



การทำนายค่าความถี่มูลฐานสำหรับการสังเคราะห์เสียงพูดภาษาไทย

โดย

เปรมวดี อุดมเรืองเกียรติ

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร  
วิทยาศาสตรมหาบัณฑิต  
สาขาวิชาวิทยาการคอมพิวเตอร์ ภาควิชาวิทยาการคอมพิวเตอร์  
คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์  
พ.ศ. 2550

การทํานายค่าความถี่มูลฐานสำหรับการสังเคราะห์เสียงพูดภาษาไทย

โดย

เปรมวดี อุดมเรืองเกียรติ

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร

วิทยาศาสตรมหาบัณฑิต

สาขาวิชาวิทยาการคอมพิวเตอร์ ภาควิชาวิทยาการคอมพิวเตอร์

คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์

พ.ศ. 2550

# F0 Prediction for Thai Speech Synthesis

By

Premvadee Udomruangkiat

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร  
วิทยาศาสตรมหาบัณฑิต  
สาขาวิชาวิทยาการคอมพิวเตอร์ ภาควิชาวิทยาการคอมพิวเตอร์  
คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์  
พ.ศ. 2550

มหาวิทยาลัยธรรมศาสตร์  
คณะวิทยาศาสตร์และเทคโนโลยี

วิทยานิพนธ์

ของ

เปรมวดี อุดมเรืองเกียรติ

เรื่อง

การทํานายค่าความถี่มูลฐานสำหรับการสังเคราะห์เสียงพูดภาษาไทย

ได้รับการตรวจสอบและอนุมัติ ให้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร

วิทยาศาสตร์มหาบัณฑิต

เมื่อ วันที่ 5 พฤษภาคม พ.ศ. 2551

ประธานกรรมการสอบวิทยานิพนธ์

\_\_\_\_\_  
(รองศาสตราจารย์ ภาวดี สมภักดี)

กรรมการและอาจารย์ที่ปรึกษาวิทยานิพนธ์

\_\_\_\_\_  
(ดร.รัชฎา คงคะจันทร์)

กรรมการสอบวิทยานิพนธ์

\_\_\_\_\_  
(ผู้ช่วยศาสตราจารย์ ดร.สุกรี สิ้นธุภิณฺโญ)

กรรมการสอบวิทยานิพนธ์

\_\_\_\_\_  
(ดร.ชัย วุฒิวิวัฒน์ชัย)

คณบดี

\_\_\_\_\_  
(รองศาสตราจารย์ สมชาย วิริยะยุทธกร)

## บทคัดย่อ

การสร้างการสังเคราะห์เสียงพูดให้มีความเป็นธรรมชาติสูงนั้น จะต้องมีส่วนของสัทสัมพันธ์ (Prosody) เป็นองค์ประกอบหลัก ค่าความถี่มูลฐานก็เป็นหนึ่งในส่วนประกอบที่สำคัญของสัทสัมพันธ์เสียงพูดความแตกต่างของค่าความถี่มูลฐานในคำพูดต่อเนื่องประโยคภาษาไทย มีผลต่อการควบคุมสัทสัมพันธ์ในระบบการสังเคราะห์เสียงภาษาไทย การทดลองนี้เลือกใช้ เทคนิคต้นไม้ตัดสินใจ (Decision Tree) ชื่อว่า Classification and Regression Trees (CART) มาช่วยในการสร้างแบบจำลองความถี่มูลฐาน และได้มีการประยุกต์เทคนิคนี้ เพื่อทำการเปรียบเทียบผลกับเทคนิคต้นไม้ตัดสินใจอื่น เช่น C4.5 และ เทคนิคข่ายงานประสาทเทียมแบบแพร่กระจายย้อนกลับหลายชั้น (Multilayer Backpropagation Neural Network) โดยการทดลองได้ทำกับคลังข้อมูลเสียงภาษาไทย ชื่อว่า 'TSynC1' ซึ่งถูกพัฒนาโดย ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ (NECTEC) ประโยคคำพูด 5,200 ประโยคถูกใช้ในการเรียนรู้และทดสอบระบบ ในอัตรา 90 ต่อ 10 ของจำนวนประโยคคำพูดในคลังข้อมูลเสียง ผลการทดลองแสดงให้เห็นว่าเทคนิคต้นไม้ตัดสินใจสามารถทำนายค่าความถี่มูลฐานได้ดีกว่าเทคนิคข่ายงานประสาทเทียม

## Abstract

For the generation of highly natural synthetic speech, the control of prosody is of primary importance. The fundamental frequency (F0) is one of the most important components of speech prosody. The variation of F0 in continuous Thai speech, with the goal of establishing an effective mechanism of prosody control in Thai text-to-speech (TTS) application. In this thesis, Classification And Regression Tree (CART) is employed for modeling the fundamental frequency (F0). This applied technique is compared with the well-known decision tree as C4.5 and the popular neural network as Multilayer Back-Propagation Neural Network (BPNN). The evaluation is done with Thai speech corpus, named TSYNC1. This corpus is developed by National Electronics and Computer Technology Center (NECTEC). Five thousands and two hundreds sentences of speech data are used for training and testing with the proportion as 90:10. The result shows that modeling F0 by decision tree technique is has higher performance than neural network technique.

## กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลงได้ด้วยความช่วยเหลือของ อาจารย์ ดร.รัชฎา คงคะจันทร์ อาจารย์ที่ปรึกษาวิทยานิพนธ์ที่ได้กรุณาให้คำปรึกษาแนะนำเป็นอย่างดี ผู้วิจัยขอขอบพระคุณอย่างจริงใจในความช่วยเหลือ ความกรุณาของท่านซึ่งได้สละเวลาอันมีค่าแก่ผู้วิจัยมาโดยตลอด และ ผู้วิจัย ขอขอบคุณ ดร.ชัย วุฒิวิวัฒน์ชัย ที่ได้ช่วยเหลือและให้ชี้แนะรายละเอียดเนื้อหาวิชาการที่มีประโยชน์ ช่วยให้งานวิจัยสำเร็จลงด้วยดีอย่างมีคุณภาพ ท้ายสุดผู้วิจัยขอขอบคุณคณะกรรมการสอบวิทยานิพนธ์ทุกท่าน ที่ได้ช่วยตรวจงานวิจัยและได้ให้ข้อเสนอแนะเพิ่มเติม

คุณความดีใดๆที่พึงมีในงานวิจัยนี้เป็นของคุณพ่อ คุณแม่ ครู อาจารย์ บุคคลและกลุ่มบุคคลที่ได้เอื้อนามข้างต้น

เปรมวดี อุดมเรืองเกียรติ

นักศึกษาระดับบัณฑิตศึกษา

ภาควิชาวิทยาการคอมพิวเตอร์ มหาวิทยาลัยธรรมศาสตร์

พ.ศ. 2551

## สารบัญ

|                                                 | หน้าที่ |
|-------------------------------------------------|---------|
| บทคัดย่อ                                        | (2)     |
| กิตติกรรมประกาศ                                 | (4)     |
| สารบัญตาราง                                     | (8)     |
| สารบัญภาพประกอบ                                 | (9)     |
| <br>บทที่                                       |         |
| 1. บทนำ .....                                   | 1       |
| 1.1 ความเป็นมาและความสำคัญของงานวิจัย .....     | 1       |
| 1.2 วัตถุประสงค์ของงานวิจัย .....               | 3       |
| 1.3 สมมติฐานของงานวิจัย .....                   | 3       |
| 1.4 ขอบเขตงานวิจัย .....                        | 3       |
| 1.5 ผลที่ได้รับจากงานวิจัย .....                | 4       |
| 1.6 รายละเอียดของวิทยานิพนธ์ .....              | 5       |
| <br>2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง .....      | 6       |
| 2.1 ทฤษฎีระบบเสียงภาษาไทย .....                 | 6       |
| 2.1.1 ลักษณะเสียงพูดในภาษาไทย .....             | 6       |
| 2.1.2 ระบบเสียงของภาษาไทยมาตรฐาน .....          | 7       |
| 2.1.3 เสียงครึ่งพยางค์ (Demisyllable) .....     | 11      |
| 2.1.4 โครงสร้างพยางค์ .....                     | 12      |
| 2.1.5 ภาษา ( Language ) .....                   | 15      |
| 2.2 ทฤษฎีระบบสังเคราะห์เสียงพูด .....           | 17      |
| 2.2.1 การวิเคราะห์ข้อความ (Text Analysis) ..... | 18      |



