การปรับปรุงแบบจำลองเสียงเพื่อเพิ่มความเป็นธรรมชาติของเสียงสังเคราะห์ภาษาไทย

นายศุภเดช ฉันจรัสวิชัย

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรดุษฎีบัณฑิต
สาขาวิชาวิศวกรรมคอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2560

An Improvement of Acoustic Model for Enhancing Naturalness in the Synthesized Thai Speech

Mr. Supadaech Chanjaradwichai

A Dissertation Submitted in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy Program in Computer Engineering
Department of Computer Engineering
Faculty of Engineering
Chulalongkorn University
Academic Year 2017

หัวข้อวิทยานิพนธ์	การปรับปรุงแบบจำลองเสียงเพื่อเพิ่มความเป็นธรรมชาติของเสียงสังเคราะห์ภาษาไทย
โดย	นายศุภเดช ฉันจรัสวิชัย
สาขาวิชา	วิศวกรรมคอมพิวเตอร์
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	รองศาสตราจารย์ ดร.อติวงศ์ สุชาโต
อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม	ผู้ช่วยศาสตราจารย์ ดร. โปรดปราน บุณยพุกกณะ

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาดุษฎีบัณฑิต

.....คณบดีคณะวิศวกรรมศาสตร์

(รองศาสตราจารย์ ดร.สุพจน์ เตชวรสินสกุล)

คณะกรรมการสอบวิทยานิพนธ์

.....ประธานกรรมการ

(ศาสตราจารย์ ดร. ประภาส จงสถิตย์วัฒนา)

.....อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

(รองศาสตราจารย์ ดร.อติวงศ์ สุชาโต)

.....อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม

(ผู้ช่วยศาสตราจารย์ ดร. โปรดปราน บุณยพุกกณะ)

.....กรรมการ

(ผู้ช่วยศาสตราจารย์ ดร. นัทที นิภานันท์)

.....กรรมการ

(ศาสตราจารย์ ดร.บุญเสริม กิจศิริกุล)

.....กรรมการภายนอกมหาวิทยาลัย

(ดร. ชัย วุฒิวิวัฒน์ชัย)

ศุภเดช ฉันจรัสวิชัย : การปรับปรุงแบบจำลองเสียงเพื่อเพิ่มความเป็นธรรมชาติของเสียงสังเคราะห์ภาษาไทย (An Improvement of Acoustic Model for Enhancing Naturalness in the Synthesized Thai Speech) อ.ที่ปรึกษาวิทยานิพนธ์หลัก: รศ. ดร.อติวงศ์ สุชาโต, อ.ที่ปรึกษาวิทยานิพนธ์ร่วม: ผศ. ดร. โปรดปราน บุณยพุกกณะ, 123 หน้า.

เสียงสังเคราะห์เป็นเทคโนโลยีทางเลือกสำหรับการรับรู้ข้อมูลประเภทข้อความ ความชัดเจน และความเป็นธรรมชาติของเสียงสังเคราะห์ส่งผลโดยตรงกับความเข้าใจของผู้ฟังที่มีต่อข้อมูลในสัญญาณเสียง ดังนั้นในงานวิจัยนี้จึงได้พัฒนาด้านความเป็นธรรมชาติ และความชัดเจนของเสียงสังเคราะห์ที่สร้างมาจากค่าพารามิเตอร์ของตัวเข้ารหัสเสียง STRAIGHT ซึ่งค่าพารามิเตอร์เหล่านั้นถูกสร้างขึ้นมาจากแบบจำลองฮิดเดนมาร์คอฟ และแบบจำลองโครงข่ายประสาทเทียมแบบลึก โดยการนำเสนอแนวคิด 3 แนวคิด ได้แก่ 1) แนวคิดการแยกกันของแบบจำลองคุณลักษณะความถี่มูลฐาน และค่าคุณลักษณะสเปกตรัม โดยทั้งสองแบบจำลองถูกฝึกฝนแยกกันเพื่อสร้างเป็นแบบจำลองฮิดเดนมาร์คอฟสำหรับสร้างค่าพารามิเตอร์ของตัวเข้ารหัสเสียง STRAIGHT ที่สอดคล้องกับแบบจำลองดังกล่าว ในวิทยานิพนธ์นี้ได้นำเสนอขั้นตอนวิธีในการปรับแนวเวลาของค่าพารามิเตอร์ที่สร้างขึ้นมาจากการใช้สองแบบจำลอง 2) เสนอการปรับเปลี่ยนค่าคุณลักษณะส่วนรับเข้าของโครงข่ายประสาทเทียมแบบลึกที่ถูกใช้ในการสร้างค่าพารามิเตอร์ของตัวเข้ารหัสเสียง STRAIGHT จากเดิมที่ใช้ค่าคุณลักษณะทางบริบท เป็นแบบจำลองฮิดเดนมาร์คอฟที่เป็นผลลัพธ์จากต้นไม้ตัดสินใจที่ใช้ในการจัดกลุ่มบริบท 3) นำเสนอวิธีการนอร์มัลไลเซชันค่าคุณลักษณะส่วนส่งออกของแบบจำลองโครงข่ายประสาทเทียมแบบลึก ที่ใช้ค่ากลาง และค่าความแปรปรวนจากแบบจำลองฮิดเดนมาร์คอฟที่เป็นผลลัพธ์จากต้นไม้ตัดสินใจ ในการทดสอบได้ทำการทดสอบ 2 รูปแบบ คือ 1) การทดสอบปรนัยที่ใช้ตัวชี้วัดค่าความเพี้ยนของเซปตรัลในระดับเมลของค่าสัมประสิทธิ์เมลเคปสตรัม (MGC_MCD) ค่าความเพี้ยนของเซปตรัลในระดับเมลของค่าแถบคลื่นความถี่ของความไม่เป็นคาบ (BAP_MCD) ความไม่สอดคล้องกันของสถานะความก้องของเสียง (LF0_UVU) และความผิดพลาดกำลังสองเฉลี่ยของค่าความถี่มูลฐาน (LF0_RMSE) 2) การทดสอบอัตนัยที่ใช้ผู้ทดสอบ 9 คน โดยวัดในด้านของความชัดเจน และความเป็นธรรมชาติของเสียงสังเคราะห์ ผลการทดสอบปรนัยการใช้แนวคิดที่ 2 และ 3 กับแบบจำลองโครงข่ายประสาทเทียมแบบลึกสามารถสังเคราะห์ค่าพารามิเตอร์ของตัวเข้ารหัสเสียง STRAIGHT ได้ใกล้เคียงกับเสียงต้นฉบับมากกว่าการใช้แนวคิดที่ 1 กับแบบจำลองฮิดเดนมาร์คอฟ และแบบจำลองดั้งเดิมทั้งในส่วนของแบบจำลองฮิดเดนมาร์คอฟ และแบบจำลองโครงข่ายประสาทเทียมแบบลึก สำหรับในการทดสอบอัตนัยพบว่าการใช้แนวคิดที่ 1 กับแบบจำลองฮิดเดนมาร์คอฟสามารถสังเคราะห์ค่าคุณลักษณะที่มีความเป็นธรรมชาติ และชัดเจนมากกว่าการใช้แนวคิดอื่น และแบบจำลองดั้งเดิมทั้งสองแบบจำลอง

ภาควิชา วิศวกรรมคอมพิวเตอร์	ลายมือชื่อนิสิต .....
สาขาวิชา วิศวกรรมคอมพิวเตอร์	ลายมือชื่อ อ.ที่ปรึกษาหลัก .....
ปีการศึกษา 2560	ลายมือชื่อ อ.ที่ปรึกษาร่วม .....

5571431321 : MAJOR COMPUTER ENGINEERING

KEYWORDS: SPEECH SYNTHESIS / HIDDEN MARKOV MODEL / DEEP NEURAL NETWORK

SUPADAECH CHANJARADWICHAI: An Improvement of Acoustic Model for Enhancing Naturalness in the Synthesized Thai Speech. ADVISOR: ASSOC. PROF. DR.ATIWONG SUCHATO, CO-ADVISOR: ASST. PROF. PROADPRAN PUNYABUKKANA, Ph.D., 123 pp.

Speech synthesis converts text to speech signals. The naturalness and intelligibility of synthesized speech affect the listeners' understanding of the content conveyed by the speech signal. This dissertation proposed 3 aspects of improving the naturalness and intelligibility of synthesized speech generating from STRAIGHT parameters. The first aspect was the separation of spectral-feature models and the fundamental-frequency models. The two types of models were trained independently to obtain the Hidden Markov Model (HMM) parameters, optimized for generating their respective STRAIGHT parameters. Algorithms handling the time-alignment of parameters, generating separately from the two models were proposed. In this work, we focused on generating STRAIGHT parameters from either HMMs or Deep Neural Networks (DNNs). The second aspect was concerned with the modification of typical inputs to DNNs used for generating STRAIGHT parameters from direct phonetic contexts, to HMMs resulting from context clustering decision trees. The third aspect was the DNN output normalization using means and variances from HMMs, which were the results of the decision trees. Tools for objective evaluations were Mel cepstral distortion for Mel cepstral coefficient of spectral filter (MGC_MCD), Mel cepstral distortion for coefficient of band aperiodicity filter (BAP_MCD), root mean square error of fundamental frequency (LF0_RMSE), and count of unmatched voicing condition between natural speech and synthesized speech (LF0_UVU). Nine participants were recruited to perform a subjective evaluation in which they were asked to evaluate the synthesized speech utterances in terms of their naturalness and intelligibility. The results of the objective test showed that applying the second and the third proposed aspects to DNN generated STRAIGHT parameters resulted in better synthesized speech than applying the first aspect to HMM models as well as using baseline HMM and DNN methods. The subjective results showed that the application of the first aspect to HMM outperformed other methods.

Department:	Computer Engineering	Student's Signature	.....
Field of Study:	Computer Engineering	Advisor's Signature	.....
Academic Year:	2017	Co-Advisor's Signature	.....

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงไปด้วยดี เนื่องด้วยได้รับคำแนะนำและการช่วยเหลืออย่างดียิ่งจากรองศาสตราจารย์ ดร.อติวงศ์ สุชาโต อาจารย์ที่ปรึกษาวิทยานิพนธ์หลักและผู้ช่วยศาสตราจารย์ ดร. โปรดปราน บุณยพุกกณะ อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม ที่ได้สละเวลาให้คำปรึกษาและคำแนะนำในการทำวิทยานิพนธ์ และให้ความช่วยเหลือผู้วิจัยทุกครั้งที่ประสบปัญญาในการทำวิทยานิพนธ์เสมอมา ผู้วิจัยรู้สึกซาบซึ้งและขอขอบพระคุณอย่างสูงมา ณ โอกาสนี้

ขอขอบพระคุณ ศาสตราจารย์ ดร. ประภาส จงสถิตย์วัฒนา ศาสตราจารย์ ประธานกรรมการสอบวิทยานิพนธ์ ศาสตราจารย์ ดร.บุญเสริม กิจศิริกุล ผู้ช่วยศาสตราจารย์ ดร. นัททีนิภานันท์ และ ดร. ชัย วุฒิวิวัฒน์ชัย กรรมการสอบวิทยานิพนธ์

ขอขอบพระคุณ คณาจารย์ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัยทุกท่าน ที่ได้ถ่ายทอดความรู้ พร้อมทั้งให้คำแนะนำในการสร้างผลงานวิชาการ และประสบการณ์อันมีค่ายิ่ง ตลอดจนชี้แนะแนวทางในการเรียนระดับบัณฑิตศึกษาและมุมมองการดำเนินชีวิต และยินดีให้ความช่วยเหลือเสมอมา

ขอขอบคุณเพื่อนๆ พี่น้องใน SLS Lab ทุกคน ที่คอยให้กำลังใจและให้ความช่วยเหลือเสมอมา

ขอขอบคุณคุณธัญญาพร เจียศิริพันธ์ ที่คอยห่วงใย ให้กำลังใจ เป็นแรงผลักดัน และสนับสนุน ตลอดจนคำแนะนำในเรื่องต่างๆ และให้ความช่วยเหลืออย่างดียิ่งเสมอมา

นอกจากนี้ ผู้วิจัยได้รับทุนสนับสนุนวิจัย ทุนอุดหนุนวิทยานิพนธ์ จุฬาลงกรณ์มหาวิทยาลัย ซึ่งผู้วิจัยขอบขอบพระคุณอย่างยิ่งและหวังว่างานวิจัยนี้จะเป็นประโยชน์ทางการศึกษา

ท้ายที่สุดนี้ขอขอบพระคุณบุคคลในครอบครัวที่คอยห่วงใย ให้กำลังใจและให้การช่วยเหลือ สนับสนุนทั้งกำลังกายและใจ กำลังทรัพย์แก่ผู้วิจัยมาโดยตลอด ทำให้ผู้วิจัยประสบความสำเร็จในวันนี้

สารบัญ

หน้า

บทคัดย่อภาษาไทย ..... ง
บทคัดย่อภาษาอังกฤษ ..... จ
กิตติกรรมประกาศ ..... ฉ
สารบัญ ..... ช
สารบัญรูปภาพ ..... ญ
สารบัญตาราง ..... ฐ
บทที่ 1 บทนำ 1
1.1 ที่มาและความสำคัญ 1
1.2 วัตถุประสงค์ของงานวิจัย 2
1.3 ขอบเขตงานวิจัย 2
1.4 แนวคิดการวิจัย 2
1.5 อภิธานศัพท์ 3
บทที่ 2 งานวิจัยที่เกี่ยวข้อง 6
2.1 ความซับซ้อนของแบบจำลอง 8
2.2 โครงสร้างแบบจำลองเสียง 10
2.2.1 การปรับเปลี่ยนโครงสร้างแบบจำลองเสียงของค่าคุณลักษณะสเปกตรัม 10
2.2.2 การปรับเปลี่ยนโครงสร้างแบบจำลองเสียงของค่าคุณลักษณะความถี่มูลฐาน 10
2.2.3 แบบจำลองโครงข่ายประสาทเทียมแบบลึก 13
บทที่ 3 ทฤษฎีที่เกี่ยวข้อง 16
3.1 ตัวเข้ารหัสเสียง STRAIGHT 16
3.2 การสร้างแบบจำลองเสียงสังเคราะห์ด้วยแบบจำลองฮิดเดนมาร์คอฟ 18
3.3 การสังเคราะห์เสียงด้วยแบบจำลองฮิดเดนมาร์คอฟ 25

3.3.1 การหาแบบจำลองที่สอดคล้องกับบริบท 26

3.3.2 การจัดเรียงและขยายช่วงเวลาของแบบจำลอง 26

3.3.3 การปรับแนวค่าคุณลักษณะ 27

3.3.4 การสังเคราะห์เสียง 28

3.4 การสร้างแบบจำลองเสียงสังเคราะห์ด้วยแบบจำลองโครงข่ายประสาทเทียมแบบลึก 28

3.4.1 การสกัดค่าคุณลักษณะส่วนส่งออก (Output Feature) 29

3.4.2 การปรับแนวในระดับสถานะ (State-level alignment) 29

3.4.3 การสร้างค่าคุณลักษณะส่วนรับเข้า (Input Feature) 30

3.4.4 การฝึกฝนแบบจำลอง 32

3.5 การสังเคราะห์เสียงด้วยแบบจำลองโครงข่ายประสาทเทียมแบบลึก 33

3.5.1 หาแบบจำลองช่วงเวลาที่สอดคล้องกับบริบท 34

3.5.2 สร้างค่าคุณลักษณะส่วนรับเข้า 34

3.5.3 การคำนวณค่าคุณลักษณะ 34

3.5.4 การปรับแนวค่าคุณลักษณะ 34

บทที่ 4 แนวคิดของการวิจัย และวิธีการดำเนินงาน 36

4.1 การปรับเปลี่ยนโครงสร้างของแบบจำลองเสียง 36

4.1.1 โครงสร้างของแบบจำลองเสียง 36

4.1.2 การคำนวณหาช่วงเวลาของเสียงสังเคราะห์ 44

4.2 ปรับเปลี่ยนค่าคุณลักษณะ และวิธีการฝึกฝนโครงข่ายประสาทเทียมแบบลึก 55

4.2.1 การปรับค่าคุณลักษณะส่วนรับเข้า 55

4.2.2 การนอร์มัลไลเซชันค่าคุณลักษณะส่วนส่งออก 62

บทที่ 5 การทดลอง และวิธีการวิเคราะห์ผลการทดลอง 65

5.1 ฐานข้อมูลเสียงภาษาไทย สำหรับสร้างระบบสังเคราะห์เสียง 65

5.2 การสกัดค่าคุณลักษณะ 68

5.3 แบบจำลองฮิดเดนมาร์คอฟ....68

5.4 แบบจำลองโครงข่ายประสาทเทียมแบบลึก....70

5.4 การทดสอบแบบปรนัย....74

5.4.1 การวัดผลช่วงเวลาของแบบจำลองเสียง....74

5.4.2 การวัดผลคุณภาพของเสียงสังเคราะห์ 75

5.5 การทดสอบแบบอัตนัย....76

บทที่ 6 ผลการทดลอง และวิเคราะห์ผลการทดลอง....78

6.1 การทดสอบแบบปรนัย....78

6.1.1 การวัดผลช่วงเวลาของแบบจำลองเสียง....78

6.1.2 การวัดผลคุณภาพของเสียงสังเคราะห์ 81

6.1.3 ความไม่สอดคล้องของสถานะความก้องของเสียง 92

6.2 การทดสอบแบบอัตนัย....95

บทที่ 7 สรุปและอภิปรายผลการวิจัย 99

รายการอ้างอิง 102

ภาคผนวก.... 107

ภาคผนวก ก คำถามของต้นไม้ตัดสินใจ 108

ประวัติผู้เขียนวิทยานิพนธ์ 123

สารบัญรูปภาพ

เรื่อง **หน้า**

รูปที่ 1 ส่วนประกอบของตัวเข้ารหัสเสียง STRAIGHT 17

รูปที่ 2 ขั้นตอนการสร้างแบบจำลองเสียง.....19

รูปที่ 3 แบบจำลองเสียงแบบฮิดเดนมาร์คอฟสำหรับสเปกตรัม และความถี่มูลฐานที่ใช้ในการสร้างเสียงสังเคราะห์.....20

รูปที่ 4 แบบจำลองฮิดเดนมาร์คอฟของช่วงเวลา.....21

รูปที่ 5 ตัวอย่างแบบจำลองแบบขึ้นกับบริบท และไม่ขึ้นกับบริบท 22

รูปที่ 6 ตัวอย่างต้นไม้ตัดสินใจของแบบจำลองเสียง.....23

รูปที่ 7 ต้นไม้ตัดสินใจของแบบจำลองเสียง 24

รูปที่ 8 ตัวอย่างต้นไม้ตัดสินใจของแบบจำลองช่วงเวลา.....25

รูปที่ 9 ขั้นตอนการสังเคราะห์เสียง 26

รูปที่ 10 ตัวอย่างการจัดเรียงและขยายช่วงเวลาของแบบจำลอง 27

รูปที่ 11 ตัวอย่างการปรับแนวค่าคุณลักษณะ.....27

รูปที่ 12 ขั้นตอนการสร้างแบบจำลองเสียงสังเคราะห์จากแบบจำลองโครงข่ายประสาทเทียมแบบลึก.....29

รูปที่ 13 ตัวอย่างการปรับแนวในระดับสถานะ 30

รูปที่ 14 ขั้นตอนการสังเคราะห์เสียงจากแบบจำลองโครงข่ายประสาทเทียมแบบลึก.....33

รูปที่ 15 ตัวอย่างผลลัพธ์การปรับแนวของค่าคุณลักษณะที่ได้จากการสังเคราะห์ 35

รูปที่ 16 แบบจำลองเสียงแบบฮิดเดนมารคอฟสำหรับสเปกตรัม และความถี่มูลฐาน.....37

รูปที่ 17 ตัวอย่างการแบ่งสถานะของค่าคุณลักษณะ 1 ค่า 37

รูปที่ 18 ตัวอย่างการแบ่งสถานะที่ให้ค่าถ่วงน้ำหนักกับค่าคุณลักษณะสเปกตรัมเพียงอย่างเดียว.....38

รูปที่ 19 ตัวอย่างการแบ่งสถานะที่ให้ค่าถ่วงน้ำหนักกับค่าคุณลักษณะความถี่มูลฐานเพียงอย่างเดียว....39
รูปที่ 20 ตัวอย่างการแบ่งสถานะที่ให้ค่าถ่วงน้ำหนักกับค่าคุณลักษณะความถี่มูลฐานเท่ากับค่าคุณลักษณะสเปกตรัม....39
รูปที่ 21 แบบจำลองสเปกตรัม....41
รูปที่ 22 แบบจำลองความถี่มูลฐาน....42
รูปที่ 23 ต้นไม้ตัดสินใจสำหรับแบบจำลองค่าสเปกตรัมที่นำเสนอ....43
รูปที่ 24 ต้นไม้ตัดสินใจสำหรับแบบจำลองค่าความถี่มูลฐานที่นำเสนอ....44
รูปที่ 25 กรณีตัวอย่างในการคำนวณช่วงเวลาของสถานะแบบไม่พิจารณาสถานะความเป็นเสียง....47
รูปที่ 26 ผลลัพธ์การหาช่วงเวลาของสถานะกรณีที่ไม่พิจารณาสถานะความเป็นเสียง....49
รูปที่ 27 ตัวอย่างกรณีที่มีสถานะความเป็นเสียงที่เหมือนกัน....50
รูปที่ 28 ตัวอย่างกรณีที่มีสถานะความเป็นเสียงที่ไม่เหมือนกัน....51
รูปที่ 29 ตัวอย่างกรณีที่มีการละทิ้งสถานะ....52
รูปที่ 30 กรณีตัวอย่างในการคำนวณช่วงเวลาของสถานะแบบพิจารณาสถานะความเป็นเสียง ..54
รูปที่ 31ผลลัพธ์การหาช่วงเวลาของสถานะกรณีที่พิจารณาสถานะความเป็นเสียง....54
รูปที่ 32 ส่วนประกอบของการสังเคราะห์เสียงด้วยแบบจำลองฮิดเดนมาร์คอฟ....57
รูปที่ 33 ส่วนประกอบการสังเคราะห์เสียงด้วยแบบจำลองโครงข่ายประสาทเทียมแบบลึก....58
รูปที่ 34 ส่วนประกอบการสังเคราะห์เสียงด้วยแบบจำลองแบบผสม....60
รูปที่ 35 กระบวนการนอร์มัลไลเซชันค่าคุณลักษณะส่วนส่งออก....64
รูปที่ 36 โครงสร้างของพยางค์ในภาษาไทย....67
รูปที่ 37 รูปแบบที่ใช้แทนค่าคุณลักษณะ....68
รูปที่ 38 ผลการทดลองช่วงเวลาของแบบจำลองเสียง....80
รูปที่ 39 ผลการทดลอง MGC_MCD....81
รูปที่ 40 ผลการทดลอง BAP_MCD....82

รูปที่ 41 ผลการทดลสอบ LF0_UVU 84
รูปที่ 42 ผลการทดลอง LF0_RMSE 85
รูปที่ 43 ตัวอย่างค่าคุณลักษณะความถี่มูลฐานที่ถูกสังเคราะห์ 87
รูปที่ 44 ผลการวิเคราะห์ความไม่สอดคล้องกันของสถานะ 94
รูปที่ 45 ผลการทดลองความเป็นธรรมชาติของเสียงสังเคราะห์.....96
รูปที่ 46 ผลการทดลองความชัดเจนของเสียงสังเคราะห์ 98

สารบัญตาราง

เรื่อง **หน้า**

ตารางที่ 1 ตัวอย่างค่าคุณลักษณะส่วนรับเข้าของแบบจำลองโครงข่ายประสาทเทียมแบบลึก ...32
ตารางที่ 2 การคำนวณหาความยาวของสถานะในกรณีไม่พิจารณาสถานะความเป็นเสียง 46
ตารางที่ 3 ตัวอย่างการคำนวณช่วงเวลาของสถานะในกรณีที่ไม่พิจารณาสถานะความเป็นเสียง 48
ตารางที่ 4 การคำนวณหาความยาวของสถานะในกรณีที่พิจารณาสถานะความเป็นเสียง.....53
ตารางที่ 5 ตัวอย่างการคำนวณช่วงเวลาของสถานะในกรณีที่พิจารณาสถานะความเป็นเสียง....55
ตารางที่ 6 ตัวอย่างโครงสร้างของต้นไม้ตัดสินใจ 61
ตารางที่ 7 กระบวนการคัดเลือกกลุ่มที่ใช้ในการทดสอบ.....66
ตารางที่ 8 รายละเอียดของข้อมูลเสียงทั้ง 3 กลุ่ม.....66
ตารางที่ 9 ตัวอย่างการแสดงค่าคุณลักษณะทางบริบทในตำแหน่งต่างๆ.....67
ตารางที่ 10 จำนวนใบในต้นไม้จากแบบจำลอง HMM_BASE.....70
ตารางที่ 11 จำนวนค่าคุณลักษณะส่วนรับเข้าของแบบจำลองโครงข่ายประสาทเทียมที่แตกต่างกัน.....71
ตารางที่ 12 ตัวแปรของแบบจำลองโครงข่ายประสาทเทียมแบบลึก 72
ตารางที่ 13 ความหมายของแต่ละระดับของประเด็นที่ใช้ในการทดสอบอัตนัย 77
ตารางที่ 14 ผลการทดลองช่วงเวลาของแบบจำลองเสียง 78
ตารางที่ 15 ผลรวมความยาวของตัวอย่างเสียงในแต่ละรูปแบบ 79
ตารางที่ 16 การเปรียบเทียบความแตกต่างของการทดลอง MGC_MGD ที่ความเชื่อมั่นระดับ 95%.....88
ตารางที่ 17 การเปรียบเทียบความแตกต่างของการทดลอง BAP_MGD ที่ความเชื่อมั่นระดับ 95%.....89
ตารางที่ 18 การเปรียบเทียบความแตกต่างของการทดลอง LF0_UVU ที่ความเชื่อมั่นระดับ 95%.....90

ตารางที่ 19 การเปรียบเทียบความแตกต่างของการทดลอง LF0_RMSE ที่ความเชื่อมั่นระดับ 95%.....91

บทที่ 1 บทนำ

1.1 ที่มาและความสำคัญ

เทคโนโลยีการสังเคราะห์เสียง (Speech Synthesis) เป็นเทคโนโลยีที่สำคัญสำหรับผู้พิการทางด้านสายตา และเข้ามามีบทบาทในชีวิตประจำวันมากขึ้นในกลุ่มคนปกติ เทคโนโลยีการสังเคราะห์เสียงช่วยให้ผู้พิการทางด้านสายตารับรู้ข้อมูลประเภทข้อความ โดยการแปลงข้อความดังกล่าวเป็นเสียงพูดที่ถูกสังเคราะห์ขึ้นมา และช่วยให้กลุ่มคนปกติรับรู้ข้อมูลประเภทข้อความได้ในสถานการณ์ที่ไม่สะดวกในการอ่านข้อความ เช่น ในกรณีที่ผู้ใช้งานอยู่ในขณะขับรถ

เทคโนโลยีการสังเคราะห์เสียงที่ใช้แบบจำลองแบบฮิดเดนมาร์คอฟ (Hidden Markov Model) เป็นที่นิยมในการใช้งานมากกว่าแบบการเลือกหน่วย (Unit Selection) เพราะมีขนาดของแบบจำลองเสียงที่เล็กกว่า ซึ่งสามารถทำให้ใช้เทคโนโลยีการสังเคราะห์เสียงบนอุปกรณ์ที่มีทรัพยากรจำกัดได้ และสามารถดัดแปลง (Adaptation) เพื่อเปลี่ยนไปเป็นเสียงต่างๆ ได้

การที่เสียงสังเคราะห์ที่ใช้แบบจำลองแบบฮิดเดนมาร์คอฟ มีขนาดของแบบจำลองเสียงที่เล็กกว่า เนื่องจากไม่ได้เก็บข้อมูลเสียงต้นฉบับ แต่ใช้การประมวลผลทางสัญญาณเพื่อคำนวณค่าตัวแปรที่ใช้แทนเสียงในช่วงเวลาหนึ่ง จากนั้นใช้การสร้างแบบจำลองทางสถิติ เพื่อเก็บลักษณะการเปลี่ยนแปลง และค่าของตัวแปรดังกล่าว

เครื่องมือที่ใช้ในการแปลงสัญญาณเสียงพูดให้กลายเป็นค่าคุณลักษณะเรียกว่าตัวเข้ารหัสเสียง (Vocoder) มีหน้าที่ในการแปลงสัญญาเสียงที่รับเข้ามาเป็นค่าคุณลักษณะ และต้องสามารถแปลงค่าคุณลักษณะดังกล่าว ให้กลับกลายเป็นสัญญาณเสียงได้ด้วย

แบบจำลองแบบฮิดเดนมาร์คอฟจะทำหน้าที่เรียนรู้ค่าลักษณะที่มาจากตัวเข้ารหัสเสียง โดยแบบจำลองแบบฮิดเดนมาร์คอฟจะประกอบด้วยหลายสถานะ ในแต่ละสถานะจะประกอบด้วยแบบจำลองทางสถิติ เช่น แบบจำลองการกระจายตัวแบบเกาส์เซียน ซึ่งแบบจำลองทางสถิติจะทำหน้าที่เรียนรู้ค่าลักษณะบัติดังกล่าว

ในการสร้างเสียงสังเคราะห์ ค่าลักษณะจะถูกสร้างจากแบบจำลองแบบฮิดเดนมาร์คอฟ โดยพิจารณาจากค่าที่ถูกเรียนรู้จากแบบจำลอง ซึ่งในกระบวนนี้ส่งผลให้ค่าลักษณะที่ใช้ในการสร้างเสียงสังเคราะห์มีความคลาดเคลื่อนไปจากต้นฉบับ

ความคลาดเคลื่อนในค่าคุณลักษณะที่สร้างมาจากแบบจำลองแบบฮิดเดนมาร์คอฟ สำหรับใช้ในการสร้างเสียงสังเคราะห์ ส่งผลกับคุณภาพของเสียงสังเคราะห์ในด้านของความเป็นธรรมชาติ และความชัดเจนของเสียงสังเคราะห์

ความเป็นธรรมชาติ และความชัดเจนของเสียงสังเคราะห์ ส่งผลกับความเข้าใจของผู้ฟังเสียงสังเคราะห์ ดังนั้นการปรับปรุงเสียงสังเคราะห์ในประเด็นดังกล่าวจึงเป็นเรื่องที่จำเป็น

ค่าคุณลักษณะที่ได้จากตัวเข้ารหัสเสียงแบ่งออกเป็น 2 ส่วนหลัก คือ ส่วนของสเปกตรัมของสัญญาณเสียง และส่วนของความถี่มูลฐาน ซึ่งทั้งสองส่วนจะถูกจำลองอยู่ในแบบจำลองแบบฮิดเดนมาร์คอฟเดียวกัน แต่อยู่ในกระแส (Stream) ที่แตกต่างกัน ซึ่งในงานวิจัยนี้ได้นำเสนอว่าการออกแบบโครงสร้างของแบบจำลองแบบฮิดเดนมาร์คอฟในลักษณะดังกล่าว ส่งผลให้การสร้างค่าคุณลักษณะได้ประสิทธิภาพไม่ดีเท่าที่ควร

ในงานวิจัยนี้ได้นำเสนอการสร้างแบบจำลองเสียงที่แยกส่วนของสเปกตรัมของสัญญาณเสียงและส่วนของความถี่มูลฐานออกจากกัน และนำเสนอวิธีการในการเชื่อมโยงกันระหว่างแบบจำลองเสียง เพื่อเน้นในการสร้างเสียงสังเคราะห์ ที่มีความเป็นธรรมชาติ และชัดเจนมากกว่าเดิม

1.2 วัตถุประสงค์ของงานวิจัย

ปรับปรุงคุณภาพของเสียงสังเคราะห์ภาษาไทยให้มีคุณภาพมากขึ้นในด้านของความเป็นธรรมชาติ และความชัดเจน

1.3 ขอบเขตงานวิจัย

1. ทำการศึกษาเฉพาะภาษาไทย

2. ใช้ข้อมูลทางบริบทเท่าที่ปรากฏในคลังข้อมูลเสียง TSYNC ที่นำเสนอโดย Hansakunbuntheung, Tesprasit และ Sornlertlamvanich (2003) [1]

3. ใช้วิธีการสังเคราะห์เสียงแบบการสังเคราะห์ค่าพารามิเตอร์ทางสถิติ (Statistical parametric synthesis)

1.4 แนวคิดการวิจัย

1. นำเสนอการปรับเปลี่ยนโครงสร้างของแบบจำลองโครงข่ายประสาทเทียมแบบลึก จากเดิมที่ทำการจำลองกระแสของค่าคุณลักษณะสเปกตรัม และความถี่มูลฐานในแบบจำลองเดียวกัน เปลี่ยนเป็นการใช้สองแบบจำลองในการจำลองกระแสสเปกตรัม และความถี่มูลฐาน เพื่อให้สามารถแบ่งสถานะของหน่วยเสียงได้ดีกว่าการใช้แบบจำลองร่วมกันของทั้งสองกระแส

2. นำเสนอการคำนวณช่วงเวลาของสถานะ จากแบบจำลองช่วงเวลาของหลายแบบจำลอง เพื่อใช้ในการคำนวณช่วงเวลาของสถานะจากแบบจำลองที่นำเสนอในหัวข้อที่ 1

3. นำเสนอค่าคุณลักษณะส่วนรับเข้าของแบบจำลองโครงข่ายประสาทเทียมแบบลึก โดยการใช้ตำแหน่งของใบไม้ของต้นไม้ตัดสินใจในแบบจำลองฮิดเดนมาร์คอฟเป็นค่าคุณลักษณะส่วนรับเข้าของแบบจำลองโครงข่ายประสาทเทียมแบบลึก

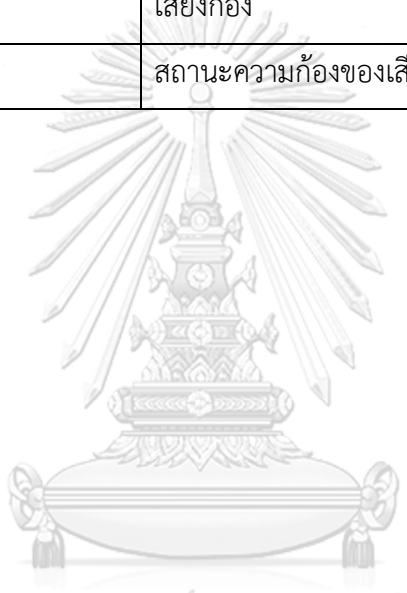
4. นำเสนอการนอร์มัลไลเซชัน (Normalization) ค่าคุณลักษณะส่วนส่งออกของโครงข่ายประสาทเทียมแบบลึก ด้วยการใช้ค่าคะแนนมาตราฐาน ที่ค่ากลาง และค่าความแปรปรวนมาจากแบบจำลองทางสถิติในใบไม้ของของต้นไม้ตัดสินใจในแบบจำลองฮิดเดนมาร์คอฟที่สอดคล้องกับค่าคุณลักษณะทางบริบท

1.5 อภิธานศัพท์

คำศัพท์	ความหมาย
Activation function	ฟังก์ชันกระตุ้น
Adaptation	การปรับเปลี่ยน
Articulatory feature	ค่าคุณลักษณะในการเปล่งเสียง
Band aperiodicity	ความไม่เป็นคาบของแถบความถี่
Binary feature	ค่าคุณลักษณะฐานสอง
Context feature	ค่าคุณลักษณะทางบริบท, ค่าคุณลักษณะบรรยายบริบทหน่วยเสียง
Cross validation	การตรวจสอบแบบไขว้
Denormalization	การลดระดับการนอร์มัลไลเซชัน
Duration model	แบบจำลองช่วงเวลา
Exclusive-or	ออร์เฉพาะ
Expectation maximization	ค่าคาดหวังสูงสุด
Factor-analyzed HMM	แบบจำลองแบบฮิดเดนมาร์คอฟกับปัจจัยที่ถูกวิเคราะห์
Global variance	ความแปรปรวนแบบครอบคลุม
Heuristic rule	กฎแบบศึกษาสำนึก
Hidden Markov Model	แบบจำลองแบบฮิดเดนมาร์คอฟ
Hidden semi-Markov	แบบจำลองประเภทฮิดเดนเซมิมาร์คอฟ

คำศัพท์	ความหมาย
Input feature	ค่าคุณลักษณะส่วนรับเข้า
Keep probability	ความน่าจะเป็นที่จะเก็บโหนด
Learning rate	อัตราการเรียนรู้
Maximum likelihood	ค่าควรจะเป็นสูงสุด
Maximum voiced frequency	ค่าความถี่เสียงก้องสูงสุด
Mel cepstrum coefficients	ค่าสัมประสิทธิ์เมลเคปสตรัม
Minimize generation error	ค่าความผิดพลาดการก่อกำเนิดต่ำสุด
Minimum description length	ความยาวเชิงการพรรณาน้อยสุด
Multi task learning	การเรียนรู้แบบหลายภารกิจ
Multi-space probability distribution HMM	การแจกแจงความน่าจะเป็นแบบหลายปริภูมิของแบบจำลองฮิดเดนมาร์คอฟ
Normal distribution	การแจกแจงแบบปกติ
Normalization	การนอร์มัลไลเซชัน
Objective test	การทดสอบแบบปรนัย
Optimizer	ตัวเพิ่มประสิทธิภาพ
Output feature	ค่าคุณลักษณะส่วนส่งออก
Over-smoothing	การปรับเรียบมากเกินไป
Part of speech	ชนิดของคำ
Positioning feature	ค่าคุณลักษณะระบุตำแหน่ง
Post filter	ตัวกรองภายหลัง
Posteriori probability	ความน่าจะเป็นภายหลัง
Probability density function	ฟังก์ชันความหนาแน่นของความน่าจะเป็น
Source – filter model	แบบจำลองแหล่งจ่าย และตัวกรอง
Speech synthesis	การสังเคราะห์เสียง
State	สถานะ
State-level alignment	การปรับแนวในระดับสถานะ

คำศัพท์	ความหมาย
Statistical parametric synthesis	การสังเคราะห์ค่าพารามิเตอร์ทางสถิติ
Stream	กระแส
Subjective test	การทดสอบแบบอัตนัย
Tri-gram	ไทรแกรม
Unit Selection	การเลือกหน่วย
Vocoder	ตัวเข้ารหัสเสียง
Voiced sound	เสียงก้อง
Voicing Condition	สถานะความก้องของเสียง



บทที่ 2 งานวิจัยที่เกี่ยวข้อง

การสร้างเสียงสังเคราะห์จากแบบจำลองแบบฮิดเดนมาร์คอฟพัฒนาขึ้นโดย Masuko และคณะ (2007) [2-4] อาศัยหลักการของตัวเข้ารหัสเสียงแบบจำลองแหล่งจ่าย และตัวกรอง (Source – filter model) ในการแปลงสัญญาณเสียงเป็นค่าคุณลักษณะที่สามารถแปลงกลับมาเป็นค่าสัญญาณเสียงแบบเดิมได้ เช่น คุณลักษณะค่าสัมประสิทธิ์เมลเคปสตรัม (Mel Cepstrum Coefficients) [5] ใช้ตัวย่อ MCEP

ค่าคุณลักษณะสามารถแบ่งออกเป็น 2 ประเภท ตามลักษณะของค่า ได้แก่ ค่าที่เป็นจำนวนต่อเนื่อง เช่น ค่าของคุณสมบัติ MCEP ที่ได้จากการคำนวณทางคณิตศาสตร์จากสัญญาณเสียง และค่าที่เป็นจำนวนไม่ต่อเนื่อง เช่น ค่าของความถี่มูลฐานของสัญญาณเสียง

ค่าความถี่มูลฐานของสัญญาณเสียง ประกอบด้วยค่า 2 รูปแบบ คือส่วนที่เป็นค่าต่อเนื่อง ได้แก่ ส่วนที่เป็นเสียงก้อง (Voiced sound) ซึ่งในบริเวณดังกล่าวจะปรากฏค่าของความถี่มูลฐาน ซึ่งเกิดจากการสั่นของเส้นเสียง และส่วนที่เสียงไม่ก้อง เช่น ส่วนเริ่มต้นของคำว่า สบาย ซึ่งในบริเวณนี้จะไม่สามารถหาค่าของความถี่มูลฐานได้ รวมถึงเสียงเงียบก็ไม่สามารถคำนวณหาค่าความถี่มูลฐานได้เช่นกัน

กระบวนการฝึกฝนเพื่อสร้างแบบจำลองเสียงสังเคราะห์จากแบบจำลองแบบฮิดเดนมาร์คอฟพัฒนาขึ้นโดย Tokuda และคณะ (2007) [2] ซึ่งใช้เครื่องมือการสร้างแบบจำลองฮิดเดนมาร์คอฟ (Hidden Markov Model Toolkit, HTK) ที่นำเสนอโดย Young และคณะ (2006) [6] เป็นต้นแบบ และมีการนำแบบจำลองที่ใช้แบบการแจกแจงความน่าจะเป็นแบบหลายปริภูมิของแบบจำลองฮิดเดนมาร์คอฟ (Multi-space probability distribution HMM, MSD-HMM) ที่นำเสนอโดย Tokuda และคณะ (1999) [7] มาใช้ในการฝึกฝนค่าคุณลักษณะที่มีค่าไม่ต่อเนื่อง (ค่าคุณลักษณะความถี่มูลฐาน)

ปัญหาที่เกิดขึ้นจากการใช้แบบจำลองแบบฮิดเดนมาร์คอฟ แบ่งออกเป็น 3 ประเด็น ตามที่นำเสนอโดย Zen และคณะ (2009) [8] เป็นดังต่อไปนี้

1. ตัวเข้ารหัสเสียง

เนื่องจากการใช้แบบจำลองฮิดเดนมาร์คอฟจำเป็นต้องทำการแปลงสัญญาณเสียงให้ออกมาเป็นค่าคุณลักษณะ และต้องสามารถแปลงค่าคุณลักษณะดังกล่าวให้กลับเป็นสัญญาณเสียงได้ ซึ่งในกระบวนการดังกล่าวอาจจะส่งผลต่อคุณภาพของเสียงสังเคราะห์ได้

2. ความถูกต้องของการจำลองเสียง

เนื่องจากการใช้แบบจำลองฮิดเดนมาร์คอฟต้องทำการเรียนรู้ค่าคุณลักษณะจากตัวอย่างเสียง และสังเคราะห์ค่าคุณลักษณะเหล่านั้นกลับออกมา ซึ่งในกระบวนการนี้จะทำให้ค่าคุณลักษณะที่สังเคราะห์ออกมาผิดเพี้ยนไปจากเดิม

3. การปรับเรียบมากเกินไป (Over-smoothing)

เนื่องจากการใช้แบบจำลองฮิดเดนมาร์คอฟไม่สามารถจัดเก็บทุกค่าของค่าคุณลักษณะได้ จึงต้องเลือกค่าใดค่าหนึ่งมาใช้เป็นตัวแทนของตัวอย่างเสียงที่อยู่ในแบบจำลองเดียวกัน ซึ่งมักจะเลือกใช้ค่ากลางในชุดข้อมูล ซึ่งส่งผลให้ในขั้นตอนการสังเคราะห์ค่าคุณลักษณะไม่สามารถสังเคราะห์ค่าคุณลักษณะที่มีความเปลี่ยนแปลงมากได้

แนวทางการแก้ปัญหาการปรับเรียบเกินไป แบ่งเป็น 4 แนวทางหลัก ดังนี้

- การปรับแต่งสัญญาณเสียงสังเคราะห์ภายหลังจากสังเคราะห์สัญญาณเสียงแล้ว โดยลดเสียงบริเวณที่เป็นเสียงดังเกินไป และเพิ่มเสียงบริเวณที่ค่อยเกินไป ตัวอย่างเช่นที่เสนอในงานวิจัยของ Yoshimura และคณะ (2005) [9]

- การเพิ่มเติมตัวกรองภายหลัง ดังเช่นในงานวิจัยของ Yoshimaru และคณะ (2001) [10] และงานวิจัยของ Ling และคณะ (2006) [11]

- การใช้แบบจำลองเสียงร่วมกับแบบจำลองอื่น ในรูปแบบของกระบวนการที่ทำเพิ่มเติมในช่วงกระบวนการสังเคราะห์เสียง เพื่อปรับเปลี่ยนให้ค่าคุณลักษณะที่ได้จากการสังเคราะห์ไม่เกิดปัญหาดังกล่าว เช่นที่นำเสนอโดย Qian และคณะ (2008) [12] , Latorre และ Akamine (2008) [13] และการเพิ่มเติมแบบจำลองของความแปรปรวนแบบครอบคลุม (Global Variance) ที่เสนอโดย Toda และคณะ (2005) [14]

- การปรับเปลี่ยนโครงสร้างของแบบจำลองเสียง เช่นที่เสนอในงานวิจัยของ Yang และ Jianhua (2014) [15] ที่ปรับเปลี่ยนให้แบบจำลองฮิดเดนมาร์คอฟแทนหน่วยของเสียงที่ใหญ่กว่าหน่วยเสียง

สำหรับในงานวิจัยนี้ได้ให้ความสำคัญกับปัญหาความถูกต้องของการจำลองเสียงเป็นหลัก และเลือกใช้ตัวเข้ารหัสเสียง STRAIGHT ที่นำเสนอโดย Kawahara และคณะ (2001) [16] ซึ่งเป็นตัวเข้ารหัสเสียงที่มีประสิทธิภาพสูง และเลือกใช้วิธีการแก้ไขปัญหาการปรับเรียบเกินไปด้วยการเพิ่มแบบจำลองความแปรปรวนแบบครอบคลุม ซึ่งค่าของแบบจำลองดังกล่าวนำไปใช้ในกระบวนการปรับแนวค่าคุณลักษณะ เพื่อปรับให้ค่าคุณลักษณะที่สังเคราะห์ออกมามีวิถีการเปลี่ยนแปลงที่มากขึ้น

วิธีการแก้ไขปัญหาความถูกต้องของการจำลองเสียงแบ่งออกเป็น 2 ประเด็นย่อย ดังต่อไปนี้

2.1 ความซับซ้อนของแบบจำลอง

เนื่องจากข้อมูลที่ใช้ในการฝึกฝนแบบจำลองเสียงสังเคราะห์มีอยู่อย่างจำกัด จึงทำให้ไม่สามารถสร้างแบบจำลองที่มีความซับซ้อนในเชิงของจำนวนตัวแปรได้มาก เพราะจะส่งผลให้เสียงที่สังเคราะห์ออกมาในกรณีของข้อมูลที่ไม่เคยพบมาก่อนได้ผลที่แย่ลง

กระบวนการที่ใช้ในการควบคุมความซับซ้อนของแบบจำลอง คือกระบวนการจัดกลุ่มแบบจำลองเสียงแบบขึ้นกับบริบทด้วยการใช้ต้นไม้ตัดสินใจ ที่ใช้หลักทางไวยากรณ์เพื่อสร้างเป็นคำถามในการแบ่งกลุ่มของบริบท โดยต้นไม้ตัดสินใจจะเลือกคำถามเหล่านั้นเป็นจุดต่อของต้นไม้ และทำการสร้างต้นไม้ให้ลึกลงไปเรื่อยๆ จนกว่าค่าของฟังก์ชันฮิวริสติกจะตรงกับเงื่อนไขที่หยุดการสร้างต้นไม้

การใช้ต้นไม้ตัดสินใจเพื่อจัดกลุ่มบริบทถูกใช้ในการสร้างระบบรู้จำเสียง ได้ใช้สมการความควรจะเป็นสูงสุด (Maximum Likelihood) [17] ตามสมการที่ 1 เป็นฟังก์ชันฮิวริสติก เพื่อใช้วัดผลกฎการแบ่งกลุ่มบริบท และเลือกกฎที่ให้ค่าสมการความควรจะเป็นมีค่าสูงสุดเป็นจุดต่อแม่ และทำการวนซ้ำขั้นตอนนี้ต่อไป เพื่อหาค่าจุดต่อในลำดับชั้นที่ลึกลงไป

$$\Delta_q = \{L(S_{q^+}) + L(S_{q^-})\} - L(S) \quad (1)$$

$$L(S) = \sum_{t=1}^{T} \sum_{m \in S} \gamma_m(t) \log(\aleph(o_t, \mu_s, \sigma_s)) \quad (2)$$

S แทนถึงกลุ่มของสถานะที่ต้องการแบ่ง, S_{q^+} แทนกลุ่มของสถานะที่ถูกแบ่งแล้วเป็นกลุ่มใช่, S_{q^-} แทนกลุ่มของสถานะที่ถูกแบ่งแล้วเป็นกลุ่มไม่ใช่, m แทนถึงสถานะของกลุ่มสถานะ S, $\gamma_m(t)$ หมายถึงความน่าจะเป็นภายหลัง (Posteriori Probability), μ_s แทนถึงค่ากลางของกลุ่มสถานะ S, σ_s แทนถึงค่าความแปรปรวนของกลุ่มสถานะ S, o_t แทนถึงตัวอย่างข้อมูลลำดับที่ t และ $\aleph(o_t, \mu_s, \sigma_s)$ แทนถึงฟังก์ชันความหนาแน่นของความน่าจะเป็น (Probability Density Function)

แต่อย่างไรก็ตาม จากพฤติกรรมของสมการความควรจะเป็นสูงสุด จะให้ค่าผลลัพธ์ของสมการที่ 1 มีค่าเป็นบวกเสมอ และการใช้ต้นไม้ตัดสินใจที่มีจำนวนจุดต่อมากเกินไป จะทำให้ได้ความแม่นยำที่ต่ำลง ดังนั้นจึงจำเป็นต้องใช้จุดสิ้นสุดในการแบ่งต้นไม้ตัดสินใจแบบเส้นแบ่งแบบเฉพาะกิจ (Ad Hoc) เช่น การกำหนดจำนวนตัวอย่างฝึกฝนน้อยสุดในจุดต่อใบ โดยข้อเสียของการหยุดการสร้างต้นไม้ด้วยการใช้เส้นแบ่งแบบเฉพาะกิจ คือไม่สามารถใช้ได้เหมาะสมกับทุกชุดข้อมูล

ในงานวิจัย Shinoda และ Watanabe (1997) [18] ได้นำการใช้หลักการความยาวเชิงการพรรณาน้อยสุด (Minimum Description Length) ใช้ตัวอักษรย่อว่า MDL ตามสมการที่ 3 เพื่อช่วย

ให้ค่า Δ_q สามารถเป็นได้ทั้งค่าบวก และลบได้ และเสนอให้ใช้หยุดการแบ่งต้นไม้เมื่อค่า Δ_q เปลี่ยนจากค่าลบเป็นค่าบวก

$$\Delta_q = \{L(S_{q^+}) + L(S_{q^-})\} - L(S) - KM \log W \quad (3)$$

เมื่อ K คือค่าคงที่ที่ใช้ในการถ่วงน้ำหนักในส่วนของความยาวของต้นไม้, M คือจำนวนของกลุ่มต้นไม้ที่ถูกแบ่ง และ W คือค่าผลรวมของการปรากฏของสถานะ (State Occupation) ของทุกๆ สถานะ

จากสมการที่ 3 และสมการที่ 1 พบว่า มีส่วนที่แตกต่างกันคือ การมีพหุนาม $KM \log W$ เพิ่มเข้ามาในสมการที่ 3 ซึ่งเสมือนกับค่าถ่วงน้ำหนักจากส่วนของความยาวของต้นไม้ โดยที่ค่าจากพหุนาม $\{L(S_{q^+}) + L(S_{q^-})\} - L(S)$ จะมีค่าเป็นบวกลดลงเรื่อยๆ เนื่องจากเมื่อแบ่งต้นไม้ให้มีจำนวนจุดต่อมากขึ้น ค่าของความแตกต่างระหว่างก่อนแบ่งจุดต่อ และหลังแบ่งจุดต่อจะลดลง แต่ในทางกลับกันค่าของพหุนาม $KM \log W$ จะมีค่าเป็นบวกมากขึ้น (แต่มีค่าเป็นลบเนื่องจากเครื่องหมายด้านหน้าเป็นลบ) ซึ่งจุดที่เหมาะสมคือจุดที่ทำให้ค่า Δ_q มีค่าเป็นศูนย์ หรือจุดที่เปลี่ยนจากบวกมาเป็นค่าลบ

ได้มีงานวิจัยมากมาย ปรับเปลี่ยนสมการที่ 3 โดยใช้หลักการของการตรวจสอบแบบไขว้ (Cross Validation) เพื่อเพิ่มความถูกต้องให้กับการสร้างต้นไม้ตัดสินใจ ดังเช่นในงานวิจัยของ Zhang และคณะ (2010) [19] ได้ทำการแบ่งชุดข้อมูลในการสร้างต้นไม้ออกเป็นกลุ่ม และใช้หลักการของการตรวจสอบแบบไขว้ในการประมาณค่า $L(S)$ แยกตามแต่ละกลุ่มข้อมูล โดยค่า $L(S)$ จะคิดจากผลรวมของทุกกลุ่มข้อมูลรวมกัน และงานวิจัย Xie และคณะ (2012) [20] ได้ทดลองเปลี่ยนจากการใช้สมการความควรจะเป็นสูงสุดเป็นสมการการสร้างค่าความผิดพลาดการก่อกำเนิดต่ำสุด (Minimize Generation Error, MGE) [21]

นอกจากการปรับเปลี่ยนสมการที่ใช้ในการสร้างต้นไม้ตัดสินใจแล้ว ยังมีแนวคิดการปรับเปลี่ยนโครงสร้างของต้นไม้ เช่นในงานวิจัย Wang และคณะ (2013) [22] ได้ปรับเปลี่ยนให้มีหลายจุดต่อแม่เชื่อมโยงมายังจุดต่อใบไม้ ซึ่งเปรียบเสมือนการสร้างความสัมพันธ์แบบ “หรือ” ในต้นไม้ตัดสินใจ

จากสูตรในสมการที่ 1 แสดงให้เห็นว่า $L(S)$ มีค่าตามจำนวนของข้อมูลที่ใช้ในการฝึกฝน ดังนั้นถ้าในกรณีที่ใช้ค่าคุณลักษณะทางบริบท (Context feature) เหมือนกัน ชุดข้อมูลที่มีข้อมูลฝึกฝนจำนวนมาก และไม่ใช่ข้อมูลซ้ำ จะส่งผลให้ค่าของพหุนาม $\{L(S_{q^+}) + L(S_{q^-})\} - L(S)$ มีค่ามากกว่าชุดข้อมูลที่มีจำนวนฝึกฝนน้อยกว่า และจะทำให้ผลลัพธ์ของต้นไม้มีค่าที่ยาวขึ้น และมีจำนวนของจุดต่อที่มากขึ้นกว่าเดิม ซึ่งการที่มีจุดต่อมากขึ้นสามารถลดความผิดพลาดจากการแบ่งจุดต่อที่ไม่

เหมาะสมอันเนื่องมาจากการนำหน่วยเสียงที่ไม่สอดคล้องกันมารวมอยู่ในจุดต่อเดียวกัน จากเหตุผลข้างต้นทำให้ผลกระทบของความผิดพลาดในกระบวนการสร้างต้นไม้ตัดสินใจมีน้อยลง เมื่อข้อมูลที่ใช้ในการฝึกฝนมีขนาดใหญ่ขึ้น

2.2 โครงสร้างแบบจำลองเสียง

การปรับเปลี่ยนโครงสร้างของแบบจำลองเสียง แบ่งออกเป็น 3 เรื่องหลัก ดังต่อไปนี้

2.2.1 การปรับเปลี่ยนโครงสร้างแบบจำลองเสียงของค่าคุณลักษณะสเปกตรัม

เนื่องจากการใช้แบบจำลองแบบฮิดเดนมาร์คอฟพัฒนาขึ้นโดย Masuko และคณะ (1996) [4] มีปัญหาในการแบ่งสถานะของหน่วยเสียง เพื่อใช้ในขั้นตอนการหาช่วงเวลาของหน่วยเสียง ตามที่เสนอโดย Moore และ Savic (2004) [23] จึงได้ใช้แบบจำลองประเภทฮิดเดนเซมิมาร์คอฟ (Hidden semi-Markov) ขึ้นมาเพื่อแก้ไขปัญหาดังกล่าว ตามที่เสนอโดย Zen และคณะ (2004) [24]

นอกจากการใช้ค่าคุณลักษณะทางบริบทในการสร้างแบบจำลองแล้ว ยังมีแนวคิดการใช้ค่าคุณลักษณะเกี่ยวกับฐานกรณ์ในการเปล่งเสียง (Articulatory features) มาใช้ร่วมในการสร้างแบบจำลองเสียง จึงจำเป็นต้องมีการปรับเปลี่ยนแบบจำลองเสียง เพื่อให้รองรับกับค่าคุณลักษณะดังกล่าว เช่นการใช้แบบจำลองแบบฮิดเดนมาร์คอฟกับปัจจัยที่ถูกวิเคราะห์ (factor-analyzed HMM) ตามที่เสนอโดย Rosti และ Gales (2004) [25]

แต่อย่างไรก็ตามใจงานวิจัยนี้ไม่ได้ปรับปรุงคุณภาพของเสียงสังเคราะห์ด้วยแนวทางนี้

2.2.2 การปรับเปลี่ยนโครงสร้างแบบจำลองเสียงของค่าคุณลักษณะความถี่มูลฐาน

เนื่องจากค่าคุณลักษณะความถี่มูลฐานเป็นค่าแบบไม่ต่อเนื่อง ซึ่งค่าดังกล่าวแบ่งออกเป็น 2 รูปแบบ ได้แก่ ช่วงที่สัญญาณเสียงเป็นเสียงก้อง โดยในช่วงนี้จะมีค่าเป็นค่าจำนวณจริงที่อยู่ในช่วง 60 ถึง 300 เฮิรตซ์ [26] และช่วงที่สัญญาณเสียงไม่ได้เป็นเสียงก้อง เป็นช่วงที่ไม่มีค่าของความถี่มูลฐาน ซึ่งมักจะนำเสนอค่าในช่วงนี้ ด้วยค่าที่ไม่อยู่ในช่วงของเสียงก้อง เช่นค่า -1

