# บทที่ ๒
# วรรณกรรมและงานวิจัยที่เกี่ยวข้อง

## ๑. เอกสารที่เกี่ยวข้อง

ในส่วนนี้จะกล่าวถึงเสียงและอวัยวะที่ใช้ในการเปล่งเสียง ระบบหน่วยเสียงในภาษาไทย ภาพรวมของระบบรู้จำเสียงพูด แบบจำลองฮิดเดนมาร์คอฟ

### ๑.๑ เสียงและอวัยวะที่ใช้ในการเปล่งเสียง

การศึกษาเรื่องเสียงจำเป็นต้องเรียนรู้เกี่ยวกับลักษณะหน้าที่และการทำงานของอวัยวะต่าง ๆ ที่ใช้ในการเปล่งเสียง เนื่องจากความรู้ดังกล่าวจะช่วยให้เข้าใจกระบวนการของการเปล่งเสียง อวัยวะที่ใช้ในการเปล่งเสียงแบ่งออกเป็น ๓ กลุ่ม ได้แก่

๑.ปอดและหลอดลม ปอดทั้ง ๒ ข้างประกอบด้วยถุงลมขนาดเล็กจำนวนมาก ปอดมีลักษณะเหมือนฟองน้ำ ไม่มีกล้ามเนื้อประกอบอยู่ ภายในปอดมีหลอดลมฝอยอยู่มากมาย หลอดลมฝอยนี้จะมารวมกันเป็นหลอดลมแขนง และหลอดลมแขนงของปอดทั้ง ๒ ข้างจะมารวมกันเป็นหลอดลม

๒. กล่องเสียงและเส้นเสียง กล่องเสียงประกอบขึ้นด้วยจมูก กระดูกอ่อน และกล้ามเนื้อหลายชิ้น ส่วนหนึ่งของกล่องเสียงคือ สิ่งที่เรียกว่าลูกกระเดือกซึ่งเห็นได้ชัดเจนในเพศชาย กล่องเสียงอยู่เหนือหลอดลม ด้านล่างของกล่องเสียงประกอบด้วยกระดูกอ่อนชื่อ ไครคอยด์ (cricoid) เป็นกระดูกอ่อนรูปคล้ายแหวน ด้านกว้างหันมาทางด้านหลังของร่างกาย ทางด้านหน้าของกล่องเสียงประกอบด้วย กระดูกอ่อนชื่อ ไทรอยด์ (thyroid) เป็นกระดูกอ่อนแผ่นเดียวรูปโค้งครึ่งวงกลม กระดูกชิ้นนี้เป็นส่วนหนึ่งของกล่องเสียงที่สัมผัสได้และเห็นได้บริเวณคอ นอกจากนี้กล่องเสียงยังประกอบด้วยกระดูกอ่อนรูปปิรามิดอีก ๑ คู่ ชื่อ อาริตินอยด์ (arytenoid) กระดูกอ่อนคู่นี้อยู่ด้านหลังของกล่องเสียงบนกระดูกอ่อนรูปวงแหวน ส่วนประกอบที่สำคัญที่สุดของกล่องเสียงได้แก่ เส้นเสียง ซึ่งมีอยู่ ๒ เส้น มีลักษณะเป็นแผ่น ประกอบด้วยเส้นเอ็นและกล้ามเนื้อ เส้นเสียงทั้งคู่พาดตัวตามแนวนอนอยู่ตรงกลางของกล่องเสียง ปลายด้านหนึ่งของเส้นเสียงทั้ง ๒ เส้นอยู่ติดกันและเชื่อมอยู่กับกระดูกอ่อนไทรอยด์ทางด้านหน้าของกล่องเสียง ปลายอีกด้านหนึ่งของเส้นเสียงแต่ละเส้นเชื่อมอยู่กับกระดูกอ่อนอาริตินอยด์แต่ละอัน

๓. อวัยวะที่อยู่เหนือเส้นเสียง ลมจากปอดที่ผ่านเส้นเสียงแล้วจะผ่านขึ้นมายังช่องคอ และจะออกสู่ภายนอกทางช่องปาก หรือ ช่องจมูก ช่องใดช่องหนึ่งหรือทั้ง ๒ ช่องพร้อม ๆ กัน

อวัยวะที่อยู่เหนือเส้นเสียงมีองค์ประกอบของอวัยวะต่าง ๆ ซึ่งอาจจำแนกตามหน้าที่ในการช่วยให้เกิดเสียงเป็น ๒ ประเภท คือ

๑. อวัยวะที่เป็นตำแหน่งที่เกิดเสียงต่าง ๆ (passive articulator) เรียกอีกชื่อหนึ่งว่า ฐานที่เกิดเสียง เป็นอวัยวะที่ไม่เคลื่อนที่ในขณะที่ออกเสียง และอยู่ทางด้านบนของช่องปาก อวัยวะเหล่านี้ได้แก่ ริมฝีปากบน ฟันบน ปุ่มเหงือก เพดานแข็ง เพดานอ่อน และ ลิ้นไก่

๒. อวัยวะที่เป็นส่วนกระทำการ (active articulator) เรียกอีกชื่อหนึ่งว่า กรณ์ เป็นอวัยวะที่เคลื่อนที่ขณะออกเสียง อยู่ด้านล่างของช่องปาก ได้แก่ ริมฝีปากล่าง ฟันล่าง ปลายลิ้น ลิ้นส่วนปลาย ลิ้นส่วนหน้าและลิ้นส่วนหลัง

อวัยวะในแต่ละส่วนมีผลต่อการเปล่งเสียงดังนี้

- ริมฝีปากบนและริมฝีปากล่าง

เป็นอวัยวะที่เคลื่อนไหวมาก ริมฝีปากจะดัดแปลงลมให้เป็นเสียงต่าง ๆ กัน โดยปิดริมฝีปาก เปิดเล็กน้อย เปิดกว้าง รูปปากแผ่ รูปปากห่อ ฯลฯ