|       |                                                               |    |
|-------|---------------------------------------------------------------|----|
| 2.2.2 | การวิเคราะห์สัทสัมพันธ์ (Prosody Analysis) .....              | 19 |
| 2.2.3 | การสังเคราะห์สัญญาณเสียงพูด (Speech synthesis) .....          | 19 |
| 2.3   | ทฤษฎีการโมเดลทำนองเสียง .....                                 | 23 |
| 2.4   | งานวิจัยที่เกี่ยวข้อง .....                                   | 26 |
| 3.    | วิธีการดำเนินการวิจัย .....                                   | 29 |
| 3.1   | ข้อมูลที่ใช้ในการวิจัย .....                                  | 29 |
| 3.2   | เครื่องมือในการเรียนรู้ระบบ .....                             | 29 |
| 3.2.1 | การเรียนรู้ต้นไม้ตัดสินใจ (Decision Tree Learning) .....      | 30 |
| 3.2.2 | ข่ายงานประสาทเทียม (Neural Network) .....                     | 32 |
| 3.3   | ขั้นตอนในการดำเนินการ .....                                   | 33 |
| 3.4   | ระยะเวลาในการดำเนินการวิจัย .....                             | 33 |
| 4.    | การทดลองและผลการทดลอง .....                                   | 34 |
| 4.1   | วิธีการทดลอง .....                                            | 34 |
| 4.1.1 | ข้อมูลนำเข้า .....                                            | 35 |
| 4.1.2 | ขั้นตอนการทดลอง.....                                          | 35 |
| 4.1.3 | การประเมินผลลัพธ์.....                                        | 40 |
| 4.2   | การคำนวณค่าผลการทดลอง.....                                    | 41 |
| 4.2.1 | การวิเคราะห์สหสัมพันธ์ (Correlation Analysis).....            | 41 |
| 4.2.2 | ค่าความผิดพลาดกำลังสองเฉลี่ย (Root Mean Square Error, RMSE).. | 41 |
| 4.3   | ผลการทดลอง .....                                              | 42 |
| 5.    | สรุปผลการวิจัยและข้อเสนอแนะ .....                             | 45 |
| 5.1   | สรุปผลการวิจัย .....                                          | 45 |
| 5.1.1 | สรุปผลการทดสอบสมมติฐานวิจัยข้อที่ 1.....                      | 47 |
| 5.1.2 | สรุปผลการทดสอบสมมติฐานวิจัยข้อที่ 2.....                      | 47 |
| 5.2   | ข้อเสนอแนะ .....                                              | 49 |

|                       |    |
|-----------------------|----|
| ภาคผนวก .....         | 50 |
| บรรณานุกรม.....       | 63 |
| ประวัติการศึกษา ..... | 65 |

## สารบัญตาราง

| ตารางที่                                                             | หน้า |
|----------------------------------------------------------------------|------|
| 2.1 สัญลักษณ์เสียงของสระไทย.....                                     | 7    |
| 2.2 สัญลักษณ์เสียงของพยัญชนะไทย .....                                | 8    |
| 2.3 สัญลักษณ์เสียงของพยัญชนะควบกล้ำ .....                            | 9    |
| 2.4 เปรียบเทียบงานวิจัยที่เกี่ยวข้อง .....                           | 26   |
| 3.1 แผนการดำเนินงาน.....                                             | 33   |
| 4.1 เปรียบเทียบผลการทดลองระหว่างค่าที่ทำนาย 2 จุดในพยางค์.....       | 42   |
| 4.2 ผลการทดลองระหว่างระบบสังเคราะห์เสียงพูดจากและค่าที่ทำนายได้..... | 43   |
| 4.3 เปรียบเทียบผลการทดลองในแต่ละวิธีการเรียนรู้ .....                | 44   |

## สารบัญภาพประกอบ

| ภาพที่                                                         | หน้าที่ |
|----------------------------------------------------------------|---------|
| 2.1 ระดับเสียงสูงต่ำในคำภาษาไทย .....                          | 10      |
| 2.2 แสดงโครงสร้างหน่วยเสียงครึ่งพยางค์ .....                   | 12      |
| 2.3 ภาษาวรรณยุกต์ .....                                        | 15      |
| 2.4 ภาษาทำนองเสียง .....                                       | 16      |
| 2.5 ขั้นตอนการสังเคราะห์เสียง .....                            | 18      |
| 2.6 จำแนกประเภทการสังเคราะห์เสียง .....                        | 21      |
| 2.7 การสังเคราะห์เสียงแบบเชื่อมต่อหน่วยเสียง .....             | 23      |
| 2.8 เปรียบเทียบประสิทธิภาพการทำนายค่าความถี่มูลฐาน .....       | 24      |
| 2.9 เปรียบเทียบเทคนิคการเรียนรู้ของเครื่องจักร .....           | 25      |
| 3.1 โครงสร้างต้นไม้เพื่อคาดเดาว่าจะเล่นกีฬาเทนนิสหรือไม่ ..... | 31      |
| 3.2 การทดลองโดยใช้เครื่องมือต้นไม้ตัดสินใจ .....               | 32      |
| 4.1 แผนภาพขั้นตอนการทดลอง .....                                | 35      |
| 4.2 การทำนายค่าความถี่มูลฐาน .....                             | 37      |
| 4.3 การเชื่อมจุดที่ทำนายค่าความถี่มูลฐาน .....                 | 38      |
| 4.4 ขั้นตอนการเรียนรู้และทดสอบระบบ .....                       | 39      |
| 4.5 การทำงานของระบบ .....                                      | 40      |
| ก.1 แสดงขอบเขตข้อมูลในคลังข้อมูล“TSynC1” .....                 | 51      |
| ข.1 แสดงขอบเขตข้อมูลในโปรแกรม“Wagon” .....                     | 59      |
| ข.2 แสดงรายละเอียดการใช้งานโปรแกรม“Wagon” .....                | 59      |
| ข.3 แสดงการทดสอบโปรแกรม .....                                  | 60      |
| ข.4 แสดงการฝึกฝนโปรแกรม .....                                  | 60      |
| ข.5 ตัวอย่างการอธิบายFeatureของชุดข้อมูล .....                 | 61      |
| ข.6 ตัวอย่างข้อมูลสำหรับTrain .....                            | 61      |
| ข.7 ตัวอย่างข้อมูลสำหรับTest .....                             | 62      |
| ข.8 ตัวอย่างส่วนแสดงผลลัพธ์ .....                              | 62      |