วิธีการสร้างแบบจำลองเพื่อให้เรียนรู้ค่าคุณลักษณะดังกล่าว แบ่งออกเป็น 3 แนวคิด ได้แก่ 1) การปรับปรุงแบบจำลองเพื่อให้สามารถจำลองค่าดังกล่าวได้ 2) แนวคิดการเปลี่ยนแปลงการนำเสนอค่าคุณลักษณะดังกล่าวเพื่อให้สามารถนำเสนอค่าความถี่มูลฐานด้วยค่าแบบต่อเนื่อง และ 3) แนวคิดที่แบบผสมที่นำแนวคิดทั้งสองมาใช้รวมกัน

1) แนวคิดการปรับปรุงแบบจำลอง

แบบจำลองที่ใช้การแจกแจงความน่าจะเป็นแบบหลายปริภูมิของแบบจำลองฮิดเดนมาร์คอฟ [7] ถูกนำเสนอเพื่อจำลองค่าที่มีคุณลักษณะที่ไม่ต่อเนื่อง และสามารถนำมาใช้ในการจำลองค่าคุณลักษณะความถี่มูลฐาน แบบจำลองดังกล่าวประกอบด้วยแบบจำลองทางสถิติจำนวน 2 แบบจำลองย่อย แบบจำลองย่อยแรก ใช้ในการตัดสินใจว่าค่าของกลุ่มตัวอย่างที่ฝึกฝนนั้น เป็นค่าแบบต่อเนื่องหรือไม่ต่อเนื่อง เรียกว่าแบบจำลองย่อยสถานะของค่า และแบบจำลองย่อยที่สอง ใช้ในการฝึกฝนจากกลุ่มตัวอย่างที่มีค่าต่อเนื่อง เรียกว่าแบบจำลองย่อยค่าความถี่มูลฐาน

ในการสังเคราะห์ค่าคุณลักษณะความถี่มูลฐานด้วยแบบจำลองที่ใช้การแจกแจงความน่าจะเป็นแบบหลายปริภูมิของแบบจำลองฮิดเดนมาร์คอฟ ค่าคุณลักษณะที่ถูกสังเคราะห์จากสถานะของแบบจำลองดังกล่าว จะต้องถูกพิจารณาว่าเป็นค่าที่ไม่ต่อเนื่อง (เสียงไม่ก้อง) หรือเป็นค่าแบบต่อเนื่อง (เสียงก้อง) โดยพิจารณาจากค่ากลางในแบบจำลองย่อยสถานะของค่า ซึ่งถ้าค่าดังกล่าวมีค่ามากกว่าค่าที่กำหนดไว้ ให้ถือว่าค่าคุณลักษณะที่สังเคราะห์จากสถานะนั้นเป็นแบบค่าต่อเนื่อง (เสียงก้อง) และในกรณีที่มีค่าน้อยกว่า ให้ถือว่าค่าคุณลักษณะที่สังเคราะห์ออกมาเป็นแบบค่าไม่ต่อเนื่อง (เสียงไม่ก้อง)

ในกรณีที่ค่าคุณลักษณะที่ถูกสังเคราะห์ออกมาเป็นแบบต่อเนื่อง (เสียงก้อง) ค่าความถี่มูลฐานจะพิจารณาจากแบบจำลองย่อยความถี่มูลฐาน โดยจะนำค่ากลาง และค่าความแปรปรวนของแบบจำลองย่อยความถี่มูลฐานไปใช้ในการสังเคราะห์ค่าคุณลักษณะ

ด้วยหลักการทำงานของการแจกแจงความน่าจะเป็นแบบหลายปริภูมิของแบบจำลองฮิดเดนมาร์คอฟ จะพบว่าค่าคุณลักษณะของความถี่มูลฐานที่สังเคราะห์ออกมาภายในสถานะเดียวกัน จะต้องเลือกว่าจะเป็นเสียงก้อง หรือเสียงไม่ก้องตลอดทั้งสถานะ ซึ่ง Yu และคณะ (2010) [27] ได้เสนอว่าการจำลองค่าความถี่มูลฐานด้วยแบบจำลองดังกล่าว จะส่งผลให้เกิดเสียงแหบ หรือเสียงรบกวนแบบต่างๆ ขึ้นมา ซึ่งเป็นผลมาจากค่าของความถี่มูลฐานในสถานะเดียวกัน จะต้องเป็นเสียงก้อง หรือไม่ก้องเท่านั้น ทั้งที่ในความเป็นจริงแล้วในสถานะเดียวกัน ค่าของความถี่มูลฐานที่สั่งเคราะห์ออกมาอาจจะไม่ได้เป็นเสียงก้อง หรือไม่ก้องทั้งสถานะ

2) แนวคิดการนำเสนอค่าความถี่มูลฐานด้วยค่าแบบต่อเนื่อง

แนวคิดนี้ใช้การประมาณค่าในช่วงที่ไม่สามารถหาค่าความถี่มูลฐานได้ และหาค่าคุณลักษณะชนิดอื่นเพื่อใช้ในการกำหนดรูปแบบของค่าคุณลักษณะความถี่มูลฐานว่าควรจะเป็นเสียงก้อง หรือเสียงไม่ก้อง

งานวิจัยที่เกี่ยวกับการประมาณค่าความถี่มูลฐานในช่วงที่ไม่ได้เป็นเสียงก้อง ได้แก่ งานวิจัยที่ใช้ฟังก์ชั่นการประมาณค่า นำเสนอโดย Lyche และ Schumaker (1973) [28] การใช้กฎแบบศึกษาสำนึก (Heuristic rule) ร่วมกับฟังก์ชั่นการประมาณค่าที่นำเสนอโดย Garner และคณะ (2013) [29] และการปรับเปลี่ยนวิธีการในการคำนวณหาค่าความถี่มูลฐาน ที่ให้ผลลัพธ์ออกมามากกว่า 1 ค่า

นำเสนอโดย Kawahara และคณะ (1999) [30] ซึ่งจะทำให้ได้ค่าของความถี่มูลฐานในช่วงที่ไม่ใช่เสียงก้องมาด้วย ซึ่งค่านั้นไม่ใช่ค่าที่ถูกต้อง

ค่าคุณลักษณะที่ใช้ในการตัดสินว่าค่าความถี่มูลฐานควรจะเป็นเสียงก้อง หรือเสียงไม่ก้อง ได้แก่ ค่าความไม่เป็นคาบของแถบความถี่ (Band Aperiodicity) นำเสนอในงานวิจัยของ Kawahara, และคณะ (2001) [16] และค่าความถี่เสียงก้องสูงสุด (Maximum voiced frequency) นำเสนอในงานวิจัยของ Drugman และคณะ (2014) [31] ซึ่งค่าทั้งสองค่านั้น เป็นค่าที่ต่อเนื่อง

แบบจำลองแบบฮิดเดนมาร์คอฟใช้ในการจำลองค่าคุณลักษณะความถี่มูลฐาน แบบที่เป็นค่าต่อเนื่อง และจำลองค่าคุณลักษณะที่ใช้ในการตัดสินว่าค่าความถี่มูลฐานควรจะเป็นเสียงก้อง หรือเสียงไม่ก้อง จากนั้นในขั้นตอนการสังเคราะห์ค่าคุณลักษณะ การกำหนดว่าค่าความถี่มูลฐานที่ถูกสังเคราะห์ขึ้นมาจะเป็นเสียงก้องหรือไม่ ขึ้นอยู่กับผลลัพธ์ของค่าคุณลักษณะที่ใช้ในการตัดสินว่าค่าความถี่มูลฐานควรจะเป็นเสียงก้อง หรือเสียงไม่ก้อง ซึ่งจะทำให้ความเป็นเสียงก้อง หรือไม่ใช่เสียงก้อง ไม่ขึ้นอยู่กับสถานะ

3) การนำแนวคิดทั้งสองมาใช้รวมกัน

Yu และคณะ (2010, 2011) [27, 32] เสนอการนำค่าคุณลักษณะทั้งสองที่เสนอในหัวข้อที่ 2 มารวมอยู่ในแบบจำลองเดียวกัน แทนที่จะแยกออกเป็น 2 แบบจำลอง โดยในแบบจำลองจะประกอบด้วย 2 แบบจำลองย่อย คล้ายกับแบบจำลองที่ใช้การกระจายตัวของความน่าจะเป็นแบบหลายปริภูมิของแบบจำลองฮิดเดนมาร์คอฟ แต่ในส่วนของแบบจำลองย่อยสถานะของค่า จะเรียนรู้จากค่าคุณลักษณะที่ใช้ในการตัดสินว่าค่าความถี่มูลฐานควรจะเป็นเสียงก้อง หรือเสียงไม่ก้อง

การใช้แนวคิดในหัวข้อที่ 2 และหัวข้อที่ 3 ได้แสดงให้เห็นว่าสามารถสร้างเสียงสังเคราะห์ ที่มีเสียงแหบ หรือเสียงรบกวนอื่นๆ ลดลงได้ โดยในงานวิจัยเหล่านั้นได้ให้เหตุผลเกี่ยวกับการที่ใช้แนวคิดในหัวข้อที่ 2 และหัวข้อที่ 3 แล้วได้เสียงสังเคราะห์ที่ดีกว่าการใช้แนวคิดที่ 1 ก็เพราะว่า การใช้การกระจายตัวของความน่าจะเป็นแบบหลายปริภูมิของแบบจำลองฮิดเดนมาร์คอฟในแนวคิดที่ 2 จะสังเคราะห์ค่าคุณลักษณะที่เป็นเสียงก้อง หรือไม่เป็นเสียงก้อง เหมือนกันทั้งสถานะ แต่ในทางกลับกันแนวคิดที่ 2 และแนวคิดที่ 3 สามารถสังเคราะห์ค่าคุณลักษณะที่เป็นได้ทั้งเสียงก้อง หรือไม่ใช่เสียงก้องได้ในสถานะเดียวกัน ซึ่งสามารถแก้ไขปัญหาเสียงแหบ หรือเสียงรบกวนอื่นๆ ได้

จากผลสรุปข้างต้นแสดงให้เห็นว่า การสร้างแบบจำลองเสียงสังเคราะห์ด้วยแบบจำลองฮิดเดนมาร์คอฟ ที่เสนอโดย Tokuda และคณะ (1999) [7] และใช้การกระจายตัวของความน่าจะเป็นแบบหลายปริภูมิของแบบจำลองฮิดเดนมาร์คอฟ เพื่อจำลองค่าคุณลักษณะความถี่มูลฐาน ก่อให้เกิดความผิดพลาดในค่าคุณลักษณะที่สังเคราะห์ออกมา ในด้านของความสอดคล้องกันระหว่างค่าคุณลักษณะทางสเปกตรัม และค่าคุณลักษณะของความถี่มูลฐาน เช่นการที่ค่าคุณลักษณะทาง

สเปกตรัมที่สังเคราะห์ออกมาเป็นเสียงก้อง แต่ค่าคุณลักษณะความถี่มูลที่สังเคราะห์ออกมาเป็นช่วงของเสียงไม่ก้อง ผลลัพธ์จะทำให้ได้เสียงสังเคราะห์ที่มีเสียงแหบ หรือเสียงรบกวน

ด้วยลักษณะของเสียงภาษาไทยที่ใช้วรรณยุกต์ในการแบ่งแยกคำ เช่นคำว่า ป่า และป้า โดยทั้งสองคำนั้นมีความหมายที่แตกต่างกัน และออกเสียงต่างกัน แต่มีหน่วยเสียงที่เหมือนกันคือ p aa แต่ที่ต่างกันคือวรรณยุกต์ที่เป็นวรรณยุกต์เอก และโทตามลำดับ

ความแตกต่างของวรรณยุกต์ คือความแตกต่างในค่าคุณลักษณะความถี่มูลฐาน ดังนั้นถ้าค่าคุณลักษณะความถี่มูลฐานที่ถูกสังเคราะห์ออกมาผิดพลาด จะส่งผลให้ผู้ฟังตีความหมายของคำที่ถูกสังเคราะห์ออกมาผิดพลาดไป จึงทำให้ในงานวิจัยได้ให้ความสำคัญกับการสังเคราะห์ค่าคุณลักษณะความถี่มูลฐานให้ถูกต้องยิ่งขึ้น

ในงานวิจัยได้นำเสนอแนวคิดที่แตกต่างจากงานวิจัยของ Yu และคณะ (2010, 2011) [27, 32] ที่นำเสนอว่าความผิดพลาดของการสร้างค่าคุณลักษณะที่ไม่สอดคล้องกันระหว่างค่าคุณลักษณะสเปกตรัม และค่าคุณลักษณะความถี่มูลฐานที่สังเคราะห์ออกมานั้นเกิดจากการใช้การกระจายตัวของความน่าจะเป็นแบบหลายปริภูมิของแบบจำลองฮิดเดนมาร์คอฟ เพราะในกรณีชุดข้อมูลฝึกฝนในสถานะเดียวกัน มีแต่กลุ่มตัวอย่าง ที่เป็นเสียงก้อง หรือเสียงไม่ก้องเพียงอย่างเดียว ค่าคุณลักษณะที่สังเคราะห์ออกมาก็ควรจะมีความสอดคล้องกัน (เป็นเสียงก้อง หรือเสียงไม่ก้อง ในทั้งสองคุณลักษณะ) แต่ในกรณีที่เกิดความผิดพลาดตามตัวอย่างข้างต้นนั้น เป็นผลมาจากตัวอย่างที่ใช้ฝึกฝนในแต่ละสถานะนั้นมีการปะปนกันของตัวอย่างเสียงที่เป็นทั้งเสียงก้อง และเสียงไม่ก้อง

ดังนั้นในงานวิจัยนี้จึงเน้นไปในด้านของการหาแนวทางที่จะแก้ไขปัญหาการแบ่งขอบเขตสถานะโดยยังคงใช้การใช้การกระจายตัวของความน่าจะเป็นแบบหลายปริภูมิของแบบจำลองฮิดเดนมาร์คอฟในการสร้างแบบจำลองเสียงเหมือนเดิม แต่ปรับปรุงในส่วนของการกำหนดสถานะของแบบจำลองฮิดเดนมาร์คอฟให้มีความถูกต้องมากยิ่งขึ้น

2.2.3 แบบจำลองโครงข่ายประสาทเทียมแบบลึก

แบบจำลองโครงข่ายประสาทเทียมแบบลึกถูกนำมาใช้แทนในส่วนการสร้างต้นไม้ตัดสินใจและแบบจำลองแบบฮิดเดนมาร์คอฟ ดังแสดงในงานวิจัยต่อไปนี้

งานวิจัยของ Hashimoto และคณะ (2015) [33], Watts และคณะ (2016) [34], Zen และคณะ (2014, 2013) [35, 36] และ Yao และคณะ (2014) [37] ได้นำเสนอการใช้แบบจำลองโครงข่ายประสาทเทียมแบบลึก โดยการเปลี่ยนคุณลักษณะทางบริบทที่กำกับมาในการถอดความไปเป็นค่าคุณลักษณะในส่วนรับเข้าของการฝึกฝนแบบจำลองโครงข่ายประสาทเทียมแบบลึก และทำการเชื่อมโยงค่าคุณลักษณะดังกล่าวกับค่าคุณลักษณะที่ใช้ในการสังเคราะห์เสียง

การเปลี่ยนคุณลักษณะทางบริบทไปเป็นค่าคุณลักษณะในส่วนรับเข้าทำโดยการสร้างชุดคำถาม (บางส่วนของชุดคำถามเป็นคำถามที่ใช้ในการสร้างต้นไม้ตัดสินใจ) และใช้ผลลัพธ์จากคำถามเป็นค่าของคุณลักษณะ โดยคำตอบของชุดคำถามมี 2 รูปแบบ คือ คำตอบที่อยู่ในรูปแบบของฐานสอง เช่นคำถามว่า บริบทของหน่วยเสียงนั้นเป็นสระหรือไม่ และคำตอบที่อยู่ในรูปแบบของเลขจำนวนจริง เช่นคำถามว่า บริบทของหน่วยเสียงนี้อยู่ในพยางค์ที่เท่าไหร่ของคำ

Zen และคณะ (2013) [35] ได้เสนอว่าการใช้ต้นไม้ตัดสินใจ ไม่สามารถสร้างความสัมพันธ์ระหว่างกฎแต่ละข้อในรูปแบบที่หลากหลายได้ เช่น ความสัมพันธ์แบบออร์เฉพาะ (Exclusive-or) หรือความสัมพันธ์ที่มีการถ่วงน้ำหนักระหว่างแต่ละกฎ และค่าคุณลักษณะทางบริบทที่ไม่เด่นชัด เช่น ค่าคุณลักษณะชนิดของคำ (Part of speech) จะถูกละทิ้งในกระบวนการสร้างต้นไม้ตัดสินใจ ตามที่เสนอโดย Yu และคณะ (2010) [38] (กฎดังกล่าวไม่ได้อยู่ในต้นไม้ตัดสินใจ)

งานวิจัยของ Lu และคณะ (2013) [39] นำเสนอค่าคุณลักษณะส่วนรับเข้าด้วยการทดลองสร้างค่าคุณลักษณะส่วนรับเข้า จากค่าคุณลักษณะทางบริบทที่ใช้หน่วยเสียง ตัวอักษร และค่าคุณลักษณะที่ Lu นำเสนอ ส่วนขั้นตอนอื่นในการสร้างแบบจำลองเสียงด้วยแบบจำลองโครงข่ายประสาทเทียมแบบลึกเหมือนกับงานวิจัยของ Zen และคณะ (2013) [35] แต่อย่างไรก็ตามในงานวิจัยนี้ พบว่าการสร้างแบบจำลองด้วยโครงข่ายประสาทเทียม มีคุณภาพของเสียงสังเคราะห์ที่แย่กว่าการใช้แบบจำลองฮิดเดนมาร์คอฟ และในงานวิจัยนี้ไม่ได้ระบุฐานข้อมูลเสียงที่ใช้ในการทดสอบ

ในงานวิจัยของ Wu และคณะ (2015) [40] ได้มีการปรับปรุงแบบจำลองของโครงข่ายประสาทเทียมแบบลึก โดยใช้การเรียนรู้แบบหลายภารกิจ (Multi task learning) และการใช้ค่าคุณลักษณะที่เป็นผลลัพธ์ของแบบจำลองของโครงข่ายประสาทเทียมแบบตัวเข้ารหัสอัตโนมัติของหลายกรอบเวลา และงานวิจัยของ Wu และคณะ (2016) [41] ได้สร้างชุดฝึกฝนแบบจำลองโครงข่ายประสาทเทียมแบบลึกที่ใช้แบบจำลองแบบ Long Short-Term Memory (LSTM) ที่สามารถนำข้อมูล จากข้อมูลลำดับก่อนหน้าเข้ามาใช้ในการประมวลผลคำตอบของข้อมูลปัจจุบันได้

ในงานวิจัยนี้ได้ให้ความสนใจในผลลัพธ์ในงานวิจัยของ Lu และคณะ (2013) [39] ที่ให้ผลลัพธ์ในทิศทางตรงกันข้ามกับผลของงานวิจัยอื่นๆ ซึ่งข้อแตกต่างระหว่างงานวิจัยของ Lu และคณะ (2013) [39] กับงานวิจัยอื่นๆ คือค่าคุณลักษณะทางบริบทของฐานข้อมูลเสียงที่ในงานวิจัยของ Lu และคณะ (2013) [39] ใช้ค่าคุณลักษณะในส่วนรับเข้าเพียงหน่วยเสียง ซึ่งแตกต่างจากงานวิจัยอื่นๆ ที่ใช้ค่าคุณลักษณะทางเสียงอย่างอื่นเพิ่มเติมเข้ามา

สำหรับฐานข้อมูลเสียงภาษาไทยสำหรับการฝึกฝนเสียงสังเคราะห์ ได้แก่ ฐานข้อมูลเสียง TSYNC นำเสนอโดย Hansakunbuntheung และคณะ (2003) [1] ซึ่งมีค่าคุณลักษณะทางบริบท คือ ตำแหน่งของคำในประโยค ชนิดของคำ หน่วยเสียง และวรรณยุกต์ โดยค่าคุณลักษณะชนิดของคำในภาษาไทยไม่ได้มีการกำหนดชนิดที่แน่นอน และรูปแบบการเขียนของภาษาไทยที่ไม่ได้มีการแยกคำ

อย่างชัดเจน ทำให้การการแบ่งคำออกจากประโยคมีความกำกวม เช่นคำว่า “ภาษาไทย” อาจถือว่าเป็น 1 คำ หรือ 2 คำที่ประกอบกันของคำว่า ภาษา และ ไทย ซึ่งความไม่แน่นอนในประเด็นดังกล่าวทำให้ในการใช้งานจำเป็นต้องใช้ส่วนระบุชนิดของคำ และส่วนตัดคำที่สอดคล้องกับลักษณะของฐานข้อมูลเสียง

จากข้อกำหนดดังกล่าวทำให้ค่าคุณลักษณะทางบริบทที่สามารถนำมาใช้ได้ มีเพียงหน่วยเสียงและวรรณยุกต์เท่านั้น ซึ่งอาจจะส่งผลให้การใช้แบบจำลองโครงข่ายประสาทเทียมแบบลึกตามรูปแบบที่เสนอในงานวิจัยของ Zen และคณะ (2013) [35] ได้ผลลัพธ์ที่แย่กว่าการใช้แบบจำลองฮิดเดนมาร์คอฟ ดังที่แสดงในงานวิจัยของ Lu และคณะ. (2013) [39]

ดังนั้นในงานวิจัยนี้จึงเน้นในการปรับปรุงกระบวนการในการใช้แบบจำลองโครงข่ายประสาทเทียมแบบลึก เพื่อให้สามารถนำไปใช้กับฐานข้อมูลเสียงสังเคราะห์ภาษาไทย ในการสร้างแบบจำลองเสียงสังเคราะห์ที่ให้คุณภาพของเสียงสังเคราะห์ที่ดีกว่ากระบวนการที่นำเสนอในงานวิจัยของ Zen และคณะ (2013) [35]

บทที่ 3 ทฤษฎีที่เกี่ยวข้อง

ในส่วนนี้จะนำเสนอทฤษฎีที่เกี่ยวข้องกับงานวิทยานิพนธ์ฉบับนี้ ซึ่งประกอบด้วยทฤษฎีที่เกี่ยวข้องกับการสร้างระบบสังเคราะห์เสียงด้วยแบบจำลองฮิดเดนมาร์คอฟ กระบวนการสังเคราะห์เสียงด้วยแบบจำลองฮิดเดนมาร์คอฟ ตัวเข้ารหัสเสียง STRAIGHT นำเสนอโดย Kawahara et al. (2001) [16] กระบวนการสร้างแบบจำลองเสียงสังเคราะห์ด้วยแบบจำลองโครงข่ายประสาทเทียมแบบลึก และกระบวนการสังเคราะห์เสียงด้วยแบบจำลองโครงข่ายประสาทเทียมแบบลึก

3.1 ตัวเข้ารหัสเสียง STRAIGHT

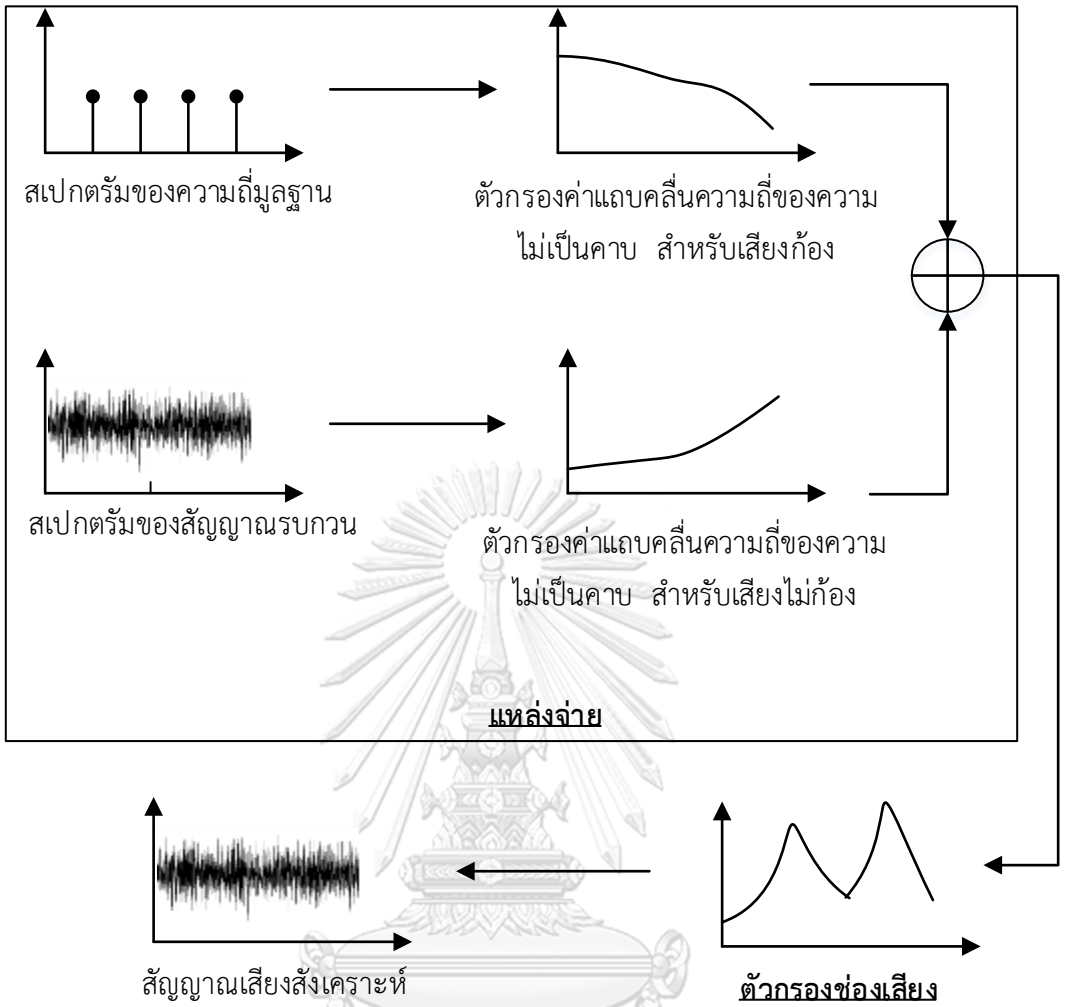
ตัวเข้ารหัสเสียง STRAIGHT ใช้ในการแปลงสัญญาณเสียงรับเข้า เป็นค่าคุณลักษณะ โดยค่าคุณลักษณะที่ STRAIGHT ใช้เพื่อสร้างสัญญาณเสียง แสดงดังรูปที่ 1 ซึ่งประกอบด้วยส่วนของแหล่งจ่าย และส่วนของตัวกรองสัญญาณเสียง

ส่วนของแหล่งจ่ายเปรียบเสมือนกับกล่องเสียง ซึ่งจะมีรูปแบบการทำงานอยู่ 2 รูปแบบ ได้แก่ เส้นเสียงสั่น และเส้นเสียงเปิดกว้าง ในกรณีที่เส้นเสียงสั่น จะทำให้เกิดค่าความถี่มูลฐานขึ้นมา ซึ่งในกรณีนี้จะแทนด้วยสเปกตรัมของความถี่มูลฐาน และในกรณีที่เส้นเสียงเปิดกว้างจะทำให้ลมจากปอดไหลผ่านหลอดลมออกมา ซึ่งมีลักษณะคล้ายกับสัญญาณเสียงรบกวนจึงแทนด้วยสเปกตรัมของสัญญาณรบกวน

เนื่องจากในทุกช่วงของแถบคลื่นความถี่ของสัญญาณเสียงไม่ได้ประกอบด้วยส่วนที่เป็นคาบทั้งหมด ดังนั้นจึงจำเป็นต้องมีตัวกรองค่าแถบคลื่นความถี่ของความไม่เป็นคาบขึ้นมา ดังแสดงในรูปที่ 1 โดยในกรณีที่เสียงเป็นเสียงก้อง ตัวกรองนี้จะสังวัตนาการกับสเปกตรัมของความถี่มูลฐาน และส่วนกลับของตัวกรองนี้จะสังวัตนาการกับสเปกตรัมของสัญญาณเสียงรบกวน จากนั้นจะนำผลลัพธ์ของทั้ง 2 ส่วนมารวมกัน เพื่อให้ได้สัญญาณเสียงของแหล่งจ่ายที่มีการปนกันของสัญญาณที่เป็นคาบ และไม่เป็นคาบในแต่ละส่วนของแถบคลื่นความถี่สัญญาณ

สำหรับในกรณีที่เป็นเสียงไม่ก้อง จะไม่มีส่วนของสเปกตรัมความถี่มูลฐาน ส่งผลให้สัญญาณเสียงของแหล่งจ่ายเป็นเสียงที่คล้ายกับสัญญาณรบกวน

สัญญาณเสียงจากส่วนของแหล่งจ่ายจะสังวัตนาการกับตัวกรองช่องเสียง ซึ่งตัวกรองช่องเสียง ทำหน้าที่ในการจำลองถึงการจัดเรียงตัวของอวัยวะภายในช่องปาก ซึ่งผลลัพธ์จากในส่วนนี้คือสัญญาณเสียงสังเคราะห์



รูปที่ 1 ส่วนประกอบของตัวเข้ารหัสเสียง STRAIGHT

ตัวเข้ารหัสเสียงมีหน้าที่ในการแปลงเสียงเป็นค่าคุณลักษณะ โดยค่าคุณลักษณะนั้นจะแทนถึงตัวกรอง และสเปกตรัมต่างๆ ในรูปที่ 1 และในกรณีที่ต้องการสร้างเสียงสังเคราะห์ ตัวเข้ารหัสเสียงก็จะทำหน้าที่ในการนำค่าคุณลักษณะแหล่งนั้น แปลงกลับมาเป็นตัวกรอง และสเปกตรัมเพื่อใช้ในสร้างเสียงสังเคราะห์

รายละเอียดของค่าคุณลักษณะที่ได้จากตัวเข้ารหัสเสียง STRAIGHT ประกอบด้วย

1. ตัวกรองช่องเสียง ซึ่งตัวกรองนี้จะถูกลดมิติลง โดยอาศัยหลักการของคุณลักษณะค่าสัมประสิทธิ์เมลเคปสตรัม เพื่อใช้ในการประมาณค่าตัวกรอง ซึ่งตามที่เสนอโดย Black และคณะ. (2007) [2] ตัวกรองของเสียงจะใช้ค่าคุณลักษณะจำนวน 35 ข้อมูล และเรียกคุณสมบัตินี้ว่า MCEP

2. ค่าความถี่มูลฐาน เนื่องจากค่าความถี่มูลฐาน แบ่งออกเป็น 2 ลักษณะ ได้แก่ ช่วงที่เป็นเสียงก้องจะมีค่าเป็นค่าความถี่มูลฐาน และช่วงที่ไม่เป็นเสียงก้องค่าความถี่มูลฐานจะไม่สามารถหาค่าได้ ซึ่งจะแทนด้วยค่าที่ไม่อยู่ในช่วงของความถี่มูลฐาน เรียกคุณสมบัตินี้ว่า f_0

3. ตัวกรองค่าแถบคลื่นความถี่ของความไม่เป็นคาบ ซึ่งตัวกรองนี้จะปรากฏอยู่ในส่วนของสเปกตรัมของความถี่มูลฐาน และสเปกตรัมของสัญญาณเสียงรบกวน แต่ตัวกรองค่าแถบคลื่นความถี่ของความไม่เป็นคาบที่ใช้สังวัตนาการกับสเปกตรัมของความถี่มูลฐาน และสเปกตรัมของสัญญาณเสียงรบกวนเป็นค่าผกผันกัน จึงสามารถใช้เพียงตัวกรองเดียวได้ ซึ่งตัวกรองนี้จะถูกลดมิติลงโดยอาศัยหลักการของคุณลักษณะค่าสัมประสิทธิ์เมลเคปสตรัม เพื่อใช้ในการประมาณค่าตัวกรอง ซึ่งตามที่เสนอโดย Black และคณะ. (2007) [2] ตัวกรองค่าความไม่เป็นคาบของแถบความถี่จะใช้ค่าคุณลักษณะจำนวน 25 ข้อมูล และเรียกคุณสมบัตินี้ว่า BAP

3.2 การสร้างแบบจำลองเสียงสังเคราะห์ด้วยแบบจำลองฮิดเดนมาร์คอฟ

การสร้างแบบจำลองเสียงสังเคราะห์ที่นำเสนอในส่วนนี้ จะทำตามขั้นตอนที่นำเสนอโดย Black และคณะ. (2007) [2] ซึ่งแบ่งเป็นส่วนต่างๆ ตามรูปที่ 2 โดยรายละเอียดในแต่ละส่วนเป็นดังต่อไปนี้

1. <u>ส่วนการสกัดค่าคุณลักษณะ</u>

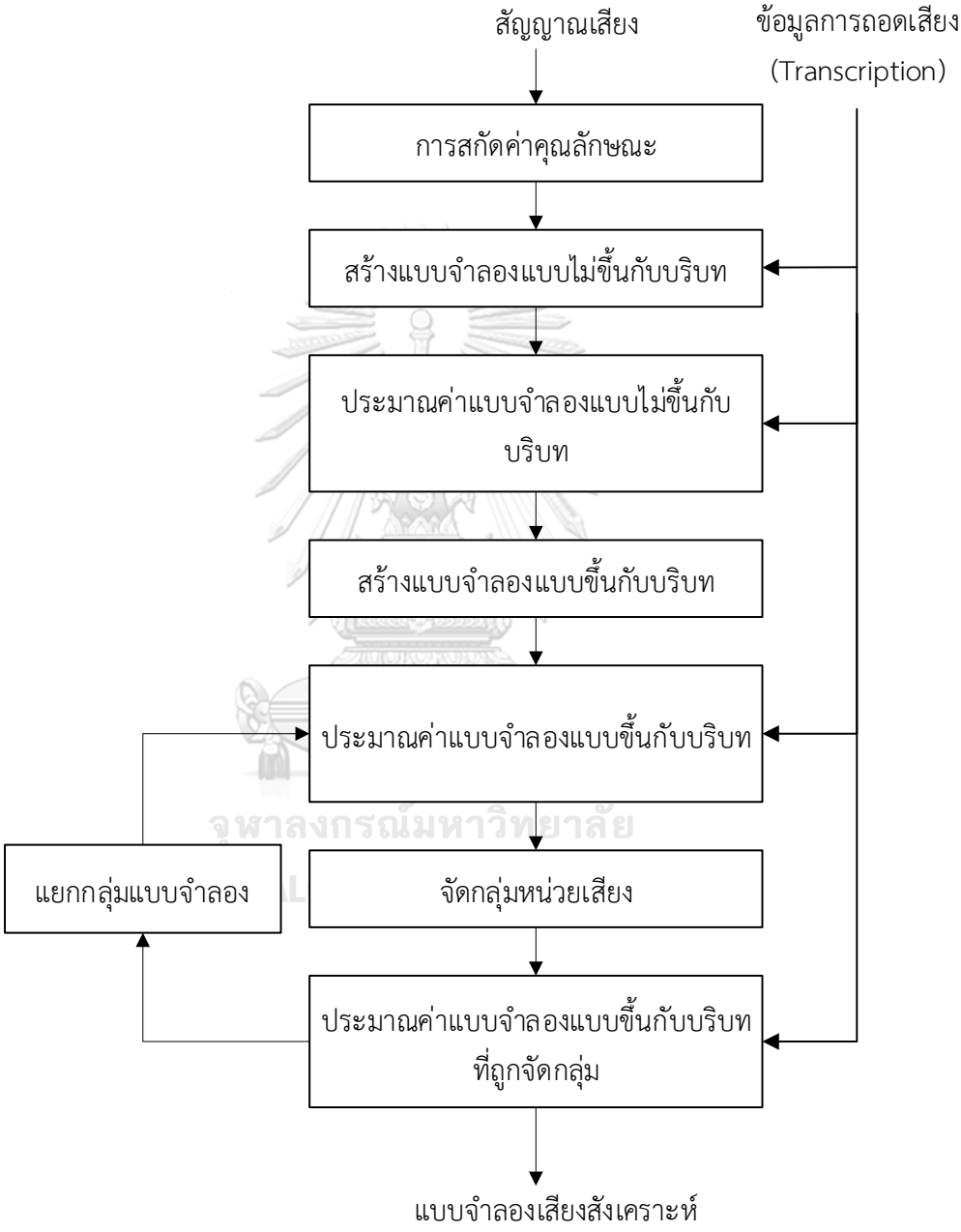
การสกัดค่าคุณลักษณะที่ใช้ในงานวิจัยนี้ จะใช้ตัวเข้ารหัสเสียงที่เรียกว่า STRAIGHT Kawahara และคณะ. (2001) [16] ซึ่งค่าคุณลักษณะที่ใช้แทนเสียงที่เป็นผลลัพธ์ของ STRAIGHT แบ่งเป็น 3 ส่วน ได้แก่ 1) ค่าคุณลักษณะที่ใช้แทนสเปกตรัมของเสียง เรียกว่า MCEP 2) ค่าคุณลักษณะความถี่มูลฐาน เรียกว่า f_0 และ3) ค่าคุณลักษณะค่าความไม่เป็นคาบของแถบความถี่ เรียกว่า BAP

2. <u>การสร้างแบบจำลองแบบไม่ขึ้นกับบริบท</u>

แบบจำลองที่ใช้ในการสร้างเสียงสังเคราะห์เป็นตาม รูปที่ 3 ซึ่งเป็นแบบจำลองฮิดเดนมาร์คอฟ โดยในหนึ่งแบบจำลองจะแบ่งออกเป็น 5 สถานะ (State) และในแต่ละสถานะจะประกอบด้วย 4 กระแส (Stream) ตามที่เสนอโดย Black และคณะ. (2007) [2] ซึ่งแต่ละกระแสจะประกอบด้วย

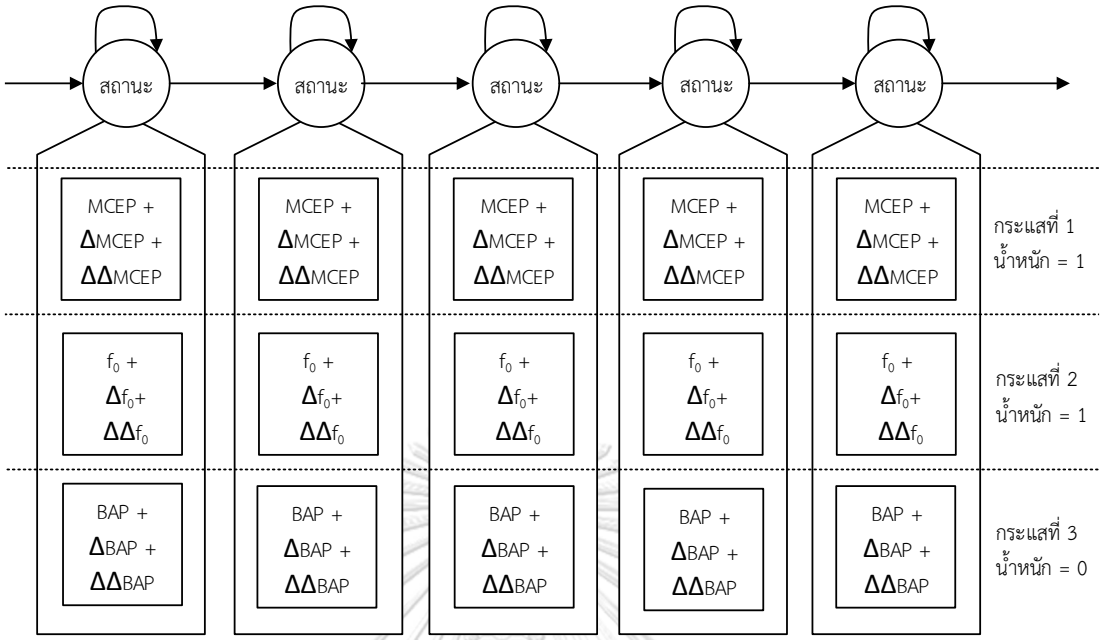
- กระแสของสเปกตรัม ประกอบด้วย ค่าคุณลักษณะสเปกตรัม รวมถึงค่าความเร็ว ค่าความเร่งของคุณลักษณะสเปกตรัม เรียกว่า MCEP, ΔMCEP และ $\Delta\Delta$MCEP ตามลำดับ
- กระแสของความถี่มูลฐาน ประกอบด้วย ค่าคุณลักษณะความถี่มูลฐาน รวมถึงค่าความเร็ว ค่าความเร่งของคุณลักษณะความถี่มูลฐาน แทนด้วยสัญลักษณ์ f_0, Δf_0, $\Delta\Delta f_0$

- กระแสของค่าคุณลักษณะค่าความไม่เป็นคาบของแถบความถี่ รวมถึงค่าความเร็วค่าความเร่งของคุณลักษณะค่าความไม่เป็นคาบของแถบความถี่ เรียกว่า BAP, ΔBAP และ ΔΔBAP ตามลำดับ



รูปที่ 2 ขั้นตอนการสร้างแบบจำลองเสียง

แบบจำลองเสียงแบบฮิดเดนมารคอฟสำหรับสเปตรัม และความถี่มูลฐาน



รูปที่ 3 แบบจำลองเสียงแบบฮิดเดนมาร์คอฟสำหรับสเปกตรัม และความถี่มูลฐานที่ใช้ในการสร้างเสียงสังเคราะห์

สำหรับแต่ละสถานะ จะประกอบด้วย แบบจำลองทางสถิติที่ใช้คำนวณค่าความน่าจะเป็นจากฟังก์ชันความหนาแน่นของความน่าจะเป็น (Probability Density Function, PDF) เรียกโดยย่อว่าแบบจำลอง PDF และมีจำนวนเท่ากับจำนวนของค่าคุณลักษณะในกระแสนั้นๆ เพื่อใช้จำลองลักษณะการกระจายตัวของค่าคุณลักษณะ โดยแบบจำลองทางสถิติที่นิยมใช้กัน ได้แก่ แบบจำลองการแจกแจงแบบปกติ (Normal Distribution) โดยตัวแปรที่จำเป็นในการจำลอง ได้แก่ ค่ากลาง (ค่าเฉลี่ย) และค่าความแปรปรวน

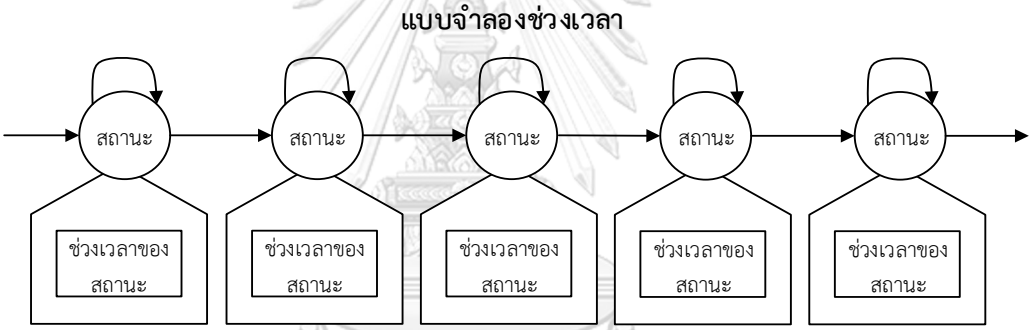
โดยค่าความน่าจะเป็นที่เป็นผลลัพธ์ของแต่ละสถานะ คิดจากผลรวมของความน่าจะเป็นที่ถูกปรับน้ำหนักของแต่ละกระแส โดยค่าตั้งต้นที่กำหนดโดย Black และคณะ. (2007) [2] กำหนดให้ค่าน้ำหนักของกระแสที่ 1 และกระแสที่ 2 มีค่าเป็น 1 และกระแสที่ 3 มีค่าเป็น 0 ซึ่งหมายความว่าค่าความน่าจะเป็นที่เป็นผลลัพธ์ของแต่ละสถานะจะไม่นำค่าความน่าจะเป็นของกระแสที่ 3 เข้ามารวมอยู่ด้วย

การสร้างแบบจำลองเริ่มต้นทำโดยการหาค่าเฉลี่ย และค่าความแปรปรวนแยกตามแต่ละหน่วยเสียงในแต่ละค่าคุณลักษณะ จากตัวอย่างฝึกฝนทั้งหมด เพื่อนำมาใช้เป็นค่าเริ่มต้นของแบบจำลองเริ่มต้นแบบไม่ขึ้นกับบริบท

3. <u>การประมาณค่าแบบจำลองฮิดเดนมาร์คอฟแบบไม่ขึ้นกับบริบท</u>

แบบจำลองที่ไม่ขึ้นกับบริบท จะใช้หนึ่งแบบจำลอง PDF ต่อหนึ่งหน่วยเสียง การประมาณค่าแต่ละแบบจำลองฮิดเดนมาร์คอฟ โดยใช้วิธีค่าคาดหวังสูงสุด (Expectation maximization) [17] จากตัวอย่างที่มีชนิดของหน่วยเสียงตรงกับชนิดแบบจำลองฮิดเดนมาร์คอฟ เพื่อทำการประมาณค่าที่ควรจะเป็นของแต่ละสถานะ

ในขั้นตอนนี้จะทำการสร้างแบบจำลองฮิดเดนมาร์คอฟของช่วงเวลา (Duration model) ของแต่ละสถานะในหน่วยเสียง โดยหนึ่งแบบจำลองเสียงที่เก็บค่าสเปกตรัมและความถี่มูลฐาน จะมีแบบจำลองช่วงเวลาควบคู่กันไป โครงสร้างแบบจำลองฮิดเดนมาร์คอฟของช่วงเวลาเป็นไปตามรูปที่ 4 โดยจะประกอบด้วย 5 สถานะ เท่ากับจำนวนสถานะแบบจำลองเสียง และจะมีเพียง 1 แบบจำลอง PDF ต่อ 1 สถานะ (ทุกกระแสมีระยะเวลาเหมือนกัน) โดยแบบจำลองช่วงเวลาจะถูกประมาณค่าพร้อมกับการประมาณค่าแบบจำลองเสียง



รูปที่ 4 แบบจำลองฮิดเดนมาร์คอฟของช่วงเวลา

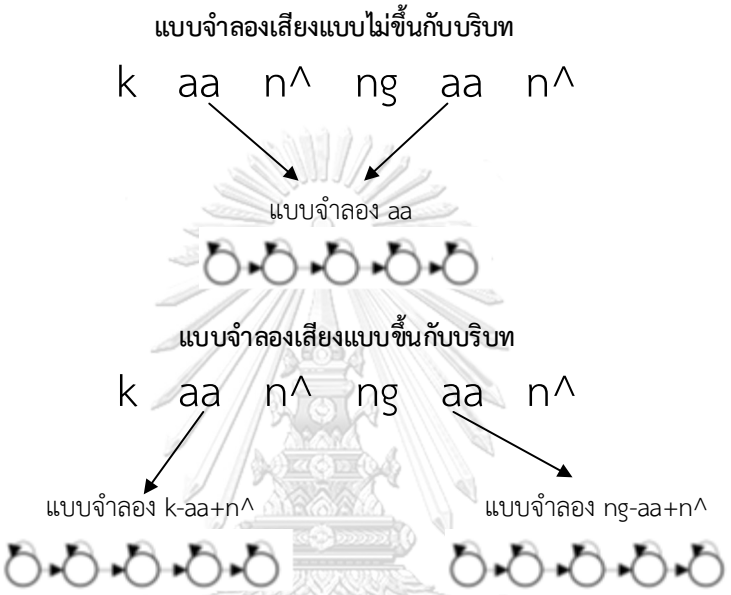
4. <u>การสร้างแบบจำลองฮิดเดนมาร์คอฟแบบขึ้นกับบริบท</u>

แบบจำลองฮิดเดนมาร์คอฟแบบขึ้นกับบริบทคือการใช้แบบจำลองฮิดเดนมาร์คอฟหลายแบบจำลองในการจำลองหน่วยเสียงเดียวกัน โดยอาศัยบริบทรอบข้าง หรือค่าคุณลักษณะทางบริบท (Context Feature) เช่น หน่วยเสียงด้านหน้า และหลังในการแบ่งแยกแบบจำลอง ดังตัวอย่างในรูปที่ 5 ที่หน่วยเสียง aa ใช้แบบจำลองเสียงแบบไม่ขึ้นกับบริบทเพียงแบบจำลองเดียว แต่ถ้าในกรณีของแบบจำลองเสียงที่ขึ้นกับบริบท หน่วยเสียง aa จะถูกแบ่งย่อยตามคุณลักษณะทางบริบท จากในตัวอย่างที่ใช้ค่าคุณลักษณะทางบริบทเพียงหน่วยเสียงก่อนหน้า และด้านหลัง จึงทำให้ต้องใช้แบบจำลองเสียง k-aa+ng^ และแบบจำลองเสียง ng-aa+n^ สำหรับจำลองหน่วยเสียง aa

การสร้างแบบจำลองแบบขึ้นกับบริบท ทำโดยการคัดลอกแบบจำลองแบบไม่ขึ้นกับบริบทที่มีหน่วยเสียงหลักเดียวกัน ทั้งในส่วนของแบบจำลองสำหรับสเปกตรัม และความถี่มูลฐาน และ

แบบจำลองช่วงเวลา ดังเช่นตัวอย่างในรูปที่ 5 โดยค่าเริ่มต้นของแบบจำลองเสียง k-aa-n^ และ ng-aa+n^ มาจากแบบจำลองเสียง aa

แต่อย่างไรก็ตาม คุณลักษณะทางบริบท สำหรับการสร้างแบบจำลองเสียงสังเคราะห์มีจำนวนมาก เช่นหน่วยเสียงก่อนหน้า และด้านหลังจำนวน 3 หน่วยเสียง จึงทำให้จำนวนแบบจำลองเสียงแบบขึ้นกับบริบทที่ใช้ในการสร้างแบบจำลองเสียงสังเคราะห์จำนวนมาก



รูปที่ 5 ตัวอย่างแบบจำลองแบบขึ้นกับบริบท และไม่ขึ้นกับบริบท

5. <u>การประมาณค่าแบบจำลองแบบขึ้นกับบริบท</u>

การประมาณค่าแบบจำลองแบบขึ้นกับบริบท ทำโดยการนำแบบจำลองฮิดเดนมาร์คอฟที่ได้จากกระบวนการที่ 4 มาทำการฝึกฝนโดยใช้วิธีเดียวกันกับการประมาณค่าแบบจำลองแบบไม่ขึ้นกับบริบท โดยใช้วิธีค่าคาดหวังสูงสุด จากตัวอย่างที่มีชนิดของหน่วยเสียง และค่าคุณลักษณะทางบริบทตรงกับชนิดของแบบจำลอง

แบบจำลองช่วงเวลาจะถูกฝึกฝนไปพร้อมกับแบบจำลองเสียงเช่นเดียวกัน

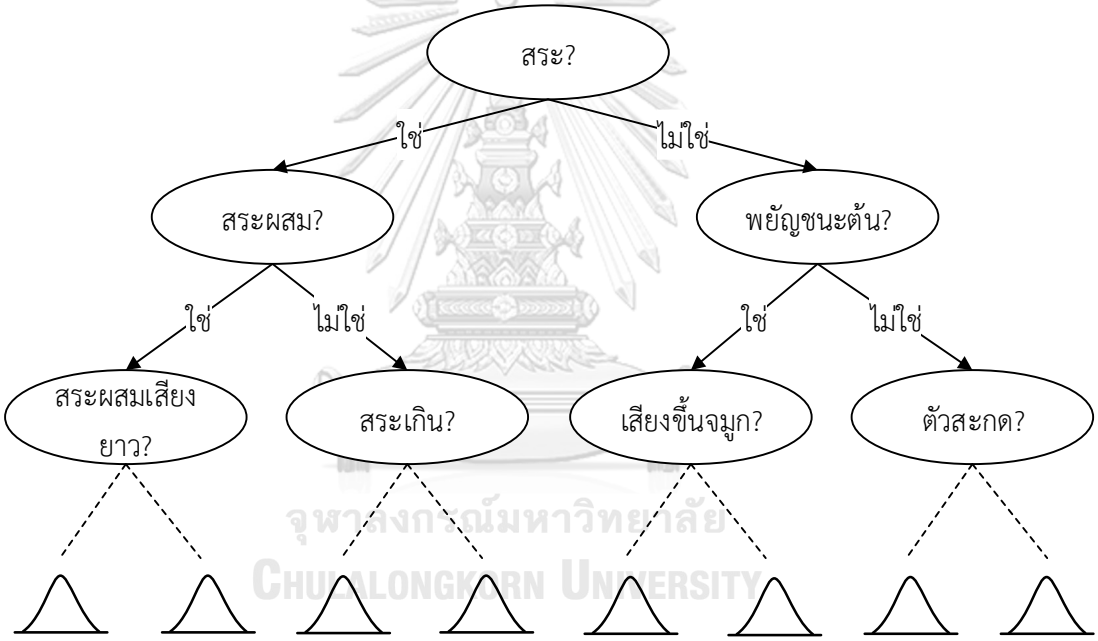
6. <u>การจัดกลุ่มหน่วยเสียง</u>

การจัดกลุ่มหน่วยเสียงจะกระทำแยกตามสถานะ และกระแสของค่าคุณลักษณะที่อยู่ในแบบจำลองเสียง โดยใช้วิธีการสร้างต้นไม้ตัดสินใจ โดยใช้กฎทางไวยากรณ์เพื่อสร้างกฎการจัดกลุ่ม โดยกฎที่สร้างขึ้นมานั้น ใช้เป็นจุดต่อในต้นไม้ตัดสินใจ ตามรูปที่ 6 โดยแต่ละจุด

ต่อจะมีคำตอบที่เป็นไปได้ คือ อยู่ และไม่อยู่ในกลุ่มที่กฎได้นิยามไว้ โดยจุดต่อที่เป็นใบจะแทนถึงฟังก์ชันความหนาแน่นของความน่าจะเป็น (PDF)

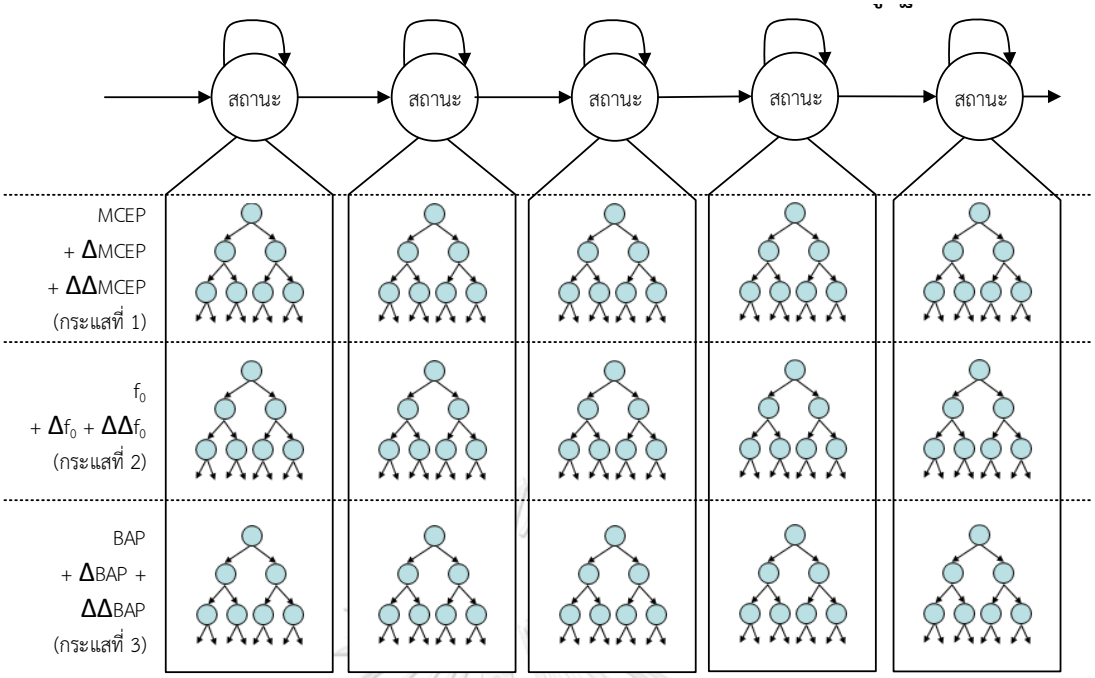
การสร้างต้นไม้ตัดสินใจจะทำในลักษณะของการเรียกซ้ำโดยเริ่มต้นจากจุดต่อบนสุด โดยการทำการวัดค่าความสามารถในการแบ่งข้อมูลทั้งหมดจากกฎทั้งหมดด้วยฮิวริสติกฟังก์ชัน จากนั้นเลือกกฎที่ได้ค่าจากฮิวริสติกฟังก์ชันที่ดีที่สุดเป็นกฎในจุดต่อนั้น จากนั้นทำการแบ่งข้อมูลออกเป็น 2 ส่วน คือ ส่วนที่อยู่ในกฎนั้น และส่วนที่ไม่อยู่ในกฎนั้น จากนั้นทำกระบวนการนี้ซ้ำไปเรื่อยๆ ในกลุ่มข้อมูลที่ถูกแบ่งออกมาแล้ว

ฮิวริสติกฟังก์ชันที่นิยมใช้ในการสร้างต้นไม้ตัดสินใจสำหรับการสังเคราะห์เสียง คือ สมการ MDL เป็นตามสมการที่ 3



