การสร้างแบบจำลองการรับรู้ความเป็นธรรมชาติของเสียงวรรณยุกต์ด้วยแบบจำลองเรขาคณิต

นายกลวัชร ตระกูลสุข

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชาวิศวกรรมคอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2556
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

บทคัดย่อและแฟ้มข้อมูลฉบับเต็มของวิทยานิพนธ์ตั้งแต่ปีการศึกษา 2554 ที่ให้บริการในคลังปัญญาจุฬาฯ (CUIR)
เป็นแฟ้มข้อมูลของนิสิตเจ้าของวิทยานิพนธ์ ที่ส่งผ่านทางบัณฑิตวิทยาลัย
The abstract and full text of theses from the academic year 2011 in Chulalongkorn University Intellectual Repository (CUIR)
are the thesis authors' files submitted through the University Graduate School.

MODELLING OF TONE NATURALNESS PERCEPTION USING GEOMETRIC MODEL

Mr. Konlawachara Trakulsuk

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

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 2013

Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์	การสร้างแบบจำลองการรับรู้ความเป็นธรรมชาติของเสียงวรรณยุกต์ด้วยแบบจำลองเรขาคณิต
โดย	นายกลวัชร ตระกูลสุข
สาขาวิชา	วิศวกรรมคอมพิวเตอร์
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	รองศาสตราจารย์ ดร.อติวงศ์ สุชาโต
อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม	ผู้ช่วยศาสตราจารย์ ดร.โปรดปราน บุณยพุกกณะ ดร.ชัย วุฒิวิวัฒน์ชัย

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

.....คณบดีคณะวิศวกรรมศาสตร์

(ศาสตราจารย์ ดร.บัณฑิต เอื้ออาภรณ์)

คณะกรรมการสอบวิทยานิพนธ์

.....ประธานกรรมการ

(ผู้ช่วยศาสตราจารย์ ดร.เศรษฐา ปานงาม)

.....อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

(รองศาสตราจารย์ ดร.อติวงศ์ สุชาโต)

.....อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม

(ผู้ช่วยศาสตราจารย์ ดร.โปรดปราน บุณยพุกกณะ)

.....อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม

(ดร.ชัย วุฒิวิวัฒน์ชัย)

.....กรรมการภายนอกมหาวิทยาลัย

(ดร.ชัชวาลย์ หาญสกุลบรรเทิง)

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

กลวัชร ตระกูลสุข : การสร้างแบบจำลองการรับรู้ความเป็นธรรมชาติของเสียงวรรณยุกต์ด้วยแบบจำลองเรขาคณิต. (MODELLING OF TONE NATURALNESS PERCEPTION USING GEOMETRIC MODEL) อ.ที่ปรึกษาวิทยานิพนธ์หลัก: รศ. ดร. อติวงศ์ สุชาโต, อ.ที่ปรึกษาวิทยานิพนธ์ร่วม: ผศ. ดร.โปรดปราน บุณยพุกกณะ, ดร.ชัย วุฒิวิวัฒน์ชัย, 83 หน้า.

ความเป็นธรรมชาติของเสียงสังเคราะห์ (Text-to-Speech System) นั้นเป็นประเด็นที่สำคัญในระบบสังเคราะห์เสียง ในการรองรับโครงร่างความถี่มูลฐาน (Fundamental Frequency Contour) ของเสียงสังเคราะห์ของพยางค์ใดๆ อย่างไม่มีกฎเกณฑ์ที่ถูกกำหนดขึ้นนั้น ระบบสังเคราะห์เสียงยังต้องสามารถคงไว้ซึ่งความเป็นธรรมชาติของเสียงสังเคราะห์ได้ วิทยานิพนธ์นี้จึงเสนอการประเมินความเป็นธรรมชาติแบบอัตโนมัติของโครงร่างความถี่มูลฐาน เพื่อพิจารณาระดับของความเป็นธรรมชาติของเสียงสังเคราะห์ของพยางค์เมื่อถูกรับรู้จากผู้ฟัง ด้วยแบบจำลองการทำนายความเป็นธรรมชาติ โดยอาศัยจุดกึ่งกลาง (Midpoint) และจุดปลาย (Endpoint) ของเส้นความถี่มูลฐานในช่วงของหน่วยตามพยางค์ (Rhyme) ที่ได้จากการวิเคราะห์ผลการทดลองการรับรู้เสียงวรรณยุกต์ในผู้ฟัง ซึ่งแบบจำลองนี้ได้ถูกใช้เพื่อการพัฒนาอัลกอริทึมในการทำนายความเป็นธรรมชาติของเสียงวรรณยุกต์จากเส้นความถี่มูลฐานโดยใช้แบบจำลองเรขาคณิต (Geometric Model) ในการประเมินผลของอัลกอริทึมในการทำนายความเป็นธรรมชาติของเสียงวรรณยุกต์ ได้ใช้ผู้ฟังรับรู้และประเมินความเป็นธรรมชาติของพยางค์จากเส้นความถี่มูลฐาน 45 รูปแบบ ซึ่งแต่ละรูปแบบ จะถูกผู้ฟังประเมินซ้ำกัน 2 ครั้ง ซึ่งอัลกอริทึมที่ได้นำเสนอนั้นมีอัตราความสอดคล้องประมาณ 80% เมื่อเทียบกับการตัดสินถึงความเป็นธรรมชาติของเสียงวรรณยุกต์ในพยางค์ของผู้ฟัง

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ภาควิชา วิศวกรรมคอมพิวเตอร์	ลายมือชื่อนิสิต .....
สาขาวิชา วิศวกรรมคอมพิวเตอร์	ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์หลัก .....
ปีการศึกษา 2556	ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์ร่วม .....
	ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์ร่วม .....

5570475821 : MAJOR COMPUTER ENGINEERING
KEYWORDS: SPEECH NATURALNESS / TONE PERCEPTION / SPEECH NATURALNESS MODEL / THAI TONES / PITCH CONTOURS / FUNDAMENTAL FREQUENCY

KONLAWACHARA TRAKULSUK: MODELLING OF TONE NATURALNESS PERCEPTION USING GEOMETRIC MODEL. ADVISOR: ASSOC. PROF. ATIWONG SUCHATO, Ph.D., CO-ADVISOR: ASST. PROF. PROADPRAN PUNYABUKKANA, Ph.D., CHAI WUTIWIWATCHAI, Ph.D., 83 pp.

Naturalness is an important issue in the Text-To-Speech (TTS) system. To support arbitrarily defined pitch contours for any synthesized syllables, a TTS should be able to maintain the naturalness of the synthetic speech. This work proposed an automatic evaluation of pitch contours in order to determine the level of naturalness of synthesized syllables when perceived by human listeners. By analyzing results, tone perception experiments conducted on human listeners in this work, a syllable tone naturalness prediction model based on the midpoint and endpoint of the syllable's rhyme part was proposed. The model was then used for developing a tone naturalness prediction algorithm using geometric models of pitch contours. The evaluation of the tone naturalness prediction algorithm involved human listeners perceiving the naturalness of syllables with 45 pitch contour patterns, each of which with 2 repetitions. The proposed algorithm achieved approximately 80% consistency rate compared against human listeners' decisions on tone naturalness of the syllables.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

Department:	Computer Engineering	Student's Signature .....
Field of Study:	Computer Engineering	Advisor's Signature .....
Academic Year:	2013	Co-Advisor's Signature .....
		Co-Advisor's Signature .....

กิตติกรรมประกาศ

ข้าพเจ้าขอขอบพระคุณ อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก รองศาสตราจารย์ ดร.อติวงศ์ สุชาโต อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม ผู้ช่วยศาสตราจารย์ ดร.โปรดปราน บุณยพุกกณะ และ ดร. ชัย วุฒิวิวัฒน์ชัย ที่ได้สละเวลาอันมีค่าเพื่อให้คำปรึกษาและให้คำแนะนำอันมีค่าตลอดระยะเวลาการศึกษาของข้าพเจ้า ข้าพเจ้าขอขอบคุณ ดร.ชัชวาลย์ หาญสกุลบรรเทิง ที่ให้คำแนะนำและชี้แนะแนวคิดในการทำวิทยานิพนธ์นี้ รวมทั้ง คุณปวันรัตน์ หิรัญกาญจน์ คุณโสรวาร ศิริพงศ์ปรีดา และคุณจิราภรณ์ ประกอบแสงสวย ที่สละเวลาอันมีค่าในการบันทึกเสียงพูดที่ใช้สำหรับการทดลองในวิทยานิพนธ์นี้ ข้าพเจ้าขอขอบคุณ หน่วยปฏิบัติการวิจัยวิทยาการมนุษยภาษา ศูนย์เทคโนโลยีและคอมพิวเตอร์แห่งชาติ: NECTEC ที่ให้เครื่องมือที่จำเป็นในการทดลองของการทำวิทยานิพนธ์ และภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ที่สนับสนุนทุนการศึกษาตลอดระยะเวลาการศึกษาของข้าพเจ้า

ข้าพเจ้าขอขอบคุณ คุณแม่ ครอบครัว และเพื่อนร่วมงานทุกคนของห้องปฏิบัติการระบบภาษาพูด ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ที่ให้คำแนะนำต่างๆ อันเป็นประโยชน์และให้กำลังใจแก่ข้าพเจ้าเสมอ ซึ่งส่งผลให้การทำวิทยานิพนธ์นี้ของข้าพเจ้าประสบความสำเร็จได้

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

สารบัญ

หน้า

บทคัดย่อภาษาไทย ..... ง

บทคัดย่อภาษาอังกฤษ ..... จ

กิตติกรรมประกาศ ..... ฉ

สารบัญ ..... ช

สารบัญตาราง ..... ฌ

สารบัญภาพ ..... ญ

บทที่ 1 บทนำ 1

1.1 ความเป็นมาและความสำคัญของปัญหา 1

1.2 วัตถุประสงค์ของการวิจัย 2

1.3 ขอบเขตของการวิจัย 2

1.4 ลำดับขั้นตอนในการเสนอผลการวิจัย 2

1.5 ประโยชน์ที่คาดว่าจะได้รับ 4

1.6 ผลงานตีพิมพ์จากวิทยานิพนธ์ 4

1.7 ลำดับการจัดเรียงเนื้อหาในวิทยานิพนธ์ 4

บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง 6

2.1 ทฤษฎีที่เกี่ยวข้อง 6

2.2 เอกสารและงานวิจัยที่เกี่ยวข้อง 17

บทที่ 3 ขั้นตอนการดำเนินงานวิจัย 22

3.1 การเลือกลักษณะสำคัญของเสียงวรรณยุกต์ภาษาไทย 22

3.2 ภาพรวมของกระบวนการในการทดลอง 24

3.3 การเตรียมเสียงตัวกระตุ้น 26

3.4 การทดลองการรับรู้ 30

3.5 การทำนายการรับรู้ความเป็นธรรมชาติโดยอาศัยแบบจำลองเรขาคณิต 32

3.6 เครื่องมือที่ใช้ในงานวิจัย 37

บทที่ 4 การทดลอง ผลการทดลองและการวิเคราะห์ผลการทดลอง 38

4.1 ข้อมูลเสียงพูดที่ใช้ในการทดลอง 38

4.1.1 เสียงตัวกระตุ้นในชุดฝึกฝน 40

หน้า

4.1.2 เสียงตัวกระตุ้นในชุดทดสอบ 41
4.2 การทดลองการรับรู้ 43
4.2.1 การระบุเสียงวรรณยุกต์และผลการทดลอง 43
4.2.2 การให้คะแนนความเป็นธรรมชาติและผลการทดลอง 47
4.2.3 การวิเคราะห์ผลการทดลองการรับรู้ 51
4.3 การสร้างแบบจำลองการทำนายการรับรู้ความเป็นธรรมชาติของเสียงวรรณยุกต์ 52
4.3.1. แบบจำลองการรับรู้ความเป็นธรรมชาติของเสียงวรรณยุกต์ 55
4.4 การวัดผลการทำนายการรับรู้ความเป็นธรรมชาติของแบบจำลอง 62
4.5 วิเคราะห์ผลการทดลอง 63
บทที่ 5 บทสรุปผลการวิจัยและข้อเสนอแนะ 72
5.1 สรุปผลการวิจัย 72
5.2 ข้อจำกัด 72
5.3 ข้อเสนอแนะ 73
รายการอ้างอิง 75
ประวัติผู้เขียนวิทยานิพนธ์ 83

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

สารบัญตาราง

ตารางที่		หน้า
2.1	เปรียบเทียบพยัญชนะกับสัญลักษณ์หน่วยเสียง	6
2.2	มาตราตัวสะกดในภาษาไทย	7
2.3	เกณฑ์การผันวรรณยุกต์ในภาษาไทย.....	9
3.1	บรรทัดฐานในการให้คะแนนความเป็นธรรมชาติของเสียงวรรณยุกต์	32
4.1	แสดงขอบเขตความถี่มูลฐานของผู้พูด.....	40
4.2	จำนวนเสียงตัวกระตุ้นที่ใช้ในการทดลองการให้คะแนนความเป็นธรรมชาติของเสียงวรรณยุกต์.....	47
4.3	ผลความแม่นยำของแบบจำลองเมื่อเทียบกับการรับรู้ของผู้ฟัง	63

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

สารบัญภาพ

ภาพที่		หน้า
2.1	ลักษณะของเสียงวรรณยุกต์ในภาษาไทย	8
2.2	ภาพรวมของระบบการแปลงข้อความเป็นเสียงพูด	10
2.3	กราฟฟังก์ชันพหุนามอันดับที่ n, $P_n(x)$	13
2.4	ตัวอย่างกราฟของฟังก์ชันพหุนามอันดับ 7	14
2.5	กราฟแสดงการปรับเส้นด้วยเส้นโค้งแบบสไปลน์อันดับหนึ่ง	15
2.6	กราฟแสดงการปรับเส้นด้วยเส้นโค้งแบบสไปลน์อันดับสอง	15
2.7	กราฟแสดงการปรับเส้นด้วยเส้นโค้งแบบสไปลน์อันดับสาม	16
2.8	กราฟแสดงเส้นโค้งแบบสไปลน์อันดับสามในรูปทั่วไป	16
2.9	แสดงผลกระทบที่เกิดจากแคร่โอเวอร์เอฟเฟกต์ในเสียงวรรณยุกต์ภาษาไทย	18
2.10	แสดงสมมติฐานในการปรับแนวของหน่วยเวลาย่อย (Moraic Alignment Hypothesis)	18
3.1	แสดงตำแหน่งของจุดเริ่ม จุดกึ่งกลาง และจุดปลาย ในพยางค์คำว่า “หน้า”	22
3.2	แสดงโครงร่างความถี่มูลฐานของพยางค์ [น-า] ที่กลุ่มผู้ฟังรับรู้เป็นเสียงวรรณยุกต์จัตวาและตรีในพยางค์เดียวกัน	23
3.3	**ภาพรวมของขั้นตอนการทดลอง**	**25**
3.4	แสดงขั้นตอนย่อยในขั้นตอนการเตรียมเสียงตัวกระตุ้น	26
3.5	แสดงตัวอย่างการปรับแต่งโครงร่างความถี่มูลฐาน	29
3.6	แสดงขั้นตอนย่อยในขั้นตอนการทดลองการรับรู้	30
3.7	แสดงขั้นตอนย่อยในขั้นตอนการทดลองการทำนายการรับรู้ความเป็นธรรมชาติโดยใช้แบบจำลองเรขาคณิต	33
3.8	ตัวอย่างข้อมูลในแผนภาพกระจาย	35
3.9	ตัวอย่างการประมาณค่าในช่วงของข้อมูลตัวอย่าง	36
3.10	ตัวอย่างการอธิบายความสัมพันธ์ด้วยแบบจำลองเรขาคณิต	36
4.1	โครงร่างความถี่มูลฐานของเสียงพูดก่อนและหลังการปรับเรียบ	39
4.2	รูปแบบโครงร่างความถี่มูลฐานของเสียงตัวกระตุ้นในชุดทดสอบ	42
4.3	หน้าจอโปรแกรมที่ใช้ในการทดลองการระบุเสียงวรรณยุกต์	44
4.4	ผลการทดลองการระบุเสียงวรรณยุกต์	45

4.5 หน้าจอโปรแกรมที่ใช้ในการทดลองการให้คะแนนความเป็นธรรมชาติ 48

4.6 ผลการทดลองการให้คะแนนความเป็นธรรมชาติ 49

4.7 แผนภาพกระจายระหว่างค่าจุดกึ่งกลางและความชันของพยางค์เป้าหมายที่มีความเป็นธรรมชาติและไม่เป็นธรรมชาติความเป็นธรรมชาติของแต่ละเสียงวรรณยุกต์จากการทดลองการรับรู้..... 53

4.8 แบบจำลองความเป็นธรรมชาติของเสียงวรรณยุกต์..... 61

4.9 ตำแหน่งของค่านอมาไลซ์ของความถี่มูลฐานที่ตำแหน่งจุดกึ่งกลางและค่าความชันของเสียงตัวกระตุ้นชุดทดสอบบนแบบจำลองการรับรู้ความเป็นธรรมชาติของเสียงวรรณยุกต์ 66

4.10 จำนวนคำตอบที่ผู้ฟังตอบตรงกับแบบจำลองในเสียงตัวกระตุ้นที่มีจุดกึ่งกลางจุดสิ้นสุดเหมือนกัน แต่มี F0 onset ต่างกันของผู้พูดเพศหญิง..... 67

4.11 จำนวนคำตอบที่ผู้ฟังตอบตรงกับแบบจำลองในเสียงตัวกระตุ้นที่มีจุดกึ่งกลางจุดสิ้นสุดเหมือนกัน แต่มี F0 onset ต่างกันของผู้พูดเพศชาย..... 68

4.12 จำนวนผู้ฟังที่ค่าออนเซ็ตมีผลต่อการรับรู้ในกลุ่มเสียงชุดทดสอบที่มีค่านอมาไลซ์ของความถี่มูลฐานที่ตำแหน่งกึ่งกลางและตำแหน่งสิ้นสุดเหมือนกันแต่มีค่านอมาไลซ์ของความถี่มูลฐานที่ตำแหน่งจุดเริ่มต้นต่างกันเล็กน้อย 70

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทที่ 1
บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ในปัจจุบัน ได้มีการนำคอมพิวเตอร์มาใช้ในด้านต่างๆ อย่างแพร่หลาย เพื่อใช้ในการแก้ปัญหาที่ซับซ้อน อาจกล่าวได้ว่า ใช้เพื่อทุ่นแรงในการคำนวณปัญหาที่ซับซ้อนและรวดเร็ว จนกระทั่งคอมพิวเตอร์นั้นเป็นที่นิยมไปถึงในระดับผู้ใช้งานทั่วไป และความต้องการที่ตามมาคือ การโต้ตอบระหว่างมนุษย์กับคอมพิวเตอร์ จึงได้มีการนำคอมพิวเตอร์มาพัฒนาเทคโนโลยีด้านภาษาศาสตร์คอมพิวเตอร์ (Computational Linguistic) ในหลายทศวรรตที่ผ่านมา หนึ่งในนั้น สาขาที่ถูกพัฒนามาอย่างต่อเนื่องก็คือเทคโนโลยีด้านการสังเคราะห์เสียงพูด (Speech Synthesis) เทคโนโลยีด้านการสังเคราะห์เสียงพูดมีความแพร่หลายไปในด้านต่างๆ อย่างกว้างขวาง ตัวอย่างเช่น ในด้านสิ่งอำนวยความสะดวกสำหรับผู้พิการ เช่น โปรแกรมอ่านหน้าจอคอมพิวเตอร์ (Non-Visual Desktop Access: NVDA) [1] และโปรแกรมแปลงข้อความเป็นเสียงพูด (Text-to-Speech) ดังนั้น จึงมีการให้ความสำคัญในด้านการสังเคราะห์เสียงพูดเป็นอย่างมาก

ในการพัฒนาการสังเคราะห์เสียงพูดนั้น ปัจจัยสำคัญที่ต้องคำนึงถึง ในการการพัฒนาระบบการสังเคราะห์เสียงให้มีคุณภาพ มีได้หลายปัจจัย เช่น คุณภาพของเสียงสังเคราะห์ และความเป็นธรรมชาติของเสียงสังเคราะห์ ซึ่งประกอบไปด้วย ความเป็นธรรมชาติ (Naturalness) ซึ่งเป็นการทำเสียงสังเคราะห์ให้มีเสียงพูดที่ใกล้เคียงเสียงพูดของมนุษย์ และ ความชัดเจนเข้าใจได้ (Intelligibility) [2] ซึ่งเมื่อพิจารณาในด้านของความเป็นธรรมชาตินั้น ปัจจัยทางสัทศาสตร์ (Prosodic Factor) มีบทบาทที่สำคัญต่อความเป็นธรรมชาติของเสียงพูดเป็นอย่างมาก สิ่งหนึ่งที่เป็นปัจจัยทางสัทศาสตร์ที่สำคัญอย่างหนึ่งในหลาย ๆ ปัจจัย ที่เกี่ยวข้องกับความเป็นธรรมชาติของเสียงพูดเป็นอย่างมาก คือค่าของความถี่มูลฐานที่เปลี่ยนไปตามแกนเวลา หรือเรียกอีกอย่างหนึ่งว่าโครงร่างความถี่มูลฐาน (Fundamental Frequency Contour; F_0 Contour) [3] สำหรับภาษาที่มีเสียงวรรณยุกต์โดยเฉพาะอย่างยิ่ง ภาษาไทยนั้น เสียงวรรณยุกต์ของคำหรือพยางค์นั้น มีผลต่อความหมายของคำนั้น ๆ เช่น ปา แปลว่า ขว้าง ป๋า แปลว่า บิดา เป็นต้น ซึ่งดังตัวอย่างข้างต้นชี้ให้เห็นว่า เสียงวรรณยุกต์มีผลต่อความหมายของคำในภาษาที่มีเสียงวรรณยุกต์ ลักษณะของโครงร่างความถี่มูลฐานของเสียงวรรณยุกต์ภาษาไทย เช่น ความสูง และรูปร่าง ของโครงร่างความถี่มูลฐานนั้น พบว่าเป็นปัจจัยที่สำคัญในการแยกแยะความแตกต่างของเสียงวรรณยุกต์ในภาษาไทยทั้งห้าเสียงออกจากกันได้ [4, 5] ด้วยเหตุนี้ การบิดเบือนใด ๆ ของโครงร่างความถี่มูลฐาน จะทำให้ความเป็นธรรมชาติของเสียงพูดนั้นลดลง ทำให้เสียงสังเคราะห์ที่สร้างขึ้น ฟังดูไม่เป็นธรรมชาติ เพราะฉะนั้น ความเป็นธรรมชาติของโครงร่างความถี่มูลฐานของแต่ละเสียงวรรณยุกต์นั้น ๆ จึงมีความสำคัญต่อระบบการสังเคราะห์เสียง เพื่อสร้างเสียงสังเคราะห์ที่เป็นธรรมชาติในแง่ของความพึงพอใจของผู้ฟัง

มีหลายงานวิจัยที่ทำการพัฒนาความเป็นธรรมชาติของเสียงพูดสังเคราะห์ให้ดีขึ้น ด้วยการควบคุมการสร้างโครงร่างความถี่มูลฐานในระบบสังเคราะห์เสียง [6-9] ซึ่งส่วนงานวิจัยใหญ่มุ่งเน้นการพัฒนาแบบจำลองของเสียงวรรณยุกต์ด้วยการเรียนรู้จากรูปร่าง และตัวแปรทางสัทศาสตร์ของโครง

ร่างความถี่มูลฐาน จากโครงร่างความถี่มูลฐานของเสียงพูดธรรมชาติ ซึ่งอาจมีความไม่คลอบคลุมรูปแบบของโครงร่างความถี่มูลฐานแบบอื่น ๆ ที่ไม่พบในเสียงพูดธรรมชาติ จากเสียงตัวอย่างที่นำมาศึกษาซึ่งมีจำนวนที่จำกัดในคลังข้อมูลเสียง ที่ยังคงมีความเป็นธรรมชาติต่อผู้ฟัง และหากโครงร่างความถี่มูลฐานที่ถูกบิดเบือนไป อาจมีความเป็นธรรมชาติต่อผู้ฟังได้เช่นกัน

ดังนั้นผู้วิจัยจึงมีแนวคิดในการศึกษารูปแบบของโครงร่างความถี่มูลฐาน ที่มีความหลากหลายมากขึ้น นอกเหนือจากโครงร่างความถี่มูลฐานของเสียงพูดธรรมชาติที่มีอยู่ ซึ่งมีผลเกี่ยวข้องต่อความเป็นธรรมชาติ นอกเหนือจากการควบคุมโครงร่างความถี่มูลฐานด้วยแบบจำลองเสียงวรรณยุกต์แล้ว ระบบเสียงสังเคราะห์ควรมีข้อจำกัดต่าง ๆ เพื่อรักษาความเป็นธรรมชาติของเสียงสังเคราะห์ เพื่อให้แน่ใจว่าเสียงสังเคราะห์ที่สร้างขึ้นนั้น รับฟังได้อย่างเป็นธรรมชาติ งานวิจัยนี้จึงได้ทำการศึกษาการรับรู้ความเป็นธรรมชาติของเสียงวรรณยุกต์ ของโครงร่างความถี่มูลฐานแบบต่าง ๆ และสร้างแบบจำลองเพื่อทำนายการรับรู้ความเป็นธรรมชาติของเสียงวรรณยุกต์ จากโครงร่างความถี่มูลฐาน ทั้งนี้เพื่อเป็นการวัดผลความเป็นธรรมชาติของเสียงวรรณยุกต์ ของโครงร่างความถี่มูลฐานแบบอัตโนมัติ

1.2 วัตถุประสงค์ของการวิจัย

1. เพื่อเสนอการศึกษาการรับรู้ความเป็นธรรมชาติของเสียงวรรณยุกต์ของพยางค์ในภาษาไทยของผู้ฟัง
2. เพื่อเสนอแบบจำลองของการรับรู้ความเป็นธรรมชาติของเสียงวรรณยุกต์ของพยางค์ในภาษาไทยของผู้ฟังจากการวิเคราะห์ผลการศึกษาการรับรู้ความเป็นธรรมชาติของเสียงวรรณยุกต์ของพยางค์ในภาษาไทยของผู้ฟัง
3. เพื่อทำนายการรับรู้ความเป็นธรรมชาติของเสียงวรรณยุกต์ของพยางค์ในภาษาไทยของผู้ฟัง

1.3 ขอบเขตของการวิจัย

1. พยางค์ที่ใช้ในการศึกษานั้น เป็นพยางค์เดี่ยวที่อยู่ในบริบทคำพูดต่อเนื่อง
2. แบบจำลองในการทำนายความเป็นธรรมชาติของเสียงวรรณยุกต์นั้น จะทำนายผลที่สอดคล้องกับการรับรู้ของคน
3. แบบจำลองในการทำนายความเป็นธรรมชาติของเสียงวรรณยุกต์นั้น จะทำนายพยางค์เดี่ยวที่อยู่ในบริบทคำพูดต่อเนื่อง

1.4 ลำดับขั้นตอนในการเสนอผลการวิจัย

1. ขั้นตอนการศึกษาเบื้องต้น
    1.1. ศึกษาทฤษฎีของเสียงวรรณยุกต์ในภาษาไทย
    1.2. ศึกษาทฤษฎีของรูปแบบโครงร่างความถี่มูลฐานของเสียงวรรณยุกต์ในภาษาไทย

1.3. ศึกษางานวิจัยเกี่ยวกับค่าลักษณะสำคัญที่ส่งผลต่อการผันวรรณยุกต์ในภาษาไทย

1.5. ศึกษางานวิจัยเกี่ยวกับค่าลักษณะสำคัญของโครงร่างความถี่มูลฐานที่ส่งผลต่อการระบุเสียงวรรณยุกต์ภาษาไทย

1.4. ศึกษางานวิจัยเกี่ยวกับการพัฒนาความเป็นธรรมชาติของเสียงสังเคราะห์ในระบบการสังเคราะห์เสียงภาษาไทย

1.6. ศึกษางานวิจัยเกี่ยวกับค่าลักษณะสำคัญของโครงร่างความถี่มูลฐานที่ส่งผลต่อการระบุเสียงวรรณยุกต์ภาษาไทย

1.7. ศึกษาเครื่องมือที่ใช้ในงานวิจัย เช่นเครื่องมือการวิเคราะห์สัญญาณเสียง (Speech Tool)

2. ขั้นตอนการออกแบบระบบ

2.1. ออกแบบกลุ่มคำที่ใช้ในการทดลองการรับรู้ (Perception Experiment)

2.2. ออกแบบขั้นตอนการทดลองการรับรู้

2.3. เลือกใช้เครื่องมือในการวิจัยที่เหมาะสม

2.4. ออกแบบการทดลอง

3. ขั้นตอนการทดลอง

3.1. เก็บข้อมูลเสียงจากการบันทึกเสียง

3.2. วิเคราะห์ข้อมูลเพื่อหาลักษณะเด่นซึ่งส่งผลของความเป็นธรรมชาติของเสียงวรรณยุกต์ในภาษาไทย

3.3. ทำการทดลองการรับรู้ความเป็นธรรมชาติของเสียงวรรณยุกต์ในภาษาไทย

3.4. วิเคราะห์ผลการทดลองการรับรู้ความเป็นธรรมชาติของเสียงวรรณยุกต์ในภาษาไทย

3.5. สร้างแบบจำลองการรับรู้ความเป็นธรรมชาติของเสียงวรรณยุกต์ในภาษาไทย

3.6. ทดสอบและวัดผลของแบบจำลองการรับรู้ความเป็นธรรมชาติของเสียงวรรณยุกต์ในภาษาไทย

3.7. วิเคราะห์ผลการทดลอง

4. สรุปผลและเรียบเรียงวิทยานิพนธ์

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

1.5 ประโยชน์ที่คาดว่าจะได้รับ

1. ทราบถึงข้อจำกัดของลักษณะโครงร่างความถี่มูลฐานที่ส่งผลถึงการรับรู้ของคน ที่มีต่อความเป็นธรรมชาติของเสียงวรรณยุกต์ในภาษาไทยของพยางค์เดี่ยวในการพูดต่อเนื่อง
2. สามารถทำนายเพื่อบอกความเป็นธรรมชาติของเสียงวรรณยุกต์ของพยางค์ในภาษาไทยซึ่งสอดคล้องกับการรับรู้ของคนได้
3. สามารถนำความรู้และข้อจำกัดที่ได้จากการศึกษา รวมทั้งแบบจำลองความเป็นธรรมชาติของเสียงวรรณยุกต์ภาษาไทย ไปประยุกต์ใช้กับระบบสังเคราะห์เสียงภาษาไทย เพื่อให้เสียงวรรณยุกต์ในเสียงสังเคราะห์เป็นธรรมชาติมากขึ้น

1.6 ผลงานตีพิมพ์จากวิทยานิพนธ์

ส่วนหนึ่งของวิทยานิพนธ์นี้ได้ตีพิมพ์เป็นบทความทางวิชาการในหัวข้อเรื่อง “Prediction of Tone Naturalness Using Geometric Model” จัดทำโดย “Konlawachara Trakulsuk, Atiwong Suchato, Proadpran Punyabukkana, Chai Wutiwiwatchai” ถูกนำเสนอในงานประชุมวิชาการ “The 11th International Joint Conference on Computer Science and Software Engineering: JCSSE’2014” จังหวัดชลบุรี ประเทศไทย ระหว่างวันที่ 14-16 พฤษภาคม 2557

1.7 ลำดับการจัดเรียงเนื้อหาในวิทยานิพนธ์

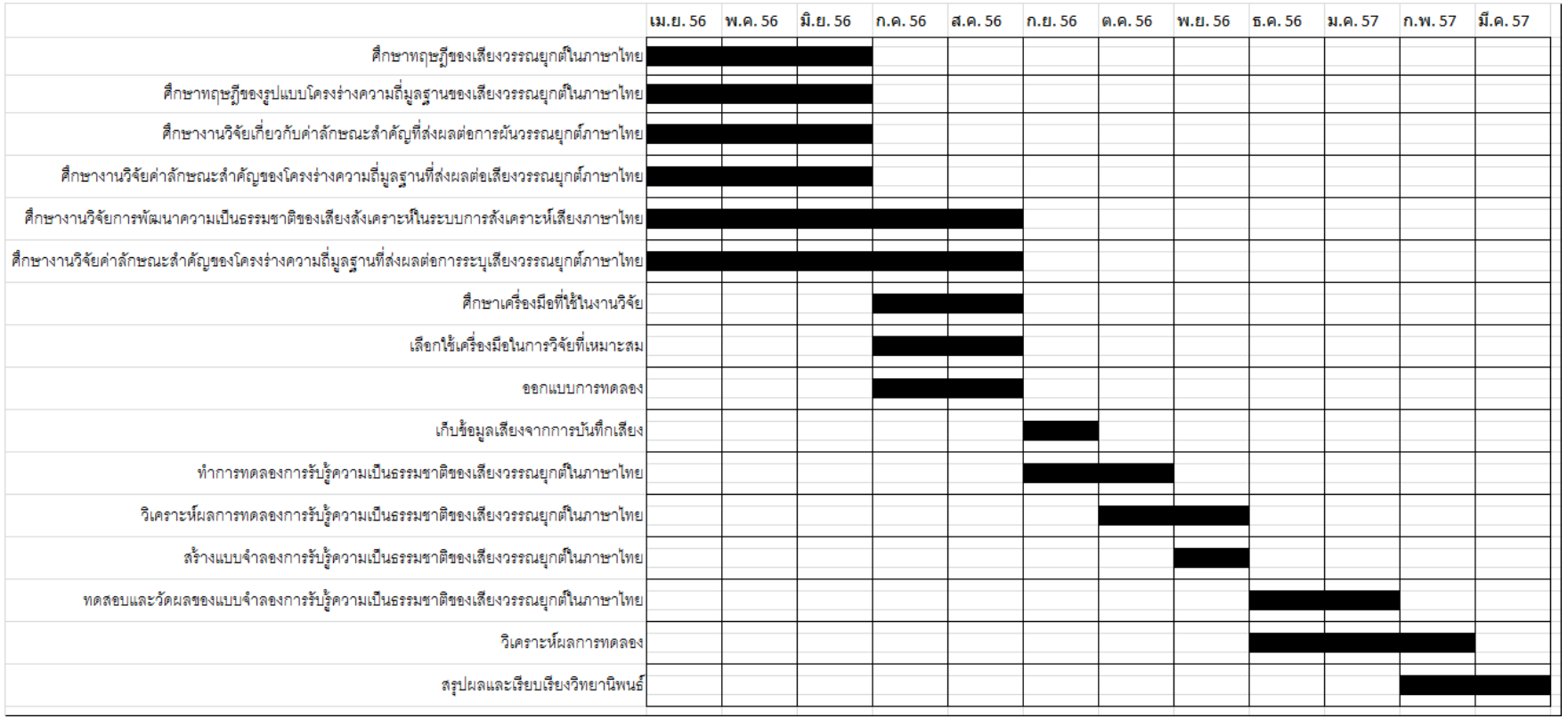
ในวิทยานิพนธ์นี้ได้แบ่งเนื้อหาออกเป็น 5 บท คือ บทที่ 1 บทนำ กล่าวถึง ความเป็นมาและความสำคัญของปัญหา, วัตถุประสงค์ของการวิจัย, ขอบเขตของการวิจัย, ประโยชน์ที่คาดว่าจะได้รับ, ลำดับขั้นตอนในการเสนอผลการวิจัย และผลงานตีพิมพ์จากวิทยานิพนธ์

บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง โดยทฤษฎีที่เกี่ยวข้องนั้น กล่าวถึง ทฤษฎีทางภาษาศาสตร์ ซึ่งประกอบไปด้วย ระบบอักษรไทยและการผันวรรณยุกต์ ทฤษฎีระบบสังเคราะห์เสียงพูด และทฤษฎีทางคณิตศาสตร์ที่เกี่ยวข้องกับงานวิจัย และงานวิจัยที่เกี่ยวข้อง

บทที่ 3 ขั้นตอนการดำเนินงานวิจัย ซึ่งประกอบด้วย การเลือกลักษณะสำคัญของโครงร่างความถี่มูลฐานของเสียงวรรณยุกต์ภาษาไทย ภาพรวมของกระบวนการในการทดลอง การเตรียมเสียงตัวกระตุ้น การทดลองการรับรู้ การทำนายการรับรู้ความเป็นธรรมชาติของเสียงวรรณยุกต์โดยใช้แบบจำลองเรขาคณิต และเครื่องมือที่ใช้ในการวิจัย

บทที่ 4 ผลการทดลองและอภิปรายผล ประกอบไปด้วย ข้อมูลเสียงพูดที่ใช้ในการทดลอง การทดลองการรับรู้ การสร้างแบบจำลองการทำนายการรับรู้ความเป็นธรรมชาติของเสียงวรรณยุกต์ การวัดผล และการวิเคราะห์ผลการทดลอง

บทที่ 5 บทสรุปผลการวิจัยและข้อเสนอแนะ กล่าวถึงการสรุปผลการวิจัย และข้อเสนอแนะต่าง ๆ รวมไปถึงงานวิจัยในอนาคต


เม.ย. 56
พ.ค. 56
มิ.ย. 56
ก.ค. 56
ส.ค. 56
ก.ย. 56
ต.ค. 56
พ.ย. 56
ธ.ค. 56
ม.ค. 57
ก.พ. 57
มี.ค. 57
ศึกษาทฤษฎีของเสียงวรรณยุกต์ในภาษาไทย
ศึกษาทฤษฎีของรูปแบบโครงร่างความถี่มูลฐานของเสียงวรรณยุกต์ในภาษาไทย
ศึกษางานวิจัยเกี่ยวกับค่าลักษณะสำคัญที่ส่งผลต่อการผันวรรณยุกต์ภาษาไทย
ศึกษางานวิจัยค่าลักษณะสำคัญของโครงร่างความถี่มูลฐานที่ส่งผลต่อเสียงวรรณยุกต์ภาษาไทย
ศึกษางานวิจัยการพัฒนาความเป็นธรรมชาติของเสียงสังเคราะห์ในระบบการสังเคราะห์เสียงภาษาไทย
ศึกษางานวิจัยค่าลักษณะสำคัญของโครงร่างความถี่มูลฐานที่ส่งผลต่อการระบุเสียงวรรณยุกต์ภาษาไทย
ศึกษาเครื่องมือที่ใช้ในงานวิจัย
เลือกใช้เครื่องมือในการวิจัยที่เหมาะสม
ออกแบบการทดลอง
เก็บข้อมูลเสียงจากการบันทึกเสียง
ทำการทดลองการรับรู้ความเป็นธรรมชาติของเสียงวรรณยุกต์ในภาษาไทย
วิเคราะห์ผลการทดลองการรับรู้ความเป็นธรรมชาติของเสียงวรรณยุกต์ในภาษาไทย
สร้างแบบจำลองการรับรู้ความเป็นธรรมชาติของเสียงวรรณยุกต์ในภาษาไทย
ทดสอบและวัดผลของแบบจำลองการรับรู้ความเป็นธรรมชาติของเสียงวรรณยุกต์ในภาษาไทย
วิเคราะห์ผลการทดลอง
สรุปผลและเรียบเรียงวิทยานิพนธ์

บทที่ 2
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ในบทนี้ จะกล่าวถึงทฤษฎีและเอกสารงานวิจัยที่เกี่ยวข้อง แบ่งออกเป็น 2 ส่วน ซึ่งในส่วนแรก จะกล่าวถึงทฤษฎีที่เกี่ยวข้อง ได้แก่ ทฤษฎีทางภาษาศาสตร์ ซึ่งประกอบไปด้วย ระบบอักษรไทยและการผันวรรณยุกต์ ทฤษฎีระบบสังเคราะห์เสียงพูด และทฤษฎีทางคณิตศาสตร์ที่เกี่ยวข้องกับงานวิจัย ในส่วนที่สอง กล่าวถึงงานวิจัยที่เกี่ยวข้องกับวิทยานิพนธ์นี้

2.1 ทฤษฎีที่เกี่ยวข้อง

2.1.1 ทฤษฎีทางภาษาศาสตร์

ในภาษาไทย มีตัวอักษรที่ใช้แทนเสียงอยู่ 3 ชนิดด้วยกัน ซึ่งได้แก่ สระ พยัญชนะ และวรรณยุกต์ [10]

