดับลง ทำให้ลมผ่านออกไปทางช่องจมูกได้ เมื่อเปิดที่กักที่ปุ่มเหงือก ลมจะออกทางปากด้วย ขณะที่เปล่งเสียงนั้นเส้นเสียงอยู่ชิดกันแต่ไม่ชิดกันสนิท ลมจากปอดจะผ่านเส้นเสียงขึ้นมาทำให้เส้นเสียงเกิดการสั่นสะเทือน

7) เสียงเสียดแทรกเกิดที่ฐานริมฝีปากและฟัน /f/

เมื่อเริ่มต้นจะเปล่งเสียง ฟันบนจะอยู่ใกล้ริมฝีปากล่าง (ฟันบนจะสัมผัสริมฝีปากล่างเล็กน้อย) ลมในช่องปากไม่ถูกกัก แต่จะผ่านออกไม่สะดวก ต้องเสียดแทรกผ่านช่องระหว่างริมฝีปากกับฟัน ขณะเปล่งเสียง เส้นเสียงจะอยู่ห่างกัน มีช่องระหว่างเส้นเสียง ลมจากปอดผ่านช่องระหว่างเส้นเสียงได้สะดวก

รูปที่ 2.6 ลักษณะการเปล่งเสียงของ /d/, /p/, /ph/, /c/, /ch/, /n/, /f/, /s/ และ /w/
(ดุษฎีพร ชำนิโรคศานต์ อ้างจากภาษาไทย 3, 2533)



คำอธิบาย
- แสดงว่าเส้นเสียงอยู่ประชิดกัน
- แสดงว่าเส้นเสียงอยู่ห่างกัน
- เส้นเสียงอยู่ห่างกัน ไม่มีกลุ่มลม
- เส้นเสียงอยู่ห่างกัน มีกลุ่มลม
- เส้นเสียงสั่นสะเทือน

8) เสียงเสียดแทรกเกิดที่ฐานปุ่มเหงือก /s/

เมื่อเริ่มต้นจะเปล่งเสียง ลิ้นส่วนปลายจะขึ้นไปใกล้ปุ่มเหงือก ลมในช่องปากไม่ถูกกัก แต่จะผ่านออกไม่สะดวก ต้องเสียดแทรกผ่านช่องระหว่างลิ้นส่วนปลายกับปุ่มเหงือก ขณะเปล่งเสียง เส้นเสียงจะอยู่ห่างกัน มีช่องระหว่างเส้นเสียง ลมจากปอดจะผ่านช่องระหว่างเส้นเสียงได้สะดวก

9) เสียงกระทบ /r/

ฐาน กับ กรณ์ จะอยู่ชิดกันชั่วครู่เท่านั้น เมื่อเริ่มต้นจะเปล่งเสียง ลิ้นจะอยู่ในระดับราบ แล้วยกปลายลิ้นขึ้นไปกระทบปุ่มเหงือก แล้วสะบัดปลายลิ้นลงอย่างรวดเร็ว ขณะที่เปล่งเสียงนั้น เส้นเสียงอยู่ชิดกันแต่ไม่ชิดกันสนิท ลมจากปอดจะผ่านเส้นเสียงขึ้นมาทำให้เส้นเสียงเกิดการสั้นสะเทือน เสียงที่เปล่งออกมาเป็นเสียงกระทบ

10) เสียงกึ่งสระที่เพดานอ่อน /w/

เมื่อเริ่มต้นจะเปล่งเสียง ยกลิ้นส่วนหลังขึ้นไปทางเพดานอ่อน แต่ไม่ใกล้มากจนเกิดเป็นช่องแคบ ๆ อันจะทำให้ลมต้องเสียดแทรกออกมา ห่อริมฝีปากขณะเปล่งเสียง เส้นเสียงอยู่ชิดกันแต่ไม่ชิดกันสนิท ลมจากปอดจะผ่านเส้นเสียงทำให้เสียงเกิดการสั่นสะเทือน เสียงที่เปล่งออกมาเป็นเสียงกึ่งสระที่เพดานอ่อน ก้อง

4. หน่วยเสียงสระ

ในภาษาไทยมีสระเดี่ยว 18 หน่วย ประกอบด้วย สระยาว 9 หน่วย สระสั้น 9 หน่วย ดังตารางที่ 2.3

ตารางที่ 2.3 หน่วยเสียงสระในภาษาไทย

ลักษณะริมฝีปาก	ริมฝีปากรี		ริมฝีปากห่อ
ตำแหน่งของลิ้น / ระดับลิ้น	หน้า	กลาง	หลัง
สูง	i,ii	ɯ,ɯɯ	u,uu
กลาง	e,ee	ɤ,ɤɤ	o,oo
ต่ำ	ɛ,ɛɛ	a,aa	ɔ,ɔɔ

นอกจากนี้ยังมีหน่วยเสียงสระผสมที่เกิดจากการใช้ลิ้นออกเสียงสระหนึ่งแล้วลิ้นยังไม่กลับเข้าที่เดิม แต่ลิ้นได้ออกเสียงสระอีกเสียงหนึ่ง สระผสมมี 2 ส่วน 6 หน่วยเสียง ประกอบด้วยสระเสียงยาว 3 หน่วยเสียง คือ iia, vva และ uua สระเสียงสั้น 3 หน่วยเสียง คือ ia, va และ ua

ในการออกเสียงสระ ซึ่งเป็นการออกเสียงที่อวัยวะในการออกเสียงอยู่ห่างกันจนไม่เกิดการเสียดสี และลมผ่านเส้นเสียงทำให้เส้นเสียงสั้น ผ่านออกมายังช่องปาก ออกมาสู่ภายนอก

ตรงส่วนกลางของช่องออกเสียง การออกเสียงลักษณะนี้จะให้คลื่นเสียงที่แสดงความถี่กำทอนชัดเจน (สุดาพร ลักษณียนาวิน, 2529)

Pickett (1980) ได้ให้กฎที่แสดงความสัมพันธ์ระหว่าง F1, F2 และ F3 กับการทำงานของอวัยวะในการออกเสียงสระ ซึ่งพอสรุปได้ดังนี้

1) กฎเกี่ยวกับความยาวของทางเดินเสียง

ความถี่กำทอนของเสียงสระจะมีค่า ตรงข้ามกับความยาวของทางเดินเสียง ถ้าช่องออกเสียงยาวค่าความถี่ฟอร์แมนท์ก็จะต่ำ

ถ้าเราแบ่งช่องปากเป็น 3 ตอนใหญ่ ๆ ตามรูปที่ 2.7 โดยที่ ก. คือบริเวณระหว่างลิ้นส่วนกลางกับเพดานแข็ง ข. คือบริเวณระหว่างลิ้นส่วนหลังกับเพดานอ่อน และ ค. คือบริเวณระหว่างโคนลิ้นกับผนังคอ สามารถเขียนกฎความสัมพันธ์ของค่าความถี่กับการออกเสียงได้ดังนี้

2) กฎเกี่ยวกับค่าของ F1

ค่าของ F1 จะลดลงถ้าบริเวณระหว่างลิ้นส่วนกลางและเพดานแข็ง (บริเวณ ก) และบริเวณระหว่างลิ้นส่วนหลังกับเพดานอ่อน (บริเวณ ข) แคบลง - สระปิด

