การสังเคราะห์เสียงอิงแบบจำลองฮิดเดนมาร์คอฟที่สามารถกำหนดสัญญาณจากเส้นเสียงและสัญญาณรบกวนลมหายใจ

นายนิพนธ์ ชินะธิมาตร์มงคล

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต
วิศวกรรมคอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2552

HMM-BASED SPEECH SYNTHESIS WITH GLOTTAL SOURCE AND ASPIRATION NOISE MODELING

Mr. Nipon Chinathimatmongkhon

A Thesis Submitted in Partial Fulfillment of the Requirements

for the Degree of Master of Engineering Program in Computer Engineering

Department of Computer Engineering

Faculty of Engineering

Chulalongkorn University

Academic Year 2009

หัวข้อวิทยานิพนธ์	การสังเคราะห์เสียงอิงแบบจำลองฮิดเดนมาร์คอฟที่สามารถกำหนดสัญญาณจากเส้นเสียง และสัญญาณรบกวนลมหายใจ
โดย	นายนิพนธ์ ชินะธิมาตร์มงคล
สาขาวิชา	วิศวกรรมคอมพิวเตอร์
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	ผู้ช่วยศาสตราจารย์ ดร.อติวงศ์ สุชาโต
อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม	ผู้ช่วยศาสตราจารย์ ดร. โปรดปราน บุณยพุกกณะ

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

..... คณบดีคณะวิศวกรรมศาสตร์
(รองศาสตราจารย์ ดร.บุญสม เลิศหิรัญวงศ์)

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ
(ศาสตราจารย์ ดร.ประภาส จงสถิตย์วัฒนา)

..... อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
(ผู้ช่วยศาสตราจารย์ ดร.อติวงศ์ สุชาโต)

..... อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม
(ผู้ช่วยศาสตราจารย์ ดร. โปรดปราน บุณยพุกกณะ)

..... กรรมการ
(ศาสตราจารย์ ดร.บุญเสริม กิจศิริกุล)

..... กรรมการ
(ดร.ชัย วุฒิวิวัฒน์ชัย)

นิพนธ์ ชินะธิมาตร์มงคล : การสังเคราะห์เสียงอิงแบบจำลองฮิดเดนมาร์คอฟที่สามารถกำหนดสัญญาณจากเส้นเสียงและสัญญาณรบกวนลมหายใจ. (HMM-BASED SPEECH SYNTHESIS WITH GLOTTAL SOURCE AND ASPIRATION NOISE MODELING) อ. ที่ปรึกษาวิทยานิพนธ์หลัก : ผศ. ดร.อติวงศ์ สุชาโต, อ.ที่ปรึกษาวิทยานิพนธ์ร่วม : ผศ. ดร. โปรดปราน บุณยพุกกณะ, 86 หน้า.

วิทยานิพนธ์นี้เสนอระบบการสังเคราะห์เสียงซึ่งดัดแปลงการสังเคราะห์เสียงที่อิงแบบจำลองฮิดเดนมาร์คอฟให้สามารถรองรับการกำหนดลักษณะของสัญญาณแหล่งกำเนิดจากเส้นเสียง และเสียงรบกวนลมหายใจได้โดยตรง ทำให้สามารถสร้างเสียงสังเคราะห์เพื่อเลียนแบบลักษณะเสียงมนุษย์ชนิดต่าง ๆ ได้โดยมิต้องทำการประมวลผลสัญญาณที่ได้จากการสังเคราะห์ในภายหลัง เพื่อสร้างแบบจำลองเสียงจากแหล่งกำเนิดเส้นเสียง วิทยานิพนธ์นี้ได้วิเคราะห์ค่าพารามิเตอร์แอลเอฟแบบแปลง ซึ่งใช้เป็นแบบจำลองของแหล่งกำเนิดจากเส้นเสียง และศึกษาการหาค่าระดับเสียงรบกวนลมหายใจโดยใช้วิธีการลดสัญญาณรบกวนในระบบโดยเวฟเลท ซึ่งเป็นวิธีการสร้างสัญญาณใหม่จากสัญญาณที่ถูกรบกวน เพื่อสกัดสัญญาณรบกวนในสัญญาณเส้นเสียงได้ นอกจากนี้วิทยานิพนธ์นี้ได้เสนอวิธีหาฟังก์ชันการหาจุดเปลี่ยนแบบใหม่ในขั้นตอนการลดสัญญาณรบกวนในระบบโดยเวฟเลท

จากการวัดความเป็นธรรมชาติของเสียงสังเคราะห์ คะแนนความเป็นธรรมชาติของระบบที่นำเสนอไม่แตกต่างกับระบบอ้างอิง ชี้ให้เห็นว่าเสียงสังเคราะห์จากการใช้แหล่งกำเนิดจากเส้นเสียงเป็นอินพุทของแบบจำลองสามารถเทียบเคียงได้กับการใช้ขบวนอิมพัลส์เป็นอินพุทดังปรากฏในการสังเคราะห์เสียงด้วยแบบจำลองฮิดเดนมาร์คอฟทั่ว ๆ ไป นอกจากนี้เสียงที่ได้จากระบบสังเคราะห์เสียงที่นำเสนอของวิทยานิพนธ์นี้สามารถเลียนแบบเสียงลมหายใจ (Breathy) และเสียงบีบ (Creaky) ได้ดีกว่าระบบอ้างอิง และสามารถสังเคราะห์ลักษณะเสียงที่มีระดับความแตกต่างหลากหลายกว่าระบบอ้างอิง

ภาควิชา..... วิศวกรรมคอมพิวเตอร์.....	ลายมือชื่อนิสิต .....
สาขาวิชา..... วิศวกรรมคอมพิวเตอร์.....	ลายมือชื่อ อ. ที่ปรึกษาวิทยานิพนธ์หลัก .....
ปีการศึกษา.. 2552.....	ลายมือชื่อ อ. ที่ปรึกษาวิทยานิพนธ์ร่วม .....

497 03918 21 : MAJOR COMPUTER ENGINEERING

KEYWORDS : HMM-BASED SPEECH SYNTHESIS/, VOICE QUALITY/, ASPIRATION NOISE/, GLOTTAL SOURCE/, ACOUSTIC-PHONETICS

NIPON CHINATHIMATMONGKHON : HMM-BASED SPEECH SYNTHESIS WITH GLOTTAL SOURCE AND ASPIRATION NOISE MODELING. THESIS ADVISOR : ATIWONG SUCHATO, Ph.D., THESIS COADVISOR : PROADPRAN PUNYABUKKANA, Ph.D., 86 pp.

This thesis proposes a modified HMM-based speech synthesis system in which characteristic of the glottal source signal and aspiration noise can be manipulated explicitly. It can synthesize speech signals with different voice qualities without post processing of the synthetic speech signals. In order to model the glottal source, the transformed LF-model was used to represent the glottal waveform, while the aspiration noise level was estimated by a wavelet denoising algorithm. This thesis also proposes a new threshold function for evaluating threshold values used during the denoising process.

Results show that the synthetic speech signals produced by applying the glottal source as the input to the system is comparable to ones from a traditional HMM-based speech synthesis system that uses a pulse train as its input in terms of their naturalness. The proposed method can also mimic the breathiness and the creakiness of the synthetic speech with more flexibility than the baseline HMM-based system.

Department : .... Computer Engineering	Student's signature .....
Field of study : . Computer Engineering	Advisor's signature .....
Academic year : ..... 2009.....	Co-advisor's signature .....

กิตติกรรมประกาศ

ในโอกาสนี้ข้าพเจ้าใคร่ขอขอบคุณ ผศ. ดร.อติวงศ์ สุชาโต อาจารย์ที่ปรึกษาวิทยานิพนธ์ และอาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม ผศ. ดร.โปรดปราน บุณยพุกกณะ ซึ่งท่านทั้งสองได้ช่วยเหลือ ให้คำแนะนำ ข้อคิดที่เป็นประโยชน์ อันเป็นส่วนสำคัญที่ทำให้วิทยานิพนธ์นี้สำเร็จลุล่วงไปได้ด้วยดี

ขอขอบคุณ ศ. ดร.ประภาส จงสถิตย์วัฒนา ศ. ดร.บุญเสริม กิจศิริกุล และดร.ชัย วุฒิวิวัฒน์ชัย ที่ให้ข้อแนะนำดี ๆ และข้อคิดที่เป็นประโยชน์ต่อการทำวิทยานิพนธ์นี้

นอกจากนี้ข้าพเจ้าขอขอบคุณ พี่ ๆ เพื่อน ๆ และน้อง ๆ ห้องปฏิบัติการ SLS ที่ได้ให้ความร่วมมือ สนับสนุน ช่วยเหลือ จนกระทั่งวิทยานิพนธ์นี้สำเร็จไปได้ด้วยดี

ท้ายที่สุดขอขอบคุณครอบครัวของข้าพเจ้า ที่ได้มอบกำลังใจ ความห่วงใย ช่วยเหลือแก่ข้าพเจ้าเสมอมา

สารบัญ

หน้า

บทคัดย่อภาษาไทย ..... ง
บทคัดย่อภาษาอังกฤษ ..... จ
กิตติกรรมประกาศ ..... ฉ
สารบัญ ..... ช
สารบัญตาราง ..... ฌ
สารบัญภาพ ..... ญ
บทที่ 1 บทนำ 1
ความเป็นมาและความสำคัญของปัญหา 1
วัตถุประสงค์ของการวิจัย 2
ขอบเขตของการวิจัย 3
ประโยชน์ที่คาดว่าจะได้รับ 3
วิธีดำเนินการวิจัย 3
ลำดับขั้นตอนในการเสนอผลการวิจัย 4
บทที่ 2 เอกสารและงานวิจัยที่เกี่ยวข้อง 5
แนวคิดและทฤษฎี 5
1. สรีรสัทศาสตร์ 5
2. สวนสัทศาสตร์ 9
3. การสังเคราะห์เสียงโดยอาศัยแบบจำลองฮิดเดนมาร์คอฟ 13
4. แบบจำลองแอลเอฟแบบแปลง (Transformed LF-model) 23
เอกสารและงานวิจัยที่เกี่ยวข้อง 26
บทที่ 3 ขั้นตอนการดำเนินการวิจัย 29
ขั้นตอนการดำเนินการวิจัย 29
ขั้นตอนการออกแบบระบบสังเคราะห์เสียง 29
1. เครื่องมือที่ใช้ในการวิจัย 29
2. ฐานข้อมูลเสียงที่ใช้พัฒนาระบบ 30
3. การออกแบบระบบสังเคราะห์เสียงที่นำเสนอ 30
ขั้นตอนการสร้างระบบสังเคราะห์เสียง 32

1. ระบบการสังเคราะห์เสียงอ้างอิง 32
2. ระบบการสังเคราะห์เสียงที่นำเสนอ 33
ขั้นตอนการประเมินระบบสังเคราะห์เสียงที่นำเสนอ 33
1. กลุ่มผู้ฟัง 33
2. การเลือกชุดเสียงเพื่อทดสอบ 34
3. การประเมินผลของระบบการสังเคราะห์เสียง 34
บทที่ 4 การวิเคราะห์ข้อมูลเสียง และการสร้างระบบที่นำเสนอ 49
การวิเคราะห์ค่าพารามิเตอร์ 49
1. การวิเคราะห์สัญญาณเส้นเสียง 49
2. การวิเคราะห์หาค่าพารามิเตอร์แบบจำลองแอลเอฟ 52
3. การวิเคราะห์ระดับของสัญญาณรบกวน 53
4. การประเมินผลขั้นตอนการวิเคราะห์ค่าพารามิเตอร์ 55
ระบบการสังเคราะห์เสียงด้วยแบบจำลองฮิดเดนมาร์คอฟที่นำเสนอ 57
1. ขั้นตอนการฝึกฝนแบบจำลองเสียง 57
2. ขั้นตอนการสังเคราะห์เสียงพูด 57
บทที่ 5 ผลการวิจัย 59
ผลการวิจัย 59
1. ผลการประเมินความเป็นธรรมชาติของเสียงสังเคราะห์ 59
2. ผลการประเมินลักษณะของเสียงด้วยการวิเคราะห์ 60
บทที่ 6 บทสรุปผลการวิจัย และข้อเสนอแนะ 78
สรุปผลการวิจัย 78
ประโยชน์ที่ได้รับจากวิทยานิพนธ์นี้ 79
ข้อเสนอแนะ 80
รายการอ้างอิง 81
ประวัติผู้เขียนวิทยานิพนธ์ 86

สารบัญตาราง

หน้า

ตารางที่ 2.1 สภาวะแวดล้อมของข้อความในภาษาไทย 15
ตารางที่ 2.2 มาตราส่วนความถี่การได้ยิน α 21
ตารางที่ 2.3 สัมประสิทธิ์การประมาณค่าของ $R_4(F(z)),\ L=4$ 23
ตารางที่ 2.4 ความสัมพันธ์พารามิเตอร์แอลเอฟ และค่า Rd 25
ตารางที่ 3.1 ตัวเลือกคะแนนระดับความชอบ 35
ตารางที่ 3.2 ประโยคที่ใช้ในการวัดความความเป็นธรรมชาติของเสียงสังเคราะห์ 36
ตารางที่ 3.3 พารามิเตอร์ที่สังเคราะห์เสียงเพื่อทดสอบความถูกต้องของลักษณะเสียง 38
ตารางที่ 3.4 ประโยคที่ใช้ในการวัดความถูกต้องของระบบ 39
ตารางที่ 3.5 ค่าพารามิเตอร์แต่ละชุดทดสอบสำหรับวัดความสามารถของระบบ 42
ตารางที่ 3.6 ชุดทดสอบการเปรียบเทียบระดับความแตกต่างของลักษณะของเสียงแต่ละชนิด..... 44
ตารางที่ 5.1 ผลคะแนน CCR เปรียบเทียบความเป็นธรรมชาติของเสียงสังเคราะห์ 59
ตารางที่ 5.2 ผลการระบุความถูกต้องของลักษณะเสียงในแต่ละประโยค 62
ตารางที่ 5.3 ผลการระบุความถูกต้องของลักษณะเสียงโดยรวม 63
ตารางที่ 5.4 ผลการวิเคราะห์ค่าลักษณะสำคัญทางเสียงในชุดทดสอบ 64
ตารางที่ 5.5 ผลการเปรียบเทียบการวัดระดับความแตกต่างของลักษณะเสียงแบบต่าง ๆ 65

สารบัญภาพ

หน้า

รูปที่ 2.1 อวัยวะภายในของระบบการพูดของมนุษย์ [5] 5
รูปที่ 2.2 กล่องเสียง [5] 8
รูปที่ 2.3 แบบจำลองแหล่งกำเนิดและตัวกรองสัญญาณ [6] 9
รูปที่ 2.4 ตำแหน่งเส้นเสียงขณะขยายตัว และบีบตัว [5] 9
รูปที่ 2.5 การสังเคราะห์เสียงตามแบบจำลองแหล่งกำเนิดเสียง และตัวกรองสัญญาณ [7] 10
รูปที่ 2.6 ลักษณะของสัญญาณเส้นเสียงตามการออกเสียงที่แตกต่างกัน 11
รูปที่ 2.7 สเปกตรัมสัญญาณเส้นเสียงลักษณะเสียงบีบ และเสียงปกติ 11
รูปที่ 2.8 สเปกตรัมของสัญญาณเส้นเสียงลักษณะเสียงลมหายใจ และเสียงปกติ 12
รูปที่ 2.9 สเปกตรัมของลักษณะเสียงแบบต่าง ๆ 13
รูปที่ 2.10 การสังเคราะห์เสียงตามแบบจำลองฮิดเดนมาร์คอฟ [4] 14
รูปที่ 2.11 ส่วนการสังเคราะห์เสียง [13] 16
รูปที่ 2.12 ต้นไม้ตัดสินใจ [13] 17
รูปที่ 2.13 ฟังก์ชันการกระจายตัวความน่าจะเป็นหลายสเปส [13] 17
รูปที่ 2.14 แบบจำลองฮิดเดนมาร์คอฟแบบฟังก์ชันการกระจายตัวความน่าจะเป็นหลายสเปส 18
รูปที่ 2.15 ตัวกรองการประมาณค่าลอการิทึมของสเปกตรัมบนเมลสเกล [13] 22
รูปที่ 2.16 สัญญาณเส้นเสียงสร้างโดยแบบจำลองแอลเอฟ [22] 24
รูปที่ 2.17 กราฟโดเมนเวลาเมื่อปรับค่า Rd 25
รูปที่ 2.18 สเปกตรัมกราฟแอลเอฟเมื่อเปลี่ยนค่า Rd 26
รูปที่ 3.1 ภาพรวมของระบบที่นำเสนอ 30
รูปที่ 3.2 แผนภาพขั้นตอนการสร้างแบบจำลองเสียง 31
รูปที่ 3.3 แผนภาพขั้นตอนการสังเคราะห์เสียง 32
รูปที่ 3.4 ความสัมพันธ์การบอกความเป็นลักษณะของเสียงจากค่า HRF 48
รูปที่ 3.5 ความสัมพันธ์การบอกความเป็นลักษณะของเสียงจากค่า HNR 48
รูปที่ 4.1 แบบจำลองการวิเคราะห์พารามิเตอร์ 49
รูปที่ 4.2 แผนผังการคำนวณตัวกรองสัญญาณย้อนกลับแบบดัดแปลงหลายรอบ [56] 51
รูปที่ 4.3 ผลลัพธ์จากการสกัดสัญญาณเส้นเสียง และผลตอบสนองช่องทางเดินเสียง 52
รูปที่ 4.4 ผลของอัลกอริทึมการประมาณค่าแอลเอฟ 52

รูปที่ 4.5 แผนผังการหาฟังก์ชันการหาจุดเปลี่ยน 54
รูปที่ 4.6 ความสัมพันธ์ของสมการฟังก์ชันการหาจุดเปลี่ยน 54
รูปที่ 4.7 ความสัมพันธ์ระหว่าง ระดับสัญญาณรบกวน และที่สกัดได้ จากรูปร่างสัญญาณเส้นเสียงแบบต่าง ๆ 55
รูปที่ 4.8 ผลการสกัดสัญญาณรบกวน ด้วยฟังก์ชันหาจุดเปลี่ยนที่นำเสนอ 55
รูปที่ 4.9 ค่าพารามิเตอร์ในการสังเคราะห์เสียง ได้จากแบบจำลองที่ฝึกฝน 56
รูปที่ 4.10 ระบบการสังเคราะห์เสียงที่นำเสนอ และระบบอ้างอิง 58
รูปที่ 5.1 สเปกโตรแกรมเปรียบเทียบเสียงสังเคราะห์ของคำว่า “วัน” ในลักษณะเสียงแบบต่าง ๆ 61
รูปที่ 5.2 สเปกตรัมของสัญญาณเสียงชุดทดสอบเสียง ลมหายใจชุดที่ 2 (P7 – P8) เมื่อปรับค่า SNR ลดลง 20 dB 67
รูปที่ 5.3 สเปกตรัมของสัญญาณเสียงชุดทดสอบเสียงปกติ-ลมหายใจชุดที่ 1A (B1-B7) เมื่อเพิ่มสัญญาณรบกวนสีขาว 68
รูปที่ 5.4 สเปกตรัมของสัญญาณเสียงชุดทดสอบเสียงปกติ-ลมหายใจชุดที่ 1B (P1 – P7) เมื่อปรับค่า SNR ลดลง 20 dB 69
รูปที่ 5.5 สเปกตรัมของสัญญาณเสียงชุดทดสอบเสียงลมหายใจชุดที่ 3 (P7 – P12) เมื่อปรับค่า SNR ลดลง 20 dB และ 40 dB 69
รูปที่ 5.6 สเปกโตรแกรมของชุดทดสอบเสียงลมหายใจ เมื่อปรับค่า SNR ลดลง 20 และ 40 dB (P7 – P12) 70
รูปที่ 5.7 สเปกตรัมของสัญญาณเสียงชุดทดสอบเสียงปกติ-บีบชุดที่ 7 (P1 – P2) เมื่อปรับค่า Rd เป็น 0.3 เท่าของค่าปกติ 72
รูปที่ 5.8 สเปกตรัมของสัญญาณเสียงชุดทดสอบเสียงบีบชุดที่ 10 (P5 – P6) เมื่อปรับ Rd เป็น 0.3 เท่าของค่าปกติ ที่ค่า F0 ลดลง 50 Hz 73
รูปที่ 5.9 สเปกตรัมของสัญญาณเสียงชุดทดสอบเสียงบีบชุดที่ 11 (P3 – P4) เมื่อปรับค่า Rd เป็น 0.3 เท่าของค่าปกติ ที่ F0 ลดลง 100 Hz 73
รูปที่ 5.10 สเปกตรัมของสัญญาณเสียงชุดทดสอบเสียงบีบชุดที่ 12 (P5 – P13) เมื่อปรับ Rd เป็น 0.3 และ 0.5 เท่าของค่าปกติ ที่ F0 ลดลง 100 Hz 74
รูปที่ 5.11 สเปกตรัมของสัญญาณเสียงชุดทดสอบเสียงบีบชุดที่ 6 (P1 – P6) เมื่อปรับ F0 ลดลง 50 Hz ที่ค่า Rd เป็นค่าปกติ 75
รูปที่ 5.12 สเปกตรัมของสัญญาณเสียงชุดทดสอบเสียงบีบชุดที่ 8 (P3 – P5) เมื่อปรับ F0 ลดลง 100 และ 50 Hz ที่ค่า Rd เป็น 0.3 เท่าของค่าปกติ 76

รูปที่ 5.13 สเปกตรัมของสัญญาณเสียงชุดทดสอบเสียงบีบชุดที่ 9 (P2 – P3) เมื่อปรับ F0 ลดลง 100 Hz ที่ค่า Rd เป็น 0.3 เท่าของค่าปกติ..... 76

บทที่ 1
บทนำ

ความเป็นมาและความสำคัญของปัญหา

ในปัจจุบันการสังเคราะห์เสียงพูดของมนุษย์ ซึ่งเป็นการสร้างสัญญาณเสียงเพื่อใช้ในการติดต่อสื่อสารระหว่างคอมพิวเตอร์และมนุษย์ได้รับความสนใจและถูกนำใช้ในโปรแกรมประยุกต์ต่าง ๆ มากขึ้น เช่น โปรแกรมอ่านจดหมายอิเล็กทรอนิกส์ ระบบสังเคราะห์เสียงพูดของหุ่นยนต์เพื่อให้หุ่นยนต์สามารถพูดโต้ตอบกับมนุษย์ และพจนานุกรมพูดได้ รวมไปถึงโปรแกรมสำหรับช่วยคนพิการ โปรแกรมอ่านหน้าจอสำหรับคนพิการทางการมองเห็น โปรแกรมแปลงข้อความเป็นเสียงพูดสำหรับคนพิการทางการพูด เป็นต้น

ระบบสังเคราะห์เสียงได้มีการวิจัยและพัฒนาหลายแนวทาง เช่นการสังเคราะห์เสียงด้วยฟอร์แมนท์ (Formant synthesis) [1] เป็นการสังเคราะห์จากการวิเคราะห์ค่าตัวแปรฟอร์แมนท์จากสัญญาณเสียงพูดของมนุษย์ แล้วนำมาสร้างเป็นกฎในการสังเคราะห์เสียง การกำหนดค่าฟอร์แมนท์ให้เสียงพูดต่อเนื่องนั้นทำได้ยากเนื่องจากความซับซ้อนของทางเดินเสียงร่วม (Co-articulation) ทำให้เสียงที่ได้จากการสังเคราะห์ไม่มีความเป็นธรรมชาติมากนัก อีกแนวทางการสังเคราะห์เสียงด้วยการจำลองอวัยวะที่ให้กำเนิดเสียง (Articulator synthesis) [2] ซึ่งเป็นการสร้างเสียงสังเคราะห์จากการค่าพารามิเตอร์ซึ่งวิเคราะห์จากอวัยวะกำเนิดเสียงพูดของมนุษย์ โดยศึกษาแบบจำลองอวัยวะกำเนิดเสียงของมนุษย์จากผลการฉายคลื่นแม่เหล็กไฟฟ้า (MRI) หรือการฉายรังสีเอ็กซ์ (X-Ray) ในการจับภาพช่องทางเสียงและตำแหน่งลิ้นขณะออกเสียง แล้วสร้างเป็นแบบจำลองการเคลื่อนไหวของอวัยวะที่ให้กำเนิดเสียงของมนุษย์ ซึ่งเป็นลักษณะแบบจำลองช่องทางเสียง แล้วนำไปคำนวณหาเสียงออกมาได้โดยผ่านฟังก์ชันโยกย้าย (Transfer function) อีกแนวทางหนึ่งคือการสังเคราะห์เสียงจากการนำหน่วยเสียงย่อยหรือส่วนเสียงที่จัดเก็บเอาไว้มาเชื่อมต่อกัน (Concatenative speech synthesis) เช่น การสังเคราะห์เสียงพูดด้วยการคัดเลือกหน่วยเสียง (Unit selection) [3] เป็นอีกเทคนิคที่สังเคราะห์เสียงโดยเลือกหน่วยเสียงจากฐานข้อมูลเสียงที่เก็บเสียงจริงของมนุษย์ แล้วนำเสียงมาเชื่อมต่อกัน เสียงที่ได้จากการสังเคราะห์จึงมีความเป็นธรรมชาติสูง แต่มีข้อจำกัดคือต้องมีการจัดเก็บข้อมูลเสียงพูดขนาดใหญ่ ต่อมาการสังเคราะห์เสียงโดยใช้แบบจำลองฮิดเดนมาร์คอฟ (HMM-based speech synthesis) [4] ได้รับความสนใจมากขึ้น ซึ่งการสังเคราะห์เสียงโดยใช้แบบจำลองฮิดเดนมาร์คอฟนี้ เป็นการสังเคราะห์เสียงจากค่าพารามิเตอร์ที่ฝึกฝนจากแบบจำลองของเสียงที่สร้างด้วยกระบวนการทางสถิติ ดังนั้นข้อดีของระบบการ

สังเคราะห์เสียงแนวทางนี้คือสามารถสร้างสัญญาณเสียงได้อย่างอัตโนมัติจากค่าพารามิเตอร์ที่ได้จากแบบจำลองที่ได้รับการฝึกฝนจากข้อมูลเสียง และเสียงสังเคราะห์ที่ได้มีความยืดหยุ่นและสามารถเปลี่ยนลักษณะของเสียงที่แตกต่างหลากหลาย โดยการแปลงแบบจำลองทางสถิติของค่าพารามิเตอร์ที่ใช้ฝึกฝน

นอกจากนี้ลักษณะของเสียงมีผลต่อความน่าสนใจในการสื่อสาร หรือมีผลต่อการฟังข้อความ และสามารถแสดงอารมณ์ในการสื่อสาร อีกทั้งสามารถบอกเอกลักษณ์ของเสียงผู้พูดได้ ดังนั้นการปรับปรุงลักษณะของเสียงให้ดีขึ้น และการให้อารมณ์กับเสียงสังเคราะห์ในการสังเคราะห์เสียง โดยศึกษาค่าพารามิเตอร์ที่เกี่ยวข้องที่มีผลต่อการปรับปรุงลักษณะของเสียง ได้แก่ ระดับของสัญญาณรบกวนลมหายใจ (Aspiration noise) ซึ่งเป็นชนิดหนึ่งของสัญญาณรบกวนจากการไหลของกระแสผ่านช่องลม (Turbulence noise) ความแข็งแรงของแหล่งกำเนิดเส้นเสียง อีกทั้งอัตราส่วนช่วงการเปิดของช่องเส้นเสียง (Open quotient) ความเร็วอัตราส่วนการเปิดของช่องเส้นเสียง มีผลต่อการผลิตลักษณะของเสียง

วิทยานิพนธ์นี้จึงเสนอการพัฒนาการสังเคราะห์เสียงที่ดัดแปลงการสังเคราะห์เสียงตามแบบจำลองฮิดเดนมาร์คอฟ ให้สามารถกำหนดสัญญาณจากแหล่งกำเนิดเส้นเสียง และสัญญาณรบกวนลมหายใจได้โดยตรง เพื่อสร้างเสียงสังเคราะห์ที่เลียนแบบลักษณะเสียงพูดของมนุษย์ โดยปรับค่าลักษณะของเสียงตามค่าพารามิเตอร์ของเส้นเสียงที่มีความสำคัญต่อลักษณะคุณภาพเสียงได้ เพื่อสร้างเสียงสังเคราะห์เลียนแบบลักษณะเสียงมนุษย์ชนิดต่าง ๆ ได้

วัตถุประสงค์ของการวิจัย

1 สร้างระบบสังเคราะห์เสียงพูดโดยอาศัยแบบจำลองฮิดเดนมาร์คอฟ ที่ใช้แบบจำลองเส้นเสียง และระดับเสียงรบกวนทางลมหายใจเป็นสัญญาณกระตุ้น
2 เพื่อสร้างระบบสังเคราะห์เสียงที่สามารถสร้างเสียงสังเคราะห์ที่มีลักษณะเสียงบีบ เสียงปกติ และเสียงลมหายใจ ซึ่งเลียนแบบลักษณะเสียงพูดของมนุษย์จากการปรับค่าพารามิเตอร์แหล่งกำเนิดเส้นเสียง และระดับเสียงรบกวนทางลมหายใจ
3 เพื่อสร้างระบบสังเคราะห์เสียงที่สามารถสังเคราะห์เสียงที่มีลักษณะเสียงบีบ และเสียงลมหายใจ ให้มีระดับความแตกต่างของระดับความเป็นเสียงบีบ และเสียงลมหายใจ จากการปรับค่าพารามิเตอร์ของสัญญาณกระตุ้นได้โดยง่าย

ขอบเขตของการวิจัย

1 ศึกษากระบวนการการสร้างเสียงพูด เพื่อมาใช้ในการหาความสัมพันธ์ของลักษณะเสียงที่เกิดจากแหล่งกำเนิดในเสียงบีบ เสียงปกติ และเสียงลมหายใจ

2 สร้างระบบการสังเคราะห์เสียงพูด ซึ่งพัฒนาเฉพาะส่วนของขั้นตอนการปรับลักษณะเสียงจากแหล่งกำเนิดเสียงบนระเบียบวิธีการสังเคราะห์เสียงพูดโดยอาศัยแบบจำลองฮิดเดนมาร์คอฟ

3 ทดลองและวัดประสิทธิภาพของเสียงสังเคราะห์ที่ได้จากแบบจำลองเสียงที่นำเสนอ (แบบจำลองที่ได้หลังจากการปรับพารามิเตอร์) เทียบกับเสียงสังเคราะห์ที่สร้างจากแบบจำลองของทฤษฎีที่เกี่ยวข้อง ซึ่งในวิทยานิพนธ์นี้หมายถึง แบบจำลองที่ใช้ขบวนอิมพัลส์ โดยวัดความคิดเห็นจากคะแนนความพอใจของผู้ฟัง

4 สังเคราะห์เสียงที่มีลักษณะของเสียงต่าง ๆ ได้ ซึ่งสามารถเลียนแบบเสียงปนลมหายใจ (Breathy) และเสียงบีบ (Creaky) ได้

ประโยชน์ที่คาดว่าจะได้รับ

ทราบถึงลักษณะของแหล่งกำเนิดเส้นเสียง และกระบวนการทำให้เกิดเสียงของมนุษย์ และหาค่าพารามิเตอร์ที่เหมาะสมในการสร้างแบบจำลองของแหล่งกำเนิดเสียงได้ อีกทั้งสามารถสร้างเสียงสังเคราะห์ได้อัตโนมัติ จากการปรับลักษณะเสียงจากแหล่งกำเนิดเส้นเสียงมีความเป็นธรรมชาติ มากกว่าเสียงที่สร้างจากแบบจำลองที่สร้างจากขบวนอิมพัลส์ นอกจากนี้สามารถสร้างเสียงที่มีลักษณะของเสียงต่าง ๆ ซึ่งเลียนแบบลักษณะของเสียงที่แตกต่างกันของมนุษย์ได้ ดังนั้นเป็นประโยชน์ในการนำไปใช้สร้างระบบสังเคราะห์เสียงโดยอาศัยแบบจำลองฮิดเดนมาร์คอฟในภาษาไทยที่มีประสิทธิภาพได้ และไปพัฒนาใช้กับโปรแกรมประยุกต์อื่น ๆ ได้ด้วย หรือนำระบบการสังเคราะห์เสียงนี้ไปใช้ประยุกต์ในการพัฒนาเครื่องมือที่ใช้ในการติดต่อสื่อสารระหว่างคอมพิวเตอร์และมนุษย์ให้ดีขึ้น เพื่อเพิ่มความหลากหลายของการสื่อสารให้ธรรมขาติ และเพิ่มอารมณ์ของการสื่อสารประโยคจากคอมพิวเตอร์ได้

วิธีดำเนินการวิจัย

1 ขั้นตอนการเตรียมตัว

1 ศึกษาข้อมูลและทฤษฎีที่เกี่ยวกับงานวิจัย เช่น ทฤษฎีสรีรสัทศาสตร์ ลักษณะของเสียงพูด การวิเคราะห์เสียงพูด และความรู้อื่น ๆ

2 ศึกษาเกี่ยวกับกระบวนการการสร้างเสียงพูด

3 ศึกษางานวิจัยที่เกี่ยวข้อง

4 เตรียมตัวอย่างเสียงต้นแบบที่จะใช้ในการศึกษาสมบัติของเสียง

2 ขั้นตอนออกแบบการปรับลักษณะเสียงสังเคราะห์

1 วิเคราะห์เสียงต้นแบบเพื่อหาสมบัติของเสียงลักษณะแบบต่าง ๆ

2 วิเคราะห์สมบัติแหล่งกำเนิดเสียงเส้นเสียง

3 ออกแบบขั้นตอนการพัฒนา และวิเคราะห์ผล

3 ขั้นตอนพัฒนาและทดสอบระบบการปรับลักษณะเสียงสังเคราะห์

1 จัดทำโปรแกรมในแต่ละส่วนที่ได้ออกแบบไว้สำหรับใช้ทำการปรับลักษณะเสียงสังเคราะห์

2 ทดสอบประสิทธิภาพของการปรับลักษณะเสียงสังเคราะห์ด้วยการเปรียบเทียบผลกับทฤษฎี และการวัดคะแนนความคิดเห็นจากผู้ฟัง แล้วจึงสรุปผลต่อไป

ลำดับขั้นตอนในการเสนอผลการวิจัย

ในวิทยานิพนธ์ฉบับนี้จะแบ่งเนื้อหาในการนำเสนอออกเป็น 6 ส่วน คือ ในบทที่ 2 จะกล่าวถึง ทฤษฎีและงานวิจัยที่เกี่ยวข้องของวิทยานิพนธ์นี้ โดยรวมทฤษฎีทางสรีรสัทศาสตร์ ทฤษฎีที่เกี่ยวกับกระบวนการทำให้เกิดเสียง ลักษณะของเสียง เสียงในภาษาไทย เป็นต้น และทฤษฎีการสังเคราะห์เสียงโดยใช้แบบจำลองฮิดเดนมาร์คอฟ

ในบทที่ 3 จะได้กล่าวถึงวิธีการดำเนินการวิจัย ฐานข้อมูลที่ใช้ในพัฒนาระบบและวัดประสิทธิภาพระบบ การสร้างระบบการสังเคราะห์เสียงที่นะเสนอ และอธิบายการประเมินผลเพื่อเปรียบเทียบคุณภาพเสียงของระบบการสังเคราะห์เสียงที่พัฒนาส่วนของขั้นตอนการปรับลักษณะเสียงที่เสนอในวิทยานิพนธ์นี้ ในบทที่ 4 อธิบายการวิเคราะห์พารามิเตอร์ที่มีผลต่อการปรับปรุงลักษณะของเสียง ได้แก่ การวิเคราะห์แหล่งกำเนิดเส้นเสียง การวิเคราะห์แบบจำลองแอลเอฟ และการวิเคราะห์ระดับของสัญญาณรบกวน โดยใช้วิธีการลดสัญญาณรบกวนด้วยการประมาณค่าจุดเปลี่ยนของเวฟเลท (Wavelet thresholding) เป็นต้น รวมทั้งอธิบายขั้นตอนการสังเคราะห์เสียงโดยรวม และวิธีการสร้างแบบจำลองที่พัฒนาส่วนของขั้นตอนการปรับลักษณะเสียงที่เสนอ

ในบทที่ 5 และบทที่ 6 จะได้กล่าวถึงผลการทดลองในการวิเคราะห์หาค่าพารามิเตอร์ที่ใช้ในการสร้างแบบจำลอง และสรุปผลที่ได้จากวิทยานิพนธ์นี้ รวมทั้งให้ข้อเสนอแนะในการพัฒนางานวิจัยการสังเคราะห์เสียงอิงแบบจำลองฮิดเดนมาร์คอฟโดยกำหนดสัญญาณจากแหล่งกำเนิดเส้นเสียง และสัญญาณรบกวนลมหายใจได้โดยตรงต่อไป

บทที่ 2
เอกสารและงานวิจัยที่เกี่ยวข้อง

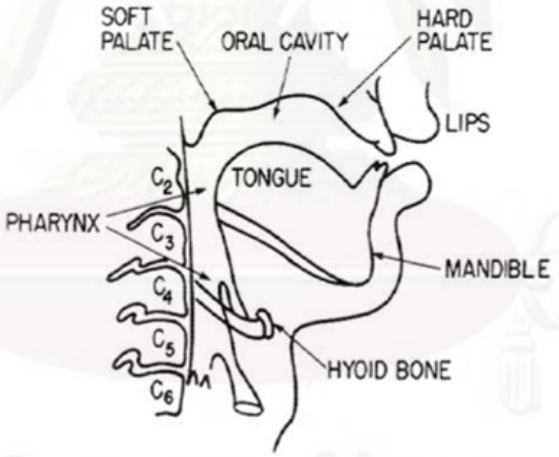
บทนี้จะนำเสนอทฤษฎีพื้นฐาน และแนวคิดที่เกี่ยวข้องกับการศึกษาแหล่งกำเนิดเส้นเสียง และการพัฒนาระบบสังเคราะห์เสียง โดยเริ่มจากทฤษฎีทางสรีรสัทศาสตร์ ซึ่งเป็นทฤษฎีที่ศึกษาปรากฏการณ์ของเสียงพูดในด้านต่าง ๆ จากนั้นจะนำเสนอทฤษฎีการสังเคราะห์เสียงโดยอาศัยแบบจำลองฮิดเดนมาร์คอฟ และแบบจำลองแอลเอฟ

แนวคิดและทฤษฎี

1. สรีรสัทศาสตร์

1.1 อวัยวะที่ทำให้เกิดเสียง (The Organs of Speech)

ตำแหน่งของอวัยวะที่ทำให้เกิดเสียงของมนุษย์สามารถแสดงได้ดังรูปที่ 2.1



รูปที่ 2.1 อวัยวะภายในของระบบการพูดของมนุษย์ [5]

คำพูดที่เราเปล่งออกมาเป็นลำดับขั้นของความถี่ของคลื่นเสียง นั่นคือตำแหน่งของช่องทางเดินของเสียงมนุษย์จะมีการเปลี่ยนแปลงเป็นลำดับขั้นอย่างต่อเนื่อง อวัยวะที่ใช้ในการออกเสียงทั้งหมดมีดังนี้

1 ริมฝีปาก (Lips) เป็นอวัยวะส่วนที่สามารถเคลื่อนไหวได้ และทำให้เสียงแตกต่างกันได้มาก เราอาจจะบังคับริมฝีปากให้ปิดสนิท ให้เปิดเล็กน้อย ให้เปิดกว้างขึ้น ให้ยื่นออกมา ให้ห่อกลมหรือทำเป็นรูปรีก็ได้ ลักษณะต่าง ๆ ของริมฝีปากล้วนมีผลต่อการ

ออกเสียงทั้งสิ้น เสียงพยัญชนะที่เกิดจากการกักที่ริมฝีปากเรียกว่าเสียงโอฐชะ (Bilabial sound)

2 ฟัน (Teeth) เป็นอวัยวะที่เป็นฐานหรือตำแหน่งที่เกิดของเสียงหลายชนิด เช่น เมื่อฟันบนกดลงบนริมฝีปากล่าง ลมที่ผ่านออกมาโดยแรงจะลอดช่องที่พอจะผ่านได้ออกมาทำให้เกิดเป็นเสียงชนิดที่เรียกว่า เสียงบีบแทรกที่เกิดระหว่างฟันกับริมฝีปาก ถ้าฟันบนกดกับฟันล่าง ลมที่ผ่านออกมาโดยแรงจะทำให้ได้เสียงบีบแทรกที่เกิดที่ฟัน เป็นต้น นอกจากนี้เนื่องจากปลายลิ้นอยู่ใกล้กับฟัน ปลายลิ้นจึงมักจะทำอาการต่าง ๆ บริเวณฟันและหลังฟันบ่อย ๆ ทำให้เกิดเสียงทันตชะ (Dental sound)

3 ปุ่มเหงือก (Alveolus, Gum ridge, Tooth ridge) เป็นส่วนที่นูนออกมาตรงบริเวณหลังฟันด้านบน ถ้าเอาลิ้นแตะดูจะรู้สึกว่ามีลักษณะเป็นคลื่น ลิ้นอาจแตะหรือวางอยู่ใกล้บริเวณปุ่มเหงือก ซึ่งทำให้เกิดเสียงมุทธชะ (Alveolar sound)

4 เพดานแข็ง หรือเพดานปาก (Palate, Hard palate) หมายถึงส่วนโค้งของเพดานปากส่วนที่เป็นกระดูกแข็ง ซึ่งอยู่ถัดจากปุ่มเหงือกเข้ามา ถ้าลิ้นแตะหรือวางใกล้เพดานแข็งจะทำให้เกิดเสียงตาลุชะ (Palatal sound)