1. พยัญชนะ

สำหรับภาษาไทย มีพยัญชนะทั้งหมด 44 รูป ได้แก่ ก ข ฃ ค ฅ ฆ ง จ ฉ ช ซ ฌ ญ ฎ ฏ ฐ ฑ ฒ ณ ด ต ถ ท ธ น บ ป ผ ฝ พ ฟ ภ ม ย ร ล ว ศ ษ ส ห ฬ อ ฮ ซึ่งประกอบไปด้วย 21 หน่วยเสียง และสามารถแบ่งตามหน้าที่ได้คือ พยัญชนะต้น และตัวสะกด ซึ่งสามารถจับคู่ระหว่างพยัญชนะกับสัญลักษณ์หน่วยเสียงได้ดังตารางที่ 2.1

ตารางที่ 2.1 เปรียบเทียบพยัญชนะกับสัญลักษณ์หน่วยเสียง [11]

พยัญชนะในภาษาไทย	สัญลักษณ์หน่วยเสียง	
	พยัญชนะต้น	ตัวสะกด
ก	k	/k/
ข ฃ ค ฅ ฆ	kh	
ง	ng	/ng/
จ	c	
ฉ ช ฌ	ch	
ซ ศ ษ ส	s	
ญ ย	j	/j/
ฎ ด	d	/t/
ฐ ฑ ฒ ถ ท ธ	t	

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

พยัญชนะในภาษาไทย	สัญลักษณ์หน่วยเสียง	
	พยัญชนะต้น	ตัวสะกด
ณ น	n	/n/
บ	b	/p/
ป	p	
พ ภ ผ	ph	
ฟ ฝ	f	
ม	m	/m/
ร	r	
ล ฬ	l	
ว	w	/w/
ห ฮ	h	-
อ	z	-

สำหรับตัวสะกดในภาษาไทยนั้น ประกอบด้วย 8 มาตรา ได้แก่ แม่กก แม่กด แม่กบ แม่กน แม่กง แม่กม แม่เกอย และแม่เกอว โดยแต่ละมาตรานั้น ประกอบด้วยพยัญชนะต่าง ๆ ดังแสดงในตารางที่ 2.2

ตารางที่ 2.2 มาตราตัวสะกดในภาษาไทย

มาตราตัวสะกด	พยัญชนะตัวสะกด
แม่กก	ก ข ฃ ค ฅ ฆ
แม่กด	จ ฉ ช ซ ฌ ฎ ฏ ฐ ฑ ฒ ด ต ถ ท ธ ศ ษ ส
แม่กบ	บ ป พ ภ ผ ฝ ฟ
แม่กน	ญ ณ น ร ล ฬ
แม่กง	ง
แม่กม	ม
แม่เกอย	ย
แม่เกอว	ว

2. สระ

ในภาษาไทย มีสระทั้งหมด 21 รูป คือ ◌ั (ไม้เอก) ◌ั้ (ไม้โท) ◌ั๊ (ไม้ตรี) ◌ั๋ (ไม้จัตวา) 32 เสียง ซึ่งเกิดจากเสียงสระเดี่ยว สระประสม และสระเกิน ดังนี้

สระเดี่ยว : มี 18 เสียง ได้แก่ อะ อา อิ อี อึ อื อุ อู เอะ เอ แอะ แอ โอะ โอ เอาะ ออ เออะ เออ

สระประสม : มี 6 เสียง ได้แก่ เอียะ เอีย เอือะ เอือ อัวะ อัว

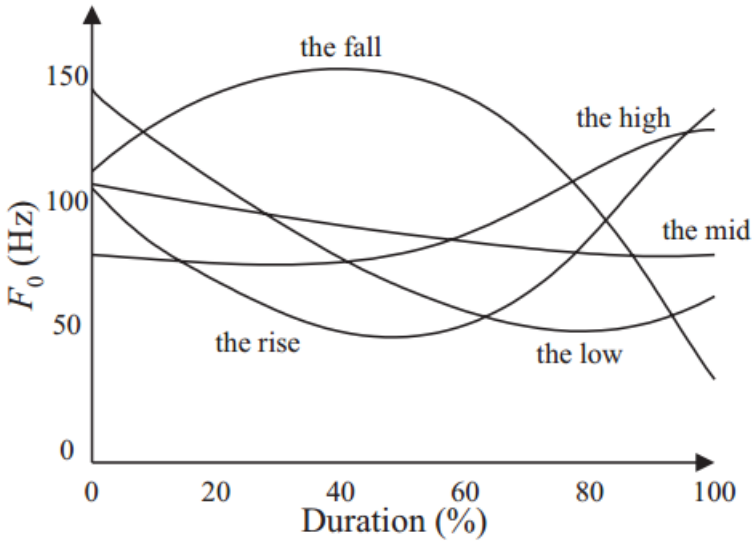
สระเกิน : มี 8 เสียง ได้แก่ อำ (อะ+ม), ไอ (อะ+ย), ใอ (อะ+ย), เอา (อะ+ว), ฤ (ร+อื), ฦ (ล+อื)

3. วรรณยุกต์

เสียงวรรณยุกต์ เกิดจากการประกอบขึ้นของเสียงสูงและต่ำ หากทำให้ความหมายของคำเปลี่ยนแปลงไป ถือว่าเป็นเสียงวรรณยุกต์ ซึ่งในภาษาไทยเป็นภาษาที่มีเสียงวรรณยุกต์ เสียงวรรณยุกต์ในภาษาไทย มีทั้งสิ้น 4 รูป และ 5 เสียง ได้แก่ เสียงสามัญ, เอก, โท, ตรี และจัตวา หากแบ่งตามลักษณะของการเปลี่ยนแปลงของระดับเสียง จะแบ่งได้เป็น 2 ประเภท คือ เสียงวรรณยุกต์แบบสตาติก (Static Tone) และเสียงวรรณยุกต์แบบไดนามิก (Dynamic Tone) ดังนี้

- เสียงวรรณยุกต์แบบสตาติก ประกอบด้วย 3 เสียง ได้แก่ เสียงสามัญ (Mid), เสียงเอก (Low) และ เสียงตรี (High)
- เสียงวรรณยุกต์แบบไดนามิก ประกอบด้วย 2 เสียง ได้แก่ เสียงโท (Falling) และ เสียงจัตวา (Rising)

ซึ่งสามารถอธิบายรูปแบบของการเปลี่ยนแปลงระดับเสียงของเสียงวรรณยุกต์ได้ด้วยโครงร่างความถี่มูลฐานของแต่ละเสียงวรรณยุกต์ ดังแสดงในภาพที่ 2.1 ดังนี้



ภาพที่ 2.1 ลักษณะของเสียงวรรณยุกต์ในภาษาไทย [12]

4. การผันวรรณยุกต์

ในการผันวรรณยุกต์ ประกอบไปด้วยหลายปัจจัยที่ต้องคำนึงถึง ได้แก่ กลุ่มของพยัญชนะต้น คำเป็นคำตาย และสระเสียงสั้น-ยาว ซึ่งมีเกณฑ์ในการผันวรรณยุกต์ได้ดังตารางที่ 2.3

ตารางที่ 2.3 เกณฑ์การผันวรรณยุกต์ในภาษาไทย [10]

กลุ่มพยัญชนะต้น	เสียงสามัญ	เสียงเอก	เสียงโท	เสียงตรี	เสียงจัตวา	หมายเหตุ
อักษรกลาง ได้แก่ ก จ ด ฎ ต ฏ บ ป อ คำเป็น คำตาย	ปา	ป่า กัด	ป้า กั้ด	ป๊า กั๊ด	ป๋า กั๋ด	คำเป็นพื้นเสียงเป็นเสียงสามัญ คำตายพื้นเสียงเป็นเสียงเอก
อักษรสูง ได้แก่ ข ฃ ฉ ฐ ถ ผ ฝ ศ ษ ส ห คำเป็น คำตาย	- -	ข่า ขัด	ข้า ขั้ด	- -	ขา -	คำเป็นพื้นเสียงเป็นเสียงจัตวา คำตายพื้นเสียงเป็นเสียงเอก
อักษรต่ำ ได้แก่ (อักษรที่เหลือ ๒๔ ตัว) คำเป็น คำตายเสียงยาว คำตายเสียงสั้น	คา - -	- - -	ค่า คาด ค่ะ	ค้า ค้าด คะ	- - -	คำเป็นพื้นเสียงเป็นเสียงสามัญ หากผันร่วมกับอักษรสูงจะผันได้ครบ ๕ เสียง เช่น คา ข่า ค่า(ข้า) ค้า ขา

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

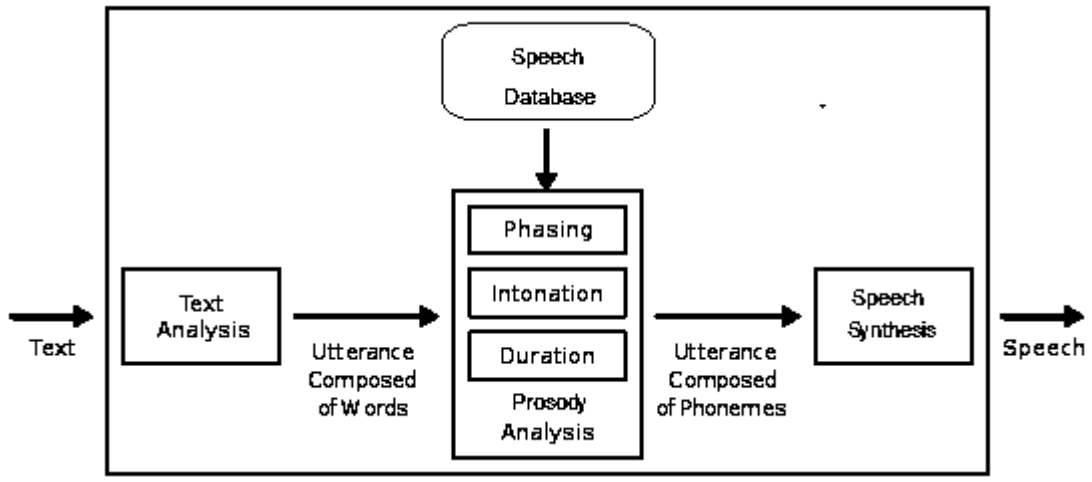
2.1.2 ทฤษฎีระบบสังเคราะห์เสียงพูด

2.1 ระบบการสังเคราะห์เสียงพูด

ระบบการสังเคราะห์เสียงพูด เป็นกระบวนการสร้างเสียงพูดสังเคราะห์ของมนุษย์ ที่ทำการแปลงจากข้อความให้เป็นเสียงพูด ตัวอย่างระบบสังเคราะห์เสียงภาษาไทย เช่น วาจา [13] ซึ่งในปัจจุบัน สามารถนำระบบสังเคราะห์เสียงพูด ไปประยุกต์ใช้ได้หลากหลาย ตัวอย่างเช่น โปรแกรมช่วยอ่านหนังสือเสียง โปรแกรมช่วยอ่านหน้าจอสำหรับผู้พิการทางสายตา ระบบเสียงพูดภายในระบบนำทาง เป็นต้น และเพื่อตอบสนองความต้องการในการสื่อสารกันระหว่างมนุษย์และคอมพิวเตอร์ ในการสร้างสัญญาณเสียงพูดส่งต่อไปยังบุคคล ความตั้งใจของการสร้างระบบนี้ก็เพื่อสร้างเสียงพูดสังเคราะห์

ที่ไม่เพียงแต่จะสามารถเข้าใจได้ง่ายเท่านั้น แต่ยังต้องไม่สามารถแยกแยะออกจากเสียงที่สร้างโดยมนุษย์เอง หรือกล่าวอีกแบบก็คือ สร้างระบบเท่าเทียมกับความสามารถของมนุษย์ในการสร้างเสียงพูด ดังนั้น คุณภาพสองสิ่งที่เป็นความต้องการของระบบการสังเคราะห์เสียงพูดคือ ความเป็นธรรมชาติ (naturalness) และ ความเข้าใจง่าย (Intelligibility) ของเสียงพูดสังเคราะห์ [2]

ระบบการสังเคราะห์เสียงพูด ถูกแบ่งออกเป็น 3 ส่วนหลัก ๆ ได้แก่ ส่วนการวิเคราะห์ข้อความ (Text Analysis) ส่วนการวิเคราะห์สัทสัมพันธ์ (Linguistic Analysis) และการสังเคราะห์เสียงพูด (Speech Synthesis) ดังแสดงในภาพที่ 2.2



ภาพที่ 2.2 ภาพรวมของระบบการแปลงข้อความเป็นเสียงพูด [14]

1. ส่วนการวิเคราะห์ข้อความ

ในส่วนนี้ จะทำการประมวลผลข้อความที่รับเข้ามาในระบบ เพื่อแปลงให้อยู่ในรูปเสียงอ่าน (Phoneme) โดยประกอบด้วยการประมวลผลต่าง ๆ คือ การตัดประโยค จากข้อความที่ยาวให้เป็นข้อความที่สั้นลง, การทำข้อความให้เป็นบรรทัดฐาน เช่นการแปลงสัญลักษณ์ คำย่อ หรือตัวเลขต่าง ๆ ให้อยู่ในรูปเสียงอ่าน และการหาขอบเขตของวลีเพื่อเว้นวรรคการอ่าน เป็นต้น

2. ส่วนการวิเคราะห์สัทสัมพันธ์

ในส่วนนี้ คือการวิเคราะห์ลักษณะสำคัญของเสียงพูด เพื่อควบคุมลักษณะของเสียงพูด เช่น ระยะเวลาของพยางค์ (Duration) ซึ่งส่งผลต่อความเร็วของเสียงพูดสังเคราะห์, ค่าความถี่มูลฐานที่เปลี่ยนไปตามเวลา หรือ โครงร่างความถี่มูลฐาน ซึ่งส่งผลโดยตรงต่อเสียงวรรณยุกต์ในภาษาที่ใช้เสียงวรรณยุกต์ดังเช่นภาษาไทย และเสียงสูงต่ำในการพูด (Intonation) ที่สะท้อนถึงประเภทของประโยคในการพูด เช่น ประโยคคำถาม ประโยคปฏิเสธ หรือ ประโยคบอกเล่า

3. ส่วนการสังเคราะห์เสียงพูด

ในส่วนนี้ จะนำข้อมูลเสียงอ่าน จากขั้นตอนการวิเคราะห์ข้อความ และข้อมูลสัทสัมพันธ์จากขั้นตอนการวิเคราะห์สัทสัมพันธ์ มาสร้างให้เกิดเป็นเสียงสังเคราะห์ขึ้น ซึ่งมี 3 วิธีหลักๆ คือ การสังเคราะห์เสียงแบบสั่นพ้อง (Formant Synthesis), การสังเคราะห์เสียงโดยอาศัยแบบจำลองอวัยวะที่ใช้ในการเปล่งเสียง (Articulatory Synthesis) และ การสังเคราะห์เสียงพูดแบบเชื่อมต่อ (Concatenative Synthesis)

3.1 การสังเคราะห์เสียงแบบสั่นพ้อง

เป็นการจำลองความถี่ของสัญญาณเสียงที่สั่นพ้องในช่องเสียง (Vocal tract) ซึ่งถูกกำหนดอยู่ในรูปของความถี่ฟอร์แมนท์ (F1, F2, F3) โดยเสียงพูดสังเคราะห์จะได้จากการประมาณค่าความถี่ฟอร์แมนท์ เพื่อให้ได้สเปคตรัมของเสียงพูดที่ต้องการ และควบคุมแหล่งกำเนิดเสียงเพื่อสร้างช่วงเสียงก้อง (Voicing) และช่วงไม่มีเสียงก้อง (Voicelessness) ซึ่งเสียงสังเคราะห์ที่ได้จากวิธีนี้นั้นไม่ค่อยเป็นธรรมชาติ มีลักษณะคล้ายหุ่นยนต์ ซึ่งไม่เป็นไปตามเป้าหมายของการทำให้เสียงมีความเป็นธรรมชาติ เนื่องจากต้องใช้กฎเกณฑ์ในการสังเคราะห์เสียงเป็นจำนวนมาก ซึ่งทำได้ยาก แต่ข้อดีคือ การใช้กฎเกณฑ์ต่าง ๆ นั้นสามารถฟังแล้วเข้าใจง่ายแม้ในการพูดที่เร็ว ซึ่งมีประโยชน์ในการใช้งานคอมพิวเตอร์โดยใช้โปรแกรมอ่านหน้าจอคอมพิวเตอร์ได้อย่างรวดเร็วสำหรับผู้พิการทางการเห็น อีกทั้งยังสามารถนำไปใช้ในระบบฝังตัว (Embedded System) ได้ดี เนื่องจากไม่มีการใช้ฐานข้อมูลของตัวอย่างเสียง ตัวอย่างระบบสังเคราะห์เสียงที่ใช้วิธีการนี้ ได้แก่ DECTALK [15]

3.2 การสังเคราะห์เสียงโดยอาศัยแบบจำลองอวัยวะที่ใช้ในการเปล่งเสียง

สร้างเสียงสังเคราะห์โดยสร้างแบบจำลองโดยตรงของพฤติกรรมทางกายภาพ ของการออกเสียงของมนุษย์ ซึ่งตามทฤษฎีแล้ว วิธีนี้เป็นวิธีที่ตอบสนองความต้องการที่สุดในการสร้างเสียงพูดสังเคราะห์ที่มีคุณภาพสูง แต่ในทางปฏิบัตินั้น เป็นวิธีที่ทำยากวิธีหนึ่ง

3.3 การสังเคราะห์เสียงพูดแบบเชื่อมต่อ

วิธีนี้เป็นวิธีที่ใช้การขับเคลื่อนด้วยข้อมูล เข้ามาช่วยในการสร้างเสียงสังเคราะห์ โดยจะสร้างเสียงพูดสังเคราะห์จากการต่อกันของหลายยูนิตของเสียงพูดจริงที่บันทึกไว้แล้ว ซึ่งยูนิตดังกล่าวนั้นสามารถเป็นได้ทั้ง คำ (Word), พยางค์ (Syllable), หน่วยครึ่งพยางค์ (Half-syllable), หน่วยเสียงเดี่ยว (Monophone), หน่วยเสียงคู่ (Diphone) หรือ หน่วยเสียงสาม (Triphone) ซึ่งต้องอาศัยฐานข้อมูลเพื่อเก็บยูนิตเสียงไว้เป็นจำนวนมาก หากต้องการให้เสียงมีคุณภาพดี ซึ่งโดยทั่วไปแล้ว ยูนิตที่นิยมใช้ในระบบการสังเคราะห์เสียงพูดแบบเชื่อมต่อคือ หน่วยเสียงคู่ ในปัจจุบันมีวิธีการสังเคราะห์เสียงแบบ Concatenative synthesis อยู่ 2 วิธีที่ได้รับความนิยมก็คือ การสังเคราะห์เสียงโดยการคัดเลือกหน่วยเสียง (Unit-selection Speech Synthesis) และ การสังเคราะห์เสียงโดยใช้แบบจำลองฮิดเดนมาร์คอฟ (HMM-based Speech Synthesis)

2.1.3 ทฤษฎีทางคณิตศาสตร์ที่เกี่ยวข้อง

ในส่วนนี้ จะกล่าวถึงทฤษฎีทางคณิตศาสตร์ที่ใช้ในงานวิจัยนี้ คือแบบจำลองเรขาคณิต (Geometric Model) และการปรับเส้นโค้ง (Curve Fitting)

1. แบบจำลองเรขาคณิต (Geometric Model)

แบบจำลองทางเรขาคณิต เป็นวิธีทางคณิตศาสตร์ประยุกต์เพื่อบรรยายรูปร่างต่าง ๆ หรือรูปเรขาคณิต ด้วยคณิตศาสตร์ เพื่อใช้ในการแสดง การอธิบายความสัมพันธ์หรือช่วยแก้ปัญหาต่าง ๆ โดยการแปลงปัญหาให้อธิบายได้ด้วยเรขาคณิต โดยรูปเรขาคณิต (Geometric Figure) นั้นประกอบด้วย จุด เส้นตรง เส้นโค้ง และระนาบ อย่างน้อยหนึ่งอย่าง โดยส่วนใหญ่ รูปร่างที่เกี่ยวข้องในแบบจำลองเรขาคณิต เป็นรูปร่างสองมิติ หรือ สามมิติ ซึ่งมีตัวอย่างของรูปเรขาคณิต ซึ่งแบ่งตามมิติได้ดังนี้

- รูปเรขาคณิตหนึ่งมิติ เช่น เส้นตรง ส่วนของเส้นตรง
- รูปเรขาคณิตสองมิติ เช่น มุม วงกลม รูปสามเหลี่ยม รูปสี่เหลี่ยม หรือ รูปหลายเหลี่ยม
- รูปเรขาคณิตสามมิติ เช่น ลูกบาศก์ ทรงกลม ปริซึม ทรงกระบอก พีระมิด

ในวิทยานิพนธ์นี้ จะเกี่ยวข้องกับการสร้างแบบจำลองเรขาคณิตสองมิติ ซึ่งเป็นบริเวณพื้นที่ปิดบนกราฟ อาจสร้างได้ด้วยแบบจำลองทางคณิตศาสตร์ เช่น สมการ หรือ อสมการ ประกอบกันเพื่อให้เกิดเป็นรูปเรขาคณิตขึ้น

2. การปรับเส้นโค้ง (Curve Fitting)

ข้อมูลส่วนใหญ่ของปัญหาทางวิศวกรรม มักจะเป็นข้อมูลที่เป็นจุดไม่ต่อเนื่อง ซึ่งมีความต้องการที่จะวาดเส้นกราฟที่ต่อเนื่องผ่านจุดต่าง ๆ เหล่านั้น ทั้งนี้ เพื่ออธิบายลักษณะของข้อมูลนั้น ๆ ซึ่งจะใช้วิธีการปรับเส้นโค้ง หรือ การเลือกเส้นกราฟที่เหมาะสมกับข้อมูล ซึ่งมี 2 วิธี ได้แก่ วิธีการประมาณค่าในช่วง (Interpolation) และการถดถอยเชิงเส้น (Regression) ซึ่งวิธีการประมาณค่าในช่วงนั้น เป็นการหาฟังก์ชัน หรือเส้นกราฟ ที่ผ่านจุดทุกจุดของข้อมูล ซึ่งวิธีนี้ เหมาะสำหรับข้อมูลที่ทราบแน่นอน และมีความถูกต้อง หรือได้จากการทำการทดลองที่แน่ชัด ส่วนวิธีการถดถอยเชิงเส้นนั้นเป็นการหาเส้นกราฟให้ผ่านจุดข้อมูลที่มีอยู่ให้ได้มากที่สุด ซึ่งจะบ่งบอกถึงแนวโน้มของข้อมูล ซึ่งเหมาะสำหรับข้อมูลที่ได้จากการทดลองที่มีความแปรปรวนสูง สำหรับวิทยานิพนธ์นี้จะกล่าวถึงวิธีการปรับเส้นโค้งด้วยวิธีการประมาณค่าในช่วง เนื่องจากข้อมูลที่ได้รับ เกิดจากการทดลองที่มีคำตอบที่แน่นอน

การประมาณค่าในช่วง

ในส่วนนี้ จะกล่าวถึงการประมาณค่าในช่วง 2 วิธี ได้แก่ การประมาณค่าในช่วงด้วยฟังก์ชันพหุนามของนิวตัน (Newton's Interpolating Polynomials) และการประมาณค่าในช่วงด้วยเส้นโค้งอันดับสาม (Cubic Spline) [16]

1 การประมาณค่าในช่วงด้วยฟังก์ชันพหุนามของนิวตัน

ฟังก์ชันพหุนามนั้นมีรูปทั่วไป ดังสมการที่ 2.1

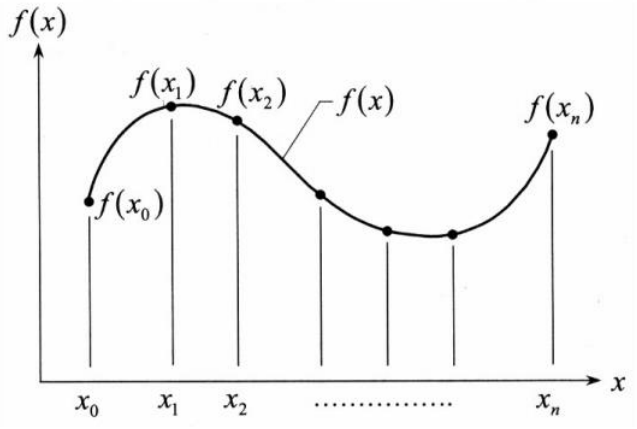
$$f(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n \qquad (2.1)$$

ฟังก์ชันพหุนามอันดับ 1 : n=1, $a_1 \neq 0$ หรือก็คือฟังก์ชันเส้นตรงนั่นเอง

ฟังก์ชันพหุนามอันดับ 2 : n=2, $a_2 \neq 0$

ฟังก์ชันพหุนามที่มีอันดับมากกว่า 2 : $n \geq 2$, $a_n \neq 0$

การประมาณค่าในช่วงฟังก์ชันพหุนามในรูปทั่วไปนั้น จะใช้ประมาณค่าของกลุ่มข้อมูลใดๆ กำหนดให้มีข้อมูลทั้งสิ้น n+1 จุด โดยจะมีเส้นฟังก์ชัน f(x) ลากผ่านทุกจุดของข้อมูล จะสามารถประมาณค่า f(x) นั้น ด้วยพหุนามอันดับ n ใด ๆ และ $f(x_i) \approx P_n(x_i)$ ซึ่งอธิบายได้ดังภาพที่ 2.3



ภาพที่ 2.3 กราฟฟังก์ชันพหุนามอันดับที่ n, $P_n(x)$ [16]

CHULALONGKORN UNIVERSITY

สมการพหุนามอันดับที่ n สามารถเขียนได้ดังสมการที่ 2.2

(2.2)

$$f_n(x) = b_0 + b_1 \cdot (x - x_0) + b_2 (x - x_0)(x - x_1) + \ldots + b_n \cdot (x - x_0)(x - x_1)(x - x_2)\ldots(x - x_{n-1})$$

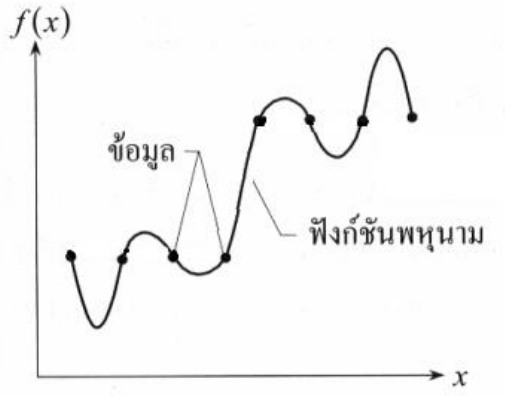
โดยที่ $b_0, b_1, b_2, \ldots, b_n$ หาได้จากสมการที่ 2.3

$$
\begin{aligned}
b_0 &= f(x_0) \\
b_1 &= f[x_1, x_0] \\
b_2 &= f[x_2, x_1, x_0] \\
&\vdots \\
b_n &= f[x_n, x_{n-1}, \ldots, x_1, x_0]
\end{aligned}
\tag{2.3}
$$

โดยที่ $f[x_i, x_j]$ คือค่าผลต่างจากการแบ่งย่อย (Finite Divided Differences) ซึ่งสามารถคำนวนได้ดังนี้

$$
\begin{aligned}
f[x_i, x_j] &= \frac{f(x_i) - f(x_j)}{x_i - x_j} \\
f[x_i, x_j, x_k] &= \frac{f[x_i, x_j] - f[x_j, x_k]}{x_i - x_k} \\
&\vdots \\
f[x_n, x_{n-1}, \ldots, x_1, x_0] &= \frac{f[x_n, x_{n-1}, \ldots, x_1] - f[x_{n-1}, x_{n-2}, \ldots, x_0]}{x_n - x_0}
\end{aligned}
\tag{2.4}
$$

ในการประมาณค่าในช่วงด้วยฟังก์ชันพหุนามนั้น ผลลัพธ์จะได้เป็นฟังก์ชันต่อเนื่องที่ลากผ่านจุดทุกจุดข้อข้อมูล หากมีจุดของข้อมูลมากขึ้น ก็ต้องใช้ฟังก์ชันอันดับในที่สูงขึ้น ซึ่งทำให้กราฟที่ได้นั้นเป็นเส้นโค้งกลับไปกลับมา ดังภาพที่ 2.4



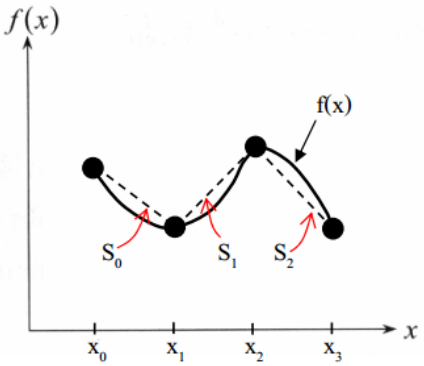
ภาพที่ 2.4 ตัวอย่างกราฟของฟังก์ชันพหุนามอันดับ 7

2 การประมาณค่าในช่วงด้วยเส้นโค้งแบบสไปลน์ (Spline Interpolation)

ในกาประมาณค่าในช่วงด้วยเส้นโค้งนั้น มีหลักการคือ การแทนจุดข้อมูล n+1 จุด ด้วยเส้นโค้งแบบสไปลน์จำนวน n เส้น ประเภทของเส้นโค้งแบบสไปลน์ เช่น เส้นโค้งแบบสไปลน์อันดับหนึ่ง (Linear Spline), เส้นโค้งแบบสไปลน์อันดับสอง (Quadratic Spline) หรือพาราโบล่า และเส้นโค้งแบบสไปลน์อันดับสาม (Cubic Spline) เป็นต้น ซึ่งในที่นี้ จะกล่าวถึงรายละเอียดของเส้นโค้งแบบสไปลน์อันดับสามเป็นหลัก ซึ่งเป็นที่นิยมใช้กันมากที่สุดในทางปฏิบัติ เนื่องจากสามารถทำการปรับเส้นโค้งได้เรียบ (Smooth)

เส้นโค้งแบบสไปลน์อันดับหนึ่ง

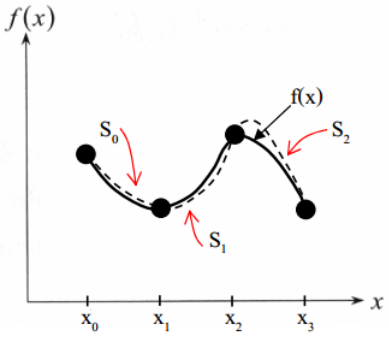
ฟังก์ชันที่ประมาณค่าระหว่างจุด จะเป็นฟังก์ชันเส้นตรง เช่น มีข้อมูล 4 จุด ที่ตำแหน่ง x_0, x_1, x_2, และ x_3 สามารถปรับเส้นด้วยเส้นโค้งแบบสไปลน์อันดับหนึ่งได้ 3 เส้น ดังภาพที่ 2.5 ซึ่ง f(x) เป็นฟังก์ชันจริงที่ผ่านข้อมูลทั้ง 4 จุด หรือจุดข้อมูลทั้ง 4 จุดถูกวัดค่าออกมาจากฟังก์ชันนั้นก็ได้ โดย S_n หมายถึงเส้นโค้งสไปลน์ที่ประมาณค่าฟังก์ชันจริง f(x) ในแต่ละช่วงของค่า x แต่สังเกตได้ว่า มีความไม่ต่อเนื่องที่จุดข้อมูล



ภาพที่ 2.5 กราฟแสดงการปรับเส้นด้วยเส้นโค้งแบบสไปลน์อันดับหนึ่ง

เส้นโค้งแบบสไปลน์อันดับสอง

เส้นโค้งแบบสไปลน์อันดับสอง แก้ปัญหาความไม่ต่อเนื่องที่จุดข้อมูล ของเส้นโค้งแบบสไปลน์อันดับหนึ่งได้ โดยการใช้เส้นโค้งแบบสไปลน์อันดับสอง ซึ่งทำให้บริเวณรอยต่อของเส้นโค้งแบบสไปลน์นั้น มีความต่อเนื่องแบบอนุพันธ์อันดับหนึ่ง ดังตัวอย่างในภาพที่ 2.6

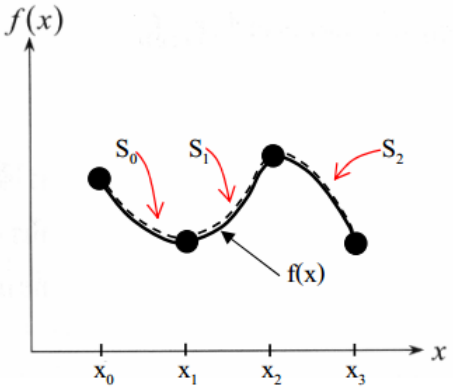


ภาพที่ 2.6 กราฟแสดงการปรับเส้นด้วยเส้นโค้งแบบสไปลน์อันดับสอง

แต่ในกรณีที่มีข้อมูลเกิดการเปลี่ยนแปลงกระทันหันในช่วงข้อมูล x ที่น้อยเกินไป เส้นโค้งแบบสไปลน์อันดับสอง จะมีความเบี่ยงเบนไปจากค่าฟังก์ชันจริง ดังส่วนโค้ง S_2 ในภาพที่ 2.6

เส้นโค้งแบบสไปลน์อันดับสาม

การประมาณค่าในช่วงด้วยเส้นโค้งแบบสไปลน์อันดับสาม จะเป็นการประมาณค่าที่มีความใกล้เคียงกับค่าของฟังก์ชันจริง f(x) มาก และมีความเรียบในบริเวณรอยต่อของเส้นโค้งแบบสไปลน์ที่อยู่ติดกันที่จุดข้อมูล โดยมีอัตราการเปลี่ยนแปลงของความชันที่จุดข้อมูลแต่ละจุด แบบอนุพันธ์อันดับสอง ซึ่งนอกจากเส้นโค้งแบบสไปลน์ที่อยู่ติดกันตรงจุดข้อมูลจะมีความชันเท่ากันแล้ว ยังมีอัตราการเปลี่ยนแปลงของความชันเท่ากันด้วย ทำให้การใช้เส้นโค้งแบบสไปลน์อันดับสาม มีความเรียบมาก ในการประมาณค่าข้อมูลในช่วง ดังตัวอย่างในภาพที่ 2.7

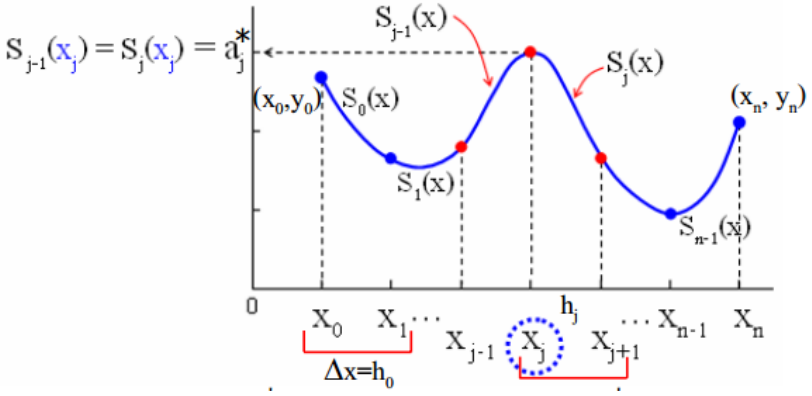


ภาพที่ 2.7 กราฟแสดงการปรับเส้นด้วยเส้นโค้งแบบสไปลน์อันดับสาม

โดยเส้นโค้งแบบสไปลน์อันดับสามในรูปทั่วไป มีสมการเป็น

$$S_j(x) = a_j + b_j(x - x_j) + c_j(x - x_j)^2 + d_j(x - x_j)^3 \qquad (2.5)$$

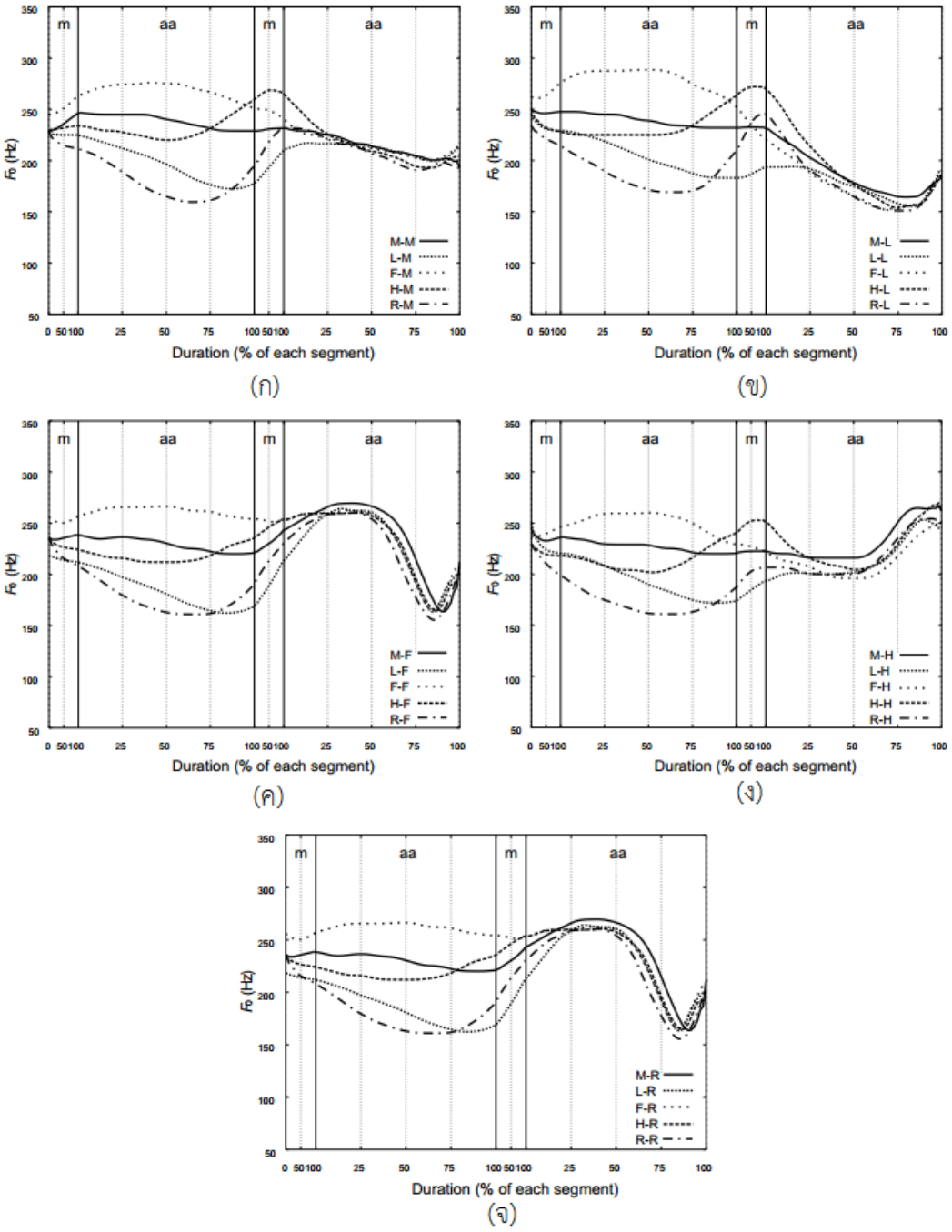
โดยที่ $0 \le j \le n-1$, และ $S_j(x)$ เป็นเส้นโค้งสไปลน์ของช่วง (x_j, x_{j+1}) และ a_j, b_j, c_j, d_j เป็นสัมประสิทธิ์ที่ไม่ทราบค่า ดังแสดงในภาพที่ 2.8



