การสังเคราะห์เสียงภาษาไทยโดยวิธีสังเคราะห์แบบฟอร์แมนท์

THAI SPEECH SYNTHESIS BASED ON A FORMANT APPROACH

สายชัญ สายยศ

SAIYAN SAIYOD

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรดุษฎีบัณฑิต

สาขาวิชาวิศวกรรมคอมพิวเตอร์

บัณฑิตวิทยาลัย

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

พ.ศ. 2543

ISBN 974-15-1998-2

สำนักหอสมุดกลาง พระจอมเกล้าลาดกระบัง

การสังเคราะห์เสียงภาษาไทยโดยวิธีสังเคราะห์แบบฟอร์แมนท์

THAI SPEECH SYNTHESIS BASED ON A FORMANT APPROACH

T060561

สายยัญ สายยศ

SAIYAN SAIYOD

เลขหมู่.....
เลขทะเบียน 60561
วัน,เดือน,ปี - 3 0 ก.ค. 2549

b 11590117
i.....

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต

สาขาวิชาวิศวกรรมคอมพิวเตอร์

บัณฑิตวิทยาลัย

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

พ.ศ. 2548

ISBN 974-15-1998-2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

THAI SPEECH SYNTHESIS BASED ON A FORMANT APPROACH

SAIYAN SAIYOD

A THESIS SUBMITTED IN PARTIAL FULFILLMENT

OF THE REQUIREMENT FOR THE DEGREE OF

MASTER OF ENGINEERING IN COMPUTER ENGINEERING

SCHOOL OF GRADUATE STUDIES

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

2005

ISBN 974-15-1998-2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

COPYRIGHT 2005

SCHOOL OF GRADUATE STUDIES

KING MONGKUT'S INSTITUTE OF TECHNOLOGY LADKRABANG

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หัวข้อวิทยานิพนธ์	การสังเคราะห์เสียงภาษาไทยโดยวิธีสังเคราะห์แบบฟอร์แมนท์
นักศึกษา	นายสายชัญ สายยศ
รหัสประจำตัว	46061018
ปริญญา	วิศวกรรมศาสตรมหาบัณฑิต
สาขาวิชา	วิศวกรรมคอมพิวเตอร์
พ.ศ.	2548
อาจารย์ผู้ควบคุมวิทยานิพนธ์	รศ.สมศักดิ์ มิตะถา
อาจารย์ผู้ควบคุมวิทยานิพนธ์ร่วม	ผศ.ดร.ศักดิ์ชัย ทิพย์จักษุรัตน์

บทคัดย่อ

การสังเคราะห์เสียงพูดภาษาไทย คือกระบวนการสร้างเสียงพูดภาษาไทย เพื่อใช้ในกระบวนการสังเคราะห์เสียงจากข้อความภาษาไทย วิทยานิพนธ์นี้ เรานำเสนอด้วยวิธีการสังเคราะห์ฟอร์แมนท์เพื่อสังเคราะห์เสียงพูดภาษาไทย วิธีการนี้มีส่วนประกอบที่ปัจจัยสำคัญ คือ การสร้างรูปแบบของการออกเสียงซึ่งเรียกว่ากฎ เริ่มแรกหลักการนี้จะเก็บค่าเสียงพูดในรูปแบบของความถี่ฟอร์แมนท์ ซึ่งประกอบด้วย กลุ่มพยัญชนะ, กลุ่มสระ และกลุ่มตัวสะกด โครงสร้างเหล่านี้จะทำงานร่วมกับ การดัดแปลงความถี่ฟอร์แมนท์ ซึ่งเรียกว่า กฎการออกเสียงภาษาไทย โดยจะเริ่มต้นด้วยการแทนที่ข้อมูลที่นำเข้าซึ่งเป็นอาเรย์ของข้อมูลหน่วยเสียงต่างๆของพยางค์ด้วยความถี่ฟอร์แมนท์ที่ถูกต้อง แล้วผลที่ได้จากการประมวลผลนี้ไปดัดแปลงด้วยกฎ ด้วยการผ่านค่าพารามิเตอร์ ด้วยวิธีการนี้ผลการทดลองที่ได้ ซึ่งเราใช้ 40 ตัวอย่างของพยางค์ภาษาไทย ให้ 40 คนฟังในห้องที่มีการควบคุม และเสียงที่สังเคราะห์ออกมาสามารถฟังแล้วเข้าใจได้ ผลการประเมินเสียงภาษาไทยที่ได้ฟังมีความถูกต้องร้อยละ 71.75

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Thesis title	Thai Speech Synthesis Based on a Formant Approach
Student	Mr.Saiyan Saiyod
Student ID.	46061018
Degree	Master of Engineering
Programme	Computer Engineering
Year	2005
Thesis Advisor	Assoc.Prof.Somsak Mitatha
Thesis Co-Advisor	Asst.Prof.Dr. Sakchai Thipchaksurat

ABSTRACT

Thai speech synthesis is a process which artificially produces Thai speech for text-to-speech synthesis. In this thesis, we propose a rule-base approach Formant synthesis to synthesis Thai speech. In this method, the core component is the construction of intonation patterns is called rules. These rules are firstly set up in the format of Formants frequencies consisting of consonant group, vowel group and final consonant group. These structures collaborate with the Formants Frequencies Adaptation called Thai-Tonal-Accent Rules. The manipulation of this approach begins with substituting each phoneme of input array of phoneme of syllable with corresponded formants frequencies. Then, outcomes of this process are adapted with the rule by pass parameter. In this way, for the experimental results, we have utilized 40 samples of Thai syllable for 40 people by listening test in control room. They showed that the speeches of syllable are correct in Thai language 71.75 percent.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กิตติกรรมประกาศ

วิทยานิพนธ์เล่มนี้สำเร็จได้ด้วยความกรุณาจากอาจารย์ผู้ควบคุมวิทยานิพนธ์ รศ.สมศักดิ์ มิตะถา และอาจารย์ผู้ควบคุมวิทยานิพนธ์ร่วม ผศ.ดร.ศักดิ์ชัย ทิพย์จักษุรัตน์ ที่ให้ความช่วยเหลือให้คำชี้แนะช่วยแก้ปัญหาตลอดจนให้ความรู้และประสบการณ์ที่ดีแก่ข้าพเจ้า

ขอขอบพระคุณ และ กรรมการสอบหัวข้อและโครงร่างวิทยานิพนธ์ที่ได้กรุณาให้คำแนะนำตลอดจนข้อชี้แนะ จนในที่สุดทำให้วิทยานิพนธ์ฉบับนี้สำเร็จลงได้

ขอขอบคุณ รศ.วิรัตน์ พงษ์ศิริ ที่คอยให้กำลังใจและช่วยหางานให้ทำเพื่อเป็นทุนในการศึกษา

ขอขอบคุณมูลนิธิคอมพิวเตอร์และการสื่อสาร ที่ให้ทุนการศึกษาในปีการศึกษาที่ 2

ขอขอบคุณบริษัท ที.เอส.ที. คอมเมอร์เชียล จำกัด ที่ให้โอกาสข้าพเจ้าได้พัฒนาความรู้ความสามารถของข้าพเจ้าให้สูงยิ่งขึ้นไป และเป็นแหล่งทุนเวลาข้าพเจ้าเดือดร้อน

ขอขอบคุณผู้ที่มีส่วนร่วมในการวิจัยและทดลอง ดังต่อไปนี้

- นางสาวพัชรินทร์ ชินวรรณ นายสัมฤทธิ์ เทวภูมิ นายพรชัย สมัครสมาน และอีกหลายท่านที่ไม่ได้เอ่ยนาม ที่เสียสละเวลาเข้าร่วมทดสอบและเป็นผู้บอกภาษา
- อาจารย์จำเนียร แก่ประโคน และคณะครู หลายโรงเรียนในอำเภอประโคนชัย จังหวัดบุรีรัมย์

สำหรับคุณงามความดีอันใดที่เกิดจากวิทยานิพนธ์ฉบับนี้ ข้าพเจ้าขอมอบให้กับบิดามารดา ซึ่งเป็นที่รักและเคารพยิ่ง ตลอดจนครูอาจารย์ที่เคารพทุกท่านที่ให้วิชาความรู้และถ่ายทอดประสบการณ์ที่ดีให้แก่ข้าพเจ้า

สายยัญ สายยศ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

หน้า

บทคัดย่อภาษาไทย I

บทคัดย่อภาษาอังกฤษ II

กิตติกรรมประกาศ III

สารบัญ IV

สารบัญตาราง VIII

สารบัญรูป IX

บทที่ 1 บทนำ 1

1.1 ความเป็นมาและความสำคัญของปัญหา 1

1.2 ความมุ่งหมายและวัตถุประสงค์ของการศึกษา 2

1.3 สมมุติฐานของการศึกษา 2

1.4 ทฤษฎีหรือแนวความคิดที่ใช้ในงานวิจัย 2

1.5 ขอบเขตของงานวิจัย 3

1.6 ขั้นตอนของการศึกษา 3

บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง 4

2.1 งานวิจัยที่เกี่ยวข้อง 4

2.1.1 งานวิจัยที่เกี่ยวข้องกับการสังเคราะห์แบบฟอร์แมนท์ในต่างประเทศ 4

2.1.2 งานวิจัยที่เกี่ยวข้องกับการสังเคราะห์เสียงแบบฟอร์แมนท์ในประเทศจีน 7

2.1.3 งานวิจัยที่เกี่ยวข้องกับการสังเคราะห์เสียงภาษาไทย 10

2.2 กายวิภาคของระบบการเปล่งเสียงของมนุษย์ 12

2.2.1 อวัยวะออกเสียง 12

2.2.2 กระบวนการผลิตเสียงพูด 14

2.2.3 เสียงพูดของมนุษย์ 15

2.3 หน่วยเสียงสำคัญในภาษาไทย 16

2.3.1 หน่วยเสียงสระ 17

2.3.2 หน่วยเสียงพยัญชนะ 19

2.3.3 หน่วยเสียงวรรณยุกต์ 20

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

หน้า

2.4 ลักษณะพยางค์และคำในภาษาไทย 22
2.5 ทฤษฎีทางด้านการวิเคราะห์เสียง 24
2.5.1 การปรับบรรทัดฐานแอมพลิจูด 24
2.5.2 วิธีการเน้นล่วงหน้า 25
2.5.3 การวางกรอบหน้าต่าง 25
2.5.4 สเปกโตรแกรม 26
2.5.5 ออโตคอร์รีเลชั่น 27
2.5.6 การประมาณการเชิงเส้น 28
2.5.7 ผลการแปลง z 29
2.5.8 โพลและซีโร 31
2.6 ทฤษฎีการวิเคราะห์คุณสมบัติของสัญญาณเสียงพูด 32
2.6.1 ความถี่มูลฐาน 32
2.6.2 ความถี่ฟอร์แมนท์ 33
2.6.3 กฎของคราเมอร์ 33
2.7 ทฤษฎีทางด้านการสังเคราะห์เสียง 35
2.7.1 การสังเคราะห์เสียงแบบฟอร์แมนท์ 37

บทที่ 3 วิธีการดำเนินการวิจัย 43
3.1 การคัดเลือกผู้บอกภาษา 43
32 การเก็บตัวอย่างข้อมูลเสียง 43
3.3 เครื่องมือและอุปกรณ์ในการวิจัย 44
3.4 ประเมินผลการสังเคราะห์เสียง 44

บทที่ 4 การวิเคราะห์คุณสมบัติของเสียงภาษาไทย 45
4.1 วิเคราะห์ความถี่ฟอร์แมนท์ของเสียงภาษาไทย 45
4.1.1 การบันทึกพยางค์ตัวอย่าง 46
4.1.2 แยกหน่วยเสียงที่ต้องการ 46
4.1.3 หาความถี่ฟอร์แมนท์ของหน่วยเสียง 47

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

หน้า

4.2 วิเคราะห์ความถี่มูลฐานสำหรับการผันวรรณยุกต์ภาษาไทย 51
4.2.1 การหาความถี่มูลฐาน 51
4.2.2 วิเคราะห์การผันวรรณยุกต์สามัญ 56
4.2.3 วิเคราะห์การผันวรรณยุกต์เอก 62
4.2.4 วิเคราะห์การผันวรรณยุกต์โท 64
4.2.5 วิเคราะห์การผันวรรณยุกต์ตรี 70
4.2.6 วิเคราะห์การผันวรรณยุกต์จัตวา 73
4.3 กฎการออกเสียงภาษาไทย 78
4.3.1 กฎการออกเสียงตามเพศ 78
4.3.2 กฎความเร็วของเสียง 81
4.3.3 กฎการแบ่งวรรคตอน 82
4.3.4 กฎระดับเสียง 83
4.3.5 กฎการผันวรรณยุกต์ 84

บทที่ 5 การสังเคราะห์เสียงภาษาไทย 86
5.1 หลักการสังเคราะห์เสียงแบบฟอร์แมนท์ 86
5.2 การสังเคราะห์เสียงภาษาไทย 87
5.2.1 การแทนที่หน่วยเสียงความถี่ฟอร์แมนท์ 88
5.2.2 การปรับปรุงความถี่ฟอร์แมนท์ของพยางค์ตามพารามิเตอร์ 91
5.2.3 การแปลงความถี่ฟอร์แมนท์ของพยางค์ให้เป็นเสียงพูด 99

บทที่ 6 ผลการทดลอง 100
6.1 การประเมินผลการสังเคราะห์เสียง 100

บทที่ 7 สรุปผลการวิจัยและข้อเสนอแนะ 104
7.1 สรุปผลการวิจัยและข้อเสนอแนะ 104
7.2 ปัญหาและอุปสรรค 104
7.3 ข้อเสนอแนะ 104

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

หน้า

เอกสารอ้างอิง 105

ภาคผนวก 108

ภาคผนวก ก. บทความเกี่ยวข้องที่ได้รับการตีพิมพ์ 109

ภาคผนวก ข. แบบสอบถามเพื่อวิเคราะห์ในงานวิจัย 110

ภาคผนวก ค. คำที่ใช้ในการทดสอบการฟัง 112

ภาคผนวก ง. แบบสอบถามที่ใช้ในการทดสอบการฟัง 113

ภาคผนวก จ. ข้อมูลความถี่ฟอร์แมนท์ 114

ประวัติผู้เขียน 116

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

ตารางที่ หน้า

2.1 การพัฒนาระบบสังเคราะห์เสียงจากข้อความในภาษาจีน 8

2.2 ลักษณะโทนเสียงภาษาจีน 8

2.3 สระเสียงเดี่ยวในภาษาไทย 18

2.4 เสียงพยัญชนะในภาษาไทย 20

2.5 การผันเสียงวรรณยุกต์ของอักษรกลาง 23

2.6 การผันเสียงวรรณยุกต์ของอักษรต่ำคู่กับอักษรสูง 23

2.7 การผันเสียงวรรณยุกต์ของอักษรต่ำเดี่ยวกับอักษรนำ 23

4.1 ค่าความถี่มูลฐานที่ได้จากการคำนวณวรรณยุกต์สามัญ 53

4.2 ค่าความถี่มูลฐานที่ได้จากการปรับปรุงของวรรณยุกต์สามัญ 55

4.3 ชุดข้อมูลการผันวรรณยุกต์เสียงสามัญ 57

4.4 การคำนวณพจน์ต่างๆของสมการจากชุดข้อมูลการผันวรรณยุกต์สามัญ 61

4.5 ชุดข้อมูลความถี่มูลฐานการผันวรรณยุกต์เอก 63

4.6 ชุดข้อมูลความถี่มูลฐานการผันวรรณยุกต์โท 68

4.7 ชุดข้อมูลความถี่มูลฐานการผันวรรณยุกต์ตรี 71

4.8 ชุดข้อมูลความถี่มูลฐานการผันวรรณยุกต์จัตวา 74

4.9 ค่าระดับความถี่มูลฐานของแต่ละวรรณยุกต์ 77

4.10 ความถี่ฟอร์แมนท์ของผู้หญิงและผู้ชาย 78

4.11 ความถี่มูลฐานของผู้หญิงและผู้ชาย 79

4.12 ค่าช่วงเวลาของเสียงสระ 81

6.1 ผลการประเมินเสียงพยัญชนะในระดับพยางค์ 101

6.2 ผลการประเมินเสียงสระ 102

6.3 ผลการประเมินเสียงวรรณยุกต์ 102

6.4 ผลการประเมินในระดับพยางค์โดยรวม 102

6.5 ผลการประเมินในระดับคำ 103

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป

รูปที่ หน้า

2.1 ตัวสังเคราะห์เสียงพูดของ OVE II 5

2.2 ตัวสังเคราะห์เสียงพูดของ John Homes 6

2.3 Rule based synthesis system using data-driven unit library 7

2.4 สเปกตรัมและขอบของความถี่มูลฐานของพยางค์ /ma1/ /ma2/ /ma3/ /ma4/ /ma5/ ภาษาจีนในภาษาจีน 7

2.5 แผนภูมิแสดงความถี่มูลฐานบนแกนเวลาเวลาของโทนเสียงภาษาจีน 9

2.6 อวัยวะในระบบการพูดของมนุษย์ 14

2.7 แผนภาพระบบเสียงพูดของมนุษย์ 15

2.8 ตำแหน่งของลิ้นเมื่ออยู่ต่ำ ส่วนหน้าสูง และส่วนหลังสูง 17

2.9 แสดงการเกิดเสียงสระผสมในภาษาไทย 19

2.10 องค์ประกอบของหน่วยเสียงต่างๆ ในพยางค์ 22

2.11 รูปแบบของฟังก์ชันกรอบหน้าต่างแบบต่างๆ 25

2.12 สเปกโตรแกรมแถบกว้าง 26

2.13 สเปกโตรแกรมแถบแคบ 27

2.14 ระนาบเชิงซ้อน z 30

2.15 สเปกโตรแกรมของเสียงสระอา 33

2.16 ระบบการสังเคราะห์เสียงจากข้อความ 35

2.17 การสังเคราะห์เสียงแบบฟอร์แมนท์ 37

2.18 โมเดลของ Formant Synthesizer 38

2.19 บล็อกไดอะแกรมของแหล่งกำเนิดเสียงก้อง 39

2.20 ตัวอย่างของสเปกตรัมความถี่ของเสียงสระ 40

2.21 การเชื่อมต่อแบบขนานของ Formant Synthesizer 41

2.22 การเชื่อมต่อแบบอนุกรมของ Formant Synthesizer 41

2.23 ผลที่ได้ของฟังก์ชั่นถ่ายโอนของฟอร์แมนท์เรโซเนเตอร์ 41

2.24 บล็อกไดอะแกรมของ Klatt Formant Synthesizer 42

4.1 สัญญาณเสียงของพยางค์ "กอ" 45

4.2 ขั้นตอนการหาความถี่ฟอร์แมนท์สำหรับเสียงพูดภาษาไทย 46

4.3 การตัดหน่วยเสียงที่ต้องการวิเคราะห์ 46

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป (ต่อ)

รูปที่ หน้า

4.4 กระบวนการหาความถี่ฟอร์แมนท์ของหน่วยเสียง 47

4.5 การแบ่งช่วงของสัญญาณที่ใช้ในการวิเคราะห์ 48

4.6 ความถี่ฟอร์แมนท์ของสระเสียง /อา/ 50

4.7 แผนภูมิแสดงความสัมพันธ์ระหว่างความถี่ฟอร์แมนท์ F1 และ F2 51

4.8 ขั้นตอนการหาความถี่มูลฐาน 52

4.9 ความถี่มูลฐานของเสียงวรรณยุกต์ 54

4.10 แผนภูมิเส้นของเสียงวรรณยุกต์หลังการปรับปรุงความถี่มูลฐาน 56

4.11 แผนภูมิเส้นแสดงการจัดกรอบเวลาใหม่ของเสียงวรรณยุกต์สามัญ 58

4.12 การถดถอยแบบเชิงเส้นโดยการประดิษฐ์ฟังก์ชั่นเส้นตรงจากผลการทดลองของเสียงวรรณยุกต์สามัญ 59

4.13 แผนภูมิเส้นความถี่มูลฐานการผันวรรณยุกต์เอกบนแกนเวลา 62

4.14 แผนภูมิเส้นแสดงการจัดกรอบเวลาใหม่ของเสียงวรรณยุกต์เอก 64

4.15 แผนภูมิเส้นความถี่มูลฐานการผันวรรณยุกต์โทบนแกนเวลา 65

4.16 แผนภูมิเส้นแสดงการจัดกรอบเวลาใหม่ของเสียงวรรณยุกต์โท 69

4.17 แผนภูมิเส้นความถี่มูลฐานการผันวรรณยุกต์ตรีบนแกนเวลา 70

4.18 แผนภูมิเส้นแสดงการจัดกรอบเวลาใหม่ของเสียงวรรณยุกต์ตรี 72

4.19 แผนภูมิเส้นความถี่มูลฐานการผันวรรณยุกต์จัตวาบนแกนเวลา 73

4.20 แผนภูมิเส้นแสดงการจัดกรอบเวลาใหม่ของเสียงวรรณยุกต์จัตวา 75

4.21 การเพิ่มลดความเร็วเสียงของพยางค์ 81

4.22 การแบ่งวรรคตอนของพยางค์ 81

4.23 การผันวรรณยุกต์ของพยางค์ 82

5.1 ระบบการสังเคราะห์เสียงแบบฟอร์แมนท์ 86

5.2 ระบบการสังเคราะห์เสียงภาษาไทย 87

5.3 โครงสร้างข้อมูลหน่วยเสียงของพยางค์ 88

5.4 องค์ประกอบของการแทนที่หน่วยเสียงด้วยความถี่ฟอร์แมนท์ 89

5.5 โครงสร้างข้อมูลความถี่ฟอร์แมนท์ของพยางค์ 90

5.6 ตัวอย่างของการแทนที่หน่วยเสียงด้วยความถี่ฟอร์แมนท์ 91

5.7 การปรับปรุงความถี่ฟอร์แมนท์ของพยางค์ตามพารามิเตอร์ 92

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป (ต่อ)

รูปที่ หน้า

5.8 การปรับปรุงความถี่ฟอร์แมนท์ของการออกเสียงเพศชาย....93
5.9 การปรับปรุงข้อมูลเวลาของการลดความเร็วเสียงลง 10%....94
5.10 การปรับปรุงข้อมูลเวลาของการเพิ่มช่วงเวลาการแบ่งวรรคตอนขึ้น 10%....95
5.11 การแทรกคอลัมน์ของจุดเปลี่ยนการผันวรรณยุกต์โท....96
5.12 การแทรกคอลัมน์ของการผันวรรณยุกต์โท....97
5.13 การคำนวณค่าความถี่มูลฐานของการผันวรรณยุกต์โท....98
5.14 องค์ประกอบของ Formant Synthesizer....99

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1
บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ปัจจุบันเทคโนโลยีการสังเคราะห์เสียงจากข้อความ (Text-to-Speech) หรือ TTS มีบทบาทสำคัญมากในหลายๆ แอพพลิเคชั่น เช่น talking dictionary และ homepage reader โดยส่วนใหญ่แล้วการสังเคราะห์ข้อความไปเป็นเสียงภาษาไทยประกอบไปด้วย 3 ส่วนหลักๆ คือ ส่วนการวิเคราะห์ข้อความ(Text Analysis), ส่วนการวิเคราะห์สัทสัมพันธ์ (Prosody Analysis) และส่วนการสังเคราะห์เสียง (Speech Synthesis) การเปล่งเสียงสำเนียงภาษาไทยที่ถูกต้องนั้นจะขึ้นอยู่กับส่วนของการสังเคราะห์เสียง ซึ่งทำหน้าที่ในการสร้างสัญญาณคลื่นเสียงจากข้อมูลเสียงอ่าน(phonetic transcription) และข้อมูลสัทสัมพันธ์ (Prosody transcription) เพื่อให้เราได้ยินเสียงของพยางค์นั้นๆ โดยทั่วไปส่วนการสังเคราะห์เสียงสามารถแบ่งตามเทคนิควิธีการสังเคราะห์เสียงได้ 3 ประเภท คือ Concatenation Synthesis, Articulation Synthesis และFormant Synthesis [1]

การวิจัยและพัฒนาด้านการสังเคราะห์เสียงเริ่มต้นมานานสำหรับภาษาหลักต่างๆ เช่น ภาษาอังกฤษ ภาษาญี่ปุ่น ภาษาจีน เป็นต้น สำหรับการสังเคราะห์เสียงภาษาไทยมีการวิจัยมานานแต่มีผู้ทำการศึกษาและวิจัยทางด้านนี้น้อย ทำให้ผลงานวิจัยทางด้านนี้สำหรับภาษาไทยมีน้อย วิธีส่วนใหญ่ที่ใช้สังเคราะห์ได้แก่ การสังเคราะห์เสียงด้วยวิธีการ Concatenation Synthesis

การสังเคราะห์เสียงด้วยวิธี Concatenation Synthesis เป็นวิธีการเก็บส่วนของเสียงที่มีลักษณะร่วมกัน เพื่อนำมาเชื่อมต่อกันให้ได้เสียงคำพูดที่ต้องการ วิธีการนี้มีคุณภาพเสียงสูงเนื่องจากเป็นเสียงที่ได้จากการตัดมาจากเสียงพูดจริง แต่วิธีการดังกล่าวมีปัญหาทางด้านคุณภาพของเสียงบริเวณรอยต่อและการปรับทำนองและจังหวะของเสียง เนื่องจากส่วนของเสียงที่เก็บไว้มีลักษณะคงตัว และภาษาไทยเป็นภาษาที่มีการออกเสียงวรรณยุกต์ ถ้าออกเสียงวรรณยุกต์ไม่ถูกต้องก็จะทำให้ความหมายผิดไปด้วย ซึ่งภาษาไทยมีเสียงพยัญชนะ 21 เสียง เสียงสระ 27 เสียง เสียงตัวสะกด 8 เสียง และแต่ละพยางค์สามารถผันเสียงวรรณยุกต์ได้อีก 5 ระดับ เพื่อให้การสังเคราะห์ได้ครอบคลุมเสียงที่เกิดขึ้นได้ทั้งหมด จำเป็นต้องเก็บส่วนของเสียงไว้ตามจำนวนลักษณะของเสียงทั้งหมด จำเป็นต้องใช้หน่วยความจำเป็นจำนวนมากและไม่สะดวกในการนำไปใช้งานจริง

จากปัญหาที่เกิดขึ้นกับการสังเคราะห์เสียงด้วยวิธี Concatenation Synthesis วิทยานิพนธ์นี้จึงนำเสนอวิธีแก้ปัญหาดังกล่าว ด้วยวิธีการสังเคราะห์เสียงแบบฟอร์แมนท์ (Formant Synthesis) ซึ่งสามารถลดขนาดข้อมูลและจำนวนหน่วยเสียง โดยที่สามารถปรับลักษณะทำนองของเสียงที่สังเคราะห์ได้ การสังเคราะห์ที่ใช้ในวิทยานิพนธ์นี้จะเก็บข้อมูลของหน่วยเสียงในรูปแบบของ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ความถี่ฟอร์แมนท์และพารามิเตอร์ แล้วนำเข้าไปประมวลผลตามกฎต่างๆ ที่จะทำให้เกิดเสียงที่ต้องการ

1.2 ความมุ่งหมายและวัตถุประสงค์ของการศึกษา

1. เพื่อศึกษาและหาหน่วยเสียงที่จำเป็นสำหรับการสังเคราะห์เสียงภาษาไทย
2. เพื่อลดขนาดของข้อมูลและจำนวนหน่วยเสียงสำหรับการสังเคราะห์เสียงภาษาไทย
3. เพื่อศึกษาและหากฎต่าง ๆ เพื่อสังเคราะห์เสียงภาษาไทยที่ต้องการ
4. เพื่อพัฒนาวิธีการสังเคราะห์เสียงภาษาไทยแบบฟอร์แมนท์

1.3 สมมุติฐานของการศึกษา

พยางค์ในภาษาไทยประกอบด้วยหน่วยเสียงพยัญชนะ หน่วยเสียงสระ และสามารถผันวรรณยุกต์ได้ 5 เสียง ซึ่งการสังเคราะห์เสียงภาษาไทยโดยวิธีสังเคราะห์แบบฟอร์แมนท์ สามารถลดขนาดของหน่วยความจำ จำนวนหน่วยเสียงที่ใช้ในการสังเคราะห์ เนื่องจากเก็บข้อมูลในรูปแบบของความถี่ฟอร์แมนท์และพารามิเตอร์ รวมทั้งใช้กฎในการสังเคราะห์เสียงที่ต้องการ

1.4 ทฤษฎีหรือแนวความคิดที่ใช้ในงานวิจัย

เมื่อพิจารณาลักษณะของเสียงพูดพบว่า ลักษณะของเสียงพูดส่วนใหญ่มีลักษณะเป็นเสียงก้องซึ่งเสียงดังกล่าวเกิดจากเสียงสระ ซึ่งมีลักษณะเป็นลูกคลื่นที่มีลักษณะซ้ำหรือใกล้เคียงกันเรียงต่อกันในเชิงเวลา และเมื่อพิจารณาในเชิงความถี่พบว่า ลักษณะของเสียงสระดังกล่าวมีลักษณะของฟอร์แมนท์ ดังนั้นในการเก็บหน่วยเสียงสำหรับการสังเคราะห์ จึงเกิดความคิดที่จะเก็บในลักษณะของความถี่ฟอร์แมนท์และค่าพารามิเตอร์ที่จำเป็นของหน่วยเสียงนั้น ๆ

สำหรับเสียงวรรณยุกต์เกิดจากการเปลี่ยนค่าของความถี่มูลฐานของเสียงก้อง ซึ่งแต่ละเสียงวรรณยุกต์มีรูปแบบการเปลี่ยนแปลงความถี่มูลฐานที่แตกต่างกัน เมื่อพิจารณาความถี่มูลฐานในเชิงเวลาพบว่า การเปลี่ยนแปลงความถี่มูลฐานของเสียงวรรณยุกต์เดียวกันจะมีลักษณะที่คล้ายคลึงกัน ดังนั้นการผันวรรณยุกต์จึงสามารถเก็บอยู่ในรูปของฟังก์ชันในเชิงเวลา เพื่อใช้เป็นกฎในการผันเสียงวรรณยุกต์ภาษาไทย ซึ่งสามารถลดจำนวนหน่วยเสียงลงได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากหลักการเก็บหน่วยเสียงและใช้กฎในการสังเคราะห์เสียงข้างต้น ต้องหาวิธีการและตัวแทนในการเก็บหน่วยเสียงที่เหมาะสำหรับการสังเคราะห์หน่วยเสียงดังกล่าว จากการศึกษาพบว่าการสังเคราะห์เสียงแบบฟอร์แมนท์ มีคุณสมบัติในการสังเคราะห์เสียงที่ใช้ความถี่ฟอร์แมนท์และพารามิเตอร์ ในการสังเคราะห์เสียงพูดออกมา จึงเกิดแนวคิดที่จะสังเคราะห์เสียงภาษาไทยโดยวิธีสังเคราะห์แบบฟอร์แมนท์

1.5 ขอบเขตของการวิจัย

1. วิจัยเฉพาะส่วนของระบบการสังเคราะห์เสียงแบบฟอร์แมนท์
2. ทำการสังเคราะห์เสียงที่มีคุณสมบัติดังนี้
    2.1 สังเคราะห์เสียงพยางค์ภาษาไทย
    2.2 สังเคราะห์โดยใช้ความถี่ฟอร์แมนท์และพารามิเตอร์
    2.3 สังเคราะห์เสียงวรรณยุกต์โดยใช้กฎ
3. พัฒนาโปรแกรมสำหรับวิเคราะห์และสังเคราะห์เสียงบนโปรแกรม MatLab

1.6 ขั้นตอนของการศึกษา

1. ศึกษาและค้นคว้าข้อมูลทางด้านการวิเคราะห์และสังเคราะห์เสียง
2. ศึกษาและค้นคว้าข้อมูลด้านภาษาศาสตร์
3. ออกแบบวิธีในการวิเคราะห์และสังเคราะห์เสียง
4. คัดเลือกผู้บอกภาษา และบันทึกเสียงเพื่อใช้ในการวิเคราะห์เสียง
5. เขียนโปรแกรมในการวิเคราะห์และสังเคราะห์เสียง
6. สร้างกฎต่างๆ ที่ใช้ในการสังเคราะห์เสียงภาษาไทย
7. ทดลองสังเคราะห์เสียงและปรับปรุง
8. ทำการทดสอบการฟัง
9. ประเมินผล สรุปผลการวิจัย และเสนอแนะแนวทาง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 2
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ในบทนี้จะกล่าวถึงทฤษฎีต่างๆ ที่เกี่ยวข้องกับงานวิจัยนี้ โดยเริ่มด้วย (1) ประวัติของการสังเคราะห์เสียงแบบฟอร์แมนท์ (2) กายวิภาคของระบบการเปล่งเสียงของมนุษย์ (3) หน่วยเสียงสำคัญในภาษาไทย (4) ลักษณะพยางค์และคำในภาษาไทย (5) ทฤษฎีทางด้านการวิเคราะห์เสียง (6) ทฤษฎีทางด้านคุณสมบัติของเสียง (7) ทฤษฎีทางด้านการสังเคราะห์เสียง

2.1 งานวิจัยที่เกี่ยวข้อง

งานวิจัยที่เกี่ยวข้องกับงานวิจัยนี้ จะกล่าวถึงงานวิจัยที่เกี่ยวการสังเคราะห์แบบฟอร์แมนท์ที่มีในต่างประเทศ และงานวิจัยที่เกี่ยวกับการสังเคราะห์เสียงภาษาไทย

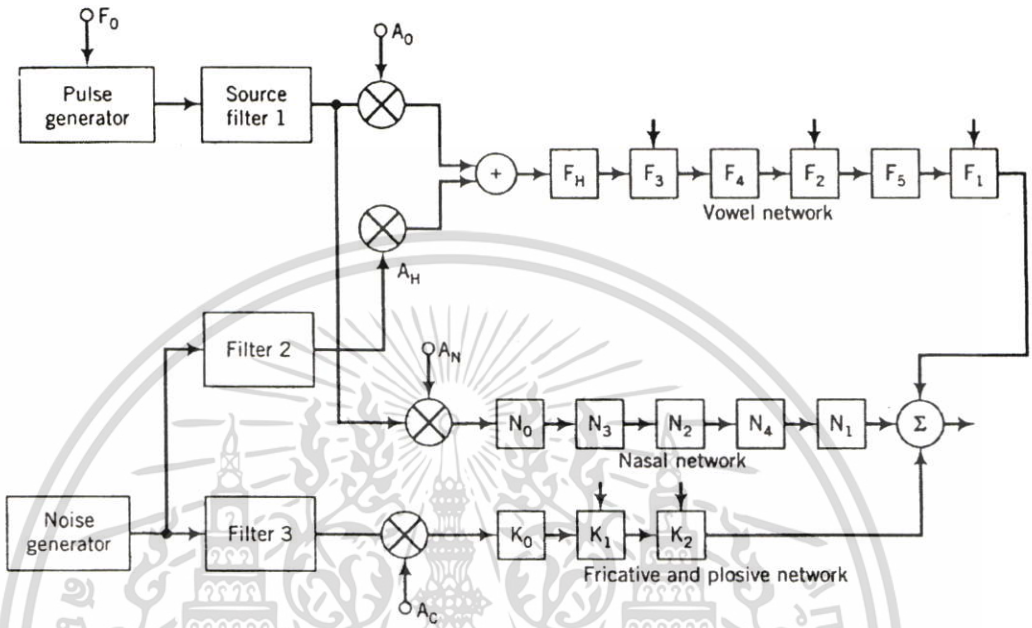
2.1.1 งานวิจัยที่เกี่ยวข้องกับการสังเคราะห์แบบฟอร์แมนท์ในต่างประเทศ

ในปี 1953 PAT [3] (Parametric Artificial Talker) ได้นำเสนอวิธีการสังเคราะห์เสียงด้วยอุปกรณ์อิเลคทรอนิคส์ ซึ่งทำการสร้างอุปกรณ์ที่ประกอบด้วยตัวสร้างความถี่เรโซแนนซ์ 3 ตัวซึ่งต่อขนานกัน ซึ่งจะทำการเปลี่ยนแปลงพารามิเตอร์ด้วยมือตามข้อมูลที่ได้จากการวิเคราะห์ความถี่ของเสียง ซึ่งนำเสนอโดย Walter LAWRENCE เนื่องจากเพิ่งมีการริเริ่มพัฒนาวิธีสังเคราะห์แบบฟอร์แมนท์ วิธีการนี้ผลที่ออกมาสามารถสังเคราะห์ได้เฉพาะเป็นเสียงสอดแทรก หรือเสียงไม่ก้องที่มีลักษณะเป็นเสียงที่ยาวต่อเนื่องตามความถี่ที่ป้อนเข้าไป เพราะว่าไม่มีการควบคุมพารามิเตอร์อย่างอื่นที่เกี่ยวข้องด้วยเช่น แอมพลิจูด และความถี่มูลฐาน เป็นต้น และในปีเดียวกันนี้ OVE I (Orator Verbis Electris) ซึ่งนำเสนอโดย Gunnar Fant นำเสนอการเชื่อมต่อแบบอนุกรมของ 3 ความถี่ฟอร์แมนท์ ซึ่งฟอร์แมนท์ที่ 1 และฟอร์แมนท์ที่ 2 ทำการควบคุมด้วยการเคลื่อนที่เชิงกล วิธีการการนี้สามารถสังเคราะห์ได้เฉพาะเสียงก้อง ซึ่งเป็นลักษณะของเสียงสระที่ยาวต่อเนื่อง เพราะไม่มีการควบคุมพารามิเตอร์อย่างอื่นที่เกี่ยวข้อง แต่จะทำการควบคุมเฉพาะค่าความถี่ฟอร์แมนที่ 1 และฟอร์แมนท์ที่ 2 ซึ่งจะเป็นคุณสมบัติเฉพาะของเสียงสระ แต่ยังไม่สามารถออกเสียงพูดได้

ในปี 1962 PAT ได้ทำการปรับปรุงตัวสังเคราะห์เสียงที่ได้พัฒนามาข้างต้น โดยเพิ่มการควบคุมแอมพลิจูดของแต่ละตัวสร้างเสียงสอดแทรก (fricative) ซึ่งเป็นการพัฒนาเสียงสอดแทรกจากของเดิม ซึ่งทำให้สามารถควบคุมการเกิดเสียงสอดแทรกได้ดีขึ้น แต่ก็ยังไม่สามารถสร้างเสียงที่มีลักษณะเป็นเสียงก้องได้ดี และในปีเดียวกัน OVE II ได้ทำการปรับปรุงตัวสังเคราะห์เสียงเช่นเดียวกัน ซึ่งประสบความสำเร็จค่อนข้างสูงในการสังเคราะห์เสียงคำพูด "I Love you" ซึ่งแสดง

เอกสารนี้มีลิขสิทธิ์ ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ในรูปที่ 2.1 [3] ซึ่งมีทั้งส่วนของการสร้างเสียงสอดแทรกและเสียงก้อง แต่เสียงที่ได้ยังไม่เป็นธรรมชาติ เนื่องจากการควบคุมแอมพลิจูดจะทำการควบคุมแต่ละสาขา ทำให้การตอบสนองด้านความถี่ไม่ดีเท่าที่ควร

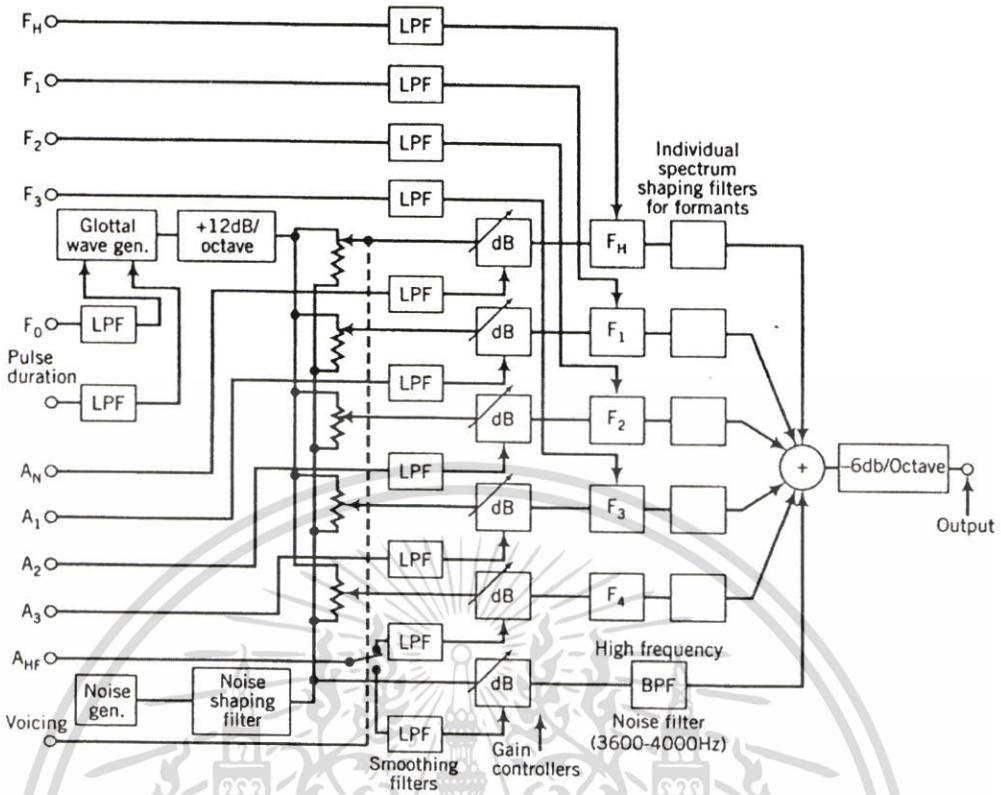


รูปที่ 2.1 ตัวสังเคราะห์เสียงพูดของ OVE II

ในปี 1972 John Holmes ทำงานอย่างหนักเพื่อทำการสังเคราะห์เสียง "I enjoy the simple life" ซึ่งได้ทำการปรับปรุงตัวสังเคราะห์เสียงให้มีประสิทธิภาพยิ่งขึ้น โดยที่เอาพุตบล็อกให้เสียงออกได้ 6 dB/octave ซึ่งใช้ตัวกรองสูงผ่าน (high-pass filter) เพื่อแสดงลักษณะการบรรจบกันของปากและลม และเพิ่มบล็อก 12 dB/octave ที่ตัวสร้างคลื่นเสียงก้องจากลำคอ โดยใช้ตัวกรองต่ำผ่าน (low-pass filter) และทำการเพิ่มความถี่ฟอร์แมนท์ขึ้นเป็น 5 ความถี่ ดังแสดงในรูปที่ 2.2 [3]

สิ่งสำคัญที่แตกต่างระหว่างตัวสังเคราะห์เสียงของ Homes และ Fant คือ แนวทางในการใช้ข้อมูลของแอมพลิจูด ในอุปกรณ์ของ Fant แต่ละแขนงแยกการควบคุมแอมพลิจูดโดย A_H, A_N, A_cและ A_o ส่วนของ Homes ตัวกรอง 5 ความถี่ฟอร์แมนท์ F_n, F_1, F_2, F_3 และ F_4 จะมีการควบคุมตัวต่อตัวในการสร้างทุกเสียง ซึ่งจะเห็นว่าโครงสร้างของ Fant ที่นำเสนอจะมีลักษณะค่อนข้างตายตัวในการสร้างเส้นทางเดินเสียงของมนุษย์ แต่ของ Homes สามารถกำหนดค่าต่างๆ ได้มากกว่า ทำให้ได้เสียงที่ค่อนข้างใกล้เคียงเสียงมนุษย์มากกว่า ระบบการสังเคราะห์เสียงที่กล่าวมาข้างต้นนี้เป็นระบบที่จำลองการพูดของมนุษย์จากการสร้างวงจรอิเล็กทรอนิกส์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



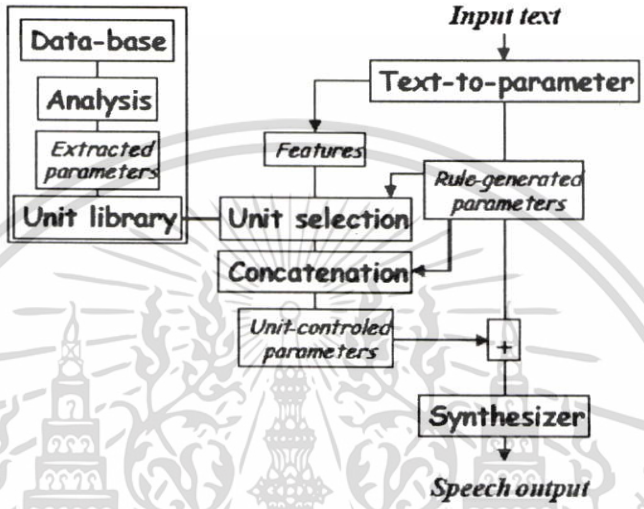
รูปที่ 2.2 ตัวสังเคราะห์เสียงพูดของ John Holmes.