## บทที่ 1

### บทนำ

#### 1.1 ความเป็นมาและความสำคัญของงานวิจัย

ปัจจุบันเทคโนโลยีทางการสังเคราะห์เสียงพูด (Speech Synthesis) ได้เข้ามามีบทบาทต่อชีวิตประจำวันของเรามากขึ้นเรื่อยๆ โดยระบบการสังเคราะห์เสียงพูด จะช่วยแปลงข้อความ ให้เป็นสัญญาณเสียงพูดซึ่งมนุษย์สามารถเข้าใจได้ การพัฒนาระบบการสังเคราะห์เสียงพูดจะมีประโยชน์ในหลายๆด้าน ไม่ว่าจะเป็นทางการเรียนรู้ภาษา การใช้งานกับผู้บกพร่องทางการมองเห็น การใช้งานกับสื่อการเรียนรู้อื่นๆ

ในการเรียนรู้ทางภาษา ระบบสังเคราะห์เสียงพูดภาษาไทย สามารถใช้เป็นต้นแบบในการศึกษาภาษาไทย ช่วยอำนวยความสะดวกต่อการเรียนรู้ ทำให้คนรู้หนังสือเพิ่มมากขึ้นในการใช้งานกับผู้บกพร่องทางการมองเห็น จะช่วยให้เกิดการรับรู้ข่าวสาร และความรู้ง่ายขึ้น เช่น โปรแกรมช่วยอ่านสำหรับผู้พิการทางสายตา โปรแกรมอ่านจดหมายอิเล็กทรอนิกส์ผ่านทางโทรศัพท์ นอกจากนี้ ระบบยังมีประโยชน์ในการประยุกต์ใช้กับงานอื่นๆอีก เช่น พจนานุกรมพูดได้ สำหรับผู้ที่ศึกษาภาษาไทย การประยุกต์ใช้เทคนิคการสังเคราะห์เสียงเพื่อให้ผู้พิการทางสายตาสามารถใช้คอมพิวเตอร์ ซึ่งจะทำให้เกิดประโยชน์กับคนโดยส่วนใหญ่ แม้กระทั่งคนไม่รู้หนังสือ คนไม่ได้รับการศึกษา ช่วยลดความเหลื่อมล้ำทางสังคม ทำให้คนสามารถดำรงชีวิตอย่างมีคุณค่า อีกทั้งยังทำให้เกิดความตื่นตัวในการพัฒนาเทคโนโลยีมีเดีย สื่อในการเรียนรู้อื่นๆ ทำให้การเรียนรู้ การรับข่าวสารข้อมูลน่าสนใจมากขึ้น เกิดความก้าวหน้าทางวิชาการด้านเทคโนโลยีการสังเคราะห์เสียงพูดภาษาไทย สามารถนำไปใช้เป็นแนวทางการศึกษา และประยุกต์ใช้กับงานวิจัยอื่นที่เกี่ยวข้อง เช่น การรู้จำเสียงพูด (Speech Recognition) ซึ่งเป็นการช่วยส่งเสริมความเจริญก้าวหน้าให้กับประเทศ อันจะนำไปสู่การพัฒนาเศรษฐกิจของประเทศต่อไป

ระบบสังเคราะห์เสียงพูดจากข้อความนี้ส่วนใหญ่จะถูกพัฒนาสำหรับภาษาหลักๆ เช่น ภาษาอังกฤษ, จีน, ญี่ปุ่น และภาษาทางยุโรป เป็นต้น สำหรับภาษาไทยยังมีเป็นส่วนน้อย และขาดการพัฒนาที่ต่อเนื่อง งานวิจัยระบบการสังเคราะห์เสียง เป็นงานที่ยาก ต้องใช้ความรู้ ความเชี่ยวชาญ อีกทั้งยังต้องใช้กำลังคน กำลังแรงงาน และเวลาอย่างมากในการพัฒนาระบบ

