การอ่าน Spectrogram ของเสียงพูด

โดย ดร. อดิวงศ์ สุชาโต

แก้ไขล่าสุด 3 สิงหาคม พ.ศ. 2549

จุดประสงค์ของการอ่าน spectrogram ของเสียงพูดคือการหาคำพูดที่สอดคล้องกับเสียงพูดใน spectrogram นั้น โดยที่ผู้อ่านไม่ต้องฟังเสียงพูดนั้นจริงๆ การอ่าน spectrogram ทำได้โดยวิเคราะห์ลักษณะต่างๆ ของสัญญาณเสียงทั้งทางเวลา และ ความถี่ หลังจากนั้นจึงคาดเดารูปร่างของช่องทางเสียง (Vocal Tract) ตลอดจนแหล่งกำเนิดสัญญาณ ในตำแหน่งเวลาที่สอดคล้องกับ ลักษณะทางเวลา หรือทางความถี่นั้นๆ ถ้าเปรียบเทียบกับแบบจำลองการสร้างเสียงพูดแบบ source-filter (source-filter model of speech production) แล้ว การคาดเดารูปร่างของช่องทางเสียงนั้น ก็สอดคล้องกับการระบุ transfer function ของ filter ในแบบจำลอง การที่จะเลือกว่าจะพิจารณาลักษณะใดๆ ของสัญญาณเสียง ผู้อ่านจะต้องมีความรู้เกี่ยวกับกลไกในการกำเนิดเสียงชนิดต่างๆ ไม่ว่าจะเป็น สระ หรือ พยัญชนะต่างๆ โดยเฉพาะอย่างยิ่ง ความรู้เกี่ยวกับความสัมพันธ์ระหว่างความถี่ธรรมชาติของแบบจำลองช่องทางเสียง และ ตำแหน่งของลิ้น

เอกสารฉบับนี้จะขอเสนอหลักการคร่าวๆ ของการอ่าน spectrogram ดังนี้

ขั้นตอนคร่าวๆ ในการอ่าน spectrogram ของเสียงพูด

1. พิจารณาตำแหน่งเวลาต่างๆ ที่เกิดความไม่ต่อเนื่อง (discontinuity) ของรูปร่างและขนาดของสัญญาณตามที่เห็นใน spectrogram ตำแหน่งทางเวลาที่เกิดความไม่ต่อเนื่องนั้นๆ เรียกว่า ขอบเขต (boundary)