ในปี 1979-1987 Denis Klatt และคณะได้นำเสนอซอฟแวร์สำหรับการสังเคราะห์เสียงแบบฟอร์แมนท์ ซึ่งสร้างจากผลที่ได้จากการออกแบบของ Fant และ Holmes [3] ซึ่ง Klatt ได้ทำการรวมตัวสร้างเสียงก้องของ Fant โดยใช้ cascade formant resonantor ส่วนการสร้างเสียงสอดแทรกเขาใช้ channel vocoder structure ด้วยการกำหนดตัวกรองตายตัวของตัวแปรอัตราขยาย ซึ่งโครงสร้างพื้นฐานนี้ได้รับการใช้เป็นตัวสังเคราะห์เสียงในหลายๆระบบการสังเคราะห์เสียงจากข้อความ ซึ่งวิธีการของ Klatt จะกล่าวโดยละเอียดในหัวข้อที่ 2.7.1 ต่อไป

ในปี 2002 Rolf Carlson, Tor Sigvardson และArvid Sjolander [4] นำเสนอวิธีการใหม่เพื่อการสังเคราะห์เสียงแบบฟอร์แมนท์ โดยใช้ทั้ง rule-based และ data-driven เข้าด้วยกัน วิธีนี้จะผสมผสานวิธีการสังเคราะห์เสียงแบบต่างๆเข้าด้วยกัน โดยพยายามสร้างตัวสังเคราะห์เสียงที่สามารถสังเคราะห์ภาษาอื่นๆได้ โดยทำการเก็บบันทึกข้อมูลเสียงพูดไว้ในส่วนของ Data-driven โดยใช้ซอฟแวร์ Label Infovox ซึ่งเป็นผลิตภัณฑ์ของ MBROLA ซึ่งจะใช้ซอฟแวร์นี้ในการวิเคราะห์ความถี่ฟอร์แมนท์ที่ 1 ถึง 4 แล้ว ความถี่ฟอร์แมนท์เหล่าจะถูกปรับปรุงในภายหลังโดยใช้กฎแล้วส่งต่อไปให้ synthesizer ดังแสดงในรูปที่ 2.3 การทดลองทำการสังเคราะห์เสียงโดยเปรียบเทียบกับวิธีอ้างอิง คือ KTH rule-based formant synthesis system โดยทำการสังเคราะห์ภาษา Swedish 4 คำพูด แล้วทำการเปรียบเทียบความความเป็นธรรมชาติของเสียง ผลที่ได้คือ วิธีการที่นำเสนอ ดีกว่าวิธีการอ้างอิง 29 เปอร์เซ็นต์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

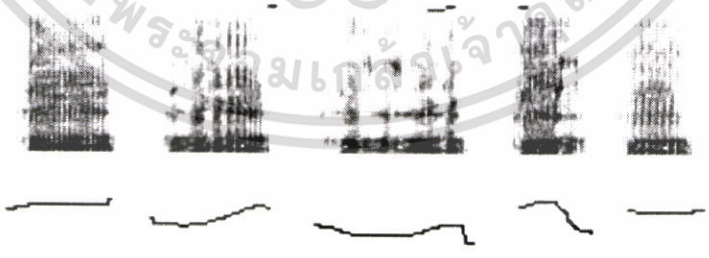
วิธีการที่นำเสนอนี้ คิดว่ามีความเหมาะสมกับกลุ่มภาษาตะวันตก เนื่องจากโครงสร้างของภาษาค่อนข้างใกล้เคียงกัน ต่างกันแค่การออกเสียงบางอย่างไม่เหมือนกัน ซึ่งวิธีนี้ทำให้สามารถนำไปใช้กับหลายๆ ภาษาได้ โดยการบันทึกเสียงของภาษานั้นๆ เข้าไป แต่วิธีนี้ไม่เหมาะกับภาษาไทย เนื่องจากภาษาไทยเป็นภาษาที่มีวรรณยุกต์ และเป็นการเปลืองทรัพยากรมากในการเก็บข้อมูลเสียงพูดเข้าไปเพื่อใช้ในการวิเคราะห์พารามิเตอร์



รูปที่ 2.3 Rule based synthesis system using data-driven unit library

2.1.2 งานวิจัยที่เกี่ยวข้องกับการสังเคราะห์เสียงแบบฟอร์แมนท์ในประเทศจีน

ภาษาจีนเป็นภาษาที่มีเสียงวรรณยุกต์ แต่ละพยางค์จะมี 5 ชนิดโทนเสียง จากโทน 1 ถึงโทน 5 เช่น /ma1/ /ma2/ /ma3/ /ma4/ ดังแสดงในรูปที่ 2.4 [5]



รูปที่ 2.4 สเปกตรัม (ด้านบน) และขอบของความถี่มูลฐาน (ด้านล่าง) ของพยางค์ /ma1/ /ma2/ /ma3/ /ma4/ /ma5/ ในภาษาจีน

การสังเคราะห์เสียงแบบฟอร์แมนท์สำหรับภาษาจีน (Mandarin) ซึ่งริเริ่มพัฒนาโดย Suen ในปี 1976 ซึ่งการพัฒนาระบบการสังเคราะห์เสียงจากข้อความในภาษาจีน ดังแสดงในตาราง ที่ 2.1

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 2.1 การพัฒนาระบบสังเคราะห์เสียงจากข้อความในภาษาจีน [5]

Inst. or author	Synth. Meth.	Unit	Text Anal.	Prosody	date
Suen	VOTRAX	phoneme	no	5 tones	1976
Lee et.al.	VOTRAX	phoneme	no	5 tones	1983
Huang et.al.	LPC	pseudo-demisyllable	no	4 tones	1983
Zhou et.al.		pseudo-demisyllable	no	7 tones	1984
Lin & Luo	LPC	toned phoneme	no	no	1985
Taiwan U.	LPC	syllable	no	rules	1985
KTH	formant		no	rules	1986
Bell Labs	LPC	diphone	no	rules	1987
Taiwan U.	formant	syllable	no	rules	1987
Tsinghua (Beijing)	LPC	demisyllable	no	4 tones	1987
Acad. of Soc. Sci.	formant	syllable	no	stat. model	1988
Qin & Hu	LPC	pseudo-demisyllable	no	4 tones	1988
U.C. London	formant	pseudo-demisyllable	no	rules	1989
Matsushita	formant	pseudo-demisyllable	no	model	1989
Telecom. Labs	LPC	syllable	yes	rules/stat. model	1989
Tsing Hua U.	hybrid	demisyllable	no	rules/stat. model	1991
Telecom. Labs	LPC	toned syllable	yes	rules	1991
Chiao Tung U.	PSOLA	syllable		stat. model	1992
Chiao Tung U.	LPC	syllable		stat. model	1992
Hong Kong U.	LPC	syllables	no	rules	1992
Xu et.al.		words	yes	rules	1993
Bell Labs	LPC	diphone	yes	rules/stat. model	1994
Inst. Acoustics	formant		yes	rules	1994
Apple	Apple	diphone	yes	rules	1994
Tsinghua (Beijing)	PSOLA	toned syllable	yes	rules	1995
Inst. Acoustics	PSOLA	toned syllable	yes	rules	1995
Cheng Kung U.	CELP	toned syllable	yes	stat. model	1995

การออกเสียงสำเนียงสูงต่ำของภาษาจีน ได้จากการคำนวณค่าของความถี่มูลฐาน (F0) ของเสียงที่สังเคราะห์ขึ้น สิ่งที่สำคัญประการหนึ่งเกี่ยวกับเสียงสูงต่ำของภาษาก็คือ การวิเคราะห์รูปแบบการผันแปรของโทนเสียง การออกเสียงวรรณยุกต์ที่ได้จากส่วนของการวิเคราะห์แล้วจะเข้าสู่กฎการออกเสียง ในขั้นตอนนี้จะแบ่งลำดับของเสียงวรรณยุกต์ที่ต้องการโดยปราศจากการเน้นเสียงของประโยคซึ่งจะทำการเพิ่มในภายหลังต่อไป

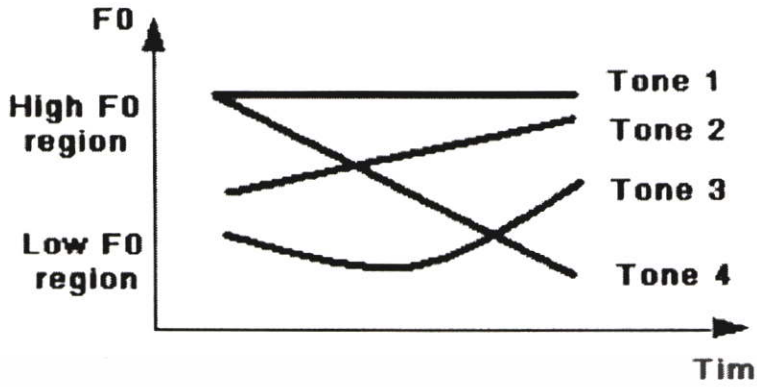
ภาษาจีนมีโทนเสียงสูงต่ำ 4 โทนเสียง และมีโทนเสียงกลางอีก 1 โทนเสียง หรือเรียกว่า Tone 0 ดังแสดงในตารางที่ 2.2 ซึ่งตัวอักษร H, M และ L หมายถึง high, mid และ low ตามลำดับ

ตารางที่ 2.2 ลักษณะโทนเสียงภาษาจีน [5]

Tone Symbol	Tone 1	Tone 2	Tone 3	Tone 4	Tone 0
Tone Name	High Level	Rising	Falling-rising	Falling	Neutral
Targets	H H	M H	M L (M+)	H+L	M

เมื่อเขียนแผนภูมิแสดงความถี่มูลฐานบนแกนเวลาของแต่ละโทนเสียง ดังแสดงในภาพที่ 2.5

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



ภาพที่ 2.5 แผนภูมิแสดงความถี่มูลฐานบนแกนเวลาของโทนเสียงภาษาจีน [5]

เนื่องจากภาษาจีนเป็นภาษาที่วรรณยุกต์เช่นเดียวกับภาษาไทย สิ่งที่น่าสนใจเกี่ยวกับการสังเคราะห์เสียงภาษาจีน คือ รูปแบบการการจัดการเกี่ยวกับวรรณยุกต์สำหรับการสังเคราะห์เสียงโดยกระบวนการคำนวณค่าความถี่มูลฐานของโทนเสียงมีดังต่อไปนี้

กำหนดค่าแรกของโทน 2 ซึ่งเป็นบริเวณใกล้ๆ กับบริเวณกึ่งกลางของระดับทำนองเสียง (เมื่อทำนองเสียงคือ ระดับพิทช์ในการออกโทนเสียงในภาษาจีน) ซึ่งเป็นตำแหน่งอิสระตำแหน่งแรกสำหรับการออกทำนองด้วยโทนก่อนหน้านี้ ที่ค่าแรกของโทน 4 มีค่าสูงกว่าระดับ H ปกติ ซึ่งจะแทนด้วยสัญลักษณ์ H+ ซึ่งเป็นตำแหน่งค่อนเข้าไปในระดับทำนองเสียงสูงสุด

รูปทรงของเสียงโทนกลาง(โทน 0) ไม่คงที่ และจะขึ้นอยู่กับค่าเดิมของโทนเสียงก่อนหน้านี้ ในกรณีทั่วไปเสียงโทนกลางจะต่อท้ายจากโทน 1, 2, และ 4 ซึ่งมีรูปทรงที่ต่ำลง แต่โทนเสียงกลางตามหลังโทน 3 จะมีรูปทรงที่สูงขึ้น สำหรับ 4 โทนเสียงนี้ เราสามารถกำหนดจุดมุ่งหมายของโทนที่ช่วงเวลาของทำนองเสียง ซึ่งแน่ใจได้ว่ามีค่าค่อนข้างคงที่ของรูปทรงของโทน ซึ่งกฎที่สำคัญสำหรับการผันวรรณยุกต์ภาษาจีน มีดังนี้

1. ถ้าโทน 4 ไม่เป็นส่วนท้ายประโยค แล้วทำการเปลี่ยน L ไปเป็นM
2. กำหนดค่าของโทน 2 ถ้าโทนที่ตามมาทีหลังเริ่มต้นด้วย H แล้ว ทำการลบค่า H ของโทน 2
3. ถ้าโทน 0 ตามหลังโทน 4 แล้วทำการเปลี่ยนค่า M ของโทน 0 ให้เป็น L
4. ถ้าโทนที่ตามหลังมาเริ่มต้นด้วย H แล้ว ค่า L ของโทน 3 จะเพิ่มค่าขึ้นเป็น L=L+0.15(M-L)
5. ถ้าโทน 3 เป็นส่วนท้ายของประโยค แล้วทำการเพิ่มค่า M โดยกำหนดค่าเป็น M=M+0.3(H-M)

จากที่กล่าวมาข้างต้นจะพบว่าการผันวรรณยุกต์ภาษาจีนแต่ละโทนเสียงจะมีค่าไม่คงที่ต้องอาศัยค่าของโทนเสียงที่ตามหลังและค่าของโทนเสียงที่อยู่ก่อนหน้ามาทำการคำนวณด้วย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ดังนั้นการสังเคราะห์เสียงภาษาจีนวิธีที่เหมาะสมที่สุดก็คือ การสังเคราะห์เสียงโดยใช้กฎ ซึ่งถ้าใช้วิธีการอื่น อย่างเช่นวิธีการ Concatenation Synthesis จะทำให้เปลืองเนื้อที่ในการเก็บข้อมูลหน่วยเสียงเป็นจำนวนมากเนื่องจากเสียงวรรณยุกต์มีการเปลี่ยนแปลงไม่คงที่ ซึ่งต่างจากวรรณยุกต์ของภาษาไทย ที่มีเสียงวรรณยุกต์ 5 เสียงเช่นเดียวกัน แต่ว่าแต่ละเสียงวรรณยุกต์จะมีทิศทางของความถี่มูลฐานคงที่ ซึ่งสามารถกำหนดกฎเป็นค่าสมการทางคณิตศาสตร์ของแต่ละเสียงวรรณยุกต์ได้

2.1.3 งานวิจัยที่เกี่ยวกับการสังเคราะห์เสียงภาษาไทย

ในปี 2539 กมลชัย โชคชัยชุติกุล [7] ได้นำเสนอระบบคอมพิวเตอร์อ่านออกเสียงภาษาไทยจากเอกสาร ซึ่งเป็นการอ่านข้อความจากสแกนเนอร์ แล้วทำการอ่านออกเสียงออกมาตามข้อความในเอกสารนั้น ในส่วนของการสังเคราะห์เสียงจะทำการค้นหาคำอ่านออกเสียงโดยใช้ต้นไม้คำศัพท์ เมื่อได้พอยเตอร์ที่ชี้ไปยังข้อมูลเสียงคำศัพท์ที่ค้นหาแล้ว ระบบจะทำการอ่านข้อมูลเสียงจากแฟ้มข้อมูลเสียง ณ จุดที่พอยเตอร์ชี้แล้วทำการอ่านออกเสียง เนื่องจากเป็นเสียงที่ได้จากการบันทึกจากคนจริงและเก็บไว้เป็นคำพยางค์เดี่ยว ดังนั้นคุณภาพของเสียงอ่านจึงค่อนข้างคมชัด แต่ จะมีปัญหาเรื่องความราบเรียบของเสียง เพราะการทำการบันทึกทีละคำในเวลาที่แตกต่างกัน ทำให้เสียงที่ได้ไม่ต่อเนื่อง ระบบที่นำเสนอนี้ต้องทำการบันทึกข้อมูลคำพูดเป็นคำนวณมากเพื่อให้ครอบคลุมเสียงภาษาไทย ซึ่งต้องใช้เนื้อที่ในการจัดเก็บข้อมูลเสียงพูดมาก และระบบไม่มีความยืดหยุ่น

และในปีเดียวกัน วิเชียร แซ่โล้ว [8] นำเสนอเรื่อง อ่านคำไทยพยางค์เดี่ยวแบบปรับความเร็วได้ ซึ่งสามารถปรับความเร็วได้ 4 ระดับ โดยไม่ทำให้ความถี่มูลฐานและคุณลักษณะทางเสียงเปลี่ยนไป โดยการสังเคราะห์เสียงใช้วิธีการบันทึกเสียงพูดทีละพยางค์ที่ต้องการ จากนั้นเมื่อระบบต้องการที่จะพูดคำใดก็จะทำการค้นหาเสียงที่ต้องการจากฐานข้อมูลที่เก็บไว้ ระบบนี้จะออกเสียงได้จำกัดเท่ากับจำนวนพยางค์ที่บันทึกไว้ และนำเสนอวิธีแก้ปัญหาพิทช์ของเสียงที่ได้จากการบันทึกเมื่อทำการปรับความเร็วแล้วทำให้พิตช์ของเสียงเปลี่ยนไป กล่าวคือ ถ้าเพิ่มความเร็วเสียงก็จะทำให้พิทช์สูงขึ้น ถ้าความลดความเร็วลงก็จะทำให้พิทช์ต่ำลง ซึ่งทำการแก้ปัญหาโดย การกำหนดการออกเสียงเป็นแต่ละรายคาบ ที่ระดับความเร็วปกติจะออกเสียงทุกลูกคลื่นที่เรียงต่อกันไม่มีการเว้น ถ้าเป็นความเร็วระดับหนึ่ง จะออกเสียงลูกคลื่น 3 ลูกติดกันและไม่ออกเสียง 1 ลูกคลื่น ที่ระดับความเร็วระดับสอง จะออกเสียงสองลูกคลื่นติดกันและไม่ออกเสียง 2 ลูกคลื่น

การแก้ปัญหาดังกล่าวจะทำให้คุณสมบัติบางส่วนของคลื่นเสียงหายไป ซึ่งจะมีผลต่อคุณภาพของเสียงได้ และทำให้เกิดความไม่ราบเรียบของเสียง และการแก้ปัญหาเรื่องความเร็วเสียงโดยไม่ให้กระทบต่อความถี่มูลฐาน โดยการตัดช่วงบางช่วงของความเป็นรายคาบทิ้งไปนั้นอาจมีผลกระทบต่อการออกเสียงวรรณยุกต์ภาษาไทยได้ เนื่องจากเสียงวรรณยุกต์จะขึ้นอยู่กับความถี่มูลฐาน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ในปี 2540 สง่า คงสุพานิช [9] ได้นำเสนอระบบแปลงหน่วยคำภาษาไทยเป็นสัญลักษณ์แทนเสียงสำหรับงานสังเคราะห์เสียงจากประโยคภาษาไทย แล้วผ่านการบวนการแยกหน่วยคำ โดยวิธี Suited Length word mapping ได้หน่วยพยางค์สำหรับการวิเคราะห์โครงสร้างพยางค์เพื่อแปลงเป็นสัญลักษณ์แทนเสียง โดยวิธี MPT-network จากนั้นทำการรวมหน่วยเสียงตามลำดับของสัญลักษณ์ที่ได้ เพื่อสร้างเป็นเสียงพูดจากหน่วยเสียงที่จัดเก็บเป็นไฟล์เสียง มีจำนวนจำกัดคือ 1,452 หน่วยเสียง เพื่อรวมเป็นหน่วยเสียงภาษาไทย

วิธีการการนี้เป็นการนำเสนอเพื่อนำไปสังเคราะห์เสียงด้วยวิธี Concatenation Synthesis สามารถแก้ปัญหาเรื่องจำนวนของคำที่ต้องบันทึกเป็นจำนวนมากที่นำเสนอโดย [8] และ[9] ได้ แต่จำนวนหน่วยเสียงที่ใช้ก็ยังมากอยู่ และมีปัญหาเรื่องการนำหน่วยเสียงมาต่อกัน ซึ่งจะเกิดความไม่ราบเรียบ จากสาเหตุจากการบันทึกเสียงที่ในเวลาที่ต่างกัน และการตัดหน่วยเสียงที่ต้องการต้องทำอย่างประณีต และการออกเสียงวรรณยุกต์ภาษาไทยที่มีระดับของพิทช์ต่างกัน ทำให้บริเวณรอยต่อของแต่ละหน่วยเสียงไม่ราบเรียบ

ในปี 2541 NECTEC [1] ได้นำเสนอระบบต้นแบบสังเคราะห์เสียงพูดภาษาไทย โดยใช้ Unit Selection ซึ่งเป็นส่วนหนึ่งของวิธี Concatenation Synthesis เป็นการบันทึกเสียงคำภาษาไทยจำนวน 5,200 คำ โดยทำการบันทึกในห้องที่มีการควบคุมสภาพแวดล้อมของเสียง โดยทำการบันทึกที่ 16 bit/sec และ sampling rate 44.1 kHz ทำการควบคุมคลื่นเสียงระหว่างการบันทึก เพื่อทำให้เสียงที่บันทึกมีคุณภาพที่สุดและทำให้เกิดความราบเรียบในการในการสังเคราะห์เสียง งานวิจัยที่พัฒนาการสังเคราะห์เสียงภาษาไทยส่วนใหญ่ของ NECTEC จะเน้นในส่วนของการวิเคราะห์ข้อความเกี่ยวกับภาษาไทย แต่ในส่วนของการสังเคราะห์เสียงใช้วิธีการที่มีอยู่แล้ว ซึ่งวิธีการนี้ต้องทำการบันทึกคำพูดเป็นจำนวนมากเพื่อให้ครอบคลุมภาษาไทยทั้งหมด และต้องทำการควบคุมคุณภาพของการบันทึกและผู้เป็นผู้บอกภาษาต้องคัดเลือกผู้ที่ออกเสียงภาษาไทยได้ถูกต้อง วิธีการนี้ใช้เนื้อที่ในการเก็บข้อมูลจำนวนมาก แต่ให้คุณภาพของเสียงที่ดี

ในปี 2542 ชัชวาล หาญสกุลบรรเทิง [10] นำเสนอการสังเคราะห์พยางค์ภาษาไทยด้วยวิธีการสังเคราะห์แบบวิเคราะห์โดยใช้คู่เส้นสเปกตรัม ซึ่งจะเก็บหน่วยเสียงของพยางค์ด้วยการบันทึกจากเสียงมนุษย์ แล้วแก้ปัญหาเกี่ยวกับเสียงวรรณยุกต์ด้วยวิธีทีดี-โซลา (Time-Domain Pitch-Synchronous Overlap Add : TD-PSOLA) และทำการเสียงให้ราบเรียบของรอยต่อของหน่วยเสียงด้วยคู่เส้นสเปกตรัม ซึ่งสามารถแก้ปัญหางานวิจัยของ [7], [8] และ[9] ได้

การสังเคราะห์นี้ใช้ฐานข้อมูลซึ่งประกอบด้วย หน่วยเสียงที่เก็บในรูปแบบคู่เส้นสเปกตรัมจำนวน 376 หน่วยเสียง รูปแบบความถี่มูลฐาน 5 รูปแบบ โดยทำการสังเคราะห์เสียงผู้ชายที่มีคุณภาพเสียงที่อัตราสุ่ม 22,050 Hz 16 bit/sec ผลการประเมินความถูกต้องของเสียงมีค่าเท่ากับ 78%

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.2 กายวิภาคของระบบการเปล่งเสียงของมนุษย์

มนุษย์สร้างเสียงพูดขึ้นมาได้โดยมนุษย์จะมีเส้นเอ็นในลำคอที่เรียกว่าเส้นเสียง (Vocal Cords) ซึ่งอยู่ในช่องหลอดลม และจากการสั่นของเส้นเสียงจะทำให้เกิดคลื่นเสียงความถี่ต่างๆผ่านเข้าสู่ลำคอ ผ่านจากลำคอเข้าสู่ปากหรือช่องจมูกออกไปภายนอกซึ่งขนาดและรูปร่างของอวัยวะภายในช่องปากนี้จะเป็นสิ่งกำหนดว่าคลื่นเสียงความถี่ไหนจะยอมให้ปรากฏออกมาให้ได้ยิน (Passed Through) หรือคลื่นเสียงความถี่ไหนที่จะถูกดูดซับไว้ไม่ยอมให้ปรากฏออกมา (Absorbed) ปรากฏการณ์ที่ทำให้เกิดเสียงพูดออกมาได้นี้ เนื่องจากรูปร่างของช่องคอและช่องปากของมนุษย์ที่สามารถเปลี่ยนแปลงไปมาได้อย่างอิสระ ทำให้ผู้พูดสามารถกำหนดระดับเสียงต่างๆ เหล่านี้ได้เองจึงทำให้เกิดการเปล่งเป็นเสียงพูดออกมา [11]

2.2.1 อวัยวะออกเสียง (The Organs of Speech)

ธรรมชาติได้กำหนดตำแหน่งของอวัยวะที่ใช้ในการทำให้เกิดเสียงของมนุษย์เอาไว้นั้นคือ ริมฝีปาก ฟัน เพดาน และลิ้น ที่อยู่ตามลำดับตลอดช่องของกล้ามเนื้อช่องคอ เพื่อทำให้เปลี่ยนขนาดและรูปร่างได้ตามต้องการ ดังรูป 2.6 เราจะเปล่งคำพูดออกมาได้โดยการเปล่งออกมาเป็นลำดับขั้นตอนของความถี่คลื่นเสียงที่ได้คัดเลือกไว้ออกมา นั่นคือ จะมีการเปลี่ยนตำแหน่งของช่องทางเดินเสียง (Vocal Tract) ของมนุษย์เป็นลำดับขั้นอย่างต่อเนื่อง

อวัยวะที่ใช้ในการออกเสียงนั้น แบ่งได้เป็น 2 พวก คือ [12]

1) อวัยวะที่เป็นส่วนกระทำอาการ (Articulator) หมายถึง อวัยวะส่วนที่เคลื่อนไหวเพื่อผลักลมไปยังส่วนต่างๆ อวัยวะที่เป็นส่วนกระทำการที่สำคัญ คือ ลิ้น ซึ่งเป็นส่วนที่เคลื่อนไหวได้มากที่สุด

2) อวัยวะที่เป็นฐานที่เกิดเสียงต่างๆ (Point of articulator) หมายถึง ตำแหน่งหรือฐานที่เกิดของเสียงต่างๆ เช่น ริมฝีปาก ฟัน เพดานส่วนต่างๆ เป็นต้น

จากรูปที่ 2.5 อวัยวะส่วนที่มีหน้าที่ในการออกเสียงพูดมีดังนี้

ริมฝีปาก (Lips) เป็นอวัยวะส่วนที่สามารถเคลื่อนไหวได้มาก และทำให้เสียงแตกต่างกันได้มาก เราอาจบังคับริมฝีปากให้อยู่ชิดกัน ห่างกัน ยื่นออกมา หรือห่อกลม ฯลฯ ก็ได้ ลักษณะริมฝีปากต่างๆ นี้ ล้วนมีผลต่อการออกเสียง และทำให้เสียงแตกต่างกันไปทั้งสิ้น

ฟัน (Teeth) เป็นอวัยวะที่เกิดของเสียงหลายชนิด เช่น เมื่อฟันบนกดลงบนริมฝีปากล่างหรือกดกับฟันล่าง ลมที่ผ่านออกมาโดยแรกจะลอดช่องที่พอผ่านได้ออกมาทำให้เกิดเป็นเสียงชนิดที่เรียกว่า เสียงเสียดแทรกที่เกิดที่ฟัน เป็นต้น

ปุ่มเหงือก (Gum ridge) คือ ส่วนที่อยู่ต่อจากฟันฟันบน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เพดานแข็ง หรือเพดานปาก (Hard palate) หมายถึง เฉพาะเพดานที่โค้งเป็นกระดูกแข็งเท่านั้น

เพดานอ่อน (Soft palate) คือ ส่วนของเพดานที่อยู่ต่อเพดานแข็งไปข้างใน มีลักษณะเป็นกระดูกอ่อนที่ขยับขึ้นลงได้เวลาหายใจเพดานอ่อนและลิ้นไก่ซึ่งอยู่ตอนปลายเพดานอ่อนจะลดระดับลงมาเปิดช่องให้ลมออกไปทางจมูก เวลาพูดส่วนใหญ่ปลายเพดานอ่อนและลิ้นไก่จะถูกยกขึ้นไปจรดกับหลังคอ นอกจากเวลาออกเสียงนาสิกเท่านั้น ที่เพดานอ่อนจะลดระดับลงมา

ลิ้น (Tongue) เป็นส่วนที่เคลื่อนไหวมากที่สุดในการออกเสียงพูด จึงต้องแบ่งเป็น 3 ส่วน ตามหน้าที่ในการออกเสียง คือ

1) ปลายลิ้น (Blade of the tongue) หมายถึง ส่วนปลายที่สามารถยกขึ้นไปแตะอวัยวะส่วนต่างๆ ในปากตอนบนได้ง่าย

2) หน้าลิ้น (Front of the tongue) หมายถึง ลิ้นที่อยู่ตรงข้ามกับเพดานแข็งขณะที่วางลิ้นราบกับปากตอนที่ไม่ได้พูด

3) หลังลิ้น (Back of the tongue) หมายถึง ส่วนของลิ้นที่อยู่ตรงข้ามกับเพดานอ่อน ขณะที่ลิ้นวางราบกับปาก

ช่องปาก (Oral cavity) ช่องปากนี้เป็นอวัยวะอีกส่วนหนึ่งที่มีความสำคัญในการพูด กล่าวคือ ช่องปากจะทำหน้าที่เป็นช่องกำทอน (Resonant chamber) ซึ่งมีบทบาทในการกำทอนเสียงที่เดินทางผ่านมาถึงบริเวณนี้ ทั้งนี้เพราะช่องปากสามารถจะเปลี่ยนแปลงเป็นช่องกำทอนรูปร่างต่างๆ กัน เนื่องจากรูปร่างของมันแปรผันไปตามการจัดท่าทางของลิ้น ริมฝีปากและขากรรไกร

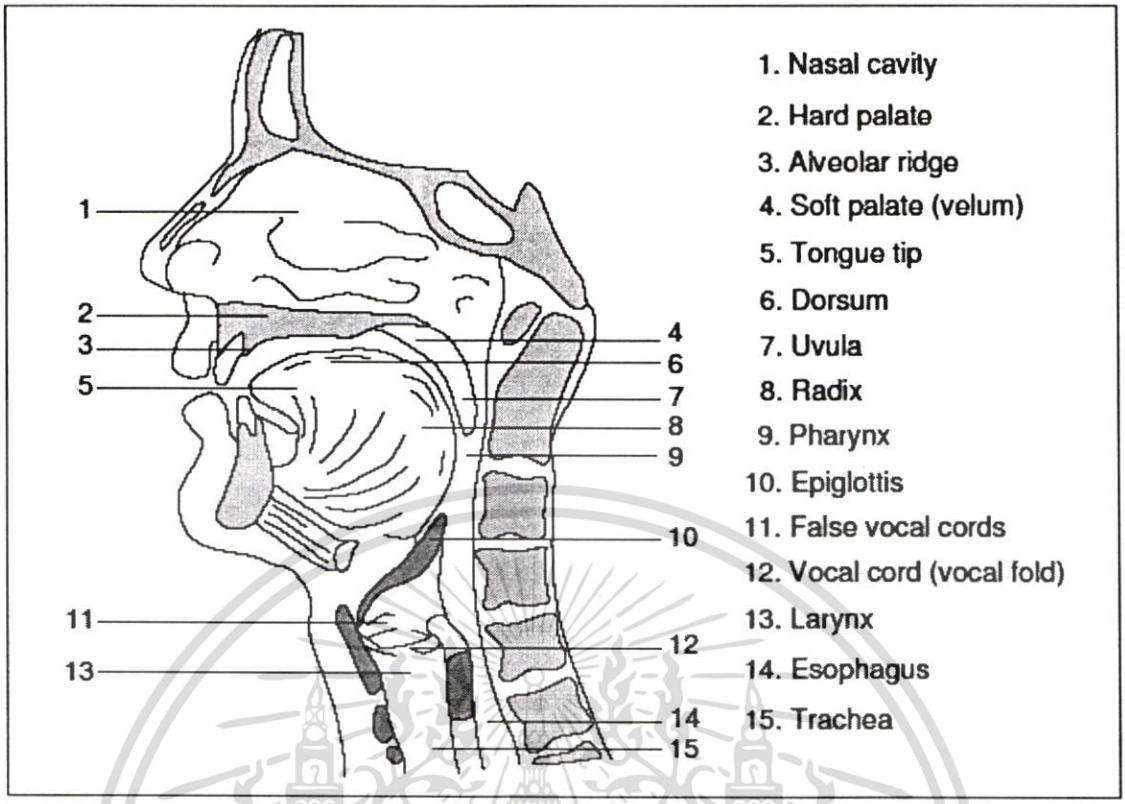
โพรงจมูก (Nasal cavity) คุณสมบัติของเสียงพูดที่เกิดขึ้นจะแปรผันไปตามการปิด-เปิดของช่องทางออกสู่โพรงจมูก ซึ่งเป็นผลจากการยกขึ้นหรือเลื่อนลงของเพดานอ่อน

ช่องคอ (Pharynx) หมายถึงโพรงคอซึ่งอยู่ถัดไปจากปากลงไปจนถึงเส้นเสียง

ลิ้นปิดหลอดลม (Epiglottis) เป็นก้อนเนื้อเล็กๆ คล้ายลิ้นไก่ อยู่ต่อจากโคนลิ้นลงไปในลำคอ มีหน้าที่ปิดช่องลมขณะรับประทานอาหาร และปิดช่องลมและจมูก

เส้นเสียง (Vocal cords) เป็นอวัยวะสำคัญที่เกิดของเสียง เส้นเสียงมีลักษณะเป็นกล้ามเนื้อ 2 แผ่น ปิดขวางอยู่ปากช่องหลอดลมจากด้านหลังมาด้านหน้า ระหว่างเส้นเสียงจะมีช่องว่างระหว่างเส้นเสียง (Glottis) เส้นเสียงทั้งสองสามารถดึงออกให้ห่างจากกัน หรือดึงเข้ามาประชิดกันก็ได้ เส้นเสียงเป็นส่วนทำให้เกิดเสียงพูดขึ้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.6 อวัยวะในระบบการพูดของมนุษย์ [3]

2.2.2 กระบวนการผลิตเสียงพูด

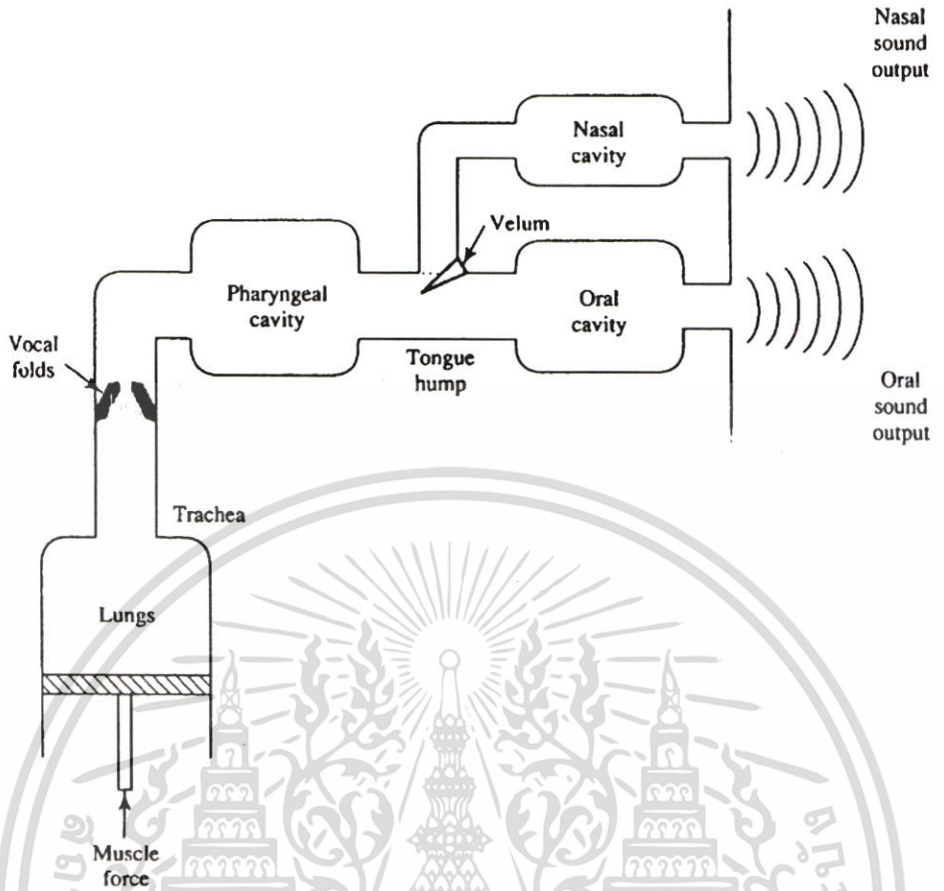
จากระบบเสียงพูดของมนุษย์ สามารถแสดงเป็นแผนภาพของระบบกำเนิดเสียงดังรูปที่ 2.6 ซึ่งสามารถจำแนกกลไกสร้างเสียงพูดของมนุษย์ได้ 3 แบบ ดังนี้

1) อากาศไหลจากปอดจะถูกมอดูเลท (Modulate) โดยการสั่นของเส้นเสียงทำให้เกิดคลื่นเสียงลักษณะคล้ายพัลส์ที่มีคาบเวลาแบบควอไซ (Quasi_periodic pulse-like excitation)

2) อากาศไหลจากปอดถูกให้ปั่นป่วนด้วยการบังคับให้ไหลผ่านช่องแคบอันเกิดจากการบีบตัวของอวัยวะในช่องปากทำให้เกิดเสียงลักษณะคล้ายเสียงรบกวน (Noise-like excitation)

3) อากาศที่ไหลถูกกัก และเกิดแรงดันอยู่ภายในส่วนของช่องปากที่ปิด จากนั้นจึงปล่อยให้อากาศให้อากาศที่มีแรงดันพุ่งออกไปอย่างรวดเร็ว ทำให้เกิดการกระตุ้นเป็นเสียงในช่วงเริ่มต้น (Transient excitation)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.7 แผนภาพระบบเสียงพูดของมนุษย์ [13]

2.2.3 เสียงพูดของมนุษย์

เสียงพูดเป็นคลื่นความยาว (Longitudinal wave) เกิดจากการสั่นของอนุภาคตัวกลาง และทิศทางการสั่นของอนุภาคจะอยู่ในทิศเดียวกันกับทิศทางของการเคลื่อนที่ คลื่นเสียงเป็นคลื่นที่เปลี่ยนแปลงไปตามเวลา แบ่งได้เป็น 2 ชนิด ตามการกำเนิดหรือโหมด (Mode) การกระตุ้น คือ

1) เสียงก้อง หรือเสียงโฆษะ (Voice sound) เกิดจากการบังคับอากาศให้ผ่านสายเสียงทำให้มีการเปลี่ยนแปลงความตึงหย่อนของเส้นเสียง โดยเส้นเสียงจะสั่นและเกิดเป็นพัลส์ (Pulse) ของอากาศไปกระตุ้นอวัยวะกำทอนเกิดเป็นเสียงก้อง ตัวอย่างเสียงก้อง ได้แก่ เสียงสระ เสียงพยัญชนะที่ออกเสียงจากลำคอ (Voiced consonants) เช่น อา อี อู ว ย เป็นต้น

2) เสียงไม่ก้อง หรือเสียงอโฆษะ (Voiceless หรือ Unvoiced sounds) เป็นเสียงที่เกิดในช่องปากหรือโพรงจมูกโดยอวัยวะภายในช่องปาก ริมฝีปาก ขวางการไหลของอากาศได้เป็นช่องเล็กๆ อากาศจึงไหลผ่านอย่างรวดเร็วและปั่นป่วนจนกระทั่งสร้างเป็นเสียงรบกวนช่วงความถี่กว้าง (Broad-spectrum noise) ตัวอย่างเสียงไม่ก้อง ได้แก่ เสียงพยัญชนะที่ไม่ได้เกิดจากลำคอ (Voiceless consonants) เช่น ซ ช ฟ เป็นต้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เสียงพูดของมนุษย์ที่เปล่งออกมา และมนุษย์ได้ยินจะมีคุณสมบัติ ดังต่อไปนี้

1) ความยาวของเสียง (Length) เป็นช่วงเวลาของการเกิดเสียงใดเสียงหนึ่งเปล่งออกมา เสียงพูดบางเสียงอาจจะเปล่งออกมาได้ติดต่อกันได้นาน เช่น เสียงสระ เสียงพยัญชนะที่เป็นเสียงสอดแทรก เป็นต้น สำหรับในภาษาไทย เสียงพูดที่มีความยาว-สั้น ก็มีเพียงเสียงสระเท่านั้น เช่น อะ อิ อุ เป็นเสียงสั้น อา อี อู เป็นเสียงยาว เป็นต้น

2) ระดับเสียงสูง-ต่ำ (Pitch) เสียงพูดจะมีระดับเสียงสูงหรือต่ำ อยู่ที่ความถี่มูลฐานของเสียง (Fundamental frequency) ถ้าความถี่ต่ำเสียงก็จะต่ำ อวัยวะส่วนที่ทำให้เสียงมีระดับ สูง-ต่ำ คือ เส้นเสียง ดังนั้น ระดับเสียงสูง-ต่ำ ก็คือ อัตราการสั่นสะบัดของเส้นเสียงนั้นเอง ในการพูดเสียงที่มีระดับสูง-ต่ำ ได้คือเสียงก้องเท่านั้น เพราะมีการสั่นสะเทือนของเส้นเสียง ที่ทำให้เกิดมีความถี่ระดับต่างๆ ได้ ในภาษาไทยระดับเสียงสูง-ต่ำ ของคำ เราเรียกว่า วรรณยุกต์

3) ความดัง (Loudness) ความดังขึ้นอยู่กับปริมาณของลม ที่ผู้พูดเปล่งเสียงออกมาในช่วงเวลาหนึ่งๆ

4) การลงน้ำหนัก (Stress) เป็นการออกเสียงพยางค์ใดพยางค์หนึ่งให้ดังเน้นมากหรือน้อยกว่าพยางค์อื่นๆที่อยู่ข้างเคียง (แสดงอารมณ์อย่างใดอย่างหนึ่ง)

5) ช่วงต่อของเสียง (Juncture) ช่วงระยะที่ผู้พูดเปล่งเสียงหนึ่งแล้วต่อไปเปล่งเสียงอีกเสียง ซึ่งเรียงกันมาเป็นลำดับเสียงที่ประกอบกันเข้าเป็นพยางค์ จะมีช่วงต่อของเสียงแนบสนิทจนไม่เห็นร่องรอย (Close juncture) แต่ถ้าเสียงปรากฏอยู่คนละพยางค์ หรือคนละคำจะมีช่วงต่อห่างจนสังเกตได้ชัด (Open juncture) ดังนั้นช่วงต่อของเสียง โดยเฉพาะช่วงต่อห่างจะเป็นลักษณะการแบ่งวรรคตอนของเสียงพูด

2.3 หน่วยเสียงสำคัญในภาษาไทย

"หน่วยเสียง" (Phoneme) เป็นหน่วยเล็กที่สุดของภาษา หน่วยดังกล่าวได้แก่เสียงสำคัญๆ ในภาษาใดภาษาหนึ่งซึ่งทำหน้าที่ให้ความหมายของคำที่ใช้ในภาษานั้น และทำให้ความหมายของคำนั้นๆ มีความหมายแตกต่างจากคำอื่นๆ หน่วยเสียงสำคัญในภาษาไทยมี 3 ประเภทใหญ่ๆ คือ เสียงพยัญชนะ เสียงสระ และเสียงวรรณยุกต์ หน่วยเสียงทั้ง 3 นี่เองที่ประกอบกันเข้าเป็นคำที่ใช้ในภาษาไทย

เสียงพูดของมนุษย์ซึ่งมีความแตกต่างกันมากมายนั้น ถ้าเราพิจารณาอย่างถ้วนๆ จะพบว่าสามารถแบ่งออกเป็น 2 ประเภทใหญ่ คือ

1. เสียงเรียง (Segmental sound) เป็นหน่วยเสียงที่แยกออกจากเสียงอื่นได้โดยเด็ดขาด เพราะมันมีลักษณะเด่นเฉพาะตัว ในภาษาไทยได้แก่เสียงสระ และเสียงพยัญชนะ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สำนักหอสมุดกลาง พระจอมเกล้าลาดกระบัง

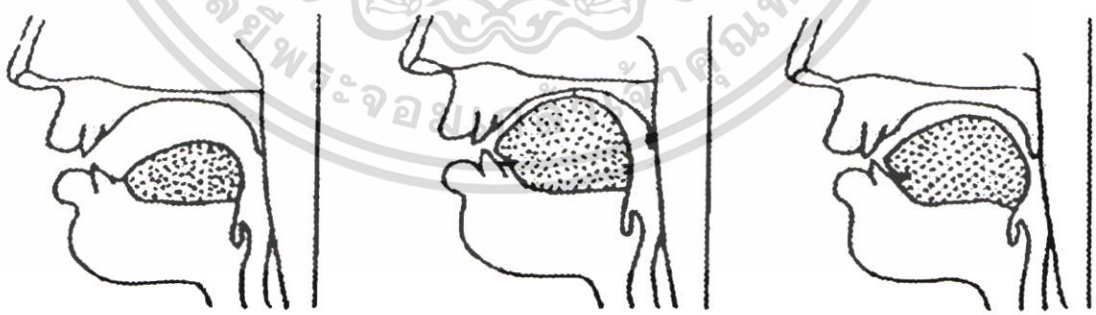
2. เสียงซ้อน (Supra-segmental feature) เป็นหน่วยเสียงที่ทำหน้าที่เป็นส่วนประกอบของเสียงอื่นเพราะไม่สามารถแยกเปล่งเสียงได้ตามลำพัง ในภาษาไทยได้แก่เสียงวรรณยุกต์ และทำนองเสียงเป็นต้น

2.3.1 หน่วยเสียงสระ

2.3.1.1 ลักษณะของเสียงสระ

ลักษณะสำคัญของเสียงสระก็คือ เป็นเสียงก้องที่เปล่งเสียงออกมาโดยให้ลมออกทางช่องปากโดยไม่ถูกลิ้นกักหรือขัดขวาง ดังนั้นเวลาเราออกเสียงสระจะออกเสียงได้สะดวกและออกเสียงได้นาน ทั้งนี้เพราะคุณสมบัติของเสียงสระมีความดังเด่นกว่าสียงอื่นๆ ที่เรียงอยู่ข้างเสมอ อวัยวะที่เกี่ยวข้องกับการออกเสียงสระได้แก่ ลิ้น กับริมฝีปาก ถ้าลิ้นส่วนใดทำหน้าที่เพียงส่วนเดียว เสียงที่เกิดขึ้นก็จะมีเพียงเส้นเสียงเดียว เสียงเช่นนี้เรียกว่า สระเดี่ยว แต่ถ้าลิ้นส่วนอื่นทำหน้าที่ร่วมด้วยเสียงสระนั้นเรียกว่า สระผสม

ฉะนั้นการอธิบายเรื่องการออกเสียงสระจึงเป็นเรื่องที่ยากกว่าการออกเสียงพยัญชนะเพราะไม่มีจุดสัมผัสในปากที่เราสามารถใช้เป็นหลักในการอธิบายได้ การที่สระมีเสียงต่างๆ กันก็เพราะช่องคอ ช่องปากมีรูปร่างต่างกันในขณะที่อากาศกำลังผ่านออกจากปากไป สิ่งที่มีส่วนทำให้ช่องปากช่องคอมีรูปร่างต่างๆ กันออกไปก็คือ ลิ้น และริมฝีปาก เพื่อความสะดวกในการอธิบายเสียงเราจะแบ่งลิ้นออกเป็น 3 ส่วน คือ ส่วนหน้า (Front) ส่วนกลาง (Center) และส่วนหลังอยู่ใต้เพดานอ่อน (Velum) และส่วนกลางอยู่ระหว่างกลางของสองส่วนนี้ รูปต่อไปนี้แสดงตำแหน่งของลิ้นเมื่ออยู่ต่ำ ส่วนหน้าสูง และส่วนหลังสูง



รูปที่ 2.8 ตำแหน่งของลิ้นเมื่ออยู่ต่ำ ส่วนหน้าสูง และส่วนหลังสูง

สำหรับภาษาไทยมีหน่วยเสียงสระทั้งหมด 24 หน่วยเสียง แยกออกเป็นสระเดี่ยว 18 หน่วยเสียง และสระผสม 6 หน่วยเสียง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
60561

เสียงสระเดี่ยว 18 หน่วยเสียง พิจารณาการเกิดเสียงได้เป็น 2 กรณีใหญ่ๆ คือ

1. การเกิดจากส่วนต่างๆ ของลิ้น หมายถึง ลมพ่นผ่านส่วนหน้า ส่วนกลาง หรือส่วนหลังของลิ้น