ค่าของ F1 จะเพิ่มขึ้นถ้าบริเวณระหว่างโคนลิ้นและผนังคอ (บริเวณ ค) แคบลง - สระเปิด

3) กฎเกี่ยวกับค่าของ F2

ค่าของ F2 จะเพิ่มขึ้นถ้าช่องระหว่างลิ้นส่วนกลางกับเพดานแข็ง (บริเวณ ก) แคบลง - สระหน้า

ค่าของ F2 จะลดลงถ้าช่องว่างระหว่างลิ้นส่วนหลังและเพดานอ่อน (บริเวณ ข) แคบลง - สระหลัง

4) กฎเกี่ยวกับการห่อริมฝีปาก

ค่าความถี่ฟอร์แมนท์ทุกตัวจะลดลงถ้ามีการห่อริมฝีปาก ช่องริมฝีปากเวลาห่อกลมยิ่งแคบเท่าใด ค่าความถี่ฟอร์แมนท์ก็จะลดต่ำลงเท่านั้น

รูปที่ 2.7 ลักษณะช่องปาก เมื่อลิ้นอยู่ในตำแหน่งต่าง ๆ (สุดาพร ลักษณียนาวิน, 2529)

จากการศึกษาความสัมพันธ์ระหว่างค่าความถี่ฟอร์แมนท์และลักษณะของเสียงสระ นักสัทศาสตร์สามารถสร้างกฎความสัมพันธ์ระหว่างลักษณะการออกเสียงสระกับค่าความถี่ฟอร์แมนท์ ดังที่กล่าวมาแล้ว หรืออาจสรุปให้ง่าย ๆ ได้ว่า

สระปิดจะมี F1 ต่ำ เช่น สระ i,ɯ

สระเปิดจะมี F1 สูง เช่น สระ a

สระหน้าจะมี F2 สูง เช่น สระ i,e

สระหลังจะมี F2 ต่ำ เช่น สระ ɯ,u

สระปากกลมจะมีค่าความถี่ฟอร์แมนท์ต่ำกว่าสระปากไม่กลม เช่น ค่าความถี่ฟอร์แมนท์ของสระ u จะต่ำกว่าสระ ɯ

Abramson (1962) ได้วัดค่าความถี่ฟอร์แมนท์ของสระในภาษาไทยมาตรฐานโดยใช้ผู้พูดเพศชาย 2 คน สรุปได้ดังตารางที่ 2.4

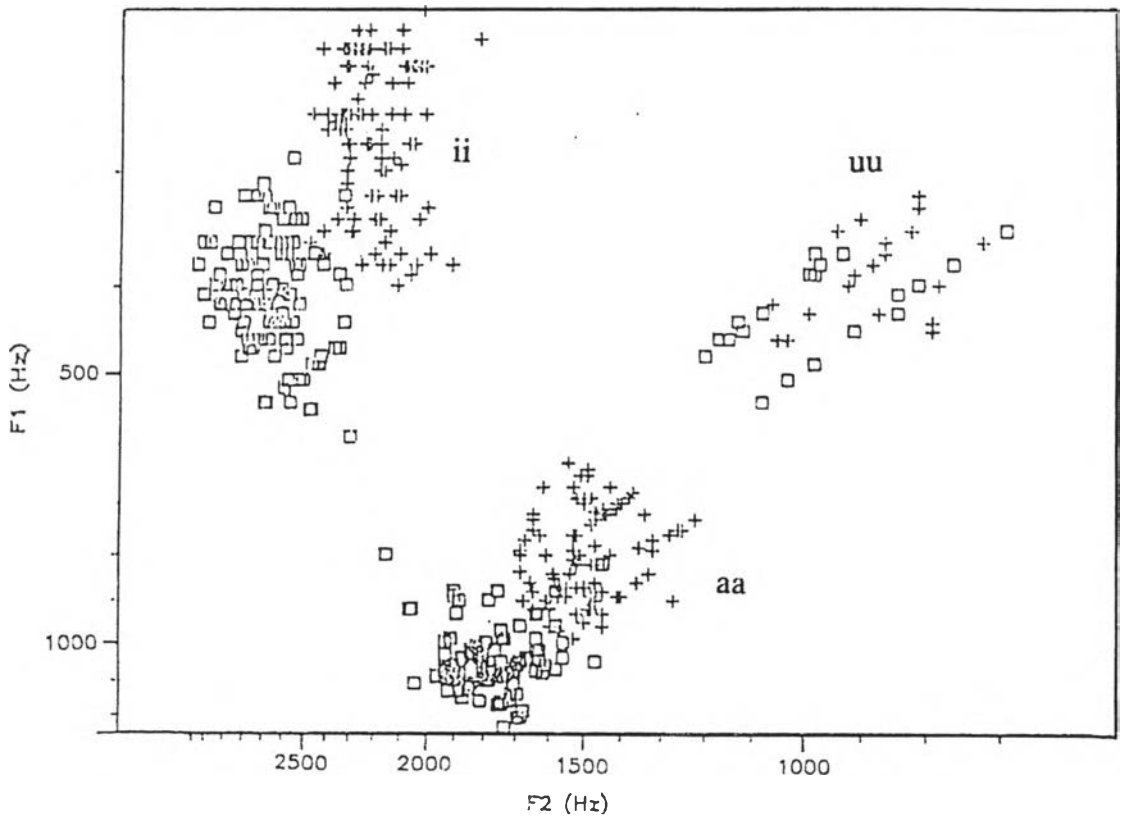
ตารางที่ 2.4 ค่าความถี่ฟอร์แมนท์ของสระในภาษาไทย (Abramson,1962)

Hight Vowel	i	ii	ɯ	ɯɯ	u	uu
Formant 2	2100	2220	1380	1380	720	660
Formant 1	360	300	300	300	360	300
Mid Vowel	e	ee	ɤ	ɤɤ	o	oo
Formant 2	1980	1980	1200	1260	840	840
Formant 1	540	480	540	540	480	480
Low Vowel	ɛ	ɛɛ	a	aa	ɔ	ɔɔ
Formant 2	1800	1800	1380	1380	960	960
Formant 1	780	720	720	780	660	660

วิสิทธิ์ ลีลาศิริวงศ์ (2535) ได้ศึกษาลักษณะทางสวนศาสตร์ของสระ โดยทำการวัดค่าความถี่ฟอร์แมนท์ที่หนึ่งและสอง ของเสียงสระ อี อา และ อู ในภาษาไทย ทั้งในเพศชายและเพศหญิง

เมื่อนำค่าความถี่ฟอร์แมนท์ที่หนึ่ง (F1) และค่าความถี่ฟอร์แมนท์ที่สอง (F2) ของสระทั้งสามมาเขียนกราฟ โดยกำหนดแกนดิ่งเป็นค่า F1 และแกนราบเป็นค่า F2 ใช้สเกลที่กลับสเกลจากปกติซึ่งเรียงค่าต่ำไปสูงในแนวแกนเป็นสเกลเรียงค่าสูงไปต่ำ ซึ่งแสดงในรูปที่ 2.8 พบว่าสามารถแยกหน่วยเสียงสระทั้งสามออกจากกันอย่างชัดเจน