รูปที่ 6 ตัวอย่างต้นไม้ตัดสินใจของแบบจำลองเสียง

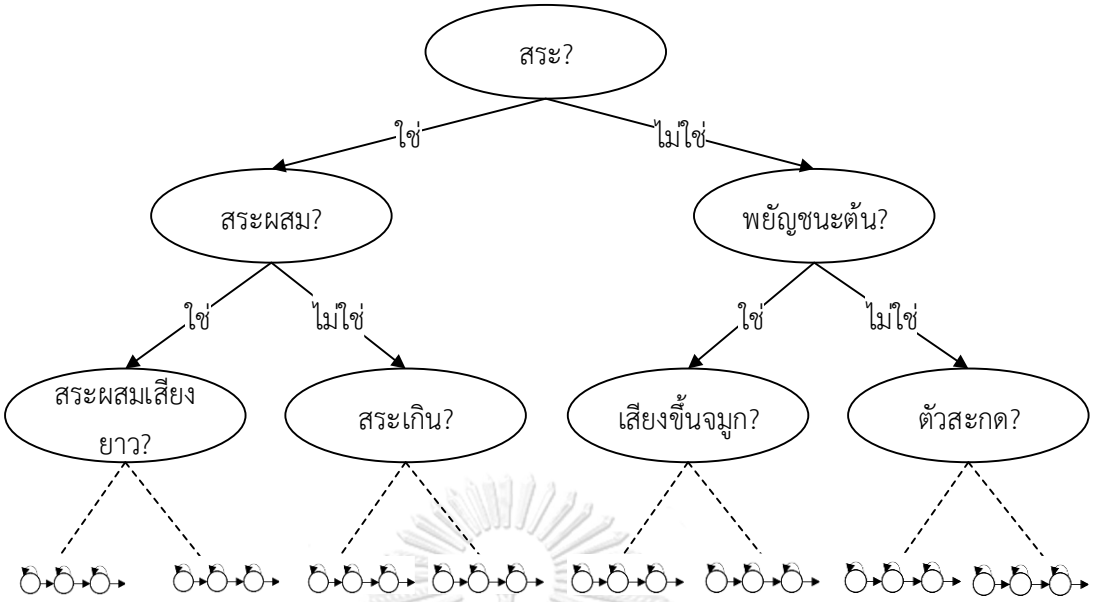
สำหรับแบบจำลองเสียงใช้ต้นไม้ตัดสินใจจำนวน 3 ต้น ต่อ 1 สถานะ โดยต้นแรกใช้ในการจัดกลุ่มกระแสสเปกตรัมในกระแสที่หนึ่ง ต้นที่สองใช้ในการจัดกลุ่มกระแสของความถี่มูลฐานในกระแสที่สอง และต้นที่สามใช้ในการจัดกลุ่มกระแสค่าคุณลักษณะค่าความไม่เป็นคาบของแถบความถี่ในกระแสที่สาม โดยในใบไม้ของต้นไม้ตัดสินใจที่อยู่ในแบบจำลองเสียง จะประกอบด้วยฟังก์ชันความหนาแน่นของความน่าจะเป็นเพียง 1 ค่าเท่านั้น โดยสรุปแล้วสำหรับแบบจำลองช่วงเวลาจะใช้ต้นไม้ถึง 15 ต้น ดังแสดงในรูปที่ 7



รูปที่ 7 ต้นไม้ตัดสินใจของแบบจำลองเสียง

การออกแบบโครงสร้างของต้นไม้ตัดสินใจในรูปแบบนี้ส่งผลให้ในแต่ละสถานะมีโครงสร้างของต้นไม้ที่ไม่เหมือนกัน ซึ่งส่งผลให้วิธีการจัดกลุ่มในแต่ละสถานะไม่เหมือนกัน เช่นตามตัวอย่างในรูปที่ 5 หน่วยเสียง k-aa+n^ และหน่วยเสียง ng-aa+n^ สามารถที่จะถูกจัดกลุ่มให้ใช้ค่า PDF ที่แตกต่างกันในสถานะแรก แต่ให้อยู่ในกลุ่มเดียวกันในสถานะที่ 2 ได้

สำหรับแบบจำลองช่วงเวลาจำเป็นต้องถูกลดจำนวนของแบบจำลองลงเช่นเดียวกัน แต่โครงสร้างของต้นไม้ตัดสินใจมีความแตกต่างกับแบบจำลองเสียง เพราะใบไม้ของต้นไม้ตัดสินใจของแบบจำลองช่วงเวลาแทนถึงทุกๆ สถานะของแบบจำลองช่วงเวลา ดังแสดงตัวอย่างในรูปที่ 8 ซึ่งแตกต่างจากแบบจำลองเสียงที่ในใบไม้ของต้นไม้ตัดสินใจ จัดเก็บฟังก์ชันความหนาแน่นของความน่าจะเป็นของหนึ่งสถานะเท่านั้น จึงส่งผลให้แบบจำลองช่วงเวลาใช้ต้นไม้ตัดสินใจเพียง 1 ต้นเท่านั้นแทนที่จะต้องใช้จำนวนต้นไม้ตัดสินใจเท่ากับจำนวนของสถานะของแบบจำลอง



รูปที่ 8 ตัวอย่างต้นไม้ตัดสินใจของแบบจำลองช่วงเวลา

7. <u>การประมาณค่าแบบจำลองแบบขึ้นกับบริบทที่ถูกจัดกลุ่ม</u>

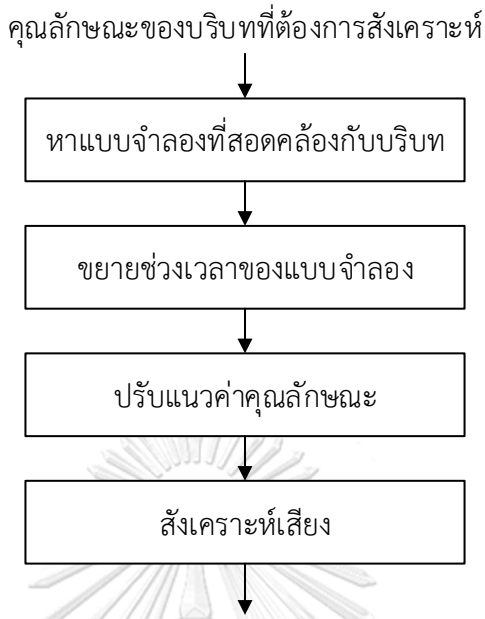
ในขั้นตอนนี้จะทำการประมาณค่าแบบจำลองเสียงแบบขึ้นกับบริบทที่ถูกจัดกลุ่ม ซึ่งจะใช้วิธีค่าคาดหวังสูงสุด จากตัวอย่างที่มีชนิดของหน่วยเสียง และบริบทตรงกับแบบจำลองเสียงขึ้นกับบริบทที่ถูกจัดกลุ่มแล้ว

หลังจากทำกระบวนการทั้ง 7 ขั้นตอนนี้เสร็จแล้ว สามารถทำการวนซ้ำกระบวนการนี้ได้ เพื่อเพิ่มความถูกต้องให้กับแบบจำลองเสียง โดยการแยกกลุ่มแบบจำลองที่ได้จากขั้นตอนที่ 7 ให้กลับมาเป็นแบบจำลองแบบขึ้นกับบริบทแบบที่ไม่ถูกจัดกลุ่ม วิธีการแยกกลุ่มทำโดยการคัดลอกแบบจำลองเสียง และช่วงเวลาที่อยู่ในกลุ่มเดียวกัน ออกมาเป็นแบบจำลองเสียงแบบขึ้นกับบริบท จากนั้นทำการวนซ้ำตั้งแต่ขั้นตอนที่ 5 อีกครั้ง

ผลลัพธ์ที่ใช้เป็นแบบจำลองเสียงคือ ต้นไม้ตัดสินใจที่เก็บแบบจำลองของค่าคุณลักษณะสเปกตรัม ความถี่มูลฐาน ค่าความไม่เป็นคาบของแถบความถี่ และช่วงเวลาไว้ที่ใบของต้นไม้ตัดสินใจจำนวน 5 ต้น 5 ต้น 5 ต้น และ 1 ต้นตามลำดับ

3.3 การสังเคราะห์เสียงด้วยแบบจำลองฮิดเดนมาร์คอฟ

ขั้นตอนการสังเคราะห์เสียงทำตามรูปที่ 9 โดยระบบจะรับค่าคุณลักษณะบริบทของข้อความที่ต้องการสังเคราะห์เป็นข้อมูลรับเข้า และใช้แบบจำลองเสียงสังเคราะห์ และโครงสร้างของต้นไม้ตัดสินใจทั้ง 16 ต้นจากส่วนการฝึกฝน ซึ่งจะมีขั้นตอนดังต่อไปนี้



รูปที่ 9 ขั้นตอนการสังเคราะห์เสียง

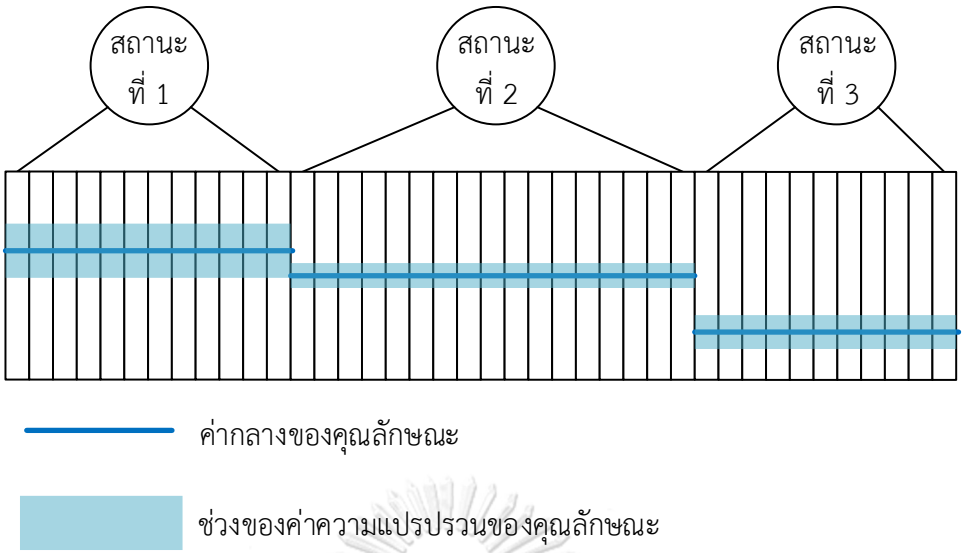
3.3.1 การหาแบบจำลองที่สอดคล้องกับบริบท

ระบบจะนำบริบทที่ต้องการสังเคราะห์เสียงไปทำการแวะผ่านต้นไม้ตัดสินใจทั้ง 16 ต้น เพื่อหาแบบจำลองที่สอดคล้องกับค่าคุณลักษณะทางบริบทที่ต้องการสังเคราะห์เสียง โดยจะได้ผลลัพธ์เป็นแบบจำลองของสเปกตรัม ค่าความถี่มูลฐาน ค่าความไม่เป็นคาบของแถบความถี่ และช่วงเวลา

วิธีการแวะผ่านต้นไม้ตัดสินใจทำโดยการเปรียบเทียบบริบทรับเข้า กับกลุ่มของบริบทที่ถูกนิยามไว้ในจุดต่อของต้นไม้ โดยเริ่มทำการเปรียบเทียบตั้งแต่จุดต่อบนสุดไล่ลงไปถึงใบไม้ของต้นไม้ ซึ่งใบไม้เป็นที่เก็บแบบจำลองไว้

3.3.2 การจัดเรียงและขยายช่วงเวลาของแบบจำลอง

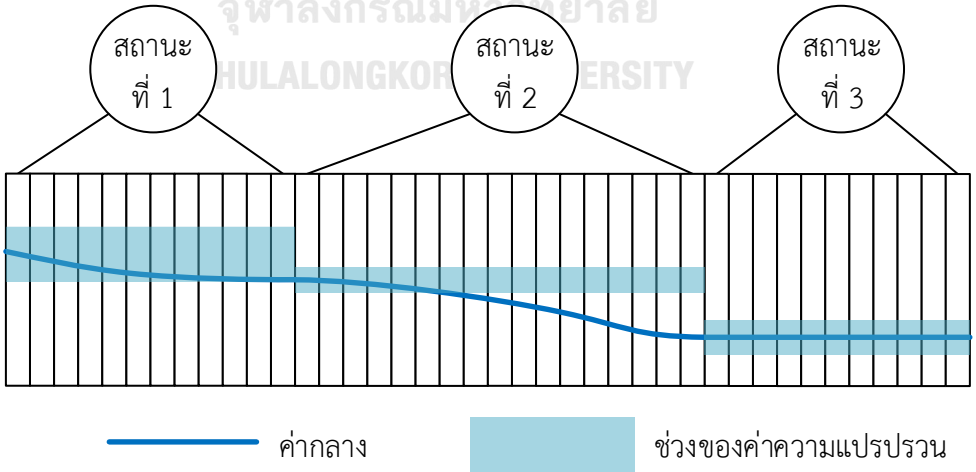
ระบบจะทำการคัดลอกค่ากลาง และค่าความแปรปรวนของแบบจำลอง PDF ของสเปกตรัมและความถี่มูลฐานตามจำนวนเท่ากับค่ากลางของแบบจำลองช่วงเวลาที่สอดคล้องกับสถานะเดียวกัน เช่น ถ้าค่ากลางของแบบจำลองช่วงเวลาของสถานะที่ 1 มีค่าเป็น 5 ระบบจะทำการคัดลอกค่ากลางของสถานะที่ 1 ของแบบจำลองสเปกตรัม และความถี่มูลฐาน ออกเป็น 5 กรอบเวลา จากนั้นนำค่าของแบบจำลองสเปกตรัม และความถี่มูลฐานของทุกๆ สถานะในบริบทมาเรียงต่อกัน ตามรูปที่ 10 แสดงถึงตัวอย่างการเรียงต่อกันของค่าคุณลักษณะเพียง 1 ค่า และใช้ 3 สถานะ



รูปที่ 10 ตัวอย่างการจัดเรียงและขยายช่วงเวลาของแบบจำลอง

3.3.3 การปรับแนวค่าคุณลักษณะ

เนื่องจากการนำค่ากลางของแบบจำลองสเปกตรัม และความถี่มูลฐานมาเรียงต่อกันจะทำให้เกิดความไม่ต่อเนื่องบริเวณรอยต่อของแต่ละสถานะ ดังนั้นจึงต้องทำการปรับแนวค่าคุณลักษณะให้มีความต่อเนื่องกัน โดยผลลัพธ์จะแสดงตัวอย่างดังรูปที่ 11 ซึ่งแสดงให้เห็นว่าค่ากลางได้ถูกปรับแต่งให้ต่อเนื่องกัน



รูปที่ 11 ตัวอย่างการปรับแนวค่าคุณลักษณะ

วิธีการปรับแนวค่าคุณลักษณะ ทำตามสมการที่เสนอโดย Tokuda และคณะ. (2000) [42] โดยจะใช้ค่าความแปรปรวน ค่าความเร็ว (Δ) และความเร่ง (ΔΔ) ของค่าคุณลักษณะมาใช้ในการคำนวณ

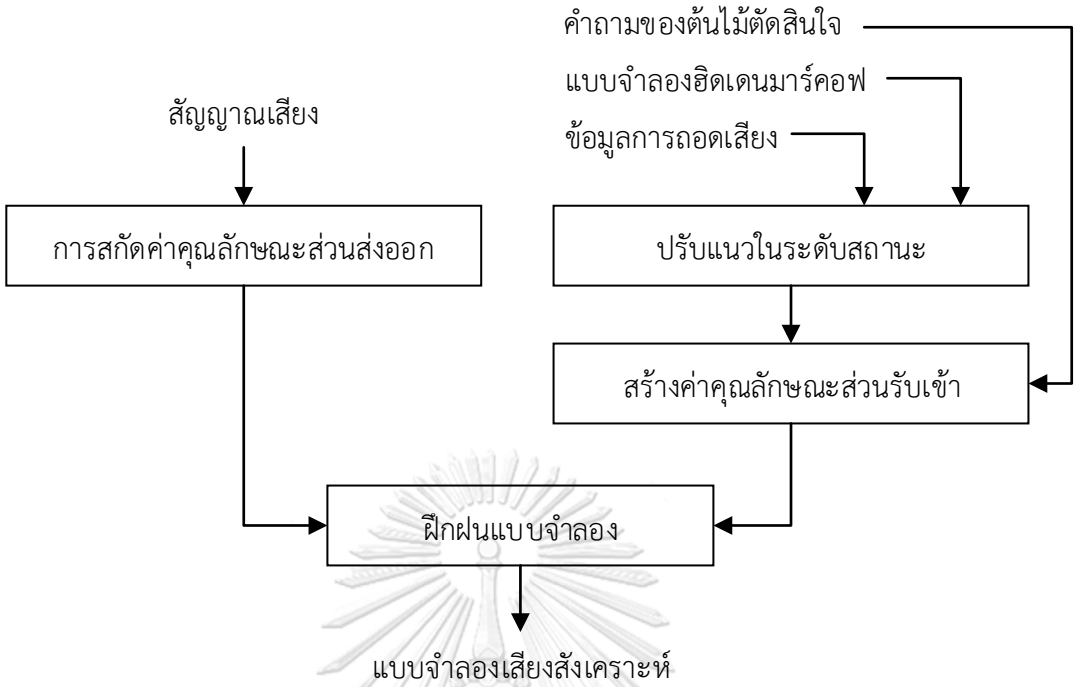
3.3.4 การสังเคราะห์เสียง

การสังเคราะห์เสียงทำโดยการนำค่ากลางของค่าคุณลักษณะที่ได้มีการปรับแต่งแล้ว จากขั้นตอนที่ 3 ไปผ่านตัวเข้ารหัสเสียงในหัวข้อที่ 3.1 เพื่อให้ได้กลับมาเป็นสัญญาณเสียง โดยรายละเอียดในส่วนนี้ จะอธิบายเพิ่มเติมในหัวข้อที่ 3.1

3.4 การสร้างแบบจำลองเสียงสังเคราะห์ด้วยแบบจำลองโครงข่ายประสาทเทียมแบบลึก

ระบบสังเคราะห์เสียงเป็นระบบที่มีข้อมูลในส่วนรับเข้า (Input) คือข้อความ และแปลงข้อความเหล่านั้นเป็นสัญญาณเสียงสังเคราะห์ และฐานข้อมูลเสียงที่ใช้สำหรับการฝึกฝนแบบจำลองเสียงจะทำการกำกับค่าคุณลักษณะทางบริบทพร้อมทั้งมีการกำกับช่วงเวลาที่สอดคล้องกับข้อมูลเสียง

ขั้นตอนการสร้างแบบจำลองเสียงด้วยแบบจำลองโครงข่ายประสาทเทียมแบบลึกอ้างอิงกับงานวิจัยที่นำเสนอโดย Zen และคณะ. (2013) [35] และมีขั้นตอนตามรูปที่ 12 ซึ่งข้อมูลรับเข้าประกอบด้วย ข้อมูลสัญญาณเสียง ข้อมูลการถอดเสียง คำถามของต้นไม้ตัดสินใจที่ใช้ในการสร้างต้นไม้ตัดสินใจเพื่อใช้ในการจัดกลุ่มแบบจำลองเสียง และแบบจำลองฮิดเดนมาร์คอฟ



รูปที่ 12 ขั้นตอนการสร้างแบบจำลองเสียงสังเคราะห์จากแบบจำลองโครงข่ายประสาทเทียมแบบลึก

ขั้นตอนในแต่ละกระบวนการเป็นดังต่อไปนี้

3.4.1 การสกัดค่าคุณลักษณะส่วนส่งออก (Output Feature)

การสกัดค่าคุณลักษณะสำหรับการสร้างแบบจำลองเสียงสังเคราะห์ด้วยแบบจำลองโครงข่ายประสาทเทียมแบบลึก จะเหมือนกับการสกัดค่าคุณลักษณะของการสร้างแบบจำลองเสียงด้วยแบบจำลองฮิดเดนมาร์คอฟ ซึ่งจะทำการแบ่งสัญญาณเสียงออกเป็นกรอบเวลา และทำการสกัดค่าคุณลักษณะ ซึ่งจะประกอบด้วยค่าคุณลักษณะ MCEP, BAP และ f_0 และรวมถึงค่าความเร็ว และความเร่งของทั้ง 3 ค่าคุณลักษณะ

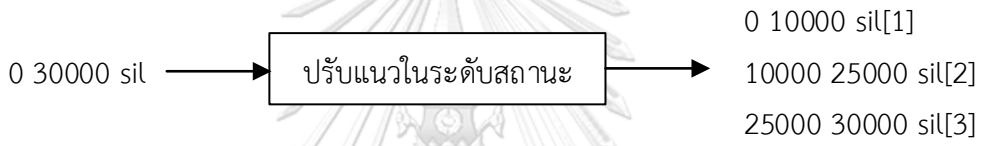
สำหรับค่าของค่าคุณลักษณะสามารถทำการนอร์มัลไลเซชันเพื่อที่จะทำให้ค่าอยู่ในช่วงระหว่าง 0 ถึง 1 เพื่อให้สามารถนำไปใช้ร่วมกับฟังก์ชันกระตุ้นประเภท Sigmoid หรือประเภท Tanh

3.4.2 การปรับแนวในระดับสถานะ (State-level alignment)

เนื่องจากในฐานข้อมูลเสียงมักจะกำกับช่วงเวลาที่ระดับของหน่วยเสียงเท่านั้น ซึ่งการระบุตำแหน่งในระดับดังกล่าวไม่เพียงพอที่จะสร้างเสียงสังเคราะห์ที่มีคุณภาพได้ ดังนั้นจึงระบุตำแหน่งในระดับของสถานะในหน่วยเสียง จะทำให้ได้เสียงสังเคราะห์ที่มีคุณภาพมากกว่า ดังนั้นจึงต้องทำการปรับแนวในระดับสถานะด้วยแบบจำลองฮิดเดนมาร์คอฟ ซึ่งแบบจำลองนี้ไม่จำเป็นต้องใช้แบบจำลอง

ที่เป็นผลลัพธ์จากส่วนของการสร้างแบบจำลองเสียงด้วยแบบจำลองฮิดเดนมาร์คอฟ อาจจะใช้แบบจำลองสำหรับการรู้จำเสียงทดแทนกันได้ แต่ต้องเป็นแบบจำลองที่มีแบบจำลองช่วงเวลาด้วย

ตัวอย่างของการปรับแนวค่าคุณลักษณะเป็นไปดังรูปที่ 13 ที่ข้อมูลการถอดความอยู่ในรูปแบบของเวลาเริ่มต้น เวลาสิ้นสุด บริบทของหน่วยเสียง โดยข้อมูลส่วนรับเข้าเป็นหน่วยเสียง sil ที่มีช่วงเวลาเริ่มต้นตั้งแต่ตำแหน่งที่ 0 ไปจนถึงตำแหน่งที่ 30000 ไมโครวินาที และสำหรับตัวอย่างตามรูปที่ 13 กำหนดให้แบบจำลองฮิดเดนมาร์คอฟมีเพียง 3 สถานะ จึงทำให้ผลลัพธ์หลังผ่านกระบวนการปรับแนวในระดับสถานะมีจำนวน 3 ค่า ซึ่งได้แก่ ช่วงเวลาที่ 0 ถึง 10000 ไมโครวินาทีแทนสถานะที่ 1 ของหน่วยเสียง sil (กำกับหมายเลขสถานะในเครื่องหมาย [ ]) ช่วงเวลาที่ 10000 ถึง 25000 แทนช่วงเวลาของหน่วยเสียง sil ในสถานะที่ 2 และช่วงเวลา 25000 ถึง 30000 ไมโครวินาที แทนหน่วยเสียง sil ในสถานะที่ 3



รูปที่ 13 ตัวอย่างการปรับแนวในระดับสถานะ

3.4.3 การสร้างค่าคุณลักษณะส่วนรับเข้า (Input Feature)

เนื่องจากวิธีการเรียนรู้ และทำนายของแบบจำลองโครงข่ายประสาทเทียมแบบลึกจำเป็นต้องกำหนดค่าคุณลักษณะให้กับทุกตัวอย่างที่ใช้ในการฝึกฝน ทั้งในส่วนที่เป็นข้อมูลส่วนรับเข้า และข้อมูลส่วนส่งออก (ข้อมูลเป้าหมาย) สำหรับในกรณีของการสังเคราะห์เสียง ข้อมูลเสียงในส่วนรับเข้า คือข้อมูลค่าคุณลักษณะทางบริบทของหน่วยเสียง ซึ่งแต่เดิมได้มีกำกับอยู่ในระดับของหน่วยเสียง และทำให้ละเอียดขึ้นไปในระดับของสถานะในหน่วยเสียงตามที่เสนอไปในหัวข้อที่ 3.4.2 และในหัวข้อนี้จะอธิบายถึงวิธีการที่จะนำข้อมูลจากหัวข้อที่ 3.4.2 ไปทำการแปลงเป็นค่าคุณลักษณะส่วนรับเข้าที่สอดคล้องกับข้อมูลตัวอย่างเสียงในแต่ละกรอบเวลา

สำหรับค่าคุณลักษณะส่วนรับเข้าถูกแบ่งออกเป็น 2 ประเภท ดังต่อไปนี้

- ค่าคุณลักษณะบรรยายบริบทหน่วยเสียง (Context Feature)

ค่าคุณลักษณะดังกล่าวจะสร้างขึ้นจากบริบทของหน่วยเสียง เช่น หน่วยเสียงของตัวอย่างข้อมูล ซึ่งแทนด้วยค่าคุณลักษณะแบบฐานสอง (Binary feature) ดังตัวอย่างเช่น หน่วยเสียงทั้งหมดมีทั้งหมด 50 หน่วยเสียง จะทำให้เกิดค่าคุณลักษณะขึ้นมา 50 ค่าคุณลักษณะตามหน่วยเสียงทั้งหมด

โดยให้ค่าคุณลักษณะที่มีค่าไม่ตรงกับหน่วยเสียงที่ต้องการมีค่าเป็น 0 และให้ค่าคุณลักษณะที่มีค่าตรงกับหน่วยเสียงมีค่าเป็น 1

แต่อย่างไรก็ตาม การใช้เพียงข้อมูลหน่วยเสียงเพียงอย่างเดียวไม่เพียงพอในการสร้างระบบสังเคราะห์เสียง จึงมีการใช้คำถามสำหรับสร้างต้นไม้ตัดสินใจมาสร้างเป็นค่าคุณลักษณะส่วนรับเข้า โดยคำถามที่ใช้ในการสร้างต้นไม้ตัดสินใจ จะอยู่ในรูปแบบของคำถามที่ถามว่า มีบริบทตรงตามรูปแบบที่กำหนดไว้หรือไม่ โดยถ้ามีตรงตามรูปแบบ จะถือว่าค่าคุณลักษณะของคำถามข้อนี้มีค่าเป็น 1 และถ้าไม่ตรงรูปแบบให้ถือว่ามีค่าเป็น 0

โดยค่าคุณลักษณะดังกล่าวจะมีค่าเหมือนกันในทุกกรอบเวลาของตัวอย่างเสียงที่สอดคล้องกับบริบทดังกล่าว

- ค่าคุณลักษณะระบุตำแหน่ง (Positioning Feature)

ค่าคุณลักษณะระบุตำแหน่งใช้ในการระบุตำแหน่งให้กับแต่ละตัวอย่างที่อยู่ในช่วงเวลาของหน่วยเสียงเดียวกัน ซึ่งประกอบด้วย

1. ตำแหน่งของสถานะของในหน่วยเสียงของตัวอย่างเสียงที่กำลังพิจารณา เรียกว่า SP ซึ่งสามารถนำเสนอค่าคุณลักษณะดังกล่าวในรูปแบบของค่าคุณลักษณะแบบฐานสอง ซึ่งจะทำให้มีจำนวนค่าคุณลักษณะเท่ากับจำนวนสถานะ และให้ค่าคุณลักษณะที่มีสถานะตรงกับตัวอย่างมีค่าเป็น 1 นอกนั้นมีค่าเป็น 0 หรือใช้การนำเสนอในรูปแบบของเลขจำนวนจริง ที่ถูกปรับสเกลให้อยู่ระหว่าง 0 ถึง 1
2. ตำแหน่งของตัวอย่างเสียงที่กำลังพิจารณาเปรียบเทียบกับตัวอย่างเสียงในสถานะ เรียกว่า FPS นำเสนอในรูปแบบของจำนวนจริงที่ปรับสเกลให้อยู่ในระหว่าง 0 ถึง 1 ซึ่งแทนถึง % ของตำแหน่งของตัวอย่างเทียบกับความยาวทั้งหมดของสถานะนั้น
3. ตำแหน่งของตัวอย่างเสียงที่กำลังพิจารณาเปรียบเทียบกับตัวอย่างเสียงในหน่วยเสียง เรียกว่า FPP นำเสนอในรูปแบบของจำนวนจริงที่ปรับสเกลให้อยู่ในระหว่าง 0 ถึง 1 ซึ่งแทนถึง % ของตำแหน่งของตัวอย่างเทียบกับความยาวทั้งหมดของหน่วยเสียงนั้น

ค่าคุณลักษณะทั้งสองประเภทจะนำมารวมกันเป็นคุณลักษณะส่วนรับเข้าของทุกตัวอย่างเพื่อใช้ในการฝึกฝนแบบจำลอง

ตัวอย่างค่าคุณลักษณะในส่วนรับเข้าตามรูปที่ 13 แสดงดังตารางที่ 1 โดยกำหนดให้หนึ่งกรอบเวลามีช่วงเวลาเป็น 5000 ไมโครวินาที ซึ่งจะทำให้ได้กรอบเวลาทั้งหมด 6 กรอบเวลา โดยในตัวอย่างกำหนดให้มีคำถามของต้นไม้ตัดสินใจเพียง 4 ข้อ โดยคำถามคือใช่หน่วยเสียงตามที่ปรากฏในหัวตารางของหัวข้อค่าคุณลักษณะบรรยายบริบทหน่วยเสียงหรือไม่ และเนื่องจากในรูปที่ 13 มีเพียงหน่วยเสียง sil จึงทำให้ค่าคุณลักษณะบรรยายบริบทหน่วยเสียงของทุกกรอบเวลาของตัวอย่างเสียง มี

ค่าเป็น 1 เฉพาะในหัวข้อของหน่วยเสียง sil และในหัวข้ออื่นมีค่าเป็น 0 ทั้งหมด และสำหรับค่าคุณลักษณะระบุตำแหน่งประเภท SP กำหนดให้ใช้เพียง 1 ค่าคุณลักษณะที่มีค่าอยู่ระหว่าง 0 ถึง 1 และในกรณีที่ในสถานะนั้นมีเพียง 1 กรอบเวลาของตัวอย่างเสียง (สถานะที่ 3 มีเพียงกรอบเวลาลำดับที่ 6) กำหนดให้ค่า FPS มีค่าเป็น 0

ตารางที่ 1 ตัวอย่างค่าคุณลักษณะส่วนรับเข้าของแบบจำลองโครงข่ายประสาทเทียมแบบลึก

ลำดับ	ค่าคุณลักษณะบรรยายบริบทหน่วยเสียง				ค่าคุณลักษณะระบุตำแหน่ง		
	aa	sil	ng	k	SP	FPS	FPP
1	0	1	0	0	0	0	0
2	0	1	0	0	0	1	0.2
3	0	1	0	0	0.5	0	0.4
4	0	1	0	0	0.5	0.5	0.6
5	0	1	0	0	0.5	1	0.8
6	0	1	0	0	1	0	1

3.4.4 การฝึกฝนแบบจำลอง

แบบจำลองโครงข่ายประสาทเทียบแบบลึกที่นิยมนำมาใช้ ได้แก่ เพอร์เซ็ปตรอนหลายชั้นแบบป้อนไปหน้า และข่ายงานแบบวนซ้ำ โดยข้อมูลรับเข้าคือคุณลักษณะส่วนรับเข้าที่เป็นผลลัพธ์ของหัวข้อที่ 3.4.3 และข้อมูลส่งออกคือค่าคุณลักษณะของตัวอย่างเสียงที่เป็นผลลัพธ์ของหัวข้อที่ 3.4.1

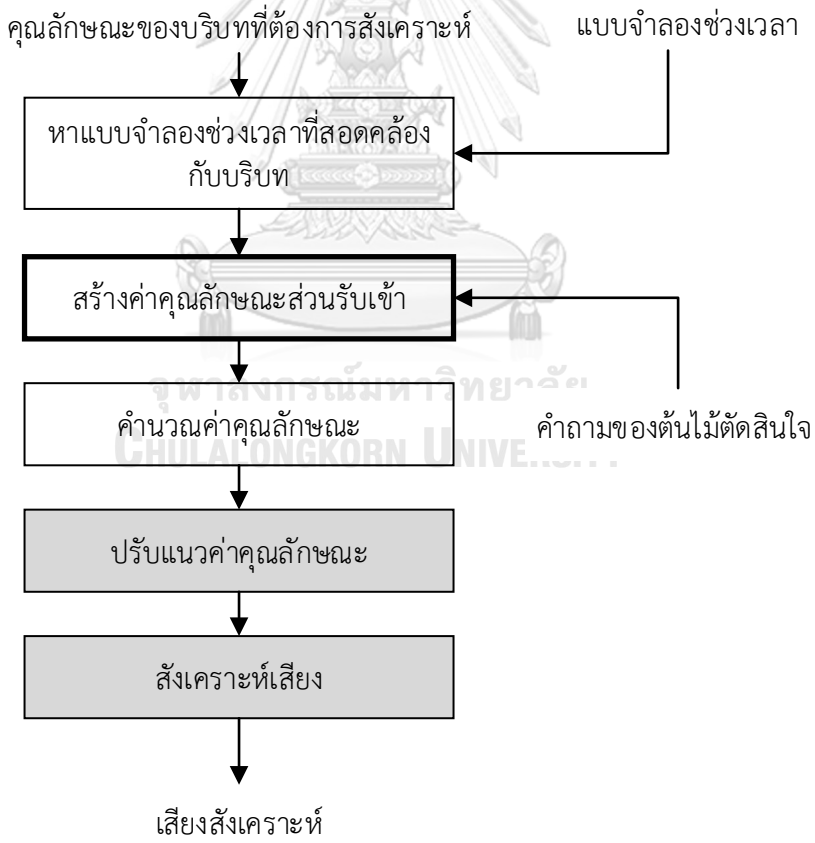
สำหรับฟังก์ชันต้นทุนที่นิยมใช้กัน ได้แก่ ค่าคลาดเคลื่อนกำลังสองเฉลี่ย และค่าคลาดเคลื่อนกำลังสองเฉลี่ยที่หารด้วยค่าความแปรปรวนของค่าคุณลักษณะนั้น โดยค่าความแปรปรวนคำนวณจากค่าคุณลักษณะทั้งหมด

การปรับค่าตัวแปรสำหรับการฝึกฝน และการเลือกตัวเพิ่มประสิทธิภาพ (Optimizer) สามารถเลือกใช้เหมือนกับการสร้างแบบจำลองโครงข่ายประสาทเทียบแบบลึกประเภทพอร์เซ็ปตรอนหลายชั้นแบบป้อนไปหน้า หรือแบบข่ายงานแบบวนซ้ำทั่วไปได้

3.5 การสังเคราะห์เสียงด้วยแบบจำลองโครงข่ายประสาทเทียมแบบลึก

กระบวนการในการสังเคราะห์เสียงแสดงในรูปที่ 14 ซึ่งประกอบด้วย กระบวนการที่มีพื้นหลังเป็นสีเทา แสดงถึงกระบวนการที่เหมือนกับการสังเคราะห์เสียงด้วยแบบจำลองฮิดเดนมาร์คอฟ และกรอบที่มีเส้นขอบหนา แสดงถึงกระบวนการที่เหมือนกับในการสร้างแบบจำลองเสียงสังเคราะห์ด้วยแบบจำลองโครงข่ายประสาทเทียมแบบลึก

ข้อมูลรับเข้าของการสังเคราะห์เสียงด้วยแบบจำลองโครงข่ายประสาทเทียมแบบลึก ประกอบด้วย คุณลักษณะของบริบทที่ต้องการสร้างเสียงสังเคราะห์ และแบบจำลองช่วงเวลาซึ่งเป็นแบบจำลองประเภทแบบจำลองฮิดเดนมาร์คอฟที่ถูกสร้างขึ้นพร้อมกับแบบจำลองฮิดเดนมาร์คอฟที่ใช้กระบวนการสร้างแบบจำลองเสียงสังเคราะห์ด้วยแบบจำลองโครงข่ายประสาทเทียมแบบลึก และคำถามของต้นไม้ตัดสินใจที่ใช้ในกระบวนการสร้างแบบจำลองเสียงสังเคราะห์ด้วยแบบจำลองโครงข่ายประสาทเทียมแบบลึก



รูปที่ 14 ขั้นตอนการสังเคราะห์เสียงจากแบบจำลองโครงข่ายประสาทเทียมแบบลึก

กระบวนการในการสังเคราะห์เสียงมีขั้นตอนเป็นดังต่อไปนี้

3.5.1 หาแบบจำลองช่วงเวลาที่สอดคล้องกับบริบท

ในกระบวนการนี้ ระบบจะนำบริบทที่ต้องการสร้างเสียงสังเคราะห์ไปทำการค้นหาแบบจำลองช่วงเวลาที่สอดคล้องกับบริบทดังกล่าว ซึ่งมีวิธีการเหมือนกับกระบวนการที่ 3.3.1 โดยเมื่อได้แบบจำลองที่สอดคล้องกับบริบทแล้ว จะใช้ค่ากลางของแบบจำลองช่วงเวลา เป็นค่าความยาวของแต่ละสถานะในหน่วยเสียง โดยผลลัพธ์ของกระบวนการนี้ จะได้ช่วงเวลาในแต่ละสถานะของแต่ละหน่วยเสียง

3.5.2 สร้างค่าคุณลักษณะส่วนรับเข้า

กระบวนการสร้างคุณลักษณะส่วนรับเข้ามีกระบวนการเหมือนกับขั้นตอนในกระบวนการที่ 3.4.3 โดยจะใช้ข้อมูลบริบทจากที่รับเข้ามา ส่วนข้อมูลช่วงเวลาจะได้รับจากผลลัพธ์ของกระบวนการหาแบบจำลองช่วงเวลาที่สอดคล้องกับบริบท

3.5.3 การคำนวณค่าคุณลักษณะ

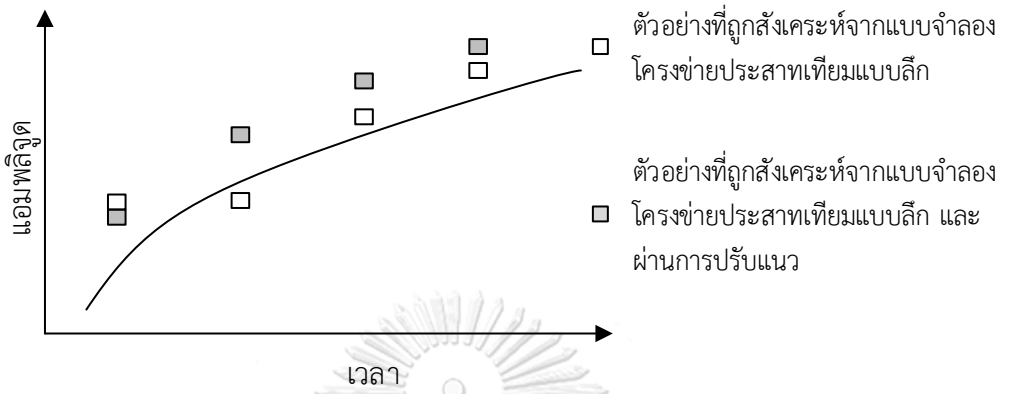
ในกระบวนการนี้จะนำค่าคุณลักษณะจากกระบวนการสร้างค่าคุณลักษณะ ส่งผ่านเข้าไปยังแบบจำลองโครงข่ายประสาทเทียมแบบลึก และได้ผลลัพธ์ออกมาเป็นค่าคุณลักษณะของเสียงสังเคราะห์

3.5.4 การปรับแนวค่าคุณลักษณะ

คุณลักษณะที่สังเคราะห์ออกมาจากแบบจำลองโครงข่ายประสาทเทียมแบบลึกสามารถจำลองวิถีการเปลี่ยนแปลงภายในสถานะได้ ซึ่งต่างจากการสังเคราะห์ค่าคุณลักษณะด้วยแบบจำลองฮิดเดนมาร์คอฟที่จะได้มาเฉพาะค่ากลางของแต่ละสถานะเท่านั้น แต่วิถีการเปลี่ยนแปลงของค่าคุณลักษณะที่สังเคราะห์จากแบบจำลองโครงข่ายประสาทเทียมนั้น มีความไม่แน่นอนเกิดขึ้น ดังแสดงในรูปที่ 15 ที่วิถีของค่าคุณลักษณะที่ต้องการมีลักษณะเป็นค่าที่เพิ่มขึ้นอย่างสม่ำเสมอ แต่ค่าที่สังเคราะห์ออกมา มีลักษณะการเปลี่ยนแปลงที่คงที่ในช่วงของตัวอย่างที่ 1 และ 2 แต่กลับมีการเพิ่มขึ้นอย่างรวดเร็วในช่วงตัวอย่างที่ 2 และ 4

วิธีการปรับแนวค่าคุณลักษณะจะเหมือนกับในหัวข้อ 3.3.3 โดยค่าความแปรปรวนในแต่ละค่าคุณลักษณะจะคำนวณจากค่าของคุณลักษณะของทุกตัวอย่างที่ใช้ในการฝึกฝน เรียกว่า ค่าความแปรปรวนแบบครอบคลุม สำหรับค่าความเร็ว และความเร่งของคุณลักษณะ ใช้จากผลลัพธ์ของที่ได้จากกระบวนการคำนวณค่าคุณลักษณะ

ผลลัพธ์ที่ได้จากกระบวนการนี้อาจจะไม่ได้ผลลัพธ์ที่ใกล้เคียงกับค่าต้นฉบับมากเท่ากับผลลัพธ์ที่ได้จากการคำนวณด้วยแบบจำลองโครงข่ายประสาทเทียมแบบลึกโดยตรง แต่จะได้วิถีการเปลี่ยนแปลงที่ใกล้เคียงกับต้นฉบับมากขึ้น ดังแสดงในรูปที่ 15



รูปที่ 15 ตัวอย่างผลลัพธ์การปรับแนวของค่าคุณลักษณะที่ได้จากการสังเคราะห์

3.5.5 กระบวนการสังเคราะห์เสียง

กระบวนการนี้เหมือนกับกระบวนการที่ 3.3.4 ของการสังเคราะห์เสียงจากแบบจำลองแบบฮิดเดนมาร์คอฟ โดยรับค่าคุณลักษณะจากส่วนของการปรับแนวค่าคุณลักษณะ และทำการสังเคราะห์ออกมาเป็นสัญญาณเสียงโดยใช้ตัวเข้ารหัสเสียง STRAIGHT

บทที่ 4 แนวคิดของการวิจัย และวิธีการดำเนินงาน

ในงานวิจัยนี้มีแนวคิดการพัฒนาคุณภาพของเสียงสังเคราะห์ในด้านของความถูกต้องในการเปล่งเสียง และความเป็นธรรมชาติ โดยจะนำเสนอแนวคิดเกี่ยวกับการปรับเปลี่ยนโครงสร้างของแบบจำลองเสียงสังเคราะห์แยกตามประเภทของแบบจำลองเสียง โดยในแบบจำลองฮิดเดนมาร์คอฟจะทำการปรับเปลี่ยนโครงสร้างของแบบจำลองเสียง และในแบบจำลองโครงข่ายประสาทเทียมแบบลึก จะทำการปรับเปลี่ยนค่าคุณลักษณะส่วนรับเข้า และการนอร์มัลไลเซชันค่าคุณลักษณะส่วนส่งออกของทั้งสองแนวคิดที่นำเสนอ มีรายละเอียดดังต่อไปนี้

4.1 การปรับเปลี่ยนโครงสร้างของแบบจำลองเสียง

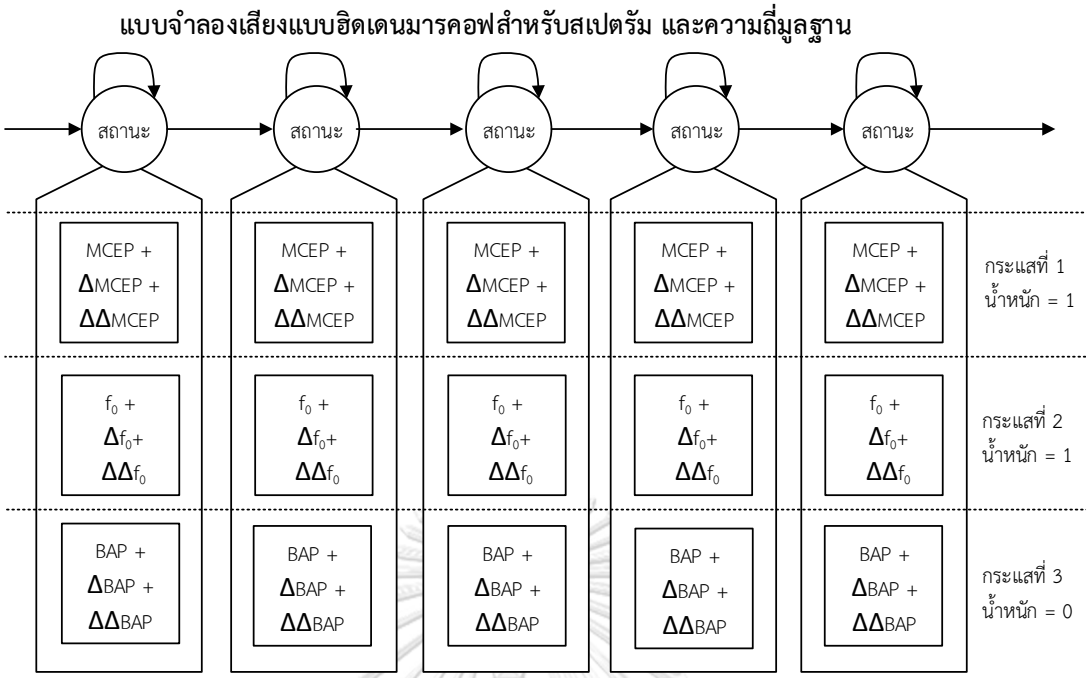
การปรับเปลี่ยนโครงสร้างของแบบจำลองเสียง ประกอบไปด้วย 2 ประเด็นย่อย ได้แก่ โครงสร้างของแบบจำลองเสียง และการคำนวณหาช่วงเวลาของเสียงสังเคราะห์

4.1.1 โครงสร้างของแบบจำลองเสียง

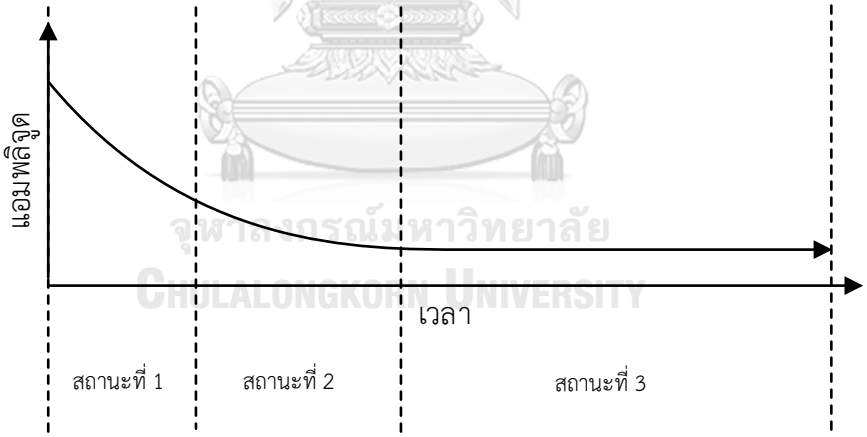
ด้วยโครงสร้างเดิมของต้นแบบของแบบจำลองเสียงสังเคราะห์แบบฮิดเดนมาร์คอฟ ที่ได้รวมค่าคุณลักษณะสเปกตรัม (กระแสที่ 1) เรียกว่า MCEP ค่าคุณลักษณะความถี่มูลฐาน (กระแสที่ 2) เรียกว่า f_0 และค่าคุณลักษณะค่าความไม่เป็นคาบของแถบความถี่ เรียกว่า BAP (กระแสที่ 3) ตามรูปที่ 16 ที่ประกอบด้วย 5 สถานะ เข้าเป็นแบบจำลองเดียวกัน ซึ่งจะส่งผลให้ทั้งสองส่วนต้องถูกถ่วงน้ำหนัก และส่งผลต่อการแบ่งสถานะในช่วงการฝึกฝน (ขั้นตอนการประมาณค่าแบบจำลอง)

การแบ่งสถานะในขั้นตอนการฝึกฝน จะทำให้ค่าของความน่าจะเป็นของผลลัพธ์มีค่ามากที่สุด โดยค่าความน่าจะเป็นของผลลัพธ์จะคำนวณจากค่าความน่าจะเป็นของกลุ่มตัวอย่างที่ใช้ในการฝึกฝนในสถานะนั้น เทียบกับแบบจำลองทางสถิติในสถานะนั้น ซึ่งแบบจำลองทางสถิติก็ถูกฝึกฝนมาจากกลุ่มตัวอย่างเช่นเดียวกัน ดังนั้นการแบ่งสถานะ จึงต้องทำให้ค่าความแปรปรวนในแต่ละสถานะมีค่าน้อยที่สุด

ตัวอย่างการแบ่งสถานะสำหรับกรณีที่มีเพียง 1 ค่าคุณลักษณะ แสดงดังตัวอย่างรูปที่ 17 ที่ต้องการจำลองการเปลี่ยนแปลงของค่าคุณลักษณะดังกล่าวด้วย 3 สถานะ ซึ่งค่าคุณลักษณะในตัวอย่างของรูปที่ 17 มีค่าลดลง และคงที่ในช่วงครึ่งสุดท้าย ผลลัพธ์การแบ่งสถานะแสดงดังรูปที่ 17 โดยจะเลือกใช้เพียง 1 สถานะเพื่อจำลองข้อมูลในช่วงครึ่งสุดท้าย เพราะในช่วงครึ่งสุดท้ายเป็นช่วงที่คงที่ จึงใช้เพียง 1 สถานะเพื่อจำลองข้อมูลในส่วนดังกล่าว และเลือกใช้ 2 สถานะที่เหลือในการจำลองส่วนเริ่มต้น ที่มีการเปลี่ยนแปลงมากกว่าช่วงท้าย



รูปที่ 16 แบบจำลองเสียงแบบฮิดเดนมารคอฟสำหรับสเปกตรัม และความถี่มูลฐาน



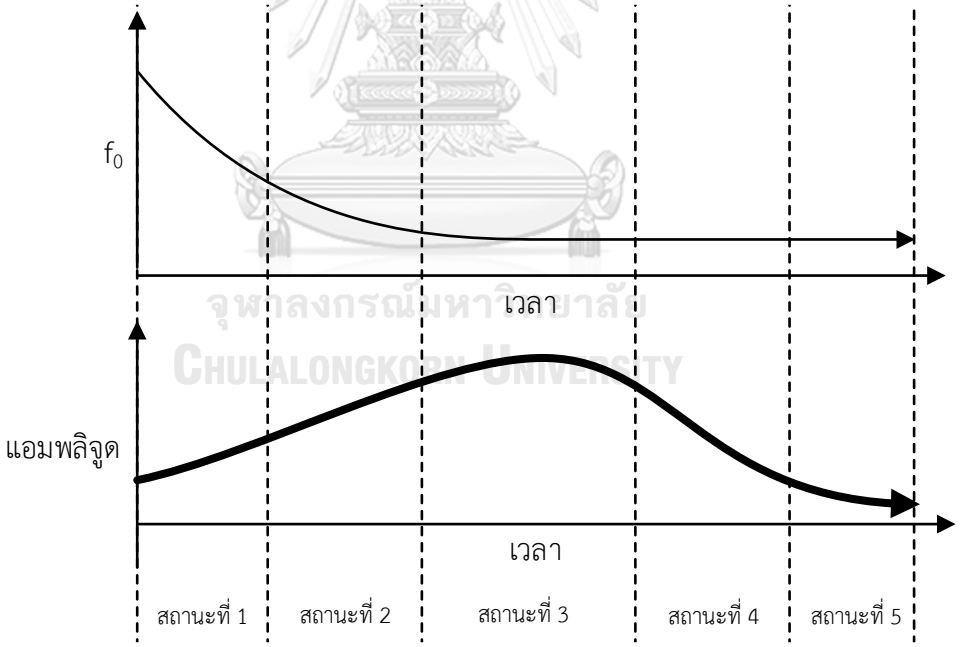
รูปที่ 17 ตัวอย่างการแบ่งสถานะของค่าคุณลักษณะ 1 ค่า

สำหรับการแบ่งสถานะของข้อมูลที่ประกอบกันมากกว่า 1 ค่าคุณลักษณะ แสดงตัวอย่างในรูปที่ 18, รูปที่ 19 และ รูปที่ 20 โดยในแต่ละตัวอย่าง ประกอบด้วยค่าคุณลักษณะของสเปกตรัมเพียง 1 คุณลักษณะ ซึ่งมีลักษณะการเปลี่ยนแปลงคือมีค่าเพิ่มขึ้น และลดลงจนมีค่ากลับมาเท่าเดิม และค่าคุณลักษณะของความถี่มูลฐาน 1 ค่าคุณลักษณะ โดยค่าคุณลักษณะมีการเปลี่ยนแปลงเหมือนกับตัวอย่างในรูปที่ 17 โดยกราฟด้านบนของตัวอย่างในรูปที่ 18, รูปที่ 19 และ รูปที่ 20 แทนถึงค่าคุณลักษณะความถี่มูลฐาน ด้านล่างแสดงถึงค่าคุณลักษณะสเปกตรัม

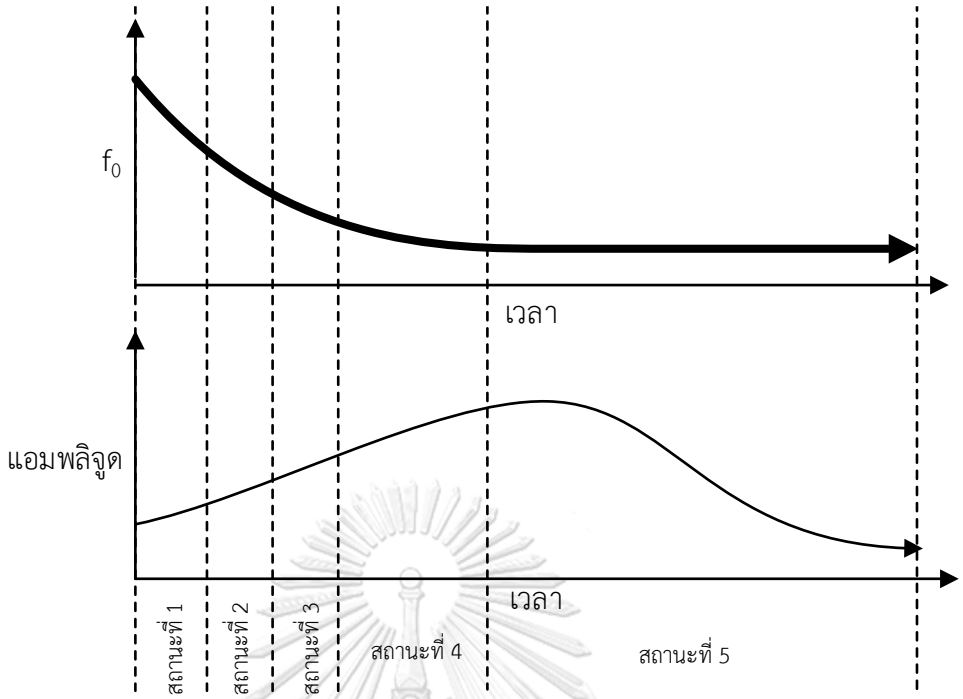
ตัวอย่างในรูปที่ 18 แทนการแบ่งสถานะที่ให้ค่าน้ำหนักกับค่าคุณลักษณะสเปกตรัมเพียงค่าเดียว ซึ่งจะสามารถแบ่งสถานะให้เหมาะสมกับค่าคุณลักษณะสเปกตรัมเป็นอย่างดี แต่ไม่เหมาะกับค่าคุณลักษณะความถี่มูลฐาน เพราะช่วงที่มีการเปลี่ยนแปลงมากกลับใช้เพียง 2 สถานะ แต่สำหรับในช่วงที่ไม่มีการเปลี่ยนแปลงเลย กลับใช้ถึง 3 สถานะ

สำหรับในตัวอย่างในรูปที่ 19 ที่ให้ค่าน้ำหนักกับค่าความถี่มูลเพียงอย่างเดียว จะเห็นว่าใช้ถึง 4 สถานะในการจำลองช่วงที่มีการเปลี่ยนแปลง และใช้เพียง 1 สถานะสำหรับช่วงที่ไม่มีการเปลี่ยนแปลง ซึ่งเหมาะกับการสร้างแบบจำของค่าคุณลักษณะความถี่มูลฐานมากกว่าในรูปที่ 18 แต่ในทางกลับกันสำหรับแบบจำลองค่าคุณลักษณะสเปกตรัมจะไม่เหมาะสมเมื่อเทียบกับรูปที่ 18