2. การเกิดจากลมผ่านลิ้นในขณะที่ลิ้นอยู่ในระดับสูง กลาง หรือ ต่ำ

ตารางที่ 2.3 เสียงสระเดี่ยวในภาษาไทย

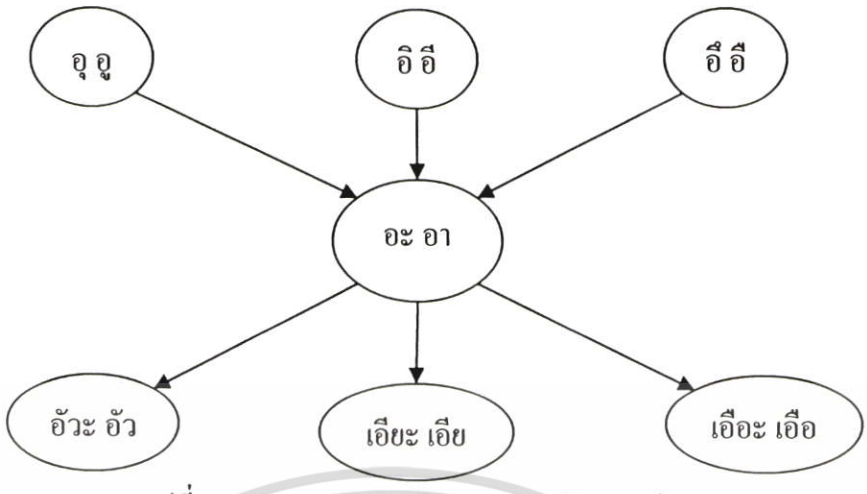
สระ	ระดับลิ้น	หน่วยเสียง
หน้า	สูง	อิ อี
	กลาง	เอะ เอ
	ต่ำ	แอะ แอ
กลาง	สูง	อึ อือ
	กลาง	เออะ เออ
	ต่ำ	อะ อา
หลัง	สูง	อุ อู
	กลาง	โอะ โอ
	ต่ำ	เอาะ ออ

นอกจากนี้ หน่วยเสียงสระเดี่ยว 18 หน่วยเสียง สามารถแบ่งตามความสั้น-ยาว ของการออกเสียงได้เป็น

- สระเดี่ยวเสียงสั้น 9 หน่วยเสียง ได้แก่ อะ อิ อึ อุ เอะ แอะ โอะ เอาะ เออะ
- สระเดี่ยวเสียงยาว 9 หน่วยเสียง ได้แก่ อา อี อือ อู เอ แอ โอ ออ เออ

เสียงผสม 6 หน่วยเสียง เกิดจากลมผ่านกระทบลิ้น 2 ส่วน คือ ส่วนบน และส่วนล่าง ในขณะออกเสียงลิ้น จะอยู่ในระดับสูง แล้วลดลงต่ำโดยเสียงหลังเป็นเสียงสระ (อะ,อา) เสมอ แสดงในแผนผัง ดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.9 แสดงการเกิดเสียงสระผสมในภาษาไทย

2.3.1.2 หน้าที่ของหน่วยเสียงสระในภาษาไทย

หน่วยเสียงสระในภาษาไทยทั้ง 24 หน่วยเสียงนี้ ทำหน้าที่เป็นแกนกลางของพยางค์ หรือคำกล่าวคือ คำทุกคำในภาษาไทยจะต้องมีเสียงสระอยู่ด้วย และเสียงสระในภาษาไทยจะสามารถเกิดกับเสียงพยัญชนะต้นได้ทุกเสียง และสามารถเกิดกับหน่วยเสียงวรรณยุกต์ได้ทุกหน่วย แต่ไม่สามารถเกิดกับหน่วยเสียงพยัญชนะสะกดได้ทุกหน่วย หน่วยเสียงสระที่ทำให้เกิดคำหรือพยางค์ใช้ได้มากที่สุดในภาษามักเป็นหน่วยเสียงยาว

2.3.2 หน่วยเสียงพยัญชนะ

เสียงพยัญชนะในภาษาไทยมีทั้งหมด 21 หน่วยเสียง (44 รูป) หน่วยเสียงพยัญชนะออกเสียงได้ไม่สะดวกเท่าหน่วยเสียง เพราะเวลาออกเสียงลมหายในที่พุ่งออกมาจากหลอดลมจะถูกขัดขวางตามส่วนต่างๆ ของปาก เสียงพยัญชนะจึงออกเสียงให้ยาวนาน อย่างเสียงสระไม่ได้และเสียงพยัญชนะก็ไม่ใช่เสียงก้องเสมอไป

2.3.2.1 ลักษณะของเสียงพยัญชนะ

หน่วยเสียงพยัญชนะ 21 หน่วยเสียงนี้จำแนกเป็นเสียงก้อง เสียงไม่ก้อง เสียงหนัก เสียงเบา และลักษณะการออกเสียง ดังนี้

เสียงก้อง (โฆษะ) มี 9 หน่วยเสียง คือ /ง/ /ย/ /บ/ /ด/ /ม/ /น/ /ร/ /ล/ /ว/

เสียงไม่ก้อง (อโฆษะ) มี 12 หน่วยเสียง คือ /ก/ /ค/ /จ/ /ช/ /ซ/ /ท/ /ต/ /ป/ /พ/ /ฟ/ /อ/ /ฮ/

เสียงหนัก (ธนิต) มี 4 หน่วยเสียง คือ /ค/ /ช/ /ท/ /พ/

เสียงเบา (ถวิล) มี 4 หน่วยเสียง คือ /ก/ /จ/ /ต/ /ป/

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 2.4 เสียงพยัญชนะในภาษาไทย

ฐานที่เกิดเสียง / ลักษณะเสียง		ริม ฝีปาก	ปุ่มเหงือก	เพดาน แข็ง	เพดาน อ่อน	เส้น เสียง
พยัญชนะ กัก	ไม่พ่นลม	/ป/	/ต/	/จ/	/ก/*	/อ/
	พ่นลม	/พ/	/ท/	/ช/	/ค/	
	ก้อง	/บ/*	/ด/*			
พยัญชนะ ไม่กัก	นาสิก	/ม/*	/น/*		/ง/*	
	เสียดแทรก	/ฟ/	/ซ/			/ฮ/
	กระทบ		/ร/			
	ข้างลิ้น		/ล/			
	กึ่งสระ	/ว/*		/ย/*		

2.3.2.2 หน้าที่ของหน่วยเสียงพยัญชนะในภาษาไทย

เสียงพยัญชนะในภาษาไทย 21 หน่วยเสียงนี้สามารถทำหน้าที่ได้ ดังนี้

1. พยัญชนะต้นของพยางค์ คือ สามารถนำหน้าเสียงสระในพยางค์หนึ่งๆ ได้ในตำแหน่งนี้เสียงพยัญชนะสามารถเกิดได้หน่วยเดียว หรือสองหน่วย ดังนี้

- เกิดได้หน่วยเดียว คือ ทำหน้าที่เป็นพยัญชนะต้นเดี่ยว หน่วยเสียงทั้ง 21 หน่วยเสียงนี้สามารถทำหน้าที่เป็นพยัญชนะต้นเดี่ยวได้
- เกิดได้สองหน่วย คือ ทำหน้าที่เป็นพยัญชนะต้นควบ โดยหน่วยเสียงแรกเป็น /ก/ /ค/ /ต/ /ป/ และ/พ/ กับหน่วยเสียงที่สองเป็น /ร/ /ล/ หรือ /ว/

2. เป็นพยัญชนะสะกดของพยางค์ ในตำแหน่งนี้เสียงพยัญชนะในภาษาไทยสามารถเกิดได้ 8 หน่วยเสียง คือ /บ/ (แม่กบ) /ด/ (แม่กด) /ก/ แม่กก /ม/ (แม่กม) /ง/ (แม่กง) /น/ (แม่กน) และ/ว/ แม่เกอว

2.3.3 หน่วยเสียงวรรณยุกต์

หน่วยเสียงวรรณยุกต์ คือ ระดับเสียงสูง-ต่ำ ของคำในภาษาไทย เช่นเดียวกับภาษาจีนและภาษาอื่นๆ ที่เป็นภาษาคำโดด ซึ่งมีการกำหนดเสียงสูง-ต่ำ ไว้ตายตัวในแต่ละคำ ถ้าออกเสียงสูง-ต่ำ ผิดไปความหมายย่อมผิดตามไปด้วย

ในภาษาไทยหน่วยเสียงวรรณยุกต์เป็นหน่วยเสียงสำคัญ มีความหมายต่างกัน ดังนั้นอาจกล่าวได้ว่าหน้าที่ของหน่วยเสียงวรรณยุกต์ ก็คือ การทำให้เกิดคำขึ้นใช้ในภาษามากขึ้น และเป็นวิธีการสร้างคำขึ้นใช้เพิ่มภาษาเป็นวิธีแรก ทั้งนี้เพราะถ้าเราเปลี่ยนเสียงวรรณยุกต์ ก็จะทำให้คำเกิดความหมายเพิ่มขึ้นใหม่นั่นเอง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เสียงสูง-ต่ำ ในภาษาพูด เกิดจากการสั่นสะเทือนของเส้นเสียงในอัตราต่างๆ กัน โดยเสียงที่เปล่งออกมาในขณะที่เส้นเสียงสั่นนั้น จะต้องเป็นเสียงก้อง ดังนั้นหน่วยเสียงวรรณยุกต์ ในภาษาไทยจึงจัดเป็นหน่วยเสียงซ้อน กับอักษรที่ใช้จึงเป็นรูปแบบเครื่องหมายเขียนซ้อนข้างบนหน่วยเสียงสระ (ซึ่งเป็นเสียงก้อง) ซึ่งรูปวรรณยุกต์อยู่ 4 แบบ แทนเสียงวรรณยุกต์ทั้งหมด 5 หน่วยเสียง โดยเสียงสามัญไม่มีรูปวรรณยุกต์

2.3.3.1 ลักษณะของเสียงวรรณยุกต์

เสียงวรรณยุกต์ในภาษาไทยมี 2 ประเภทคือ เสียงวรรณยุกต์ ระดับและเสียงวรรณยุกต์เปลี่ยนระดับ [2]

วรรณยุกต์ระดับ (Level tone) เป็นเสียงที่มีความถี่ค่อนข้างคงที่ตลอดต้นพยางค์จะมีลักษณะเสียงสูงกว่าปลายเสียงพยางค์เล็กน้อย เสียงวรรณยุกต์ระดับในภาษาไทย มี 3 เสียงได้แก่

1. เสียงวรรณยุกต์ ต่ำ-ระดับ (Low tone) คือ เสียงวรรณยุกต์เอก เสียงวรรณยุกต์นี้เริ่มออกเสียงในระดับประมาณ 118 Hz. แล้วเสียงจะค่อย ๆ ลดระดับลงมาเล็กน้อยเมื่อใกล้จบพยางค์ คือประมาณ 110 Hz.

2. เสียงวรรณยุกต์ กลาง-ต่ำ (Mid tone) คือ เสียงวรรณยุกต์สามัญ เสียงวรรณยุกต์นี้เริ่มออกเสียงในระดับประมาณ 120 Hz. แล้วเสียงจะค่อย ๆ ลดระดับต่ำลงเล็กน้อยเมื่อใกล้จบพยางค์ประมาณ 112 Hz.

3. เสียงวรรณยุกต์ สูง-ระดับ (High tone) คือ เสียงวรรณยุกต์ตรี เสียงวรรณยุกต์นี้เริ่มออกเสียงในระดับประมาณ 125 Hz. แล้วเสียงจะค่อย ๆ สูงขึ้นประมาณ 135-140 Hz. จนกระทั่งใกล้จบพยางค์เสียงจึงต่ำลงเล็กน้อย

เสียงวรรณยุกต์เปลี่ยนระดับ (Contour tone) เป็นเสียงที่มีความถี่เปลี่ยนแปลงมากระหว่างต้นพยางค์และท้ายพยางค์ ในภาษาไทยมี 2 เสียง ได้แก่

1. เสียงวรรณยุกต์ ต่ำ-ขึ้น (Rising tone) หรือเสียงต่ำเลื่อนสูง คือ เสียงวรรณยุกต์จัตวา เสียงวรรณยุกต์นี้ เริ่มต้นออกเสียงในระดับต่ำประมาณ 110 Hz. เสียงจะลดระดับลงเล็กน้อยแล้วเปลี่ยนเป็นสูงขึ้นอย่างรวดเร็ว และสูงสุดเมื่อจบพยางค์ประมาณ 140 Hz.

2. เสียงวรรณยุกต์ สูง-ตก (Falling tone) คือ เสียงวรรณยุกต์โท เสียงวรรณยุกต์นี้เริ่มต้นออกเสียงในระดับสูงประมาณ 140 Hz. เสียงจะสูงขึ้นเล็กน้อยแล้วต่ำลงอย่างรวดเร็วจนถึงระดับต่ำสุด เมื่อจบพยางค์ต่ำกว่า 100 Hz

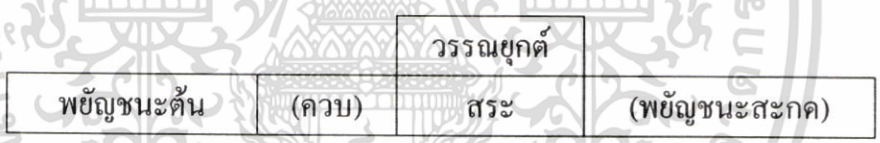
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.4 ลักษณะพยางค์และคำในภาษาไทย

กาณจนา นาคสกุล [12] ได้ให้ความหมายของพยางค์ในระบบเสียงภาษาไทยว่า "พยางค์" หมายถึง จำนวนเสียงที่ดังเด่นซึ่งปรากฏในกลุ่มเสียงที่เรียงกันเป็นคำพูดส่วนเสียงอื่นๆ ที่อยู่ข้างเคียงก็จะประกอบเข้าเป็นส่วนของพยางค์ เสียงที่ดังเด่นในกลุ่มเสียงก็คือ เสียงสระ ซึ่งมีลักษณะประจำตัวก็คือ เป็นเสียงก้อง ซึ่งดังเด่นกว่าเสียงอื่นๆ ดังนั้นเสียงสระจึงมักเป็นเสียงที่ทำให้เสียงพยางค์ ถ้ามีเสียงสระเด่นอยู่กี่เสียง พยางค์ก็จะมีจำนวนเท่านั้นด้วย

พยางค์ที่เปล่งออกมาครั้งหนึ่งๆ อาจมีความหมายหรือไม่ก็ได้ แต่เมื่อใดพยางค์ที่ประกอบขึ้นจากเสียงสระพยัญชนะ และวรรณยุกต์ เป็นอย่างน้อยที่สุด และกลุ่มเสียงเหล่านี้มีความหมายและสามารถปรากฏได้โดยลำพัง พยางค์นั้นๆ ก็กลายเป็นคำในภาษาไทย

คำในภาษาไทยส่วนใหญ่จะเป็นคำพยางค์เดียว ซึ่งเป็นคำพื้นฐานของภาษา ภาษาไทยจึงจัดอยู่ในตระกูลภาษา คำโดด หรือ คำพยางค์เดียว (Monosyllabic language) หน่วยเสียงที่ประกอบกันเข้าเป็นพยางค์ จะต้องมีอย่างน้อย 3 หน่วยเสียงคือ หน่วยเสียงพยัญชนะต้น 1 หน่วย หน่วยเสียงสระ 1 หน่วย และหน่วยเสียงวรรณยุกต์ 1 หน่วย และมีหน่วยเสียงอย่างมากไม่เกิน 5 หน่วย คือ เพิ่มหน่วยเสียงที่เป็นพยัญชนะต้นที่เป็นเสียงควบกล้ำอีก 1 หน่วย และหน่วยเสียงพยัญชนะสะกด อีก 1 หน่วย โดยมีองค์ประกอบของหน่วยเสียงต่างๆ ในพยางค์แสดงได้ ดังรูปที่ 2.10



รูปที่ 2.10 องค์ประกอบของหน่วยเสียงต่างๆ ในพยางค์

การผันเสียงวรรณยุกต์ในพยางค์ไทยคำเดี่ยว

ปัญหาใหญ่ในการออกเสียงพยางค์ต่างๆ ในภาษาไทยอยู่ตรงที่ส่วนของวรรณยุกต์ ซึ่งมีรูปแบบและเสียงไม่จำเป็นต้องตรงกันเสมอไป ในทางปฏิบัติจะยึดหลักไตรยางค์ และเรื่องคำเป็นคำตายตามหลักไวยากรณ์ไทยดั้งเดิม [2]

คำเป็น คือพยางค์ที่ประสมสระเสียงยาวในแม่ก.กา เช่น กา กี กือ กู และพางค์ที่ประสมสระเสียงยาวหรือสั้นในแม่ กง กน กม เกย และ เกอว

คำตาย คือพยางค์ที่ประสมสระเสียงสั้นในแม่ ก.กา เช่น กะ กิ กึ กุ (เว้นสระ อำ ไอ ใอ เอา) และพยางค์ที่ประสมสระเสียงยาวหรือสั้นแม่ กก กด กบ

ไตรยางค์ คือการแบ่งพยัญชนะออกเป็น 3 พวก เพื่อให้สอดคล้องกับวิธีการผันวรรณยุกต์ โดยกำหนดให้อักษรกลุ่มหนึ่งมีจำนวน 9 ตัว ผันวรรณยุกต์ได้ครบ 5 เสียง เรียกว่าอักษรกลาง ได้แก่ ก จ ฎ ฏ ด ต บ ป และ อ เช่น กา ก่า ก้า ก๊า ก๋า

เอกสารนี้มีไว้เพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

อักษรอีกกลุ่มหนึ่งจำนวน 11 ตัว ได้แก่ ข ฃ ฉ ฐ ถ ผ ฝ ส ศ ษ ห มีพื้นเสียงเป็นเสียงสูง (เสียงจัตวา) จึงเรียกอักษรสูง อักษรประเภทนี้ผันวรรณยุกต์ได้เพียง 3 เสียงคือ จัตวา เอก โท (ขา ข่า ข้า)

อักษรอีก 24 ตัวที่เหลือ เรียกว่าอักษรต่ำ จะแบ่งได้เป็น 2 กลุ่มย่อย คือ อักษรต่ำที่มีคู่กับอักษรสูงมีอยู่ 14 ตัวคือ ค ฅ ฆ ช ฌ ซ ฑ ฒ ธ พ ภ ฟ และ ฮ และอักษรต่ำเดี่ยวมีอยู่ทั้งหมด 10 ตัวคือ ง ญ ณ น ม ย ร ล ว ฬ

การผันเสียงของแต่ละกลุ่มอักษรเป็นดังตารางที่ 2.5 2.6 และ 2.7

ตารางที่ 2.5 การผันเสียงวรรณยุกต์ของอักษรกลาง

	เสียง	สามัญ	เอก	โท	ตรี	จัตวา
คำเป็น		กา	ก่า	ก้า	ก๊า	ก๋า
คำตาย	รัสสะ		กะ	ก๊ะ	ก๊ะ	
	ทีฆะ		กาก	ก้าก	ก๊าก	

ตารางที่ 2.6 การผันเสียงวรรณยุกต์ของอักษรต่ำคู่กับอักษรสูง

	เสียง	สามัญ	เอก	โท	ตรี	จัตวา
คำเป็น		คา	ข่า	ข้า	ค้า	ขา
คำตาย	รัสสะ	-	ขะ	ข๊ะ, ค่ะ	คะ	-
	ทีฆะ	-	ขาก	ข้าก, คาก	ค้าก	-

ตารางที่ 2.7 การผันเสียงวรรณยุกต์ของอักษรต่ำเดี่ยวกับอักษรนำ

		เสียง	สามัญ	เอก	โท	ตรี	จัตวา
คำเป็น		กลางนำ	ยา	อย่า	อย้า, ย่า	ย้า	อย๋า
		สูงนำ	ยา	หย่า	หย้า, ย่า	ย้า	หยา
คำตาย	รัสสะ	กลางนำ	-	อยะ	อย๊ะ	อย๊ะ, ยะ	-
		สูงนำ	-	หยะ	หย๊ะ, ย่ะ	ยะ	-
	ทีฆะ	กลางนำ	-	อยาก	อย้าก, ยาก	อย๊าก	-
		สูงนำ	-	หยอก	หย้อก, ยอก	ย้อก	-

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.5 ทฤษฎีทางด้านการวิเคราะห์เสียง

โดยธรรมชาติของสัญญาณเสียงพูดจะไม่เสถียร และเปลี่ยนแปลงตามเวลา (Non-stationary) ดังนั้น เมื่อต้องการนำสัญญาณเสียงพูดมาประมวลผลสัญญาณดิจิตอล (Digital Singal Processing) จึงจำเป็นต้องแบ่งสัญญาณเสียงพูดออกเป็นช่วงเวลาสั้นๆ (Shot Time) เพื่อให้สัญญาณเสียงมีความเสถียรและไม่เปลี่ยนแปลงตามเวลา (Stationary) จากนั้นจึงจะสามารถนำสัญญาณเสียงไปประมวลผลต่อไปได้ กรอบเสียงพูด (Speech Frame) ความยาวประมาณ 20-30 มิลลิวินาที ทำให้สัญญาณเสียงพูดในแต่ละกรอบเสียงพูดเป็นสัญญาณที่มีความเสถียรและไม่เปลี่ยนแปลงตามเวลา การเหลื่อมกรอบเสียงพูด (Frame overlap) จะทำให้รอยต่อของลักษณะสำคัญของเสียงพูด จากกรอบเสียงพูดหนึ่งไปยังอีกกรอบเสียงพูดหนึ่งเรียบ (Smooth) ขึ้น [15] กรรมวิธีการวิเคราะห์สัญญาณเสียงพูดเบื้องต้น ประกอบด้วย

2.5.1 การปรับบรรทัดฐานแอมพลิจูด (Amplitude Normalization)

สัญญาณเสียงพูดเป็นสัญญาณต่อเนื่อง (Continuous Signal) จะต้องทำการสุ่มตัวอย่าง (Sampling) สัญญาณเสียงพูด โดยอัตราการสุ่มตัวอย่าง (Sampling Rate) เท่ากับ อย่างน้อย 2 เท่าของความถี่ของสัญญาณเสียงพูด [14] ซึ่งจะมีตัวอย่างของสัญญาณเสียงพูด (Speech Sample) อย่างน้อย 8,000-10,000 ตัวอย่างต่อวินาที ตัวอย่างสัญญาณเสียงจะถูกนำมาปรับระดับสัญญาณ (Quantization) ให้เป็นสัญญาณดิจิตอล (Digital Signal) กรรมวิธีการปรับบรรทัดฐานแอมพลิจูดของสัญญาณเสียงพูดเป็นการเพิ่มหรือลดขนาดของสัญญาณเสียงพูด เพื่อให้ขนาดของสัญญาณเสียงพูดมีความเหมาะสม เนื่องจากสัญญาณเสียงพูดของแต่ละบุคคลมีขนาดไม่เท่ากัน จึงจำเป็นต้องปรับให้ขนาดของสัญญาณเสียงพูดอยู่ในบรรทัดฐานเดียวกัน เพื่อง่ายต่อการวัดค่าคุณลักษณะและเปรียบเทียบสัญญาณเสียงกับค่ากำหนด(Threshold) การปรับบรรทัดฐานแอมพลิจูดแสดงในสมาการที่ (2.1)

$$\tilde{X} = \frac{x[i]}{2^{N+1}} \qquad (2.1)$$

โดยที่ $\tilde{x}$ คือสัญญาณเสียงพูดที่ปรับบรรทัดฐานแอมพลิจูดแล้ว

$x[i]$ คือสัญญาณเสียงพูดดิจิตอล

N คือจำนวนบิตที่ใช้แทนค่าสัญญาณเสียงพูด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.5.2 วิธีการเน้นล่วงหน้า

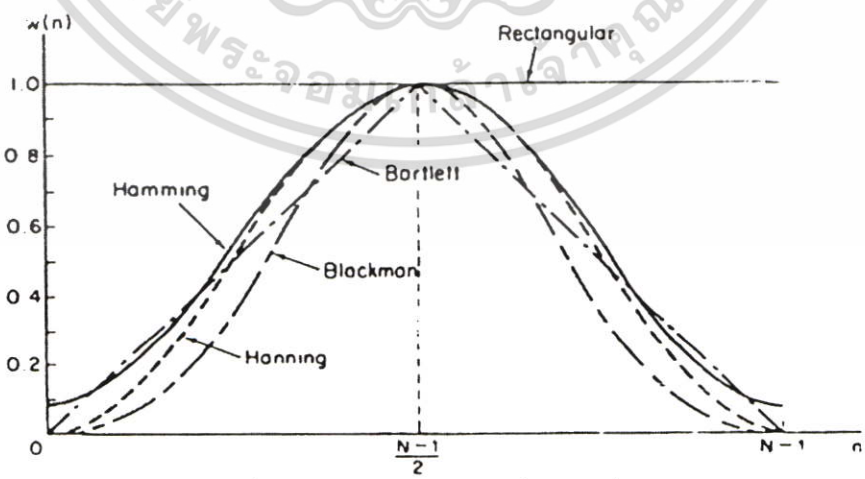
การเน้นล่วงหน้าจะทำให้ความลาดเอียงในเชิงความถี่แบบราบลง [16] และทำให้อัตราส่วนของสัญญาณเสียงพูดต่อสัญญาณรบกวน (Signal to Noise Ratio) มีค่าสูงขึ้น โดยนำสัญญาณเสียงพูด มาผ่านตัวกรองดิจิตอลลำดับหนึ่ง (First Order Digital Filter) ดังแสดงในสมการที่ 2.2

$$H(z) = 1 - a \cdot z^{-1} \quad \text{โดยที่} \quad 0.9 < a < 1.0 \tag{2.2}$$

2.5.3 การวางกรอบหน้าต่าง (Windowing)

การวางรูปแบบกรอบหน้าต่าง เป็นการเตรียมข้อมูลในแต่ละกรอบเสียงพูด เพื่อนำไปวิเคราะห์ในขั้นตอนการวัดค่าคุณลักษณะของสัญญาณเสียง โดยสัญญาณเสียงพูดจะถูกคูณด้วยฟังก์ชันกรอบหน้าต่าง (Window Function) ซึ่งฟังก์ชันกรอบนั้นมีหลายประเภท เช่น Rectangular window, Hamming window, Blackman window, Kaiser เป็นต้น [17] รูปที่ 2.11 แสดงรูปแบบหน้าต่างประเภทต่างๆ ในงานวิจัยนี้เลือกฟังก์ชันกรอบแบบ Hamming Window [16] ดังแสดงในสมการที่ (2.3) เนื่องจาก Hamming Window จะทำให้ขนาดของสัญญาณลดลงอย่างช้าๆ ที่บริเวณปลายของกรอบหน้าต่าง จึงเป็นการช่วยป้องกันการเปลี่ยนแปลงอย่างกะทันหันที่บริเวณปลายของกรอบเสียงพูด สมการการวางกรอบหน้าต่างกับสัญญาณเสียงพูด แสดงในสมการที่ (2.3)

$$w(n) = 0.54 - 0.5\cos\left(\frac{2\pi n}{N-1}\right) \quad \text{เมื่อ} \; 0 < n < N-1 \tag{2.3}$$



รูปที่ 2.11 รูปแบบของฟังก์ชันกรอบหน้าต่างแบบต่างๆ

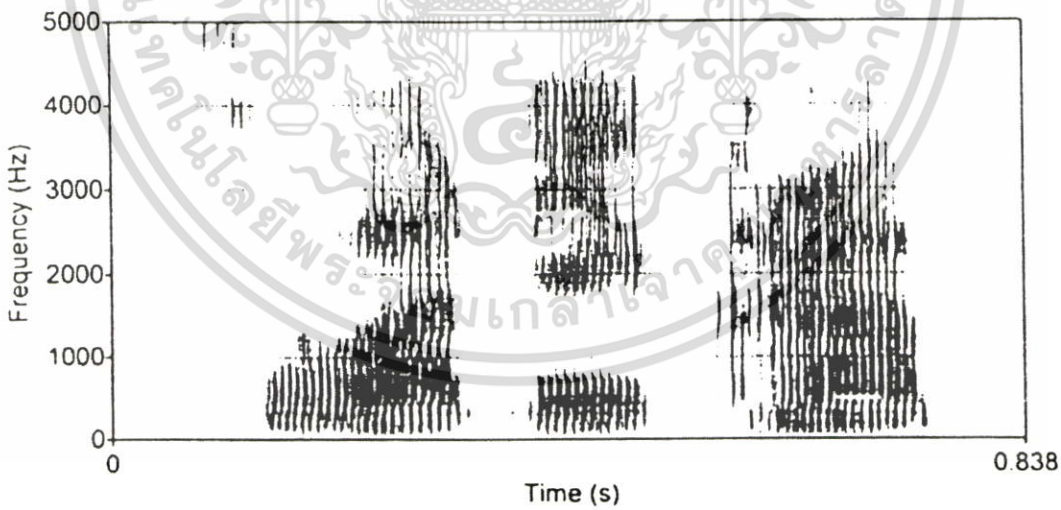
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.5.4 สเปกโตรแกรม (Spectrogram)

สเปกโตรแกรมเป็นรูปแบบหนึ่งของการแสดงสัญญาณเสียงเพื่อนำไปใช้ในการวิเคราะห์ที่นิยมใช้และมีความสำคัญสำหรับงานด้านนี้เป็นอย่างมาก การแสดงผลชนิดนี้ได้จากการคำนวณการแปลงฟูเรียร์ที่ขึ้นกับเวลา ผลที่แสดงคือกราฟสองมติ ที่แสดงความสัมพันธ์ระหว่างสเปกตรัมที่ขึ้นกับเวลา สำหรับขนาดสเปกตรัมแสดงด้วยสีที่มีความเข้มแตกต่างกันซึ่งสามารถนำไปใช้ในการวิเคราะห์หาลักษณะที่สำคัญของเสียงและลักษณะเด่นสำหรับการรับรู้เสียง เช่น ความก้องของเสียง โครงสร้างฟอร์แมนท์ การเน้นเสียงและพิทช์ของเสียง เป็นต้น ในการวิเคราะห์โดยโดยใช้ สเปกโตรแกรมโดยทั่วไปแบ่งเป็น 2 แบบคือ

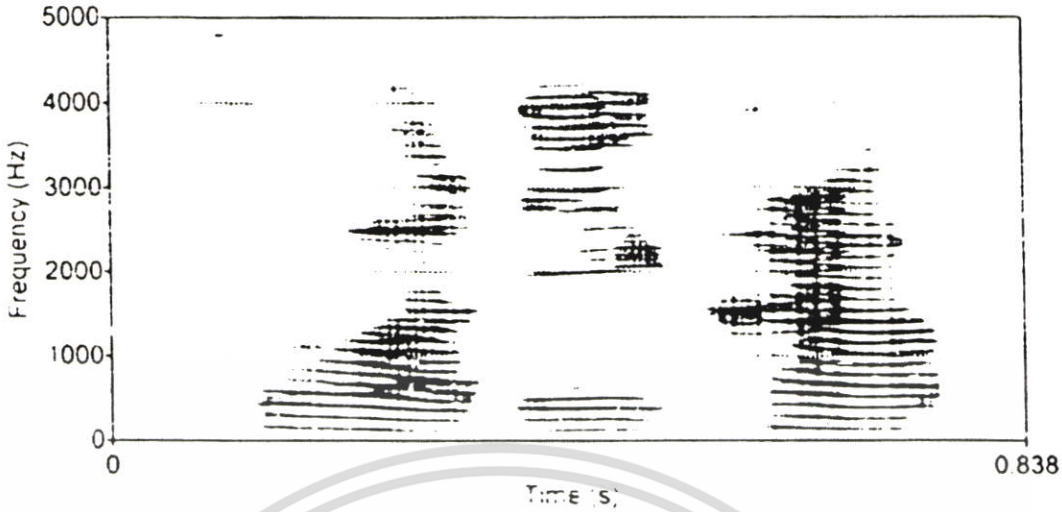
สเปกโตรแกรมแถบกว้าง (Wideband Spectrogram) เป็นการวิเคราะห์ที่แสดงรายละเอียดชั่วขณะในเชิงเวลาได้ดี แต่มีความละเอียดทางแกนความถี่ต่ำ [14] ดังแสดงในรูปที่ 2.12 กราฟชนิดนี้แสดงถึงการเปลี่ยนแปลงของฟอร์แมนท์และแสดงคาบพิทช์โดยดูจากลายเส้นในนาวตั้งได้ [18]

สเปกโตรแกรมแถบแคบ (Narrowband Spectrogram) เป็นการวิเคราะห์ที่แสดงรายละเอียดทางด้านฮาร์โมนิคได้ดี แต่แสดงรายละเอียดชั่วขณะในเชิงเวลาไม่ชัดเจน [14] ดังรูปที่ 2.13 กราฟชนิดนี้สามารถนำไปหาความถี่มูลฐานได้โดยการวิเคราะห์จากอาร์โมนิคของความถี่มูลฐานและใช้ในการแยกช่วงที่เป็นเสียงก้องออกจากเสียงไม่ก้องได้ [18]



รูปที่ 2.12 สเปกโตรแกรมแถบกว้าง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.13 สเปกโตรแกรมแถบแคบ

2.5.5 ออโตคอร์รีเลชั่น (Autocorrelation)

กรณีพิเศษที่มีความสำคัญมากของฟังก์ชั่นคอร์รีเลชั่น (Correlation) ก็คือ คอร์รีเลชั่นของฟังก์ชั่นด้วยตัวมันเอง คอร์รีเลชั่นชนิดนี้เรียกว่า ฟังก์ชั่นออโตคอร์รีเลชั่น (autocorrelation) ถ้า x(n) เป็นสัญญาณพลังงาน(energy signal), ออโตคอร์รีเลชั่นของมัน คือ [19]

$$R_{xx}[m] = \sum_{n=-\infty}^{\infty} x[n]x[n+m] \qquad (2.4)$$

เมื่อเลื่อนเข้าใกล้ 0 จะได้

$$R_{xx}[0] = \sum_{0}^{\infty} x^2[n] \qquad (2.5)$$

ซึ่งเป็นผลรวมสัญญาณพลังงาน ของสัญญาณ

ถ้า x[*n*] เป็นสัญญาณสัญญาณกำลัง ออโตคอร์รีเลชั่นเมื่อเลื่อนเข้าใกล้ 0 คือ

$$R_{xx}[0] = \lim_{N\to\infty} \sum_{0}^{\infty} x^2[n] \qquad (2.5)$$

ซึ่งคือค่าเฉลี่ยสัญญาณกำลังของสัญญาณ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.5.6 การประมาณการเชิงเส้น (LPC: Linear Predictive Coding)

การประมาณการเชิงเส้นเป็นเทคนิคที่สำคัญทางด้านการวิเคราะห์และการสังเคราะห์เสียง เนื่องจากมีความแม่นยำสูงในการประมาณค่าพารามิเตอร์ของเสียงพูดเมื่อเทียบกับความเร็วในการประมวลผล หลักการพื้นฐานของการประมาณค่าแบบเชิงเส้นอาศัยแนวคิดว่า ตัวอย่างสัญญาณเสียงพูดสามารถประมาณค่าได้จากผลรวมของตัวอย่างสัญญาณเสียงพูดจากอดีต [20] การวิเคราะห์เพื่อหาพารามิเตอร์เพื่อใช้การประมาณค่าโดยทั่วไปเรียกว่า การเข้ารหัสการประมาณค่าแบบเชิงเส้น หรือแอลพีซี (Linear Predictive Coding: LPC) ในทางด้านการประมวลผลสัญญาณเสียง การเข้ารหัสการประมาณค่าแบบเชิงเส้นนี้ถูกนำไปใช้ในสองทาง คือ

1. การเข้ารหัสสัญญาณเสียง โดยที่การเข้ารหัสการประมาณค่าแบบเชิงเส้นถูกนำไปใช้เป็นวงจรกรองวิเคราะห์การประมาณค่าแบบเชิงเส้น (LP Analysis Filter) เพื่อแยกส่วนซ้ำซ้อน (Redundancy) ของสัญญาณเสียงออก ส่วนที่เหลือเรียกว่าสัญญาณตกค้าง (Residual Signal)

2. การสังเคราะห์สัญญาณเสียงพูด โดยการเข้ารหัสการกประมาณค่าแบบเชิงเส้นถูกนำไปใช้เป็นวงจรกรองการประมาณค่าแบบเชิงเส้นผกผัน (Inverse LP Filter) หรือวงจรกรองสังเคราะห์การประมาณค่าแบบเชิงเส้น (LP Synthesis Filter) โดยที่ฟังก์ชั่นการถ่ายโอนของ วงจรกรองดังกล่าวแสดงกรอบสเปกตรัมของสัญญาณเสียงพูด จงจรกรองสังเคราะห์การทำนายแบบเชิงเส้นถูกใช้แทนช่องทางเดินเสียงของมนุษย์และใช้หาสัญญาณกระตุ้นที่เหมาะสมสำหรับวงจรกรองดังกล่าว

ในการวิเคราะห์การเข้ารหัสการประมาณค่าแบบเชิงเส้นเริ่มต้นจากพิจารณากรอบสัญญาณเสียงที่มี N ตัวอย่าง คือ $x_1, x_2, ..., x_n$ โดยอ้างว่าตัวอย่างสัญญาณปัจจุบันสามารถประมาณค่าได้จากผลรวมของตัวอย่างสัญญาณในอดีต p ตัวอย่าง ดังสมการที่ (2.6)

$$\tilde{x}_n = -a_1 x_{n-1} - a_2 x_{n-2} - a_3 x_{n-3} - ... - a_p x_{n-p} = -\sum_{k=1}^{p} a_k x_{n-1} \qquad (2.6)$$

เมื่อ p คืออันดับของการวิเคราะห์การประมาณค่าแบบเชิงเส้นและ $a_1, a_2, a_3,...,a_p$ คือสัมประสิทธิ์การเข้ารหัสการประมาณค่าเชิงแบบเชิงเส้น กำหนด e_n แทนค่าผิดพลาดระหว่างค่าจริงและค่าที่ทำนายได้จะได้ตามสมการที่ (2.7)

$$e_n = x_n - \tilde{x}_n$$

$$e_n = x_n - \sum_{k=1}^{p} a_k x_{n-1} \qquad (2.7)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สัญญาณ e_n เรียกว่าสัญญาณตกค้าง เนื่องจากสัญญาณ e_n ได้จากผลการลบสัญญาณ s_n ด้วย $\tilde{x}_n$ เนื่องจากค่าสหสัมพันธ์ช่วงสั้น (Short-term Correlation) ระหว่างตัวอย่างของสัญญาณตกค้างมีค่าต่ำ ดังนั้นประมาณได้ว่ากรอบสเปกตรัมกำลังของสัญญาณตกค้างมีลักษณะเรียบ (flat) เมื่อทำการแปลงแซด (z Transform) ของสมการที่ (2.8) ได้ดังสมการที่ (2.9)

$$E(z) = A(z)S(z) \tag{2.8}$$

โดยที่ $S(z)$ และ $E(z)$ เป็นผลการแปลงแซดของสัญญาณเสียงและสัญญาณตกค้างตามลำดับ

$$E(z) = 1 + \sum_{k=1}^{p} a_k z^{-k} \tag{2.9}$$

โดยที่ A(z) หรือวงจรกรองไวเทนนิง (Whitening Filter) มีหน้าที่แยกความสหพันธ์ (Correlation) ช่วงสั้นที่ปรากฏในสัญญาณเสียงพูดซึ่งก็คือการทำให้สเปกตรัมเรียบ เนื่องจาก E(z) ประมาณได้ว่ามีสเปกตรัมเรียบดังนั้นสามารถออกแบบกรอบสเปกตรัมของสัญญาณช่วงสั้นได้จากการวิเคราะห์การประมาณค่าแบบเชิงเส้นรูปแบบออล์โพล (All-pole Model) ดังสมการที่ (2.10)

$$H(z) = \frac{1}{A(z)} \tag{2.10}$$

วงจรกรอง A(z) เรียกได้อีกชื่อหนึ่งว่าวงจรกรองผกผัน (Inverse Filter) เนื่องจากเป็นส่วนผกผันของรูปแบบออล์โพล H(z) ของสัญญาณเสียงพูด และรากของ A(z) ทำให้เกิดโพลใน H(z)

2.5.7 ผลการแปลง z (z-transform)

วิธีทางคณิตศาสตร์อย่างหนึ่งที่เรียกว่าผลการแปลง z (z-transform) วิธีการนี้เป็นการแปลงลำดับซึ่งเป็นตัวเลขที่เขียนเรียงกันให้เป็นฟังก์ชันของตัวแปรเชิงซ้อน(complex variable) [21]

ให้ลำดับที่เราสนใจคือ $\{x[n]\}$ และลำดับนี้มีค่าเป็น 0 เมื่อเวลาเต็มหน่วยมีค่าเป็นลบ ผลการแปลง z ของลำดับนี้คือ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

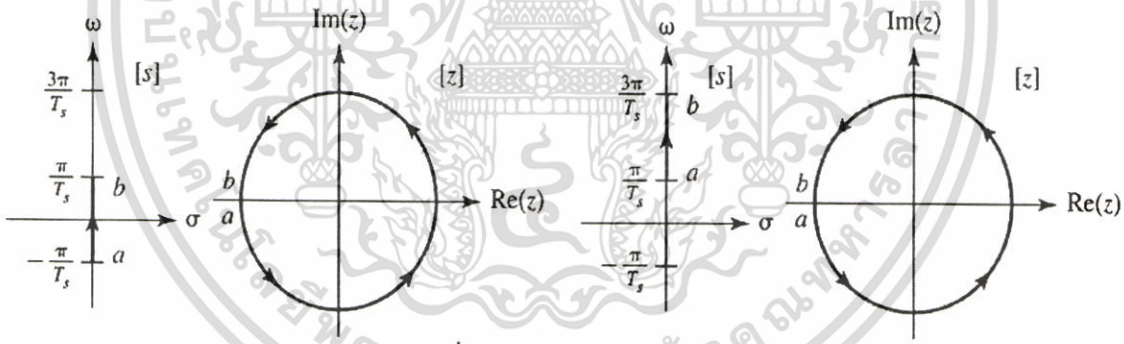
$$Z\{x[n]\} = X(z) = x[0] + \frac{x[1]}{z} + \frac{x[2]}{z^2} + \frac{x[3]}{z^3} + \ldots \qquad (2.11)$$

เมื่อ z คือตัวแปรซึ่งเป็นจำนวนเชิงซ้อนใดๆ Z เป็นสัญลักษณ์ใช้แทนคำว่าการทำผลการแปลง z ของลำดับ $\{x[n]\}$ หลังจากทำการแปลง (transformation) จะได้ฟังก์ชันที่มีตัวแปรเป็น z ได้ดังสมการที่ (2.12)

$$X(z) = \sum_{n=0}^{\infty} x[n] z^{-n} \qquad (2.12)$$

เนื่องจากผลการแปลง z เป็นการสร้างฟังก์ชัน z โดยใช้ลำดับเป็นข้อมูล จากฟังก์ชัน z นั้นสามารถแปลงกลับปาลำดับเดิมได้ ซึ่งเรียกว่า ผลการแปลง z ผกผัน (inverse z-transform)

ผลการแปลง z คือ ฟังก์ชันที่มี z เป็นตัวแปร ตัวแปรนี้เป็นจำนวนเชิงซ้อนใดๆ เพื่อสะดวกในการแสดงทุกๆค่าของ z จึงมักใช้ระนาบเชิงซ้อน z (complex z-plane) ดังแสดงในรูป 2.14



รูปที่ 2.14 ระนาบเชิงซ้อน z

z = a+jb

โดยที่ $j = \sqrt{-1}$

a คือส่วนจริงของจำนวนเชิงซ้อน z (real part)

b คือส่วนจินตภาพของจำนวนเชิงซ้อน (imaginary part)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.5.8 โพลและซีโร

ผลการแปลง z ของลำดับเฉพาะส่วนที่มีขอบเขตจะสามารถเขียนอยู่ในรูปของนิพจน์ที่ดูง่ายขึ้นมาก และพบว่านิพจน์นี้จะอยู่ในรูปของอัตราส่วนของพหุนาม (polynomail) 2 จำนวน ดังนี้

$$X(z) = \frac{P(z)}{Q(z)} \tag{2.13}$$

โดยที่ P(z) และ Q(z) เป็นพหุนามของตัวแปรเชิงซ้อน z เรียกว่าพหุนามตัวเศษ (numberator polynomial) และพหุนามตัวส่วน (denominator polynomial) ตามลำดับ ฟังก์ชัน X(z) ที่เขียนในรูปแบบนี้เรียกว่า ฟังก์ชันตรรกยะ(rational function) และพอที่จะกล่าวได้ว่าผลการแปลง z ในส่วนที่มีขอบเขตของลำดับทุกตัวสามารถจะเขียนอยู่ในรูปของฟังก์ชันตรรกยะได้ ซึ่งพหุนามของทั้งตัวเศษและตัวส่วนจะมีลำดับมากน้อยขึ้นอยู่กับตัวลำดับที่ใช้แปลง ถ้าให้พหุนามของตัวเศษมีอันดับ M และพหุนามของตัวส่วนมีอันดับ N ผลการแปลง z ของลำดับ x[n] ใดๆในส่วนที่มีขอบเขตสามารถเขียนอยู่ในรูปฟังก์ชันตรรกยะ ซึ่งเป็นอัตราส่วนของพหุนามของตัวแปร z สองจำนวนดังนี้ [21]

$$X(z) = \frac{b_0 z^{M} + b_1 z^{M-1} + \ldots + b_M}{z^{N} + a_1 z^{N-1} + \ldots + a_N} \tag{2.14}$$

เมื่อเป็นพหุนามจะต้องแยกตัวประกอบได้ เมื่อแยกตัวประกอบทั้งตัวเศษและตัวส่วนแล้วจะได้ผลการแปลง z ในรูปแยกตัวประกอบดังนี้

$$X(z) = \frac{b_0 (z - z_1)(z - z_2)\ldots(z - z_M)}{(z - p_1)(z - p_2)\ldots(z - p_N)} \tag{2.15}$$

ทั้งนี้ $z_1, z_2, \ldots, z_M$ คือ ราก(root) ของพหุนามตัวเศษ และ $p_1, p_2, \ldots, p_N$ คือรากของพหุนามตัวส่วน

จากสมการ (2.15) เมื่อแทนค่า z ที่มีค่าเท่ากับค่าหนึ่งค่าใดของ $z_1, z_2, \ldots, z_M$ ลงในฟังก์ชัน X(z) จะทำให้เศษส่วนของ X(z) มีค่าเท่ากับ 0 ซึ่งจะทำให้ X(z) มีค่าเท่ากับ 0 ด้วย ดังนั้นเราจึงเรียก $z_1, z_2, \ldots, z_M$ ว่าเป็น ซีโร ของ X(z) ในทำนองเดียวกัน เมื่อพิจารณาจากสมการเดียวกันจี้จะพบว่าเมื่อ z มีค่าเท่ากับค่าใดค่าหนึ่งต่อไปนี้ $p_1, p_2, \ldots, p_N$ จะทำให้ตัวส่วนของ X(z) มีค่าเท่ากับ 0 ซึ่งจะทำให้ X(z) ไม่มีขอบเขตหรือมีค่าเข้าหาอนันต์ ดังนั้น $p_1, p_2, \ldots, p_N$ จึงเรียกว่า โพล (Pole) ของ X(z)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.6 ทฤษฎีการวิเคราะห์คุณสมบัติของสัญญาณเสียงพู ด