- ฟัน

ช่วยกักลมทำให้เกิดเสียงหลายประเภท เช่น เมื่อฟันบนกดลงบนฟันล่างและปล่อยให้ลมเสียดแทรกช่องฟันออกมาเป็นเสียง ฟ[f] หรือใช้ลิ้นส่วนปลายแตะหลังฟันบน ทำให้เกิดเสียงต่าง ๆ เช่น ด[d] ต[t] ฯลฯ

- ปุ่มเหงือก

คือส่วนนูนออกมาจากแผ่นเพดานบน อยู่หลังฟันบน เป็นส่วนสำคัญในการเปล่งเสียงพยัญชนะอีกส่วนหนึ่ง เช่น ร[r] ล[l] ฯลฯ

- เพดานแข็ง

คือส่วนโค้งที่เป็นกระดูกแข็งอยู่ถัดจากปุ่มเหงือกเข้าไป เป็นบริเวณที่เกิดเสียงพยัญชนะอีกส่วนหนึ่ง เช่น จ[c] ช[ch] ฯลฯ

- เพดานอ่อน

คือ กระดูกที่อยู่ต่อกับเพดานแข็ง สามารถเคลื่อนขยับขึ้นลงได้เล็กน้อย เป็นบริเวณที่เกิดเสียงพยัญชนะอีกส่วนหนึ่ง เช่น ก[k] ค[kh] ฯลฯ เพดานอ่อน เป็นส่วนที่เคลื่อนไหวได้เล็กน้อย โดยอาจยกขึ้นไปติดกับผนังคอทำให้ลมจากปอดเคลื่อนไปสู่ช่องจมูกไม่ได้ หรืออาจลดตัวลงปล่อยให้ลมจากปอดเข้าสู่ช่องจมูกหากเพดานอ่อนและลิ้นไก่เคลื่อนไปติดกับผนังคอ ลมจากปอดจะผ่านไปทางช่องจมูกไม่ได้ ต้องผ่านทางช่องปากเท่านั้น เสียงที่เกิดขึ้นก็เป็นเสียงที่เกิดจากช่องปาก แต่ถ้าเพดานอ่อนและลิ้นไก่อยู่ในลักษณะพักธรรมดา ลมจากปอดจะผ่านไปทางช่องจมูกได้ เสียงที่เกิดขึ้นจะมี ๒ ประเภท คือ

๑.ถ้ามีการกักปิดลม ณ ที่ใดที่หนึ่งในช่องปาก ลมจะผ่านออกได้เฉพาะทางจมูก เสียงที่เกิดขึ้นเป็นเสียงทางจมูก (nasal sounds) ซึ่งเกิดขึ้นได้ที่เสียงท้ายพยางค์ เช่นคำว่า ลม เสียง “ม” จะปรากฏที่พยางค์ท้าย โดยลักษณะของเพดานอ่อนและลิ้นไก่จะลดระดับลงมาปิดกั้นไม่ให้ลมผ่านออกทางปาก

๒.ถ้าไม่มีการปิดกักลมในช่องปาก ลมจะผ่านออกได้ทั้งทางปากและจมูก เสียงที่เกิดขึ้นเป็น เสียงขึ้นจมูก ซึ่งจะเกิดขึ้นกับเสียงสระที่ชิดกับเสียงทางจมูกและในบางภาษาจะมีสระทั้งที่เป็นเสียงสระขึ้นจมูก และเสียงสระที่ออกทางช่องปาก โดยการเปล่งเสียง ริมฝีปากทั้งคู่อยู่ชิดกันกักลมไว้ ในขณะเดียวกันเพดานอ่อนและลิ้นไก่ลดระดับลงอยู่ในลักษณะพัก ทำให้ลมผ่านออกทางช่องจมูกเมื่อเปิดที่กักลมที่ริมฝีปาก ลมจะออกทางปากด้วย

- ลิ้นไก่

เป็นก้อนเนื้อเล็ก ๆ อยู่ปลายเพดานอ่อนตรงกลางช่องปาก สามารถขยับขึ้นลงได้ เป็นอวัยวะที่สำคัญที่ช่วยในการเปล่งเสียงทางจมูกและเสียงที่ไม่ใช่เสียงทางจมูก

- ลิ้น

เป็นอวัยวะที่เคลื่อนไหวได้มากในการเปล่งเสียงพูด และเคลื่อนไหวได้อย่างคล่องแคล่วในหลายทิศทาง ทั้งกระดกขึ้น กระดกลง แลบยื่นไปข้างหน้า ดึงไปข้างหลัง นอกจากนี้ปลายลิ้นยังเคลื่อนตัวได้มากกว่าส่วนอื่น อาจกระดกขึ้นไปในทิศทางของฟันบนและปุ่มเหงือก หรืองอปลายลิ้นกลับเข้าไปด้านในและด้านล่างของลิ้น กระดกขึ้นไปในทิศทางของฟันบน ปุ่มเหงือก เพดานแข็ง โดยลิ้นแบ่งออกเป็น ๕ ส่วนตามตำแหน่งที่ช่วยให้เกิดเสียงต่าง ๆ กัน ได้แก่

๑.ปลายลิ้น คือ ส่วนปลายสุดของลิ้น สามารถยกขึ้นไปแตะอวัยวะต่าง ๆ ในปากส่วนบนได้ง่าย ช่วยให้เกิดเสียงพยัญชนะ ร[r] ล[l] ฯลฯ

๒. ลิ้นส่วนปลาย คือ ส่วนที่ติดกับปลายลิ้น สามารถขึ้นไปแตะอวัยวะในปากส่วนบนได้ ช่วยให้เกิดเสียงพยัญชนะ ต[t] ด[d] ฯลฯ