ภาพที่ 2.8 กราฟแสดงเส้นโค้งแบบสไปลน์อันดับสามในรูปทั่วไป

2.2 เอกสารและงานวิจัยที่เกี่ยวข้อง

ในภาษาไทยนั้นประกอบไปด้วยเสียงวรรณยุกต์ 5 เสียง ได้แก่เสียงวรรณยุกต์สามัญ, เอก, โท , ตรี และจัตวา ซึ่งสามารถแยกแยะได้โดยใช้การเปลี่ยนแปลงของค่าความถี่มูลฐานตามแกนเวลา หรือเรียกว่า ค่าโครงร่างความถี่มูลฐาน (Fundamental Frequency Contour; F_0 Contour) ภายในขอบเขตของพยางค์ ซึ่งโดยทั่วไปแล้ว สามารถแบ่งเสียงวรรณยุกต์ตามลักษณะการเปลี่ยนแปลงของโครงร่างความถี่มูลฐานได้เป็น 2 ประเภท คือ 1. เสียงวรรณยุกต์แบบสตาติก (Static Tone) คือเสียงวรรณยุกต์ที่มีอัตราการเปลี่ยนแปลงของโครงร่างความถี่มูลฐานค่อนข้างคงที่ ได้แก่ เสียงวรรณยุกต์สามัญ, เอก และตรี 2. เสียงวรรณยุกต์แบบไดนามิก (Dynamic Tone) คือเสียงวรรณยุกต์ที่มีอัตราการเปลี่ยนแปลงของโครงร่างความถี่มูลฐานสูง ได้แก่เสียงวรรณยุกต์โท และจัตวา [12] แต่ว่าต่อมา ผณินทรา ธีรานนท์ [17] ได้เสนอว่า เสียงวรรณยุกต์ตรีนั้น มีความคล้ายคลึงกับเสียงวรรณยุกต์แบบไดนามิกมากกว่าแบบสตาติก ในการรับรู้เสียงวรรณยุกต์นั้น ปัจจัยที่สำคัญในการแยกแยะเสียงวรรณยุกต์คือระดับ และรูปร่างของโครงร่างความถี่มูลฐาน [4, 5] และยังมีปัจจัยที่สำคัญที่ทำให้โครงร่างความถี่มูลฐานของพยางค์เดี่ยว กับพยางค์ที่อยู่ในการพูดต่อเนื่อง มีการเปลี่ยนแปลงไปเล็กน้อย ซึ่งก็คือ พยางค์ที่อยู่ด้านหน้า และด้านหลังของพยางค์นั้น ๆ โดย Gandour และคณะ [5] ได้ทำการวาดโครงร่างความถี่มูลฐานของคำพูดสองพยางค์ ที่เป็นการจับคู่ของเสียงวรรณยุกต์ทั้งหมด พบว่าปัจจัยดังกล่าวคือโคอาติคูเลชัน เอฟเฟกต์ (Coarticulation Effect) ได้แก่ แคร์รี่โอเวอร์ เอฟเฟกท์ ซึ่งมีผลต่อช่วงต้นของโครงร่างความถี่มูลฐานของพยางค์เนื่องจากเสียงวรรณยุกต์ก่อนหน้า ซึ่งมีผลกระทบต่อพยางค์ถัดไปถึง 75% ของช่วงเวลาของพยางค์ถัดไป และเอนติซิพาทอรี่ เอฟเฟกท์ (Anticipatory Effect) ซึ่งมีผลต่อช่วงท้ายของโครงร่างความถี่มูลฐานของพยางค์เนื่องจากเสียงวรรณยุกต์ถัดไป ซึ่งมีผลกระทบต่อพยางค์ก่อนหน้าถึง 50% ของช่วงเวลาของพยางค์ก่อนหน้านับจากท้ายพยางค์ และเนื่องจากในช่วงท้ายของพยางค์มีการเปลี่ยนแปลงของโครงร่างความถี่มูลฐานมากกว่าในช่วงต้นพยางค์ในแต่ละเสียงวรรณยุกต์ ซึ่ง [18] ได้กำหนดบริเวณที่มีข้อมูลของเสียงวรรณยุกต์ที่สำคัญว่า บริเวณวิกฤติของเสียงวรรณยุกต์ (Tone Critical Segment) ซึ่งได้อ้างอิงบทความของ [19, 20] ว่า พยางค์ในส่วนของหน่วยตามพยางค์ (Rhyme) นั้นมีข้อมูลที่สำคัญในการรับรู้เสียงวรรณยุกต์ ในขณะที่ส่วนของหน่วยเริ่มพยางค์ (Onset) นั้น มีความแปรปรวนสูงเนื่องจากโครงร่างความถี่มูลฐานถูกรบกวนโดยเสียงวรรณยุกต์จากพยางค์ที่อยู่ทั้งด้านหน้าและด้านหลัง ซึ่งได้มีการศึกษาผลกระทบจากโคอาติคูเลชัน เอฟเฟกต์ ใน [21, 22] ดังแสดงในภาพที่ 2.9



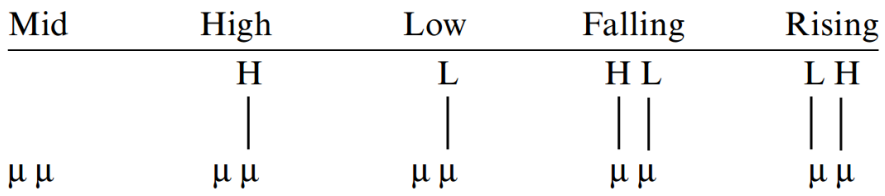
ภาพที่ 2.9 แสดงผลกระทบที่เกิดจากแคร่ีโอเวอร์เอฟเฟกต์ในเสียงวรรณยุกต์ภาษาไทย
คัดลอกจาก [22]

จากภาพที่ 2.9 (ก), (ข), (ค), (ง) และ(จ) แสดงโครงร่างความถี่มูลฐานที่เกิดแคร่ีโอเวอร์เอฟเฟกต์ที่มีเสียงวรรณยุกต์ของพยางค์ก่อนหน้าเป็นเสียงสามัญ, เอก, โท, ตรี และจัตวา ตามลำดับ โดยบันทึกจากคำว่า “/ma-ma/” ซึ่งแสดงผลกระทบของ ma/ พยางค์ที่สอง ที่เกิดจากเสียงวรรณยุกต์จาก /ma พยางค์ก่อนหน้า จะสังเกตได้ว่า มีความแปรปรวนเกิดขึ้นบริเวณช่วงต้นของพยางค์ที่ 2 เมื่อคำว่า /ma พยางค์แรก มีเสียงวรรณยุกต์ที่แตกต่างกัน ซึ่งสามารถสรุปได้ว่า เกิดความแปรปรวนในช่วง

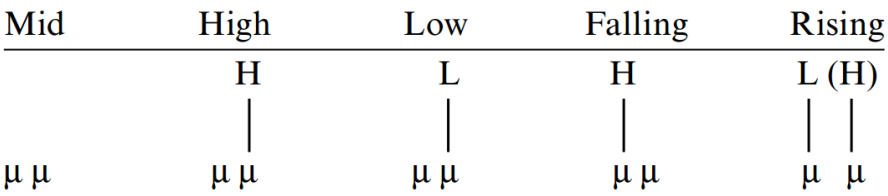
ต้นพยางค์ (Onset) หากพยางค์ด้านหน้าเปลี่ยนเสียงวรรณยุกต์ไป แต่หากสังเกตตั้งแต่ 25% ของระยะเวลาของพยางค์เป็นต้นไป พบว่าโครงร่างความถี่มูลฐานถูกคงลักษณะของรูปร่าง ซึ่งเป็นการเปลี่ยนแปลงโครงร่างความถี่มูลฐานที่เป็นไปในรูปแบบเดียวกันในแต่ละเสียงวรรณยุกต์ ซึ่งจะเป็นได้ว่า มีการเปลี่ยนแปลงที่คงที่ คล้ายๆ กัน ถึงแม้ว่า tone ของพยางค์หน้าจะเปลี่ยนไปก็ตาม

สำหรับงานวิจัยในการรับรู้เสียงวรรณยุกต์นั้น Zsiga E.C. และ Nitisaroj R. [23] ได้ทำการทดลองการรับรู้ในการระบุเสียงวรรณยุกต์ภาษาไทย ในบริบททั้งแบบพยางค์เดี่ยวหรือพยางค์ที่อยู่ท้ายประโยค (Phrase-final Position) และพยางค์ที่ไม่ได้อยู่ท้ายประโยค (Nonphrase-final Position) ด้วยเสียงตัวกระตุ้น ซึ่งเสียงตัวกระตุ้นนั้น ได้จากการปรับแต่งโครงร่างพื้นฐานตามการแทนเสียงวรรณยุกต์ด้วยหน่วยเสียงย่อย ดังที่ [24] ได้เสนอสมมติฐานในการปรับแนวของหน่วยเวลาย่อย (Moraic Alignment Hypothesis) ดังภาพที่ 2.10

(ก)



(ข)



ภาพที่ 2.10 แสดงสมมติฐานในการปรับแนวของหน่วยเวลาย่อย (Moraic Alignment Hypothesis)
(ก) แสดงสมมติฐานในการปรับแนวของหน่วยเวลาย่อยในตำแหน่งพยางค์ที่อยู่ท้ายประโยค
(ข) แสดงสมมติฐานในการปรับแนวของหน่วยเวลาย่อยในตำแหน่งพยางค์ที่อยู่กลางประโยค
คัดลอกจาก [24]

ความแตกต่างของแต่ละเสียงวรรณยุกต์ในภาษาไทยนั้นสามารถอธิบายได้ด้วยส่วนย่อยของหน่วยพื้นฐาน ซึ่งแทนด้วย สูง (H) และ ต่ำ (L) ในหน่วยเวลาย่อยของพยางค์ซึ่งได้นำมาประยุกต์ใช้กับลักษณะโครงร่างความถี่มูลฐานของเสียงวรรณยุกต์ภาษาไทย โดยสำหรับภาษาไทย ได้แบ่งพยางค์ออกเป็น 2 หน่วยเวลาย่อยที่เท่ากัน หรือเรียกว่า โมรา (Moras of syllable rhyme) ดังแสดงในภาพที่ 2.10

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

สำหรับเสียงวรรณยุกต์แบบสตาติก (Static Tone)

เสียงวรรณยุกต์สามัญ : ไม่มีหน่วยเสียงย่อยสูงและต่ำ ทั้งในตำแหน่งพยางค์ที่อยู่กลางประโยคและท้ายประโยค

เสียงวรรณยุกต์ตรี : มีหน่วยเสียงย่อยสูง ในโมราที่ 2 ทั้งในตำแหน่งพยางค์ที่อยู่กลางประโยคและท้ายประโยค

เสียงวรรณยุกต์เอก : มีหน่วยเสียงย่อยต่ำ ในโมราที่ 2 ทั้งในตำแหน่งพยางค์ที่อยู่กลางประโยคและท้ายประโยค

สำหรับเสียงวรรณยุกต์แบบไดนามิก (Dynamic Tone)

เสียงวรรณยุกต์โท : ประกอบด้วยหน่วยเสียงย่อยสูงและต่ำตามลำดับ สำหรับตำแหน่งพยางค์ที่อยู่ท้ายประโยค และ มีเฉพาะหน่วยเสียงย่อยสูงในโมราที่ 1 สำหรับตำแหน่งพยางค์ที่อยู่กลางประโยค

เสียงวรรณยุกต์จัตวา : ประกอบด้วยหน่วยเสียงย่อยต่ำและสูงตามลำดับ สำหรับตำแหน่งพยางค์ที่อยู่ท้ายประโยค และ มีเฉพาะหน่วยเสียงย่อยต่ำในโมราที่ 1 สำหรับตำแหน่งพยางค์ที่อยู่กลางประโยค

จากการแปลงวิธีการแทนช่วงเวลา (Timing) พยางค์ด้วยการแทนด้วยการแบ่งหน่วยเวลาย่อยในพยางค์ (Mora-based representation) เข้ากับโครงร่างความถี่มูลฐานของพยางค์นั้นอย่างตรงไปตรงมา สามารถแทนลักษณะของโครงร่างความถี่มูลฐานได้อย่างคร่าว ๆ ด้วย ตำแหน่ง 3 ตำแหน่ง ในส่วนของหน่วยตามพยางค์ (Syllable Rhyme) ของโครงร่างความถี่มูลฐานในการทดลองการระบุเสียงวรรณยุกต์ของ Zsiga E.C. และ Nitisaroj R. ดังนี้

1. จุดเริ่ม (Beginning Point) : จุดเริ่มต้นในส่วนของหน่วยตามพยางค์
2. จุดกึ่งกลาง (Midpoint) : จุดกึ่งกลางในส่วนของหน่วยตามพยางค์
3. จุดสิ้นสุด (Endpoint) : จุดสิ้นสุดของหน่วยตามพยางค์

เมื่อเปรียบเทียบกับการแทนด้วยการแบ่งหน่วยเวลาย่อยในพยางค์เป็น 2 โมรา นั้น โมราที่ 1 คือช่วงเวลาระหว่างจุดเริ่มและจุดกึ่งกลาง ส่วนโมราที่ 2 นั้น คือช่วงเวลาระหว่างจุดกึ่งกลางและจุดสิ้นสุด ซึ่งการอธิบายโครงร่างความถี่มูลฐานของพยางค์ด้วยค่าความถี่มูลฐาน ณ ตำแหน่งจุดเริ่ม, จุดกึ่งกลาง และจุดสิ้นสุด เพียงพอในการอธิบายโครงร่างความถี่มูลฐานทั้งหมดของพยางค์ และผู้ฟังสามารถแยกแยะเสียงวรรณยุกต์ออกจากกันได้ จากการทดลองการระบุเสียงวรรณยุกต์ในพยางค์เดี่ยว และ Zsiga และNitisaroj ยังได้ทำการทดลองโดยให้ผู้ฟังระบุเสียงวรรณยุกต์ของพยางค์เดี่ยวเป้าหมายในประโยคพูดต่อเนื่องจากเสียงตัวกระตุ้น โดยที่เสียงตัวกระตุ้นนั้นถูกกำหนดด้วยค่าความถี่

มูลฐานที่จุดกึ่งกลาง และจุดสิ้นสุด เพียง 2 จุด ผู้ฟังยังคงสามารถแยกแยะและระบุเสียงวรรณยุกต์ภาษาไทยได้ แต่กลุ่มผู้ฟังยังคงมีการระบุเสียงวรรณยุกต์ ที่สับสนกันในกลุ่มผู้ฟัง ในเสียงตัวกระตุ้นที่มีความกำกวม หรือมีลักษณะของการมี 2 เสียงวรรณยุกต์ในพยางค์เดียว ซึ่งเกิดจากปรับการกำหนดค่าความถี่มูลฐานที่จุดกึ่งกลาง และจุดสิ้นสุดไปในรูปแบบต่าง ๆ

ต่อมา Ramadoss, D. [25] ได้นำผลจากการศึกษาการระบุเสียงวรรณยุกต์ของ Zsiga และ Nitisaroj ซึ่งอยู่บนพื้นฐานของทฤษฎีหน่วยเสียงย่อย ไปสร้างแบบจำลองเพื่อทำนายเสียงวรรณยุกต์ของโครงร่างความถี่มูลฐาน โดยใช้ค่าลักษณะสำคัญเดียวกัน ด้วยวิธีความน่าจะเป็นของการกระจายของค่าจุดเริ่ม จุดกึ่งกลาง และจุดสิ้นสุด ซึ่งได้สร้างแบบจำลองขึ้นมา 2 แบบจำลอง ซึ่งให้ผลความถูกต้องมากที่สุดที่ 63.6% ในการระบุเสียงวรรณยุกต์ จากนั้น Ramadoss D. ได้ทดลองการระบุเสียงวรรณยุกต์ซึ่งใช้ จุดเป้าหมาย (Target Point) กำหนดตำแหน่งในพยางค์ โดยทดสอบว่าหาก เปลี่ยนโครงร่างความถี่มูลฐาน ไว้ การรับรู้เสียงวรรณยุกต์จะเปลี่ยนไปหรือไม่ [26] และให้ผู้ฟังระบุเสียงวรรณยุกต์ที่รับรู้ และให้คะแนนความดี (goodness rating) พบว่า การปรับแต่งโครงร่างความถี่มูลฐานในแบบต่าง ๆ นั้น มีผลกระทบต่อการแยกแยะเสียงวรรณยุกต์อย่างมีนัยสำคัญในเสียงวรรณยุกต์สามัญ, ตรี และจัตวา และมีผลกระทบต่อ goodness rating อย่างมีนัยสำคัญในเสียงวรรณยุกต์สามัญ, โท, ตรี และจัตวา

เนื่องจากผลการศึกษาการระบุเสียงวรรณยุกต์ที่มีพื้นฐานบนทฤษฎีหน่วยเสียงย่อย ที่ผู้ฟังสามารถแยกแยะเสียงวรรณยุกต์ได้ด้วยค่าของจุดกึ่งกลาง และจุดปลาย ในเสียงพูดของพยางค์ในการพูดต่อเนื่อง แต่ยังเกิดความกำกวมในบางเสียงตัวกระตุ้น ที่เกิดจากการปรับแต่งโครงร่างความถี่มูลฐาน ซึ่งเกิดความไม่เป็นธรรมชาติขึ้น และการสร้างแบบจำลองเพื่อทำนายการรับรู้ของผู้ฟัง ดังนั้น ในวิทยานิพนธ์นี้ จึงทำการศึกษาการรับรู้ความเป็นธรรมชาติของโครงร่างความถี่มูลฐาน โดยอยู่บนพื้นฐานของทฤษฎีหน่วยเสียงย่อยเช่นกัน ซึ่งมีสิ่งที่น่าสนใจคือ การรับรู้และความสามารถในการแยกแยะหรือตัดสินความเป็นธรรมชาติของเสียงวรรณยุกต์ของผู้ฟัง ของเสียงที่มีรูปแบบโครงร่างความถี่มูลฐานแบบต่าง ๆ เมื่อทราบเสียงวรรณยุกต์ และสร้างแบบจำลองเพื่อทำนายการรับรู้ความเป็นธรรมชาติของผู้ฟัง เพื่อใช้ประเมินความเป็นธรรมชาติของโครงร่างความถี่มูลฐานของเสียงวรรณยุกต์แบบอัตโนมัติ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทที่ 3
ขั้นตอนการดำเนินงานวิจัย

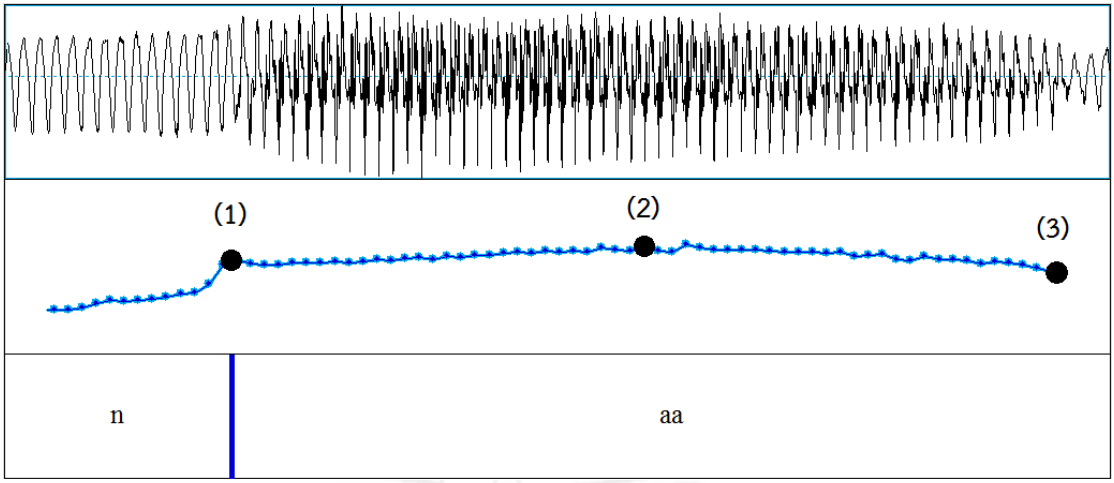
ในบทนี้จะกล่าวถึงการเลือกลักษณะสำคัญของเสียงวรรณยุกต์ภาษาไทยที่ใช้ในงานวิจัย ภาพรวมของขั้นตอนในการออกแบบการทดลองเพื่อศึกษาลักษณะของโครงร่างความถี่มูลฐานที่ส่งผลต่อการรับรู้ความเป็นธรรมชาติของเสียงวรรณยุกต์ ของพยางค์เดี่ยวภาษาไทยในการพูดต่อเนื่องของผู้ฟัง รวมไปถึงกระบวนการในการทดลองของการศึกษาการรับรู้ความเป็นธรรมชาติของเสียงวรรณยุกต์ของพยางค์เดี่ยวภาษาไทยในการพูดต่อเนื่องของผู้ฟัง และการสร้างแบบจำลองในการทำนายการรับรู้ความเป็นธรรมชาติของเสียงวรรณยุกต์ของพยางค์เดี่ยวภาษาไทยในการพูดต่อเนื่องของผู้ฟัง ของเสียงวรรณยุกต์ในภาษาไทยทั้ง 5 เสียง ซึ่งได้แก่ สามัญ, เอก, โท, ตรี และจัตวา ซึ่งขั้นตอนของการศึกษาและสร้างแบบจำลองนั้น ประกอบไปด้วยสามขั้นตอนหลัก ได้แก่ ขั้นตอนการเตรียมเสียงตัวกระตุ้น (Stimuli Preparation) ซึ่งจะกล่าวถึงการเตรียมประโยคและพยางค์เป้าหมายเพื่อใช้เป็นเสียงพูดที่ใช้ในการทดลองการรับรู้ความเป็นธรรมชาติ, ขั้นตอนการทดลองการรับรู้ (Perception Experiment) และขั้นตอนการทำนายความเป็นธรรมชาติโดยอาศัยแบบจำลองเรขาคณิต (Naturalness Prediction Using Geometric Model)

3.1 การเลือกลักษณะสำคัญของเสียงวรรณยุกต์ภาษาไทย

จากการแปลงวิธีการแทนช่วงเวลา (Timing) พยางค์ด้วยการแทนด้วยการแบ่งหน่วยเวลาย่อยในพยางค์ (Mora-based representation) เข้ากับโครงร่างความถี่มูลฐานของพยางค์นั้นอย่างตรงไปตรงมา ซึ่งแทนลักษณะของโครงร่างความถี่มูลฐานได้อย่างคร่าว ๆ ด้วย ตำแหน่ง 3 ตำแหน่งในส่วนของหน่วยตามพยางค์ (Syllable Rhyme) ของโครงร่างความถี่มูลฐาน ในการทดลองการระบุเสียงวรรณยุกต์ของ [23] ซึ่งได้แก่

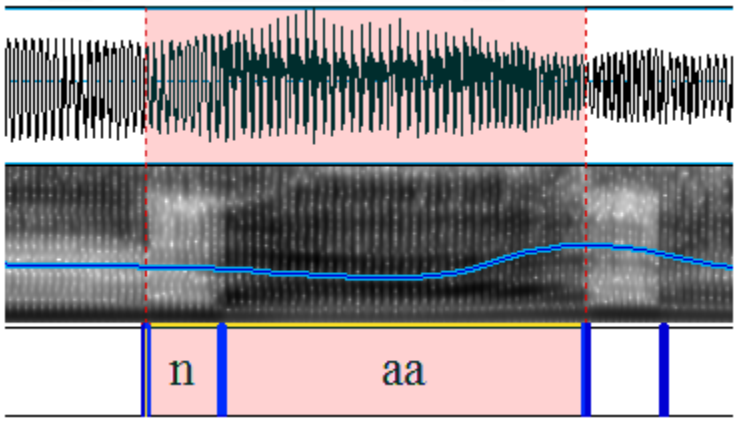
- จุดเริ่ม (Beginning Point) : จุดเริ่มต้นในส่วนของหน่วยตามพยางค์
- จุดกึ่งกลาง (Midpoint) : จุดกึ่งกลางในส่วนของหน่วยตามพยางค์
- จุดสิ้นสุด (Endpoint) : จุดสิ้นสุดของหน่วยตามพยางค์

ซึ่งแสดงตำแหน่งต่าง ๆ บนโครงร่างความถี่มูลฐานของพยางค์ได้ ดังภาพที่ 3.1



ภาพที่ 3.1 แสดงตำแหน่งของจุดเริ่ม จุดกึ่งกลาง และจุดปลาย ในพยางค์คำว่า “หน้า”

และจากการทดลองการระบุเสียงวรรณยุกต์ในพยางค์เดี่ยวของ Zsiga E.C. และNitisaroj R. ซึ่งทำการทดลองโดยให้ผู้ฟังระบุเสียงวรรณยุกต์ของพยางค์เดี่ยวเป้าหมายในประโยคพูดต่อเนื่องจากเสียงตัวกระตุ้น โดยที่เสียงตัวกระตุ้นนั้นถูกกำหนดด้วยค่าความถี่มูลฐานที่จุดกึ่งกลาง และจุดสิ้นสุดเพียง 2 จุด ผู้ฟังนั้นสามารถแยกแยะและระบุเสียงวรรณยุกต์ภาษาไทยได้ แต่กลุ่มผู้ฟังมีการระบุเสียงวรรณยุกต์ ที่สับสนกัน ในเสียงตัวกระตุ้นที่มีความกำกวม ซึ่งรับรู้ได้เป็น 2 เสียงวรรณยุกต์ในพยางค์เดียว ซึ่งเกิดจากปรับการกำหนดค่าความถี่มูลฐานที่จุดกึ่งกลาง และจุดสิ้นสุดไปในรูปแบบต่าง ๆ ดังตัวอย่างในภาพที่ 3.2 ซึ่งกลุ่มผู้ฟัง เกิดความกำกวมในการรับรู้ โดยมีจำนวนผู้ฟังที่ระบุเป็นเสียงวรรณยุกต์จัตวา และตรีในจำนวนที่ใกล้เคียงกัน ซึ่งเกิดจากการสุ่มตอบหนึ่งในสองเสียงวรรณยุกต์ที่ผู้ฟังได้ยิน



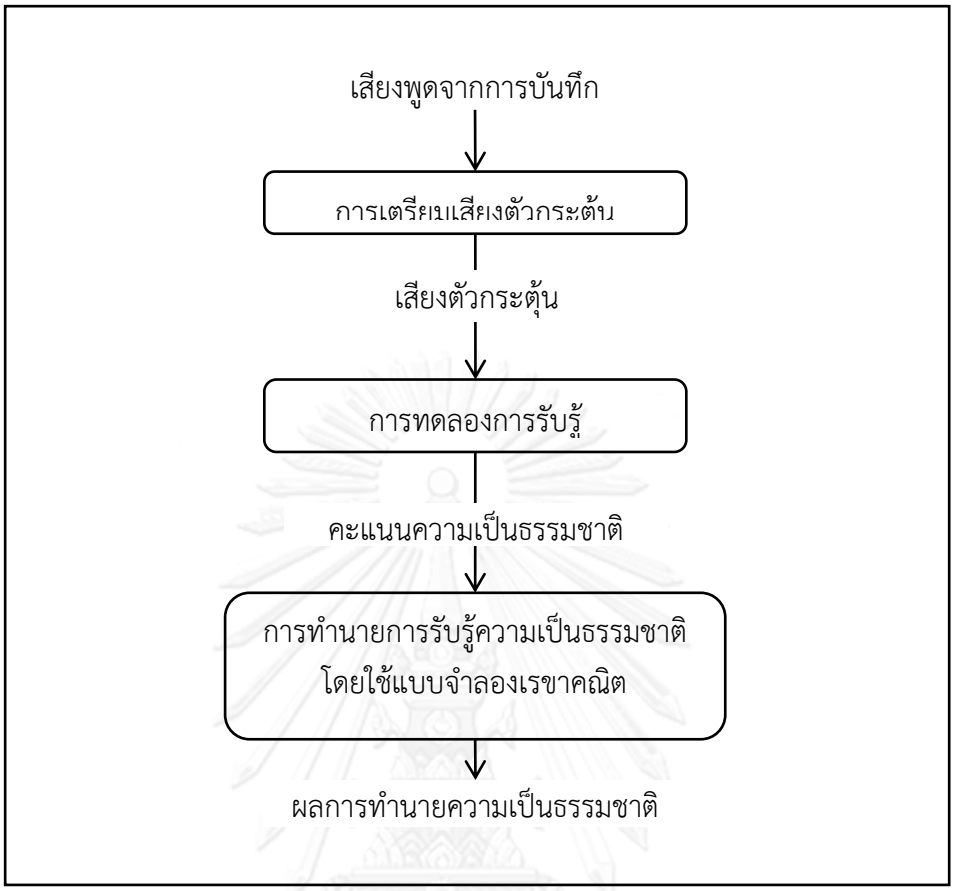
รูปที่ 3.2 แสดงโครงร่างความถี่มูลฐานของพยางค์ [น-า]
ที่กลุ่มผู้ฟังรับรู้เป็นเสียงวรรณยุกต์จัตวาและตรีในพยางค์เดียวกัน

ผู้วิจัยจึงเลือกใช้ตำแหน่งจุดกึ่งกลาง และจุดสิ้นสุด เป็นค่าลักษณะสำคัญในการในการปรับค่าโครงร่างความถี่มูลฐาน เพื่อใช้ในการทดลองการรับรู้ความเป็นธรรมชาติของเสียงวรรณยุกต์ภาษาไทย เนื่องจากทั้ง 2 ตำแหน่งในพยางค์นี้ มีแนวโน้มที่ผู้ฟังสามารถใช้เพื่อประเมินความเป็นธรรมชาติของเสียงวรรณยุกต์ภาษาไทยได้ หรือกล่าวอีกนัยหนึ่งคือ ผู้ฟังสามารถแยกแยะเสียงวรรณยุกต์ของพยางค์ที่มีความเป็นธรรมชาติ ออกจากเสียงวรรณยุกต์ของพยางค์ที่ไม่มีความเป็นธรรมชาติได้ นอกเหนือจากการใช้เพื่อระบุเสียงวรรณยุกต์เพียงอย่างเดียว

3.2 ภาพรวมของกระบวนการในการทดลอง

กระบวนการในการศึกษาการรับรู้ความเป็นธรรมชาติและการสร้างแบบจำลองการรับรู้ความเป็นธรรมชาติของเสียงวรรณยุกต์ของพยางค์เดี่ยวภาษาไทยในการพูดต่อเนื่องนั้น ประกอบไปด้วยสามขั้นตอนหลัก ขั้นตอนแรก คือ การเตรียมเสียงตัวกระตุ้น ซึ่งเป็นการนำเสียงพูดที่บันทึกไว้ มาสกัดค่าลักษณะสำคัญ และปรับแต่งเสียงพูดตามรูปแบบที่กำหนดเพื่อใช้เป็นเสียงกระตุ้นในการทดลองการรับรู้ความเป็นธรรมชาติ จากนั้น ขั้นตอนที่สอง คือ การทดลองการรับรู้ ซึ่งจะให้ผู้ฟัง ฟังเสียงตัวกระตุ้น เพื่อให้คะแนนความเป็นธรรมชาติของแต่ละเสียงกระตุ้นซึ่งมีรูปแบบของเส้นความถี่มูลฐานที่แตกต่างกัน และขั้นตอนที่สาม คือ การทำนายการรับรู้ความเป็นธรรมชาติโดยอาศัยแบบจำลองเรขาคณิต ซึ่งขั้นตอนนี้จะอาศัยความรู้ที่ได้จากการทดลองลองการรับรู้ในขั้นตอนที่สอง มาวิเคราะห์และสร้างเป็นแบบจำลองเพื่อทำนายความเป็นธรรมชาติที่สอดคล้องกับการรับรู้ของผู้ฟังของพยางค์เป้าหมาย ดังที่แสดงในภาพที่ 3.3

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

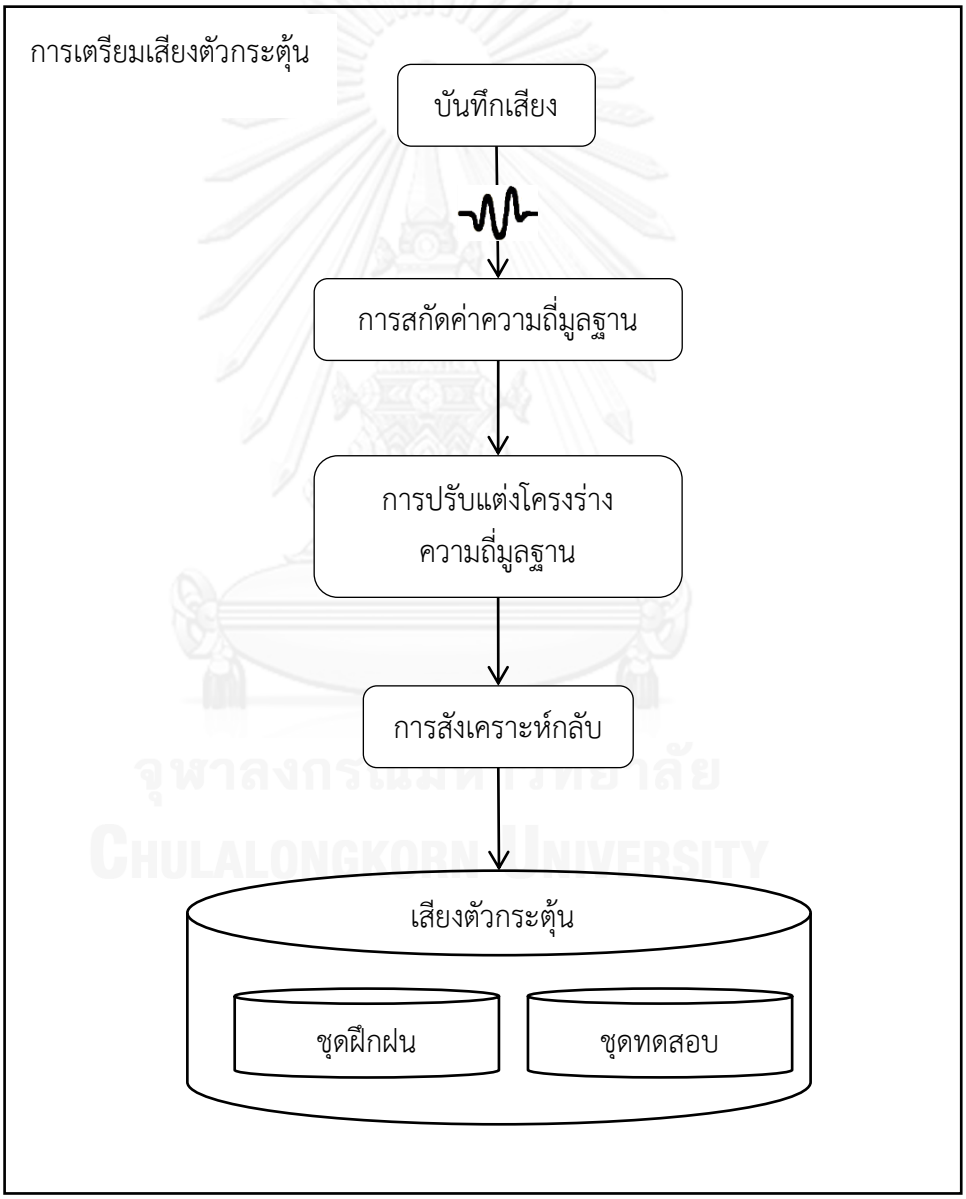


ภาพที่ 3.3 ภาพรวมของขั้นตอนการทดลอง

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

3.3 การเตรียมเสียงตัวกระตุ้น

ในขั้นตอนการเตรียมเสียงตัวกระตุ้นนี้ จะเป็นการเตรียมเสียงตัวกระตุ้นที่ใช้ในการทดลองการรับรู้ความเป็นธรรมชาติ ซึ่งมีขั้นตอนย่อยต่าง ๆ ได้แก่ การบันทึกเสียง (Recording), การสกัดค่าความถี่มูลฐาน (Fundamental Frequency Extraction), การปรับแต่งโครงร่างความถี่มูลฐาน (Fundamental Frequency Contour Configuration), และการสังเคราะห์กลับ (Resynthesis) ดังแสดงในภาพที่ 3.4



ภาพที่ 3.4 แสดงขั้นตอนย่อยในขั้นตอนการเตรียมเสียงตัวกระตุ้น

1. ขั้นตอนการบันทึกเสียง (Recording)

ในขั้นตอนนี้ จะทำการบันทึกเสียงพูด ที่ใช้ในการทดลองการรับรู้ ซึ่งจะกล่าวในหัวข้อที่ 3.3 ต่อไป โดยจะทำการบันทึกเสียงของพยางค์เป้าหมาย (Target Syllable) ซึ่งพยางค์เป้าหมายนั้น จะถูกแทรกไว้ในวลีตัวนำ (Carrier Phrase) พยางค์เป้าหมายคือ “นา” หรือเขียนในรูปแบบของโฟนีม (Phoneme) ได้เป็น “n-aa” ซึ่งพยางค์เป้าหมายนี้ จะถูกผันไปตามเสียงวรรณยุกต์ทั้ง 5 เสียงในภาษาไทย ได้แก่เสียงวรรณยุกต์ สามัญ, เอก, โท, ตรี และจัตวา โดยพยางค์เป้าหมายจะได้เป็น นา, หน่า, หน้า, น้า, หนา ตามลำดับ ทั้งนี้ พยางค์เป้าหมายจะถูกแทรกอยู่ในวลีตัวนำคือ “มี [พยางค์เป้าหมาย] ในประโยค” เพราะฉะนั้น วลีตัวนำที่ใช้ในการบันทึกเสียงพูด มีทั้งหมด 5 ประโยค ดังนี้

1. มี [นา] ในประโยค
2. มี [หน่า] ในประโยค
3. มี [หน้า] ในประโยค
4. มี [น้า] ในประโยค
5. มี [หนา] ในประโยค

การใช้พยางค์เป้าหมายคำว่า “นา” เนื่องจากเมื่อผันเสียงวรรณยุกต์ทั้ง 5 เสียงแล้วมีความหมาย ซึ่งลดการเกิดความลำเอียงของผู้ฟังในเชิงความหมายของคำลง และสามารถผันวรรณยุกต์ได้ครบทั้ง 5 เสียง โดยเสียงวรรณยุกต์เอก และจัตวาต้องใช้ ห นำหน้าเพื่อให้ผันได้ครบทุกเสียง และ สระอา เป็นสระเสียงยาวและเป็นคำเป็น จึงทำให้ผันวรรณยุกต์ได้ครบทั้ง 5 เสียงเช่นกัน (หากเป็นคำตาย จะสามารถผันวรรณยุกต์ได้เพียง 3 เสียง ได้แก่ เสียงวรรณยุกต์เอก, ตรี และโท) ในการทดลองนี้ สำหรับเสียงวรรณยุกต์โท จะใช้คำว่า “หน้า” แทนคำว่า “น่า” เพื่อไม่ให้เกิดการสับสนในขั้นตอนการบันทึกเสียงของผู้พูด โดยที่พยางค์เป้าหมายจะถูกแทรกไว้ระหว่างคำที่มีเสียงวรรณยุกต์สามัญ ได้แก่คำว่า มี และ ใน เพื่อลดผลกระทบที่เกิดจากพยางค์ก่อนหน้า (Carry-over Effect) และผลกระทบที่เกิดจากพยางค์ถัดไป (Anticipatory Effect) ที่มีต่อพยางค์เป้าหมาย ซึ่งจะทำให้โครงร่างของพยางค์เป้าหมายนั้นเปลี่ยนแปลงไปเล็กน้อย โดยขึ้นอยู่กับพยางค์ก่อนหน้า และพยางค์ถัดไป [18]

2. ขั้นตอนการสกัดโครงร่างความถี่มูลฐาน (Fundamental Frequency Contour Extraction)

ขั้นตอนนี้เป็นการสกัดเอาโครงร่างความถี่มูลฐานออกมาจากเสียงที่บันทึก ด้วยออโตคอรีเลชัน อัลกอริทึม (Autocorrelation Algorithm) เพื่อนำไปสู่การหาค่าต่าง ๆ ที่จำเป็นในการปรับแต่งโครงร่างความถี่มูลฐาน ดังนี้

- ขอบเขตความถี่มูลฐานของผู้พูด (Speaker’s F_0 Range)
    - ค่าจุดสูงสุดของขอบเขตความถี่มูลฐานของผู้พูด (Top Speaker’s F_0 Range)