5 เพดานอ่อน (Velum, Soft palate) คือ ส่วนของเพดานที่อยู่ต่อเพดานแข็งเข้าไปข้างในเป็นกระดูกอ่อนที่ขยับขึ้นลงได้เล็กน้อย เวลาหายใจเพดานอ่อนและลิ้นไก่ซึ่งอยู่ปลายเพดานอ่อนจะลดระดับลงมาเปิดช่องให้ลมออกทางจมูก ฉะนั้นเวลาที่ไม่พูด เพดานอ่อนและลิ้นไก่จะลดระดับลงมา เวลาพูดส่วนใหญ่เพดานอ่อนและลิ้นไก่จะถูกยกขึ้นไปจดกับผนังคอ จะมีแต่เวลาออกเสียงนาสิกเท่านั้นที่เพดานอ่อนจะลดระดับลงมาเพื่อให้ลมออกไปทางจมูกได้ ถ้าลิ้นแตะหรือวางใกล้เพดานอ่อนจะทำให้เกิดเสียงที่เกิดที่เพดานอ่อน (Velar sound)

6 ลิ้นไก่ (Uvula) เป็นก้อนเนื้อเล็ก ๆ อยู่ต่อจากปลายเพดานอ่อนเข้าไปข้างใน และห้อยอยู่ตรงกลางปาก สามารถสั่นรัวได้ เวลาอ้าปากมักจะเห็น ลิ้นไก่ใช้ออกเสียงในบางภาษาเช่น ภาษาเยอรมัน ฝรั่งเศส นอร์เวย์ อาหรับ และอิสราเอล เป็นต้น

7 ช่องจมูก (Nasal cavity) หมายถึง โพรงในช่องจมูกซึ่งอยู่เหนือลิ้นไก่ขึ้นไป เป็นช่องที่ลมซึ่งผ่านเส้นเสียงขึ้นมาจะผ่านออกไปทางจมูกได้เมื่อเวลาหายใจและเวลาออกเสียงนาสิก ในเวลาเปล่งเสียงอื่น ๆ ลิ้นไก่จะถูกยกขึ้นไปปิดช่องจมูกเพื่อให้ลมออกมาทางช่องปาก

8 ลิ้น (Tongue) เป็นส่วนที่เคลื่อนไหวได้มากที่สุดในการออกเสียงพูด ส่วนที่เคลื่อนไหวของลิ้นแต่ละส่วนมีผลต่อการออกเสียง เราจึงแบ่งลิ้นออกเป็น 3 ส่วน ด้วยกันตามหน้าที่ที่มีในการออกเสียงคือ

1 ปลายลิ้น (Tip of the tongue) หรือ ลิ้นส่วนปลายสุด หมายถึงส่วนปลายของลิ้น ซึ่งสามารถจะยกขึ้นไปแตะอวัยวะส่วนต่าง ๆ ในปากตอนบนได้โดยง่าย

2 หน้าลิ้น (Blade of the tongue) หรือ ลิ้นส่วนหน้า หมายถึงลิ้นส่วนที่อยู่ตรงข้ามกับเพดานแข็ง ในขณะที่วางลิ้นราบกับปากตอนไม่ได้พูด

3 หลังลิ้น (Back of the tongue) หรือ ลิ้นส่วนหลัง หมายถึงส่วนของลิ้นที่อยู่ตรงข้ามกับเพดานอ่อน ในขณะที่วางลิ้นราบกับปากตอนไม่ได้พูด

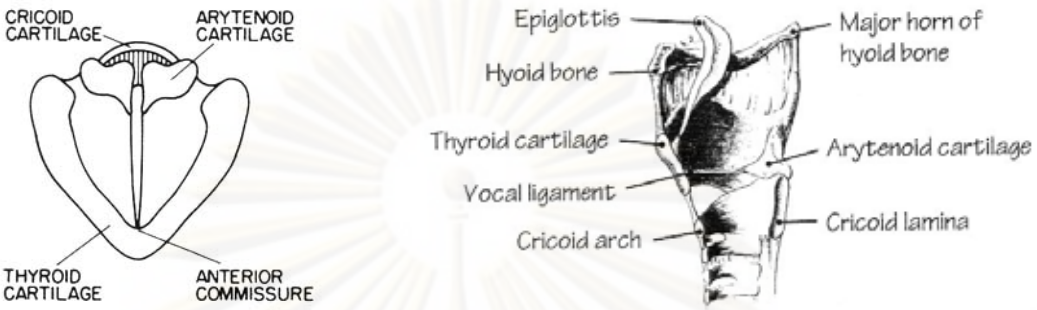
9 แผ่นเนื้อปากหลอดลม (Epiglottis) หรือ ลิ้นปิดกล่องเสียง เป็นก้อนเนื้อเล็ก ๆ คล้ายลิ้นไก่อยู่ต่อโคนลิ้นลงไปในคอ มีหน้าที่ปิดเปิดช่องหลอดลม เพื่อป้องกันมิให้อาหารตกลงไปในหลอดลม ในเวลาที่กลืนอาหาร แผ่นเนื้อปากหลอดลมปิดลงให้อาหารผ่านไปลงหลอดอาหาร แต่ในเวลาที่พูด แผ่นเนื้อนี้จะเปิดออกเพื่อให้ลมจากหลอดลมออกมา

10 โพรงคอ (Pharynx) เป็นโพรงซึ่งอยู่ถัดปากลงไปจากช่องปากจนถึงเส้นเสียงหรือสายเสียง

11 เส้นเสียง หรือสายเสียง (Vocal cords) เป็นอวัยวะสำคัญที่ทำให้เกิดเสียง เส้นเสียงประกอบด้วยเส้นเอ็นและกล้ามเนื้อเป็นแผ่น 2 แผ่น มีความยาวประมาณ 1.2-1.7 เซนติเมตร กว้างประมาณ 0.2-0.3 เซนติเมตร ปิดขวางอยู่ตรงปากของช่องหลอดลม โดยจะวางตัวจากด้านหลังมายังด้านหน้าอยู่ตรงกลางของกล่องเสียง เส้นเสียงทั้งสองสามารถที่จะดึงออกให้ห่างจากกันหรือดึงเข้ามาให้ชิดกันก็ได้ เส้นเสียงเป็นส่วนสำคัญที่ทำให้เกิดเสียงพูด โดยจะเปิดให้ลมผ่านในเวลาหายใจตามปกติ แต่จะอยู่ชิดกันเมื่อมีการเปล่งเสียง

12 กล่องเสียง (Larynx) ตั้งอยู่ตอนบนของหลอดลมตรงตำแหน่งที่เรียกว่าลูกกระเดือก (Adam's apple) กล่องเสียงประกอบด้วยกระดูกอ่อนหลายส่วนด้วยกัน ส่วนที่อยู่ด้านหน้า คือ กระดูกอ่อนไทรอยด์ (Thyroid cartilage) ปลายด้านหนึ่งของเส้นเสียงทั้งสองจะเชื่อมอยู่กับกระดูกอ่อนไทรอยด์นี้และอยู่ชิดกัน ส่วนปลายอีกด้านหนึ่งของเส้นเสียงทั้งสอง จะเชื่อมอยู่กับกระดูกอ่อนอาริตินอยด์ (Arytenoids cartilages) ซึ่งเป็นกระดูกอ่อนอีกสองชิ้น กระดูกอ่อนอาริตินอยด์และกล้ามเนื้อในกล่องเสียงจะทำให้เส้นเสียงทั้งสองอยู่ชิดติดกันหรือห่างจากกันได้ เมื่อเส้นเสียงอยู่ห่างจากกันจะเกิดเป็น

ช่องสามเหลี่ยม ซึ่งเป็นทางให้ลมผ่านเข้าไปถึงปอด หรือผ่านออกมาจากปอดได้ ดังรูปที่ 2.2



ก. มุมมองจากด้านบน ข. มุมมองจากด้านข้าง

รูปที่ 2.2 กล่องเสียง [5]

13 ช่องระหว่างเส้นเสียง (Glottis) จะเปิดอยู่ระหว่างที่หายใจเข้าออกตามปกติ แต่จะปิดลงเมื่อมีการเปล่งเสียง ก่อให้เกิดการสั่น และเป็นเสียงดังขึ้น

14 ช่องปาก (Oral cavity) ทำหน้าที่เป็นช่องกำทอน (Resonant chamber) ซึ่งสามารถเปลี่ยนให้มีรูปร่างต่าง ๆ กัน ตามท่าทางของอวัยวะภายในช่องปาก โดยอวัยวะภายในช่องปากอาจสามารถแบ่งได้ดังนี้

1 อวัยวะส่วนกระทำอาการ (Articulator) หมายถึงอวัยวะส่วนที่เคลื่อนไหวเพื่อผลักหรือกักลมในที่ต่าง ๆ อวัยวะส่วนกระทำอาการที่สำคัญคือลิ้น ซึ่งเคลื่อนไหวได้มากที่สุด อวัยวะส่วนกระทำอาการอาจเรียกว่า “กรณ์”

2 อวัยวะส่วนเกิดอาการ (Point of articulation) หมายถึง ตำแหน่งที่อวัยวะส่วนกระทำอาการเคลื่อนไหวไป เพื่อผลักหรือกักลมไว้ อาจเรียกอวัยวะส่วนนี้ว่า “ฐาน” ที่เกิดของหน่วยเสียงต่าง ๆ ฐานภายในช่องปากที่สำคัญได้แก่ ริมฝีปาก ฟัน ปุ่มเหงือก เพดานแข็ง และเพดานอ่อน

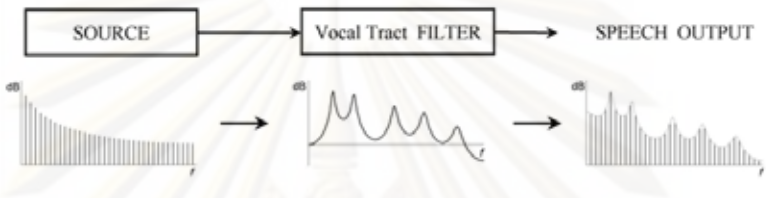
15 หลอดลม (Trachea) เป็นทางเดินอากาศจากปอดถึงกล่องเสียง

ความรู้เกี่ยวกับสรีรสัทศาสตร์ ในวิทยานิพนธ์นี้จะถูกนำไปใช้ในการศึกษาและวิเคราะห์กระบวนการทำให้เกิดเสียงของมนุษย์

2. สวนสัทศาสตร์

2.1 กระบวนการทำให้เกิดเสียง (Speech production)

กระบวนการทำให้เกิดเสียงของมนุษย์เป็นผลลัพธ์ของการรวมกันของแหล่งกำเนิด (Source) ได้แก่ กล่องเสียง และฟังก์ชันโยกย้าย (Transfer function) ได้แก่ รูปร่างของช่องทางเดินเสียง แบบจำลองนี้เรียกว่าแบบจำลองแหล่งกำเนิดและตัวกรองสัญญาณ ดังแสดงตามรูปที่ 2.3



รูปที่ 2.3 แบบจำลองแหล่งกำเนิดและตัวกรองสัญญาณ [6]

ภายในกล่องเสียงประกอบด้วยเส้นเสียง (Vocal cords) ซึ่งเป็นกลุ่มของกล้ามเนื้อ และเส้นเอ็นสองชุดยึดติดอยู่ทั้งสองข้างของกล่องเสียง ช่องระหว่างเส้นเสียงทั้งสองเรียกว่าช่องเส้นเสียง (Glottis) ในระหว่างการหายใจ หรือการพูด เส้นเสียงจะมีรูปร่างแตกต่างกันออกไป เช่น ในขณะหายใจช่องเส้นเสียงจะขยายออก (Abduct) และจะบีบตัวแคบลงในขณะที่ออกเสียงพูด (Adduct) แสดงตัวอย่างในรูปที่ 2.4

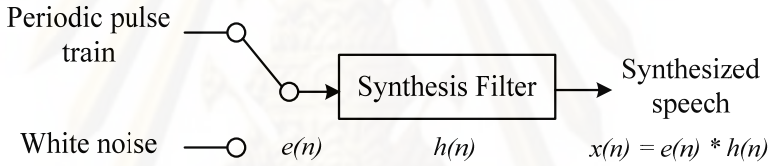
เปิดออก (abduct) บีบตัวเข้า (adduct)

รูปที่ 2.4 ตำแหน่งเส้นเสียงขณะขยายตัว และบีบตัว [5]

ในการออกเสียงพูดแหล่งกำเนิดเสียงเกิดจากการสั่นสะเทือนของเส้นเสียง โดยความดันอากาศจากกะบังลมจะพยายามไหลผ่านเส้นเสียงที่ปิดอยู่ เมื่อมีแรงดันมากพอเส้นเสียงจะถูกแยกออก และอากาศสามารถไหลผ่านได้เรียกว่าช่วงกำลังเปิด (Opening phase) ต่อมาเมื่อแรงดันจากกะบังลมเริ่มลดลง เส้นเสียงจะเริ่มปิดตัวลง เรียกช่วงนี้ว่าช่วงกำลังปิด (Closing phase) จะเส้นเสียงจะเปิดอีกครั้งเมื่อมีแรงดันมากพอ กระบวนการดังกล่าวทำให้เกิดเสียงนี้มีลักษณะเป็นรายคาบ ซึ่งเรียกว่าเสียงโฆษะ (Voiced sound) เป็นเสียงที่เกิดจากการสั่นของเส้นเสียง ประกอบด้วยความถี่มูล

ฐาน และฮาร์โมนิกส์ต่าง ๆ เมื่อสัญญาณกระตุ้นนี้ผ่านช่องทางเดินเสียงซึ่งทำหน้าที่เป็นตัวกรองสัญญาณ ความถี่ฟอร์แมนท์ซึ่งเป็นส่วนสำคัญของสัญญาณเสียง อีกด้านหนึ่งจะมีเสียงที่ไม่มีความเป็นคาบ ซึ่งเรียกว่าเสียงอโฆษะ (Unvoiced sound) เป็นเสียงที่ไม่ได้เกิดจากการสั่นของเส้นเสียงสั่น แต่เกิดจากแหล่งกำเนิดเสียงอื่นเช่นสัญญาณรบกวนลมหมุนตามช่องปิด (Turbulence noise) ซึ่งถูกพิจารณาเป็นสัญญาณรบกวนสีขาว (White noise) ในสัญญาณเสียงเสียงอโฆษะนี้ไม่มีความถี่มูลฐาน

จากกระบวนการการเกิดเสียงพูดของมนุษย์สามารถอธิบายกับการสังเคราะห์เสียงได้ตามรูปที่ 2.5 โดยมีสัญญาณเสียงส่วนที่เป็นคาบ และสัญญาณรบกวนสีขาว เป็นสัญญาณกระตุ้น (Excitationม, $e(n)$) ซึ่งเป็นอินพุทของตัวกรองสังเคราะห์ $h(n)$ เมื่อสัญญาณกระตุ้นผ่านตัวกรองสัญญาณเสียงจะถูกสร้างตามสเปกตรัมของตัวกรองสัญญาณ โดยมีลักษณะของความเป็นคาบตามสัญญาณกระตุ้น

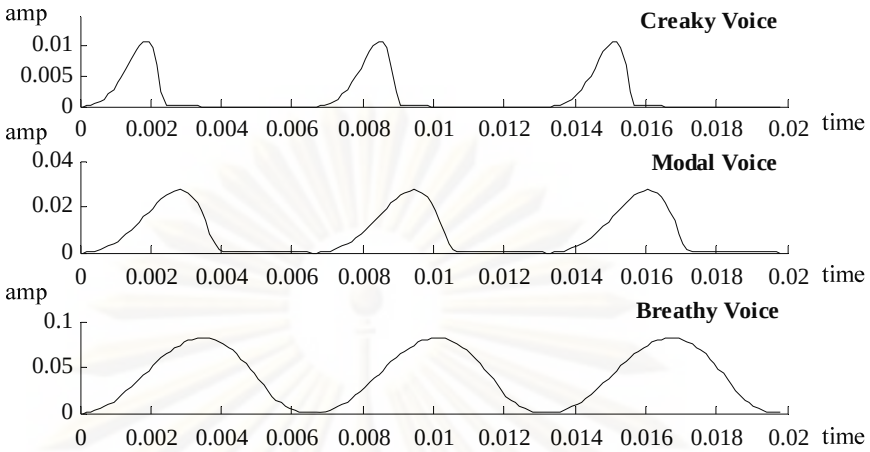


รูปที่ 2.5 การสังเคราะห์เสียงตามแบบจำลองแหล่งกำเนิดเสียง และตัวกรองสัญญาณ [7]

ความรู้ในเรื่องกระบวนการการทำให้เกิดเสียงของมนุษย์จะถูกนำไปใช้ในพัฒนาแบบจำลองการสังเคราะห์เสียง เพื่อให้เข้าใจและสามารถนำสัญญาณเส้นเสียงใช้เป็นสัญญาณกระตุ้นได้

2.2 ลักษณะของเสียง (Voice quality)

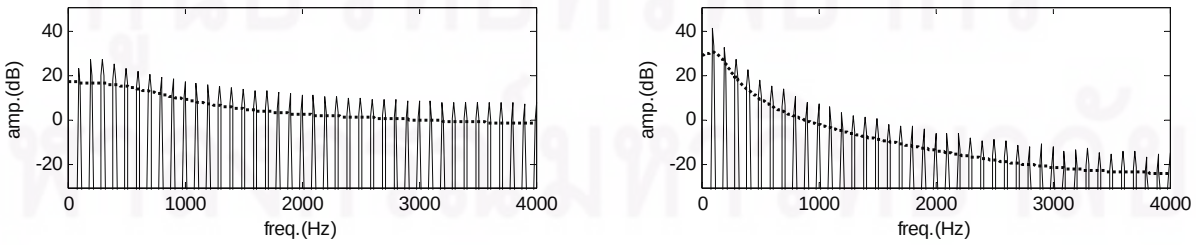
ลักษณะของเสียงเป็นรูปเอกลักษณ์ของแต่ละบุคคล เกิดจากความหลากหลายของกล่องเสียง และทางเดินเสียง ในการศึกษาลักษณะของเสียงสามารถวิเคราะห์ได้ 2 แนวทาง [8] คือ จากความผิดปกติทางด้านการพูด [9] และจากลักษณะการวางตัวของกล่องเสียง (Laryngeal setting) สำหรับการออกเสียงแบบปกติ [10] ซึ่งในการออกเสียงแบบปกติ การสั่นของช่องคอส่งผลให้เกิดรูปร่างของช่องทางเดินเสียงที่แตกต่างกัน เกิดเป็นชนิดลักษณะของเสียงที่แตกต่างกัน รูปที่ 2.6 แสดงตัวอย่างลักษณะของสัญญาณเส้นเสียงในการออกเสียงที่แตกต่างกัน



รูปที่ 2.6 ลักษณะของสัญญาณเส้นเสียงตามการออกเสียงที่แตกต่างกัน

เสียงพูดปกติ (Modal Voice): เป็นการพูดแบบสนทนาโดยทั่วไปจะถือว่าการพูดปกตินี้ เส้นเสียงจะมีความตึงในขณะเปิดปานกลาง แรงการบีบของเส้นเสียงปานกลาง และมีแรงตึงของเส้นเสียงตามความยาวปานกลาง ทำให้เกิดการสั่นของกล่องเสียงมีความเป็นคาบอย่างปกติ โดยไม่เกิดเสียงแทรกที่เกิดจากการที่เส้นเสียงปิดไม่สนิท ในแต่ละบุคคลค่าความถี่ และความตึงของเส้นเสียงจะแตกต่างกันไป

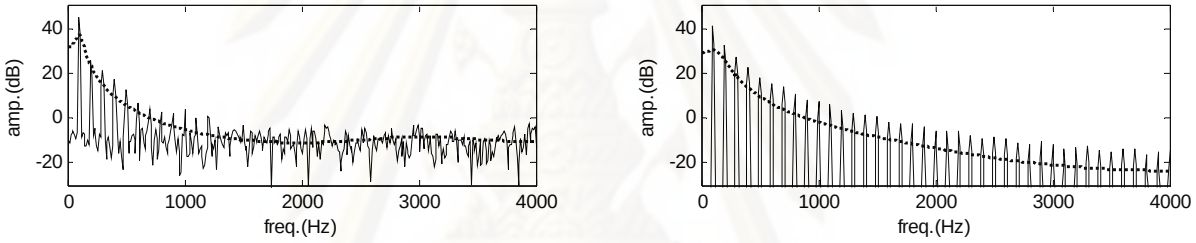
เสียงบีบ (Creaky voice): เกิดจากการเกร็งตัวของเส้นเสียง มีลักษณะเป็นของแข็งขณะสั่นสะเทือน ทำให้มีช่วงเปิดแคบ และลมไหลผ่านช่องเส้นเสียงได้ในปริมาณที่น้อยกว่าปกติ โดยทั่วไปแล้วเสียงพูดลักษณะนี้จะมีค่าความถี่มูลฐานต่ำลง พร้อมกับมีค่าแอมพลิจูดของสัญญาณเส้นเสียงลดลง และอาจจะเกิดการหายไปของเส้นเสียงบางลูกคลื่น (Diplophonic) เมื่อพิจารณาในโดเมนความถี่แล้วจะพบว่าในการออกเสียงบีบ จะมีค่าฮาร์โมนิกส์ที่ 1 (H1) ต่ำกว่าการพูดแบบปกติประมาณ 6 เดซิเบล [11] ปัจจัยที่ผลในการวัดระดับการรับรู้เสียงบีบคือ การแคบลงของสัญญาณเส้นเสียงและ การลดต่ำลงของความถี่มูลฐาน ดังแสดงในรูปที่ 2.7 เปรียบเทียบสเปกตรัมสัญญาณเส้นเสียงลักษณะเสียงบีบ และเสียงปกติ เส้นประแสดงซองของสัญญาณสเปกตรัม



ก. เสียงบีบ ข. เสียงปกติ

รูปที่ 2.7 สเปกตรัมสัญญาณเส้นเสียงลักษณะเสียงบีบ และเสียงปกติ

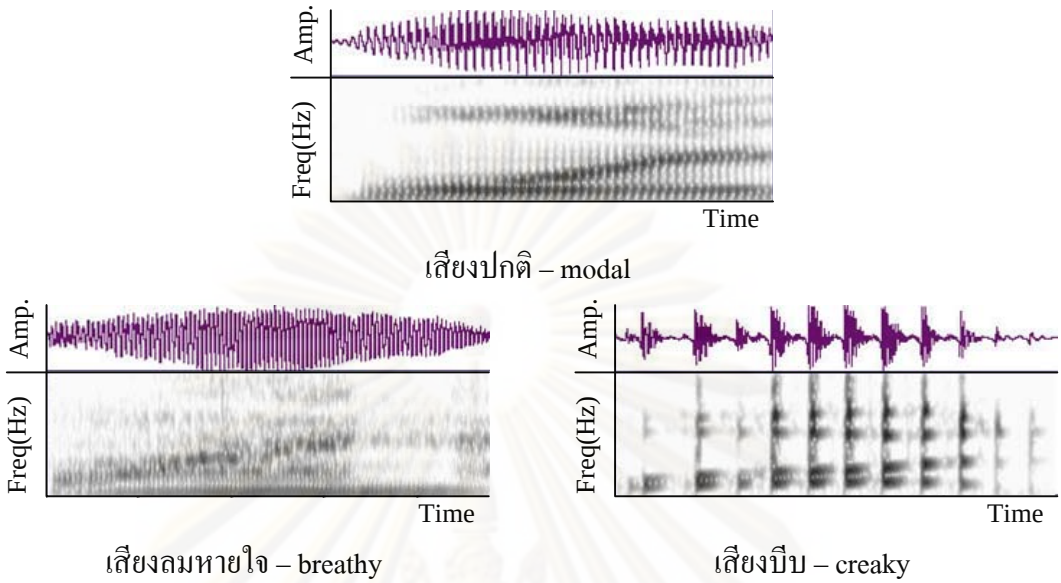
เสียงลมหายใจ (Breathy voice): การออกเสียงประเภทนี้เกิดจากแรงตึงเส้นเสียงอ่อนแรงทำให้มีปริมาณลมหายใจผสมกับเสียงพูด ซึ่งเกิดขณะที่เส้นเสียงปิดไม่สนิทตลอดความยาวขณะที่เส้นเสียงสั่น ทำให้อากาศสามารถแทรกผ่านช่องเปิด สัญญาณในช่วงเปิดของเส้นเสียงจะนานขึ้น สัญญาณพัลส์ของเส้นเสียงมีลักษณะสมมาตร เมื่อพิจารณาในโดเมนความถี่แล้วจะพบว่าในการออกเสียงลมหายใจ จะมีฮาร์โมนิกส์ที่ 1 สูงกว่าฮาร์โมนิกส์ที่ 2 และองค์ประกอบความถี่สูงในสเปกโตรแกรมจะถูกแทนที่ด้วยสัญญาณรบกวนจากลมหายใจ (Aspiration noise) ซึ่งเป็นชนิดหนึ่งของสัญญาณรบกวนลมหมุน ในการวัดระดับการรับรู้ของเสียงประเภทนี้ ระดับของสัญญาณรบกวน จะมีผลต่อการรับรู้มากกว่าค่าความถี่มูลฐานและ รูปร่างของแหล่งกำเนิดเส้นเสียง ดังแสดงในรูปที่ 2.8 เปรียบเทียบสเปกตรัมสัญญาณเส้นเสียงลักษณะเสียงลมหายใจ และเสียงปกติ เส้นประแสดงซองของสัญญาณสเปกตรัม



ก. เสียงลมหายใจ ข. เสียงปกติ

รูปที่ 2.8 สเปกตรัมของสัญญาณเส้นเสียงลักษณะเสียงลมหายใจ และเสียงปกติ

ลักษณะของเสียงสังเคราะห์เช่น เสียงแหบ หรือเสียงกระซิบ มีผลทำให้การฟังข้อความว่ามีความน่าสนใจมากขึ้น และสามารถแสดงอารมณ์การพูดได้ด้วย การให้อารมณ์กับเสียงสังเคราะห์ในการสังเคราะห์เสียงโดยอาศัยแบบจำลองฮิดเดนมาร์คอฟสร้างจากแบบจำลองที่ฝึกฝนจากตัวอย่างเสียงในรูปแบบต่าง ๆ ของการพูด แต่การเก็บตัวอย่างเสียงเพื่อฝึกฝนสร้างแบบจำลองนั้นมีความซับซ้อน และใช้เวลามาก ในขณะที่ลักษณะของเสียงบางชนิดเช่น เสียงแหบ เสียงกระซิบ นั้นสามารถสังเคราะห์ได้โดยการปรับลักษณะของแหล่งกำเนิดเสียง ดังนั้นลักษณะเสียงเหล่านี้จึงสามารถสังเคราะห์ได้โดยใช้วิธีการปรับสัญญาณเสียงที่แหล่งกำเนิดเสียงได้โดยตรง ทำให้ช่วยลดจำนวนเสียงต้นแบบที่ต้องเก็บ และสามารถดัดแปลงให้ใช้ได้กับทุกภาษาอีกด้วย ซึ่งกราฟตัวอย่างแรงดันอากาศ และสเปกตรัมของลักษณะเสียงแบบต่างดังรูปที่ 2.9 แสดงตัวอย่างสเปกตรัมของเสียงพูดจากการออกเสียงคำว่า “วัน” ในลักษณะเสียงลมหายใจ เสียงปกติ และเสียงบีบตามลำดับจากการเก็บตัวอย่างของ Childers [12]

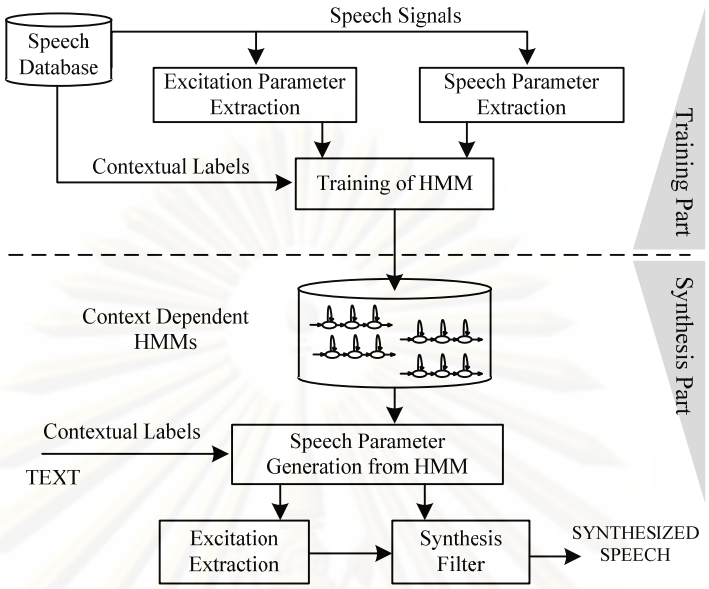


รูปที่ 2.9 สเปกตรัมของลักษณะเสียงแบบต่าง ๆ

ในวิทยานิพนธ์นี้นำความรู้เกี่ยวกับลักษณะของเสียงไปใช้ในการปรับค่าพารามิเตอร์แบบจำลองแหล่งกำเนิดเสียงที่จะเพิ่มเข้าไปในระบบสังเคราะห์เสียงที่นำเสนอ เพื่อให้ระบบสามารถสังเคราะห์เสียงที่มีลักษณะเสียงปกติ เสียงบีบ และเสียงลมหายใจได้ตามทฤษฎี และเสมือนกับลักษณะของเส้นเสียงในการออกเสียงของมนุษย์

3. การสังเคราะห์เสียงโดยอาศัยแบบจำลองฮิดเดนมาร์คอฟ

การสังเคราะห์เสียงโดยอาศัยแบบจำลองฮิดเดนมาร์คอฟ [13] (HMM-based speech synthesis) ซึ่งมีแผนภาพแสดงตามรูปที่ 2.10 ประกอบด้วยสองส่วน คือ กระบวนการฝึกฝน (Training part) ของแบบจำลองฮิดเดนมาร์คอฟ ซึ่งเป็นการสร้างแบบจำลองเสียงจากพารามิเตอร์ที่สกัดจากเสียงที่เตรียมไว้ในฐานข้อมูลเสียง มีการนำสภาวะแวดล้อมของหน่วยเสียง (Contextual factors) มาใช้ในการเรียนรู้ด้วย และ กระบวนการสังเคราะห์เสียง (Synthesis part) โดยสร้างประโยคแบบจำลองฮิดเดนมาร์คอฟ ตามข้อความที่ได้รับเข้ามา และประมาณค่าพารามิเตอร์ที่ใกล้เคียงกับข้อความ โดยทั้งสองขั้นตอนมีรายละเอียดดังนี้



รูปที่ 2.10 การสังเคราะห์เสียงตามแบบจำลองฮิดเดนมาร์คอฟ [4]

3.1 ขั้นตอนการฝึกฝน (Training part)

ขั้นตอนการฝึกฝน เสียงต้นแบบที่จัดเก็บไว้ในฐานข้อมูลเสียง ไฟล์เสียงแต่ละไฟล์จะถูกแบ่งเป็นเฟรม (Frames) และถูกแยกพารามิเตอร์ในแต่ละเฟรมได้แก่ พารามิเตอร์สเปกตรัม (Spectrum parameters) ประกอบด้วยสัมประสิทธิ์เมลเซปสตรอล (Mel-cepstral coefficient) [14] รวมกับ ผลต่าง (Delta) ผลต่างของผลต่าง (delta-delta), พารามิเตอร์กระตุ้น (Excitation parameters) ประกอบด้วย ค่าลอการิทึมของความถี่มูลฐาน และพารามิเตอร์พลวัต (Dynamic feature) และค่าช่วงเวลาของแต่ละสถานะ (State-duration density)

พารามิเตอร์ต่าง ๆ ที่สกัดได้จากฐานข้อมูลเสียง จะถูกนำมาเข้าสู่กระบวนการฝึกฝนเพื่อสร้างแบบจำลองทางสถิติด้วยวิธีการที่แตกต่างกันไปในแต่ละชนิด สำหรับพารามิเตอร์สเปกตรัมเนื่องจากมีค่าเป็นค่าต่อเนื่อง (Continuous values) ทั้งหมดจึงใช้การกระจายตัวแบบเกาซมิติเดียว (Single Gaussian distributions) ในการสร้างแบบจำลองทางสถิติสำหรับพารามิเตอร์กระตุ้นซึ่งประกอบด้วยค่าความถี่มูลฐาน และพารามิเตอร์พลวัต ที่เป็นค่าต่อเนื่องสำหรับเสียงโฆษะ และเป็นค่าไม่ต่อเนื่อง (Discrete values) สำหรับเสียงอโฆษะ จึงต้องใช้การกระจายตัวแบบเกาซหลายมิติ (Multi-dimensional Gaussian distributions) [15] และสำหรับช่วงเวลาของแต่ละสถานะ ซึ่งเป็นช่วงเวลาของสถานะในแต่ละหน่วยเสียง จะใช้แบบจำลองการกระจายตัวแบบเกาซที่มีหลายความแปรผัน (Multivariate Gaussian distribution)

กระบวนการเรียนรู้จะใช้ต้นไม้ตัดสินใจ (Decision trees) ร่วมกับสภาวะแวดล้อมกับหน่วยเสียง ซึ่งประกอบด้วย ลักษณะการออกเสียง ชนิดของตัวอักษร การเน้นคำ ตำแหน่งหน่วยเสียงในระดับต่าง ๆ เสียงวรรณยุกต์ และหน้าที่ของคำ เป็นต้น ซึ่งส่วนประกอบเหล่านี้ได้มาจากการวิเคราะห์ข้อความ (Text analysis) ต้นไม้ตัดสินใจเป็นต้นไม้ประเภทสองกิ่ง (Binary tree) ซึ่งคำถามที่ถามต้องการตอบคำถามจะเป็น ใช่/ไม่ใช่ เช่น “หน่วยเสียงนี้มีเสียงวรรณยุกต์เอกหรือไม่” การที่มีข้อมูลสภาวะแวดล้อมของหน่วยเสียงมากช่วยให้แบบจำลองเสียงมีความหลากหลายและถูกต้องมากขึ้น แต่ยิ่งทำให้ต้องการจำนวนเสียงตัวอย่างมากขึ้นด้วย เพื่อเป็นข้อมูลในการฝึกฝน ในการลดปริมาณการใช้ตัวอย่างเสียงในการฝึกฝน อัลกอริทึมการจัดกลุ่มข้อมูล (Clustering algorithm) [16] จึงถูกนำมาใช้ด้วย

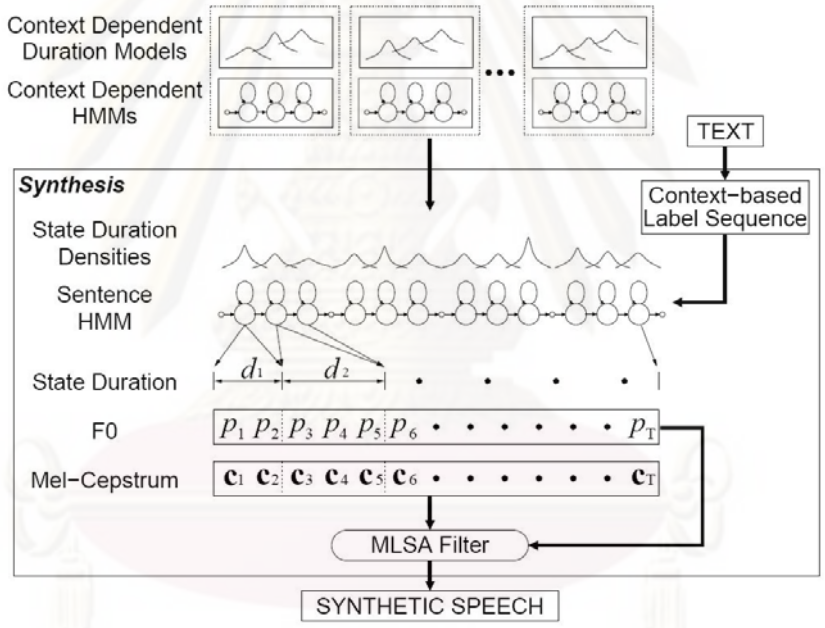
สภาวะแวดล้อมของข้อความในภาษาไทย [17] (Contextual factors in Thai) ที่ใช้ในการสังเคราะห์เสียงโดยอาศัยแบบจำลองฮิดเดนมาร์คอฟแบ่งได้เป็น 13 กลุ่ม แสดงตามตารางที่ 2.1 ซึ่งจะถูกใช้เป็นข้อมูลในการฝึกฝนข้อมูลเสียงในอัลกอริทึมการจัดกลุ่มข้อมูล

ตารางที่ 2.1 สภาวะแวดล้อมของข้อความในภาษาไทย

ระดับของสภาวะแวดล้อม	**ค่าที่ใช้พิจารณา**
ระดับหน่วยเสียง	ชนิดของหน่วยเสียง ตำแหน่งของหน่วยเสียงในพยางค์
ระดับพยางค์	ชนิดของวรรณยุกต์ จำนวนหน่วยเสียงในพยางค์ ตำแหน่งของหน่วยเสียงปัจจุบันในพยางค์
ระดับคำ	ตำแหน่งของพยางค์ปัจจุบันในคำ หน้าที่ของคำ จำนวนพยางค์ในคำ
ระดับวลี	ตำแหน่งของคำปัจจุบันในวลี จำนวนพยางค์ในวลี
ระดับประโยค	ตำแหน่งของวลีปัจจุบันในประโยค จำนวนพยางค์ในประโยค จำนวนคำในประโยค

3.2 ขั้นตอนการสังเคราะห์เสียง (Synthesis part)

ในขั้นตอนนี้ข้อความที่ถูกด้วยตัววิเคราะห์ข้อความแล้วจะถูกป้อนให้แก่ระบบ ซึ่งจะถูกนำไปสร้างเป็นประโยคของแบบจำลองฮิดเดนมาร์คอฟ โดยต้นไม้ตัดสินใจโดยพารามิเตอร์ที่ได้จะค่าความหนาแน่นการกระจายตัว ประกอบด้วยพารามิเตอร์สเปกตรัม และค่าพารามิเตอร์กระตุ้นซึ่งประกอบด้วยสัญญาณพัลส์และสัญญาณรบกวน จากนั้นค่าพารามิเตอร์ทั้งหมดจะถูกสังเคราะห์ออกเป็นเสียงพูดด้วยตัวกรองการประมาณค่าลอการิทึมของสเปกตรัมบนเมลสเกล (Mel-log spectrum approximation filter: MLSA) [14] ระยะเวลาของหน่วยเสียงถูกกำหนดด้วยช่วงเวลาของแต่ละสถานะ แบบจำลองที่ใช้ในการสังเคราะห์เสียงจะใช้ แบบจำลองแหล่งกำเนิดและตัวกรองสัญญาณ แผนภาพการสังเคราะห์เสียงแสดงในรูปที่ 2.11

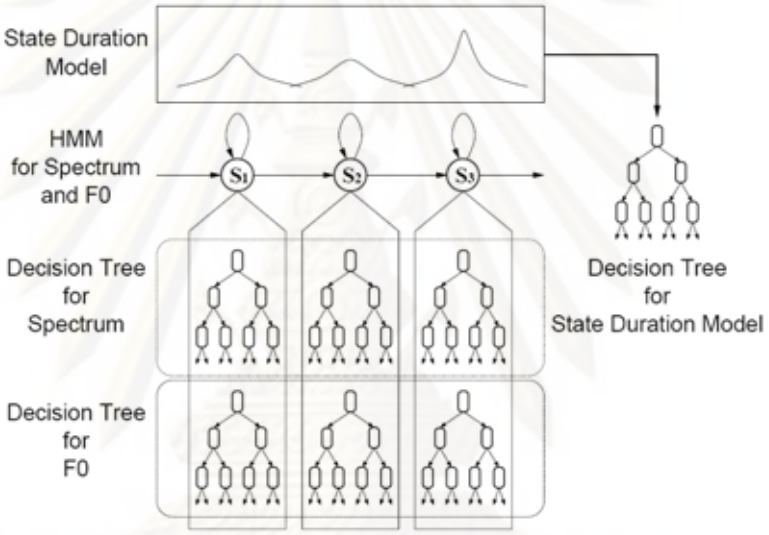


รูปที่ 2.11 ส่วนการสังเคราะห์เสียง [13]

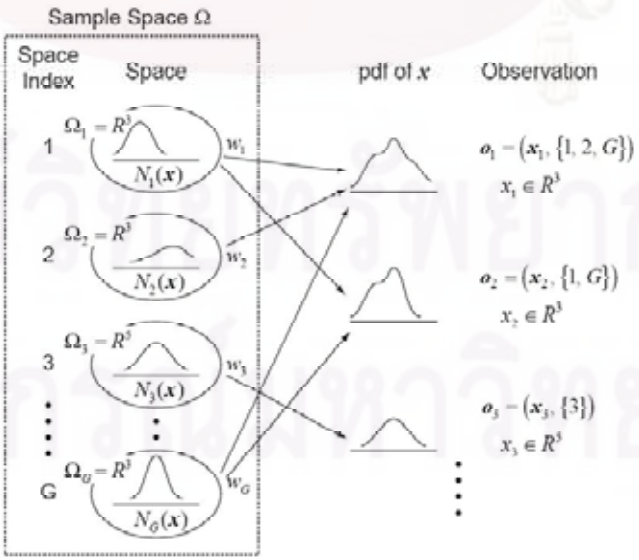
3.2.1 การจัดกลุ่มข้อมูล (Context Clustering)

ในการสร้างแบบจำลองเสียงด้วยแบบจำลองฮิดเดนมาร์คอฟ นอกเหนือจากข้อมูลหน่วยเสียงแล้ว ยังได้นำข้อมูลแวดล้อมของหน่วยเสียงซึ่งมีผลต่อฉันทลักษณ์ของเสียง แต่เนื่องจากข้อมูลที่ใช้ในการฝึกฝนมีจำนวนจำกัด การที่จะสร้างแบบจำลองที่มีความถูกต้องจึงใช้การจัดกลุ่มข้อมูลในต้นไม้ตัดสินใจ โดยพารามิเตอร์สเปกตรัม พารามิเตอร์ความถี่มูลฐาน และช่วงเวลาของหน่วยเสียง จะถูกจัดกลุ่มแยกกันอย่างอิสระดังแสดงในรูปที่ 2.12