๓.ลิ้นส่วนหน้า คือ ส่วนที่อยู่ตรงกับเพดานแข็ง เมื่อวางลิ้นราบ ช่วยให้เกิดเสียงพยัญชนะ จ[c] ช[ch] และสระหน้า ฯลฯ

๔.ลิ้นส่วนหลัง คือ ส่วนของลิ้นที่อยู่ตรงข้ามกับเพดานอ่อน เมื่อวางลิ้นราบ ช่วยในการเกิดเสียงพยัญชนะ ก[k] ค[kh] และสระหลัง ฯลฯ

๕.โคนลิ้น คือ ส่วนที่อยู่ต่อจากลิ้นส่วนหลังลงไปในลำคอ

- แผ่นเนื้อปากหลอดลม

เป็นก้อนเนื้อเล็ก ๆ คล้ายลิ้นไก่อยู่ต่อจากโคนลิ้นลงไปในช่องคอ มีหน้าที่ปิดช่องลมเมื่อกลืนอาหารและเปิดช่องลมเมื่อพูด

- กรวยคอ

คือ โพรงคอที่อยู่ถัดจากช่องปากลงไปถึงเส้นเสียง

- เส้นเสียง

**๑.๒ ระบบหน่วยเสียงในภาษาไทย**

หน่วยเสียงในภาษาไทยจะแบ่งออกเป็น ๓ กลุ่มคือ กลุ่มหน่วยเสียงพยัญชนะ กลุ่มหน่วยเสียงสระ และกลุ่มหน่วยเสียงวรรณยุกต์ โดยมีรายละเอียดของแต่ละชนิดดังนี้

๑.๒.๑ กลุ่มหน่วยเสียงพยัญชนะ

หน่วยเสียงพยัญชนะประกอบด้วยหน่วยเสียงพยัญชนะต้น หน่วยเสียงควบกล้ำ และ หน่วยเสียงพยัญชนะท้าย [๑]

เสียงพยัญชนะต้นในภาษาไทยแบ่งได้ ๒๑ หน่วยเสียง ดังตารางที่ ๑

ตารางที่ ๑ หน่วยเสียงพยัญชนะต้น

| เสียงที่ | สัทอักษร | พยัญชนะ |
|---|---|---|
| ๑ | p | ป |
| ๒ | t | ต ฏ |
| ๓ | c | จ |
| ๔ | k | ก |
| ๕ | z หรือ ? | อ |
| ๖ | ph | พ ภ ผ |
| ๗ | th | ท ธ ถ ฐ ฑ ฒ |
| ๘ | ch | ช ฌ ฉ |
| ๙ | kh | ข ค ฆ |
| ๑๐ | b | บ |
| ๑๑ | d | ด ฎ ฑ |
| ๑๒ | m | ม หม |
| ๑๓ | n | น ณ หน |
| ๑๔ | ng | ง หง |
| ๑๕ | f | ฟ ฝ |
| ๑๖ | s | ส ซ ศ ษ |
| ๑๗ | h | ห ฮ |
| ๑๘ | r | ร หร |
| ๑๙ | l | ล ฬ หล |
| ๒๐ | w | ว หว |
| ๒๑ | j | ย ญ หญ หย |

เสียงพยัญชนะควบกล้ำในภาษาไทยแบ่งได้ ๑๒ หน่วยเสียง ดังตารางที่ ๒

ตารางที่ ๒ หน่วยเสียงพยัญชนะควบกล้ำ

| เสียงที่ | สัทอักษรอย่างง่าย | พยัญชนะ |
|---|---|---|
| ๑ | kr | กร |
| ๒ | kl | กล |
| ๓ | kw | กว |
| ๔ | khr | คร |
| ๕ | khl | คล |
| ๖ | khw | คว |
| ๗ | pr | ปร |
| ๘ | pl | ปล |
| ๙ | phr | พร |
| ๑๐ | phl | พล |
| ๑๑ | tr | ตร |
| ๑๒ | thr | ทร |

เสียงพยัญชนะท้ายหรือตัวสะกดในภาษาไทยแบ่งได้ ๘ หน่วยเสียง ดังตารางที่ ๓

ตารางที่ ๓ หน่วยเสียงพยัญชนะท้าย

| เสียงที่ | สัทอักษร | พยัญชนะท้าย |
|---|---|---|
| ๑ | k^ | แม่กก |
| ๒ | t^ | แม่กด |
| ๓ | p^ | แม่กบ |
| ๔ | n^ | แม่กน |
| ๕ | ng^ | แม่กง |
| ๖ | m^ | แม่กม |
| ๗ | j^ | แม่เกย |
| ๘ | w^ | แม่เกยว |

ซึ่งแต่ละหน่วยเสียงของพยัญชนะในภาษาไทยมีลักษณะของเสียง (Manner) และตำแหน่งที่เกิดของเสียง (Place) ที่แตกต่างกัน เช่นมีลักษณะเป็นเสียงแบบไม่ก้องไม่พ่นลม (voiceless aspirated stop) และตำแหน่งที่เกิดเสียงเป็นริมฝีปาก (bilabial) ดังตารางที่ ๔

ตารางที่ ๔ ลักษณะของเสียงและตำแหน่งที่เกิดของเสียงของพยัญชนะต้นภาษาไทย

| ลักษณะของเสียง /ตำแหน่งที่เกิดของเสียง | | | ริมฝีปาก | ปุ่มเหงือก | เพดานแข็ง | เพดานอ่อน | เส้นเสียง |
|---|---|---|---|---|---|---|---|
| เสียงกัก | ก้อง | | b | d | c | k | z |
| | ไม่ก้อง | มีลม | ph | th | ch | kh | |
| | ไม่ก้อง | ไม่มีลม | p | t | | | |
| ไม่กัก | นาสิก | | m | n | | ng | |
| | เสียดแทรก | | f | s | | | h |
| | ข้างลิ้น | | | l | | | |
| | รัว | | | r | | | |
| | กึ่งสระ | | w | | j | | |