2. หา “segment” จากขอบเขตที่หาได้ในข้อ 1 และ พิจารณา class ของเสียง ที่สอดคล้องกับแต่ละ segment โดย segment หนึ่งๆ จะสอดคล้องกับหนึ่งในสิ่งต่อไปนี้: ความเงียบ (silence), สระ (vowels), พยัญชนะต่างๆ (consonants) ซึ่งได้แก่ พยัญชนะกัก (stop consonants), เสียงเสียดแทรก (fricatives), เสียงกึ่งเสียดแทรก (affricates), และ พยัญชนะนาสิก (nasal consonants) , หรือ เสียงกึ่งสระ (semi-vowels) โดยทั่วไปนั้น สัญญาณ 1 segment มักจะอยู่ระหว่างขอบเขต 2 ขอบเขตที่ติดกัน ยกเว้น segment ที่เกิดจากเสียงพยัญชนะกัก (stop consonant) และ เสียงกึ่งเสียดแทรก (affricate) โดยที่ segment ของเสียงทั้งสอง class นั้น มักจะมีขอบเขตอีก 1 ขอบเขตที่อยู่ระหว่าง segment นั้น ขอบเขตนี้สอดคล้องกับความไม่ต่อเนื่องซึ่งเกิดจากการปล่อยช่องปิด (closure release) ในกระบวนการกำเนิดเสียงทั้งสองประเภท การกำหนด class ของเสียงนั้น ให้พิจารณาจาก พลังงานของเสียงที่ปรากฏใน spectrogram ตลอดจนรูปร่าง และการเปลี่ยนแปลงของสัญญาณเสียงที่ความถี่ต่างๆ ข้อสังเกตในการกำหนด class ของเสียง (และความเงียบ) มีดังนี้
    a. ความเงียบ (ไม่ใช่เสียง แต่เป็นส่วนหนึ่งของทุกๆ สัญญาณเสียงพูด)
        i. ไม่มีพลังงานปรากฏให้เห็นใน spectrogram ตลอดช่วงความถี่ทั้งหมด (อาจจะมีพลังงานที่ความถี่ต่ำมาก อันเกิดมาจาก background noise ซึ่งจะคงที่ตลอดทั้งความยาวของสัญญาณ)
        ii. พลังงานที่หายไปซึ่งเกิดจากการสร้างช่องปิดเพื่อเปล่งเสียงกัก หรือ เสียงกึ่งเสียดแทรก ไม่ถือว่าเป็นส่วนหนึ่งของ segment ความเงียบ
    b. สระ
        i. มีพลังงานสูง มีการไหลของลมที่ทำให้เกิดเสียงมากที่สุด
        ii. เห็นโครงสร้างของ Formant ชัดเจน
        iii. เป็นสัญญาณที่มีลักษณะเป็นคาบ (periodic) อาจสังเกตจากรูปร่างของคลื่นเสียงทางเวลา
    c. พยัญชนะกัก (รูปที่ 2)
        i. พลังงานตลอดทั้งช่วงความถี่หายไป ซึ่งการที่พลังงานหายไปนี้เกิดจากการสร้างช่องปิด แต่อาจจะมีพลังงานที่ความถี่ต่ำอันเกิดจากการสั่นของเส้นเสียงในขณะที่เกิดช่องปิดในกรณีของ voiced stop consonant พลังงานที่ถูกส่งผ่านออกมาในอากาศนี้ ผ่านออกมาจากการแผ่รังสี (radiation) จากกระพุ้งแก้ม มิได้เกิดจากลมจากช่องปากโดยตรง
        ii. ในขณะที่ช่องปิดถูกปล่อยออกอย่างรวดเร็ว อาจเกิด noise สั้นๆ ซึ่งสังเกตได้จากพลังงานที่มีรูปร่าง ยาวออกไปทางแนวตั้ง ใน spectrogram หลังจากช่วงที่เป็นช่องปิด
        iii. อาจสังเกตเห็นสัญญาณที่มีลักษณะเป็น noise ซึ่งถูก filter โดย transfer function ของ filter ที่กำลังเปลี่ยนแปลงไปสู่ filter ของสระที่ตามพยัญชนะกักนั้นมา ถ้าช่องระหว่างเส้นเสียงเปิดกว้าง (spread glottis) ในขณะที่กลไกกำลังเลื่อนเข้าสู่การสร้างเสียงสระตัวถัดมานั้น
    d. เสียงเสียดแทรก (รูปที่ 3)
        i. ลักษณะของสัญญาณเป็น noise ที่เกิดจากกระแสลมถูกขับผ่านช่องแคบ พลังงานของ noise นี้จะหนาแน่นที่ช่วงความถี่ใดนั้นขึ้นอยู่กับตำแหน่งของช่องแคบที่ใช้ในการสร้างเสียงเสียดแทรกนั้นๆ
    e. เสียงกึ่งเสียดแทรก
        i. มีลักษณะใน spectrogram เช่นเดียวกับ ลักษณะที่เกิดจากการสร้างช่องปิด เพื่อเตรียมเปล่งเสียงพยัญชนะกัก (stop consonant) แล้วตามด้วยเสียงเสียดแทรก (fricative) หลังจากที่ช่องปิดถูกปล่อยออก
    f. พยัญชนะนาสิก (รูปที่ 4)
        i. พลังงานในความถี่กลางถึงสูงจะลดต่ำลงจากระดับพลังงานของสระที่อยู่ใกล้เคียงเล็กน้อย เนื่องจาก การสูญเสียพลังงานในโพรงจมูก

ii. เสียงนาสิกที่ตามหลังสระ จะมีลูกคลื่นในทางเวลามีขนาดต่ำลงกว่าลูกคลื่นของสระที่นำหน้าอยู่ และ มีการหายไปของส่วนประกอบความถี่สูง (ลูกคลื่นมีรูปร่างขรุขระน้อยลง) ถ้าพิจารณา spectrum ของสระในส่วนที่ติดกับพยัญชนะนาสิกซึ่งถูก nasalized นั้น จะพบว่า Bandwidth ของ F1 ของสระนั้นจะกว้างขึ้นกว่าสระในส่วนที่ไม่ถูก nasalized

g. <u>เสียงกึ่งสระ</u> (รูปที่ 5)

i. มีการเปลี่ยนแปลงของลักษณะโครงสร้าง formant ที่รวดเร็วและมากกว่าการเปลี่ยนแปลงที่เกิดในสระ เนื่องจากการสร้างช่องแคบที่แคบกว่า แต่มักจะไม่มีความไม่ต่อเนื่องของ formant ต่างๆ

ii. อาจมีการเปลี่ยนแปลงทางขนาดของสัญญาณที่กะทันหันเหมือนกับการเปลี่ยนแปลงที่เกิดในพยัญชนะ