พลังงานของสัญญาณเสียงพูด ถูกนำมาใช้ในการวิเคราะห์เสียงพูดตั้งแต่ ปี ค.ศ. 1970 และเป็นคุณลักษณะที่นิยมนำมาใช้กันอย่างแพร่หลาย เนื่องจากเป็นวิธีที่คำนวณง่าย และรวดเร็ว พลังงานของสัญญาณเสียงพูดเป็นคุณสมบัติที่แสดงให้เห็นว่ามีสัญญาณเสียง (รวมทั้งสัญญาณรบกวน) เกิดขึ้น ณ เวลานั้นหรือไม่ การคำนวณหาค่าพลังงาน จะทำทีละกรอบเสียงพูด โดย E_n คือ ค่าพลังงานของกรอบเสียงพูดที่ n เมื่อ $x_n[i]$ คือสัญญาณเสียงพูดที่ i ในกรอบเสียงพูด n และในแต่ละกรอบเสียงพูดจะมีสัญญาณเสียงพูดจำนวน K ซึ่งวิธีการคำนวณหาค่าพลังงานของสัญญาณเสียงพูด มีดังนี้ คือ [15]

1) พลังงานสมบูรณ์ (Absolute Energy)

เป็นการหาผลรวมของสัญญาณเสียงพูดสมบูรณ์ ในแต่ละกรอบเสียงพูด ดังสมการที่ 2.16

$$E_n = \sum_{i=1}^{k} \left| x_n^2[i] \right| \tag{2.16}$$

2) พลังงานเฉลี่ย (Root Mean Square Energy)

เป็นการหาพลังงานของสัญญาณเสียงพูดจากรากที่สองของผลรวมกำลังสองเฉลี่ย ดังสมการ 2.17

$$E_n = \frac{1}{K} \sum_{i=1}^{k} \left| x_n^2[i] \right|^2 \tag{2.17}$$

2.6.1 ความถี่มูลฐาน (Fundamental Frequency)

เสียงพูด (Voicing) เกิดจากลมหายใจที่ถูกดัดแปลงไป (Modified Breathing) โดยอาศัยการทำงานร่วมกันของอวัยวะในการออกเสียงต่างๆ (Articulators) เมื่อกระแสอากาศจากแหล่งพลังงานงานต่างๆ ซึ่งส่วนใหญ่หมายถึงปอด เคลื่อนที่มาสู่กล่องเสียงและถูกดัดแปลงให้เป็นเสียงแบบต่างๆ การสั่นของเส้นเสียงแบบธรรมดา (Normal Vibration) ทำให้เกิดเสียงก้อง (Voiced Sound) ส่วนเสียงไม่ก้อง (Voiceless Sound) จะไม่มีการสั่นสะเทือนของเส้นเสียง ต่อมาอากาศก็จะเดินทางเข้าสู่ช่องปาก ซึ่งประกอบด้วยอวัยวะแปรเสียงหรือฐานกรณ์ (Articulators) มากมายเพื่อทำหน้าที่ในการกล่อมเกลาเสียงให้ออกมามีคุณสมบัติแตกต่างกัน สระซึ่งเป็นแกนกลางพยางค์มีลักษณะเป็นเสียงก้อง ซึ่งเสียงก้องจะมีความเป็นรายคาบของระดับเสียง (Pitch Period) และส่วนกลับของระยะเวลารายคาบของสัญญาณเสียงคือความถี่มูลฐาน (Fundamental Frequency) วิธีการตรวจหาระดับเสียงที่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

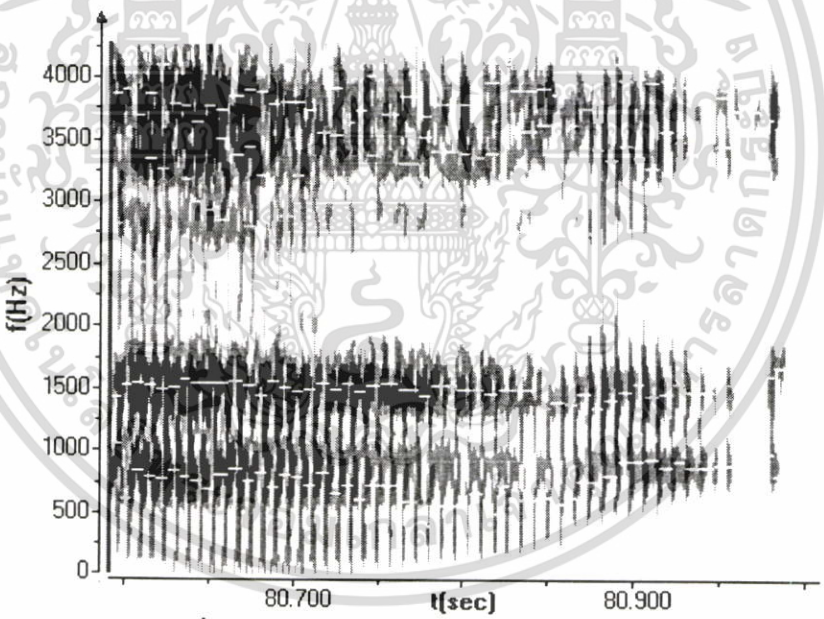
เป็นรายคาบมีดังต่อไปนี้

1. วิธี Cepstrum (CEP)
2. วิธี Simplified Inverse Filtering Technique (SIFT)
3. วิธี Modified Autocorrelation Using Clipping (AUTOC)
4. วิธี Average Magnitude Difference Function (AMDF)[13]

ในงานวิจัยนี้ใช้วิธี Modified Autocorrelation Using Clipping

2.6.2 ความถี่ฟอร์แมนท์ (Formant Frequency)

ความถี่ฟอร์แมนท์ คือความถี่กำทอน (Resonance frequency) ของกลุ่มอวัยวะกำทอนเสียง ความถี่ฟอร์แมนท์ของเสียงจะมีค่าเท่าใดขึ้นอยู่กับการบังคับขนาดและรูปทรงของอวัยวะส่วนนี้ อันเกิดเป็นเสียงความถี่ที่แตกต่างกันนั่นคือ เสียงพูดต่างๆ ความถี่ฟอร์แมนท์สำหรับเสียงพูดคำหนึ่งๆ อาจจะมีหลายค่า ดังรูปที่ 2.15 เป็นสเปกตรัมของเสียง /อา/ ซึ่งมีความถี่ฟอร์แมนท์ที่ 1 2 3 และ 4 อยู่ ณ ตำแหน่งความถี่ ที่ 845 1544 2495 และ3658 เฮิรตซ์ ตามลำดับ



รูปที่ 2.15 สเปกโตรแกรมของเสียงสระอา

2.6.3 กฎของคราเมอร์ (Cramer's rule)

งานวิจัยนี้.ใช้กฎของคราเมอร์ในการการแก้สมการที่เขียนให้อยู่ในรูปแบบของเมตริกซ์ ที่ได้จากการหากำลังสองน้อยที่สุดของข้อมูลความถี่มูลฐานที่ได้จากการทดลอง เพื่อหากฎการผันวรรณยุกต์ภาษาไทย ดังสมการที่ (2.18)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$\underset{(nxn)(nx1)}{[A]}\ \underset{(nx1)}{[X]} = [B] \tag{2.18}$$

โดยที่ n แทนจำนวนสมการ เมตริกซ์ [A] เป็นเมตริกจัตุรัสขนาด nxn ที่ประกอบด้วยสัมประสิทธิ์ต่างๆ ที่รู้ค่า เมตริกซ์ [X] เป็นเมตริกซ์แถวตั้งที่ประกอบด้วยตัวไม่รู้ค่าจำนวน n ค่า และ เมตริกซ์ [B] เป็นเมตริกซ์แถวตั้งเช่นกัน แต่ประกอบด้วยตัวรู้ค่าทั้งหมด n ค่า สามารถค่าตัวไม่รู้ค่าโดยใช้กฎของคราเมอร์เพื่อแก้สมการได้

การใช้กฎของคราเมอร์ เป็นวิธีการหนึ่งที่สามารถใช้แก้สมการขนาดเล็กๆ ได้โดยง่าย หัวใจของกฎคราเมอร์คือการหาค่าดีเทอร์มิแนนต์ (determinant) โดยกฎของคราเมอร์กล่าวไว้ว่า ตัวไม่รู้ค่า x_i ของสมการ

$$x_i = \frac{\det[A]_i}{\det[A]} \tag{2.19}$$

โดย det[A] แทนค่ากำหนดของเมตริกซ์ [A] และ $[A]_i$ แทนค่าตัวกำหนดของเมตริกซ์ [A] หลังจากที่เมตริกซ์ [A] นี้ ได้เปลี่ยนค่าในแนวแถวตั้ง i ด้วยค่าในเมตริกซ์ [B] แล้ว การหาผลลัพธ์ด้วยวิธีคราเมอร์โดยใช้สมการ (2.19) [22]

ตัวอย่างการใช้กฎของคราเมอร์ในการแก้สมการต่อไปนี้

$$\begin{bmatrix} 2 & 1 \\ 1 & -1 \end{bmatrix} \begin{Bmatrix} x_1 \\ x_2 \end{Bmatrix} = \begin{Bmatrix} 4 \\ -1 \end{Bmatrix} \tag{2.20}$$

จากสมการที่ (2.20) ค่าของ [A] และ [B] คือ

$$[A] = \begin{bmatrix} 2 & 1 \\ 1 & -1 \end{bmatrix} \qquad [B] = \begin{Bmatrix} 4 \\ -1 \end{Bmatrix}$$

ดังนั้นจากกฎของคราเมอร์สมการที่ (2.19) ค่าของ x_1 และ x_2 คือ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

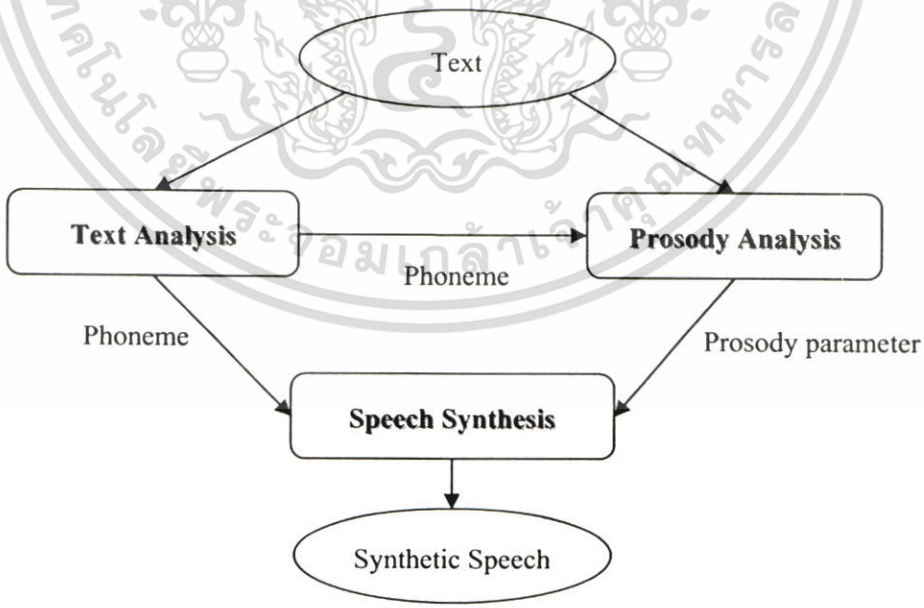
$$x_1 = \frac{\det[A]_1}{\det[A]} = \frac{\begin{vmatrix} 4 & 1 \\ -1 & -1 \end{vmatrix}}{\begin{vmatrix} 2 & 1 \\ 1 & -1 \end{vmatrix}} = \frac{-4+1}{-2-1} = 1$$

$$x_2 = \frac{\det[A]_2}{\det[A]} = \frac{\begin{vmatrix} 2 & 4 \\ 1 & -1 \end{vmatrix}}{\begin{vmatrix} 2 & 1 \\ 1 & -1 \end{vmatrix}} = \frac{-2-4}{-2-1} = 2$$

ซึ่งคือผลลัพธ์ของสมการที่ (2.20) กฎของคราเมอร์นี้มีประโยชน์ในการแก้สมการที่เขียนอยู่ในรูปของเมตริกที่มีขนาดไม่ใหญ่มากนัก

2.7 ทฤษฎีทางด้านการสังเคราะห์เสียง

โครงสร้างของระบบสังเคราะห์เสียงโดยทั่วไป สามารถแบ่งการทำงาน ภายใน ได้เป็น 3 ส่วนดังรูป 2.16



รูปที่ 2.16 ระบบการสังเคราะห์เสียงจากข้อความ [1]

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ส่วนการวิเคราะห์ข้อความ (Text Analysis) ส่วนนี้จะมีหน้าที่วิเคราะห์ข้อความอินพุท เพื่อแปลงเป็นข้อมูล เสียงอ่าน (phoneme) ของคำนั้น และส่งต่อให้ส่วนของการ สังเคราะห์เสียง (Speech synthesis) ต่อไป

ส่วนการวิเคราะห์สัทสัมพันธ์ (Prosody Analysis) ส่วนนี้ทำหน้าที่ในการวิเคราะห์และสังเคราะห์ข้อมูล สัทสัมพันธ์ (prosody) ของประโยคใดๆ จากข้อมูลเสียงอ่าน และข้อความ ข้อมูลสัทสัมพันธ์ ที่วิเคราะห์ ออกมาได้ในระบบทั่วไปได้แก่

1) Segment Duration คือความยาวของเสียงย่อยที่ต้องการสังเคราะห์ ค่านี้จะ มีผลต่อจังหวะของเสียงที่ทำการสังเคราะห์ เช่น ถ้ากำหนด ให้ค่า ความ ยาว ของเสียงย่อยที่ต้องการสังเคราะห์มี ขนาดสั้น เสียงที่ทำ การสังเคราะห์ก็จะเหมือนกับการพูดเร็ว

2) Pitch Contour คือ ค่าความสัมพันธ์ของความถี่มูลฐานกับเวลา ค่านี้จะมีผลต่อเสียงสูงต่ำ (intonation) ของประโยคนั้นๆ

ส่วนการสังเคราะห์เสียง (Speech synthesis) ส่วนนี้ทำหน้าที่ในการสร้างสัญญาณคลื่นเสียง จากข้อมูลเสียง อ่าน (phonetic transcription) และข้อมูลสัทสัมพันธ์ (Prosody transcription) จากข้อ 1 และ 2 และส่งออกสู่ลำโพง เพื่อให้เราได้ยินเสียงพูดประโยคนั้นๆ โดยทั่วไป ส่วนนี้สามารถแบ่งตามเทคนิควิธีการสังเคราะห์เสียง ได้ 3 ประเภท คือ

1) Articulation Synthesis สำหรับวิธีนี้ข้อมูลเสียงที่ต้องการสังเคราะห์จะ โมเดล อยู่ในรูปของค่า พารามิเตอร์ของ โครงสร้าง ทางกายภาพของการเคลื่อนไหวของ อวัยวะ ในช่อง ปากที่ทำให้เกิดเสียงต่างๆวิธีการนี้ค่อนข้างยาก ในแง่การโมเดล เสียงต่างๆ ซึ่งจะต้องศึกษาจากอวัยวะ ในการออกเสียงจริงๆ

2) Concatenation synthesis เทคนิควิธีการนี้เป็นวิธีที่ได้รับความนิยมมากที่สุดในปัจจุบัน โดยเสียง ที่ทำการสังเคราะห์ขึ้น นี้เกิดจากการนำหน่วยเสียงย่อย ที่ทำการเก็บไว้ก่อน แล้วมาต่อกันเป็น เสียงพูดที่ต้องการ โดยทั่วไปหน่วยเสียงย่อยที่ทำเก็บ ไว้จะอยู่ระดับต่ำกว่าคำ เช่น หน่วยของเสียงพยางค์ หน่วยของเสียง ครึ่งพยางค์ (demisyllable) หน่วยของเสียงคู่เสียง (diphone) เป็นต้น ซึ่งในงานวิจัยนี้ได้เลือกใช้เทคนิควิธีการนี้ โดยใช้หน่วยของ เสียงครึ่งพยางค์ เป็นพื้นฐาน

3) Formant Synthesis เทคนิคการสังเคราะห์วิธีแบบนี้ ข้อมูลเสียงอ่านใดๆ จะถูกกำหนดไว้อยู่ในรูป ของความถี่ฟอร์แมนท์ต่างๆ (F1, F2, F3) ของเสียงนั้นๆ เมื่อต้องการ สังเคราะห์ เสียงใดๆ ก็นำข้อมูลเหล่านี้มาทำการสังเคราะห์ให้เป็นสัญญาณ เสียง ซึ่งวิธีการนี้จะมีข้อดี ที่สามารถควบคุมค่าความเปลี่ยนแปลงของ ความถี่ฟอร์แมนท์ (Formant transition) ที่บริเวณรอย ต่อระหว่างเสียงได้ง่าย แต่มีข้อเสีย คือ การจะแทนเสียงใดๆด้วยค่าฟอร์แมนท์ทำได้ยาก จะต้องมีกฎในการสังเคราะห์เสียงใดๆจำนวนมาก และเสียงที่ สังเคราะห์ออก มาได้จะไม่ค่อยเป็นธรรมชาติ ตัวอย่างของระบบแบบนี้ได้แก่ MITALK และ DECTALK [1] รายละเอียดของวิธีการสังเคราะห์แบบฟอร์แมนท์ดังแสดงในหัวข้อถัดไป

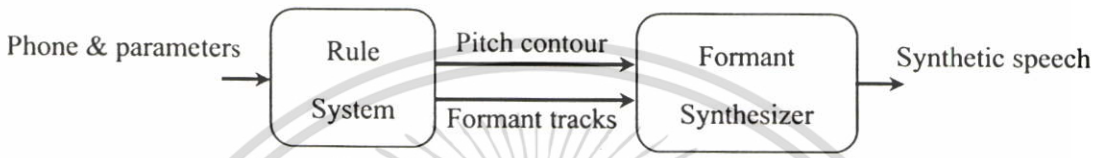
เอกสารนี้สงวนไว้สำหรับใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.7.1 การสังเคราะห์เสียงแบบฟอร์แมนท์ (Formant Synthesis)

การสังเคราะหการสังเคราะห์เสียงแบบฟอร์แมนท์ประกอบไปด้วย 2 ส่วนหลักๆ คือ

1) Rule System คือ ส่วนที่ทำหน้าที่ในการสร้างกฎต่างๆ ส่วนนี้จะเป็นส่วนที่เกี่ยวข้องการสร้างกฎต่างๆ ของแต่ละภาษา ซึ่งในที่นี้จะเป็นการสร้างกฎสำหรับการออกเสียงภาษาไทย

2) Formant Synthesizer คือ ส่วนที่ทำหน้าที่สังเคราะห์เสียง ซึ่งนำข้อมูลคุณสมบัติของเสียงที่จากส่วนของ Rule System ไปทำการสังเคราะห์เสียง ดังแสดงรูปที่ 2.17



รูปที่ 2.17 การสังเคราะห์เสียงแบบฟอร์แมนท์

7.2.1.1 Rule System

การออกเสียงพูดใดๆ เป็นส่วนหนึ่งของ Rule System ซึ่งเป็นผลที่ได้จากการวิเคราะห์จากการทดลอง ซึ่งจะนำค่าที่จำเป็นต่อการสังเคราะห์เสียง เช่น ความถี่ฟอร์แมนท์ ความถี่มูลฐาน มาสร้างให้อยู่ในรูปแบบของกฎ ซึ่งจะเป็นตัวอธิบายถึงข้อมูลของเสียงพูด ซึ่งจะทำการวิเคราะห์เพื่อให้อยู่ในรูปแบบของพารามิเตอร์ ที่สามารถคำนวณได้โดยกฎ ซึ่งกฎต่างๆ สำหรับการสังเคราะห์เสียงแบบฟอร์แมนท์ มีดังนี้

1) กฎความถี่มูลฐาน (Fundamental frequency rule) เป็นกฎที่กำหนดลักษณะของระดับเสียงสูงต่ำ หรือ พิตท์ (Pitch) ซึ่งเป็นข้อมูลที่สำคัญ และเป็นลักษณะเฉพาะของแต่ละภาษา เช่น ภาษาไทย การเปลี่ยนระดับเสียงของพยางค์ จะทำให้เกิดความหมายที่แตกต่างกัน แต่ในภาษาตะวันตก พิตท์ไม่ช่วยให้สามารถระบุคำได้โดยตรง แต่จะจัดการเกี่ยวกับข้อมูลเพิ่มเติม เช่น มีคำไหนในประโยคที่มีลักษณะเด่น เป็นต้น

2) กฎความหนาแน่น (Intensity rule) เป็นกฎที่จัดการกับความดังของเสียง เช่น เสียงไม่ก้อง จะมีความดังน้อยกว่าเสียงก้อง การผันแปรเหล่านี้สามารถรวมเข้าไว้ในข้อมูลของหน่วยเสียงและไม่ต้องกำหนดที่อินพุต ซึ่งค่าความดังจะถูกตั้งเป็นค่าเริ่มต้นสำหรับแต่ละหน่วยเสียง และสามารถนำค่ามาเปลี่ยนแปลงได้ในภายหลัง

3) กฎระยะเวลา (Duration Rule) เป็นกฎที่จัดการระยะเวลาของหน่วยเสียงและพยางค์ ระยะเวลาของหน่วยเสียงจะถูกกำหนดไว้เป็นค่าเริ่มต้นของแต่ละหน่วยเสียง ซึ่งได้จากจากวิเคราะห์หน่วยเสียง และสามารถเปลี่ยนแปลงค่าได้ แต่ค่าที่เปลี่ยนแปลงควรอยู่ในระยะเวลาที่ควรจะเป็น เช่น เสียงสระและสามารถเปลี่ยนแปลงค่าได้ แต่ค่าที่เปลี่ยนแปลงควรอยู่ในระยะเวลาที่ควร

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

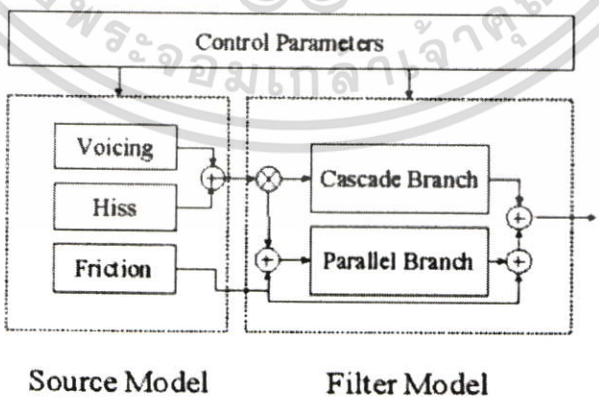
จะเป็น เช่น เสียงสระภาษาไทย ประกอบไปด้วย สระเสียงสั้น และสระเสียงยาว ซึ่งสระเสียงสั้นควรมีระยะไม่เกินกว่าค่าสูงสุดที่สามารถกำหนดได้

4) กฎความถี่ฟอร์แมนท์ (Formant frequency rule) ความถี่ฟอร์แมนท์ของแต่ละหน่วยเสียงจะมีค่าเฉพาะของแต่ละหน่วยเสียง ซึ่งได้จากการวิเคราะห์ความถี่ฟอร์แมนท์ของแต่ละหน่วยเสียง กฎการเปลี่ยนแปลงความถี่ฟอร์แมนท์จะเกี่ยวข้องกับการออกเสียงของเพศชายและเพศหญิง เนื่องจากความถี่ฟอร์แมนท์ของผู้หญิงจะสูงกว่าเสียงผู้ชาย

7.2.1.2 Formant Synthesizer

ก่อนจะกล่าวถึงรายละเอียดของ Formant Synthesizer มาดูหลักการของการเกิดเสียงรูปจากแบบจำลองอย่างง่ายของช่องทางเดินเสียง เป็นลักษณะของท่อทรงกลมตันที่มีต้นกำเนิดเสียงอยู่ที่ปลายข้างหนึ่ง (ส่วนของกล่องเสียง) และปลายอีกข้างหนึ่งจะปิด (ส่วนของปาก) ดังนั้นมักจะเกิดการกำทอนในท่อได้ที่ความยาวคลื่นเท่ากับ 4L, 4L/3, 4L/5, ... โดยที่ L คือความยาวของท่อ ถ้าคิดเป็นความถี่ที่เกิดการกำทอน จะได้ความถี่ที่ c/4L, 3c/4L, 5c/4L, ... Hz โดยที่ c คือค่าความเร็วเสียงในอากาศ และถ้าจะคำนวณหาความถี่การกำทอนของช่องทางเดินเสียงของคน ซึ่งปกติช่องทางเดินเสียงของคนเราจะมีความยาวประมาณ 7 นิ้ว หรือ 17 นิ้ว และ c มีค่าเท่ากับ 340 เมตรต่อวินาที ดังนั้นจึงมีการกำทอนที่ความถี่ประมาณ 500 Hz, 1500 Hz, 2500 Hz, ... เป็นต้น

เมื่อกล่องเสียงกระตุ้นให้เกิดคลื่นที่ประกอบไปด้วยฮาร์มอนิก (Harmonic) ต่างๆ การกำทอนของช่องทางเดินเสียงนี้จะสร้างรูปคลื่นเสียงที่มียอดสูงเด่น ซึ่งจะเรียกว่า ฟอร์แมนท์ (formants) หรือเรโซแนนซ์ (Resonance) ของเสียง จากหลักการที่กล่าวมาข้างต้น เป็นแนวคิดในการสร้าง Formant Synthesizer ซึ่งพยายามสังเคราะห์เสียงจากการจำลองเสียงพูดของมนุษย์ ดังแสดงในรูปที่ 2.18



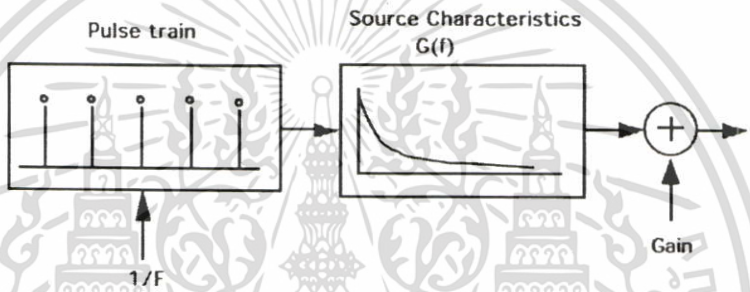
รูปที่ 2.18 โมเดลของ Formant Synthesizer

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Source-filler model เป็นการเปลี่ยนแปลงเส้นทางเดินเสียงด้วย linear filter ที่แปรผันตามเวลา ซึ่งแหล่งกำเนิดเสียงได้มากจาก การจำลองการสั่นของหลอดลมซึ่งทำให้เกิดเสียงก้อง และเสียงซ่าที่จำลองจากการหดคลายตัวบริเวณทางเดินเสียง

1) Source Model เสียงพูดสามารถแบ่งจากการแหล่งกำหนดการสร้างเสียงเป็น ความเป็นรายคาบของการสั่นของเส้นเสียง (เสียงก้อง) และเกิดจากการสร้างโดยปราศจากการสั่นของเส้นเสียง ซึ่งประกอบด้วย เสียงพ่นลม หรือเสียงสอดแทรก (เสียงไม่ก้อง) ซึ่งประกอบด้วยแหล่งกำเนิดเสียงหลักๆ 2 ชนิดคือ Voicing Source และ Fraction Source มีรายละเอียดดังต่อไปนี้

- Voicing Source เป็นแหล่งกระตุ้นสำหรับสร้างเสียงก้อง ซึ่งจากการสั่นของเส้นเสียง ซึ่งแสดงดังรูปที่ 2.19



รูปที่ 2.19 บล็อกไดอะแกรมแหล่งกำเนิดเสียงก้อง

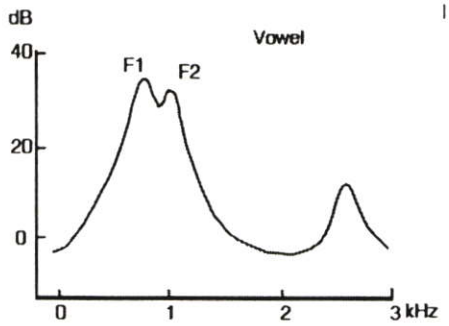
โมเดลนี้ประกอบด้วยการก่อกำเนิด Impulse train ซึ่งจะสร้างพัลส์ในอัตรา 1 ต่อคาบมูลฐาน (fundamental period) การกำเนิดอิมพัลส์นี้เป็นการจำลองการเปิดของเส้นเสียง ซึ่งสัญญาณนี้จะถูกทำงานต่อโดยลิเนียร์ฟิลเตอร์ซึ่งตอบสนองความถี่ G(f) ซึ่งจะจัดให้อยู่ในรูปแบบใกล้เคียงกับ Glottal wave form

Source spectral characteristics เป็นความเกี่ยวข้องโดยตรงกับชนิดการทำงานของเส้นเสียง ซึ่งเป็นฟังก์ชั่นถ่ายโอน (Transfer function)

- Friction Source เป็นการสร้างเสียงไม่ก้องเมื่อเส้นเสียงไม่เกิดการสั่น และหลอดเสียงเปิด ซึ่งพ่นลมออกทางเดินลมโดยไม่มีสิ่งกีดขวางตลอดทางเดินเสียง เมื่อมีการขัดขวางในบางจุดจะทำให้เกิดการสอดแทรก (friction) หรือการระเบิด (Plosion)

2) Filter Model หรือ Vocal tract modeling เป็นพฤติกรรมของเส้นทางเดินเสียงที่จำลองลักษณะของความถี่ฟอร์แมนท์ของเสียงพูด ลักษณะของรีโซแนนซ์จะขึ้นอยู่กับการกำหนดค่าลำดับอย่างเดียวเท่านั้น และจะไม่ขึ้นอยู่กับแหล่งกำเนิดเสียง (Source)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 2.20 ตัวอย่างของสเปกตรัมความถี่ของเสียงสระ

สำหรับเสียงที่เปล่งออกมาเช่น เสียงสระ คลื่นเสียงจะถูกสร้างโดยแหล่งกำเนิดเสียงตลอดทั้งเส้นทางเดินเสียง ความถี่บางความถี่จะสูงขึ้น บางความถี่ก็จะต่ำลง ซึ่งบางความถี่ก็จะได้รับการเพิ่มประสิทธิภาพขึ้นอยู่กับการลักษณะการเสียงของอวัยวะมนุษย์ นั่นก็คือตำแหน่งของขากรรไกร ลิ้น และริมฝีปาก ความถี่จะเกิดขึ้นบริเวณเรโซแนนซ์เด่นของเส้นทางเดินเสียงซึ่งจะเรียกว่า ฟอร์แมนท์ (Formants) เรโซแนนซ์เหล่านี้อาจจะอยู่ในรูปของ all-pole filter ซึ่งโพลเหล่านี้จะเป็นจุดสูงสุดของสเปกตรัมความถี่ ดังแสดงในรูปที่ 2.20 ฟังก์ชั่นถ่ายโอน (transfer function) ของตลอดเส้นทางเดินเสียงได้มาจากการเชื่อมต่อของวงจรต่างๆ ซึ่งประกอบด้วย 2 วงจรเข้าด้วยกัน คือ การเชื่อมต่อแบบขนานและการเชื่อมต่อแบบอนุกรม ดังแสดงในรูปที่ 2.21 และรูปที่ 2.22

- Parallel vocal tract transformation เป็นการเชื่อมต่อแบบขนานของ เรโซเนเตอร์ (resonators) ซึ่งจะประกอบด้วยไป ฟังก์ชั่นถ่ายโอน ดังสมการที่ (2.21) แต่ละเรโซเนเตอร์จะมีการปรับอัตราขยาย (Gain) ก่อน ซึ่งเป็นแอมพลิจูดจุดสูงสุดของสเปกตรัล

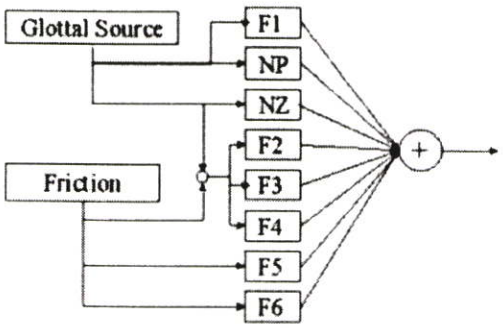
$$H_i(z) = \frac{1}{1 - 2e^{-\pi b_i}\cos(2\pi f_i)z^{-1} + e^{-2\pi b_i}z^{-2}} \qquad (2.21)$$

เมื่อ $H_i(z)$ คือ ฟังก์ชั่นถ่ายโอน

b_i คือแบนวิดต์

f_i คือ ความถี่ฟอร์แมนท์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

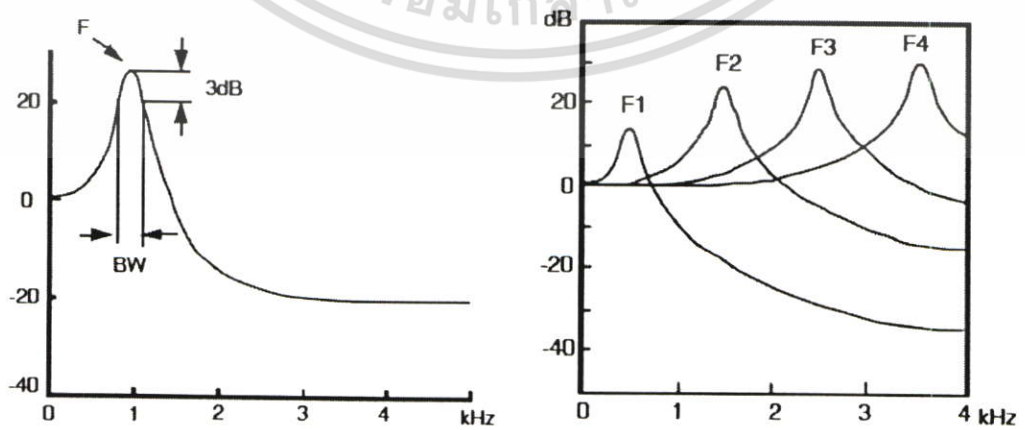


รูปที่ 2.21 การเชื่อมต่อแบบขนานของ Formant Synthesizer

- Cascade vocal track transformation เป็นการเชื่อมต่อเรโซเนเตอร์แบบอนุกรม ดังแสดงในรูปที่ 2.22 เป็นการคูณของสเปกตรัมโดยฟังก์ชั่นถ่ายโอน ดังสมการที่ (2.21) ของแต่ละตัวตามลำดับ การต่ออนุกรมของฟอร์แมนท์เรโซเนเตอร์ ได้ผลดีมากสำหรับการออกเสียงสระ ซึ่งเป็นการเสียงก้อง และเสียงใกล้เคียงธรรมชาติ

รูปที่ 2.22 การเชื่อมต่อแบบอนุกรมของ Formant Synthesizer

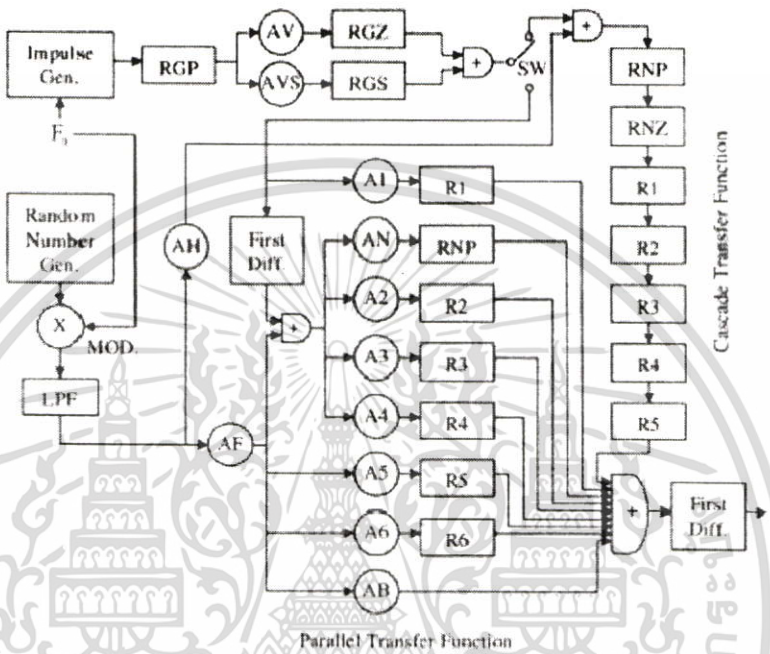
การคำนวณด้วยฟังก์ชั่นถ่ายโอนด้วยความถี่ฟอร์แมนท์และแบนวิดต์ จากรูปที่ 2.23 ด้านซ้ายมือ และเมื่อนำค่าของฟังก์ชันถ่ายโอนมาต่ออนุกรมกัน จะได้ดังรูปที่ 2.23 ด้านขวามือ



รูปที่ 2.23 ผลที่ได้ของฟังก์ชั่นถ่ายโอนของฟอร์แมนท์เรโซเนเตอร์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากหลักการที่กล่าวมาข้างต้น ซึ่งเป็นหลักการที่พัฒนาขึ้นโดย Denish Klatt ดังแสดงในรูปที่ 2.24 ซึ่งเป็นโมเดลที่ได้รับความนิยมและเป็นต้นแบบในการพัฒนาการสังเคราะห์เสียงแบบฟอร์แมนท์มาอย่างต่อเนื่องไม่น้อยกว่า 25 ปี



รูปที่ 2.24 บล็อกไดอะแกรมของ Klatt Formant Synthesizer. [23]

งานวิจัยนี้ใช้วิธีสังเคราะห์แบบฟอร์แมนท์ (Formant Synthesis) โดยใช้ซอฟแวร์ Klatt Synthesizer ในการแปลงความถี่ฟอร์แมนท์และพารามิเตอร์ให้เป็นเสียงพูด โดยใช้ผลที่ได้จากการวิเคราะห์คุณสมบัติของเสียงพูดภาษาไทย แล้วกำหนดให้อยู่ในรูปแบบของกฎ ซึ่งจะได้อธิบายในบทต่อๆไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3
วิธีการดำเนินการวิจัย

3.1 การคัดเลือกผู้บอกภาษา

การคัดเลือกผู้บอกภาษาในวิทยานิพนธ์นี้ แบ่งผู้บอกภาษาเป็น 2 กลุ่ม คือ กลุ่มผู้บอกภาษาเพื่อวิเคราะห์คุณสมบัติหน่วยเสียง และกลุ่มผู้บอกภาษาเพื่อวิเคราะห์การผันวรรณยุกต์

3.1.1 กลุ่มผู้บอกภาษาเพื่อวิเคราะห์คุณสมบัติหน่วยเสียง

กลุ่มผู้บอกภาษากลุ่มนี้ จะแบ่งตามวัยเป็น 3 กลุ่ม คือ วัยเด็ก วัยรุ่น และวัยผู้ใหญ่ โดยกำหนดอายุวัยเด็ก 7-12 ปี วัยรุ่น13-25 ปี และวัยผู้ใหญ่อายุ 26 ปีขึ้นไป โดยเลือกเด็กหญิง 3 คน เด็กชาย 3 คน วัยรุ่นชาย 3 คน วัยรุ่นหญิง 3 คน วัยผู้ใหญ่ชาย 3 คน และวัยผู้ใหญ่หญิง 3 คน รวมทั้งหมด18 คน

3.1.2 กลุ่มผู้บอกภาษาเพื่อวิเคราะห์การผันวรรณยุกต์

ผู้บอกภาษากลุ่มนี้ทำการคัดเลือกผู้บอกภาษาที่มีความรู้ทางด้านภาษาศาสตร์เป็นอย่างดี เป็นผู้ที่สำเร็จการศึกษาทางด้านภาษาศาสตร์ หรือได้รับใบผู้ประกาศ โดยเป็นผู้หญิง 2 คน และผู้ชาย 2 คน รวมเป็น 4 คน

3.2 การเก็บตัวอย่างข้อมูลเสียงพูด

การเก็บตัวอย่างข้อมูลเสียงพูดจากผู้บอกภาษา ในการบันทึกเสียงจะบันทึกอยู่ในรูปแบบ wave form ซึ่งเป็นไฟล์ข้อมูล ".wav" อัตราการสุ่มตัวอย่าง 22.050 kHz, ขนาดตัวอย่าง 8 บิต

3.2.1 ตัวอย่างเสียงพยัญชนะสระและตัวสะกด

ทำการบันทึกเสียงของพยัญชนะ 21 เสียง สระ 24 เสียง และตัวสะกด 8 เสียง

3.2.2 ตัวอย่างเสียงการผันวรรณยุกต์

ทำการบันทึกเสียงการผันวรรณยุกต์สามัญ เอก โท ตรี และจัตวา ในคราวเดียวกัน โดยใช้เสียงของสระในการผันวรรณยุกต์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.3 เครื่องมือและอุปกรณ์ในการวิจัย

3.3.1 ฮาร์ดแวร์

1. เครื่องคอมพิวเตอร์โน้ตบุค รุ่น Acer TravelMate 230,Mobile Intel Celeron CPU 2.00 GHz, RAM 256 MB, Crystal WDM Audio
2. ไมโครโฟน 1 ตัว
3. ชุดลำโพงรุ่น AV-2126L, Output: RMS 15Wx2+30W woofer, Frequency response: Woofer: 30Hz-125Hz, Satellites: 70Hz-20kHz

3.3.2 ซอฟแวร์

1. ระบบปฏิบัติการ Microsoft Windows XP Home Edition
2. โปรแกรมชุด Microsoft Office XP
3. โปรแกรม MATLAB 6.5
4. โปรแกรม Speech Analyzer (Summer Institute of Linguistics) version 1.5
5. โปรแกรม Klatt Synthesizer

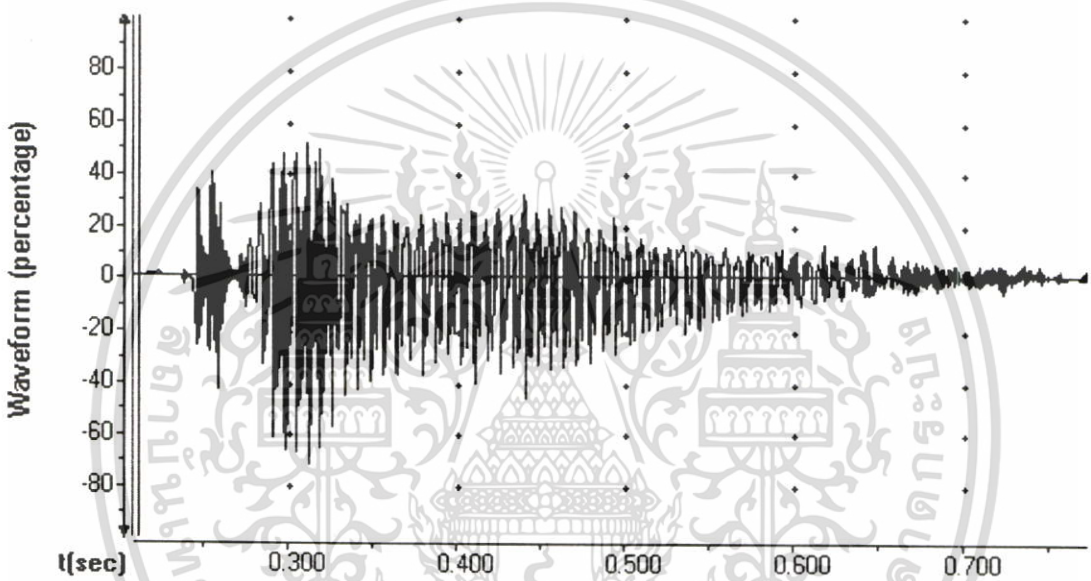
3.4 ประเมินผลการสังเคราะห์เสียง

ประเมินผลการสังเคราะห์เสียงทำเพื่อวัดคุณภาพเสียงที่สังเคราะห์ขึ้น เพื่อวัดความสามารถในการรับรู้ของเสียงที่สังเคราะห์ โดยคัดเลือกผู้เข้ารับการประเมินจากกลุ่มของนักเรียนระดับประถมศึกษาปีที่ 4 ถึง ระดับมัธยมศึกษาปีที่ 3 ทั้งหญิงและชาย โดยผู้เข้ารับการประเมินไม่เป็นบุคคลที่มีปัญหาทางด้านการได้ยิน ไม่เป็นผู้ที่มีปัญหาด้านการเรียนรู้ และสามารถอ่านภาษาไทยได้เป็นอย่างดี ผู้เข้ารับการประเมินจำนวน 40 คน ฟังเสียงพยางค์ภาษาไทยที่สังเคราะห์ขึ้นจำนวน 40 พยางค์ ในห้องที่มีการควบคุมสภาพแวดล้อมเสียง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4
วิเคราะห์คุณสมบัติของเสียงภาษาไทย

ในการวิเคราะห์สัญญาณเสียง โดยพื้นฐานสัญญาณเสียงมีคุณสมบัติแปรเปลี่ยนไปตามเวลา ดังนั้นการวิเคราะห์หาคุณคุณสมบัติต่างๆ ของเสียงจึงต้องทำในช่วงเวลาสั้น (Short time) ซึ่งคุณสมบัติชั่วคราวของเสียงในระยะเวลาสั้นสามารถอ้างได้ว่าคงที่ในช่วงระยะเวลา 10 ถึง 30 มิลลิวินาที [20] จากคุณสมบัติของเสียงดังกล่าวนำไปสู่วิธีการวิเคราะห์ต่างๆ ดังต่อไปนี้

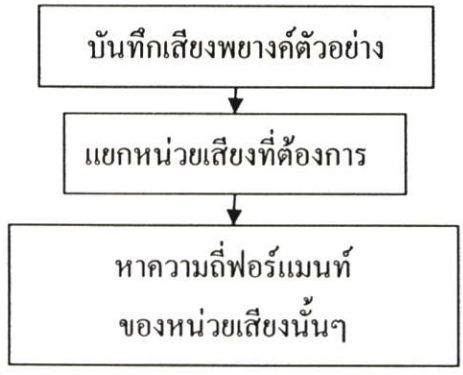


รูปที่ 4.1 สัญญาณเสียงของพยางค์ "กอ"

4.1 วิเคราะห์ความถี่ฟอร์แมนท์ของเสียงภาษาไทย

ความถี่ฟอร์แมนท์เป็นคุณสมบัติที่สำคัญอย่างหนึ่งของเสียง การวิเคราะห์ความถี่ฟอร์แมนท์ทำได้โดยการนำสัญญาณเสียงตัวอย่างที่ต้องการวิเคราะห์ ไปแยกเอาส่วนที่ต้องการวิเคราะห์ แล้วนำไปเข้าฟังก์ชั่นเพื่อหาความถี่ฟอร์แมนท์ ดังแสดงในรูปที่ 4.2