การที่จะทำให้ระบบสังเคราะห์เสียงพูดภาษาไทย สามารถใช้งานได้จริงนั้น จะต้องทำการพัฒนาระบบการสังเคราะห์เสียงให้มีความเป็นธรรมชาติ (Naturalness) ใกล้เคียงกับเสียงพูดของมนุษย์ ฟังแล้วเข้าใจในการสื่อสารความหมายของประโยคได้อย่างชัดเจน (Intelligibility) ซึ่งระบบนั้นจะต้องมีการพัฒนาในส่วนของสัทสัมพันธ์ (Prosody)

ในการพัฒนาทางด้านสัทสัมพันธ์ (Prosody) จะช่วยให้ระบบสังเคราะห์เสียงมีความสัมพันธ์ในระดับเหนือหน่วยเสียง ทำให้เสียงมีความเป็นธรรมชาติ มีทำนองเสียงใกล้เคียงเสียงใกล้เคียงเสียงพูดของมนุษย์ ในการวิเคราะห์และสังเคราะห์ข้อมูลสัทสัมพันธ์ (Prosody) ของประโยคนั้น จะเกี่ยวข้องกับ การกำหนดช่วงเวลาของพยางค์ ความยาวของเสียงย่อย ความถี่มูลฐาน (F0 Contour) ของหน่วยเสียง ซึ่งจะมีผลต่อระดับเสียงสูงต่ำของประโยค (Intonation)

งานวิจัยทางด้านการสังเคราะห์เสียง วิธีที่นิยมกันจะทำการทดลองกับคลังข้อมูลเสียงขนาดใหญ่ โดยการฝึกฝนระบบจากการเรียนรู้ตัวอย่างประโยคภาษาพูดโดยตรง จะทำให้ระบบมีความฉลาดที่จะสร้างเสียงพูดที่มีความชัดเจนมีความเป็นธรรมชาติใกล้เคียงเสียงมนุษย์ แต่เนื่องจากภาษาไทยยังมีคลังข้อมูลเสียงที่จำกัด เนื่องจากต้องใช้ทรัพยากร และเวลาอย่างมากในการพัฒนา ทำให้มีความจำเป็นที่จะต้องพัฒนาเทคนิคเพื่อมาปรับใช้กับระบบการสังเคราะห์เสียงภาษาไทย

ดังนั้น วิทยานิพนธ์นี้จึงจัดทำขึ้นเพื่อหาเทคนิคใหม่ในการปรับปรุงระบบสังเคราะห์เสียงให้มีความเป็นธรรมชาติใกล้เคียงเสียงมนุษย์มากขึ้น ด้วยวิธีการทำนายค่าความถี่มูลฐานเพื่อใช้ในการสังเคราะห์เสียงพูดภาษาไทย โดยคำในภาษาไทยมีองค์ประกอบ 3 ส่วน คือ พยัญชนะ สระ และ วรรณยุกต์ โดยเสียงในภาษาไทยประกอบด้วยพยัญชนะ 21 หน่วยเสียง สระ 24 หน่วยเสียง และวรรณยุกต์ 5 หน่วยเสียง

งานวิจัยนี้ได้ทำการทดลองกับคลังข้อมูลเสียงขนาดใหญ่ ซึ่งถูกพัฒนาโดยศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ (NECTEC) ในฐานข้อมูลเสียงนี้ประกอบไปด้วยประโยค 5,200 ประโยค ซึ่งครอบคลุมเสียงทั้งคำไทย และคำที่ยืมมาจากภาษาต่างประเทศ โดยการทดลองได้นำเทคนิคทางด้านปัญญาประดิษฐ์ (Artificial Intelligence) มาปรับใช้ในการเรียนรู้ระบบ ซึ่งในการสร้างระบบทำนายจะใช้ เทคนิคต้นไม้ตัดสินใจ (Decision Tree) ฝึกฝนร่วมกับข้อมูลอินพุตทางด้านภาษา ในการทำนายค่าความถี่มูลฐาน (F0 Contour) ของหน่วยเสียงได้ใกล้เคียงกับเสียงพูดของมนุษย์ ซึ่งจะทำให้ระบบสังเคราะห์เสียงสามารถสร้างเสียงที่มีระดับเสียงสูงต่ำของประโยค (Intonation) มีความเป็นธรรมชาติมากขึ้น