๑.๒.๒ กลุ่มหน่วยเสียงสระ

หน่วยเสียงสระในภาษาไทยประกอบด้วยเสียงสระแท้จำนวน ๑๘ หน่วยเสียง และสระประสมจำนวน ๖ หน่วยเสียง โดยแบ่งเป็นสระแท้เสียงยาวจำนวน ๙ หน่วยเสียง สระแท้เสียงสั้นจำนวน ๙ หน่วยเสียง ดังตารางที่ ๕

ตารางที่ ๕ หน่วยเสียงสระแท้

| เสียงที่ | สัทอักษร | สระแท้ |
|---|---|---|
| ๑ | i | อิ |
| ๒ | ii | อี |
| ๓ | e | เอะ |
| ๔ | ee | เอ |
| ๕ | x | แอะ |
| ๖ | xx | แอ |
| ๗ | v | อึ |
| ๘ | vv | อือ |
| ๙ | q | เออะ |
| ๑๐ | qq | เออ |
| ๑๑ | a | อะ |
| ๑๒ | aa | อา |
| ๑๓ | u | อุ |
| ๑๔ | uu | อู |
| ๑๕ | o | โอะ |
| ๑๖ | oo | โอ |
| ๑๗ | @ | เอาะ |
| ๑๘ | @@ | ออ |

ส่วนสระประสมแบ่งเป็นสระประสมเสียงยาวจำนวน ๓ หน่วยเสียง และสระประสมเสียงสั้นจำนวน ๓ หน่วยเสียง ดังตารางที่ ๖

ตารางที่ ๖ หน่วยเสียงสระประสม

| เสียงที่ | สัทอักษร | สระประสม | หน่วยเสียงที่ประสม |
|---|---|---|---|
| ๑ | @@ | ออ | สระอี+ สระอา |
| ๒ | ia | เอียะ | สระอู+ สระอา |
| ๓ | iia | เอีย | สระอือ + สระอา |
| ๔ | va | เอือะ | สระอี+ สระอะ |
| ๕ | vva | เอือ | สระอู+ สระอะ |
| ๖ | ua | อัวะ | สระอือ + สระอะ |

ซึ่งแต่ละหน่วยเสียงสระจะเกิดจากส่วนของลิ้น ระดับของลิ้น และลักษณะของริมฝีปากที่แตกต่างกัน เช่น หน่วยเสียงสระอี จะเกิดจากลิ้นส่วนหน้า อยู่ในระดับสูงคือ ลิ้นส่วนหน้ายกสูงเกือบจรดเพดานแข็ง ริมฝีปากไม่ห่อกลม ดังตารางที่ ๗

ตารางที่ ๗ ส่วนของลิ้นและระดับลิ้นในการเปล่งเสียงสระ

| *ตำแหน่งของระดับลิ้น* | ลักษณะริมฝีปาก | | |
|---|---|---|---|
| | ริมฝีปากไม่ห่อกลม | | ริมฝีปากห่อกลม |
| | *ลิ้นส่วนหน้า* | *กลาง* | *หลัง* |
| ระดับสูง | อิ อี | อึ อือ | อุ อู |
| ระดับกลาง | เอะ เอ | เออะ เออ | โอะ โอ |
| ระดับต่ำ | แอะ แอ | อะ อา | เอาะ ออ |

๑.๒.๓ กลุ่มหน่วยเสียงวรรณยุกต์

หน่วยเสียงวรรณยุกต์ในภาษาไทยมีจำนวน ๕ หน่วยเสียง ดังแสดงในตารางที่ ๘

ตารางที่ ๘ หน่วยเสียงวรรณยุกต์

| เสียงที่ | สัทอักษร | วรรณยุกต์ |
|---|---|---|
| ๑ | ๐ | สามัญ |
| ๒ | ๑ | เอก |
| ๓ | ๒ | โท |
| ๔ | ๓ | ตรี |
| ๕ | ๔ | จัตวา |

**๑.๓ ภาพรวมของระบบรู้จำเสียงพูด**

ระบบรู้จำเสียงพูด [๒] ส่วนใหญ่ จะเกี่ยวข้องกับขั้นตอนต่างๆ ดังนี้

๑.๓.๑ การดักจับสัญญาณเสียงพูด (Utterance Detection)

การดักจับสัญญาณเสียงพูด โดยทั่วไปจะทำการอัดเก็บเสียงในรูปแบบของสัญญาณดิจิตอล ด้วยอุปกรณ์ไมโครโฟนที่ต่อกับคอมพิวเตอร์ผ่านทางการ์ดเสียง และทำการเก็บเสียงในรูปแบบของไฟล์ WAV ซึ่งจุดเริ่มต้นของเสียงสามารถดักจับได้โดยการเปรียบเทียบระดับเสียงที่บุคคลสามารถฟังได้ (ที่ความถี่ ๑๕ – ๒๐,๐๐๐ Hz.) กับตัวอย่างที่ได้เก็บไว้ แต่ในส่วนของจุดสิ้นสุดของเสียงอาจจะจับได้ยากกว่า เพราะผู้พูดอาจจะจบการพูดด้วยเสียงหายใจ ถอนหายใจ เสียงแหลมของผู้หญิง เสียงขบฟันที่ดังมาก หรือเสียงอาจจะก้องกังวานออกไป ซึ่งฐานข้อมูลที่ใช้ในการจัดเก็บข้อมูลสัญญาณเสียงนั้น จะแบ่งได้เป็น ๒ กลุ่มคือ