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



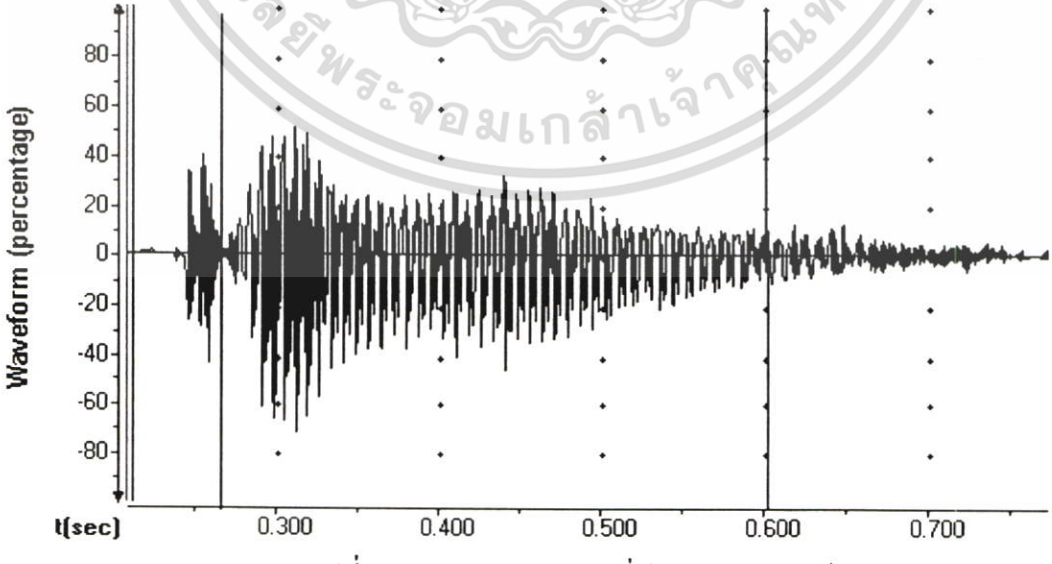
รูปที่ 4.2 ขั้นตอนการหาความถี่ฟอร์แมนท์สำหรับเสียงพูดภาษาไทย

4.1.1 การบันทึกพยางค์ตัวอย่าง

เพื่อให้ได้ข้อมูลตัวอย่างที่หลากหลาย และครอบคลุมเสียงที่เกิดจากคนไทยพูด เพื่อนำข้อมูลความถี่ฟอร์แมนท์มาทำการวิเคราะห์ข้อมูลทางสถิติ โดยใช้ประชากร(Population) คือคนไทย แล้วทำการเลือกกลุ่มตัวอย่าง (Sample) โดยใช้มาตราวัดแบบแบ่งกลุ่มคือ เพศชายและเพศหญิง โดยแต่ละกลุ่มจะเลือกกลุ่มอายุ คือ เด็ก วัยรุ่น และวัยผู้ใหญ่ แล้วทำการบันทึกเสียงตัวอย่าง โดยให้อ่านออกเสียงพยัญชนะ สระ และตัวสะกด

4.1.2 แยกหน่วยเสียงที่ต้องการ

นำไฟล์เสียงที่ได้จากการบันทึกของบอกภาษา มาทำการแยกเอาเฉพาะหน่วยเสียงที่ต้องการวิเคราะห์ ในรูปที่ 4.1 เป็นสัญญาณเสียงของพยางค์ "กอ" ถ้าต้องการวิเคราะห์เสียงสระออก็ตัดเอาเฉพาะส่วน สระออ ดังรูปที่ 4.3

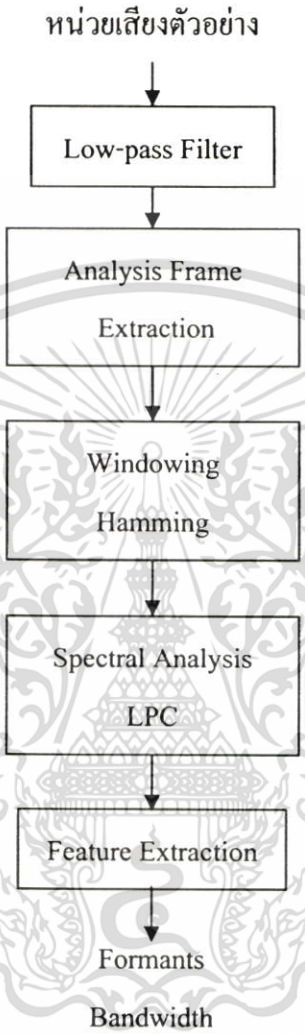


รูปที่ 4.3 การตัดหน่วยเสียงที่ต้องการวิเคราะห์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.1.3 หาความถี่ฟอร์แมนท์ของหน่วยเสียง

การหาความถี่ฟอร์แมนท์ของหน่วยเสียงภาษาไทย



รูปที่ 4.4 กระบวนการหาความถี่ฟอร์แมนท์ของหน่วยเสียง [20]

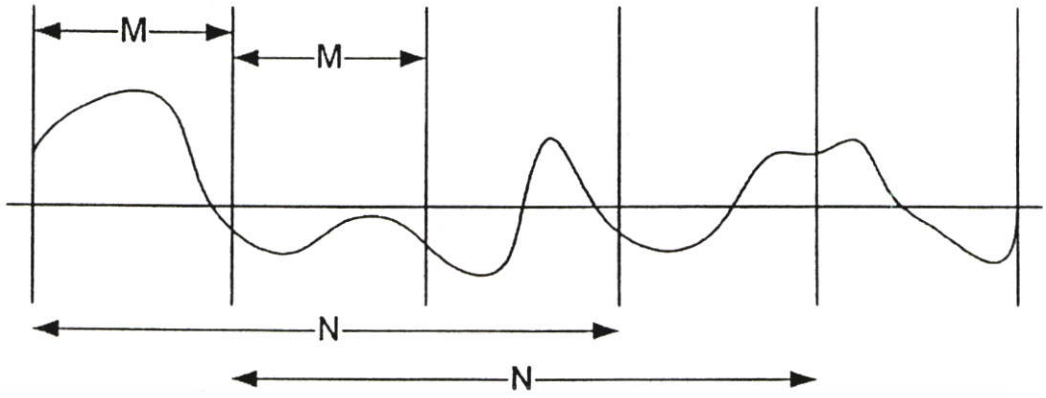
4.1.3.1 กรองต่ำผ่าน (Low pass Filter)

เนื่องจากสัญญาณเสียงพูดของมนุษย์ มีองค์ประกอบส่วนใหญ่อยู่บริเวณความถี่ต่ำ จึงทำการกรองความความถี่ต่ำผ่าน ที่ 8 kHz [20] เพื่อทำการกรองความถี่รบกวนทิ้งไป

4.1.3.2 การแบ่งช่วงสัญญาณในการวิเคราะห์ (Analysis Frame Extraction)

สัญญาณที่ผ่านการกรองต่ำผ่านแล้ว จะถูกตัดออกเป็นช่วงๆ หรือ เฟรม ช่วงละ N ตัวอย่างสัญญาณ การวิเคราะห์จะวิเคราะห์ทีละช่วงของแต่ละ N ตัวอย่างสัญญาณ ดังรูปที่ 4.5

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.5 การแบ่งช่วงของสัญญาณที่ใช้ในการวิเคราะห์

โดยช่วงการวิเคราะห์แต่ละช่วงจะถูกเลื่อนไปเป็นระยะ M ช่วงสัญญาณ จะเห็นได้ว่า ถ้าค่า M โตกว่าค่า N ในการเลื่อนของช่วงในการวิเคราะห์จะทำให้บางสัญญาณไม่ถูกใช้ในการวิเคราะห์ ก็จะเป็นการสูญเสียส่วนหนึ่งทำให้ผลที่ได้ไม่ถูกต้องเท่าที่ควร ถ้าค่า M เล็กกว่า N ก็จะทำให้ตัวอย่างสัญญาณทุกตัวถูกนำมาวิเคราะห์ ยิ่งค่า M เล็กเท่าใด ความแม่นยำในการวิเคราะห์ก็จะยิ่งสูงขึ้นเท่านั้น แต่ก็จะทำให้การคำนวณช้าลงตามไปด้วย

หลักการกำหนดขนาดของเฟรม

1. เฟรมจะต้องสั้นพอที่คุณสมบัติของเสียงที่สนใจจะวิเคราะห์นั้น ยังไม่เปลี่ยนแปลงในเฟรม
2. เฟรมจะจะต้องยาวพอที่จำนวนตัวอย่างสัญญาณในเฟรม สามารถนำมาคำนวณหาคุณสมบัติที่ต้องการได้
3. เฟรมที่ติดกัน ไม่ควรจะสั้นจนกระโดดข้ามข้อมูลบางส่วนไป แต่ควรเลื่อนเฟรมให้น้อยกว่าขนาดของเฟรม

4.1.3.3 การวินโดว์ (Windowing)

ถ้าพิจารณาช่วงสัญญาณ N ตัวอย่างสัญญาณของช่วงใดๆ ที่นำมาวิเคราะห์ จะเห็นได้ว่าที่ขอบของเฟรมที่ตัดมานี้มีความไม่ต่อเนื่องของสัญญาณ ถ้ามองในโดเมนความถี่สูง ก็จะมีความถี่สูงเกิดขึ้น ดังนั้นเพื่อลดองค์ประกอบความถี่ที่สูงเหล่านี้ จะคูณด้วยฟังก์ชั่นวินโดว์เพื่อลดความไม่ต่อเนื่องของสัญญาณที่ขอบ และไม่ทำให้สเปกตรัมของสัญญาณในช่วงความถี่ต่ำเปลี่ยนไปมากนัก ในวิทยานิพนธ์นี้เลือกใช้ฟังก์ชันวินโดว์แฮมมิ่ง (Hamming window function)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.1.3.4 ประมาณการเชิงเส้น (Spectral Analysis LPC)

การหากรอบสเปกตรัมช่วงสั้นของเสียงพูดด้วยวิธีการวิเคราะห์การประมาณการเชิงเส้น คำนวณได้จาก H(z) จากสมการที่ (2.10) บนวงกลมหนึ่งหน่วย โดยขั้นแรกต้องหาสัมประสิทธิ์การการประมาณค่าแบบเชิงเส้นของเสียงพูดก่อน โดยกำหนดให้ LPC คือ ฟังก์ชันการหาสัมประสิทธิ์ประมาณการเชิงเส้น ซึ่งแสดงในในการที่ (4.1)

$$a = LPC(x, p) \tag{4.1}$$

เมื่อ a คือ สัมประสิทธิ์พหุนามนามประมาณแบบเชิงเส้น a =[1, a(2), a(3), ...,a(p+1)]

x คือ สัญญาณเสียง

p คือ อันดับของการวิเคราะห์ประมาณเชิงเส้น

สัมประสิทธิ์พหุนามนามประมาณแบบเชิงเส้นที่ได้จากการคำนวณนี้จะนำไปใช้ในขั้นตอนของการหาคุณสมบัติของเสียงต่อไป

4.1.3.5 การหาคุณสมบัติของเสียง (Feature Extraction)

การหาคุณสมบัติของเสียงในที่นี้จะหาค่าความถี่ฟอร์แมนท์(Formants) และแบนวิดต์ (Bandwidth) ของสัญญาณเสียง โดยนำค่าสัมประสิทธิ์พหุนามประมาณแบบเชิงเส้น (LPC polynomial) ทำการคำนวณหาราก (Root) ดังแสดงในสมการ(4.2)

$$z_i = Roots(a) \tag{4.2}$$

เมื่อ z_i คือ รากของสัมประสิทธิ์พหุนามนามประมาณแบบเชิงเส้น

a คือ สัมประสิทธิ์พหุนามนามประมาณแบบเชิงเส้น

โพลของสเปกตรัมจะอยู่ใน z_i (i=1, 2, ...,p) ซึ่งเป็นจำนวนเชิงซ้อนที่ตอบสนองต่อความถี่ฟอร์แมนท์และแบนวิดต์ โดยหาได้จากสมการ [20]

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$F_i = \frac{\arg z_i}{2\pi\Delta T} \tag{4.3}$$

เมื่อ F_i คือ ความถี่ฟอร์แมนท์

$\arg z_i$ คือ มุมของ z_i

ΔT คือ คาบของ sampling

และสามารถหาแบนวิดต์ได้จากสมการต่อไปนี้

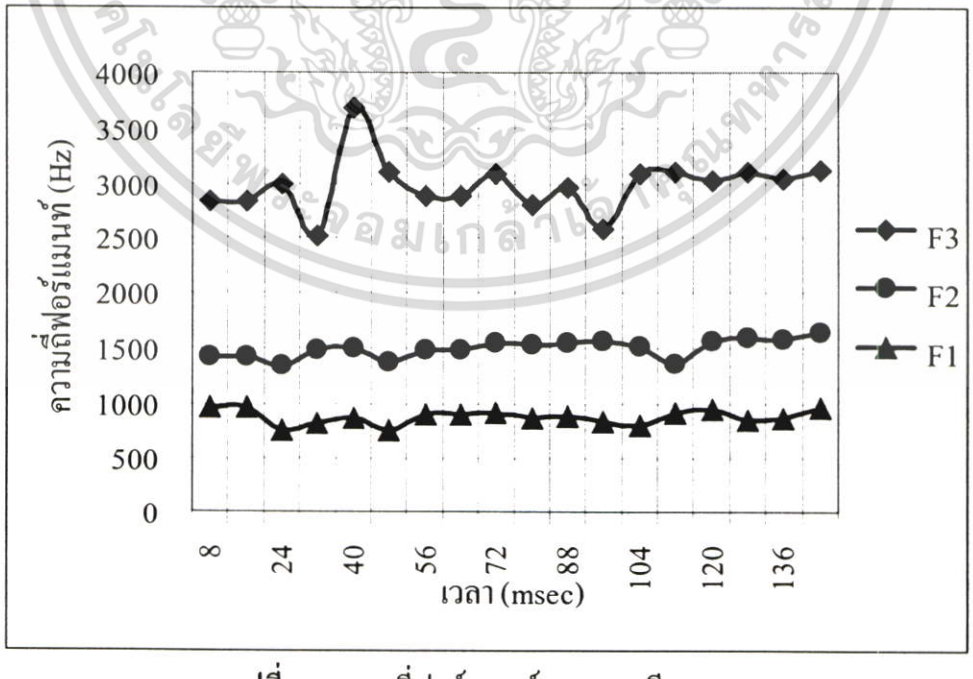
$$B_i = \frac{\log |z_i|}{\pi\Delta T} \tag{4.3}$$

เมื่อ B_i คือ แบนวิดต์

$|z_i|$ คือ ขนาดของ z_i

ΔT คือ คาบของ sampling

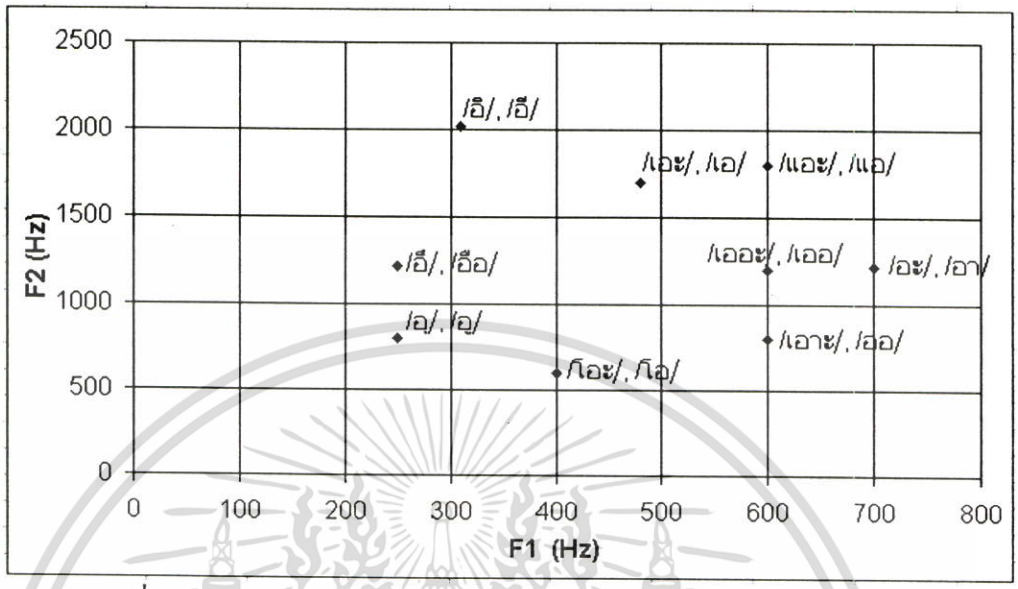
ผลที่ได้จากการคำนวณนี้จะได้ลำดับของความถี่ฟอร์แมนท์และแบนวิดต์ที่ต้องการดังแสดงในรูปที่ 4.6



รูปที่ 4.6 ความถี่ฟอร์แมนท์ของสระเสียง /อา/

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ในรูปที่ 4.7 ความสัมพันธ์ระหว่างความถี่ฟอร์แมนท์ F1 และ F2 ที่มีผลต่อการเกิดเสียงสระภาษาไทย



รูปที่ 4.7 แผนภูมิแสดงความสัมพันธ์ระหว่างความถี่ฟอร์แมนท์ F1 และ F2

4.2 วิเคราะห์ความถี่มูลฐานสำหรับการผันวรรณยุกต์ภาษาไทย

เนื่องจากภาษาไทยเป็นภาษาที่มีเสียงวรรณยุกต์คือ มีระดับโทนเสียง สามัญ เอก โท ตรี และจัตวา ซึ่งแต่ละโทนเสียงก็จะให้ความหมายที่แตกต่างกันออกไป ดังนั้นความถี่มูลฐานจึงเป็นสิ่งที่สำคัญในการสังเคราะห์เสียงภาษาไทย

สระซึ่งเป็นแกนกลางพยางค์มีลักษณะเป็นเสียงก้อง ซึ่งเสียงก้องจะมีความเป็นรายคาบของระดับเสียง (Pitch Period) และส่วนกลับของระยะเวลารายคาบของสัญญาณเสียงคือความถี่มูลฐาน (Fundamental Frequency) เราใช้คุณสมบัตินี้เพื่อวิเคราะห์เสียงวรรณยุกต์ จากคุณสมบัติของความถี่มูลฐานนี้เอง ที่จะนำไปวิเคราะห์เพื่อหากฎการผันวรรณยุกต์ซึ่งจะสร้างให้อยู่ในรูปแบบของสมการทางคณิตศาสตร์ โดยเริ่มแรกจะทำการวิเคราะห์สัญญาณเสียงการผันวรรณยุกต์เพื่อหาความถี่มูลฐาน และนำไปวิเคราะห์หารูปแบบของสมการ ดังกล่าวในรายละเอียดต่อไปนี้

4.2.1 การหาความถี่มูลฐาน

ขั้นตอนการหาความถี่มูลฐาน เริ่มต้นด้วยการบันทึกเสียงการผันวรรณยุกต์จากผู้บอกภาษา แล้วทำการคำนวณหาความถี่มูลฐาน สุดท้ายทำการปรับปรุงความถี่มูลฐาน เพื่อนำไปใช้ในกระบวนการอื่นต่อไป ดังแสดงในรูปที่ 4.8

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บันทึกเสียงการผันวรรณยุกต์

คำนวณความถี่มูลฐาน
ของเสียงวรรณยุกต์

ปรับปรุงความถี่มูลฐาน

ความถี่มูลฐาน

รูปที่ 4.8 ขั้นตอนการหาความถี่มูลฐาน

4.2.1.1 บันทึกเสียงการผันวรรณยุกต์

บันทึกข้อมูลเสียงการผันวรรณยุกต์จากผู้บอกบอกภาษาที่ได้คัดเลือกแล้วในห้องที่มีการควบสภาพแวดล้อมของเสียง โดยบันทึกเสียงการผันวรรณยุกต์ทั้ง 5 เสียงในคราวเดียวกัน

4.2.1.2 การคำนวณความถี่มูลฐานของเสียงวรรณยุกต์

เสียงวรรณยุกต์แต่ละเสียงที่เกิดขึ้นเกิดจากการเปลี่ยนแปลงของความถี่มูลฐาน ซึ่งการคำนวณหาความถี่มูลฐานนั้นทำได้หลายวิธี วิทยานิพนธ์นี้ใช้วิธีออโตคอร์รีเลชั่น ผลที่ได้จากการคำนวณความถี่มูลฐานของไฟล์เสียงตัวอย่าง ตัวอย่างของค่าที่ได้จากการคำนวณความถี่มูลฐานดังแสดงในตารางที่ 4.1

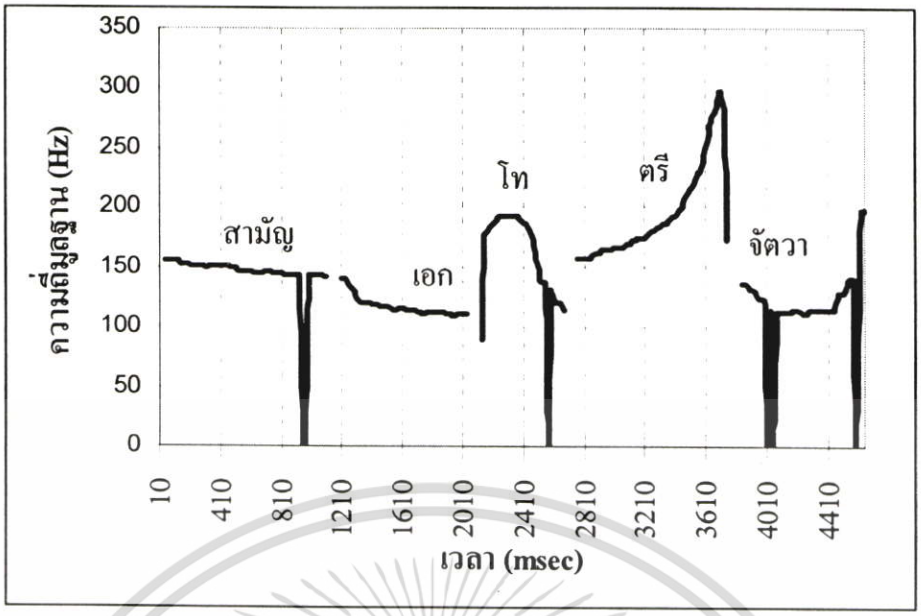
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.1 ค่าความถี่มูลฐานที่ได้จากการคำนวณวรรณยุกต์สามัญ

เวลา (msec)	ความถี่มูลฐาน (Hz)	เวลา (msec)	ความถี่มูลฐาน (Hz)	เวลา (msec)	ความถี่มูลฐาน (Hz)
10	155.28	410	151.03	810	143.18
30	155.28	430	148.99	830	143.18
50	155.28	450	148.99	850	143.18
70	155.28	470	148.99	870	143.18
90	155.28	490	147	890	143.18
110	153.13	510	147	910	0
130	153.13	530	147	930	0
150	153.13	550	147	950	143.18
170	151.03	570	147	970	143.18
190	151.03	590	145.07	990	143.18
210	151.03	610	145.07	1010	143.18
230	151.03	630	145.07	1030	143.18
250	151.03	650	145.07	1050	141.35
270	148.99	670	147	1070	141.35
290	148.99	690	147		
310	151.03	710	145.07		
330	151.03	730	145.07		
350	151.03	750	145.07		
370	151.03	770	145.07		
390	151.03	790	143.18		

เมื่อนำข้อมูลที่ได้จากการคำนวณไปสร้างแผนภูมิจะเห็นว่า ข้อมูลความถี่มูลฐานจะเกิดการกระโดด ซึ่งเป็นผลจากความผิดพลาดจากการคำนวณ เมื่อนำข้อมูลการคำนวณความถี่มูลฐานของเสียงวรรณยุกต์ทั้ง 5 เสียง มาสร้างเป็นแผนภูมิเส้น ดังแสดงในรูปที่ 4.9

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.9 ความถี่มูลฐานของเสียงวรรณยุกต์

4.2.1.3 การปรับปรุงความถี่มูลฐาน

เนื่องจากการหาสมการของเสียงวรรณยุกต์อาศัยแผนภูมิค่าความถี่มูลฐานแปรผันตามเวลา จึงจำเป็นต้องปรับปรุงให้แผนภูมิเส้นมีคุณลักษณะของการผันวรรณยุกต์เรียบขึ้น เพื่อให้เหมาะสมก่อนนำไปวิเคราะห์หาสมการของแต่ละเสียงวรรณยุกต์ โดยใช้วิธีการปรับเรียบโดยค่ากลาง ณ จุดข้อมูลที่เราสนใจ ทำการกำหนดขนาดหน้าต่างขึ้น โดยจุดข้อมูลที่เราสนใจจะเป็นจุดกึ่งกลางของข้อมูลภายในกรอบหน้าต่างที่กำหนดขึ้น จากนั้นแทนค่ากลางที่หาได้ลงที่จุดกึ่งกลางหน้าต่าง ในวิทยานิพนธ์นี้เลือกขนาดหน้าต่างเท่ากับ 5 กรรมวิธีนี้ทำให้ข้อมูลไม่เกิดการกระโดด ดังแสดงในตารางที่ 4.2

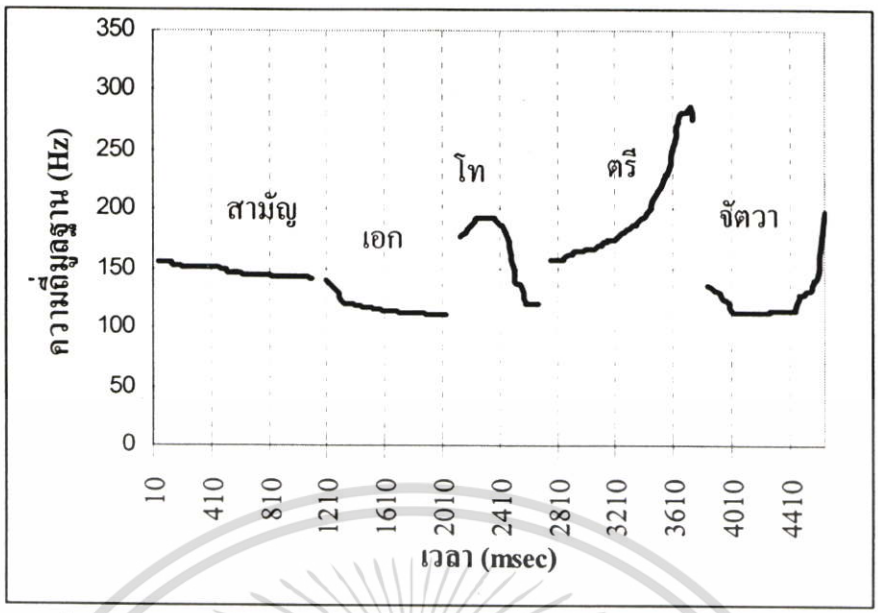
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.2 ค่าความถี่มูลฐานที่ได้จากการปรับปรุงของวรรณยุกต์สามัญ

เวลา (msec)	ความถี่มูลฐาน (Hz)	เวลา (msec)	ความถี่มูลฐาน (Hz)	เวลา (msec)	ความถี่มูลฐาน (Hz)
10	155.28	410	151.03	810	143.18
30	155.28	430	148.99	830	143.18
50	155.28	450	148.99	850	143.18
70	155.28	470	148.99	870	143.18
90	155.28	490	147	890	143.18
110	153.13	510	147	910	143.18
130	153.13	530	147	930	143.18
150	153.13	550	147	950	143.18
170	151.03	570	147	970	143.18
190	151.03	590	145.07	990	143.18
210	151.03	610	145.07	1010	143.18
230	151.03	630	145.07	1030	143.18
250	151.03	650	145.07	1050	142.265
270	151.03	670	145.07	1070	141.35
290	151.03	690	145.07		
310	151.03	710	145.07		
330	151.03	730	145.07		
350	151.03	750	145.07		
370	151.03	770	145.07		
390	151.03	790	143.18		

เมื่อนำข้อมูลที่ได้จากการปรับปรุงความถี่มูลฐานของเสียงวรรณยุกต์ทั้ง 5 เสียง มาสร้างเป็นแผนภูมิเส้น ดังแสดงในรูปที่ 4.10

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.10 แผนภูมิเส้นของเสียงวรรณยุกต์หลังการปรับปรุงความถี่มูลฐาน

นำแผนภูมิเส้นแสดงความถี่มูลฐานบนแกนเวลาของแต่ละเสียงวรรณยุกต์มาทำการวิเคราะห์เพื่อหาสมการที่แปรผันตามเวลา จากผลการทดลองพบว่าความถี่มูลฐานของเสียงแต่ละบุคคลมีค่าไม่เท่ากัน แต่เมื่อพิจารณารูปทรงทางเรขาคณิตแล้วพบว่ามีรูปทรงที่คล้ายคลึงกัน ซึ่งสามารถหาสมการทางคณิตศาสตร์แต่ละเสียงวรรณยุกต์ดังนี้

4.2.2 วิเคราะห์การผันวรรณยุกต์สามัญ

ทำการวิเคราะห์การผันวรรณยุกต์สามัญจากข้อมูลความถี่มูลฐานที่ได้จากกระบวนการข้างต้น จากการพิจารณาแผนภูมิเส้นของความถี่มูลฐานบนแกนเวลา พบว่าแผนภูมิเส้นมีแนวโน้มเป็นเส้นตรง ดังนั้นการผันวรรณยุกต์สามัญจะมีลักษณะเป็นสมการเส้นตรง ทั้งนี้สืบเนื่องมาจากข้อมูลที่วัดมาได้นั้นเป็นข้อมูลที่เกิดขึ้นจริงในทางปฏิบัติซึ่งไม่มีความเที่ยงตรงสูงพอ จากข้อมูลที่ได้จากลักษณะเช่นนี้ เราจะเห็นได้ว่า วิธีที่ดีที่สุดควรเป็นการประดิษฐ์เส้นตรง ซึ่งลากผ่านจุดเหล่านี้โดยพยายามให้เกิดค่าความผิดพลาดเฉลี่ยที่น้อยที่สุด หากเราทำการลากลากเส้นตรงโดยการประมาณจากสายตา บุคคลต่างกันก็จะลากเส้นที่ต่างกัน ซึ่งวิธีนี้จะไม่ก่อให้เกิดเส้นตรงที่เป็นมาตรฐานจากข้อมูลชุดเดียวกัน ซึ่งเสียงสามัญค่าของข้อมูลค่อนข้างเป็นข้อมูลแบบเชิงเส้น จึงเลือกใช้การถดถอยแบบเชิงเส้น [24]

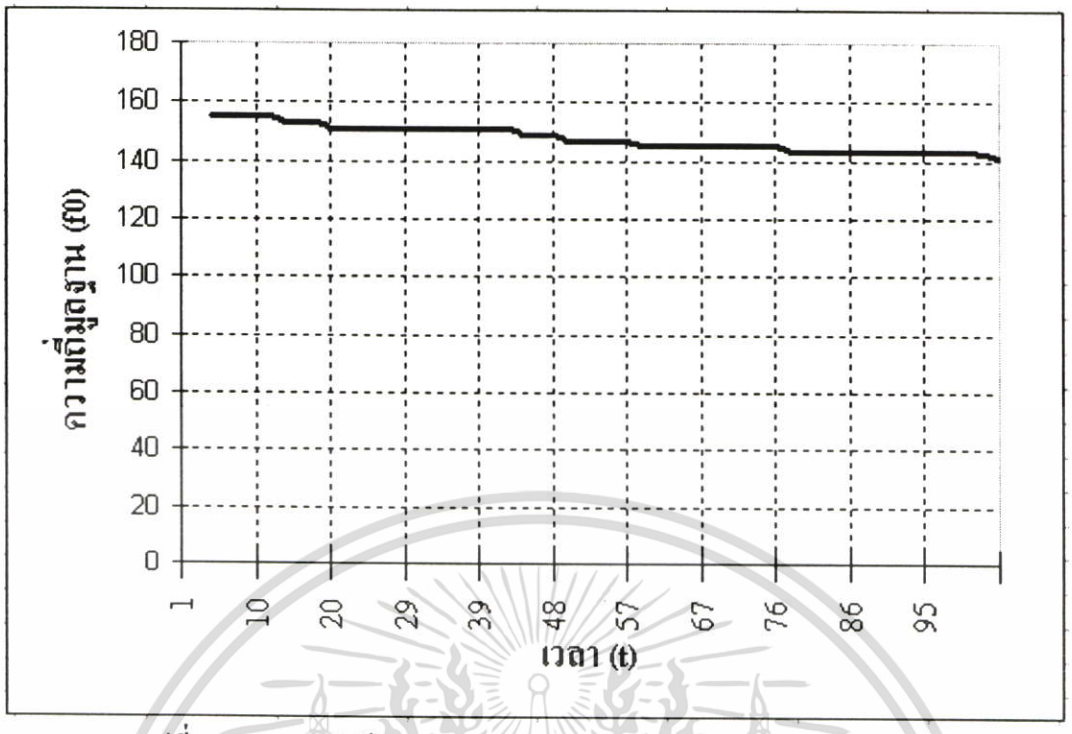
แต่เนื่องจากเสียงวรรณยุกต์จะเกิดขึ้นกับเสียงภาษาไทยหนึ่งพยางค์ ซึ่งช่วงเวลาของแต่ละพยางค์จะมีค่าไม่เท่ากัน ดังนั้นจึงต้องช่วงเวลาซึ่งมีหน่วยเป็นมิลลิวินาทีให้อยู่ในกรอบเป็น 100 หน่วย ซึ่งแสดงในตารางที่ 4.3

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.3 ชุดข้อมูลการผันวรรณยุกต์เสียงสามัญ

ลำดับ (i)	จัดกรอบ (t)	ความถี่มูลฐาน (f0)	ลำดับ (i)	จัดกรอบ (t)	ความถี่มูลฐาน (f0)
1	1	155.28	28	51.76	147
2	2.88	155.28	29	53.64	147
3	4.76	155.28	30	55.52	145.07
4	6.64	155.28	31	57.4	145.07
5	8.52	155.28	32	59.28	145.07
6	10.4	153.13	33	61.16	145.07
7	12.28	153.13	34	63.04	145.07
8	14.16	153.13	35	64.92	145.07
9	16.04	151.03	36	66.8	145.07
10	17.92	151.03	37	68.68	145.07
11	19.8	151.03	38	70.56	145.07
12	21.68	151.03	39	72.44	145.07
13	23.56	151.03	40	74.32	143.18
14	25.44	151.03	41	76.2	143.18
15	27.32	151.03	42	78.08	143.18
16	29.2	151.03	43	79.96	143.18
17	31.08	151.03	44	81.84	143.18
18	32.96	151.03	45	83.72	143.18
19	34.84	151.03	46	85.6	143.18
20	36.72	151.03	47	87.48	143.18
21	38.6	151.03	48	89.36	143.18
22	40.48	148.99	49	91.24	143.18
23	42.36	148.99	50	93.12	143.18
24	44.24	148.99	51	95	143.18
25	46.12	147	52	96.88	143.18
26	48	147	53	98.76	142.265
27	49.88	147	54	100.64	141.35

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



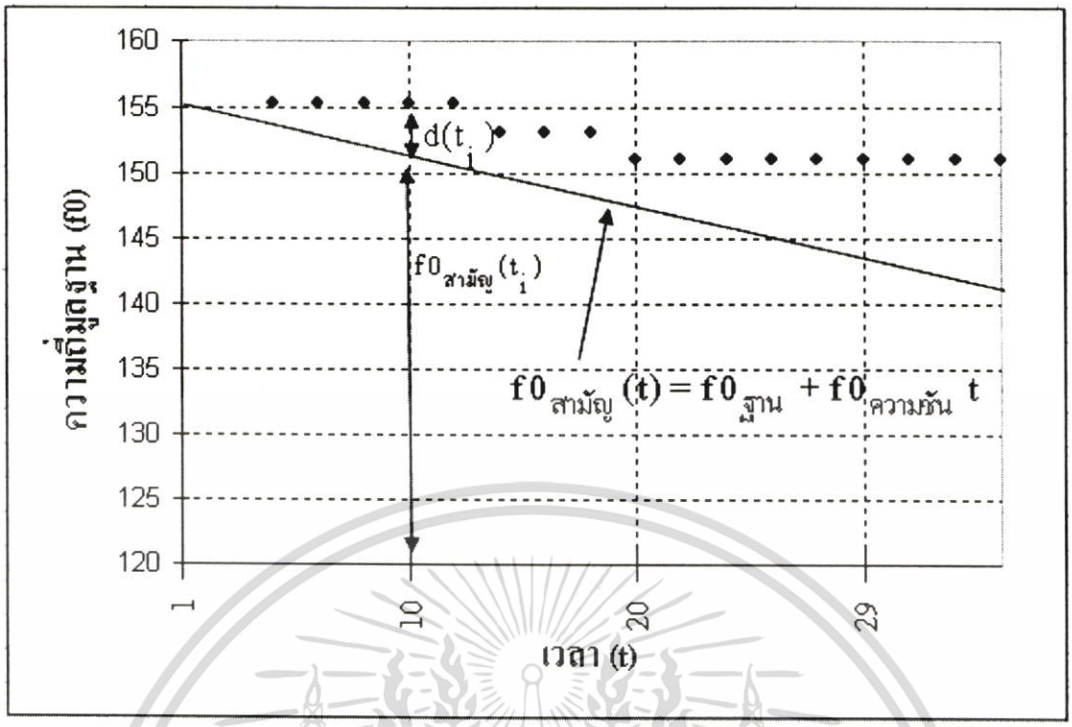
รูปที่ 4.11 แผนภูมิเส้นแสดงการจัดกรอบเวลาใหม่ของเสียงวรรณยุกต์สามัญ

จากตารางที่ 4.3 แสดงชุดข้อมูลที่ประกอบด้วย $f0_i$, t_i, i=1, 2, ..., n นั่นคือมีจำนวนข้อมูลทั้งสิ้น n ข้อมูล ซึ่งจะประดิษฐ์สมการเส้นตรงในรูปแบบของฟังก์ชั่น

$$f0_{\text{สามัญ}}(t) = f0_{\text{ฐาน}} + f0_{\text{ความชัน}}\, t \qquad (4.4)$$

โดย $f0_{\text{ฐาน}}$ และ $f0_{\text{ความชัน}}$ เป็นค่าคงตัวที่ไม่รู้ค่าและจำเป็นต้องคำนวณหาจากเงื่อนไขที่ว่าสมการเส้นตรงที่ประดิษฐ์ขึ้นมานี้จะก่อให้เกิดค่าความผิดพลาดโดยเฉลี่ยที่น้อยที่สุดจากข้อมูลที่ได้จากการทดลอง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.12 การถดถอยแบบเชิงเส้นโดยการประดิษฐ์ฟังก์ชั่นเส้นตรงจากผลการทดลองของเสียงวรรณยุกต์สามัญ

จากรูป 4.12 จะเห็นได้ว่า ณ ตำแหน่ง t_i ใดๆ ค่าของฟังก์ชัน $f0_{สามัญ}(t_i)$ ที่ประดิษฐ์ขึ้นจะมีค่าที่แตกต่างไปจากค่าของข้อมูล $f0_i$ เท่ากับ $d(t_i)$ ที่ตำแหน่งนั้น นั่นหมายความว่า ค่าความผิดพลาด E ทั้งหมดที่เกิดขึ้นมาจากข้อมูลทั้งหมด n ข้อมูล สามารถเขียนได้ดังนี้

$$E = \sum_{i=1}^{n} \left[d(t_i) \right]^2 \tag{4.5}$$

ซึ่งในที่นี้เราทำการยกกำลังสองของค่าแตกต่าง $d(t_i)$ เพื่อกำจัดค่าที่อาจมีเครื่องหมายเป็นลบ ดังสมการ (4.5) จะให้ความหมายของค่าความผิดพลาดทั้งหมด สมการ (4.5) สามารถเขียนได้ว่า

$$E = \sum_{i=1}^{n} \left[f0_i - f0_{สามัญ}(t_i) \right]^2 \tag{4.6}$$

แทนสมการ (4.4) ที่ $t=t_i$ ลงในสมการ (4.6) จะได้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$E = \sum_{i=1}^{n} \left[f0_i - (f0_{\text{ฐาน}} + f0_{\text{ความชัน}}\, t_i \right]^2 \tag{4.7}$$

จากสมการ (4.7) นี้ เราสามารถหาตัวไม่รู้ค่า $f0_{\text{ฐาน}}$ และ $f0_{\text{ความชัน}}$ ที่ต้องการได้โดยวิธีกำลังน้อยสุด (least-squares) ซึ่งทำจากวิธีการหาค่าต่ำสุด (minimization) ของค่าความผิดพลาดโดยเกี่ยวข้องกับตัวไม่รู้ค่า นั้นคือ

$$\frac{\partial E}{\partial f0_{\text{ฐาน}}} = 0 \tag{4.8}$$

และ

$$\frac{\partial E}{\partial f0_{\text{ความชัน}}} = 0 \tag{4.9}$$

และเงื่อนไขในสมการ (4.8) ให้ผลดังนี้

$$2\sum_{i=1}^{n} \left[f0_i - (f0_{\text{ฐาน}} + f0_{\text{ความชัน}}\, t_i \right](-1) = 0$$

$$\sum_{i=1}^{n} f0_i - \sum_{i=1}^{n} f0_{\text{ฐาน}} - \sum_{i=1}^{n} f0_{\text{ความชัน}}\, t_i = 0$$

$$nf0_{\text{ฐาน}} + (\sum_{i=1}^{n} t_i) f0_{\text{ความชัน}} = \sum_{i=1}^{n} f0_i \tag{4.10}$$

และเงื่อนไขในสมการ (4.9) ให้ผลดังนี้

$$2\sum_{i=1}^{n} \left[f0_i - (f0_{\text{ฐาน}} + f0_{\text{ความชัน}}\, t_i \right](-t_i) = 0$$

$$\sum_{i=1}^{n} f0_i\, t_i - \sum_{i=1}^{n} f0_{\text{ฐาน}}\, t_i - \sum_{i=1}^{n} f0_{\text{ความชัน}}\, t_i^2 = 0$$

$$(\sum_{i=1}^{n} t_i) f0_{\text{ฐาน}} + (\sum_{i=1}^{n} t_i^2) f0_{\text{ความชัน}} = \sum_{i=1}^{n} f0_i t_i \tag{4.11}$$

ทั้งสองสมการ (4.10) และ (4.11) นี้สามารถเขียนให้อยู่ในรูปของเมตริกซ์ได้ดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$\begin{bmatrix} n & \sum_{i=1}^{n} t_i \\ \sum_{i=1}^{n} t_i & \sum_{i=1}^{n} t_i^2 \end{bmatrix} \begin{bmatrix} f0_{\text{ฐาน}} \\ f0_{\text{ความชัน}} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} f0_i \\ \sum_{i=1}^{n} f0_i t_i \end{bmatrix} \quad (4.12)$$

ซึ่งสามารถใช้กฎของคราเมอร์ ในการแก้สมการนี้เพื่อหาค่าคงที่ $f0_{\text{ฐาน}}$ และ $f0_{\text{ความชัน}}$ ได้ดังนี้

$$f0_{\text{ฐาน}} = \frac{(\sum_{i=1}^{n} f0_i)(\sum_{i=1}^{n} t_i^2) - (\sum_{i=1}^{n} f0_i t_i)(\sum_{i=1}^{n} t_i)}{n(\sum_{i=1}^{n} t_i^2) - (\sum_{i=1}^{n} t_i)^2} \quad (4.13)$$

$$f0_{\text{ความชัน}} = \frac{n(\sum_{i=1}^{n} f0_i t_i) - (\sum_{i=1}^{n} t_i)(\sum_{i=1}^{n} f0_i)}{n(\sum_{i=1}^{n} t_i^2) - (\sum_{i=1}^{n} t_i)^2} \quad (4.14)$$

ทำการคำนวณพจน์ต่างๆของสมการ (4.10) และสมการ (4.11) จากชุดข้อมูลการผันวรรณยุกต์สามัญ ตามตารางที่ 4.3 ผลการคำนวณแสดงในตารางที่ 4.4

ตารางที่ 4.4 การคำนวณพจน์ต่างๆของสมการจากชุดข้อมูลการผันวรรณยุกต์สามัญ

N	$\sum_{i=1}^{n} f0_i$	$\sum_{i=1}^{n} t_i^2$	$\sum_{i=1}^{n} f0_i t_i$	$\sum_{i=1}^{n} t_i$	$(\sum_{i=1}^{n} t_i)^2$
54	7,977	185,827	398,994	2,744	7,531,073

แทนที่สมการ (4.13) และสมการ (4.14) ด้วยข้อมูลตามตารางที่ 4.4

$$f0_{\text{ฐาน}} = \frac{(7,977)(185,827) - (398,994)(2,744)}{54(7,531,073) - (7,531,073)}$$

$$f0_{\text{ฐาน}} = 154.7196 \quad (4.15)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$f0_{\text{ความชัน}} = \frac{54(398,994) - (2,774)(7,977)}{54(7,531,073) - (7,531,073)}$$

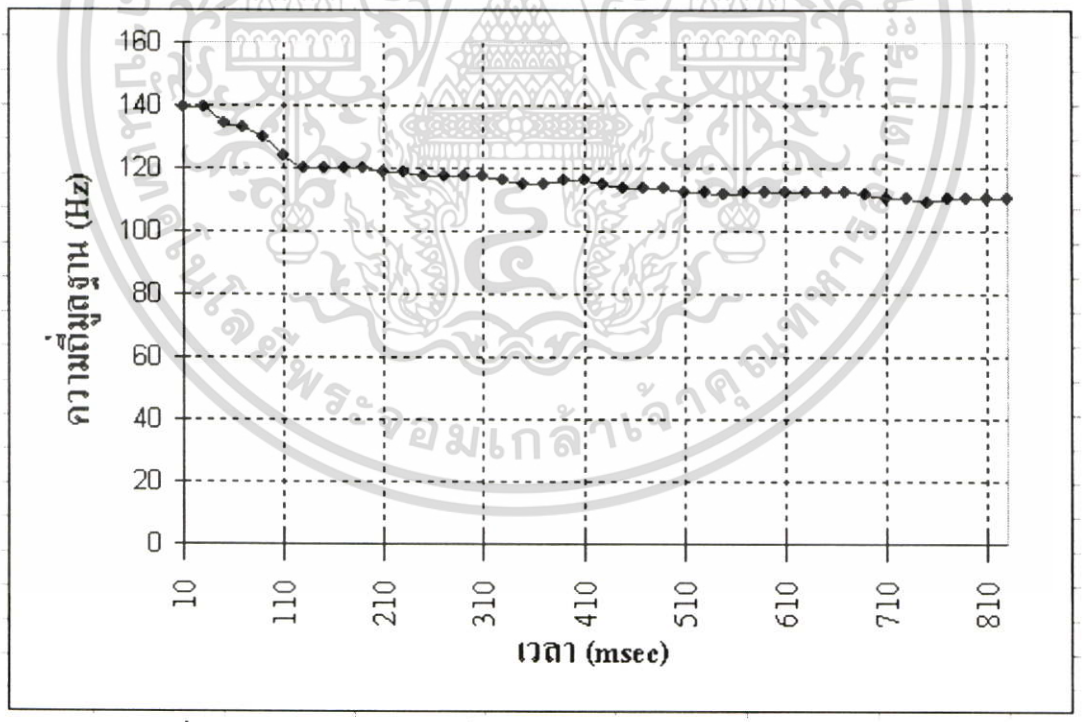
$$f0_{\text{ความชัน}} = -0.13776 \qquad (4.16)$$