รูปที่ 2.8 ความถี่ฟอร์แมนท์ที่หนึ่ง (แกนดิ่ง) และฟอร์แมนท์ที่สอง (แกนราบ) ของสระ อี อา และ อู จำแนกเพศชาย (เครื่องหมาย +) และเพศหญิง (เครื่องหมาย □)
(วิสิทธิ์ ลีลาศิริวงศ์, 2535)

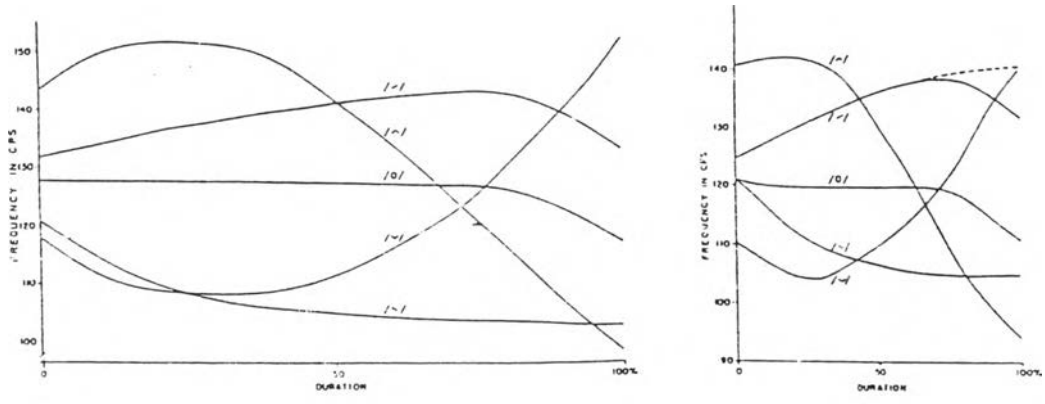


5. หน่วยเสียงวรรณยุกต์

ในภาษาไทยมี 5 เสียงวรรณยุกต์ แบ่งออกเป็น 2 กลุ่มคือ วรรณยุกต์ที่มีเสียงคงระดับ (Static tones) 3 เสียง คือ เสียงสามัญ (mid tone) เสียงเอก (low tone) และเสียงตรี (high tone) และวรรณยุกต์ที่มีเสียงไม่คงระดับ (Dynamic tones) 2 เสียง คือ เสียงโท (falling tone) และ เสียงจัตวา (rising tone)

Abramson (1962) ได้ศึกษาวรรณยุกต์ในภาษาไทย โดยใช้เครื่องวิเคราะห์สัญญาณเสียง (Sound spectrograph) ทำการวิเคราะห์ค่าความถี่มูลฐานของวรรณยุกต์ทั้ง 5 เมื่อปรากฏในคำพูดเดี่ยว จำนวน 88 คำ เป็นคำสระเสียงสั้น 39 คำ และคำสระเสียงยาวหรือสระผสม 49 คำ ใช้ผู้บอกภาษาเพศชาย 1 คน พบว่าลักษณะความถี่มูลฐานเป็นดังรูปที่ 2.9

รูปที่ 2.9 ค่าความถี่มูลฐานของเสียงวรรณยุกต์ทั้ง 5 เสียงในภาษาไทย (Abramson,1962)

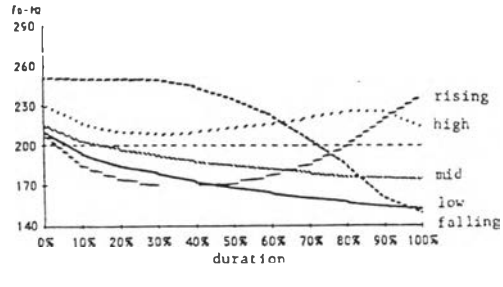


สระยาว สระสั้น

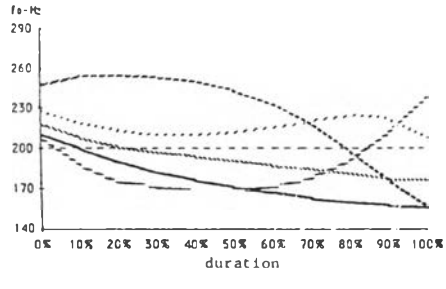
ปิยฉัตร ปานโรจน์ (2534) ได้ศึกษาลักษณะเชิงสวนสัทศาสตร์ของวรรณยุกต์ อันได้แก่ ทิศทางของความถี่มูลฐาน ค่าความถี่มูลฐานเฉลี่ย และพิสัยของค่าความถี่มูลฐาน ในภาษาไทยกรุงเทพฯ จากผู้พูด 3 กลุ่มอายุ คือ 50 - 60 ปี, 30 - 40 ปี และ 10 - 20 ปี พบว่าลักษณะดังกล่าวมีการแปรตามกลุ่มอายุอย่างชัดเจนในวรรณยุกต์ตรีและโท มีความแตกต่างกันมากระหว่างกลุ่มเด็กและกลุ่มผู้ใหญ่ ซึ่งแสดงในรูปที่ 2.10

รูปที่ 2.10 ค่าความถี่มูลฐานของเสียงวรรณยุกต์ทั้ง 5 เสียงในภาษาไทย
ของผู้พูด 3 กลุ่มอายุ (ปิยฉัตร ปานโรจน์, 2534)

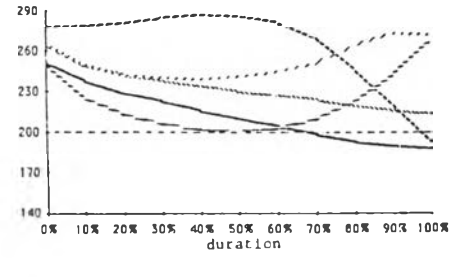
50-60 ปี



30-40 ปี



15-20 ปี



ทฤษฎีการวิเคราะห์เสียง

การวิเคราะห์เสียงพูด คือการหาค่าพารามิเตอร์ต่าง ๆ ที่เป็นลักษณะเฉพาะเพื่อเป็นตัวแทนของเสียงต่าง การวิเคราะห์เสียงพูดนั้นมีอยู่หลายวิธี มีทั้งในด้านโดเมนเวลา (time domain) คือ วิเคราะห์จากรูปคลื่นของสัญญาณเสียงตามแกนเวลาโดยตรง และในด้านโดเมนความถี่ (frequency domain) คือวิเคราะห์จากสเปกตรัมของสัญญาณเสียง

1. การวิเคราะห์สัญญาณเสียงในช่วงเวลาสั้น ๆ

เนื่องจากว่าสัญญาณเสียงเป็นสัญญาณที่แปรตามเวลา มีการแปรเปลี่ยนที่ไม่แน่นอน เช่น ในขณะที่พูดช้า ๆ รูปร่างของโพรงเสียงรวมทั้งลักษณะรูปแบบของการกระตุ้นอาจจะไม่มีการเปลี่ยนแปลงในช่วงเวลาที่นานที่สุดประมาณ 200 มิลลิวินาที แต่ในขณะที่พูดอย่างรวดเร็ว อาจจะมีช่วงที่ไม่มีการเปลี่ยนแปลงสั้นมากคือประมาณ 80 มิลลิวินาที ก็ได้ (O'shaughnessy, 1987)