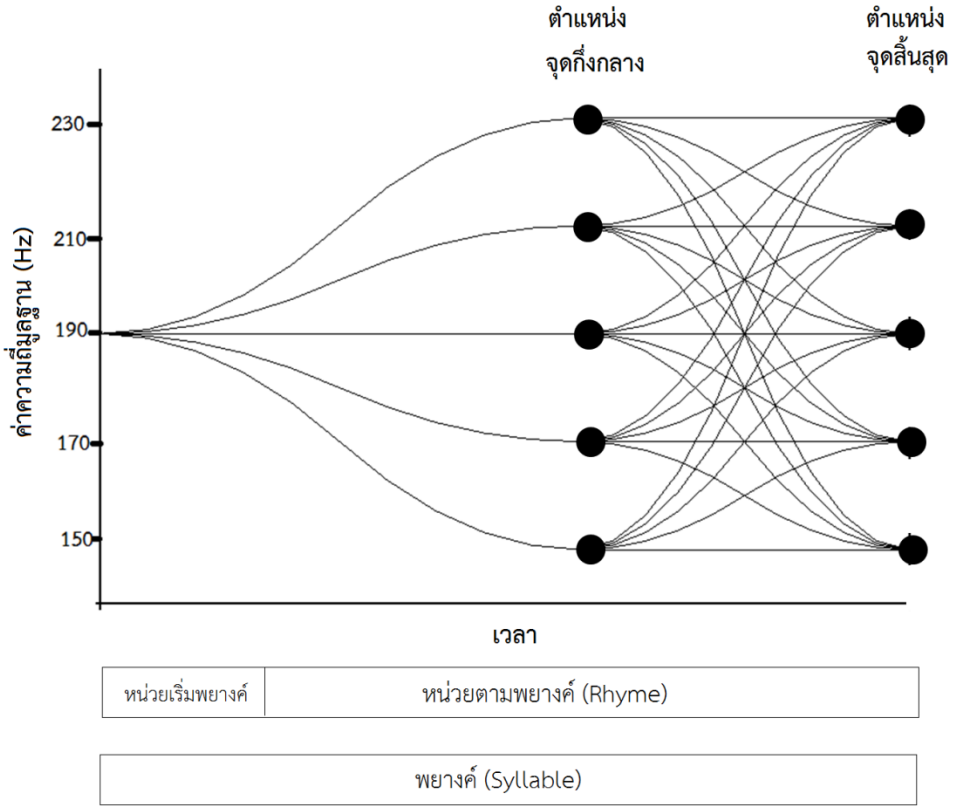
- ค่าจุดกึ่งกลางของขอบเขตความถี่มูลฐานของผู้พูด (Middle Speaker's F_0 Range)
- ค่าจุดต่ำสุดของขอบเขตความถี่มูลฐานของผู้พูด (Bottom Speaker's F_0 Range)

ขอบเขตความถี่มูลฐานของผู้พูดนั้น จะถูกนำมาใช้ในการพิจารณาในการทำการปรับแต่งโครงร่างความถี่มูลฐาน เพื่อใช้ในการเตรียมเสียงตัวกระตุ้นที่ใช้ในการทดลองการรับรู้ในงานวิจัยนี้

3. การปรับแต่งโครงร่างความถี่มูลฐาน (Fundamental Frequency Contour Configuration)

ในขั้นตอนนี้ เป็นการปรับแต่งค่าความถี่มูลฐานที่ตำแหน่งจุดกึ่งกลาง และจุดสิ้นสุดของช่วงหน่วยตามพยางค์ของพยางค์เป้าหมาย ของเสียงตัวกระตุ้นในชุดฝึกฝน จากขั้นตอนการเลือกลักษณะสำคัญของวรรณยุกต์ในหัวข้อ 3.1 โดยค่าความถี่มูลฐานจะถูกกำหนดที่ตำแหน่งจุดกึ่งกลาง และจุดสิ้นสุดของช่วงหน่วยตามพยางค์ของพยางค์เป้าหมายอย่างเป็นระบบ กระจายอยู่ทั่วขอบเขตความถี่มูลฐานของผู้พูด เช่น สมมติให้ผู้พูดมีขอบเขตความถี่มูลฐาน 80 Hz ซึ่งมีค่าจุดสูงสุดของขอบเขตความถี่มูลฐานเท่ากับ 230 Hz, ค่าจุดกึ่งกลางของขอบเขตความถี่มูลฐานเท่ากับ 190 Hz และค่าจุดต่ำสุดของขอบเขตความถี่มูลฐานเท่ากับ 150 Hz กำหนดให้ค่าความถี่มูลฐาน ณ ตำแหน่งจุดกึ่งกลางและจุดสิ้นสุด ให้มีระยะห่างขั้นละ 20 Hz โดยจะกำหนดค่าจุดกึ่งกลางได้เป็น [150, 170, 190, 210, 230] และกำหนดค่าจุดสิ้นสุดได้เป็น [150, 170, 190, 210, 230] เพราะฉะนั้น จะมีรูปแบบโครงร่างความถี่มูลฐานทั้งหมด 5 x 5 = 25 รูปแบบ (จำนวนของค่าความถี่มูลฐาน ณ ตำแหน่งจุดกึ่งกลางที่กำหนด x จำนวนของค่าความถี่มูลฐาน ณ ตำแหน่งจุดสิ้นสุดที่กำหนด) ดังตัวอย่างในภาพที่ 3.5

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University



ภาพที่ 3.5 แสดงตัวอย่างการปรับแต่งโครงร่างความถี่มูลฐาน

4. ขั้นตอนการสังเคราะห์กลับ (Resynthesis)

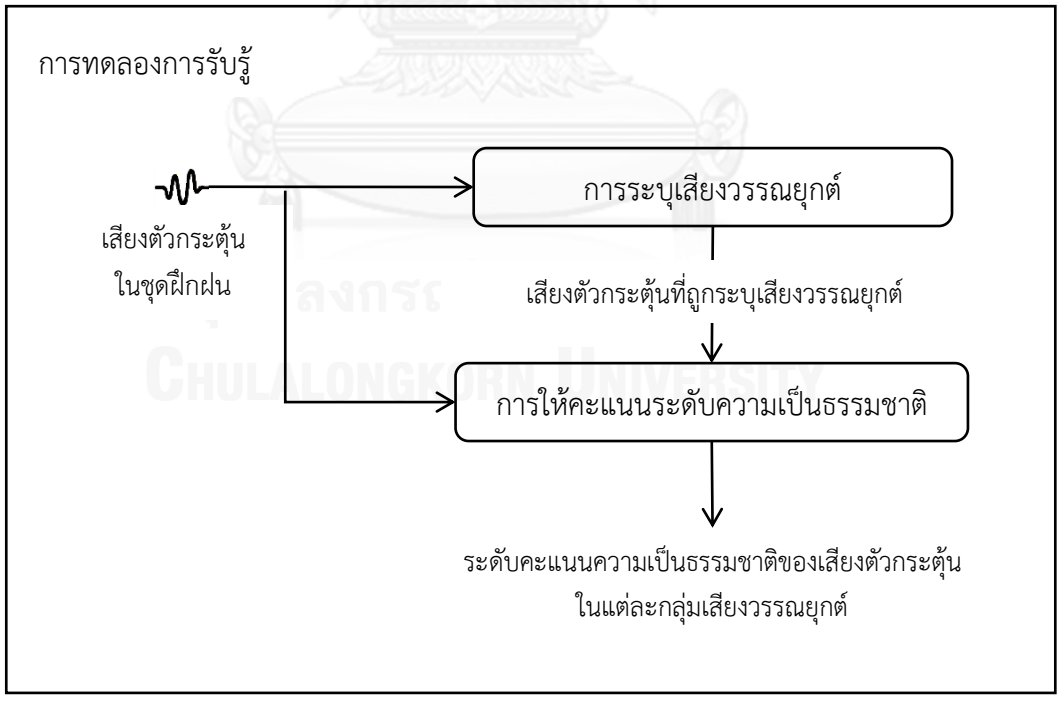
หลังจากกำหนดรูปแบบการปรับแต่งโครงร่างพื้นฐานได้เป็นรูปแบบต่าง ๆ แล้ว จะทำการประมาณค่าของค่าความถี่มูลฐาน ให้เป็นโครงร่างความถี่มูลฐาน โดยใช้เทคนิคการประมาณค่าในช่วงเชิงเส้น (Linear Interpolation) และทำการสังเคราะห์กลับ เพื่อสร้างเสียงตัวกระตุ้น โดยใช้เทคนิคการแก้ไขโครงร่างความถี่มูลฐานด้วยวิธี ไทม์โดเมนพิทช์ซิงโครนัสโอเวอร์แลปแอด (Time-domain Pitch Synchronous Overlap-add) หรือ TD-PSOLA [27]

ในการทดลองนี้ จะแบ่งข้อมูลเสียงออกเป็น 2 ชุด คือ ชุดฝึกฝน (Training Set) และ ชุดทดสอบ (Testing Set) ชุดฝึกฝนจะใช้ในการศึกษาการรับรู้ความเป็นธรรมชาติของเสียงวรรณยุกต์ซึ่งมีการปรับแต่งโครงร่างพื้นฐานดังที่ได้กล่าวไว้ข้างต้น และนำผลลัพธ์จากการศึกษาการรับรู้ความเป็นธรรมชาติ ไปสร้างแบบจำลองความเป็นธรรมชาติของเสียงวรรณยุกต์ภาษาไทย ส่วนชุดทดสอบจะใช้ในการวัดประสิทธิภาพของแบบจำลองความเป็นธรรมชาติของเสียงวรรณยุกต์ที่ถูกสร้างขึ้น ซึ่งรายละเอียดการปรับแต่งโครงร่างความถี่มูลฐานสำหรับเสียงตัวกระตุ้นในชุดทดสอบ จะกล่าวถึงในบทที่ 4 ต่อไป

3.4 การทดลองการรับรู้

กระบวนการการทดลองการรับรู้ ทำขึ้นเนื่องจากในการศึกษาความเป็นธรรมชาติของเสียงวรรณยุกต์นั้น ปัจจัยหลักในการตัดสินความเป็นธรรมชาติหรือไม่เป็นธรรมชาติของเสียงวรรณยุกต์ขึ้นอยู่กับการรับรู้ของคน ดังนั้น เพื่อสะท้อนคำตอบที่แท้จริงจากการรับรู้ของคนที่มีต่อความเป็นธรรมชาติของเสียงวรรณยุกต์ของพยางค์ กระบวนการการทดลองการรับรู้จึงถูกจัดขึ้นเพื่อศึกษาหาความสัมพันธ์ของค่าความถี่มูลฐาน ณ ตำแหน่งจุดกึ่งกลาง (Midpoint) และค่าความถี่มูลฐาน ณ ตำแหน่งจุดสิ้นสุด (Endpoint) ที่ส่งผลต่อการรับรู้ความเป็นธรรมชาติของเสียงวรรณยุกต์ในระดับต่าง ๆ โดยใช้รูปแบบของโครงร่างความถี่มูลฐานรูปแบบต่าง ๆ ที่ถูกปรับแต่งอย่างเป็นระบบ ดังที่กล่าวถึงใน 3.3. จุดมุ่งหมายของการศึกษาการรับรู้ความเป็นธรรมชาตินี้ ก็เพื่อหาคำตอบว่า แม้ว่าจากการที่ค่าความถี่มูลฐาน ณ ตำแหน่งจุดกึ่งกลางและจุดปลาย สามารถใช้แยกแยะเสียงวรรณยุกต์แต่ละเสียงได้จากการรับรู้ของคนแล้ว ในแต่ละรูปแบบโครงร่างความถี่มูลฐานที่ถูกระบุเป็นเสียงวรรณยุกต์ใด ๆ แล้วนั้น จะยังคงมีความแตกต่างกันของระดับความเป็นธรรมชาติซึ่งเกิดจากการรับรู้ และคนจะสามารถแยกแยะเสียงวรรณยุกต์ที่มีความเป็นธรรมชาติและไม่มีความเป็นธรรมชาติออกจากกันได้ด้วยค่าความถี่มูลฐาน ณ ตำแหน่งจุดกึ่งกลางและจุดปลาย

การทดลองการรับรู้ ประกอบไปด้วย 2 การทดลอง ได้แก่ การระบุเสียงวรรณยุกต์ และ การให้คะแนนความเป็นธรรมชาติ ดังแสดงในภาพที่ 3.6



ภาพที่ 3.6 แสดงขั้นตอนย่อยในขั้นตอนการทดลองการรับรู้

1. ระบุเสียงวรรณยุกต์

ในการทดลองนี้ เสียงพูดของพยางค์เป้าหมายของเสียงตัวกระตุ้นในชุดฝึกฝน จะถูกระบุเสียงวรรณยุกต์ในภาษาไทย ได้แก่เสียงวรรณยุกต์สามัญ, เอก, โท, ตรี และจัตวา โดยผู้ฟัง เพื่อแยกแยะเสียงพูดของพยางค์เป้าหมาย ของเสียงตัวกระตุ้นในชุดฝึกฝน ที่ถูกรับฟังเป็นเสียงวรรณยุกต์เดียวกันออกเป็นกลุ่ม ซึ่งจะมีทั้งหมด 5 กลุ่ม ตามหมวดที่แบ่งด้วยเสียงวรรณยุกต์ เนื่องจากจุดมุ่งหมายของการทดลองการรับรู้นั้น คือการแบ่งระดับความเป็นธรรมชาติของของพยางค์เสียงพูดด้วยการรับรู้ของคน ดังนั้นจึงมีความจำเป็นที่จะทำการทดลองเพื่อระบุเสียงวรรณยุกต์เพื่อแยกแยะกลุ่มของเสียงตัวกระตุ้น ให้อยู่ในกลุ่มเสียงวรรณยุกต์ต่าง ๆ ทั้งนี้ เพื่อให้ผู้ฟังสนใจเฉพาะการให้คะแนนระดับความเป็นธรรมชาติเพียงอย่างเดียวเท่านั้น เพื่อไม่ให้เกิดความลำเอียง และเพิ่มความถูกต้องแม่นยำ ของการให้คะแนนระดับความเป็นธรรมชาติ

2. ให้คะแนนระดับความเป็นธรรมชาติ

ในการทดลองการให้คะแนนระดับความเป็นธรรมชาตินั้น ผู้ฟังจะได้ฟังเสียงกระตุ้นเป็นกลุ่มมีทั้งหมด 5 กลุ่ม ได้แก่ กลุ่มเสียงวรรณยุกต์สามัญ, กลุ่มเสียงวรรณยุกต์เอก, กลุ่มเสียงวรรณยุกต์โท, กลุ่มเสียงวรรณยุกต์ตรี และกลุ่มเสียงวรรณยุกต์จัตวา เสียงตัวกระตุ้นในแต่ละกลุ่มวรรณยุกต์นั้น มาจากเสียงตัวกระตุ้นที่ผู้ฟังได้ระบุเสียงวรรณยุกต์ไว้ ในขั้นตอนการระบุเสียงวรรณยุกต์ และให้ผู้ฟังให้คะแนนระดับความเป็นธรรมชาติของเสียงพูดของพยางค์เป้าหมายที่ได้ยิน โดยมีการกำหนดเสียงวรรณยุกต์ให้ ซึ่งก็คือกลุ่มเสียงวรรณยุกต์ของเสียงตัวกระตุ้นนั้น ๆ ผู้ฟังจะต้องระบุว่า เสียงพูดของพยางค์เป้าหมายที่ได้ยิน มีความเป็นธรรมชาติที่สอดคล้องกับเสียงวรรณยุกต์ที่กำหนดให้ ในระดับใด ซึ่งได้มีการกำหนดบรรทัดฐานในการให้คะแนนความเป็นธรรมชาติของเสียงวรรณยุกต์ของแต่ละระดับความเป็นธรรมชาติ โดยระดับความเป็นธรรมชาติ จะถูกแบ่งออกเป็น 5 ระดับ ดังตารางที่ 3.1 ดังนี้

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ตารางที่ 3.1 บรรทัดฐานในการให้คะแนนความเป็นธรรมชาติของเสียงวรรณยุกต์

ระดับความเป็นธรรมชาติของเสียงวรรณยุกต์	คำอธิบายบรรทัดฐาน
5	มีความเป็นธรรมชาติมากที่สุด, สำเนียงถูกต้อง, เสียงวรรณยุกต์ตรงตามที่กำหนด
4	มีความไม่เป็นธรรมชาติเล็กน้อย แต่ยอมรับได้, สำเนียงแตกต่างไปจากสำเนียงที่ถูกต้องเล็กน้อย, เสียงวรรณยุกต์ตรงตามที่กำหนด
3	มีความไม่เป็นธรรมชาติ และยอมรับไม่ได้, สำเนียงแตกต่างไปจากสำเนียงที่ถูกต้องพอสมควร, เสียงวรรณยุกต์ตรงตามที่กำหนด
2	มีความไม่เป็นธรรมชาติมาก และยอมรับไม่ได้, สำเนียงแตกต่างไปจากสำเนียงที่ถูกต้องมาก, เสียงวรรณยุกต์ตรงตามที่กำหนด
1	มีความไม่เป็นธรรมชาติมาก และยอมรับไม่ได้, เสียงวรรณยุกต์ไม่ตรงตามที่กำหนด

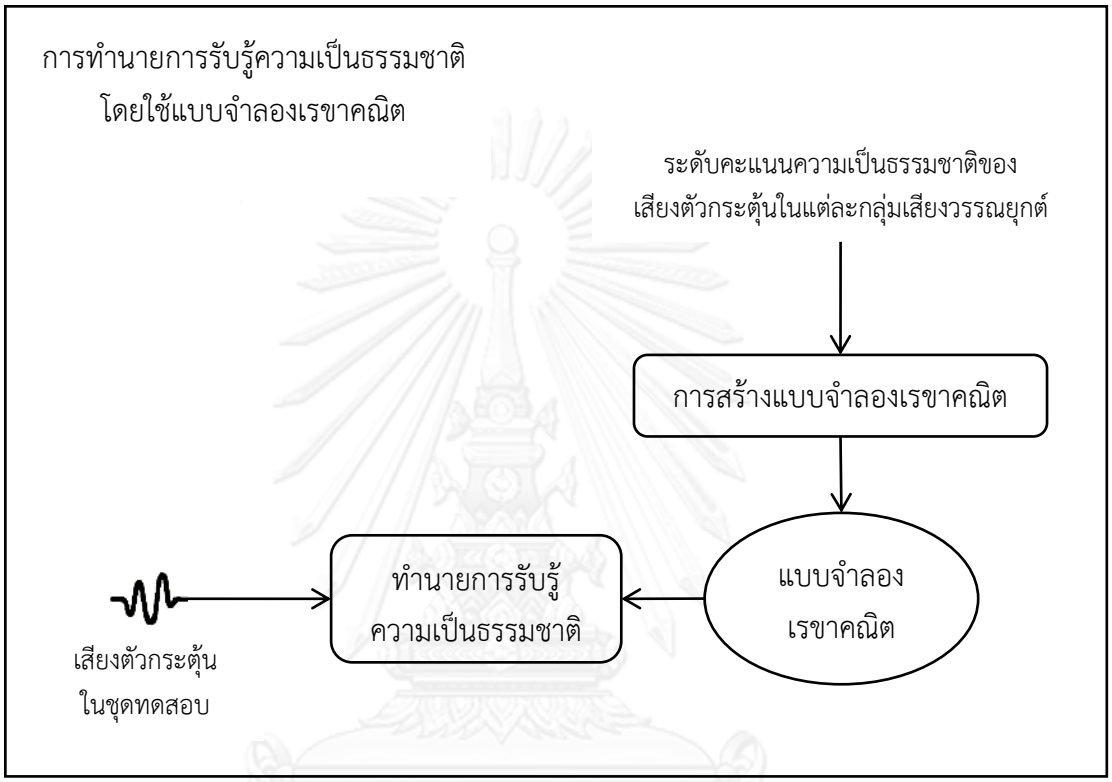
จุดประสงค์หลักของการกำหนดบรรทัดฐานในการให้คะแนนความเป็นธรรมชาติของเสียงวรรณยุกต์ของแต่ละระดับความเป็นธรรมชาตินั้น เพื่อลดขนาดของความผิดพลาดของคำตอบซึ่งเกิดจากการมีแนวโน้มในการให้คำตอบที่อยู่ตรงกลางของช่วงระดับคะแนน (Error of Central Tendency) [28, 29]

ผลลัพท์ที่ได้จากการทดลองการรับรู้ความเป็นธรรมชาติของเสียงวรรณยุกต์นี้ คือระดับคะแนนความเป็นธรรมชาติของเสียงตัวกระตุ้นในแต่ละกลุ่มเสียงวรรณยุกต์ ซึ่งจะเป็นตัวบ่งบอกว่าแต่ละรูปแบบของโครงร่างความถี่มูลฐานที่ถูกปรับแต่งโดยกำหนดค่าความถี่มูลฐาน ณ ตำแหน่งจุดกึ่งกลาง และจุดสิ้นสุดในช่วงหน่วยตามพยางค์นั้น มีความเป็นธรรมชาติของเสียงวรรณยุกต์ในระดับใด จากนั้นจึงนำผลจากการศึกษาความเป็นธรรมชาติของเสียงวรรณยุกต์ ไปสร้างแบบจำลองเพื่อทำนายความเป็นธรรมชาติของเสียงวรรณยุกต์ในพยางค์ต่อไป

3.5 การทำนายการรับรู้ความเป็นธรรมชาติโดยอาศัยแบบจำลองเรขาคณิต

ในขั้นตอนนี้ เป็นการนำผลจากการศึกษาความเป็นธรรมชาติของเสียงวรรณยุกต์ ซึ่งได้มาจากการทดลองการรับรู้ มาสร้างเป็นแบบจำลองเพื่อทำนายความเป็นธรรมชาติของเสียงวรรณยุกต์ เพื่อทำนายความเป็นธรรมชาติของเสียงวรรณยุกต์ในพยางค์ ซึ่งในขั้นตอนนี้ เมื่อได้ระดับคะแนนความเป็นธรรมชาติของเสียงตัวกระตุ้นในแต่ละกลุ่มเสียงวรรณยุกต์แล้ว จะนำข้อมูลไปสร้างแบบจำลองของการรับรู้ความเป็นความเป็นธรรมชาติโดยอาศัยแบบจำลองเรขาคณิต (Geometric Model) และประเมินผลแบบจำลองด้วยเสียงตัวกระตุ้นในชุดทดสอบ เพื่อวัดผลความถูกต้องของการ

ทำนายความเป็นธรรมชาติของเสียงวรรณยุกต์ โดยเทียบกับการรับรู้ของคนในการตัดสินความเป็นธรรมชาติของพยางค์เป้าหมายของเสียงตัวกระตุ้นในชุดทดสอบ ภาพรวมของขั้นตอนการทำนายความเป็นธรรมชาติโดยอาศัยแบบจำลองเรขาคณิต แสดงในภาพที่ 3.7



ภาพที่ 3.7 แสดงขั้นตอนย่อยในขั้นตอนการทดลองการทำนายการรับรู้ความเป็นธรรมชาติ
โดยใช้แบบจำลองเรขาคณิต

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

1. การสร้างแบบจำลองเรขาคณิต

ลักษณะสำคัญที่ใช้ในแบบจำลองเรขาคณิตนั้น ได้แก่ ค่าความถี่มูลฐานที่ถูกทำให้เป็นบรรทัดฐาน ณ ตำแหน่งจุดกึ่งกลาง และค่าความชันของโครงร่างความถี่มูลฐาน ระหว่างตำแหน่งจุดกึ่งกลางถึงตำแหน่งจุดสิ้นสุด ในช่วงของหน่วยตามพยางค์ของของพยางค์เป้าหมาย ซึ่งได้มาจากการกำหนดค่า 2 ค่านี้ คือ ค่าความถี่มูลฐาน ณ ตำแหน่งจุดกึ่งกลาง และ ค่าความถี่มูลฐาน ณ ตำแหน่งจุดสิ้นสุด ในขั้นตอนการทดลองการรับรู้ความเป็นธรรมชาติของเสียงวรรณยุกต์ ที่ทำการศึกษาผลกระทบของการเปลี่ยนแปลงโครงร่างความถี่มูลฐานที่มีต่อการรับรู้ความเป็นธรรมชาติของเสียงวรรณยุกต์ โดยการกำหนดค่าของความถี่มูลฐานของทั้ง 2 ตำแหน่งนี้ อย่างเป็นระบบ ภายในช่วงขอบเขตความถี่มูลฐานของผู้พูด และเนื่องจากผู้พูดแต่ละคน มีช่วงขอบเขตความถี่มูลฐานที่แตกต่างกัน ซึ่งขึ้นอยู่กับเพศ และอายุ ของแต่ละบุคคล ดังตัวอย่างการรายงานผลการวิจัยที่บ่งบอกว่า

ค่าเฉลี่ยของค่าความถี่มูลฐานของเพศชาย มีค่าประมาณ 120 Hz และ ค่าเฉลี่ยของค่าความถี่มูลฐานของเพศหญิง มีค่าประมาณ 210 Hz [30] ดังนั้นในงานวิจัยนี้ จึงทำการแปลงค่าความถี่มูลฐานของผู้พูดให้เป็นบรรทัดฐานเสียก่อน เพื่อปรับค่าความถี่มูลฐานของแต่ละผู้พูด ให้เป็นบรรทัดฐานเดียวกัน โดยการแปลงหน่วยความถี่ เฮิร์ตซ์ (Hertz) ให้อยู่ในรูปสเกลของอัตราส่วนในช่วงขอบเขตความถี่มูลฐาน (Percent Scale) ดังสมการที่ 3.1

$$F_0 norm = \left(\frac{F_0 - bottom F_0 Range}{F_0 Range} \right) \times 100 \qquad (3.1)$$

F_0 norm : ค่าความถี่มูลฐานที่ถูกแปลงให้เป็นบรรทัดฐาน

F_0 : ค่าความถี่มูลฐาน (Hz)

bottom F_0 Range : ค่าความถี่มูลฐานที่จุดต่ำสุดของขอบเขตความถี่มูลฐานของผู้พูด (Hz)

F_0 Range : ค่าความกว้างของขอบเขตความถี่มูลฐานของผู้พูด (Hz)

และเนื่องจากความชัน (Slope) ของโครงร่างความถี่มูลฐานของพยางค์ บ่งบอกถึงอัตราในการเปลี่ยนแปลงความเร็วในการเคลื่อนที่ของเส้นความถี่มูลฐาน ซึ่งเป็นอีกปัจจัยที่สำคัญอย่างหนึ่ง ที่มีผลโดยตรงต่อการรับรู้เสียงวรรณยุกต์ ดังนั้นจึงใช้ค่าความชันของโครงร่างความถี่มูลฐานระหว่างตำแหน่งจุดกึ่งกลางถึงตำแหน่งจุดสิ้นสุด ในช่วงหน่วยตามพยางค์ของพยางค์เป้าหมาย เป็นลักษณะสำคัญ ซึ่งใช้ในการพิจารณาร่วมกับค่าความถี่มูลฐานที่ถูกทำให้เป็นบรรทัดฐาน ณ ตำแหน่งจุดกึ่งกลาง ในการสร้างแบบจำลองการรับรู้ความเป็นธรรมชาติ โดยสามารถหาค่าความชันได้จากสมการที่ 3.2

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

$$Slope = \frac{F_{0endpoint}(\mathrm{Hz}) - \mathrm{F}_{0midpoint}(Hz)}{T_{endpoint}(s) - T_{midpoint}(s)} \qquad (3.2)$$

Slope : ค่าความชันของโครงร่างความถี่มูลฐานระหว่างตำแหน่งจุดกึ่งกลางถึงตำแหน่งจุดสิ้นสุด ของพยางค์เป้าหมาย

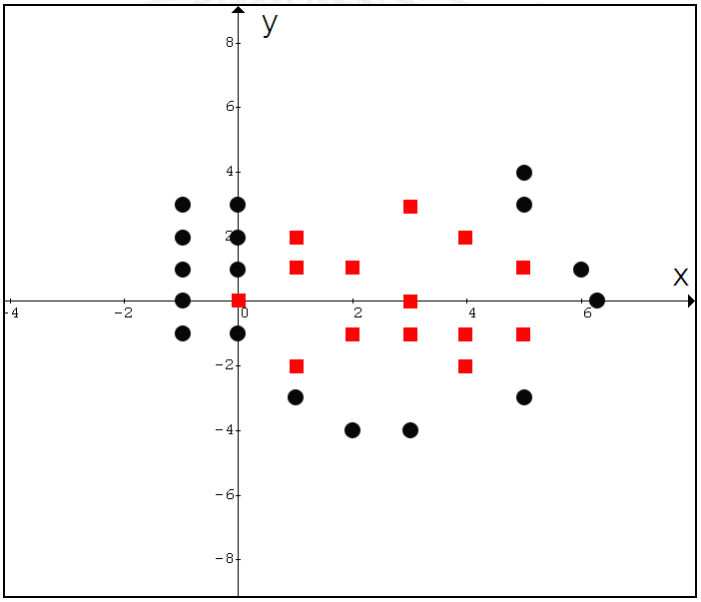
$F_{0\ endpoint}$: ค่าความถี่มูลฐาน ณ ตำแหน่งจุดสิ้นสุด ของพยางค์เป้าหมาย (Hz)

$F_{0\ midpoint}$: ค่าความถี่มูลฐาน ณ ตำแหน่งจุดกึ่งกลาง ของพยางค์เป้าหมาย (Hz)

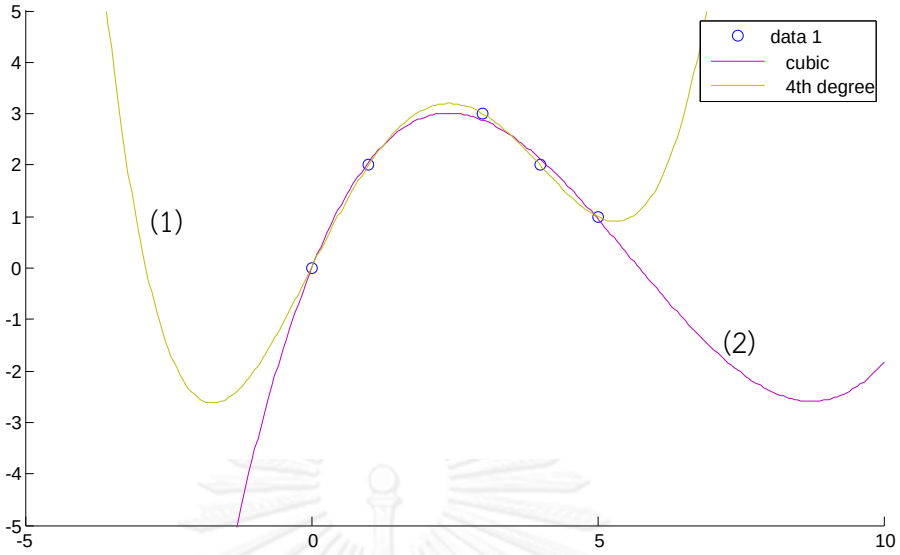
$T_{endpoint}$: เวลา ณ ตำแหน่งจุดสิ้นสุด ของพยางค์เป้าหมาย (s)

$T_{midpoint}$: เวลา ณ ตำแหน่งจุดกึ่งกลาง ของพยางค์เป้าหมาย (s)

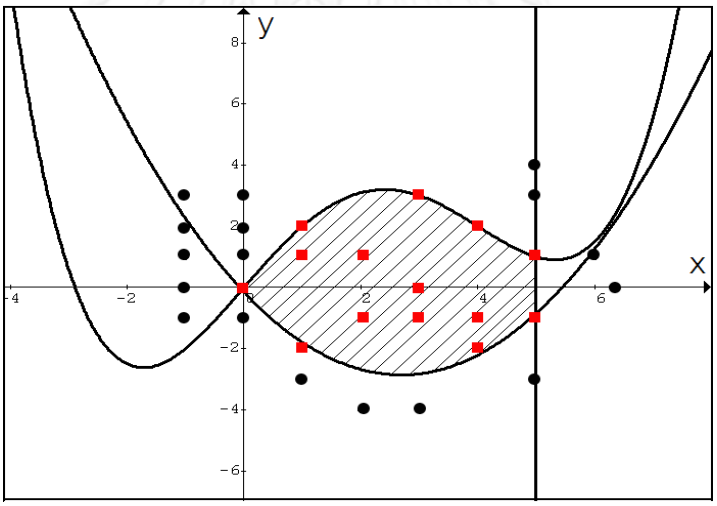
จากนั้นจะนำข้อมูลที่ได้ วาดลงบนแผนภาพกระจาย (Scatter Plot) เพื่อหาความสัมพันธ์ระหว่างค่าความถี่มูลฐานที่ตำแหน่งกึ่งกลางและค่าความชัน และใช้แบบจำลองเรขาคณิตมาใช้เพื่ออธิบายบริเวณที่มีความเป็นธรรมชาติของเสียงวรรณยุกต์ที่ได้จากผลการทดลอง โดยในการสร้างรูปเรขาคณิตบนแผนภาพกระจายนั้น จะใช้เทคนิคการประมาณค่าในช่วง ของแต่ละขอบเขตความเป็นธรรมชาติของข้อมูล ซึ่งประกอบด้วยการประมาณค่าในช่วงด้วยฟังก์ชันพหุนาม หรือการประมาณค่าในช่วงด้วยเส้นโค้งสไปลน์อันดับสาม ของจุดข้อมูล ตามความเหมาะสมของข้อมูลที่ได้จากการทดลองในการประกอบกันมาเป็นแบบจำลอง ซึ่งผลลัพธ์ของการประมาณค่าในช่วง จะอยู่ในรูปของฟังก์ชันพหุนาม แต่เนื่องจากต้องการรูปเรขาคณิตบนแผนภาพกระจาย จึงนำฟังก์ชันที่ได้ มาอยู่ในรูปอสมการ เพื่ออธิบายบริเวณของความสัมพันธ์นั้น ๆ ดังตัวอย่างข้อมูลในภาพที่ 3.8 ซึ่งกำหนดให้หาความสัมพันธ์เพื่อสร้างแบบจำลองของจุดข้อมูลสี่เหลี่ยม ที่ต้องการแยกออกจากข้อมูลจุดวงกลม บนความสัมพันธ์ x กับ y ขั้นตอนคือ หารูปเรขาคณิตเพื่อกำหนดขอบเขตและบริเวณในกลุ่มข้อมูลของจุดสี่เหลี่ยม ในขั้นแรก ทำการเลือกจุดสี่เหลี่ยมที่อยู่บริเวณขอบของกลุ่มข้อมูล จากนั้นทำการปรับเส้นโค้งให้เหมาะสมกับข้อมูล (Curve Fitting) เพื่อหาสมการความสัมพันธ์ ดังภาพที่ 3.9 ซึ่งแสดงการปรับเส้นโค้ง 2 แบบคือ (1) การปรับเส้นโค้งด้วยวิธีการประมาณค่าในช่วงด้วยสมการพหุนามอันดับสี่ (เนื่องจากมีจุดข้อมูล 5 จุด) (2) การปรับเส้นโค้งด้วยวิธีการประมาณค่าในช่วงด้วยเส้นโค้งแบบสไปลน์อันดับสาม ซึ่งจากรูป แสดงให้เห็นว่าการปรับเส้นโค้งแบบที่ (2) เหมาะสมกับข้อมูลมากกว่าแบบที่ (1) แต่ให้ความถูกต้อง ก็ต่อเมื่อค่า $x \leq 5$ และ $x \geq 0$ ทั้งนี้สามารถกำหนดโดเมนได้เนื่องจากบริเวณจุดสี่เหลี่ยมในภาพที่ 3.8 อยู่ในช่วง $0 \leq x \leq 5$ อยู่แล้ว ในตัวอย่างนี้ จึงเลือกวิธีการที่ (1) ในการปรับเส้นโค้ง เส้นโค้งด้านล่างของรูปร่างเรขาคณิตที่ต้องการ ทำได้เช่นเดียวกัน ซึ่งสามารถสร้างแบบจำลองเรขาคณิตกับข้อมูลตัวอย่างจากภาพที่ 3.8 ได้ดังภาพที่ 3.10



ภาพที่ 3.8 ตัวอย่างข้อมูลในแผนภาพกระจาย



ภาพที่ 3.9 ตัวอย่างการประมาณค่าในช่วงของข้อมูลตัวอย่าง



ภาพที่ 3.10 ตัวอย่างการอธิบายความสัมพันธ์ด้วยแบบจำลองเรขาคณิต

จากภาพที่ 3.10 บริเวณที่แรเงา แสดงแบบจำลองเรขาคณิตที่ธิบายความสัมพันธ์ของจุดข้อมูลสี่เหลี่ยม จากกลุ่มข้อมูลตัวอย่าง ซึ่งได้จากพื้นที่ปิดภายในเส้นโค้งที่ได้จากการปรับเส้นโค้งจากจุดข้อมูลบริเวณขอบของขอบเขตกลุ่มข้อมูลจุดสี่เหลี่ยม ซึ่งอธิบายได้ด้วยแบบจำลองดังนี้

$$y \geq -0.05x^3 + 0.75x^2 - 2.7x - 3.8892e-15 \quad (3.3)$$

$$y \leq 0.025x^4 - 0.2x^3 - 0.025x^2 + 2.2x - 2.1775e-15 \quad (3.4)$$

$$0 \leq x \leq 5 \quad (3.5)$$

แบบจำลองเรขาคณิตสำหรับอธิบายรูปร่างนี้คือ $(3.3) \wedge (3.4) \wedge (3.5)$

3.6 เครื่องมือที่ใช้ในงานวิจัย

สำหรับงานวิจัยนี้ ได้ใช้เครื่องมือต่าง ๆ ในการดำเนินงานวิจัย ดังนี้

1. เครื่องมือสำหรับบันทึกเสียงพูด ซึ่งในงานวิจัยนี้จะใช้ โปรแกรม ออดาซิตี้ (Audacity) เวอร์ชัน 2.0 ซึ่งใช้สำหรับบันทึกเสียง และตัดต่อไฟล์เสียง
2. เครื่องมือที่ใช้สำหรับการถอดเสียงของเสียงที่บันทึก ซึ่งในงานวิจัยนี้ใช้โปรแกรม เวฟเซอเฟอร์ (Wavesurfer)
3. เครื่องมือในการวิเคราะห์สัญญาณเสียง, สัทสัมพันธ์ และปรับแต่งโครงร่างความถี่มูลฐานของเสียงพูด ซึ่งในงานวิจัยนี้ใช้ โปรแกรมพราท (Praat) เวอร์ชัน 5.3.48 [31]
4. โปรแกรมที่ใช้ในการคำนวนสมการของแบบจำลองเรขาคณิต ซึ่งในงานวิจัยนี้ใช้โปรแกรมแมทแลป (Matlab) เวอร์ชัน 7.13.0.564
5. โปแกรมที่ใช้ในการแสดงแบบจำลองเรขาคณิต ซึ่งในงานวิจัยนี้ใช้โปรแกรมกราฟเมติกา (Graphmatica) เวอร์ชัน 2.2e
6. โปรแกรม Microsoft Excel 2010
7. เครื่องมือพัฒนาโปรแกรมภาษา C#, Microsoft Visual Studio 2010

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

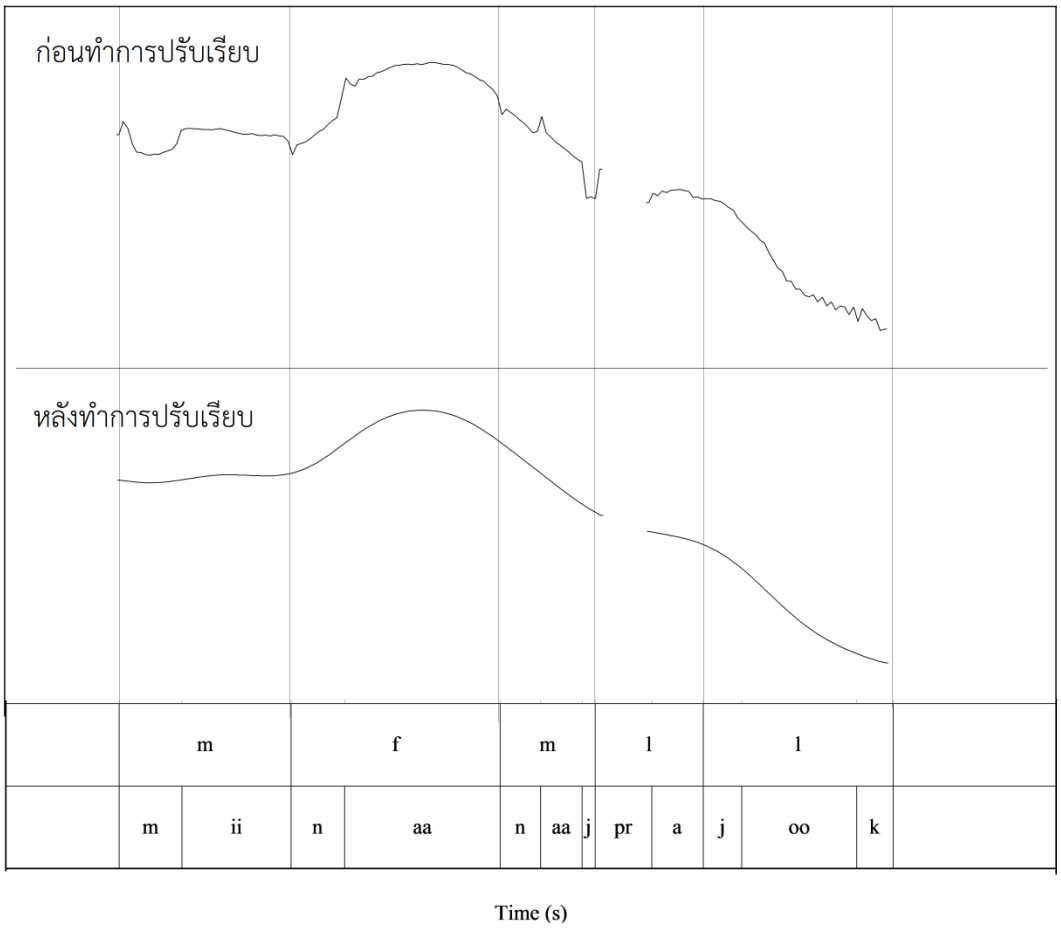
บทที่ 4
การทดลอง ผลการทดลองและการวิเคราะห์ผลการทดลอง