ค่าคงที่ $f0_{\text{ฐาน}}$ และ $f0_{\text{ความชัน}}$ ที่คำนวณได้นี้ เมื่อแทนกลับลงไปในสมการ (4.4) จะได้สมการเส้นตรงของความถี่มูลฐานการผันวรรณยุกต์สามัญ ดังสมการ (4.17)

$$f0_{\text{สามัญ}}(t) = 154 - 0.13t \qquad (4.17)$$

4.2.3 วิเคราะห์การผันวรรณยุกต์เอก

ทำการวิเคราะห์การผันวรรณยุกต์เอก จากผลการทดลอง เมื่อพิจารณาแผนภูมิเส้นความถี่มูลฐานบนแกนเวลา พบว่ามีลักษณะเป็นสมการเชิงเส้นแบ่งเป็นสองส่วนที่ต่อเนื่องกัน แต่มีความชันไม่เท่ากัน ดังรูป 4.13



รูปที่ 4.13 แผนภูมิเส้นความถี่มูลฐานการผันวรรณยุกต์เอกบนแกนเวลา

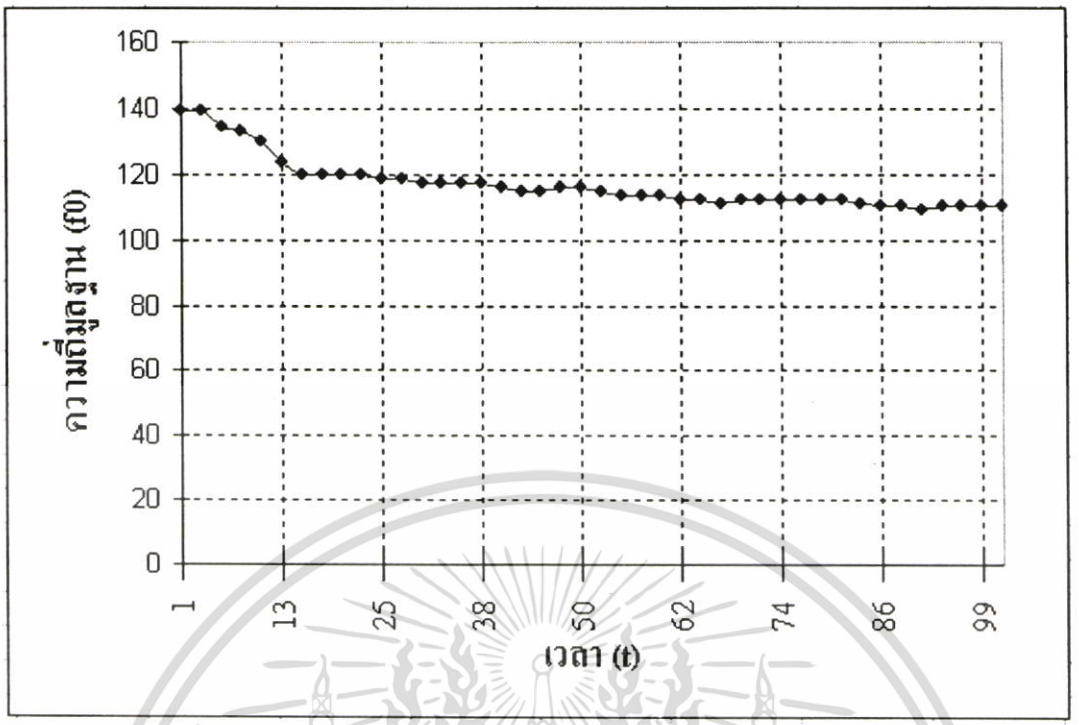
ดังนั้นการหาสมการของการผันวรรณยุกต์จึงใช้วิธีเช่นเดียวกับการหาสมการของการผันวรรณยุกต์สามัญ แตกต่างกันเพียงแต่มีสมการเส้นตรงสองสมการ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.5 ชุดข้อมูลความถี่มูลฐานการผันวรรณยุกต์เอก

ลำดับ (i)	จัดกรอบ (t)	ความถี่มูลฐาน (f0)	ลำดับ (i)	จัดกรอบ (t)	ความถี่มูลฐาน (f0)
1	1	139.56	22	52	114.84
2	3	137.005	23	55	113.66
3	6	134.45	24	57	113.66
4	8	132.83	25	60	113.66
5	11	129.71	26	62	112.5
6	13	123.88	27	64	112.5
7	16	119.84	28	67	112.5
8	18	119.84	29	69	112.5
9	21	119.84	30	72	112.5
10	23	119.84	31	74	112.5
11	25	118.55	32	77	112.5
12	28	118.55	33	79	112.5
13	30	117.29	34	81	112.5
14	33	117.29	35	84	111.36
15	35	117.29	36	86	110.25
16	38	117.29	37	89	110.25
17	40	116.05	38	91	110.25
18	42	116.05	39	94	110.25
19	45	116.05	40	96	110.25
20	47	114.84	41	99	110.25
21	50	114.84	42	101	110.25

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.14 แผนภูมิเส้นแสดงการจัดกรอบเวลาใหม่ของเสียงวรรณยุกต์เอก

จากตาราง 4.5 และรูปที่ 4.14 ซึ่งพบว่า ณ เวลา t ใดๆ ที่ทำให้ความชันของเส้นตรงเปลี่ยนไป ซึ่งสามารถประดิษฐ์สมการเส้นตรงได้ดังนี้

$$F0_{เอก}(t) = \begin{vmatrix} F0_{ฐาน} + F0_{ความชัน1}\, t & 0 < t <= A \\ F0_{เอก}(A) + F0_{ความชัน2}\,(t - A) & A < t <= 100 \end{vmatrix} \quad (4.18)$$

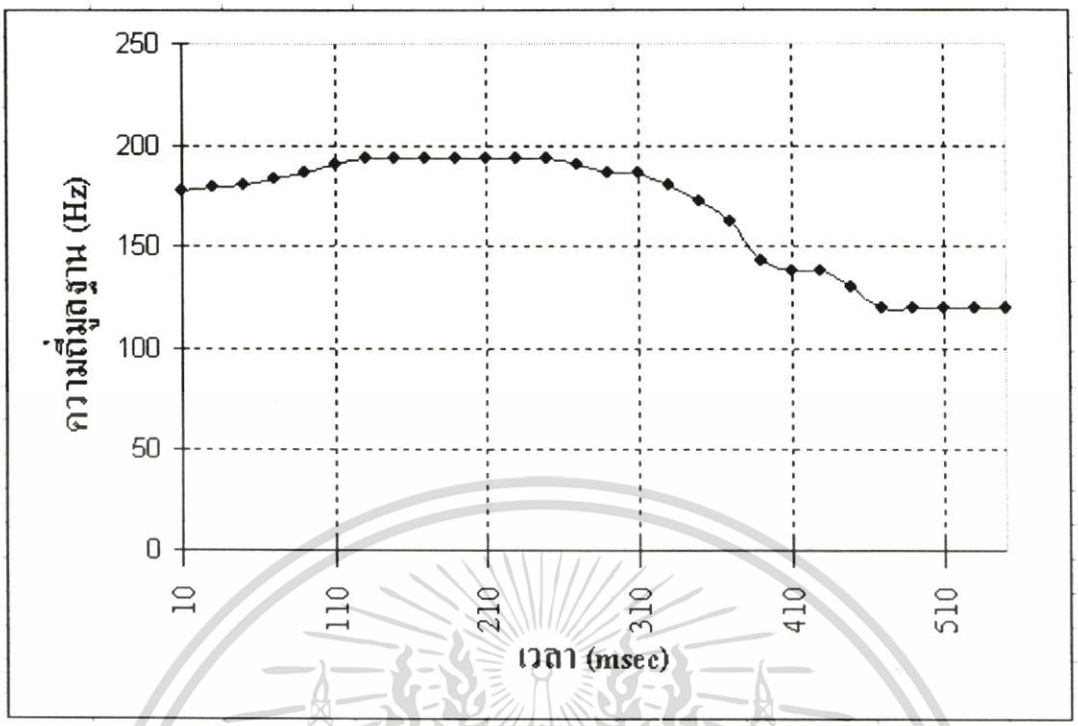
เมื่อทำการคำนวณแล้วจะได้สมการการผันวรรณยุกต์ของวรรณยุกต์เอก ดังสมการ (4.19)

$$F0_{เอก}(t) = \begin{vmatrix} 140 - 1.25t & 0 < t <= 15 \\ F0_{เอก}(15) - 0.12(t - 15) & 15 < t <= 100 \end{vmatrix} \quad (4.19)$$

4.2.4 วิเคราะห์การผันวรรณยุกต์โท

ทำการวิเคราะห์การผันวรรณยุกต์โท จากผลการทดลอง เมื่อพิจารณาแผนภูมิเส้นความถี่มูลฐานบนแกนเวลา พบว่ามีลักษณะเป็นสมการควอดราติคแบ่งเป็นสองส่วนที่ต่อเนื่องกัน ดังรูปที่ 4.15

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.15 แผนภูมิเส้นความถี่มูลฐานการผันวรรณยุกต์โทบนแกนเวลา

วิทยานิพนธ์นี้เลือกใช้วิธีการหาสมการควอดราติคด้วยวิธีการแบบถดถอยแบบพหุนาม (polynomial regression) ในที่นี้เราจะประดิษฐ์ฟังก์ชันพหุนามอันดับ 2 สำหรับข้อมูลชุดนี้

$$f0_{\text{ไท}}(t) = f0_{\text{ฐาน}} + f0_1 t + f0_2 t^2 \tag{4.20}$$

โดยที่ $f0_{\text{ฐาน}}$, $f0_1$ และ $f0_2$ เป็นค่าคงที่ ซึ่งจะคำนวณหาจากเงื่อนไขสมการพหุนามที่ประดิษฐ์ขึ้นมานี้จะก่อให้เกิดค่าความผิดพลาดโดยเฉลี่ยที่น้อยที่สุดจากข้อมูลที่ได้จากการทดลอง ขั้นตอนการประดิษฐ์สมการพหุนาม เริ่มจากการหาค่าความผิดพลาด E ทั้งหมดที่เกิดขึ้นจาก n ข้อมูล ในรูปแบบดังสมการ (4.17) ซึ่งสามารถเขียนให้ประกอบด้วยฟังก์ชันพหุนามดังแสดงในสมการ (4.20) ได้ดังนี้

$$E = \sum_{i=1}^{n} \left[f0_i - f0_{\text{ไท}}(t_i) \right]^2$$

$$E = \sum_{i=1}^{n} \left[f0_i - (f0_{\text{ฐาน}} + f0_1 t_i + f0_2 (t_i)^2 \right]^2 \tag{4.21}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ในการหาตัวไม่รู้ค่า $f0_{\text{ฐาน}}$, $f0_1$ และ $f0_2$ ซึ่งใช้วิธีกำลังสองน้อยสุด (least-squares) ซึ่งทำจากวิธีการหาค่าต่ำสุด (minimization) ของค่าความผิดพลาด E โดยเกี่ยวข้องกับตัวไม่รู้ค่า ก่อให้เกิดระบบสมการที่ประกอบด้วย 3 สมการย่อย คือ

$$\left.\begin{aligned}\frac{\partial E}{\partial f0_{\text{ฐาน}}} &= 0\\ \frac{\partial E}{\partial f0_1} &= 0\\ \frac{\partial E}{\partial f0_2} &= 0\end{aligned}\right\} \text{3 สมการ} \qquad (4.22)$$

สมการแรกสามารถประดิษฐ์ขึ้นได้ดังนี้

$$2\sum_{i=1}^{n}\left[f0_i-(f0_{\text{ฐาน}}+f0_1 t_i+f0_2(t_i)^2\right](-1)=0$$

$$-\sum_{i=1}^{n}f0_i-\sum_{i=1}^{n}f0_{\text{ฐาน}}-\sum_{i=1}^{n}f0_1 t_i-\sum_{i=1}^{n}f0_2(t_i)^2=0$$

$$n(f0_i)+(\sum_{i=1}^{n}t_i)f0_1+(\sum_{i=1}^{n}(t_i)^2)f0_2=\sum_{i=1}^{n}f0_i \qquad (4.23)$$

และเช่นเดียวกับสมการที่ 2

$$2\sum_{i=1}^{n}\left[f0_i-(f0_{\text{ฐาน}}+f0_1 t_i+f0_2(t_i)^2\right](-t_i)=0$$

$$-\sum_{i=1}^{n}f0_i t_i-\sum_{i=1}^{n}f0_{\text{ฐาน}}t_i-\sum_{i=1}^{n}f0_1(t_i)^2-\sum_{i=1}^{n}f0_2(t_i)^3=0$$

$$(\sum_{i=1}^{n}t_i)f0_{\text{ฐาน}}+(\sum_{i=1}^{n}(t_i)^2)f0_1+(\sum_{i=1}^{n}(t_i)^3)f0_2=\sum_{i=1}^{n}f0_i t_i \qquad (4.24)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

และเช่นเดียวกับสมการที่ 3

$$2\sum_{i=1}^{n}\left[f0_i - (f0_{\text{ฐาน}} + f0_1 t_i + f0_2 (t_i)^2\right](-t_i)^2 = 0$$

$$-\sum_{i=1}^{n} f0_i (t_i)^2 - \sum_{i=1}^{n} f0_{\text{ฐาน}} (t_i)^2 - \sum_{i=1}^{n} f0_1 (t_i)^3 - \sum_{i=1}^{n} f0_2 (t_i)^4 = 0$$

$$(\sum_{i=1}^{n}(t_i)^2) f0_{\text{ฐาน}} + (\sum_{i=1}^{n}(t_i)^3) f0_1 + (\sum_{i=1}^{n}(t_i)^4) f0_2 = \sum_{i=1}^{n} f0_i (t_i)^2 \qquad (4.25)$$

ทั้งสามสมการที่ประดิษฐ์ขึ้นมาได้นี้ สามารถเขียนให้อยู่ในรูปแบบสมการที่ประกอบ 3 สมการย่อยได้ดังนี้

$$\begin{bmatrix} n & \sum_{i=1}^{n} t_i & \sum_{i=1}^{n} t_i^2 \\ \sum_{i=1}^{n} t_i & \sum_{i=1}^{n} t_i^2 & \sum_{i=1}^{n} t_i^3 \\ \sum_{i=1}^{n} t_i^2 & \sum_{i=1}^{n} t_i^3 & \sum_{i=1}^{n} t_i^4 \end{bmatrix} \begin{bmatrix} f0_{\text{ฐาน}} \\ f0_1 \\ f0_2 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} f0_i \\ \sum_{i=1}^{n} f0_i t_i \\ \sum_{i=1}^{n} f0_i t_i^2 \end{bmatrix} \qquad (4.26)$$

เพื่อให้ง่ายต่อการเขียนสมการและง่ายต่อความเข้าใจ กำหนดสมการให้อยู่ในรูปของตัวแปร โดยกำหนดให้

$$A = n, B = \sum_{i=1}^{n} t_i, C = \sum_{i=1}^{n} t_i^2, D = \sum_{i=1}^{n} t_i^3, E = \sum_{i=1}^{n} t_i^4, F = \sum_{i=1}^{n} f0_i, I = \sum_{i=1}^{n} f0_i t_i$$

และ $J = \sum_{i=1}^{n} f0_i t_i^2$

เมื่อแทนที่ตัวแปรเหล่านี้ในสมการ (4.26) จะได้

$$\begin{bmatrix} A & B & C \\ B & C & D \\ C & D & E \end{bmatrix} \begin{bmatrix} f0_{\text{ฐาน}} \\ f0_1 \\ f0_2 \end{bmatrix} = \begin{bmatrix} F \\ I \\ J \end{bmatrix} \qquad (4.27)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ซึ่งสามารถใช้กฎของคราเมอร์ ในการแก้สมการนี้เพื่อหาค่าคงที่ $f0_{\text{ฐาน}}$, $f0_1$ และ $f0_2$ ได้ดังนี้

$$f0_{\text{ฐาน}} = \frac{C(EF + DI - CJ) + B(DJ - IE) - D^2F}{C(AE + 2(BD) - C^2) - B^2E - AD^2} \tag{4.28}$$

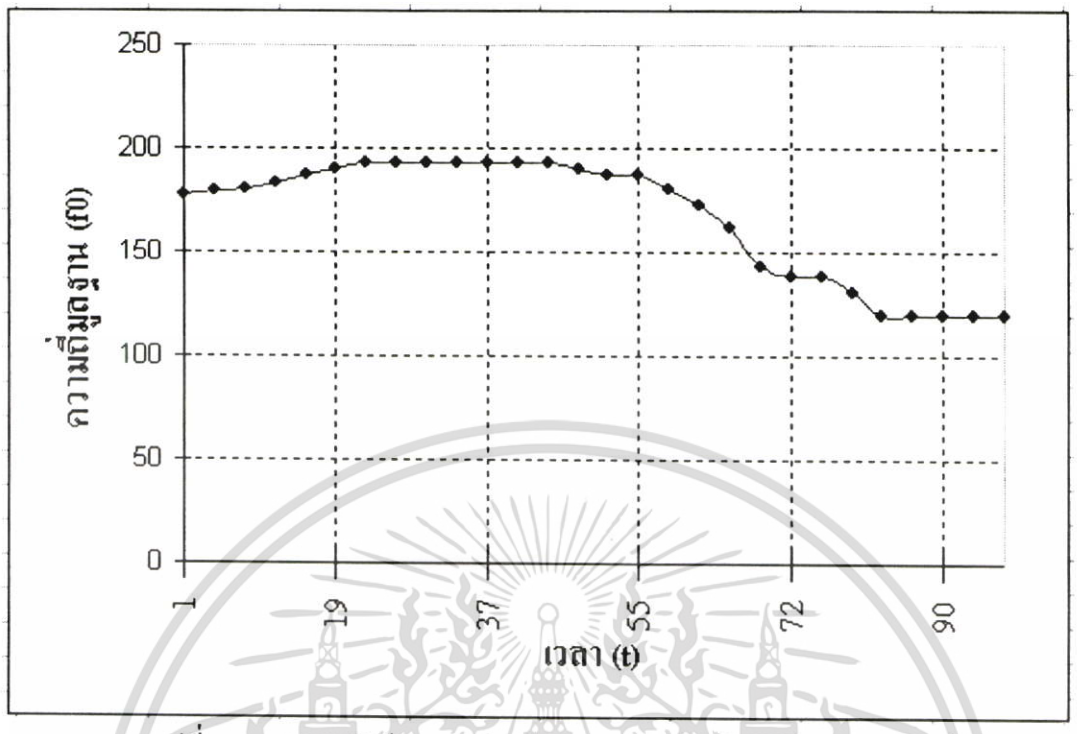
$$f0_1 = \frac{C(BJ + DF - CI) + A(IE - DJ) + BEF}{C(AE + 2(BD) - C^2) - B^2E - AD^2} \tag{4.29}$$

$$f0_2 = \frac{C(AJ + BI - CF) + D(BF - AI) + B^2J}{C(AE + 2(BD) - C^2) - B^2E - AD^2} \tag{4.30}$$

ตารางที่ 4.6 ชุดข้อมูลความถี่มูลฐานการผันวรรณยุกต์ไท

ลำดับ (i)	จัดกรอบ (t)	ความถี่มูลฐาน (f0)	ลำดับ (i)	จัดกรอบ (t)	ความถี่มูลฐาน (f0)
1	1	177.82	15	51	186.86
2	5	179.28	16	55	186.86
3	8	180.74	17	58	180.74
4	12	183.75	18	62	172.27
5	15	186.86	19	65	162.13
6	19	190.09	20	69	143.18
7	22	193.42	21	72	137.81
8	26	193.42	22	76	137.81
9	30	193.42	23	80	130
10	33	193.42	24	83	120
11	37	193.42	25	87	119.84
12	40	193.42	26	90	119.84
13	44	193.42	27	94	119.84
14	47	190.09	28	97	119.84

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.16 แผนภูมิเส้นแสดงการจัดกรอบเวลาใหม่ของเสียงวรรณยุกต์โท

จากตารางที่ 4.6 และรูปที่ 4.16 ซึ่งพบว่าแผนภูมิเส้นความถี่มูลฐานบนแกนเวลา มีลักษณะเป็น พหุนาม 2 สมการ ซึ่งจะเชื่อมต่อกันที่ t = A ซึ่งสามารถประดิษฐ์สมการควอดราติคได้ดังนี้

$$F0_{\text{โท}}(t) = \left| \begin{array}{ll} F0_{\text{มูลฐาน}} + F0_1 t + F0_2 t^2 & 0 < t <= A \\ F0_{\text{โท}}(A) + F0_3 (t - A) + F0_4 (t - A)^2 & A < t <= 100 \end{array} \right| \quad (4.31)$$

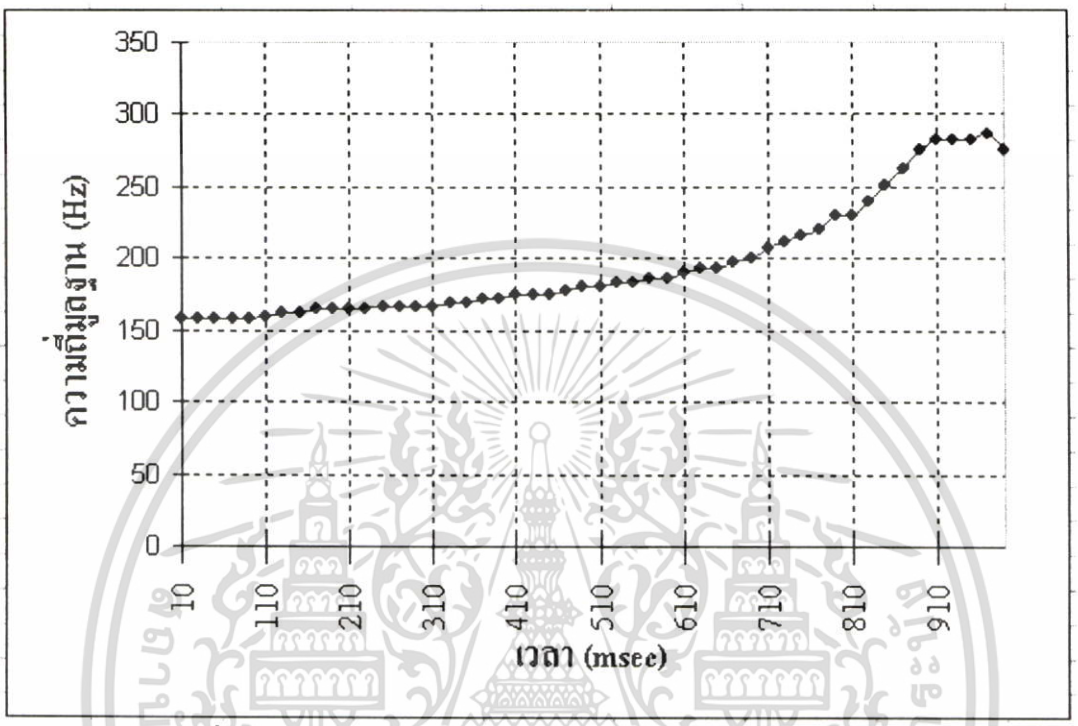
เมื่อทำการคำนวณแล้วจะได้สมการการผันวรรณยุกต์ของวรรณยุกต์โท ดังสมการ (4.32)

$$F0_{\text{โท}}(t) = \left| \begin{array}{ll} 178 + 1.18t - 0.02^2 & 0 < t <= 56 \\ F0_{\text{โท}}(56) - 3.5(t - 56) - 0.05(t - 56)^2 & 56 < t <= 100 \end{array} \right| \quad (4.32)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2.5 วิเคราะห์การผันวรรณยุกต์ตรี

ทำการวิเคราะห์การผันวรรณยุกต์ตรี จากผลการทดลอง เมื่อพิจารณาแผนภูมิเส้นความถี่มูลฐานบนแกนเวลา พบว่ามีลักษณะเป็นสมการควอดราติคแบ่งเป็นสองส่วนที่ต่อเนื่องกัน ดังรูป 4.17



รูปที่ 4.17 แผนภูมิเส้นความถี่มูลฐานการผันวรรณยุกต์ตรีบนแกนเวลา

วิทยานิพนธ์นี้เลือกใช้วิธีการหาสมการควอดราติคด้วยวิธีการแบบถดถอยแบบพหุนาม (polynomial regression) ในที่นี้เราจะประดิษฐ์ฟังก์ชันพหุนามอันดับ 2 สำหรับข้อมูลชุดนี้

$$F0_{\text{ตรี}}(t) = F0_{\text{ฐาน}} + F0_1 t + F0_2 t^2 \qquad (4.33)$$

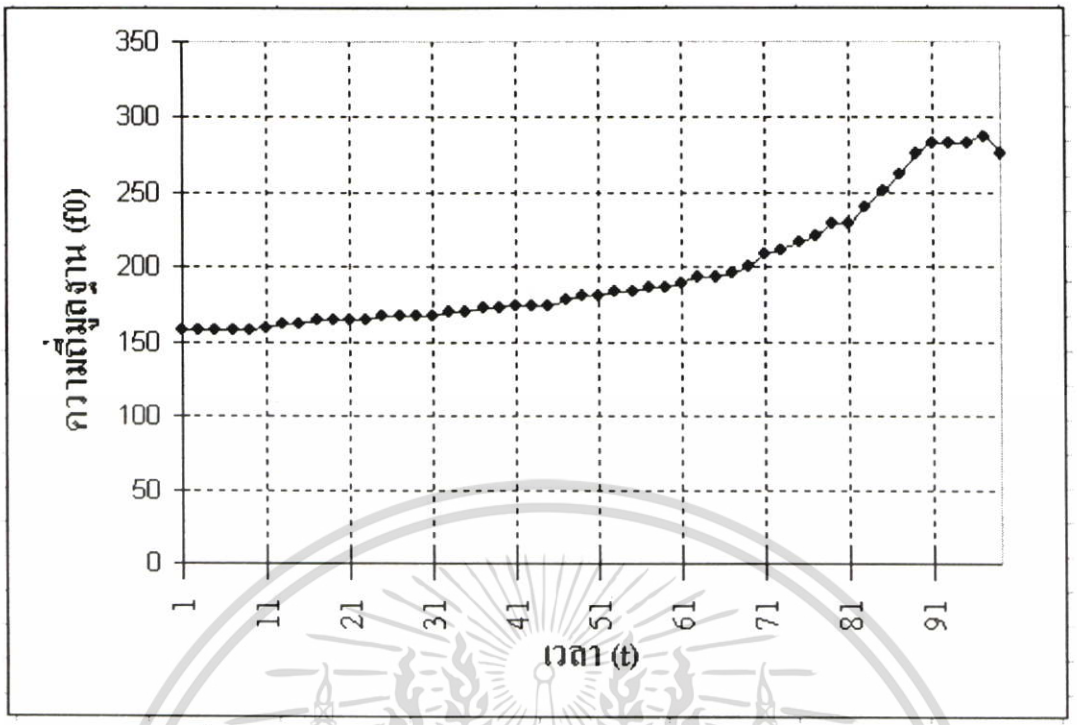
โดยที่ $F0_{\text{ฐาน}}$, $F0_1$ และ $F0_2$ เป็นค่าคงที่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.7 ชุดข้อมูลความถี่มูลฐานการผันวรรณยุกต์ตรี

ลำดับ (i)	จัดกรอบ (t)	ความถี่มูลฐาน (f0)	ลำดับ (i)	จัดกรอบ (t)	ความถี่มูลฐาน (f0)
1	1	157.5	26	51	180.74
2	3	157.5	27	53	183.75
3	5	157.5	28	55	183.75
4	7	157.5	29	57	186.86
5	9	157.5	30	59	186.86
6	11	159.78	31	61	190.09
7	13	162.13	32	63	193.42
8	15	162.13	33	65	193.42
9	17	164.55	34	67	196.88
10	19	164.55	35	69	200.45
11	21	164.55	36	71	208.02
12	23	164.55	37	73	212.02
13	25	167.05	38	75	216.18
14	27	167.05	39	77	220.5
15	29	167.05	40	79	229.69
16	31	167.05	41	81	229.69
17	33	169.62	42	83	239.67
18	35	169.62	43	85	250.57
19	37	172.27	44	87	262.5
20	39	172.27	45	89	275.63
21	41	175	46	91	282.69
22	43	175	47	93	282.69
23	45	175	48	95	282.69
24	47	177.82	49	97	286.8
25	49	180.74	50	99	275.63

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.18 แผนภูมิเส้นแสดงการจัดกรอบเวลาใหม่ของเสียงวรรณยุกต์ตรี

จากตารางที่ 4.7 และรูปที่ 4.18 ซึ่งพบว่าแผนภูมิเส้นความถี่มูลฐานบนแกนเวลา มีลักษณะเป็น พหุนาม 2 สมการ ซึ่งจะเชื่อมต่อกันที่ t = A ซึ่งสามารถประดิษฐ์สมการควอดราติคได้ดังนี้

$$F0_{\text{ตรี}}(t) = \begin{cases} F0_{\text{ฐาน}} + F0_1 t + F0_2 t^2 & 0 < t <= A \\ f0_{\text{ตรี}}(A) + F0_3 (t - A) + f0_4 (t - A)^2 & A < t <= 100 \end{cases} \quad (4.34)$$

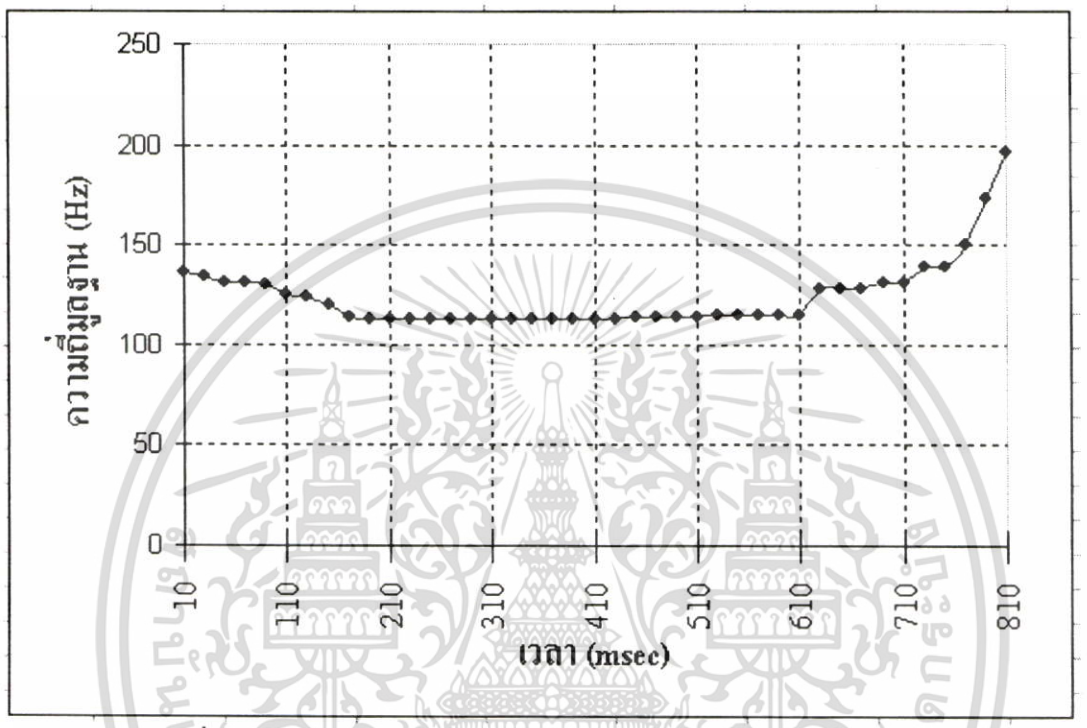
เมื่อทำการคำนวณแล้วจะได้สมการการผันวรรณยุกต์ของวรรณยุกต์ตรี ดังสมการ (4.35)

$$F0_{\text{ตรี}}(t) = \begin{cases} 158 + 0.001t + 0.009t^2 & 0 < t <= 70 \\ F0_{\text{ตรี}}(70) - 2(t - 70) + 0.1(t - 70)^2 & 70 < t <= 100 \end{cases} \quad (4.35)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2.6 วิเคราะห์การผันวรรณยุกต์จัตวา

ทำการวิเคราะห์การผันวรรณยุกต์จัตวา จากผลการทดลอง เมื่อพิจารณาแผนภูมิเส้นความถี่มูลฐานบนแกนเวลา พบว่ามีลักษณะเป็นสมการควอดราติคแบ่งเป็นสองส่วนที่ต่อเนื่องกัน ดังรูป 4.19



รูปที่ 4.19 แผนภูมิเส้นความถี่มูลฐานการผันวรรณยุกต์จัตวาบนแกนเวลา

วิทยานิพนธ์นี้เลือกใช้วิธีการหาสมการการผันวรรณยุกต์จัตวา ด้วยวิธีการแบบถดถอยแบบพหุนาม (polynomial regression) ในที่นี้เราจะประดิษฐ์ฟังก์ชันพหุนามอันดับ 2 สำหรับข้อมูลชุดนี้

$$F0_{\text{จัตวา}}(t) = F0_{\text{ฐาน}} + F0_1 t + F0_2 t^2 \qquad (4.36)$$

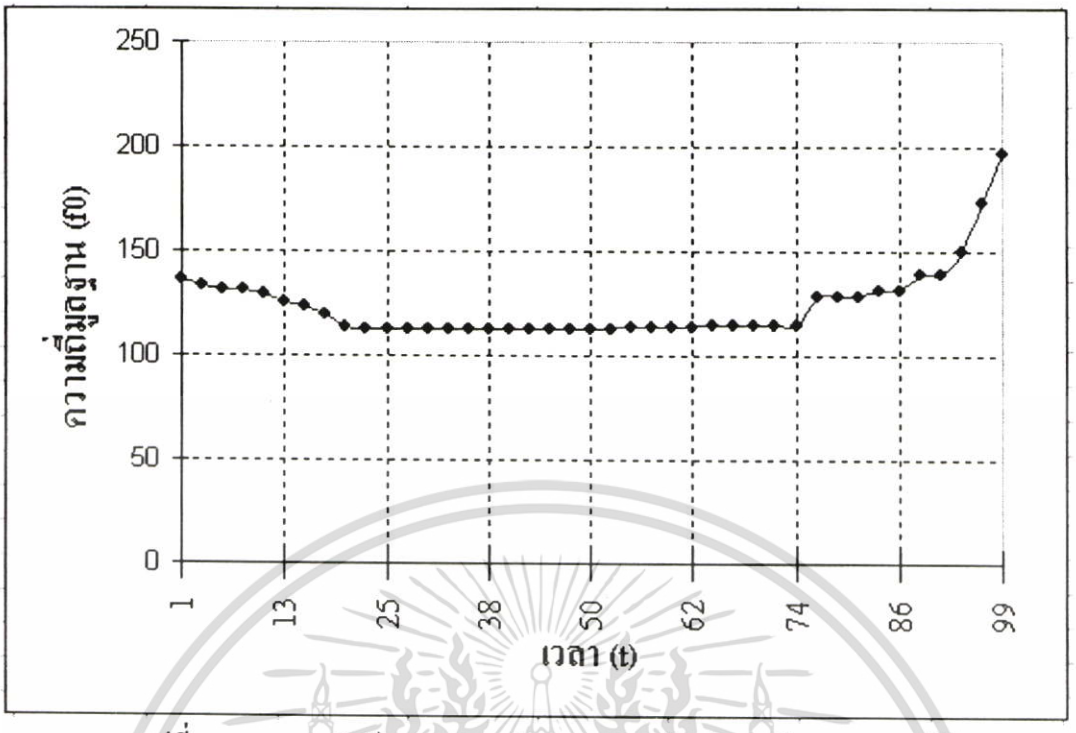
โดยที่ $F0_{\text{ฐาน}}$, $F0_1$ และ$F0_2$ เป็นค่าคงที่

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 4.8 ชุดข้อมูลความถี่มูลฐานการผันวรรณยุกต์จัตวา

ลำดับ (i)	จัดกรอบ (t)	ความถี่มูลฐาน (f0)	ลำดับ (i)	จัดกรอบ (t)	ความถี่มูลฐาน (f0)
1	1	136.11	22	52	112.5
2	3	133.68	23	55	113.66
3	6	131.25	24	57	113.66
4	8	131.25	25	60	113.66
5	11	129.71	26	62	113.66
6	13	125.28	27	64	114.84
7	16	123.88	28	67	114.84
8	18	119.84	29	69	114.84
9	21	113.66	30	72	114.84
10	23	112.5	31	74	114.84
11	25	112.5	32	77	128.2
12	28	112.5	33	79	128.2
13	30	112.5	34	81	128.2
14	33	112.5	35	84	131.25
15	35	112.5	36	86	131.25
16	38	112.5	37	89	139.56
17	40	112.5	38	91	139.56
18	42	112.5	39	94	150
19	45	112.5	40	96	173.44
20	47	112.5	41	99	196.88
21	50	112.5			

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 4.20 แผนภูมิเส้นแสดงการจัดกรอบเวลาใหม่ของเสียงวรรณยุกต์จัตวา

จากตารางที่ 4.8 และรูปที่ 4.20 ซึ่งพบว่าแผนภูมิเส้นความถี่มูลฐานบนแกนเวลา มีลักษณะเป็น พหุนาม 2 สมการ ซึ่งจะเชื่อมต่อกันที่ t = A ซึ่งสามารถประดิษฐ์สมการควอดราติคได้ดังนี้

$$F0_{\text{จัตวา}}(t) = \begin{cases} F0_{\text{ฐาน}} + F0_1 t + F0_2 t^2 & 0 < t <= A \\ F0_{\text{จัตวา}}(A) + F0_3 (t - A) + f0_4 (t - A)^2 & A < t <= 100 \end{cases} \quad (4.37)$$

เมื่อทำการคำนวณแล้วจะได้สมการการผันวรรณยุกต์ของวรรณยุกต์จัตวา ดังสมการ (4.38)

$$F0_{\text{จัตวา}}(t) = \begin{cases} 136 - t + 0.01t^2 & 0 < t <= 78 \\ F0_{\text{จัตวา}}(78) + 0.16(t - 78)^2 & 78 < t <= 100 \end{cases} \quad (4.38)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ความถี่มูลฐานของการผันวรรณยุกต์ของแต่ละเพศมีค่าไม่เท่ากัน การเปล่งเสียงวรรณยุกต์จะขึ้นอยู่กับทิศทางของความถี่มูลฐานบนแกนเวลา และระดับของความถี่มูลฐาน

ทิศทางของความถี่มูลฐานบนแกนเวลาซึ่งได้จากสมการที่ได้จากการวิเคราะห์คือ $F0_{ความชัน}$ สำหรับการผันวรรณยุกต์สามัญ และวรรณยุกต์เอก และ $F0_1$, $F0_2$ สำหรับสียงวรรณยุกต์โท, ตรี และจัตวา

ระดับของความถี่มูลฐาน ซึ่งก็คือค่าของ $F0_{ฐาน}$ ซึ่งมีผลทำให้เกิดระดับเสียงที่แตกต่างกันขึ้น และเนื่องจากค่า $F0_{ฐาน}$ ของแต่ละวรรณยุกต์มีค่าไม่เท่ากันดังนั้น จึงกำหนดให้

$$F0_{ฐาน} = F0_{base} + F0_{ฐานเสียงวรรณยุกต์} \tag{4.39}$$

เมื่อ $F0_{ฐาน}$ คือ ระดับของความถี่มูลฐานของแต่ละวรรณยุกต์

$F0_{base}$ คือ ค่าตัวแปรความถี่มูลฐานหลักที่เปลี่ยนค่าได้ตามพารามิเตอร์

$F0_{ฐานเสียงวรรณยุกต์}$ คือ ค่าต่ำสุดของระดับความถี่ของวรรณยุกต์-ระดับความถี่มูลฐานของวรรณยุกต์นั้น

ดังนั้นเมื่อแทนที่ $F0_{base}$ และ$F0_{ฐานเสียงวรรณยุกต์}$ โดยเปลี่ยน $F0_{ฐานเสียงวรรณยุกต์}$ เป็น $F0_{ฐานสามัญ}$, $F0_{ฐานเอก}$, $F0_{ฐานโท}$, $F0_{ฐานตรี}$, $F0_{ฐานจัตวา}$ ดังแสดงในสมการที่ (4.40), (4.42), (4.43), (4.44), (4.45) ตามลำดับ

$$F0_{สามัญ}(t) = F0_{base} + F0_{ฐานสามัญ} - 0.13t \tag{4.40}$$

$$F0_{เอก}(t) = \left| \begin{array}{ll} F0_{base} + F0_{ฐานเอก} - 1.25t & 0 < t <= 15 \\ F0_{เอก}(15) - 0.12(t-15) & 15 < t <= 100 \end{array} \right| \tag{4.41}$$

$$F0_{โท}(t) = \left| \begin{array}{ll} F0_{base} + F0_{ฐานโท} + 1.18t - 0.02^2 & 0 < t <= 56 \\ F0_{โท}(56) - 3.5(t-56) - 0.05(t-56)^2 & 56 < t <= 100 \end{array} \right| \tag{4.42}$$

$$F0_{ตรี}(t) = \left| \begin{array}{ll} F0_{base} + F0_{ฐานตรี} + 0.001t + 0.009t^2 & 0 < t <= 70 \\ F0_{ตรี}(70) - 2(t-70) + 0.1(t-70)^2 & 70 < t <= 100 \end{array} \right| \tag{4.43}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$F0_{\text{จัตวา}}(t) = \begin{cases} F0_{\text{base}} + F0_{\text{ฐานจัตวา}} - t + 0.01t^2 & 0 < t <= 78 \\ F0_{\text{จัตวา}}(78) + 0.16(t-78)^2 & 78 < t <= 100 \end{cases} \quad (4.44)$$

สามารถหาค่าของ $F0_{\text{ฐานสามัญ}}$, $F0_{\text{ฐานเอก}}$, $F0_{\text{ฐานโท}}$, $F0_{\text{ฐานตรี}}$, $F0_{\text{ฐานจัตวา}}$ ดังแสดงในตารางที่ 4.9

ตารางที่ 4.9 ค่าระดับความถี่มูลฐานของแต่ละวรรณยุกต์

วรรณยุกต์ / ชนิดของ F0	วรรณยุกต์สามัญ	วรรณยุกต์เอก	วรรณยุกต์โท	วรรณยุกต์ตรี	วรรณยุกต์จัตวา
$F0_{\text{ฐาน}}$	154	140	178	158	136
$F0_{\text{ฐานวรรณยุกต์}}$	18	4	42	22	0

เมื่อแทนค่า $F0_{\text{ฐานวรรณยุกต์}}$ ด้วย $F0_{\text{ฐานสามัญ}}$, $F0_{\text{ฐานเอก}}$, $F0_{\text{ฐานโท}}$, $F0_{\text{ฐานตรี}}$, $F0_{\text{ฐานจัตวา}}$ ลงในสมการที่ (4.40), (4.42), (4.43), (4.44), (4.45) จะได้สมการของการผันวรรณยุกต์ดังสมการที่ (4.45), (4.46), (4.47), (4.48), (4.49) ตามลำดับ

$$F0_{\text{สามัญ}}(t) = F0_{\text{base}} + 18 - 0.13t \quad (4.45)$$

$$F0_{\text{เอก}}(t) = \begin{cases} F0_{\text{base}} + 4 - 1.25t & 0 < t <= 15 \\ F0_{\text{เอก}}(15) - 0.12(t-15) & 15 < t <= 100 \end{cases} \quad (4.46)$$

$$F0_{\text{โท}}(t) = \begin{cases} F0_{\text{base}} + 42 + 1.18t - 0.02t^2 & 0 < t <= 56 \\ F0_{\text{โท}}(56) - 3.5(t-56) - 0.05(t-56)^2 & 56 < t <= 100 \end{cases} \quad (4.47)$$

$$F0_{\text{ตรี}}(t) = \begin{cases} F0_{\text{base}} + 22 + 0.001t + 0.009t^2 & 0 < t <= 70 \\ F0_{\text{ตรี}}(70) - 2(t-70) + 0.1(t-70)^2 & 70 < t <= 100 \end{cases} \quad (4.48)$$

$$F0_{\text{จัตวา}}(t) = \begin{cases} F0_{\text{base}} - t + 0.01t^2 & 0 < t <= 78 \\ F0_{\text{จัตวา}}(78) + 0.16(t-78)^2 & 78 < t <= 100 \end{cases} \quad (4.49)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.3 กฎการออกเสียงภาษาไทย

กฎการออกเสียงภาษาไทย เป็นการนำผลจากการวิเคราะห์คุณสมบัติที่เป็นลักษณะเด่นของการเกิดเสียงภาษาไทย ซึ่งประกอบด้วย ความถี่ฟอร์แมนท์ ความถี่มูลฐาน และเวลา กำหนดให้อยู่ในรูปแบบของกฎในระดับพยางค์ภาษาไทย โดยค่าเหล่านี้ได้มาจากการวิเคราะห์คุณสมบัติของเสียงภาษาไทย กฎการออกเสียงภาษาไทย ประกอบไปด้วย 1) กฎการออกเสียงตามเพศ 2) กฎการแบ่งวรรคตอน 3) กฎความเร็วของเสียง 4) กฎการผันวรรณยุกต์ รายละเอียดดังต่อไปนี้

4.3.1 กฎการออกเสียงตามเพศ

ลักษณะของเสียงพูดระหว่างผู้หญิงและผู้ชายที่แตกต่างกันคือ ความถี่ฟอร์แมนท์และความถี่มูลฐาน ซึ่งเสียงของผู้หญิงจะมีลักษณะเป็นเสียงแหลมและเล็กมากกว่าเสียงผู้ชาย กล่าวคือ ความถี่ฟอร์แมนท์และความถี่มูลฐานของผู้หญิงจะสูงกว่าของผู้ชาย ดังแสดงในตารางที่ 4.10 และตารางที่ 4.11