3. หลังจากแบ่งเสียงเป็น segment ต่างๆ พร้อมทั้งกำหนด class ของเสียงแล้ว ผู้อ่าน spectrogram จะต้องคาดเดาตำแหน่งของลิ้น หรือ ช่องแคบช่องปิดต่างๆ ตามแต่ที่จะสอดคล้องกับ class ของเสียงนั้นๆ ถ้าเป็นพยัญชนะกัก, เสียงเสียดแทรก, หรือ เสียงกึ่งเสียดแทรก ผู้อ่านจะต้องคาดเดาการสั่นของเส้นเสียงด้วย (voiced หรือ unvoiced)

a. <u>สระ</u> ต้องหาตำแหน่งของลิ้น

i. วัดค่า F1 และ F2 ของสระนั้น แล้วนำค่าความถี่ formant ทั้งสองนั้นมาพิจารณาตำแหน่งของลิ้น (Font/Back, High/Mid/Low) ตาม vowel chart (รูปที่ 1.1-1.2)

ii. ถ้าโครงสร้างของ formant มีการเปลี่ยนแปลงจากสระหนึ่งไปอีกสระหนึ่ง สระนั้นเป็นสระประสม (diphthong)

b. <u>พยัญชนะกัก</u> ต้องหาตำแหน่งของการเปล่งเสียง (place of articulation) และ voicing

i. ถ้ามีการปล่อยช่องปิดที่ทำให้เกิด noise (release burst) ให้พิจารณารูปร่าง spectrum ของ release burst นั้น
รูปร่างดังกล่าวจะขึ้นอยู่กับเสียงที่อยู่รอบข้างเป็นอย่างมาก แต่โดยทั่วไปแล้วรูปร่างของ noise spectrum คร่าวๆ จะมีลักษณะดังนี้

1. Release burst ของ Labial stop consonant มักจะอ่อน มีพลังงานไม่มาก รูปร่างมักจะแบนในตลอดทุกความถี่ (ไม่มี peak ที่เด่นชัดในความถี่ช่วงใดช่วงหนึ่ง)
2. Release burst ของ Alveolar stop consonant มักจะมีพลังงานหนาแน่นบริเวณความถี่สูง
ถ้าเสียงรอบข้างมีตำแหน่งของลิ้นที่อยู่ค่อนไปทางด้านหลัง (เช่น back vowel) ความถี่ที่มีพลังงานหนาแน่นนี้ จะต่ำกว่าในกรณีที่เสียงรอบข้างมีตำแหน่งของลิ้นที่ค่อน ไปทางด้านหน้า (เช่น front vowel)
3. Release burst ของ Velar stop consonant มักจะมีพลังงานหนาแน่นบริเวณความถี่กลาง แต่อย่างไรก็ตามรูปร่างของ release burst ในกรณีนี้ จะเปลี่ยนแปลงไปตามเสียงรอบข้างอย่างมาก ถ้าเสียงรอบข้างมีตำแหน่งของลิ้น ที่อยู่ค่อนไปทางด้านหลัง (เช่น back vowel) ความถี่ที่มีพลังงานหนาแน่นนี้ อาจจะต่ำกว่าในกรณีของ alveolar stop consonant ในทางตรงกันข้าม
ถ้าเสียงรอบข้างมีตำแหน่งของลิ้นที่ค่อนไปทางด้านหน้า (เช่น front vowel) พลังงานอาจจะหนาแน่นในบริเวณความถี่ที่สูงกว่าของ alveolar

ii. ถ้าการปล่อยช่องปิดไม่ทำให้เกิด release burst การหาตำแหน่งของการเปล่งเสียง (place of articulation) จะต้องดูจากการเปลี่ยนแปลงของ formant ของสระที่อยู่ข้างหน้าหรือข้างหลังของเสียงกักนั้น หรือ แม้ในกรณีที่มี release burst ผู้อ่านก็ควรจะพิจารณาการเคลื่อนที่ของ formant ด้วย เพื่อเป็นข้อมูลที่ช่วยในการตัดสินใจ ควบคู่ไปกับรูปร่างของ release burst การพิจารณาโดยทั่วไปอาจใช้หลักดังนี้