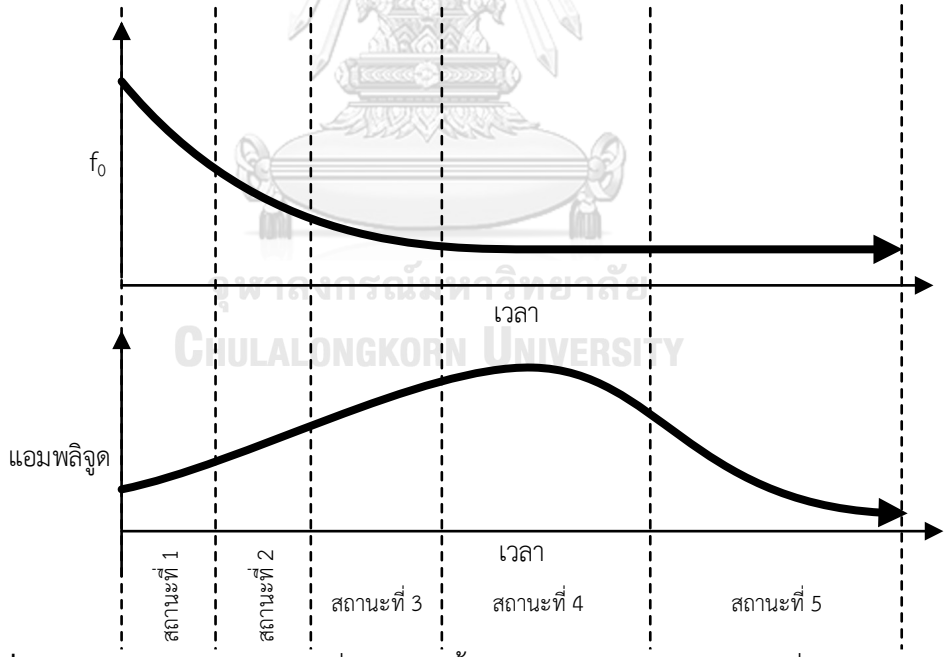
ตัวอย่างในรูปที่ 20 แสดงถึงการแบ่งค่าคุณลักษณะโดยมีการถ่วงน้ำหนักระหว่างค่าคุณลักษณะความถี่มูลฐาน และค่าคุณลักษณะสเปกตรัมที่เท่ากัน ซึ่งจะยังคงแสดงให้เห็นว่า ถึงแม้ว่าการให้น้ำหนักที่เท่ากัน ก็ไม่สามารถแบ่งสถานะได้อย่างเหมาะสม เป็นเพียงการหาค่าที่ดีสุดระหว่างทั้งสองค่าคุณลักษณะเท่านั้น ค่าคุณลักษณะความถี่มูลฐาน ก็ยังคงใช้ถึง 2 สถานะในการจำลองส่วนที่เป็นค่าคงที่ในช่วงท้าย



รูปที่ 18 ตัวอย่างการแบ่งสถานะที่ให้ค่าถ่วงน้ำหนักกับค่าคุณลักษณะสเปกตรัมเพียงอย่างเดียว



รูปที่ 19 ตัวอย่างการแบ่งสถานะที่ให้ค่าถ่วงน้ำหนักกับค่าคุณลักษณะความถี่มูลฐานเพียงอย่างเดียว



รูปที่ 20 ตัวอย่างการแบ่งสถานะที่ให้ค่าถ่วงน้ำหนักกับค่าคุณลักษณะความถี่มูลฐานเท่ากับค่าคุณลักษณะสเปกตรัม

ผลลัพธ์การแบ่งสถานะตามรูปที่ 20 อาจจะทำให้เกิดเหตุการณ์ที่แบบจำลองสามารถจำลองข้อมูลฝึกฝนได้ดีเกินไป (Over fitting) เป็นเพราะค่าความแปรปรวนของแบบจำลองที่ต่ำมาก ในสถานะที่ 4 และ 5 ของค่าคุณลักษณะความถี่มูลฐาน

การที่ได้ค่าความแปรปรวนต่ำ จะส่งผลอย่างมากในกระบวนการปรับแนวค่าคุณลักษณะ เพราะถ้าค่าความแปรปรวนต่ำ (ปัญหาการปรับเรียบมากเกินไป) จะทำให้ค่าคุณลักษณะหลังการปรับแนวมีการเปลี่ยนแปลงน้อยกว่าที่ควรจะเป็น ทำให้น้ำเสียงของเสียงสังเคราะห์ที่ได้ออกมามีลักษณะไม่เป็นธรรมชาติ

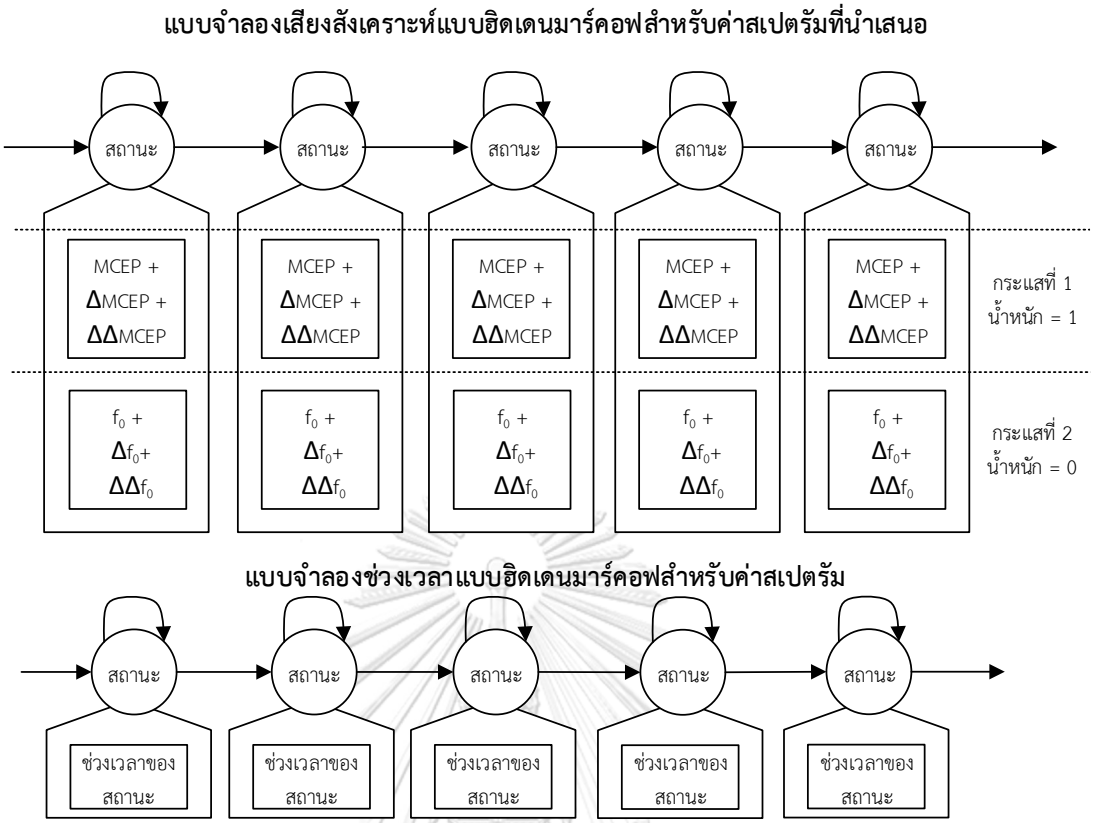
ดังนั้น ในงานวิจัยนี้จึงได้นำเสนอแนวคิดการปรับเปลี่ยนโครงสร้างของแบบจำลองเสียง โดยแยกแบบจำลองออกเป็น 2 แบบจำลอง ได้แก่

1. แบบจำลองที่ใช้ในการจำลองค่าคุณลักษณะความถี่มูลฐานร่วมกับค่าความไม่เป็นคาบของแถบความถี่ โดยเรียกว่าแบบจำลองค่าความถี่มูลฐาน
2. แบบจำลองที่สองใช้ในการจำลองค่าคุณลักษณะสเปกตรัม เรียกว่า แบบจำลองสเปกตรัม

ในแบบจำลองความถี่มูลฐาน ประกอบด้วย ค่าคุณลักษณะค่าความไม่เป็นคาบของแถบความถี่ และค่าคุณลักษณะความถี่มูลฐาน โดยทำการแยกเป็นคนละกระแส และกำหนดให้กระแสของค่าความไม่เป็นคาบของแถบความถี่มีค่าเป็น 0

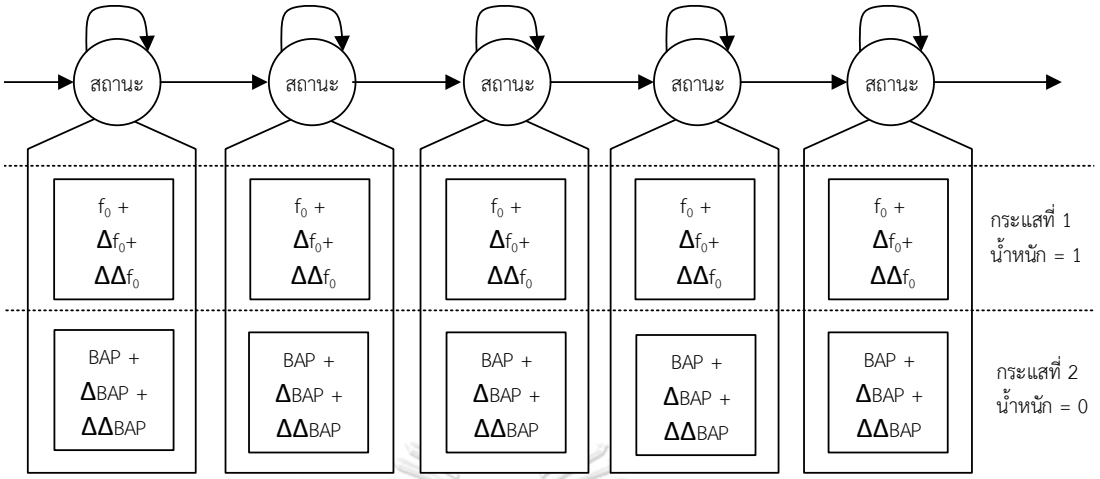
ในแบบจำลองสเปกตรัมมีการเพิ่มกระแสของค่าคุณลักษณะความถี่มูลฐาน แต่มีการกำหนดค่าน้ำหนักของกระแสเป็น 0 ซึ่งจะทำให้ค่าการเปลี่ยนแปลงของค่าความถี่มูลฐานไม่ส่งผลใดๆ กับการเปลี่ยนแปลงของสถานะ เพราะจะนำค่าที่ได้จากค่าคุณลักษณะความถี่มูลฐานไปใช้ในการตรวจสอบความสอดคล้องกันของประเภทของเสียงที่สังเคราะห์มาจากแบบจำลองของสเปกตรัม และแบบจำลองค่าความถี่มูลฐาน ว่าเป็นค่าคุณลักษณะจากช่วงที่เป็นเสียงก้อง หรือช่วงที่เป็นเสียงไม่ก้องเหมือนกันหรือไม่

แบบจำลองสเปกตรัม และแบบจำลองค่าความถี่มูลฐาน แสดงดังรูปที่ 21 และรูปที่ 22 ตามลำดับ

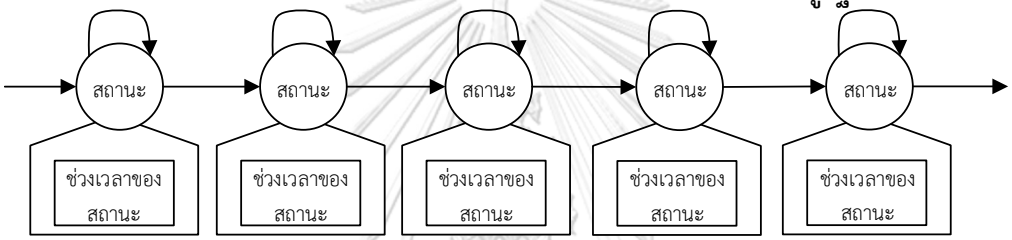


รูปที่ 21 แบบจำลองสเปกตรัม

แบบจำลองเสียงสังเคราะห์แบบฮิดเดนมาร์คอฟสำหรับค่าความถี่มูลฐานที่นำเสนอ



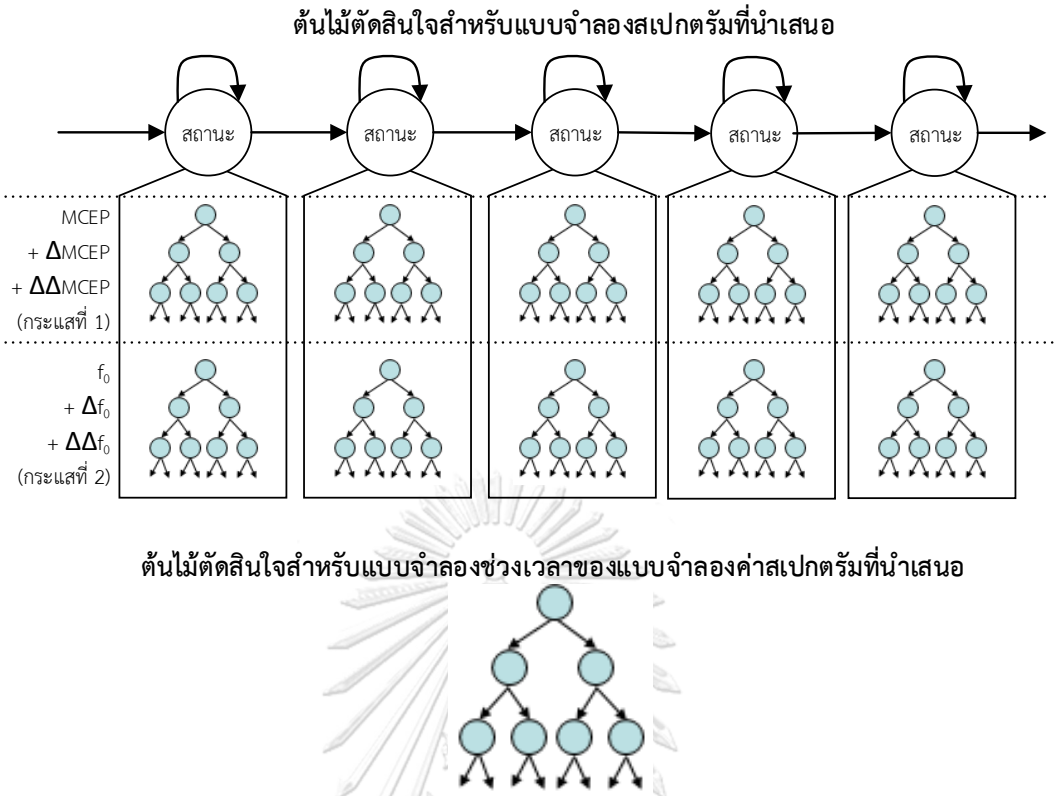
แบบจำลองช่วงเวลาแบบฮิดเดนมาร์คอฟสำหรับค่าความถี่มูลฐาน



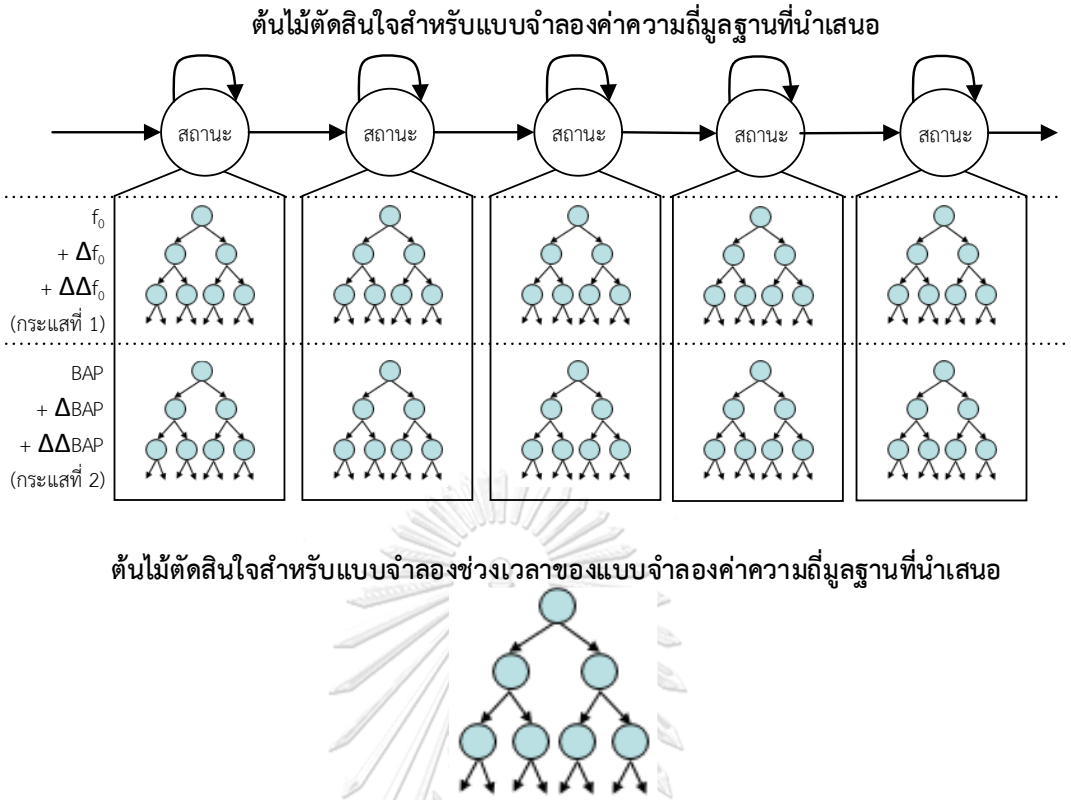
รูปที่ 22 แบบจำลองความถี่มูลฐาน

เนื่องจากแบบจำลองทั้งสองถูกฝึกฝนแยกกัน จึงจำเป็นต้องมีแบบจำลองช่วงเวลาอยู่ในแบบจำลองความถี่มูลฐาน และแบบจำลองสเปกตรัม ดังแสดงในรูปที่ 21 และรูปที่ 22 ซึ่งแบบจำลองช่วงเวลาทั้ง 2 นั้น มีค่าไม่เหมือนกัน

สำหรับต้นไม้ตัดสินใจที่ใช้ในการจัดกลุ่มแบบจำลองเสียง ยังคงใช้หลักการเหมือนกับที่เสนอโดย Black และคณะ. (2007) [2] ที่กำหนดให้ใช้ต้นไม้ตัดสินใจหนึ่งต้นต่อหนึ่งกระแสต่อหนึ่งสถานะสำหรับแบบจำลองสเปกตรัม แบบจำลองค่าความถี่มูลฐาน และหนึ่งต้นสำหรับแบบจำลองช่วงเวลา จึงทำให้ต้นไม้ตัดสินใจสำหรับแบบจำลองสเปกตรัมมีจำนวน 11 ต้น ซึ่งประกอบด้วยต้นไม้ตัดสินใจสำหรับค่าคุณลักษณะสเปกตรัมจำนวน 5 ต้น ค่าความถี่มูลฐานจำนวน 5 ต้น และแบบจำลองช่วงเวลาจำนวน 1 ต้น ดังแสดงในรูปที่ 23 และต้นไม้ตัดสินใจสำหรับแบบจำลองค่าความถี่มูลฐานมีจำนวน 11 ต้น ซึ่งประกอบด้วย ต้นไม้ตัดสินใจสำหรับค่าคุณลักษณะความไม่เป็นคาบของแถบความถี่จำนวน 5 ต้น ค่าความถี่มูลฐานจำนวน 5 ต้น และแบบจำลองช่วงเวลาจำนวน 1 ต้น ดังแสดงในรูปที่ 24



รูปที่ 23 ต้นไม้ตัดสินใจสำหรับแบบจำลองค่าสเปกตรัมที่นำเสนอ



รูปที่ 24 ต้นไม้ตัดสินใจสำหรับแบบจำลองค่าความถี่มูลฐานที่นำเสนอ

4.1.2 การคำนวณหาช่วงเวลาของเสียงสังเคราะห์

ในกรณีที่ใช้แบบจำลองแบบรวมตามที่เสนอโดย Black และคณะ (2007) [2] จะมีแบบจำลองช่วงเวลาเพียง 1 แบบจำลอง ในกรณีที่ต้องการสังเคราะห์เสียงหน่วยเสียงใด ระบบจะทำการค้นหาแบบจำลองช่วงเวลาที่สอดคล้องกับหน่วยเสียงที่ต้องการสังเคราะห์ ซึ่งในแบบจำลองนั้นจะประกอบด้วย 5 สถานะ จากนั้นเลือกใช้ค่ากลางที่อยู่ในแบบจำลองทางสถิติของแบบจำลองช่วงเวลาเป็นค่าความยาวในหน่วยกรอบเวลาของสถานะ

แต่ในกรณีที่ใช้แบบจำลองตามที่งานวิจัยนี้นำเสนอ จะประกอบด้วยแบบจำลองช่วงเวลาจำนวน 2 แบบจำลอง ดังนั้นจึงมีโอกาสที่ผลรวมของช่วงเวลาในทุกๆ สถานะในหน่วยเสียงเดียวกันที่ได้จากแบบจำลองช่วงเวลาของค่าความถี่มูลฐาน และค่าสเปกตรัมมีค่าไม่เท่ากัน และยิ่งไปกว่านั้นการสังเคราะห์ค่าคุณลักษณะจากทั้งสองแบบจำลอง อาจจะได้ค่าคุณลักษณะที่มีสถานะความก้องของเสียง (Voicing Condition) ที่ไม่เหมือนกัน ดังเช่นในตัวอย่างเสียงสังเคราะห์เดียวกัน ค่าคุณลักษณะที่สังเคราะห์มาจากแบบจำลองสเปกตรัมเป็นเสียงที่ไม่ก้อง แต่ค่าคุณลักษณะที่สังเคราะห์มาจากแบบจำลองความถี่มูลฐานเป็นเสียงก้อง

การพิจารณาสถานะความก้องของเสียงของค่าคุณลักษณะสเปกตรัมที่สังเคราะห์ออกมา จะพิจารณาจากค่าความถี่มูลฐานที่สังเคราะห์ออกมาพร้อมกันจากแบบจำลองสเปกตรัม (แต่ไม่ได้นำค่าความถี่มูลฐานส่วนนี้ไปใช้ในการสังเคราะห์เสียง) และค่าความเป็นเสียงของค่าคุณลักษณะความถี่มูลฐาน โดยจะพิจารณาจากค่าความถี่มูลฐานที่สังเคราะห์มาจากแบบจำลองความถี่มูลฐาน

จากข้อกำหนดดังกล่าว ทำให้การสร้างช่วงเวลาจากการใช้หลายแบบจำลองสามารถแบ่งออกตามเงื่อนไขได้ ดังต่อไปนี้

1. กรณีที่ไม่พิจารณาสถานะความก้องของเสียง

ในกรณีนี้สามารถแบ่งแยกออกเป็น 3 วิธีการ ได้แก่ 1) การเลือกใช้แบบจำลองช่วงเวลาจากแบบจำลองของสเปกตรัมเป็นหลัก 2) การเลือกใช้แบบจำลองช่วงเวลาจากแบบจำลองความถี่มูลฐานเป็นหลัก และ3) การใช้แบบจำลองช่วงเวลาจากทั้งสองแบบจำลอง

แนวคิดที่ใช้แบบจำลองช่วงเวลาของแบบจำลองสเปกตรัมเป็นหลัก จะคำนวณหาช่วงเวลาของหน่วยเสียงจากแบบจำลองช่วงเวลาของแบบจำลองสเปกตรัม จากนั้นให้แบบจำลองช่วงเวลาของแบบจำลองความถี่มูลฐานคำนวณความยาวของแต่ละสถานะโดยการใช้กฎความน่าจะเป็นสูงสุด และให้ผลรวมของความยาวในทุกสถานะมีค่าเท่ากับความยาวของหน่วยเสียงที่คำนวณไว้จากแบบจำลองสเปกตรัม

แนวคิดที่ใช้แบบจำลองช่วงเวลาของแบบจำลองความถี่มูลฐานเป็นหลัก จะคำนวณหาช่วงเวลาของหน่วยเสียงจากแบบจำลองช่วงเวลาของแบบจำลองความถี่มูลฐาน จากนั้นให้แบบจำลองช่วงเวลาของแบบจำลองสเปกตรัมคำนวณความยาวของแต่ละสถานะโดยการใช้กฎความน่าจะเป็นสูงสุด และให้ผลรวมของความยาวในทุกสถานะมีค่าเท่ากับความยาวของหน่วยเสียงที่คำนวณไว้จากแบบจำลองความถี่มูลฐาน

แนวคิดการใช้แบบจำลองช่วงเวลาจากทั้งสองแบบจำลอง จะคำนวณค่าความยาวของแต่ละหน่วยเสียง และความยาวของแต่ละสถานะของทั้งสองแบบจำลองจากสมการที่ 4 และ 5 โดยที่ d_{si} คือค่าของช่วงเวลาในสถานะที่ i ของแบบจำลองสเปกตรัม, d_{fi} คือค่าของช่วงเวลาในสถานะที่ i ของแบบจำลองค่าความถี่มูลฐาน, λ_{ds_i} คือแบบจำลองทางสถิติของแบบจำลองช่วงเวลาในสถานะที่ i ของแบบจำลองสเปกตรัม และ λ_{df_i} คือแบบจำลองทางสถิติของแบบจำลองช่วงเวลาในสถานะที่ i ของแบบจำลองความถี่มูลฐาน

$$P(\{D_s, D_f\}) = \prod_{i=1}^{n} P(d_{si}|\lambda_{ds_i}) P\left(d_{fi}|\lambda_{d_{fi}}\right) \quad (4)$$

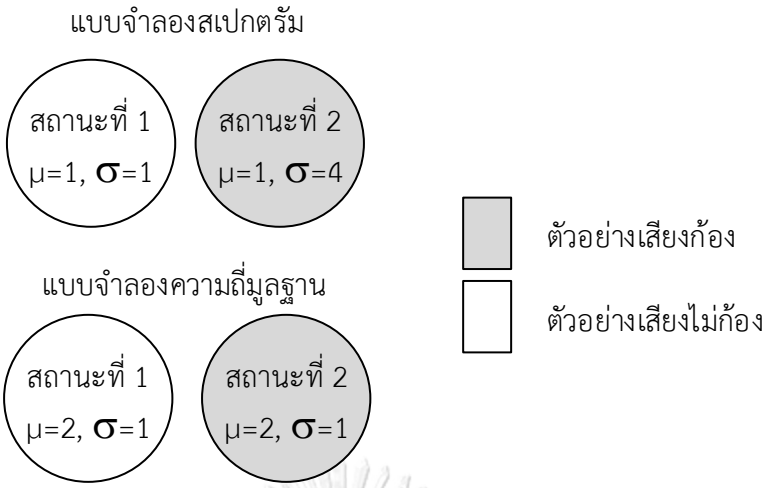
โดยที่ $D_s = \{d_{si}, \dots, d_{sn}\}$ และ $D_f = \{d_{fi}, \dots, d_{fn}\}$

$$\{D_s, D_f\} = \underset{\{D_s, D_f\}}{\arg\max}\, P\big(\{D_s, D_f\}\big) \quad \text{โดยที่ } \sum_{i=1}^{n} d_{si} = \sum_{i=1}^{n} d_{fi} \tag{5}$$

วิธีการคำนวณค่าของตัวแปร $d_{s1}, \dots, d_{sn}, d_{f1}, \dots, d_{fn}$ มีขั้นตอน ดังตารางที่ 2

ตารางที่ 2 การคำนวณหาความยาวของสถานะในกรณีไม่พิจารณาสถานะความก้องของเสียง

ขั้นตอน	รายละเอียด
1	กำหนดค่าเริ่มต้นของ $d_{s1}, \dots, d_{sn}, d_{f1}, \dots, d_{fn}$ ให้มีค่าเท่ากับค่ากลางของแบบจำลองช่วงเวลาที่สอดคล้องกัน
2	ตรวจสอบผลรวมของ $d_{s1}, \dots, d_{sn}$ เปรียบเทียบกับ $d_{f1}, \dots, d_{fn}$ ถ้ากรณีที่มีค่าความยาวเท่ากันให้ถือว่าสิ้นสุดกระบวนการ แต่ถ้าความยาวไม่เท่ากันให้ทำขั้นตอนต่อไป
3	ในกรณีที่ผลรวมของ $d_{s1}, \dots, d_{sn}$ มากกว่าค่าผลรวมของ $d_{f1}, \dots, d_{fn}$ ให้ลดค่า $d_{s1}, \dots, d_{sn}$ จำนวน 1 กรอบเวลา และในกรณีที่ผลรวมมีค่าน้อยกว่า ให้เพิ่มค่า $d_{s1}, \dots, d_{sn}$ จำนวน 1 กรอบเวลา โดยทำการเพิ่มค่าหรือลดค่าทีละตัวแปร และทำการบันทึกรูปแบบทั้งหมด เช่นในกรณีที่มี d_{s1} ถึง d_{s5} จะทำให้เกิดรูปแบบใหม่ขึ้นมา จำนวน 5 รูปแบบ
4	ในกรณีที่ผลรวมของ $d_{s1}, \dots, d_{sn}$ มากกว่าค่าผลรวมของ $d_{f1}, \dots, d_{fn}$ ให้เพิ่มค่า $d_{f1}, \dots, d_{fn}$ จำนวน 1 กรอบเวลา และในกรณีที่ผลรวมมีค่าน้อยกว่า ให้ลดค่า $d_{f1}, \dots, d_{fn}$ จำนวน 1 กรอบเวลา โดยทำการเพิ่มค่าหรือลดค่าทีละตัวแปร และทำการบันทึกรูปแบบทั้งหมด เช่นในกรณีที่มี d_{f1} ถึง d_{f5} จะทำให้เกิดรูปแบบใหม่ขึ้นมา จำนวน 5 รูปแบบ
5	นำรูปแบบที่ได้จากขั้นตอนที่ 3 และ 4 มาทำการคำนวณค่าความน่าจะเป็นตามสมการที่ 4 และสมการที่ 5
6	เลือกรูปแบบที่ทำให้ค่าความน่าจะเป็นในกระบวนการที่ 5 มีค่ามากที่สุด เป็นค่าตัวแปร $d_{s1}, \dots, d_{sn}, d_{f1}, \dots, d_{fn}$ แล้วทำการวนซ้ำในขั้นตอนที่ 2



รูปที่ 25 กรณีตัวอย่างในการคำนวณช่วงเวลาของสถานะแบบไม่พิจารณาสถานะความก้องของเสียง

ตัวอย่างของการคำนวณวิธีการหาช่วงเวลาของสถานะตามตารางที่ 2 โดยมีแบบจำลองช่วงเวลาของแบบจำลองสเปกตรัม และแบบจำลองความถี่แบบจำลองละ 2 สถานะ ดังแสดงในรูปที่ 25 เป็นดังตารางที่ 3

ความหมายของค่าในหัวแถวของตารางที่ 3 เป็นดังต่อไปนี้

- d_{s1} และ d_{s2} แทนถึงช่วงเวลาของสถานะของแบบจำลองสเปกตรัมในสถานะที่ 1 และ 2 ตามลำดับ
- d_{f1} และ d_{f2}แทนถึงช่วงเวลาของสถานะของแบบจำลองความถี่มูลฐานในสถานะที่ 1 และ 2 ตามลำดับ
- P_{s1} และ P_{s2} แทนถึงค่าค่าความน่าจะเป็นที่จากการที่ใช้ค่า d_{s1} และ d_{s2}เป็นช่วงเวลาของสถานะของแบบจำลองสเปกตรัมในสถานะที่ 1 และ 2 ตามลำดับ
- P_{f1} และ P_{f2} แทนถึงค่าค่าความน่าจะเป็นที่จากการที่ใช้ค่า d_{f1} และ d_{f2}เป็นช่วงเวลาของสถานะของแบบจำลองความถี่มูลฐานในสถานะที่ 1 และ 2 ตามลำดับ
- P หมายถึงความน่าจะเปิดรวมที่เกิดจากการนำค่า P_{s1}, P_{s2}, P_{f1} และ P_{f2} มาทำการคูณกัน

ในรอบที่ 0 หมายถึง รอบเริ่มต้น ซึ่งกำหนดให้ค่าเริ่มต้นเป็นค่ากลางในแต่ละสถานะ และเมื่อเริ่มรอบที่ 1 พบว่า ผลรวมของสถานะของแบบจำลองสเปกตรัม (d_{s1} และ d_{s2}) มีค่าน้อยกว่าผลรวมของสถานะของแบบจำลองความถี่มูลฐาน (d_{f1} และ d_{f2}) ดังนั้น จึงต้องทำการเพิ่มสถานะให้กับสถานะของแบบจำลองสเปกตรัมรูปแบบละ 1 กรอบเวลา ซึ่งทำให้เกิดรูปแบบขึ้นมา 2 รูปแบบ (ตามจำนวนสถานะ) คือ รูปแบบที่ 1 ของรอบที่ 1 ที่เพิ่มค่า d_{s1} และรูปแบบที่ 2 ของรอบที่ 1 ที่เพิ่มค่า

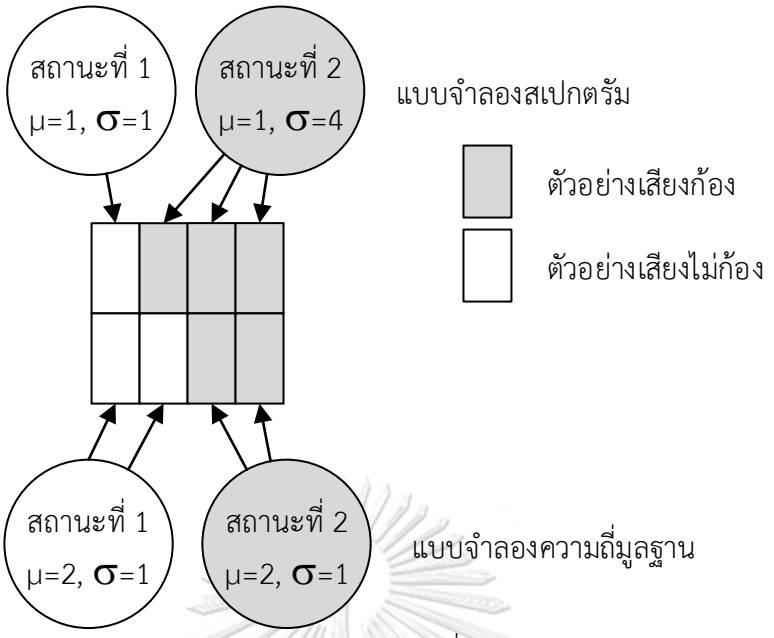
d_{s2} และสำหรับในกรณีของช่วงเวลาของสถานะในแบบจำลองความถี่มูลฐานต้องทำการลดค่าลงรูปแบบละ 1 กรอบเวลา ซึ่งจะทำให้เกิดรูปแบบขึ้นมา 2 รูปแบบ (ตามจำนวนสถานะ) คือ รูปแบบที่ 3 ของรอบที่ 1 ที่ทำการลดช่วงเวลาของสถานะ d_{f1} และรูปแบบที่ 4 ของรอบที่ 1 ที่ทำการลดช่วงเวลาของสถานะ d_{f2}

หลังจากสร้างรูปแบบทั้ง 4 ของรอบที่ 1 แล้วเสร็จให้ทำการคำนวณหาค่า P ของทุกรูปแบบ และเลือกรูปแบบที่มีค่าความน่าจะเป็นสูงที่สุดนำมาใช้ต่อ โดยหลังจากจบรอบที่ 1 พบว่าค่าผลรวมของกรอบเวลาในแบบจำลองสเปกตรัม และแบบจำลองความถี่มูลฐานมีค่าไม่เท่ากัน ดังนั้นจึงจำเป็นต้องทำซ้ำในกระบวนการที่เหมือนกับรอบที่ 1 ใหม่ โดยใช้รูปแบบที่ 2 ที่ได้ความน่าจะเป็นสูงสุด เป็นค่าเริ่มต้นของแบบที่ 2

เมื่อคำนวณค่า P ของรอบที่ 2 แล้วเสร็จ พบว่าผลรวมของกรอบเวลาในแบบจำลองสเปกตรัม และแบบจำลองความถี่มูลฐานมีค่าเท่ากัน ดังนั้นจึงใช้รูปแบบที่ได้ค่าความน่าจะเป็นมากที่สุดมาใช้เป็นคำตอบ ซึ่งคือรูปแบบที่ 2 ของรอบที่ 2 ซึ่งมีรายละเอียดดังรูปที่ 26

ตารางที่ 3 ตัวอย่างการคำนวณช่วงเวลาของสถานะในกรณีที่ไม่พิจารณาสถานะความก้องของเสียง

รอบ	รูปแบบ	d_{s1}	d_{s2}	d_{f1}	d_{f2}	P_{s1}	P_{s2}	P_{f1}	P_{f2}	P
0	1	1	1	2	2	-	-	-	-	-
1	1	2	1	2	2	0.24	0.10	0.40	0.40	0.0038
	2	1	2	2	2	0.40	0.10	0.40	0.40	0.0064
	3	1	1	1	2	0.40	0.10	0.24	0.40	0.0038
	4	1	1	2	1	0.40	0.10	0.40	0.24	0.0038
2	1	2	2	2	2	0.24	0.10	0.40	0.40	0.0038
	2	1	3	2	2	0.40	0.09	0.40	0.40	0.0058
	3	1	2	1	2	0.40	0.10	0.24	0.40	0.0038
	4	1	2	2	1	0.40	0.10	0.40	0.24	0.0038

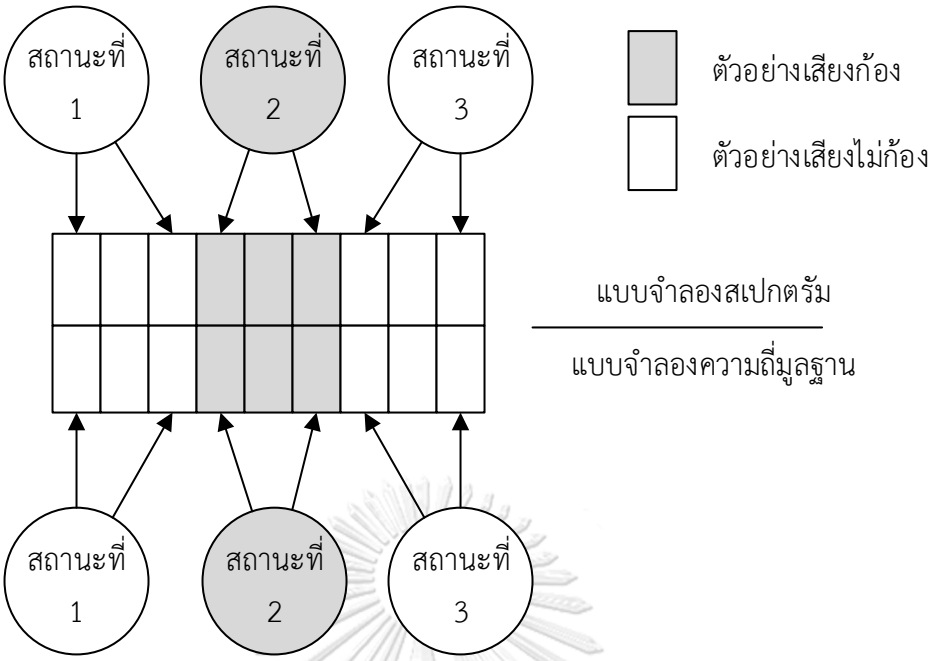


รูปที่ 26 ผลลัพธ์การหาช่วงเวลาของสถานะกรณีที่ไม่พิจารณาสถานะความก้องของเสียง

2. กรณีที่พิจารณาสถานะความก้องของเสียง

ในกรณีที่พิจารณาสถานะความก้องของเสียง จำเป็นต้องมีเงื่อนไขในการตรวจสอบสถานะความก้องของเสียงเพิ่มเข้ามาในสมการที่ 2 โดยเงื่อนไขที่เพิ่มเข้ามา คือ ค่าคุณลักษณะที่สังเคราะห์จากทุกตัวอย่างของทั้งสองแบบจำลองต้องมีสถานะความก้องของเสียงที่เหมือนกัน

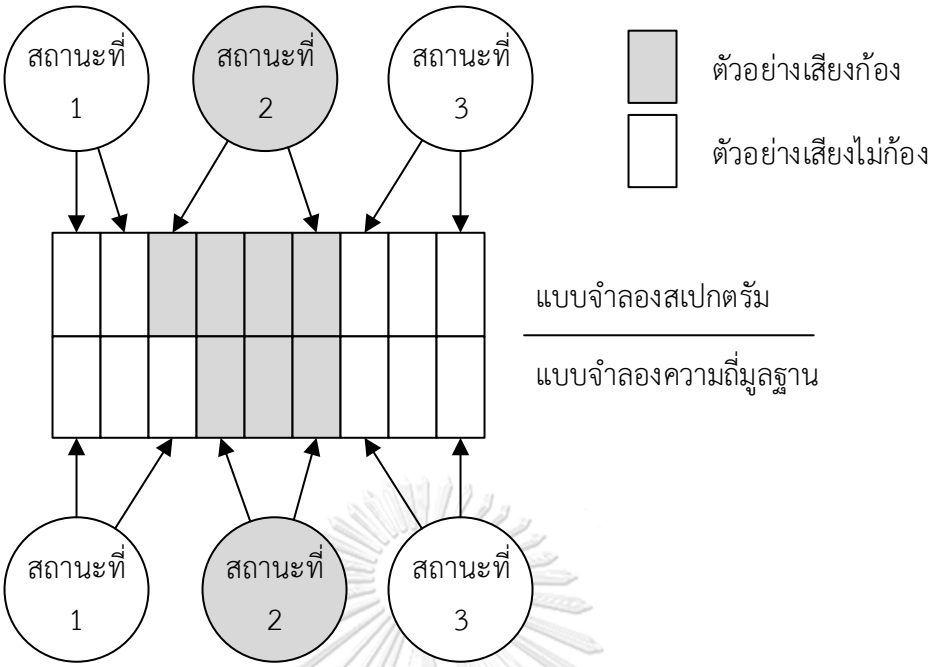
ในรูปที่ 27 ที่ตัวอย่างที่ค่าคุณลักษณะที่สังเคราะห์จากทุกตัวอย่างของทั้งสองแบบจำลองมีสถานะความก้องของเสียงที่เหมือนกัน โดยในรูปที่ 27 แสดงการสังเคราะห์ค่าคุณลักษณะออกมาจากทั้ง 2 แบบจำลอง โดยในทุกๆ สถานะของทั้ง 2 แบบจำลอง สังเคราะห์ออกมาเป็นจำนวน 3 ตัวอย่าง และในสถานะที่ 2 ของทั้ง 2 แบบจำลองเป็นเสียงก้อง ส่วนสถานะอื่นเป็นเสียงไม่ก้อง โดยเมื่อนำตัวอย่างที่สังเคราะห์ออกมาทั้งหมดมาเรียงกัน จะได้ตัวอย่างที่มีสถานะของเสียงตรงกันทั้งหมด



รูปที่ 27 ตัวอย่างกรณีที่มีสถานะความก้องของเสียงที่เหมือนกัน

ในรูปที่ 28 ที่ค่าคุณลักษณะมีสถานะความก้องของเสียงที่ต่างกันในตัวอย่างที่ 3 (จากทางซ้าย) โดยในตัวอย่างที่ 3 ค่าคุณลักษณะของแบบจำลองสเปกตรัมจะมาจากสถานะที่ 2 ที่เป็นเสียงก้อง แต่ค่าคุณลักษณะความถี่มูลฐานมาจากสถานะที่ 1 ที่เป็นตัวอย่างเสียงไม่ก้อง ซึ่งถ้าเกิดในกรณีตามรูปที่ 28 ต้องทำการแก้ไขเพื่อให้ได้ผลลัพธ์ที่มีสถานะความก้องของเสียงเหมือนกัน

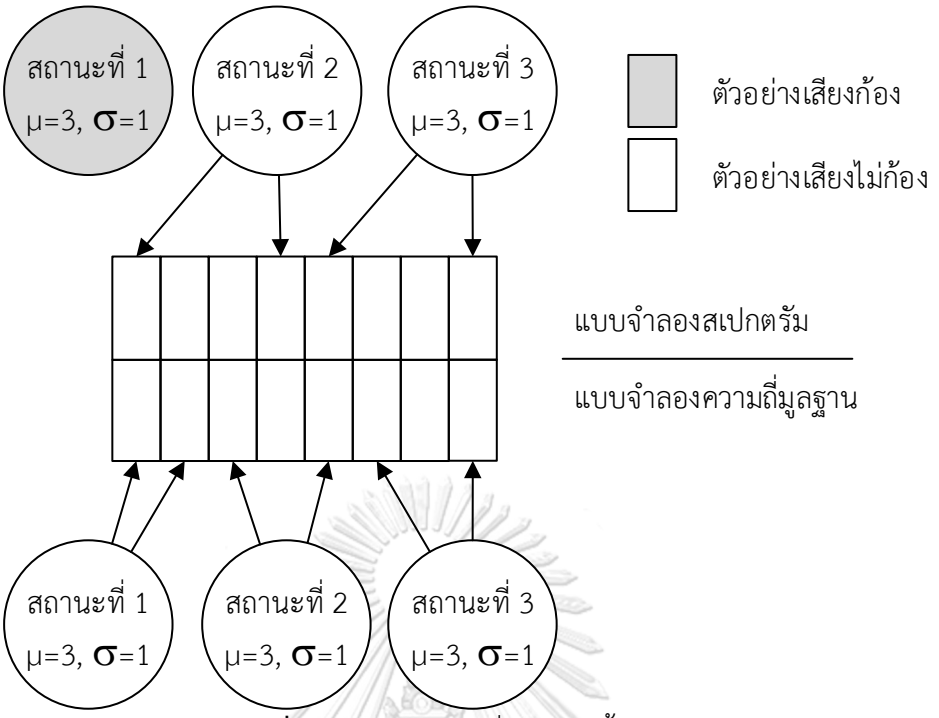
วิธีการแก้ไขอาจจะเป็นได้ 2 กรณี คือ รูปแบบแรกให้กรอบเวลาลำดับที่ 3 ของแบบจำลองสเปกตรัม ให้ใช้ค่าคุณลักษณะจากสถานะที่ 1 แทน ซึ่งจะทำให้ได้ผลลัพธ์เหมือนกับรูปที่ 27 หรือรูปแบบที่สองให้กรอบเวลาลำดับที่ 3 ของแบบจำลองความถี่มูลฐาน ให้ใช้ค่าคุณลักษณะจากสถานะที่ 2 แทน โดยการตัดสินใจว่าจะเลือกใช้รูปแบบแรก หรือรูปแบบที่สอง ให้พิจารณาจากค่าความจะเป็นจากสมการที่ 3 ว่ากรณีใดที่ให้ผลความน่าจะเป็นที่ดีกว่า และเลือกใช้กรณีนั้น



รูปที่ 28 ตัวอย่างกรณีที่มีสถานะความก้องของเสียงที่ไม่เหมือนกัน

มีบางกรณีที่ไม่สามารถสังเคราะห์ตัวอย่างจากทุกสถานะที่สอดคล้องกับเงื่อนไขดังกล่าวได้ ดังเช่นตัวอย่างในรูปที่ 29 ที่มีเพียงสถานะที่ 1 ของแบบจำลองสเปกตรัมที่เป็นเสียงก้องเท่านั้น ทำให้ไม่สามารถสังเคราะห์เสียงจากสถานะนั้นได้ เพราะถ้ามีกรอบเวลาจากสถานะนั้นถูกสังเคราะห์ออกมา จะส่งผลให้ค่าสถานะความก้องของเสียงไม่สอดคล้องกัน โดยเรียกกรณีที่สถานะมีช่วงเวลาเป็น 0 ว่ามีการละทิ้งสถานะ

เมื่อพิจารณากรณีที่เกิดการละทิ้งสถานะตามรูปที่ 29 โดยสมมติให้แบบจำลองสเปกตรัมมีเพียงสถานะที่ 2 และ 3 เพราะสถานะที่ 1 ถูกละทิ้ง โดยทุกสถานะมีค่ากลางเท่ากับ 3 และมีค่าความแปรปรวนเท่ากับ 1 และใช้กระบวนการคำนวณระยะเวลาของสถานะตามในกรณีของการไม่ละทิ้งสถานะ ตามตารางที่ 2 จะได้ผลลัพธ์ที่มีความยาวของทั้งหน่วยเสียงเป็น 8 กรอบเวลา แต่ถ้านำสถานะที่ 1 ไปคิดรวมด้วย (ในกรณีไม่พิจารณาสถานะความก้องของเสียง) ได้จะได้ผลลัพธ์ที่มีความยาวของทั้งหน่วยเสียงเป็น 9 กรอบเวลา ซึ่งแสดงให้เห็นว่าถ้ามีการละทิ้งของสถานะ และใช้วิธีการคำนวณช่วงเวลาของสถานะตามที่เสนอในตารางที่ 2 จะส่งผลให้ได้ความยาวของทั้งหน่วยเสียงสั้นลง



รูปที่ 29 ตัวอย่างกรณีที่มีการละทิ้งสถานะ

เพื่อไม่ให้ค่าความยาวของตัวอย่างเสียงสังเคราะห์มีค่าน้อยกว่าที่ควรจะเป็นเมื่อเกิดกรณีที่ต้องละทิ้งบางสถานะ ดังนั้นในกรณีที่พิจารณาสถานะความก้องของเสียง จึงจำเป็นต้องมีการกำหนดค่าความยาวเป้าหมายไว้ โดยค่าความยาวเป้าหมายนั้นคำนวณจากกระบวนการตามขั้นตอนในตารางที่ 2 ในกรณีที่ไม่พิจารณาสถานะความก้องของเสียง

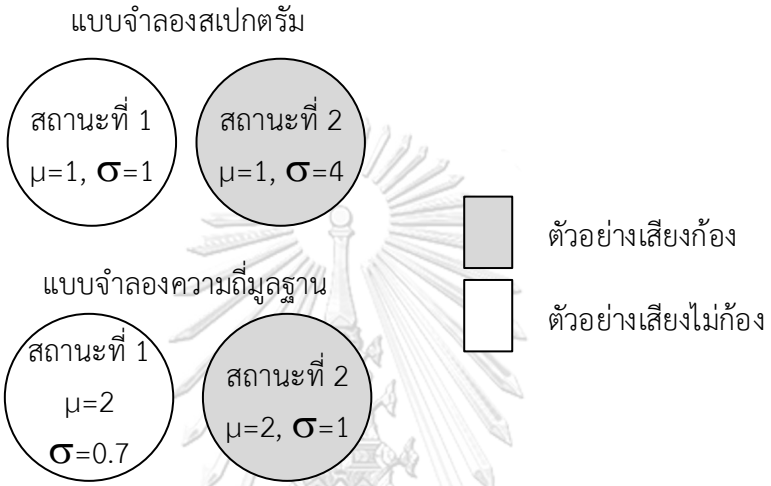
หลังจากที่ได้ความยาวเป้าหมายแล้ว ขั้นตอนต่อไปคือการกำหนดความยาวให้กับแต่ละสถานะ ซึ่งมีกระบวนการดังตารางที่ 3 ซึ่งมีแนวคิดการคำนวณที่แตกต่างจากที่นำเสนอในตารางที่ 2 ของกรณีไม่พิจารณาสถานะความเป็นเสียง โดยกำหนดให้ช่วงเวลาเริ่มต้นของทุกสถานะจากแบบจำลองสเปกตรัม และความถี่มูลฐานมีค่าเป็น 0 จากนั้นทำการเพิ่มค่าของกรอบเวลาในแต่ละสถานะของทั้งสองแบบจำลองพร้อมๆ กัน โดยทำการเพิ่มรอบละ 1 กรอบเวลา และทำการตรวจสอบสถานะความก้องของเสียงของแต่ละรูปแบบที่สร้างขึ้นมาจากแต่ละรอบ และละทิ้งรูปแบบที่มีสถานะความก้องของเสียงไม่สอดคล้องกัน จากนั้นเลือกรูปแบบที่ได้ค่าความน่าจะเป็นสูงที่สุด เพื่อใช้เป็นรูปแบบเริ่มต้นของรอบถัดไป และทำซ้ำกระบวนการนี้จนกว่าความยาวของหน่วยเสียงเท่ากับความยาวเป้าหมาย

ตารางที่ 4 การคำนวณหาความยาวของสถานะในกรณีที่พิจารณาสถานะความก้องของเสียง

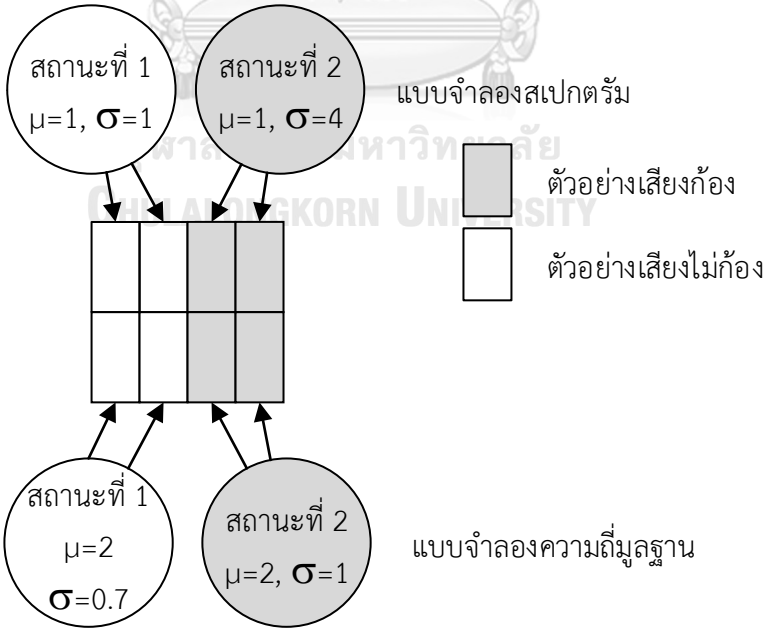
ลำดับ	รายละเอียด
1	หาความยาวของหน่วยเสียงจากวิธีการในหัวข้อที่ 1 แต่ไม่ได้นำผลลัพธ์ในส่วนของความยาวในแต่ละสถานะมาใช้ โดยเรียกความยาวของหน่วยเสียงที่คำนวณจากขั้นตอนนี้ว่าความยาวเป้าหมาย
2	สร้างรายการเพื่อใช้ในการเก็บรูปแบบความยาวของแต่ละสถานะของทั้ง 2 แบบจำลอง โดยกำหนดให้ค่าเริ่มต้นมีเพียง 1 รายการ ที่มีความยาวของทุกสถานะของทั้ง 2 แบบจำลองให้มีค่าเท่ากับ 0 โดยเรียกรายการนี้ว่ารายการรูปแบบของรอบก่อนหน้า
3	ตรวจสอบผลรวมของความยาวในทุกสถานะของทั้ง 2 แบบจำลองว่ามีความยาวเท่ากับความยาวเป้าหมายหรือไม่ ถ้ามีขนาดเท่ากับความยาวเป้าหมายให้สิ้นสุดกระบวนการ ถ้ามีความยาวน้อยกว่าความยาวเป้าหมายให้ทำขั้นตอนต่อไป
4	นำทุกรายการของรายการรูปแบบของรอบก่อนหน้า มาทำการสร้างรูปแบบใหม่ โดยทำการเพิ่มความยาวให้กับทั้งสองแบบจำลอง แบบจำลองละ 1 กรอบเวลา ตัวอย่างเช่นในกรณีที่รายการรูปแบบของรอบก่อนหน้ามีข้อมูลอยู่ 5 รายการ และทั้ง 2 แบบจำลองมี 3 สถานะ จะทำให้สร้างรูปแบบใหม่ขึ้นมาได้ 45 (3 x 3 x 5) รูปแบบ จากการเลือกที่จะเพิ่มในแบบจำลองแรก 3 รูปแบบ และเลือกที่จะเพิ่มในแบบจำลองที่ 2 ได้ 3 รูปแบบ และมีรูปแบบในรายการรูปแบบของรอบก่อนหน้าอยู่ 5 รายการ
5	นำรายการทั้งหมดที่สร้างจากรายการที่ 4 มาทำการตรวจสอบเงื่อนไขของความเป็นเสียง โดยทิ้งรูปแบบที่ไม่ผ่านเงื่อนไขของความเป็นเสียง
6	นำรายการจากขั้นตอนที่ 5 ไปคำนวณค่าความน่าจะเป็นตามสมการที่ 3 และ 4
7	จัดเก็บรายการจากขั้นตอนที่ 6 จำนวน n รายการเรียงตามค่าความน่าจะเป็น ลงในรายการรูปแบบของรอบก่อนหน้าแทนที่รายการเดิม โดยค่า n นี้เรียกว่าจำนวนจัดเก็บรูปแบบ
8	วนซ้ำขั้นตอนที่ 3

ตัวอย่างการคำนวณด้วยกระบวนการตามตารางที่ 4 สถานะของแบบจำลองในกรณีที่มีแบบจำลองช่วงเวลาเป็นตามรูปที่ 30 และมีความยาวเป้าหมายที่ 4 กรอบเวลา มีขั้นตอนแสดงดังตารางที่ 5 โดยในรอบที่ 0 กำหนดให้ค่าความยาวของทุกสถานะเป็น 0 จากนั้นในรอบที่ 1 กำหนดให้เพิ่มความยาวของสถานะในแต่ละแบบจำลอง แบบจำลองละ 1 กรอบเวลา ซึ่งสามารถสร้างรูปแบบออกมาได้ทั้งหมด 4 รูปแบบ จากนั้นทำการตรวจสอบรูปแบบว่าสอดคล้องกับเงื่อนไขของสถานะ

ความก้องของเสียงหรือไม่ โดยจะทำการคำนวณค่าความน่าจะเป็นให้กับกรณีที่มีสถานะความก้องของเสียงสอดคล้องกันเท่านั้น (กรณีที่เป็นตัวอักษรหนา) จากนั้นเลือกรูปแบบที่ให้ค่าความน่าจะเป็นสูงสุดไปใช้ไปต้นแบบของรอบถัดไป โดยจะต้องทำการวนซ้ำกระบวนการนี้เป็นจำนวน 4 รอบ (วนซ้ำเท่ากับความยาวเป้าหมาย) จะทำให้ได้ผลลัพธ์ตามรูปที่ 31 ซึ่งมีความสอดคล้องกันของสถานะความก้องของเสียงในทุกกรอบเวลา