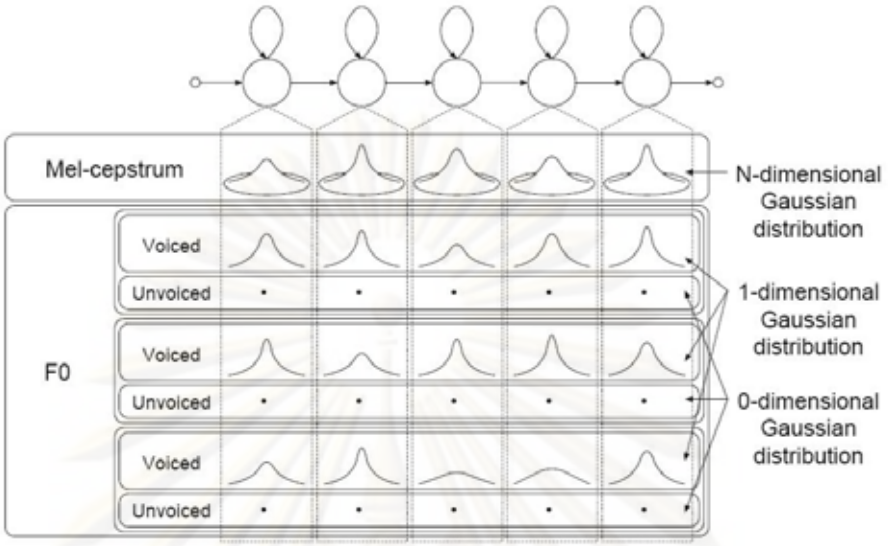
ก่อนที่พารามิเตอร์ของเสียงจะถูกแบ่งกลุ่มจะถูกคำนวณเป็นฟังก์ชันการกระจายตัวความน่าจะเป็น (Probability distribution function, pdf) ซึ่งในกรณีของความถี่มูลฐานที่มีค่าเป็นค่าต่อเนื่องในกรณีที่เป็นเสียงโฆษะ และเป็นค่าไม่ต่อเนื่องในกรณีที่เป็นเสียงอโฆษะจึงจำเป็นต้องใช้ฟังก์ชันการกระจายตัวความน่าจะเป็นหลายสเปส [13] (Multi-space probability distribution function, MSDs) สามารถแสดงได้ดังรูปที่ 2.13 และ รูปที่ 2.14 แบบจำลองฮิดเดนมาร์คอฟแบบฟังก์ชันการกระจายตัวความน่าจะเป็นหลายสเปส สำหรับค่าความถี่มูลฐานในย่านโฆษะจะมี MSDs เป็นแบบ 1 มิติ และเป็นแบบ 0 มิติในย่านอโฆษะ



รูปที่ 2.12 ต้นไม้ตัดสินใจ [13]



รูปที่ 2.13 ฟังก์ชันการกระจายตัวความน่าจะเป็นหลายสเปส [13]



รูปที่ 2.14 แบบจำลองฮิดเดนมาร์คอฟแบบฟังก์ชันการกระจายตัวความน่าจะเป็นหลายสเปส

ในการสร้างต้นไม้ตัดสินใจสำหรับฟังก์ชันการกระจายตัวความน่าจะเป็นหลายสเปสนี้ จะใช้วิธีหาค่าความควรจะเป็นสูงสุด (Maximum likelihood) [18] และใช้ความยาวของลักษณะที่สั้นที่สุด (Minimum description length, MDL) [19] เป็นเงื่อนไขในการหยุดการสร้างต้นไม้ในขั้นตอนการสร้างต้นไม้ตัดสินใจให้เซตของกลุ่ม (Cluster) ซึ่งเป็นผลของการแบ่งกลุ่มแทนด้วย $\mathbf{S} = \{S_1, S_2, ..., S_i, ..., S_M\}$ ใน S_i จะประกอบด้วยเซตของการกระจายตัวความน่าจะเป็นจากพารามิเตอร์เสียงจำนวน G "สเปส" (Space) จากตัวอย่างเช่นรูปที่ 2.14 ค่าความถี่มูลฐานมีจำนวนสเปส G=3 สำหรับค่าลอการิทึมของความถี่มูลฐาน และพารามิเตอร์พลวัต (Dynamic feature) ค่าลอการิทึมของค่าความควรจะเป็น (Log likelihood, L) แสดงได้ดังสมการ (1)

$$L = -\sum_{s \in S} \sum_{g=1}^{G} \frac{1}{2} (n_g (\log(2\pi) + 1) + \log\left|\Sigma_{sg}\right| - 2\log w_{sg}) \sum_{t \in T(O,g)} \Upsilon_t(s,g) \qquad (1)$$

เมื่อ

g เป็นลำดับของสเปสในกลุ่มตัวอย่าง S_i

w_{sg} แทนค่าน้ำหนักของสเปส g ในกลุ่ม s

$T(O,g)$ แทนเซตของเวลา t ซึ่งทำให้เซตลำดับของเวกเตอร์สำรวจ (Observation vector) o_t มีสเปสลำดับ g รวมอยู่ด้วย

$\Upsilon_t(s,g)$ แทนความน่าจะเป็นของสเปส g ของกลุ่ม s ณ เวลา s

Σ_{sg} คือเมตริกซ์แปรปรวนร่วม (Covariance matrix) ของสเปส g ของกลุ่ม s ในกรณีที่เป็นสเปสมิติศูนย์ซึ่งมีค่า $\log\left|\Sigma_{sg}\right|$ เป็น 0

โดยจากค่าความควรจะเป็น L จะสามารถหาค่าความยาวของลักษณะ (Description length, DL) l ได้ตามสมการ (2)

$$\begin{aligned} l = & \sum_{s\in S}\sum_{g=1}^{G}\frac{1}{2}(n_g(\log(2\pi)+1)+\log\left|\Sigma_{sg}\right|-2\log w_{sg})\sum_{t\in T(O,g)}\Upsilon_t(s,g) \\ & +\left(\sum_{s\in S}\sum_{g=1}^{G}\frac{1}{2}(2n_g+1)\right)\cdot\left(\log\sum_{s\in S}\sum_{g=1}^{G}\sum_{t\in T(O,g)}\Upsilon_t(s,g)\right) \end{aligned} \tag{2}$$

สมมุติให้ค่าความยาวของลักษณะเป็น l' เมื่อกลุ่ม S_i ถูกแบ่งเป็นสองกลุ่ม S_{i+} และ S_{i-} จะสามารถวัดค่าการเปลี่ยนแปลงค่าความยาวลักษณะ ∂l ได้ตามสมการ (3)

$$\begin{aligned} \partial l &= l'-l \\ &= \sum_{s\in\{S_{i+},S_{i-}\}}\sum_{g=1}^{G}\frac{1}{2}(\log\left|\Sigma_{sg}\right|-2\log w_{sg})\sum_{t\in T(O,g)}\Upsilon_t(s,g) \\ &\quad -\sum_{s\in\{S_i\}}\sum_{g=1}^{G}\frac{1}{2}(\log\left|\Sigma_{sg}\right|-2\log w_{sg})\sum_{t\in T(O,g)}\Upsilon_t(s,g) \\ &\quad +\left(\sum_{g=1}^{G}\frac{1}{2}(2n_g+1)\right)\cdot\left(\log\sum_{s\in S}\sum_{g=1}^{G}\sum_{t\in T(O,g)}\Upsilon_t(s,g)\right) \end{aligned} \tag{3}$$

ถ้าค่า $\partial l<0$ แสดงว่าสามารถแบ่งกลุ่มได้ แต่ถ้า $\partial l\geq 0$ ให้หยุดการแบ่งกลุ่ม

3.2.2 การสร้างค่าพารามิเตอร์จากแบบจำลองฮิดเดนมาร์คอฟ

พารามิเตอร์ที่ได้จากต้นไม้ตัดสินใจจากทุกสถานะ จะเรียงตัวเป็นเวกเตอร์ค่าสังเกต (Observation vector) เมื่อเราพิจารณาแบบจำลองฮิดเดนมาร์คอฟแบบความหนาแน่นต่อเนื่อง (Continuous density Hidden Markov Model, CDHMM) ที่มีจำนวน N สถานะ ในแต่ละสถานะมี M ตัวผสม (Mixtures) และเวกเตอร์ค่าสังเกต $O=[o_1,o_2,...,o_T]$ จะได้ค่าความเหมือน (Likelihood) ของค่าสังเกต o_t ในสถานะ S_j คือ

$$\begin{aligned} \mathrm{P}(o_t\mid q_j=S_j) &= \sum_{m=1}^{M}c_{jm}b_{jm}(o_t) \\ &= \sum_{m=1}^{M}c_{jm}N(o_t,\mu_{jm}\Sigma_{jm}) \end{aligned} \tag{4}$$

เมื่อพิจารณาเวกเตอร์ค่าสังเกตโดยไม่สนใจค่าพารามิเตอร์พลวัตใช้แต่ค่าพารามิเตอร์สถิต นั่นคือ $o_t=c_t$ ดังนั้นค่าสังเกตที่ทำให้ $p(O\mid\lambda)$ มีค่ามากที่สุด จะเป็น $\mathrm{P}(O\mid\lambda)=\mathrm{P}(o\mid Q,\lambda)\mathrm{P}(Q\mid\lambda)$ โดยที่ Q คือลำดับของสถานะและตัวผสม $Q=(q,i),\ q=\{q_1,q_2,...,q_T\},\ i=\{i_1,i_2,...,i_T\}$ จะได้เป็น

$$\begin{aligned}\log \mathrm{P}(O \mid Q, \lambda) &= \log \prod_{t=1}^{T} b_{qt}, i_t(o_t) \\ &= -\frac{1}{2}(O-\mu)'\Sigma^{-1}(O-\mu) - \frac{1}{2}\sum_{t=1}^{T}\log\left|\Sigma_{qt}\right| - \frac{1}{2}TD\log 2\pi\end{aligned} \tag{5}$$

ซึ่ง

$\mu = [\mu'_{q_1,i_1}, \mu'_{q_2,i_2}, ..., \mu'_{q_T,i_T}]'$ เป็นเวกเตอร์ค่าเฉลี่ย

$\Sigma = diag[\Sigma_{q_1,i_1}, \Sigma_{q_2,i_2}, ..., \Sigma_{q_T,i_T}]$ เป็นเวกเตอร์ค่าความแปรปรวนร่วม

T เป็นค่าความยาวของเวกเตอร์ค่าสังเกตในเฟรม

D เป็นมิติของพารามิเตอร์สถิต

จากสมการที่ (5) ค่า $\log \mathrm{P}(O \mid Q, \lambda)$ จะมีค่าสูงที่สุด เมื่อค่าอนุพันธ์ของสมการมีค่าเป็น 0 จะได้ว่า

$$\frac{\partial(\log P(O \mid Q, \lambda))}{\partial c} = -\Sigma^{-1}c + \Sigma^{-1}\mu = 0 \tag{6}$$

แสดงว่าค่าที่สูงที่สุดจะหาได้ก็ต่อเมื่อ $c = \mu$ นั่นคือ ลำดับของค่าสังเกตที่ใกล้เคียงที่สุดคือเวกเตอร์ค่าเฉลี่ย ซึ่งเป็นอิสระกับค่าความแปรปรวนร่วม Σ แต่ผลที่ได้จากการคำนวณยังขาดความต่อเนื่อง ณ จุดเปลี่ยนแปลงเฟรม เพื่อแก้ปัญหาข้างต้น จึงใช้พารามิเตอร์พลวัตเข้ามาร่วมพิจารณาด้วย [20, 21] เพื่อให้การเปลี่ยนแปลงมีความต่อเนื่องมากขึ้น

เมื่อพิจารณาพารามิเตอร์พลวัตด้วย จะใช้หน้าต่างค่าอัตราการเปลี่ยนแปลงในช่วงเวลา W โดยค่าสังเกตใหม่จะกลายเป็น $o_t = [c'_t, \Delta c'_t, \Delta^2 c'_t]$ เมื่อ $\Delta^{(n)}c'_t = \sum_{\tau=-L(n)}^{L(n)} w^{(n)}(\tau)c_{T+\tau}$; $(n = 0, 1, 2)$ ทำให้ได้นิยามของเวกเตอร์ลักษณะสำคัญใหม่คือ

$$\begin{aligned}\log \mathrm{P}(O \mid Q, \lambda) &= \frac{1}{2}(O-\mu)'\Sigma^{-1}(O-\mu) - \frac{1}{2}\log\left|\Sigma_{qt}\right| - \frac{3TD}{2}\log 2\pi \\ &= \frac{1}{2}(Wc-\mu)'\Sigma^{-1}(Wc-\mu) - \frac{1}{2}\log\left|\Sigma_{qt}\right| - \frac{3TD}{2}\log 2\pi \\ &= \frac{1}{2}\varepsilon(c) - \frac{1}{2}\log\left|\Sigma_{qt}\right| - \frac{3TD}{2}\log 2\pi\end{aligned} \tag{7}$$

ซึ่ง

ค่า μ และค่า Σ ตามที่นิยามไว้ในสมการ (5)

$\varepsilon(c) = (Wc-\mu)'\Sigma^{-1}(Wc-\mu)$

$W = [w_1, w_2, ..., w_T]'$

$w_t = [w^{(0)}{}_t, w^{(1)}{}_t, w^{(2)}{}_t]$;

เมื่อหาค่าต่ำที่สุดโดย $\partial \log P(O \mid Q, \lambda)/\partial c = 0$ จะได้

$$(W'\Sigma^{-1}W)c - W'\Sigma^{-1}\mu = 0 \quad (8)$$

หรือสามารถเขียนให้อยู่ในรูปของ

$$\begin{aligned} Rc &= r \\ R &= W'\Sigma^{-1}W \\ r &= W'\Sigma^{-1}\mu \end{aligned} \quad (9)$$

จากสมการที่ (9) สามารถหาค่าพารามิเตอร์ได้โดยใช้วิธีของ Tokuda [20]

3.2.3 ตัวกรองสัญญาณในการสังเคราะห์เสียง (Synthesis filter)

ในการสังเคราะห์เสียงโดยอาศัยแบบจำลองฮิดเดนมาร์คอฟนี้ ตัวกรองสัญญาณ $D(z)$ ที่ใช้ในการสังเคราะห์เสียงจะคำนวณจากค่าสัมประสิทธิ์เมลเซปตรัมที่ได้จากผลของการฝึกฝนแบบจำลองเสียง สมการของตัวกรองสัญญาณแสดงดังสมการ (10)

$$D(z) = \exp F(z) = \exp \sum_{m=0}^{M} b(m)\Phi_m(z) \quad (10)$$

โดยที่ M คือจำนวนลำดับของสัมประสิทธิ์เมลเซปตรัม และ $c(m)$ เป็นสัมประสิทธิ์เมลเซปตรัมลำดับที่ m

$$b(m) = \begin{cases} c(m) & m = M \\ c(m) - \alpha b(m+1) & 0 \le m \le M \end{cases}$$
$$\Phi_m(z) = \begin{cases} 1 & m = 0 \\ \dfrac{(1-\alpha^2)z^{-1}}{1-\alpha z^{-1}} \tilde{z}^{-(m-1)} & m \ge 1 \end{cases} \quad (11)$$

ให้ค่า α เป็นค่ามาตราส่วนความถี่การได้ยิน (Auditory frequency scale) ซึ่งมีค่าตามตารางที่ 2.2

ตารางที่ 2.2 มาตราส่วนความถี่การได้ยิน α

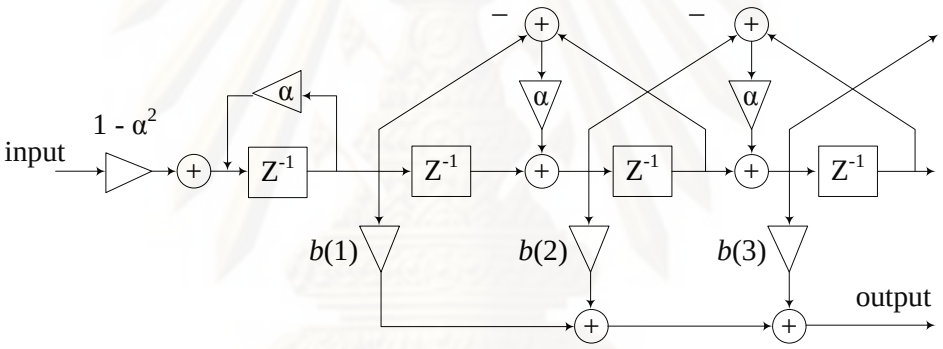
ค่า Sampling Frequency	มาตราส่วนเมล (Mel scale)	มาตราส่วนบาร์ค (Bark scale)
8 kHz	0.31	0.42
10 kHz	0.35	0.47
12 kHz	0.37	0.50
16 kHz	0.42	0.55

สมการตัวกรองสัญญาณที่ใช้ในการสังเคราะห์เสียงตามสมการที่ (10) ไม่อยู่รูปของฟังก์ชันตรรกยะ (Rational function) ทำให้ยากในการนำมาใช้ จึงได้นำการประมาณค่าลอการิทึมของสเปกตรัมบนเมลสเกล (Mel log spectrum approximation, MLSA) มาใช้ในการประมาณค่าตัว

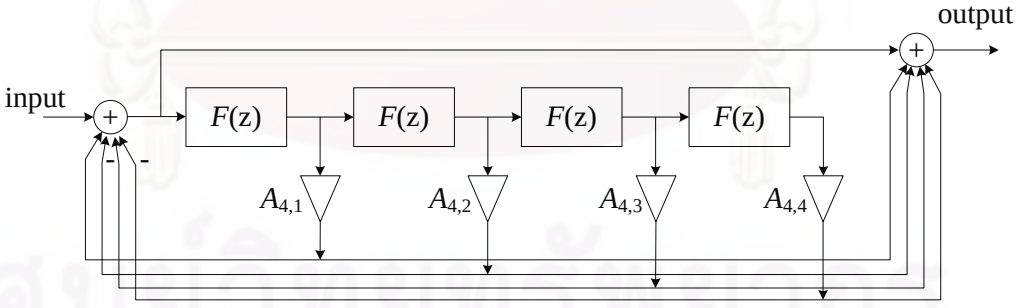
กรองสัญญาณ ซึ่งมีความแม่นยำสูง โดยการประมาณค่าฟังก์ชันเอ็กโพเนนเชียลเชิงซ้อน (Complex exponential function, $\exp F(z)$) จะอยู่ในรูปของฟังก์ชันตรรกยะ (Rational function, $R_L(F(z))$) ตามสมการ (12)

$$D(z) = \exp F(z) \approx R_L(F(z)) = \frac{1+\sum_{l=1}^{L} A_{L,1}\omega^l}{1+\sum_{l=1}^{L} A_{L,1}(-\omega)^l} \tag{12}$$

$R_L(F(z))$ แทนการประมาณฟังก์ชันตรรกยะของฟังก์ชันเอ็กโพเนนเชียลเชิงซ้อน $\exp\omega$ โดยที่ $A_{L,1}$ เป็นสัมประสิทธิ์การประมาณค่าลำดับที่ L ซึ่งสามารถแสดงแผนภาพการประมาณตัวกรองสัญญาณได้ดังรูปที่ 2.15 ที่มีค่าลำดับเป็น L = 4 และใช้ค่าสัมประสิทธิ์ $A_{4,1}$ ตามตารางที่ 2.3



(a) Basic filter $F(z)$



(b) $R_L(F(z)) \approx D(z)$, $L = 4$

รูปที่ 2.15 ตัวกรองการประมาณค่าลอการิทึมของสเปกตรัมบนเมลสเกล [13]

ตารางที่ 2.3 สัมประสิทธิ์การประมาณค่าของ $R_4(F(z))$, $L = 4$

l	$A_{4,l}$
1	4.999273×10^{-1}
2	1.067005×10^{-1}
3	1.170221×10^{-2}
4	5.656279×10^{-4}

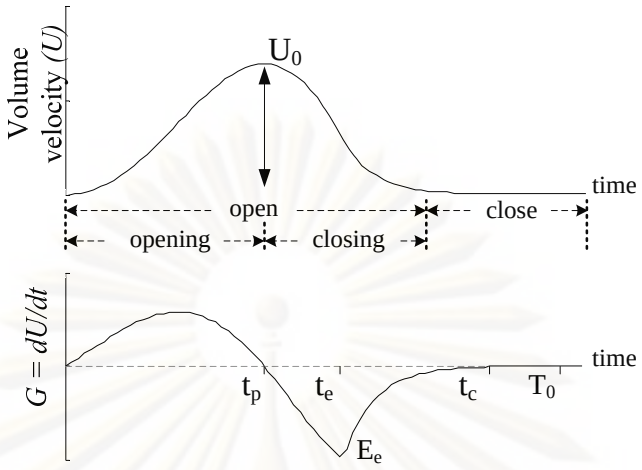
การสังเคราะห์เสียงโดยอาศัยแบบจำลองฮิดเดนมาร์คอฟนี้จะเป็นการงาน และระบบอ้างอิงในการปรับระบบการสังเคราะห์เสียงให้สามารถใช้สัญญาณเส้นเสียง และสัญญาณรบกวนทางลมหายใจ เป็นสัญญาณกระตุ้นได้

4. แบบจำลองแอลเอฟแบบแปลง (Transformed LF-model)

แบบจำลองแอลเอฟ [22] เป็นแบบจำลองที่ใช้ในการจำลองอนุพันธ์ของสัญญาณเส้นเสียงทางเวลา (time-domain derivative glottal waveform) การประมาณสัญญาณเสียงประกอบด้วยสองส่วนคือส่วนเปิดและส่วนปิด ซึ่งแทนด้วยค่าพารามิเตอร์ทางเวลา 4 ค่า ได้แก่ t_p, t_e, t_a, และ t_c แทนค่าสูงสุดของช่องทางเดินเสียง (glottal flow) ค่าสูงสุดทางลบของอนุพันธ์ของช่องทางเดินเสียง ค่าคงที่ทางเวลาของเส้นโค้งเอกซ์โพเนนเชียลของส่วนที่สองของแบบจำลอง และค่าคงที่ที่ช่วงที่เส้นเสียงปิดสมบูรณ์ ตามลำดับ โดยค่าพารามิเตอร์ถูกประมาณโดยตัวกรองย้อนกลับ (Inverse filtering) ของเสียงที่บันทึก สัญญาณเสียงถูกปรับเปลี่ยนค่าพารามิเตอร์ให้เกิดลักษณะของเสียงได้หลากหลายแตกต่างกัน ดังนั้นแบบจำลองแอลเอฟแบบแปลง จะทำให้ 4 ค่าพารามิเตอร์ เหลือเพียง 1 ค่า (Rd) ซึ่งทำให้ง่ายต่อการเปลี่ยนลักษณะของเสียงรูปที่ 2.16 แสดงสัญญาณแบบจำลองแอลเอฟและค่าพารามิเตอร์เส้นเสียง ภาพบนคือสัญญาณเสียงเส้นเสียงและภาพล่างแสดงสัญญาณอนุพันธ์ของช่องทางเดินเส้นเสียง โดยให้ E_e แทนค่าขนาดของช่วงปิดเส้นเสียงของสัญญาณกระตุ้น

ค่าพารามิเตอร์ของสัญญาณเส้นเสียง $g(t)$ สามารถแสดงได้ดัง สมการ (13)

$$\begin{aligned} g(t) &= E_0 e^{at} \sin(w_g t) && ; \ 0 < t < t_e \\ &= \frac{E_e}{e t_a}\left[e^{-e(t-t_e)} - e^{e(t_c - t_e)} \right] && ; \ t_e < t < t_c < T_0 \end{aligned} \qquad (13)$$



รูปที่ 2.16 สัญญาณเส้นเสียงสร้างโดยแบบจำลองแอลเอฟ [22]

เมื่อค่าพารามิเตอร์ทางเวลา t_p, t_e, t_a, และ t_c แทนค่าสูงสุดของช่องทางเดินเสียง (glottal flow) ค่าสูงสุดทางลบของอนุพันธ์ของช่องทางเดินเสียง ค่าคงที่ทางเวลาของเส้นโค้งเอ็กโปเนนเชียลของส่วนที่สองของแบบจำลอง (Exponential curve of the second segment of the model) และค่าคงที่ที่ช่วงที่เส้นเสียงปิดสมบูรณ์ (Complete glottal closure) ตามลำดับ และให้ E_e แทนค่า ขนาดของช่วงปิดเส้นเสียงของสัญญาณกระตุ้น ลักษณะรูปร่างของส่วนแรกของแบบจำลองแอลเอฟเป็นสัญญาณไซน์ (Sine) ที่เพิ่มขึ้นแบบเอ็กโปเนนเชียล ซึ่งอธิบายอนุพันธ์ของช่องทางเดินเส้นเสียง (Derivative glottal flow) บนช่วงเปิดเส้นเสียงถึงจุดสูงสุดทางลบของสัญญาณเสียง ลักษณะรูปร่างของส่วนที่สองของแบบจำลองแอลเอฟเป็นช่วงปิด แสดงฟังก์ชันเอ็กโปเนนเชียลแบบลดข้อกำหนดของค่าพารามิเตอร์แสดงดังสมการที่ (14)

$$
\begin{aligned}
& 0 \le t_p \le t_e \le t_c; \quad t_a \le 0 \\
& \textstyle\int_0^T g(t)\,dt = 0 \\
& \omega_g = \frac{\pi}{t_p} \\
& \varepsilon t_a = 1 - e^{-\varepsilon(t_c - t_e)} \\
& E_0 = -\frac{E_e}{e^{\alpha t}\sin(\omega_g t_e)}
\end{aligned}
\qquad (14)
$$

ค่าพารามิเตอร์ในแบบจำลองแอลเอฟแบบแปลง สามารถอธิบายได้ในรูปของค่าพารามิเตอร์ทางเวลาที่ถูกนอร์มอลไลซ์ดังสมการที่ (15) ซึ่งมีค่าผิดพลาดที่สูงสุดคือ 1.5 เดซิเบล ที่ค่า Rd = 2.7

$$R_a = t_a / T_0; \qquad R_g = T_0 / 2t_p; \qquad R_k = (t_e - t_p)/t_p$$
$$R_d = (1/0.11)(0.5 + 1.2R_k)(R_k / 4R_g + R_a) \tag{15}$$

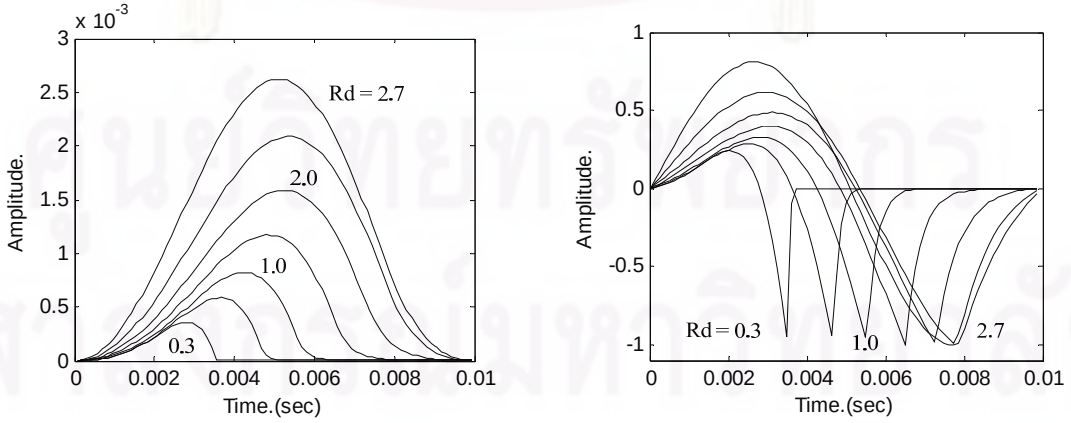
ในการสร้างสัญญาณเส้นเสียง ค่าพารามิเตอร์ทางเวลาถูกประมาณจาก ค่า Rd ด้วยสมการที่ (16) โดยสัญลักษณ์ p หมายถึงค่าพารามิเตอร์ที่ประมาณค่านั้น ๆ จากค่า Rd

$$R_{ap} = (-1 + 4.8R_d)/100$$
$$R_{kp} = (22.4 + 11.8R_d)/100 \tag{16}$$

ในการสังเคราะห์สัญญาณเส้นเสียงด้วยแบบจำลองแอลเอฟ การแปลงค่าตัวแปร Rd มีความสัมพันธ์แสดงได้ในตารางที่ 2.4 เมื่อกำหนดค่าความถี่มูลฐานมีค่าเป็น 100 Hz และกราฟสัญญาณเส้นเสียง 1 ลูกคลื่นในโดเมนเวลาแสดงในรูปที่ 2.17

ตารางที่ 2.4 ความสัมพันธ์พารามิเตอร์แอลเอฟ และค่า Rd

Rd	**Ra (%)**	**Fa (Hz)**	**Rk (%)**	**R (%)g**	**OQ (%)**
0.3	0.44	3600	26	179	35
0.5	0.71	1590	28.3	137	47
0.7	2.36	674	30.7	118	55.5
1.0	3.8	420	34.2	103	65
1.4	5.7	280	39.0	95	73
2.0	8.6	185	46.0	93.5	78
2.7	12.0	133	54.3	98	79

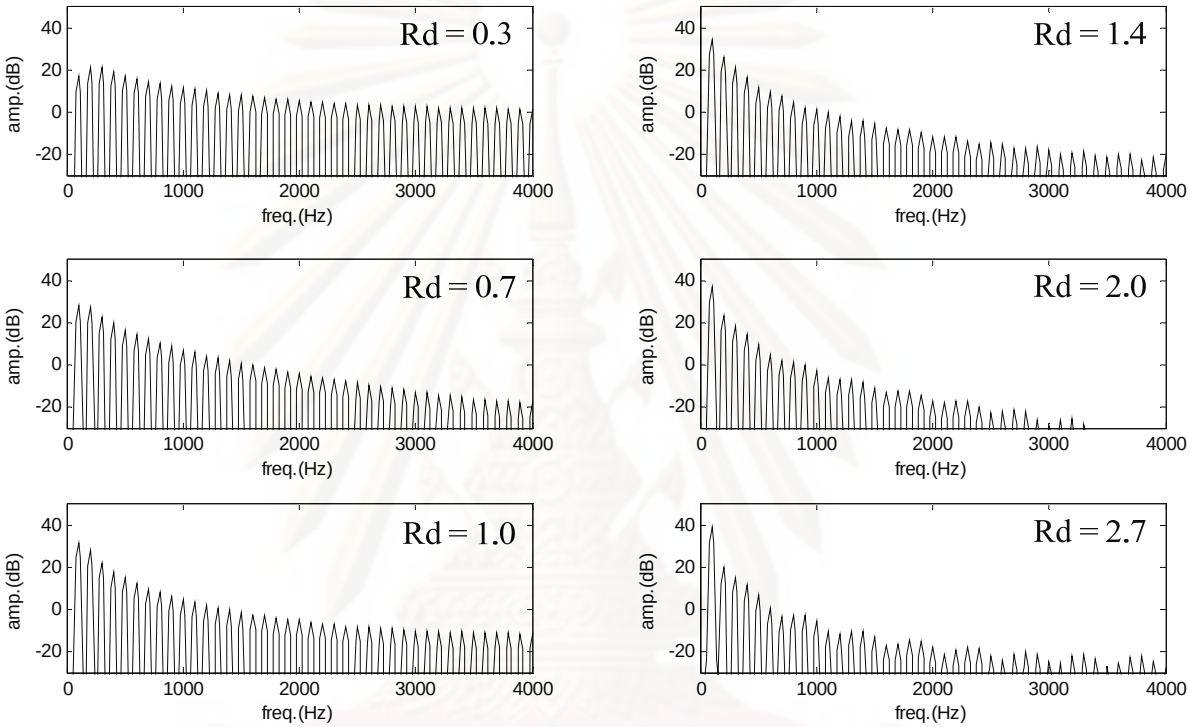


ก. สัญญาณเส้นเสียง ข. อนุพันธ์ของสัญญาณเส้นเสียง

รูปที่ 2.17 กราฟโดเมนเวลาเมื่อปรับค่า Rd

โดยที่ $F_a = F_0 / (2\pi R_a)$ เป็นค่าดัชนีการสั่นของสเปกตรัม (Spectral tilt) และค่า $OQ = (1+R_k)/2R_g$ เป็นค่าอัตราส่วนช่วงเปิด (Open quotient)

ในโดเมนความถี่ สามารถแสดงสเปกตรัมของแบบจำลองแอลเอฟเมื่อกำหนดค่าความถี่คงที่เป็น 100 Hz. ดังรูปที่ 2.18



รูปที่ 2.18 สเปกตรัมกราฟแอลเอฟเมื่อเปลี่ยนค่า Rd

การจำลองสัญญาณเส้นเสียงด้วยแบบจำลองแอลเอฟแบบแปลง จะใช้ในการวิเคราะห์การแยกสัญญาณเส้นเสียงออกจากสัญญาณเสียง และยังใช้สำหรับสัญญาณกระตุ้นในการสังเคราะห์เสียงอีกด้วย

เอกสารและงานวิจัยที่เกี่ยวข้อง

การสังเคราะห์เสียงโดยอาศัยแบบจำลองฮิดเดนมาร์คอฟเป็นสังเคราะห์เสียงที่ประสบความสำเร็จมากในช่วงสิบปีที่ผ่านมา เนื่องจากการที่ระบบเก็บหน่วยของเสียงเป็นแบบจำลองทางสถิติ ทำให้มีขนาดของระบบเล็ก เสียงที่สังเคราะห์มีความเป็นธรรมชาติ และสามารถปรับรูปแบบเสียงสังเคราะห์ได้โดยการปรับแบบจำลองทางสถิติ ระบบการสังเคราะห์นี้ถูกนำไปพัฒนาใช้กับหลายภาษา [23-25] รวมถึงภาษาไทย [17] และเนื่องจากความสามารถในการปรับแบบจำลองทางสถิติในขั้นตอนการสังเคราะห์เสียงจึงมีงานวิจัยเกี่ยวกับการเปลี่ยนเสียงพูดระหว่างเสียงผู้ชาย และ

ผู้หญิง [26-28] การใส่อารมณ์ในเสียงสังเคราะห์ [29, 30] ซึ่งทำโดยการเพิ่มตัวอย่างเสียงชาย หญิง หรือ อารมณ์ของการพูด เช่น ดีใจ เศร้า หรือโกรธ ในการเรียนรู้แบบจำลองเสียง

อย่างไรก็ดีถึงแม้เสียงที่สังเคราะห์ได้จากระบบการสังเคราะห์เสียงโดยอาศัยแบบจำลองฮิดเดนมาร์คอฟนี้ จะมีความเป็นธรรมชาติของเสียงสังเคราะห์สูงก็ตาม แต่ก็ยังลักษณะเป็นเสียงหึ่ง (Buzzy noise) เนื่องจากระบบนี้ใช้กระแสพัลส์ (Pulse-train) หรือสัญญาณรบกวนสีขาว (White noise) เป็นสัญญาณกระตุ้นในแบบจำลองแหล่งกำเนิด และตัวกรองสัญญาณ งานวิจัยที่เสนอการปรับปรุงคุณภาพเสียงสังเคราะห์เช่น ในช่วงแรกของงานวิจัยมีการใช้สัญญาณรบกวนแบบผสม (Mixed excitation) [31] สัญญาณกระตุ้นแบบผสมหลายช่วง (Multi-band excitation) [32] และวิธีสัญญาณฮาร์โมนิกส์และสัญญาณรบกวน (Harmonic plus noise) [33] ซึ่งทำการรวมสัญญาณเสียงรบกวนเข้ากับกระแสพัลส์ทำให้คุณภาพเสียงดีขึ้น ต่อมามีการใช้สัญญาณหลงเหลือ (Residual signal) จากตัวกรองสัญญาณย้อนกลับในการเรียนรู้เป็นสัญญาณกระตุ้น [34] มีผลทำให้ได้เสียงสังเคราะห์ที่คล้ายกับเสียงต้นแบบ และลดความเสียงหึ่งในเสียงสังเคราะห์ลงได้

ยังมีงานวิจัยที่มุ่งปรับปรุงคุณภาพของเสียงสังเคราะห์โดยใช้สัญญาณแหล่งกำเนิดเส้นเสียง ซึ่งจากงานวิจัยของ Childers [35] พบว่าสัญญาณเส้นเสียงสามารถเพิ่มความเป็นธรรมชาติของเสียงสังเคราะห์ได้ ในงานวิจัยเกี่ยวกับการสังเคราะห์เสียงด้วยแบบจำลองทางสถิติ Cabral [36] เสนอการรวมแบบจำลองแหล่งกำเนิดเส้นเสียง (Glottal source) เข้าไว้กับระบบการสังเคราะห์เสียงด้วยแบบจำลองทางสถิต โดยใช้แบบจำลองแอลเอฟ (LF-model) แทนสัญญาณพัลส์ และสเปกตรัมของสัญญาณกระตุ้นถูกทำให้เรียบโดยตัวกรองตัวแบบโพส (Post-filter) ต่อมา Cabral [37] ได้เสนอการแยกสเปกตรัมเส้นเสียง (Glottal spectral separation, GSS) ซึ่งเป็นวิธีการหาค่าแหล่งกำเนิดเส้นเสียงแทนการใช้ตัวกรองแบบโพส ที่เสนอในงานก่อนหน้า และพบว่าคุณภาพเสียงดีขึ้น แต่ทั้งนี้ Cabral ยังไม่ได้ทำแบบจำลองสัญญาณเส้นเสียงมาใช้ในการสังเคราะห์เสียงโดยอาศัยแบบจำลองฮิดเดนมาร์คอฟ นอกจากการใช้สัญญาณแหล่งกำเนิดเส้นเสียงเป็นสัญญาณกระตุ้นจะช่วยเพิ่มคุณภาพของเสียงสังเคราะห์แล้ว ยังสามารถปรับเปลี่ยนเสียงสังเคราะห์เป็นลักษณะต่าง ๆ (Voice quality)

แบบจำลองสัญญาณเส้นเสียงที่ใช้เป็นสัญญาณกระตุ้นในการสังเคราะห์เสียงจะมีสองกลุ่มหลักคือ แบบจำลองที่อ้างอิงอากาศพลศาสตร์และกลไกการทำงานของเส้นเสียง [38-40] แบบจำลองกลุ่มนี้จะถูกควบคุมโดยพารามิเตอร์ที่เกี่ยวกับลักษณะทางกายภาพของระบบการออกเสียงของมนุษย์ แต่เนื่องจากการวิเคราะห์พารามิเตอร์ที่ใช้มีความซับซ้อน จึงเกิดการประมาณแบบจำลองที่มากเกินไป จึงต้องมีการปรับแต่งค่าบางค่าเพื่อให้สามารถสังเคราะห์สัญญาณเส้นเสียงที่เหมือนจริง [41] และแบบจำลองเส้นเสียงอีกกลุ่มเป็นแบบจำลองที่ศึกษารูปร่างสัญญาณเส้นเสียง

โดยตรง [8, 42, 43] ในแบบจำลองกลุ่มนี้จะศึกษาพารามิเตอร์จากลักษณะการทำงานของเส้นเสียงและการเปลี่ยนแปลงปริมาณอากาศที่ไหลผ่าน โดยวิเคราะห์จากการออกเสียงจริง และใช้ตัวกรองสัญญาณย้อนกลับ ในการเลือกแบบจำลองเพื่อใช้สำหรับการสังเคราะห์เสียงนี้มีปัจจัยที่ต้องพิจารณาคือ ความซับซ้อนของแบบจำลอง กับความสะดวกในการวิเคราะห์ค่าพารามิเตอร์ ในงานวิจัยนี้ใช้แบบจำลองเส้นเสียงที่อ้างอิงจากสัญญาณเสียง โดยใช้แบบจำลองแอลเอฟ [22] เนื่องจากเป็นแบบจำลองที่ประมาณอนุพันธ์ของสัญญาณเส้นเสียง และมีการศึกษาและใช้งานอย่างแพร่หลายในงานวิจัยด้านการสังเคราะห์เสียง

งานวิจัยที่วิเคราะห์สัญญาณรบกวน เช่น Hu [44] เสนอวิธีการลดสัญญาณรบกวนในระบบโดยเวฟเลท (Wavelet denoising) ซึ่งเป็นการสร้างสัญญาณใหม่จากสัญญาณที่ถูกรบกวน ในการสกัดสัญญาณรบกวนในสัญญาณเส้นเสียง และ Hu ได้สรุปว่าวิธีการใช้รากฐานที่ดีที่สุด (Best-basis) ให้ผลการสร้างสัญญาณใหม่ดีที่สุด