1. การเคลื่อนที่ไปข้างหน้าของลิ้นจะทำให้ F2 เคลื่อนที่สูงขึ้น ในทางตรงกันข้าม F2 จะเคลื่อนที่ต่ำลง ถ้าลิ้นมีการเคลื่อนที่ไปข้างหลัง ในขณะที่การเคลื่อนที่ไปสูงขึ้นของลิ้น จะทำให้ F1 ต่ำลง และ F1 จะสูงขึ้นถ้าลิ้นลดต่ำลง
2. การลดพื้นที่หน้าตัดของช่องทางเสียงบริเวณริมฝีปาก ตัวอย่างเช่น การหุบริมฝีปาก เพื่อเตรียมเปล่งเสียง Labial stop consonant จะทำให้ F1 และ F2 เคลื่อนที่ต่ำลง (F1 และ F2 เคลื่อนที่สูงขึ้น เมื่อเคลื่อนจาก Labial stop consonant เข้าสู่สระ)
3. F2 และ F3 มักจะเคลื่อนที่เข้าชิดกัน ในขณะที่มีการสร้างช่องปิดที่เพดานแข็ง เพื่อเปล่งเสียง Alveolar stop consonant

iii. การสั่นของเส้นเสียงในกรณีของ voiced stop consonant อาจจะสามารถเห็นได้จากพลังงานความถี่ต่ำ ที่ปรากฏอยู่ในช่วงเวลาที่เป็นช่องปิด (voice bar) แต่การแบ่งแยก voiced และ unvoiced จาก voice bar นี้สามารถดูได้ยาก และไม่แน่นอน

iv. ค่า Voicing Onset Time (VOT) ซึ่งคือ ระยะเวลานับตั้งแต่ช่องปิดถูกปล่อย จนถึง เวลาที่เส้นเสียงเริ่มสั่นเพื่อเปล่งเสียงสระ สามารถบอกความเป็น voiced หรือ unvoiced ได้ในระดับหนึ่ง ค่า VOT ของ voiced stop consonant มักจะสั้นกว่า unvoiced stop consonant

v. การตัดสินใจความเป็น voiced หรือ unvoiced จากลักษณะของสัญญาณเพียงอย่างเดียวนั้น ทำได้ยาก ผู้อ่าน spectrogram ควรตัดสินใจ voicing จากบริบท (context) ด้วย

c. <u>เสียงเสียดแทรก</u> ต้องหาตำแหน่งของการเปล่งเสียง (place of articulation) และ voicing

i. ตำแหน่งของการเปล่งเสียงเสียดแทรก หรือ ตำแหน่งของช่องแคบที่เกิดขึ้นนั้น สามารถหาได้จากรูปร่างของ noise ที่ปรากฏบน spectrogram

1. labial และ dental fricatives จะมีรูปร่างของ noise spectrum ที่ค่อนข้างแบนในทุกความถี่ (ไม่มี peak ที่เด่นชัดในความถี่ช่วงใดช่วงหนึ่ง) พลังงานมักจะต่ำกว่า fricative ประเภทอื่น
2. alveolar fricatives มี noise spectrum ที่มีพลังงานสูง รูปร่างของ spectrum จะเป็นลักษณะที่มีพลังงานหนาแน่นอยู่ในช่วงความถี่สูง
3. palatal fricatives มี noise spectrum ที่มีพลังงานสูง รูปร่างของ spectrum จะเป็นลักษณะที่มีพลังงานหนาแน่นอยู่ในช่วงต่ำกว่า และกว้างกว่า noise spectrum ของ alveolar fricative

ii. การสั่นของเส้นเสียงในกรณีของ voiced fricative อาจจะสามารถเห็นได้จากพลังงานความถี่ต่ำ ที่ปรากฏอยู่ในช่วงเวลาเดียวกับที่เกิด noise แต่การแบ่งแยก voiced และ unvoiced จาก voice bar นี้สามารถดูได้ยาก และไม่แน่นอน เช่นเดียวกับ stop consonant

iii. voiced fricative มักจะสั้นกว่า unvoiced fricative

iv. การตัดสินใจความเป็น voiced หรือ unvoiced จากลักษณะของสัญญาณเพียงอย่างเดียวนั้น ทำได้ยาก ผู้อ่าน spectrogram ควรตัดสินใจ voicing จากบริบท (context) ด้วย