- เสียงพูดที่จัดเก็บในสภาพแวดล้อมการทำงานทั่วไป
- เสียงพูดที่จัดเก็บผ่านสายโทรศัพท์ ทั้งนี้เพราะมีงานประยุกต์ใช้หลายอย่างที่รับอินพุตจาก สายโทรศัพท์ ซึ่งมักเกี่ยวเนื่องกับงานธนาคาร งานเครดิตการ์ด

๑.๓.๒ การกรองความถี่เบื้องต้น (Pre-Filtering)

การกรองความถี่สามารถทำได้หลายวิธี ขึ้นอยู่กับการทำงานของระบบรู้จำนั้นๆ ตัวอย่างเช่น ใช้ตัวกรองแบบดิจิตอลชนิดผ่านความถี่ (Cut-off frequency) เป็นการกำหนดค่าความถี่ เช่น ที่ ๒๐๐ Hz. เพื่อป้องกันสัญญาณรบกวนความถี่ต่ำที่เกิดจากแหล่งกำเนิดไฟฟ้า ซึ่งในขั้นตอนนี้อาจรวมไปถึงการตัดหัวและท้ายของสัญญาณเสียง (End-point Detection) หรือการกรองสัญญาณรบกวนทางความถี่(Frequency Filtering) และการปรับความยาวเสียง (Time Normalization) ซึ่งการปรับความยาวเสียงนั้น แม้ว่าจะมีผลให้สัญญาณเสียงผิดเพี้ยนไปจากเดิม แต่ก็อาจจะหลีกเลี่ยงไม่ได้ในบางระบบ เช่น ถ้าเลือกโครงข่ายประสาทเทียมแบบการเรียนรู้แพร่กระจายย้อนหลัง (Back-Propagation Learning Algorithm) อินพุตที่ใส่เข้ามาทุกตัวต้องมีขนาดเท่าเทียมกัน ซึ่งต้องทำโดยการปรับความยาวของเสียงก่อนการสกัดค่าลักษณะสำคัญ

๑.๓.๓ การแบ่งเฟรม (Framing / Windowing)

เป็นการแบ่งข้อมูลสัญญาณเสียงให้อยู่ในขนาดที่กำหนดไว้ (ส่วนใหญ่ใช้ที่ขนาดความยาวประมาณ ๑๐ – ๔๐ มิลลิวินาที) เพื่อนำไปใช้ในระบบรู้จำเสียงพูด อีกทั้งยังเป็นการจัดเตรียมขอบเขตสำหรับการวิเคราะห์ ซึ่งระบบส่วนใหญ่จะรวมขั้นตอนนี้ไว้กับ ขั้นตอนการกรองความถี่เบื้องต้น

๑.๓.๔ การสกัดค่าคุณลักษณะเด่น (Feature Extraction)

การสกัดค่าคุณลักษณะเด่นเป็นการวิเคราะห์หาค่าที่จะใช้แทนสัญญาณเสียง โดยการเน้นสัญญาณเบื้องต้น (Preemphasis) ขั้นตอนนี้เป็นการเตรียมความพร้อมขั้นสุดท้ายของแต่ละเฟรมก่อนที่จะนำไปสู่ขั้นตอนการเปรียบเทียบและจับคู่ โดยในขั้นตอนนี้สามารถแบ่งได้เป็น ๓ กลุ่มหลัก ได้แก่

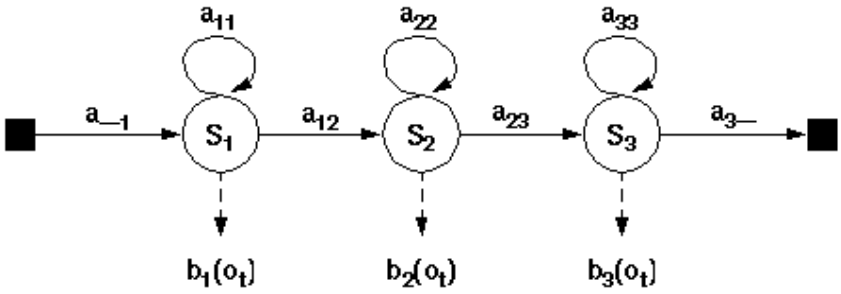
(๑) กลุ่มค่าคุณลักษณะเด่นระดับสูง (High level feature) ได้แก่ สำเนียงการพูด รูปแบบในการพูด และความเร็วในการพูด เป็นต้น

(๒) กลุ่มค่าคุณลักษณะเด่นทางฉันทลักษณ์ (Prosodic feature) เช่น ค่าความถี่มูลฐาน (Fundamental frequency) ความถี่ฟอร์แมนท์ (Formant frequency) เป็นต้น ค่าคุณลักษณะเด่นกลุ่มนี้มีประสิทธิภาพสูงในการรู้จำ แต่ยากที่จะสกัดค่าจากสัญญาณเสียง

(๓) กลุ่มค่าคุณลักษณะเด่นแบบเอนเวโลปของสเปกตรัม (Spectrum envelop feature) เป็นกลุ่มที่นิยมใช้กันมาก เนื่องจากค่าคุณลักษณะเด่นส่วนใหญ่ที่ใช้สำหรับการรู้จำเสียงจะรวมอยู่ในข้อมูลเชิงสเปกตรัมนี้ อีกทั้งยังง่ายและสะดวกในการคำนวณหาค่าด้วย ตัวอย่าง ค่าคุณลักษณะเด่นกลุ่มนี้ ได้แก่ สัมประสิทธิ์การประมาณพันธะเชิงเส้น (Linear prediction coefficients: LPC), สัมประสิทธิ์เซปสตรัล (Cepstral coefficient) และพัฒนาการอีกมากมายจากเซปสตรัลปกติ อาทิ เช่นสัมประสิทธิ์เซปสตรัลที่คำนวณบนแกนความถี่แบบเมล (Mel frequency cepstral coefficients: MFCC) เซปสตรัลแบบผ่านตัวกรองภายหลัง (Post filtered Cepstral: PFL) เป็นต้น นอกจากนั้น ยังมีการคำนวณค่าการเปลี่ยนแปลง (Derivative หรือ Delta) ของสัมประสิทธิ์เหล่านี้มาใช้เป็นค่าคุณลักษณะเด่นเพิ่มเติมได้