ในบทนี้จะกล่าวถึงการเตรียมข้อมูลเสียงสำหรับทำการทดลองตามขั้นตอนการเตรียมเสียงกระตุ้น การทดลองการรับรู้ และการสร้างแบบจำลองการทำนายการรับรู้ความเป็นธรรมชาติของเสียงวรรณยุกต์ที่นำเสนอ รวมไปถึงการวัดผลของแบบจำลองที่นำเสนอ โดยจะทำการทดสอบผลของการทำนายการรับรู้ความเป็นธรรมชาติของแบบจำลอง เทียบกับการรับรู้ของกลุ่มผู้ฟัง ด้วยเสียงตัวกระตุ้นในชุดทดสอบ ซึ่งในบทนี้จะประกอบไปด้วยเนื้อหาดังต่อไปนี้ 1. ข้อมูลเสียงพูดที่ใช้ในการทดลอง 2. การทดลองการรับรู้ 3. การสร้างแบบจำลองเรขาคณิตจากผลการทดลองการรับรู้ 4. การวัดผลการทำนายความเป็นธรรมชาติของเสียงวรรณยุกต์ด้วยแบบจำลองเรขาคณิต 5. การวิเคราะห์ผลารทดลอง

4.1 ข้อมูลเสียงพูดที่ใช้ในการทดลอง

เสียงพูดที่ใช้ในการทดลองนี้ ได้จากการบันทึกเสียง โดยมีผู้พูดที่เป็นคนไทยเจ้าของภาษาทั้งหมด 4 คน แบ่งเป็นเพศชาย 2 คน และเพศหญิง 2 คน ทั้งนี้ ข้อมูลเสียงพูดจะแบ่งออกเป็น 2 ชุด ได้แก่ ชุดฝึกฝน (Training Set) และ ชุดทดสอบ (Testing Set) ซึ่งในแต่ละชุด ประกอบไปด้วยผู้พูดที่เป็นเพศชาย และ เพศหญิง อย่างละหนึ่งคน ในชุดฝึกฝนประกอบด้วยผู้พูดที่ 1 เพศหญิง และผู้พูดที่ 2 เพศชาย ส่วนชุดทดสอบประกอบด้วยผู้พูดที่ 3 เพศชาย และผู้พูดที่ 4 เพศหญิง มีอายุ 27, 23, 22 และ 24 ปี ตามลำดับ เสียงพูดจะถูกบันทึกในห้องบันทึกเสียง ด้วยคอนเดนเซอร์ ไมโครโฟน (Condenser Microphone) ยี่ห้อ แซมสัน รุ่น โกมิค (Samson Go Mic) และบันทึกในรูปแบบไฟล์ .wav ด้วยอัตราการชักตัวอย่างที่ 44.1 kHz ผู้พูดแต่ละคน จะพูดวลีตัวนำทั้ง 5 วลี ดังที่กล่าวถึงใน 3.3.1 เป็นจำนวน 3 ครั้ง ดังนั้น แต่ผู้พูด จะบันทึกวลีตัวนำทั้งสิ้น 15 วลี ผู้พูดแต่ละคนจะได้รับคำสั่งให้ทำการพูดบันทึกเสียงทีละวลี โดยวลีจะอยู่ในบัตรคำซึ่งเขียนเป็นภาษาไทยซึ่งเรียงลำดับแบบสุ่ม และพูดด้วยความเร็วในการพูดแบบการสนทนา

เมื่อทำการบันทึกเสียงวลีตัวนำแล้ว จะนำเสียงพูดมาสกัดโครงร่างความถี่มูลฐาน โดยใช้ออโตคอรีเลชัน อัลกอริทึม ด้วยโปรแกรมพรัท (Praat) [31] เมื่อได้โครงร่างความถี่มูลฐานแล้ว จะทำการปรับโครงร่างความถี่มูลฐานให้เรียบ โดยใช้พารามิเตอร์ความกว้างแถบความถี่ 5 Hz ด้วยโปรแกรมพรัท เพื่อลดการกระโดดของโครงร่างความถี่มูลฐานของเสียง ดังตัวอย่างภาพโครงร่างความถี่มูลฐานก่อนและหลังการทำให้เรียบ แสดงในภาพที่ 4.1 ซึ่งแสดงอย่างโครงร่างความถี่มูลฐานของเสียงพูดของผู้พูดที่หนึ่ง ในวลี “มีหน้าในประโยค” ก่อนและหลังการปรับเรียบ



ภาพที่ 4.1 โครงร่างความถี่มูลฐานของเสียงพูดก่อนและหลังการปรับเรียบ

หลังจากที่ได้โครงร่างความถี่มูลฐานของเสียงพูดแล้ว จึงทำการหาขอบเขตความถี่มูลฐานของผู้พูด ซึ่งประกอบไปด้วย ค่าจุดสูงสุดของขอบเขตความถี่มูลฐานของผู้พูด, ค่าจุดกึ่งกลางสุดของขอบเขตความถี่มูลฐานของผู้พูด และค่าจุดต่ำสุดของขอบเขตความถี่มูลฐานของผู้พูด แต่ละค่าสามารถหาได้ดังนี้ ค่าจุดสูงสุดของขอบเขตความถี่มูลฐานของผู้พูด หาได้จากค่าเฉลี่ยของจุดสูงสุดจากเสียงวรรณยุกต์โทของพยางค์เป้าหมาย ค่าจุดกึ่งกลางของขอบเขตความถี่มูลฐานของผู้พูด หาได้จากค่าเฉลี่ยของโครงร่างความถี่มูลฐานในช่วงหน่วยตามพยางค์จากเสียงวรรณยุกต์สามัญของพยางค์เป้าหมาย และค่าจุดต่ำสุดของขอบเขตความถี่มูลฐานของผู้พูด หาได้จากค่าเฉลี่ยของจุดต่ำสุดจากเสียงวรรณยุกต์จัตวาของพยางค์เป้าหมาย ซึ่งขอบเขตความถี่มูลฐานของผู้พูดทั้งสี่คน แสดงในตารางที่ 4.1

ตารางที่ 4.1 แสดงขอบเขตความถี่มูลฐานของผู้พูด

ผู้พูด	ค่าจุดสูงสุดของขอบเขตความถี่มูลฐาน (Hz)	ค่าจุดกึ่งกลางของขอบเขตความถี่มูลฐาน (Hz)	ค่าจุดต่ำสุดของขอบเขตความถี่มูลฐาน (Hz)
ผู้พูดที่ 1 (หญิง)	290	230	170
ผู้พูดที่ 2 (ชาย)	160	130	100
ผู้พูดที่ 3 (ชาย)	155	125	95
ผู้พูดที่ 4 (หญิง)	290	240	190

จากขอบเขตความถี่มูลฐานของแต่ละผู้พูด พบว่าผู้พูดที่ 1, ผู้พูดที่ 2, ผู้พูดที่ 3 และผู้พูดที่ 4 มีความกว้างของขอบเขตความถี่มูลฐานเป็น 120 Hz, 60Hz, 60Hz และ 100 Hz ตามลำดับ ซึ่งสังเกตได้ว่า ความกว้างของขอบเขตความถี่มูลฐานของผู้พูดเพศหญิง มากกว่าของผู้พูดเพศชาย ค่าความกว้างของขอบเขตความถี่มูลฐานนั้น จะถูกนำมาพิจารณาในการกำหนดค่าระยะห่างของแต่ละขั้นความที่มูลฐานที่ใช้ปรับแต่งโครงร่างความถี่มูลฐานเพื่อใช้เป็นเสียงตัวกระตุ้นในชุดฝึกฝน ทั้งนี้เสียงตัวกระตุ้นถูกแบ่งออกเป็น 2 ชุด ได้แก่ชุดฝึกฝน และชุดทดสอบ ซึ่งมีรายละเอียดในการปรับแต่งโครงร่างความถี่มูลฐานดังนี้ดังนี้

4.1.1 เสียงตัวกระตุ้นในชุดฝึกฝน

สำหรับเสียงตัวกระตุ้นในชุดฝึกฝน จะใช้ในการศึกษาการรับรู้ความเป็นธรรมชาติของเสียงวรรณยุกต์ ซึ่งจะใช้เสียงพูดจากผู้พูดที่ 1 (หญิง) และผู้พูดที่ 2 (ชาย) โดยการเลือกวลีที่บันทึกที่มีคุณภาพดีที่สุด ตามความเห็นของผู้ทำวิจัยมาหนึ่งเสียงของแต่ละผู้พูด จากวลีที่มีพยางค์เป้าหมายเป็นเสียงวรรณยุกต์สามัญ (นา) จากนั้น จะนำเสียงพูดที่เลือกมานั้น มาปรับแต่งโครงร่างความถี่มูลฐานที่ตำแหน่งจุดกึ่งกลาง และตำแหน่งจุดสิ้นสุดของช่วงหน่วยตามพยางค์ ของพยางค์เป้าหมาย ซึ่งอยู่ในในพยางค์ที่สองของวลี ภายในขอบเขตความถี่มูลฐานของแต่ละผู้พูด สำหรับเสียงจากผู้พูดที่ 1 นั้น มีความกว้างของขอบเขตความถี่มูลฐาน 120 Hz ฉะนั้นจึงกำหนดค่าระยะห่างของแต่ละขั้นของการปรับค่าความถี่มูลฐาน ขั้นละ 10 Hz เพราะฉะนั้นค่าความถี่มูลฐานจะถูกกำหนดที่ตำแหน่งจุดกึ่งกลาง และตำแหน่งจุดสิ้นสุดของช่วงหน่วยตามพยางค์ ซึ่งมีทั้งสิ้น 13 ขั้น ได้แก่ 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280 และ 290 Hz เพราะฉะนั้น เสียงตัวกระตุ้นของผู้พูดที่ 1 ใช้ชุดฝึกฝนมีทั้งสิ้น 13 x 13 = 169 เสียงตัวกระตุ้น (จำนวนขั้นของจุดกึ่งกลาง x จำนวนขั้นของจุดสิ้นสุด) สำหรับเสียงจากผู้พูดที่ 2 นั้น มีความกว้างของขอบเขตความถี่มูลฐาน 60 Hz ซึ่งแคบกว่าของผู้พูดที่ 1 ดังนั้นเพื่อให้มีจำนวนขั้นในการปรับแต่งที่เท่ากัน จึงกำหนดค่าระยะห่างของแต่ละขั้นของการปรับค่าความถี่มูลฐาน ขั้นละ 5 Hz เพราะฉะนั้นค่าความถี่มูลฐานจะถูกกำหนดที่ตำแหน่งจุดกึ่งกลาง และตำแหน่งจุดสิ้นสุดของช่วงหน่วยตามพยางค์ จะมีทั้งสิ้น 13 ขั้น ได้แก่ 100, 105,

110, 115, 120, 125, 130, 135, 140, 145, 150, 155, และ 160 Hz เพราะฉะนั้น เสียงตัวกระตุ้นของผู้พูดที่ 2 ใช้ชุดฝึกฝนจะมีทั้งสิ้น 13 x 13 = 169 เสียงตัวกระตุ้น (จำนวนขั้นของจุดกึ่งกลาง x จำนวนขั้นของจุดสิ้นสุด) สำหรับพยางค์แรกของวลี คำว่า “มี” ถูกปรับแต่งให้โครงร่างความถี่มูลฐานค่อย ๆ ลดลงจนมีค่าเท่ากับค่ากึ่งกลางของขอบเขต (230 Hz สำหรับผู้พูดที่ 1 และ130 Hz สำหรับผู้พูดที่ 2) ส่วนพยางค์ที่ 3 คำว่า “ใน” นั้น จะถูกกำหนดค่าความถี่มูลฐานทั้งตำแหน่งจุดกึ่งกลางและจุดสิ้นสุดให้เท่ากับค่ากึ่งกลางของขอบเขตเช่นกัน และช่วงต้นพยางค์เกิดจากการประมาณค่าในช่วงระหว่างจุดสิ้นสุดของพยางค์เป้าหมายกับพยางค์ที่ 3 เพื่อให้เกิดการค่อยๆ ปรับโครงร่างความถี่มูลฐานให้มาสู่เสียงวรรณยุกต์สามัญองพยางค์ที่ 3 ส่วนพยางค์ที่ 4 และ 5 คำว่า “ประโยค” นั้น ไม่ถูกปรับแต่งใด ๆ

4.1.2 เสียงตัวกระตุ้นในชุดทดสอบ

เสียงตัวกระตุ้นในชุดทดสอบนั้น จะใช้สำหรับประเมินความถูกต้องของแบบจำลองการทำนายการรับรู้ความเป็นธรรมชาติที่สร้างขึ้น ซึ่งจะใช้เสียงพูดจากผู้พูดที่ 3 (ชาย) และผู้พูดที่ 4 (หญิง) สำหรับการปรับแต่งรูปแบบของโครงร่างพื้นฐานนั้น ลักษณะของโครงร่างความถี่มูลฐานของพยางค์เป้าหมาย จะแบ่งตามประเภทของรูปร่างเส้นใดๆ ที่เป็นไปได้ โดยมีข้อจำกัดคือ 1. มีจุดสะท้อน (Reflection Point) ไม่เกินหนึ่งจุด 2. ค่าความถี่มูลฐานมีค่าอยู่ภายในขอบเขตความถี่มูลฐานของแต่ละผู้พูด โดยแบ่งประเภทของรูปร่างของโครงร่างที่ใช้ทดสอบ เป็น 11 ประเภท และมีทั้งหมด 45 รูปแบบ ดังนี้

ก. ชันลงมากในครึ่งแรกของพยางค์ : 4 รูปแบบ

ข. ชันขึ้นมากในครึ่งแรกของพยางค์ : 4 รูปแบบ

ค. ชันลงมากในครึ่งหลังของพยางค์ : 2 รูปแบบ

ง. ชันขึ้นมากในครึ่งหลังของพยางค์ : 2 รูปแบบ

จ. รูปร่างแบบระฆังคว่ำ : 5 รูปแบบ

ฉ. รูปร่างแบบตัวยู : 5 รูปแบบ

ช. เส้นตรง : 3 รูปแบบ

ซ. เส้นตรงในครึ่งแรก / ชันขึ้นอย่างช้า ๆ ในครึ่งหลังของพยางค์ : 5 รูปแบบ

ฌ. เส้นตรงในครึ่งแรก / ชันลงอย่างช้า ๆ ในครึ่งหลังของพยางค์ : 5 รูปแบบ

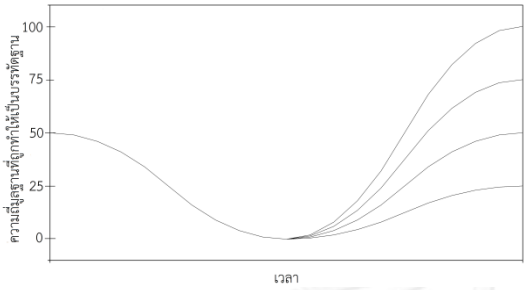
ญ. ชันลงอย่างช้า ๆ ในครึ่งแรก / เส้นตรงในครึ่งหลังของพยางค์ : 5 รูปแบบ

ฎ. ชันขึ้นอย่างช้า ๆ ในครึ่งแรก / เส้นตรงในครึ่งหลังของพยางค์ : 5 รูปแบบ

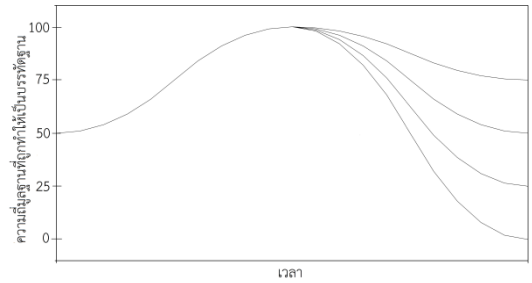
จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

รูปร่างของโครงร่างความถี่มูลฐานในแต่ละประเภท ที่ใช้ปรับแต่งเสียงพูดของพยางค์เป้าหมายของเสียงตัวกระตุ้นในชุดทดสอบ แสดงในภาพที่ 4.2 ซึ่งแต่ละผู้พูด จะมีทั้งหมด 45 รูปแบบ ดังนั้นจะมีเสียงตัวกระตุ้นในชุดทดสอบทั้งสิ้น 90 เสียงตัวกระตุ้น (2 ผู้พูด)

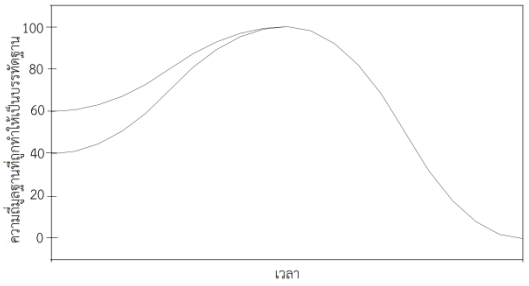
(ก)



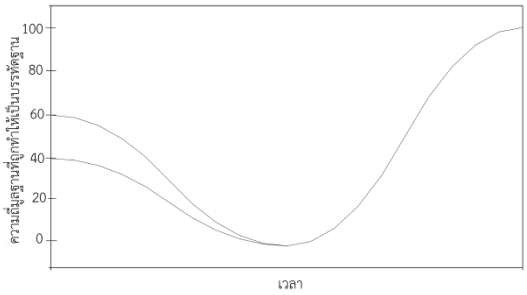
(ข)



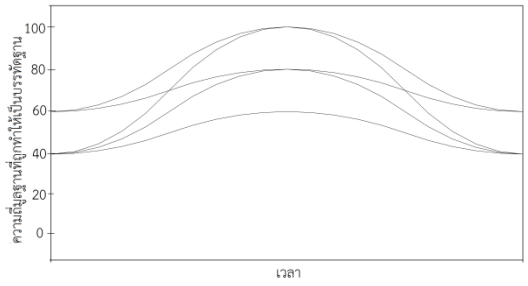
(ค)



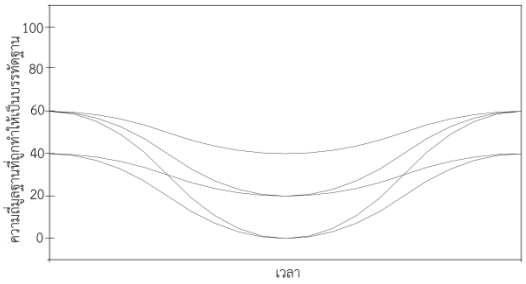
(ง)



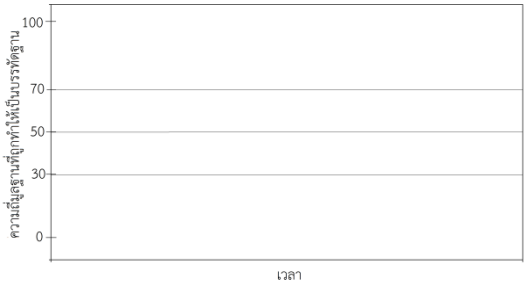
(จ)



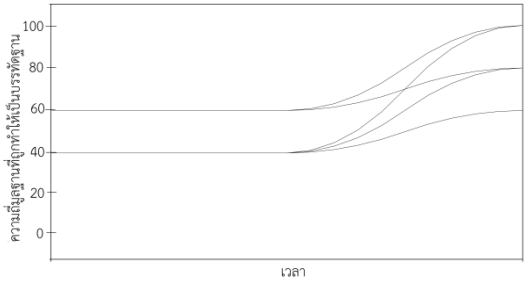
(ฉ)



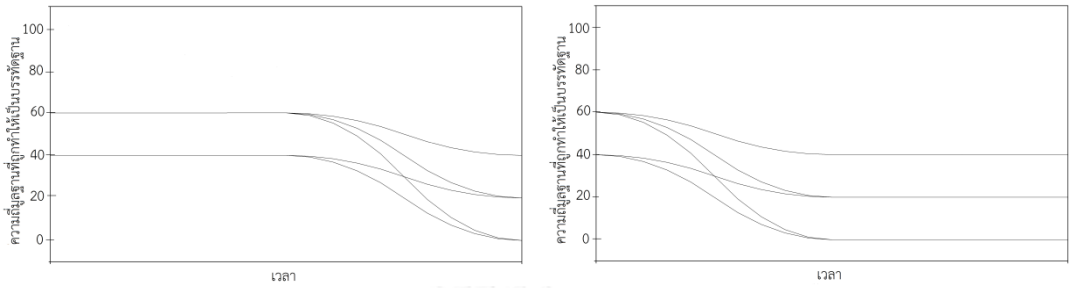
(ช)



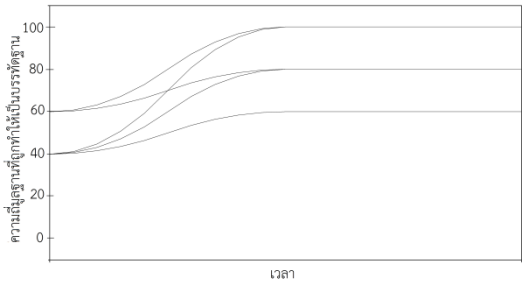
(ซ)



(ฌ) (ญ)



(ฎ)



ภาพที่ 4.2 รูปแบบโครงร่างความถี่มูลฐานของเสียงตัวกระตุ้นในชุดทดสอบ

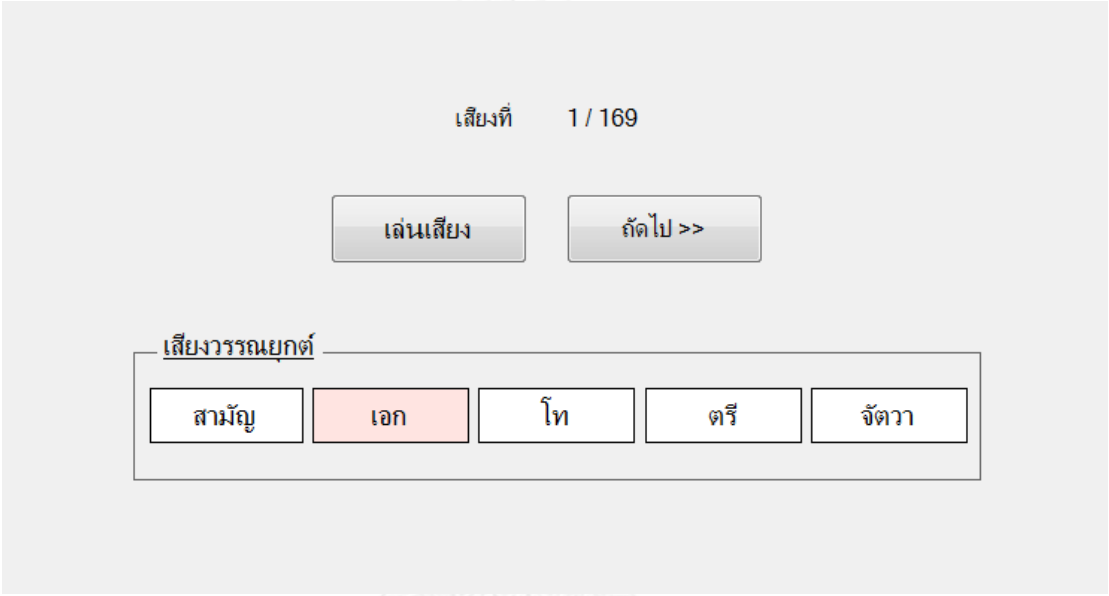
4.2 การทดลองการรับรู้

ในการทดลองการรับรู้นั้นจะแบ่ง 2 การทดลอง ได้แก่ การระบุเสียงวรรณยุกต์ เพื่อแยกแยะเสียงตัวกระตุ้นที่มีเสียงวรรณยุกต์เดียวกัน ซึ่งตัดสินจากการรับรู้ด้วยคนก่อน จากนั้นจึงทำการให้คะแนนความเป็นธรรมชาติ ของเสียงตัวกระตุ้นที่ซึ่งสอดคล้องกับเสียงวรรณยุกต์นั้น ๆ ดังนี้

4.2.1 การระบุเสียงวรรณยุกต์และผลการทดลอง

กลุ่มผู้ฟังที่เข้าร่วมการทดลองการระบุเสียงวรรณยุกต์นั้น มีทั้งหมด 20 คน เป็นเพศชาย 10 คน และเพศหญิง 10 คน มีอายุระหว่าง 22 – 30 ปี มีค่าเฉลี่ยของอายุเท่ากับ 23.9 ปี ซึ่งกลุ่มผู้ฟัง มีความเข้าใจในระบบเสียงวรรณยุกต์ของภาษาไทยเป็นอย่างดี การทดลองนั้นแบ่งออกเป็น 2 รอบ ได้แก่ รอบที่ 1 เป็นการทดลองโดยใช้เสียงตัวกระตุ้นจากชุดฝึกฝนของผู้พูดที่ 1 (เพศหญิง) และรอบที่ 2 เป็นการทดลองโดยใช้เสียงตัวกระตุ้นจากชุดฝึกฝนของผู้พูดที่ 2 (เพศชาย) แต่ละรอบของการทดลอง ถูกจัดขึ้นในระยะเวลาที่ห่างกัน 1 สัปดาห์ โดยแต่ละชุดจะมีเสียงตัวกระตุ้นทั้งหมด 169 เสียง ในการทำการทดลอง จะให้ผู้ฟัง ฟังเสียงตัวกระตุ้น ผ่านโปรแกรมคอมพิวเตอร์ที่สร้างขึ้นสำหรับการทดลองนี้โดยเฉพาะ โดยมีตัวอย่างหน้าจอของโปรแกรมแสดงดังรูปที่ 4.3 และฟังเสียงผ่านหูฟังที่จัดเตรียมไว้ให้ในระดับเสียงที่เหมาะสม กลุ่มผู้ฟังจะได้รับคำชี้แจงต่าง ๆ จากผู้วิจัยก่อนที่จะเริ่มต้นทำการทดลอง และให้ผู้ฟัง ฟังเสียงตัวอย่างอ้างอิง ของแต่ละเสียงวรรณยุกต์ก่อนทำการทดลอง โดยให้

ผู้ฟัง ตอบเสียงวรรณยุกต์ของพยางค์เป้าหมาย คำว่า “[น-า]” ในพยางค์ที่สองของเสียงตัวกระตุ้นที่ได้ยิน จากตัวเลือกเสียงวรรณยุกต์ทั้งห้าเสียงในภาษาไทย จากการรับฟังเสียงของพยางค์เป้าหมาย ทั้งนี้ผู้ฟังสามารถฟังซ้ำเสียงคำถามได้ตามที่ต้องการ เพื่อให้คำตอบที่ได้ เกิดความผิดพลาดน้อยที่สุด อันเนื่องมาจากการที่ผู้ฟังตัดสินใจไม่ได้ จากการฟังเพียงครั้งเดียว โดยเสียงตัวกระตุ้นจะถูกเรียงลำดับแบบสุ่ม และเนื่องจากจำนวนเสียงตัวกระตุ้นที่มีปริมาณมาก ดังนั้นจึงให้กลุ่มผู้ฟังได้พักเป็นระยะเวลา 2 นาที ทุก ๆ 50 เสียงกระตุ้น เพื่อหลีกเลี่ยงอาการล้าที่เกิดจากการฟังและตอบคำถาม



ภาพที่ 4.3 หน้าจอโปรแกรมที่ใช้ในการทดลองการระบุเสียงวรรณยุกต์

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ผลการทดลองการระบุเสียงวรรณยุกต์ ดังแสดงในภาพที่ 4.4

(ก) หญิง

1) สามัญ

จุดสิ้นสุด (Hz)

จุดกึ่งกลาง (Hz) M	290	280	270	260	250	240	230	220	210	200	190	180	170
290													
280											1		
270													
260									1				
250					1	1		2	4	2	1	2	
240					3	13	14	9	8	1	2		
230					5	15	20	20	20	4	4		
220					1	11	20	20	20	14	1	3	
210				1		9	17	20	18	15	3		
200						3	14	14	14	3	1		
190								1		2			
180													
170													

2) เอก

จุดสิ้นสุด (Hz)

จุดกึ่งกลาง (Hz) L	290	280	270	260	250	240	230	220	210	200	190	180	170
290							1			2			1
280												1	1
270							2					1	1
260							1		1				2
250								2	2	3	3	3	5
240						1		2	2	3	9	8	6
230									1	10	9	12	14
220									1	13	16	18	17
210									1	8	18	16	18
200	1	1					1	1	4	13	19	18	16
190					1		1		2	10	17	19	19
180						1	1			2	6	15	19
170	1									3	3	7	12

3) โท

จุดสิ้นสุด (Hz)

จุดกึ่งกลาง (Hz) F	290	280	270	260	250	240	230	220	210	200	190	180	170
290	14	16	20	20	20	20	19	20	19	18	19	19	16
280	5	12	17	20	20	20	18	18	19	20	19	18	17
270	4	6	12	20	20	20	17	20	18	20	19	18	18
260	1	1	11	12	16	19	19	20	19	19	20	18	16
250	1	3	3	6	11	17	17	17	17	12	13	15	14
240	1	1		2	4	6	6	12	13	16	9	9	12
230	2	1		1	1	2			5	7	7	6	4
220	2		2				1			1	1		2
210	1		1	1	1	1	1					1	
200				1	2		2	1					1
190	1	1					1						
180		1											
170													

(ข) ชาย

1) สามัญ

จุดสิ้นสุด (Hz)

จุดกึ่งกลาง (Hz) M	160	155	150	145	140	135	130	125	120	115	110	105	100
160													
155											1		
150													
145									1				
140					1	1		2	4	2	1	2	
135					3	13	14	9	8	1	2		
130					5	15	20	20	20	4	4		
125					1	11	20	20	20	14	1	3	
120				1		9	17	20	18	15	3		
115						3	14	14	14	3	1		
110								1		2			
105													
100													

2) เอก

จุดสิ้นสุด (Hz)

จุดกึ่งกลาง (Hz) L	160	155	150	145	140	135	130	125	120	115	110	105	100
160						1		1	3		4	2	3
155							1	3			1	3	1
150								1	1			1	2
145						1			1		1		3
140										4	3	4	5
135									3	2	6	6	10
130										7	13	16	16
125									3	6	17	17	19
120									1	5	17	18	18
115					1	1	1	1	3	13	17	16	20
110							2		1	12	18	18	20
105				1		2	2			6	8	19	18
100										3	2	4	14

3) โท

จุดสิ้นสุด (Hz)

จุดกึ่งกลาง (Hz) F	160	155	150	145	140	135	130	125	120	115	110	105	100
160	15	17	19	20	20	19	20	19	17	19	16	18	17
155	6	18	16	20	18	20	19	17	20	19	18	17	19
150	3	14	16	19	20	20	20	19	18	20	20	18	18
145	2	1	13	16	17	19	20	20	17	20	19	20	17
140		2	3	11	16	17	20	18	16	14	16	14	15
135	2	1	1		1	3	2	11	11	17	8	14	10
130	1	1	1							4	2	2	4
125									1				1
120	3	1		1									1
115					2		1						
110													
105						1							
100	1												

(ก) หญิง

4) ตรี

จุดสิ้นสุด (Hz)

จุดกึ่งกลาง (Hz)	290	280	270	260	250	240	230	220	210	200	190	180	170
290	6	4							1		1	1	1
280	15	8	3				2	2			1	1	2
270	16	14	8				1		2		1		
260	19	19	9	8	4	1				1		1	1
250	19	17	17	14	9	1	2	1		2		1	
240	19	19	20	18	16	9	2						
230	18	18	20	19	19	13							
220	18	20	18	20	20	19	2						
210	18	20	19	19	18	19	17						
200	11	11	13	14	16	13	10	1	1				
190	3	3	3	4	5	5	3	4		1			
180	3	3	2	2	1	1		1					
170	1	1											

5) จัตวา

จุดสิ้นสุด (Hz)

จุดกึ่งกลาง (Hz)	290	280	270	260	250	240	230	220	210	200	190	180	170
290													1
280													
270													1
260													1
250											3		1
240									1		1	1	2
230		1									1	1	2
220										1	1		1
210	1				1					1	1	1	2
200	8	8	7	5	2	7	7	11	4	3		1	3
190	16	16	17	16	14	15	15	15	16	8	3	1	1
180	17	16	18	18	19	18	19	19	20	18	14	5	1
170	18	19	20	20	20	20	20	20	20	17	17	13	8

(ข) ชาย

4) ตรี

จุดสิ้นสุด (Hz)

จุดกึ่งกลาง (Hz)	160	155	150	145	140	135	130	125	120	115	110	105	100
160	5	3	1							1			
155	14	2	4		2						1		
150	17	6	4	1					1				
145	18	19	7	4	3				1				
140	20	18	18	9	3	2							
135	18	19	19	20	16	4	4						
130	19	20	20	20	15	5							
125	20	19	18	20	19	9							
120	17	19	20	17	20	11	3						
115	17	16	16	17	15	4							
110	3	2	3	1	4	3	1	1	1				
105	4	4	2	1	2	3	2						
100	2												

5) จัตวา

จุดสิ้นสุด (Hz)

จุดกึ่งกลาง (Hz)	160	155	150	145	140	135	130	125	120	115	110	105	100
160													
155													
150												1	
145													
140													
135													
130										4	2	2	
125			1						2		2		
120				1					1			2	1
115	3	4	4	3	2	13	4	4	3	4	2	4	
110	16	16	18	19	16	17	17	18	18	6	2	2	
105	16	16	18	19	18	17	18	20	20	14	12	1	2
100	17	20	20	20	20	20	20	20	20	17	18	16	6

ภาพที่ 4.4 ผลการทดลองการระบุเสียงวรรณยุกต์

จากภาพที่ 4.4 (ก) แสดงผลการระบุเสียงวรรณยุกต์ของเสียงผู้พูดที่ 1 (ข) แสดงผลการระบุเสียงวรรณยุกต์ของเสียงผู้พูดที่ 2 และตารางที่ 1-5 ของแต่ละชุด หมายถึงตารางบอกผลของการระบุเสียงวรรณยุกต์ของเสียงวรรณยุกต์สามัญ, เอก, โท, ตรี และจัตวา ตามลำดับ แถวของตาราง คือค่าความที่มูลฐานของโครงร่างความถี่มูลฐานที่ตำแหน่งจุดกึ่งกลางของพยางค์เป้าหมาย ส่วนคอลัมน์ของตาราง คือค่าความที่มูลฐานของโครงร่างความถี่มูลฐานที่ตำแหน่งจุดสิ้นสุดของพยางค์เป้าหมาย ในหน่วยเฮิร์ตซ์ และตัวเลขในแต่ละช่องของตารางแสดงจำนวนครั้งที่กลุ่มผู้ฟังระบุเป็นเสียงวรรณยุกต์นั้น ๆ และเนื่องจากเสียงตัวกระตุ้นที่ถูกระบุเป็นเสียงวรรณยุกต์นั้น ๆ จะถูกนำไปเป็นเสียงตัวกระตุ้นในการทดลองการให้คะแนนความเป็นธรรมชาติของเสียงวรรณยุกต์ต่อไป โดยที่เสียงตัวกระตุ้นที่ถูกระบุให้เป็นเสียงวรรณยุกต์ใด จะถูกจัดอยู่ในกลุ่มเสียงวรรณยุกต์นั้น ๆ ซึ่งจากผลการทดลอง พบว่าเสียงกระตุ้นที่ใช้ในการทดลองการให้คะแนนความเป็นธรรมชาติ จะมีจำนวนที่มากเกินไป หากใช้เสียงตัวกระตุ้นทุกเสียงที่ถูกระบุเสียงวรรณยุกต์จากกลุ่มผู้ใช้ทั้งหมด และอาจเกิดความผิดพลาดของผู้ฟังในการตอบคำถาม หรือมีการระบุเสียงวรรณยุกต์เดียวกันเป็นจำนวนน้อยมากในบางเสียง

ตัวกระตุ้น ดังนั้น จึงจะนำเสียงตัวกระตุ้นที่ถูกกลุ่มผู้ฟังระบุเป็นเสียงวรรณยุกต์เดียวกันจำนวน 80% ขึ้นไป ของจำนวนกลุ่มผู้ฟัง (5 คนจาก 20 คนในกลุ่มผู้ฟัง) มาเป็นเสียงตัวกระตุ้นของการทดลองการให้คะแนนความเป็นธรรมชาติของเสียงวรรณยุกต์ ซึ่งแสดงผลเป็นช่องสีทึบในภาพที่ 4.4 และจำนวนเสียงตัวกระตุ้นดังแสดงในตารางที่ 4.2

ตารางที่ 4.2 จำนวนเสียงตัวกระตุ้นที่ใช้ในการทดลองการให้คะแนนความเป็นธรรมชาติของเสียงวรรณยุกต์

เสียงวรรณยุกต์	จำนวนเสียงตัวกระตุ้น	
	ผู้พูดที่ 1 (เพศหญิง)	ผู้พูดที่ 2 (เพศชาย)
สามัญ	15	22
เอก	29	29
โท	71	65
ตรี	49	39
จัตวา	42	35
รวม	206	190

4.2.2 การให้คะแนนความเป็นธรรมชาติและผลการทดลอง

กลุ่มผู้ฟังที่เข้าร่วมการทดลองการระบุเสียงวรรณยุกต์นั้น มีทั้งหมด 10 คน เป็นเพศชาย 8 คน และเพศหญิง 2 คน มีอายุระหว่าง 21 – 30 ปี มีค่าเฉลี่ยของอายุเท่ากับ 22.1 ปี ซึ่งกลุ่มผู้ฟัง มีความเข้าใจในระบบเสียงวรรณยุกต์ของภาษาไทยเป็นอย่างดี โดยกลุ่มผู้ฟังที่เข้าร่วมการทดลองการให้คะแนนความเป็นธรรมชาติ เป็นกลุ่มที่ไม่ซ้ำกับกลุ่มผู้ฟังที่ทำการระบุเสียงวรรณยุกต์ทั้งหมด การทดลองนั้นแบ่งออกเป็น 2 รอบ ได้แก่ รอบที่ 1 เป็นการทดลองโดยใช้เสียงตัวกระตุ้นจากชุดฝึกฝนของผู้พูดที่ 1 (เพศหญิง) และรอบที่ 2 เป็นการทดลองโดยใช้เสียงตัวกระตุ้นจากชุดฝึกฝนของผู้พูดที่ 2 (เพศชาย) แต่ละรอบของการทดลอง ถูกจัดขึ้นในระยะเวลาที่ห่างกัน 1 สัปดาห์ ในการทำการทดลอง จะให้ผู้ฟัง ฟังเสียงตัวกระตุ้น ผ่านโปรแกรมคอมพิวเตอร์ที่สร้างขึ้นสำหรับการทดลองนี้โดยเฉพาะ โดยมีตัวอย่างหน้าจอของโปรแกรมแสดงดังรูปที่ 4.5 เสียงตัวกระตุ้นจะถูกนำเสนอเป็นกลุ่มตามที่ได้จากการระบุเสียงวรรณยุกต์ ได้แก่กลุ่มเสียงที่ถูกระบุเป็นเสียงวรรณยุกต์สามัญ, เอก, โท, ตรี และจัตวา ตามลำดับ และภายในกลุ่มเสียงตัวกระตุ้นของแต่ละเสียงวรรณยุกต์ จะถูกเรียงลำดับแบบสุ่ม ซึ่งจำนวนของเสียงตัวกระตุ้นได้แสดงไว้ในตารางที่ 4.2 กลุ่มผู้ฟังได้รับคำสั่งให้ตัดสินระดับคะแนนความเป็นธรรมชาติของพยางค์เป้าหมายในเสียงตัวกระตุ้น ว่ามีความเป็นธรรมชาติในระดับใดกับเสียงวรรณยุกต์ของกลุ่มนั้น ๆ ซึ่งเกณฑ์การให้คะแนนของแต่ละระดับความเป็นธรรมชาติถูกชี้แจงให้กลุ่มผู้ฟังก่อนเริ่มต้นการทดลอง ดังตารางที่ 3.1 และเพื่อลดขนาดของความผิดพลาดของคำตอบ