การศึกษางานวิจัยเกี่ยวกับการสังเคราะห์เสียงโดยอาศัยแบบจำลองฮิดเดนมาร์คอฟที่สามารถกำหนดสัญญาณจากเส้นเสียงและสัญญาณรบกวนลมหายใจ จะแบ่งขั้นตอนการทำงานเป็นสองส่วน โดยขั้นตอนแรกคือการสกัดค่าแบบจำลองเส้นเสียง และระดับของสัญญาณรบกวนจากฐานข้อมูลเสียง ขั้นต่อมาคือการนำค่าที่สกัดได้เข้าขั้นตอนการเรียนรู้ด้วยแบบจำลองฮิดเดนมาร์คอฟ ซึ่งการการศึกษางานวิจัยที่เกี่ยวข้องพบว่า งานวิจัยที่เกี่ยวข้องนั้นได้เสนอวิธีการสังเคราะห์เสียงในขอบเขตที่แตกต่างกันเช่นในงานของ Cabral เป็นการสังเคราะห์เสียงซึ่งใช้แบบจำลองแอลเอฟเป็นสัญญาณกระตุ้น แต่ยังไม่ได้นำพารามิเตอร์ที่ได้ใช้ร่วมกับแบบจำลองฮิดเดนมาร์คอฟ และยังมีการศึกษาระดับของสัญญาณรบกวน ในงานวิจัยของ Hu ซึ่งใช้ wavelet packet analysis ในการวิเคราะห์ระดับของสัญญาณรบกวน ในงานวิจัยนี้เสนอการนำสัญญาณแอลเอฟแบบแปลง และระดับของสัญญาณรบกวนเข้าเรียนรู้ด้วยแบบจำลองฮิดเดนมาร์คอฟ ส่วนในการศึกษาระดับเสียงรบกวนในสัญญาณแหล่งกำเนิดเส้นเสียง ในงานวิจัยนี้เสนอวิธีหาค่าฟังก์ชันการหาจุดเปลี่ยน

บทที่ 3
ขั้นตอนการดำเนินการวิจัย

ขั้นตอนการดำเนินการวิจัย

ในบทนี้จะกล่าวถึงขั้นตอนการดำเนินการวิจัยในวิทยานิพนธ์ ซึ่งประกอบด้วย

1 ขั้นตอนการออกแบบระบบสังเคราะห์เสียง

เพื่อการศึกษาและพัฒนาวิธีการปรับสังเคราะห์เสียงที่มีแหล่งกำเนิดจากเส้นเสียง และเพื่อให้เสียงสังเคราะห์มีความเป็นธรรมชาติมากขึ้นตามวัตถุประสงค์ของวิทยานิพนธ์ ในขั้นตอนนี้จึงกล่าวถึงขั้นตอนการออกแบบระบบสังเคราะห์เสียง เครื่องมือที่ใช้ในการวิจัย และเพื่อสร้างระบบสังเคราะห์เสียงภาษาไทยตามวัตถุประสงค์ของวิทยานิพนธ์ จึงอธิบายชุดข้อมูลฝึกฝนจากฐานข้อมูลเสียงภาษาไทย สำหรับใช้ในการสร้างแบบจำลองเสียงของระบบสังเคราะห์เสียงที่นำเสนอ และระบบสังเคราะห์เสียงอ้างอิง

2 ขั้นตอนการสร้างระบบสังเคราะห์เสียง

เพื่อพัฒนาระบบสังเคราะห์เสียงที่สามารถสังเคราะห์เสียงที่มีลักษณะเสียงต่าง ๆ ได้ โดยอาศัยแบบจำลองแหล่งกำเนิดเส้นเสียง และระดับสัญญาณเสียงรบกวนจากลมหายใจตามวัตถุประสงค์ของวิทยานิพนธ์ ในขั้นตอนนี้อธิบายการวิเคราะห์ค่าพารามิเตอร์แหล่งกำเนิดเส้นเสียงและระดับสัญญาณเสียงรบกวนจากลมหายใจจากชุดข้อมูลฝึกฝน และอธิบายขั้นตอนการฝึกฝนแบบจำลองเสียงที่นำเสนอ และการสังเคราะห์เสียง

3 ขั้นตอนการประเมินระบบสังเคราะห์เสียงที่นำเสนอ

เพื่อการประเมินความสามารถและประสิทธิภาพของระบบสังเคราะห์เสียงที่นำเสนอวิทยานิพนธ์นี้ จึงทดสอบและประเมินผลของระบบการสังเคราะห์เสียงที่นำเสนอเปรียบเทียบกับเสียงจากระบบสังเคราะห์เสียงอ้างอิง

ขั้นตอนการออกแบบระบบสังเคราะห์เสียง

1. เครื่องมือที่ใช้ในการวิจัย

1 MATLAB 7.6 [45] ใช้เป็นเครื่องมือใช้เพื่อพัฒนาและทดสอบระบบ

2 เครื่องมือแบบจำลองฮิดเดนมาร์คอฟ (Hidden Markov Model toolkit: HTK) [46]

3 การสังเคราะห์เสียงโดยอาศัยแบบจำลองฮิดเดนมาร์คอฟ (HMM-based speech synthesis version 2.0) [47]

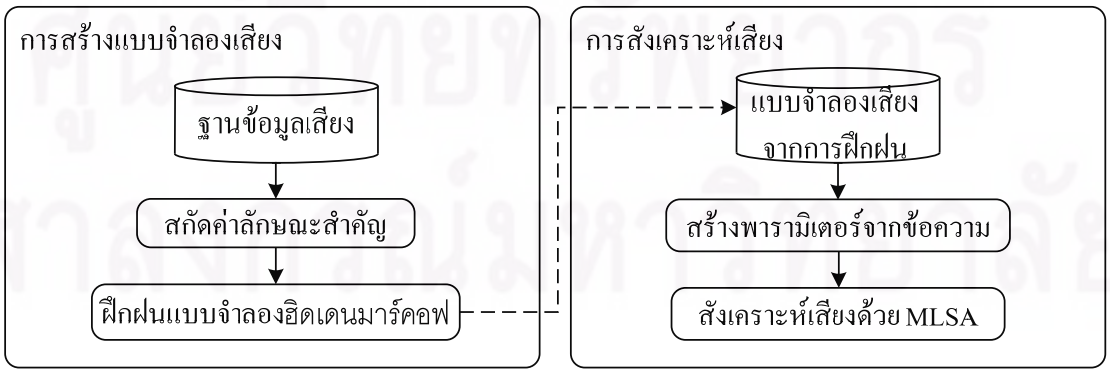
4 เครื่องมือทีเคเคอาพรัด (TKK Aparat) [48] เป็นเครื่องมือสำหรับใช้ศึกษาและประมาณค่าพารามิเตอร์แอลเอฟ

2. ฐานข้อมูลเสียงที่ใช้พัฒนาระบบ

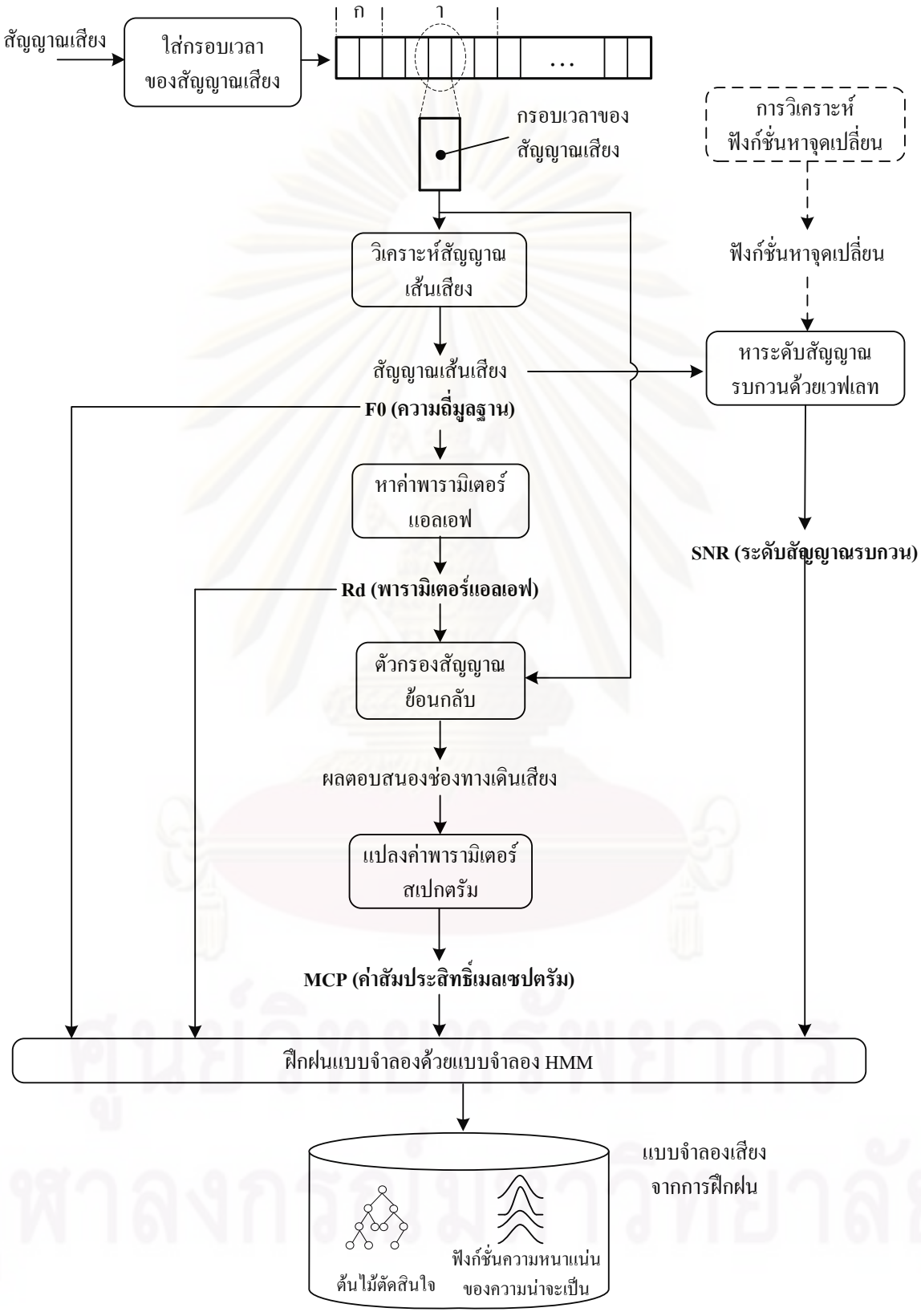
ฐานข้อมูลเสียงที่ใช้ในวิทยานิพนธ์นี้ เป็นฐานข้อมูลเสียงแบบระบุชนิดข้อมูลสำหรับการพัฒนาระบบการสังเคราะห์เสียงในภาษาไทยแล้ว (Thai tagged speech corpus for speech synthesis version 1: TSync1) [49] ซึ่งเป็นฐานข้อมูลหน่วยเสียงสมดุล (Phonetically balanced) จัดทำโดยศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ หรือเนคเทค โครงสร้างของชุดข้อมูลที่ระบุไว้ในฐานข้อมูลเสียงอยู่จัดเก็บในรูปแบบของเอ็กเอ็มแอล (Extensible Markup Language: XML) ที่จะบอกโครงสร้างของไวยากรณ์ และข้อมูลแวดล้อมของข้อความคือ ย่อหน้า, ประโยค, คำ, หน้าที่ของคำ, เสียงวรรณยุกต์ และหน่วยเสียง ชุดข้อความที่ใช้บันทึกในฐานข้อมูลเสียงนี้ถูกเลือกมาจากฐานข้อมูลชุดคำ ORCHID [50] ซึ่งครอบคลุมคำศัพท์ในภาษาไทยทั้งหมดประมาณ 5,000 คำ ชุดเสียงทั้งหมดถูกบันทึกด้วยวิธีการอ่านข้อความเป็นเสียงผู้หญิงที่มีทักษะทางด้านการอ่านเป็นอย่างดี ในการพัฒนาระบบการสังเคราะห์เสียงในวิทยานิพนธ์นี้เลือกประโยค 1500 ประโยคแบบสุ่มจากฐานข้อมูลเสียง

3. การออกแบบระบบสังเคราะห์เสียงที่นำเสนอ

จากการออกแบบระบบ สามารถแสดงภาพรวมการระบบสังเคราะห์เสียงที่นำเสนอได้ดังรูปที่ 3.1 ซึ่งภาพรวมของระบบที่นำเสนอประกอบด้วย การสร้างแบบจำลองเสียง และการสังเคราะห์เสียง ซึ่งมีรายละเอียดดังนี้



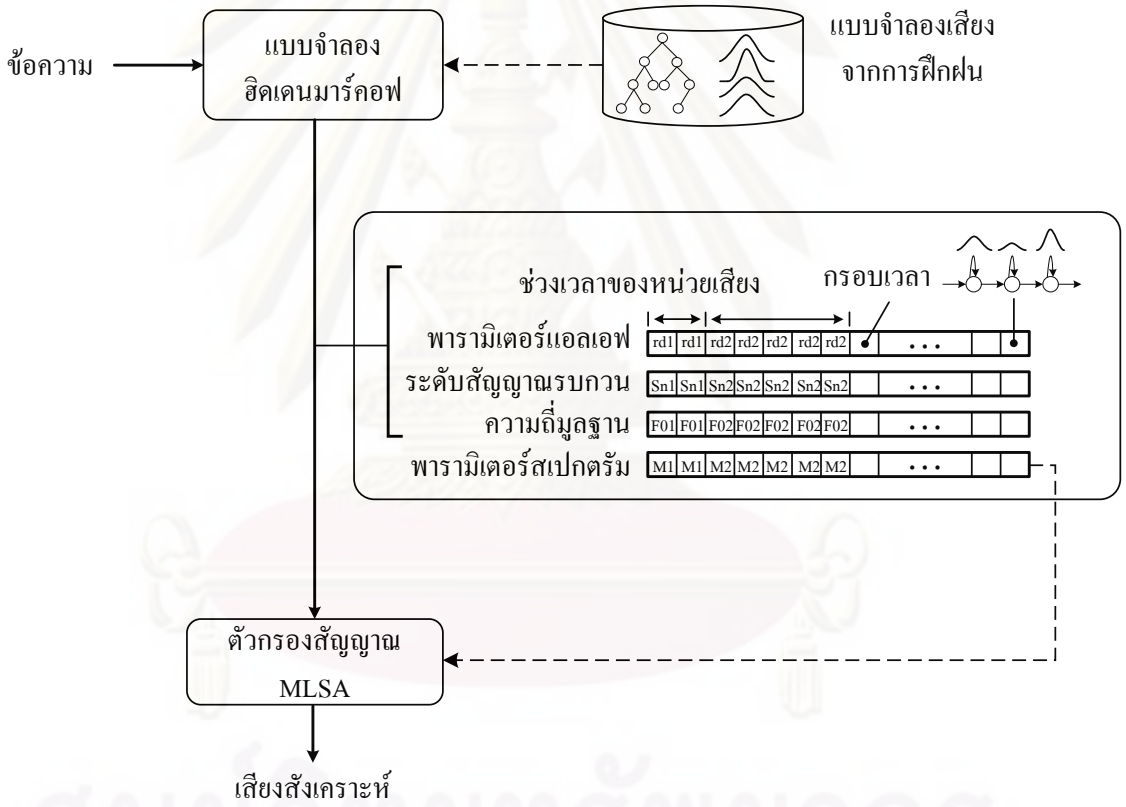
รูปที่ 3.1 ภาพรวมของระบบที่นำเสนอ



รูปที่ 3.2 แผนภาพขั้นตอนการสร้างแบบจำลองเสียง

1 ขั้นตอนการสร้างแบบจำลองเสียงประกอบด้วยการสกัดค่าลักษณะสำคัญจากฐานข้อมูลเสียงได้แก่ ค่าความถี่มูลฐาน ค่าพารามิเตอร์แอลเอฟ ระดับสัญญาณเสียงรบกวน และค่าสัมประสิทธิ์เมลเซปตรัม โดยจะอธิบายรายละเอียดในบทที่ 4 จากนั้นนำพารามิเตอร์ที่ได้จากการวิเคราะห์ไปฝึกฝนแบบจำลองเสียงด้วยกรอบงานการสังเคราะห์เสียงโดยอาศัยแบบจำลองฮิดเดนมาร์คอฟ ดังรูปที่ 3.2

2 ขั้นตอนการสังเคราะห์เสียง ใช้ข้อความจากชุดข้อมูลทดสอบ และใช้แบบจำลองเสียงที่เป็นผลลัพธ์จากขั้นตอนการฝึกฝน มาสร้างเสียงสังเคราะห์ในลักษณะเสียงต่าง ๆ ตามทฤษฎีกระบวนการทำให้เกิดเสียง ดังรูปที่ 3.3



รูปที่ 3.3 แผนภาพขั้นตอนการสังเคราะห์เสียง

ขั้นตอนการสร้างระบบสังเคราะห์เสียง

1. ระบบการสังเคราะห์เสียงอ้างอิง

ระบบการสังเคราะห์เสียงที่ใช้เป็นระบบอ้างอิง ฝึกฝนจากค่าพารามิเตอร์ที่ใช้ค่าเวกเตอร์ลักษณะสำคัญที่สกัดแต่ละเฟรมของสัญญาณเสียง ประกอบด้วยเมลเซปสตรัม (Mel-cepstrum) ค่าผลต่าง และค่าผลต่างของผลต่าง 39 ค่า และค่าลอการิทึมของความถี่มูลฐานพร้อมกับค่าผลต่างของ

ค่าลอการิทึมของความถี่มูลฐาน และผลต่างของผลต่างของค่าลอการิทึมของความถี่มูลฐาน พร้อมกับค่าช่วงเวลาของหน่วยเสียง ได้มาจากการระบุขอบเขตหน่วยเสียงด้วยมือจากฐานข้อมูลเสียง เข้าฝึกฝนด้วยกรอบงานการสังเคราะห์เสียงโดยอาศัยแบบจำลองฮิดเดนมาร์คอฟ [47] ซึ่งแบบจำลองฮิดเดนมาร์คอฟมี 7 สถานะเรียงตัวจากซ้ายไปขวา ไม่มีการกระโดดข้าม หรือย้อนกลับ ตามที่ได้อธิบายไว้แล้วในบทที่ 2

2. ระบบการสังเคราะห์เสียงที่นำเสนอ

ระบบการสังเคราะห์เสียงที่นำเสนอ แบบจำลองเสียงฝึกฝนจากค่าพารามิเตอร์ที่ใช้ค่าเวกเตอร์ลักษณะสำคัญที่สกัดแต่ละเฟรมของสัญญาณเสียง ประกอบด้วย เมลเซปสตรัม (Mel-cepstrum) ค่าผลต่าง และค่าผลต่างของผลต่าง 39 ค่า ค่าลอการิทึมของความถี่มูลฐานพร้อมกับค่าผลต่างของค่าลอการิทึม และผลต่างของผลต่างของค่า ลอการิทึมของความถี่มูลฐาน พร้อมกับค่าช่วงเวลาของหน่วยเสียง นอกจากนี้ในระบบสังเคราะห์เสียงที่นำเสนอใช้ค่าพารามิเตอร์ Rd (LF-model) และค่าสัญญาณเสียงรบกวน ซึ่งสกัดจากฐานข้อมูลเสียง เพิ่มจากค่าพารามิเตอร์ที่ใช้ในระบบสังเคราะห์เสียงอ้างอิง โดยปรับค่าพารามิเตอร์มิติของเวกเตอร์ลักษณะสำคัญในการฝึกฝนแบบจำลองฮิดเดนมาร์คอฟให้รองรับกับค่าพารามิเตอร์ Rd และค่าสัญญาณเสียงรบกวนที่เพิ่มขึ้น ซึ่งจะกล่าวถึงรายละเอียดระบบที่นำเสนอในบทที่ 4

ขั้นตอนการประเมินระบบสังเคราะห์เสียงที่นำเสนอ

ระบบสังเคราะห์เสียงที่นำเสนอในวิทยานิพนธ์นี้มีจุดมุ่งหมาย เพื่อนำแบบจำลองสัญญาณเส้นเสียงมาใช้เป็นสัญญาณกระตุ้น และสามารถสังเคราะห์เสียงที่มีลักษณะเสียงบีบ เสียงปกติ เสียงลมหายใจได้ ดังนั้นในการประเมินประสิทธิภาพของระบบสังเคราะห์เสียงที่นำเสนอนี้จึงต้องทำการทดสอบกับผู้ฟังเพื่อวัดระดับการรับรู้ของลักษณะเสียงต่าง ๆ จากเสียงสังเคราะห์ ซึ่งมีรายละเอียดการประเมินระบบสังเคราะห์เสียงที่นำเสนอดังต่อไปนี้

1. กลุ่มผู้ฟัง

กลุ่มผู้ฟังที่เป็นผู้ประเมินผลคุณภาพเสียงสังเคราะห์ของวิทยานิพนธ์นี้ เป็นอาสาสมัครจากนักศึกษามหาวิทยาลัย 15 คน มีอายุระหว่าง 18 -25 ปี ซึ่งสามารถฟังและสื่อสารได้อย่างปกติ และไม่มีประสบการณ์หรือรู้จักการแบ่งแยกลักษณะของเสียง และระบบเสียงสังเคราะห์ทั้งสองมาก่อนมาก่อน

2. การเลือกชุดเสียงเพื่อทดสอบ

วิทยานิพนธ์นี้เลือกการสร้างประโยคที่สุ่มเลือกจากฐานข้อมูลเสียงเพื่อสามารถใช้เปรียบเทียบกับเสียงจริงที่เป็นเป้าหมายได้ และมีหน่วยเสียงครอบคลุมทั้งเสียงที่เป็นโฆษะและเสียงอโฆษะ เพื่อสอดคล้องกับต้นแบบที่ต้องการสร้างเสียงสังเคราะห์ที่สามารถจำลองได้จากพารามิเตอร์ของแหล่งกำเนิดเส้นเสียงค่า Rd และเสียงที่สามารถแสดงการสังเคราะห์จากสัญญาณรบกวนได้ชัดเจน

3. การประเมินผลของระบบการสังเคราะห์เสียง

ในการประเมินคุณภาพ และลักษณะของเสียงสังเคราะห์ จากระบบสังเคราะห์เสียงที่นำเสนอ เปรียบเทียบกับระบบสังเคราะห์เสียงอ้างอิง จะประเมินในสองลักษณะคือ การประเมินความเป็นธรรมชาติของเสียงสังเคราะห์ และการประเมินลักษณะของเสียงสังเคราะห์ด้วยการวิเคราะห์ ทั้งในด้านการรับรู้จากผู้ฟัง และอ้างอิงจากทฤษฎีการกำเนิดเสียงพูด ซึ่งมีรายละเอียดดังนี้

3.1 ขั้นตอนการทดสอบ และลงความเห็นจากผู้ฟัง

ในการวัดความเป็นธรรมชาติ และการรับรู้ลักษณะของเสียงจากผู้ฟัง มีรายละเอียดดังนี้

3.1.1 สภาวะแวดล้อม

1. การทดสอบกระทำในสภาพแวดล้อมแบบเปิด เสียงรบกวนจากภายนอกน้อย
2. การฟังเสียงสังเคราะห์ จะฟังผ่านหูฟังยี่ห้อฝ PHILIPS รุ่น SHG 2100

3.1.2 การทดสอบและลงความเห็น

ในการทดสอบ จะทำการทดสอบผู้ฟังทีละคนซึ่งมีขั้นตอนในการทดสอบดังนี้

1. ผู้ฟังจะได้รับการอธิบายเกี่ยวกับลักษณะของเสียง ได้ฟังเสียงต้นแบบที่บันทึกจากเสียงจริง เพื่อให้ผู้ฟังเรียนรู้ลักษณะของเสียง ทั้งในแบบเสียงปกติ เสียงบีบ และเสียงลมหายใจ ประเภทละ 5 เสียง โดยแต่ละเสียงจะระบุลักษณะของเสียงเพื่อให้ผู้ฟังทราบถึงประเภทเสียง
2. จากนั้นผู้ฟังจะได้รับทราบถึงจุดประสงค์การทดสอบแต่ละชุดทดสอบ
3. จากนั้นผู้ฟังจะได้ฟังเสียงจากชุดทดสอบทีละเสียงเป็นลำดับ และขอฟังใหม่ได้
4. ผู้ฟังจะถูกขอให้ลงความเห็นเชิงบังคับ โดยให้ตอบคำถามตามวัตถุประสงค์การทดสอบ
5. จากนั้นนำผลที่ได้จากการประเมินมาวิเคราะห์ และสรุปผลการทดสอบ

3.2 การประเมินความเป็นธรรมชาติของเสียงสังเคราะห์

วิทยานิพนธ์นี้เสนอการสังเคราะห์เสียงอิงแบบจำลองฮิดเดนมาร์คอฟโดยกำหนดสัญญาณจากแหล่งกำเนิดเส้นเสียง และสัญญาณรบกวนลมหายใจได้โดยตรง เพื่อให้ระบบสามารถสร้างเสียงสังเคราะห์ให้มีความเป็นธรรมชาติ โดยการใช้สัญญาณกระตุ้นจากแหล่งกำเนิดเส้นเสียง และระดับสัญญาณเสียงรบกวน แทนการใช้สัญญาณพัลส์ตามการสังเคราะห์เสียงอิงแบบจำลองฮิดเดนมาร์คอฟในระบบอ้างอิง ดังนั้นเพื่อประเมินความเป็นธรรมชาติของเสียงสังเคราะห์จึงพิจารณาจากการวัดคะแนนแบบซีซีอาร์ (CCR) [51] โดยประเมินผลระบบการสังเคราะห์เสียงใช้การเปรียบเทียบเสียงสังเคราะห์ของระบบสังเคราะห์เสียง 2 ระบบ ทดสอบจากผู้ฟัง 15 คน ซึ่งผู้ฟังแต่ละคนจะได้ฟังประโยคที่สังเคราะห์ 15 ชุด ชุดละสองเสียงสังเคราะห์ซึ่งเป็นประโยคเดียวกันสังเคราะห์ด้วยระบบที่นำเสนอและระบบอ้างอิง (สุ่มข้อความจากฐานข้อมูลเสียง Tsync1 และไม่ได้ใช้ในชุดข้อมูลการฝึกฝน) และจะตั้งชื่อประโยคเป็นประโยค A และประโยค B แบบสุ่ม โดยผู้ทดสอบจะไม่ทราบว่าเป็นประโยคที่สังเคราะห์มาจากระบบใด ในการทดสอบผู้ฟังจะได้ฟังเสียงทีละชุดทดสอบทีละชุดโดนฟังประโยค A ตามด้วยประโยค B จากนั้นผู้ฟังให้คะแนนความรู้สึกความเป็นธรรมชาติเทียบกับเสียงในประโยคเดียวกันที่สังเคราะห์จากระบบอ้างอิง โดยมีระดับคะแนนความชอบแสดงดังตารางที่ 3.1

ตารางที่ 3.1 ตัวเลือกคะแนนระดับความชอบ

คะแนน	ระดับความชอบ
1	ดีกว่า (Better)
0	เหมือนกัน (Similar)
-1	แย่กว่า (Worse)

นอกจากการเปรียบเทียบคะแนนซีซีอาร์ ซึ่งใช้วัดระดับความชอบของความเป็นเสียงธรรมชาติของเสียงสังเคราะห์ทั้งสองระบบแล้ว เพื่อเปรียบเทียบความแตกต่างของคะแนนความเป็นธรรมชาติของทั้งสองระบบ วิทยานิพนธ์นี้จึงทำการทดสอบนัยสำคัญทางสถิติ (Significance Test) ของค่าคะแนนซีซีอาร์ของทั้งสองระบบ โดยใช้ทดสอบเครื่องหมาย (Sign test) [52] ซึ่งเป็นการวัดความเชื่อมั่นจากชุดทดสอบแบบไม่มีรูปแบบ (Non-parametric test) โดยวัดความเชื่อมั่นจากทิศทางของเครื่องหมายของคู่ทดสอบ ประกอบด้วยเครื่องหมายบวก (+) เครื่องหมายลบ (-) และศูนย์ (0) ที่ได้มาจากผลต่างของคะแนนจากคู่ทดสอบ เพื่อวัดความน่าจะเป็นที่มีระดับความเชื่อมั่นอย่างมีนัยสำคัญทางสถิติที่ 0.05 สมมุติฐานนี้ความแตกต่างกันอย่างมีนัยสำคัญก็ต่อเมื่อ $h \le \alpha$ โดยที่ h มีค่าตามสมการ (17)

$$h = F(x \mid n, p) = \sum_{i=0}^{x} \binom{n}{i} p^i q^{(n-i)} I_{(0,1,\ldots,n)}(i) \quad (17)$$

โดย n คือจำนวนคู่ตัวอย่างทั้งหมดที่ผลต่างไม่เป็นศูนย์ โดย x เป็นจำนวนเครื่องหมายที่มีจำนวนน้อยกว่าที่เกิดขึ้นจากคู่ตัวอย่างทั้งหมด p คือค่าความเป็นไปได้ของการเกิดเครื่องหมายบวกและลบ ซึ่งมีค่าเท่ากับ 0.5 และ $q = 1 - p$ และ $I_{(0,1,\ldots,n)}(i)$ คือฟังก์ชันตัวชี้วัด (Indicator function) เพื่อให้แน่ใจว่าค่า x อยู่ระหว่าง $0,1,\ldots,n$ เท่านั้น

ชุดประโยคที่ใช้ในการวัดค่าความเป็นธรรมชาติ แสดงได้ดังตารางที่ 3.2 ซึ่งแต่ละประโยคจะถูกสังเคราะห์ในลักษณะเสียงปกติ (ไม่มีการปรับพารามิเตอร์ใด ๆ) จากทั้งระบบที่นำเสนอและระบบอ้างอิง

ตารางที่ 3.2 ประโยคที่ใช้ในการวัดความความเป็นธรรมชาติของเสียงสังเคราะห์

ประโยคทดสอบ	**ข้อความ**
ประโยคที่ 1	โครงสร้างจุลภาพพื้นผิวของผิวชุบเคลือบเรียบ
ประโยคที่ 2	ทำการชุบเคลือบทองแดงแบบไม่ใช้ไฟฟ้าด้วยสภาวะการชุบเคลือบต่าง ๆ แสดงดังตารางที่ 2
ประโยคที่ 3	จะเห็นได้ว่าการตรวจคำในภาษาไทย ซึ่งเขียนติดกันเป็นพืดนั้น มีอยู่ทางเดียว คือ จากจุดเริ่มต้น
ประโยคที่ 4	เนื่องจากภาษาไทยเขียนเรียงติดกันไปหมด
ประโยคที่ 5	นอกจากนั้นแล้ว คณะวิจัยก็กำลังพิจารณาที่จะนำผลพลอยได้จากนี้มาใช้ในอีกโครงการหนึ่ง คือการ "พิจารณา" ภาษาไทย หรือ และอังกฤษ โดยอัตโนมัติ
ประโยคที่ 6	2.2.2 คำที่มีตัวควบกล้ำ เช่น มากกว่า วิธีของผู้วิจัยบอกได้ว่าไม่ใช่คำผิด
ประโยคที่ 7	โดยได้ทำการออกแบบโปรแกรมซอฟต์แวร์ส่วนหนึ่ง ซึ่งเป็นโปรแกรมพื้นฐานเบื้องต้นที่ใช้ในการประมวลผลข้อมูลภาพถ่ายดาวเทียม
ประโยคที่ 8	ความสำเร็จของศูนย์เทคโนโลยีอีเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาตินั้นอาจจะกล่าวได้ว่าเกิดจากความร่วมมือในลักษณะของเครือข่ายระหว่างศูนย์ฯ มหาวิทยาลัย และภาคธุรกิจเอกชน
ประโยคที่ 9	ตลอดจนการเปิดโอกาสให้บรรดานักวิจัยทั้งในภาครัฐบาลและเอกชนได้มีโอกาสมาพบปะแลกเปลี่ยนความรู้และประสบการณ์เกี่ยวกับการวิจัยทางด้านเทคโนโลยีอีเล็กทรอนิกส์และคอมพิวเตอร์
ประโยคที่ 10	การพัฒนาหัววัดน้ำตาลกูลโคสและการประยุกต์ใช้งาน

ประโยคทดสอบ	ข้อความ
ประโยคที่ 11	กิจกรรมคือการผลิตหนังสือ/ตำรา รายงานสถานภาพทางเศรษฐกิจเทคโนโลยี การศึกษาเชิงนโยบาย การผลิตสื่อความรู้เพื่อสาธารณะ และการจัดสัมมนาและฝึกอบรมเฉพาะทาง
ประโยคที่ 12	เพื่อเป็นการถ่ายทอดความรู้ แลกเปลี่ยนประสบการณ์และความคิดเห็นระหว่างนักวิชาการและผู้ประกอบการอิเล็กทรอนิกส์และคอมพิวเตอร์ ได้จัดการสัมมนาและฝึกอบรม ดังนี้คือ
ประโยคที่ 13	(1) ชนะ โสภารักษ์ 2533 ศัพท์ไมโครคอมพิวเตอร์ กรุงเทพฯ : บริษัทอมรินทร์พริ้นติ้งกรุ๊ปจำกัด 153 หน้า
ประโยคที่ 14	ถ้าการรบกวนเกิดขึ้นที่กล้ามเนื้อของอวัยวะภายในที่สำคัญๆ อาจมีอาการรุนแรงถึงตายได้
ประโยคที่ 15	จากหลักการทำงานของพีโซอิเลกตริกคริสตัลทำให้สามารถนำคริสตัลนี้มาสร้างเป็นเซนเซอร์สำหรับตรวจวัดปริมาณสารได้

ขั้นตอนการสอบถามความคิดเห็นจากผู้ฟังในการวัดความเป็นธรรมชาติมีรายละเอียดดังนี้

1 คำสั่งสำหรับผู้ทดสอบคือ “ต่อจากนี้จะเป็นการทดสอบเพื่อวัดความเป็นธรรมชาติของเสียงสังเคราะห์ ท่านจะได้ฟังเสียงจากระบบสังเคราะห์เป็น ทั้งหมด 15 ชุด ชุดละ 2 ประโยคซึ่งเป็นข้อความเดียวกัน โดยให้เสียงแรกที่ได้ยิน ให้มีชื่อว่าเสียง A และเสียงที่ 2 มีชื่อว่าเสียง B เมื่อท่านได้ฟังครบทั้งสองเสียงแล้ว กรุณาระบุความเป็นธรรมชาติของเสียงสังเคราะห์ทั้งสองเทียบกัน ว่าเสียงใดมีความเป็นธรรมชาติมากกัน”

2 จากนั้นสุ่มเล่นเสียงโดยไม่เรียงลำดับประโยคทั้ง 15 ชุดทดสอบ พร้อมทั้งบันทึกคำตอบของผู้ทดสอบในแต่ละประโยคทดสอบโดยแปลงคำตอบของผู้ทดสอบให้เป็นคะแนนตามตารางที่ 3.1

3.3 การประเมินลักษณะของเสียงด้วยการวิเคราะห์

3.3.1 จากการวิเคราะห์ลักษณะของเสียงจากสเปกโตรแกรม

การสังเคราะห์เสียงที่เสนอในวิทยานิพนธ์นี้นอกจากต้องการให้สามารถสร้างเสียงสังเคราะห์ที่มีความเป็นธรรมชาติแล้ว ยังต้องการสร้างเสียงสังเคราะห์ที่มีลักษณะของเสียงที่แตกต่างกัน ดังนั้นเพื่อการประเมินลักษณะของเสียง จึงวิเคราะห์ลักษณะของเสียงจากสเปกโตรแก

รม ซึ่งวิเคราะห์เสียงสังเคราะห์ที่ได้การปรับค่าพารามิเตอร์ต่าง ๆ ของระบบตามที่เสนอเปรียบเทียบกับสเปกโตรแกรมเสียงจริงที่มีลักษณะเสียงต่าง ๆ

3.3.2 จากการวิเคราะห์ลักษณะของเสียงจากการวัดความถูกต้อง (Validity)

เพื่อการวัดความถูกต้องของลักษณะของเสียงจากระบบที่เสนอซึ่งสามารถสังเคราะห์เสียงลมหายใจ เสียงปกติ และเสียงบีบได้ ดังนั้นประเมินการระบุความถูกต้องของลักษณะของเสียงจากการทดสอบกับผู้ฟัง 15 คน โดยฟังประโยคของชุดเสียงทดสอบจากลักษณะเสียงทั้ง 3 ชนิด ชนิดละ 5 ประโยครวมเป็น 15 ประโยคแบบไม่เรียงลำดับ แล้วให้ผู้ฟังตอบชนิดของลักษณะของเสียงซึ่งคำตอบที่ถูกต้องจากผู้ฟังในแต่ละประโยคนับเป็น 1 คะแนน จากนั้นวัดเปอร์เซ็นต์ความถูกต้องการระบุชนิดของลักษณะของเสียง โดยคำนวณจากคะแนนเต็มในแต่ละลักษณะเสียงคือ $15 \times 5 = 75$ คะแนน รายละเอียดการออกแบบค่าพารามิเตอร์ และการทดสอบในการวัดความถูกต้องของลักษณะเสียงมีดังนี้

1 ค่าพารามิเตอร์สำหรับการสังเคราะห์เสียงที่มีลักษณะของเสียงต่าง ๆ

เพื่อการวัดความถูกต้องของการสังเคราะห์เสียงลมหายใจ เสียงปกติ และเสียงบีบ ให้ตรงตามทฤษฎีของลักษณะของเสียง วิทยานิพนธ์นี้จึงปรับค่าพารามิเตอร์ที่มีความสัมพันธ์กับแต่ละชนิดเสียง ตรงตามค่าที่ถูกต้องตามทฤษฎีลักษณะของเสียงซึ่งได้อธิบายแล้วในบทที่ 2 โดยค่าพารามิเตอร์ที่ใช้ในการสังเคราะห์เสียงแต่ละชนิด ในวิทยานิพนธ์นี้เพื่อสังเคราะห์ประโยค แสดงดังตารางที่ 3.3

ตารางที่ 3.3 พารามิเตอร์ที่สังเคราะห์เสียงเพื่อทดสอบความถูกต้องของลักษณะเสียง

ประโยคที่ใช้ทดสอบ	ค่าพารามิเตอร์			คำตอบ
	Rd	F0	SNR	
ประโยคที่ 1	ค่าปกติ *	ค่าปกติ	ค่าปกติ	เสียงปกติ
ประโยคที่ 2	ค่าปกติ	ค่าปกติ	ค่าปกติ	เสียงปกติ
ประโยคที่ 3	ค่าปกติ	ค่าปกติ	ค่าปกติ	เสียงปกติ
ประโยคที่ 4	ค่าปกติ	ค่าปกติ	ค่าปกติ	เสียงปกติ
ประโยคที่ 5	ค่าปกติ	ค่าปกติ	ค่าปกติ	เสียงปกติ
ประโยคที่ 6	ค่าปกติ	ค่าปกติ	-20 dB	เสียงลมหายใจ
ประโยคที่ 7	ค่าปกติ	ค่าปกติ	-20 dB	เสียงลมหายใจ
ประโยคที่ 8	ค่าปกติ	ค่าปกติ	-40 dB	เสียงลมหายใจ

ประโยคที่ใช้ทดสอบ	ค่าพารามิเตอร์			คำตอบ
	Rd	F0	SNR	
ประโยคที่ 9	x 2	ค่าปกติ	-20 dB	เสียงลมหายใจ
ประโยคที่ 10	x 2	ค่าปกติ	-40 dB	เสียงลมหายใจ
ประโยคที่ 11	x 0.3	ค่าปกติ	ค่าปกติ	เสียงบีบ
ประโยคที่ 12	x 0.5	-50	ค่าปกติ	เสียงบีบ
ประโยคที่ 13	x 0.5	-100	ค่าปกติ	เสียงบีบ
ประโยคที่ 14	x 0.3	-50	ค่าปกติ	เสียงบีบ
ประโยคที่ 15	x 0.3	-100	ค่าปกติ	เสียงบีบ

* ค่าปกติ : เป็นค่าที่เป็นผลลัพธ์จากต้นไม้ตัดสินใจ

จากตารางข้างต้น ความสัมพันธ์ของการปรับค่าพารามิเตอร์นี้กับลักษณะของเสียงที่เกิดขึ้นตรงตามทฤษฎี นั่นคือ

1.1 จากทฤษฎีเสียงบีบช่วงเปิดของสัญญาณเส้นเสียงแคบกว่าค่าปกติ และเสียงบีบมีค่าความถี่มูลฐานของเสียงผู้หญิงประมาณ 100-200 Hz [10] ดังนั้นในการสังเคราะห์เสียงในลักษณะบีบ จึงลดค่าความถี่มูลฐานจากระดับปกติลง 50-100 Hz และปรับค่า Rd ซึ่งมีความสัมพันธ์โดยตรงกันช่วงเปิดของเส้นเสียง ลดลงจากค่าปกติประมาณ 0.3 ถึง 0.5 เท่าของค่า Rd ปกติ

1.2 จากทฤษฎีเสียงลมหายใจ มีช่วงเปิดของสัญญาณเส้นเสียงกว้างกว่าค่าปกติ [10] และมีค่าสัญญาณรบกวนมากกว่าปกติ ดังนั้นในการสังเคราะห์เสียงลักษณะลมหายใจจึงทำการปรับค่า Rd เพิ่มขึ้น ประมาณ 2 เท่าของค่า Rd ปกติเพื่อให้ช่วงเปิดของเส้นเสียงกว้างขึ้น และเพิ่มระดับสัญญาณรบกวน

2 ประโยคที่ใช้ในการทดสอบ

ประโยคที่ใช้ในการทดสอบการวัดความถูกต้องมีแสดงได้ดังตารางที่ 3.4

ตารางที่ 3.4 ประโยคที่ใช้ในการวัดความถูกต้องของระบบ

ประโยคทดสอบ	ข้อความ
ประโยคที่ 1	คณะทรัพยากรชีวภาพและเทคโนโลยี, สายวิชาเทคโนโลยีวัสดุ คณะพลังงานและวัสดุ