รูปที่ 30 กรณีตัวอย่างในการคำนวณช่วงเวลาของสถานะแบบพิจารณาสถานะความก้องของเสียง



รูปที่ 31 ผลลัพธ์การหาช่วงเวลาของสถานะกรณีที่พิจารณาสถานะความก้องของเสียง

ตารางที่ 5 ตัวอย่างการคำนวณช่วงเวลาของสถานะในกรณีที่พิจารณาสถานะความก้องของเสียง

รอบ	รูปแบบ	d_{s1}	d_{s2}	d_{f1}	d_{f2}	P_{s1}	P_{s2}	P_{f1}	P_{f2}	P
0	1	0	0	0	0	-	-	-	-	-
1	**1**	**1**	**0**	**1**	**0**	**0.40**	**0.10**	**0.21**	**0.05**	**0.0004**
	2	1	0	0	1	-	-	-	-	-
	3	0	1	1	0	-	-	-	-	-
	4	**0**	**1**	**0**	**1**	**0.24**	**0.10**	**0.01**	**0.24**	**0.0000**
2	**1**	**2**	**0**	**2**	**0**	**0.24**	**0.10**	**0.57**	**0.05**	**0.0007**
	2	2	0	1	1	-	-	-	-	-
	3	1	1	2	0	-	-	-	-	-
	4	**1**	**1**	**1**	**1**	**0.40**	**0.10**	**0.21**	**0.24**	**0.0020**
3	**1**	**2**	**1**	**2**	**1**	**0.24**	**0.10**	**0.57**	**0.24**	**0.0033**
	2	2	1	1	2	-	-	-	-	-
	3	1	2	2	1	-	-	-	-	-
	4	**1**	**2**	**1**	**2**	**0.40**	**0.10**	**0.21**	**0.40**	**0.0034**
4	**1**	**2**	**2**	**2**	**2**	**0.24**	**0.10**	**0.57**	**0.40**	**0.0055**
	2	2	2	1	3	-	-	-	-	-
	3	1	3	2	2	-	-	-	-	-
	4	**1**	**3**	**1**	**3**	**0.40**	**0.09**	**0.21**	**0.24**	**0.0018**

4.2 ปรับเปลี่ยนค่าคุณลักษณะ และวิธีการฝึกฝนโครงข่ายประสาทเทียมแบบลึก

เนื่องจากการสร้างแบบจำลองเสียงสังเคราะห์ด้วยแบบจำลองโครงข่ายประสาทเทียมแบบลึกยังไม่มีการนำความรู้ และข้อมูลจากแบบจำลองฮิดเดนมาร์คอฟมาใช้เท่าที่ควร ทั้งที่ในการสร้างเสียงสังเคราะห์ด้วยแบบจำลองโครงข่ายประสาทเทียมแบบลึกจำเป็นต้องใช้ข้อมูลบางส่วนจากแบบจำลองฮิดเดนมาร์คอฟ ดังนั้นในงานวิจัยนี้ จึงได้นำเสนอแนวคิดการใช้วิธีการนำข้อมูลจากแบบจำลองฮิดเดนมาร์คอฟมาใช้ในการฝึกฝนโครงข่ายประสาทเทียมแบบลึกเพิ่มมากขึ้น ซึ่งมีรายละเอียดดังต่อไปนี้

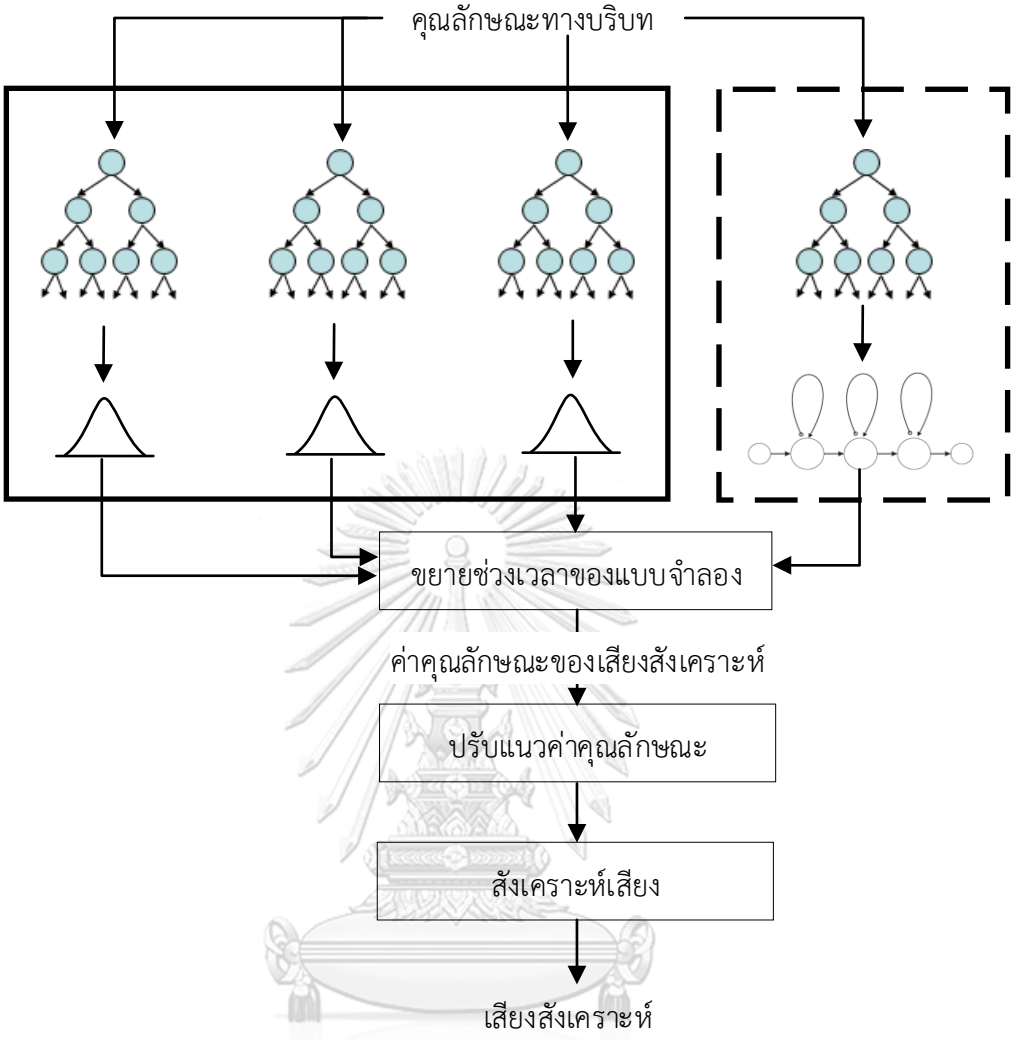
4.2.1 การปรับค่าคุณลักษณะส่วนรับเข้า

เมื่อเปรียบเทียบส่วนประกอบในการสังเคราะห์เสียงของแบบจำลองฮิดเดนมาร์คอฟตามรูปที่ 28 และแบบจำลองโครงข่ายประสาทเทียมแบบลึกตามรูปที่ 29 พบว่าส่วนที่อยู่ในกรอบสี่เหลี่ยมที่มีขอบเป็นเส้นทึบหนา (ด้านซ้าย) คือส่วนที่แตกต่างกันระหว่างแบบจำลองฮิดเดนมาร์คอฟ และแบบจำลองโครงข่ายประสาทเทียมแบบลึก ซึ่งส่วนดังกล่าวทำหน้าที่ในการเชื่อมโยงค่าคุณลักษณะ

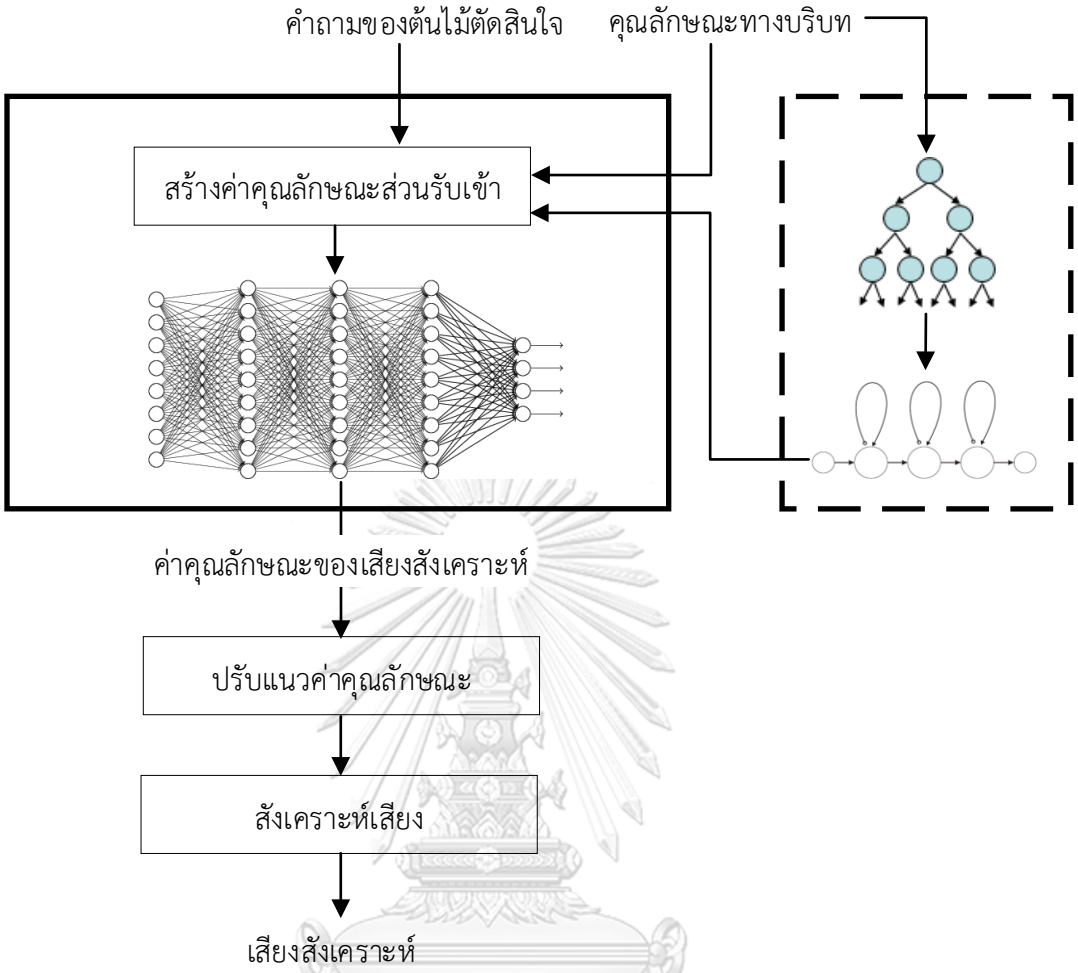
ทางบริบท ไปยังค่าคุณลักษณะทางเสียง และส่วนของการขยายช่วงเวลาของแบบจำลองที่จะมีเฉพาะแบบจำลองฮิดเดนมาร์คอฟ สำหรับในกรอบที่มีขอบเป็นเส้นประหนาของรูปที่ ***28***32 และ รูปที่ ***29***33 แทนถึงการค้นหาช่วงเวลาของแต่ละสถานะที่สอดคล้องกับบริบทของเสียงที่ต้องการสังเคราะห์ผ่านทางต้นไม้ตัดสินใจ

ในกรณีของแบบจำลองฮิดเดนมาร์คอฟค่าคุณลักษณะทางบริบทของเสียงที่ต้องการสังเคราะห์ จะถูกนำไปค้นหาแบบจำลองที่สอดคล้องกับค่าบริบทดังกล่าวผ่านทางต้นไม้ตัดสินใจ โดยใช้ค่ากลางของแบบจำลองแทนค่าของทั้งสถานะนั้น และในกระบวนการขยายช่วงเวลาของแบบจำลอง จะทำการคัดลอกค่ากลางของแต่ละสถานะให้มีจำนวนเท่ากับค่ากลางในสถานะที่สอดคล้องกันของแบบจำลองช่วงเวลา

แต่สำหรับแบบจำลองโครงข่ายประสาทเทียมแบบลึก จะนำผลลัพธ์ของแบบจำลองช่วงเวลาไปใช้ในการสร้างเป็นส่วนหนึ่งของค่าคุณลักษณะส่วนรับเข้า และใช้ในการกำหนดจำนวนของกรอบเวลาของเสียงสังเคราะห์ โดยถ้าในกรณีที่แบบจำลองโครงข่ายประสาทเทียมแบบลึก และแบบจำลองฮิดเดนมาร์คอฟใช้แบบจำลองช่วงเวลาเหมือนกัน จะทำให้ได้เสียงสังเคราะห์ที่มีช่วงเวลาในแต่ละสถานะที่เหมือนกัน



รูปที่ 32 ส่วนประกอบของการสังเคราะห์เสียงด้วยแบบจำลองฮิดเดนมาร์คอฟ



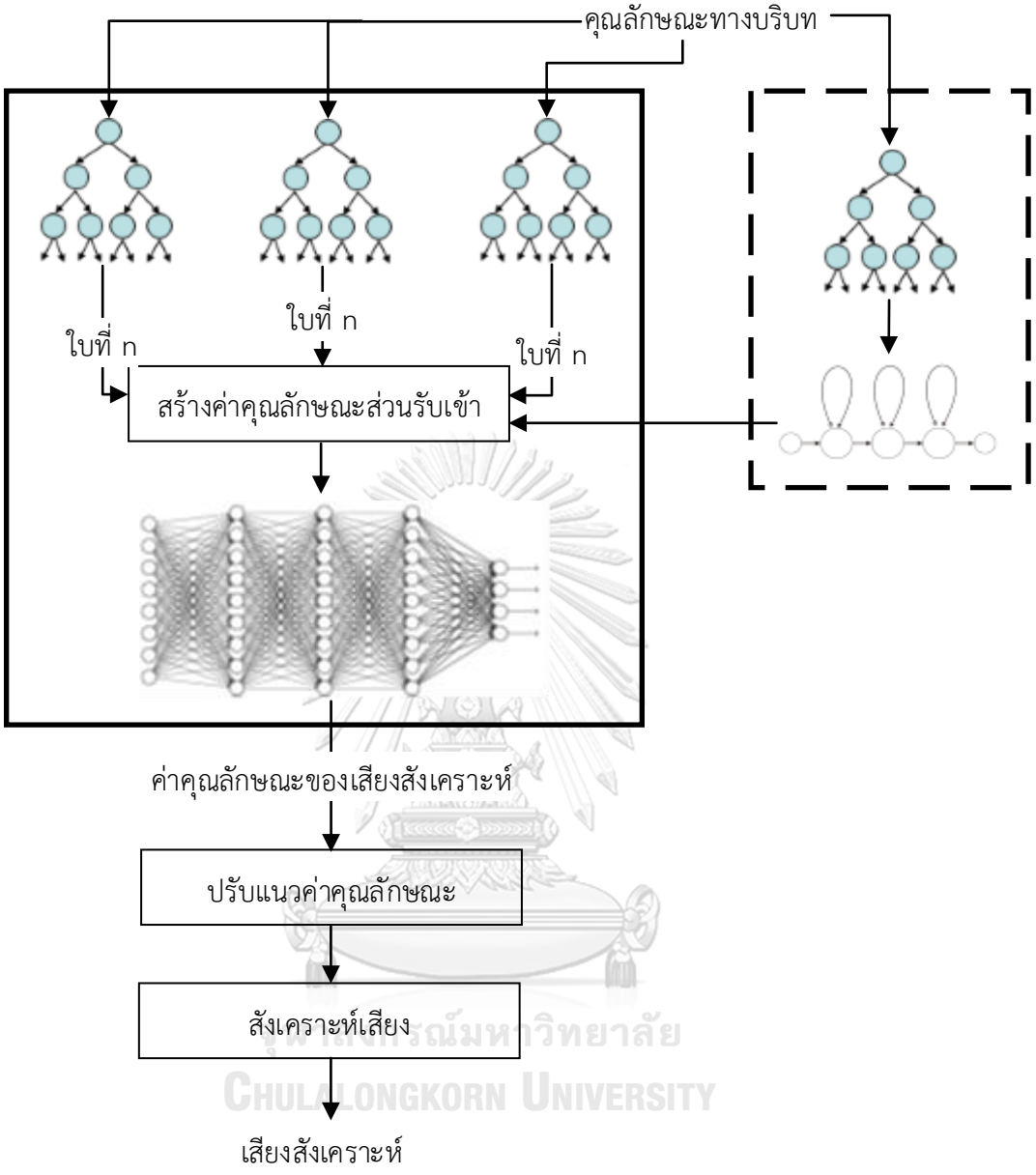
รูปที่ 33 ส่วนประกอบการสังเคราะห์เสียงด้วยแบบจำลองโครงข่ายประสาทเทียมแบบลึก

กระบวนการเชื่อมโยงค่าคุณลักษณะทางบริบท กับค่าคุณลักษณะทางเสียงของแบบจำลองฮิดเดนมาร์คอฟ กระทำผ่านทางต้นไม้ตัดสินใจ ที่จะทำการจัดกลุ่มของแบบจำลองเสียงที่มีค่าคุณลักษณะของเสียงที่คล้ายกันเข้ามาอยู่ในกลุ่มเดียวกัน และสร้างความสัมพันธ์ดังกล่าวออกมาในรูปแบบของต้นไม้ตัดสินใจ โดยรูปแบบของค่าคุณลักษณะของบริบทที่ไม่เคยพบมาก่อน จะถูกเชื่อมโยงกับแบบจำลองที่ถูกฝึกฝนผ่านทางต้นไม้ตัดสินใจ โดยการเลือกแบบจำลองของค่าคุณลักษณะของสัญญาณเสียง ที่อยู่ในกลุ่มของบริบทที่ใกล้เคียงกันมาใช้แทน

แต่ในทางกลับกันการสร้างแบบจำลองเสียงจากแบบจำลองโครงข่ายประสาทเทียมแบบลึกเป็นการสร้างสมการที่เชื่อมโยงความสัมพันธ์ระหว่างคุณลักษณะส่วนรับเข้า กับค่าคุณลักษณะทางเสียง แต่อย่างไรก็ตาม ในกรณีที่เป็นคุณลักษณะทางบริบทที่ไม่เคยพบมาก่อนปรากฏขึ้น จะทำให้ได้ผลลัพธ์ที่ไม่สามารถคาดเดาได้

เนื่องจากคุณลักษณะทางบริบทมีจำนวนมาก ส่งผลให้ในการสร้างเสียงสังเคราะห์มีโอกาสพบรูปแบบของบริบทที่ไม่เคยพบมาก่อนเป็นจำนวนมาก ซึ่งจะส่งผลกับการสร้างเสียงสังเคราะห์จากแบบจำลองโครงข่ายประสาทเทียมแบบลึก เพราะไม่สามารถคาดเดาผลลัพธ์ของค่าคุณลักษณะทางบริบทที่ไม่เคยปรากฏมาก่อนได้

ดังนั้น ในงานวิจัยนี้จึงได้นำแนวคิดของการใช้ต้นไม้ตัดสินใจ ควบคู่กับแบบจำลองโครงข่ายประสาทเทียมแบบลึก เพื่อใช้ข้อดีของต้นไม้ตัดสินใจในการลดจำนวนรูปแบบของคุณลักษณะที่ไม่เคยปรากฏมาก่อน และใช้ข้อดีของโครงข่ายประสาทเทียมแบบลึกที่สามารถเชื่อมโยงความสัมพันธ์ระหว่างคุณลักษณะส่วนรับเข้า กับค่าคุณลักษณะทางเสียง ได้ดีกว่าการใช้แบบจำลองฮิดเดนมาร์คอฟ เพราะการใช้แบบจำลองฮิดเดนมาร์คอฟสามารถสังเคราะห์ค่าคุณลักษณะได้โดยใช้ค่ากลางแทนทั้งสถานะ แต่แบบจำลองโครงข่ายประสาทเทียมแบบลึกสามารถสังเคราะห์ค่าคุณลักษณะที่แตกต่างกันตามแต่ละกรอบเวลาของตัวอย่างในสถานะ



รูปที่ 34 ส่วนประกอบการสังเคราะห์เสียงด้วยแบบจำลองแบบผสม

แบบจำลองที่นำเสนอเป็นไปตามรูปที่ 34 โดยคุณลักษณะทางบริบทที่ใช้เป็นข้อมูลขาเข้าของแบบจำลองโครงข่ายประสาทเทียมแบบลึกมาจากข้อมูล 2 ประเภท ดังนี้

1. ค่าคุณลักษณะบรรยายบริบทหน่วยเสียง

การหาค่าคุณลักษณะบรรยายบริบทหน่วยเสียงของวิธีที่นำเสนอ ทำโดยการกำหนดหมายเลขให้กับแต่ละใบไม้ในต้นไม้ตัดสินใจ และนำคุณลักษณะทางบริบทไปผ่านเข้าไปในต้นไม้ตัดสินใจ จะได้หมายเลขของใบไม้ในแต่ละสถานะของแต่ละค่าคุณลักษณะทางเสียง ซึ่งประกอบด้วย

ค่าคุณลักษณะสเปกตรัม ค่าคุณลักษณะความถี่มูลฐาน และค่าคุณลักษณะความไม่เป็นคาบของแถบความถี่

ค่าคุณลักษณะจะมีจำนวนเท่ากับจำนวนใบไม้ที่มากที่สุดของกลุ่มต้นไม้ตัดสินใจที่ใช้จำลอง ค่าคุณลักษณะสเปกตรัม ค่าคุณลักษณะความถี่มูลฐาน และค่าคุณลักษณะค่าความไม่เป็นคาบของแถบความถี่ ดังตัวอย่างของจำนวนใบของต้นไม้ตัดสินใจในตารางที่ ***4***6 ที่กำหนดให้แต่ละคุณลักษณะมี 3 สถานะ จะมีจำนวนค่าคุณลักษณะเท่ากับ 16 ซึ่งเป็นค่าคุณลักษณะที่มาจากต้นไม้ตัดสินใจของคุณลักษณะสเปกตรัมจำนวน 5 คุณลักษณะ ค่าความถี่มูลฐานจำนวน 7 คุณลักษณะ และความไม่เป็นคาบของแถบความถี่จำนวน 4 คุณลักษณะ

การออกแบบค่าคุณลักษณะส่วนรับเข้าในรูปแบบที่นำเสนอมีข้อดีที่ดีกว่าการใช้คำถามจากต้นไม้ตัดสินใจมาใช้เป็นค่าคุณลักษณะโดยตรง เพราะสามารถลดจำนวนรูปแบบของค่าคุณลักษณะส่วนรับเข้าที่ไม่เคยพบในการฝึกฝนได้ และการนำตำแหน่งของใบจากต้นไม้ของทั้ง 3 กระแสเข้ามาใช้เป็นค่าคุณลักษณะส่วนรับเข้าพร้อมกัน สามารถสร้างความสัมพันธ์ที่หลายรูปแบบกว่าการใช้แบบจำลองฮิดเดนมาร์คอฟ เช่น แบบจำลองฮิดเดนมาร์คอฟของกระแสสเปกตรัมในใบลำดับที่ 2 ของสถานะที่ 1 จะมีค่าเหมือนกันทั้งหมดโดยไม่สนใจลำดับของใบไม้จากต้นไม้ในกระแสอื่น ซึ่งต่างจากการใช้แบบจำลองโครงข่ายประสาทเทียมแบบลึกที่สามารถนำลำดับของใบไม้จากต้นไม้ในกระแสอื่นเข้ามาพิจารณา และสร้างเป็นค่าคุณลักษณะที่แตกต่างกันออกมา

ตารางที่ 6 ตัวอย่างโครงสร้างของต้นไม้ตัดสินใจ

ค่าคุณลักษณะ	สถานะ	จำนวนใบ
สเปกตรัม	1	3
	2	**5**
	3	4
ความถี่มูลฐาน	1	**7**
	2	6
	3	5
ความไม่เป็นคาบของแถบความถี่	1	2
	2	**4**
	3	3

2. ค่าคุณลักษณะระบุตำแหน่ง

ค่าคุณลักษณะที่ใช้ในการระบุตำแหน่ง ยังคงใช้เหมือนกับที่นำเสนอในงานวิจัยของ Zen และคณะ. (2013) [35] ดังที่อธิบายไว้ในหัวข้อ 3.5.2

4.2.2 การนอร์มัลไลเซชันค่าคุณลักษณะส่วนส่งออก

ค่าคุณลักษณะส่วนส่งออกมีค่าอยู่ในระดับที่ไม่เหมือนกัน และมีความแปรปรวนที่ไม่เท่ากัน ดังนั้นการนอร์มัลไลเซชันจึงเป็นกระบวนการที่จำเป็น เพื่อให้กระบวนการฝึกฝนไม่มีการโน้มเอียงไปยังค่าคุณลักษณะใดมากเกินไป

การนอร์มัลไลเซชันที่นิยมใช้ ได้แก่

1. การนอร์มัลไลเซชันให้ค่าต่ำสุด และสูงสุดมีค่าอยู่ระหว่าง 0 ถึง 1 (Wu และคณะ. (2016) [41]) ทำโดยการหาค่าสูงสุด และค่าต่ำสุดของแต่ละค่าคุณลักษณะ และทำการปรับระดับค่าสูงสุดมีค่าเป็น 1 และค่าต่ำสุดมีค่าเป็น 0 ซึ่งการนอร์มัลไลเซชันในรูปแบบนี้จะมีปัญหากับกรณีที่มีตัวอย่างที่ไม่ถูกต้อง และทำให้ช่วงของค่าคุณลักษณะที่ผิดเพี้ยนไป เช่น ในกรณีที่มีตัวอย่างเพียงไม่มากที่มีค่าความถี่มูลฐานสูงเกินกว่าที่ควรจะเป็น
2. การนอร์มัลไลเซชันด้วยคะแนนมาตรฐาน (Zen และคณะ. (2013) [35]) ทำโดยการหาค่ากลาง และค่าความแปรปรวนของแต่ละค่าคุณลักษณะจากข้อมูลทั้งหมด เพื่อใช้ในการคำนวณคะแนนมาตรฐาน ซึ่งการนอร์มัลไลเซชันนี้สามารถป้องกันปัญหากรณีที่มีตัวอย่างที่ไม่ถูกต้องได้ แต่เนื่องจากค่ามาตรฐานมีช่วงของลัพธ์ที่เป็นจำนวนจริงทั้งหมด จึงจำเป็นต้องใช้กับฟังก์ชันกระตุ้น (Activation function) ประเภทเส้นตรง

เนื่องจากค่าความแปรปรวนในแต่ละค่าคุณลักษณะมีค่าไม่เท่ากันดังนั้นการนอร์มัลไลเซชันด้วยการกำหนดให้ค่าสูงสุดมีค่าอยู่ระหว่าง 0 ถึง 1 จึงไม่เหมาะสมเท่าที่ควร

จากการนอร์มัลไลเซชันด้วยวิธีทั้ง 2 พบว่าเป็นการนอร์มัลไลเซชันแบบไม่คำนึงถึงค่าคุณลักษณะทางบริบท เช่น ในหน่วยเสียงที่แตกต่างกัน อาจจะมีค่ากลาง และค่าความแปรปรวนที่แตกต่างกันออกไป ดังนั้นการนอร์มัลไลเซชันด้วยรูปแบบที่เหมือนกันในทุกตัวอย่างในคลังข้อมูลเสียงจึงไม่เหมาะสม

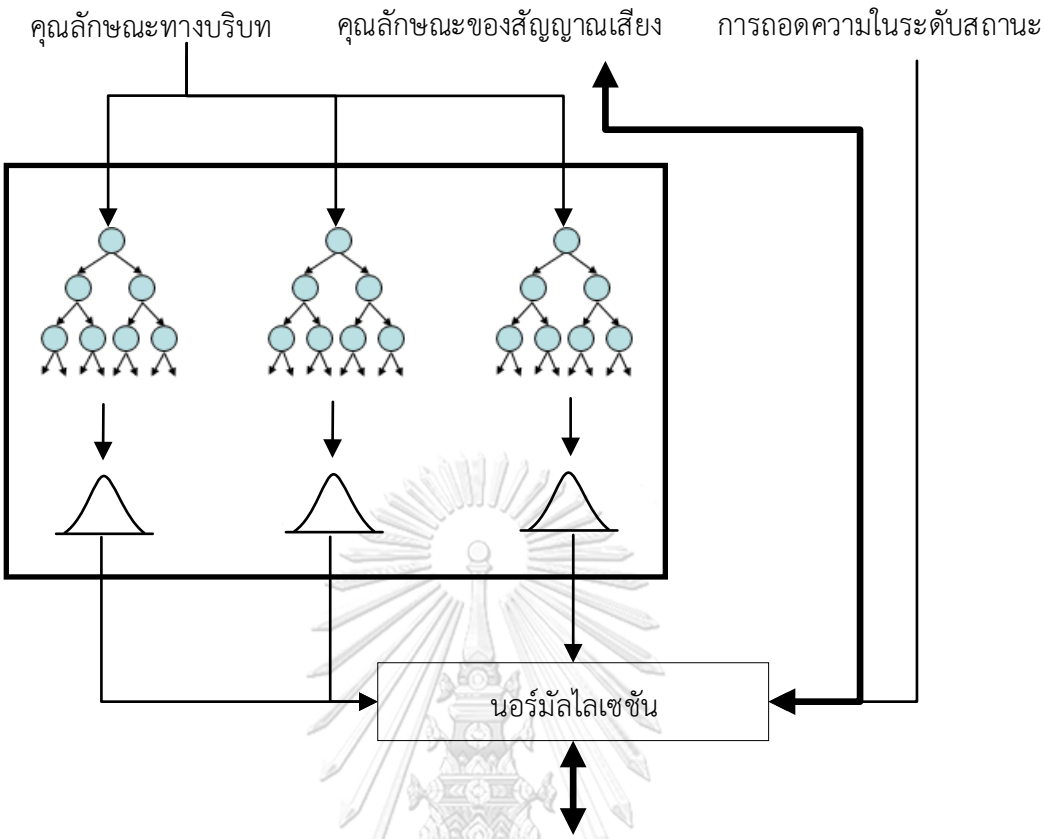
ดังนั้น ในงานวิจัยนี้จึงได้นำเสนอการนอร์มัลไลเซชันด้วยการใช้ค่าคะแนนมาตรฐานอ้างอิงจากแบบจำลองฮิดเดนมาร์คอฟ โดยจะนำค่าคุณลักษณะทางบริบทไปทำการค้นหาใบของต้นไม้ตัดสินใจที่อยู่ในกระแสต่างๆ ซึ่งในใบของต้นไม้ตัดสินใจของแต่ละกระแส จะมีแบบจำลองฮิดเดนมาร์คอฟอยู่ ซึ่งแบบจำลองฮิดเดนมาร์คอฟจะqจัดเก็บค่ากลาง และค่าความแปรปรวน โดยจะนำค่ากลาง และค่าความแปรปรวนดังกล่าวไปใช้ในการนอร์มัลไลเซชันด้วยการใช้ค่าคะแนนมาตรฐาน

ข้อดีของการนอร์มัลไลเซชันด้วยการใช้ค่าคะแนนมาตรฐานอ้างอิงจากแบบจำลองฮิดเดนมาร์คอฟ คือ จะทำให้ค่าคุณลักษณะส่วนส่งออกทั้งหมดมีค่าอยู่ในช่วงหนึ่ง เสมือนกับการนอร์มัลไลเซชันด้วยค่าสูงสุด และต่ำสุด ถ้าอ้างอิงตามหลักสถิติของการกระจายตัวแบบปกติซึ่งใช้ในการสร้างแบบจำลองฮิดเดนมาร์คอฟ จะพบว่ามีข้อมูลที่มีค่าปกติอยู่ในช่วง -3 ถึง 3 มากกว่า 99% และนอกจากนั้นการนอร์มัลไลเซชันด้วยการใช้ค่าคะแนนมาตรฐานอ้างอิงจากแบบจำลองฮิดเดนมาร์คอฟสามารถกำหนดช่วงของผลลัพธ์ของหน่วยเสียงที่มีค่าคุณลักษณะทางบริบทสอดคล้องกันได้ เพราะค่ากลาง และค่าความแปรปรวนที่ใช้มาจากแบบจำลองฮิดเดนมาร์คอฟที่อยู่ในใบของต้นไม้ตัดสินใจ

วิธีการดังกล่าวเปรียบเสมือนกับเป็นการนำความรู้ของโครงสร้างต้นไม้ตัดสินใจมาใช้รวมในการฝึกฝน โดยไม่ได้ใช้โดยตรงแบบวิธีการปรับค่าคุณลักษณะส่วนรับเข้า แต่เป็นการสร้างค่าคุณลักษณะส่วนส่งออกที่โน้มเอียงความสัมพันธ์ระหว่างค่าคุณลักษณะรับเข้า และส่งออก ให้เป็นไปตามโครงสร้างของต้นไม้ตัดสินใจ

ขั้นตอนในการนอร์มัลไลเซชันด้วยการใช้ค่าคะแนนมาตรฐานอ้างอิงจากแบบจำลองฮิดเดนมาร์คอฟมีขั้นตอนแสดงดังรูปที่ 35 โดยจำเป็นต้องใช้ค่าคุณลักษณะทางบริบท และข้อมูลการถอดความในระดับสถานะที่สอดคล้องกับกรอบเวลาของตัวอย่างเสียงที่ต้องการนอร์มัลไลเซชันค่าคุณลักษณะ โดยค่าคุณลักษณะทางบริบทจะใช้ในการค้นหาใบไม้ของต้นไม้ตัดสินใจ โดยในใบไม้ของต้นไม้ตัดสินใจจะเก็บค่ากลาง และค่าความแปรปรวน ข้อมูลการถอดความในระดับสถานะ จะใช้ในการระบุสถานะของกรอบเวลาของตัวอย่างเสียง เพื่อเลือกใช้ต้นไม้ที่สอดคล้องกับสถานะดังกล่าว

การลดระดับการนอร์มัลไลเซชัน (Denormalization) เพื่อแปลงค่าคุณลักษณะที่สังเคราะห์มาจากโครงข่ายประสาทเทียมแบบลึก ให้กลับมาเป็นค่าคุณลักษณะที่ใช้ในการสังเคราะห์เสียง ให้ทำตามกระบวนการนอร์มัลไลเซชันในทิศทางย้อนกลับ โดยค่าคุณลักษณะทางบริบทมาจากข้อความที่ต้องการสังเคราะห์เสียง และการถอดความในระดับสถานะ คือ ผลลัพธ์จากแบบจำลองช่วงเวลาที่สอดคล้องกับค่าคุณลักษณะทางบริบทของข้อความที่ต้องการสังเคราะห์เสียง



รูปที่ 35 กระบวนการนอร์มัลไลเซชันค่าคุณลักษณะส่วนส่งออก

บทที่ 5 การทดลอง และวิธีการวิเคราะห์ผลการทดลอง

ในบทนี้อธิบายถึงฐานข้อมูลเสียงที่ใช้ในการทดลอง การตั้งค่าของแบบจำลองเสียงสังเคราะห์ชนิดต่างๆ วิธีการดำเนินการทดลองที่แบ่งเป็นการวัดผลเป็นแบบการทดสอบแบบปรนัย (Objective test) และการทดสอบแบบอัตนัย (Subjective test) และชนิดของเครื่องมือทางสถิติ ที่ใช้ในการวัดผลการทดลอง

5.1 ฐานข้อมูลเสียงภาษาไทย สำหรับสร้างระบบสังเคราะห์เสียง

ฐานข้อมูลเสียงภาษาไทยที่ใช้การทดสอบคือฐานข้อมูลเสียง TSYNC [1] ที่ประกอบด้วยข้อมูลเสียงจำนวน 5,200 ประโยค ที่ถูกคัดเลือกจากบทความต่างๆ มีความยาวทั้งหมด 13.94 ชั่วโมง โดยใช้นักข่าวเพศหญิงเป็นผู้อ่านบทความดังกล่าว โดยฐานข้อมูล TSYNC ออกแบบให้มีความสมดุลกันของจำนวนหน่วยเสียงในระดับไทรแกรม (Tri-gram)

ค่าคุณลักษณะทางบริบทที่ฐานข้อมูลเสียง TSYNC ได้กำกับมาให้ด้วย ได้แก่ หน่วยเสียงวรรณยุกต์ และชนิดของคำ แต่อย่างไรก็ตามค่าชนิดของคำไม่สามารถนำมาใช้ในการสร้างระบบสังเคราะห์เสียงได้ เพราะชนิดของคำที่กำกับในฐานข้อมูล TSYNC ถูกแบ่งหมวดหมู่เป็นกลุ่มที่กำหนดขึ้นเอง ตามที่เสนอในงานวิจัยของ Sornlertlamvanich และคณะ (2009) [43] และไม่ได้เป็นกลุ่มของชนิดของคำที่เป็นมาตรฐานในภาษาไทย

ในการทดลองนี้ได้แบ่งฐานข้อมูลเสียงออกเป็น 2 กลุ่ม ได้แก่ กลุ่มที่ใช้ในการทดสอบแบบปรนัยจำนวน 500 ประโยค และที่เหลือเป็นกลุ่มที่ใช้ในการฝึกฝน

กระบวนการคัดเลือกประโยคที่ใช้ในการทดสอบแบบปรนัย 500 ประโยค มีขั้นตอนตามตารางที่ 7 ซึ่งจะเน้นเลือกประโยคโดยใช้เกณฑ์ที่คำนึงถึงค่าคุณลักษณะทางบริบทที่แตกต่างกันมากที่สุด

จากประโยคที่ใช้ในการทดสอบแบบปรนัยจำนวน 500 ประโยค ได้ทำการสุ่มเลือกประโยคอีกจำนวน 25 ประโยค เพื่อใช้ในการทดสอบการฟังของการทดสอบแบบอัตนัย เรียกว่า กลุ่มทดสอบการฟัง

รายละเอียดของข้อมูลเสียงทั้ง 3 กลุ่ม สรุปอยู่ในตารางที่ 8

ตารางที่ 7 กระบวนการคัดเลือกกลุ่มที่ใช้ในการทดสอบ

ลำดับ	รายละเอียด
1	กำหนดกลุ่มของคำตอบ เป็นรายการที่เก็บประโยคที่ต้องการ
2	นับจำนวนหน่วยเสียงที่มีค่าคุณลักษณะทางบริบทที่ไม่เหมือนกันในแต่ละประโยค และค่าคุณลักษณะทางบริบทดังกล่าวต้องไม่ปรากฏในประโยคที่ถูกจัดเก็บอยู่ในกลุ่มของคำตอบ
3	เลือกประโยคที่มีค่าผลลัพธ์ในลำดับที่ 2 มากที่สุด ไปใส่ในกลุ่มของคำตอบ เพียง 1 ประโยคเท่านั้น
4	ทำการตรวจสอบกลุ่มของคำตอบว่ามีจำนวนเท่ากับจำนวนที่ต้องการหรือไม่ ถ้ามีจำนวนเท่ากันให้สิ้นสุดกระบวนการ ถ้าไม่เท่ากันให้เริ่มต้นทำซ้ำตั้งแต่กระบวนการที่ 2

ตารางที่ 8 รายละเอียดของข้อมูลเสียงทั้ง 3 กลุ่ม

คุณสมบัติ	กลุ่มฝึกฝน	กลุ่มทดสอบแบบปรนัย	กลุ่มทดสอบฟัง
จำนวนประโยค	4,700	500	25
ความยาว	13.94 ชั่วโมง	1.05 ชั่วโมง	3.13 นาที
จำนวนหน่วยเสียง	422,261	24,002	1,239
จำนวนค่าคุณลักษณะทางบริบทที่แตกต่างกัน	233,041	22,192	1,157
% ของค่าคุณลักษณะทางบริบทที่ทับซ้อนกับกลุ่มฝึกฝน	---	11.34%	10.63%

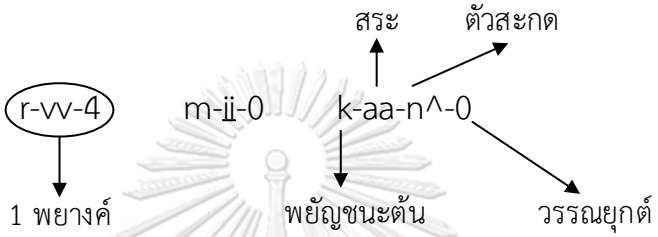
ค่าคุณลักษณะทางบริบทที่สร้างขึ้นจากข้อมูลในฐานข้อมูลเสียงนี้ ได้แก่

- หน่วยเสียงจำนวน 7 หน่วยเสียงที่อยู่ติดกัน ในรูปแบบของ 3 หน่วยเสียงในลำดับก่อนหน้า หน่วยเสียงปัจจุบัน และ 3 หน่วยเสียงที่ในลำดับถัดไป
- วรรณยุกต์ในระดับของหน่วยเสียงจำนวน 7 หน่วยเสียง ในรูปแบบของ 3 วรรณยุกต์ที่อยู่ในลำดับก่อนหน้า วรรณยุกต์ปัจจุบัน และ 3 วรรณยุกต์ที่อยู่ในลำดับถัดไป

โดยทั่วไปแล้ววรรณยุกต์จะเป็นคุณสมบัติของพยางค์ แต่สำหรับค่าคุณลักษณะทางบริบทของวรรณยุกต์ที่ใช้ในการทดลองนี้ได้กำหนดให้วรรณยุกต์เป็นคุณสมบัติในระดับหน่วยเสียง โดยให้ค่า

ของวรรณยุกต์ที่อยู่ในหน่วยเสียงของพยางค์เดียวกันมีค่าเหมือนกันทั้งหมด และเหมือนกับค่าวรรณยุกต์ในระดับพยางค์

โครงสร้างของพยางค์ในภาษาไทยมี 2 รูปแบบคือ พยัญชนะต้น-สระ-ตัวสะกด และพยัญชนะต้น-สระ ตามที่แสดงในรูปที่ 36 โดยค่าคุณลักษณะทางบริบทของหน่วยเสียง ii ที่ขีดเส้นใต้ในรูปที่ 36 เป็นดังตารางที่ 9 โดยให้ค่าตำแหน่งที่ติดลบ หมายถึง ตำแหน่งในลำดับก่อนหน้า และตำแหน่งที่เป็นบวก หมายถึง ตำแหน่งในลำดับถัดไป

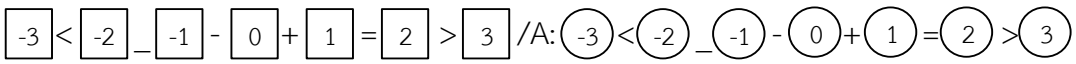


รูปที่ 36 โครงสร้างของพยางค์ในภาษาไทย

ตารางที่ 9 ตัวอย่างการแสดงค่าคุณลักษณะทางบริบทในตำแหน่งต่างๆ

ตำแหน่ง	หน่วยเสียง	วรรณยุกต์
-3	r	4
-2	vv	4
-1	m	0
0	ii	0
1	k	0
2	aa	0
3	n^	0

โดยจะนำค่าของหน่วยเสียง และวรรณยุกต์ในแต่ละตำแหน่งที่แสดงในตารางที่ 9 เข้าไปใส่ตามรูปแบบในรูปที่ 37 ที่มีการใช้สัญลักษณะเพื่อขั้นระหว่างแต่ละตำแหน่ง และใช้กลุ่มตัวอักษร /A: เพื่อแบ่งระหว่างส่วนที่เป็นหน่วยเสียง และส่วนที่เป็นวรรณยุกต์ ซึ่งจะทำให้ได้รูปแบบที่ใช้แทนค่าคุณลักษณะทางบริบทเป็น r<vv_mii+k=aa>n^/A:4<4_0-0+0=0>0



n หน่วยเสียงตำแหน่งที่ n

n วรรณยุกต์ตำแหน่งที่ n

รูปที่ 37 รูปแบบที่ใช้แทนค่าคุณลักษณะ

คำถามสำหรับต้นไม้ตัดสินใจสร้างขึ้นมาจากกฎทางไวยกรณ์ของภาษาไทย และนำกฎนั้นไปใช้กับทุกตำแหน่งของหน่วยเสียง และวรรณยุกต์ โดยมีคำถามทั้งสิ้น 742 ข้อ รายการคำถามทั้งหมดแสดงในภาคผนวก ก

5.2 การสกัดค่าคุณลักษณะ

สัญญาณเสียงที่อยู่ในฐานข้อมูลเสียงจะถูกแปลงให้มีความถี่ 16,000 เฮิรตซ์ และทำการแบ่งเป็นกรอบเวลา โดยแต่ละกรอบเวลามีระยะเวลา 25 มิลลิวินาที และแต่ละกรอบเวลามีระยะเวลาห่างกันอยู่ 5 มิลลิวินาที ซึ่งจะทำให้ช่วงเวลาเพียง 1 วินาที มีตัวอย่างเสียงถึง 200 กรอบเวลา

ตัวเข้ารหัสเสียงที่ใช้ในการทดลองคือ ตัวเข้ารหัสเสียง STRAIGHT [16] ค่าคุณลักษณะที่ใช้ในงานวิจัยนี้ ได้แก่

- คุณลักษณะค่าสัมประสิทธิ์เมลเคปสตรัม จำนวน 35 คุณลักษณะ และค่าคุณลักษณะของความเร็ว และความเร่ง อีก 70 คุณลักษณะ รวมทั้งสิ้น 105 คุณลักษณะ
- ค่าความถี่มูลฐาน จำนวน 1 คุณลักษณะ และค่าคุณลักษณะของความเร็ว และความเร่ง อีก 2 คุณลักษณะ รวมทั้งสิ้น 3 คุณลักษณะ
- คุณลักษณะค่าสัมประสิทธิ์ค่าความไม่เป็นคาบของแถบความถี่ จำนวน 25 คุณลักษณะ และค่าคุณลักษณะของความเร็ว และความเร่ง อีก 50 คุณลักษณะ รวมทั้งสิ้น 75 คุณลักษณะ

5.3 แบบจำลองฮิดเดนมาร์คอฟ

แบบจำลองฮิดเดนมาร์คอฟที่ใช้ในการทดลองนี้ ประกอบด้วย 7 แบบจำลอง (รูปแบบ) ได้แก่

1. แบบจำลอง HMM_BASE ประกอบด้วยแบบจำลอง 1 แบบจำลอง ที่รวมกระแสของค่าคุณลักษณะสเปกตรัม และค่าคุณลักษณะความถี่มูลฐานเข้ามาอยู่ในแบบจำลองเดียวกัน ซึ่งเหมือนกับแบบจำลองที่นำเสนอโดย Zen และคณะ (2007) [2]
2. รูปแบบ HMM_NOREJ ประกอบด้วย 2 แบบจำลอง ที่แยกกระแสค่าคุณลักษณะสเปกตรัม และค่าคุณลักษณะความถี่มูลฐานออกเป็นคนละแบบจำลอง ตามที่นำเสนอ

ในงานวิจัยนี้ แต่ในขั้นตอนการหาระยะเวลาของสถานะไม่ได้มีการพิจารณาสถานะความก้องของเสียง และรูปแบบนี้ถูกแบ่งออกเป็น 3 แบบจำลองย่อย โดยแยกตามวิธีการหาระยะเวลาของหน่วยเสียง ดังต่อไปนี้

2.1. แบบจำลอง HMM_NOREJ_SPEC ใช้แบบจำลองช่วงเวลาจากแบบจำลองสเปกตรัมเป็นหลัก

2.2. แบบจำลอง HMM_NOREJ_F0 ใช้แบบจำลองช่วงเวลาจากแบบจำลองค่าความถี่มูลฐานเป็นหลัก

2.3. แบบจำลอง HMM_NOREJ_AVG ใช้แบบจำลองช่วงเวลาจากทั้งสองแบบจำลองและทำการคำนวณหาช่วงเวลาตามกระบวนการในตารางที่ 2

3. รูปแบบ HMM_REJ มีโครงสร้างของแบบจำลองเสียงเหมือนกับรูปแบบ HMM_NOREJ แต่ในขั้นตอนการหาระยะเวลาของสถานะมีการพิจารณาสถานะความก้องของเสียง โดยกระบวนการหาระยะเวลาของสถานะเป็นไปตามตารางที่ 3 โดยช่วงเวลาเป้าหมายของแต่ละหน่วยเสียง แบ่งการคำนวณได้เป็น 3 แบบจำลอง ดังนี้

3.1. แบบจำลอง HMM_REJ_SPEC ใช้ช่วงเวลาเป้าหมายจากการใช้แบบจำลองช่วงเวลาจากแบบจำลองสเปกตรัมเป็นหลัก

3.2. แบบจำลอง HMM_REJ_F0 ใช้ช่วงเวลาเป้าหมายจากการใช้แบบจำลองช่วงเวลาจากแบบจำลองค่าความถี่มูลฐานเป็นหลัก

3.3. แบบจำลอง HMM_REJ_AVG ใช้ช่วงเวลาเป้าหมายจากการใช้แบบจำลองช่วงเวลาจากทั้งสองแบบจำลอง ตามกระบวนการในตารางที่ 2

สำหรับกระบวนการในการฝึกฝนแบบจำลองทั้ง 3 แบบจำลอง เป็นไปตามค่าตั้งต้นตามที่นำเสนอโดย Zen และคณะ (2007) [2] ซึ่งกำหนดให้มีทั้งหมด 5 สถานะต่อหนึ่งแบบจำลองฮิดเดนมาร์คอฟ และค่าน้ำหนักของกระแสค่าคุณลักษณะสเปกตรัม และค่าคุณลักษณะความถี่มูลฐานมีค่าเป็น 1 เท่ากันในรูปแบบ HMM_BASE และค่าน้ำหนักของกระแสค่าคุณลักษณะความไม่เป็นคาบของแถบความถี่มีค่าเป็น 0 ในทุกรูปแบบ

5.4 แบบจำลองโครงข่ายประสาทเทียมแบบลึก

ตัวแปรในการทดลองของแบบจำลองโครงข่ายประสาทเทียมแบบลึก แบ่งออกเป็น 3 ประเด็น ดังต่อไปนี้

1. ค่าคุณลักษณะส่วนรับเข้า

ค่าคุณลักษณะส่วนรับเข้าของระบบประกอบด้วย 2 รูปแบบ ได้แก่ รูปแบบที่นำเสนอโดยงานวิจัย Zen และคณะ (2013) [35] และรูปแบบที่นำเสนอในงานวิจัยนี้

รูปแบบที่นำเสนอโดย Zen และคณะ (2013) เรียกว่าค่าคุณลักษณะส่วนรับเข้าจากคำถามต้นไม้ตัดสินใจ มีคุณลักษณะทั้งสิ้น 745 คุณลักษณะ ซึ่งประกอบด้วยค่าคุณลักษณะทางบริบทจำนวน 742 คุณลักษณะที่มาจากข้อคำถามที่ใช้ในการสร้างต้นไม้ตัดสินใจ แสดงในภาคผนวก ก และค่าคุณลักษณะระบุตำแหน่ง 3 คุณลักษณะ ตามที่นำเสนอในหัวข้อ 4.2.1

รูปแบบที่นำเสนอในงานวิจัยนี้ เรียกว่าค่าคุณลักษณะส่วนรับเข้าจากตำแหน่งใบไม้ของต้นไม้ตัดสินใจ ใช้ข้อมูลจากแบบจำลอง HMM_BASE ซึ่งมีจำนวนของใบไม้ในต้นไม้ตัดสินใจในแต่ละสถานะแสดงดังตารางที่ 10 ซึ่งจะทำให้มีค่าคุณลักษณะทางบริบทจำนวน 5,678 ค่าคุณลักษณะ ซึ่งมาจากกระแสของค่าคุณลักษณะสเปกตรัม 1,315 คุณลักษณะ กระแสของค่าคุณลักษณะค่าความถี่มูลฐาน 3,870 คุณลักษณะ และกระแสของค่าคุณลักษณะความไม่เป็นคาบของแถบความถี่ 493 คุณลักษณะ และค่าคุณลักษณะระบุตำแหน่ง 3 คุณลักษณะ รวมทั้งสิ้น 5,681 คุณลักษณะ

ตารางที่ 10 จำนวนใบในต้นไม้จากแบบจำลอง HMM_BASE

สถานะ	สเปกตรัม	ค่าความถี่มูลฐาน	ความไม่เป็นคาบของแถบความถี่
1	811	2,748	380
2	1,004	3,389	464
3	**1,315**	**3,870**	**493**
4	1,015	2,682	417
5	953	2,658	383

ถึงแม้ว่าจำนวนค่าคุณลักษณะด้วยรูปแบบที่นำเสนอในงานวิจัยนี้มีมากถึง 5,681 คุณลักษณะ มากกว่ารูปแบบที่นำเสนอในงานวิจัยของ Zen และคณะ (2013) [35] ที่มีเพียง 745 คุณลักษณะ แต่เมื่อพิจารณาจำนวนรูปแบบของค่าคุณลักษณะส่วนรับเข้าทางบริบทที่ซ้ำกันชุดข้อมูลฝึกฝน (ไม่รวมค่าคุณลักษณะระบุตำแหน่ง 3 คุณลักษณะ) พบว่า ค่าคุณลักษณะส่วนรับเข้าที่นำเสนอ

มีชุดข้อมูลที่ซ้ำกับข้อมูลฝึกฝนถึง 91.11% เปรียบเทียบกับค่าคุณลักษณะส่วนรับเข้าที่ นำเสนอในงานวิจัยของ Zen และคณะ (2013) [35] ที่มีข้อมูลซ้ำกันเพียง 17.02% ซึ่งแสดงให้เห็นว่ารูปแบบค่าคุณลักษณะส่วนรับเข้าที่นำเสนอ สามารถลดการเกิดรูปแบบที่ไม่เคยพบมาก่อนได้

สำหรับจำนวนค่าคุณลักษณะทางบริบทที่แตกต่างกันของรูปแบบที่นำเสนอมีค่ามากกว่าจำนวนค่าคุณลักษณะทางบริบทที่นำเสนอในตารางที่ ***6***11 เพราะตำแหน่งของใบไม้ในสถานะที่แตกต่างกันในหน่วยเสียงเดียวกันมีค่าไม่เหมือนกัน จึงทำให้ค่าคุณลักษณะส่วนรับเข้าทางบริบทมีความแตกต่างกันไปแม้ว่าจะมีหน่วยเสียงเดียวกัน ซึ่งแตกต่างจากการใช้ค่าคุณลักษณะส่วนรับเข้าจากคำถามของต้นไม้ตัดสินใจที่จะได้ค่าคุณลักษณะเหมือนกันทุกสถานะในหน่วยเสียง

ตารางที่ 11 จำนวนค่าคุณลักษณะส่วนรับเข้าของแบบจำลองโครงข่ายประสาทเทียมที่แตกต่างกัน

รูปแบบ	คุณสมบัติ	กลุ่มฝึกฝน	กลุ่มทดสอบแบบปรนัย	กลุ่มทดสอบฟัง
Zen และคณะ (2013) [35]	จำนวนค่าคุณลักษณะทางบริบทที่แตกต่างกัน	209,185	21,819	1,156
	% ของค่าคุณลักษณะทางบริบทที่ทับซ้อนกับกลุ่มฝึกฝน	---	17.02%	16.52%
ที่นำเสนอ	จำนวนค่าคุณลักษณะทางบริบทที่แตกต่างกัน	69,107	35,004	4,756
	% ของค่าคุณลักษณะทางบริบทที่ทับซ้อนกับกลุ่มฝึกฝน	---	91.11%	95.98%

2. การนอร์มัลไลเซชันค่าคุณลักษณะส่วนส่งออก

การนอร์มัลไลเซชันค่าคุณลักษณะมี 2 แนวคิด โดยเลือกใช้การนอร์มัลไลเซชันด้วยการถ่วงค่าน้ำหนักด้วยค่าความแปรปรวนตามที่เสนอในชุดเครื่องมือฝึกฝนแบบจำลองเสียง HTS รุ่น 2.3.1 ที่นำเสนอโดย Black และคณะ (2007) [2]และแนวคิดการนอร์มัลไลเซชันด้วยการใช้ค่าคะแนนมาตรฐานอ้างอิงจากแบบจำลองฮิดเดนมาร์คอฟ ที่นำเสนอในหัวข้อ 4.2.2

3. ค่าความแปรปรวนในกระบวนการปรับแนวคุณลักษณะ

ถึงแม้ว่าในวิจัยของ Watts และคณะ (2016) [34] ได้นำเสนอว่าการใช้ค่าความแปรปรวนที่คำนวณจากข้อมูลทั้งหมดของแต่ละค่าคุณลักษณะ เรียกว่า ค่าความแบบครอบคลุม ให้ผลไม่แตกต่างจากการใช้ค่าความแปรปรวนจากแบบจำลองฮิดเดนมาร์คอฟที่อยู่ในใบไม้ของต้นไม้ตัดสินใจ เรียกว่า

ค่าความแปรปรวนตามบริบท แต่อย่างไรก็ตามในงานวิจัยนี้ได้นำไปใช้กับฐานข้อมูลเสียงที่มีค่าคุณลักษณะทางบริบทไม่มาก เมื่อเปรียบเทียบกับที่ใช้ในงานวิจัยของ Watts และคณะ (2016) [34] ดังนั้นจึงนำประเด็นนี้มาทดสอบในการทดลองนี้อีกครั้ง

จากปัจจัยทั้งหมดสามารถสรุปแบบจำลอง และกำหนดชื่อของแต่ละแบบจำลองดังแสดงในตารางที่ 12 ซึ่งประกอบด้วยแบบจำลองทั้งหมด 8 แบบจำลอง