## 1.2 วัตถุประสงค์ของงานวิจัย

งานวิจัยนี้มีเป้าหมายเพื่อทำให้ระบบสังเคราะห์เสียงพูดภาษาไทย สามารถผลิตเสียงพูดให้มีทำนองเสียงที่เป็นธรรมชาติมากขึ้น โดยการออกแบบการทำนายค่าความถี่มูลฐานสำหรับการสังเคราะห์เสียงพูดภาษาไทย และทำการทดลองกับคลังข้อมูลเสียงขนาดใหญ่ แล้วใช้เทคนิคทางด้านปัญญาประดิษฐ์เข้าช่วยในการเรียนรู้ และทดสอบระบบ ซึ่งจะช่วยให้ประสิทธิภาพในการสังเคราะห์เสียง ทั้งนี้ เพื่อให้งานวิจัยดำเนินไปตามเป้าหมายดังกล่าว จึงได้กำหนดวัตถุประสงค์ของการวิจัยไว้ ดังนี้

- 1.2.1 เพื่อศึกษาระบบเสียงภาษาไทย
- 1.2.2 เพื่อศึกษาระบบการสังเคราะห์เสียงพูดและการโมเดลทำนองเสียง
- 1.2.3 เพื่อศึกษาเทคนิคทางด้านปัญญาประดิษฐ์ที่เหมาะสมต่อการสังเคราะห์ทำนองเสียง
- 1.2.4 เพื่อออกแบบการทำนายค่าความถี่มูลฐานสำหรับการสังเคราะห์เสียงพูดภาษาไทย

## 1.3 สมมติฐานของงานวิจัย

- 1.3.1 ข้อมูลความถี่มูลฐานจากคลังข้อมูลเสียง สามารถใช้ทำนายค่าความถี่มูลฐานสำหรับการสังเคราะห์เสียงพูดภาษาไทย
- 1.3.2 เทคนิคทางด้านปัญญาประดิษฐ์ที่ใช้โดยการเรียนรู้ต้นไม้ตัดสินใจ (Decision Tree) ในการทำนายค่าความถี่มูลฐานมีความถูกต้องมากกว่าเทคนิคข่ายงานประสาทเทียม (Neural Network)

## 1.4 ขอบเขตงานวิจัย

เพื่อเป็นกรอบในการดำเนินการวิจัย ให้สามารถดำเนินตามประเด็นที่สำคัญของวัตถุประสงค์ดังกล่าว งานวิจัยจึงได้กำหนดขอบเขตของงานวิจัยไว้ดังต่อไปนี้

- 1.4.1 ข้อมูลนำเข้า  
ประโยคจากคลังข้อมูลเสียงขนาดใหญ่ (Corpus) ชื่อว่า "TSynC1" เพื่อใช้ในการสังเคราะห์สัญญาณเสียงพูด ซึ่งในฐานข้อมูลเสียงนี้ประกอบด้วยคำ 5,200 คำ ซึ่งครอบคลุมหมด

ทุกเสียงทั้งคำไทย และคำที่ยืมมาจากภาษาต่างประเทศโดยจะแบ่งเป็นข้อมูลที่ใช้ในการทดลอง ดังนี้

1.4.1.1 ข้อมูลใช้ในการฝึกฝน (Training Data) 90%

1.4.1.2 ข้อมูลใช้ในการทดสอบ (Testing Data) 10%

#### 1.4.2 ขั้นตอนดำเนินการ

ในการสังเคราะห์เสียง จะทำการสังเคราะห์เสียงทีละประโยค ซึ่งการสังเคราะห์เสียงนี้ จะใช้หลักของการเชื่อมต่อแต่ละหน่วยเสียงย่อย (Concatenation Synthesis)

ในการสังเคราะห์เสียง เพื่อที่จะให้คอมพิวเตอร์รู้จักรูปแบบของเสียง ซึ่งในวิทยานิพนธ์นี้ ใช้การฝึกฝนแบบมีผู้สอน (Supervised Learning) โดยใช้เทคนิคทางด้านปัญญาประดิษฐ์ คือ เทคนิคต้นไม้ตัดสินใจ (Decision Tree) มาช่วยในการเรียนรู้

ในวิทยานิพนธ์นี้ จะใช้ระบบการสังเคราะห์เสียงพูดภาษาไทย “วาจา” (Vaja) เป็นเครื่องมือในการสังเคราะห์เสียง

#### 1.4.3 การประเมินผลลัพธ์

การวิจัยนี้ได้ทำการออกแบบการประเมินผลการทดลองอย่างสมบูรณ์ในเชิงปริมาณ โดยมีรายละเอียดดังนี้

การทดลองนี้ได้ทำการออกแบบการประเมินผลลัพธ์โดยการวิเคราะห์สหสัมพันธ์ (Correlation Analysis) ระหว่างทำนองเสียงที่สังเคราะห์ออกมา กับทำนองเสียงของเสียงต้นฉบับ

อีกทั้งยังได้ทำการเปรียบเทียบผลการทดลองในการทำนายค่าความถี่มูลฐาน กับระบบการสังเคราะห์เสียงพูดภาษาไทย “วาจา” (Vaja) ซึ่งเป็นสังเคราะห์เสียงพูดภาษาไทยที่มีการใช้งานในปัจจุบัน เพื่อวัดประสิทธิภาพในการใช้งาน