d. <u>เสียงกึ่งเสียดแทรก</u> ต้องหาตำแหน่งของการเปล่งเสียง (place of articulation) และ voicing ซึ่งทำได้เช่นเดียวกับ ในกรณีของพยัญชนะกัก และเสียงเสียดแทรก

e. <u>พยัญชนะนาสิก</u> ต้องหาตำแหน่งของการเปล่งเสียง (place of articulation) ซึ่งสามารถพิจารณาได้ จากการเคลื่อนที่ของ formant ในสระที่ติดกัน

f. <u>เสียงกึ่งสระ</u> ได้แก่ เสียง /l/, /w/, /j / ,/ ɻ /, /r/, /h/ การแบ่งแยกระหว่างเสียงทั้ง 6 ทำได้โดยพิจารณา การเปลี่ยนแปลงของตำแหน่งและขนาดของ F1, F2, และ F3

i. /l/ มี F1 ที่ต่ำ, F2 ที่ต่ำถึงปานกลาง มีการเปลี่ยนแปลงกะทันหันของขนาดของ spectrum ในช่วง F2 จาก /l/ สู่สระ

ii. /w/ มี F1 ที่ต่ำ, F2 ที่ต่ำมาก ขนาดของ formant ต่างๆ ไม่เปลี่ยนแปลงจาก ขนาดของ formant ในสระมาก

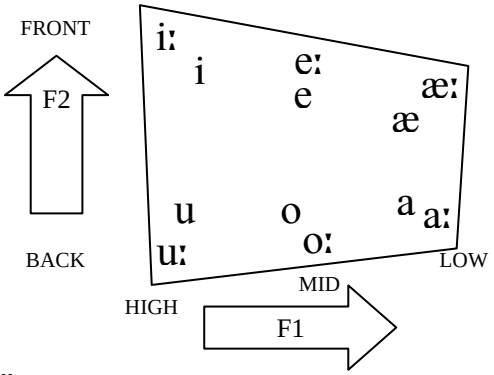
iii. /j/ มี F1 ที่ต่ำ, แต่ F2 และ F3 สูง ขนาดของ formant ต่างๆ ไม่เปลี่ยนแปลงจาก ขนาดของ formant ในสระมาก

iv. /ɻ / ข้อสังเกตที่เด่นชัดมาก คือ F3 ต่ำ

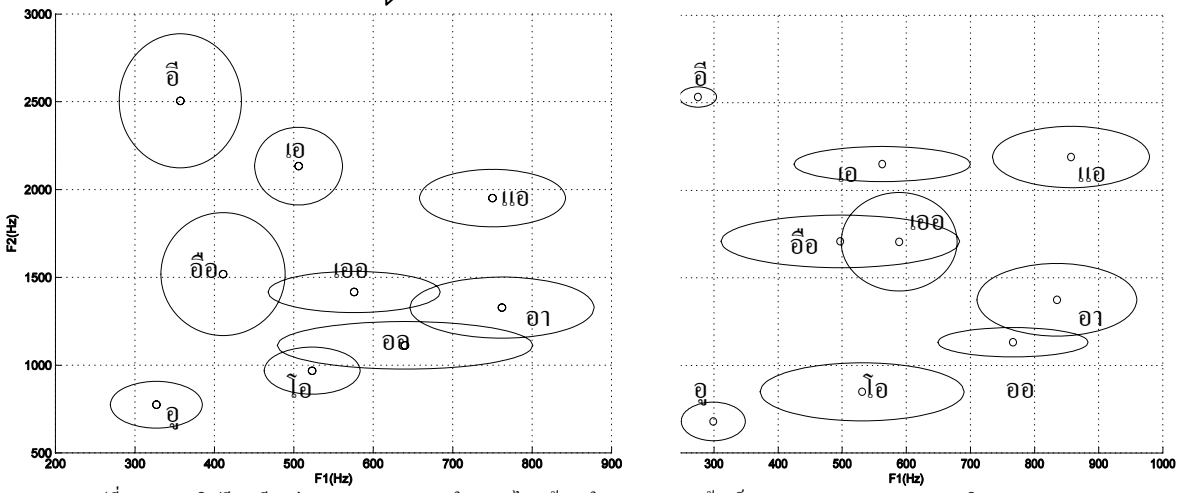
v. /r/ มี F3 ที่ต่ำ เช่นเดียวกับ /ɻ/ แต่มีหลักฐานของการรัวลิ้นเกิดขึ้น (การขาดช่วงของพลังงาน) ในช่วงที่ F3 ลดต่ำลง

vi. /h/ สามารถสังเกตได้จากลักษณะของสัญญาณที่เกิดจาก noise ถูก filter ด้วย transfer function ของ vocal tract ที่มีลักษณะเป็นโครงสร้าง formant คล้ายกับสระ noise ดังกล่าวอาจจะไม่ได้มีพลังงานสูง