**๑.๔ แบบจำลองฮิดเดนมาร์คอฟ**

แบบจำลองฮิดเดนมาร์คอฟ เป็นวิธีการจำแนกรูปแบบโดยอาศัยวิธีการทางสถิติในการรู้จำ เช่นการรู้จำเสียง ระบบจะทำการเก็บรวบรวมรายละเอียดทางสถิติเกี่ยวกับเสียงพูด โดยเก็บข้อมูลการกระจายที่สมบูรณ์ของลักษณะสำคัญของเสียงไว้ในข้อมูลฝึกฝน ซึ่งเมื่อมีการทดสอบกับชุดทดสอบ ก็จะสามารถจำแนกความแตกต่างระหว่างเสียงพูดได้เป็นอย่างดี

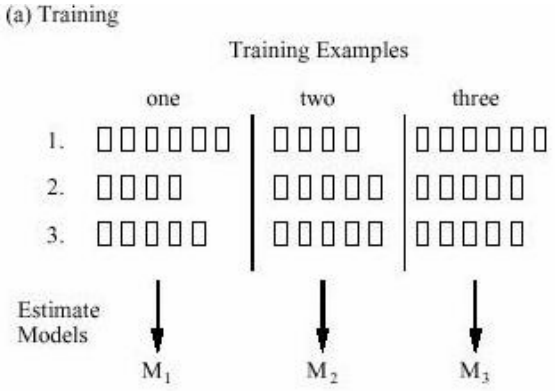



ภาพที่ ๑ โมเดลทางสถิติแบบจำลองฮิดเดนมาร์คอฟ

โมเดลทางสถิติแบบจำลองฮิดเดนมาร์คอฟ [๓] นี้ ใช้ได้ดีในการจำแนกรูปแบบของลำดับเหตุการณ์หนึ่งๆ โดยที่มีเพียงผลลัพธ์ที่ได้จากการสังเกตเท่านั้น แบบจำลองนี้มีลักษณะคร่าวๆคือ

- มีการย้ายสถานะทุกๆช่วงเวลาหนึ่งๆ
- จำนวนสถานะขึ้นอยู่กับจำนวนของสัญลักษณ์การสังเกต (Observation symbols)ทั้งหมด
- การเปลี่ยนสถานะจะมีความสัมพันธ์กับค่าความน่าจะเป็น โดยจะมีพารามิเตอร์เป็นอักษร a หมายถึง การกระจายของความน่าจะเป็นในการเปลี่ยนแปลงสถานะ (State Transition Probability Distribution)
- แต่ละสถานะมีความสัมพันธ์กับค่าความน่าจะเป็นของผลลัพธ์ที่เกิดขึ้น ค่าความน่าจะเป็นของผลลัพธ์ที่เกิดขึ้นแทนด้วยตัวอักษร b (Observation Symbol Probability Distribution)
- เมื่อเข้าสู่สถานะอื่นๆ จะมีการกำเนิดผลลัพธ์ขึ้น (feature vector)
- ในหนึ่งโมเดลจะประกอบด้วยสถานะหลายๆสถานะ แต่ละสถานะสามารถให้ผลลัพธ์ออกมาได้อีกจำนวนหนึ่งเท่าๆกัน

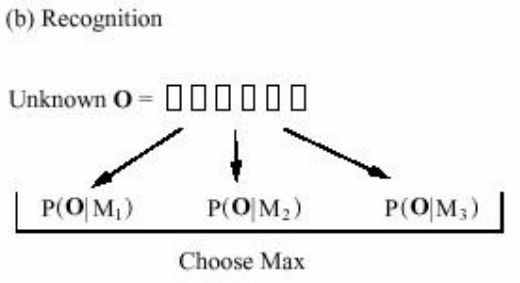
การใช้แบบจำลองฮิดเดนมาร์คอฟเพื่อรู้จำเสียง จะประกอบด้วยขั้นตอน ๒ ขั้นตอน ดังภาพที่ ๒ และ ๓




ภาพที่ ๒ ขั้นตอนการฝึกฝนแบบจำลอง

๑.๔.๑ ขั้นตอนการฝึกฝนแบบจำลอง

ในขั้นตอนนี้จะต้องแบ่งสัญญาณเสียงออกเป็นแต่ละหน่วยเสียงย่อยๆ และวัดค่าลักษณะสำคัญ แล้วนำค่าลักษณะสำคัญที่ได้ ไปเปลี่ยนให้อยู่ในรูปของสัญลักษณ์ที่ใช้กับแบบจำลอง ด้วยวิธีควอนไทซ์ค่าที่ได้ออกเป็นระดับๆ ผลที่ได้คือลำดับของสัญลักษณ์การสังเกต (Observation Sequence) ลำดับที่ได้จะมีลักษณะแตกต่างกันไปตามหน่วยเสียงแต่ละเสียง ลำดับการสังเกตของหน่วยเสียงเดียวกัน จะส่งผ่านเข้าไปในแบบจำลอง แบบจำลองจะปรับค่าความน่าจะเป็นต่างๆให้เข้ากับลำดับการสังเกตนั้นๆ ด้วยวิธี Forward-Backward Algorithm และ Baum-Welch และกำหนดลำดับสถานะที่เกิดขึ้นด้วยวิธี Viterbi ให้แบบจำลองมีความสามารถจำลองเสียงที่พูดเข้าไปได้ ผลที่ได้คือแบบจำลองที่มีค่าพารามิเตอร์เข้ากับหน่วยเสียงที่นำมาฝึกฝน ๑ แบบจำลองต่อหน่วยเสียง ๑ เสียง