ซึ่งเกิดจากการมีแนวโน้มในการให้คำตอบที่อยู่ตรงกลางของช่วงระดับคะแนน (Error of Central Tendency) [28, 29] จึงให้ผู้ฟัง ฟังเสียงอ้างอิงซึ่งเป็นเสียงที่เป็นธรรมชาติของแต่ละเสียงวรรณยุกต์นั้น ๆ ทุกครั้งก่อนการฟังเสียงคำถามในแต่ละคำถาม

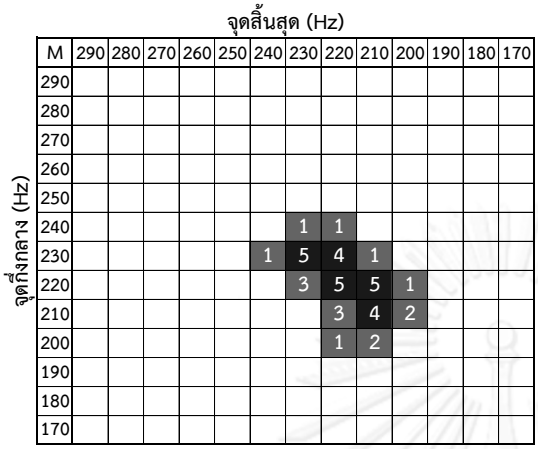
ภาพที่ 4.5 หน้าจอโปรแกรมที่ใช้ในการทดลองการให้คะแนนความเป็นธรรมชาติ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

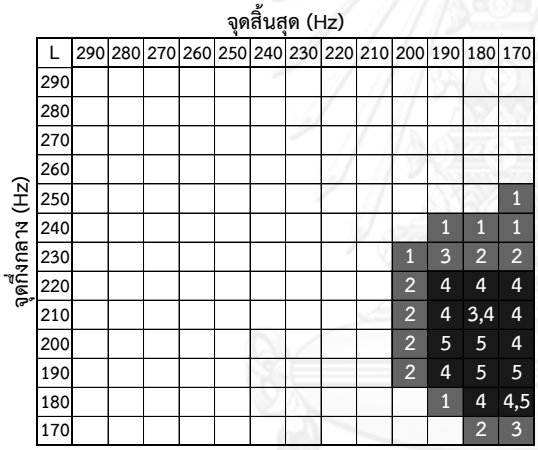
ผลการทดลองการให้คะแนนความเป็นธรรมชาติ ดังแสดงในภาพที่ 4.6

(ก) หญิง

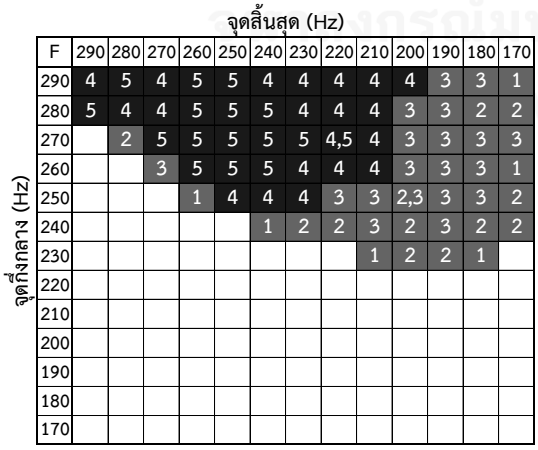
1) สามัญ



2) เอก

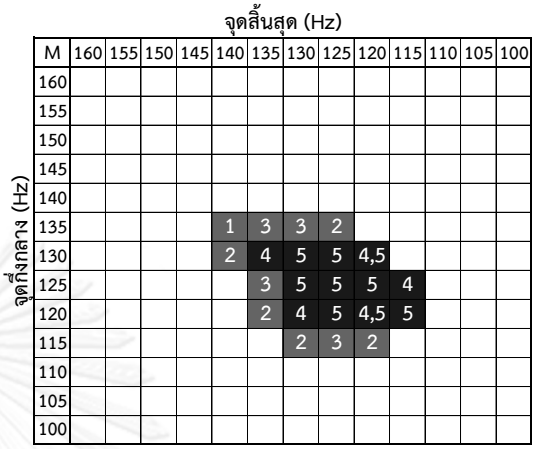


3) โท

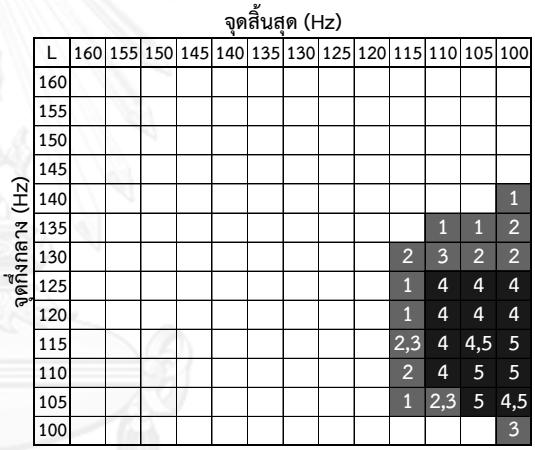


(ข) ชาย

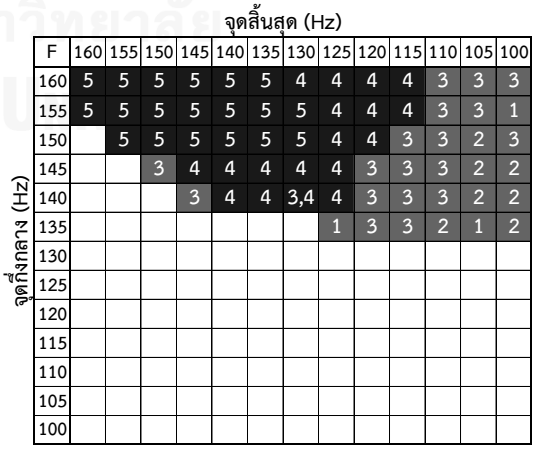
1) สามัญ



2) เอก



3) โท



(ก) หญิง

4) ตรี

จุดสิ้นสุด (Hz)

จุดกึ่งกลาง (Hz)	290	280	270	260	250	240	230	220	210	200	190	180	170
290	2												
280	2	1,2											
270	2	1,2	1										
260	3	2	1	1									
250	4	4	4	2	1								
240	4	5	4	4	2	1							
230	4	5	5	5	5	3							
220	3	4	5	5	5	4							
210	3	4	4	4,5	4	4	3						
200	3	3	2	2	2	1	2						
190					1	1							
180													
170													

5) จัตวา

จุดสิ้นสุด (Hz)

จุดกึ่งกลาง (Hz)	290	280	270	260	250	240	230	220	210	200	190	180	170
290													
280													
270													
260													
250													
240													
230													
220													
210													
200	1	1	1	2		1	2	1					
190	2	3	3	3	3	3	4	4	2	2			
180	2	3	3	3	3	4	4	5	5	5	5	3	
170	2	2	3	3	3	3	4	5	5	5	5	5	4

(ข) ชาย

4) ตรี

จุดสิ้นสุด (Hz)

จุดกึ่งกลาง (Hz)	160	155	150	145	140	135	130	125	120	115	110	105	100
160	1												
155	2												
150	2	2											
145	3	3	2										
140	5	4	2	2									
135	5	5	5	4	2								
130	5	5	5	4	4	1							
125	5	5	5	4,5	4	4							
120	5	5	5	5	5	4							
115	3	3	3	4	3								
110													
105													
100													

5) จัตวา

จุดสิ้นสุด (Hz)

จุดกึ่งกลาง (Hz)	160	155	150	145	140	135	130	125	120	115	110	105	100
160													
155													
150													
145													
140													
135													
130													
125													
120													
115						2							
110	2	2	3	3	3	3	3	2	1,2	2			
105	3	3	3	4	4	4,5	5	5	5	5	5		
100	3	3	4	4	4	5	5	5	5	5	5	5	5

ภาพที่ 4.6 ผลการทดลองการให้คะแนนความเป็นธรรมชาติ

จากภาพที่ 4.6 (ก) แสดงผลการให้คะแนนความเป็นธรรมชาติของเสียงผู้พูดที่ 1 (ข) แสดงผลการให้คะแนนความเป็นธรรมชาติของเสียงผู้พูดที่ 2 และตารางที่ 1-5 ของแต่ละชุด หมายถึง ตารางบอกผลของการให้คะแนนความเป็นธรรมชาติของเสียงวรรณยุกต์สามัญ, เอก, โท, ตรี และจัตวา ตามลำดับ แถวของตาราง คือค่าความที่มูลฐานของโครงร่างความถี่มูลฐานที่ตำแหน่งจุดกึ่งกลางของพยางค์เป้าหมาย ส่วนคอลัมน์ของตาราง คือค่าความที่มูลฐานของโครงร่างความถี่มูลฐานที่ตำแหน่งจุดสิ้นสุดของพยางค์เป้าหมาย ในหน่วยเฮิร์ตซ์ ตัวเลขในช่องตารางแสดงถึงระดับคะแนนความเป็นธรรมชาติส่วนใหญ่ (Majority) จากการตอบระดับระแนนของกลุ่มผู้ฟัง ซึ่งมีค่าตั้งแต่ 1 (มีความไม่เป็นธรรมชาติมากที่สุด) ถึง 5 (มีความเป็นธรรมชาติมากที่สุด) ในช่องตารางที่มีสองระดับคะแนน หมายถึง กลุ่มผู้ฟังตอบระดับคะแนนนั้น ๆ เป็นจำนวนเท่ากัน

4.2.3 การวิเคราะห์ผลการทดลองการรับรู้

จากภาพรวมของผลการทดลองการระบุเสียงวรรณยุกต์ จะเห็นได้ว่า ทั้งผลของผู้พูดที่ 1 (เพศหญิง) และผู้พูดที่ 2 (เพศชาย) เสียงตัวกระตุ้นที่กลุ่มผู้ฟังตอบเสียงวรรณยุกต์ที่เหมือนกันในจำนวนมากนั้น อยู่ในบริเวณที่ใกล้ ๆ กันของตาราง อีกทั้งยังเห็นการเกาะกลุ่มของคำตอบที่ชัดเจนในแต่ละเสียงวรรณยุกต์ หากพิจารณาค่าของความถี่มูลฐานที่จุดกึ่งกลางและจุดสิ้นสุด เป็นอัตราส่วนภายในช่วงขอบเขตความถี่มูลฐานของผู้พูด ในแต่ละเสียงจะมีค่าของความถี่มูลฐานที่จุดกึ่งกลางและจุดสิ้นสุดที่แตกต่างกัน ซึ่งเป็นข้อบ่งชี้ในการยืนยันว่าค่าของความถี่มูลฐานที่ตำแหน่งจุดกึ่งกลางและจุดสิ้นสุด เพียงพอในการอธิบายโครงร่างอย่างคร่าว ของโครงร่างความถี่มูลฐานของเสียงวรรณยุกต์ต่าง ๆ ในภาษาไทย และผู้ฟังสามารถแยกแยกเสียงวรรณยุกต์ออกจากกันได้ เมื่อพิจารณาช่องสีเทาของตารางในภาพที่ 4.4 ซึ่งเป็นบริเวณขอบของกลุ่มคำตอบที่กลุ่มผู้ฟังระบุเสียงวรรณยุกต์ของพยางค์เป้าหมายตรงกันเกือบทั้งหมด กลุ่มผู้ฟังมีการระบุเสียงวรรณยุกต์ที่ไม่ตรงกัน ในจำนวนที่ใกล้เคียงกันนั่นแสดงให้เห็นถึงความไม่แน่ใจของผู้ฟังในการระบุเสียงวรรณยุกต์ และดังที่ได้กล่าวไว้ข้างต้น เสียงของพยางค์เป้าหมายอาจเกิดการมี 2 เสียงวรรณยุกต์อยู่ในพยางค์เดียว เนื่องจากการปรับแต่งโครงร่างความถี่มูลฐานของเสียงตัวกระตุ้น ทำให้ผู้ฟังระบุหนึ่งในสองเสียงวรรณยุกต์ที่รับรู้อย่างสุ่ม ซึ่งตรงกับผู้ฟังบางรายที่รายงานกลับมายังผู้ทำวิจัย หลังจากทำการทดลอง

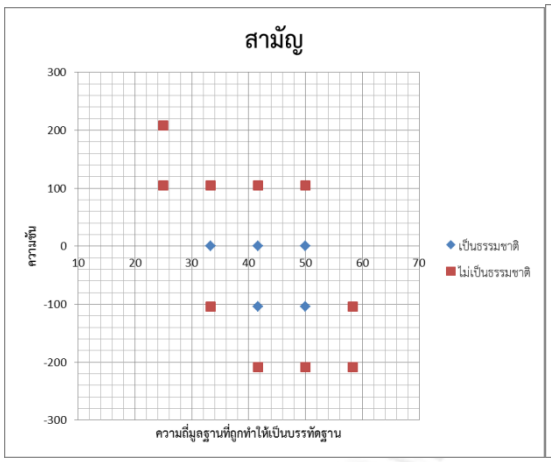
จากผลการทดลองการให้คะแนนระดับความเป็นธรรมชาติของเสียงวรรณยุกต์ ดังภาพที่ 4.6 พบว่าบางส่วนของบริเวณสีดำเข้มในตารางคำตอบของการทำลองการระบุเสียงวรรณยุกต์ ซึ่งเป็นเสียงของพยางค์เป้าหมายที่ผู้ฟังเกือบทั้งหมดในกลุ่มที่รับรู้เป็นเสียงวรรณยุกต์นั้น ๆ กลับได้คะแนนระดับความเป็นธรรมชาติน้อย สาเหตุเกิดจากการที่ผู้ฟังตัดสินความเป็นธรรมชาติจากสำเนียงแบบธรรมชาติ และเสียงของพยางค์เป้าหมายเหล่านั้นมีความเหน่อที่ไม่เป็นธรรมชาติ และจากผลการทดลองนั้น เมื่อพิจารณาตามเกณฑ์การให้คะแนนความเป็นธรรมชาติ ที่ระบุให้ระดับคะแนน 4 และ 5 ถือเป็นเสียงพูดที่ยอมรับได้ ถึงแม้ว่าในระดับคะแนน 4 จะมีความไม่เป็นธรรมชาติเกิดขึ้นเล็กน้อยก็ตาม จะสังเกตได้ว่า บริเวณคำตอบของระดับคะแนน 4 และ 5 (บริเวณช่องคำตอบในตารางคำตอบที่มีสีดำ) ถูกแยกออกอย่างชัดเจนกับคำตอบของระดับคะแนน 1, 2 และ 3 (บริเวณช่องคำตอบในตารางคำตอบที่มีสีเทา) ซึ่งกำหนดให้เป็นเสียงพูดที่ไม่เป็นธรรมชาติ และยอมรับไม่ได้ ซึ่งอยู่บริเวณขอบของกลุ่มคำตอบ ที่ซึ่งมีลักษณะโครงร่างที่คลุมเครือระหว่างสองเสียงวรรณยุกต์ ดังนั้นจึงทำการรวมคำตอบของระดับคะแนน 4 และ 5 เข้าไว้ด้วยกันเพื่อหาความสัมพันธ์ระหว่างค่าจุดกึ่งกลาง และจุดสิ้นสุด เพื่อนำไปสร้างแบบจำลองเพื่อทำนายการรับรู้ความเป็นธรรมชาติของของเสียงวรรณยุกต์ต่อไป

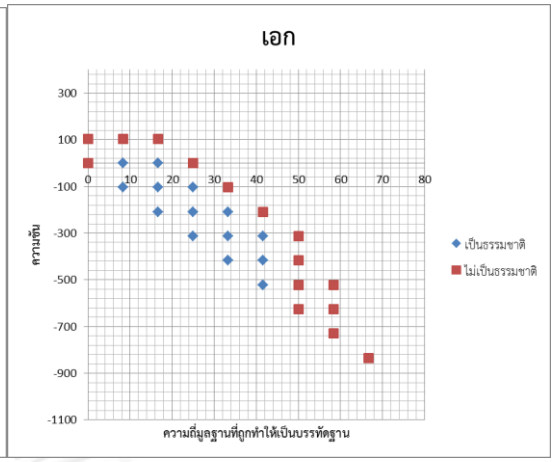
4.3 การสร้างแบบจำลองการทำนายการรับรู้ความเป็นธรรมชาติของเสียงวรรณยุกต์

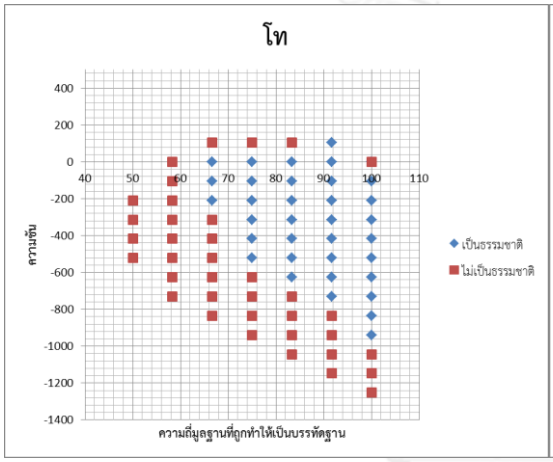
จากผลการทดลองการรับรู้ความเป็นธรรมชาติของเสียงวรรณยุกต์ จึงได้นำค่าความถี่มูลฐานที่ถูกแปลงให้เป็นบรรทัดฐานของจุดกึ่งกลาง และค่าความชันของโครงร่างความถี่มูลฐานระหว่างตำแหน่งจุดกึ่งกลางถึงตำแหน่งจุดสิ้นสุด ของพยางค์เป้าหมาย ซึ่งหาได้ดังสมการที่ 3.2 และ 3.3 ของพยางค์เป้าหมายที่มีระดับคะแนนความเป็นธรรมชาติ 4 และ 5 ของเสียงวรรณยุกต์นั้น ๆ ซึ่งแสดงถึงพยางค์เป้าหมายที่มีโครงร่างความถี่มูลฐานที่เป็นธรรมชาติและยอมรับได้ จากการรับรู้ของคน มาวาดลงบนแผนภาพกระจาย (Scatter Plot) ดังแสดงในภาพที่ 4.7

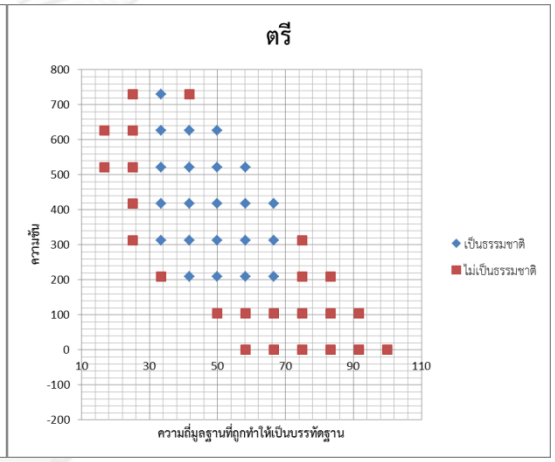
จากภาพที่ 4.7 (ก) แสดงแผนภาพกระจายที่แสดงความสัมพันธ์ระหว่างค่าความถี่มูลฐานที่ถูกแปลงให้เป็นบรรทัดฐานของจุดกึ่งกลางกับค่าความชันของโครงร่างความถี่มูลฐานระหว่างตำแหน่งจุดกึ่งกลางถึงตำแหน่งจุดสิ้นสุดของพยางค์เป้าหมาย ของผู้พูดเพศหญิง และ (ข) แสดงแผนภาพกระจายที่แสดงความสัมพันธ์ระหว่างค่าความถี่มูลฐานที่ถูกแปลงให้เป็นบรรทัดฐานของจุดกึ่งกลางกับค่าความชันของโครงร่างความถี่มูลฐานระหว่างตำแหน่งจุดกึ่งกลางถึงตำแหน่งจุดสิ้นสุดของพยางค์เป้าหมาย ของผู้พูดเพศชาย ซึ่งสังเกตได้ว่า ค่าความชันของเพศหญิง จะมีค่ามากกว่าค่าความชันของเพศชาย ในทุกๆเสียงวรรณยุกต์ที่ถูกตัดสินว่าเป็นธรรมชาติ กล่าวได้อีกนัยหนึ่ง คือ เสียงพูดของผู้พูดเพศหญิงที่ยังคงรับฟังเป็นธรรมชาตินั้น สามารถมีค่าสัมบูรณ์ของความชันของครึ่งพยางค์หลังได้มากกว่าเสียงพูดของผู้พูดเพศชาย ทั้งนี้เนื่องจากความกว้างของขอบเขตความถี่มูลฐานของเพศหญิงมากกว่าเพศชาย อีกทั้งค่าความถี่มูลฐานนั้น มีความสัมพันธ์อย่างมาก กับขอบเขตความถี่มูลฐานของเสียงพูด ของผู้พูด ซึ่งการรับรู้ระดับเสียงของเสียงพูดนั้น มีความสัมพันธ์กับตำแหน่งภายในขอบเขตเสียงพูด [32]

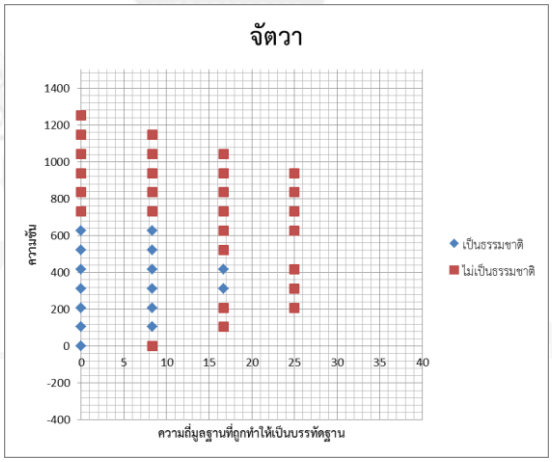
จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University


สามัญ
ความถี่มูลฐานที่ถูกทำให้เป็นบรรทัดฐาน
เป็นธรรมชาติ
ไม่เป็นธรรมชาติ

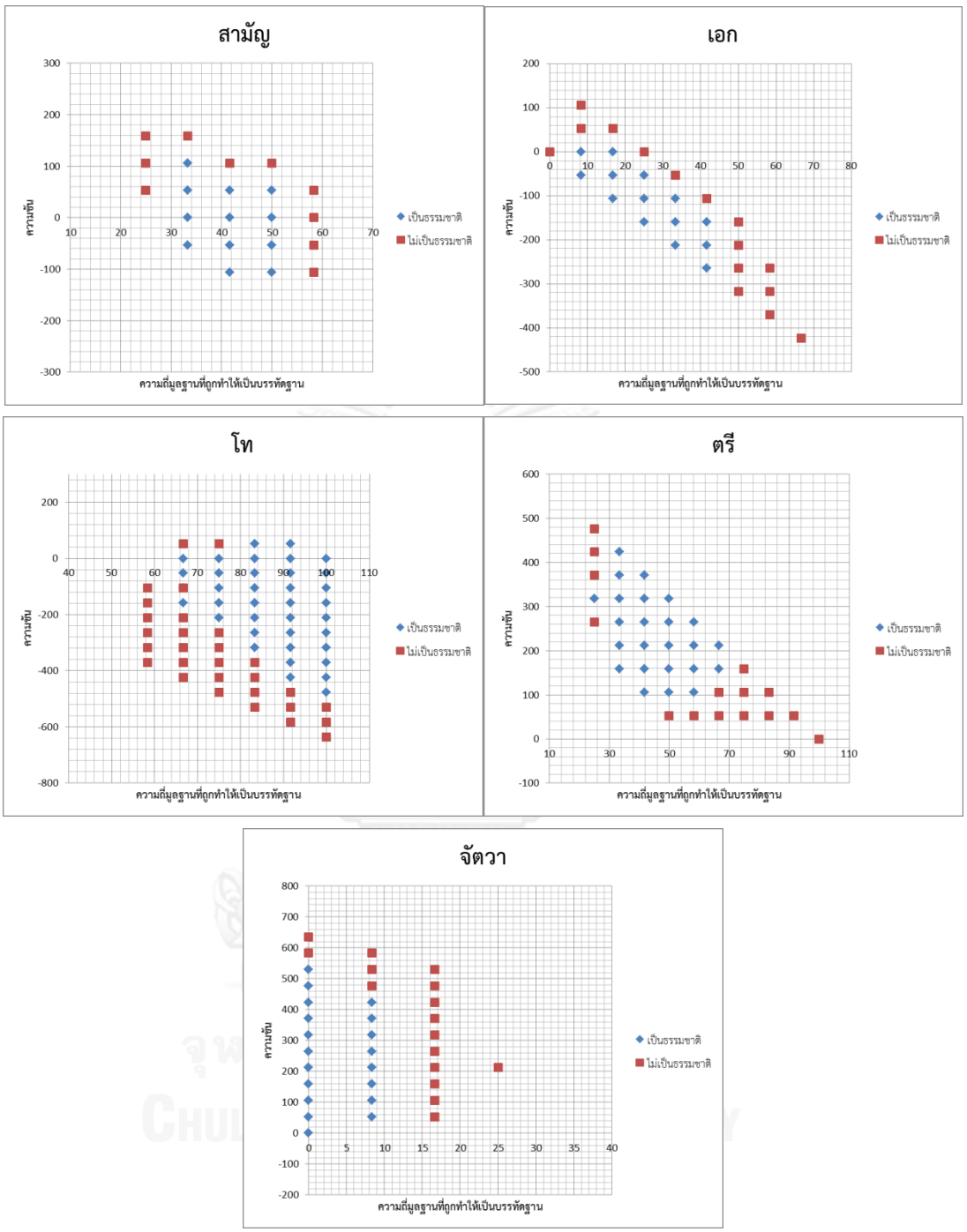

เอก
ความถี่มูลฐานที่ถูกทำให้เป็นบรรทัดฐาน
เป็นธรรมชาติ
ไม่เป็นธรรมชาติ


โท
ความถี่มูลฐานที่ถูกทำให้เป็นบรรทัดฐาน
เป็นธรรมชาติ
ไม่เป็นธรรมชาติ


ตรี
ความถี่มูลฐานที่ถูกทำให้เป็นบรรทัดฐาน
เป็นธรรมชาติ
ไม่เป็นธรรมชาติ


จัตวา
ความถี่มูลฐานที่ถูกทำให้เป็นบรรทัดฐาน
เป็นธรรมชาติ
ไม่เป็นธรรมชาติ

(ก)



(ข)

ภาพที่ 4.7 แผนภาพกระจายระหว่างค่าจุดกึ่งกลางและความชันของพยางค์เป้าหมายที่มีความเป็นธรรมชาติและไม่เป็นธรรมชาติความเป็นธรรมชาติของแต่ละเสียงวรรณยุกต์จากการทดลองการรับรู้

(ก) ผู้พูดที่ 1 (เพศหญิง)

(ข) ผู้พูดที่ 2(เพศชาย)

จากภาพที่ 4.7 (ก) และ (ข) แสดงแผนภาพกระจายของความสัมพันธ์ระหว่างค่าจุดกึ่งกลางและความชันในครึ่งหลังของช่วงหน่วยตามพยางค์ของพยางค์เป้าหมาย ที่ส่งผลต่อการรับรู้ความเป็นธรรมชาติของแต่ละวรรณยุกต์ของผู้พูดที่ 1 (เพศหญิง) และผู้พูดที่ 2 (เพศชาย) ตามลำดับ จุดสี่เหลี่ยมจัตุรัสในแผนภาพกระจาย หมายถึงความสัมพันธ์ที่กลุ่มผู้ฟังส่วนใหญ่ ตัดสินว่ามีความเป็นธรรมชาติ (ระดับคะแนนความเป็นธรรมชาติ 4 และ 5) ส่วนจุดสี่เหลี่ยมข้าวหลามตัด หมายถึงความสัมพันธ์ที่กลุ่มผู้ฟังส่วนใหญ่ ตัดสินว่ามีความไม่เป็นธรรมชาติ (ระดับคะแนนความเป็นธรรมชาติ 1, 2 และ 3)

เมื่อนำข้อมูลจากการทดลองมาวาดลงบนแผนภาพกระจายแล้ว จึงทำการสร้างแบบจำลองการรับรู้ความเป็นธรรมชาติของเสียงวรรณยุกต์ ด้วยแบบจำลองเรขาคณิต โดยการหาสมการที่พอดี (fit) กับบริเวณขอบ ของขอบเขตคำตอบที่มีความเป็นธรรมชาติ เพื่อกำหนดบริเวณของความเป็นธรรมชาติ จากความสัมพันธ์ของค่าความถี่มูลฐานที่ถูกแปลงให้เป็นบรรทัดฐานของจุดกึ่งกลาง กับค่าความชันของโครงร่างความถี่มูลฐานระหว่างตำแหน่งจุดกึ่งกลางถึงตำแหน่งจุดสิ้นสุด ซึ่งใช้ตัดสินความเป็นธรรมชาติ หากเสียงพูดใด ๆ มีค่าความถี่มูลฐานที่ถูกแปลงให้เป็นบรรทัดฐานของจุดกึ่งกลางกับค่าความชันของโครงร่างความถี่มูลฐานมีความสัมพันธ์อยู่ภายในบริเวณของความเป็นธรรมชาติของแบบจำลองเรขาคณิตที่สร้างขึ้น จากการศึกษาการรับรู้ความเป็นธรรมชาติของแต่ละเสียงวรรณยุกต์ ซึ่งแบบจำลอง จะแบ่งตามเพศ และเสียงวรรณยุกต์ โดยแต่ละเพศและเสียงวรรณยุกต์สามารถสร้างแบบจำลองเรขาคณิตได้ดังต่อไปนี้

4.3.1. แบบจำลองการรับรู้ความเป็นธรรมชาติของเสียงวรรณยุกต์

ผลของการสร้างแบบจำลองเรขาคณิตในการทำนายการรับรู้ความเป็นธรรมชาติของเสียงวรรณยุกต์ แบ่งออกเป็นแบบจำลองของเสียงเพศหญิง และเพศชาย ซึ่งมี 5 เสียงวรรณยุกต์ ดังนั้น จึงมีแบบจำลองทั้งสิ้น 10 แบบจำลอง ดังนี้

โดยสัญลักษณ์ในสมการแสดงความสัมพันธ์ในแบบจำลอง มีดังนี้

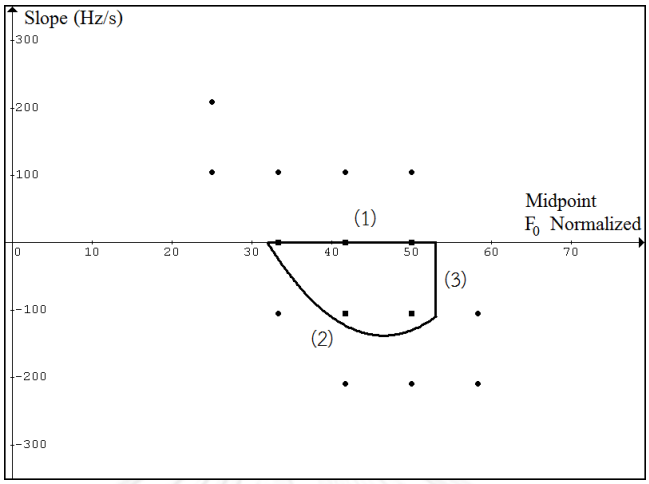
s หมายถึง ค่าความชันของโครงร่างความถี่มูลฐานระหว่างตำแหน่งจุดกึ่งกลางถึงตำแหน่งจุดสิ้นสุด

f หมายถึง ค่าความถี่มูลฐานที่ถูกแปลงให้เป็นบรรทัดฐานของจุดกึ่งกลาง

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

1. แบบจำลองการรับรู้ความเป็นธรรมชาติของเสียงวรรณยุกต์ของเสียงเพศหญิง

1.1 เสียงวรรณยุกต์สามัญ



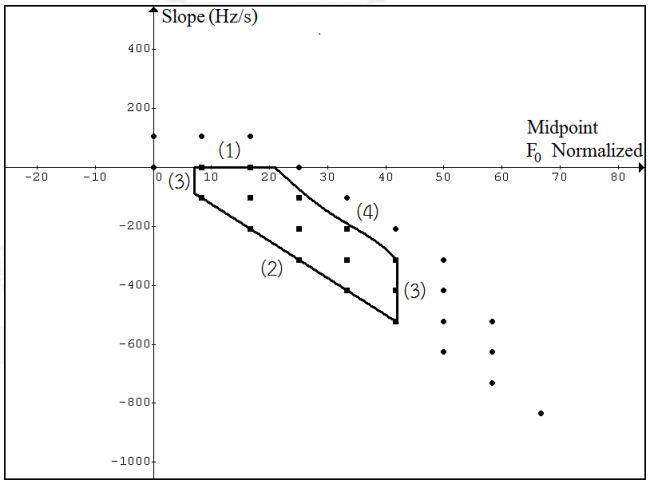
ระบบบอสมการของแบบจำลองเรขาคณิต

$$s \leq 0 \tag{4.1}$$

$$s \geq 0.6546f^2 - 60.8795f + 1277.8195 \tag{4.2}$$

$$f \leq 53 \tag{4.3}$$

1.2 เสียงวรรณยุกต์เอก



ระบบบอสมการของแบบจำลองเรขาคณิต

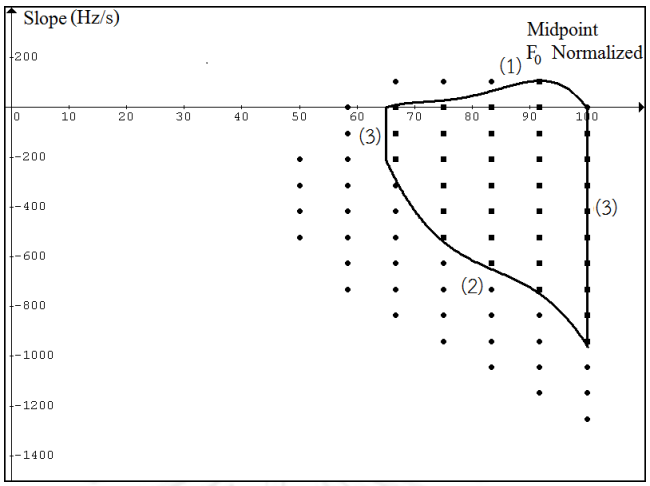
$$s \leq 0 \tag{4.4}$$

$$s \geq -12.5231f - (3.4197 \times 10^{-7}) \tag{4.5}$$

$$7 \leq s \leq 42 \tag{4.6}$$

$$s \leq -0.00177f^4 + 0.20107f^3 - 8.1699f^2 + 124.96f - 539.09 \tag{4.7}$$

1.3 เสียงวรรณยุกต์โท



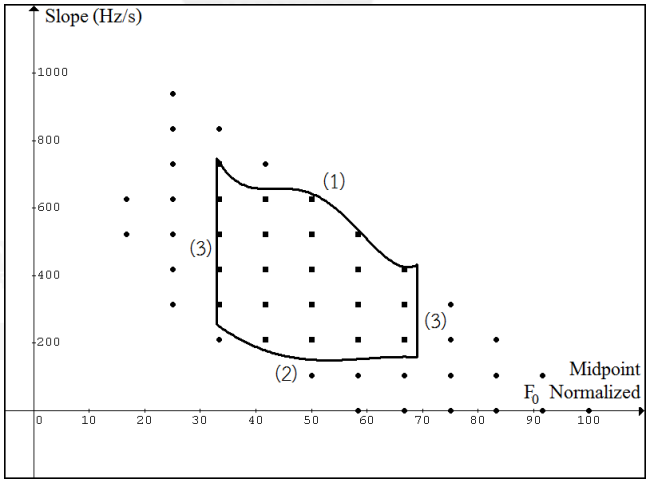
ระบบอสมการของแบบจำลองเรขาคณิต

$$s \leq -0.0012161 f^4 + 0.38083 f^3 - 44.427 f^2 + 2292 f - 44154 \quad (4.8)$$

$$s \geq -0.03678 f^3 + 9.2897 f^2 - 791.99 f + 22122 \quad (4.9)$$

$$65 \leq s \leq 100 \quad (4.10)$$

1.4 เสียงวรรณยุกต์ตรี



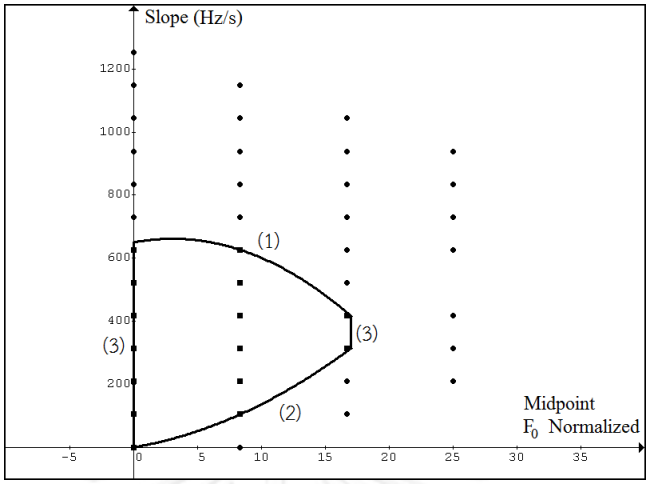
ระบบอสมการของแบบจำลองเรขาคณิต

$$s \leq 0.0020982 f^4 - 0.43004 f^3 + 32.232 f^2 - 1053.5 f + 13379 \quad (4.11)$$

$$s \geq (-6.4090 \times 10^{-5} f^4) + 0.0071 f^3 + 0.1106 f^2 - 33.4404 f + 1058.9678 \quad (4.12)$$

$$33 \leq s \leq 69 \quad (4.13)$$

1.5 เสียงวรรณยุกต์จัตวา



ระบบบอสมการของแบบจำลองเรขาคณิต

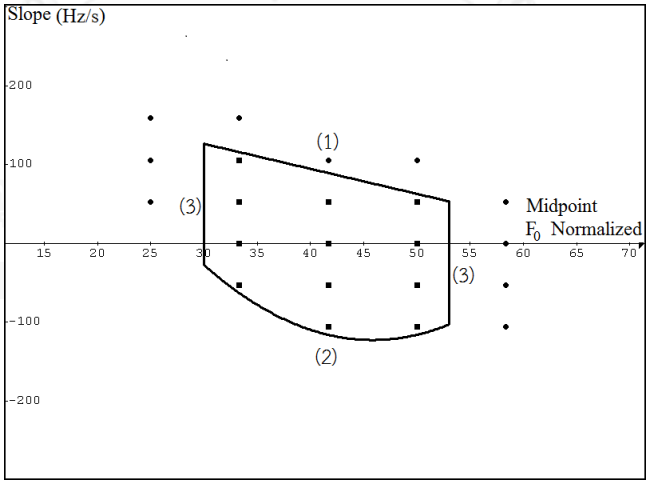
$$s \leq -1.2483 f^2 + 7.5417 f + 650.0 \quad (4.14)$$