ตารางที่ 4.10 ความถี่ฟอร์แมนท์ของผู้หญิงและผู้ชาย

หน่วยเสียง	ผู้หญิง			ผู้ชาย		
	F1 (Hz)	F2 (Hz)	F3 (Hz)	F1 (Hz)	F2 (Hz)	F3 (Hz)
/อา/	958	1356	2769	807	1151	2403
/อี/	355	2119	3148	210	1860	2767
/อู/	331	825	2209	270	706	2002
/แอ/	771	1963	2985	558	1740	2528
/เอ/	529	1863	2796	428	1654	2473
/โอ/	482	731	2719	405	686	2563
/เออ/	471	1724	3212	328	1520	2797
/เออ/	743	1421	3410	593	1132	2873
/ออ/	694	895	2875	435	785	2456

จากตารางที่ 4.10 จะพบว่าความถี่ฟอร์แมนท์ของหญิงจะสูงกว่าผู้ชาย ความถี่ฟอร์แมนท์ที่ 1 ของผู้หญิงสูงกว่าของผู้ชายประมาณ 144 Hz ความถี่ฟอร์แมนท์ที่ 2 ของผู้หญิงสูงกว่าของผู้ชายประมาณ 184 Hz และความถี่ฟอร์แมนท์ที่ 3 ของผู้หญิงสูงกว่าผู้ชายประมาณ 362 Hz ซึ่งจะได้ผลต่างระหว่างความถี่ฟอร์แมนท์เฉลี่ยกับความถี่ฟอร์แมนท์ของผู้หญิงหรือผู้ชายคือ ความถี่ฟอร์แมนท์ที่ 1 มีค่าประมาณ 72 Hz ซึ่งจะกำหนดค่าของ $F1_{male}$ และ $F1_{female}$ ในสมการที่ (4.53) ซึ่ง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ค่าของ $F1_{male}$ จะมีค่าเป็นลบ ความถี่ฟอร์แมนท์ที่ 2 มีค่าประมาณ 92 Hz ซึ่งจะกำหนดค่าของ $F2_{male}$ และ $F2_{female}$ ในสมการที่ (4.54) ซึ่งค่าของ $F2_{male}$ จะมีค่าเป็นลบ และความถี่ฟอร์แมนท์ที่ 3 มีค่าประมาณ 181 Hz ซึ่งจะกำหนดค่าของ $F3_{male}$ และ $F3_{female}$ ในสมการที่ (4.55) ซึ่งค่าของ $F3_{male}$ จะมีค่าเป็นลบ

ตารางที่ 4.11 ความถี่มูลฐานของผู้หญิงและผู้ชาย

กลุ่มผู้บอกภาษา	หญิง			ชาย		
	F0 ต่ำสุด	F0 สูงสุด	F0 เฉลี่ย	F0 ต่ำสุด	F0 สูงสุด	F0 เฉลี่ย
วัยเด็ก	236	250	244	208	248	222
วัยรุ่น	180	217	202	85	130	107
ผู้ใหญ่	192	210	203	116	125	119

จากตารางที่ 4.11 จะพบว่าความมูลฐานของหญิงจะสูงกว่าผู้ชาย และเสียงเด็กผู้ชายจะมีความถี่มูลฐานค่อนข้างสูงมีค่าเฉลี่ยประมาณ 222 Hz ซึ่งมีค่าความถี่มูลฐานสูงกว่าความถี่มูลฐานของผู้หญิงวัยรุ่นและวัยผู้ใหญ่ ดังนั้นการออกเสียงผู้หญิงและผู้ชายจะใช้ความถี่มูลฐานของเสียงวัยรุ่นและผู้ใหญ่ ในการกำหนดค่า การเสียงตามเพศจะกำหนดค่าของ $F0_{base_male}$ มีค่าประมาณ 113 Hz และความถี่มูลฐาน $F0_{base_female}$ มีค่าประมาณ 202 Hz การกำหนดค่าดังแสดงในสมการที่ (4.50)

$$F0_{base} = \left| \begin{array}{ll} F0_{base_male} & \text{เมื่อ sex = male} \\ F0_{base_female} & \text{เมื่อ sex = female} \end{array} \right| \qquad (4.50)$$

เมื่อ $F0_{base}$ คือ ค่าตัวแปรความถี่มูลฐานหลัก

$F0_{base_male}$ คือ ค่าความถี่มูลฐานเฉลี่ยของเสียงผู้ชาย

$F0_{base_female}$ คือ ค่าความถี่มูลฐานเฉลี่ยของเสียงผู้หญิง

แล้วกำหนดค่าช่วงของความถี่มูลฐานของเสียงต่ำสุด ซึ่งกำหนดค่าไปที่ $F0_{min_male}$ ซึ่งมีค่าประมาณ 85 Hz และ $F0_{min_female}$ มีค่าประมาณ 180 Hz ดังสมการที่ 4.51

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$F0_{min} = \left| \begin{array}{ll} F0_{min_male} & \text{เมื่อ sex = male} \\ F0_{min_female} & \text{เมื่อ sex = female} \end{array} \right| \quad (4.51)$$

เมื่อ $F0_{min}$ คือ ตัวแปรค่าความถี่มูลฐานต่ำสุด

$F0_{min_male}$ คือ ค่าความถี่มูลฐานต่ำสุดของเสียงผู้ชาย

$F0_{min_female}$ คือ ค่าความถี่มูลฐานต่ำสุดของเสียงผู้หญิง

กำหนดค่าช่วงของความถี่มูลฐานของเสียงสูงสุด ซึ่งกำหนดค่าไปที่ $F0_{max_male}$ ซึ่งมีค่าประมาณ 130 Hz และ $F0_{max_female}$ มีค่าประมาณ 250 Hz ดังสมการที่ 4.52

$$F0_{max} = \left| \begin{array}{ll} F0_{max_male} & \text{เมื่อ sex = male} \\ F0_{max_female} & \text{เมื่อ sex = female} \end{array} \right| \quad (4.52)$$

เมื่อ $F0_{max}$ คือ ตัวแปรค่าความถี่มูลฐานสูงสุด

$F0_{max_male}$ คือ ค่าความถี่มูลฐานสูงสุดของเสียงผู้ชาย

$F0_{max_female}$ คือ ค่าความถี่มูลฐานสูงสุดของเสียงผู้หญิง

$$F1 = \left| \begin{array}{ll} F1 + F1_{male} & \text{เมื่อ sex = male} \\ F1 + F1_{female} & \text{เมื่อ sex = female} \end{array} \right| \quad (4.53)$$

เมื่อ F1 คือ ค่าความถี่ฟอร์แมนท์ที่ 1

$F1_{male}$ คือ ค่าคงตัวของความถี่ฟอร์แมนท์ที่ 1 ของเสียงผู้ชาย

$F1_{female}$ คือ ค่าคงตัวของความถี่ฟอร์แมนท์ที่ 1 ของเสียงผู้หญิง

$$F2 = \left| \begin{array}{ll} F2 + F2_{male} & \text{เมื่อ sex = male} \\ F2 + F2_{female} & \text{เมื่อ sex = female} \end{array} \right| \quad (4.54)$$

เมื่อ F2 คือ ค่าความถี่ฟอร์แมนท์ที่ 2

$F2_{male}$ คือ ค่าคงตัวของความถี่ฟอร์แมนท์ที่ 2 ของเสียงผู้ชาย

$F2_{female}$ คือ ค่าคงตัวของความถี่ฟอร์แมนท์ที่ 2 ของเสียงผู้หญิง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$F3 = \begin{cases} F3 + F3_{male} & \text{เมื่อ sex = male} \\ F3 + F3_{female} & \text{เมื่อ sex = female} \end{cases} \quad (4.55)$$

เมื่อ F3 คือ ค่าความถี่ฟอร์แมนท์ที่ 3

$F3_{male}$ คือ ค่าคงตัวของความถี่ฟอร์แมนท์ที่ 3 ของเสียงผู้ชาย

$F3_{female}$ คือ ค่าคงตัวของความถี่ฟอร์แมนท์ที่ 3 ของเสียงผู้หญิง

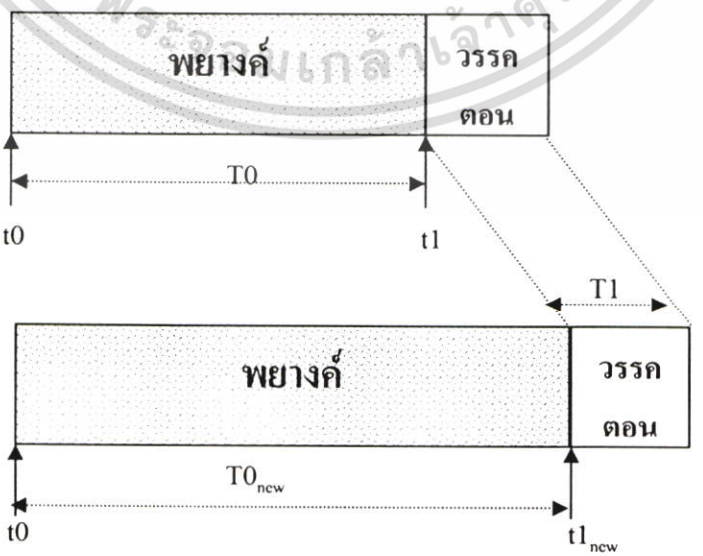
4.3.2 กฎความเร็วเสียง

ความเร็วเสียงเกิดจากระยะเวลาในการเปล่งเสียงในหนึ่งพยางค์ เมื่อมีการเพิ่มหรือลดระยะเวลาในการเปล่งเสียงในพยางค์ จะทำให้เกิดเสียงที่มีลักษณะการพูดช้าหรือพูดเร็ว ค่าระยะเวลาในการเปล่งเสียงในหนึ่งพยางค์มีหน่วยเป็นมิลลิวินาที

ตารางที่ 4.12 ค่าช่วงเวลาของเสียงสระ

ประเภทสระ	ช่วงเวลา (มิลลิวินาที)		
	สระเดี่ยว	สระประสม	สระเกิน
สระเสียงสั้น	80-150	90-250	-
สระเสียงยาว	270-800	380-690	200-400

ค่าของระยะเวลาที่เหมาะสมสำหรับการกำหนดค่า ซึ่งเป็นค่าเวลาของเสียงสระ ซึ่งเป็นแกนกลางของพยางค์ ดังแสดงในตารางที่ 4.12



ภาพที่ 4.21 การเพิ่มลดความเร็วเสียงของพยางค์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากภาพที่ 4.12 แสดงหลักการเพิ่มลดความเสียงของพยางค์ ซึ่งสามารถคำนวณช่วงเวลาใหม่ของพยางค์ได้จากสมการที่ (4.56) และสามารถคำนวณเวลา ณ ตำแหน่งใดๆ ในพยางค์ได้จากสมการที่ (4.57)

$$T0_{new} = \frac{(100 + Speed)xT0}{100} \tag{4.56}$$

$$t1_{new} = \frac{t1xT0_{new}}{T0} \tag{4.57}$$

เมื่อ $T0_{new}$ คือ ช่วงเวลาใหม่ของพยางค์

Speed คือ เปอร์เซ็นต์ความเร็วที่เพิ่มขึ้นหรือลดลง

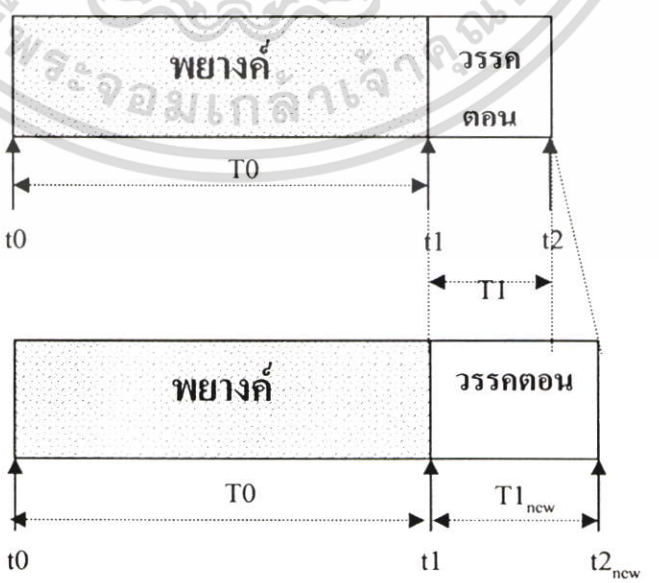
T0 คือ ช่วงเวลาเก่าของพยางค์

t1 คือ เวลา ณ ตำแหน่งใดๆ ในพยางค์เก่า

$t1_{new}$ คือ เวลา ณ ตำแหน่งใดๆ ในพยางค์ใหม่

4.3.3 กฎการแบ่งวรรคตอน

การแบ่งวรรคตอน เป็นการแบ่งช่วงในหนึ่งพยางค์ออกจากกัน โดยจะเพิ่มเสียงเงียบต่อท้ายที่พยางค์ตามระยะเวลาที่กำหนด เมื่อมีการเพิ่มหรือลดระยะเวลาในช่วงของเสียงเงียบที่ท้ายพยางค์จะทำให้เกิดการแบ่งวรรคตอนของพยางค์ ค่าระยะเวลาของเสียงเงียบมีหน่วยเป็นมิลลิวินาที



ภาพที่ 4.22 การแบ่งวรรคตอนของพยางค์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูปที่ 4.22 แสดงหลักการแบ่งวรรคตอนของพยางค์ ซึ่งสามารถคำนวณช่วงเวลาใหม่ของเสียงเงียบท้ายพยางค์ได้จากสมการที่ (4.58) และสามารถคำนวณเวลา ณ ตำแหน่งท้ายสุดได้จากสมการที่ (4.59)

$$T1_{new} = \frac{(100 + Space)xT1}{100} \tag{4.58}$$

$$t2_{new} = t1 + T1_{new} \tag{4.59}$$

เมื่อ $T1_{new}$ คือ ช่วงเวลาใหม่ของเสียงเงียบ

Space คือ เปอร์เซ็นต์ของช่วงเวลาที่เพิ่มขึ้นหรือลดลง

T1 คือ ช่วงเวลาเก่าของเสียงเงียบ

t1 คือ เวลา ณ ตำแหน่งท้ายพยางค์

$t2_{new}$ คือ เวลา ณ ตำแหน่งท้ายสุดของเสียงเงียบใหม่

4.3.4 กฎระดับเสียง

ระดับเสียงคือ เสียงสูงต่ำ ที่เราได้ยินในการเปล่งเสียพูดของมนุษย์ เกิดจากการสั่นของเส้นเสียง ซึ่งเป็นลักษณะเด่นของความถี่มูลฐาน เนื่องจากระดับเสียงของเพศหญิงและเพศชายจะมีระดับเสียงไม่เท่ากัน ดังนั้นระดับเสียงของแต่ละเพศจึงควรไม่ควรต่ำกว่า $F0_{min}$ และไม่สูงกว่า $F0_{max}$ การเพิ่มหรือลดระดับเสียง ดังแสดงในสมการที่ (4.60)

$$F0_{base} = \frac{(100 + Pitch)xF0_{base}}{100} \tag{4.60}$$

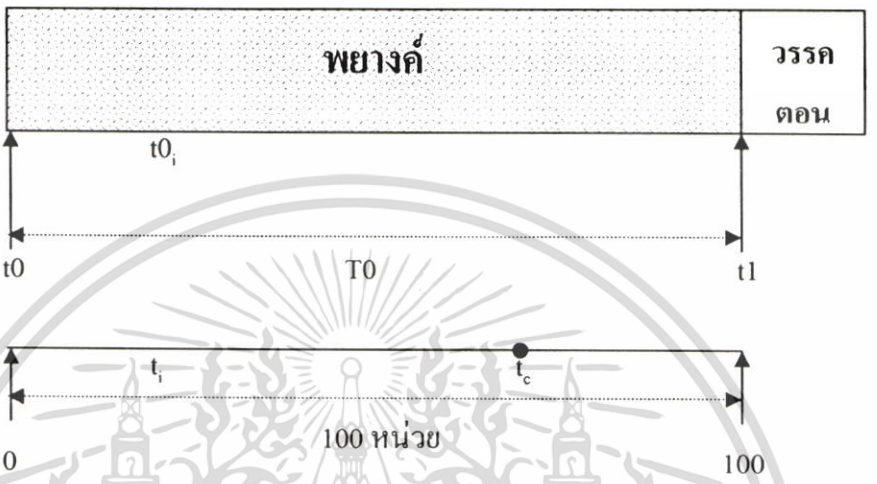
เมื่อ $F0_{base}$ คือ ค่าตัวแปรความถี่มูลฐานหลัก

Pitch คือ เปอร์เซ็นต์ของระดับเสียงที่เพิ่มขึ้นหรือลดลง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.3.5 กฎการผันวรรณยุกต์

กฎการผันวรรณยุกต์เป็นการคำนวณเกี่ยวกับค่าความถี่มูลฐาน ณ เวลาใดๆ ในพยางค์ โดยทำการคำนวณจากสมการการผันวรรณยุกต์ โดยจะทำการแทรกค่าความถี่มูลฐานทุกๆ วงรอบของเฟรม (FrameInterval) ที่มีลักษณะของความเป็นรายคาบ



ภาพที่ 4.23 การผันวรรณยุกต์ของพยางค์

เนื่องจากสมการการผันวรรณยุกต์เป็นสมการที่แบ่งเป็นสองส่วนที่ต่อเนื่องกัน ดังนั้นสามารถหาเวลาของจุดเปลี่ยนสมการดังแสดงในสมการที่ (4.61)

$$t0_c = \frac{t_c \times T0}{100} \tag{4.61}$$

เมื่อ $t0_c$ คือ เวลา ณ ตำแหน่งที่เปลี่ยนสมการมีหน่วยเป็นมิลลิวินาที

t_c คือ จุดเปลี่ยนสมการ

T0 คือ ช่วงเวลาของพยางค์

การแทรกเฟรมเข้าไปในแต่ละช่วงของพยางค์ สามารถคำนวณได้จากสมการที่ (4.62) โดยจะทำการหารแบบปัดเศษทิ้ง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

$$\text{Frame}_j = \frac{\text{T0}_j}{\text{FrameInterval}} \tag{4.62}$$

เมื่อ Frame_j คือ จำนวนเฟรมในช่วง j

T0_j คือ ช่วงเวลาในช่วง j

FrameInterval คือ วงรอบของเฟรมมีหน่วยเป็นมิลลิวินาที

วงรอบของเฟรมในวิทยานิพนธ์นี้ใช้ 20 มิลลิวินาที จากนั้นทำการคำนวณค่า F0 ของการผันวรรณยุกต์ โดยคำนวณได้จากสมการที่ (4.45-4.49) โดยแทนค่า t ด้วย t_i ซึ่งสามารถคำนวณได้จากสมการที่ (4.63)

$$t_i = \frac{t0_i \text{ x100}}{\text{T0}} \tag{4.63}$$

เมื่อ $t0_i$ คือ ค่าเวลาของแต่ละเฟรม

T0 คือ ช่วงเวลาของพยางค์

การคำนวณค่า F0 ของการผันวรรณยุกต์จะใช้ค่า F0_{base} ซึ่งได้จากการคำนวณจากสมการที่ (4.50) และ (4.60) ตามลำดับ

ในบทนี้เราได้นำเสนอการวิเคราะห์สัญญาณเสียง เพื่อวิเคราะห์หาคุณคุณสมบัติต่างๆ ของเสียงภาษาไทย แล้วนำมาเขียนอยู่ในรูปแบบของกฎการออกเสียงภาษาไทย จากกฎการออกเสียงภาษาไทยดังกล่าวนำไปสู่วิธีการสังเคราะห์เสียงภาษาไทย ซึ่งจะได้กล่าวในบทถัดไป

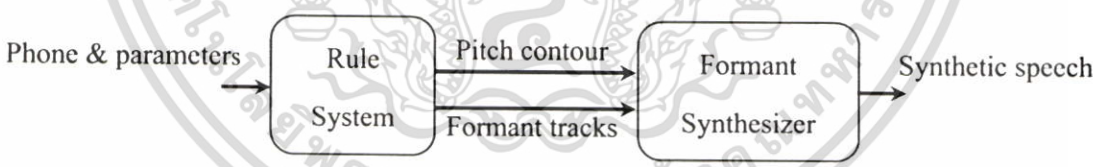
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5
การสังเคราะห์เสียงภาษาไทย

ในบทที่ผ่านมาได้กล่าวถึงการวิเคราะห์คุณสมบัติของเสียงภาษาไทย แล้วทำการวิเคราะห์เพื่อหากฎการออกเสียงภาษาไทย ในบทนี้จะนำเสนอการนำกฎการออกเสียงภาษาไทยมาทำการสังเคราะห์เสียงโดยวิธีสังเคราะห์เสียงแบบฟอร์แมนท์ เริ่มแรกจะกล่าวถึงวิธีสังเคราะห์แบบฟอร์แมนท์ แล้วจะกล่าวถึงการสังเคราะห์เสียงภาษาไทย โดยใช้กฎการออกเสียงภาษาไทยเพื่อสังเคราะห์เสียงที่ต้องการออกมา

5.1 หลักการสังเคราะห์เสียงแบบฟอร์แมนท์

หลักการสังเคราะห์เสียงแบบฟอร์แมนท์ เป็นหลักการที่เกี่ยวกับการสังเคราะห์เสียงของมนุษย์ที่สามารถกำหนดให้อยู่ในรูปแบบจำลองแหล่งกำเนิดเสียงที่สามารถสร้างขึ้นได้จากพื้นฐานทางสัญญาณไฟฟ้า และกำหนดให้ช่วงเวลาแปรผันตามฟอร์แมนท์ โครงสร้างของฟอร์แมนท์และเปลี่ยนฟอร์แมนท์เพื่อให้เกิดเสียงพูดธรรมชาติ ข้อมูลเสียงใดๆ จะถูกกำหนดไว้ในรูปของความถี่ฟอร์แมนท์ต่างๆ (F1, F2, F3) คุณสมบัตินี้สามารถลดจำนวนของข้อมูลที่ใช้ในการสร้างสัญญาณเสียงขึ้นเป็นจำนวนมาก รูปที่ 5.1 [23] แสดงระบบการสังเคราะห์เสียงแบบฟอร์แมนท์



รูปที่ 5.1 ระบบการสังเคราะห์เสียงแบบฟอร์แมนท์

ระบบสังเคราะห์เสียงแบบฟอร์แมนท์ประกอบด้วย 2 ส่วนหลัก คือ 1) Rule System 2) Formant Synthesizer

Rule System คือ ส่วนที่ทำหน้าที่ในการสร้างกฎต่างๆ ซึ่งในที่นี้จะเป็นการสร้างกฎสำหรับการออกเสียงภาษาไทย

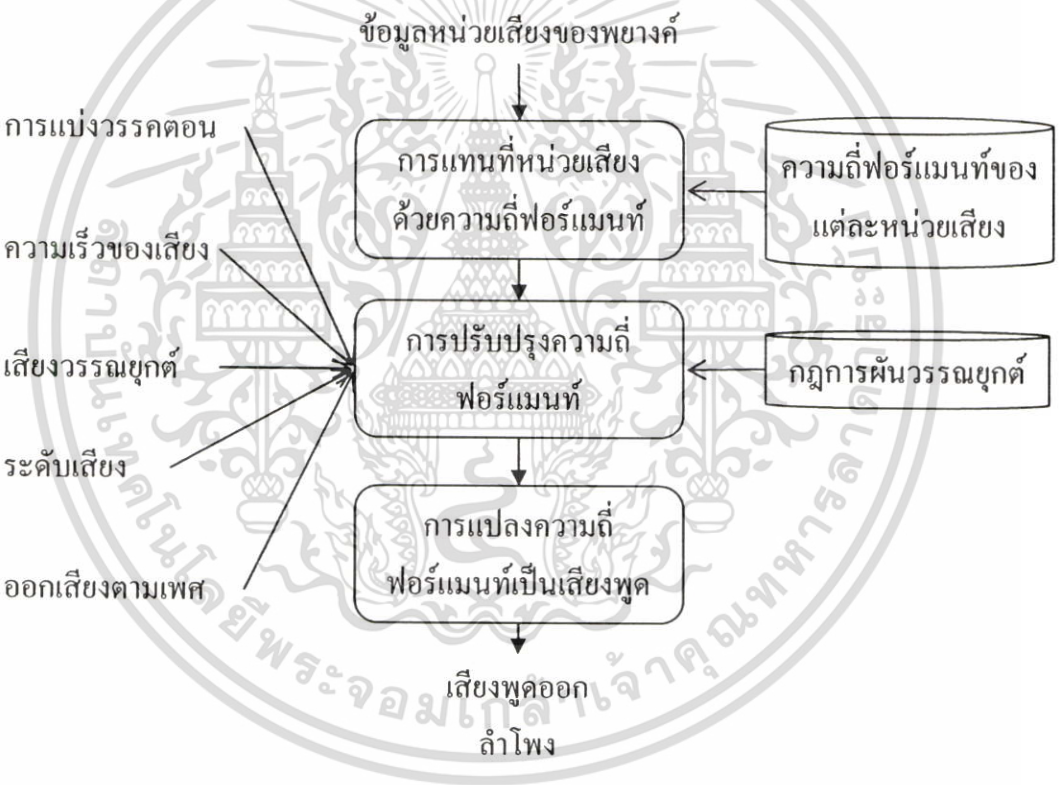
Formant Synthesizer คือ ส่วนที่ทำหน้าที่สังเคราะห์เสียง ซึ่งนำข้อมูลคุณสมบัติของเสียงที่จะสังเคราะห์ จากส่วนของ Rule System

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.2 การสังเคราะห์เสียงภาษาไทย

หลักการที่ใช้ในการสังเคราะห์เสียงคือ การนำกฎการออกเสียงภาษาไทยที่ได้จากการวิเคราะห์ มาทำการประมวลเพื่อให้ได้เสียงพยางค์ภาษาไทยที่สามารถเปลี่ยนแปลงได้ตามพารามิเตอร์ คือ การแบ่งวรรคตอน การกำหนดความเร็วเสียง การออกเสียงตามเพศ การกำหนดระดับเสียง และการผันวรรณยุกต์ภาษาไทย

ระบบสังเคราะห์เสียงภาษาไทยประกอบด้วย 3 ส่วนหลัก คือ1) การแทนหน่วยเสียงด้วยความถี่ฟอร์แมนท์ 2) การปรับปรุงความถี่ฟอร์แมนท์ และ 3) การเปลี่ยนความถี่ฟอร์แมนท์เป็นเสียงพูด ดังแสดงในรูปที่ 5.2 และเพื่อให้ง่ายต่อความเข้าใจ จะขอยกตัวอย่างการสังเคราะห์พยางค์ “ก้าน” ซึ่งมีรายละเอียดดังต่อไปนี้



รูปที่ 5.2 ระบบการสังเคราะห์เสียงภาษาไทย

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กำหนดให้ Space คือ เปอร์เซ็นต์การเพิ่มหรือลดการแบ่งวรรคตอน

Speed คือ เปอร์เซ็นต์การเพิ่มหรือลดความเร็วเสียง

Pitch คือ เปอร์เซ็นต์การเพิ่มหรือลดระดับเสียง

Tone คือ เสียงวรรณยุกต์

Sex คือ เพศ

InputPhoneme[] คือ อาเรย์หน่วยเสียงของพยางค์

เมื่อต้องการสังเคราะห์เสียงพยางค์ภาษาไทย จะนำเข้าหน่วยเสียงที่ต้องการสร้าง และกำหนดพารามิเตอร์ที่ต้องการ ในที่นี้ยกตัวอย่าง InputPhoneme[/ก/, /า/, /น/, s] โดยออกเสียงเพศชาย ทำการผันวรรณยุกต์โท ทำการลดความเร็วเสียง 10% ทำการเพิ่มช่วงเวลาการแบ่งวรรคตอนขึ้น 10% และเพิ่มระดับเสียงขึ้น 10% โดยมีขั้นตอนและวิธีการดังต่อไปนี้

5.2.1 การแทนที่หน่วยเสียงความถี่ฟอร์แมนท์

การแทนที่หน่วยเสียงด้วยความถี่ฟอร์แมนท์ ซึ่งจะรวมถึงพารามิเตอร์อื่นๆ ที่เกี่ยวข้องกับการสังเคราะห์ของแต่ละหน่วยเสียงนั้นๆ ด้วย ซึ่งเป็นกระบวนการแรกสำหรับระบบการสังเคราะห์เสียงภาษาไทย เราจะใช้ความถี่ฟอร์แมนท์และพารามิเตอร์อื่นๆ แทนหน่วยเสียงของพยางค์ ทุกๆ หน่วยเสียงของพยางค์จะจับคู่กับความถี่ฟอร์แมนท์ของหน่วยเสียงนั้นๆ ซึ่งข้อมูลหน่วยเสียงของพยางค์จะได้มาจากการวิเคราะห์ข้อความ โครงสร้างข้อมูลหน่วยเสียงของพยางค์ดังรูปที่ 5.3

รูปที่ 5.3 โครงสร้างข้อมูลหน่วยเสียงของพยางค์

โดยที่

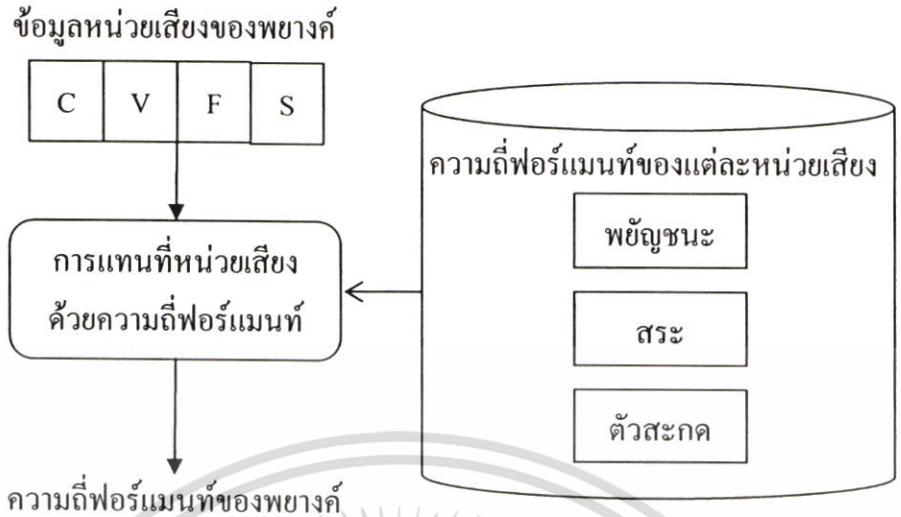
C คือ หน่วยเสียงของเสียงพยัญชนะต้นพยางค์หรือพยัญชนะควบกล้ำ

V คือ หน่วยเสียงของสระ

F คือ หน่วยเสียงของตัวสะกด

S คือ ไม่มีเสียง (เสียงเงียบ)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 5.4 องค์ประกอบของการแทนที่หน่วยเสียงด้วยความถี่ฟอร์แมนท์

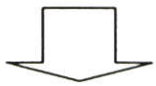
ซึ่งค่าความถี่ฟอร์แมนท์ของแต่ละหน่วยเสียง เป็นค่าคงตัวที่ได้จากการวิเคราะห์ในบทที่แล้ว ซึ่งประกอบไปด้วยกลุ่มของพยัญชนะ สระ และตัวสะกด ซึ่งการทำงานดังต่อไปนี้

```
// InputPhoneme = Input Phoneme of syllable  ;
// N = Integer variable for the length of InputPhoneme;
// Ph = A Phoneme is cut from an InputPhoneme
// F_Array = A buffer for Formant frequencies;
for(I=0; I<=N; I++)
    Ph= InPhoneme[I];
    F_Array = Search Ph from formant frequencies database
    F_Array= F_Array & F_Array
```

ผลที่ได้จากการแทนที่หน่วยเสียงด้วยความถี่ฟอร์แมนท์ จะเป็นข้อมูลความถี่ฟอร์แมนท์และพารามิเตอร์ต่างๆ ที่จำเป็นต่อการสังเคราะห์พยางค์นั้นๆ ซึ่งจะเก็บอยู่ในเป็นอาเรย์หรือเมตริก ดังแสดงในรูปที่ 5.5

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

C	V	F	S



TIME	TIME	TIME	TIME
F0	F0	F0	F0
F1	F1	F1	F1
F2	F2	F2	F2
F3	F3	F3	F3
AV	AV	AV	AV
AH	AH	AH	AH
AF	AF	AF	AF

รูปที่ 5.5 โครงสร้างข้อมูลความถี่ฟอร์แมนท์ของพยางค์

โดยที่

TIME คือ เวลามีหน่วยเป็นมิลลิวินาที

F0 คือ ความถี่มูลฐาน

F1 คือ ความถี่ฟอร์แมนท์ที่ 1

F2 คือ ความถี่ฟอร์แมนท์ที่ 2

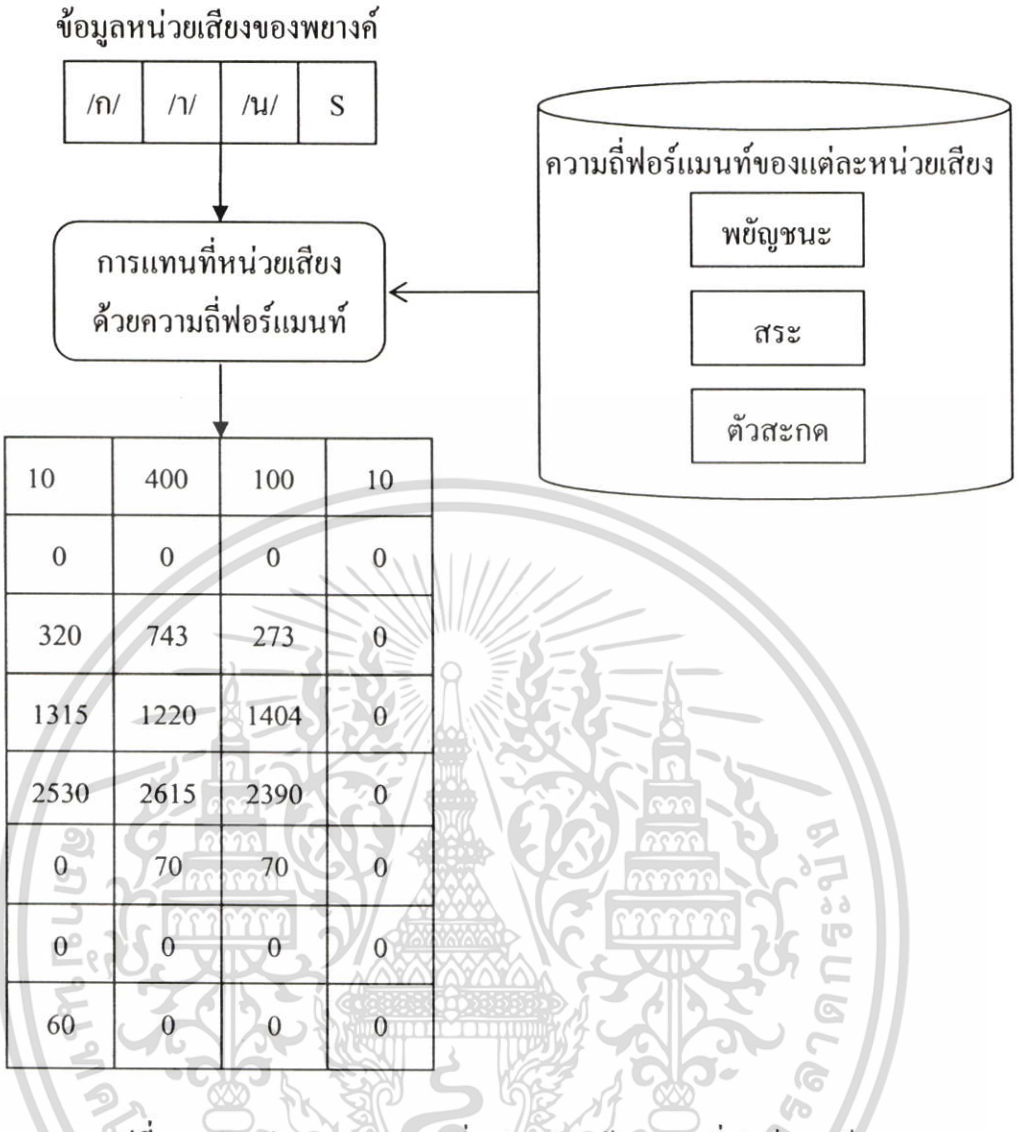
F3 คือ ความถี่ฟอร์แมนท์ที่ 3

AV คือ แอมพลิจูดเสียงก้อง

AH คือ แอมพลิจูดเสียงพ่น

AF คือ แอมพลิจูดเสียงสอดแทรก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



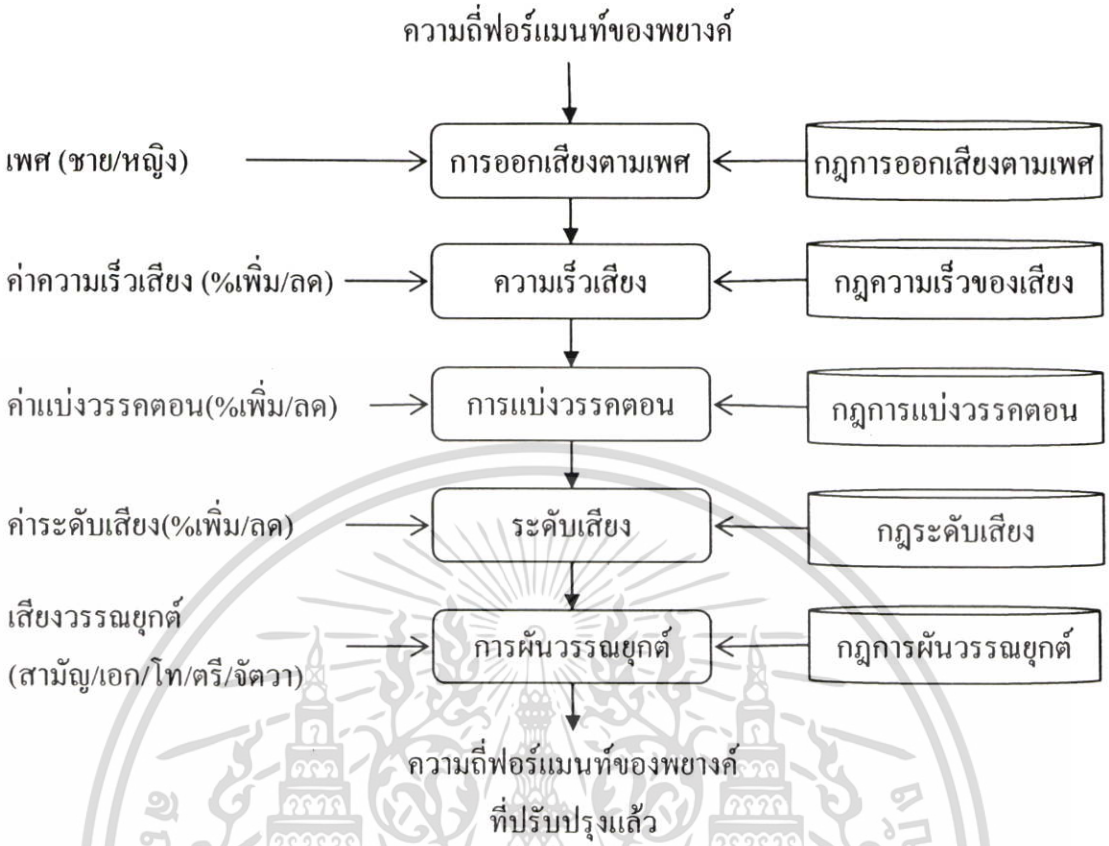
รูปที่ 5.6 ตัวอย่างของการแทนที่หน่วยเสียงด้วยความถี่ฟอร์แมนท์

ผลที่ได้จากการแทนที่หน่วยเสียงด้วยความถี่ฟอร์แมนท์ จะถูกนำไปใช้ในการปรับปรุงความถี่ฟอร์แมนท์ของพยางค์ตามพารามิเตอร์ ดังกล่าวถึงในหัวข้อถัดไป

5.2.2 การปรับปรุงความถี่ฟอร์แมนท์ของพยางค์ตามพารามิเตอร์

เป็นกระบวนการที่ทำการปรับปรุงความถี่ฟอร์แมนท์ของพยางค์ตามพารามิเตอร์ที่ถูกส่งเข้ามา พารามิเตอร์เหล่านี้ประกอบด้วย การแบ่งวรรคตอน ความเร็วเสียง, ระดับเสียง, เสียงวรรณยุกต์ และออกเสียงตามเพศ โดยจะทำการดึงข้อมูลของกฎต่างๆ ที่ใช้สำหรับแต่ละพารามิเตอร์มาเป็นข้อมูลในการปรับปรุงความถี่ฟอร์แมนท์ให้ได้ตามที่ต้องการ ซึ่งมีรายละเอียดดังต่อไปนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูปที่ 5.7 การปรับปรุงความถี่ฟอร์แมนท์ของพยางค์ตามพารามิเตอร์

การออกเสียงตามเพศ เป็นการประมวลผลตามค่าพารามิเตอร์เพศ คือ เพศชาย และเพศหญิง ซึ่งกำหนดโดยผู้ใช้ ดังได้อธิบายในหัวข้อที่ 4.3.1 ซึ่งเป็นการประมวลผลค่าความถี่มูลฐานและความถี่ฟอร์แมนท์ของพยางค์ที่ได้จากกระบวนการก่อนหน้า ตัวอย่างนี้จะแสดงการประมวลผลการออกเสียงของเพศชาย

เริ่มแรกทำการปรับค่าช่วงเวลา(TIME) ให้เป็นเวลา ณ ตำแหน่งนั้นๆแทน จากนั้นทำการกำหนดค่าตัวแปรความถี่มูลฐานหลัก ($F0_{base}$) โดยจะนำค่าความถี่มูลฐานเฉลี่ยของเพศชายกำหนดเป็นค่าเริ่มต้น ซึ่งมีค่าเท่ากับ 115 Hz จากนั้นทำการกำหนดค่าช่วงของความถี่มูลฐานต่ำสุดสูงสุด ที่สามารถปรับค่าได้ โดยกำหนดค่าของตัวแปรค่าความถี่มูลฐานต่ำสุด ($F0_{min}$) และกำหนดค่าของตัวแปรค่าความถี่มูลฐานสูงสุด ($F0_{max}$) เพื่อนำไปใช้ในขั้นตอนการปรับระดับเสียง
จากนั้นทำการปรับปรุงความถี่ฟอร์แมนท์ที่ 1, 2 และ 3 ดังแสดงในสมการที่ (4.53-4.55) ซึ่งค่าคงตัวของความถี่ฟอร์แมนท์ที่ 1 ($F1_{male}$) ค่าคงตัวของความถี่ฟอร์แมนท์ที่ 2 ($F2_{male}$) และค่าคงตัวของความถี่ฟอร์แมนท์ที่ 3 ($F3_{male}$) ของเสียงผู้ชายมีค่าเท่ากับ -100

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

10	400	100	10
0	0	0	0
320	743	273	0
1315	1220	1404	0
2530	2615	2390	0
0	70	70	0
0	0	0	0
60	0	0	0

ความถี่ฟอร์แมนท์ก่อนการปรับปรุง

10	410	510	520
0	0	0	0
220	643	173	0
1215	1120	1304	0
2430	2515	2290	0
0	70	70	0
0	0	0	0
60	0	0	0

ความถี่ฟอร์แมนท์หลังการปรับปรุง

รูปที่ 5.8 การปรับปรุงความถี่ฟอร์แมนท์ของการออกเสียงเพศชาย

ความเร็วเสียง เป็นกระบวนการที่ประมวลผลเกี่ยวกับความเร็วของแต่ละเสียงพยางค์ นำค่าความเร็วที่ผู้ใช้กำหนด ซึ่งมีการเพิ่มหรือลดความเร็วเป็นเปอร์เซ็นต์ ในตัวอย่างนี้จะทำการลดความเร็วเสียงลง 10% นั่นหมายความว่าช่วงระยะเวลาของพยางค์จะเพิ่มมากขึ้น ซึ่งได้อธิบายไว้ในหัวข้อที่ 4.3.2 มีวิธีการคำนวณดังต่อไปนี้

ทำการคำนวณช่วงเวลาของพยางค์ ($T0_{new}$)

$$T0_{new} = \frac{(100+10)x510}{100} \tag{5.1}$$

$$T0_{new} = 561 \tag{5.2}$$

จากนั้นทำการคำนวณค่าเวลาใดๆ ($t1_{new}$) ในพยางค์ด้วยสมการที่ (4.56) จนครบทุกช่องเวลาดังตัวอย่างต่อไปนี้

$$t1_{new} = \frac{10x561}{510} = 11 \tag{5.3}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

10	410	510	520
0	0	0	0
220	643	173	0
1215	1120	1304	0
2430	2515	2290	0
0	70	70	0
0	0	0	0
60	0	0	0

ข้อมูลเวลาก่อนการปรับปรุง

11	451	561	572
0	0	0	0
220	643	173	0
1215	1120	1304	0
2430	2515	2290	0
0	70	70	0
0	0	0	0
60	0	0	0

ข้อมูลเวลาหลังการปรับปรุง

รูปที่ 5.9 การปรับปรุงข้อมูลเวลาของการลดความเร็วเสียงลง 10%

การแบ่งวรรคตอน เป็นกระบวนการที่ประมวลผลเกี่ยวกับช่วงเวลาของเสียงเงียบที่ท้ายพยางค์ นำค่าการแบ่งวรรคตอนที่ผู้ใช้กำหนด ที่มีการการเพิ่มหรือลดช่วงเวลาเป็นเปอร์เซ็นต์ ในตัวอย่างนี้จะทำการเพิ่มช่วงเวลาการแบ่งวรรคตอนขึ้น 10% มีวิธีการคำนวณดังต่อไปนี้

ทำการคำนวณช่วงเวลาของวรรคตอน ($T1_{new}$) ดังสมการที่ (4.58)

$$T1_{new} = \frac{(100+10)x(572-561)}{100} \tag{5.4}$$

$$T1_{new} = 12.1 \tag{5.5}$$

จากนั้นทำการคำนวณค่าเวลาใดๆ ($t2_{new}$) ของท้ายวรรคตอนด้วยสมการที่ (4.59) ดังตัวอย่างต่อไปนี้

$$t2_{new} = 561 + 12 = 573 \tag{5.6}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

11	451	561	572
0	0	0	0
220	643	173	0
1215	1120	1304	0
2430	2515	2290	0
0	70	70	0
0	0	0	0
60	0	0	0

ข้อมูลเวลาก่อนการปรับปรุง

11	451	561	573
0	0	0	0
220	643	173	0
1215	1120	1304	0
2430	2515	2290	0
0	70	70	0
0	0	0	0
60	0	0	0