ประโยคทดสอบ	ข้อความ
ประโยคที่ 2	ในปัจจุบันเครื่องวัดวิเคราะห์น้ำตาลกลูโคสที่ใช้ในทางการแพทย์ อุตสาหกรรมอาหาร รวมทั้งที่ใช้ในกระบวนการทางเทคโนโลยีชีวภาพและอื่นๆ ทั้งหมดต้องนำเข้ามาจากต่างประเทศ ซึ่งมีราคาแพง
ประโยคที่ 3	หัววัดน้ำตาลกลูโคส ระบบโฟลว์อินเจ็กชัน และระบบเก็บข้อมูล-แสดงและวิเคราะห์ผลแบบเวลาจริง เข้าด้วยกัน
ประโยคที่ 4	จากรูปจะเห็นได้ว่าหัวน้ำตาลที่พัฒนาขึ้นสามารถวัดได้ดีในช่วงความเข้มข้นของน้ำตาลกลูโคสถึง 4,000
ประโยคที่ 5	3) จากการวิจัยและพัฒนาในปีที่สองนี้ ได้มีการปรับปรุงประสิทธิภาพของเซลล์ให้สูงขึ้น
ประโยคที่ 6	และชดเชยกระแสไฟฟ้ารั่ว
ประโยคที่ 7	แต่เราจะพบว่าไม่ว่าอย่างไรก็ตามตำแหน่งแรกที่เป็นจุดเริ่มต้นไม่มีทางเปลี่ยนเสมอ
ประโยคที่ 8	อันจะเป็นประโยชน์แก่ครูที่ลดเวลาในการตรวจทำผลและวิเคราะห์ข้อสอบลง
ประโยคที่ 9	อุปสรรคที่สำคัญอย่างหนึ่งในประเทศของเราที่ทำให้การเรียนรู้และการค้นคว้าทดลองในส่วนของเทคโนโลยีที่ทันสมัยไม่สามารถทำได้กว้างขวางเท่าที่ควร คือ การขาดเครื่องมืออุปกรณ์ตลอดจนตำรับตำรา
ประโยคที่ 10	นอกจากบทเรียนสำหรับเรียนด้วยตัวเองบนไมโครคอมพิวเตอร์ดังกล่าวแล้ว
ประโยคที่ 11	จากนั้นจะผ่านวงจรลดทอนสัญญาณความถี่สูงแบบที่สามารถปรับค่าการลดทอนได้ด้วยสัญญาณดิจิตอล
ประโยคที่ 12	ในปีแรกของงานวิจัยและพัฒนาของห้องปฏิบัติการวิจัยเทคโนโลยีเลเซอร์ เพื่อ " พัฒนาระบบฮีเลียม-นีออนเลเซอร์ "
ประโยคที่ 13	ในทางการแพทย์ก็มีการใช้เลเซอร์แบบนี้ในการกระตุ้นกล้ามเนื้อ ซึ่งเรียกว่าวิธีการฝังเข็มด้วยเลเซอร์
ประโยคที่ 14	ทั้งนี้เนื่องจากว่าหลอดเลเซอร์ทุกชนิดที่เราผลิตเองได้ ยังมีชิ้นส่วนของหลอดเลเซอร์ที่ยังต้องสั่งซื้อ ก็คือ กระจกเลเซอร์
ประโยคที่ 15	ปี 2537 สามารถให้บริการแก่ภาคอุตสาหกรรมในการตัด, เจาะวัสดุต่างๆ ทั้งโลหะและอโลหะ

3 ขั้นตอนการทดสอบการวัดความถูกต้องของเสียงที่สังเคราะห์

ในการทดสอบนี้ผู้ฟังจะได้ฟังตัวอย่างเสียงในลักษณะเสียงต่าง ๆ ซึ่งประกอบด้วย เสียงพูดปกติ เสียงพูดแบบเสียงบีบ และเสียงพูดแบบเสียงลมหายใจ ตามที่ได้ออกแบบไว้ในตารางที่ 3.3 ขั้นตอนการสอบถามความคิดเห็นจากผู้ฟังมีรายละเอียดดังนี้

3.1 คำสั่งสำหรับผู้ทดสอบคือ “ต่อจากนี้จะเป็นการทดสอบเพื่อวัดถูกต้องของเสียงสังเคราะห์ ท่านจะได้ฟังเสียงจากระบบสังเคราะห์ทั้งหมด 15 ประโยค ปนกันระหว่างเสียงบีบ เสียงปกติ และเสียงลมหายใจ ตามที่ได้อธิบายความหมายของลักษณะของเสียงไปแล้วก่อนเริ่มการทดลอง เมื่อท่านได้ฟังแต่ละประโยคจบแล้ว กรุณาระบุลักษณะของเสียงที่ได้ฟังว่าเป็นเสียงลักษณะใด ระหว่าง เสียงบีบ เสียงลมหายใจ และเสียงปกติ”

3.2 จากนั้นสุ่มเล่นเสียงโดยไม่เรียงลำดับประโยคทั้ง 15 ประโยคให้ผู้ทดสอบได้ฟัง พร้อมทั้งบันทึกคำตอบของผู้ทดสอบในแต่ละประโยคทดสอบ

3.3.3 จากการวิเคราะห์ลักษณะของเสียงซึ่งวัดความสามารถในการทำงานของระบบ

การประเมินความแตกต่างของเสียงนี้ วัดระดับความแตกต่างของลักษณะของเสียง ซึ่งแต่ละชุดทดสอบใช้เสียงที่มีค่าพารามิเตอร์ที่แตกต่างกัน นอกจากนี้เพื่อทดสอบความสามารถของระบบสังเคราะห์เสียงที่สามารถสร้างเสียงได้อัตโนมัติ และหลากหลายตามค่าพารามิเตอร์ จึงเพิ่มการสังเคราะห์เสียงแบบผสมระหว่างชนิดของลักษณะของเสียงที่แตกต่างกัน การวัดความสามารถของระบบทำโดยใช้ผู้ฟัง 15 คน โดยฟังประโยคของเสียงจากลักษณะเสียงของแต่ละระบบ ซึ่งค่าพารามิเตอร์ที่ใช้ในการสังเคราะห์ชนิดของลักษณะของเสียงของแต่ละระบบแสดงดังตารางที่ 3.5 ประโยคที่ใช้ในการทดสอบ เป็นประโยคที่สุ่มเลือกจากฐานข้อมูลเสียง ที่ไม่ได้ใช้สำหรับฝึกฝนแบบจำลองเสียง มีข้อความว่า “แต่ที่เลือกใช้โปรเซสเซอร์เบอร์นี้ ก็เนื่องจากนักวิจัยมีประสบการณ์กับโปรเซสเซอร์เบอร์นี้อยู่มาก” เหตุผลที่เลือกใช้ประโยคจากฐานข้อมูลเสียง เพราะจะทำให้สามารถฟังเสียงสังเคราะห์เทียบกับเสียงพูดจริง เพื่อยืนยันความถูกต้องของระบบสังเคราะห์เสียงที่นำเสนอได้ การออกแบบการทดลองแบ่งได้เป็นสองขั้นตอนคือ การปรับค่าพารามิเตอร์ และการจับคู่เสียงสังเคราะห์จากชุดพารามิเตอร์ ซึ่งมีรายละเอียดดังนี้

ตารางที่ 3.5 ค่าพารามิเตอร์แต่ละชุดทดสอบสำหรับวัดความสามารถของระบบ

ค่าทดสอบ	ระบบอ้างอิง			ระบบที่นำเสนอ			
	ชุดที่	F0	SNR	ชุดที่	F0	Rd	SNR
เสียงปกติ	**B1**	ค่าปกติ	--	**P1**	ค่าปกติ	ค่าปกติ	ค่าปกติ
เสียงบีบ เมื่อ Rd เป็น 0.3 เท่า		--	--	**P2**	ค่าปกติ	x0.3	ค่าปกติ
เสียงบีบที่ F0 ลดลง 100 Hz จากปกติ เมื่อ Rd เป็น 0.3 เท่า	**B3**	-100	--	**P3**	-100	x0.3	ค่าปกติ
เสียงบีบที่ F0 ลดลง 100 Hz จากปกติ เมื่อ Rd ปกติ		--	--	**P4**	-100	ค่าปกติ	ค่าปกติ
เสียงบีบที่ F0 ลดลง 50 Hz จากปกติ (ที่ Rd เป็น 0.3 เท่า)	**B5**	-50	--	**P5**	-50	x0.3	ค่าปกติ**
เสียงบีบ เมื่อ F0 ลดลง 50 Hz จากปกติ (ที่ Rd ปกติ)		--	--	**P6**	-50	ค่าปกติ	ค่าปกติ
เสียงบีบ เมื่อ F0 ลดลง 50 Hz จากปกติ (ที่ Rd 0.5 เท่า)		--	--	**P13**	-50	x0.5	ค่าปกติ
เสียงลมหายใจที่ SNR ลดลง 20 dB จากปกติ	**B7**	ค่าปกติ	G-wn*	**P7**	ค่าปกติ	ค่าปกติ	-20 dB
เสียงลมหายใจที่ Rd เป็น 2 เท่า เมื่อ SNR ลดลง 20 dB จากปกติ		--	--	**P8**	ค่าปกติ	x2	-20 dB
เสียงลมหายใจที่ SNR ลดลง 40 dB จากปกติ		--	--	**P12**	ค่าปกติ	ค่าปกติ	-40 dB
เสียงบีบที่ปนเสียงลมหายใจ เมื่อค่า Rd เป็น 0.3 เท่า และ SNR ลดลง 20 dB จากปกติ		--	--	**P11**	ค่าปกติ	x0.3	-20 dB
เสียงบีบที่ปนเสียงลมหายใจที่ F0 ลดลง 100 Hz จากปกติ เมื่อ Rd เป็น 0.3 เท่า SNR ลดลง 20 dB จากปกติ	**B9**	-100	G-wn	**P9**	-100	x0.3	-20 dB

ค่าทดสอบ	ระบบอ้างอิง			ระบบที่นำเสนอ			
	ชุดที่	F0	SNR	ชุดที่	F0	Rd	SNR
เสียงบีบที่ปนเสียงลมหายใจที่ F0 ลดลง 50 Hz จากปกติ เมื่อ Rd เป็น 0.3 เท่า และ SNR ลดลง 20 dB จากปกติ	**B10**	-50	G-wn	**P10**	-50	x0.3	-20 dB

* G-wn : สัญญาณรบกวนสีขาวกระจายตัวแบบเกาซ (Gaussian white noise)

** ค่าปกติ: เป็นค่าที่เป็นผลลัพธ์จากต้นไม้ตัดสินใจ

1 หลักการปรับค่าพารามิเตอร์

การปรับค่าพารามิเตอร์ระบบสังเคราะห์เสียง เพื่อสร้างเสียงชุดทดสอบดังตารางที่ 3.5 โดยมีหลักการตามทฤษฎีลักษณะของเสียง [10] ซึ่งให้รายละเอียดของเสียงแต่ละชนิดดังนี้

1.1 จากทฤษฎี ลักษณะเสียงบีบจะมีค่าความถี่มูลฐานของเสียงผู้หญิงลดลงจากค่าปกติ อยู่ที่ประมาณ 100-200 Hz และมีช่วงเปิดของสัญญาณเส้นเสียงแคบลงจากระดับปกติ ดังนั้นการสังเคราะห์เสียงสำหรับระบบที่นำเสนอ และระบบอ้างอิง จึงปรับค่าพารามิเตอร์ค่าความถี่มูลฐานลดลง 50 Hz และ 100 Hz จากค่าปกติที่ได้จากต้นไม้ตัดสินใจ และปรับค่า Rd เป็น 0.5 และ 0.3 เท่าของค่าปกติ เพื่อเปรียบเทียบความแตกต่างของช่วงเปิดสัญญาณเส้นเสียงที่แคบลงเป็นลำดับ

1.2 สำหรับการปรับลักษณะเสียงลมหายใจซึ่งมีลักษณะช่วงเปิดสัญญาณเส้นเสียงกว้างกว่าจากระดับปกติ และมีระดับสัญญาณรบกวนสูงขึ้น ดังนั้นสำหรับการสังเคราะห์เสียงจากระบบอ้างอิงใช้วิธีเพิ่มค่าสัญญาณรบกวนสีขาวกระจายตัวแบบเกาซ เพื่อสร้างปรับค่าสัญญาณรบกวนสำหรับเสียงลมหายใจ ขณะที่ในระบบสังเคราะห์เสียงที่นำเสนอใช้วิธีปรับค่า SNR เพื่อปรับระดับสัญญาณรบกวนลมหายใจ และใช้การปรับค่า Rd เพิ่มขึ้นเป็นสองเท่า เพื่อปรับความกว้างของสัญญาณเส้นเสียง

2 การออกแบบชุดการทดลอง

เพื่อวัดประสิทธิภาพของระบบในการประเมินความแตกต่างของระดับเสียงสังเคราะห์ในลักษณะเสียงต่าง ๆ จึงได้ทำการออกแบบคู่เสียงสังเคราะห์ที่ได้สังเคราะห์ตามพารามิเตอร์แต่ละชุดในตารางที่ 3.5 และจับคู่เสียงสังเคราะห์จากพารามิเตอร์ที่มีผลในการออกเสียงลักษณะเสียงบีบและเสียงลมหายใจตามที่ได้กล่าวไว้แล้วข้างต้น คู่ชุดการทดลองสามารถสรุปได้ตามตารางที่ 3.6

ตารางที่ 3.6 ชุดทดสอบการเปรียบเทียบระดับความแตกต่างของลักษณะของเสียงแต่ละชนิด

ชุดที่	ความแตกต่างของลักษณะเสียง	คู่ชุดการทดสอบ	
		ระบบอ้างอิง	ระบบที่นำเสนอ
1	ชุดเสียงปกติ – เสียงลมหายใจ โดยปรับระดับสัญญาณรบกวน	(1A) B1 – B7	(1B) P1 – P7
2	ชุดเสียงลมหายใจเมื่อ Rd ต่างกัน (โดยมี SNR ลดลง 20 dB จากปกติ)	-	(2B) P7 – P8
3	ชุดเสียงลมหายใจที่ค่า SNR ต่างกัน (โดยมี F0 และ Rd เป็นค่าปกติ)	-	P7 – P12
4	ชุดเสียงบีบที่ปนเสียงลมหายใจที่ Rd ต่างกัน (โดยมี F0 ปกติ และ SNR ลดลง 20 dB จากปกติ)	-	P7 – P11
5	เสียงบีบที่ปนเสียงลมหายใจที่ F0 ต่างกัน (โดยมี SNR ลดลง 20 dB จากปกติ เป็นค่าคงที่และระบบที่เสนอให้ Rd เป็น 0.3 เท่า)	(5A) B9 – B10	(5B) P9 – P10
6	ชุดเสียงปกติ-บีบเมื่อ F0 ต่างกัน (โดยมี F0 ลดลง 50 Hz จากปกติ)	(6A) B1 – B5	(6B) P1 – P6
7	ชุดเสียงปกติ-บีบเมื่อ Rd ต่างกัน (โดยมี F0 เป็นค่าปกติ)	-	P1 – P2
8	ชุดเสียงบีบเมื่อ F0 ต่างกัน 50 Hz (ระบบที่เสนอให้ Rd เป็น 0.3 เท่า)	B3 – B5	P3 – P5
9	ชุดเสียงบีบเมื่อ F0 ต่างกัน 100 Hz (โดยมี Rd เป็น 0.3 เท่า)	-	P2 – P3
10	ชุดเสียงบีบเมื่อ Rd ต่าง กัน (โดยมี F0ลดลง 50 Hz จากปกติ)	-	P5 – P6
11	ชุดเสียงบีบเมื่อ Rd ต่าง (โดยมี F0 ลดลง 100 Hz จากปกติ)	-	P3 – P4
12	ชุดเสียงบีบเมื่อ Rd ต่างกัน (โดยมี F0 ลดลง 50 Hz จากปกติ)	-	P5 – P13

โดยพิจารณาตามปัจจัยที่มีผลกระทบต่อลักษณะของเสียงนั้น ๆ ดังนี้

2.1 การเปรียบเทียบระดับความแตกต่างของลักษณะเสียงลมหายใจ

เนื่องจากเสียงลมหายใจคือ เสียงที่เกิดจากการสั่นของเส้นเสียงในขณะที่เส้นเสียงปิดไม่สนิท ทำให้มีอากาศสามารถแทรกผ่านเส้นเสียงขณะที่ออกเสียง (ในช่วงเปิดของเส้นเสียง) ดังนั้นปัจจัยที่มีผลต่อความแตกต่างเสียงลมหายใจ เกิดจากส่วนเปิดของสัญญาณเส้นเสียงซึ่งมีความสัมพันธ์กับค่า Rd และเกิดจากระดับความสัญญาณรบกวนซึ่งมีความสัมพันธ์กับค่า SNR ดังนั้นวิทยานิพนธ์นี้จึงสร้างชุดการทดลองเพื่อเปรียบเทียบความแตกต่างของทั้งสองปัจจัยดังนี้

- พิจารณาผลกระทบของ Rd บนเสียงลมหายใจ โดยใช้ชุดการทดลองที่เปรียบเทียบเสียงลมหายใจที่มีค่าความถี่มูลฐาน และค่า SNR เท่ากัน แต่ปรับให้ค่า Rd แตกต่างกัน ดังชุดการทดลองที่ 2B ให้เสียงลมหายใจมีการปรับค่า Rd เป็น 2 เท่าของค่าปกติ (P8) เทียบกับเสียงลมหายใจที่มีค่า Rd เป็นค่าปกติ (P7)
- พิจารณาผลกระทบของค่า SNR บนเสียงลมหายใจ โดยใช้ชุดการทดลองที่เปรียบเทียบเสียงลมหายใจสองเสียงที่มีค่าความถี่มูลฐานและ Rd เท่ากัน แต่มีค่า SNR แตกต่างกัน ดังชุดการทดลองที่ 3 ซึ่งเปรียบเทียบเสียงลมหายใจซึ่งมีค่า SNR ลดลง 20 เดซิเบล จากปกติ (P7) กับเสียงลมหายใจที่มีการปรับค่า SNR ลดลง 40 เดซิเบล (P12)

2.2 การเปรียบเทียบความแตกต่างของระดับความแตกต่างของลักษณะเสียงบีบ

เนื่องจากเสียงบีบมีลักษณะของเส้นเสียงเกร็ง ช่วงเปิดสัญญาณเส้นเสียงแคบลง ทำให้ลมไหลผ่านช่องเส้นเสียงได้ในปริมาณที่น้อยกว่าปกติ ส่งผลให้ความถี่มูลฐานต่ำลง ดังนั้นปัจจัยที่มีผลต่อความแตกต่างลักษณะเสียงบีบ คือช่วงเปิดของสัญญาณเส้นเสียงซึ่งมีความสัมพันธ์กับค่า Rd และค่าความถี่มูลฐาน ดังนั้นวิทยานิพนธ์นี้จึงสร้างชุดการทดลองเพื่อเปรียบเทียบความแตกต่างของทั้งสองปัจจัยดังนี้

- พิจารณาผลกระทบของค่า Rd บนเสียงบีบ โดยใช้ชุดการทดลองที่เปรียบเทียบเสียงบีบสองเสียงที่มีค่าค่าความถี่มูลฐานเท่ากัน แต่มีค่า Rd

แตกต่างกัน ดังชุดการทดลองที่ 7 10 และ 11 ซึ่งเสียงบีบมีการปรับค่า Rd เป็น 0.3 เท่าของค่าปกติ แตกต่างจากเสียงบีบที่ไม่ปรับค่า Rd และชุดการทดลองที่ 12 ที่มีการปรับค่า Rd เป็น 0.3 เท่าของค่าปกติเทียบกับเสียงบีบที่มีการปรับค่า Rd เป็น 0.5 เท่าของค่าปกติ

- พิจารณาผลกระทบของค่าความถี่มูลฐานบนเสียงบีบ โดยใช้ชุดการทดลองที่เปรียบเทียบเสียงบีบที่มีค่า Rd เท่ากัน แต่มีค่าความถี่มูลฐานแตกต่างกัน ดังชุดการทดลองที่ 6 8 และ 9 ซึ่งเปรียบเทียบเสียงบีบที่มีความถี่มูลฐานต่ำลงกว่าระดับเสียงปกติ โดยปรับค่าความถี่มูลฐาน ลดลง 50 Hz และ 100 Hz เทียบกับเสียงปกติ

3 เกณฑ์การตัดสินใจ

เพื่อประเมินการวัดระดับความแตกต่างของลักษณะของเสียงจากผู้ฟังมีดังนี้ ผู้ฟังสามารถบอกลักษณะของเสียงชนิดเดียวกัน ภายใต้ระบบสังเคราะห์เสียงเดียวกัน ยกตัวอย่างเช่นการทดสอบจากเสียงสังเคราะห์ที่มาจากระบบที่นำเสนอ ผู้ฟังสามารถบอกความแตกต่างระหว่างเสียงลมหายใจที่เพิ่มระดับสัญญาณรบกวนแตกต่างกันได้ จากการทดสอบการรับรู้จากผู้ฟังผู้ฟังจะฟังเสียงในชุดการทดสอบละ 1 เสียง ถ้าผู้ฟังระบุว่าสามารถรับรู้ถึงการเปลี่ยนแปลงได้ จะถือว่าชุดการทดสอบนั้นได้ 1 คะแนน คะแนนเต็มในแต่ละชุดการทดสอบ 15 คะแนน จากผู้ฟังทั้งหมด 15 คน จากนั้นแสดงผลเป็นเปอร์เซ็นต์ที่ผู้ฟังสามารถรับรู้ได้ถึงการเปลี่ยนแปลง ขั้นตอนการสอบถามความคิดเห็นจากผู้ฟังมีรายละเอียดดังนี้

3.1 คำสั่งสำหรับผู้ทดสอบคือ “ต่อจากนี้จะเป็นการทดสอบเพื่อวัดระดับการเปลี่ยนแปลงของลักษณะในเสียงสังเคราะห์ ท่านจะได้ฟังเสียงจากระบบสังเคราะห์ทั้งหมด 16 ชุด ชุดละ 2 ประโยค ทุกประโยคจะมีข้อความว่า ‘แต่ที่เลือกใช้โปรเซสเซอร์เบอร์นี้ ก็เนื่องจากนักวิจัยมีประสบการณ์กับโปรเซสเซอร์เบอร์นี้อยู่มาก’ ก่อนที่จะเล่นเสียงในแต่ละชุดการทดสอบ ท่านจะทราบถึงลักษณะของเสียงในชุดนั้น เมื่อท่านฟังเสียงครบในแต่ละชุดแล้ว ให้ท่านพิจารณาว่าท่านสามารถรับรู้ถึงว่าแตกต่างของระดับในลักษณะเสียงชุดที่ได้ฟังไปแล้วหรือไม่ เช่าเสียงบีบ ที่รู้สึกว่ามีลักษณะของเสียงบีบ มากขึ้นหรือน้อยลงหรือไม่”

3.2 จากนั้นเล่นเสียงโดยเรียงลำดับชุดการทดสอบทั้ง 16 ชุดให้ผู้ทดสอบได้ฟัง พร้อมทั้งบันทึกคำตอบของผู้ทดสอบในแต่ละชุดทดสอบ

3.3.4 วิธีการประเมินลักษณะของเสียงโดยการวิเคราะห์ค่าอะคูสติกจากเสียงสังเคราะห์

1 ภาพตัดขวางสเปกตรัม (Spectral cross section)

ภาพตัดขวางของสเปกตรัมได้มาจากการนำสัญญาณเสียงช่วงสั้นในกรอบหน้าต่างทำการเปลี่ยนรูปฟูเรียร์ไม่ต่อเนื่องแบบช่วงสั้น (Short-term discrete Fourier transform) ผลลัพธ์ที่ได้คือสเปกตรัมของสัญญาณเสียง ประโยชน์ของภาพตัดขวางสเปกตรัมใช้ประกอบการพิจารณาการประมาณของการรับรู้เสียงสระได้เป็นอย่างดี และใช้ระบุตำแหน่งความถี่ และแอมปลิจูดสัมพัทธ์ของความถี่ฟอร์แมนท์ นอกจากนี้ภาพตัดขวางของสเปกตรัมสามารถนำมาใช้วิเคราะห์ค่าฮาร์โมนิกส์ และใช้เพื่อเปรียบเทียบรูปร่างสเปกตรัมของสระจากแต่ละสัญญาณเสียงอีกด้วย

วิทยานิพนธ์นี้ได้วิเคราะห์ภาพตัดขวางสเปกตรัมของชุดเสียงทดสอบต่าง ๆ โดยเลือกอธิบายภาพตัดขวางของสเปกตรัมของเสียงสระที่ความถี่ในการปรากฏมากในประโยคที่ทดสอบ ซึ่งสระที่ใช้พิจารณาภาพวิเคราะห์ตัดขวาง ในการทดสอบนี้เลือกเสียงสระอะ /a/ เพื่อใช้เป็นข้อมูลช่วยพิจารณาประกอบกับผลการทดสอบการรับรู้จากผู้ฟัง

2 การวิเคราะห์ค่าแอมปลิจูดของฮาร์โมนิกส์

ในการวิเคราะห์วิธีการออกเสียง ค่าแอมปลิจูดฮาร์โมนิกส์ที่ 1 (H1) มีความสัมพันธ์กับช่วงเปิดของสัญญาณเส้นเสียง (Open Quotient) [53, 54] ซึ่งเป็นปัจจัยในการทำให้เกิดลักษณะเสียงแบบต่าง ๆ ดังนั้นค่า H1 ถือว่ามีความเหมาะสมในการวิเคราะห์ลักษณะของเสียง การเปรียบเทียบค่า H1 ที่ใช้ในวิทยานิพนธ์นี้ได้แก่

2.1 ค่าผลต่างฮาร์โมนิกส์ที่หนึ่งและสอง (H1- H2)

ค่า H1 จะใช้ในการเปรียบเทียบกับค่าแอมปลิจูดในองค์ประกอบความถี่อื่น ๆ แต่เนื่องจากในองค์ประกอบความถี่สูง เช่นค่าแอมปลิจูดของความถี่ฟอร์แมนท์ที่ 1 หรือ 2 ถูกพิจารณาว่ามีความสัมพันธ์กับความเร็วในช่วงปิดของเส้นเสียง ในการวิเคราะห์ค่าช่วงเปิดของเส้นเสียงจึงนิยมนำค่า H1 ใช้เปรียบเทียบกับค่าฮาร์โมนิกส์ที่ 2 (H2) ที่อยู่ถัดจาก H1 และจากการทดลองของ Esposito [55] พบว่าค่าผลต่างฮาร์โมนิกส์ที่หนึ่ง และสองสามารถใช้แยกลักษณะเฉพาะของเสียงได้ถูกต้องจาก 8 ใน 10 ค่า

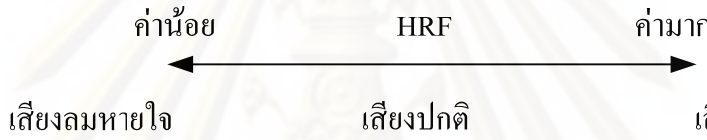
2.2 ค่าปัจจัย HRF (Harmonic richness factor)

จากการวิเคราะห์สเปกตรัมของสัญญาณเส้นเสียงของเสียงชนิดต่าง ๆ พบว่าฮาร์โมนิกส์ที่หนึ่งถึงห้าของสเปกตรัมเส้นเสียงของลักษณะเสียงชนิดต่าง ๆ มีความสัมพันธ์ของแอมปลิจูดระหว่างฮาร์โมนิกส์มูลฐาน และฮาร์โมนิกส์ลำดับที่สูงขึ้นไป แตกต่างกัน ซึ่ง

สามารถคำนวณเป็นผลต่างของความชันสเปกตรัม ในรูปของค่าปัจจัย HRF (Harmonic richness factor) ตามสมการที่ (18) ซึ่งใช้วัดความสัมพันธ์ฮาร์โมนิกส์ลำดับต่างๆและแอมปลิจูดของความถี่มูลฐาน

$$HRF = \frac{\sum_{i\geq 2} H_i}{H_1} \tag{18}$$

โดยที่ Hi เป็นค่าแอมปลิจูดของฮาร์โมนิกส์ลำดับที่ i และ H1 คือค่าแอมปลิจูดของความถี่มูลฐาน ตามทฤษฎีลักษณะของเสียง ค่า HRF สามารถบอกลักษณะของเสียงต่าง ๆ โดยเสียงบีบจะมีค่า HRF สูง (มากกว่า 2 เดซิเบล) ตามด้วยเสียงปกติ และเสียงลมหายใจ (น้อยกว่า -16 เดซิเบล) ตามลำดับสามารถแสดงความสัมพันธ์ได้ดังรูปที่ 3.4



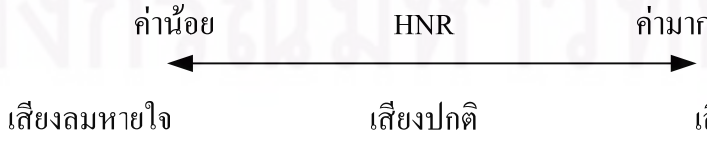
รูปที่ 3.4 ความสัมพันธ์การบอกความเป็นลักษณะของเสียงจากค่า HRF

2.3 การประเมินสัญญาณรบกวน

ระดับของสัญญาณรบกวนที่เส้นเสียง มีความสำคัญต่อการรับรู้ถึงสักษณะเสียงลมหายใจ [38] ในวิทยานิพนธ์นี้ ค่าอัตราส่วนฮาร์โมนิกส์ต่อสัญญาณรบกวน (Harmonic to noise ratio, HNR) ถูกใช้เป็นเครื่องมือวัดปริมาณของสัญญาณรบกวนบนสัญญาณเส้นเสียงในองค์ประกอบความถี่สูง สามารถคำนวณได้ตามสมการที่ (19)

$$HNR = \frac{\sum H_i}{\sum N_i} \tag{19}$$

โดยที่ค่า Hi แทนพลังงาน ณ ฮาร์โมนิกส์ที่ i และ Ni แทนพลังงานรอบตำแหน่งฮาร์โมนิกส์ที่ i มีความกว้างไม่เกินครึ่งหนึ่งของความถี่มูลฐาน ซึ่งค่า HNR นี้สามารถสะท้อนถึงความสัมพันธ์ระหว่างคุณภาพเสียงได้ดังรูปที่ 3.5



รูปที่ 3.5 ความสัมพันธ์การบอกความเป็นลักษณะของเสียงจากค่า HNR

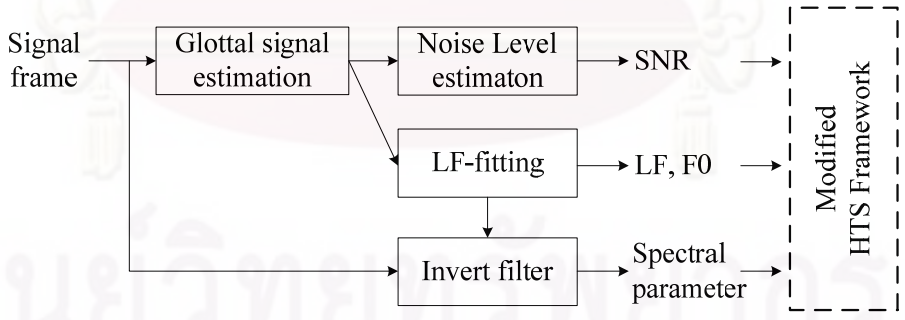
บทที่ 4
การวิเคราะห์ข้อมูลเสียง และการสร้างระบบที่นำเสนอ

ในบทนี้อธิบายการวิเคราะห์พารามิเตอร์เพื่อใช้ในการฝึกฝน และการปรับเปลี่ยนลักษณะของเสียง ประกอบด้วยการวิเคราะห์แหล่งกำเนิดเส้นเสียง การวิเคราะห์แบบจำลองแอลเอฟ และการวิเคราะห์ระดับของสัญญาณรบกวนซึ่งใช้วิธีการลดสัญญาณรบกวนด้วยการประมาณค่าจุดเปลี่ยนของเวฟเลท รวมทั้งอธิบายขั้นตอนการสังเคราะห์เสียง และวิธีการสร้างแบบจำลองที่พัฒนาส่วนของขั้นตอนการปรับลักษณะเสียงที่เสนอ

การวิเคราะห์ค่าพารามิเตอร์

การวิเคราะห์หาค่าพารามิเตอร์นี้มีจุดประสงค์เพื่อสกัดสัญญาณเส้นเสียง เพื่อใช้แทนที่สัญญาณกระตุ้นจากเดิมที่ใช้สัญญาณพัลส์ จากนั้นเปลี่ยนให้เป็นค่าพารามิเตอร์ที่สามารถใช้กับการสังเคราะห์เสียงด้วยแบบจำลองฮิดเดนมาร์คอฟแบบปรับปรุง

ขั้นตอนการวิเคราะห์พารามิเตอร์นี้ประกอบด้วยการแยกสัญญาณเส้นเสียง และลักษณะของช่องทางเดินเสียงออกมาจากสัญญาณเสียง จากนั้นสกัดระดับสัญญาณรบกวนจากสัญญาณเส้นเสียงโดยการใช้การวิเคราะห์ด้วยเวฟเลท และใช้แบบจำลองแอลเอฟเพื่อประมาณสัญญาณเส้นเสียง ซึ่งแผนผังแบบจำลองการวิเคราะห์พารามิเตอร์แสดงตามรูปที่ 4.1



รูปที่ 4.1 แบบจำลองการวิเคราะห์พารามิเตอร์

1. การวิเคราะห์สัญญาณเส้นเสียง

ตามทฤษฎีของแหล่งกำเนิดเสียง และตัวกรองสัญญาณ [7] นั้น กระบวนการการเกิดเสียงพูดนั้นจะสามารถแบ่งได้เป็นสามส่วนได้แก่ สัญญาณกระตุ้นเส้นเสียง (Glottal excitation, $G(\omega)$) , ตัวกรองสัญญาณทางเดินเสียง (Vocal tract response, $V(\omega)$) และผลจากการกระจาย ณ

ริมฝีปาก (Lip radiation effect, $R(\omega)$) ซึ่งสามารถแทนกระบวนการการเกิดเสียงพูดได้ด้วยสมการ (20) เมื่อ $S(\omega)$ แทนสัญญาณเสียงพูด

$$S(\omega) = G(\omega)V(\omega)R(\omega) \quad (20)$$

ผลการกระจาย ณ ริมฝีปากนี้ สามารถประมาณได้เป็นตัวกรองอนุพันธ์สัญญาณ (Differentiating filter) เมื่อนำมารวมกับสัญญาณกระตุ้นเส้นเสียงจะได้เป็นสมการ (21) ซึ่ง $G'(\omega)$ แทนอนุพันธ์ของสัญญาณเส้นเสียง

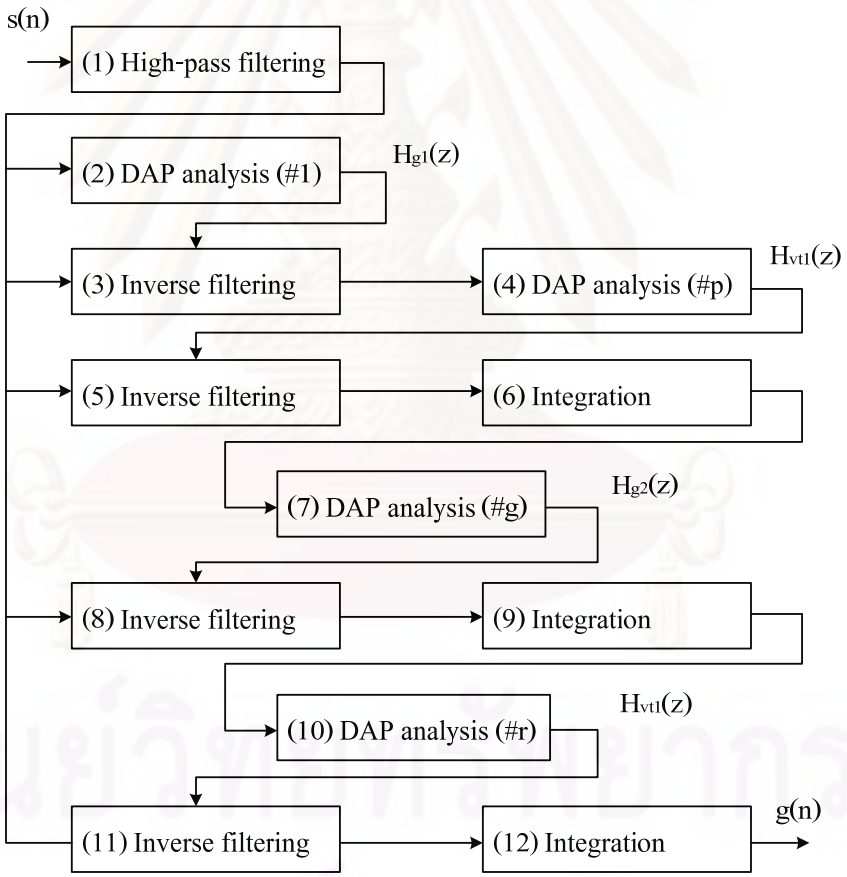
$$S(\omega) = G'(\omega)V(\omega) \quad (21)$$

แบบจำลองแอลเอฟซึ่งเป็นแบบจำลองที่ประมาณอนุพันธ์ของสัญญาณเส้นเสียง จึงถูกนำมาใช้เป็นพารามิเตอร์แทนสัญญาณเส้นเสียง ในการวิเคราะห์แหล่งกำเนิดเส้นเสียง

แบบจำลองข้อมูลเสียงในระบบสังเคราะห์เสียงด้วยแบบจำลองฮิดเดนมาร์คอฟจะวิเคราะห์เสียงพูดให้เป็นสัมประสิทธิ์เมลเซปสตรอล ซึ่งเป็นการประมาณซองของสเปกตรัม (Spectral envelope) ของสัญญาณเสียง แต่การประมาณนี้รวมเอาลักษณะของแหล่งกำเนิดเส้นเสียงเข้าไว้ด้วย ดังนั้นเพื่อนำแหล่งกำเนิดเส้นเสียงเข้าเป็นสัญญาณกระตุ้นในสังเคราะห์เสียงด้วยแบบจำลองฮิดเดนมาร์คอฟ จึงจำเป็นต้องแยกสัญญาณเส้นเสียงออกก่อนที่จะนำไปเข้าสู่ขั้นตอนการฝึกฝนตามกรอบงานของระบบสังเคราะห์เสียงด้วยแบบจำลองฮิดเดนมาร์คอฟโดยทั่วไปการแยกสัญญาณแหล่งกำเนิดเสียงจะใช้ตัวกรองสัญญาณย้อนกลับ เช่นการวิเคราะห์แอลพีซี หนึ่งในวิธีการแยกสัญญาณเส้นเสียงจากเสียงพูดที่เป็นที่นิยมคือ ตัวกรองสัญญาณย้อนกลับแบบดัดแปลงหลายรอบ (Iterative Adaptive Inverse Filter, IAIF) [56] แผนผังของวิธีการคำนวณนี้แสดงในรูปที่ 4.2 ในการคำนวณหาค่าพารามิเตอร์ทางด้านความถี่ในหลาย ๆ ขั้นตอน จะใช้วิธีการสร้างแบบจำลองค่าหลักทั้งหมดแบบไม่ต่อเนื่อง (Discrete all-pole model, DAP) [57] แทนการใช้การวิเคราะห์การประมาณค่าเชิงเส้นที่เป็นวิธีการเดิม

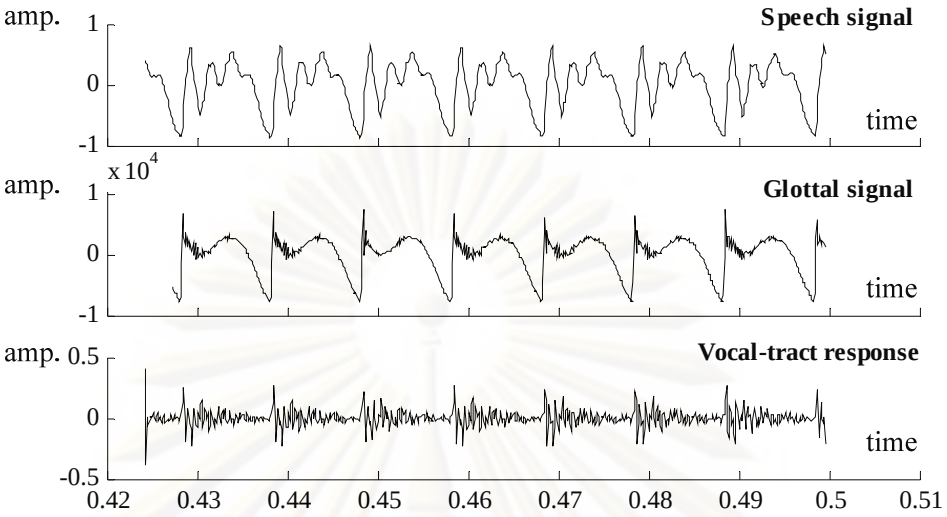
การขั้นตอนการคำนวณเพื่อสกัดสัญญาณเส้นเสียงออกจากสัญญาณเสียงตามรูปที่ 4.2 เริ่มจากการลดความเพี้ยนในองค์ประกอบความถี่ต่ำที่เกิดจากไมโครโฟนในขั้นตอนของการอัด ด้วยตัวกรองสัญญาณความถี่สูงผ่านในขั้นที่ 1 ค่าจุดตัดความถี่ที่ใช้จะน้อยกว่าความถี่มูลฐานเล็กน้อย ในขั้นที่ 2 ตัวกรองสัญญาณค่าหลักทั้งหมดลำดับที่หนึ่ง (First-order all pole filter) ถูกใช้เพื่อประมาณผลจากสัญญาณเส้นเสียง และการกระจายตัวที่ปาก จะได้ผลลัพธ์เป็นฟังก์ชันถ่ายเทตัวกรองสัญญาณศูนย์ทั้งหมด $H_{gl}(z)$ ต่อมาในขั้นที่ 3 นำผลลัพธ์ที่ได้ก่อนหน้านี้มาทำตัวกรองสัญญาณย้อนกลับซึ่งจะเป็นการหักล้างผลจากสัญญาณเส้นเสียงและผลจากการกระจายตัวที่ริมฝีปากออกจากสัญญาณเสียง จากนั้นประมาณตัวกรองสัญญาณทางเดินเสียง ด้วยดีเอพีอันดับที่ p ในขั้นที่ 4