ดังนั้นเทคนิคในการวิเคราะห์เสียงพูดส่วนใหญ่แล้ว จะสมมติให้สัญญาณเสียงมีคุณสมบัติที่เปลี่ยนแปลงสัมพันธ์กับเวลาอย่างเชื่องช้า นั่นก็คือเราจะต้องแบ่งทำการวิเคราะห์หาพารามิเตอร์ของสัญญาณเสียงพูดในช่วงเวลาสั้น ๆ เหมือนมองผ่านช่องแคบ ๆ ที่เรียกว่า ช๊อต-ไทม์วินโดว์ (short-time window) เมื่อเทียบตามเวลาที่เสียงจะมีการเปลี่ยนแปลงได้ เพื่อจะได้มองเห็นเหมือนกับว่า เราหาพารามิเตอร์นั้น ๆ ได้มาจากสัญญาณเสียงที่อยู่ภายในช่องแคบ ๆ และมีความเสถียรภายในช่วงเวลาสั้น ๆ

เทคนิคส่วนใหญ่จะกำหนดให้พารามิเตอร์ได้มาจากค่าเฉลี่ยจากการวิเคราะห์ภายในช่วงเวลาแคบ ๆ นั้น สำหรับกรณีที่ต้องพิจารณาพารามิเตอร์ต่าง ๆ ที่มีการเปลี่ยนแปลง ก็จะทำการแบ่งสัญญาณเสียงออกเป็นช่องหลาย ๆ ช่องหรืออาจจะเรียกว่า กรอบการวิเคราะห์ (analysis frame) ดังนั้นพารามิเตอร์ต่าง ๆ จะสามารถหาได้ทันเพียงพอที่จะติดตามการเปลี่ยนแปลงของสัญญาณ สำหรับในช่วงที่สัญญาณมีการเปลี่ยนแปลงช้า อาจจะกำหนดให้ช่องแคบมีขนาดใหญ่ประมาณ 100 มิลลิวินาที แต่ในทางตรงข้ามถ้าสัญญาณมีการเปลี่ยนแปลงเร็ว ก็ต้องใช้ช่องแคบที่มีขนาดเล็กมาก ๆ ประมาณ 5 - 10 มิลลิวินาที เพื่อป้องกันการสูญหายของรายละเอียดของสัญญาณถัดไป

1.1 รูปแบบของช่องแคบ (Windows)

การกำหนดขนาดของช่องแคบที่จะใช้ขึ้นอยู่กับ

1) ช่องจะต้องสั้นพอ ที่จะทำให้คุณสมบัติของเสียงที่กำลังพิจารณาไม่มีการเปลี่ยนแปลงอย่างมีนัยสำคัญในช่องแคบนั้น

2) ช่องแคบจะต้องยาวพอ ที่จะทำให้การจัดเตรียมตัวอย่างของเสียงเพื่อจะนำไปคำนวณหาพารามิเตอร์ให้ได้ตามต้องการอย่างเช่น ในกรณีที่มีสัญญาณรบกวนเข้ามาแทรกอยู่บางช่วงในสัญญาณเสียงด้วย ถ้าเราเลือกใช้ช่องแคบที่มีขนาดใหญ่กว่า เมื่อทำการหาค่าพารามิเตอร์โดยเฉลี่ยแล้ว ก็จะทำให้ส่วนประกอบของสัญญาณรบกวนถูกตัดทิ้งหรือมองข้ามไป

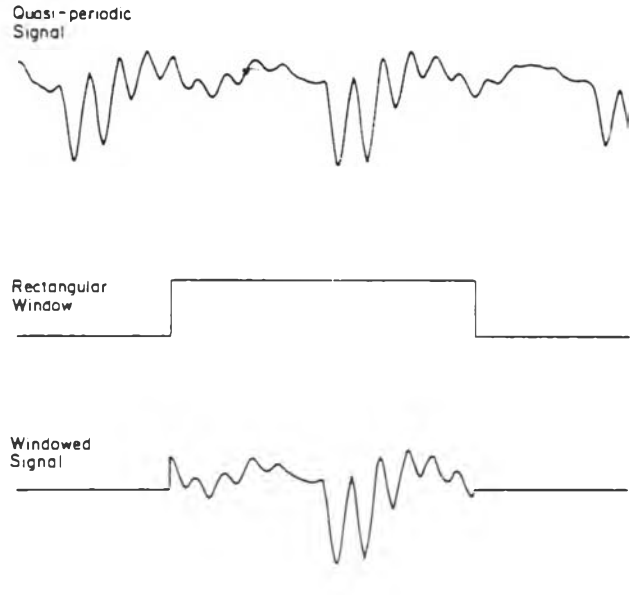
3) ช่องแคบที่เหมาะสม ไม่ควรสั้นเกินกว่าช่วงหนึ่งคาบของสัญญาณเสียงในช่วงที่กำลังวิเคราะห์ เงื่อนไขนี้จะมีผลต่อค่าเฟรมเรท (frame rate ซึ่งก็คือ จำนวนครั้งต่อวินาทีที่ทำการวิเคราะห์สัญญาณเสียง โดยการขยับช่องแคบไปเป็นคาบ ๆ ตามแกนเวลา) ตามปกติเฟรมเรทจะมีค่าประมาณ 2 เท่าของส่วนกลับของขนาดช่องแคบ นั่นก็คือช่องแคบถัด ๆ กันไปจะมีการซ้อนทับกัน 50 เปอร์เซ็นต์

การนำฟังก์ชันของช่องแคบที่มีช่วงขนาดจำกัด w(n) มาคูณเข้ากับสัญญาณ s(n) จะทำให้ได้กลุ่มตัวอย่างของเสียงพูดที่ถูกกำหนดน้ำหนักให้แปรไปตามรูปร่างของช่องแคบ รูปแบบของช่องแคบที่ง่ายที่สุดก็คือ กรอบแบบสี่เหลี่ยม (rectangular window) ซึ่งมีนิยามดังนี้

$$w(n) = \begin{cases} 1, & n = 0,1,\ldots,N-1 \\ 0, & n \quad otherwise \end{cases} \qquad (2.1)$$

ในสมการ (2.1) นี้ คือการกำหนดช่องของการวิเคราะห์ให้มีจำนวนตัวอย่าง N ตัวอย่าง

รูปที่ 2.11 ช่องแคบแบบสี่เหลี่ยม (Witten, 1982)