ข้อมูลเวลาหลังการปรับปรุง

รูปที่ 5.10 การปรับปรุงข้อมูลเวลาของการเพิ่มช่วงเวลาการแบ่งวรรคตอนขึ้น 10%

ระดับเสียง เป็นการกำหนดค่าตัวแปรความถี่มูลฐานหลัก ($F0_{base}$) ที่ถูกกำหนดในกระบวนการออกเสียงตามเพศ การเพิ่มหรือลดความถี่มูลฐานหลักคิดเป็นเปอร์เซ็นต์ ในตัวอย่างนี้จะทำการเพิ่มความระดับเสียงขึ้น 10% มีวิธีการคำนวณดังต่อไปนี้

$$F0_{base} = \frac{(100 + Pitch)xF0_{base}}{100} \tag{5.7}$$

$$F0_{base} = \frac{(100+10)x115}{100} \tag{5.8}$$

$$F0_{base} = 126.5 \tag{5.9}$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การผันวรรณยุกต์ เป็นการประมวลผลค่าของความถี่มูลฐานตามค่าพารามิเตอร์เสียงวรรณยุกต์ โดยจะทำการเลือกกฎดังสมการที่ (4.45-4.49) ที่จะใช้ผันวรรณยุกต์จากค่าของเสียงวรรณยุกต์ ในตัวอย่างนี้จะทำการผันวรรณยุกต์โท มีขั้นตอนดังต่อไปนี้

เลือกสมการการผันวรรณยุกต์จากค่าของเสียงวรรณยุกต์ ในตัวอย่างนี้เป็นเสียงวรรณยุกต์โท ดังนั้นสมการคือ

$$F0_{ln}(t) = \begin{cases} F0_{basc} + 42 + 1.18t - 0.02t^2 & 0 < t <= 56 \\ F0_{ln}(56) - 3.5(t-56) - 0.05(t-56)^2 & 56 < t <= 100 \end{cases} \quad (5.10)$$

ทำการแทรกคอลัมน์ระหว่างเวลาที่ความถี่มูลฐานเปลี่ยนทิศทาง วรรณยุกต์โทมีค่าเท่ากับ 56 แล้วทำการสำเนาค่าต่างๆ ยกเว้นค่าเวลาของคอลัมน์ถัดไป แล้วทำการคำนวณเวลาของคอลัมน์ที่แทรก ตามสมกาที่ (4.61)

$$t0_c = \frac{56 \times 561}{100} = 314.16 \quad (5.11)$$

แทรกคอลัมน์ ณ เวลา 314

11	451	561	573
0	0	0	0
220	643	173	0
1215	1120	1304	0
2430	2515	2290	0
0	70	70	0
0	0	0	0
60	0	0	0

ข้อมูลเวลาก่อนการปรับปรุง

11	314	451	561	573
0	0	0	0	0
220	643	643	173	0
1215	1120	1120	1304	0
2430	2515	2515	2290	0
0	70	70	70	0
0	0	0	0	0
60	0	0	0	0

ข้อมูลเวลาหลังการปรับปรุง

รูปที่ 5.11 การแทรกคอลัมน์ของจุดเปลี่ยนการผันวรรณยุกต์โท

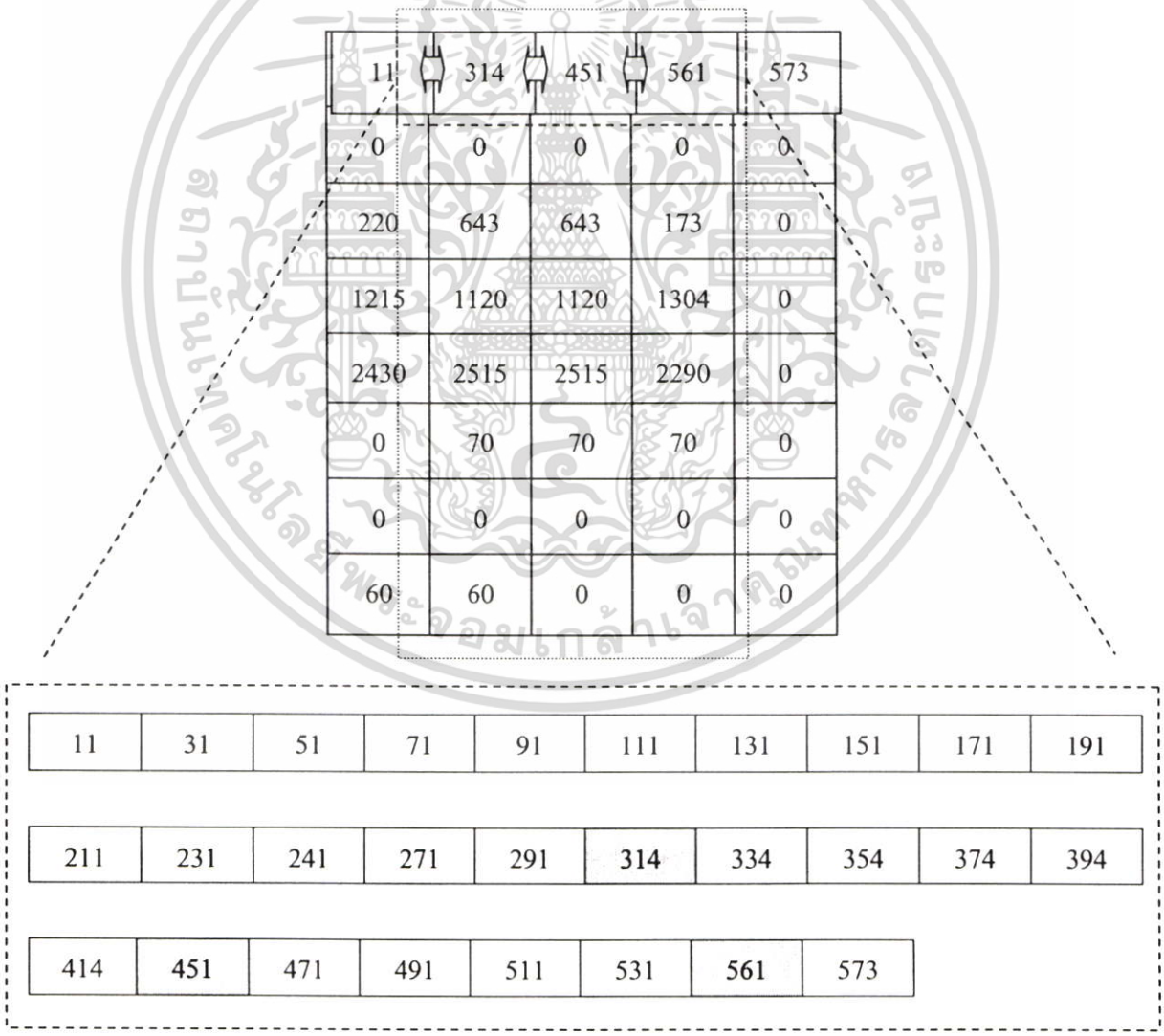
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ทำการคำนวณจำนวนเฟรมที่จะใช้ในแต่ละช่วงคอลัมน์ ซึ่งจำนวนเฟรมจะขึ้นอยู่กับค่าของเวลาที่ใช้ในแต่ละเฟรม ผลที่ได้จากการหารจะปัดเศษทิ้งไป ดังแสดงในสมการที่ (5.12-5.14)

$$\text{Frame}_1 = \frac{314 - 11}{20} = 15 \tag{5.12}$$

$$\text{Frame}_2 = \frac{451 - 314}{20} = 6 \tag{5.13}$$

$$\text{Frame}_3 = \frac{561 - 451}{20} = 5 \tag{5.14}$$



รูปที่ 5.12 การแทรกคอลัมน์ของการผันวรรณยุกต์โท

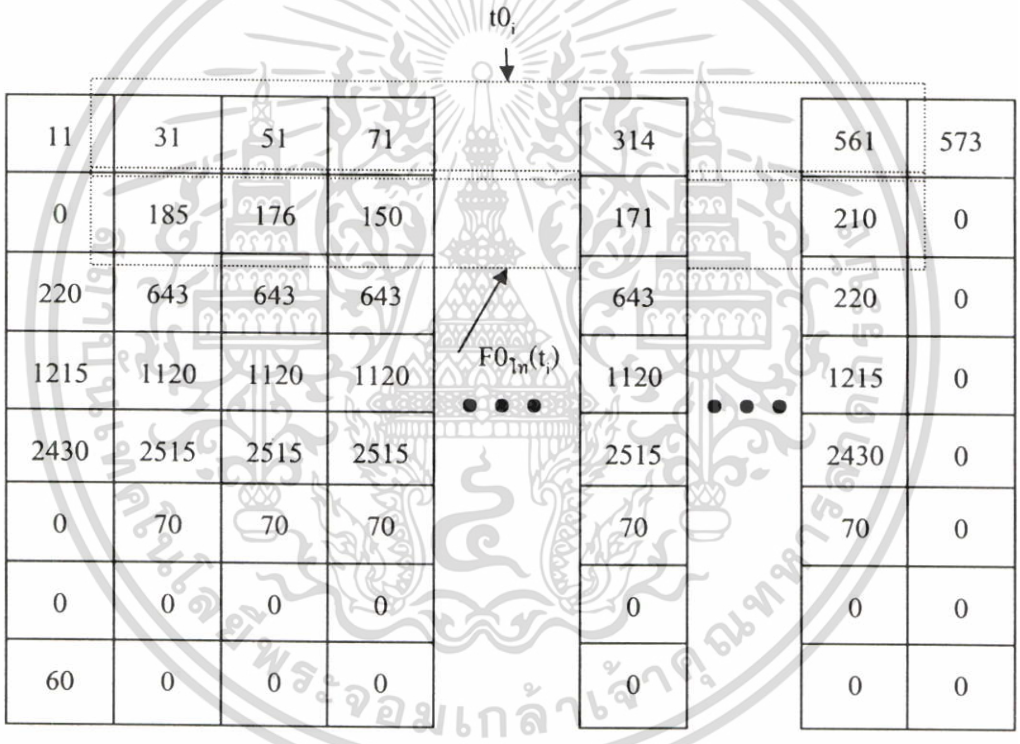
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ทำการคำนวณหาค่าความถี่มูลฐานของแต่ละคอลัมน์ จากฟังก์ชัน $F0_{โท}(t_i)$ เมื่อ t_i คือ ค่าเวลา ณ t ใดๆ ซึ่งสามารถคำนวณได้จากสมการที่ (5.15) ซึ่งจะทำการคำนวณเฉพาะคอลัมน์ที่มีลักษณะความเป็นรายคาบหรือค่าในแถว AV ไม่เท่ากับ 0 แล้วนำค่าที่ได้กลับเข้าคอลัมน์เดิมในแถว F0

$$t_i = \frac{t0_i \times 100}{T0} \qquad (5.15)$$

เมื่อ $t0_i$ คือ ค่าเวลาในแถว TIME

T0 คือ ช่วงเวลาของพยางค์



11	31	51	71	314	561	573
0	185	176	150	171	210	0
220	643	643	643	643	220	0
1215	1120	1120	1120	1120	1215	0
2430	2515	2515	2515	2515	2430	0
0	70	70	70	70	70	0
0	0	0	0	0	0	0
60	0	0	0	0	0	0

รูปที่ 5.13 การคำนวณค่าความถี่มูลฐานของการผันวรรณยุกต์โท

ค่าที่ได้จากการปรับปรุงความถี่ฟอร์แมนท์นี้ จะถูกนำไปสังเคราะห์เป็นเสียง 1 พยางค์ โดย Formant Synthesizer โดยจะนำค่าแต่ละคอลัมน์ไปประมวลผลตามลำดับ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.2.3 การแปลงความถี่ฟอร์แมนท์ของพยางค์ให้เป็นเสียงพูด

ผลที่ได้จากการปรับปรุงความถี่ฟอร์แมนท์ ของพยางค์ตามพารามิเตอร์ จะนำไปใช้ในการแปลงให้เป็นเสียงพูด โดยส่งให้ Formant synthesizer เป็นตัวสังเคราะห์เสียงขึ้นมา ซึ่งจะต้องทำการแปลงข้อมูลให้ได้ตามที่ Formant Synthesizer ต้องการ โดยข้อมูลของการส่งเข้า Formant Synthesizer จะเป็นข้อมูลที่เปลี่ยนพารามิเตอร์ตามเวลา ดังแสดงต่อไปนี้

TIME = 0

F1=220; F2=1215; F3=2430; AF=60;

TIME = 11

F0=185; F1=643; F2=1120; F3=2515; AV=70;

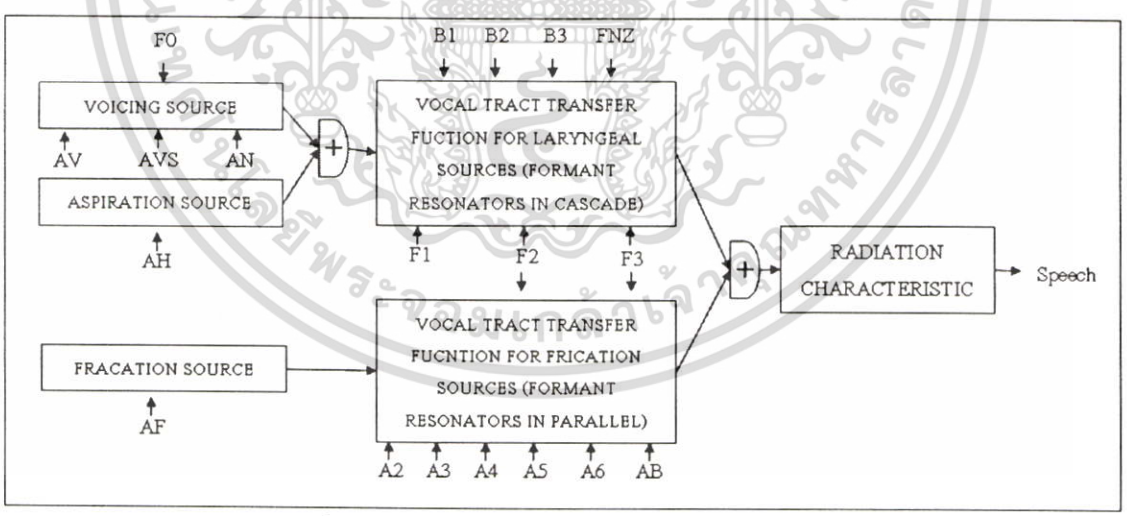
TIME =51

F0=176; F1=643; F2=1120; F3=2515; AV=70;

TIME =573

F0=0; F1=0; F2=0; F3=5; AV=0;

ซึ่งข้อมูลเหล่านี้จะจะถูกส่งไปให้ Formant Synthesizer ดังรูปที่ 5.14 ตามลำดับเวลาที่ได้จากการประมวลผล แล้วจึงสังเคราะห์เป็นเสียงพูดออกมา



รูปที่ 5.14 องค์ประกอบของ Formant Synthesizer

ในบทนี้เราได้นำเสนอการสังเคราะห์เสียงภาษาไทย พร้อมทั้งยกตัวอย่าง ซึ่งผลที่ได้จากการสังเคราะห์เสียงดังกล่าวนำไปสู่การประเมินผลการสังเคราะห์เสียงภาษาไทย ซึ่งจะได้กล่าวในบทถัดไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 6

ผลการทดลอง

จากวิธีการสังเคราะห์เสียงภาษาไทยโดยวิธีสังเคราะห์แบบฟอร์แมนท์ที่นำเสนอไป เมื่อนำมาทำการสังเคราะห์พยางค์ต่างๆ ในภาษาไทยได้ผลการสังเคราะห์ดังที่จะแสดงในบทนี้

6.1 การประเมินผลการสังเคราะห์เสียง

การประเมินผลการสังเคราะห์เสียงทำเพื่อวัดคุณภาพของเสียงที่สังเคราะห์ขึ้นโดยการประเมินที่จะทำไปนี้ทำเพื่อวัดความสามารถในการรับรู้ (Intelligibility) การประเมินนี้มุ่งไปที่คุณภาพระดับพยางค์เนื่องจากวิธีการที่นำเสนอตั้งอยู่บนฐานเสียงสังเคราะห์ระดับพยางค์โดยวิธีการวิเคราะห์แบบเอ็มอาร์ที

เอ็มอาร์ที [10], [27] เป็นวิธีที่ใช้วัดความสามารถในการแยกแยะเสียงที่สังเคราะห์ขึ้นกลุ่มคำหรือพยางค์ที่ใช้ในการทดสอบเป็นคำหรือพยางค์ที่มีเสียงคล้ายกัน (Rhyme) เสียงที่ใช้ทดสอบอยู่บนฐานเสียงระดับพยางค์หรือคำ เนื่องจากการทดสอบระดับพยางค์หรือคำนี้มีความไวต่อการผิดพลาดที่จะเกิดขึ้นในเสียงสังเคราะห์ได้ง่าย ทำให้สามารถนำไปใช้วิเคราะห์หาข้อผิดพลาดในการสังเคราะห์ได้ง่ายและเสียงที่ใช้ไม่จำเป็นต้องมีความหมาย ซึ่งจะทำให้ไวต่อข้อผิดพลาดได้มากขึ้น เนื่องจากไม่สามารถใช้การคาดเดาได้ ข้อดีประการหนึ่ง คือวิธีการนี้อาสาสมัครที่ทำแบบทดสอบไม่จำเป็นต้องได้รับการฝึกมาก่อน สำหรับวิธีการทดสอบ ใช้อาสาสมัครฟังชุดเสียงสังเคราะห์ทีละพยางค์ โดยไม่ทราบว่าเสียงที่ได้ยินคือเสียงอะไร แล้วให้เลือกตัวเลือกว่าเสียงที่ได้ยินเป็นเสียงใด โดยทั่วไปมี 6 ตัวเลือก ตัวเลือกแต่ละตัวจะมีเสียงที่แตกต่างกันเพียงจุดเดียวและครอบคลุมเสียงที่เกิดขึ้นได้ทั้งหมด การทดสอบนี้โดยทั่วไปมุ่งเน้นไปที่เสียงพยัญชนะ เนื่องจากเสียงพยัญชนะเป็นเสียงที่สังเคราะห์ได้ยากกว่าเสียงสระ

การประเมินผลใช้อาสาสมัครในการทำแบบประเมินจำนวน 40 คน การเลือกพยางค์ที่นำมาทดสอบมีกฎดังต่อไปนี้

1. เสียงพยัญชนะ ได้จากการสุ่มพยัญชนะในแต่ละตำแหน่งการออกเสียง เนื่องจากแต่ละตำแหน่งมีการสังเคราะห์ที่แตกต่างกัน

2. เสียงสระ ได้จากการสุ่มจากสระเสียงสั้น และสระเสียงยาว เนื่องจากสระเสียงสั้นและเสียงยาวใช้ความถี่ฟอร์แมนท์เดียวกัน จึงเลือกใช้ทั้งสอง เพื่อวิเคราะห์หาความแตกต่าง

3. เสียงวรรณยุกต์ ทดสอบด้วยการนำเสียงสระเท่านั้นมาทำการผันเสียงวรรณยุกต์ทั้ง 5 เสียง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ในแบบประเมินผลดังกล่าวแบ่งเป็นการประเมินระดับพยางค์เสียงพยัญชนะและสระ 30 พยางค์ และเสียงวรรณยุกต์ 10 พยางค์ การทดลองและหาค่าร้อยละความถูกต้องของวิธีการสังเคราะห์เสียงพูดภาษาไทย โดยใช้สูตร

$$a = \frac{C}{N} \times 100 \tag{6.1}$$

โดยที่ a คือ ร้อยละความถูกต้องของแต่ละตัวอย่าง

C คือ จำนวนเสียงที่ฟังแล้วถูกต้อง

N คือ จำนวนเสียงตัวอย่างทั้งหมดที่ฟัง

ร้อยละความถูกต้องเฉลี่ยหาได้จากสูตร

$$O = \frac{1}{N}\sum_{i=1}^{N} a_i \tag{6.2}$$

โดยที่ O คือ ร้อยละความถูกต้องเฉลี่ย

a_i คือ ร้อยละความถูกต้องของแต่ละตัวอย่าง

N คือ จำนวนตัวอย่างทั้งหมด

ซึ่งผลการทดลองดังแสดงในตารางที่ 6.1-6.4

ตารางที่ 6.1 ผลการประเมินเสียงพยัญชนะในระดับพยางค์

ผลการประเมินเสียงพยัญชนะ (ค่าเฉลี่ย)		
ฐานที่เกิดเสียง	จำนวนคำที่ถูก	ร้อยละความถูกต้อง
ริมฝีปาก	64 (120)	53.33
ปุ่มเหงือก	44 (120)	36.66
เพดานแข็ง	64 (120)	53.33
เพดานอ่อน	88 (120)	73.33
รวม	260 (480)	54.16

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ตารางที่ 6.2 ผลการประเมินเสียงสระ

ผลการประเมินเสียงสระ (ค่าเฉลี่ย)		
ประเภท	จำนวนคำที่ถูก	ร้อยละความถูกต้อง
สระเสียงสั้น	336 (360)	93.33
สระเสียงยาว	347 (360)	96.39
รวม	683 (720)	94.86

ตารางที่ 6.3 ผลการประเมินเสียงวรรณยุกต์

ผลการประเมินเสียงวรรณยุกต์ (ค่าเฉลี่ย)		
วรรณยุกต์	จำนวนคำที่ถูก	ร้อยละความถูกต้อง
สามัญ	46 (80)	57.50
เอก	41 (80)	51.25
โท	47 (120)	39.17
ตรี	34 (120)	42.50
จัตวา	37 (80)	92.50
รวม	205 (400)	51.25

ตารางที่ 6.4 ผลการประเมินในระดับพยางค์โดยรวม

ผลการประเมินในระดับพยางค์โดยรวม (เฉลี่ย)		
ผลการประเมิน	จำนวนคำที่ถูก	ร้อยละความถูกต้อง
เสียงพยัญชนะ	206	54.16
เสียงสระ	603	94.86
เสียงวรรณยุกต์	205	51.25
รวม	1,014	71.75

สรุปผลประเมินการสังเคราะห์เสียงภาษาไทยโดยวิธีสังเคราะห์แบบฟอร์แมนท์ ให้ผลโดยรวมของร้อยละความถูกต้องเฉลี่ยเท่ากับ 71.75

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

การประเมินการสังเคราะห์เสียงเมื่อทำการสังเคราะห์เสียงจากคำโดดในภาษาไทยที่มีความหมาย โดยใช้คำที่มีเสียงผสมทั้งพยัญชนะ สระ และวรรณยุกต์ จำนวน 30 คำ การประเมินใช้อาสาสมัคร 5 คน โดยใช้แบบสอบถามที่เป็นอัตนัย ให้ฟังทีละคำ แล้วเขียนคำที่ตอบลงในแบบสอบถาม

ตารางที่ 6.5 ผลการประเมินในระดับคำ

ผลการประเมินในระดับคำ		
ผลการประเมิน	จำนวนคำที่ถูก	ร้อยละความถูกต้อง
อาสาสมัครคนที่ 1	25 (30)	83.33
อาสาสมัครคนที่ 2	21 (30)	70.00
อาสาสมัครคนที่ 3	22 (30)	73.33
อาสาสมัครคนที่ 4	20 (30)	66.67
อาสาสมัครคนที่ 5	23 (30)	76.67
รวม	111 (150)	74.00

ทำการเปรียบเทียบพยัญชนะที่มีเสียงใกล้เคียงกัน โดยการสังเคราะห์เป็นคำเดี่ยวภาษาไทย โดยทำการเปรียบเทียบเป็นคู่พยัญชนะดังต่อไปนี้ กิน เคย, รั่ว ลอย, แพ ป่า, โม้ เงา, ไว ย่าม, หอ อา ซึ่งผลการทดลองร้อยละความถูกต้องที่ได้คือ 90, 85, 60, 80, 70 และ 50 ตามลำดับ

สรุปผลประเมินการสังเคราะห์เสียงภาษาไทยเมื่อทำการสังเคราะห์เสียงจากคำโดดในภาษาไทยที่มีความหมาย โดยใช้คำที่มีเสียงผสมทั้งพยัญชนะ สระ และวรรณยุกต์ จำนวน 30 คำ ให้ผลโดยรวมของร้อยละความถูกต้องเฉลี่ยเท่ากับ 74

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 7
สรุปผลการวิจัยและข้อเสนอแนะ

7.1 สรุปผลการทดลอง

วิทยานิพนธ์นี้นำเสนอวิธีการและผลของการสังเคราะห์เสียงพูดภาษาไทยด้วยวิธีสังเคราะห์ฟอร์แมนท์ ซึ่งเรามีกระบวนการเบื้องต้น 2 กระบวนการคือ 1) กระบวนการวิเคราะห์เพื่อหาความถี่ฟอร์แมนท์และพารามิเตอร์ที่เกี่ยวข้อง 2) กระบวนการสังเคราะห์เสียงพูดภาษาไทย กระบวนการแรก เป็นการหาความถี่ฟอร์แมนท์และพารามิเตอร์ที่เหมาะสมสำหรับการแทนหน่วยเสียงภาษาไทย และกระบวนการที่สอง เป็นการสังเคราะห์เสียงพูดภาษาไทยด้วยวิธีสังเคราะห์ฟอร์แมน ซึ่งประกอบไปด้วย 3 ส่วนหลักๆ คือ 1) การแทนที่หน่วยเสียงด้วยความถี่ฟอร์แมนท์, 2) การปรับปรุงความถี่ฟอร์แมนท์ตามพารามิเตอร์ และ 3) การแปลงความถี่ฟอร์แมนท์ให้เป็นเสียงพูด ผลการทดลองด้วยวิธีที่ใช้ในวิทยานิพนธ์ โดยสำรวจจากอาสาสมัครจำนวน 40 คน พบว่าความถูกต้องของเสียงพยางค์ภาษาไทย ร้อยละความถูกต้องของเสียงภาษาไทยเท่ากับ 71.75

7.2 ปัญหาและอุปสรรค

สำหรับปัญหาในส่วนของการวิจัยนั้น ส่วนใหญ่เป็นปัญหาเกี่ยวกับการเก็บตัวอย่างเสียง ซึ่งต้องเลือกผู้บอกภาษาที่มีความหลากหลาย และสามารถพูดภาษาไทยได้ชัดเจน

7.3 ข้อเสนอแนะ

เพื่อให้ได้เสียงสังเคราะห์ที่มีคุณภาพ จึงควรหาวิธีการวิเคราะห์เสียงพยัญชนะภาษาไทยที่เป็นลักษณะเด่นของพยัญชนะแต่ละเสียง ที่ออกเสียงที่คล้ายกัน และหาพารามิเตอร์ที่เกี่ยวข้องที่มีผลต่อการสังเคราะห์ให้ได้เสียงที่มีลักษณะคล้ายเสียงมนุษย์มากที่สุด สุดท้ายในการเก็บตัวอย่างเสียงจากผู้บอกภาษาควรลดสัญญาณรบกวนให้ได้มากที่สุด เพื่อให้ได้ผลการวิเคราะห์ที่ถูกต้องแม่นยำที่สุด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

เอกสารอ้างอิง

[1] NECTEC. "**ซอฟท์แวร์ประมวลผลเพื่อสังเคราะห์เสียงพูดภาษาไทย.**" [Online]. Available : http://vaja.nectec.or.th/report/spsths2000.html. 2005.

[2] จินดา เฮงสมบูรณ์. **ภาษาศาสตร์เบื้องต้น.** กรุงเทพ : สุวีริยาสาส์น. 2542.

[3] Ben Gold and Nelson Morgan. **Speech and audio signal processing.** USA : John Wiley & Sons. 1999.

[4] Roft C., Tor S. and Arvid S. "Data-driven formant synthesis" **TMH-QPSR** Vol. 44. Fornetik, 2002.

[5] Chao Huang, Yu Shi, Jianlai Zhou, Min Chu, Terry Wang and Eric Chang. "**Segmental Tonal Modeling for Phone Set Design in Mandarin LVCSR.**" proc. of the 15th International Congress of Phonetic Sciences. Barcelona. 2003.

[6] Chilin Shih and Richard Sproat. "**Issues in Text-to-Speech Conversion for Mandarin.**" [Online]. Available : http://www.bell-labs.com/project/tts/synth.ps. 2005.

[7] กมลชัย โชคชัยชุติกุล. "**ระบบคอมพิวเตอร์อ่านออกเสียงภาษาไทยจากเอกสาร.**" วิทยานิพนธ์ วิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาศาสตรคอมพิวเตอร์ บัณฑิตวิทยาลัย, จุฬาลงกรณ์ มหา วิทยาลัย. 2539.

[8] วิเชียร แซ่โล้ว. "**อ่านคำไทยพยางค์เดี่ยวแบบปรับความเร็วได้.**" วิทยานิพนธ์วิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมไฟฟ้า บัณฑิตวิทยาลัย, จุฬาลงกรณ์มหาวิทยาลัย. 2539.

[9] สง่า คงสุพานิช. "**การแปลงหน่วยคำในภาษาไทยเป็นสัญลักษณ์แทนเสียงสำหรับงานสังเคราะห์เสียงจากข้อความ.**" วิทยานิพนธ์วิศวกรรมศาสตรมหาบัณฑิต สาขาวิศวกรรมไฟฟ้า บัณฑิตวิทยาลัย, สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาด กระบัง. 2540.

[10] ชัชวาล หาญสกุลบรรเทิง. "**การสังเคราะห์เสียงพยางค์ภาษาไทยด้วยวิธีการสังเคราะห์แบบวิเคราะห์โดยใช้คู่เส้นสเปกตรัม.**" วิทยานิพนธ์วิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมไฟฟ้า บัณฑิตวิทยาลัย, จุฬาลงกรณ์มหาวิทยาลัย. 2542.

[11] นงนุช สุขตั้งมั่น. "**การรู้จำเสียงสระเดี่ยว และเสียงสระประสมในภาษาไทยโดยการใช้ สเปกตรัมแอลพีซีบนสเกลบาร์ก.**" วิทยานิพนธ์วิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมไฟฟ้า บัณฑิตวิทยาลัย, สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง. 2547.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

[12] กาญจนา นาคสกุล. **ระบบเสียงภาษาไทย**. กรุงเทพมหานคร : โรงพิมพ์จุฬาลงกรณ์มหาวิทยาลัย. 2524.

[13] Deller, J.R., Jr., Proakis, J.G. and Hansen, J.H. "**Discrete-Time Processing of Speech Signals.**" New Jersey, Prentice-Hall, 1993.

[14] Li Deng and Douglas O'Shaughnessy. **Speech Processing A Dynamic and Optimization-Oriented Approach.** New York : Marcel Dekker, Inc. 2004.

[15] ณัฏฐชา จิตติวรางกูล. "**ขั้นตอนการหาขอบเขตพยางค์สำหรับคำพูดต่อเนื่องภาษาไทย.**" วิทยานิพนธ์วิทยาศาสตร์มหาบัณฑิต สาขาวิชาวิศวกรรมไฟฟ้า บัณฑิตวิทยาลัย, จุฬาลงกรณ์มหาวิทยาลัย. 2541.

[16] วิรุศรุต อาขุบุตร. "**ระบบการรู้จำคำไทยหลายพยางค์แบบไม่ขึ้นกับผู้พูดโดยใช้แบบ จำลองฮิดเดนมาร์คอฟ.**" วิทยานิพนธ์วิศวกรรมศาสตรมหาบัณฑิต สาขาวิศวกรรมไฟฟ้า บัณฑิตวิทยาลัย, จุฬาลงกรณ์มหาวิทยาลัย. 2539.

[17] Alan V. Oppenheim and Ronald W. Schafer. **Discrete-Time Signal Processing**. USA : Prentice-hall International Inc. 1989.

[18] Michael D. Riley. **Speech Time-Frequency Representations**. USA : Kluwwer Academic Publishers. 1989.

[19] Michael, J. Roberts. **Signal and System Analysis Using Transform Methods and MATLAB**. Singapore : McGraw-Hill. 2004.

[20] Rabiner, L.R. and Schafer, R.W. **Digital Processing of Speech Signals**. New Jersy : Prentice-Hall. 1989.

[21] กิตติพัฒน์ ตันตระรุ่งโรจน์. **สัญญาณและระบบเวลาต่อเนื่อง**. กรุงเทพมหานคร : วิทยาพัฒน์. 2540.

[22] ปราโมทย์ เดชะอำไพ. **ระเบียบวิธีเชิงตัวเลขในงานวิศวกรรม**. พิมพ์ครั้งที่ 4. กรุงเทพมหานคร : สำนัก พิมพ์แห่งจุฬาลงกรณ์ มหาวิทยาลัย. 2546.

[23] Torbjorn Svendsen. "**Speech synthesis.**" [Online]. Available : http://www.tele.ntnu.no/signal/courses/taletek/Speech_ synthesis.pdf. 2004.

[24] ทรงศิริ แต้สมบัติ. **การวิเคราะห์การถดถอย**. พิมพ์ครั้งที่ 2. สำนักพิมพ์มหาวิทยาลัยเกษตรศาสตร์. 2542.

[25] ณัฐกร ทับทอง. "**การรู้จำคำพูดภาษาไทย โดยใช้ลักษณะบ่งความต่างของหน่วยเสียง.**" วิทยานิพนธ์วิทยาศาสตร์มหาบัณฑิต สาขาวิชาวิทยาศาสตร์คอมพิวเตอร์ บัณฑิตวิทยาลัย, จุฬาลงกรณ์มหาวิทยาลัย. 2538.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

[26] P.M. Hughes. **Formant based Speech Synthesis.** Speech and LanguageProcessing, USA, 1990.

[27] Ince, A.N. **Digital speech processing speech coding, synthesis and recognition.** USA : Kluwer Academic Publishers. 1992.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ก.
บทความเกี่ยวข้องที่ได้รับการตีพิมพ์

1. Saiyan Saiyod,.Sakchai Thipchaksurat and Somsak Mitatha. "**Thai Speech Synthesis for Text-to-Speech based on Formant Synthesis Technique.**" ECTI-CON2005, pp. 562-565, Pattaya, Thailand, May 12-13, 2005.
2. สายยัญ สายยศ, ศักดิ์ชัย ทิพย์จักษุรัตน์ และ สมศักดิ์ มิตะถา. "**กฎการผันวรรณยุกต์สำหรับการสังเคราะห์เสียงภาษาไทยโดยวิธีสังเคราะห์เสียงแบบฟอร์แมนท์.**" การประชุมวิชาการทางวิศวกรรมไฟฟ้า ครั้งที่ 28, ภูเก็ต, หน้า 1049-1052, 20-21 ตุลาคม 2548.
3. สายยัญ สายยศ, ศักดิ์ชัย ทิพย์จักษุรัตน์ และ สมศักดิ์ มิตะถา. "**กฎการออกเสียงภาษาไทยสำหรับการสังเคราะห์เสียงจากข้อความโดยวิธีสังเคราะห์แบบฟอร์แมนท์.**" The 9th National Computer Science and Engineering Conference(NCSEC 2005), มหาวิทยาลัยหอการค้าไทย, หน้า 519-529, 27-28 ตุลาคม 2548.

บทความเกี่ยวข้องที่รอตีพิมพ์

1. สายยัญ สายยศ, ศักดิ์ชัย ทิพย์จักษุรัตน์ และ สมศักดิ์ มิตะถา. "**การสังเคราะห์เสียงพูดภาษาไทยสำหรับการสังเคราะห์เสียงจากข้อความโดยวิธีสังเคราะห์ฟอร์แมนท์.**" วารสารพระจอมเกล้าลาดกระบัง.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ข

แบบสอบถามเพื่อวิเคราะห์ในงานวิจัยเรื่อง
การสังเคราะห์เสียงภาษาไทยโดยวิธีสังเคราะห์เสียงแบบฟอร์แมนท์
ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ชื่อ.....ชั้น..... อายุ..... เพศ (ชาย/หญิง).....

จงวงกลม◯ในข้อที่คิดว่าเป็นเสียงที่นักเรียนได้ยิน

1)	1. กอ	2. คอ	3. งอ	4. จอ	5. ชอ	6. ซอ
2)	1. กอ	2. คอ	3. งอ	4. จอ	5. ชอ	6. ซอ
3)	1. กอ	2. คอ	3. งอ	4. จอ	5. ชอ	6. ซอ
4)	1. กอ	2. คอ	3. งอ	4. จอ	5. ชอ	6. ซอ
5)	1. กอ	2. คอ	3. งอ	4. จอ	5. ชอ	6. ซอ
6)	1. กอ	2. คอ	3. งอ	4. จอ	5. ชอ	6. ซอ
7)	1. ปอ	2. พอ	3. งอ	4. มอ	5. ยอ	6. รอ
8)	1. ปอ	2. พอ	3. งอ	4. มอ	5. ยอ	6. รอ
9)	1. ปอ	2. พอ	3. งอ	4. มอ	5. ยอ	6. รอ
10)	1. ปอ	2. พอ	3. งอ	4. มอ	5. ยอ	6. รอ
11)	1. ปอ	2. พอ	3. งอ	4. มอ	5. ยอ	6. รอ
12)	1. ปอ	2. พอ	3. งอ	4. มอ	5. ยอ	6. รอ
13)	1. อะ	2. อา	3. อิ	4. อี	5. อุ	6. อู
14)	1. อะ	2. อา	3. อิ	4. อี	5. อุ	6. อู
15)	1. อะ	2. อา	3. อิ	4. อี	5. อุ	6. อู
16)	1. อะ	2. อา	3. อิ	4. อี	5. อุ	6. อู
17)	1. อะ	2. อา	3. อิ	4. อี	5. อุ	6. อู
18)	1. อะ	2. อา	3. อิ	4. อี	5. อุ	6. อู
19)	1. อึ	2. อือ	3. เอะ	4. เอ	5. เออะ	6. เออ
20)	1. อึ	2. อือ	3. เอะ	4. เอ	5. เออะ	6. เออ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

21)	1. อึ	2. อือ	3. เอะ	4. เอ	5. เออะ	6. เออ
22)	1. อึ	2. อือ	3. เอะ	4. เอ	5. เออะ	6. เออ
23)	1. อึ	2. อือ	3. เอะ	4. เอ	5. เออะ	6. เออ
24)	1. อึ	2. อือ	3. เอะ	4. เอ	5. เออะ	6. เออ
25)	1. แอะ	2. แอ	3. โอะ	4. โอ	5. เอาะ	6. ออ
26)	1. แอะ	2. แอ	3. โอะ	4. โอ	5. เอาะ	6. ออ
27)	1. แอะ	2. แอ	3. โอะ	4. โอ	5. เอาะ	6. ออ
28)	1. แอะ	2. แอ	3. โอะ	4. โอ	5. เอาะ	6. ออ
29)	1. แอะ	2. แอ	3. โอะ	4. โอ	5. เอาะ	6. ออ
30)	1. แอะ	2. แอ	3. โอะ	4. โอ	5. เอาะ	6. ออ
31)	1. ออ	2. อ่อ	3. อ้อ	4. อ๊อ	5. อ๋อ	
32)	1. ออ	2. อ่อ	3. อ้อ	4. อ๊อ	5. อ๋อ	
33)	1. ออ	2. อ่อ	3. อ้อ	4. อ๊อ	5. อ๋อ	
34)	1. ออ	2. อ่อ	3. อ้อ	4. อ๊อ	5. อ๋อ	
35)	1. ออ	2. อ่อ	3. อ้อ	4. อ๊อ	5. อ๋อ	
36)	1. แอ	2. แอ่	3. แอ้	4. แอ๊	5. แอ๋	
37)	1. อี	2. อี่	3. อี้	4. อี๊	5. อี๋	
38)	1. อา	2. อ่า	3. อ้า	4. อ๊า	5. อ๋า	
39)	1. อะ	2. อ่ะ	3. อ้ะ	4. อ๊ะ	5. อ๋ะ	
40)	1. อุ	2. อุ่	3. อุ้	4. อุ๊	5. อุ๋	

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ค.

คำที่ใช้ในการทดสอบการฟัง

1. กิน
2. เตะ
3. โต๊ะ
4. จีน
5. คืน
6. ดูด
7. โม้
8. เลอะ
9. ดุ
10. เคย
11. ย่าม
12. ฟาง
13. ทำ
14. เงา
15. แซะ
16. แพ
17. ไว
18. เสพ
19. จ๊ะ
20. เอาะ
21. อึ
22. ชาว
23. ยั้วะ
24. รั่ว
25. เสือ
26. เบียร์
27. หอ
28. เอียะ
29. เอือะ
30. ป่า

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก ง.

แบบสอบถามที่ใช้ในการทดสอบการฟัง

แบบสอบถามเพื่อวิเคราะห์ในงานวิจัยเรื่อง
การสังเคราะห์เสียงภาษาไทยโดยวิธีสังเคราะห์แบบฟอร์แมนท์
ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์
สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ชื่อ.....

จงเขียนคำที่ท่านได้ยิน

1. .....
2. .....
3. .....
4. .....
5. .....
6. .....
7. .....
8. .....
9. .....
10. .....
11. .....
12. .....
13. .....
14. .....
15. .....
16. .....
17. .....
18. .....
19. .....
20. .....
21. .....
22. .....
23. .....
24. .....
25. .....
26. .....
27. .....
28. .....
29. .....
30. .....

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก จ.

ข้อมูลความถี่ฟอร์แมนท์

หน่วยเสียง	ความถี่ฟอร์แมนท์			
	ช่วงเวลา (msec)	F1 (Hz)	F2 (Hz)	F3 (Hz)
/ก/	20	420	1,477	2,030
/ค/	60	500	1,220	1,900
/ง/	200	345	1,169	2,088
/จ/	50	320	2,200	2,830
	370	310	2,020	2,960
/ช/	80	803	1,860	2,861
	40	803	1,260	2,561
/ซ/	170	303	1,860	3,261
/ด/	40	253	1,460	2,561
/ต/	30	403	1,460	2,561
/ท/	10	253	1,260	2,686
/น/	100	360	1,170	2,350
/บ/	25	227	1,261	2,373
	25	204	1,152	2,354
/ป/	10	340	1,100	2,080
	10	440	900	2,080
/พ/	40	103	660	2,561
/ฟ/	160	330	1,233	3,254
/ม/	100	250	821	2,019
/ย/	100	311	2,070	2,920
/ร/	20	488	1,185	3,216
	20	488	1,185	3,685
	20	488	1,188	1,685

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

หน่วยเสียง	ความถี่ฟอร์แมนท์			
	ช่วงเวลา (msec)	F1 (Hz)	F2 (Hz)	F3 (Hz)
/ล/	50	310	1,250	2,880
/ว/	100	220	610	2,180
/ย/	60	400	620	2,645
/อ/	70	350	1,320	2,624

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ประวัติผู้เขียน

ชื่อ-นามสกุล	นายสายยัญ สายยศ
วัน เดือน ปี	15 มีนาคม 2520 ที่สุรินทร์
ที่อยู่	30 หมู่1 ต.บึง กิ่ง อ.เขวาสินรินทร์ จ.สุรินทร์ 32000 โทร.0-9620-4227
ประวัติการศึกษา	2543 วิทยาศาสตรบัณฑิต สาขาวิทยาการคอมพิวเตอร์ มหาวิทยาลัยมหาสารคาม

ประสบการณ์การทำงานและผลงานวิจัย

พ.ศ.2542-2543	ตำแหน่งอาจารย์ผู้ช่วยสอน ภาควิชาคอมพิวเตอร์ คณะวิศวกรรมศาสตร์และวิทยาการคอมพิวเตอร์ มหาวิทยาลัยมหาสารคาม - สอนในรายวิชาคอมพิวเตอร์โปรแกรมมิ่ง
พ.ศ.2543-2544	ตำแหน่งโปรแกรมเมอร์บริษัทอัลฟ่าออฟฟิชออโตเมชั่น จำกัด - เขียนโปรแกรมระบบนวสาร - ดูแลระบบคอมพิวเตอร์โรงงาน บริษัทไทยเปอร์ออกไซด์ จำกัด - ติดตั้งและดูแลระบบคอมพิวเตอร์ของลูกค้า รับราชการทหารกองประจำการ ศูนย์ฝึกทหารใหม่ กรมยุทธศึกษาทหารเรือ กองทัพเรือสัตหีบ - เป็นหัวหน้าหมวดที่ 3 กองร้อยที่ 4 กองพันที่ 2 - เขียนโปรแกรมระบบจ่ายเงินเดือนทหารใหม่ กองร้อยที่ 4 กองพันที่ 2
พ.ศ.2544-2545	ตำแหน่งเจ้าหน้าที่คอมพิวเตอร์ ภัณฑุปกรณ์ ศูนย์ฝึกทหารใหม่ กรมยุทธศึกษาทหารเรือ กองทัพเรือสัตหีบ - ติดตั้งระบบ Inventory - เขียนโปรแกรมระบบงานภัณฑุปกรณ์ ศูนย์ฝึกทหารใหม่
พ.ศ.2545-2546	ตำแหน่งผู้ช่วยนักวิจัย คณะวิทยาการสารสนเทศ มหาวิทยาลัยมหาสารคาม - วิจัยเรื่อง ระบบสารสนเทศชุมชน - ออกแบบและเขียนโปรแกรมระบบการคลังองค์การบริหารส่วนตำบล - ออกแบบและเขียนโปรแกรมสารสนเทศชุมชน - ออกแบบและเขียนโปรแกรมระบบ e-learning - วิทยากรอบรมคอมพิวเตอร์ให้แก่ 35 องค์การบริหารส่วนตำบล 3 จังหวัด คือ ร้อยเอ็ด กาฬสินธุ์ และมหาสารคาม

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

พ.ศ.2546-2547 ตำแหน่งโปรแกรมเมอร์ บริษัท เอมไอเอส โซลูชั่น จำกัด

- พัฒนาและติดตั้งระบบภายองค์การบริหารส่วนตำบล
- ออกแบบและเขียนโปรแกรมระบบแผนที่ภายองค์การบริหารส่วนท้องถิ่น

พ.ศ.2547-ปัจจุบัน ตำแหน่งวิศวกร บริษัท ที.เอส.ที. คอมเมอร์เชียล จำกัด

- ออกแบบและเขียนโปรแกรมระบบรักษาความปลอดภัย ธนาคารกสิกรไทย จำกัด (มหาชน)
- ติดตั้งและดูแลระบบรักษาความปลอดภัย ธนาคารกสิกรไทย จำกัด (มหาชน)
- ติดตั้งและดูแลระบบรักษาความปลอดภัย ศูนย์เงินสด ธนาคารกสิกรไทย จำกัด (มหาชน)
- ติดตั้งและดูแลระบบรักษาความปลอดภัย ธนาคารกรุงเทพ
- ติดตั้งและดูแลระบบรักษาความปลอดภัย ธนาคารอาคารสงเคราะห์
- ติดตั้งและดูแลระบบDVR โรงงานบริษัทดัชมิลล์ จำกัด
- ดูแลระบบเตือนภัย โรงกลั่นบริษัทไทยออยล์
- ดูแลระบบรักษาความปลอดภัย บริษัทเทสโกโลตัส จำกัด
- ออกแบบและเขียนโปรแกรมระบบบัตรคิวทัชสกรีน บริษัทกรุงไทยบัตรเครดิต จำกัด
- ติดตั้งและดูแลระบบ DVR บริษัทกรุงไทยบัตรเครดิต จำกัด
- ออกแบบและเขียนโปรแกรมระบบสารสนเทศ บริษัท ที.เอส.ที. คอมเมอร์เชียล จำกัด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้