แล้วหักล้างลักษณะทางเดินเสียงออกจากสัญญาณเสียงโดยการทำตัวกรองสัญญาณย้อนกลับที่หามาได้ตัวกรองสัญญาณอันดับที่ p การประมาณสัญญาณเส้นเสียงครั้งแรกจะสิ้นสุดที่ขั้นที่ 6 ซึ่งเป็นการหักล้างผลจากกระจายที่ริมฝีปาก การประมาณสัญญาณเส้นเสียงครั้งที่สอง เริ่มที่ขั้น 7 โดยการใช้ดีเอพีคำนวณค่าตัวกรองสัญญาณอันดับที่ g ของสัญญาณเส้นเสียง $H_{g2}(z)$ ที่ได้จากผลการประมาณครั้งแรก จากนั้นหักล้างสัญญาณเส้นเสียงออกจากสัญญาณเสียงในขั้นที่ 8 และหักล้างการกระจายตัวที่ริมฝีปากในขั้น 9 ผลลัพธ์ที่ได้จากเป็นผลตอบสนองของทางเดินเสียง ซึ่งจะถูกประมาณเป็นตัวกรองสัญญาณ $H_{vt2}(z)$ อันดับที่ r ด้วยดีเอพีในขั้นที่ 10 สุดท้ายหักล้างผลของแบบจำลองทางเดินเสียงใหม่ที่ได้ และผลจากการกระจายที่ริมฝีปากจากสัญญาณเสียงในขั้นที่ 11 และขั้นที่ 12 จะได้สัญญาณเสียงเสียง g(n) จากสัญญาณเสียง



รูปที่ 4.2 แผนผังการคำนวณตัวกรองสัญญาณย้อนกลับแบบดัดแปลงหลายรอบ [56]

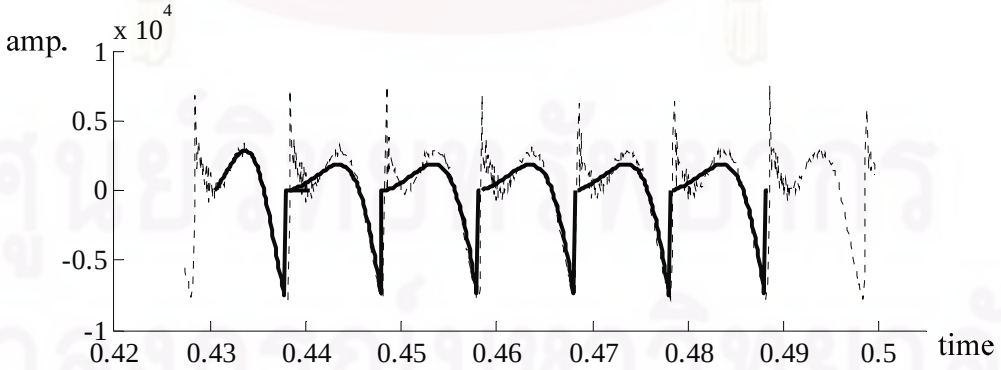
ผลลัพธ์จากการแยกสัญญาณเส้นเสียงแสดงในรูปที่ 4.3 โดยรูปบนแสดงสัญญาณเสียง รูปกลางเป็นอนุพันธ์ของสัญญาณเส้นเสียง และรูปล่างเป็นรูปตอบสนองตัวกรองสัญญาณทางเดินเสียง อนุพันธ์ของสัญญาณเส้นเสียงที่สกัดได้ จะถูกนำไปประมาณค่าโดยแบบจำลองแอลเอฟต่อไป



รูปที่ 4.3 ผลลัพธ์จากการสกัดสัญญาณเส้นเสียง และผลตอบสนองช่องทางเดินเสียง

2. การวิเคราะห์หาค่าพารามิเตอร์แบบจำลองแอลเอฟ

อัลกอริทึมการประมาณค่าแอลเอฟ (LF-fitting) [58] เป็นการหาเส้นโค้งที่เหมาะสมสำหรับสัญญาณเส้นเสียงซึ่งใช้ค่าจุดตัดทางเวลาเป็นค่าเริ่มต้นการประมาณ และใช้การประมาณไม่เชิงเส้น (Non-linear optimization) ในวิทยานิพนธ์นี้ใช้การวิเคราะห์สัญญาณเส้นเสียงได้ใช้เครื่องมือทีเคเคอาพรัด (TKK Aparat) [48] ซึ่งเป็นเครื่องมือสำหรับการศึกษาและประมาณค่าพารามิเตอร์แอลเอฟ ซึ่งผลลัพธ์จากการหาค่าพารามิเตอร์แอลเอฟแสดงดังรูปที่ 4.4 โดยที่เส้นประแสดงสัญญาณอนุพันธ์ซึ่งเป็นสัญญาณเป้าหมาย และเส้นทึบคือผลจากการประมาณ ซึ่งพบว่าอัลกอริทึมสามารถประมาณค่าสัญญาณได้เป็นอย่างดี



รูปที่ 4.4 ผลของอัลกอริทึมการประมาณค่าแอลเอฟ

3. การวิเคราะห์ระดับของสัญญาณรบกวน

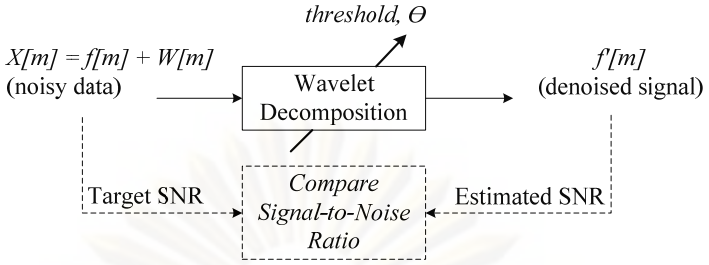
จากกาวิเคราะห์สัญญาณเส้นเสียง และการวิเคราะห์แบบจำลองแอลเอฟที่อธิบายข้างต้น พบว่า สัญญาณเส้นเสียงถูกสกัดออกมาจากสัญญาณเสียงและถูกประมาณค่าโดยแบบจำลองแอลเอฟ ในสัญญาณเส้นเสียงที่สกัดมายังมีสัญญาณรบกวนจากการไหลเวียนของลม (Turbulence noise) ซึ่งสัญญาณรบกวนนี้เป็นส่วนประกอบหนึ่งของสัญญาณกระตุ้นแบบผสม (Mixed excitation) สัญญาณรบกวนนี้จะมีระดับต่างกันในการออกเสียงแต่ละหน่วยเสียง เช่น การออกเสียงอา /aa/ และเสียงอี /ii/ การสกัดระดับเสียงรบกวนในสัญญาณเส้นเสียงจึงเป็นสิ่งจำเป็นเพื่อให้สามารถสร้างสัญญาณกระตุ้นแบบผสมได้อย่างถูกต้อง

การศึกษาการหาค่าระดับสัญญาณรบกวน ซึ่งมีระดับต่างกันในแต่ละหน่วยเสียง เทคนิคการแยกส่วนประกอบสัญญาณ (Signal decomposition) ถูกนำมาใช้ในการหาระดับสัญญาณรบกวน ซึ่งมีสองเทคนิคที่ใช้กันโดยทั่วไปได้แก่ การแปลงฟูเรียร์ (Fourier transform) และ การแปลงเวฟเลท (Wavelet transform) เมื่อพิจารณาดูลักษณะสัญญาณแอลเอฟแล้ว พบว่าสัญญาณมีส่วนประกอบของมุมแหลม การลดสัญญาณรบกวนในระบบโดยเวฟเลท (Wavelet denoising) เป็นการสร้างสัญญาณใหม่จากสัญญาณที่ถูกรบกวน ในการสกัดสัญญาณรบกวนในสัญญาณเส้นเสียง

การลดสัญญาณรบกวนด้วยการประมาณค่าจุดเปลี่ยนของเวฟเลท (Wavelet thresholding) เป็นวิธีการลดสัญญาณรบกวนที่ขึ้นกับค่าจุดเริ่มเปลี่ยน (Threshold) การหาจุดเริ่มเปลี่ยน มีหลายวิธี [15] และแต่ละวิธีให้ผลแตกต่างกัน ซึ่งการหาจุดเริ่มเปลี่ยนนี้มีความสำคัญต่อการลดสัญญาณรบกวนด้วยวิธีการประมาณค่าจุดเปลี่ยนของเวฟเลท ในงานวิจัยนี้สนใจการเลือกค่าจุดเริ่มเปลี่ยนสำหรับการลดสัญญาณรบกวน โดยพิจารณาจากระดับของสัญญาณรบการต่างกัน โดยกำหนดให้สัญญาณที่ถูกรบกวนเป็น X[m] เป็นสัญญาณอินพุท ให้ f[m] คือสัญญาณต้นแบบที่ต้องการสร้างขึ้นใหม่ (Reconstruct) และ W[n] เป็นสัญญาณรบกวนสีขาว (White noise) สามารถแสดงดังตามสมการ (22)

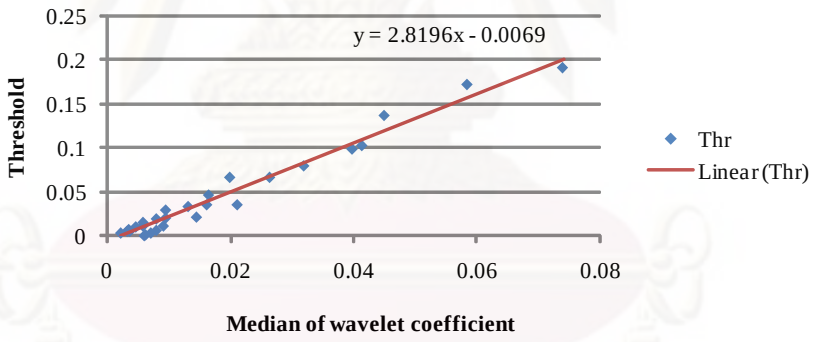
$$X[m] = f[m] + W[m] \tag{22}$$

ที่สัญญาณอินพุท X[m] มีค่าระดับสัญญาณต่อสัญญาณรบกวน (Signal-to-noise ratio, SNR) จะถูกคำนวณ และที่เอาต์พุทค่าระดับสัญญาณต่อสัญญาณรบกวนสามารถคำนวณได้จากค่าสัญญาณต้นแบบและค่าสัญญาณรบกวนสีขาว ดังนั้นในการหาค่าจุดเริ่มเปลี่ยนในงานวิทยานิพนธ์นี้ จึงพิจารณาจากสัญญาณที่สร้างขึ้นใหม่และสัญญาณรบกวนที่สกัดได้ มีระดับสัญญาณรบกวนเท่ากันกับอินพุท X[m] ซึ่งสามารถอธิบายดังรูป



รูปที่ 4.5 แผนผังการหาฟังก์ชันการหาจุดเปลี่ยน

วิทยานิพนธ์นี้เสนอวิธีหาฟังก์ชันการหาจุดเปลี่ยน (Thresholding function) ดังรูปที่ 4.5 โดยพิจารณาจากการเปรียบเทียบระดับสัญญาณรบกวนของอินพุท X[m] กับระดับสัญญาณรบกวนที่ลดสัญญาณรบกวน $f'[m]$ หากมีค่าใกล้เคียงกันสูงสุด สำหรับทุกสัญญาณเส้นเสียง เริ่มต้นด้วยการหาความสัมพันธ์โดยปรับค่าจุดเปลี่ยนของเสียงที่ทำการลดสัญญาณรบกวนในระบบโดยเวฟเลท จนให้ค่าระดับสัญญาณรบกวนที่ลดสัญญาณรบกวนมีค่าใกล้เคียงกันกับระดับสัญญาณรบกวนของอินพุท X[m] มากที่สุด และเลือกจุดเปลี่ยน ณ ระดับสัญญาณรบกวนนี้ เพื่อหาความสัมพันธ์ของสมการฟังก์ชันการหาจุดเปลี่ยน สามารถแสดงความสัมพันธ์ได้ดังรูปที่ 4.6



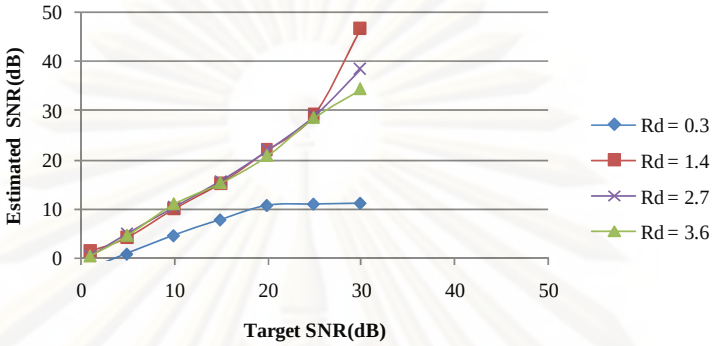
รูปที่ 4.6 ความสัมพันธ์ของสมการฟังก์ชันการหาจุดเปลี่ยน

ซึ่งจากผลการทดลองสามารถหาสมการความสัมพันธ์ของค่าจุดเปลี่ยนซึ่งแทนด้วย θ กับค่ามัธยฐานของสัมประสิทธิ์เวฟเลทแทนด้วย x มีความสัมพันธ์กันแบบสมการเส้นตรงดังสมการ (23) ซึ่งนำไปใช้ในการหาค่าจุดเปลี่ยนของการหาค่าประสิทธิ์เวฟเลทในการลดสัญญาณรบกวน

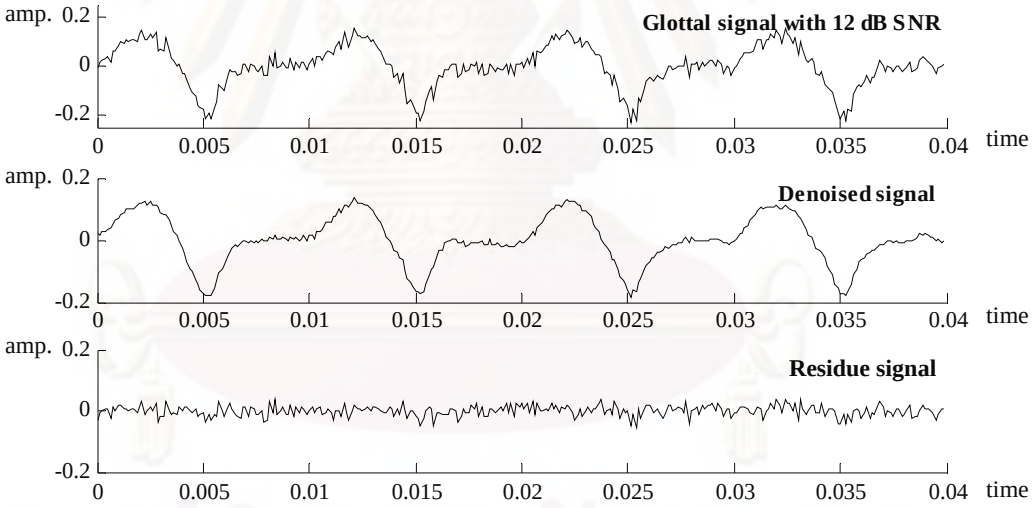
$$\theta = 2.8196x - 0.0069 \quad (23)$$

นอกจากนี้เพื่อทดสอบฟังก์ชันการหาจุดเปลี่ยน จึงนำอนุพันธ์ของสัญญาณเส้นเสียงในรูปแบบต่างซึ่งเป็นสัญญาณที่จะใช้ในการฝึกฝนแบบจำลองการสังเคราะห์เสียง มาใส่สัญญาณรบกวนระดับต่าง ๆ กันตั้งแต่ 1 เดซิเบล จนถึง 30 เดซิเบล จากนั้นนำสัญญาณดังกล่าวไปทำการแยกสัญญาณรบกวนโดยใช้ฟังก์ชันการหาจุดเปลี่ยนตามสมการ (23) ซึ่งผลลัพธ์ที่ได้มี

ความสัมพันธ์ของระดับสัญญาณรบกวนเป้าหมาย และระดับสัญญาณรบกวนที่แยกได้นั้นดังรูปที่ 4.7 แสดงให้เห็นว่า ฟังก์ชันดังกล่าวสามารถใช้ในการหาจุดเปลี่ยนของอนุพันธ์ของสัญญาณเส้นเสียงได้เป็นอย่างดี ตัวอย่างสัญญาณที่ได้จากการแยกสัญญาณรบกวน แสดงในรูปที่ 4.8



รูปที่ 4.7 ความสัมพันธ์ระหว่าง ระดับสัญญาณรบกวน และที่สกัดได้จากรูปร่างสัญญาณเส้นเสียงแบบต่าง ๆ

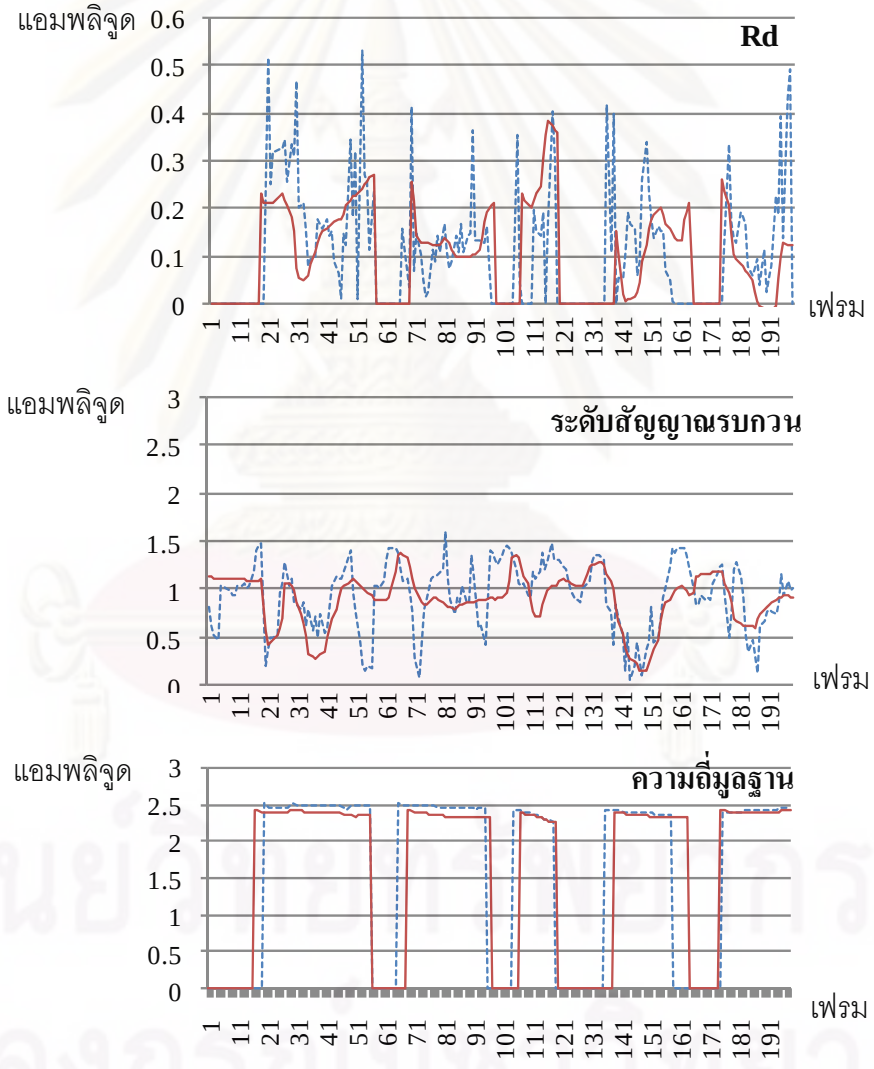


รูปที่ 4.8 ผลการสกัดสัญญาณรบกวน ด้วยฟังก์ชันหาจุดเปลี่ยนที่นำเสนอ

4. การประเมินผลขั้นตอนการวิเคราะห์ค่าพารามิเตอร์

การประเมินผลขั้นตอนการวิเคราะห์ค่าพารามิเตอร์เพื่อวัดผลความเหมาะสมของค่าพารามิเตอร์ที่ใช้ในการสร้างแบบจำลอง โดยพิจารณาจากความสัมพันธ์ของค่าพารามิเตอร์ SNR ค่า Rd และค่าความถี่มูลฐานจากการวิเคราะห์สัญญาณเสียงในฐานข้อมูลเสียงในระบบการสังเคราะห์เสียงจากขั้นตอนการสร้างแบบจำลองของหน่วยเสียง เทียบกับผลลัพธ์ของเสียงที่สังเคราะห์ได้จากระบบที่ใช้แบบจำลองที่สร้างมาจากค่าพารามิเตอร์ที่มาจากขั้นตอนการวิเคราะห์

พารามิเตอร์ พบว่าค่าพารามิเตอร์ที่ได้จากสัญญาณเสียงแต่ละค่า ณ ทุกตำแหน่งเฟรมที่สกัดในขั้นตอนการวิเคราะห์ มีความสัมพันธ์ไปในทิศทางเดียวกันกับค่าพารามิเตอร์ที่สกัดได้จากสัญญาณเสียงของระบบสังเคราะห์ที่นำเสนอ ดังรูปที่ 4.9 (เส้นประคือค่าพารามิเตอร์ที่วิเคราะห์ได้ เส้นทึบเป็นค่าพารามิเตอร์ที่ได้จากการสังเคราะห์) จึงสามารถสรุปได้ว่า ค่าพารามิเตอร์ที่วิเคราะห์เพื่อนำไปใช้สังเคราะห์เสียงมีความสอดคล้องกัน แสดงกระบวนการการวิเคราะห์ค่าพารามิเตอร์ การสกัดค่าพารามิเตอร์ และการสร้างแบบจำลองของเสียงเพื่อใช้สร้างแบบจำลองฮิดเดนมาร์คอฟโมเดลทำได้ถูกต้อง



รูปที่ 4.9 ค่าพารามิเตอร์ในการสังเคราะห์เสียง ได้จากแบบจำลองที่ฝึกฝน

ระบบการสังเคราะห์เสียงด้วยแบบจำลองฮิดเดนมาร์คอฟที่นำเสนอ

1. ขั้นตอนการฝึกฝนแบบจำลองเสียง

กระบวนการฝึกฝนแบบจำลองฮิดเดนมาร์คอฟ (ซึ่งอธิบายรายละเอียดการฝึกฝนแล้วในบทที่ 2 หัวข้อ 3.1 โดยใช้ชุดค่าพารามิเตอร์จากเสียงที่เตรียมไว้จากฐานข้อมูลเสียงตามวิธีที่ได้กล่าวไว้แล้วข้างต้นในบทนี้ ในการฝึกฝนแบบจำลองเสียงในระบบที่นำเสนอนี้ ค่าพารามิเตอร์แอลเอฟและระดับสัญญาณเสียงรบกวนซึ่งเป็นค่าต่อเนื่อง (Continuous value) จะทำในลักษณะเดียวกันฝึกฝนค่าความถี่มูลฐานในกรอบงานการสังเคราะห์เสียงด้วยแบบจำลองฮิดเดนมาร์คอฟ

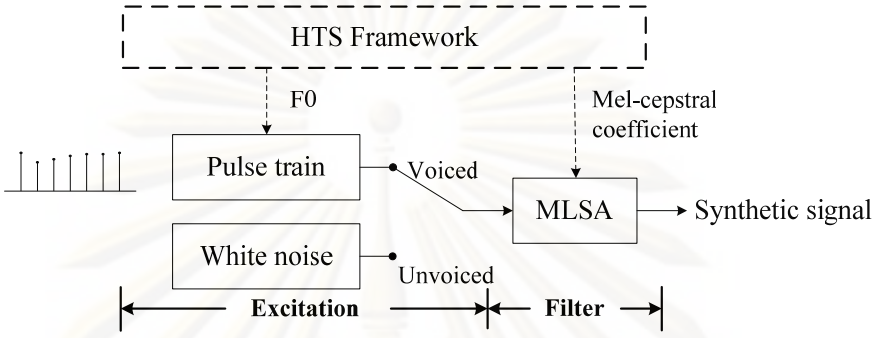
2. ขั้นตอนการสังเคราะห์เสียงพูด

การสังเคราะห์เสียงในระบบอ้างอิง สัญญาณกระตุ้นมาจากกระแสพัลส์สำหรับเสียงโฆษะและสลับกับเป็นสัญญาณรบกวนสีขาวสำหรับเสียงอโฆษะ ซึ่งพารามิเตอร์ที่ใช้ในการสร้างสัญญาณกระตุ้นในระบบอ้างอิงคือ ค่าความถี่มูลฐาน สำหรับระบบที่นำเสนอตามที่ได้ออกแบบไว้ในบทที่ 3 สัญญาณกระตุ้นได้จากสัญญาณเส้นเสียง รวมกับสัญญาณรบกวนตามอัตราส่วนสัญญาณต่อสัญญาณรบกวนในแต่ละเฟรม พารามิเตอร์ที่ใช้ในการสร้างสัญญาณกระตุ้นในระบบที่นำเสนอประกอบด้วย ค่าความถี่มูลฐาน พารามิเตอร์แอลเอฟ และระดับสัญญาณเสียงรบกวน แผนผังในส่วนการสังเคราะห์เสียงของระบบที่นำเสนอ และระบบอ้างอิงแสดงดังรูปที่ 4.10

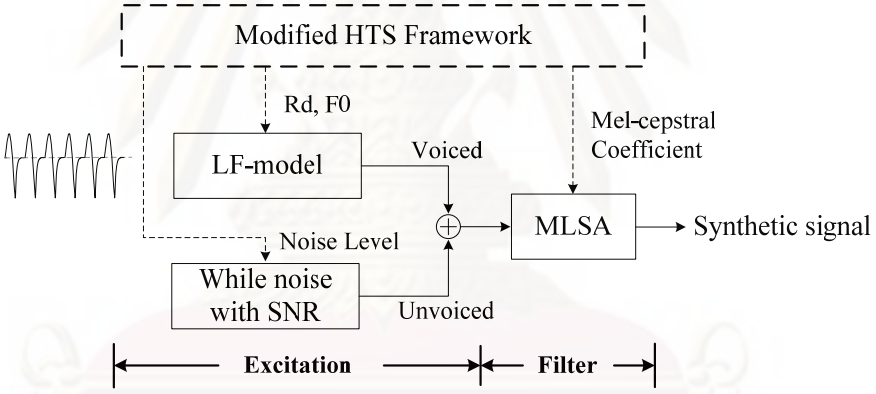
ในรูปที่ 4.10 (b) แสดงภาพรวมของส่วนการสังเคราะห์เสียงของระบบที่นำเสนอ เมื่อได้ค่าพารามิเตอร์จากแบบจำลองฮิดเดนมาร์คอฟในแต่เฟรมจากแบบจำลองเสียงที่ฝึกฝนไว้แล้ว ซึ่งประกอบด้วย ค่าพารามิเตอร์แอลเอฟ ระดับสัญญาณรบกวน ค่าความถี่มูลฐาน และสัมประสิทธิ์เมลเซปตรัม จากนั้นค่าพารามิเตอร์แอลเอฟ และค่าความถี่มูลฐานจะถูกแปลงเป็นสัญญาณเส้นเสียงโดยใช้สมการ (16) และ (15) ตามที่ได้กล่าวไว้แล้วในบทที่ 2 ลำดับต่อมาเพิ่มสัญญาณรบกวนสีขาวให้กับสัญญาณเส้นเสียงที่ได้ ให้มีปริมาณเท่ากับอัตราส่วนสัญญาณต่อสัญญาณรบกวนที่ได้จากแบบจำลอง ในขั้นตอนนี้จะได้สัญญาณกระตุ้นจากสัญญาณเส้นเสียง

สำหรับตัวกรองสัญญาณในการสังเคราะห์เสียง ใช้สัมประสิทธิ์เมลเซปตรัมเป็นอินพุทของอัลกอริทึมการประมาณค่าลอการิทึมของสเปกตรัมบนเมลสเกล (MLSA) ซึ่งจะได้ผลลัพธ์เป็นสเปกตรัมของช่องทางเดินเสียง ในการสังเคราะห์เสียงเมื่อนำสัญญาณกระตุ้นจากเส้นเสียงกระตุ้นไปกระตุ้นสเปกตรัมของช่องทางเดินเสียงจะได้สัญญาณเสียงตามพารามิเตอร์ที่ได้ในแต่เฟรม เมื่อนำทุกเฟรมมาเรียงต่อกันจะได้เป็นสัญญาณเสียงแบบต่อเนื่อง

สัมประสิทธิ์เมลเซปตรัมที่ใช้ในระบบที่นำเสนอนี้ ได้จากสัญญาณเสียงที่ถูกแยกส่วนประกอบจากสัญญาณเส้นเสียงออกแล้ว ซึ่งต่างกับสัมประสิทธิ์เมลเซปตรัมที่ใช้ในระบบอ้างอิงที่ได้จากสัญญาณเส้นเสียงโดยตรง



(a) Baseline System



(b) Proposed System

รูปที่ 4.10 ระบบการสังเคราะห์เสียงที่นำเสนอ และระบบอ้างอิง

2.1 การปรับลักษณะของสัญญาณเส้นเสียง

ค่าพารามิเตอร์ที่ได้จากต้นไม่ตัดสินใจ เมื่อนำไปใช้ในการสังเคราะห์เสียง เสียงสังเคราะห์ที่ได้จะมีลักษณะเสียงปกติ (Modal voice) ในการปรับรูปร่างของเส้นเสียง พารามิเตอร์ที่เกี่ยวข้องคือ ค่า Rd ของพารามิเตอร์แอลเอฟซึ่งจะมีผลต่อช่วงเปิดของสัญญาณเส้นเสียง ทำให้เกิดลักษณะเสียงบีบ เมื่อช่วงเปิดของสัญญาณเสียงลดลง และมีลักษณะเสียงลมหายใจเมื่อช่วงเปิดของสัญญาณเส้นเสียงเพิ่มขึ้น พารามิเตอร์อีกตัวหนึ่งที่เกี่ยวข้องกับสัญญาณเส้นเสียงคือ ระดับสัญญาณรบกวน ซึ่งเมื่อปรับให้มีระดับเพิ่มขึ้นจะส่งผลให้สัญญาณรบกวนทางลมหายใจเพิ่มขึ้น ทำให้สามารถรับรู้ลักษณะเสียงลมหายใจได้

บทที่ 5
ผลการวิจัย

ผลการวิจัย

วิทยานิพนธ์นี้ได้เสนอการสังเคราะห์เสียงโดยอาศัยแบบจำลองฮิดเดนมาร์คอฟให้สามารถกำหนดสัญญาณจากแหล่งกำเนิดเส้นเสียง และสัญญาณรบกวนลมหายใจโดยตรง โดยทำการเปรียบเทียบผลการสังเคราะห์เสียงจากวิธีที่นำเสนอกับเสียงสังเคราะห์จากระบบอ้างอิงได้แก่ การประเมินความเป็นธรรมชาติของเสียงสังเคราะห์ และการประเมินลักษณะของเสียง ด้วยการวิเคราะห์สัญญาณเสียงและการทดสอบจากผู้ฟัง ในวิทยานิพนธ์นี้ได้รายงานผลการวิเคราะห์ประกอบด้วย

1 ผลการประเมินความเป็นธรรมชาติของเสียงสังเคราะห์

2 ผลการประเมินลักษณะของเสียงด้วยการวิเคราะห์

ซึ่งผลการวิเคราะห์รายงานอย่างละเอียดดังนี้

1. ผลการประเมินความเป็นธรรมชาติของเสียงสังเคราะห์

การวัดความเป็นธรรมชาติด้วยผลการวัดคะแนนซีซีอาร์ตามกระบวนการที่ได้กล่าวไว้แล้วในตารางที่ 3.1 จากประโยค 15 ประโยคสุ่มเลือกจากฐานข้อมูลเสียงที่ไม่ถูกใช้การฝึกฝนแบบจำลองเสียง พารามิเตอร์ในการสังเคราะห์เสียงใช้พารามิเตอร์สำหรับสังเคราะห์เสียงปกติจากระบบสังเคราะห์เสียงที่นำเสนอเทียบกับระบบเสียงอ้างอิง ได้ผลตามตารางที่ 5.1 จากผลการทดลองแสดงให้เห็นว่าเสียงสังเคราะห์ที่ได้จากระบบที่นำเสนอมีความเป็นธรรมชาติกว่าเสียงสังเคราะห์ที่ได้จากระบบอ้างอิงเล็กน้อย

ตารางที่ 5.1 ผลคะแนน CCR เปรียบเทียบความเป็นธรรมชาติของเสียงสังเคราะห์

	ดีกว่า	เหมือนกัน	แย่กว่า
ระบบที่นำเสนอ	63	112	50

เพื่อเปรียบเทียบความแตกต่างของคะแนนที่ใช้วัดความเป็นธรรมชาติของเสียงสังเคราะห์จากทั้งสองระบบจากการพิจารณาคะแนนความพอใจด้วยการทดสอบนัยสำคัญทางสถิติตามที่ได้กล่าวไว้แล้วในบทที่ 3 โดยมีการทดสอบสมมติฐานคือ เปรียบเทียบความแตกต่างของคะแนนความเป็นธรรมชาติจากเสียงสังเคราะห์ทั้งสองระบบ วัดค่าความเชื่อมั่นโดยอย่างมีนัยสำคัญทางสถิติที่

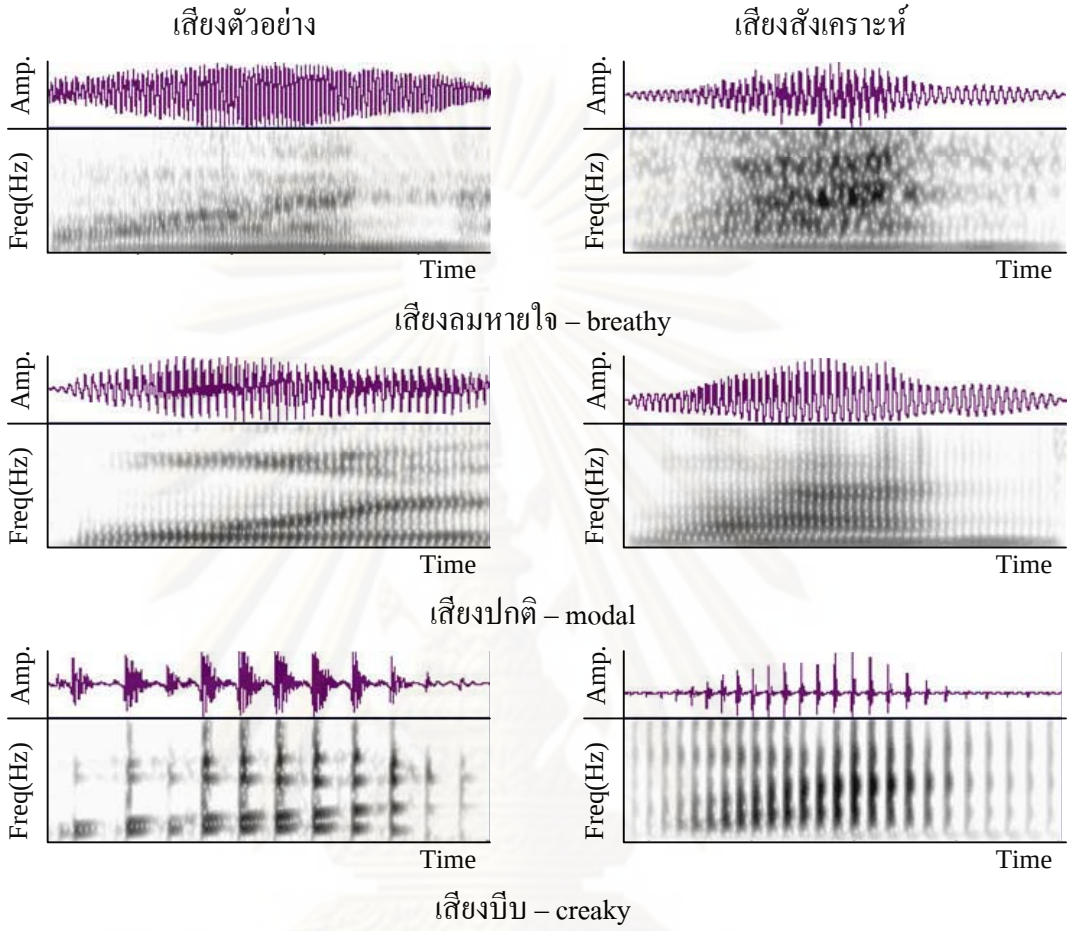
ระดับ 0.05 (ซึ่งระดับนี้ใช้เป็นเกณฑ์เปรียบเทียบการทดสอบความแตกต่างของคะแนนความเป็นธรรมชาติของทั้งสอง) จากการทดสอบพบว่า ค่าความเชื่อมั่น 0.3018 มากกว่า 0.05 ดังนั้นจึงสามารถสรุปได้ว่า ค่าเฉลี่ยระหว่างคะแนนที่วัดความเป็นธรรมชาติของเสียงสังเคราะห์จากระบบอ้างอิงและระบบที่นำเสนอไม่มีความแตกต่างกัน ชี้ให้เห็นว่าการใช้ค่าพารามิเตอร์จากแบบจำลองแหล่งกำเนิดเส้นเสียง ให้ผลการสังเคราะห์เสียงมีความเป็นธรรมชาติเช่นเดียวกับการใช้สัญญาณกระตุ้นแบบพัลส์ซึ่งใช้ทั่วไปในการสังเคราะห์เสียงด้วยแบบจำลองฮิดเดนมาร์คอฟ

2. ผลการประเมินลักษณะของเสียงด้วยการวิเคราะห์

2.1 จากการวิเคราะห์ลักษณะของเสียงจากสเปกโตรแกรม

เพื่อวิเคราะห์ลักษณะเสียงของเสียงสังเคราะห์จากระบบที่นำเสนอ จึงวิเคราะห์สัญญาณเสียงสังเคราะห์ด้วยสเปกโตรแกรมเพื่อดูสมบัติต่าง ๆ ของแต่ละชนิดของเสียงเปรียบเทียบลักษณะเสียงต่าง ๆ จากเสียงต้นแบบที่บันทึกจากการพูดของมนุษย์ ดังแสดงในรูปที่ 5.1 โดยคอลัมน์ด้านซ้ายมือแสดงสเปกโตรแกรมของเสียงต้นแบบได้จากเสียงจริงที่เลียนแบบเสียงลักษณะลมหายใจ เสียงปกติ และเสียงบีบ ตามลำดับ และในคอลัมน์ด้านขวาแสดงสเปกโตรแกรมของเสียงสังเคราะห์เรียงลำดับตามลักษณะเสียงต่าง ๆ

จากรูปที่ 5.1 พบว่าสเปกโตรแกรมของเสียงลมหายใจ (คู่บน) ซึ่งเปรียบเทียบเสียงต้นแบบที่เป็นเสียงจริงที่ออกเสียงลักษณะเสียงลมหายใจกับเสียงสังเคราะห์พบว่า สเปกโตรแกรมของเสียงสังเคราะห์มีลักษณะเป็นสัญญาณรบกวนกระจายอยู่ในทุกองค์ประกอบความถี่ บริเวณความถี่ฟอร์แมนท์ที่องค์ประกอบความถี่สูงไม่ชัด ซึ่งคล้ายกับสเปกโตรแกรมของเสียงลมหายใจที่เป็นเสียงจริง ขณะที่เมื่อเปรียบเทียบสเปกโตรแกรมของเสียงปกติ (คู่กลาง) พบว่าแม้ความถี่ฟอร์แมนท์ของเสียงจริงแสดงความชัดเจนมากกว่าสเปกโตรแกรมของเสียงสังเคราะห์ แต่เส้นการเคลื่อนที่ของความถี่ฟอร์แมนท์ของเสียงสังเคราะห์ยังคงมีตำแหน่ง และลักษณะคล้ายกับสเปกโตรแกรมของเสียงจริง และเมื่อเปรียบเทียบสเปกโตรแกรมเสียงบีบ (คู่ล่าง) พบว่าสเปกโตรแกรมของเสียงบีบที่ได้จากเสียงจริงมีลักษณะความชัดของแถบความกว้าง มีความกว้างมากขึ้น ความเข้มของความถี่ฟอร์แมนท์จะปรากฏเป็นช่วง ๆ ระยะห่างมากขึ้นเมื่อเทียบกับเสียงปกติ (ลักษณะความชัดของแถบกว้างที่ห่างมากขึ้นนี้เป็นสมบัติเด่นสำคัญของเสียงบีบ ที่เกิดจากช่วงเปิดของสัญญาณเส้นเสียงที่แคบลง และช่วงปิดของสัญญาณเป็นเวลานาน จึงทำให้จุดเริ่มต้นของช่วงเปิดสัญญาณเส้นเสียงที่อยู่ถัดไป มีระยะห่างมากขึ้น) ซึ่งเหมือนกับสเปกโตรแกรมของเสียงสังเคราะห์ลักษณะเสียงบีบมีความกว้างของแถบมากขึ้นกว่าเสียงสังเคราะห์ลักษณะเสียงปกติเช่นกัน



รูปที่ 5.1 สเปกโตรแกรมเปรียบเทียบเสียงสังเคราะห์ของคำว่า “วัน” ในลักษณะเสียงแบบต่าง ๆ

2.2 จากการวิเคราะห์ลักษณะของเสียงจากการวัดความถูกต้อง (Validity)