ภาพที่ ๓ ขั้นตอนการทดสอบผลการรู้จำ

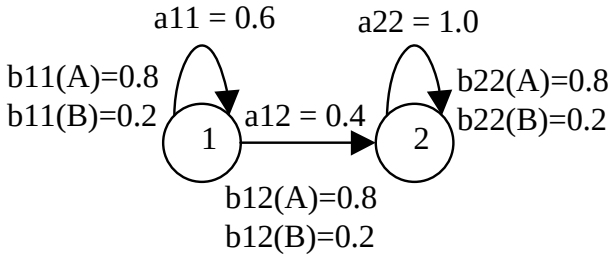
๑.๔.๒.ขั้นตอนการทดสอบผลการรู้จำ

โดยนำแบบจำลองของทุกๆหน่วยเสียงมาเชื่อมขนานกัน แล้วนำสัญญาณเสียงที่ต้องการทราบความหมาย ผ่านการวัดค่าลักษณะสำคัญ และควอนไทซ์ให้อยู่ในรูปลำดับการสังเกต ลำดับการสังเกตที่ได้จะผ่านแบบจำลองทุกๆแบบ เพื่อเปรียบเทียบหาความน่าจะเป็นว่าแบบจำลองใดให้ความน่าจะเป็นที่แบบจำลองนั้นๆจะเกิดลำดับการสังเกตแบบนี้ ได้มากที่สุด ตัวอย่างของแบบจำลองฮิดเดนมาร์คอฟ [๔] ที่มี ๒ เอาท์พุตคือ A และ B และ ๒ สถานะ คือ ๑ และ ๒ แสดงดังภาพที่ ๔

กำหนดให้

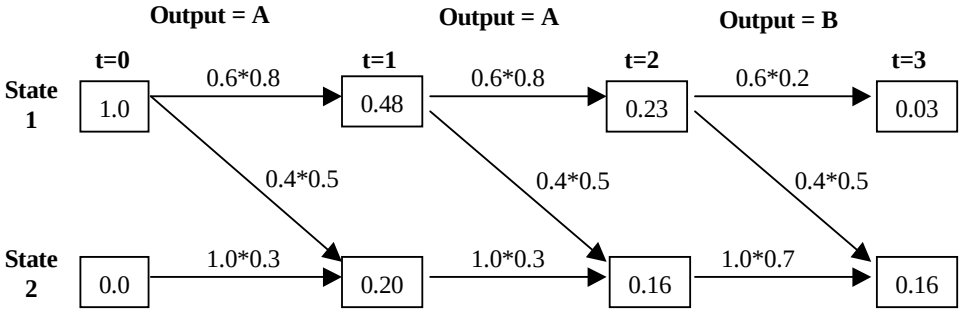
$a_{ij} = P(X_{t+1} = j \mid X_t = i)$ คือ a ความน่าจะเป็นของสถานะถัดไปจะเป็น j เมื่อสถานะปัจจุบันเป็น i

$b_{ij}(k) = P(Y_t = k \mid X_t = i, X_{t+1} = j)$ คือ ความน่าจะเป็นของผลลัพธ์เป็น k เมื่อมีการเปลี่ยนสถานะจากสถานะ-i ไปหาสถานะ-j




ภาพที่ ๔ ตัวอย่าง แบบจำลองฮิดเดนมาร์คอฟ ที่มี ๒ สถานะ และมี ๒ เครื่องหมายผลลัพธ์

ภาพที่ ๔ เป็นตัวอย่างโมเดลแบบง่ายๆ ที่ผ่านการฝึกฝนหรือมีการกำหนดค่าความน่าจะเป็นของ a และ b มาแล้ว ค่า a๑๒=๐.๔ หมายความว่า เมื่อสถานะปัจจุบันเป็น ๑ สถานะ ต่อไปจะเป็นสถานะ ๒ มีความน่าจะเป็นอยู่ ๐.๔ ส่วน b๑๒ (A) =๐.๘ คือ เมื่อมีการเปลี่ยนสถานะจากสถานะ๑ ไปหาสถานะ ๒ ความน่าจะเป็นที่จะมีผลลัพธ์เป็น A คือ ๐.๘ (เมื่อมีการเปลี่ยนสถานะจะมีการสร้างผลลัพธ์หนึ่งออกมาเท่านั้น)




ภาพที่ ๕ การคำนวณ trellisตามรูปแบบ forward computation

ภาพที่ ๕ แสดงถึงการคำนวณของค่า *trellis* โดยการใช้ทฤษฎีของ แบบจำลองฮิดเด็นมาร์คอฟ ตามภาพที่ ๕ โดยสมมุติว่าลำดับการสังเกตเป็น A A B ตามลำดับ ได้ผลลัพธ์เป็นลำดับของสถานะ ๑ ๑ ๑ ๒

**๒. งานวิจัยที่เกี่ยวข้อง**

Ali J.A. Soleymani, Dr. Martin J. McCutcheon and Dr. M. H. Southwoo (ค.ศ. ๑๙๙๗) ศึกษาเรื่อง โปรแกรมช่วยฝึกฝนการเปล่งเสียงที่ชื่อ SIM Speech Illumina Mentor หรือ SIM จากการศึกษาพบว่า การพัฒนาโปรแกรมช่วยฝึกฝนการเปล่งเสียงที่ชื่อ SIM Speech Illumina Mentor หรือ SIM เพื่อลดปัญหาในเรื่องการขาดแคลนบุคลากรและเวลาในการสอน โดยจะใช้กลไกการ