$$s \geq 0.68 f^2 + 6.8566 f \quad (4.15)$$

$$0 \geq s \leq 17 \quad (4.16)$$

2. แบบจำลองการรับรู้ความเป็นธรรมชาติของเสียงวรรณยุกต์ ของเสียงพูดเพศชาย

2.1 เสียงวรรณยุกต์สามัญ



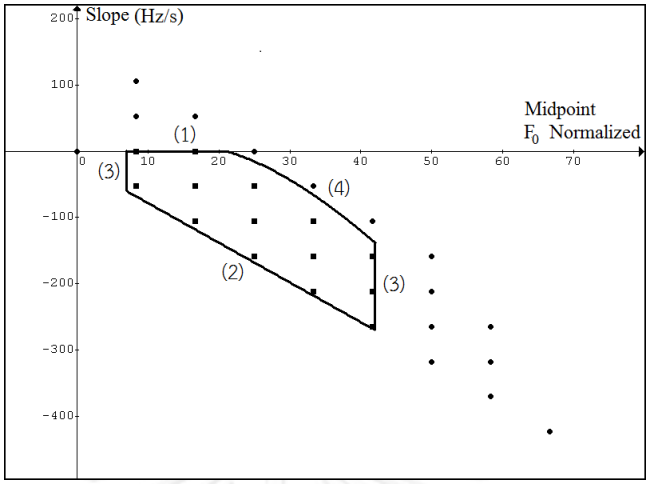
ระบบบอสมการของแบบจำลองเรขาคณิต

$$s \geq -3.1775 f + 221.8324 \quad (4.17)$$

$$s \geq 0.3813 f^2 - 34.9512 f + 678.4319 \quad (4.18)$$

$$30 \leq s \leq 53 \quad (4.19)$$

2.2 เสียงวรรณยุกต์เอก



ระบบบอสมการของแบบจำลองเรขาคณิต

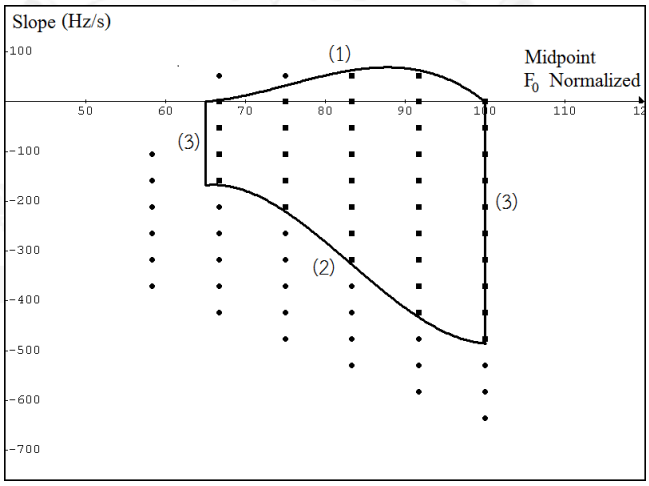
$$s \leq 0 \tag{4.20}$$

$$s \geq -6f - 18 \tag{4.21}$$

$$7 \leq s \leq 42 \tag{4.22}$$

$$s \leq 0.0018162f^3 - 0.31489f^2 + 7.66f - 38.813 \tag{4.23}$$

2.3 เสียงวรรณยุกต์โท



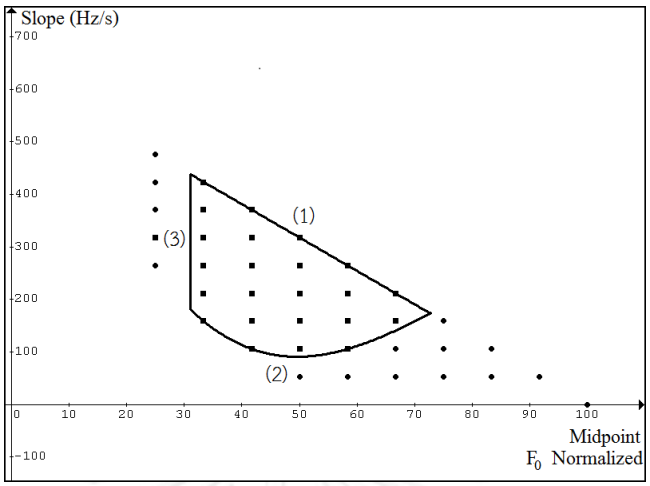
ระบบบอสมการของแบบจำลองเรขาคณิต

$$s \leq -0.0092782f^3 + 2.0979f^2 - 153.86f + 3685.5 \tag{4.24}$$

$$s \geq 0.015217f^3 - 3.8036f^2 + 303.14f - 7980.6 \tag{4.25}$$

$$65 \leq s \leq 100 \tag{4.26}$$

2.4 เสียงวรรณยุกต์ตรี



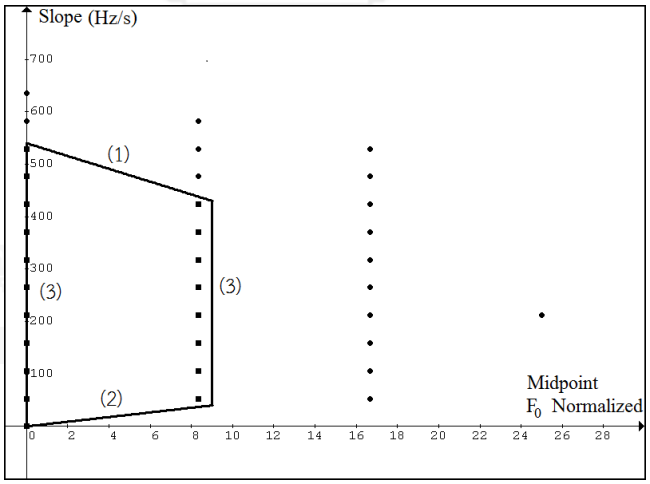
ระบบอสมการของแบบจำลองเรขาคณิต

$$s \leq -6.355f + 635.4971 \tag{4.27}$$

$$s \geq -0.0025633f^3 + 0.59598f^2 - 40.195f + 931.44 \tag{4.28}$$

$$s \geq 31 \tag{4.29}$$

2.5 เสียงวรรณยุกต์จัตวา



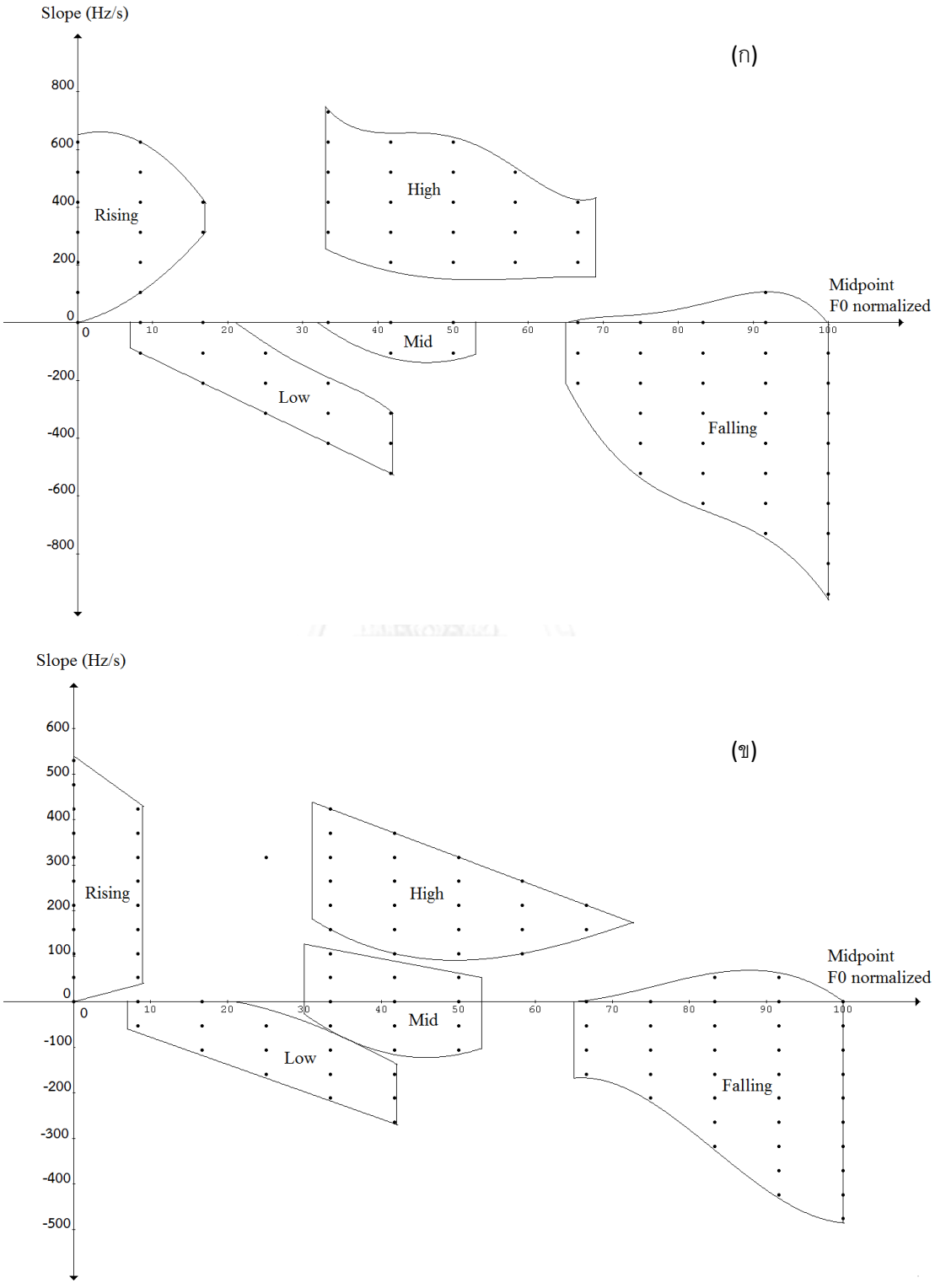
ระบบอสมการของแบบจำลองเรขาคณิต

$$s \leq -12.2222f + 540 \tag{4.30}$$

$$s \leq 4.4444f \tag{4.31}$$

$$0 \leq s \leq 9 \tag{4.32}$$

และเมื่อนำแบบจำลองเรขาคณิตของแต่ละเสียงวรรณยุกต์มาวาดบนระนาบเดียวกัน โดย จะแสดงได้ดังภาพที่ 4.8



ภาพที่ 4.8 แบบจำลองความเป็นธรรมชาติของเสียงวรรณยุกต์
(ก).ผู้พูดที่ 1เพศหญิง (ข). ผู้พูดที่ 2 เพศชาย

4.4 การวัดผลการทำนายการรับรู้ความเป็นธรรมชาติของแบบจำลอง

ในการวัดผลของแบบจำลองการรับรู้ความเป็นธรรมชาติของเสียงวรรณยุกต์นั้น จะทำการเปรียบเทียบผลลัพธ์ของผลการทำนายด้วยแบบจำลองที่สร้างขึ้น กับผลการทดลองการรับรู้ โดยใช้เสียงตัวกระตุ้นในชุดทดสอบ ที่ถูกปรับแต่งโครงร่างความถี่มูลฐานดังที่กล่าวไว้ใน 3.3 สำหรับการทำนายด้วยแบบจำลอง จะให้แบบจำลอง ทำนายความเป็นธรรมชาติของพยางค์เป้าหมายของเสียงตัวกระตุ้นในชุดทดสอบ ซึ่งคำตอบของการทำนาย มี 2 คำตอบ คือ เป็นธรรมชาติ และ ไม่เป็นธรรมชาติ และทำการทดลองการรับรู้ โดยมีรายละเอียดดังนี้

กลุ่มผู้ฟังที่เข้าร่วมการทดลอง มีทั้งหมด 10 คน เป็นเพศชาย 4 คน และเพศหญิง 6 คน และมีผู้ฟัง 3 คน เคยเข้าร่วมการทดลองการให้คะแนนระดับความเป็นธรรมชาติ ซึ่งกลุ่มผู้ฟัง มีความเข้าใจในระบบเสียงวรรณยุกต์ของภาษาไทยเป็นอย่างดี ผู้ฟังแต่ละคนจะได้ฟังเสียงตัวกระตุ้นด้วยโปรแกรมที่ถูกสร้างขึ้น โดยเสียงตัวกระตุ้น ถูกเรียงลำดับแบบสุ่ม ซึ่งมีทั้งหมด 180 เสียง (45 รูปแบบ x 2 ผู้พูดชายและหญิง x ฟังซ้ำ 2 รอบ) เสียงตัวกระตุ้นแต่ละแบบ ผู้ฟังจะได้ฟังซ้ำกัน 2 ครั้งในชุดเสียงตัวกระตุ้น และมีการกำหนดเสียงวรรณยุกต์ของแต่ละเสียงวรรณยุกต์ไว้ ซึ่งผู้ฟังจะต้องตอบว่าเสียงของพยางค์เป้าหมายที่ได้ยินนั้น มีความเป็นธรรมชาติสำหรับเสียงวรรณยุกต์ที่กำหนดให้หรือไม่ระหว่างเป็นธรรมชาติ และไม่เป็นธรรมชาติ

การวัดผลของแบบจำลอง จะมีการวัดผลตามสมการที่ 4.33

$$Accuracy(\%) = \frac{N_{consistent}}{N_{Total}} \times 100\% \qquad (4.33)$$

Accuracy (%) คือ ความแม่นยำของแบบจำลองเมื่อเปรียบเทียบกับการตัดสินใจของผู้ฟัง

$N_{consistent}$ คือ จำนวนเสียงกระตุ้นทั้งหมดที่คำตอบจากการตัดสินใจของผู้ฟังกับผลทำนายของแบบจำลองนั้นตรงกัน ของเสียงตัวกระตุ้นนั้น ๆ

N_{Total} คือ จำนวนเสียงตัวกระตุ้นทั้งหมดในการทดลองการรับรู้ (45 รูปแบบ x 2 ผู้พูด x ฟังซ้ำ 2 ครั้ง x จำนวนผู้ฟัง 10 คน)

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ผลการวัดผลแบบจำลอง ซึ่งแสดงเป็นอัตราส่วนของความแม่นยำเมื่อเทียบกับการรับรู้ของผู้ฟังของผู้พูดเพศหญิงและผู้พูดเพศชาย และแบ่งตามเสียงวรรณยุกต์ แสดงในตารางที่ 4.3

ตารางที่ 4.3 ผลความแม่นยำของแบบจำลองเมื่อเทียบกับการรับรู้ของผู้ฟัง

เสียงวรรณยุกต์	ความแม่นยำของแบบจำลองเมื่อเปรียบเทียบกับการตัดสินใจของผู้ฟัง (%)	
	ผู้พูดเพศหญิง	ผู้พูดเพศชาย
สามัญ	82.5	85
เอก	75	74
โท	75.26	79.74
ตรี	93	79.58
จัตวา	77.08	78.33
เฉลี่ย	79.78	78.33

4.5 วิเคราะห์ผลการทดลอง

วิเคราะห์ผลการสร้างแบบจำลองเรขาคณิต

จากภาพที่ 4.8 ความแตกต่างที่เห็นได้อย่างชัดเจนอันดับแรก ระหว่างเสียงที่เป็นธรรมชาติของเสียงผู้หญิง (ก) กับเสียงผู้ชาย (ข) คือช่วงของค่าความชันของครึ่งหลังในหน่วยตามพยางค์ ซึ่งเสียงผู้หญิงจะมีค่าที่กว้างกว่าผู้ชาย ทั้งนี้ เป็นไปตามข้อสรุปของ [32] ว่า การรับรู้ระดับเสียงนั้น มีความสัมพันธ์อย่างมาก กับตำแหน่งภายในขอบเขตความถี่มูลฐานของเสียงพูดของผู้พูด ไม่ใช่ค่าความถี่มูลฐานที่แท้จริง

สำหรับเสียงวรรณยุกต์สามัญ เสียงผู้หญิง ค่าตำแหน่งกึ่งกลางจะอยู่บริเวณช่วงกึ่งกลางของขอบเขตระดับเสียงของผู้พูด และค่าความชันจะเท่ากับ 0 หรือน้อยกว่า 0 เล็กน้อย กล่าวคือ โครงร่างความถี่มูลฐานจะเป็นเส้นตรง หรือโค้งลงเล็กน้อย จึงจะฟังเป็นธรรมชาติ ส่วนเสียงผู้ชายนั้น ค่าตำแหน่งกึ่งกลางจะอยู่บริเวณช่วงกึ่งกลางของขอบเขตระดับเสียงของผู้พูดเช่นเดียวกัน แต่ความชันจะปรากฏการชันขึ้นของโครงร่างเล็กน้อย เป็นเพราะว่า ช่วงขอบเขตระดับเสียงของผู้ชายนั้นต่ำ (100 – 160 Hz) และแคบกว่าเสียงผู้หญิง ดังนั้น ผู้ฟังอาจไม่สามารถแยกแยะความแตกต่างเพียงเล็กน้อยนี้ได้เนื่องจากการชันขึ้นเล็กน้อยของโครงร่างความถี่มูลฐาน ทำให้ผู้ฟัง ยังคงฟังเป็นธรรมชาติอยู่ แต่ทั้งนี้ เมื่อถึงระดับที่ผู้ฟังสามารถแยกแยะความแตกต่างของระดับเสียงนี้ได้แล้ว กลุ่มผู้ฟังจะเริ่มมีการตอบว่าไม่เป็นธรรมชาติในขั้นตอนการให้คะแนนระดังความเป็นธรรมชาติ หรือตอบเป็นเสียงวรรณยุกต์ตรีในขั้นตอนการระบุเสียงวรรณยุกต์

สำหรับเสียงวรรณยุกต์เอกนั้น รูปร่างเรขาคณิตของแบบจำลองทั้งเสียงผู้หญิงและผู้ชาย มีความคล้ายกัน โดยค่าตำแหน่งกึ่งกลางจะกระจายอยู่ตั้งแต่ช่วงล่างของขอบเขตระดับเสียง แต่ไม่ถึง

ด้านล่างสุด ถึงกึ่งกลางของขอบเขตระดับเสียง และโครงร่างความถี่มูลฐานจะโค้งลงเล็กน้อย (ค่าความชันน้อยว่า 0) เมื่อค่ากึ่งกลางอยู่ในตำแหน่งที่ต่ำลง ค่าความชันจะมีแนวโน้มเพิ่มไปในทิศทางบวกมากขึ้น หรือกล่าวอีกนัยหนึ่งคือ โครงร่างจะโค้งลงน้อยลงลงจนเรียบในบริเวณที่ค่าความถี่มูลฐานในตำแหน่งกึ่งกลางอยู่ในช่วงล่างของขอบเขตระดับเสียง และค่าความชันจะมีค่าน้อยกว่าเท่ากับ 0 เสมอ

สำหรับเสียงวรรณยุกต์โท ค่าความถี่มูลฐานที่ตำแหน่งกึ่งกลางจะมีค่าตั้งแต่ช่วงกึ่งกลางของช่วงระดับเสียงไปจนถึงบนสุดของช่วงระดับเสียง และโครงร่างจะโค้งลงได้มาก หรือค่าความชัน เป็นลบได้มากนั่นเอง และจะมีความโค้งลงได้น้อยลง เมื่อค่าความถี่มูลฐานที่ตำแหน่งกึ่งกลาง มีค่าต่ำลงภายในช่วงขอบเขตระดับเสียง แต่พบว่าค่าค่าความถี่มูลฐานที่ตำแหน่งกึ่งกลางที่ช่วงบนของขอบเขตระดับเสียง มีการโค้งขึ้นเล็กน้อย ที่ยังฟังเป็นธรรมชาติ ทั้งนี้เนื่องจากเป็นผลของการพูดในบริบทที่ต่อเนื่อง (Connected Speech)

สำหรับเสียงวรรณยุกต์ตรี ค่าความถี่มูลฐานที่ตำแหน่งกึ่งกลางจะมีค่าอยู่ในช่วงประมาณ 30%-70% ของช่วงขอบเขตระดับเสียง ทั้งเสียงผู้หญิงและผู้ชาย และมีค่าความชันเป็นบวกเสมอ โครงร่างความถี่มูลฐานจะสามารถโค้งขึ้นได้มาก หากค่าความถี่มูลฐานที่ตำแหน่งจุดกึ่งกลางอยู่บริเวณที่ต่ำกว่าช่วงกึ่งกลางของขอบเขตระดับเสียง และจะมีความชันจะมีแนวโน้มลดลงเมื่อค่าความถี่มูลฐานที่ตำแหน่งจุดกึ่งกลางเพิ่มขึ้น

สำหรับเสียงวรรณยุกต์จัตวา ค่าความถี่มูลฐานที่ตำแหน่งกึ่งกลางจะมีค่าอยู่ในช่วงล่างสุดของขอบเขตเสียงพูด ซึ่งจะอยู่ต่ำกว่าค่าความถี่มูลฐานที่ตำแหน่งกึ่งกลางของเสียงวรรณยุกต์เอก และโครงร่างความถี่มูลฐานจะโค้งขึ้นเล็กน้อยถึงมาก

ในภาพรวมนั้น ทั้งเสียงพูดผู้หญิงและผู้ชาย ไม่มีบริเวณของแบบจำลองเรขาคณิตของเสียงวรรณยุกต์ใดที่ทับซ้อนกัน ซึ่งแต่ละเสียงวรรณยุกต์มีบริเวณที่มีความเป็นธรรมชาติที่ชัดเจน นั่นหมายความว่า ในความสัมพันธ์ระหว่างค่าความถี่มูลฐานที่ตำแหน่งจุดกึ่งกลางของหน่วยตามพยางค์กับค่าความชันครึ่งหลังของหน่วยตามพยางค์ของแต่ละเสียงวรรณยุกต์มีความเฉพาะในแต่ละเสียงวรรณยุกต์ และไม่เกิดความกำกวมในการทำนายความเป็นธรรมชาติของแต่ละเสียงวรรณยุกต์

วิเคราะห์ผลการวัดผลแบบจำลองการรับรู้ความเป็นธรรมชาติ

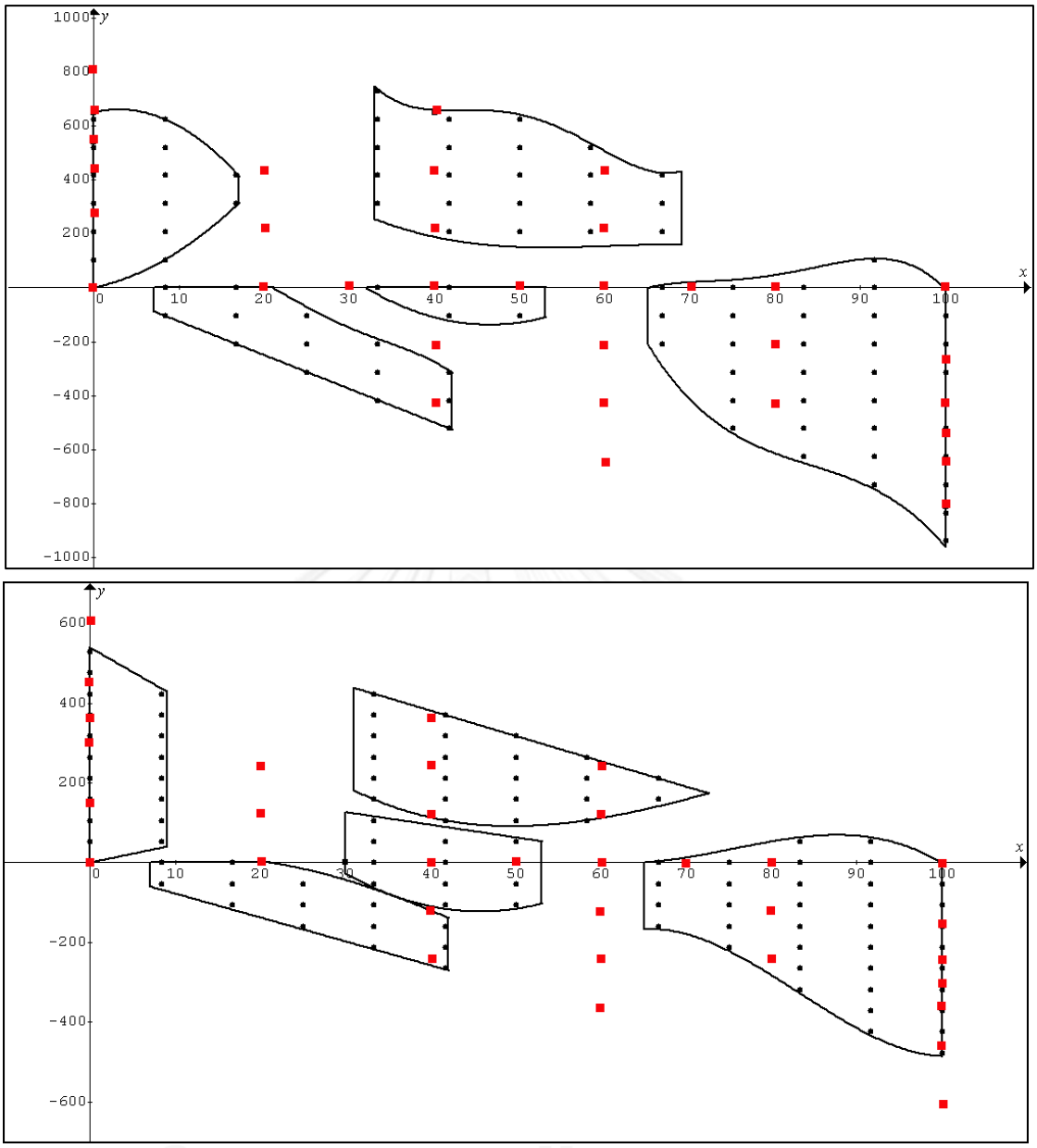
การวัดผลความถูกต้องของแบบจำลอง จะเป็นการเปรียบเทียบผลการทำนายความเป็นธรรมชาติที่ได้จากแบบจำลอง ซึ่งเรียนรู้ความสัมพันธ์จากการทดลองการรับรู้ความเป็นธรมชาติจากผู้พูดที่ 1 (เพศหญิง) และผู้พูดที่ 2 (เพศชาย) กับผลจากการตัดสินความเป็นธรรมชาติของผู้ฟังแต่ละคน ระหว่างความเป็นธรรมชาติ และความไม่เป็นธรรมชาติของเสียงวรรณยุกต์ ของพยางค์เป้าหมายในเสียงตัวกระตุ้นชุดทดสอบ ซึ่งความถูกต้องโดยรวมของแบบจำลองอยู่ที่ 79.78% สำหรับเสียงพูดผู้หญิง และ 78.33% สำหรับเสียงพูดผู้ชาย เมื่อเทียบกับการตัดสินใจของผู้ฟัง และเมื่อแยกตามเสียงวรรณยุกต์ จะได้ความถูกต้องที่ 82.5%, 75.0%, 75.26%, 93.0% และ 77.08% สำหรับเสียงวรรณยุกต์สามัญ, เอก, โท, ตรี และจัตวา ตามลำดับ ในเสียงพูดผู้หญิง และ 85.0%, 74.0%,

Chulalongkorn University

79.74%, 79.58% และ 78.33% สำหรับเสียงวรรณยุกต์สามัญ, เอก, โท, ตรี และจัตวา ตามลำดับ ในเสียงพูดผู้ชาย เมื่อพิจารณาในแต่ละผู้พูด สำหรับเสียงพูดผู้หญิง เสียงวรรณยุกต์ที่มีเปอร์เซ็นต์ความถูกต้องสูงที่สุด คือเสียงตรี และเสียงวรรณยุกต์ที่มีเปอร์เซ็นต์ความถูกต้องต่ำที่สุด คือเสียงต่ำและเสียงโท ซึ่งใกล้เคียงกัน สำหรับเสียงโทนั้น เมื่อพิจารณาลักษณะโครงร่างความถี่มูลฐานของเสียงกระตุ้นที่คำตอบของแบบจำลองกับการตัดสินใจของผู้ฟังที่มักจะไม่ตรงกันเป็นจำนวนมากนั้น ได้แก่ โครงร่างที่มีค่าความชันของครึ่งหลังของหน่วยตามพยางค์มาก ซึ่งในแบบจำลองนั้นตอบว่าไม่เป็นธรรมชาติ แต่ผู้ฟังที่ตอบว่าไม่เป็นธรรมชาติ ก็มีพอสมควร ฉะนั้นจึงยังมีความกำกวมอยู่ และไม่แน่ชัดถึงค่าจำกัดของความชัน สำหรับเสียงวรรณยุกต์เอก ค่าความชันที่แบบจำลองจะตัดสินให้มีความเป็นธรรมชาตินั้น แคบกว่าเสียงวรรณยุกต์อื่นๆ ซึ่งโครงร่างความถี่มูลฐานที่มีค่าลักษณะสำคัญที่อยู่ใกล้เคียงรูปเรขาคณิตเสียงเอก อาจยังคงถูกรับฟังเป็นธรรมชาติจากผู้ฟัง ดังนั้นหากทำการปรับแต่งแบบจำลองเพิ่มเติม อาจทำให้มีเปอร์เซ็นต์ความถูกต้องเพิ่มขึ้น และสำหรับเสียงพูดผู้ชายเสียงวรรณยุกต์ที่มีเปอร์เซ็นต์ความถูกต้องสูงที่สุด คือเสียงสามัญ ซึ่งเป็นเพราะค่าความชันของแบบจำลองมีช่วงของความเป็นธรรมชาติที่กว้าง ทั้งนี้ เป็นการได้ผลมาจากการทดลองการรับรู้ ซึ่งเมื่อทดสอบแบบจำลอง ผลการรับรู้จึงสอดคล้องกัน คือ ความกว้างของขอบเขตระดับเสียงของผู้ชายนั้นน้อย ทำให้เมื่อมีการเปลี่ยนแปลงค่าความชันเพียงเล็กน้อย ผู้ฟังจึงไม่สามารถแยกแยะได้นั่นเอง

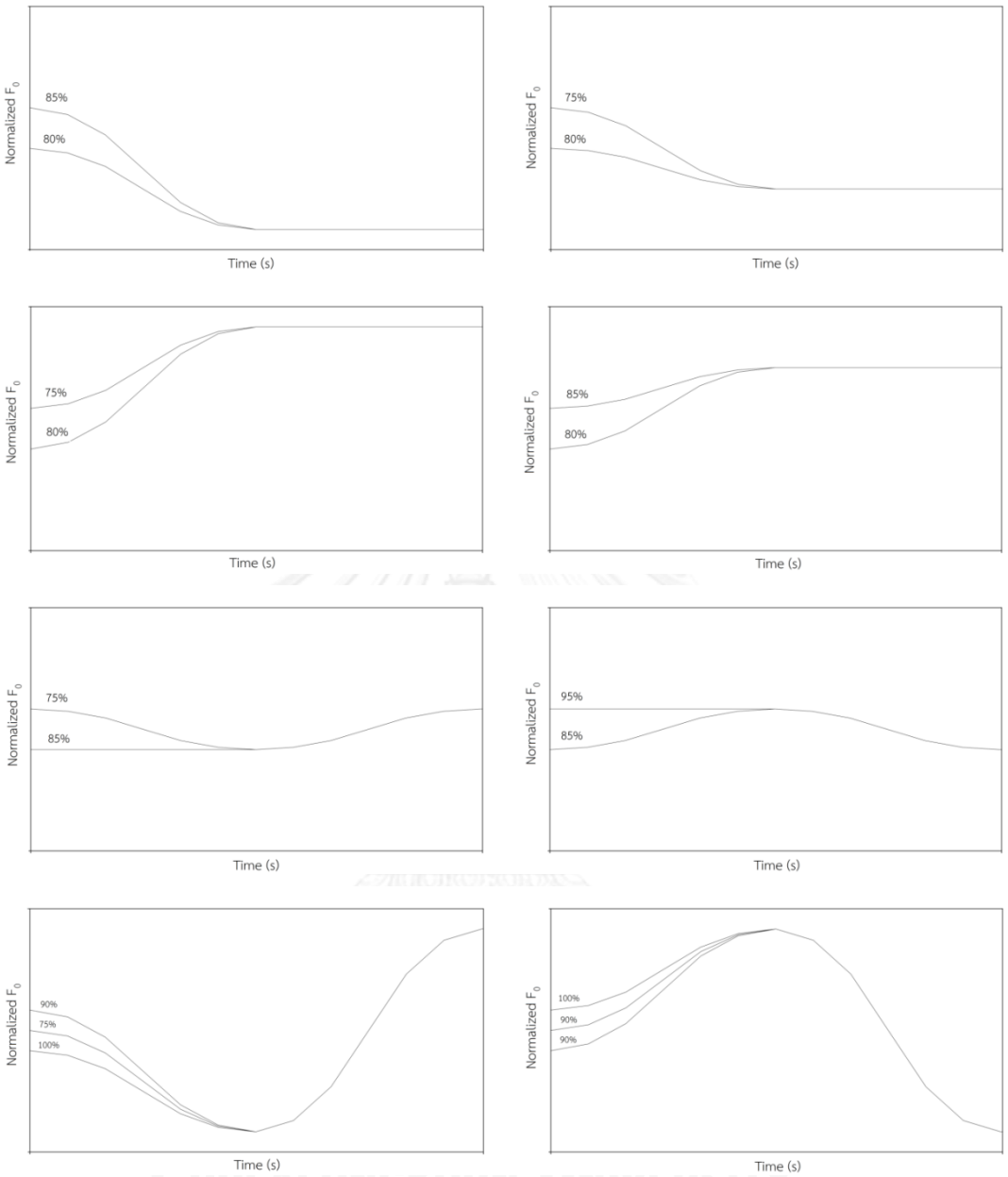
เมื่อนำค่านอมาไลซ์ของความถี่มูลฐานที่ตำแหน่งจุดกึ่งกลางและค่าความชันของเสียงตัวกระตุ้นชุดทดสอบ มากำหนดลงบนความสัมพันธ์ในแบบจำลองการรับรู้ความเป็นธรรมชาติของเสียงวรรณยุกต์ จะแสดงได้ดังภาพที่ 4.9 ภาพ 4.9 (บน) แสดงค่านอมาไลซ์ของความถี่มูลฐานที่ตำแหน่งจุดกึ่งกลางและค่าความชันของเสียงตัวกระตุ้นชุดทดสอบของผู้พูดเพศหญิง และภาพ 4.9 (ล่าง) แสดงค่านอมาไลซ์ของความถี่มูลฐานที่ตำแหน่งจุดกึ่งกลางและค่าความชันของเสียงตัวกระตุ้นชุดทดสอบของผู้พูดเพศชาย จุดสี่เหลี่ยม แสดงตำแหน่งภายในแบบจำลองเรขาคณิตของพารามิเตอร์เสียงตัวกระตุ้นชุดทดสอบ

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

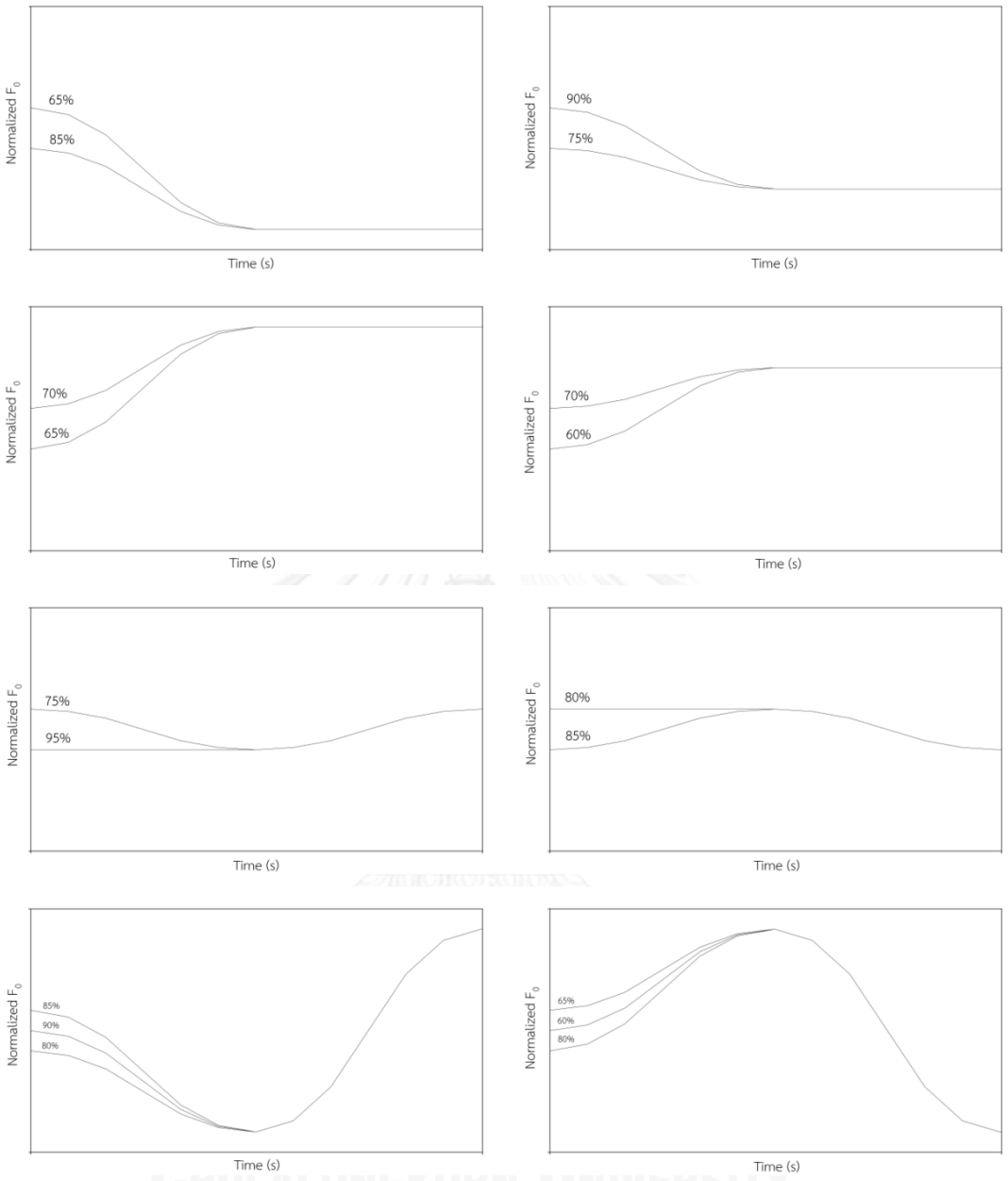


ภาพที่ 4.9 แสดงตำแหน่งของค่านอมาไลซ์ของความถี่มูลฐานที่ตำแหน่งจุดกึ่งกลางและค่าความชันของเสียงตัวกระตุ้นชุดทดสอบบนแบบจำลองการรับรู้ความเป็นธรรมชาติของเสียงวรรณยุกต์ของ ผู้พูดเพศหญิง (ภาพบน) ผู้พูดเพศชาย (ภาพล่าง)

เมื่อพิจารณาถึงโครงร่างความถี่มูลฐานที่ใช้ปรับแต่งเสียงตัวกระตุ้นในชุดทดสอบนั้น พบว่ามีรูปแบบของการปรับแต่งที่มีค่านอมาไลซ์ของความถี่มูลฐานที่ตำแหน่งกึ่งกลางและตำแหน่งสิ้นสุดเหมือนกันแต่มีค่านอมาไลซ์ของความถี่มูลฐานที่ตำแหน่งจุดเริ่มต้นต่างกันเล็กน้อย (ไม่เกิน 20% ของช่วงความถี่มูลฐานของผู้พูด) อยู่ 8 รูปแบบ เมื่อแยกพิจารณาเฉพาะ 8 รูปแบบ เพื่อหาผลกระทบของช่วงต้นพยางค์ (Onset) ที่มีต่อการรับรู้ความเป็นธรรมชาติของผู้ฟัง พบว่าค่าความถี่มูลฐานที่เปลี่ยนไปนั้น มีผลต่อการรับรู้ความเป็นธรรมชาติเล็กน้อย ดังแสดงในภาพที่ 4.10 และ ภาพที่ 4.11



ภาพที่ 4.10 จำนวนคำตอบที่ผู้ฟังตอบตรงกับแบบจำลองในเสียงตัวกระตุ้นที่มีจุดกึ่งกลางจุดสิ้นสุดเหมือนกัน แต่มี F0 onset ต่างกันของผู้พูดเพศหญิง