ผลการวัดความถูกต้องของการรับรู้ลักษณะเสียงของเสียงสังเคราะห์ตามกระบวนการที่กล่าวไว้ในบทที่ 3 หัวข้อ 3.3.2 จาก 15 ประโยคที่ถูกปรับค่าพารามิเตอร์ในการสังเคราะห์เสียงตามตารางที่ 3.3 ประกอบด้วย เสียงพูดปกติ เสียงพูดลักษณะบีบ และเสียงพูดแบบเสียงลมหายใจ แสดงดังตารางที่ 5.2 ประกอบด้วยร้อยละความถูกต้องของผลการตัดสินใจของผู้ฟังที่ได้ฟังเสียงประโยคชุดเสียงทดสอบที่ตรงกับตามคำตอบชนิดลักษณะเสียงของประโยคทดสอบที่กำหนดไว้เปรียบเทียบกับค่า HRF และค่า HNR ซึ่งเป็นค่าลักษณะสำคัญที่บอกความแตกต่างของลักษณะเสียงบีบ เสียงลมหายใจ และเสียงปกติ ตามที่ได้กล่าวไว้แล้วในบทที่ 3 เมื่อวิเคราะห์ผลการทดสอบพบว่า ผู้ฟังสามารถรับรู้ และตัดสินใจในเสียงลมหายใจได้ถูกต้องร้อยเปอร์เซ็นต์ ดังแสดงในตารางที่ 5.3 ซึ่งอาจจะเกิดจากเสียงลมหายใจมีลักษณะของสัญญาณรบกวนในปริมาณมากแตกต่างชัดเจนกับชนิดลักษณะเสียงปกติและเสียงบีบจึงสามารถตอบได้ถูกต้อง ขณะที่ผู้ฟังมีความสามารถในการ

รับรู้และตัดสินใจระบุเสียงปกติและเสียงบีบมีความถูกต้องเพียง 66.67 และ 58.67 เปอร์เซ็นต์ ตามลำดับ ทั้งนี้อาจเป็นเพราะความคล้ายคลึงของลักษณะเสียงสังเคราะห์ ผู้ฟังเกิดความลังเลในการตัดสินใจระบุชนิดของลักษณะเสียงระหว่างเสียงปกติและเสียงบีบ จากตารางที่ 5.3 เป็นตารางแสดงการปะปนระหว่างประเภทข้อมูล (Confusion matrix) ซึ่งแจกแจงจากผลการทดสอบจากตารางที่ 5.2 ชี้ให้เห็นว่าประโยคทดสอบที่เป็นเสียงปกติถูกพิจารณาเป็นเสียงบีบ 29.33 เปอร์เซ็นต์ และถูกระบุเป็นเสียงลมหายใจเพียง 4 เปอร์เซ็นต์ เช่นเดียวกันกับประโยคทดสอบที่เป็นเสียงบีบที่ถูกพิจารณาเป็นเสียงปกติ 36 เปอร์เซ็นต์ และถูกระบุเป็นเสียงลมหายใจเพียง 5.33 เปอร์เซ็นต์ แสดงให้เห็นว่าผู้ฟังไม่สามารถแยกความแตกต่างระหว่างเสียงบีบ และเสียงปกติได้ดีนัก แต่สามารถแยกเสียงลมหายใจออกจากเสียงอื่นได้

ตารางที่ 5.2 ผลการระบุความถูกต้องของลักษณะเสียงในแต่ละประโยค

ประโยคที่ใช้ทดสอบ	**คำตอบ**	HRF	HNR	% ความถูกต้องการรับรู้จากผู้ฟัง
ประโยคที่ 1	เสียงปกติ	-3.70	17.57	73.33
ประโยคที่ 2	เสียงปกติ	-5.22	18.17	46.67
ประโยคที่ 3	เสียงปกติ	-2.15	17.05	80.00
ประโยคที่ 4	เสียงปกติ	-2.61	17.27	73.33
ประโยคที่ 5	เสียงปกติ	-4.72	17.88	60.00
ประโยคที่ 6	เสียงลมหายใจ	-0.87	1.58	100.00
ประโยคที่ 7	เสียงลมหายใจ	-1.26	1.97	100.00
ประโยคที่ 8	เสียงลมหายใจ	-2.70	-3.44	100.00
ประโยคที่ 9	เสียงลมหายใจ	-9.60	5.53	100.00
ประโยคที่ 10	เสียงลมหายใจ	-10.13	-0.68	100.00
ประโยคที่ 11	เสียงบีบ	7.87	17.12	53.33
ประโยคที่ 12	เสียงบีบ	8.88	20.10	33.33
ประโยคที่ 13	เสียงบีบ	11.38	22.51	73.33
ประโยคที่ 14	เสียงบีบ	14.38	21.74	60.00
ประโยคที่ 15	เสียงบีบ	17.03	25.14	73.33

ตารางที่ 5.3 ผลการระบุความถูกต้องของลักษณะเสียงโดยรวม

		คำตอบที่ถูกต้อง		
		เสียงลมหายใจ	**เสียงปกติ**	**เสียงบีบ**
ผลการระบุจากผู้ฟัง	**เสียงลมหายใจ**	75 (100%)	3 (4%)	4 (5.33%)
	เสียงปกติ	-	50 (66.67%)	27 (36%)
	เสียงบีบ	-	22 (29.33%)	44 (58.67)
	รวม	75	75	75

เมื่อพิจารณาค่าลักษณะสำคัญของเสียงสังเคราะห์ตารางที่ 5.2 จากพบว่าค่า HNR ของเสียงลมหายใจในประโยคที่ทดสอบนี้มีค่าต่ำกว่าลักษณะเสียงชนิดอื่น ๆ อย่างเห็นได้ชัด เพราะระดับสัญญาณรบกวนที่เพิ่มขึ้นในสัญญาณเสียง ในขณะที่ค่า HNR ของสัญญาณเสียงปกติ และเสียงบีบจะมีค่าที่ไม่ต่างกันมากนัก เมื่อวิเคราะห์ค่า HRF พบว่าในเสียงสังเคราะห์ลักษณะบีบมีค่า HRF สูงกว่าลักษณะเสียงอื่น ๆ ซึ่งในเสียงปกติ และเสียงลมหายใจ จะมีค่า HRF ไม่ต่างกันอย่างชัดเจน จากการวิเคราะห์ค่า HRF และค่า HNR ของเสียงสังเคราะห์ประโยคที่เป็นเสียงปกติ เสียงลมหายใจและเสียงบีบทีละประโยค พบว่ามีค่าสอดคล้องตามทฤษฎีลักษณะเสียง และจากเปอร์เซ็นต์ความถูกต้องจากการรับรู้จากผู้ฟังในการฟังเสียงลักษณะต่าง ๆ สามารถระบุลักษณะเสียงได้อย่างถูกต้อง ดังนั้นจึงสามารถสรุปได้ว่าระบบเสียงสังเคราะห์ที่นำเสนอนี้ สามารถสังเคราะห์เสียงลักษณะเสียงต่าง ๆ ตามเป้าหมายได้ถึงแม้ผลการวัดเปอร์เซ็นต์ความถูกต้องจากผู้ฟังจะมีค่าเปอร์เซ็นต์ความถูกต้องไม่สูงนัก

2.3 จากการวิเคราะห์ลักษณะของเสียงซึ่งวัดความสามารถในการทำงานของระบบ

ผลการวิเคราะห์ค่าลักษณะสำคัญทางเสียงในชุดทดสอบดังตารางที่ 5.4 ซึ่งประเมินการรับรู้ความแตกต่างลักษณะของเสียงตามกระบวนการที่กล่าวแล้วในบทที่ 3 หัวข้อ 3.3.3 เพื่อวัดความสามารถในการทำงานของระบบที่สังเคราะห์เสียงลักษณะแบบต่าง ๆ โดยการปรับพารามิเตอร์ที่แตกต่างกันในแต่ละชุดทดสอบตามตารางที่ 3.5 ผลการวิเคราะห์ของแต่ละชุดการทดสอบ ประกอบด้วย ค่าความสัมพันธ์ของฮาร์โมนิกส์มูลฐานกับค่าฮาร์โมนิกส์อื่น ๆ (HRF) ค่าความสัมพันธ์ของฮาร์โมนิกส์กับสัญญาณรบกวนคือ HNR และผลต่างฮาร์โมนิกส์ที่หนึ่งและสอง (H1- H2) ซึ่งความหมายและของค่าวิเคราะห์ลักษณะสำคัญได้อธิบายแล้วในบทที่ 3

ตารางที่ 5.4 ผลการวิเคราะห์ค่าลักษณะสำคัญทางเสียงในชุดทดสอบ

ค่าทดสอบ	ระบบอ้างอิง				ระบบที่นำเสนอ			
	ชื่อ	**HRF**	**HNR**	**H1-H2**	**ชื่อ**	**HRF**	**HNR**	**H1-H2**
เสียงปกติ	**B1**	19.92	93.00	0.86	**P1**	-4.08	12.53	11.55
เสียงบีบ เมื่อ Rd เป็น 0.3 เท่า					**P2**	12.39	13.90	0.11
เสียงบีบที่ F0 ลดลง 100 Hz จากปกติ เมื่อ Rd เป็น 0.3 เท่า	**B3**	26.08	50.83	0.77	**P3**	13.52	15.71	0.27
เสียงบีบที่ F0 ลดลง 100 Hz จากปกติ เมื่อ Rd ปกติ					**P4**	-1.91	13.20	10.47
เสียงบีบที่ F0 ลดลง 50 Hz จากปกติ (ที่ Rd เป็น 0.3 เท่า)	**B5**	22.57	61.94	0.73	**P5**	15.77	15.47	-4.41
เสียงบีบ เมื่อ F0 ลดลง 50 Hz จากปกติ (ที่ Rd ปกติ)					**P6**	-0.63	18.92	7.65
เสียงบีบ เมื่อ F0 ลดลง 50 Hz จากปกติ (ที่ Rd 0.5 เท่า)					**P13**	9.49	14.5	-0.93
เสียงลมหายใจที่ SNR ลดลง 20 dB	**B7**	19.73	10.28	1.02	**P7**	-3.89	1.15	11.43
เสียงลมหายใจที่ Rd เป็น 2 เท่า เมื่อ SNR ลดลง 20 dB จากปกติ					**P8**	-9.23	-0.23	20.92
เสียงลมหายใจที่ SNR ลดลง 40 dB					**P12**	12.42	0.86	7.11
เสียงบีบที่ปนเสียงลมหายใจที่ F0 ปกติ เมื่อค่า Rd เป็น 0.3 เท่า และ SNR ลดลง 20 dB จากปกติ					**P11**	2.89	4.55	0.14
เสียงบีบที่ปนเสียงลมหายใจที่ F0 ลดลง 100 Hz จากปกติ เมื่อ Rd เป็น 0.3 เท่า SNR ลดลง 20 dB จากปกติ	**B9**	26.04	5.00	0.95	**P9**	13.49	9.13	0.28
เสียงบีบที่ปนเสียงลมหายใจที่ F0 ลดลง 50 Hz เมื่อ Rd เป็น 0.3 เท่า และ SNR ลดลง 20 dB	**B10**	23.15	8.43	0.72	**P10**	-8.13	3.35	16.09

ตารางที่ 5.5 ผลการเปรียบเทียบการวัดระดับความแตกต่างของลักษณะเสียงแบบต่าง ๆ

ชุดที่	ความแตกต่างของลักษณะเสียง	% ความสามารถบอกความแตกต่างในการรับรู้	
		ระบบอ้างอิง	ระบบที่นำเสนอ
1	ชุดเสียงปกติ – เสียงลมหายใจ โดยปรับระดับสัญญาณรบกวน	(1A) ชุดที่ B1 – B7 100%	(1B) ชุดที่ P1 – P7 100%
2	ชุดเสียงลมหายใจเมื่อ Rd ต่างกัน (โดยมี SNR ลดลง 20 dB จากปกติ)		ชุดที่ P7 – P8 53.33%
3	ชุดเสียงลมหายใจที่ค่า SNR ต่างกัน (โดยมี F0 และ Rd เป็นค่าปกติ)		ชุดที่ P7 – P12 100%
4	ชุดเสียงบีบที่ปนเสียงลมหายใจที่ Rd ต่างกัน (โดยมี F0 ปกติ และ SNR ลดลง 20 dB จาก)		ชุดที่ P7 – P11 86.67%
5	เสียงบีบที่ปนเสียงลมหายใจที่ F0 ต่างกัน (โดยมี SNR ลดลง 20 dB จากปกติ เป็นค่าคงที่ และระบบที่เสนอให้ Rd เป็น 0.3 เท่า)	(5A) ชุดที่ B9 – B10 60%	(5B) ชุดที่ P9 – P10 60%
6	ชุดเสียงปกติ-บีบเมื่อ F0 ต่างกัน (โดยมี F0 ลดลง 50 Hz จากปกติ)	(6A) ชุดที่ B1 – B5 60%	(6B) ชุดที่ P1 – P6 60%
7	ชุดเสียงปกติ-บีบเมื่อ Rd ต่างกัน (โดยมี F0 เป็นค่าปกติ)		ชุดที่ P1 – P2 80%
8	ชุดเสียงบีบเมื่อ F0 ต่างกัน (ระบบที่เสนอให้ Rd เป็น 0.3 เท่า)	(8A) ชุดที่ B3 – B5 40%	(8B) ชุดที่ P3 – P5 66.67%
9	ชุดเสียงบีบเมื่อ F0 ต่างกัน (โดยมี Rd เป็น 0.3 เท่า)	-	ชุดที่ P2 – P3 80%
10	ชุดเสียงบีบเมื่อ Rd ต่างกัน (โดยมี F0 ลดลง 50 Hz จากปกติ)	-	ชุดที่ P5 – P6 66.67%
11	ชุดเสียงบีบเมื่อ Rd ต่างกัน (โดยมี F0 ลดลง 100 Hz จากปกติ)	-	ชุดที่ P3 – P4 86.67%
12	ชุดเสียงบีบเมื่อ Rd ต่างกัน (โดยมี F0 ลดลง 50 Hz จากปกติ)	-	ชุดที่ P5 – P13 60%

จากตารางที่ 5.4 เมื่อพิจารณาลักษณะเสียงแบบต่าง ๆ ตามค่าพารามิเตอร์ที่ใช้วัด พบว่าเสียงสังเคราะห์ที่มีการปรับค่า Rd เช่นในเสียงบีบที่ได้จากระบบที่นำเสนอ P2 P3 P5 และ P7 จะมีค่า HRF สูง ขณะที่เสียงบีบได้จากระบบที่อ้างอิง (ไม่สามารถปรับค่า Rd ได้) และเมื่อพิจารณาระบบอ้างอิง โดยใช้เสียงสังเคราะห์ที่มีการปรับพารามิเตอร์ที่แตกต่างกันของเสียงแต่ละชนิด และความคิดเห็นของผู้ฟังที่รู้สึกต่อการระบุความแตกต่างลักษณะของเสียงแสดงดังตารางที่ 5.5

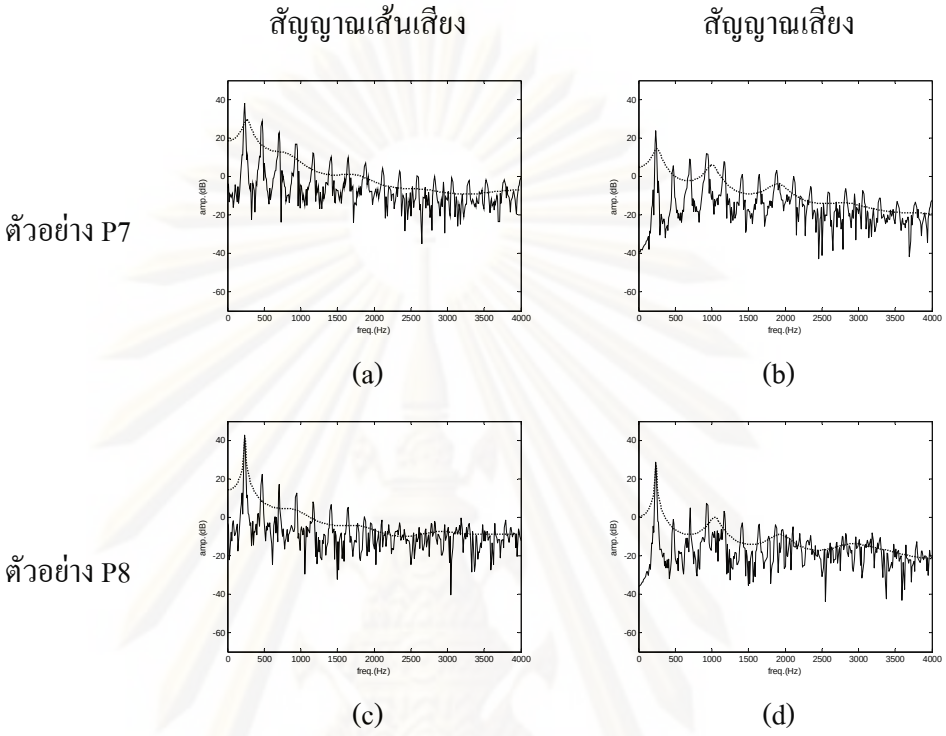
โดยพิจารณาตามปัจจัยที่มีผลกระทบต่อลักษณะของเสียงนั้น ๆ ดังนี้

1 ผลการเปรียบเทียบระดับความแตกต่างของลักษณะเสียงลมหายใจ

1.1 ผลการพิจารณาผลกระทบของ Rd บนเสียงลมหายใจ

จากผลการทดลองการรับรู้ของการเปรียบเทียบความแตกต่างของชุดทดสอบที่ 2 (P7-P8) เสียงลมหายใจเปรียบเทียบค่า Rd จากค่าปกติ และเมื่อเพิ่มค่า Rd เป็นสองเท่า ในขณะที่ระดับสัญญาณรบกวน SNR ลดลง 20 เดซิเบลเท่ากัน พบว่าความสามารถของผู้ฟังในการรับรู้ความแตกต่างของระดับเสียงลมหายใจของสองเสียงนี้เป็น 53.33% ตามตารางที่ 5.5 ซึ่งไม่แตกต่างกันมากนัก แสดงให้เห็นว่าผลจากการปรับจากค่า Rd ที่ทำให้เกิดการเปลี่ยนแปลงของรูปร่างของเส้นเสียง ไม่ส่งผลในด้านการรับรู้ถึงความแตกต่างลักษณะของเสียงลมหายใจได้อย่างชัดเจน

เมื่อวิเคราะห์ความแตกต่างลักษณะเสียงลมหายใจในโดเมนความถี่ โดยพิจารณาภาพตัดขวางของสเปกตรัมของสัญญาณเสียงสระ /a/ ดังรูปที่ 5.2 จากการเปรียบเทียบสเปกตรัมของเสียงลมหายใจเมื่อปรับค่า Rd เป็น 2 เท่าดังรูปที่ 5.2(c) พบว่าฮาร์โมนิกส์ที่ 1 สูงกว่าฮาร์โมนิกส์ที่ 1 ของสเปกตรัมสัญญาณเสียงที่ไม่มีการปรับค่า Rd ในรูปที่ 5.2(a) นอกจากนี้จากตารางที่ 5.4 ค่าผลต่างของฮาร์โมนิกส์ที่ 1 และ 2 (H1-H2) ของเสียงลมหายใจที่ปรับค่า Rd มีค่าเป็น 20.92 เดซิเบล ซึ่งมีค่าสูงกว่าผลต่างของฮาร์โมนิกส์ที่ 1 และ 2 ของเสียงลมหายใจที่ไม่ปรับค่า Rd ซึ่งมีค่าเป็น 11.43 เดซิเบล แตกต่างอย่างเห็นได้ชัด สำหรับค่า HNR มีค่าใกล้เคียงกันในชุดการทดลอง



รูปที่ 5.2 สเปกตรัมของสัญญาณเสียงชุดทดสอบเสียง ลมหายใจชุดที่ 2 (P7 – P8) เมื่อปรับค่า SNR ลดลง 20 dB

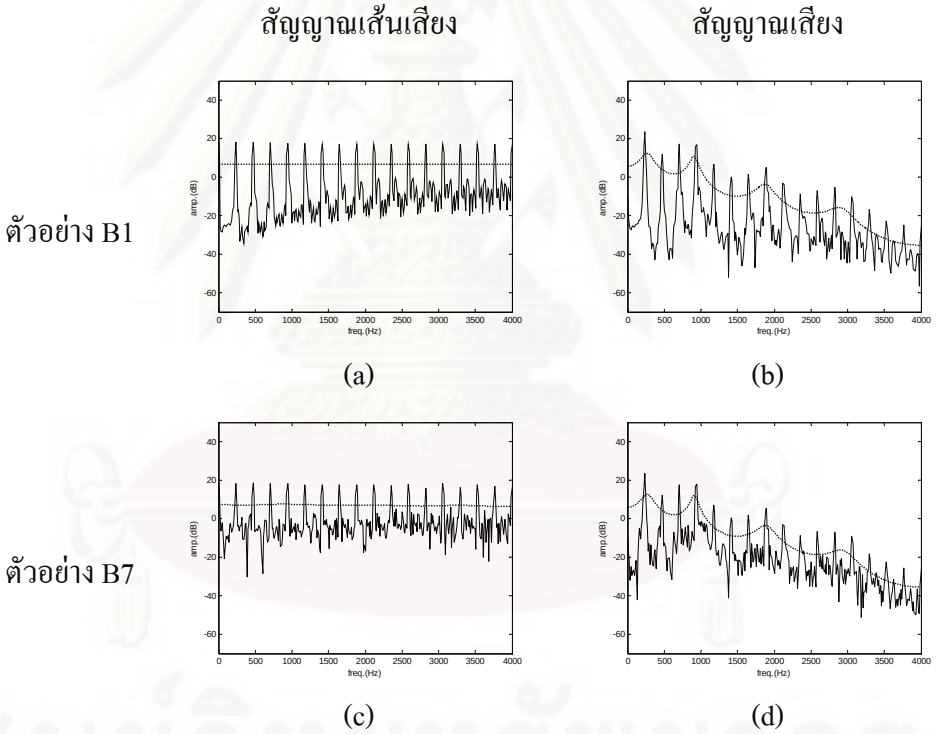
1.2 ผลการพิจารณาผลกระทบของค่า SNR บนเสียงลมหายใจ

จากผลการทดลองการเปรียบเทียบความแตกต่างลักษณะเสียงลมหายใจจากการปรับระดับสัญญาณรบกวน ในชุดที่ 1 (B1 - B7) และ (P1 - P7) ที่เปรียบเทียบเสียงปกติกับเสียงลมหายใจ ระหว่างระบบอ้างอิงและระบบที่นำเสนอตามลำดับ และการทดลองชุดที่ 3 (P7 – P12) ที่เปรียบเทียบเสียงลมหายใจที่มีค่า SNR ต่างกันจากระบบที่นำเสนอ โดยที่ทั้งสองชุดการทดลองใช้ค่า Rd เป็นค่าปกติ พบว่าสามารถรับรู้ความแตกต่างได้ 100% ตามตารางที่ 5.5 สรุปได้ว่าปริมาณของสัญญาณรบกวนที่เพิ่มขึ้นในเสียงสังเคราะห์ สามารถบอกความแตกต่างของลักษณะเสียงลมหายใจได้อย่างชัดเจน อีกทั้งผู้ฟังให้ความคิดเห็นว่า สามารถรับรู้ลักษณะเสียงลมหายใจได้จากระดับสัญญาณรบกวน

เมื่อวิเคราะห์ในโดเมนความถี่จากภาพตัดขวางของสเปกตรัมของสัญญาณเสียงสระ /a/ เพื่อบอกความแตกต่างลักษณะเสียงลมหายใจ ดังรูปที่ 5.3

รูปที่ 5.4 และรูปที่ 5.5 พบว่ามีลักษณะของสัญญาณรบกวนปรากฏในเส้นสเปกตรัมของสัญญาณเส้นเสียง และสัญญาณรบกวนทำให้รูปร่างซองของสเปกตรัม (Spectral envelope) ของสัญญาณเสียงในองค์ประกอบความถี่สูงเกิดเป็นลักษณะแบน (flat) มากขึ้น หรือรูปร่างของความถี่ฟอร์แมนท์ไม่ปรากฏชัดเจนเมื่อเปรียบเทียบกับเสียงปกติ

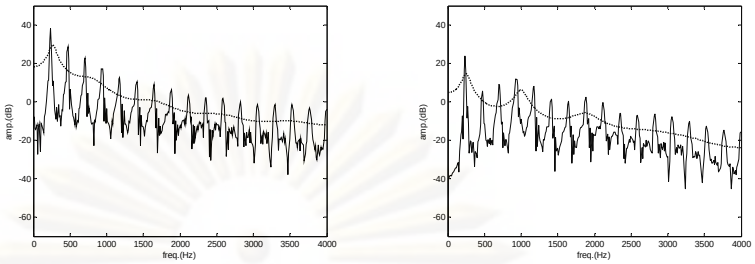
จากรูปที่ 5.3 ที่แสดงสเปกตรัมของเสียงปกติ – ลมหายใจจากระบบอ้างอิง (ชุดที่ 1A) โดยการเพิ่มสัญญาณรบกวนสีขาวเข้าไปในสัญญาณพัลส์ พบว่ารูปร่างของสัญญาณกระตุ้นที่เพิ่มสัญญาณรบกวนสีขาวเข้าไปแล้วในรูปที่ 5.3(c) มีลักษณะของสัญญาณรบกวนบนสเปกตรัม ต่างจากสัญญาณกระตุ้นในรูปที่ 5.3(a)



รูปที่ 5.3 สเปกตรัมของสัญญาณเสียงชุดทดสอบเสียงปกติ-ลมหายใจชุดที่ 1A (B1-B7) เมื่อเพิ่มสัญญาณรบกวนสีขาว

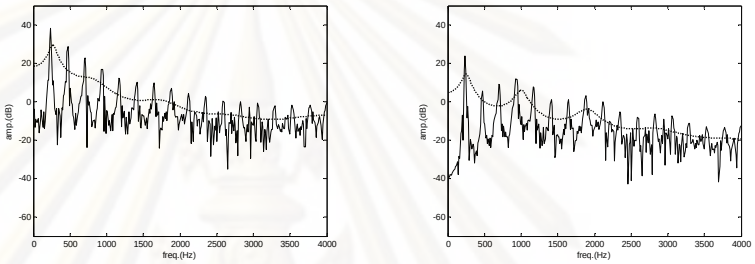
สัญญาณเส้นเสียง สัญญาณเสียง

ตัวอย่าง P1



(a) (b)

ตัวอย่าง P7

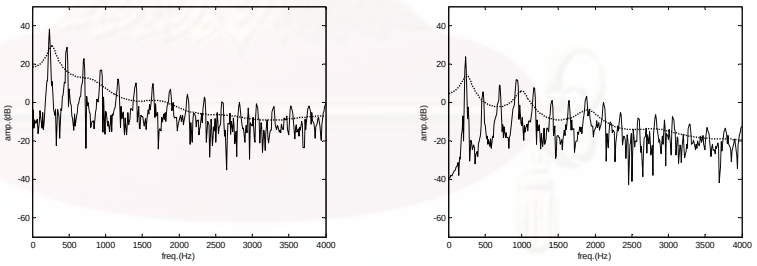


(c) (d)

รูปที่ 5.4 สเปกตรัมของสัญญาณเสียงชุดทดสอบเสียงปกติ-ลมหายใจชุดที่ 1B (P1 – P7) เมื่อปรับค่า SNR ลดลง 20 dB

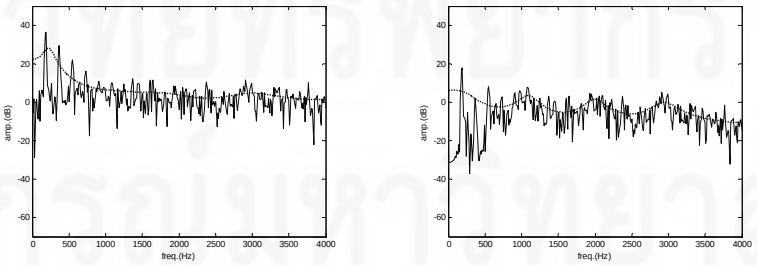
สัญญาณเส้นเสียง สัญญาณเสียง

ตัวอย่าง P7



(a) (b)

ตัวอย่าง P12

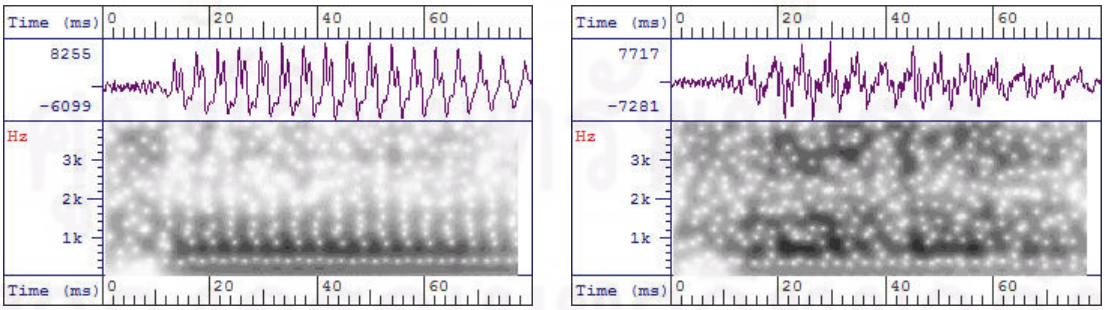


(c) (d)

รูปที่ 5.5 สเปกตรัมของสัญญาณเสียงชุดทดสอบเสียงลมหายใจชุดที่ 3 (P7 – P12) เมื่อปรับค่า SNR ลดลง 20 dB และ 40 dB

ขณะที่ชุดการทดลอง 1B ซึ่งเปรียบเทียบเสียงปกติ – เสียงลมหายในจากระบบที่นำเสนอโดยการปรับ SNR ลดลง 20 เดซิเบล จากค่าปกติ และชุดที่ 3 ที่เปรียบเทียบเสียงลมหายใจโดยปรับ SNR ต่างกับ (ลดลง 20 และ 40 เดซิเบลจากค่าปกติ) พบว่าสเปกตรัมของเสียงลมหายใจในรูปที่ 5.4(d) และรูปที่ 5.5(d) มีสัญญาณรบกวนเด่นชัดที่องค์ประกอบความถี่สูง ขณะที่ไม่ปรากฏในสเปกตรัมของเสียงปกติรูปที่ 5.4(b) และรูปที่ 5.5 (b) ซึ่งสัญญาณรบกวนนี้ จะส่งผลกระทบทำให้ความถี่ฟอร์แมนท์ที่องค์ประกอบความถี่สูงของเสียงลมหายใจไม่ชัดเจน ดังรูปที่ 5.6 แสดงสเปกโตรแกรมของเสียงลมหายใจที่มีสัญญาณรบกวนลดลง 20 และ 40 เดซิเบล ตามลำดับ

นอกจากนี้ยังได้เปรียบเทียบความแตกต่างการรับรู้ของลักษณะเสียงลมหายใจที่มีการเพิ่มสัญญาณรบกวน โดยพิจารณาจากค่า HRF และค่า HNR ของเสียงสังเคราะห์แต่ละคู่ประโยคในชุดทดสอบ เพื่อวิเคราะห์ความเป็นลักษณะเสียงลมหายใจซึ่งพบว่า ชุดทดสอบเสียงลมหายใจเมื่อปรับค่า SNR ลดลง ต่างกัน 20 และ 40 เดซิเบล (P7-P12) มีค่า HRF และค่า HNR แตกต่างกันอย่างชัดเจน ซึ่งเกิดจากผลกระทบของการเพิ่มระดับปริมาณสัญญาณรบกวนในปริมาณมาก ต่างกับชุดทดสอบที่เปรียบเทียบเสียงปกติและเสียงลมหายใจที่มีค่า HRF และค่า HNR ไม่แตกต่างกันมากนัก ดังนั้นจึงสามารถสรุปได้ว่า ปริมาณสัญญาณรบกวนที่เพิ่มเข้าไปในสัญญาณเสียงนี้มีผลกระทบต่อการสังเคราะห์เสียงลมหายใจอย่างมาก แต่ทั้งนี้ต้องคำนึงถึงปริมาณสัญญาณรบกวนที่เพิ่มให้ไม่เกินค่าที่เหมาะสมที่สามารถรับรู้และได้ยินเสียงสังเคราะห์ได้



ตัวอย่าง P7 ตัวอย่าง P12

รูปที่ 5.6 สเปกโตรแกรมของชุดทดสอบเสียงลมหายใจ
เมื่อปรับค่า SNR ลดลง 20 และ 40 dB (P7 – P12)

จากผลการรับรู้ความแตกต่างของลักษณะเสียงลมหายใจ และการวิเคราะห์จากสเปกตรัมสอดคล้องกับ ทฤษฎีการรับรู้ลักษณะเสียงลมหายใจที่ได้อธิบายแล้วในบทที่ 2 สามารถสรุปได้ว่าการเพิ่มระดับของสัญญาณรบกวน มีผลต่อการรับรู้ความแตกต่างลักษณะสัญญาณเสียงลมหายใจ ได้ดีกว่าการปรับช่วงเปิดของสัญญาณเส้นเสียงให้กว้างขึ้นด้วยการปรับค่า Rd

2 การเปรียบเทียบระดับความแตกต่างของลักษณะเสียงบีบ

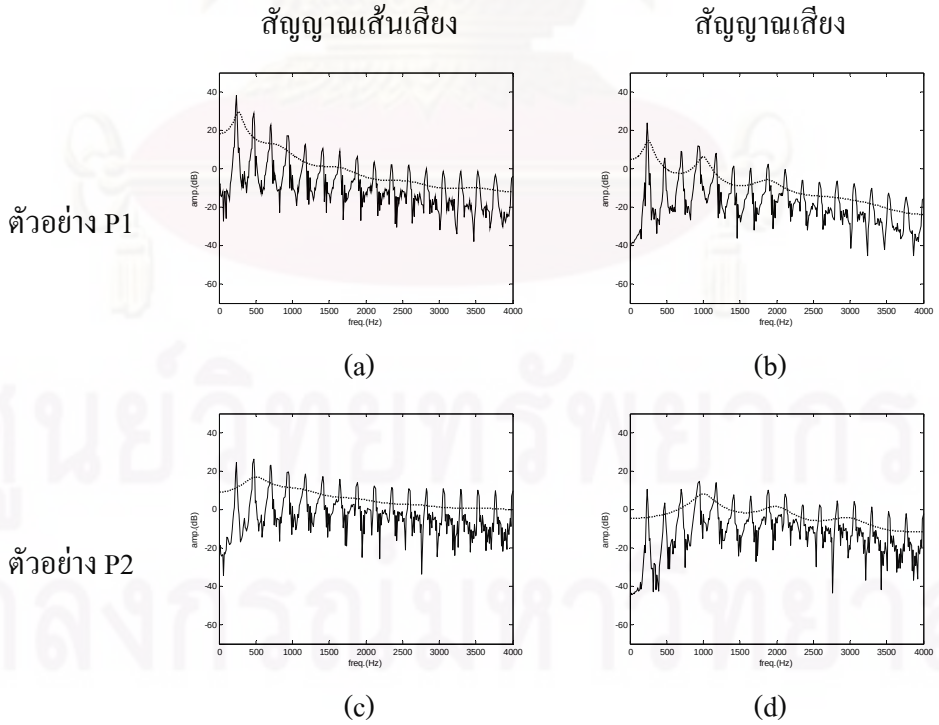
2.1 ผลการพิจารณาผลกระทบของ Rd บนเสียงบีบ

จากผลการทดลองการเปรียบเทียบความแตกต่างของเสียงปกติและเสียงบีบที่มีการปรับช่วงเปิดสัญญาณเส้นเสียงของชุดที่ 7 (P1 – P2) และชุดการทดลองที่ 10 (P5 – P6) การทดลองที่ 11 (P3 – P4) และการทดลองที่ 12 (P5 – P13) ที่มีการปรับค่าช่วงเปิดสัญญาณเส้นเสียง เมื่อกำหนดให้ค่า F0 เป็นค่าคงที่ พบว่าความสามารถของบอกการรับรู้ความแตกต่างของลักษณะเสียงบีบเมื่อปรับค่า Rd สามารถบอกได้เป็น 80% 66.67% 86.67% และ 60% ตามตารางที่ 5.5 ซึ่งแตกต่างกันอย่างชัดเจน โดยเฉพาะชุดการทดลองที่ปรับค่าความถี่มูลฐานให้มีค่าต่ำลงมาก ๆ ส่งผลให้ความสามารถการบอกความแตกต่างลักษณะเสียงบีบชัดเจนมากขึ้น และเมื่อพิจารณาชุดการทดลองที่ 10 และชุดการทดลองที่ 12 ซึ่งเป็นการเปรียบเทียบการปรับค่า Rd ที่มีความถี่มูลฐานลดลงจากค่าปกติ 50 Hz พบว่าผู้ฟังสามารถรับรู้ความแตกต่างของเสียงสังเคราะห์ในชุดการทดลองที่มีการเปลี่ยนแปลงค่า Rd สูง (จากค่าปกติเป็นลดลง 0.3 เท่า) ได้ดีกว่า สรุปได้ว่าผลกระทบจากค่า Rd ทำให้เกิดการเปลี่ยนแปลงของรูปร่างของเส้นเสียง โดยจะมีช่วงเปิดสัญญาณเสียงจะแคบลง ซึ่งค่า Rd นี้เป็นปัจจัยสำคัญของการสังเคราะห์และรับรู้ลักษณะของเสียงบีบ นอกจากนี้ความเห็นของผู้ฟังขณะทดสอบให้ความเห็นว่า สามารถได้ยินเสียงสังเคราะห์ที่ทดสอบมีลักษณะการบีบของเสียงได้ยินเด่นชัด มีลักษณะการเกร็ง บีบของเส้นเสียงต่างจากเสียงปกติ แต่การตัดสินใจเพื่อบอกระดับความแตกต่างของการรับรู้จากการฟังว่ามีความเกร็งของเส้นเสียงมากหรือน้อยต่างกันนั้น ตัดสินใจยาก

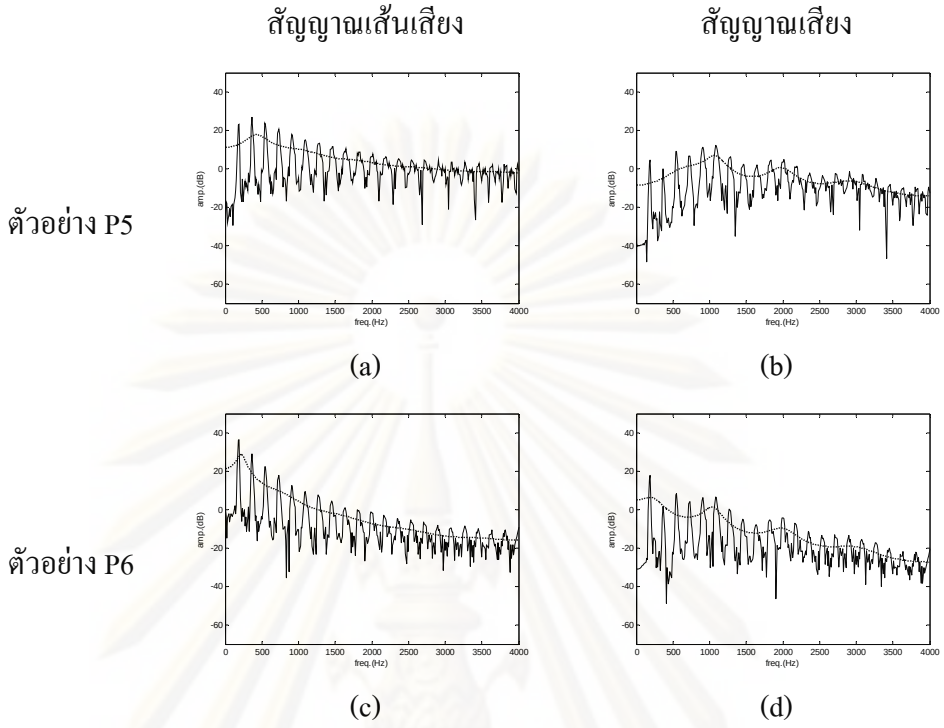
จากการวิเคราะห์ผลในโดเมนความถี่ของความแตกต่างลักษณะเสียงบีบจากการปรับลักษณะสัญญาณเส้นเสียงของชุดการทดลองข้างต้นพบว่า เสียงบีบที่

มีการปรับค่า Rd เป็น 0.3 พบว่าสเปกตรัมของเสียงบีบในรูปที่ 5.7(c) รูปที่ 5.8(c) รูปที่ 5.9(c) และรูปที่ 5.10(c) มีค่าฮาร์โมนิกส์ที่ 1 (H1) แตกต่างอย่างชัดเจนกับสเปกตรัมของเสียงบีบในรูปที่ 5.7(a) รูปที่ 5.8(a) รูปที่ 5.9(a) และรูปที่ 5.10(a) ที่ใช้ค่า Rd ต่างกัน