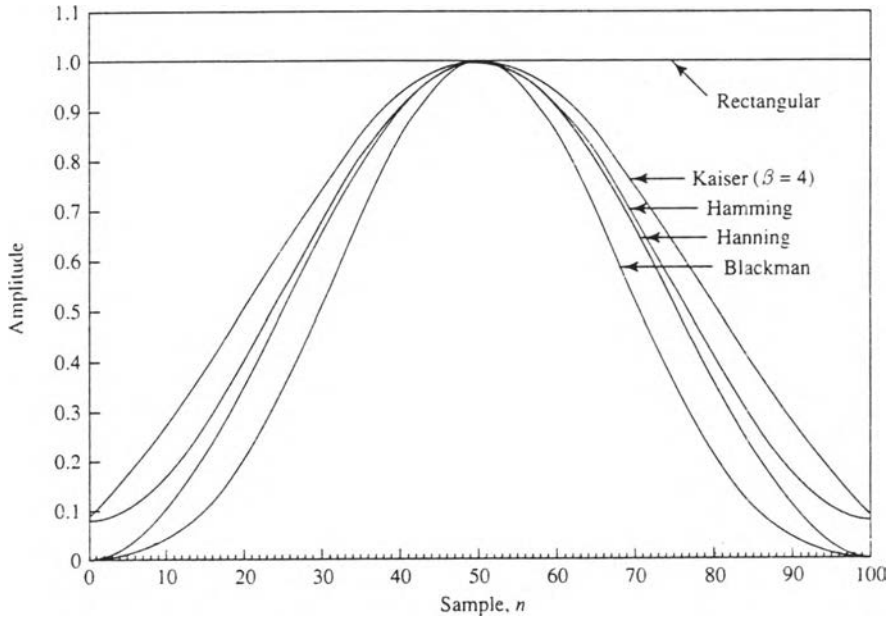
โดยปกติ การประยุกต์ส่วนมากจะใช้ช่องแคบที่มีช่วงกว้างกว่าช่องของข้อมูลที่คงที่ และเน้นเฉพาะช่องตรงกลางของช่องแคบให้เป็นส่วนของข้อมูลที่จะทำการวิเคราะห์ เช่น ข้อมูลสัญญาณเสียงที่มีลักษณะคงที่ในช่วงเวลา 10 มิลลิวินาที ก็อาจจะใช้ช่องแคบขนาด 20 มิลลิวินาที โดยช่วงกึ่งกลางช่องแคบขนาด 10 มิลลิวินาที จะมีการเน้นเพิ่มน้ำหนักให้มากกว่าช่วง 5 วินาที ที่ริมทั้งสองข้างของช่องแคบ

รูปแบบของช่องแคบในลักษณะนี้ ก็มีฟังก์ชันของช่องแคบหลายลักษณะด้วยกัน ตัวอย่างเช่น Blackman, Barlett, Hamming, Hanning เป็นต้น โดยฟังก์ชันของช่องแคบที่นิยมใช้กันมากในการวิเคราะห์สัญญาณเสียงก็คือ ฟังก์ชัน Hamming ซึ่งมีรูปร่างตามลักษณะของโคซายน์พลัส (cosine pulse) มีนิยามดังนี้

$$w(n) = \begin{cases} 0.54 - 0.46\cos(2\pi n)/(N-1), & n = 0,1,\ldots,N-1 \\ 0, & n \quad otherwise \end{cases} \qquad (2.2)$$

ส่วนที่เรียวลงไปตามขอบของช่องแคบ มีผลทำให้เกิดการเลื่อนออกไปแบบเป็นคาบของเฟรมวิเคราะห์ตามสัญญาณอินพุท (input) แต่ไม่มีผลกระทบต่อพารามิเตอร์ของเสียงที่เนื่องมากจากขอบเขตของคาบของพิทช์ (pitch periodic boundaries) หรือการเปลี่ยนแปลงอื่น ๆ ที่มีลักษณะเดียวกันนี้ในสัญญาณเสียง

รูปที่ 2.12 ลักษณะของช่องแคบชนิดต่าง ๆ (Deller, Proakis และ Hansen, 1993)



2. การวิเคราะห์ในโดเมนเวลา

การวิเคราะห์โดเมนเวลา คือวิธีการประมวลสัญญาณเสียงพูดโดยยึดหลักการวิเคราะห์จากรูปคลื่นสัญญาณที่แปรตามแกนเวลาซึ่งมีอยู่หลายวิธีด้วยกัน ผลการวิเคราะห์สัญญาณเสียงในโดเมนเวลา เช่น ค่าพลังงาน (acoustic energy) ของเสียง ค่าจุดตัดศูนย์ (Zero Crossing) เป็นต้น วิธีต่าง ๆ ที่เป็นพื้นฐานในการวิเคราะห์สัญญาณเสียงในโดเมนเวลาบางวิธีจะกล่าวดังต่อไปนี้

2.1 การหาค่าพลังงานของสัญญาณเสียง

พลังงานของสัญญาณเป็นตัวแทนอันหนึ่งที่เรามักจะนำมาใช้ในการวิเคราะห์ลักษณะต่าง ๆ ของสัญญาณทั่ว ๆ ไป โดยพลังงานของสัญญาณ s(n) ใด ๆ ที่แปรตามเวลาสามารถนิยามได้ว่า

$$E = \sum_{n=-\infty}^{\infty} s^2(n) \qquad (2.3)$$

แต่สำหรับสัญญาณเสียงซึ่งเป็นสัญญาณที่แปรเปลี่ยนอยู่ตลอด ไม่มีเสถียรภาพตามเวลา เราจะต้องแบ่งสัญญาณออกมาพิจารณาเป็นช่วงเล็ก ๆ ตามแกนเวลาหรือเรียกว่าแบ่งออกเป็นเฟรม เช่น เฟรมละประมาณ 10-30 วินาที ดังนั้นก็จะสามารถหาพลังงานของเสียงในแต่ละเฟรมได้เป็น

$$E(m) = \sum_{n=0}^{N-1} [w(m)s(m-n)]^2 \qquad (2.4)$$

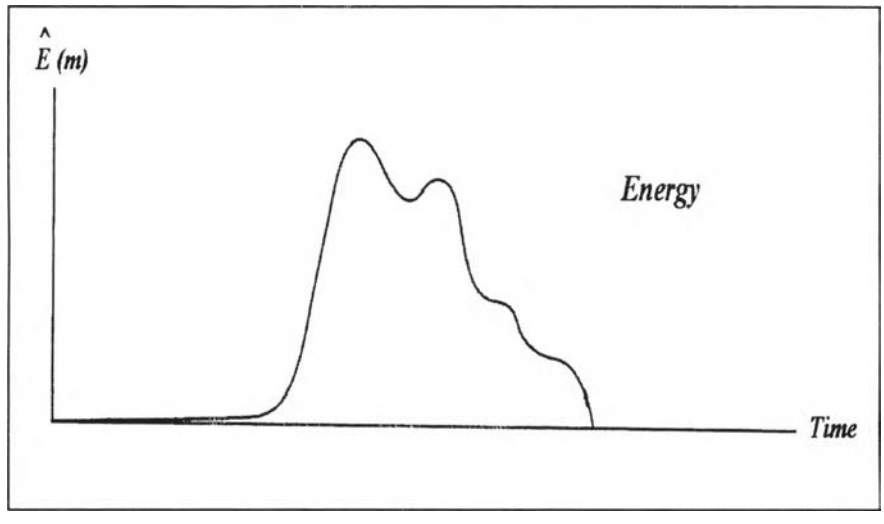
โดยที่ w(m) คือ วินโดว์ฟังก์ชัน (windows function) ที่ใช้กำหนดรูปร่างในการพิจารณาของสัญญาณเสียง s(n) ในหนึ่งเฟรม และ N คือจำนวนตัวอย่างของสัญญาณเสียงที่อยู่ในเฟรมหรือภายในกรอบของวินโดว์ฟังก์ชัน