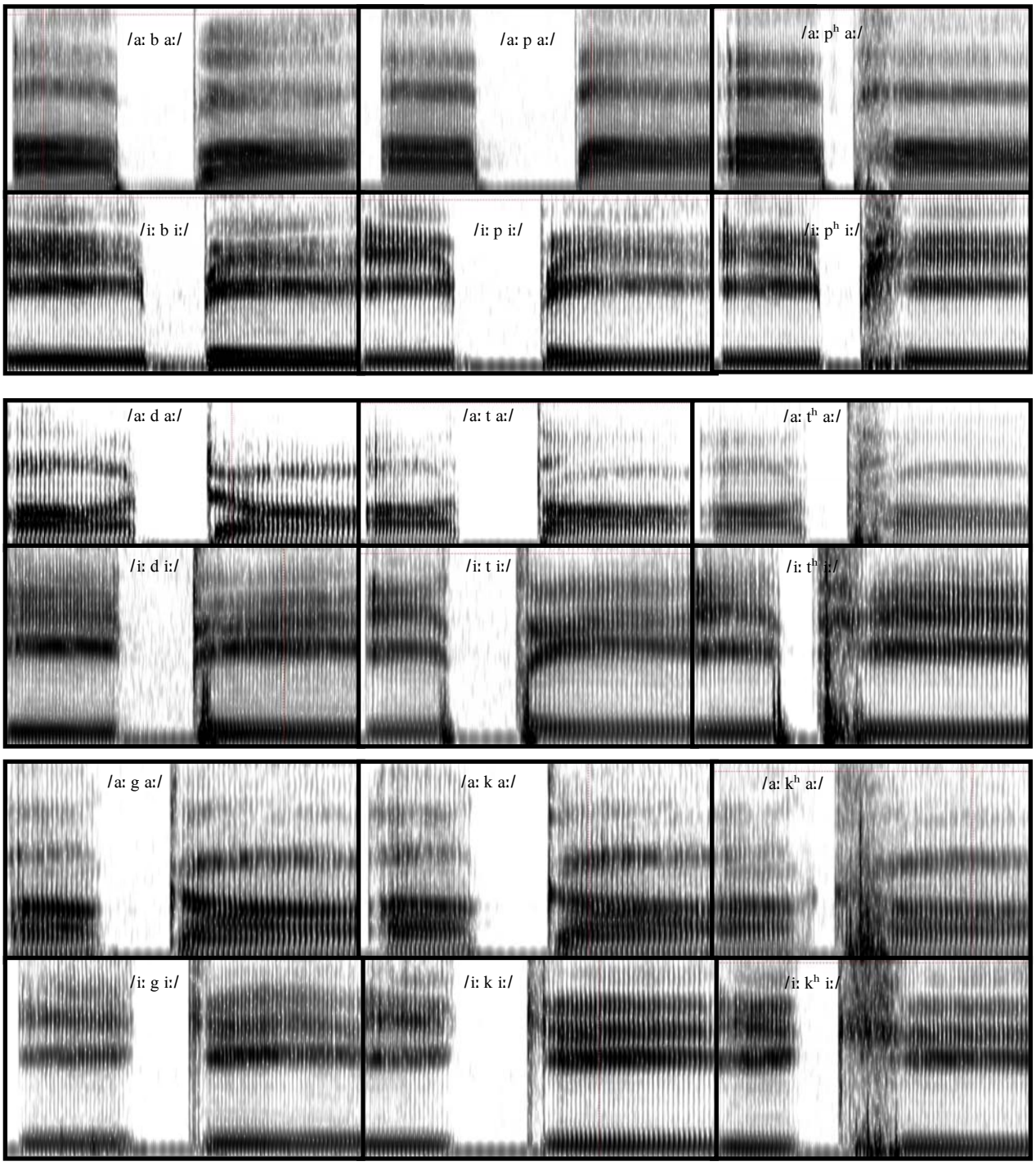
4. ขั้นตอนสุดท้ายในการอ่าน spectrogram ของเสียงพูด คือ การรวมเสียงต่างๆ ที่คาดคะเนไว้ในขั้นตอนต่างๆ ข้างต้น ให้เป็นคำ การรวมเสียงต่างๆ เป็นคำจะต้องใช้ความหมายของประโยคเข้ามาช่วยในการพิจารณา การใช้ความหมาย (semantic) เข้ามาช่วยจะทำให้เราตัดสินใจเกี่ยวกับเสียงบางเสียงที่ยังไม่แน่ใจจากการดูแค่ลักษณะทางสัญญาณได้ดีขึ้น ในภาษาไทยนั้น นอกจาก ลักษณะการออกเสียง (manner of articulation), ตำแหน่งของการเปล่งเสียง (place of articulation), และลักษณะของ voicing แล้ว การที่จะระบุเสียงได้อย่างสมบูรณ์นั้นต้องอาศัยข้อมูลของ เสียงวรรณยุกต์ (tone) ด้วย แต่อย่างไรก็ตาม การหาเสียงวรรณยุกต์จากสัญญาณนั้น ไม่สามารถทำได้สะดวกจากการอ่าน spectrogram เนื่องจากเสียงวรรณยุกต์นั้นมีความสัมพันธ์โดยตรงกับความถี่มูลฐาน (Fundamental Frequency หรือ pitch frequency) ซึ่งมองเห็นได้ยากใน spectrogram ดังนั้นการที่จะอ่าน spectrogram ภาษาไทยนั้น ส่วนของเสียงวรรณยุกต์มักจะไม่มีการคาดเดาโดยตรงจากสัญญาณ แต่จะใช้ความหมายของคำเข้ามาช่วย หรือ วัดความถี่มูลฐานจากรูปคลื่นเสียงในทางเวลา