ตารางที่ 12 ตัวแปรของแบบจำลองโครงข่ายประสาทเทียมแบบลึก

ชื่อเรียก	ค่าคุณลักษณะส่วนรับเข้า	การนอร์มัลไลเซชันค่าคุณลักษณะส่วนส่งออก	ค่าความแปรปรวนในกระบวนการปรับแนวคุณลักษณะ
DNN_G_G	คำถามของต้นไม้ตัดสินใจ	ถ่วงค่าน้ำหนักด้วยค่าความแปรปรวน	ความแปรปรวนแบบครอบคลุม
DNN_G_L	คำถามของต้นไม้ตัดสินใจ	ถ่วงค่าน้ำหนักด้วยค่าความแปรปรวน	ความแปรปรวนตามบริบท
DNN_Z_G	คำถามของต้นไม้ตัดสินใจ	คะแนนมาตรฐานอ้างอิงจากแบบจำลองฮิดเดนมาร์คอฟ	ความแปรปรวนแบบครอบคลุม
DNN_Z_L	คำถามของต้นไม้ตัดสินใจ	คะแนนมาตรฐานอ้างอิงจากแบบจำลองฮิดเดนมาร์คอฟ	ความแปรปรวนตามบริบท
DNNHMM_G_G	ตำแหน่งใบไม้ของต้นไม้ตัดสินใจ	ถ่วงค่าน้ำหนักด้วยค่าความแปรปรวน	ความแปรปรวนแบบครอบคลุม
DNNHMM_G_L	ตำแหน่งใบไม้ของต้นไม้ตัดสินใจ	ถ่วงค่าน้ำหนักด้วยค่าความแปรปรวน	ความแปรปรวนตามบริบท
DNNHMM_Z_G	ตำแหน่งใบไม้ของต้นไม้ตัดสินใจ	คะแนนมาตรฐานอ้างอิงจากแบบจำลองฮิดเดนมาร์คอฟ	ความแปรปรวนแบบครอบคลุม
DNNHMM_Z_L	ตำแหน่งใบไม้ของต้นไม้ตัดสินใจ	คะแนนมาตรฐานอ้างอิงจากแบบจำลองฮิดเดนมาร์คอฟ	ความแปรปรวนตามบริบท

สำหรับค่าคุณลักษณะความถี่มูลฐานจะทำการนำเสนอโดยใช้ 2 ค่าคุณลักษณะ โดยคุณลักษณะแรก คือ ค่าความถี่มูลฐานที่เป็นค่าต่อเนื่องทั้งหมด โดยทำการประมาณค่าในช่วงที่ไม่

สามารถหาค่าความถี่มูลฐานได้ (ช่วงที่เป็นเสียงไม่ก้อง) และค่าคุณลักษณะที่ 2 คือ ค่าที่กำหนดว่าตัวอย่างนั้นเป็นเสียงก้อง หรือเสียงไม่ก้อง เรียกว่าค่าคุณลักษณะสถานะความก้องของเสียง

ค่าคุณลักษณะส่งออกประกอบด้วยทั้งหมด 184 ค่าคุณลักษณะ ประกอบด้วยค่าคุณลักษณะดังต่อไปนี้

- คุณลักษณะค่าสัมประสิทธิ์เมลเคปสตรัม จำนวน 35 คุณลักษณะ และค่าคุณลักษณะของความเร็ว และความเร่ง อีก 70 คุณลักษณะ รวมทั้งสิ้น 105 คุณลักษณะ

- คุณลักษณะค่าสัมประสิทธิ์ค่าความไม่เป็นคาบของแถบความถี่ จำนวน 25 คุณลักษณะ และค่าคุณลักษณะของความเร็ว และความเร่ง อีก 50 คุณลักษณะ รวมทั้งสิ้น 75 คุณลักษณะ

- ค่าความถี่มูลฐานที่เป็นค่าต่อเนื่องทั้งหมด จำนวน 1 คุณลักษณะ และค่าคุณลักษณะของความเร็ว และความเร่ง อีก 2 คุณลักษณะ รวมทั้งสิ้น 3 คุณลักษณะ

- ค่าคุณลักษณะสถานะความก้องของเสียง จำนวน 1 คุณลักษณะ

ทุกตัวแปรของแบบจำลองโครงข่ายประสาทเทียมแบบลึกมีโครงสร้างของแบบจำลองเหมือนกัน คือ มีจำนวน 3 ชั้นซ้อนเร้น ที่มีจำนวนโหนดชั้นละ 512 โหนด โดยใช้ฟังก์ชันกระตุ้น (Activation Function) ประเภท Sigmoid ในชั้นรับเข้า และชั้นซ้อนเร้นทั้งหมด และใช้ฟังก์ชันกระตุ้นประเภทเส้นตรงสำหรับชั้นส่งออก

ฟังก์ชันต้นทุนที่เลือกใช้คือค่าคลาดเคลื่อนกำลังสองเฉลี่ย และตัวเพิ่มประสิทธิภาพที่เลือกใช้ คือ ADAM นำเสนอโดย Kingma และ Ba (2014) [44] และกำหนดให้มีค่าอัตราการเรียนรู้ (Learning rate) ที่ 0.001

เพื่อป้องกันการเรียนรู้ข้อมูลที่โน้มเอียงเข้ากับชุดฝึกฝนมากเกินไป (Over fitting) ในระหว่างการเรียนรู้ของแบบจำลองโครงข่ายประสาทเทียมจึงมีการใช้ Dropout ตามที่นำเสนอโดย Srivastava และคณะ (2014) [45] โดยกำหนดให้มีค่าความน่าจะเป็นที่จะเก็บโหนด (Keep probability) ที่ 0.5

ในการเรียนรู้จะใช้วิธีการเรียนรู้แบบกลุ่ม (Batch) โดยกำหนดให้ขนาดของกลุ่มมีค่าเป็น 256 และทำการเรียนรู้ทั้งสิ้น 10 รอบข้อมูล

5.4 การทดสอบแบบปรนัย

การทดสอบแบบปรนัยประกอบด้วย 2 การทดลองย่อย ดังต่อไปนี้

5.4.1 การวัดผลช่วงเวลาของแบบจำลองเสียง

การวัดผลช่วงเวลาของเสียงคำนวณจากการหาความผิดพลาดกำลังสองเฉลี่ยของความยาวของหน่วยเสียง (DUR_RMSE) ตามสมการที่ 6 โดยที่ dn_i แทนถึงช่วงเวลาของหน่วยเสียงจากคลังข้อมูลเสียงในหน่วยวินาที และ ds_i แทนถึงช่วงเวลาของหน่วยเสียงที่ได้จากการสังเคราะห์เสียงในหน่วยวินาที โดยจะทำการคำนวณเฉพาะหน่วยเสียงที่ไม่ใช่เสียงเงียบเท่านั้น

$$DUR\_RMSE = \sqrt{\frac{\sum_{i=1}^{n}(dn_i - ds_i)^2}{n}} \quad (6)$$

การวัดผลช่วงเวลาจะทำการทดสอบกับแบบจำลองฮิดเดนมาร์คอฟ เพื่อทำการประเมินว่าการคำนวณหาช่วงเวลาตามที่นำเสนอในหัวข้อที่ 4.1.2 ว่ารูปแบบใดได้ผลลัพธ์ที่ดีที่สุด

เนื่องจากการวิเคราะห์ด้วยการใช้ DUR_RMSE สามารถแสดงได้เพียงความคลาดเคลื่อนโดยรวมเท่านั้น ดังนั้น ในงานวิจัยนี้จึงได้ทำการนำเสนอผลต่างของแต่ละหน่วยเสียงตามสมการที่ 7 จากนั้นนำค่าผลลัพธ์ไปทำการวาดภาพแท่งความถี่ โดยจะทำการวัดผลเฉพาะหน่วยเสียงที่ไม่ใช่เสียงเงียบเท่านั้น

$$DUR\_DIFF(dn_i) = \frac{ds_i}{dn_i} \quad (7)$$

ค่า DUR_DIFF แสดงถึง % ความแตกต่างระหว่างความยาวช่วงเวลาของหน่วยเสียงที่สังเคราะห์ออกมา เปรียบเทียบกับเสียงต้นฉบับจากฐานข้อมูลเสียง โดยจะมีค่าระหว่าง 0 ถึงค่าบวกอนันต์ โดยค่าในช่วงระหว่าง 0 – 1 แทนถึงเสียงที่สังเคราะห์ออกมา มีช่วงเวลาที่สั้นกว่าเสียงจากฐานข้อมูลเสียง และค่าที่มากกว่า 1 แทนถึงเสียงที่สังเคราะห์ออกมามีช่วงเวลาที่ยาวกว่าเสียงจากฐานข้อมูลเสียง

5.4.2 การวัดผลคุณภาพของเสียงสังเคราะห์

เนื่องจากค่าคุณลักษณะของเสียงที่สังเคราะห์ออกมามีความยาวไม่เท่ากัน ดังนั้นในงานวิจัยนี้จึงต้องทำการปรับแนวทางเวลาด้วยวิธีการไดนามิคไทมวารปพิง (2007) [46] โดยจะทำการปรับพร้อมกันทุกค่าคุณลักษณะ และใช้ฟังก์ชันต้นทุนที่ใช้ในการค้นหาเส้นทางจากค่าความเพี้ยนของเซปตรัลในระดับเมลของค่าสัมประสิทธิ์เมลเคปสตรัม ตามที่เสนอโดย Toda และคณะ (2007) [47] ซึ่งจะใช้เพียงค่าคุณลักษณะของสเปกตรัมเพียงอย่างเดียวเท่านั้นในการปรับแนวทางเวลา โดยหลังจากการปรับแนวทางเวลาแล้ว จะทำให้ความยาวของค่าคุณลักษณะที่สังเคราะห์ออกมาเท่ากับความยาวของค่าคุณลักษณะของเสียงที่อยู่ในฐานข้อมูลเสียง

การวัดผลคุณภาพของแบบจำลองเสียงใช้เครื่องมือวัดดังต่อไปนี้

- ค่าความเพี้ยนของเซปตรัลในระดับเมลของค่าสัมประสิทธิ์เมลเคปสตรัม (MGC_MCD)
  วิธีการคำนวณความเพี้ยนของเซปตรัลในระดับเมลคำนวณจากงานวิจัยของ Toda และคณะ (2007) [47] โดยจะนำค่าสัมประสิทธิ์ของค่าคุณลักษณะสเปกตรัมที่ได้จากการสังเคราะห์เสียง และปรับแนวทางเวลาแล้ว ไปทำการเปรียบเทียบกับค่าคุณลักษณะจากเสียงในฐานข้อมูลเสียง โดยจะเลือกเฉพาะส่วนที่ไม่ใช่เสียงเงียบเท่านั้น
- ค่าความเพี้ยนของเซปตรัลในระดับเมลของค่าความไม่เป็นคาบของแถบความถี่ (BAP_MCD)
  วิธีการคำนวณความเพี้ยนของเซปตรัลในระดับเมลคำนวณจากงานวิจัยของ Toda, และคณะ (2007) [47] โดยจะนำค่าสัมประสิทธิ์ของค่าคุณลักษณะค่าความไม่เป็นคาบของแถบความถี่ที่ได้จากการสังเคราะห์เสียง และปรับแนวทางเวลาแล้ว ไปทำการเปรียบเทียบกับค่าคุณลักษณะจากเสียงในฐานข้อมูลเสียง โดยจะเลือกเฉพาะส่วนที่ไม่ใช่เสียงเงียบเท่านั้น
- ความไม่สอดคล้องกันของสถานะความก้องของเสียง (LF0_UVU)
  คำนวณจากการเปรียบเทียบค่าสถานะความก้องของเสียงของเสียงที่สังเคราะห์ และปรับแนวทางเวลาแล้ว กับค่าคุณลักษณะจากเสียงในฐานข้อมูลเสียงที่อยู่ในลำดับเดียวกัน โดยนับจำนวนตัวอย่างที่มีสถานะความเป็นเสียงไม่สอดคล้องกัน แล้วทำการหารด้วยจำนวนตัวอย่าง โดยจะพิจารณาเฉพาะหน่วยเสียงที่ไม่ใช่เสียงเงียบ
- ความผิดพลาดกำลังสองเฉลี่ยของค่าความถี่มูลฐาน (LF0_RMSE)
  คำนวณจากการเปรียบเทียบค่าคุณลักษณะความถี่มูลฐานของเสียงที่สังเคราะห์ และปรับแนวทางเวลาแล้ว กับค่าคุณลักษณะจากเสียงในฐานข้อมูลเสียง โดยเลือกเฉพาะ

ตัวอย่างที่เป็นเสียงก้องเหมือนกันเท่านั้น และไม่ใช่หน่วยเสียงเงียบ เพื่อไปคำนวณหาความผิดพลาดกำลังสองเฉลี่ยในหน่วยลอการิทึมของความถี่

สำหรับการวิเคราะห์ผลของตัวชี้วัดทั้ง 4 ตัว จากทั้ง 15 รูปแบบเสียง ด้วยกระบวนการทางสถิติ จะใช้การทดลอง F-test ด้วยสถิติทดสอบความแปรปรวนทางเดียว (One-way ANOVA) เพื่อเปรียบเทียบความแตกต่างของผลการทดลองจากตัวชี้วัดทั้ง 4 ตัว ของทั้ง 15 รูปแบบเสียง ที่ระดับความเชื่อมั่น 95% และหากมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติ จะทำการทดสอบเทียบพหุคูณ โดยเปรียบเทียบรายคู่ (Post Hoc comparison) ใช้วิธี Least Significant Difference (LSD)

5.5 การทดสอบแบบอัตนัย

การทดสอบแบบอัตนัยจะใช้ผู้ฟัง จำนวน 9 คน ประกอบด้วย ผู้ทดสอบที่เป็นผู้ชาย 2 คน และผู้หญิง 7 คน มีอายุอยู่ในช่วงอายุ 19 – 22 ปี โดยผู้ทดสอบทั้งหมดต้องฟังเสียงสังเคราะห์ทั้งหมด 25 ประโยค ต่อ 1 รูปแบบ โดยในการทดสอบจะใช้ทั้งหมด 15 รูปแบบ รวมแล้วผู้ฟังจะต้องฟังเสียงทั้งหมด 375 ไฟล์ โดยจะต้องทำการทดสอบให้แล้วเสร็จในครั้งเดียว ไม่สามารถแบ่งช่วงการทดสอบได้ โดยไฟล์เสียงที่เป็นเสียงต้นฉบับจากฐานข้อมูลเสียง (เรียกว่าเสียงอ้างอิง) ถูกบรรจุอยู่ในการทดสอบ เพื่อให้ผู้ฟังใช้เป็นเกณฑ์อ้างอิง

สำหรับประเด็นในเรื่องจำนวนผู้ฟังที่ใช้ตัวอย่างเพียง 9 คนนั้น ซึ่งเป็นข้อจำกัดมาจากจำนวนรูปแบบของเสียงที่ต้องฟังเป็นจำนวนมาก แต่เมื่อทำการเปรียบเทียบกับงานวิจัยอื่น พบว่ามีงานวิจัยที่ใช้ผู้ทดสอบไม่เกิน 15 คน ได้แก่ งานวิจัยของ Wu และ King (2016) [48] ที่ใช้ผู้ทดสอบเพียง 12 คน งานวิจัยของ Zen และคณะ (2004) [24] ที่ใช้ผู้ทดสอบเพียง 8 คน และงานวิจัยของ Tamura และคณะ (1998) [49] ที่ใช้ผู้ทดสอบเพียง 7 คน

เนื่องจากรูปแบบทั้งหมดมีมากถึง 15 รูปแบบ ในการทดลองจึงแนะนำให้ผู้ฟัง ฟังเสียงทั้ง 15 รูปแบบก่อน แล้วจึงให้ฟังใหม่อีกรอบ พร้อมกับทำการให้คะแนน โดยผู้ฟังสามารถฟังได้ไม่จำกัดจำนวนครั้ง และไม่จำกัดเวลาที่ใช้ในการทดสอบ

ในการทดสอบได้ประเมินคุณภาพของเสียงสังเคราะห์ใน 2 ประเด็น ได้แก่

- ความชัดเจนของเสียงสังเคราะห์ เรียกว่า INT

  ในส่วนนี้ประเมินความสามารถในการรับรู้ของผู้ฟังว่าสามารถรับรู้ข้อความจากเสียงสังเคราะห์มากน้อยแค่ไหน โดยในประเด็นนี้รวมถึงการออกเสียงผิดพลาด ทั้งในส่วนของหน่วยเสียง และวรรณยุกต์

- ความเป็นธรรมชาติของเสียงสังเคราะห์ เรียกว่า NAT

ในส่วนนี้ประเมินในด้านของจังหวะในการออกเสียง และระดับความดังของเสียง ว่ามีการเน้นคำถูกต้อง การเว้นวรรค เหมือนกับเสียงอ้างอิงมากน้อยแค่ไหน

ระดับของคะแนนที่ใช้ในการประเมินทั้ง 2 ด้าน มีระดับคะแนนตั้งแต่ 1 ถึง 10 โดยความหมายของระดับ 10 คือระดับที่เสียงที่สังเคราะห์ออกมามีคุณภาพมากที่สุด และ 1 คือมีคุณภาพน้อยที่สุด ตามรายละเอียดในตารางที่ 13

ตารางที่ 13 ความหมายของแต่ละระดับของประเด็นที่ใช้ในการทดสอบอัตนัย

ประเด็น	ระดับ	ความหมาย
Int	1	ผู้ฟังไม่สามารถรับรู้ข้อความในเสียงสังเคราะห์ได้ทั้งหมด
	10	ผู้ฟังสามารถรับรู้ข้อความในเสียงสังเคราะห์ได้ทั้งหมด
Nat	1	ผู้ฟังรับรู้ว่าเสียงที่สังเคราะห์มีจังหวะ และทำนองการเปล่งเสียงไม่เหมือนเสียงอ้างอิงทั้งหมด
	10	ผู้ฟังรับรู้ว่าเสียงที่สังเคราะห์มีจังหวะ และทำนองการเปล่งเสียงเหมือนเสียงอ้างอิงทั้งหมด

ผู้ฟังจะได้รับความหมายของระดับที่ดีที่สุด และระดับน้อยที่สุดเท่านั้น ส่วนระดับอื่นไม่ได้มีการนิยามไว้ แต่มีการแจ้งว่าให้ระดับคะแนนที่เท่ากันถ้าผู้ฟังรับรู้ว่าเสียงในรูปแบบที่แตกต่างกัน มีคุณภาพในด้านที่ประเมินเหมือนกัน และให้ระดับที่แตกต่างกันในรูปแบบเสียงที่มีคุณภาพแตกต่างกัน

บทที่ 6 ผลการทดลอง และวิเคราะห์ผลการทดลอง

ผลการทดลองนำเสนอแยกตามประเภทของการทดลองได้ ดังต่อไปนี้

6.1 การทดสอบแบบปรนัย

การทดสอบแบบปรนัยแบ่งออกเป็น 2 การทดลองย่อย ตามที่นำเสนอในบทที่ 5 ซึ่งประกอบด้วย

6.1.1 การวัดผลช่วงเวลาของแบบจำลองเสียง

ผลการทดสอบในส่วนของค่า DUR_RMSE แสดงในตารางที่ 14 และเนื่องจากเป็นการวัดผลในระดับหน่วยเสียงจึงทำให้รูปแบบที่มีการพิจารณาสถานะความก้องของเสียง และไม่พิจารณาสถานะความก้องของเสียงมีผลการทดลองที่เท่ากัน

จากผลการทดลองแสดงให้เห็นว่า แบบจำลอง HMM_BASE มีค่าความคลาดเคลื่อนน้อยที่สุด และช่วงเวลาที่สังเคราะห์จากแบบจำลองช่วงเวลาของแบบจำลองสเปกตรัมเพียงอย่างเดียว (รูปแบบที่ลงท้ายด้วย SPEC) และแบบจำลองความถี่มูลฐานเพียงอย่างเดียว (รูปแบบที่ลงท้ายด้วย LF0) ให้ผลลัพธ์ที่ใกล้เคียงกัน และแย่กว่าแบบจำลอง HMM_BASE อยู่มาก แต่เมื่อนำทั้งสองแบบจำลองมาใช้ร่วมกัน (รูปแบบที่ลงท้ายด้วย AVG) กลับทำให้ช่วงเวลาที่สังเคราะห์ออกมามีคะแนนดีขึ้น แต่ยังไม่สามารถดีกว่ารูปแบบ HMM_BASE

ตารางที่ 14 ผลการทดลองช่วงเวลาของแบบจำลองเสียง

แบบจำลอง	RMSE_DUR (วินาที)
HMM_BASE	0.03109
HMM_REJ_AVG, HMM_NOREJ_AVG	0.03231
HMM_REJ_LF0, HMM_NOREJ_LF0	0.03546
HMM_REJ_SPEC, HMM_NOREJ_SPEC	0.03545

เมื่อทำการพิจารณาภาพแห่งความถี่ของค่า DUR_DIFF (สมการที่ 7) แสดงตามรูปที่ ***34***38 โดยค่า DUR_DIFF แสดงในแกน X ในรูปแบบของ % โดยในช่องขวาสุดแทนถึงจำนวนตัวอย่างทั้งหมดที่มีความยาวของเสียงที่สังเคราะห์ออกมายาวมากกว่า 3 เท่าของเสียงในฐานข้อมูลเสียง และแกน Y แทนถึงจำนวนตัวอย่างในหน่วยของหน่วยเสียง โดยเส้น BASE แทนถึงแบบจำลอง HMM_BASE, เส้น AVG แทนถึงแบบจำลอง HMM_REJ_AVG และ HMM_NOREJ_AVG, เส้น LF0

แทนถึงแบบจำลอง HMM_REJ_LF0 และ HMM_NO_REJ_LF0 และเส้น SPEC แทนถึงแบบจำลอง HMM_REJ_SPEC และ HMM_NOREJ_SPEC

ทุกรูปแบบของการสังเคราะห์เสียงมีช่วงที่มีจำนวนตัวอย่างอยู่สูงสุดคือช่วง 90% - 95% เหมือนกันหมด แต่สำหรับรูปแบบ LF0 มีจำนวนตัวอย่างที่อยู่ในช่วงดังกล่าวน้อยกว่าแบบจำลองอื่นอย่างชัดเจน

ในช่วงที่มีค่า 0-70% พบว่ารูปแบบ LF0 มีจำนวนหน่วยเสียงที่ตกอยู่ในช่วงเวลาดังกล่าวมากที่สุด ซึ่งสะท้อนให้เห็นว่าเสียงที่สังเคราะห์จากรูปแบบ LF0 มีแนวโน้มที่มีหน่วยเสียงสั้นกว่ารูปแบบอื่น และจึงทำให้รูปแบบ AVG ที่นำรูปแบบจาก LF0 และ SPEC มาถ่วงน้ำหนักเข้าด้วยกัน มีจำนวนหน่วยเสียงในช่วงดังกล่าวน้อยกว่ารูปแบบ LF0 แต่มากกว่ารูปแบบ SPEC

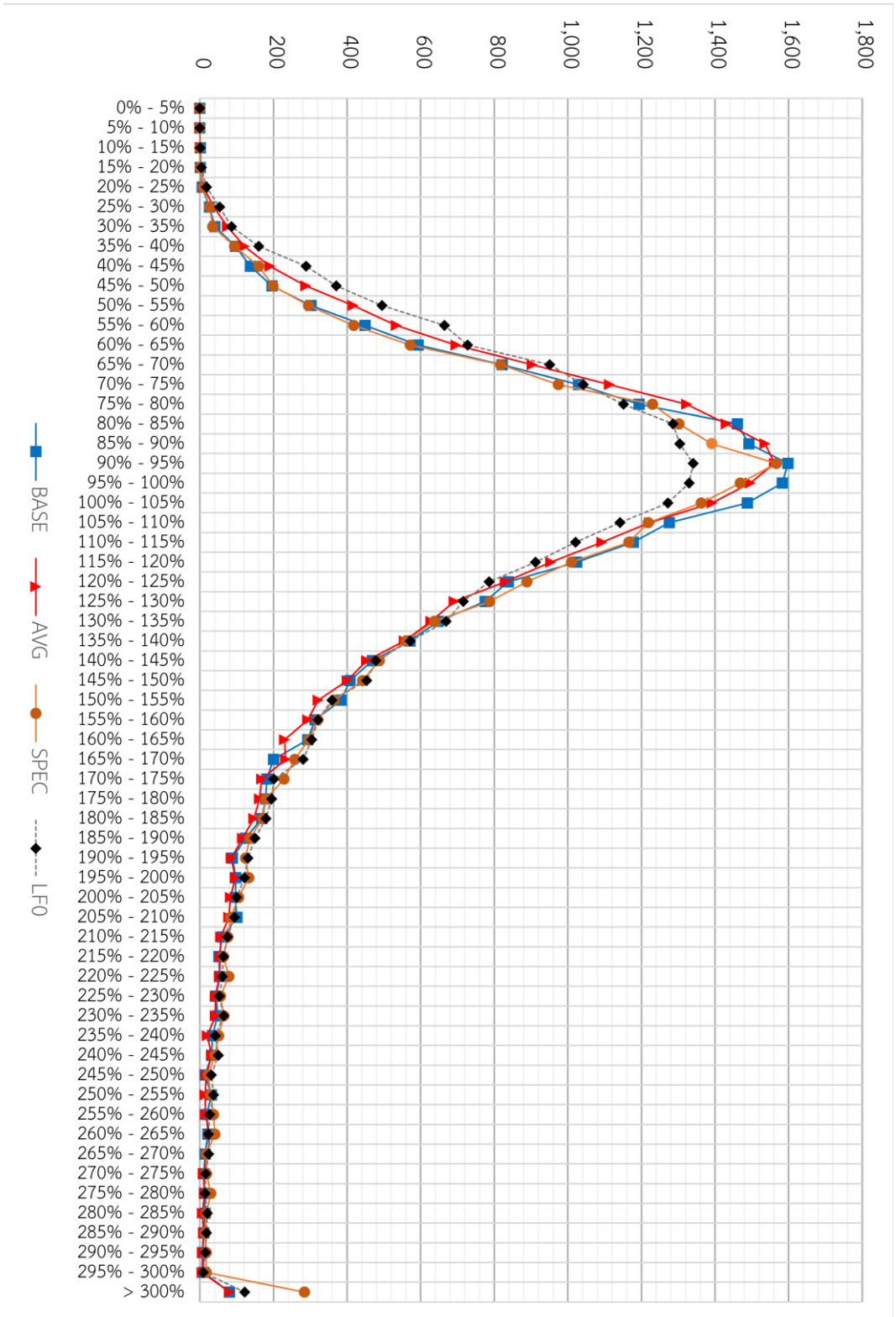
เมื่อเปรียบเทียบรูปแบบ BASE และรูปแบบ SPEC พบว่ามีการกระจายตัวที่คล้ายคลึงกันเกือบทุกช่วงเวลายกเว้นช่วง 75% - 110% ที่รูปแบบ BASE มีจำนวนตัวอย่างในบริเวณนั้นมากกว่า และช่วงที่มากกว่า 300% ที่รูปแบบ SPEC มีจำนวนตัวอย่างมากกว่ารูปแบบ BASE รวมถึงมากกว่ารูปแบบ LF0 และ AVG เป็นจำนวนมาก

เมื่อพิจารณาเวลารวมของทุกตัวอย่างตามตารางที่ 15 พบว่าเวลารวมของรูปแบบ LF0 น้อยกว่ารูปแบบ SPEC และ BASE และรูปแบบ SPEC เป็นรูปแบบที่ยาวที่สุด และมากกว่ารูปแบบ BASE ซึ่งสอดคล้องกับผลลัพธ์ของรูปที่ 38

ถึงแม้ว่ารูปแบบ AVG จะมีจำนวนตัวอย่างในช่วงที่มีค่า 0-70% (บริเวณที่เสียงสังเคราะห์จะสั้นลง) น้อยกว่ารูปแบบ LF0 แต่ในช่วงที่ 125% เป็นต้นไป (บริเวณที่เสียงสังเคราะห์จะยาวขึ้น) รูปแบบ AVG กลับมีตัวอย่างน้อยที่สุด และการนำเสนอในรูปที่ 38 อยู่ในรูปแบบของ % ซึ่งมีความเป็นไปได้ว่า ส่วนที่หน่วยเสียงสั้นลงมีค่าความยาวมากกว่า ส่วนที่ยาวขึ้น จึงทำให้ความยาวทั้งหมดของรูปแบบ AVG มีระยะเวลาที่สั้นที่สุด

ตารางที่ 15 ผลรวมความยาวของตัวอย่างเสียงในแต่ละรูปแบบ

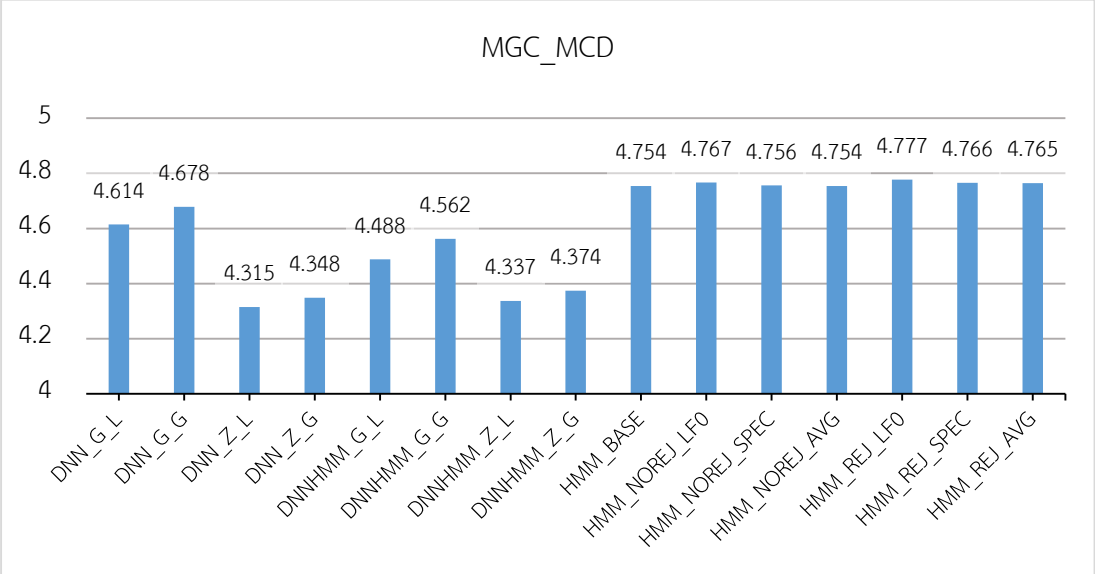
รูปแบบ	แบบจำลอง	ระยะเวลา (วินาที)
BASE	HMM_BASE	1,878
AVG	HMM_REJ_AVG, HMM_NOREJ_AVG	1,810
LF0	HMM_REJ_LF0, HMM_NOREJ_LF0	1,834
SPEC	HMM_REJ_SPEC, HMM_NOREJ_SPEC	1,943



รูปที่ 38 ผลการทดลองช่วงเวลาของแบบจำลองเสียง

6.1.2 การวัดผลคุณภาพของเสียงสังเคราะห์

ผลการวัดผลคุณภาพของแบบจำลองเสียงด้วยตัวชี้วัด MGC_MCD แสดงดังรูปที่ 39 ที่แสดงในรูปแบบของกราฟแท่งที่แกน X แสดงถึงรูปแบบของเสียงสังเคราะห์ และแกน Y แทนถึงค่าผลลัพธ์ที่ได้จากการคำนวณ โดยถ้ามีค่ามากหมายถึงค่าคุณลักษณะที่สังเคราะห์ออกมาไม่สอดคล้องกับค่าคุณลักษณะในฐานข้อมูลเสียง และค่าคุณลักษณะ MGC จะสะท้อนถึงการจัดเรียงกันของอวัยวะในช่องปาก และผลการวิเคราะห์ความแตกต่างกันของแต่ละรูปแบบแสดงในตารางที่ 16



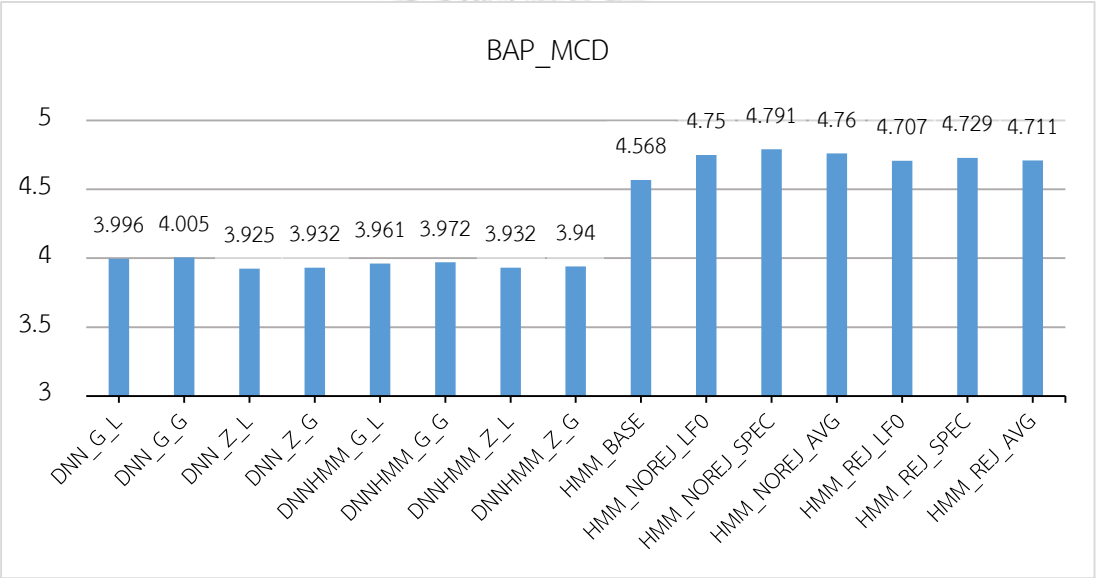
รูปที่ 39 ผลการทดลอง MGC_MCD

ค่าเฉลี่ยของแบบจำลองเสียงสังเคราะห์ฮิดเดนมาร์คอฟมีค่าสูงกว่า (แย่กว่า) แบบจำลองโครงข่ายประสาทเทียมแบบลึกทั้งหมดอย่างมีนัยสำคัญทางสถิติดังแสดงในตารางที่ 16 แสดงให้เห็นว่าผลลัพธ์ที่ได้จากแบบจำลองโครงข่ายประสาทเทียมแบบลึกดีกว่าการใช้แบบจำลองฮิดเดนมาร์คอฟ สำหรับทุกรูปแบบของเสียงสังเคราะห์ที่มาจากแบบจำลองฮิดเดนมาร์คอฟ มีค่าที่อยู่ที่ประมาณ 4.7 ถึง 4.8 ทั้งหมด และความแตกต่างของทุกรูปแบบที่มาจากแบบจำลองฮิดเดนมาร์คอฟไม่มีนัยสำคัญทางสถิติดังแสดงในตารางที่ 16

สำหรับค่าของแบบจำลองโครงข่ายประสาทเทียมแบบลึก แบ่งออกเป็น 2 ช่วง ได้แก่ ช่วงตั้งแต่ 4.4 ถึง 4.7 ที่เป็นช่วงของแบบจำลองโครงข่ายประสาทเทียมแบบลึกที่ใช้การนอร์มัลไลเซชันค่าคุณลักษณะส่งออกด้วยการหารด้วยค่าความแปรปรวนแบบครอบคลุม และช่วงตั้งแต่ 4.3 ถึง 4.4 ที่ใช้การนอร์มัลไลเซชันค่าคุณลักษณะส่วนส่งออกด้วยการใช้คะแนนมาตรฐานอ้างอิงจากแบบจำลอง

ฮิดเดนมาร์คอฟ ซึ่งแสดงให้เห็นว่าวิธีการนอร์มัลไลเซชันด้วยวิธีการที่นำเสนอช่วยเพิ่มคะแนนในตัวชี้วัด MGC_MCD ได้เป็นอย่างมาก

ผลการวัดผลคุณภาพของแบบจำลองเสียงด้วย BAP_MCD แสดงดังรูปที่ 40 ที่แสดงในรูปแบบของกราฟแท่ง ที่แกน X แสดงถึงรูปแบบของเสียงสังเคราะห์ และแกน Y แทนถึงค่าผลลัพธ์ที่ได้จากการคำนวณ โดยถ้ามีค่ามากหมายถึงค่าคุณลักษณะที่สังเคราะห์ออกมาไม่สอดคล้องกับค่าคุณลักษณะในฐานข้อมูลเสียง และค่าคุณลักษณะ BAP ใช้ในการควบคุมความไม่เป็นคาบในช่วงความถี่ต่างๆ ในกรณีที่กรอบเวลานั้นเป็นเสียงก้อง โดยถ้าค่าคุณลักษณะ BAP สังเคราะห์ออกมาไม่ถูกต้อง อาจจะทำให้ได้สัญญาณเสียงสังเคราะห์ที่มีสัญญาณเสียงรบกวนปนออกมาด้วย และผลการวิเคราะห์ความแตกต่างกันของแต่ละรูปแบบแสดงในตารางที่ 17 โดยช่องตารางที่มีพื้นหลังเป็นสีเทาแสดงถึงคู่ของรูปแบบที่มีความแตกต่างอย่างไม่มีนัยสำคัญทางสถิติ



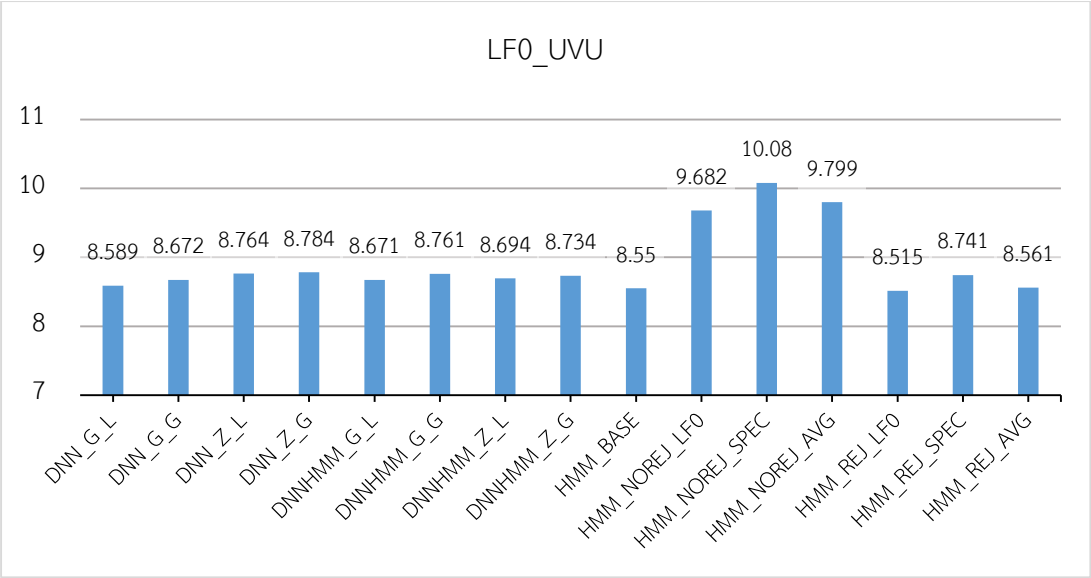
รูปที่ 40 ผลการทดลอง BAP_MCD

โดยสำหรับแบบจำลองเสียงสังเคราะห์จากแบบจำลองฮิดเดนมาร์คอฟที่ใช้แบบจำลองที่รวมกันของหลายกระแส (HMM_BASE) ได้ผลลัพธ์ที่ดีกว่า (คะแนนน้อยกว่า) แบบจำลองที่มีการใช้หลายแบบจำลองของแบบจำลองฮิดเดนมาร์คอฟทุกรูปแบบอย่างมีนัยสำคัญทางสถิติ และในรูปแบบมีการพิจารณาความสอดคล้องของสถานะความก้องของเสียงในช่วงการคำนวณหาระยะเวลาของสถานะ พบว่ามีผลลัพธ์ที่ดีขึ้นอย่างมีนัยสำคัญทางสถิติ แต่อย่างไรก็ตามก็ยังไม่มีผลลัพธ์ที่ดีเท่าการใช้แบบจำลองรวม

การใช้แบบจำลองโครงข่ายประสาทเทียมแบบลึกทุกรูปแบบ ได้ผลลัพธ์ที่ดีกว่าทุกรูปแบบของแบบจำลองฮิดเดนมาร์คอฟอย่างมีนัยสำคัญทางสถิติ โดยการใช้แบบจำลองโครงข่ายประสาทเทียมแบบลึกมีคะแนนอยู่ในระดับ 3.90 ถึง 4.01 ทั้งหมด เทียบกับการใช้แบบจำลองฮิดเดนมาร์คอฟที่มีคะแนนอยู่ในระดับ 4.5 ถึง 4.8

สำหรับความแตกต่างกันในแต่ละรูปแบบโครงข่ายประสาทเทียมแบบลึกพบว่ามีรูปแบบที่ไม่มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ได้แก่ รูปแบบที่มีการใช้การนอร์มัลไลเซชันค่าคุณลักษณะส่วนส่งออกด้วยการใช้คะแนนมาตรฐานอ้างอิงจากแบบจำลองฮิดเดนมาร์คอฟจำนวน 4 รูปแบบ ซึ่งประกอบด้วย DNN_Z_L, DNN_Z_G, DNNHMM_Z_L และ DNNHMM_Z_G ซึ่งแสดงให้เห็นว่าการนอร์มัลไลเซชันค่าคุณลักษณะส่วนส่งออกด้วยวิธีการที่นำเสนอช่วยทำให้คุณภาพของเสียงสังเคราะห์ที่ดีขึ้น เพราะ 4 รูปแบบดังกล่าวมีระดับคะแนนที่ดีกว่ารูปแบบที่เหลือ และรูปแบบที่มีการใช้ความแปรปรวนแบบครอบคลุมกับใช้ความแปรปรวนตามบริบทในการปรับแนวค่าคุณลักษณะ ที่ไม่มีความแตกต่างกัน ซึ่งแสดงให้เห็นว่าตัวแปรเรื่องของความแปรปรวนในการปรับแนวค่าคุณลักษณะไม่ส่งผลต่อตัวชี้วัด BAP_MCD

ผลการทดลอง LF0_UVU แสดงในรูปที่ 41 ผลการทดลสอบ LF0_UVUที่แสดงในรูปแบบของกราฟแท่ง ที่แกน X แสดงถึงรูปแบบของเสียงสังเคราะห์ และแกน Y แทนถึงค่าผลลัพธ์ที่ได้จากการคำนวณ ในรูปหน่วย % ของจำนวนตัวอย่างเสียงทั้งหมด โดยถ้ามีค่ามากจะหมายถึงมีความไม่สอดคล้องกันของสถานะความก้องของเสียงระหว่างเสียงสังเคราะห์ เปรียบเทียบกับเสียงจากฐานข้อมูลเสียงเป็นจำนวนมาก ซึ่งแทนถึงค่าที่ไม่ดี และผลการวิเคราะห์ความแตกต่างกันของแต่ละรูปแบบแสดงในตารางที่ 18 โดยช่องตารางที่มีพื้นหลังเป็นสีเทา แสดงถึงคู่ของรูปแบบที่มีความแตกต่างแบบไม่มีนัยสำคัญทางสถิติ



รูปที่ 41 ผลการทดลสอบ LF0_UVU

สำหรับแบบจำลองเสียงสังเคราะห์ที่ใช้แบบจำลองฮิดเดนมาร์คอฟ พบว่าการใช้แบบจำลองที่รวมกันของหลายกระแส (HMM_BASE) ได้ผลลัพธ์ในการทดลองนี้ดีกว่าการใช้แบบหลายแบบจำลองที่ไม่มีการพิจารณาสถานะความก้องของเสียงในช่วงของการคำนวณระยะเวลาของสถานะอย่างมีค่านัยสำคัญทางสถิติ แต่เมื่อเปรียบเทียบกับรูปแบบที่มีการพิจารณาสถานะความก้องของเสียงในช่วงของการคำนวณระยะเวลาของสถานะของเสียงสังเคราะห์ พบว่าความแตกต่างนั้นไม่มีนัยสำคัญทางสถิติ ซึ่งแสดงให้เห็นว่าการพิจารณาสถานะความก้องของเสียงในช่วงของการคำนวณระยะเวลาของสถานะของเสียงสังเคราะห์ ช่วยทำให้ผลลัพธ์ในการทดลองนี้ดีขึ้นอย่างมีนัยสำคัญทางสถิติดังแสดงในตารางที่ 18

สำหรับแบบจำลองโครงข่ายประสาทเทียมแบบลึกทุกรูปแบบมีผลลัพธ์ที่แย่กว่าการใช้แบบจำลองฮิดเดนมาร์คอฟที่ใช้แบบจำลองที่รวมกันของหลายกระแส (HMM_BASE) ในทุกรูปแบบ แต่ความแตกต่างในส่วนนี้ไม่มีนัยสำคัญทางสถิติดังแสดงในตารางที่ 18

เมื่อพิจารณาความแตกต่างของแต่ละรูปแบบที่เป็นแบบจำลองโครงข่ายประสาทเทียมแบบลึก พบว่ามีความแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติในทุกรูปแบบดังแสดงในตารางที่ 18

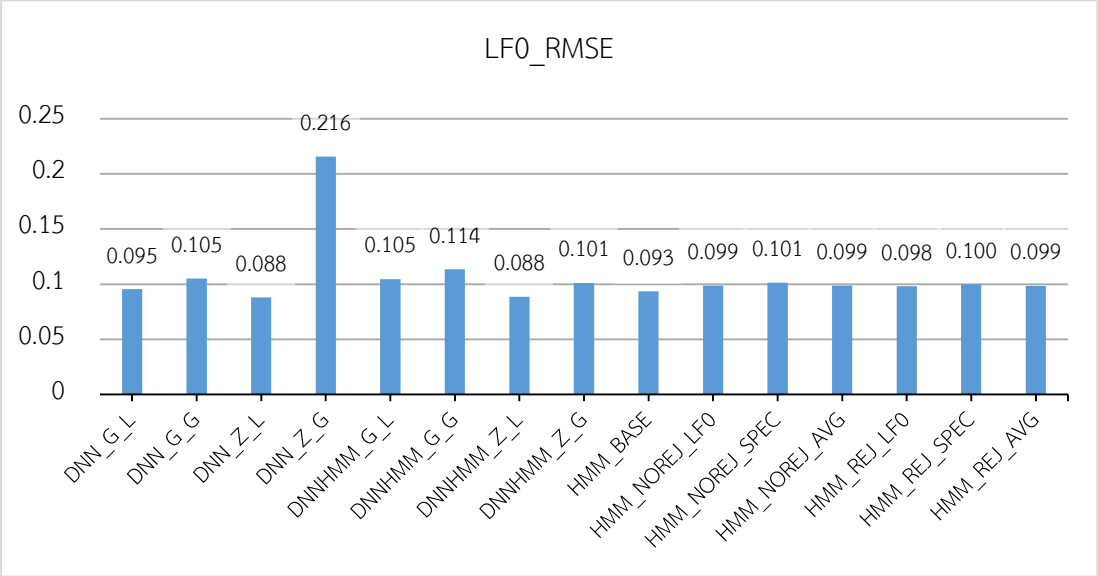
ผลการทดลอง LF0_RMSE แสดงในรูปที่ 42 ที่แสดงในรูปแบบของกราฟแท่ง ที่แกน X แสดงถึงรูปแบบของเสียงสังเคราะห์ และแกน Y แทนถึงค่าผลลัพธ์ที่ได้จากการคำนวณ โดยถ้ามีค่ามากจะหมายถึงมีความคลาดเคลื่อนของค่าคุณลักษณะความถี่มูลฐานระหว่างค่าคุณลักษณะที่สังเคราะห์ออกมาเปรียบเทียบกับเสียงจากฐานข้อมูลเสียงเป็นจำนวนมาก ซึ่งแทนถึงค่าที่ไม่ดี โดย

หน่วยของความถี่ของค่าความถี่มูลฐานที่ใช้ในการคำนวณคือ Log ของหน่วยเฮิรตซ์ และผลการวิเคราะห์ความแตกต่างกันของแต่ละรูปแบบแสดงในตารางที่ 19

โดยช่องตารางที่มีพื้นหลังเป็นสีเทา แสดงถึงคู่ของรูปแบบที่มีความแตกต่างอย่างไม่มีนัยสำคัญทางสถิติ

จากผลการทดลองแสดงให้เห็นว่าในทุกรูปแบบของแบบจำลองเสียงสังเคราะห์มีระดับคะแนนที่เท่ากัน ยกเว้นในรูปแบบ DNNHMM_G_G ที่มีระดับคะแนนมากกว่า (แย่กว่า) รูปแบบอื่นๆ อย่างมีนัยสำคัญทางสถิติ ดังแสดงในตารางที่ 19

ความแตกต่างในแต่ละรูปแบบของแบบจำลองฮิดเดนมาร์คอฟมีค่าน้อยมากจนไม่มีนัยสำคัญ



รูปที่ 42 ผลการทดลอง LF0_RMSE

จากการวัดผลคุณภาพของเสียง ถ้าพิจารณาระหว่างรูปแบบที่สร้างมาจากแบบจำลองโครงข่ายประสาทเทียมแบบลึกเปรียบเทียบกับรูปแบบที่สร้างมาจากแบบจำลองฮิดเดนมาร์คอฟ พบว่า รูปแบบที่สร้างขึ้นมาจากโครงข่ายประสาทเทียมแบบลึกมีคุณภาพดีกว่าในตัวชี้วัด MGC_MCD, BAP_MCD และ LF0_UVU (ยกเว้นในรูปแบบ HMM_REJ_LF0 และ HMM_REJ_AVG ที่ไม่แตกต่างกัน) และมีคุณภาพไม่แตกต่างกันในตัวชี้วัด LF0_RMSE ยกเว้นในกรณีของรูปแบบ DNNHMM_G_G ที่มีคุณภาพแย่กว่ารูปแบบอื่น อย่างมีนัยสำคัญทางสถิติ

เมื่อเปรียบเทียบคุณภาพเสียงสังเคราะห์ที่สังเคราะห์มาจากแบบจำลองฮิดเดนมาร์คอฟ พบว่าผลลัพธ์ของทุกรูปแบบในตัวชี้วัด MGC_MCD และ LF0_RMSE มีค่าไม่แตกต่างกัน ยกเว้นในการทดลอง BAP_MCD และ LF0_UVU ที่การใช้รูปแบบ HMM_BASE ได้ผลลัพธ์ที่ดีกว่ารูปแบบที่

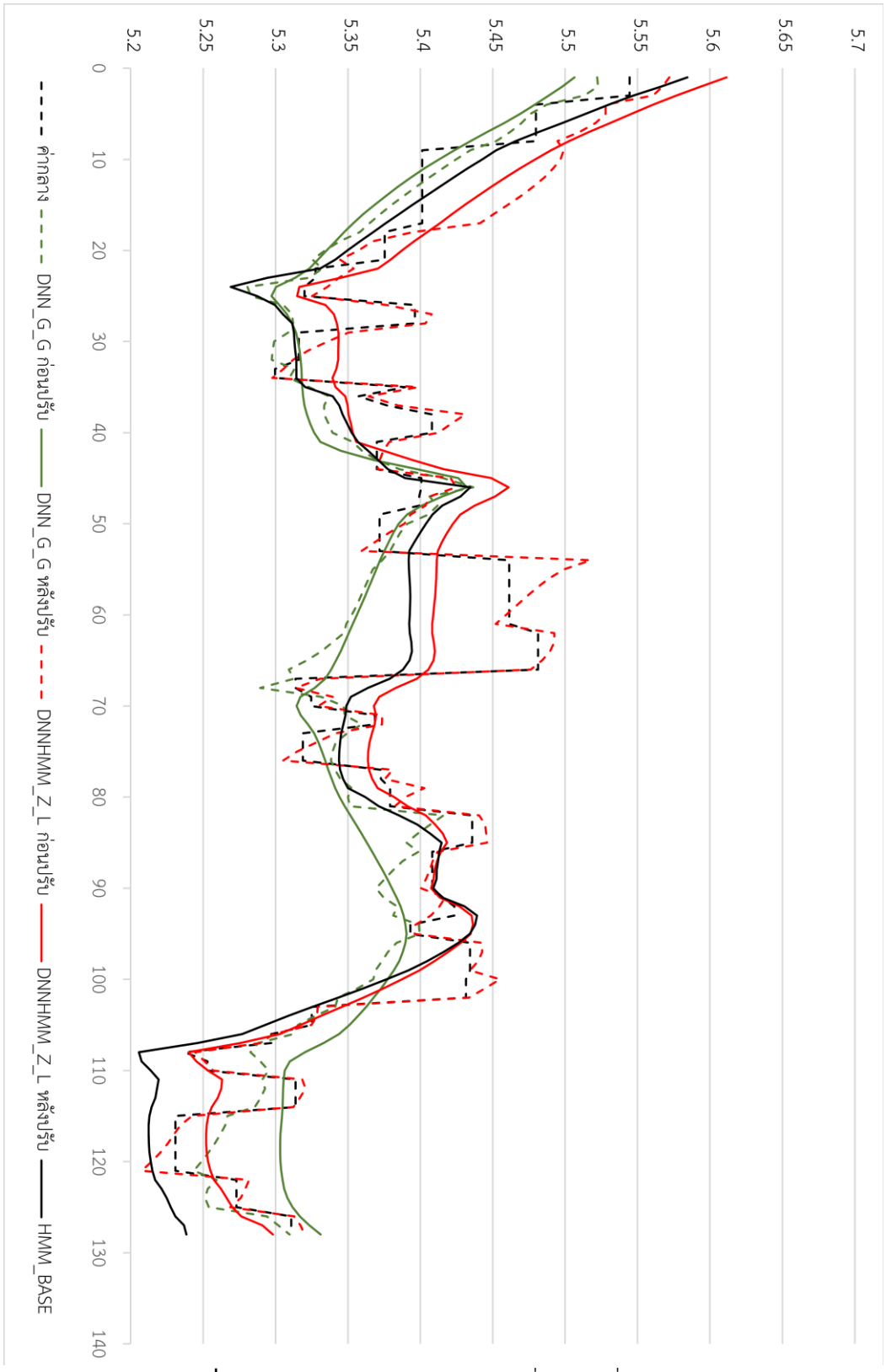
นำเสนอในงานวิจัยนี้ และวิธีการคำนวณหาช่วงเวลาของสถานะด้วยการพิจารณาสถานะความก้องของเสียง ช่วยให้ผลลัพธ์ในตัวชี้วัด BAP_MCD และ LF0_UVU ดีขึ้น เมื่อเทียบกับรูปแบบที่ไม่พิจารณาสถานะความก้องของเสียงอย่างมีนัยสำคัญทางสถิติ

เมื่อเปรียบเทียบคุณภาพเสียงสังเคราะห์ที่สังเคราะห์มาจากแบบจำลองโครงข่ายประสาทเทียมแบบลึก พบว่า ในตัวชี้วัด LF0_RMSE และ LF0_UVU (ยกเว้นรูปแบบ DNNHMM_G_G) มีค่าไม่แตกต่างกัน และในตัวชี้วัด MGC_MCD และ BAP_MCD ผลลัพธ์ที่ได้จากแต่ละรูปแบบแตกต่างกันอย่างมีนัยสำคัญทางสถิติ และรูปแบบที่มีการนอร์มัลไลเซชันค่าคุณลักษณะส่งออกด้วยคะแนนมาตรฐานอ้างอิงจากแบบจำลองฮิดเดนมาร์คอฟได้ผลลัพธ์ที่ดีกว่ารูปแบบอื่น

ตัวอย่างของการสังเคราะห์ค่าคุณลักษณะของแบบจำลองโครงข่ายประสาทเทียมแบบลึกเปรียบเทียบกับแบบจำลองฮิดเดนมาร์คอฟของค่าคุณลักษณะความถี่มูลฐาน แสดงดังรูปที่ 43 โดยในกรณีของโครงข่ายประสาทเทียมแบบลึกได้แสดงตัวอย่างของรูปแบบ DNN_G_G ที่ไม่มีการใช้ความรู้ของแบบจำลองฮิดเดนมาร์คอฟ และตัวอย่างของรูปแบบ DNNHMM_Z_L ที่มีการใช้ความรู้ของแบบจำลองฮิดเดนมาร์คอฟ และเส้นของค่ากลางได้มาจากแบบจำลอง HMM_BASE

ในกรณีของรูปแบบโครงข่ายประสาทเทียมแบบลึก ได้แสดงค่าของคุณลักษณะความถี่มูลฐานทั้งในช่วงก่อนปรับแนวค่าคุณลักษณะ (แสดงเป็นเส้นประ) และช่วงหลังปรับแนวค่าคุณลักษณะ (แสดงเป็นเส้นทึบ)

จากผลลัพธ์ที่แสดงในรูปที่ 43 แสดงให้เห็นชัดเจนว่าค่าคุณลักษณะที่สังเคราะห์จากแบบจำลอง DNNHMM_Z_L ในช่วงก่อนปรับค่าคุณลักษณะ มีค่าเข้าใกล้ค่ากลางของแบบจำลองฮิดเดนมาร์คอฟ มากกว่าแบบจำลอง DNN_G_G และเมื่อปรับค่าคุณลักษณะทำการปรับแนวค่าคุณลักษณะแล้ว ผลลัพธ์ของ DNNHMM_Z_L ยังคงมีแนวโน้มของการเปลี่ยนแปลงค่าคุณลักษณะเข้าใกล้ค่ากลางของแบบจำลอง HMM_BASE มากกว่าผลของค่าคุณลักษณะที่สังเคราะห์มาจากแบบจำลอง HMM_BASE และแบบจำลอง DNN_G_G ซึ่งสอดคล้องกับผลการทดสอบ LF0_RMSE



รูปที่ 43 ตัวอย่างค่าคุณลักษณะความถี่มูลฐานที่ถูกสังเคราะห์

ตารางที่ 16 การเปรียบเทียบความแตกต่างของการทดลอง MGC_MGD ที่ความเชื่อมั่นระดับ 95%