การวัดค่าพลังงานดังในสมการที่ (2.4) นั้นมีข้อจำกัดตรงที่ว่ามันจะมีความไวต่อสัญญาณที่มีขนาดใหญ่ ๆ เนื่องจากเราใช้วิธียกกำลังสองค่าของสัญญาณอินพุท ดังนั้นการแก้ปัญหาอย่างหนึ่งก็คือวัดพลังงานของเสียงโดยใช้สมการดังนี้

$$\hat{E}(m) = \sum_{n=0}^{N-1} |w(m)s(m-n)| \qquad (2.5)$$

นั้นคือ การใช้ผลบวกของค่าสัมบูรณ์หรือเรียกว่าค่าแมกนิจูด (magnitude) แทนผลรวมของค่ายกกำลังสอง จากรูปที่ 2.13 แสดงถึงตัวอย่างรูปฟังก์ชันของพลังงานของคำว่า "four" จะสังเกตได้ว่าเราสามารถเห็นช่วงเริ่มต้นของคำก่อนจะเริ่มเสียง /f/ ซึ่งเป็นเสียงเสียดแทรก และช่วงท้ายเป็นเสียง /r/ ซึ่งมีระดับของพลังงานไม่คงที่

รูปที่ 2.13 รูปพลังงานของคำว่ า "four" (ดังแปลงจาก Deller, Proakis และ Hansen, 1993)



2.2 การหาค่าจุดตัดศูนย์ (Zero Crossing)

การเกิดจุดตัดศูนย์จะเกิดขึ้นจากการที่รูปคลื่นของสัญญาณมีการตัดกับแกนเวลา นั่นคือค่าของสัญญาณจะมีการเปลี่ยนสัญลักษณ์ทางคณิตศาสตร์ (Algebraic Sign) นั่นเอง อัตราการเกิดจุดตัดศูนย์เป็นเครื่องมืออย่างง่ายที่ใช้อธิบายการเปลี่ยนแปลงข้อมูลของสัญญาณ ค่าจุดตัดศูนย์นั้นสามารถนำมาใช้ในการตัดสินว่าสัญญาณเสียงนั้นเป็นเสียงก้อง (voiced) หรือเสียงไม่ก้อง (unvoiced) เนื่องจากเสียงก้องส่วนใหญ่จะมีค่าพลังงานอยู่ในช่วงความถี่ต่ำ ส่วนเสียงไม่ก้องจะมีพลังงานอยู่ในช่วงความถี่สูง และค่าจุดตัดศูนย์ก็มีความสัมพันธ์โดยตรงกับค่าความถี่ของสัญญาณ ดังนั้นจึงอาจสรุปได้ว่า สัญญาณเสียงที่มีค่าจุดตัดศูนย์สูงจะเป็นเสียงไม่ก้อง และสัญญาณที่มีค่าจุดตัดศูนย์ต่ำจะเป็นเสียงก้อง แต่อย่างไรก็ตามการกำหนดขนาดของค่าจุดตัดศูนย์ที่แน่นอนเพื่อจำแนกชนิดของเสียงนั้น จะต้องอาศัยผลจากการทดลองเป็นหลัก ในรูปที่ 2.14 จะแสดงให้เห็นถึงการกระจายของค่าจุดตัดศูนย์ของเสียงไม่ก้อง และเสียงก้อง

จุดตัดศูนย์สามารถนิยามได้ดังสมการที่ (2.6)

$$Z_s = \sum \left\{ \frac{|\text{sgn}\{s(n)\} - \text{sgn}\{s(n-1)\}|}{2} \right\} \qquad (2.6)$$

เมื่อ

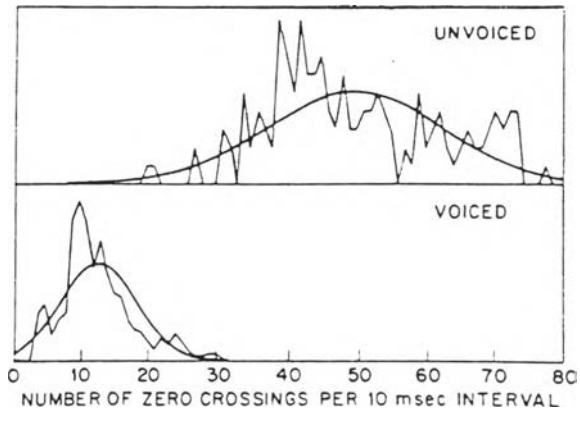
$$\text{sgn}\{s(n)\} = \begin{cases} +1, & s(n) \geq 0 \\ -1, & s(n) < 0 \end{cases} \qquad (2.7)$$

ในการวิเคราะห์สัญญาณเสียง จำเป็นต้องแบ่งสัญญาณออกมาพิจารณาเป็นเฟรมตามแกนเวลา ดังนั้นสมการในการหาค่าจัดตัดศูนย์ของสัญญาณในแต่ละเฟรมจะเป็นดังสมการที่ (2.8)

$$Z_s(m) = \frac{1}{N} \sum_{n=m-N+1}^{m} \frac{|\text{sgn}\{s(n)\} - \text{sgn}\{s(n-1)\}|}{2} w(m-n) \qquad (2.8)$$

โดยที่ w(m) คือ วินโดว์ฟังก์ชัน (windows function) ที่ใช้กำหนดรูปร่างในการพิจารณาของสัญญาณเสียง s(n) ในหนึ่งเฟรม และ N คือจำนวนตัวอย่างของสัญญาณเสียงที่อยู่ในเฟรมหรือภายในกรอบของวินโดว์ฟังก์ชัน

รูปที่ 2.14 การกระจายของค่าจุดตัดศูนย์ของเสียงไม่ก้อง และเสียงก้อง
(Rabiner และ Schafer, 1978)



3. การวิเคราะห์ตามแกนความถี่

การวิเคราะห์ตามแกนความถี่ ก็คือการวิเคราะห์ทางสเปกตรัมนั่นเอง โดยจะวิเคราะห์หรือพิจารณาสัญญาณเสียงในด้านของโดเมนความถี่ เหตุที่ต้องมีการวิเคราะห์ทางด้านสเปกตรัมนั้นก็เนื่องจากพารามิเตอร์ในการประมวลผลสัญญาณเสียงส่วนใหญ่ จะสามารถหาได้จากโดเมนความถี่ และสัญญาณที่สร้างออกมาจากการเปลี่ยนแปลงรูปร่างของโพรงเสียงของคนก็จะสามารถพิจารณาเป็นรูปแบบในโดเมนของความถี่ได้ง่าย และคงที่กว่าวิเคราะห์ในโดเมนเวลา

การวิเคราะห์ทางสเปกตรัมในช่วงเวลาสั้น ๆ เป็นวิธีการดั้งเดิมอันหนึ่งของการประมวลผลสัญญาณเสียงที่สำคัญ ด้วยข้อสมมติฐานที่ว่าสัญญาณเสียงในช่วงเวลายาว ๆ จะมีการเปลี่ยนแปลงอยู่ตลอดเวลาไม่มีสภาพของการคงที่ แต่ถ้าวิเคราะห์ในช่วงเวลาสั้น ๆ ช่วงหนึ่งจะสามารถแทนเป็นค่าสเปกตรัมของสัญญาณเสียงในช่วงเวลาดังกล่าวได้เป็นอย่างดี การวิเคราะห์สเปกตรัมของสัญญาณเสียงในช่วงเวลาอันสั้นนั้น จะเป็นพื้นฐานของเทคนิคการวิเคราะห์เสียงต่าง ๆ เช่น สเปกโทรกราฟ (spectrograph) เป็นต้น