เปล่งเสียง speech articulators feedback เป็นข้อมูลแสดงกลับมาให้ผู้เรียนได้ทราบว่าผู้เรียนมีกลไกในการเปล่งเสียงอย่างไร แทนการแสดงเสียงที่ผู้เรียนเปล่งออกมาด้วยข้อมูลทางเสียง (auditory feedback) ที่เคยใช้มาแต่เดิม ซึ่งข้อมูลทางเสียงจะต้องอาศัยผู้เชี่ยวชาญในการวิเคราะห์และตีความ ซึ่งเหมาะกับนักแก้ไขการพูด (speech therapy) มากกว่าผู้เรียน

Massaro, D (ค.ศ. ๒๐๐๓) ศึกษาเรื่อง โปรแกรมที่ช่วยฝึกฝนการออกเสียงสระภาษาอังกฤษ BALDI จากการศึกษาพบว่า โปรแกรมนี้จะใช้โมเดลแบบสามมิติในการจำลองศีรษะของมนุษย์ (๓D model of a human head ) เพื่อเปล่งเสียงสระและแสดงตำแหน่งของกลไกในการออกเสียงของสระภาษาอังกฤษ ซึ่งในแต่ละสระจะมีตำแหน่งของกลไกในการออกเสียงที่แตกต่างกัน ซึ่งผิวหน้าของแบบจำลองสามมิติของศีรษะมนุษย์จะสามารถแสดงในลักษณะแบบโปร่งใสได้ เพื่อให้ผู้มีความบกพร่องทางการได้ยินสามารถเห็นการเคลื่อนไหวที่อยู่ภายในปาก เช่น ลิ้น ฟัน เพดานแข็ง นอกจากนี้ยังมีการใช้ภาพเพื่อสื่อถึงลักษณะของเสียงที่เปล่งออกมา เช่น ใช้ภาพวงแหวนจากคอเพื่อแสดงว่ามีการสั่นของเส้นเสียง(vocal cord) ภาพกลุ่มของจุดจากปากเพื่อแสดงว่าเป็นเสียงแบบเสียดแทรก(fricative) และใช้ภาพปีกจมูกสีแดงเพื่อแสดงว่าเป็นเสียงแบบนาสิก (nasal) สำหรับในส่วนของแบบฝึกหัดจะใช้ลักษณะของวิดีโอเกมในการนำเสนอ เช่น แสดงเป็นรูปภาพของผัก จากนั้นแบบจำลองสามมิติศีรษะมนุษย์จะถามว่าเป็นผักอะไร เพื่อให้ผู้มีความบกพร่องทางการได้ยินได้ตอบชื่อผักกลับมาเพื่อที่โปรแกรมจะได้ทำการวิเคราะห์ว่า ผู้มีความบกพร่องทางการได้ยินเปล่งเสียงได้ถูกต้องหรือไม่

วิศรุต อาขุบุตร (พ.ศ.๒๕๓๙) ศึกษาเรื่อง ระบบการรู้จำคำไทยหลายพยางค์แบบไม่ขึ้นกับผู้พูด โดยใช้แบบจำลองฮิดเดนมาร์คอฟ จากการศึกษาพบว่า การเปรียบเทียบประสิทธิภาพวิธีต่างๆที่ใช้ในการรู้จำเสียงพูดตัวเลขไทยตั้งแต่ ๐-๙ โดยใช้ตัวอย่างเสียงจากผู้ชายและผู้หญิงจำนวน ๖๐ คน ทั้งหมด ๑๒๐๐ เสียง โดยให้พูดภายใต้สภาพแวดล้อมเดียวกัน เทคนิคที่ใช้ในการเปรียบเทียบได้แก่ Hidden Markov Model ซึ่งได้ความถูกต้อง ๘๔.๒๕ % Neural Network ได้ความถูกต้อง ๗๘.๐๐% และ Fuzzy-Neural Network ได้ความถูกต้อง ๗๘.๓๐% ทำให้เห็นได้ว่าแบบจำลองฮิดเดนมาร์คอฟได้ประสิทธิภาพดีที่สุด นอกจากนี้เมื่อเทียบกับวิธี neural network แล้ววิธีแบบจำลองฮิดเดนมาร์คอฟ ยังไม่จำกัดความยาวของเสียงอินพุต และไม่จำเป็นต้องตัดเสียงประโยคให้เป็นคำก่อนเข้าสู่แบบจำลอง

จิตรลดา จารุมิศร์, ไกรสิน ส่งวัฒนา, อิทธิชัย อรุณศรีแสงชัย (พ.ศ.๒๕๔๑) ศึกษาเรื่อง วิธีการรู้จำเสียงวรรณยุกต์แบบไม่ต่อเนื่อง โดยใช้วิธีหาค่าคาบความถี่มูลฐานแบบออโต้คอรีเลชัน (Autocorrelation, AUTOC) จากการศึกษาพบว่า จากการกรองสัญญาณเสียงให้สัญญาณเสียงที่มีความถี่ต่ำผ่านเพื่อกำจัดความถี่ฮาร์โมนิคที่ไม่ต้องการออกไป จากนั้นจึงแบ่งสัญญาณออกเป็น

เฟรมๆ ทำการตัดสัญญาณที่มีขนาดแอมปลิจูดต่ำซึ่งคาดว่าจะเป็นสัญญาณรบกวนออก จากนั้นนำสัญญาณที่ผ่านการตัดสัญญาณรบกวนออก มาคำนวณหาค่า Autocorrelation ระยะห่างจากจุดเริ่มต้นของสัญญาณไปยังตำแหน่งที่มี Autocorrelation peak สูงสุด จะเป็นค่าคาบเวลาพิทช์ จากนั้นนำค่าคาบเวลาพิทช์มาคำนวณเพื่อหาค่าความถี่มูลฐาน Fo (Pitch) หรือเสียงวรรณยุกต์ แล้วจึงนำความถี่มูลฐานที่ได้นี้มาเข้าสู่ระบบรู้จำเสียงที่สร้างด้วยแบบจำลองฮิดเดนมาร์คอฟ ได้ผลความถูกต้องอยู่ที่ ๙๐%