	DNN_G_G	DNN_Z_L	DNN_Z_G	DNNHMM_G_L	DNNHMM_G_G	DNNHMM_Z_L	DNNHMM_Z_G	HMM_BASE	HMM_NOREJ_LF0	HMM_NOREJ_SPEC	HMM_NOREJ_AVG	HMM_REJ_LF0	HMM_REJ_SPEC	HMM_REJ_AVG
DNN_G_L	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
DNN_G_G	-	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
DNN_Z_L	-	-	0.013	<0.001	<0.001	0.042	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
DNN_Z_G	-	-	-	<0.001	<0.001	0.449	0.037	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
DNNHMM_G_L	-	-	-	-	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
DNNHMM_G_G	-	-	-	-	-	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
DNNHMM_Z_L	-	-	-	-	-	-	0.011	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
DNNHMM_Z_G	-	-	-	-	-	-	-	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
HMM_BASE	-	-	-	-	-	-	-	-	0.209	0.499	0.483	0.071	0.188	0.237
HMM_NOREJ_LF0	-	-	-	-	-	-	-	-	-	0.210	0.197	0.254	0.462	0.462
HMM_NOREJ_SPEC	-	-	-	-	-	-	-	-	-	-	0.482	0.071	0.183	0.238
HMM_NOREJ_AVG	-	-	-	-	-	-	-	-	-	-	-	0.065	0.172	0.224
HMM_REJ_LF0	-	-	-	-	-	-	-	-	-	-	-	-	0.285	0.225
HMM_REJ_SPEC	-	-	-	-	-	-	-	-	-	-	-	-	-	0.425

ตารางที่ 17 การเปรียบเทียบความแตกต่างของการทดลอง BAP_MGD ที่ความเชื่อมั่นระดับ 95%

	DNN_G_L	DNN_G_G	DNN_Z_L	DNN_Z_G	DNNHMM_G_L	DNNHMM_G_G	DNNHMM_Z_L	DNNHMM_Z_G	HMM_BASE	HMM_NOREJ_LF0	HMM_NOREJ_SPEC	HMM_NOREJ_AVG	HMM_REJ_LF0	HMM_REJ_SPEC	HMM_REJ_AVG
DNN_G_G	0.165	-	-	-	-	-	-	-	-	-	-	-	-	-	-
DNN_Z_L	<0.001	<0.001	-	-	-	-	-	-	-	-	-	-	-	-	-
DNN_Z_G	<0.001	<0.001	0.231	-	-	-	-	-	-	-	-	-	-	-	-
DNNHMM_G_L	<0.001	<0.001	<0.001	<0.001	-	-	-	-	-	-	-	-	-	-	-
DNNHMM_G_G	0.017	0.001	<0.001	<0.001	<0.001	0.113	-	-	-	-	-	-	-	-	-
DNNHMM_Z_L	<0.001	<0.001	0.216	0.480	<0.001	0.004	<0.001	-	-	-	-	-	-	-	-
DNNHMM_Z_G	<0.001	<0.001	0.057	0.199	<0.001	0.04	0.001	0.214	-	-	-	-	-	-	-
HMM_BASE	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	-	-	-	-	-	-	-
HMM_NOREJ_LF0	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	-	-	-	-	-	-
HMM_NOREJ_SPEC	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	-	-	-	-	-
HMM_NOREJ_AVG	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.157	-	-	-	-
HMM_REJ_LF0	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	-	-	-
HMM_REJ_SPEC	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.001	0.033	-	-
HMM_REJ_AVG	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.401	0.056	-

ตารางที่ 18 การเปรียบเทียบความแตกต่างของการทดลอง LF0_UVU ที่ความเชื่อมั่นระดับ 95%

	DNN_G_L	DNN_G_G	DNN_Z_L	DNN_Z_G	DNNHMM_G_L	DNNHMM_G_G	DNNHMM_Z_L	DNNHMM_Z_G	HMM_BASE	HMM_NOREJ_LF0	HMM_NOREJ_SPEC	HMM_NOREJ_AVG	HMM_REJ_LF0	HMM_REJ_SPEC
DNN_G_G	0.736	-												
DNN_Z_L	0.315	0.505	-											
DNN_Z_G	0.287	0.468	0.952	-										
DNNHMM_G_L	0.729	0.993	0.511	0.473	-									
DNNHMM_G_G	0.431	0.653	0.828	0.782	0.659	-								
DNNHMM_Z_L	0.677	0.937	0.557	0.517	0.944	0.711	-							
DNNHMM_Z_G	0.529	0.770	0.708	0.664	0.777	0.875	0.831	-						
HMM_BASE	0.779	0.536	0.199	0.179	0.531	0.286	0.486	0.363	-					
HMM_NOREJ_LF0	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	-				
HMM_NOREJ_SPEC	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.026	-			
HMM_NOREJ_AVG	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	0.544	0.106	-		
HMM_REJ_LF0	0.453	0.277	0.079	0.070	0.273	0.124	0.243	0.168	0.639	<0.001	<0.001	<0.001	-	
HMM_REJ_SPEC	0.602	0.854	0.630	0.588	0.860	0.791	0.916	0.914	0.422	<0.001	<0.001	<0.001	0.203	-
HMM_REJ_AVG	0.602	0.390	0.127	0.113	0.385	0.191	0.348	0.250	0.809	<0.001	<0.001	<0.001	0.819	0.297

ตารางที่ 19 การเปรียบเทียบความแตกต่างของการทดลอง LF0_RMSE ที่ความเชื่อมั่นระดับ 95%

	DNN_G_L	DNN_G_G	DNN_Z_L	DNN_Z_G	DNNHMM_G_L	DNNHMM_G_G	DNNHMM_Z_L	DNNHMM_Z_G	HMM_BASE	HMM_NOREJ_LF0	HMM_NOREJ_SPEC	HMM_NOREJ_AVG	HMM_REJ_LF0	HMM_REJ_SPEC	HMM_REJ_AVG
DNN_G_G	0.168	-	-	-	-	-	-	-	-	-	-	-	-	-	-
DNN_Z_L	0.185	0.032	-	-	-	-	-	-	-	-	-	-	-	-	-
DNN_Z_G	0.054	0.258	0.006	-	-	-	-	-	-	-	-	-	-	-	-
DNNHMM_G_L	0.164	0.493	0.03	0.264	-	-	-	-	-	-	-	-	-	-	-
DNNHMM_G_G	<0.001	<0.001	<0.001	0.003	<0.001	-	-	-	-	-	-	-	-	-	-
DNNHMM_Z_L	0.207	0.038	0.468	0.008	0.036	<0.001	-	-	-	-	-	-	-	-	-
DNNHMM_Z_G	0.496	0.171	0.182	0.055	0.166	<0.001	0.204	-	-	-	-	-	-	-	-
HMM_BASE	0.404	0.114	0.257	0.032	0.110	<0.001	0.283	0.400	-	-	-	-	-	-	-
HMM_NOREJ_LF0	0.365	0.269	0.107	0.103	0.262	<0.001	0.123	0.369	0.278	-	-	-	-	-	-
HMM_NOREJ_SPEC	0.291	0.341	0.074	0.145	0.334	<0.001	0.086	0.294	0.213	0.419	-	-	-	-	-
HMM_NOREJ_AVG	0.341	0.291	0.095	0.115	0.285	<0.001	0.110	0.344	0.256	0.474	0.445	-	-	-	-
HMM_REJ_LF0	0.370	0.265	0.109	0.101	0.258	<0.001	0.125	0.374	0.282	0.495	0.414	0.469	-	-	-
HMM_REJ_SPEC	0.369	0.266	0.109	0.101	0.259	<0.001	0.125	0.373	0.281	0.497	0.415	0.47	0.499	-	-
HMM_REJ_SPEC	0.364	0.270	0.106	0.103	0.264	<0.001	0.122	0.368	0.277	0.499	0.420	0.475	0.494	0.495	-

6.1.3 ความไม่สอดคล้องของสถานะความก้องของเสียง

ในกรณีที่สถานะความก้องของเสียงไม่สอดคล้องกันระหว่างตัวอย่างเสียงสังเคราะห์ และตัวอย่างเสียงจากฐานข้อมูลเสียงไม่สอดคล้องกัน สามารถแบ่งได้เป็น 2 กรณี ได้แก่

1. ข้อมูลในฐานข้อมูลเสียงเป็นเสียงก้อง แต่ในเสียงสังเคราะห์เป็นเสียงไม่ก้อง (NVSU)
2. ข้อมูลในฐานข้อมูลเสียงเป็นเสียงไม่ก้อง แต่ในเสียงสังเคราะห์เป็นเสียงก้อง (NUSV)

โดยปกติแล้วตัวอย่างเสียงที่เป็นเสียงก้อง คือเสียงในหน่วยเสียงที่เป็นสระ และบางหน่วยเสียงของพยัญชนะต้น และตัวสะกด ดังนั้นถ้าเสียงในส่วนที่เป็นเสียงก้องกลายเป็นเสียงไม่ก้อง จะทำให้ได้ผลลัพธ์ที่เหมือนสัญญาณเสียงรบกวนที่มีพลังงานอย่างมาก เพราะเป็นช่วงของหน่วยเสียงพยัญชนะ หรือตัวสะกด

ส่วนที่เป็นตัวอย่างเสียงแบบไม่ก้อง ได้แก่ ส่วนที่เป็นพยัญชนะต้น หรือตัวสะกดในบางหน่วยเสียงที่เป็นประเภทเสียงไม่ก้อง ซึ่งในหน่วยเสียงประเภทนี้ทุกตัวอย่างเสียงในหน่วยเสียงจะเป็นเสียงไม่ก้องทั้งหมด แต่อย่างไรก็ตามช่วงท้ายของหน่วยเสียงอาจจะมีบางสถานะที่เป็นเสียงก้อง เพราะเป็นส่วนที่ต้องเปลี่ยนแปลงรูปแบบการออกเสียงเพื่อเชื่อมโยงกับเสียงสระ แต่สำหรับตัวสะกดที่ประเภทเสียงกักที่เป็นแบบเสียงก้อง ก็ประกอบด้วยตัวอย่างเสียงที่เป็นเสียงไม่ก้องปนอยู่ด้วย เพราะต้องมีช่วงที่เป็นเสียงเงียบเพื่อใช้ในการกักลมก่อนที่จะทำการปล่อยลมออกมา

ถ้าเสียงที่สังเคราะห์มีความผิดพลาดโดยการเปลี่ยนเป็นเสียงก้องในบริเวณดังกล่าว (ไม่รวมส่วนที่เป็นสัญญาณเสียงเงียบ) อาจจะทำให้ผู้ฟังรับรู้สัญญาณเสียงที่ผิดไป เพราะในบางหน่วยเสียงมีฐานกรณ์ และรูปแบบการออกเสียงที่เหมือนกันต่างกันแค่เป็นเสียงก้อง หรือไม่ก้อง

นอกจากเหตุผลข้างต้นแล้วการเกิดเหตุการณ์ NUSV อาจจะเกิดจากความผิดพลาดในการถอดความ เพราะผู้พูดอาจจะมีการหยุดพูดเพื่อชั่วขณะ แต่ในการถอดความไม่ได้มีการระบุไว้ประกอบกับตัวสะกด และพยัญชนะต้นของหน่วยเสียงที่ติดกันเป็นเสียงก้องในทุกตัวอย่างเสียง ซึ่งถ้าไม่มีการระบุว่ามีหยุดพูดชั่วขณะ ระบบจะทำการสังเคราะห์ค่าที่เป็นเสียงก้องทุกตัวอย่างเสียง แต่ในความเป็นจริงแล้ว มีตัวอย่างเสียงที่เป็นเสียงเงียบปนอยู่ด้วย ซึ่งจะส่งผลต่อกระบวนการทดสอบปรนัยที่ต้องทำการปรับแนวค่าคุณลักษณะให้สอดคล้องกันโดยไม่มีการละทิ้งตัวอย่างในเสียงจากฐานข้อมูลเสียง เลยต้องทำการเลือกตัวอย่างเสียงที่เป็นเสียงพยัญชนะ หรือตัวสะกด ไปจับคู่กับเสียงเงียบ

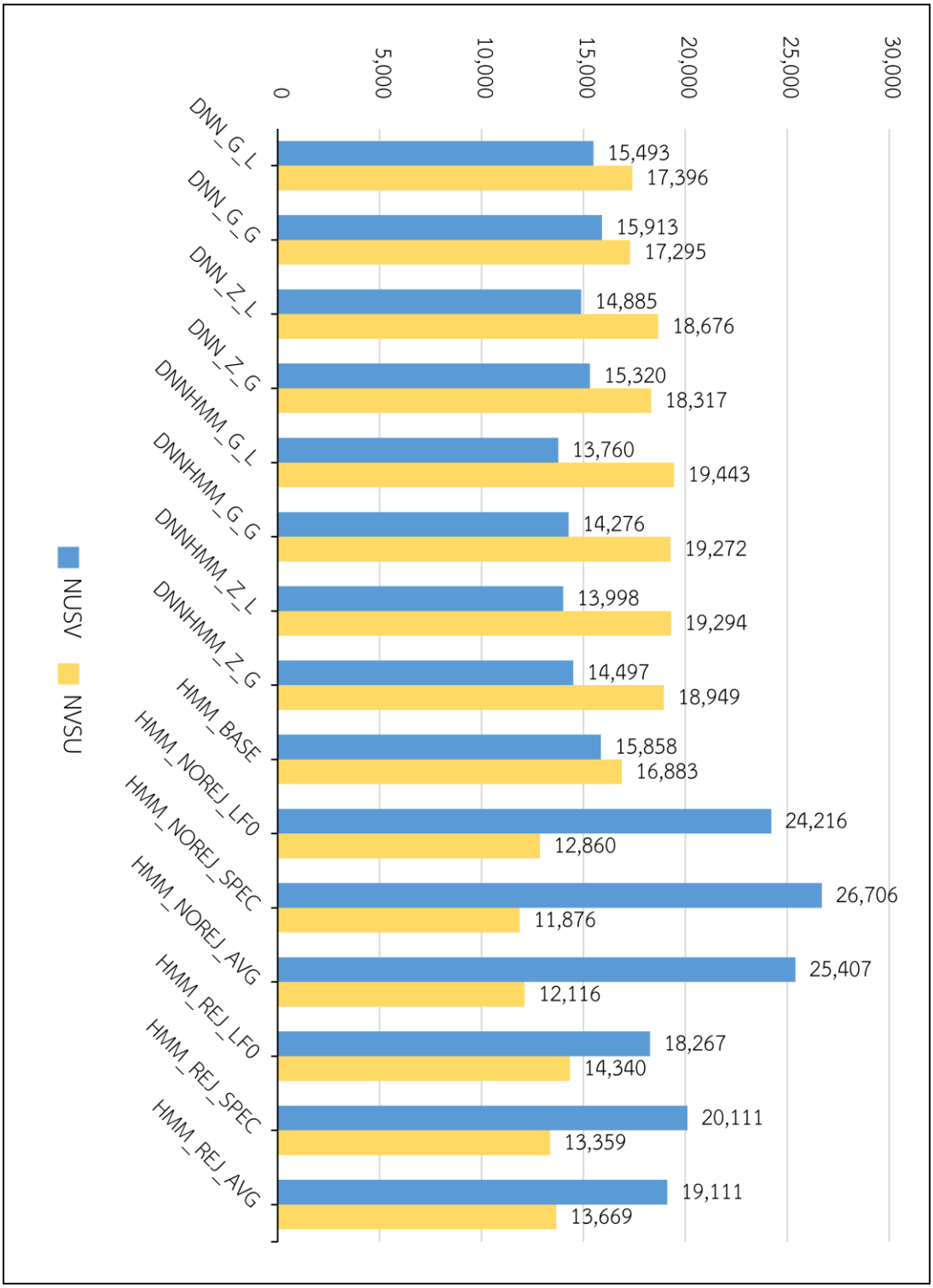
ผลการวิเคราะห์ในส่วนนี้แสดงดังรูปที่ 44 โดยสามารถสรุปได้เป็นประเด็น ดังต่อไปนี้

- ทุกรูปแบบของแบบจำลองโครงข่ายประสาทเทียมแบบลึก และแบบจำลองฮิดเดนมาร์คอฟที่ใช้แบบจำลองรวมกันของ 2 กระแสค่าคุณลักษณะ HMM_BASE มีผลลัพธ์ที่คล้ายกัน โดยที่มีตัวอย่างที่เป็นรูปแบบ NVSU มากกว่า หรือใกล้เคียงกับ NUSV ซึ่งต่างจากแบบจำลองฮิดเดนมาร์คอฟที่ใช้หลายแบบจำลองที่มีตัวอย่างที่เป็นรูปแบบ NUSV มากกว่ารูปแบบ NVSU

- รูปแบบของแบบจำลองโครงข่ายประสาทเทียมแบบลึก ที่มีตัวแปรที่นำเสนอในงานวิจัยนี้ พบว่ามีตัวอย่างเสียงที่เป็นรูปแบบ NVSU มากกว่าตัวอย่างเสียงที่เป็นรูปแบบ NUSV เมื่อเปรียบเทียบกับรูปแบบที่ไม่ได้มีตัวแปรที่นำเสนอในงานวิจัยนี้ (DNN_G_G และ DNN_G_L)

- รูปแบบที่ไม่มีการพิจารณาสถานะความก้องของเสียงมีรูปแบบ NUSV มากกว่า NVSU ถึง 2 เท่า และเมื่อพิจารณาสถานะความก้องของเสียง พบว่าสามารถลดรูปแบบ NUSV ได้อย่างมาก แต่กลับทำให้มีรูปแบบ NVSU มากขึ้น แต่ก็ยังไม่มากเท่ากับรูปแบบ HMM_BASE

- เมื่อเปรียบเทียบกับผลการทดลอง LF0_UVU พบว่าการที่รูปแบบของแบบจำลองฮิดเดนมาร์คอฟแบบแยกหลายแบบจำลองที่นำเสนอมีผลการทดลอง LF0_UVU ที่แย่กว่าแบบจำลองฮิดเดนมาร์คอฟที่ใช้แบบจำลองรวมกันของหลายกระแส เพราะมีรูปแบบ NUSV ที่มากกว่า แต่กลับมีรูปแบบ NVSU ที่น้อยกว่า



รูปที่ 44 ผลการวิเคราะห์ความไม่สอดคล้องกันของสถานะ

6.2 การทดสอบแบบอัตนัย

ผลลัพธ์การทดสอบอัตนัยแบ่งออกเป็น 2 ประเด็นย่อย ได้แก่ เรื่องความเป็นธรรมชาติของเสียงสังเคราะห์ และความชัดเจนของเสียงสังเคราะห์ ซึ่งผลลัพธ์แสดงดังรูปที่ 45 และรูปที่ 46 ตามลำดับ โดยค่าที่แสดงอยู่ในกราฟทั้งสองคือค่าเฉลี่ยของผู้ทดสอบทั้ง 9 คน

ผลการทดลองในส่วนของความเป็นธรรมชาติของเสียงสังเคราะห์ พบว่ารูปแบบที่เป็นแบบจำลองโครงข่ายประสาทเทียมแบบลึกมีเพียงรูปแบบ DNNHMM_Z_L และรูปแบบ DNNHMM_Z_G เท่านั้นที่มีระดับคะแนนมากกว่ารูปแบบ HMM_BASE ซึ่งผลการทดลองในส่วนนี้ได้ขัดแย้งกับผลที่ได้จากการทดสอบแบบปรนัย

สำหรับรูปแบบที่ใช้แบบจำลองฮิดเดนมาร์คอฟ พบว่ารูปแบบที่นำเสนอ และมีการพิจารณาสถานะความก้องของเสียงให้ผลลัพธ์ที่ดีกว่า HMM_BASE และดีกว่าการไม่พิจารณาสถานะความก้องของเสียงโดยรูปแบบ HMM_REJ_SPEC ให้ผลลัพธ์ที่ดีที่สุดในระดับคะแนน 6.99 และลำดับต่อมาคือรูปแบบ HMM_REJ_LF0 ที่ระดับคะแนน 6.94 และรูปแบบ HMM_REJ_AVG ให้ระดับคำแนนที่ต่ำที่สุดที่ระดับคะแนน 6.91

จากผลการทดสอบการวัดผลช่วงเวลาของแบบจำลองเสียง พบว่าแบบจำลองที่หาช่วงเวลาของหน่วยเสียงจากการคำนวณร่วมกันของแบบจำลองช่วงเวลาของค่าคุณลักษณะสเปกตรัม และค่าคุณลักษณะความถี่มูลฐานได้ผลลัพธ์ที่ดีกว่าการหาช่วงเวลาของหน่วยเสียงที่ใช้เพียงแบบจำลองใดแบบจำลองหนึ่ง เมื่อนำผลการทดลองจากการทดลองดังกล่าวมาเปรียบเทียบกับการทดสอบอัตนัย พบว่ามีผลลัพธ์ที่ไม่สอดคล้องกัน เพราะระดับความเป็นธรรมชาติที่ได้จากรูปแบบ HMM_REJ_AVG และรูปแบบ HMM_NOREJ_AVG มีค่าน้อยกว่ารูปแบบ HMM_REJ_SPEC, HMM_REJ_LF0 และ HMM_NOREJ_SPEC, HMM_NOREJ_LF0 ตามลำดับ

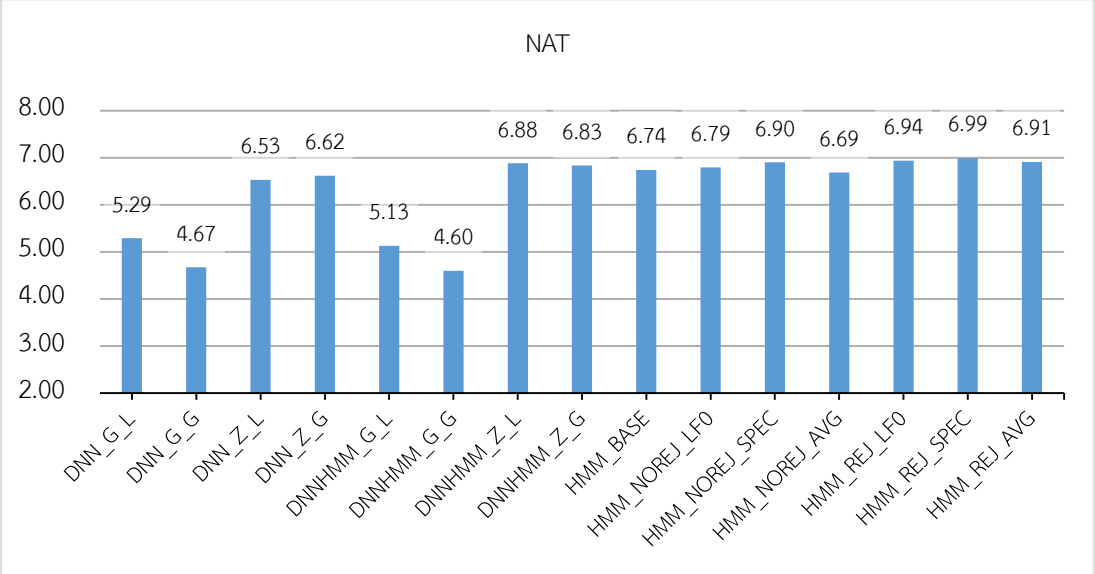
จากประเด็นดังกล่าวแสดงให้เห็นว่าผลต่างเฉลี่ยที่ค่าประมาณ 0.003 วินาทีต่อหน่วยเสียง ไม่ส่งผลต่อระดับความเป็นธรรมชาติของเสียงสังเคราะห์

สำหรับรูปแบบที่ใช้แบบจำลองโครงข่ายประสาทเทียมแบบลึก พบว่ารูปแบบที่มีการใช้การนอร์มัลไลเซชันค่าคุณลักษณะส่งออกด้วยคะแนนมาตรฐานอ้างอิงจากแบบจำลองฮิดเดนมาร์คอฟได้คะแนนมากกว่ารูปแบบที่ใช้การนอร์มัลไลเซชันด้วยการหารด้วยค่าความแปรปรวนอย่างมาก และรูปแบบที่ใช้ข้อมูลจากแบบจำลองที่ใช้ความรู้จากแบบจำลองฮิดเดนมาร์คอฟมากที่สุด (DNNHMM_Z_L) ได้ผลลัพธ์ที่ดีที่สุดที่ระดับคะแนน 6.88

สำหรับปัจจัยค่าคุณลักษณะส่วนรับเข้า พบว่าการใช้ตำแหน่งของต้นไม้ตัดสินใจร่วมกับการนอร์มัลไลเซชันค่าคุณลักษณะส่งออกถ่วงค่าน้ำหนักด้วยค่าความแปรปรวน พบว่าทำให้มีระดับคะแนนความเป็นธรรมชาติลดลงเมื่อเทียบกับการใช้ค่าคุณลักษณะส่วนรับเข้าที่มาจากคำถามของ

ต้นไม้ตัดสินใจ แต่เมื่อนำไปใช้กับการนอร์มัลไลเซชันค่าคุณลักษณะส่งออกด้วยคะแนนมาตรฐานอ้างอิงจากแบบจำลองฮิดเดนมาร์คอฟ พบว่ามีระดับคะแนนที่ดีขึ้น เมื่อเทียบกับการใช้ค่าคุณลักษณะส่วนรับเข้าที่มาจากคำถามของต้นไม้ตัดสินใจ

ปัจจัยค่าความแปรปรวนในกระบวนการปรับแนวคุณลักษณะของแบบจำลองโครงข่ายประสาทเทียมแบบลึก ส่งผลกับระดับคะแนนความเป็นธรรมชาติของเสียงสังเคราะห์อย่างมากเมื่อใช้กับรูปแบบที่มีการนอร์มัลไลเซชันค่าคุณลักษณะส่งออกด้วยการหารด้วยค่าความแปรปรวน โดยพบว่าการปรับแนวค่าคุณลักษณะด้วยค่าความแปรปรวนตามบริบทให้ผลลัพธ์ที่ดีกว่า แต่เมื่อนำไปใช้กับการนอร์มัลไลเซชันค่าคุณลักษณะส่งออกด้วยคะแนนมาตรฐานอ้างอิงจากแบบจำลองฮิดเดนมาร์คอฟ พบว่ามีระดับของความเป็นธรรมชาติที่ไม่แตกต่างกันมาก



รูปที่ 45 ผลการทดลองความเป็นธรรมชาติของเสียงสังเคราะห์

ผลการทดลองในส่วนของความชัดเจนของเสียงสังเคราะห์ พบว่ารูปแบบที่เป็นแบบจำลองโครงข่ายประสาทเทียมแบบลึกทุกรูปแบบมีระดับความชัดเจนของเสียงสังเคราะห์น้อยกว่ารูปแบบที่ใช้แบบจำลองฮิดเดนมาร์คอฟทั้งหมด ซึ่งผลการทดลองในส่วนนี้ได้ขัดแย้งกับผลที่ได้จากการทดสอบแบบปรนัย

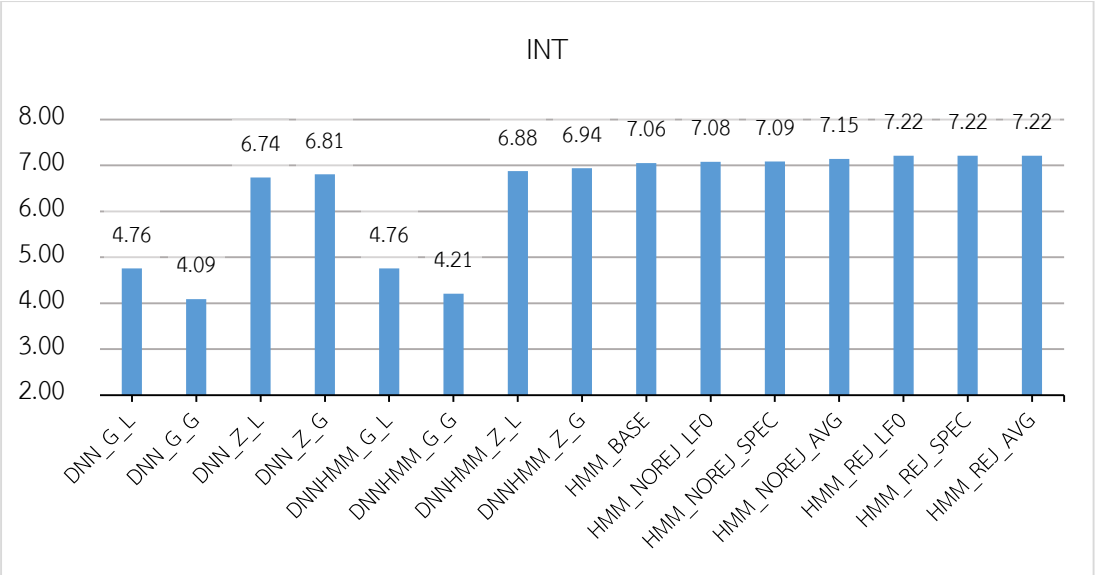
สำหรับรูปแบบที่ใช้แบบจำลองฮิดเดนมาร์คอฟ พบว่ารูปแบบที่นำเสนอ และมีการพิจารณาสถานะความก้องของเสียงให้ผลลัพธ์ที่ดีกว่า HMM_BASE และดีกว่าการไม่พิจารณาสถานะความก้องของเสียงโดยทั้ง 3 รูปแบบ ซึ่งประกอบด้วยรูปแบบ HMM_REJ_SPEC, HMM_REJ_AVG และ HMM_REJ_LF0 ให้ผลลัพธ์ที่ดีที่สุดในระดับคะแนนประมาณ 7.22 เท่ากันทั้ง 3 รูปแบบ และในกรณี

ที่ไม่พิจารณาสถานะความก้องของเสียง พบว่ารูปแบบ HMM_NOREJ_AVG ได้ระดับคะแนนที่ดีที่สุดที่ระดับ 7.15 รองลงมาคือ HMM_NOREJ_SPEC ที่ระดับ 7.09 และที่แย่ที่สุดคือ HMM_NOREJ_LF0 ที่ระดับ 7.08 โดยแบบจำลองที่ได้ระดับคะแนนน้อยที่สุดของรูปแบบที่ใช้แบบจำลองฮิดเดนมาร์คอฟคือ HMM_BASE ที่ระดับคะแนน 7.06

สำหรับรูปแบบที่ใช้แบบจำลองโครงข่ายประสาทเทียมแบบลึกพบว่ารูปแบบที่มีการใช้การนอร์มัลไลเซชันค่าคุณลักษณะส่งออกด้วยคะแนนมาตรฐานอ้างอิงจากแบบจำลองฮิดเดนมาร์คอฟได้ มีระดับความชัดเจนของเสียงสังเคราะห์อยู่ที่ 6.7 – 7.0 คะแนน ซึ่งมากกว่ารูปแบบที่ใช้การนอร์มัลไลเซชันค่าคุณลักษณะส่งออกด้วยการหารด้วยค่าความแปรปรวนที่มีระดับความชัดเจนของเสียงเพียง 4.0 - 4.8 คะแนน

สำหรับปัจจัยค่าคุณลักษณะส่วนรับเข้า พบว่าการใช้ตำแหน่งของต้นไม้ตัดสินใจช่วยเพิ่มระดับความชัดเจนของเสียงสังเคราะห์เพียงเล็กน้อยเท่านั้น และไม่เพิ่มขึ้นในรูปแบบ DNN_G_L เปรียบเทียบกับ DNNHMM_G_L

ปัจจัยค่าความแปรปรวนในกระบวนการปรับแนวคุณลักษณะของแบบจำลองโครงข่ายประสาทเทียมแบบลึก ส่งผลกับระดับคะแนนความชัดเจนของเสียงสังเคราะห์ เมื่อใช้กับรูปแบบที่มีการนอร์มัลไลเซชันค่าคุณลักษณะส่งออกด้วยการหารด้วยค่าความแปรปรวน มากกว่าการนอร์มัลไลเซชันค่าคุณลักษณะส่งออกด้วยคะแนนมาตรฐานอ้างอิงจากแบบจำลองฮิดเดนมาร์คอฟ โดยในรูปแบบที่มีการนอร์มัลไลเซชันค่าคุณลักษณะส่งออกด้วยการหารด้วยค่าความแปรปรวน พบว่าการใช้ความแปรปรวนในกระบวนการปรับแนวคุณลักษณะด้วยความแปรปรวนตามบริบทให้ระดับความชัดเจนของเสียงสังเคราะห์ที่ดีกว่าใช้ความแปรปรวนแบบครอบคลุม แต่ในทางกลับกันเมื่อนำไปใช้กับการนอร์มัลไลเซชันค่าคุณลักษณะส่งออกด้วยคะแนนมาตรฐานอ้างอิงจากแบบจำลองฮิดเดนมาร์คอฟกลับพบว่ามีระดับความชัดเจนของเสียงที่แย่ลง แต่แย่ลงเพียงเล็กน้อยเมื่อเปรียบเทียบกับการเพิ่มขึ้นในกรณีก่อนหน้า



รูปที่ 46 ผลการทดลองความชัดเจนของเสียงสังเคราะห์

บทที่ 7 สรุปและอภิปรายผลการวิจัย

งานวิจัยนี้ได้นำเสนอแนวคิดการพัฒนาเสียงสังเคราะห์ภาษาไทยที่ใช้แบบจำลองฮิดเดนมาร์คอฟ และแบบจำลองโครงข่ายประสาทเทียมแบบลึก เพื่อเพิ่มความเป็นธรรมชาติของเสียงสังเคราะห์ภาษาไทย

สำหรับแบบจำลองฮิดเดนมาร์คอฟได้นำเสนอแนวคิดการใช้หลายแบบจำลองของแบบจำลองฮิดเดนมาร์คอฟ ซึ่งแยกค่าคุณลักษณะสเปกตรัม และค่าคุณลักษณะความถี่มูลฐานออกเป็นคนละแบบจำลอง และนำเสนอวิธีการคำนวณช่วงเวลาของหน่วยเสียงที่ใช้หลายแบบจำลอง พร้อมทั้งวิธีการคำนวณช่วงเวลาของสถานะภายในหน่วยเสียง ในกรณีที่มีการพิจารณาสถานะความก้องของเสียงของค่าคุณลักษณะที่สังเคราะห์ออกมาจากแบบจำลองทั้งสอง

สำหรับแบบจำลองโครงข่ายประสาทเทียมแบบลึกได้นำเสนอแนวคิดการสร้างค่าคุณลักษณะส่วนรับเข้าจากตำแหน่งของต้นไม้ตัดสินใจที่ได้จากแบบจำลองฮิดเดนมาร์คอฟ และการนอร์มัลไลเซชันค่าคุณลักษณะส่งออกด้วยคะแนนมาตรฐานอ้างอิงจากแบบจำลองฮิดเดนมาร์คอฟ

จากผลการทดลองในส่วนของการทดสอบแบบปรนัย แสดงให้เห็นว่าผลการทดลองในตัวชี้วัด MGC_MCD และ BAP_MCD ของแบบจำลองที่เป็นประเภทโครงข่ายประสาทเทียมแบบลึกทุกรูปแบบ ได้ผลลัพธ์ที่ดีกว่าแบบจำลองฮิดเดนมาร์คอฟ และในตัวชี้วัด LF0_UVU และ LF0_RMSE ของรูปแบบประเภทโครงข่ายประสาทเทียมได้คะแนนในระดับที่ไม่แตกต่างกันมากเมื่อเทียบกับแบบจำลองฮิดเดนมาร์คอฟ แต่เมื่อพิจารณาผลการทดลองอัตนัยกลับพบว่าเสียงสังเคราะห์ที่ได้จากโครงข่ายประสาทเทียมแบบลึกทุกรูปแบบ มีระดับของความชัดเจนของเสียงสังเคราะห์ และความเป็นธรรมชาติของเสียงสังเคราะห์ที่แย่กว่า หรือใกล้เคียงกับเสียงสังเคราะห์ที่ได้จากแบบจำลองฮิดเดนมาร์คอฟเท่านั้น

ซึ่งประเด็นดังกล่าวอาจจะเกิดขึ้นจากการที่ทำการเปรียบเทียบเสียงสังเคราะห์ที่มีรูปแบบของเสียงที่แตกต่างกันมากเกินไป ดังตัวอย่างในงานวิจัยของ Hashimoto และคณะ. (2016) [50] ที่แสดงให้เห็นว่ามีโอกาสที่ผลจากตัวชี้วัดประเภทปรนัยไม่สอดคล้องกับผลที่ได้จากการทดลองอัตนัย ประกอบกับประสิทธิภาพของแบบจำลองโครงข่ายประสาทเทียมแบบลึกที่สามารถเรียนรู้ และสังเคราะห์ค่าที่ใกล้เคียงกับข้อมูลฝึกฝนได้เป็นอย่างดี จึงทำให้ผลลัพธ์ในส่วนของการทดสอบปรนัยดีกว่าการใช้แบบจำลองฮิดเดนมาร์คอฟ แต่อย่างไรก็ตามความรับรู้ถึงความเป็นธรรมชาติ และความชัดเจนของเสียงสังเคราะห์ ต้องพิจารณาถึงวิถีการเปลี่ยนแปลงของค่าคุณลักษณะ และความสอดคล้องกันของแต่ละลำดับในค่าคุณลักษณะ ซึ่งปัจจัยในส่วนนี้ไม่ได้ถูกรวมเข้าไปในกระบวนการวัดผลปรนัย

เมื่อพิจารณาเฉพาะผลการทดลองปรนัย และการทดลองอัตนัยของแบบจำลองโครงข่ายประสาทเทียมแบบลึก พบว่ามีส่วนที่สอดคล้องกันในส่วนของการนอร์มัลไลเซชันค่าคุณลักษณะส่งออก โดยเมื่อใช้การการนอร์มัลไลเซชันค่าคุณลักษณะส่งออกด้วยการใช้ค่าคะแนนมาตรฐานอ้างอิงจากแบบจำลองฮิดเดนมาร์คอฟ ช่วยให้คุณภาพเสียงสังเคราะห์มีคุณภาพที่ดีกว่าการนอร์มัลไลเซชันด้วยการหารด้วยค่าความแปรปรวนแบบครอบคลุม ซึ่งแสดงให้เห็นว่าการใช้ความรู้ของแบบจำลองฮิดเดนมาร์คอฟในรูปแบบของการนอร์มัลไลเซชันค่าคุณลักษณะส่งออกให้ผลลัพธ์ที่ดีที่สุด

สำหรับปัจจัยในด้านของค่าคุณลักษณะส่วนรับเข้า พบว่าการใช้ค่าคุณลักษณะจากตำแหน่งของต้นไม้ตัดสินใจจะช่วยให้ได้คะแนนที่ดีขึ้นในการทดสอบแบบปรนัยเมื่อใช้ร่วมกับการนอร์มัลไลเซชันค่าคุณลักษณะส่วนส่งออกด้วยการหารด้วยค่าความแปรปรวนแบบครอบคลุม ซึ่งขัดแย้งการผลการทดสอบอัตนัยที่การใช้ค่าคุณลักษณะจากตำแหน่งของต้นไม้ตัดสินใจจะช่วยให้ได้คุณภาพของเสียงสังเคราะห์ที่ดีขึ้นเมื่อใช้ร่วมกับการนอร์มัลไลเซชันด้วยคะแนนมาตรฐานอ้างอิงจากแบบจำลองฮิดเดนมาร์คอฟ

เนื่องจากการนอร์มัลไลเซชันด้วยการหารด้วยค่าความแปรปรวนแบบครอบคลุม ทำให้ได้ค่าคุณลักษณะส่งออกที่มีช่วงแคบกว่าการนอร์มัลไลเซชันด้วยคะแนนมาตรฐานอ้างอิงจากแบบจำลองฮิดเดนมาร์คอฟที่มีค่าความแปรปรวนน้อยกว่า ประกอบกับการใช้การตั้งค่าการฝึกฝนที่เหมือนกัน ทำให้ผลลัพธ์ของฟังก์ชันต้นทุนที่ได้จากใช้การนอร์มัลไลเซชันด้วยการหารด้วยค่าความแปรปรวนแบบครอบคลุมมีค่าน้อยกว่าการนอร์มัลไลเซชันด้วยคะแนนมาตรฐานอ้างอิงจากแบบจำลองฮิดเดนมาร์คอฟ ซึ่งอาจจะทำให้ฝึกฝนข้อมูลได้ดีกว่า แต่อย่างไรก็ตามการสังเคราะห์ค่าที่เหมือนกับข้อมูลฝึกฝนมาก ไม่ได้ส่งผลกับการรับรู้ของผู้ฟังในการทดสอบอัตนัย

สำหรับปัจจัยด้านความแปรปรวนในกระบวนการปรับแนวคุณลักษณะเมื่อใช้ค่าความแปรปรวนแบบครอบคลุม จะได้ผลคะแนนในการทดสอบแบบปรนัยที่แย่กว่าการใช้ความแปรปรวนตามบริบทในทุกรูปแบบ แต่ในการทดสอบแบบอัตนัยกลับมีบางรูปแบบที่ค่าความแปรปรวนแบบครอบคลุมจะให้คุณภาพของเสียงสังเคราะห์ที่ดีกว่า โดยเป็นรูปแบบที่มีการนอร์มัลไลเซชันด้วยคะแนนมาตรฐานอ้างอิงจากแบบจำลองฮิดเดนมาร์คอฟ

โดยในประเด็นนี้อธิบายได้จากการที่กระบวนการปรับแนวคุณลักษณะจะใช้ค่าความแปรปรวนในการกำหนดขอบเขตที่สามารถเปลี่ยนแปลงค่าได้ โดยค่าความแปรปรวนจากค่าความแปรปรวนตามบริบทจะมีค่าน้อยกว่าค่าความแปรปรวนแบบครอบคลุม ทำให้มีการเปลี่ยนแปลงค่าคุณลักษณะที่น้อยกว่า ซึ่งส่งผลให้ได้ค่าคุณลักษณะที่ใกล้เคียงกับข้อมูลฝึกฝนมากกว่าจึงทำให้ได้คะแนนในการทดสอบปรนัยที่ดีกว่า

สำหรับแบบจำลองฮิดเดนมาร์คอฟพบว่าการทดสอบแบบปรนัยเปรียบเทียบกับการทดสอบอัตนัยมีผลลัพธ์ที่ตรงข้ามกันทุกรูปแบบ โดยเมื่อทำการวิเคราะห์ผลในการทดสอบปรนัยที่ตัวชี้วัด

MGC_MCD กับตัวชี้วัด LF0_RMSE พบว่ามีระดับความแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ แต่สำหรับตัวชี้วัด BAP_MCD และ LF0_UVU มีผลลัพธ์ที่แตกต่างกันอย่างมีนัยสำคัญทางสถิติโดยรูปแบบ HMM_BASE มีผลลัพธ์ที่ดีกว่า แต่สำหรับการทดสอบอัตนัยกลับพบว่ารูปแบบ HMM_BASE ได้ระดับคะแนนในส่วนของความชัดเจน และความเป็นธรรมชาติของเสียงสังเคราะห์ที่แย่ที่สุด

โดยในกรณีที่ BAP_MCD ของรูปแบบที่นำเสนอมีค่าระดับคะแนนที่แย่กว่า เพราะในขั้นตอนการทำการทดสอบแบบปรนัยต้องทำการปรับแนวทางเวลาให้จำนวนตัวอย่างของเสียงสังเคราะห์และเสียงจากฐานข้อมูลเสียงมีช่วงเวลาที่เท่ากัน โดยในขั้นตอนการปรับแนวทางเวลาได้พิจารณาเฉพาะค่าคุณลักษณะสเปกตรัมเท่านั้น และในกรณีที่มีหลายตัวอย่างของเสียงสังเคราะห์เชื่อมโยงไปยังหนึ่งตัวอย่างของเสียงต้นฉบับจากฐานข้อมูลเสียง จะเลือกใช้ตัวอย่างเสียงสังเคราะห์ที่มีค่า MGC_MCD ดีที่สุด (น้อยที่สุด) เพื่อใช้ในการคำนวณตัวชี้วัดอื่นๆ ซึ่งการเลือกตัวอย่างที่ทำให้ได้ค่า MGC_MCD ดีที่สุด อาจจะไม่ได้ทำให้ได้ค่า BAP_MCD ที่ดีที่สุด

สำหรับประเด็นของตัวชี้วัด LF0_UVU ที่พบว่าค่าของรูปแบบที่นำเสนอมีค่าแย่กว่า (มากกว่า) รูปแบบ HMM_BASE ทุกรูปแบบ แต่ในการทดสอบอัตนัยทั้งในด้านของความเป็นธรรมชาติของเสียงสังเคราะห์ และความชัดเจนของเสียงสังเคราะห์กลับทำคะแนนได้ดีกว่าในทุกรูปแบบ ซึ่งรวมถึงรูปแบบที่ไม่พิจารณาสถานะความก้องของเสียง (รูปแบบที่ขึ้นต้นด้วย HMM_NOREJ) ซึ่งเหตุผลที่อธิบายประเด็นนี้ มีดังต่อไปนี้

1. การปรับแนวทางเวลาส่งผลกับตัวชี้วัด LF0_UVU เพราะการปรับแนวทางเวลาจะทำการจับคู่ค่าคุณลักษณะทางสเปกตรัมระหว่างเสียงที่สังเคราะห์ออกมาเทียบกับเสียงต้นฉบับ ซึ่งอาจจะเกิดความคลาดเคลื่อนในการจับคู่ โดยเฉพาะในช่วงการต่อกันระหว่างตัวสะกดหรือพยัญชนะที่ไม่ได้เป็นเสียงก้องกับสระ

2. เมื่ออ้างอิงจากผลของความไม่สอดคล้องกันระหว่างสถานะความก้องของเสียงในหัวข้อที่ 6.1.4 พบว่ารูปแบบ HMM_BASE มีตัวอย่างเสียงประเภท NVSU มากที่สุดในทุกรูปแบบที่เป็นแบบจำลองฮิดเดนมาร์คอฟ ซึ่งการที่มีตัวอย่างประเภท NVSU มาก อาจจะส่งผลกับคุณภาพของเสียงสังเคราะห์มากกว่ารูปแบบ NUSV แต่อย่างไรก็ตามเมื่อเปรียบเทียบรูปแบบที่มีการพิจารณาสถานะความก้องของเสียงในขั้นตอนการคำนวณหาค่าระยะเวลาของสถานะ กลับพบว่ารูปแบบที่มีการพิจารณาสถานะความก้องของเสียง มีค่าความชัดเจนของเสียงสังเคราะห์มากกว่ารูปแบบที่ไม่มีการพิจารณาสถานะความก้องของเสียง ทั้งที่มีรูปแบบ NVSU มากกว่า ซึ่งอาจเป็นเพราะรูปแบบที่มีการพิจารณาสถานะความก้องของเสียงลดตัวอย่างเสียงที่เป็น NUSV ลงอย่างมากเมื่อเปรียบเทียบกับรูปแบบที่ไม่มีการพิจารณาสถานะความก้องของเสียง จึงทำให้รูปแบบที่มีการพิจารณาสถานะความก้องของเสียงได้เสียงสังเคราะห์ที่มีคุณภาพมากกว่า

รายการอ้างอิง

1. Hansakunbuntheung, C., V. Tesprasit, and V. Sornlertlamvanich, *Thai tagged speech corpus for speech synthesis.* The Oriental COCOSDA 2003, 2003: p. 97-104.
2. Black, A.W., H. Zen, and K. Tokuda. *Statistical parametric speech synthesis*. in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2007. IEEE.
3. Tokuda, K., et al., *Speech synthesis based on hidden markov models.* Proceedings of the IEEE, 2013. **101**(5): p. 1234-1252.
4. Masuko, T., et al. *Speech synthesis using HMMs with dynamic features*. in *Acoustics, Speech, and Signal Processing, 1996. ICASSP-96. Conference Proceedings., 1996 IEEE International Conference on*. 1996. IEEE.
5. Fukada, T., et al. *An adaptive algorithm for mel-cepstral analysis of speech*. in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 1992. IEEE.
6. Young, S., et al., *The HTK book (for HTK version 3.4).* 2006.
7. Tokuda, K., et al. *Hidden Markov models based on multi-space probability distribution for pitch pattern modeling*. in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 1999. IEEE.
8. Zen, H., K. Tokuda, and A.W. Black, *Statistical parametric speech synthesis.* Speech Communication, 2009. **51**(11): p. 1039-1064.
9. Yoshimura, T., et al., *Incorporating a mixed excitation model and postfilter into HMM-based text-to-speech synthesis.* Systems and Computers in Japan, 2005. **36**(12): p. 43-50.
10. Yoshimura, T., et al. *Mixed excitation for HMM-based speech synthesis*. in *Seventh European Conference on Speech Communication and Technology*. 2001.
11. Ling, Z.-H., et al. *USTC system for Blizzard Challenge 2006 an improved HMM-based speech synthesis method*. in *Blizzard Challenge Workshop*. 2006.

12. Qian, Y., H. Liang, and F.K. Soong. *Generating natural F0 trajectory with additive trees*. in *Ninth Annual Conference of the International Speech Communication Association*. 2008.

13. Latorre, J. and M. Akamine. *Multilevel parametric-base f0 model for speech synthesis*. in *Ninth Annual Conference of the International Speech Communication Association*. 2008.

14. Toda, T., A.W. Black, and K. Tokuda. *Spectral Conversion Based on Maximum Likelihood Estimation Considering Global Variance of Converted Parameter*. in *IEEE International Conference on Acoustics, Speech and Signal Processing*. 2005.

15. Yang, W. and T. Jianhua. *Evaluation of parameter generation using high order dynamic features and long span windows for HMM based speech synthesis*. in *Chinese Spoken Language Processing (ISCSLP), 2014 9th International Symposium on*. 2014.

16. Kawahara, H., J. Estill, and O. Fujimura. *Aperiodicity extraction and control using mixed mode excitation and group delay manipulation for a high quality speech analysis, modification and synthesis system STRAIGHT*. in *Second International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications*. 2001.

17. Young, S.J., J.J. Odell, and P.C. Woodland. *Tree-based state tying for high accuracy acoustic modelling*. in *Proceedings of the workshop on Human Language Technology*. 1994. Association for Computational Linguistics.

18. Shinoda, K. and T. Watanabe. *Acoustic modeling based on the MDL criterion for speech recognition*. in *EuroSpeech*. 1997.

19. Zhang, Y., Z.-J. Yan, and F.K. Soong. *Cross-validation based decision tree clustering for HMM-based TTS*. in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2010. IEEE.

20. Xie, F.-L., Y.-J. Wu, and F.K. Soong. *Cross validation and Minimum Generation Error for improved model clustering in HMM-based TTS*. in *Chinese Spoken Language Processing (ISCSLP)*. 2012. IEEE.

21. Wu, Y.-J. and R.-H. Wang. *Minimum generation error training for HMM-based speech synthesis*. in *IEEE International Conference on Acoustics, Speech and Signal Processing*. 2006. IEEE.

22. Wang, Y., et al. *Extended Decision Tree with OR Relationship for HMM-based Speech Synthesis*. in *Pattern Recognition (ACPR)*. 2013. IEEE.

23. Moore, M.D. and M.I. Savic, *Speech reconstruction using a generalized HSMM (GHSMM).* Digital Signal Processing, 2004. **14**(1): p. 37-53.

24. Zen, H., et al. *Hidden semi-Markov model based speech synthesis*. in *Interspeech*. 2004.

25. Rosti, A.I. and M. Gales, *Factor analysed hidden Markov models for speech recognition.* Computer Speech & Language, 2004. **18**(2): p. 181-200.

26. Huang, X., A. Acero, and H.-W. Hon, *Spoken Language Processing: A Guide to Theory, Algorithm, and System Development*. 2001: Prentice Hall PTR. 960.

27. Yu, K., et al. *From discontinuous to continuous F0 modelling in HMM-based speech synthesis*. in *SSW*. 2010.

28. Lyche, T. and L.L. Schumaker, *On the convergence of cubic interpolating splines*, in *Spline functions and approximation theory*. 1973, Springer. p. 169-189.

29. Garner, P.N., M. Cernak, and P. Motlicek, *A simple continuous pitch estimation algorithm.* IEEE Signal Processing Letters, 2013. **20**(1): p. 102-105.

30. Kawahara, H., I. Masuda-Katsuse, and A. De Cheveigne, *Restructuring speech representations using a pitch-adaptive time–frequency smoothing and an instantaneous-frequency-based F0 extraction: Possible role of a repetitive structure in sounds.* Speech communication, 1999. **27**(3): p. 187-207.

31. Drugman, T. and Y. Stylianou, *Maximum voiced frequency estimation: Exploiting amplitude and phase spectra.* IEEE Signal Processing Letters, 2014. **21**(10): p. 1230-1234.

32. Yu, K. and S. Young, *Continuous F0 modeling for HMM based statistical parametric speech synthesis.* IEEE Transactions on Audio, Speech, and Language Processing, 2011. **19**(5): p. 1071-1079.

33. Hashimoto, K., et al. *The effect of neural networks in statistical parametric speech synthesis*. in *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on*. 2015. IEEE.

34. Watts, O., et al. *From HMMs to DNNs: where do the improvements come from?* in *Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on*. 2016. IEEE.

35. Zen, H., A. Senior, and M. Schuster. *Statistical parametric speech synthesis using deep neural networks*. in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*. 2013. IEEE.

36. Zen, H. and A. Senior. *Deep mixture density networks for acoustic modeling in statistical parametric speech synthesis*. in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*. 2014. IEEE.

37. Qian, Y., et al. *On the training aspects of deep neural network (DNN) for parametric TTS synthesis*. in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*. 2014. IEEE.

38. Yu, K., F. Mairesse, and S. Young. *Word-level emphasis modelling in HMM-based speech synthesis*. in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*. 2010. IEEE.

39. Lu, H., S. King, and O. Watts. *Combining a Vector Space Representation of Linguistic Context with a Deep Neural Network for Text-To-Speech Synthesis*. in *Eighth ISCA Workshop on Speech Synthesis*. 2013.

40. Wu, Z., et al. *Deep neural networks employing multi-task learning and stacked bottleneck features for speech synthesis*. in *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on*. 2015. IEEE.

41. Wu, Z., O. Watts, and S. King, *Merlin: An open source neural network speech synthesis system.* Proc. SSW, Sunnyvale, USA, 2016.

42. Tokuda, K., et al. *Speech parameter generation algorithms for HMM-based speech synthesis*. in *IEEE International Conference on Acoustics, Speech and Signal Processing*. 2000. IEEE.

43. Sornlertlamvanich, V., T. Charoenporn, and H. Isahara, *ORCHID: Thai part-of-speech tagged corpus.*

44. Kingma, D. and J. Ba, *Adam: A method for stochastic optimization.* arXiv preprint arXiv:1412.6980, 2014.

45. Srivastava, N., et al., *Dropout: a simple way to prevent neural networks from overfitting.* Journal of machine learning research, 2014. **15**(1): p. 1929-1958.

46. Salvador, S. and P. Chan, *Toward accurate dynamic time warping in linear time and space.* Intelligent Data Analysis, 2007. **11**(5): p. 561-580.

47. Toda, T., A.W. Black, and K. Tokuda, *Voice conversion based on maximum-likelihood estimation of spectral parameter trajectory.* IEEE Transactions on Audio, Speech, and Language Processing, 2007. **15**(8): p. 2222-2235.

48. Wu, Z. and S. King, *Improving trajectory modelling for dnn-based speech synthesis by using stacked bottleneck features and minimum generation error training.* IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2016. **24**(7): p. 1255-1265.

49. Tamura, M., et al. *Speaker adaptation for HMM-based speech synthesis system using MLLR*. in *the third ESCA/COCOSDA Workshop (ETRW) on Speech Synthesis*. 1998.

50. Hashimoto, K., et al. *Trajectory training considering global variance for speech synthesis based on neural networks*. in *Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on*. 2016. IEEE.