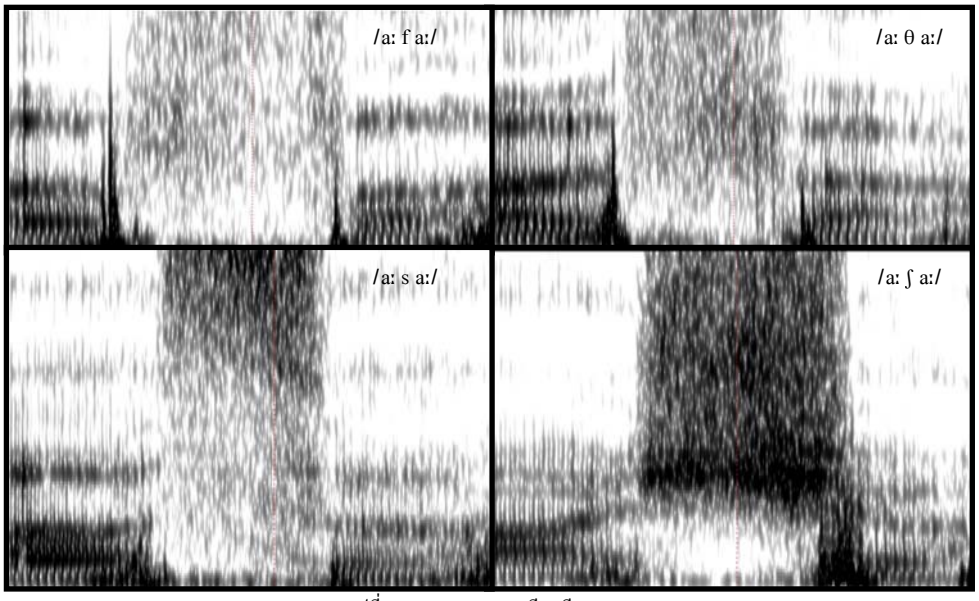
รูปที่ 1.1 แผนภูมิเปรียบเทียบค่า F1 และ F2 ของ tense-lax cardinal vowels