### 1.5 ผลที่ได้รับจากงานวิจัย

เมื่อได้ดำเนินการวิจัยตามวัตถุประสงค์ ภายใต้ขอบเขตของงานวิจัยที่ได้กำหนดไว้ ผู้วิจัยมุ่งหวังว่าจะได้รับประโยชน์ดังนี้

1.5.1 เข้าใจระบบเสียงภาษาไทย

1.5.2 เข้าใจระบบการสังเคราะห์เสียงพูดและการโมเดลทำนองเสียง

1.5.3 สามารถนำเทคนิคด้านปัญญาประดิษฐ์ที่เหมาะสมต่อการสังเคราะห์เสียงมาประยุกต์ใช้



1.5.4 สามารถทำนายค่าความถี่มูลฐานสำหรับการสังเคราะห์ เสียงพูดภาษาไทย และนำผลลัพธ์ที่ได้จากงานวิจัยไปพัฒนาเป็นงานวิจัยอื่นๆต่อไป

## 1.6 รายละเอียดของวิทยานิพนธ์

รายงานนำเสนอวิทยานิพนธ์ฉบับนี้ ได้แบ่งเนื้อหาของวิทยานิพนธ์เป็น 4 ส่วนหลักคือ บทนำ ทฤษฎีและงานวิจัยที่เกี่ยวข้อง วิธีดำเนินการวิจัย การทดลองและผลการทดลอง รวมทั้งสรุปผลการวิจัยและข้อเสนอแนะ โดยมีเนื้อหารายละเอียดดังนี้

บทที่ 1 บทนำ ประกอบไปด้วย ความเป็นมาและความสำคัญของงานวิจัย วัตถุประสงค์ของงานวิจัย สมมติฐานของงานวิจัย ขอบเขตงานวิจัย ผลที่ได้รับจากงานวิจัย และรายละเอียดของวิทยานิพนธ์

บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง ประกอบไปด้วย ทฤษฎีระบบเสียงภาษาไทย ทฤษฎีระบบสังเคราะห์เสียงพูด ทฤษฎีการโมเดลทำนองเสียง และงานวิจัยที่เกี่ยวข้อง

บทที่ 3 วิธีดำเนินการวิจัย ประกอบไปด้วย ข้อมูลที่ใช้ในการวิจัย เครื่องมือในการเรียนรู้ระบบ ขั้นตอนในการดำเนินการ และระยะเวลาในการดำเนินการวิจัย

บทที่ 4 การทดลองและผลการทดลอง ประกอบไปด้วย วิธีการทดลอง การคำนวณค่า ผลการทดลอง และ ผลการทดลอง

บทที่ 5 สรุปผลการวิจัยและข้อเสนอแนะ ประกอบไปด้วย สรุปผลการวิจัย และ ข้อเสนอแนะ

## บทที่ 2

### ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

บทนี้จะกล่าวถึง ทฤษฎีและงานวิจัยที่เกี่ยวข้องในวิทยานิพนธ์ฉบับนี้ ทฤษฎีที่นำมาใช้ในการทำวิจัย ประกอบไปด้วย ทฤษฎีระบบเสียงภาษาไทย ทฤษฎีระบบสังเคราะห์เสียงพูด และทฤษฎีการโมเดลทำนองเสียง

### 2.1 ทฤษฎีระบบเสียงภาษาไทย

#### 2.1.1 ลักษณะเสียงพูดในภาษาไทย

เสียงก้องหรือโห่ (Voiced) และ เสียงไม่ก้องหรืออโห่ (Voiceless) ซึ่งเสียงก้องเป็นเสียงที่ตั้ง

เสียงสูง เสียงต่ำ ซึ่งเสียงสูง หรือต่ำนี้ทำให้ความหมายของคำเปลี่ยนไปจัดว่าเป็นหน่วยเสียงที่เราเรียกกันว่า วรรณยุกต์ (Tone) ส่วนเสียงสูงต่ำที่ไม่ผูกพันกับคำ แต่เปลี่ยนไปตามรูปประโยคก็จะทำให้ความหมายเปลี่ยนไปตามทำนองของประโยค เช่นจากประโยคบอกเล่า กลายเป็นประโยคคำถาม เรียกกันว่า ทำนองเสียง

ความยาวของเสียง หมายถึงการที่เสียงใดเสียงหนึ่งเปล่งออกมายาวนานเท่าใด ซึ่งก็มักจะพูดกันแค่สั้น กับ ยาวเท่านั้น ซึ่งการกล่าวเสียงสั้น ยาวนั้น จะกล่าวถึงในเสียงสระ