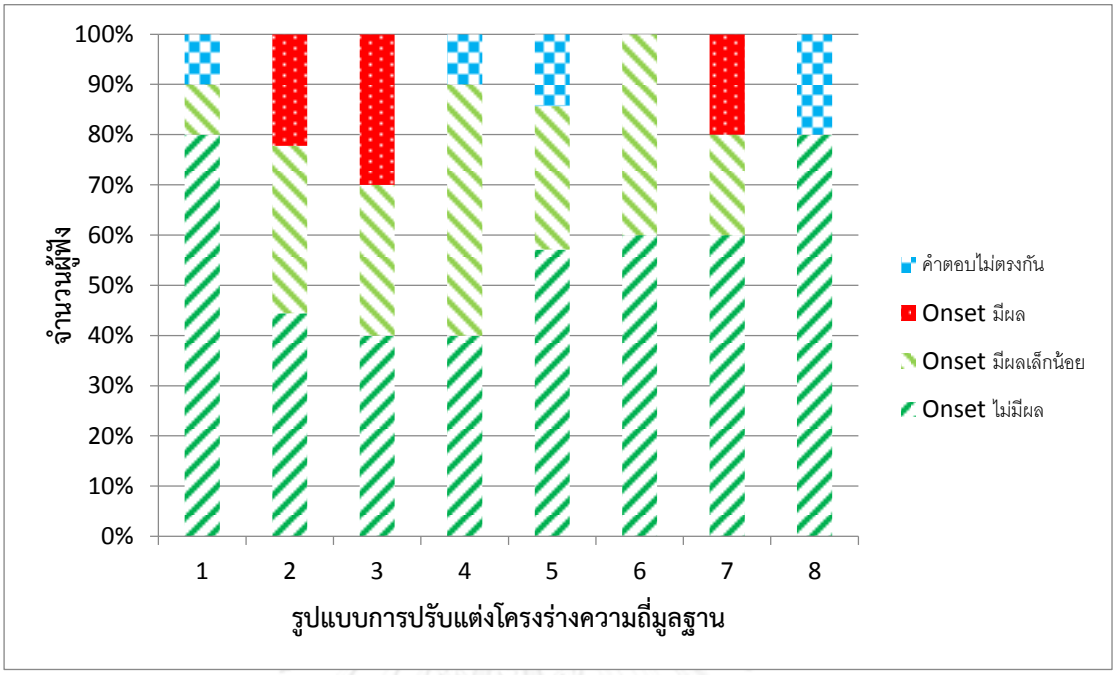
ภาพที่ 4.11 จำนวนคำตอบที่ผู้ฟังตอบตรงกับแบบจำลองในเสียงตัวกระตุ้นที่มีจุดกึ่งกลางจุดสิ้นสุดเหมือนกัน แต่มี F0 onset ต่างกันของผู้พูดเพศชาย

ซึ่งแสดงโครงร่างความถี่มูลฐานที่มีค่านอมาไลซ์ของความถี่มูลฐานที่ตำแหน่งกึ่งกลางและตำแหน่งสิ้นสุดเหมือนกันแต่มีค่านอมาไลซ์ของความถี่มูลฐานที่ตำแหน่งจุดเริ่มต้นต่างกันเล็กน้อยในชุดทดสอบ ตัวเลขเปอร์เซ็นต์แสดงถึงจำนวนคำตอบระหว่างเสียงชุดทดสอบที่มีค่านอมาไลซ์ของความถี่มูลฐานที่ตำแหน่งกึ่งกลางและตำแหน่งสิ้นสุดเหมือนกันแต่มีค่านอมาไลซ์ของความถี่มูลฐานที่ตำแหน่งจุดเริ่มต้นต่างกันเล็กน้อยของผู้ฟังทั้งหมดที่ต้องตรงกับคำตอบที่ได้จากแบบจำลอง สำหรับผู้พูดเพศหญิง จำนวนคำตอบมีความแตกต่างกันโดยเฉลี่ย 9.375% (จาก 20 คำตอบ) สำหรับผู้พูดเพศชาย จำนวนคำตอบระหว่างเสียงชุดทดสอบที่มีค่านอมาไลซ์ของความถี่มูลฐานที่ตำแหน่งกึ่งกลางและตำแหน่งสิ้นสุดเหมือนกันแต่มีค่านอมาไลซ์ของความถี่มูลฐานที่ตำแหน่งจุดเริ่มต้นต่างกันเล็กน้อยนั้น มีความแตกต่างกันโดยเฉลี่ย 13.125% (จาก 20 คำตอบ) ซึ่งบอกได้ว่าการที่เสียงตัวกระตุ้นมีค่านอมาไลซ์ของความถี่มูลฐานที่ตำแหน่งจุดเริ่มต้นต่างกันเล็กน้อยนั้น มีส่วนในการส่งผลกระทบต่อการรับรู้ความเป็นธรรมชาติของผู้ฟัง

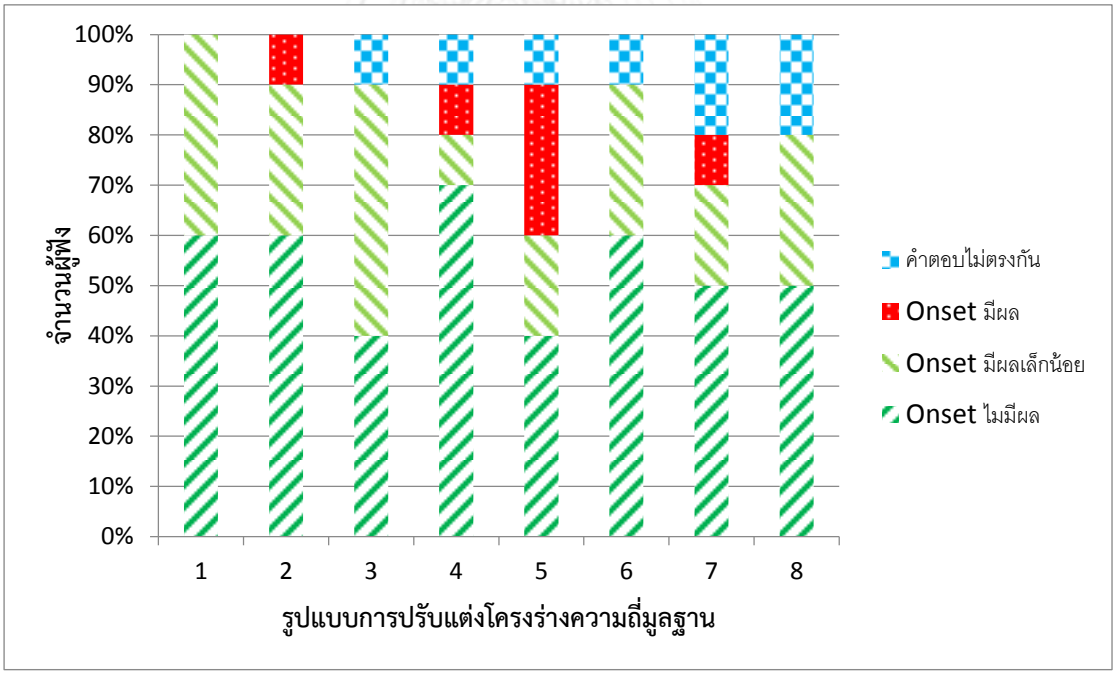
เมื่อพิจารณาในแง่ของค่าความถี่มูลฐานที่ตำแหน่งจุดเริ่มต้นนั้น ว่ามีผลกระทบต่อการรับรู้ของผู้ฟังหรือไม่ จึงได้พิจารณาผลการทดลองโดยสนใจเฉพาะการตัดสินใจของความเป็นธรรมชาติในการรับรู้เสียงตัวกระตุ้นของผู้ฟัง ดังภาพที่ 4.12

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

(ก) ผู้พูดเพศหญิง



(ข) ผู้พูดเพศชาย



ภาพที่ 4.12 จำนวนผู้ฟังที่ค่าออนเซ็ตมีผลต่อการรับรู้ในกลุ่มเสียงชุดทดสอบที่มีค่านอมาไลซ์ของความถี่มูลฐานที่ตำแหน่งกึ่งกลางและตำแหน่งสิ้นสุดเหมือนกันแต่มีค่านอมาไลซ์ของความถี่มูลฐานที่ตำแหน่งจุดเริ่มต้นต่างกันเล็กน้อย (ก) ผู้พูดเพศชาย (ข) ผู้พูดเพศหญิง

จากภาพที่ 4.12 จะแบ่งกลุ่มของผู้ฟังเป็น 4 กลุ่มคือ 1). กลุ่มผู้ฟังที่ออนเซ็ตไม่มีผลต่อการรับรู้ของผู้ฟัง ได้แก่ กลุ่มผู้ฟังที่ตอบว่า ระหว่างเสียงตัวกระตุ้นที่มีค่านอมาไลซ์ของความถี่มูลฐานที่ตำแหน่งกึ่งกลางและตำแหน่งสิ้นสุดเหมือนกันแต่มีค่านอมาไลซ์ของความถี่มูลฐานที่ตำแหน่งจุดเริ่มต้นต่างกันเล็กน้อยนั้น เป็นธรรมชาติหรือไม่เป็นธรรมชาติเหมือน ๆ กัน 2). กลุ่มผู้ฟังที่ออนเซ็ตมีผลต่อการรับรู้ของผู้ฟังเล็กน้อย ได้แก่ กลุ่มผู้ฟังที่ตอบความเป็นธรรมชาติเหมือนกัน 3 ใน 4 ครั้งของจำนวนเสียงตัวกระตุ้นที่มีค่านอมาไลซ์ของจุดกึ่งกลางและจุดสิ้นสุดตรงกันทั้งหมด 3.) กลุ่มผู้ฟังที่ออนเซ็ตมีผลต่อการรับรู้ ได้แก่ กลุ่มผู้ฟังที่ตอบว่า ระหว่างเสียงตัวกระตุ้นที่มีค่านอมาไลซ์ของความถี่มูลฐานที่ตำแหน่งกึ่งกลางและตำแหน่งสิ้นสุดเหมือนกันแต่มีค่านอมาไลซ์ของความถี่มูลฐานที่ตำแหน่งจุดเริ่มต้นต่างกันเล็กน้อยนั้น เป็นธรรมชาติหรือไม่เป็นธรรมชาติต่างกัน 4.) กลุ่มผู้ฟังที่ตัดสินความเป็นธรรมชาติไม่เหมือนกันในเสียงเดียวกัน ทั้ง 2 เสียงตัวกระตุ้นที่มีค่านอมาไลซ์ของความถี่มูลฐานที่ตำแหน่งกึ่งกลางและตำแหน่งสิ้นสุดเหมือนกันแต่มีค่านอมาไลซ์ของความถี่มูลฐานที่ตำแหน่งจุดเริ่มต้นต่างกัน จะเห็นได้ว่า จำนวนผู้ฟังที่อยู่ในกลุ่มผู้ฟังที่ออนเซ็ตมีผลนั้น ปรากฏให้เห็นอยู่เล็กน้อย ดังนั้นค่าออนเซ็ตอาจเป็นปัจจัยหนึ่ง ที่มีผลต่อการรับรู้ความเป็นธรรมชาติของผู้ฟัง

ทั้งนี้ เมื่อวัดผลความถูกต้องโดยรวม ซึ่งมีความถูกต้องประมาณเกือบ 80% นั้น สามารถนำไปประยุกต์ใช้เพื่อช่วยในการประเมินความเป็นธรรมชาติของเสียงวรรณยุกต์ และเป็นข้อจำกัดของรูปแบบโครงร่างความถี่มูลฐาน ในระบบการปรับแต่งโครงร่างความถี่มูลฐานโดยผู้ใช้ ให้คงความเป็นธรรมชาติในการรับรู้ได้ หรือเพื่อเป็นข้อจำกัดในการสร้างโครงร่างความถี่มูลฐานของพยางค์ ในระบบสังเคราะห์เสียงให้มีความเป็นธรรมชาติได้

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

บทที่ 5
บทสรุปผลการวิจัยและข้อเสนอแนะ

5.1 สรุปผลการวิจัย

วิทยานิพนธ์ฉบับนี้ได้เสนอวิธีการสร้างแบบจำลองของการรับรู้ความเป็นธรรมชาติของเสียงวรรณยุกต์ภาษาไทย จากความสัมพันธ์ระหว่างค่าความถี่มูลฐานที่ถูกทำให้เป็นบรรทัดฐานที่ตำแหน่งจุดกึ่งกลางของช่วงหน่วยตามพยางค์ กับค่าความชันของค่าความถี่มูลฐานระหว่างจุดกึ่งกลางและจุดสิ้นสุดของช่วงหน่วยตามพยางค์ ในโครงร่างความถี่มูลฐาน ซึ่งเป็นช่วงที่มีผลต่อลักษณะโครงร่างความถี่มูลฐานของแต่ละเสียงวรรณยุกต์ โดยศึกษาการทดลองการรับรู้ความเป็นธรรมชาติในการให้คะแนนระดับของความเป็นธรรมชาติจากกลุ่มผู้ฟัง ซึ่งพบว่ากลุ่มผู้ฟังสามารถแยกแยะระดับความเป็นธรรมชาติของพยางค์เป้าหมาย ของเสียงตัวกระตุ้น ซึ่งถูกปรับแต่งโครงร่างความถี่มูลฐานโดยการกำหนดค่าความถี่มูลฐานที่ตำแหน่งจุดกึ่งกลางและค่าความถี่มูลฐานที่ตำแหน่งจุดสิ้นสุด อย่างเป็นระบบ โดยมีรูปแบบในการปรับแต่งทั้งหมด 169 แบบ ภายในช่วงขอบเขตความถี่มูลฐานของผู้พูด ทั้งผู้พูดที่เป็นเพศหญิงและเพศชาย ออกเป็นระดับต่าง ๆ ได้ และได้นำผลจากการพิจารณาศึกษาการรับรู้ความเป็นธรรมชาตินั้น มาสร้างเป็นแบบจำลองการรับรู้ความเป็นธรรมชาติของเสียงวรรณยุกต์ของแต่ละเพศของเสียงผู้พูดและเสียงวรรณยุกต์ ด้วยแบบจำลองเรขาคณิต ซึ่งเป็นการกำหนดบริเวณปิดภายในแผนภาพการกระจายของความสัมพันธ์ระหว่างค่าความถี่มูลฐานที่ถูกทำให้เป็นบรรทัดฐานที่ตำแหน่งจุดกึ่งกลางของช่วงหน่วยตามพยางค์ กับค่าความชันของค่าความถี่มูลฐานระหว่างจุดกึ่งกลางและจุดสิ้นสุดของช่วงหน่วยตามพยางค์ ซึ่งใช้ในการตัดสินความเป็นธรรมชาติ และจากการวัดประสิทธิภาพของแบบจำลองโดยการเปรียบเทียบผลลัพธ์ที่ได้จากการทำนายของแบบจำลอง กับการตัดสินใจจากการรับรู้ของกลุ่มผู้ฟัง ด้วยเสียงตัวกระตุ้นในชุดทดสอบ ซึ่งมีรูปแบบของโครงร่างความถี่มูลฐานซึ่งอ้างอิงกับรูปแบบของเส้นโค้งแบบต่าง ๆ โดยสามารถมีจุดสะท้อนได้หนึ่งจุด ทั้งหมด 45 รูปแบบ แบบจำลองสามารถทำนายผลความเป็นธรรมชาติได้ตรงกับการตัดสินใจของกลุ่มผู้ฟังประมาณ 80%

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

5.2 ข้อจำกัด

ข้อจำกัดด้านการทดลอง

- เนื่องจากข้อมูลการรับรู้ความเป็นธรรมชาติที่ใช้ในการทดลอง ขึ้นกับการทดลองการรับรู้ของผู้ฟัง ดังนั้นจึงมีความจำกัดในรูปแบบต่าง ๆ ของโครงร่างความถี่มูลฐานที่ใช้ในการทดลองการรับรู้ ให้อยู่ในปริมาณที่เหมาะสม
- ในขั้นตอนการทดลองการระบุเสียงวรรณยุกต์และการรับรู้ความเป็นธรรมชาติ ผู้ฟังจะได้ฟังและตอบจากเสียงตัวกระตุ้นนั้น ๆ 1 ครั้ง เนื่องจากหากเพิ่มจำนวนเป็น 2 ครั้งต่อ 1 เสียงตัวกระตุ้น เพื่อเพิ่มความมั่นใจของข้อมูลนั้น จะทำให้เสียงในชุดพัฒนามีจำนวนมาก ผู้ฟังไม่สามารถตอบคำถามได้จนจบการทดลอง

- ในขั้นตอนการแบ่งขั้นในการปรับแต่งค่าจุดกึ่งกลางและจุดสิ้นสุดนั้น เมื่อเพิ่มความละเอียดมากขึ้น จะทำให้แบบจำลองมีความละเอียดเพิ่มขึ้น แต่ปริมาณเสียงตัวกระตุ้น จะเพิ่มขึ้นมาก จึงต้องควบคุมปริมาณเสียงตัวกระตุ้นได้เหมาะสม
- ในการออกแบบกรอบประโยคในการทดลอง มีการจำกัดพยางค์ที่อยู่หน้าพยางค์เป้าหมายเป็นเสียงวรรณยุกต์สามัญ แต่โดยทั่วไปนั้น กรอบประโยคสามารถมีได้หลากหลาย ไม่จำกัด หากเพิ่มความหลากหลายของเสียงวรรณยุกต์ ครบ ทั้ง 5 เสียง ปริมาณเสียงตัวกระตุ้นจากเดิม 206 เสียงในชุดพัฒนา จะมีจำนวน 1030 เสียง ซึ่งผู้ฟังไม่สามารถทำการทดลองได้ในเวลาที่จำกัด ทั้งนี้ รวมไปถึงพยัญชนะและสระที่เปลี่ยนไปของพยางค์เป้าหมายเช่นกัน

ข้อจำกัดด้านแบบจำลอง

- แบบจำลองการรับรู้ความเป็นธรรมชาติของเสียงวรรณยุกต์นี้ ศึกษาการรับรู้จากพยางค์ในคำพูดต่อเนื่องในภายบริบทที่ถูกจำกัดไว้ โดยมีพยางค์ก่อนหน้าเป็นเสียงสามัญ จึงสามารถนำแบบจำลองนี้ไปใช้ได้ กับพยางค์ที่มีพยางค์ก่อนหน้าเป็นเสียงวรรณยุกต์สามัญได้ตามผลการทดลอง
- แบบจำลองนี้ มีแนวโน้มที่อาจใช้ได้กับบริบทที่มีพยางค์ก่อนหน้าเป็นเสียงวรรณยุกต์อื่น ๆ แต่อาจมีความถูกต้องแม่นยำที่ไม่ตรงกับที่ระบุไว้ในผลการทดลอง
- แบบจำลองนี้สามารถประยุกต์ใช้กับพยางค์เป้าหมายที่มีสระต่าง ๆ ได้
- เนื่องจากตัวสะกดนั้น มีผลต่อโครงร่างความถี่มูลฐานโดยรวมของพยางค์ ในการใช้แบบจำลองนี้จึงควรระวัง เมื่อพยางค์เป้าหมายมีตัวสะกดแบบเสียงกัก (Plosive Stop Consonant) เนื่องจากส่งผลต่อโครงร่างความถี่มูลฐานที่แตกต่างจากพยางค์ที่ไม่มีตัวสะกด ซึ่งอาจเกิดการรับรู้ที่ต่างกัน
- ขอบเขตของความเป็นธรรมชาติของเสียงวรรณยุกต์ในแบบจำลองการรับรู้ความเป็นธรรมชาตินั้น คำนวณโดยผู้วิจัย ซึ่งอ้างอิงกับผลการทดลองการรับรู้ เพื่อหาความสัมพันธ์

5.3 ข้อเสนอแนะ

แบบจำลองการทำนายการรับรู้ความเป็นธรรมชาติของเสียงวรรณยุกต์นี้ สร้างขึ้นจากการศึกษาความเป็นธรรมชาติของเสียงวรรณยุกต์ด้วยการทดลองการรับรู้ ซึ่งสามารถสะท้อนการตัดสินใจของคนได้เป็นอย่างดี ทั้งนี้ การศึกษาเป็นการศึกษารูปแบบโครงร่างความถี่มูลฐานของพยางค์เป้าหมายที่มีพยางค์ก่อนหน้า และพยางค์ถัดไป เป็นเสียงวรรณยุกต์สามัญเท่านั้น ซึ่งการเพิ่มประสิทธิภาพของแบบจำลองให้ดีขึ้นนั้น ควรศึกษาในกรณีเสียงวรรณยุกต์ของพยางค์ก่อนหน้าและพยางค์ถัดไปเป็นเสียงวรรณยุกต์อื่นเพิ่มเติม ซึ่งเป็นปัจจัยที่เกิดผลกระทบต่อโครงร่างความถี่มูลฐาน

ของพยางค์ ที่เรียกว่า แคร์รี่โอเวอร์ เอฟเฟกท์ ซึ่งมีผลต่อช่วงต้นของโครงร่างความถี่มูลฐานของพยางค์เป้าหมายเนื่องจากเสียงวรรณยุกต์ก่อนหน้า และเอนติซิพาทอรี่ เอฟเฟกท์ (Anticipatory Effect) ซึ่งมีผลต่อช่วงท้ายของโครงร่างความถี่มูลฐานของพยางค์เป้าหมายเนื่องจากเสียงวรรณยุกต์ถัดไป และเสียงตัวสะกดที่เป็นเสียงกักเพิ่มเติม เนื่องจากอาจส่งผลกระทบต่อการรับรู้ความเป็นธรรมชาติของเสียงวรรณยุกต์ได้ ทั้งนี้เพื่อเพิ่มความสมบูรณ์ของแบบจำลอง อีกทั้งการเพิ่มจำนวนของเสียงผู้พูดในการศึกษาการรับรู้ความเป็นธรรมชาตินั้น จะช่วยเพิ่มความน่าเชื่อถือของแบบจำลองมากขึ้น แต่เนื่องจากข้อเสียของวิธีการศึกษาจากการทดลองการรับรู้นั้น ใช้เวลาเป็นจำนวนมาก และมีหลายปัจจัยที่ต้องคำนึง ได้แก่ ปริมาณเสียงตัวกระตุ้นที่มากเกินไป เนื่องจากการต้องการความละเอียดของข้อมูล หรือแม้กระทั่งการเพิ่มลักษณะสำคัญที่ใช้ในการทดลอง ปัจจัยต่าง ๆ เหล่านี้ มีผลอย่างมากต่อจำนวนเสียงตัวกระตุ้นที่ใช้ในการทดลองการรับรู้ ทำให้ผู้ฟัง ไม่สามารถทำการทดลองได้ทั้งหมด จึงต้องทำการลดจำนวนเสียงตัวกระตุ้นลง แต่ยังคงไว้ซึ่งความคลอบคลุมถึงวัตถุประสงค์ที่ต้องการศึกษา อีกทั้งการทดลองนำข้อมูลที่ได้จากการศึกษาความเป็นธรรมชาติของเสียงวรรณยุกต์ ไปทดลองสร้างแบบจำลองด้วยวิธีต่าง ๆ เพิ่มเติม เช่น แบบจำลองความน่าจะเป็น (Probabilistic Model) อาจทำให้ความถูกต้องแม่นยำของแบบจำลองที่เพิ่มขึ้น และการนำแบบจำลองการรับรู้ความเป็นธรรมชาติของเสียงวรรณยุกต์นี้ ไปประยุกต์ใช้กับระบบสังเคราะห์เสียง จะสามารถช่วยเพิ่มความถูกต้องและเป็นธรรมชาติของเสียงสังเคราะห์ในแง่ของโครงร่างความถี่มูลฐานให้เป็นธรรมชาติมากขึ้น หรือแม้กระทั่งนำไปใช้ในการประเมินความเป็นธรรมชาติหรือกำหนดขอบเขตในการปรับแต่งโครงร่างความถี่มูลฐานที่ถูกปรับแต่งโดยคนต่อไป

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

รายการอ้างอิง

[1] *NonVisual Desktop Access (NVDA),* http://www.nvaccess.org/

[2] Viswanathan, M. and M. Viswanathan, *Measuring speech quality for text-to-speech systems: development and assessment of a modified mean opinion score (MOS) scale.* Computer Speech & Language, 2005. 19(1): p. 55-83.

[3] Bulyko, I., M. Ostendorf, and P. Price, *On The Relative Importance Of Different Prosodic Factors For Improving Speech Synthesis.* In Proc. of ICPhS '99, 1999: p. 81-84.

[4] Abramson, A.S., *The Tones of Central Thai: Some Perceptual Experiments.* J. G. Harris & J. R. Chamberlain (Eds.), Studies in Thai linguistics in honor of William J. Gedney, 1975: p. 1-16.

[5] Potisuk, S., J.T. Gandour, and M.P. Harper, *F0 Correlates of Stress in Thai.* Linguistics of the Tibeto-Burman Area, 1994(17(2)): p. 1-27.

[6] Hirose, K., H. Hashimoto, J. Ikeshima, and N. Minematsu, *Use of generation process model for synthesizing fundamental frequency contours in HMM-based speech synthesis.* Signal Processing (ICSP) ,IEEE 11th International Conference on 2012. 1: p. 575-578.

[7] Li, Y., T. Lee, and Y. Qian, *F0 analysis and modelling for Cantonese text-to-speech.* Speech Prosody, 2004: p. 467-470.

[8] Mittrapiyanurak, P., C. Hansakunbuntheung, V. Tesprasit, and V. Sornlertlamvanich, *Improving naturalness of Thai text-to-speech synthesis by prosodic rule.* ISCA, 2000: p. 334-337.

[9] Suphattharachai, C. and K. Takao, *Tone correctness improvement in speaker-independent average-voice-based Thai speech synthesis.* Speech Communication, 2009(4): p. 330-343.

[10] อักษรไทยและการผันวรรณยุกต์, http://www.st.ac.th/bhatips/.

[11] Mittrapiyanurak, P., C. Harnsakunbuntheung, V. Tesprasit, and V. Sornlertlamvanich, *Issues in Thai text-to-speech synthesis: the NECTEC approach.* NECTEC Annual Conference, Bangkok, 2000: p. 483-495.

[12] Mixdorff, H., S. Luksaneeyanawin, H. Fujisaki, and P. Charnvivit, *Perception of tone and vowel quality in Thai.* ICSLP, 2002: p. 753-756.

[13] ระบบสังเคราะห์เสียงพูดภาษาไทย "วาจา", http://vaja.nectec.or.th/.

[14] *Speech Synthesis,* http://en.wikipedia.org/wiki/Speech_synthesis.

[15] Klatt, D., *How Klattalk became DECtalk: An Academic's Experiences in the Business World.* The Official Proceedings of Speech Tech '87 (New York: Media Dimensions/Penn State), 1987.

[16] เดชะอำไพ, ป. and น. วรรณโสภาคย์, ระเบียบวิธีเชิงตัวเลขในงานวิศวกรรม. สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย, 2008: p. 679.

[17] Phanintra, T. and R. Rungwimon, *Change in the Standard Thai High Tone: An Acoustic Study.* MANUSYA: Journal of Humanities, 2009(Special Issue No. 17): p. 34-44.

[18] Thubthong, N. and B. Kijsrikul, *Tone recognition of continuous Thai speech under tonal assimilation and declination effects using halftone model.* Electrical and Computer Engineering, 2004.

[19] Xu, Y., *Production and Perception of Coarticulated Tones.* Journal of the Acoustical Society of America, 1994(95(4)): p. 2240-2253.

[20] Xu, Y., *Contextual Tonal Variations in Mandarin.* Journal of Phonetics, 1997(25): p. 61-83.

[21] Abramson, A.S., *The coarticulation of tones: An acoustic study of Thai.* 1979.

[22] Thubthong, N., Kijsirikul, B., *Tone Recognition Of Continuous Thai Speech Under Tonal Assimilation And Declination Effects Using Half-Tone Model.* International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 2001: p. 815-825.

[23] Zsiga, E. and R. Nitisaroj, *Tone Features, Tone Perception, and Peak Alignment in Thai.* Language and Speech, 2007. 50(3): p. 343-383.

[24] Morén, B. and E.C. Zsiga, *The lexical and post-lexical phonology of Thai tones.* Natural Language and Linguistic Theory, 2006: p. (24)1: 113-178.

[25] Ramadoss, D., *Probabilistic Modelling of Tone Perception: Autosegmental 'target' are insufficient.* Proceedings of the Linguistic Society of America LSA, 2011.

[26] Ramadoss, D. and P. Smolensky, *Tone perception cues: Pitch targets, trajectories or both?* Acoustical Society of America, 2011.

[27] Charpentier, F. and E. Moulines, *Pitch synchronous waveform processing techniques for text-to-speech synthesis using diphones.* European Conference on Speech Communication and Technology. **I**: p. 013-019.

[28] Grether, C. and R. Stroh, *Subjective evaluation of differential pulse-code modulation using the speech "Goodness" rating scale.* Audio and Electroacoustics, IEEE Transactions on, 1973. 21(3): p. 179-184.

[29] Guilford, J.B., *Psychometric Method.* New York: McGraw-Hill, 1954.

[30] Eriksson, H.T.A., *The frequency range of the voice fundamental in the speech of male and female adults.* Department of Linguistics, University of Stockholm, 1994. 97: p. 1905191--5.

[31] Boersma, P. and D. Weenink, *Praat* 5.3.48*, A system for doing phonetics by computer.* 2013.

[32] Honorof, D.N. and D.H. Whalen, *Perception of pitch location within a speaker's F*0 *range.* The Journal of the Acoustical Society of America, 2005(0001-4966).

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ภาคผนวก

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ภาคผนวก ก

ตารางเปรียบเทียบผลการทำนายความเป็นธรรมชาติของเสียงวรรณยุกต์ด้วยแบบจำลองกับการตัดสินใจของผู้ฟัง

Model คือ ค่าที่ทำนายจากแบบจำลอง: nat = เป็นธรรมชาติ, unnat = ไม่เป็นธรรมชาติ

Listener คือ ผู้ฟัง : 0 = ผู้ฟังตอบไม่ตรงกับผลการทำนาย , 1 = ผู้ฟังตอบตรงกับผลการทำนาย

1. เสียงผู้พูดเพศชาย

F0 Pattern	model	Listener																				sum
		1		2		3		4		5		6		7		8		9		10		
bell_40_100_40	nat	1	1	1	1	1	1	1	0	1	1	0	1	0	0	0	1	1	1	1	0	14
bell_40_60_40+F	unnat	1	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	17
bell_40_80_40+F	nat	1	1	1	1	1	0	1	1	1	1	0	0	0	0	0	1	1	1	1	1	14
bell_60_100_60+F	nat	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	19
bell_60_80_60+F	nat	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	19
fall1St2_40_0_0+R	nat	1	1	1	1	1	1	0	0	1	1	1	1	0	1	1	1	1	1	1	1	17
fall1St2_40_20_20+L	nat	0	0	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1	15
fall1St2_60_0_0+R	nat	1	0	0	1	1	1	0	0	1	1	1	0	0	0	1	1	1	1	1	1	13
fall1St2_60_20_20+L	nat	1	1	1	1	0	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	18
fall1St2_60_40_40+M	nat	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	19
rise1St2_40_100_100+F	nat	0	0	1	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1	0	1	13
rise1St2_40_60_60+F	unnat	0	0	1	0	1	1	1	1	0	1	1	0	0	1	0	0	1	1	1	1	12
rise1St2_40_80_80+F	nat	0	1	1	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	14
rise1St2_60_100_100+F	nat	1	0	0	1	1	0	1	0	1	1	1	1	1	0	1	1	1	1	0	1	14
rise1St2_60_80_80+F	nat	1	1	0	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1	0	0	13
sFall1_50_0_100+R	unnat	1	1	1	0	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	18
sFall1_50_0_25+R	nat	1	1	1	1	0	0	1	0	1	1	1	1	1	1	1	1	1	1	1	1	17
sFall1_50_0_50+R	nat	1	1	1	1	0	0	1	1	1	1	1	1	0	1	0	1	1	1	0	0	14
sFall1_50_0_75+R	nat	1	0	1	1	0	0	1	1	1	1	1	1	0	1	1	1	1	1	1	1	16
sFall2_40_100_0+F	unnat	1	0	1	1	1	0	1	1	1	0	1	1	1	1	1	0	1	1	1	1	16
sFall2_60_100_0+F	unnat	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	13
sRise1_50_100_0+F	unnat	0	0	0	0	0	0	1	0	0	1	1	1	1	1	1	1	1	1	1	1	12
sRise1_50_100_25+F	nat	1	1	1	1	1	1	1	1	1	1	0	1	0	0	0	0	1	1	1	1	15
sRise1_50_100_50+F	nat	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1	1	1	1	17
sRise1_50_100_75+F	nat	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	19
sRise2_40_0_100+R	unnat	1	1	1	1	1	1	0	0	1	1	1	0	1	1	1	1	1	0	1	1	16
sRise2_60_0_100+R	unnat	1	1	1	1	1	1	1	0	1	1	1	1	1	1	0	1	0	1	1	1	17
st1Fall2_40_40_0+L	nat	1	0	1	1	1	1	0	1	1	0	0	0	0	1	0	1	1	1	1	1	13
st1Fall2_40_40_20+L	nat	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0	0	0	0	1	0	14
st1Fall2_60_60_0+L	unnat	1	0	0	0	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	0	15
st1Fall2_60_60_20+F	unnat	0	0	1	0	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	16

F0 Pattern	model	Listener																				sum
		1		2		3		4		5		6		7		8		9		10		
st1Rise2_40_40_100+H	nat	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	19
st1Rise2_40_40_60+H	nat	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	19
st1Rise2_40_40_80+H	nat	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	20
st1Rise2_60_60_100+H	nat	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	18
st1Rise2_60_60_80+H	nat	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	17
straight_30_30_30+M	nat	1	1	1	1	0	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	18
straight_50_50_50+M	nat	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	18
straight_70_70_70+F	nat	1	0	0	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	0	13
u_40_0_40+R	nat	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
u_40_20_40+R	unnat	1	0	0	1	1	1	0	0	1	0	0	0	0	1	0	0	1	1	0	1	9
u_60_0_60+R	nat	0	0	1	1	1	0	1	1	0	1	1	1	0	1	1	1	1	1	0	0	13
u_60_20_60+R	unnat	1	1	0	1	1	1	1	1	1	1	1	0	1	1	0	1	1	1	1	1	17
u_60_40_60+M	nat	0	1	0	1	1	1	0	0	0	0	1	1	0	0	1	1	1	0	1	1	11
	%	71		71		68		70		90		86		76		80		93		79		705

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

2. เสียงผู้พูดเพศหญิง

F0 Pattern	model	Listener										sum
		1	2	3	4	5	6	7	8	9	10	
bell_40_100_40+F	nat	0 1	1 1	0 1	0 0	1 1	1 1	1 1	0 0	1 1	0 1	13
bell_40_60_40+F	unnat	0 1	1 1	0 1	1 1	0 1	1 1	1 1	1 1	1 1	1 1	17
bell_40_80_40+F	nat	1 0	1 0	1 1	0 0	0 0	0 0	1 0	0 0	0 1	0 1	7
bell_60_100_60+F	nat	1 1	1 1	1 1	1 1	1 1	0 1	1 1	1 1	1 1	1 1	19
bell_60_80_60+F	nat	1 1	1 1	1 1	1 1	1 1	0 1	1 1	1 1	1 1	1 0	18
fall1St2_40_0_0+R	nat	1 1	1 0	0 1	0 0	1 1	1 1	1 1	1 1	1 1	1 1	16
fall1St2_40_20_20+L	nat	0 1	1 1	1 1	1 1	1 1	0 0	1 0	1 1	1 1	1 1	16
fall1St2_60_0_0+R	nat	1 1	1 1	0 1	0 0	1 1	1 1	1 1	1 1	1 1	1 1	17
fall1St2_60_20_20+L	nat	0 1	1 1	0 0	1 1	0 1	1 1	1 1	1 1	1 1	0 1	15
fall1St2_60_40_40+M	nat	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	20
rise1St2_40_100_100+F	nat	1 1	1 1	1 1	1 1	1 1	1 1	1 1	0 0	0 0	1 1	16
rise1St2_40_60_60+F	unnat	1 1	1 1	1 1	1 1	1 0	1 1	1 0	1 1	0 0	1 1	16
rise1St2_40_80_80+F	nat	1 1	0 1	1 1	1 1	1 1	1 0	0 1	1 1	1 1	1 0	16
rise1St2_60_100_100+F	nat	1 1	0 1	0 0	1 0	1 1	1 1	1 1	1 1	1 1	1 0	15
rise1St2_60_80_80+F	nat	1 1	1 1	1 0	1 0	1 1	1 1	1 1	1 1	1 1	1 0	17
sFall1_50_0_100+R	unnat	1 1	1 0	1 1	1 1	0 0	1 1	0 1	1 0	1 1	1 1	15
sFall1_50_0_25+R	nat	1 1	1 1	0 1	1 1	0 1	1 1	1 1	1 1	1 1	1 1	18
sFall1_50_0_50+R	nat	1 1	1 1	0 0	1 1	1 1	1 0	1 0	1 1	0 1	1 0	14
sFall1_50_0_75+R	unnat	1 1	0 1	1 1	0 0	0 1	1 1	1 1	0 1	1 1	1 1	15
sFall2_40_100_0+F	unnat	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 0	1 0	18
sFall2_60_100_0+F	unnat	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	20
sRise1_50_100_0+F	unnat	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 0	1 0	18
sRise1_50_100_25+F	nat	0 0	1 1	0 1	0 0	1 1	0 0	0 0	1 0	1 1	0 1	9
sRise1_50_100_50+F	nat	1 1	1 1	1 0	0 1	1 1	1 1	0 1	0 1	1 1	1 0	15
sRise1_50_100_75+F	nat	1 1	1 0	1 1	1 1	1 1	0 1	1 1	1 1	1 1	1 1	18
sRise2_40_0_100+R	unnat	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	20
sRise2_60_0_100+R	unnat	1 1	0 0	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	18
st1Fall2_40_40_0+L	nat	0 1	1 1	1 0	1 1	0 1	1 1	0 1	1 1	1 1	1 1	16
st1Fall2_40_40_20+L	unnat	1 1	0 1	1 0	0 0	0 0	1 1	1 0	1 1	0 0	1 0	10
st1Fall2_60_60_0+L	unnat	1 1	1 1	1 1	1 1	0 1	1 1	1 0	0 1	1 1	1 1	17
st1Fall2_60_60_20+F	unnat	1 1	1 0	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	19
st1Fall2_60_60_40+F	unnat	1 1	0 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	19
st1Rise2_40_40_100+H	nat	1 1	1 1	1 1	1 1	1 1	0 0	1 1	1 1	1 1	1 1	18
st1Rise2_40_40_60+H	nat	1 0	1 1	0 0	1 1	1 1	1 1	1 1	1 1	1 1	1 1	17
st1Rise2_40_40_80+H	nat	1 1	1 1	0 1	1 1	1 1	1 1	1 1	1 1	1 1	1 1	19
st1Rise2_60_60_100+H	nat	1 1	1 1	1 1	1 1	1 1	0 0	0 1	1 1	1 1	1 1	17
st1Rise2_60_60_80+H	nat	0 0	1 1	1 1	1 1	1 1	1 0	0 1	0 0	1 1	1 0	13

straight_30_30_30+M	unnat	1	1	1	0	0	0	1	1	0	1	1	0	1	1	1	0	0	0	1	1	12
straight_50_50_50+M	nat	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	19
straight_70_70_70+F	nat	0	0	1	1	1	0	0	0	1	1	0	1	1	1	1	1	1	1	1	0	13
u_40_0_40+R	nat	1	0	1	1	0	0	1	1	1	1	1	0	0	0	1	1	1	1	1	1	14
u_40_20_40+R	unnat	1	1	0	0	1	1	0	0	1	0	1	1	0	0	1	1	1	1	1	1	13
u_60_0_60+R	unnat	0	1	0	1	1	0	1	0	0	0	1	1	1	1	1	0	1	1	1	1	13
u_60_20_60+R	unnat	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	18
u_60_40_60+H	nat	1	0	1	0	1	1	1	1	1	1	0	0	1	1	0	1	1	1	1	1	15
	%	82		78		72		76		81		77		80		82		87		83		718

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ประวัติผู้เขียนวิทยานิพนธ์

นายกลวัชร ตระกูลสุข เกิดวันจันทร์ที่ 19 มีนาคม พ.ศ.2533 ที่กรุงเทพมหานคร สำเร็จการศึกษาระดับมัธยมศึกษาตอนต้นจากโรงเรียนแม่พระฟาติมา สำเร็จการศึกษาระดับมัธยมศึกษาตอนปลายจาก โรงเรียนสามเสนวิทยาลัย กรุงเทพมหานคร สำเร็จการศึกษาระดับปริญญาบัณฑิต สาขาวิศวกรรมคอมพิวเตอร์ จากคณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย กรุงเทพมหานคร ในปีการศึกษา 2554

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University