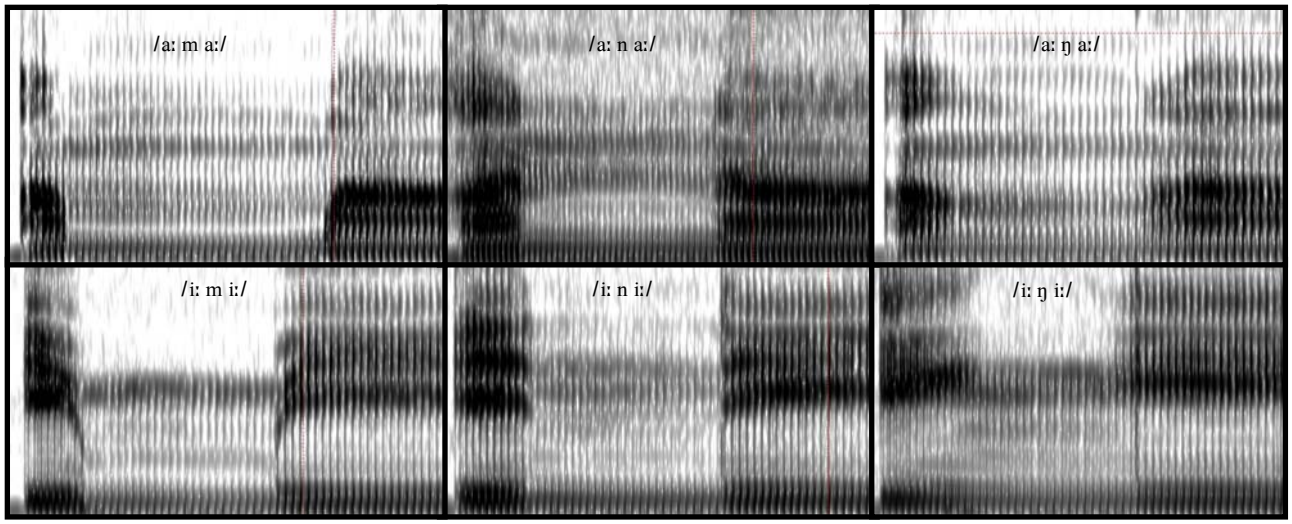
รูปที่ 1.2 แผนภูมิเปรียบเทียบค่า F1 และ F2 ของสระในภาษาไทย ข้อมูลในแผนภาพทางซ้ายเก็บจากชาย 19 คน ทางขวาจากหญิง 5 คน



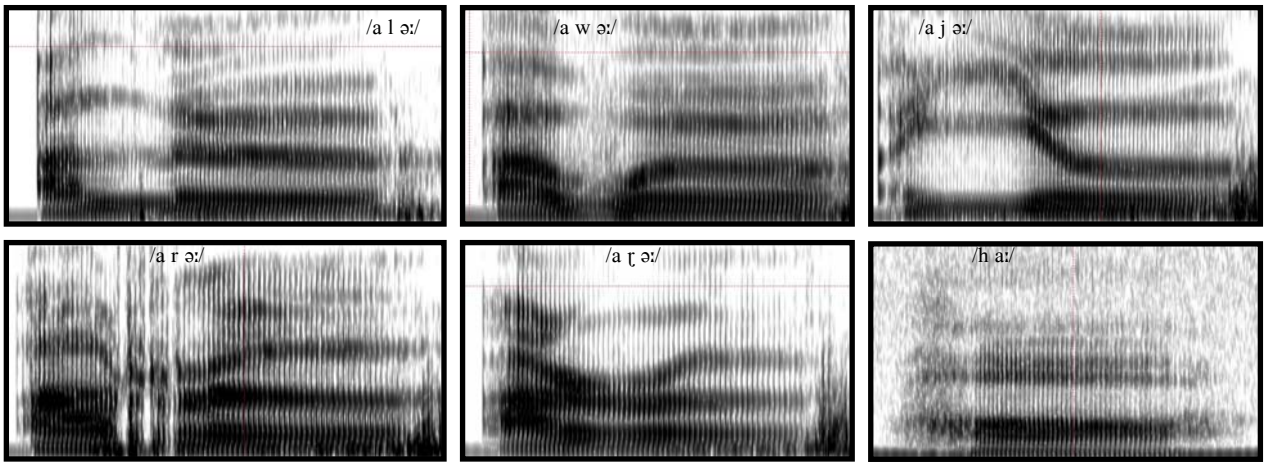
รูปที่ 2 Spectrogram ของเสียงพยัญชนะกัก



รูปที่ 3 Spectrogram ของเสียงเสียดแทรก



รูปที่ 4 Spectrogram ของเสียงพยัญชนะนาสิกในบริบทของ front vowel และ back vowel



รูปที่ 4 Spectrogram ของเสียงกึ่งสระ