3.1 การวิเคราะห์ฟูเรียร์ในช่วงเวลาสั้น ๆ

เป็นวิธีการวิเคราะห์สัญญาณทางด้านสเปกตรัมอย่างหนึ่ง การวิเคราะห์แบบฟูเรียร์นี้ จะสามารถแทนค่าสัญญาณเสียงออกมาในเทอมของอำพนและความถี่ ถ้าเรามองโพรงเสียงของคนเป็นระบบเชิงเส้นอันหนึ่ง การแปลงฟูเรียร์ของสัญญาณเสียงก็คือ ผลลัพธ์ที่เกิดจากการกระตุ้นของเส้นเสียงกับการตอบสนองของโพรงเสียงนั่นเอง นิยามของการแปลงฟูเรียร์ของสัญญาณ g(t) ใด ๆ ดังนี้

$$g(t) = \int_{-\infty}^{\infty} G(f) e^{+j2\pi ft}\, df \qquad (2.9)$$

$$G(f) = \int_{-\infty}^{\infty} g(t) e^{-j2\pi ft}\, dt \qquad (2.10)$$

อนุกรมฟูเรียร์ (Fourier Series) คือการปฏิบัติการกับรูปคลื่นที่มีลักษณะเป็นคาบตามแกนเวลา ถ้ารูปคลื่นมีคาบอยู่ในช่วง [0,b) จะได้ฟังก์ชันการแปลงที่มีนิยามดังนี้

$$g(t) = \sum_{r=-\infty}^{\infty} G(r) e^{+j2\pi rt/b} \qquad (2.11)$$

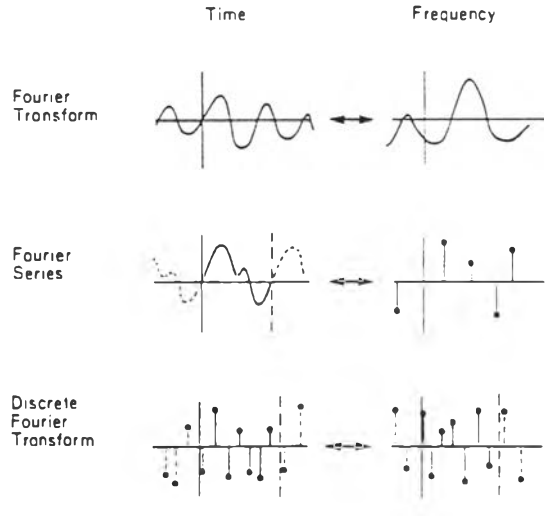
$$G(r) = \int_{0}^{b} g(t) e^{-j2\pi rt/b}\, dt \qquad (2.12)$$

อนุกรมฟูเรียร์ เป็นการแปลงระหว่างฟังก์ชันที่เป็นคาบเวลาใน โดเมนเวลา กับฟังก์ชันแบบไม่ต่อเนื่องในโดเมนความถี่ ถ้าเราทำการแปลงรูปคลื่นในโดเมนเวลาที่มีลักษณะเป็นฟังก์ชันแบบไม่ต่อเนื่องและมีช่วงที่จำกัด เราจะได้รูปคลื่นในโดนเมนของความถี่ซึ่งมีช่วงที่จำกัด มีลักษณะเป็นคาบและเป็นฟังก์ชันที่ไม่ต่อเนื่อง ลักษณะการแปลงแบบนี้ก็คือ การแปลงฟูเรียร์แบบไม่ต่อเนื่อง (Discrete Fourier Transform) ซึ่งก็คือการแปลงฟังก์ชันที่เป็นคาบแบบไม่ต่อเนื่องในโดเมนเวลา ให้เป็นฟังก์ชันที่เป็นคาบและไม่ต่อเนื่องในโดเมนความถี่นั่นเอง โดยมีนิยามกำหนดได้ดังนี้

$$g(n) = \frac{1}{N}\sum_{r=0}^{N-1} G(r)\, e^{+j2\pi rn/N} \qquad (2.13)$$

$$G(r) = \sum_{n=0}^{N-1} g(n)\, e^{-j2\pi rn/N} \qquad (2.14)$$

รูปที่ 2.15 การแปลงฟูเรียร์ อนุกรมฟูเรียร์ และการแปลงฟูเรียร์แบบไม่ต่อเนื่อง (Witten,1982)