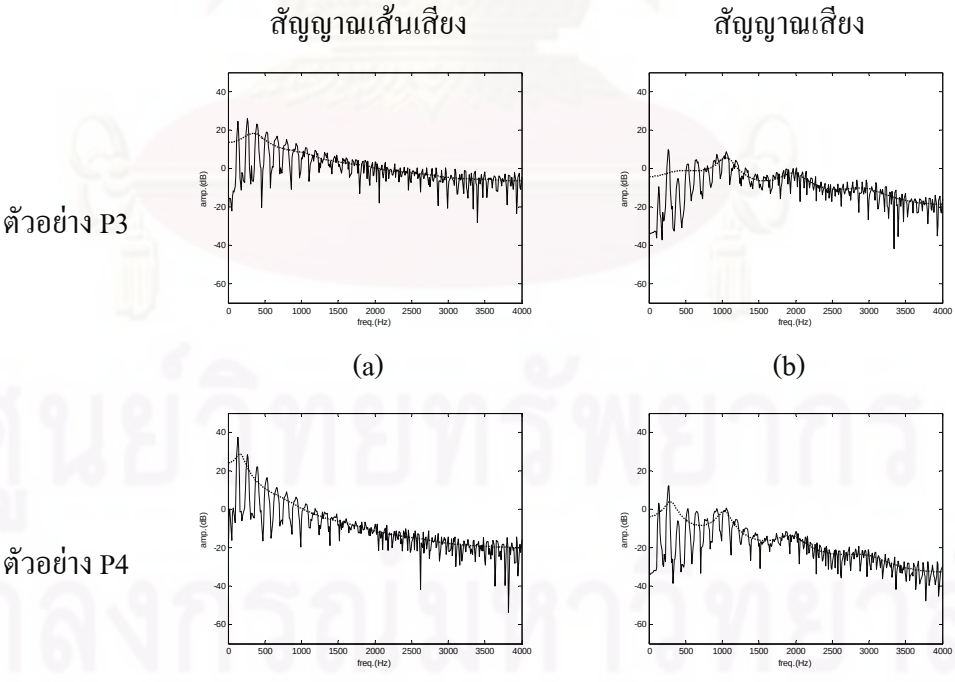
นอกจากนี้จากการวิเคราะห์ค่า HRF และผลต่างของฮาร์โมนิกส์ที่หนึ่งและสอง พบว่าเสียงที่มีการปรับค่า Rd ลดลงจะมีค่า HRF สูง และมีค่าผลต่างของฮาร์โมนิกส์ที่หนึ่งและสองต่ำ ซึ่งแสดงสมบัติของลักษณะเสียงบีบ มากกว่าเสียงที่ไม่มีการปรับค่า Rd ดังนั้นในชุดการทดลองที่เปรียบเทียบเสียงบีบที่ปรับค่า Rd เทียบกับเสียงบีบที่ใช้ค่า Rd ปกติ จึงสามารถบอกความแตกต่างได้ชัดเจน ตัวอย่างเช่นในชุดการทดลองที่ 11 (P3 – P4) ซึ่งเปรียบเทียบเมื่อค่า Rd ต่างโดยมีค่าความถี่มูลฐานลดลง 100 Hz จากปกติ พบว่าค่า HRF ของเสียงบีบมีความแตกต่างกันอย่างมาก ซึ่งทำให้สามารถสรุปได้ว่าค่า HRF จะมีการเปลี่ยนแปลงขึ้นอยู่กับค่า Rd ในกรณีที่ค่าความถี่มูลฐานต่ำมากพอ ผู้ฟังจะสามารถรับรู้ลักษณะเสียงบีบได้



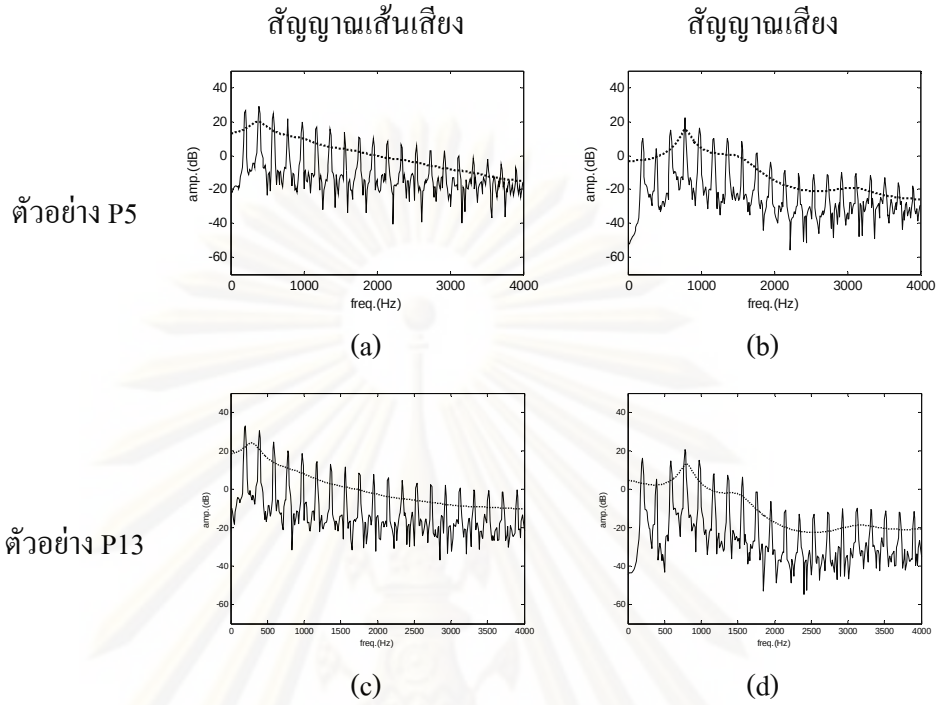
รูปที่ 5.7 สเปกตรัมของสัญญาณเสียงชุดทดสอบเสียงปกติ-บีบชุดที่ 7 (P1 – P2) เมื่อปรับค่า Rd เป็น 0.3 เท่าของค่าปกติ



รูปที่ 5.8 สเปกตรัมของสัญญาณเสียงชุดทดสอบเสียงบีบชุดที่ 10 (P5 – P6)
เมื่อปรับ Rd เป็น 0.3 เท่าของค่าปกติ ที่ค่า F0 ลดลง 50 Hz



รูปที่ 5.9 สเปกตรัมของสัญญาณเสียงชุดทดสอบเสียงบีบชุดที่ 11 (P3 – P4)
เมื่อปรับค่า Rd เป็น 0.3 เท่าของค่าปกติ ที่ F0 ลดลง 100 Hz



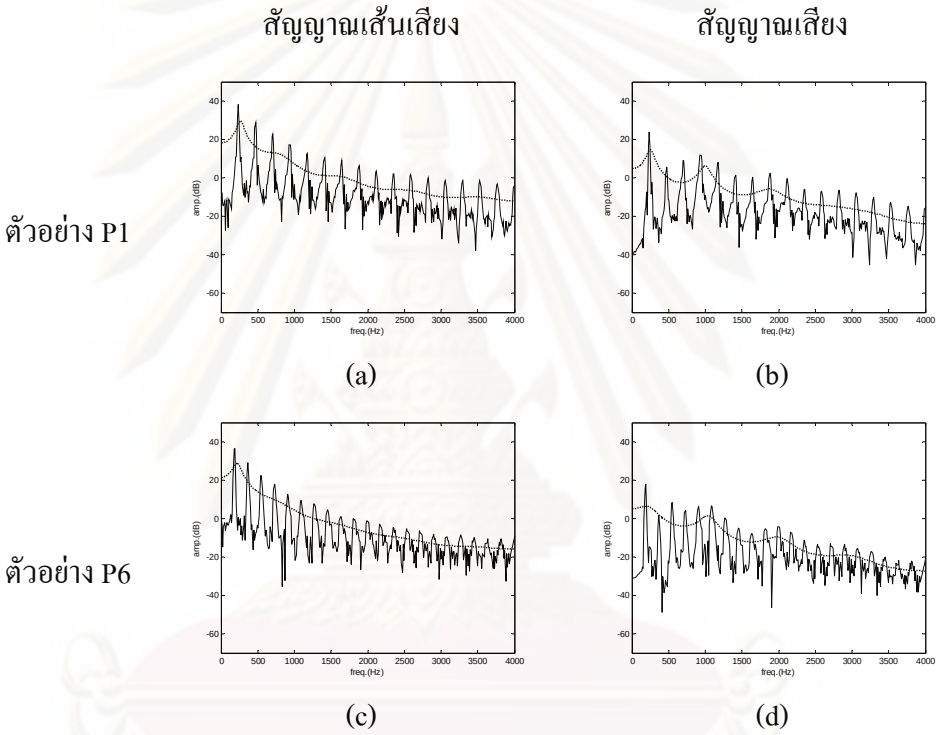
รูปที่ 5.10 สเปกตรัมของสัญญาณเสียงชุดทดสอบเสียงบีบชุดที่ 12 (P5 – P13) เมื่อปรับ Rd เป็น 0.3 และ 0.5 เท่าของค่าปกติ ที่ F0 ลดลง 100 Hz

2.2 ผลการพิจารณาผลกระทบของความถี่มูลฐานบนเสียงบีบ

จากผลการทดลองการเปรียบเทียบเสียงปกติและเสียงบีบที่มีการปรับค่าความถี่มูลฐานของชุดที่ 6 (P1 – P6) ซึ่งมีค่า Rd เป็นค่าปกติและชุดที่ 8 (P3 – P5) กับ ชุดที่ 9 (P2 – P5) ซึ่งมีค่า Rd เป็น 0.3 เท่า พบว่าความถี่มูลฐานสามารถบอกการรับรู้ความแตกต่างของลักษณะเสียงบีบได้ดี ซึ่งค่าความสามารถการบอกระดับความแตกต่างของลักษณะเสียงมีค่าเป็น 60%, 66.67% และ 80% ตามลำดับ ดังแสดงในตารางที่ 5.5 สรุปได้ว่า ผลกระทบจากค่าความถี่มูลฐานที่แตกต่างกันทำให้เกิดการสั่นที่ความถี่ที่ต่ำกว่าปกติ รับรู้ได้ถึงลักษณะของเสียงบีบที่ได้อย่างชัดเจน และสามารถรับรู้การเปลี่ยนแปลงได้มากขึ้นเมื่อค่าความถี่มูลฐานมีความแตกต่างกันมากตามชุดการทดลองที่ 9 ที่มีค่าความถี่มูลฐานต่างกัน 100 Hz

สำหรับผลการทดสอบกับระบบอ้างอิงในชุดการทดลองที่ 6A และชุดการทดลองที่ 8A พบว่าสามารถรับรู้ความแตกต่างของเสียงสังเคราะห์ได้บ้าง แต่เมื่อความถี่ฐานมีค่าต่ำลงมาก ความสามารถในการบอกความแตกต่างจะลดลง

จากการวิเคราะห์ผลในโดเมนความถี่ของความแตกต่างลักษณะเสียงบีบ จากการปรับลักษณะสัญญาณเส้นเสียงของชุดการทดลองข้างต้นพบว่า รูปที่ 5.11(a) รูปที่ 5.12(a) และรูปที่ 5.13(a) ตำแหน่งของฮาร์โมนิกส์ที่ 1 (H1) และฮาร์โมนิกส์ที่ 2 (H2) จะแตกต่างกับในรูปที่ 5.11(c) รูปที่ 5.12(c) และรูปที่ 5.13(c) ซึ่งเป็นผลเนื่องจากการปรับค่าความถี่มูลฐาน ในขณะที่ตำแหน่งของความถี่ฟอร์แมนท์มีค่าไม่เปลี่ยนแปลง

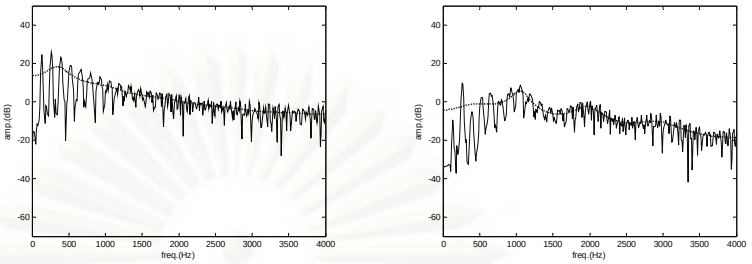


รูปที่ 5.11 สเปกตรัมของสัญญาณเสียงชุดทดสอบเสียงบีบชุดที่ 6 (P1 – P6) เมื่อปรับ F0 ลดลง 50 Hz ที่ค่า Rd เป็นค่าปกติ

จากผลการรับรู้ความแตกต่างของลักษณะเสียงบีบ และการวิเคราะห์จากสเปกตรัมสอดคล้องกับทฤษฎีการรับรู้ลักษณะเสียงบีบ สามารถสรุปการสังเคราะห์เสียงในลักษณะเสียงบีบทำได้โดยการปรับค่า Rd ลงซึ่งจะส่งผลให้ช่วงเปิดของสัญญาณเส้นเสียงแคบลง และการลงค่าความถี่มูลฐาน โดยการปรับค่าพารามิเตอร์ทั้งสองส่งผลถึงการรับรู้ลักษณะเสียงบีบของผู้ฟังในปริมาณที่ต่างกัน ทั้งนี้ขึ้นอยู่กับประสบการณ์ และการตัดสินใจของผู้ฟังแต่ละคน ขณะที่ค่า HRF ซึ่งมีลักษณะของค่าเด่นชัดในเสียงสังเคราะห์ลักษณะบีบที่มีปรับค่า Rd แต่ในกรณีที่ความถี่มูลฐานของเสียงสังเคราะห์ลดลงมากระดับหนึ่ง ผู้ฟังจะสามารถรับรู้ความเป็นเสียงบีบได้ โดยที่ค่า HRF จะมีค่าไม่สอดคล้องตามทฤษฎีเสียงบีบ

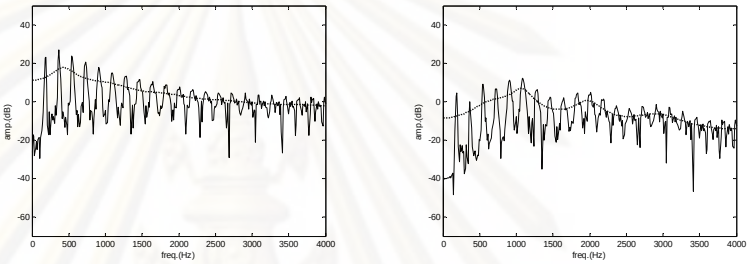
สัญญาณเส้นเสียง สัญญาณเสียง

ตัวอย่าง P3



(a) (b)

ตัวอย่าง P5

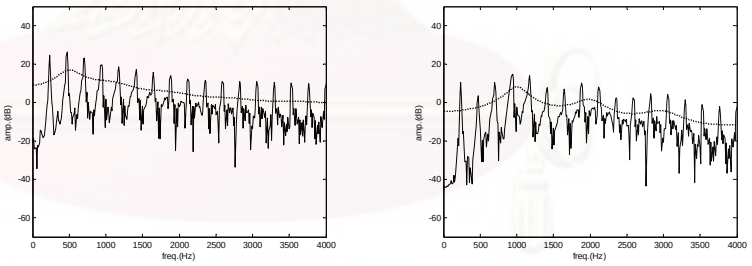


(c) (d)

รูปที่ 5.12 สเปกตรัมของสัญญาณเสียงชุดทดสอบเสียงบีบชุดที่ 8 (P3 – P5)
เมื่อปรับ F0 ลดลง 100 และ 50 Hz ที่ค่า Rd เป็น 0.3 เท่าของค่าปกติ

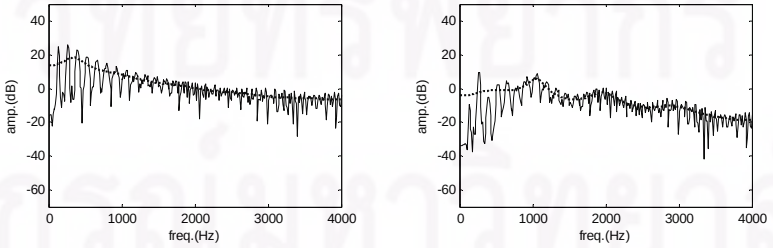
สัญญาณเส้นเสียง สัญญาณเสียง

ตัวอย่าง P2



(a) (b)

ตัวอย่าง P15



(c) (d)

รูปที่ 5.13 สเปกตรัมของสัญญาณเสียงชุดทดสอบเสียงบีบชุดที่ 9 (P2 – P3)
เมื่อปรับ F0 ลดลง 100 Hz ที่ค่า Rd เป็น 0.3 เท่าของค่าปกติ

จากตารางผลการเปรียบเทียบระดับความแตกต่างลักษณะของเสียง ชี้ว่าการบอกความแตกต่างของเสียงลมหายใจที่ได้รับผลกระทบจากการเพิ่มสัญญาณรบกวน ซึ่งเสียงสังเคราะห์ลมหายใจที่มาจากทั้งระบบสังเคราะห์เสียงที่นำเสนอที่ใช้การปรับระดับสัญญาณเสียงรบกวน ช่วยให้เสียงที่สังเคราะห์มานั้นมีความเป็นเสียงลมหายใจ และระบบอ้างอิงใช้การเพิ่มสัญญาณรบกวนสีขาวนั้น สามารถบอกความแตกต่างระหว่างเสียงปกติและเสียงลมหายใจได้ในเกณฑ์ที่ดี ซึ่งสามารถแยกความแตกต่างได้อย่างชัดเจนทั้งสองระบบ สามารถสรุปได้ว่า เสียงสังเคราะห์ลมหายใจได้รับผลกระทบจากการเพิ่มปริมาณสัญญาณรบกวนในการสังเคราะห์เสียง ถึงแม้สำหรับระบบที่นำเสนออาศัยค่าพารามิเตอร์ Rd เพื่อให้ลักษณะเส้นเสียงมีความหลากหลายแตกต่างกัน แต่พบว่าไม่ได้รับผลกระทบต่อการบอกความแตกต่างของเสียงสังเคราะห์แบบลมหายใจ ระบบสังเคราะห์เสียงในวิทยานิพนธ์นี้ เพื่อช่วย SNR ให้เสียงที่สังเคราะห์มานั้นมีความเป็นเสียงลมหายใจ และบีบ สามารถสรุปได้ว่า การรับรู้ของคนไม่สามารถแยกความแตกต่างของปริมาณสัญญาณรบกวนที่แตกต่างภายในเสียงลมหายใจได้

วิทยานิพนธ์นี้ได้วัดความค่าความแตกต่างของเสียงบีบ พบว่าเสียงบีบที่สังเคราะห์จากระบบที่นำเสนอมีความแตกต่างของลักษณะของความเป็นเสียงบีบมากกว่าเสียงที่สังเคราะห์จากระบบอ้างอิง และระบบการสังเคราะห์เสียงที่นำเสนอสามารถสังเคราะห์ระดับเสียงบีบได้หลากหลายมากกว่าระบบสังเคราะห์เสียงอ้างอิง เนื่องจากระบบที่นำเสนอสามารถปรับค่าพารามิเตอร์ค่าความถี่มูลฐาน และค่า Rd ซึ่งมีผลต่อการสร้างการบีบตัวของรูปร่างของเส้นเสียงได้ ขณะที่ระบบอ้างอิงสามารถปรับได้เพียงค่าความถี่มูลฐาน

นอกจากนี้ วิทยานิพนธ์นี้ได้สอบถามความคิดเห็นจากผู้ฟังเกี่ยวกับประเมินผลการวัดระดับความแตกต่างลักษณะของเสียง ความคิดเห็นของผู้ฟังจากการทดสอบการรับรู้ความแตกต่างของเสียงลมหายใจ ผู้ฟังให้ความเห็นว่าสามารถรับรู้ว่ามีปริมาณเสียงรบกวน หรือเสียง “ซ่า” ในเสียงที่ทดสอบมากน้อยได้ ขณะที่การทดสอบการรับรู้เสียงบีบและเสียงปกติ พบว่าผู้ฟังบางคนให้ความเห็นสามารถรับรู้ว่าเป็นลักษณะของเสียงดังกล่าว แต่ยากต่อการวัดหรือแยกการรับรู้ของส่วนที่แตกต่างระหว่างลักษณะของเสียงได้ ผู้ฟังมีความลังเลในการตอบแบบสอบถามบ้าง เนื่องจากการขาดประสบการณ์ของการรับฟังเสียงสังเคราะห์ และขาดประสบการณ์การแยกแยะลักษณะของเสียงปกติ เสียงบีบ เช่น มีความเห็นว่าเสียงปกติที่ได้มีลักษณะคล้ายเสียงบีบ เป็นต้น

บทที่ 6
บทสรุปผลการวิจัย และข้อเสนอแนะ

สรุปผลการวิจัย

วิทยานิพนธ์นี้ได้พัฒนาระบบสังเคราะห์เสียงที่ดัดแปลงการสังเคราะห์เสียงตามแบบจำลองฮิดเดนมาร์คอฟที่สามารถกำหนดสัญญาณจากแหล่งกำเนิดเส้นเสียง และสัญญาณรบกวนลมหายใจได้โดยตรง ซึ่งศึกษาจากการสร้างแบบจำลองของเสียงโดยเลียนแบบลักษณะของเสียงพูดมนุษย์ ทำให้ได้เสียงสังเคราะห์ที่มีความเป็นธรรมชาติ และสามารถสังเคราะห์เสียงได้ลักษณะต่าง ๆ ที่เกิดจากการเปลี่ยนแปลงรูปร่างของกล่องเสียง (โดยการปรับแบบจำลองแหล่งกำเนิดเส้นเสียง) หรือเพิ่มปรับระดับสัญญาณรบกวนในแหล่งกำเนิดเสียงเพื่อสังเคราะห์เสียงลมหายใจ เป็นต้น ตามวัตถุประสงค์ของวิทยานิพนธ์ที่ต้องการศึกษาและพัฒนาวิธีการปรับสังเคราะห์เสียงที่มีแหล่งกำเนิดจากเส้นเสียง และเพื่อให้เสียงสังเคราะห์มีความเป็นธรรมชาติมากขึ้น และเพื่อพัฒนาระบบสังเคราะห์เสียงที่สามารถสังเคราะห์เสียงที่มีลักษณะเสียงต่าง ๆ ได้

เพื่อประเมินความสามารถ และประสิทธิภาพของระบบสังเคราะห์เสียงที่นำเสนอวิทยานิพนธ์นี้ได้ประเมินเสียงสังเคราะห์ที่ได้จากวิธีที่นำเสนอเปรียบเทียบกับเสียงสังเคราะห์ที่ได้จากระบบแบบอ้างอิงที่ใช้สัญญาณกระตุ้นแบบพัลส์ ซึ่งผลการวัดความเป็นธรรมชาติของเสียงพบว่าคุณภาพของเสียงสังเคราะห์ที่ได้จากวิธีที่นำเสนอนี้มีความเป็นธรรมชาติใกล้เคียงกัน เมื่อเทียบกับระบบสังเคราะห์จากระบบอ้างอิง นอกจากนี้ระบบสังเคราะห์เสียงที่นำเสนอยังสามารถสังเคราะห์เสียงลักษณะต่าง ๆ ที่มีความแตกต่างกันหลากหลายมากกว่าเสียงสังเคราะห์ที่ได้จากระบบอ้างอิง ตามวัตถุประสงค์ที่ต้องการสร้างระบบสังเคราะห์เสียงที่สามารถสังเคราะห์เสียงที่มีลักษณะเสียงต่าง ๆ ได้ โดยการปรับค่าพารามิเตอร์ซึ่ง พารามิเตอร์ที่มีผลต่อการสังเคราะห์เสียงลมหายใจคือการเพิ่มปริมาณสัญญาณรบกวน แต่ระดับของปริมาณสัญญาณรบกวนที่ใช้สังเคราะห์เสียงลมหายใจในระบบที่นำเสนอนี้ใช้การลดระดับอัตราสัญญาณต่อสัญญาณรบกวน (SNR) ลดลงจากระดับปกติ สำหรับการสังเคราะห์เสียงบีบจากระบบที่นำเสนอ พบว่าพารามิเตอร์ที่มีผลต่อการสังเคราะห์เสียงคือ 1) ช่วงเปิดสัญญาณเสียง Rd เมื่อวิเคราะห์ได้จากสเปกโตรแกรมเป็นลักษณะของการปรากฏช่วงแถบเสียงที่มีระยะห่างระหว่างแถบกว้างมากขึ้นกว่ากว่าเสียงปกติ และ 2) ความถี่มูลฐาน ที่ส่งผลต่อความถี่ และการสั่นของฮาร์โมนิกส์ ซึ่งทั้งสองค่าพารามิเตอร์มีผลต่อการสมบัติการรับรู้เสียงบีบ

ระบบการสังเคราะห์เสียงโดยใช้แบบจำลองฮิดเดนมาร์คอฟที่นำเสนอในวิทยานิพนธ์นี้ ใช้แบบจำลองแหล่งกำเนิดเส้นเสียง และสัญญาณรบกวนลมหายใจเพื่อใช้ในการสร้างแบบจำลองแทนการใช้สัญญาณกระตุ้นแบบพัลส์และสัญญาณรบกวนสีขาว ที่ใช้ทั่วไปสำหรับแนวทางการสร้างระบบการสังเคราะห์เสียงที่ใช้แบบจำลองฮิดเดนมาร์คอฟ ดังนั้นวิทยานิพนธ์นี้จึงได้ศึกษาแบบจำลองแอลเอฟเพื่อใช้ประมาณสัญญาณเส้นเสียง และวิเคราะห์หาค่าพารามิเตอร์โดยใช้อัลกอริทึมการประมาณค่าพารามิเตอร์แอลเอฟ สำหรับการหาค่าระดับสัญญาณรบกวน ใช้การลดสัญญาณรบกวนในสัญญาณเส้นเสียงโดยเวฟเลท ซึ่งในวิทยานิพนธ์นี้เสนอวิธีการหาจุดเริ่มเปลี่ยนซึ่งมีความสำคัญต่อการลดสัญญาณรบกวนด้วยวิธีการประมาณค่าจุดเปลี่ยนของเวฟเลท และได้ทดสอบฟังก์ชันการหาจุดเปลี่ยนโดยวิเคราะห์ความสัมพันธ์ของระดับสัญญาณรบกวนเป้าหมายเทียบกับระดับสัญญาณรบกวนที่แยกได้จากฟังก์ชัน เมื่อมีการปรับสัญญาณรบกวนระดับต่าง ๆ พบว่าวิธีที่นำเสนอมีประสิทธิภาพในการประมาณระดับสัญญาณรบกวน ซึ่งได้อธิบายการวิเคราะห์ค่าพารามิเตอร์ในบทที่ 4

นอกจากการวัดคุณภาพเสียงสังเคราะห์แล้ว วิทยานิพนธ์ได้ทำการประเมินผลการทำงานของระบบที่นำเสนอโดยใช้วิธีการประเมินจากการรับรู้และตัดสินใจจากผู้ฟัง และใช้การวิเคราะห์สเปกตรัม วิเคราะห์สเปกโตรแกรม และการวิเคราะห์ค่าอะคูสติกทางเสียง เพื่อประเมินการวัดความถูกต้อง และการวิเคราะห์ความแตกต่างของการรับรู้ลักษณะเสียงของเสียงสังเคราะห์แต่ละแบบได้ถูกต้องมากขึ้น ซึ่งจากการศึกษาผลวิเคราะห์ที่ได้จากการรับรู้ของผู้ฟัง และผลจากการวิเคราะห์ค่าอะคูสติกที่บอกสมบัติลักษณะเสียงแบบต่าง ๆ พบว่า ผลส่วนใหญ่ที่ได้จากการรับรู้จากผู้ฟังมีความสอดคล้องกับผลที่ได้จากการวิเคราะห์ค่าทางอะคูสติก แต่อาจจะมีผลของการวิเคราะห์แตกต่างกันบ้าง ซึ่งอาจจะเกิดจากค่าที่ใช้วัดอะคูสติกที่ใช้ในวิทยานิพนธ์นี้ที่เลือกนำมาวิเคราะห์ เช่น การวัด HNR ที่ต้องการวัดค่าอะคูสติกที่สะท้อนปริมาณสัญญาณรบกวนในเสียงสระ อาจจะไม่เพียงพอในการบอกคุณภาพของการรับรู้ได้สัญญาณรบกวนของแต่ละเสียงได้ หรือการวัดค่า HRF ซึ่งต้องการศึกษาลักษณะเสียงแบบต่าง ๆ จากความสัมพันธ์ของฮาร์โมนิกส์และความถี่มูลฐานอาจจะเป็นค่าวัดที่ไม่คงทน เมื่อค่าความถี่มูลฐานมีค่าในช่วงที่ต่ำกว่าปกติมากเกินไป หรือได้รับผลกระทบจากสัญญาณรบกวน จึงทำให้ค่าที่คำนวณได้นั้น ไม่ตรงตามทฤษฎีการรับรู้ลักษณะเสียง

ประโยชน์ที่ได้รับจากวิทยานิพนธ์นี้

วิทยานิพนธ์นี้ นำเสนอการสกัดค่าลักษณะของแหล่งกำเนิดเสียงจากสัญญาณเสียงสำหรับใช้ในการสร้างระบบสังเคราะห์เสียงพูดโดยอาศัยแบบจำลองฮิดเดนมาร์คอฟที่ใช้แบบจำลองเส้นเสียง และระดับเสียงรบกวนทางลมหายใจเป็นสัญญาณกระตุ้น

อีกทั้งได้นำเสนอการสร้างเสียงสังเคราะห์ที่มีลักษณะเสียงบีบ เสียงปกติ และเสียงลมหายใจ ซึ่งเลียนแบบลักษณะเสียงพูดของมนุษย์จากการปรับค่าพารามิเตอร์แหล่งกำเนิดเส้นเสียงและระดับเสียงรบกวนทางลมหายใจ ซึ่งเป็นสัญญาณกระตุ้นของระบบสังเคราะห์เสียง และยังสามารถสังเคราะห์เสียงที่มีลักษณะเสียงบีบ และเสียงลมหายใจ ให้มีระดับความแตกต่างของระดับความเป็นเสียงบีบ และเสียงลมหายใจ จากการปรับค่าพารามิเตอร์ของสัญญาณกระตุ้นได้โดยง่าย

นอกจากนี้ วิทยานิพนธ์นี้ยังได้นำเสนอวิธีการหาค่าจุดเปลี่ยนจากค่าสัมประสิทธิ์เวฟเลทเพื่อใช้ในประมาณระดับสัญญาณรบกวนบนสัญญาณเส้นเสียงอีกด้วย

ข้อเสนอแนะ

จากการวิเคราะห์เสียงพูดตัวอย่างในลักษณะเสียงบีบ และเสียงลมหายใจ ยังมีปัจจัยที่มีผลต่อการรับรู้ในลักษณะเสียงต่าง ๆ คือ ความยาวของหน่วยเสียง และระดับความดัง โดยถ้ามีการปรับปัจจัยทั้งสองให้เหมาะสมในแต่ละลักษณะเสียงแล้วจะมีผลต่อรับรู้ของผู้ฟัง ซึ่งถ้าสามารถวิเคราะห์หาความสัมพันธ์ของปัจจัยดังกล่าวได้น่าจะช่วยให้เสียงสังเคราะห์มีคุณภาพดียิ่งขึ้น

จากผลการทดลอง จากการปรับค่า Rd ให้เพิ่มขึ้น ไม่ส่งผลต่อการรับรู้เสียงลมหายใจมากนักเนื่องจากว่าขั้นตอนการวิเคราะห์สัญญาณรบกวนมุ่งเน้นการหาสัญญาณรบกวนในแต่ละหน่วยเสียงจากสัญญาณเสียงพูด ยังไม่ได้ศึกษาความสัมพันธ์ระหว่างความกว้างของช่วงเปิดบนสัญญาณเส้นเสียงต่อสัญญาณรบกวน ซึ่งถ้าสามารถหาความสัมพันธ์ของความสัมพันธ์นี้ได้จะทำให้ระบบสังเคราะห์เสียงสมบูรณ์ยิ่งขึ้น และสามารถสังเคราะห์เสียงลมหายใจจากการปรับค่าพารามิเตอร์เส้นเสียงได้

รายการอ้างอิง

[1] G. Fant, Acoustic theory of speech production, Mouton De Gruyter, 1970.

[2] C. H. Coker, Speech synthesis with a parametric articulatory model Speech Synthesis, pp. 135-139, 1973.

[3] A. J. Hunt, and A. W. Black. Unit selection in a concatenative speech synthesis system using alarge speech database In Porceeding of ICASSP (1996):

[4] K. Tokuda, H. Zen, and A. W. Black. An HMM-based speech synthesis system applied to English IEEE Speech Synthesis Workshop (2002):

[5] K. N. Stevens, Acoustic Phonetics, MIT Press, 1999.

[6] C. Gobl, The voice source in speech communication, Ph. D. dissertation, KTH, 2003, 2003.

[7] A. V. Oppenheim, and R. W. Schafer, Discrete-time signal processing, Prentice-Hall, Inc. Upper Saddle River, NJ, USA, 1989.

[8] C. Gobl, and A. Ní Chasaide, Techniques for analysing the voice source Coarticulation: Theory, Data and Techniques, pp. 300-321, 1999.

[9] B. Hammarberg, Perceptual and acoustic analysis of dysphonia, Department of Logopedics and Phoniatrics, Huddinge University Hospital, Stockholm, 1986.

[10] J. Laver, The phonetic description of voice quality, Cambridge: Cambridge University Press, 1980.

[11] P. Kirk, P. Ladefoged, and J. Ladefoged, The linguistic use of different phonation types Vocal fold physiology: Contemporary research and clinical issues, pp. 351-360, 1983.

[12] D. G. Childers, Speech processing and synthesis toolboxes, John Wiley & Sons, Inc, 2000.

[13] T. Yoshimura, Simultaneous modeling of phonetic and prosodic parameters, and characteristic conversion for HMM-based Text-to-Speech systems, Ph. D. Thesis, Department of Electrical and Computer Engineering, Nagoya Institute of Technology, 2001.

[14] T. Fukada, K. Tokuda, T. Kobayashi, and S. Imai. An adaptive algorithm for mel-cepstral analysis of speech In Proceeding of ICASSP (1992): 137-140.

[15] K. Tokuda, T. Masuko, N. Miyazaki, and T. Kobayashi, Multi-space probability distribution HMM IEICE TRANSACTIONS on Information and Systems, vol. 85, no. 3, pp. 455-464, 2002.

[16] J. Dines, and S. Sridharan. Trainable speech synthesis with trended hidden Markov models In Proceeding of ICASSP'01 (2001):

[17] S. Chomphan, and T. Kobayashi. Implementation and evaluation of an HMM-based Thai speech synthesis system In Proceeding of Interspeech (2007): 2849-2852.

[18] J. J. Odell, The use of context in large vocabulary speech recognition, University of Cambridge, 1995.

[19] K. Shinoda, and T. Watanabe. Speaker adaptation with autonomous model complexity control by MDLprinciple In Proceeding of ICASSP-96 (1996):

[20] K. Tokuda, T. Kobayashi, and S. Imai. Speech parameter generation from HMM using dynamic features In Proceeding of ICASSP (1995):

[21] K. Tokuda, T. Yoshimura, T. Masuko, T. Kobayashi, and T. Kitamura. Speech parameter generation algorithms for HMM-based speech synthesis In Proceeding of ICASSP (2000):

[22] G. Fant, The LF-model revisited. Transformations and frequency domain analysis Speech Trans. Lab. Q. Rep., Royal Inst. of Tech. Stockholm, vol. 2, pp. 3, 1995.

[23] R. S. Maia, H. Zen, K. Tokuda, T. Kitamura, and F. G. V. Resende Jr. Towards the development of a Brazilian Portuguese text-to-speech system based on HMM Eighth European Conference on Speech Communication and Technology (2003):

[24] Y. Qian, F. Soong, Y. Chen, and M. Chu, An HMM-based Mandarin Chinese text-to-speech system Lecture Notes in Computer Science, vol. 4274, pp. 223, 2006.

[25] S. J. Kim, J. J. Kim, and M. Hahn, Implementation and evaluation of an HMM-based Korean speech synthesis system IEICE TRANSACTIONS on Information and Systems, no. 3, pp. 1116-1119, 2006.

[26] J. Yamagishi, T. Kobayashi, Y. Nakano, K. Ogata, and J. Isogai, Analysis of speaker adaptation algorithms for HMM-based speech synthesis and a constrained SMAPLR adaptation algorithm IEEE Transactions on Audio, Speech, and Language Processing, vol. 17, no. 1, pp. 66-83, 2009.

[27] J. Yamagishi, and T. Kobayashi, Average-voice-based speech synthesis using HSMM-based speaker adaptation and adaptive training IEICE TRANSACTIONS on Information and Systems, vol. 90, no. 2, pp. 533-543, 2007.

[28] M. Tamura, T. Masuko, K. Tokuda, and T. Kobayashi. Speaker adaptation for HMM-based speech synthesis system using MLLR In Proceeding of COCOSDA (1998):

[29] T. Nose, J. Yamagishi, T. Masuko, and T. Kobayashi, A style control technique for HMM-based expressive speech synthesis IEICE TRANSACTIONS on Information and Systems, vol. 90, no. 9, pp. 1406-1413, 2007.

[30] M. Tachibana, J. Yamagishi, T. Masuko, and T. Kobayashi, Speech synthesis with various emotional expressions and speaking styles by style interpolation and morphing IEICE TRANSACTIONS on Information and Systems, no. 11, pp. 2484-2491, 2005.

[31] T. Yoshimura, K. Tokuda, T. Masuko, T. Kobayashi, and T. Kitamura. Mixed excitation for HMM-based speech synthesis Seventh European Conference on Speech Communication and Technology (2001):

[32] S. J. Kim, and M. Hahn, Two-band excitation for HMM-based speech synthesis IEICE TRANSACTIONS on Information and Systems, vol. 90, no. 1, pp. 378-381, 2007.

[33] C. Hemptinne, Integration of the Harmonic plus Noise Model (HNM) into the Hidden Markov Model-Based Speech Synthesis System (HTS) Master thesis, IDIAP Research Institute, 2006.

[34] R. Maia, T. Toda, H. Zen, Y. Nankaku, and K. Tokuda. An excitation model for HMM-based speech synthesis based on residual modeling In Proceeding of ISCA SSW6 (2007): 131–136.

[35] D. G. Childers, and H. T. Hu, Speech synthesis by glottal excited linear prediction Journal of the Acoustical Society of America, vol. 96, pp. 2026-2036, 1994.

[36] J. P. Cabral, S. Renals, K. Richmond, and J. Yamagishi. Towards an improved modeling of the glottal source in statistical parametric speech synthesis In Proceeding of the 6th ISCA Workshop on Speech Synthesis (2007):

[37] J. P. Cabral, S. Renals, K. Richmond, and J. Yamagishi. Glottal Spectral Separation for Parametric Speech Synthesis In Proceeding of the Interspeech (2008):

[38] K. Ishizaka, and J. Flanagan, Synthesis of voiced sounds from a two-mass model of the vocal cords Bell System Technical Journal, vol. 51, pp. 1233-1268, 1972.

[39] J. L. Flanagan, K. Ishizaka, and K. L. Shipley, Synthesis of speech from a dynamic model of the vocal cords and vocal tract Bell System Technical Journal, vol. 54, no. 3, pp. 485-506, 1975.

[40] N. J. C. Lous, G. C. J. Hofmans, R. N. J. Veldhuis, and A. Hirschberg, A symmetrical two-mass vocal-fold model coupled to vocal tract and trachea, with application to prosthesis design Acustica, vol. 84, pp. 1135-1150, 1998.

[41] C. d'Alessandro, and D. Sciamarella. A study of the Two-Mass Model in terms of Acoustic parameters (2002): 2313-2316.

[42] T. V. Ananthapadmanabha, Acoustic analysis of voice source dynamics Speech Transmission Laboratory, Q. Prog. Status Rep, pp. 2-3, 1984.

[43] B. Doval, and C. d'Alessandro. Spectral correlates of glottal waveform models: an analytic study (1997):

[44] H. L. Lu, and J. O. Smith Iii. Estimating glottal aspiration noise via wavelet thresholding andbest-basis thresholding IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics (2001): 11-14.

[45] C. Moler, and J. Little, The MathWorks - MATLAB and Simulink for Technical Computing, www.mathworks.com.

[46] S. Young, J. Odell, D. Ollason, V. Valtchev, and P. Woodland, The HTK book (for HTK Version 3.4), 2006.

[47] H. Zen, T. Nose, J. Yamagishi, S. Sako, T. Masuko, A. W. Black, and K. Tokuda. The HMM-based speech synthesis system version 2.0 In Proceeding of ISCA SSW6 (2007):

[48] M. Airas, TKK Aparat: An environment for voice inverse filtering and parameterization Logopedics Phoniatrics Vocology, vol. 33, no. 1, pp. 49-64, 2008.

[49] C. Hansakunbuntheung, V. Tesprasit, and V. Sornlertlamvanich, Thai tagged speech corpus for speech synthesis The Oriental COCOSDA 2003, pp. 97-104, 2003.

[50] T. Charoenporn, V. Sornlertlamvanich, and H. Isahara. Building a large Thai text corpus-part of speech tagged corpus: ORCHID (1997): 509-512.

[51] X. Huang, A. Acero, and H. W. Hon, Spoken language processing: A guide to theory, algorithm, and system development, Prentice Hall PTR Upper Saddle River, NJ, USA, 2001.

[52] M. Hollander, and D. A. Wolfe, Nonparametric statistical methods, Wiley, 1973.

[53] E. B. Holmberg, R. E. Hillman, J. S. Perkell, P. C. Guiod, and S. L. Goldman, Comparisons among aerodynamic, electroglottographic, and acoustic spectral measures of female voice Journal of Speech and Hearing Research, vol. 38, no. 6, pp. 1212-1223, 1995.

[54] A. Ní Chasaide, and C. Gobl, Voice source variation The handbook of phonetic sciences, vol. 5, pp. 427–461, 1997.

[55] C. M. Esposito, The effects of linguistic experience on the perception of phonation, University of California, Los Angeles, 2006.

[56] P. Alku, Glottal wave analysis with pitch synchronous iterative adaptive inverse filtering Speech Communication, vol. 11, no. 2-3, pp. 109-118, 1992.

[57] A. El-Jaroudi, and J. Makhoul. Discrete all-pole modeling IEEE Transactions on signal processing (1991): 411-423.

[58] H. Strik, and L. Boves. Automatic estimation of voice source parameters In Proceeding of ICSLP (1994): 155-158.

ประวัติผู้เขียนวิทยานิพนธ์

นายนิพนธ์ ชินะธิมาตร์มงคล เกิดเมื่อวันที่ 18 สิงหาคม พ.ศ. 2524 สำเร็จการศึกษาระดับมัธยมศึกษาจากโรงเรียนเซนต์คาเบรียล และสำเร็จการศึกษาระดับปริญญาตรีจากภาควิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์ จากมหาวิทยาลัยเกษตรศาสตร์ เข้าศึกษาต่อระดับปริญญาโทที่จุฬาลงกรณ์มหาวิทยาลัย