ภาคผนวก

ภาคผนวก ก คำถามของต้นไม้ตัดสินใจ

ชื่อคำถาม	กลุ่มของหน่วยเสียงที่อยู่ในคำถาม
LLL-Consonants	{k<*, kh<*, ng<*, c<*, ch<*, s<*, j<*, d<*, t<*, th<*, n<*, b<*, p<*, ph<*, f<*, m<*, r<*, l<*, w<*, h<*, z<*, pr<*, phr<*, tr<*, kr<*, khr<*, pl<*, phl<*, thr<*, kl<*, khl<*, kw<*, khw<*, br<*, bl<*, fr<*, fl<*, dr<*, p^<*, t^<*, k^<*, n^<*, m^<*, ng^<*, j^<*, w^<*, f^<*, l^<*, s^<*, ch^<*}
LLL-InitialConsonants	{k<*, kh<*, ng<*, c<*, ch<*, s<*, j<*, d<*, t<*, th<*, n<*, b<*, p<*, ph<*, f<*, m<*, r<*, l<*, w<*, h<*, z<*, pr<*, phr<*, tr<*, kr<*, khr<*, pl<*, phl<*, thr<*, kl<*, khl<*, kw<*, khw<*, br<*, bl<*, fr<*, fl<*, dr<*}
LLL-FinalConsonants	{p^<*, t^<*, k^<*, n^<*, m^<*, ng^<*, j^<*, w^<*, f^<*, l^<*, s^<*, ch^<*}
LLL-SingleConsonants	{k<*, kh<*, ng<*, c<*, ch<*, s<*, j<*, d<*, t<*, th<*, n<*, b<*, p<*, ph<*, f<*, m<*, r<*, l<*, w<*, h<*, z<*}
LLL-DoubleConsonants	{pr<*, phr<*, tr<*, kr<*, khr<*, pl<*, phl<*, thr<*, kl<*, khl<*, kw<*, khw<*, br<*, bl<*, fr<*, fl<*, dr<*}
LLL-VoicelessUnaspirated	{p<*, t<*, c<*, k<*, z<*}
LLL-VoiceAspirated	{ph<*, th<*, ch<*, kh<*}
LLL-Voiced	{b<*, d<*, ng<*}
LLL-StopConsonants	{p<*, t<*, c<*, k<*, z<*, ph<*, th<*, ch<*, kh<*, b<*, d<*, ng<*}
LLL-NonStopConsonants	{m<*, n<*, h<*, f<*, s<*, r<*, l<*, w<*, j<*}
LLL-Nasal	{m<*, n<*, h<*}
LLL-Fricative	{f<*, s<*}
LLL-Trill	{r<*}
LLL-Lateral	{l<*}
LLL-Approximant	{w<*, j<*}
LLL-Vowels	{a<*, aa<*, i<*, ii<*, v<*, vv<*, u<*, uu<*, e<*, ee<*, x<*, xx<*, o<*, oo<*, @<*, @@<*, q<*, qq<*, ia<*, iia<*, va<*, vva<*, ua<*, uua<*}
LLL-SingleVowels	{a<*, aa<*, i<*, ii<*, v<*, vv<*, u<*, uu<*, e<*, ee<*, x<*, xx<*, o<*, oo<*, @<*, @@<*, q<*, qq<*}
LLL-CloseVowels	{i<*, ii<*, v<*, vv<*, u<*, uu<*}
LLL-MidVowels	{e<*, ee<*, q<*, qq<*, o<*, oo<*}
LLL-OpenVowels	{x<*, xx<*, a<*, aa<*, @<*, @@<*}
LLL-FrontVowels	{i<*, ii<*, e<*, ee<*, x<*, xx<*, ia<*, iia<*}
LLL-CentralVowels	{v<*, vv<*, q<*, qq<*, a<*, aa<*, va<*, vva<*}
LLL-BackVowels	{u<*, uu<*, o<*, oo<*, a<*, aa<*, ua<*, uua<*}
LLL-Diphthongs	{ia<*, iia<*, va<*, vva<*, ua<*, uua<*}
LLL-ShortVowels	{a<*, i<*, v<*, u<*, e<*, x<*, o<*, @<*, q<*, ia<*, va<*, ua<*}
LLL-LongVowels	{aa<*, ii<*, vv<*, uu<*, ee<*, xx<*, oo<*, @@<*, qq<*, iia<*, vva<*, uua<*}
LLL-aVowel	{a<*, aa<*}
LLL-iVowel	{i<*, ii<*}
LLL-vVowel	{v<*, vv<*}
LLL-uVowel	{u<*, uu<*}
LLL-xVowel	{x<*, xx<*}
LLL-oVowel	{o<*, oo<*}
LLL-@Vowel	{@<*, @@<*}
LLL-qVowel	{q<*, qq<*}
LLL-iaVowel	{ia<*, iia<*}
LLL-vaVowel	{va<*, vva<*}
LLL-uaVowel	{ua<*, uua<*}
LLL-Silience	{sil<*, pau<*, sp<*}
LLL-k	{k<*}
LLL-kh	{kh<*}
LLL-ng	{ng<*}
LLL-c	{c<*}
LLL-ch	{ch<*}
LLL-s	{s<*}

ชื่อคำถาม	กลุ่มของหน่วยเสียงที่อยู่ในคำถาม
LLL-j	{j<*}
LLL-d	{d<*}
LLL-t	{t<*}
LLL-th	{th<*}
LLL-n	{n<*}
LLL-b	{b<*}
LLL-p	{p<*}
LLL-ph	{ph<*}
LLL-f	{f<*}
LLL-m	{m<*}
LLL-r	{r<*}
LLL-l	{l<*}
LLL-w	{w<*}
LLL-h	{h<*}
LLL-z	{z<*}
LLL-pr	{pr<*}
LLL-tr	{tr<*}
LLL-kr	{kr<*}
LLL-khr	{khr<*}
LLL-pl	{pl<*}
LLL-thr	{thr<*}
LLL-khw	{khw<*}
LLL-dr	{dr<*}
LLL-p^	{p^<*}
LLL-t^	{t^<*}
LLL-k^	{k^<*}
LLL-n^	{n^<*}
LLL-m^	{m^<*}
LLL-ng^	{ng^<*}
LLL-j^	{j^<*}
LLL-w^	{w^<*}
LLL-f^	{f^<*}
LLL-l^	{l^<*}
LLL-s^	{s^<*}
LLL-ch^	{ch^<*}
LLL-a	{a<*}
LLL-aa	{aa<*}
LLL-i	{i<*}
LLL-ii	{ii<*}
LLL-v	{v<*}
LLL-vv	{vv<*}
LLL-uu	{uu<*}
LLL-ee	{ee<*}
LLL-x	{x<*}
LLL-xx	{xx<*}
LLL-oo	{oo<*}
LLL-@	{@<*}
LLL-@@	{@@<*}
LLL-qq	{qq<*}
LLL-vva	{vva<*}
LLL-uua	{uua<*}
LLL-sil	{sil<*}
LL-Consonants	{*<k_*, *<kh_*, *<ng_*, *<c_*, *<ch_*, *<s_*, *<j_*, *<d_*, *<t_*, *<th_*, *<n_*, *<b_*, *<p_*, *<ph_*, *<f_*, *<m_*, *<r_*, *<l_*, *<w_*, *<h_*, *<z_*, *<pr_*, *<phr_*, *<tr_*, *<kr_*, *<khr_*, *<pl_*, *<phl_*, *<thr_*, *<kl_*, *<khl_*, *<kw_*, *<khw_*, *<br_*, *<bl_*, *<fr_*, *<fl_*, *<dr_*, *<p^_*, *<t^_*, *<k^_*, *<n^_*, *<m^_*, *<ng^_*, *<j^_*, *<w^_*, *<f^_*, *<l^_*, *<s^_*, *<ch^_*}

ชื่อคำถาม	กลุ่มของหน่วยเสียงที่อยู่ในคำถาม
LL-InitialConsonants	{*<k_*, *<kh_*, *<ng_*, *<c_*, *<ch_*, *<s_*, *<j_*, *<d_*, *<t_*, *<th_*, *<n_*, *<b_*, *<p_*, *<ph_*, *<f_*, *<m_*, *<r_*, *<l_*, *<w_*, *<h_*, *<z_*, *<pr_*, *<phr_*, *<tr_*, *<kr_*, *<khr_*, *<pl_*, *<phl_*, *<thr_*, *<kl_*, *<khl_*, *<kw_*, *<khw_*, *<br_*, *<bl_*, *<fr_*, *<fl_*, *<dr_*}
LL-FinalConsonants	{*<p^_*, *<t^_*, *<k^_*, *<n^_*, *<m^_*, *<ng^_*, *<j^_*, *<w^_*, *<f^_*, *<l^_*, *<s^_*, *<ch^_*}
LL-SingleConsonants	{*<k_*, *<kh_*, *<ng_*, *<c_*, *<ch_*, *<s_*, *<j_*, *<d_*, *<t_*, *<th_*, *<n_*, *<b_*, *<p_*, *<ph_*, *<f_*, *<m_*, *<r_*, *<l_*, *<w_*, *<h_*, *<z_*}
LL-DoubleConsonants	{*<pr_*, *<phr_*, *<tr_*, *<kr_*, *<khr_*, *<pl_*, *<phl_*, *<thr_*, *<kl_*, *<khl_*, *<kw_*, *<khw_*, *<br_*, *<bl_*, *<fr_*, *<fl_*, *<dr_*}
LL-VoicelessUnaspirated	{*<p_*, *<t_*, *<c_*, *<k_*, *<z_*}
LL-VoiceAspirated	{ph_*, *<th_*, *<ch_*, *<kh_*}
LL-Voiced	{*<b_*, *<d_*, *<ng_*}
LL-StopConsonants	{*<p_*, *<t_*, *<c_*, *<k_*, *<z_*, *<ph_*, *<th_*, *<ch_*, *<kh_*, *<b_*, *<d_*, *<ng_*}
LL-NonStopConsonants	{*<m_*, *<n_*, *<h_*, *<f_*, *<s_*, *<r_*, *<l_*, *<w_*, *<j_*}
LL-Nasal	{*<m_*, *<n_*, *<h_*}
LL-Fricative	{*<f_*, *<s_*}
LL-Trill	{*<r_*}
LL-Lateral	{*<l_*}
LL-Approximant	{*<w_*, *<j_*}
LL-Vowels	{*<a_*, *<aa_*, *<i_*, *<ii_*, *<v_*, *<vv_*, *<u_*, *<uu_*, *<e_*, *<ee_*, *<x_*, *<xx_*, *<o_*, *<oo_*, *<@_*, *<@@_*, *<q_*, *<qq_*, *<ia_*, *<iia_*, *<va_*, *<vva_*, *<ua_*, *<uua_*}
LL-SingleVowels	{*<a_*, *<aa_*, *<i_*, *<ii_*, *<v_*, *<vv_*, *<u_*, *<uu_*, *<e_*, *<ee_*, *<x_*, *<xx_*, *<o_*, *<oo_*, *<@_*, *<@@_*, *<q_*, *<qq_*}
LL-CloseVowels	{*<i_*, *<ii_*, *<v_*, *<vv_*, *<u_*, *<uu_*}
LL-MidVowels	{*<e_*, *<ee_*, *<q_*, *<qq_*, *<o_*, *<oo_*}
LL-OpenVowels	{*<x_*, *<xx_*, *<a_*, *<aa_*, *<@_*, *<@@_*}
LL-FrontVowels	{*<i_*, *<ii_*, *<e_*, *<ee_*, *<x_*, *<xx_*, *<ia_*, *<iia_*}
LL-CentralVowels	{*<v_*, *<vv_*, *<q_*, *<qq_*, *<a_*, *<aa_*, *<va_*, *<vva_*}
LL-BackVowels	{*<u_*, *<uu_*, *<o_*, *<oo_*, *<a_*, *<aa_*, *<ua_*, *<uua_*}
LL-Diphthongs	{*<ia_*, *<iia_*, *<va_*, *<vva_*, *<ua_*, *<uua_*}
LL-ShortVowels	{*<a_*, *<i_*, *<v_*, *<u_*, *<e_*, *<x_*, *<o_*, *<@_*, *<q_*, *<ia_*, *<va_*, *<ua_*}
LL-LongVowels	{*<aa_*, *<ii_*, *<vv_*, *<uu_*, *<ee_*, *<xx_*, *<oo_*, *<@@_*, *<qq_*, *<iia_*, *<vva_*, *<uua_*}
LL-aVowel	{*<a_*, *<aa_*}
LL-iVowel	{*<i_*, *<ii_*}
LL-vVowel	{*<v_*, *<vv_*}
LL-uVowel	{*<u_*, *<uu_*}
LL-xVowel	{*<x_*, *<xx_*}
LL-oVowel	{*<o_*, *<oo_*}
LL-@Vowel	{*<@_*, *<@@_*}
LL-qVowel	{*<q_*, *<qq_*}
LL-iaVowel	{*<ia_*, *<iia_*}
LL-vaVowel	{*<va_*, *<vva_*}
LL-uaVowel	{*<ua_*, *<uua_*}
LL-Silience	{*<sil_*, *<pau_*, *<sp_*}
LL-k	{*<k_*}
LL-kh	{*<kh_*}
LL-ng	{*<ng_*}
LL-c	{*<c_*}
LL-ch	{*<ch_*}
LL-s	{*<s_*}
LL-j	{*<j_*}
LL-d	{*<d_*}
LL-t	{*<t_*}

ชื่อคำถาม	กลุ่มของหน่วยเสียงที่อยู่ในคำถาม
LL-th	{*<th_*}
LL-n	{*<n_*}
LL-b	{*<b_*}
LL-p	{*<p_*}
LL-ph	{*<ph_*}
LL-f	{*<f_*}
LL-m	{*<m_*}
LL-r	{*<r_*}
LL-l	{*<l_*}
LL-w	{*<w_*}
LL-h	{*<h_*}
LL-z	{*<z_*}
LL-pr	{*<pr_*}
LL-tr	{*<tr_*}
LL-kr	{*<kr_*}
LL-khr	{*<khr_*}
LL-pl	{*<pl_*}
LL-thr	{*<thr_*}
LL-khw	{*<khw_*}
LL-dr	{*<dr_*}
LL-p^	{*<p^_*}
LL-t^	{*<t^_*}
LL-k^	{*<k^_*}
LL-n^	{*<n^_*}
LL-m^	{*<m^_*}
LL-ng^	{*<ng^_*}
LL-j^	{*<j^_*}
LL-w^	{*<w^_*}
LL-f^	{*<f^_*}
LL-l^	{*<l^_*}
LL-s^	{*<s^_*}
LL-ch^	{*<ch^_*}
LL-a	{*<a_*}
LL-aa	{*<aa_*}
LL-i	{*<i_*}
LL-ii	{*<ii_*}
LL-v	{*<v_*}
LL-vv	{*<vv_*}
LL-uu	{*<uu_*}
LL-ee	{*<ee_*}
LL-x	{*<x_*}
LL-xx	{*<xx_*}
LL-oo	{*<oo_*}
LL-@	{*<@_*}
LL-@@	{*<@@_*}
LL-qq	{*<qq_*}
LL-vva	{*<vva_*}
LL-uua	{*<uua_*}
LL-sil	{*<sil_*}
L-SingleConsonants	{*_k-*, *_kh-*, *_ng-*, *_c-*, *_ch-*, *_s-*, *_j-*, *_d-*, *_t-*, *_th-*, *_n-*, *_b-*, *_p-*, *_ph-*, *_f-*, *_m-*, *_r-*, *_l-*, *_w-*, *_h-*, *_z-*}
L-DoubleConsonants	{*_pr-*, *_phr-*, *_tr-*, *_kr-*, *_khr-*, *_pl-*, *_phl-*, *_thr-*, *_kl-*, *_khl-*, *_kw-*, *_khw-*, *_br-*, *_bl-*, *_fr-*, *_fl-*, *_dr-*}
L-VoicelessUnaspirated	{*_p-*, *_t-*, *_c-*, *_k-*, *_z-*}
L-VoiceAspirated	{*_ph-*, *_th-*, *_ch-*, *_kh-*}
L-Voiced	{*_b-*, *_d-*, *_ng-*}
L-StopConsonants	{*_p-*, *_t-*, *_c-*, *_k-*, *_z-*, *_ph-*, *_th-*, *_ch-*, *_kh-*, *_b-*, *_d-*, *_ng-*}

ชื่อคำถาม	กลุ่มของหน่วยเสียงที่อยู่ในคำถาม
L-NonStopConsonants	{*_m-*, *_n-*, *_h-*, *_f-*, *_s-*, *_r-*, *_l-*, *_w-*, *_j-*}
L-Nasal	{*_m-*, *_n-*, *_h-*}
L-Fricative	{*_f-*, *_s-*}
L-Trill	{*_r-*}
L-Lateral	{*_l-*}
L-Approximant	{*_w-*, *_j-*}
L-Vowels	{*_a-*, *_aa-*, *_i-*, *_ii-*, *_v-*, *_vv-*, *_u-*, *_uu-*, *_e-*, *_ee-*, *_x-*, *_xx-*, *_o-*, *_oo-*, *_@-*, *_@@-*, *_q-*, *_qq-*, *_ia-*, *_iia-*, *_va-*, *_vva-*, *_ua-*, *_uua-*}
L-SingleVowels	{*_a-*, *_aa-*, *_i-*, *_ii-*, *_v-*, *_vv-*, *_u-*, *_uu-*, *_e-*, *_ee-*, *_x-*, *_xx-*, *_o-*, *_oo-*, *_@-*, *_@@-*, *_q-*, *_qq-*}
L-CloseVowels	{*_i-*, *_ii-*, *_v-*, *_vv-*, *_u-*, *_uu-*}
L-MidVowels	{*_e-*, *_ee-*, *_q-*, *_qq-*, *_o-*, *_oo-*}
L-OpenVowels	{*_x-*, *_xx-*, *_a-*, *_aa-*, *_@-*, *_@@-*}
L-FrontVowels	{*_i-*, *_ii-*, *_e-*, *_ee-*, *_x-*, *_xx-*, *_ia-*, *_iia-*}
L-CentralVowels	{*_v-*, *_vv-*, *_q-*, *_qq-*, *_a-*, *_aa-*, *_va-*, *_vva-*}
L-BackVowels	{*_u-*, *_uu-*, *_o-*, *_oo-*, *_a-*, *_aa-*, *_ua-*, *_uua-*}
L-Diphthongs	{*_ia-*, *_iia-*, *_va-*, *_vva-*, *_ua-*, *_uua-*}
L-ShortVowels	{*_a-*, *_i-*, *_v-*, *_u-*, *_e-*, *_x-*, *_o-*, *_@-*, *_q-*, *_ia-*, *_va-*, *_ua-*}
L-LongVowels	{*_aa-*, *_ii-*, *_vv-*, *_uu-*, *_ee-*, *_xx-*, *_oo-*, *_@@-*, *_qq-*, *_iia-*, *_vva-*, *_uua-*}
L-aVowel	{*_a-*, *_aa-*}
L-iVowel	{*_i-*, *_ii-*}
L-vVowel	{*_v-*, *_vv-*}
L-uVowel	{*_u-*, *_uu-*}
L-eVowel	{*_e-*, *_ee-*}
L-xVowel	{*_x-*, *_xx-*}
L-oVowel	{*_o-*, *_oo-*}
L-@Vowel	{*_@-*, *_@@-*}
L-iaVowel	{*_ia-*, *_iia-*}
L-vaVowel	{*_va-*, *_vva-*}
L-uaVowel	{*_ua-*, *_uua-*}
L-Silience	{*_sil-*, *_pau-*, *_sp-*}
L-k	{*_k-*}
L-kh	{*_kh-*}
L-ng	{*_ng-*}
L-c	{*_c-*}
L-ch	{*_ch-*}
L-s	{*_s-*}
L-j	{*_j-*}
L-d	{*_d-*}
L-t	{*_t-*}
L-th	{*_th-*}
L-n	{*_n-*}
L-b	{*_b-*}
L-p	{*_p-*}
L-ph	{*_ph-*}
L-f	{*_f-*}
L-m	{*_m-*}
L-r	{*_r-*}
L-l	{*_l-*}
L-w	{*_w-*}
L-h	{*_h-*}
L-z	{*_z-*}
L-pr	{*_pr-*}
L-phr	{*_phr-*}
L-tr	{*_tr-*}

ชื่อคำถาม	กลุ่มของหน่วยเสียงที่อยู่ในคำถาม
L-kr	{*_kr-*}
L-khr	{*_khr-*}
L-pl	{*_pl-*}
L-phl	{*_phl-*}
L-thr	{*_thr-*}
L-kl	{*_kl-*}
L-khl	{*_khl-*}
L-p^	{*_p^-*}
L-t^	{*_t^-*}
L-k^	{*_k^-*}
L-n^	{*_n^-*}
L-m^	{*_m^-*}
L-ng^	{*_ng^-*}
L-j^	{*_j^-*}
L-w^	{*_w^-*}
L-f^	{*_f^-*}
L-s^	{*_s^-*}
L-a	{*_a-*}
L-aa	{*_aa-*}
L-i	{*_i-*}
L-ii	{*_ii-*}
L-v	{*_v-*}
L-u	{*_u-*}
L-uu	{*_uu-*}
L-e	{*_e-*}
L-ee	{*_ee-*}
L-x	{*_x-*}
L-xx	{*_xx-*}
L-o	{*_o-*}
L-oo	{*_oo-*}
L-@	{*_@-*}
L-@@	{*_@@-*}
L-qq	{*_qq-*}
L-iia	{*_iia-*}
L-uua	{*_uua-*}
L-sil	{*_sil-*}
C-Consonants	{*-k+*, *-kh+*, *-ng+*, *-c+*, *-ch+*, *-s+*, *-j+*, *-d+*, *-t+*, *-th+*, *-n+*, *-b+*, *-p+*, *-ph+*, *-f+*, *-m+*, *-r+*, *-l+*, *-w+*, *-h+*, *-z+*, *-pr+*, *-phr+*, *-tr+*, *-kr+*, *-khr+*, *-pl+*, *-phl+*, *-thr+*, *-kl+*, *-khl+*, *-kw+*, *-khw+*, *-br+*, *-bl+*, *-fr+*, *-fl+*, *-dr+*, *-p^+*, *-t^+*, *-k^+*, *-n^+*, *-m^+*, *-ng^+*, *-j^+*, *-w^+*, *-f^+*, *-l^+*, *-s^+*, *-ch^+*}
C-FinalConsonants	{*-p^+*, *-t^+*, *-k^+*, *-n^+*, *-m^+*, *-ng^+*, *-j^+*, *-w^+*, *-f^+*, *-l^+*, *-s^+*, *-ch^+*}
C-SingleConsonants	{*-k+*, *-kh+*, *-ng+*, *-c+*, *-ch+*, *-s+*, *-j+*, *-d+*, *-t+*, *-th+*, *-n+*, *-b+*, *-p+*, *-ph+*, *-f+*, *-m+*, *-r+*, *-l+*, *-w+*, *-h+*, *-z+*}
C-DoubleConsonants	{*-pr+*, *-phr+*, *-tr+*, *-kr+*, *-khr+*, *-pl+*, *-phl+*, *-thr+*, *-kl+*, *-khl+*, *-kw+*, *-khw+*, *-br+*, *-bl+*, *-fr+*, *-fl+*, *-dr+*}
C-VoicelessUnaspirated	{*-p+*, *-t+*, *-c+*, *-k+*, *-z+*}
C-VoiceAspirated	{*-ph+*, *-th+*, *-ch+*, *-kh+*}
C-Voiced	{*-b+*, *-d+*, *-ng+*}
C-StopConsonants	{*-p+*, *-t+*, *-c+*, *-k+*, *-z+*, *-ph+*, *-th+*, *-ch+*, *-kh+*, *-b+*, *-d+*, *-ng+*}
C-NonStopConsonants	{*-m+*, *-n+*, *-h+*, *-f+*, *-s+*, *-r+*, *-l+*, *-w+*, *-j+*}
C-Nasal	{*-m+*, *-n+*, *-h+*}
C-Fricative	{*-f+*, *-s+*}
C-Trill	{*-r+*}
C-Lateral	{*-l+*}
C-Approximant	{*-w+*, *-j+*}

ชื่อคำถาม	กลุ่มของหน่วยเสียงที่อยู่ในคำถาม
C-Vowels	{*-a+*, *-aa+*, *-i+*, *-ii+*, *-v+*, *-vv+*, *-u+*, *-uu+*, *-e+*, *-ee+*, *-x+*, *-xx+*, *-o+*, *-oo+*, *-@+*, *-@@+*, *-q+*, *-qq+*, *-ia+*, *-iia+*, *-va+*, *-vva+*, *-ua+*, *-uua+*}
C-SingleVowels	{*-a+*, *-aa+*, *-i+*, *-ii+*, *-v+*, *-vv+*, *-u+*, *-uu+*, *-e+*, *-ee+*, *-x+*, *-xx+*, *-o+*, *-oo+*, *-@+*, *-@@+*, *-q+*, *-qq+*}
C-CloseVowels	{*-i+*, *-ii+*, *-v+*, *-vv+*, *-u+*, *-uu+*}
C-MidVowels	{*-e+*, *-ee+*, *-q+*, *-qq+*, *-o+*, *-oo+*}
C-OpenVowels	{*-x+*, *-xx+*, *-a+*, *-aa+*, *-@+*, *-@@+*}
C-FrontVowels	{*-i+*, *-ii+*, *-e+*, *-ee+*, *-x+*, *-xx+*, *-ia+*, *-iia+*}
C-CentralVowels	{*-v+*, *-vv+*, *-q+*, *-qq+*, *-a+*, *-aa+*, *-va+*, *-vva+*}
C-BackVowels	{*-u+*, *-uu+*, *-o+*, *-oo+*, *-a+*, *-aa+*, *-ua+*, *-uua+*}
C-Diphthongs	{*-ia+*, *-iia+*, *-va+*, *-vva+*, *-ua+*, *-uua+*}
C-ShortVowels	{*-a+*, *-i+*, *-v+*, *-u+*, *-e+*, *-x+*, *-o+*, *-@+*, *-q+*, *-ia+*, *-va+*, *-ua+*}
C-LongVowels	{*-aa+*, *-ii+*, *-vv+*, *-uu+*, *-ee+*, *-xx+*, *-oo+*, *-@@+*, *-qq+*, *-iia+*, *-vva+*, *-uua+*}
C-aVowel	{*-a+*, *-aa+*}
C-iVowel	{*-i+*, *-ii+*}
C-vVowel	{*-v+*, *-vv+*}
C-uVowel	{*-u+*, *-uu+*}
C-eVowel	{*-e+*, *-ee+*}
C-xVowel	{*-x+*, *-xx+*}
C-oVowel	{*-o+*, *-oo+*}
C-@Vowel	{*-@+*, *-@@+*}
C-qVowel	{*-q+*, *-qq+*}
C-iaVowel	{*-ia+*, *-iia+*}
C-vaVowel	{*-va+*, *-vva+*}
C-uaVowel	{*-ua+*, *-uua+*}
C-Silience	{*-sil+*, *-pau+*, *-sp+*}
C-k	{*-k+*}
C-ng	{*-ng+*}
C-c	{*-c+*}
C-j	{*-j+*}
C-d	{*-d+*}
C-t	{*-t+*}
C-th	{*-th+*}
C-n	{*-n+*}
C-b	{*-b+*}
C-p	{*-p+*}
C-m	{*-m+*}
C-r	{*-r+*}
C-l	{*-l+*}
C-w	{*-w+*}
C-h	{*-h+*}
C-z	{*-z+*}
C-pr	{*-pr+*}
C-phr	{*-phr+*}
C-kr	{*-kr+*}
C-khr	{*-khr+*}
C-phl	{*-phl+*}
C-thr	{*-thr+*}
C-kl	{*-kl+*}
C-khl	{*-khl+*}
C-khw	{*-khw+*}
C-bl	{*-bl+*}
C-dr	{*-dr+*}
C-p^	{*-p^+*}
C-t^	{*-t^+*}

ชื่อคำถาม	กลุ่มของหน่วยเสียงที่อยู่ในคำถาม
C-k^	{*-k^+*}
C-n^	{*-n^+*}
C-m^	{*-m^+*}
C-ng^	{*-ng^+*}
C-j^	{*-j^+*}
C-w^	{*-w^+*}
C-f^	{*-f^+*}
C-l^	{*-l^+*}
C-s^	{*-s^+*}
C-ch^	{*-ch^+*}
C-a	{*-a+*}
C-aa	{*-aa+*}
C-i	{*-i+*}
C-ii	{*-ii+*}
C-v	{*-v+*}
C-vv	{*-vv+*}
C-u	{*-u+*}
C-uu	{*-uu+*}
C-e	{*-e+*}
C-ee	{*-ee+*}
C-x	{*-x+*}
C-xx	{*-xx+*}
C-o	{*-o+*}
C-oo	{*-oo+*}
C-@	{*-@+*}
C-@@	{*-@@+*}
C-qq	{*-qq+*}
C-iia	{*-iia+*}
C-vva	{*-vva+*}
C-uua	{*-uua+*}
R-Consonants	{*+k=*, *+kh=*, *+ng=*, *+c=*, *+ch=*, *+s=*, *+j=*, *+d=*, *+t=*, *+th=*, *+n=*, *+b=*, *+p=*, *+ph=*, *+f=*, *+m=*, *+r=*, *+l=*, *+w=*, *+h=*, *+z=*, *+pr=*, *+phr=*, *+tr=*, *+kr=*, *+khr=*, *+pl=*, *+phl=*, *+thr=*, *+kl=*, *+khl=*, *+kw=*, *+khw=*, *+br=*, *+bl=*, *+fr=*, *+fl=*, *+dr=*, *+p^=*, *+t^=*, *+k^=*, *+n^=*, *+m^=*, *+ng^=*, *+j^=*, *+w^=*, *+f^=*, *+l^=*, *+s^=*, *+ch^=*}
R-InitialConsonants	{*+k=*, *+kh=*, *+ng=*, *+c=*, *+ch=*, *+s=*, *+j=*, *+d=*, *+t=*, *+th=*, *+n=*, *+b=*, *+p=*, *+ph=*, *+f=*, *+m=*, *+r=*, *+l=*, *+w=*, *+h=*, *+z=*, *+pr=*, *+phr=*, *+tr=*, *+kr=*, *+khr=*, *+pl=*, *+phl=*, *+thr=*, *+kl=*, *+khl=*, *+kw=*, *+khw=*, *+br=*, *+bl=*, *+fr=*, *+fl=*, *+dr=*}
R-FinalConsonants	{*+p^=*, *+t^=*, *+k^=*, *+n^=*, *+m^=*, *+ng^=*, *+j^=*, *+w^=*, *+f^=*, *+l^=*, *+s^=*, *+ch^=*}
R-SingleConsonants	{*+k=*, *+kh=*, *+ng=*, *+c=*, *+ch=*, *+s=*, *+j=*, *+d=*, *+t=*, *+th=*, *+n=*, *+b=*, *+p=*, *+ph=*, *+f=*, *+m=*, *+r=*, *+l=*, *+w=*, *+h=*, *+z=*}
R-DoubleConsonants	{*+pr=*, *+phr=*, *+tr=*, *+kr=*, *+khr=*, *+pl=*, *+phl=*, *+thr=*, *+kl=*, *+khl=*, *+kw=*, *+khw=*, *+br=*, *+bl=*, *+fr=*, *+fl=*, *+dr=*}
R-VoicelessUnaspirated	{*+p=*, *+t=*, *+c=*, *+k=*, *+z=*}
R-VoiceAspirated	{*+ph=*, *+th=*, *+ch=*, *+kh=*}
R-Voiced	{*+b=*, *+d=*, *+ng=*}
R-StopConsonants	{*+p=*, *+t=*, *+c=*, *+k=*, *+z=*, *+ph=*, *+th=*, *+ch=*, *+kh=*, *+b=*, *+d=*, *+ng=*}
R-NonStopConsonants	{*+m=*, *+n=*, *+h=*, *+f=*, *+s=*, *+r=*, *+l=*, *+w=*, *+j=*}
R-Nasal	{*+m=*, *+n=*, *+h=*}
R-Fricative	{*+f=*, *+s=*}
R-Trill	{*+r=*}
R-Lateral	{*+l=*}
R-Approximant	{*+w=*, *+j=*}

ชื่อคำถาม	กลุ่มของหน่วยเสียงที่อยู่ในคำถาม
R-Vowels	{*+a=*, *+aa=*, *+i=*, *+ii=*, *+v=*, *+vv=*, *+u=*, *+uu=*, *+e=*, *+ee=*, *+x=*, *+xx=*, *+o=*, *+oo=*, *+@=*, *+@@=*, *+q=*, *+qq=*, *+ia=*, *+iia=*, *+va=*, *+vva=*, *+ua=*, *+uua=*}
R-SingleVowels	{*+a=*, *+aa=*, *+i=*, *+ii=*, *+v=*, *+vv=*, *+u=*, *+uu=*, *+e=*, *+ee=*, *+x=*, *+xx=*, *+o=*, *+oo=*, *+@=*, *+@@=*, *+q=*, *+qq=*}
R-CloseVowels	{*+i=*, *+ii=*, *+v=*, *+vv=*, *+u=*, *+uu=*}
R-MidVowels	{*+e=*, *+ee=*, *+q=*, *+qq=*, *+o=*, *+oo=*}
R-OpenVowels	{*+x=*, *+xx=*, *+a=*, *+aa=*, *+@=*, *+@@=*}
R-CentralVowels	{*+v=*, *+vv=*, *+q=*, *+qq=*, *+a=*, *+aa=*, *+va=*, *+vva=*}
R-BackVowels	{*+u=*, *+uu=*, *+o=*, *+oo=*, *+a=*, *+aa=*, *+ua=*, *+uua=*}
R-Diphthongs	{*+ia=*, *+iia=*, *+va=*, *+vva=*, *+ua=*, *+uua=*}
R-ShortVowels	{*+a=*, *+i=*, *+v=*, *+u=*, *+e=*, *+x=*, *+o=*, *+@=*, *+q=*, *+ia=*, *+va=*, *+ua=*}
R-LongVowels	{*+aa=*, *+ii=*, *+vv=*, *+uu=*, *+ee=*, *+xx=*, *+oo=*, *+@@=*, *+qq=*, *+iia=*, *+vva=*, *+uua=*}
R-aVowel	{*+a=*, *+aa=*}
R-iVowel	{*+i=*, *+ii=*}
R-uVowel	{*+u=*, *+uu=*}
R-xVowel	{*+x=*, *+xx=*}
R-oVowel	{*+o=*, *+oo=*}
R-@Vowel	{*+@=*, *+@@=*}
R-uaVowel	{*+ua=*, *+uua=*}
R-Silience	{*+sil=*, *+pau=*, *+sp=*}
R-k	{*+k=*}
R-kh	{*+kh=*}
R-ng	{*+ng=*}
R-c	{*+c=*}
R-ch	{*+ch=*}
R-s	{*+s=*}
R-j	{*+j=*}
R-d	{*+d=*}
R-t	{*+t=*}
R-th	{*+th=*}
R-n	{*+n=*}
R-b	{*+b=*}
R-p	{*+p=*}
R-ph	{*+ph=*}
R-f	{*+f=*}
R-m	{*+m=*}
R-r	{*+r=*}
R-l	{*+l=*}
R-w	{*+w=*}
R-h	{*+h=*}
R-z	{*+z=*}
R-pr	{*+pr=*}
R-phr	{*+phr=*}
R-tr	{*+tr=*}
R-kr	{*+kr=*}
R-khr	{*+khr=*}
R-pl	{*+pl=*}
R-thr	{*+thr=*}
R-kl	{*+kl=*}
R-khl	{*+khl=*}
R-kw	{*+kw=*}
R-khw	{*+khw=*}
R-fr	{*+fr=*}
R-dr	{*+dr=*}
R-p^	{*+p^=*}

ชื่อคำถาม	กลุ่มของหน่วยเสียงที่อยู่ในคำถาม
R-t^	{*+t^=*}
R-k^	{*+k^=*}
R-n^	{*+n^=*}
R-m^	{*+m^=*}
R-ng^	{*+ng^=*}
R-j^	{*+j^=*}
R-w^	{*+w^=*}
R-f^	{*+f^=*}
R-s^	{*+s^=*}
R-a	{*+a=*}
R-aa	{*+aa=*}
R-i	{*+i=*}
R-ii	{*+ii=*}
R-vv	{*+vv=*}
R-u	{*+u=*}
R-uu	{*+uu=*}
R-e	{*+e=*}
R-ee	{*+ee=*}
R-x	{*+x=*}
R-xx	{*+xx=*}
R-o	{*+o=*}
R-@	{*+@=*}
R-@@	{*+@@=*}
R-vva	{*+vva=*}
R-sil	{*+sil=*}
RR-Consonants	{*=k>*, *=kh>*, *=ng>*, *=c>*, *=ch>*, *=s>*, *=j>*, *=d>*, *=t>*, *=th>*, *=n>*, *=b>*, *=p>*, *=ph>*, *=f>*, *=m>*, *=r>*, *=l>*, *=w>*, *=h>*, *=z>*, *=pr>*, *=phr>*, *=tr>*, *=kr>*, *=khr>*, *=pl>*, *=phl>*, *=thr>*, *=kl>*, *=khl>*, *=kw>*, *=khw>*, *=br>*, *=bl>*, *=fr>*, *=fl>*, *=dr>*, *=p^>*, *=t^>*, *=k^>*, *=n^>*, *=m^>*, *=ng^>*, *=j^>*, *=w^>*, *=f^>*, *=l^>*, *=s^>*, *=ch^>*}
RR-InitialConsonants	{*=k>*, *=kh>*, *=ng>*, *=c>*, *=ch>*, *=s>*, *=j>*, *=d>*, *=t>*, *=th>*, *=n>*, *=b>*, *=p>*, *=ph>*, *=f>*, *=m>*, *=r>*, *=l>*, *=w>*, *=h>*, *=z>*, *=pr>*, *=phr>*, *=tr>*, *=kr>*, *=khr>*, *=pl>*, *=phl>*, *=thr>*, *=kl>*, *=khl>*, *=kw>*, *=khw>*, *=br>*, *=bl>*, *=fr>*, *=fl>*, *=dr>*}
RR-FinalConsonants	{*=p^>*, *=t^>*, *=k^>*, *=n^>*, *=m^>*, *=ng^>*, *=j^>*, *=w^>*, *=f^>*, *=l^>*, *=s^>*, *=ch^>*}
RR-SingleConsonants	{*=k>*, *=kh>*, *=ng>*, *=c>*, *=ch>*, *=s>*, *=j>*, *=d>*, *=t>*, *=th>*, *=n>*, *=b>*, *=p>*, *=ph>*, *=f>*, *=m>*, *=r>*, *=l>*, *=w>*, *=h>*, *=z>*}
RR-DoubleConsonants	{*=pr>*, *=phr>*, *=tr>*, *=kr>*, *=khr>*, *=pl>*, *=phl>*, *=thr>*, *=kl>*, *=khl>*, *=kw>*, *=khw>*, *=br>*, *=bl>*, *=fr>*, *=fl>*, *=dr>*}
RR-VoicelessUnaspirated	{*=p>*, *=t>*, *=c>*, *=k>*, *=z>*}
RR-VoiceAspirated	{*=ph>*, *=th>*, *=ch>*, *=kh>*}
RR-Voiced	{*=b>*, *=d>*, *=ng>*}
RR-StopConsonants	{*=p>*, *=t>*, *=c>*, *=k>*, *=z>*, *=ph>*, *=th>*, *=ch>*, *=kh>*, *=b>*, *=d>*, *=ng>*}
RR-NonStopConsonants	{*=m>*, *=n>*, *=h>*, *=f>*, *=s>*, *=r>*, *=l>*, *=w>*, *=j>*}
RR-Nasal	{*=m>*, *=n>*, *=h>*}
RR-Fricative	{*=f>*, *=s>*}
RR-Trill	{*=r>*}
RR-Lateral	{*=l>*}
RR-Approximant	{*=w>*, *=j>*}
RR-Vowels	{*=a>*, *=aa>*, *=i>*, *=ii>*, *=v>*, *=vv>*, *=u>*, *=uu>*, *=e>*, *=ee>*, *=x>*, *=xx>*, *=o>*, *=oo>*, *=@>*, *=@@>*, *=q>*, *=qq>*, *=ia>*, *=iia>*, *=va>*, *=vva>*, *=ua>*, *=uua>*}
RR-SingleVowels	{*=a>*, *=aa>*, *=i>*, *=ii>*, *=v>*, *=vv>*, *=u>*, *=uu>*, *=e>*, *=ee>*, *=x>*, *=xx>*, *=o>*, *=oo>*, *=@>*, *=@@>*, *=q>*, *=qq>*}

ชื่อคำถาม	กลุ่มของหน่วยเสียงที่อยู่ในคำถาม
RR-CloseVowels	{*=i>*, *=ii>*, *=v>*, *=vv>*, *=u>*, *=uu>*}
RR-MidVowels	{*=e>*, *=ee>*, *=q>*, *=qq>*, *=o>*, *=oo>*}
RR-OpenVowels	{*=x>*, *=xx>*, *=a>*, *=aa>*, *=@>*, *=@@>*}
RR-FrontVowels	{*=i>*, *=ii>*, *=e>*, *=ee>*, *=x>*, *=xx>*, *=ia>*, *=iia>*}
RR-CentralVowels	{*=v>*, *=vv>*, *=q>*, *=qq>*, *=a>*, *=aa>*, *=va>*, *=vva>*}
RR-BackVowels	{*=u>*, *=uu>*, *=o>*, *=oo>*, *=a>*, *=aa>*, *=ua>*, *=uua>*}
RR-Diphthongs	{*=ia>*, *=iia>*, *=va>*, *=vva>*, *=ua>*, *=uua>*}
RR-ShortVowels	{*=a>*, *=i>*, *=v>*, *=u>*, *=e>*, *=x>*, *=o>*, *=@>*, *=q>*, *=ia>*, *=va>*, *=ua>*}
RR-LongVowels	{*=aa>*, *=ii>*, *=vv>*, *=uu>*, *=ee>*, *=xx>*, *=oo>*, *=@@>*, *=qq>*, *=iia>*, *=vva>*, *=uua>*}
RR-aVowel	{*=a>*, *=aa>*}
RR-iVowel	{*=i>*, *=ii>*}
RR-vVowel	{*=v>*, *=vv>*}
RR-uVowel	{*=u>*, *=uu>*}
RR-eVowel	{*=e>*, *=ee>*}
RR-xVowel	{*=x>*, *=xx>*}
RR-oVowel	{*=o>*, *=oo>*}
RR-@Vowel	{*=@>*, *=@@>*}
RR-qVowel	{*=q>*, *=qq>*}
RR-iaVowel	{*=ia>*, *=iia>*}
RR-vaVowel	{*=va>*, *=vva>*}
RR-uaVowel	{*=ua>*, *=uua>*}
RR-Silience	{*=sil>*, *=pau>*, *=sp>*}
RR-k	{*=k>*}
RR-kh	{*=kh>*}
RR-ng	{*=ng>*}
RR-c	{*=c>*}
RR-ch	{*=ch>*}
RR-s	{*=s>*}
RR-j	{*=j>*}
RR-d	{*=d>*}
RR-t	{*=t>*}
RR-th	{*=th>*}
RR-n	{*=n>*}
RR-b	{*=b>*}
RR-p	{*=p>*}
RR-ph	{*=ph>*}
RR-f	{*=f>*}
RR-m	{*=m>*}
RR-r	{*=r>*}
RR-l	{*=l>*}
RR-w	{*=w>*}
RR-h	{*=h>*}
RR-z	{*=z>*}
RR-pr	{*=pr>*}
RR-phr	{*=phr>*}
RR-kr	{*=kr>*}
RR-khr	{*=khr>*}
RR-thr	{*=thr>*}
RR-kl	{*=kl>*}
RR-khl	{*=khl>*}
RR-kw	{*=kw>*}
RR-khw	{*=khw>*}
RR-t^	{*=t^>*}
RR-k^	{*=k^>*}
RR-n^	{*=n^>*}
RR-m^	{*=m^>*}

ชื่อคำถาม	กลุ่มของหน่วยเสียงที่อยู่ในคำถาม
RR-ng^	{*=ng^>*}
RR-j^	{*=j^>*}
RR-w^	{*=w^>*}
RR-s^	{*=s^>*}
RR-ch^	{*=ch^>*}
RR-a	{*=a>*}
RR-aa	{*=aa>*}
RR-i	{*=i>*}
RR-ii	{*=ii>*}
RR-v	{*=v>*}
RR-vv	{*=vv>*}
RR-u	{*=u>*}
RR-uu	{*=uu>*}
RR-e	{*=e>*}
RR-ee	{*=ee>*}
RR-x	{*=x>*}
RR-xx	{*=xx>*}
RR-o	{*=o>*}
RR-oo	{*=oo>*}
RR-@	{*=@>*}
RR-@@	{*=@@>*}
RR-q	{*=q>*}
RR-qq	{*=qq>*}
RR-iia	{*=iia>*}
RR-vva	{*=vva>*}
RR-uua	{*=uua>*}
RR-sil	{*=sil>*}
RRR-Consonants	{*>k/A:*, *>kh/A:*, *>ng/A:*, *>c/A:*, *>ch/A:*, *>s/A:*, *>j/A:*, *>d/A:*, *>t/A:*, *>th/A:*, *>n/A:*, *>b/A:*, *>p/A:*, *>ph/A:*, *>f/A:*, *>m/A:*, *>r/A:*, *>l/A:*, *>w/A:*, *>h/A:*, *>z/A:*, *>pr/A:*, *>phr/A:*, *>tr/A:*, *>kr/A:*, *>khr/A:*, *>pl/A:*, *>phl/A:*, *>thr/A:*, *>kl/A:*, *>khl/A:*, *>kw/A:*, *>khw/A:*, *>br/A:*, *>bl/A:*, *>fr/A:*, *>fl/A:*, *>dr/A:*, *>p^/A:*, *>t^/A:*, *>k^/A:*, *>n^/A:*, *>m^/A:*, *>ng^/A:*, *>j^/A:*, *>w^/A:*, *>f^/A:*, *>l^/A:*, *>s^/A:*, *>ch^/A:*}
RRR-InitialConsonants	{*>k/A:*, *>kh/A:*, *>ng/A:*, *>c/A:*, *>ch/A:*, *>s/A:*, *>j/A:*, *>d/A:*, *>t/A:*, *>th/A:*, *>n/A:*, *>b/A:*, *>p/A:*, *>ph/A:*, *>f/A:*, *>m/A:*, *>r/A:*, *>l/A:*, *>w/A:*, *>h/A:*, *>z/A:*, *>pr/A:*, *>phr/A:*, *>tr/A:*, *>kr/A:*, *>khr/A:*, *>pl/A:*, *>phl/A:*, *>thr/A:*, *>kl/A:*, *>khl/A:*, *>kw/A:*, *>khw/A:*, *>br/A:*, *>bl/A:*, *>fr/A:*, *>fl/A:*, *>dr/A:*}
RRR-FinalConsonants	{*>p^/A:*, *>t^/A:*, *>k^/A:*, *>n^/A:*, *>m^/A:*, *>ng^/A:*, *>j^/A:*, *>w^/A:*, *>f^/A:*, *>l^/A:*, *>s^/A:*, *>ch^/A:*}
RRR-SingleConsonants	{*>k/A:*, *>kh/A:*, *>ng/A:*, *>c/A:*, *>ch/A:*, *>s/A:*, *>j/A:*, *>d/A:*, *>t/A:*, *>th/A:*, *>n/A:*, *>b/A:*, *>p/A:*, *>ph/A:*, *>f/A:*, *>m/A:*, *>r/A:*, *>l/A:*, *>w/A:*, *>h/A:*, *>z/A:*}
RRR-DoubleConsonants	{*>pr/A:*, *>phr/A:*, *>tr/A:*, *>kr/A:*, *>khr/A:*, *>pl/A:*, *>phl/A:*, *>thr/A:*, *>kl/A:*, *>khl/A:*, *>kw/A:*, *>khw/A:*, *>br/A:*, *>bl/A:*, *>fr/A:*, *>fl/A:*, *>dr/A:*}
RRR-VoicelessUnaspirated	{*>p/A:*, *>t/A:*, *>c/A:*, *>k/A:*, *>z/A:*}
RRR-VoiceAspirated	{*>ph/A:*, *>th/A:*, *>ch/A:*, *>kh/A:*}
RRR-Voiced	{*>b/A:*, *>d/A:*, *>ng/A:*}
RRR-StopConsonants	{*>p/A:*, *>t/A:*, *>c/A:*, *>k/A:*, *>z/A:*, *>ph/A:*, *>th/A:*, *>ch/A:*, *>kh/A:*, *>b/A:*, *>d/A:*, *>ng/A:*}
RRR-NonStopConsonants	{*>m/A:*, *>n/A:*, *>h/A:*, *>f/A:*, *>s/A:*, *>r/A:*, *>l/A:*, *>w/A:*, *>j/A:*}
RRR-Nasal	{*>m/A:*, *>n/A:*, *>h/A:*}
RRR-Fricative	{*>f/A:*, *>s/A:*}
RRR-Trill	{*>r/A:*}
RRR-Lateral	{*>l/A:*}
RRR-Approximant	{*>w/A:*, *>j/A:*}

ชื่อคำถาม	กลุ่มของหน่วยเสียงที่อยู่ในคำถาม
RRR-Vowels	{*>a/A:*, *>aa/A:*, *>i/A:*, *>ii/A:*, *>v/A:*, *>vv/A:*, *>u/A:*, *>uu/A:*, *>e/A:*, *>ee/A:*, *>x/A:*, *>xx/A:*, *>o/A:*, *>oo/A:*, *>@/A:*, *>@@/A:*, *>q/A:*, *>qq/A:*, *>ia/A:*, *>iia/A:*, *>va/A:*, *>vva/A:*, *>ua/A:*, *>uua/A:*}
RRR-SingleVowels	{*>a/A:*, *>aa/A:*, *>i/A:*, *>ii/A:*, *>v/A:*, *>vv/A:*, *>u/A:*, *>uu/A:*, *>e/A:*, *>ee/A:*, *>x/A:*, *>xx/A:*, *>o/A:*, *>oo/A:*, *>@/A:*, *>@@/A:*, *>q/A:*, *>qq/A:*}
RRR-CloseVowels	{*>i/A:*, *>ii/A:*, *>v/A:*, *>vv/A:*, *>u/A:*, *>uu/A:*}
RRR-MidVowels	{*>e/A:*, *>ee/A:*, *>q/A:*, *>qq/A:*, *>o/A:*, *>oo/A:*}
RRR-OpenVowels	{*>x/A:*, *>xx/A:*, *>a/A:*, *>aa/A:*, *>@/A:*, *>@@/A:*}
RRR-FrontVowels	{*>i/A:*, *>ii/A:*, *>e/A:*, *>ee/A:*, *>x/A:*, *>xx/A:*, *>ia/A:*, *>iia/A:*}
RRR-CentralVowels	{*>v/A:*, *>vv/A:*, *>q/A:*, *>qq/A:*, *>a/A:*, *>aa/A:*, *>va/A:*, *>vva/A:*}
RRR-BackVowels	{*>u/A:*, *>uu/A:*, *>o/A:*, *>oo/A:*, *>a/A:*, *>aa/A:*, *>ua/A:*, *>uua/A:*}
RRR-Diphthongs	{*>ia/A:*, *>iia/A:*, *>va/A:*, *>vva/A:*, *>ua/A:*, *>uua/A:*}
RRR-ShortVowels	{*>a/A:*, *>i/A:*, *>v/A:*, *>u/A:*, *>e/A:*, *>x/A:*, *>o/A:*, *>@/A:*, *>q/A:*, *>ia/A:*, *>va/A:*, *>ua/A:*}
RRR-LongVowels	{*>aa/A:*, *>ii/A:*, *>vv/A:*, *>uu/A:*, *>ee/A:*, *>xx/A:*, *>oo/A:*, *>@@/A:*, *>qq/A:*, *>iia/A:*, *>vva/A:*, *>uua/A:*}
RRR-aVowel	{*>a/A:*, *>aa/A:*}
RRR-iVowel	{*>i/A:*, *>ii/A:*}
RRR-vVowel	{*>v/A:*, *>vv/A:*}
RRR-uVowel	{*>u/A:*, *>uu/A:*}
RRR-eVowel	{*>e/A:*, *>ee/A:*}
RRR-xVowel	{*>x/A:*, *>xx/A:*}
RRR-oVowel	{*>o/A:*, *>oo/A:*}
RRR-@Vowel	{*>@/A:*, *>@@/A:*}
RRR-qVowel	{*>q/A:*, *>qq/A:*}
RRR-iaVowel	{*>ia/A:*, *>iia/A:*}
RRR-vaVowel	{*>va/A:*, *>vva/A:*}
RRR-uaVowel	{*>ua/A:*, *>uua/A:*}
RRR-Silience	{*>sil/A:*, *>pau/A:*, *>sp/A:*}
RRR-k	{*>k/A:*}
RRR-kh	{*>kh/A:*}
RRR-ng	{*>ng/A:*}
RRR-c	{*>c/A:*}
RRR-ch	{*>ch/A:*}
RRR-s	{*>s/A:*}
RRR-j	{*>j/A:*}
RRR-d	{*>d/A:*}
RRR-t	{*>t/A:*}
RRR-th	{*>th/A:*}
RRR-n	{*>n/A:*}
RRR-b	{*>b/A:*}
RRR-p	{*>p/A:*}
RRR-ph	{*>ph/A:*}
RRR-f	{*>f/A:*}
RRR-m	{*>m/A:*}
RRR-r	{*>r/A:*}
RRR-l	{*>l/A:*}
RRR-w	{*>w/A:*}
RRR-h	{*>h/A:*}
RRR-z	{*>z/A:*}
RRR-pr	{*>pr/A:*}
RRR-phr	{*>phr/A:*}
RRR-kr	{*>kr/A:*}
RRR-khr	{*>khr/A:*}
RRR-thr	{*>thr/A:*}

ชื่อคำถาม	กลุ่มของหน่วยเสียงที่อยู่ในคำถาม
RRR-kl	{*>kl/A:*}
RRR-khl	{*>khl/A:*}
RRR-kw	{*>kw/A:*}
RRR-khw	{*>khw/A:*}
RRR-t^	{*>t^/A:*}
RRR-k^	{*>k^/A:*}
RRR-n^	{*>n^/A:*}
RRR-m^	{*>m^/A:*}
RRR-ng^	{*>ng^/A:*}
RRR-j^	{*>j^/A:*}
RRR-w^	{*>w^/A:*}
RRR-s^	{*>s^/A:*}
RRR-ch^	{*>ch^/A:*}
RRR-a	{*>a/A:*}
RRR-aa	{*>aa/A:*}
RRR-i	{*>i/A:*}
RRR-ii	{*>ii/A:*}
RRR-v	{*>v/A:*}
RRR-vv	{*>vv/A:*}
RRR-u	{*>u/A:*}
RRR-uu	{*>uu/A:*}
RRR-e	{*>e/A:*}
RRR-ee	{*>ee/A:*}
RRR-x	{*>x/A:*}
RRR-xx	{*>xx/A:*}
RRR-o	{*>o/A:*}
RRR-oo	{*>oo/A:*}
RRR-@	{*>@/A:*}
RRR-@@	{*>@@/A:*}
RRR-q	{*>q/A:*}
RRR-qq	{*>qq/A:*}
RRR-iia	{*>iia/A:*}
RRR-vva	{*>vva/A:*}
RRR-uua	{*>uua/A:*}
RRR-sil	{*>sil/A:*}
Syllable_A_Tone_0	{*0-0+0*}
Syllable_A_Tone_1	{*1-1+1*}
Syllable_A_Tone_2	{*2-2+2*}
Syllable_A_Tone_3	{*3-3+3*}
Syllable_A_Tone_4	{*4-4+4*}
Syllable_AL_Tone_0	{*0-0*}
Syllable_AL_Tone_1	{*1-1*}
Syllable_AL_Tone_2	{*2-2*}
Syllable_AL_Tone_3	{*3-3*}
Syllable_AL_Tone_4	{*4-4*}
Syllable_ALL_Tone_0	{*0_0-0*}
Syllable_ALL_Tone_1	{*1_1-1*}
Syllable_ALL_Tone_2	{*2_2-2*}
Syllable_ALL_Tone_3	{*3_3-3*}
Syllable_ALL_Tone_4	{*4_4-4*}
Syllable_AR_Tone_0	{*0+0*}
Syllable_AR_Tone_1	{*1+1*}
Syllable_AR_Tone_2	{*2+2*}
Syllable_AR_Tone_3	{*3+3*}
Syllable_AR_Tone_4	{*4+4*}
Syllable_ARR_Tone_0	{*0+0=0*}
Syllable_ARR_Tone_1	{*1+1=1*}
Syllable_ARR_Tone_2	{*2+2=2*}

ชื่อคำถาม	กลุ่มของหน่วยเสียงที่อยู่ในคำถาม
Syllable_ARR_Tone_3	{*3+3=3*}
Syllable_ARR_Tone_4	{*4+4=4*}
Syllable_LL_Tone_sp	{*<st_*}
Syllable_LL_Tone_0	{*<0_*}
Syllable_LL_Tone_1	{*<1_*}
Syllable_LL_Tone_2	{*<2_*}
Syllable_LL_Tone_3	{*<3_*}
Syllable_LL_Tone_4	{*<4_*}
Syllable_LLL_Tone_0	{*^/A:0<*}
Syllable_LLL_Tone_1	{*^/A:1<*}
Syllable_LLL_Tone_2	{*^/A:2<*}
Syllable_LLL_Tone_3	{*^/A:3<*}
Syllable_LLL_Tone_4	{*^/A:4<*}
Syllable_L_Tone_sp	{*_st-*}
Syllable_L_Tone_0	{*_0-*}
Syllable_L_Tone_1	{*_1-*}
Syllable_L_Tone_2	{*_2-*}
Syllable_L_Tone_3	{*_3-*}
Syllable_L_Tone_4	{*_4-*}
Syllable_M_Tone_sp	{*-st+*}
Syllable_M_Tone_0	{*-0+*}
Syllable_M_Tone_1	{*-1+*}
Syllable_M_Tone_2	{*-2+*}
Syllable_M_Tone_3	{*-3+*}
Syllable_M_Tone_4	{*-4+*}
Syllable_R_Tone_sp	{*+st=*}
Syllable_R_Tone_0	{*+0=*}
Syllable_R_Tone_1	{*+1=*}
Syllable_R_Tone_2	{*+2=*}
Syllable_R_Tone_3	{*+3=*}
Syllable_R_Tone_4	{*+4=*}
Syllable_RR_Tone_sp	{*=st>*}
Syllable_RR_Tone_0	{*=0>*}
Syllable_RR_Tone_1	{*=1>*}
Syllable_RR_Tone_2	{*=2>*}
Syllable_RR_Tone_3	{*=3>*}
Syllable_RR_Tone_4	{*=4>*}
Syllable_RRR_Tone_0	{*>0}
Syllable_RRR_Tone_1	{*>1}
Syllable_RRR_Tone_2	{*>2}
Syllable_RRR_Tone_3	{*>3}
Syllable_RRR_Tone_4	{*>4}

ประวัติผู้เขียนวิทยานิพนธ์

นายศุภเดช ฉันจรัสวิชัย เกิดเมื่อวันที่ 22 กันยายน 2532 ที่จังหวัดกรุงเทพฯ สำเร็จการศึกษาระดับปริญญาวิศวกรรมศาสตรบัณฑิต (เกียรติยมอันดับ 1) สาขาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยมหิดล เมื่อปีการศึกษา 2550 จากนั้นสำเร็จการศึกษาระดับปริญญาวิศวกรรมศาสตรมหาบัณฑิต สาขาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย เมื่อปีการศึกษา 2553 และเข้าศึกษาในหลักสูตรปริญญาวิศวกรรมศาสตรดุษฎีบัณฑิต สาขาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย เมื่อปีการศึกษา 2555