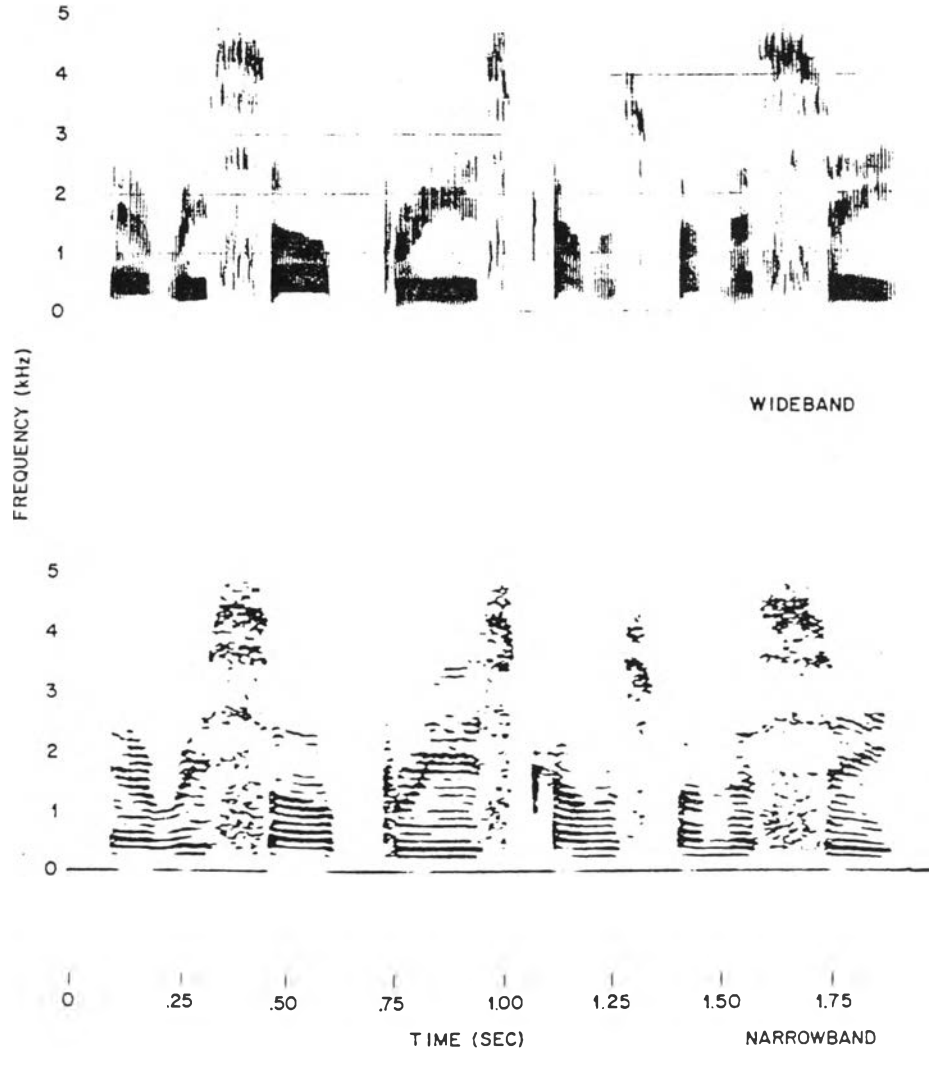
3.2 การแสดงสเปกตรัมของเสียง

อุปกรณ์ที่สำคัญในการวิเคราะห์สัญญาณเสียงซึ่งมีขึ้นเมื่อประมาณ 40 ปีที่แล้วก็คือ สเปกโทรแกรม (spectrogram) หรือ ซาว์นสเปกโทรกราฟ (sound spectrograph) ซึ่งจะแสดงให้เห็นถึงลักษณะเฉพาะของสัญญาณเสียงในรูปสามมิติภายในช่วงเวลาสั้น ๆ ประมาณ 2-3 วินาที โดยการใช้การแปลงฟูเรียร์ในช่วงเวลาสั้น ๆ ของสัญญาณเสียงมาพล๊อตรูปกราฟโดยมีแกนนอนเป็นแกนของเวลา และแกนตั้งเป็นแกนของความถี่ และแสดงค่าของแมกนิจูดด้วยความเข้มของภาพ ดังรูปที่ 2.16 ซึ่งแสดงสเปกโทรแกรมแบบช่วงกรองกว้าง และช่วงกรองแคบ

สเปกโทรแกรมแบบช่วงกรองกว้าง (wide-band spectrogram) จะแสดงให้เห็นถึงคาบของพิทช์ออกมาในแนวตั้ง ซึ่งจะสัมพันธ์กับช่วงที่อำพนของเสียงมีค่ามาก ขณะที่เส้นเสียงกำลังจะปิด ดังนั้นในกรณีของเสียงก้องเราสามารถที่จะตรวจพบได้โดยง่ายจากการค้นหาเส้นในแนวตั้งที่มีลักษณะเป็นคาบนี้ นอกจากนี้การพิจารณาในด้านความถี่ฟอร์แมนท์ อำพนของฮาร์มอนิกส์ภายใต้ฟอร์แมนท์แต่ละฟอร์แมนท์จะอยู่ในช่วงแถบความถี่ประมาณ 300 Hz แสดงให้เห็นเป็นแถบสีดำครอบคลุมในแต่ละฟอร์แมนท์ทางแนวตั้ง ความถี่ตรงกึ่งกลางของแต่ละแถบสีดำก็คือ ความถี่ฟอร์แมนท์โดยประมาณนั่นเอง

สเปกโทรแกรมแบบช่วงกรองแคบ (narrowband spectrogram) จะแสดงให้เห็นถึงฮาร์มอนิกแต่ละอันแทนคาบของพิทช์

รูปที่ 2.16 แสดงลักษณะสเปกโทรแกรมแบบช่วงกรองกว้าง และช่วงกรองแคบ (Rabiner และ Schafer, 1978)



ทฤษฎีทางสถิติพื้นฐานที่ใช้ในงานวิจัย

1. มัชฌิมเลขคณิต (Arithmetic mean)

คือจุดสมดุลย์ (balance point) ของคะแนนข้อมูลในกลุ่ม หาได้จาก ผลบวกของคะแนนทั้งหมดของข้อมูลหารด้วยจำนวนข้อมูลนั้น จะได้ค่าเฉลี่ย (average) ซึ่งเป็นตัวเลขจำนวนเดียวซึ่งเป็นตัวแทนของคะแนนทั้งหมดในข้อมูล สูตรคำนวณเป็นดังสมการที่ 2.15

$$\bar{X} = \frac{X_1 + X_2 + \ldots + X_N}{N} \qquad (2.15)$$

2. ส่วนเบี่ยงเบนมาตรฐาน (Standard deviation)

เป็นวิธีหนึ่งที่ใช้วัดการกระจายของข้อมูล โดยค่านี้จะทำให้ทราบว่ากลุ่มข้อมูลมีการเบี่ยงเบนจากค่าเฉลี่ยมากน้อยอย่างไร สูตรคำนวณเป็นดังสมการที่ 2.16

$$S.D. = \sqrt{\frac{N\sum x^2 - \bar{x}^2}{N(N-1)}} \qquad (2.16)$$

3. สัดส่วน (proportion) และร้อยละ (percent)

สัดส่วน คือ เศษส่วนของจำนวนย่อยกับจำนวนรวมทั้งหมด กล่าวคือให้ถือจำนวนรวมทั้งหมดเป็น 1 ส่วน

ร้อยละ คือ อัตราส่วนที่มีส่วนเป็น 100 ในงานวิจัยนี้ได้นำสัดส่วนและร้อยละมาใช้ในการเปรียบเทียบค่าทางสวนสัทศาสตร์ เช่น ค่า F2-Trans เฉลี่ย ของ /d/, /n/ และ /w/ มีค่า 2216, 2120 และ 1092 ตามลำดับ ค่า F2-Trans จะมีสัดส่วนต่อจำนวนรวมทั้งหมดคือ 5428 เท่ากับ $\frac{2216}{5428} : \frac{2120}{5428} : \frac{1092}{5428}$ แต่ในงานวิจัยนี้ได้ถือจำนวนทั้งหมดเป็น 100 ดังนั้นจึงมีสัดส่วนเท่ากับ 40.57 : 39.81 : 19.62

4. การจัดอันดับ (Ranking) หมายถึงการเรียงค่าจากสูงไปต่ำหรือต่ำไปสูง เพื่อศึกษาการกระจายของคะแนนหรือค่าร้อยละ ตามลำดับ

การเข้ากลุ่ม

ถ้ามีกลุ่มสิ่งของ m กลุ่ม แต่ละกลุ่มจะมีลักษณะที่บ่งบอกว่าเป็นกลุ่มของมันอยู่ n ลักษณะ การนำสิ่งของใด ๆ มารวมในกลุ่มที่มีลักษณะเหมือนมันที่สุด จะใช้วิธีการดังนี้

1. หาลักษณะทั้ง n ของสิ่งของทดสอบ

2. นำลักษณะทั้ง n มาหาค่าความต่าง (Δ_i) จากกลุ่มสิ่งของทั้ง m กลุ่ม โดยใช้สมการ (2.17)

$$\Delta_i = \sum_{j=1}^{n}\left[\left|\frac{S_{ij} - T_j}{T_j}\right| \times w_j\right] \qquad i = 1,2,..,m \qquad (2.17)$$

โดยที่ T_j แทนค่าลักษณะที่ j ของสิ่งของที่ต้องการทดสอบ

S_{ij} แทนค่าลักษณะที่ j ของสิ่งของจากกลุ่มสิ่งของที่ i

w_j แทนค่าน้ำหนักของลักษณะที่ j

i แทนกลุ่มสิ่งของใด ๆ

j แทนลักษณะใด ๆ

3. นำค่าความต่าง (Δ_i) ทั้ง m ค่า มาเปรียบเทียบกัน และเลือกใส่สิ่งของทดสอบในกลุ่มที่มีค่า Δ_i น้อยที่สุด