การลงน้ำหนัก (Stress) คือ การออกเสียงพยางค์ใดพยางค์หนึ่งให้นั่นมากหรือน้อยกว่าพยางค์ที่อยู่ข้างเคียง ซึ่งการลงน้ำหนักนี้มีความสัมพันธ์กับจังหวะในการพูดด้วย

ความดังของเสียง

ช่วงต่อของเสียง มีช่วงต่อแนบชิดในลักษณะของพยางค์ กับช่วงต่อห่าง ที่เกิดขึ้นระหว่างคำหรือประโยค

ทำนองเสียง (Intonation) มนุษย์มีความแตกต่างกับเครื่องดนตรีที่มนุษย์ใช้ท่วงทำนองของเสียงในการสังเคราะห์เสียงด้วย เช่น ประโยคที่ลงท้ายด้วยเสียงสูง หมายถึง คำถาม และประโยคที่ลงท้ายด้วยเสียงต่ำ หมายถึง คำสั่ง เป็นต้น

## 2.1.2 ระบบเสียงของภาษาไทยมาตรฐาน

ระบบเสียงของภาษาไทยมาตรฐาน โดยเฉพาะภาษาที่พูดในกรุงเทพมหานคร และบริเวณรอบชานเมืองหลวง ประกอบด้วยหน่วยเสียงสระ 24 หน่วยเสียง , หน่วยเสียงพยัญชนะ 21 หน่วยเสียง , และหน่วยเสียงวรรณยุกต์ 5 หน่วยเสียง โดยมีรายละเอียดเกี่ยวกับระบบเสียงดังนี้

### 2.1.2.1. หน่วยเสียงสระ (Vowel Phonemes)

ประกอบด้วยหน่วยเสียงสระเดี่ยว (Monophthong Vowels) 18 หน่วยเสียง คือหน่วยเสียงสระสั้น (Short Vowels) 9 หน่วยเสียง หน่วยเสียงสระยาว (Long Vowels) 9 หน่วยเสียง และหน่วยเสียงสระประสม (Diphthong Vowels) 6 หน่วยเสียง ดังนี้

#### ตารางที่ 2.1

สัญลักษณ์เสียงของสระไทย

| Monophthong |     |            |      | Diphthong   |      |            |       | Vowel Letter |      |            |   |
|-------------|-----|------------|------|-------------|------|------------|-------|--------------|------|------------|---|
| Short Vowel |     | Long Vowel |      | Short Vowel |      | Long Vowel |       | Short Vowel  |      | Long Vowel |   |
| ะ           | /a/ | า          | /aː/ | เียะ        | /ia/ | เีย        | /iːa/ | ำ            | /am/ | -          | - |
| ิ           | /i/ | ี          | /iː/ | เือะ        | /va/ | เือ        | /vːa/ | ไ, ใ         | /aj/ | -          | - |
| ุ           | /u/ | ู          | /uː/ | ัวะ         | /ua/ | ัว         | /uːa/ | เา           | /aw/ | -          | - |
| ุ           | /u/ | ู          | /uː/ |             |      |            |       |              |      |            |   |
| ะ           | /e/ | เ          | /eː/ |             |      |            |       |              |      |            |   |
| แะ          | /x/ | แ          | /xː/ |             |      |            |       |              |      |            |   |
| โะ          | /o/ | โ          | /oː/ |             |      |            |       |              |      |            |   |
| เาะ         | /@/ | -อ         | /@ː/ |             |      |            |       |              |      |            |   |
| เอะ         | /#/ | -เ         | /#ː/ |             |      |            |       |              |      |            |   |

ที่มา : Mittrapiyanurak et al., 2000

### 2.1.2.2. หน่วยเสียงพยัญชนะ (Consonant Phonemes)

ภาษาไทยมาตรฐานมีหน่วยเสียงพยัญชนะ 21 หน่วยเสียง มีรายละเอียดดังนี้

ตารางที่ 2.2  
สัญลักษณ์เสียงของพยัญชนะไทย

| Thai Letter      | Phonetic Symbol |       |
|------------------|-----------------|-------|
|                  | Initial         | Final |
| ก                | /k/             | /k/   |
| ข, ข, ค, ค, ฅ    | /kh/            |       |
| ง                | /ng/            | /ng/  |
| จ                | /c/             |       |
| ฉ, ช, ฌ          | /ch/            |       |
| ซ, ศ, ษ, ส       | /s/             |       |
| ญ, ย             | /j/             | /j/   |
| ฎ, ด             | /d/             | /t/   |
| ฏ, ต             | /t/             |       |
| ฐ, ฑ, ฒ, ณ, ฑ, ฒ | /th/            |       |
| ณ, น             | /n/             | /n/   |
| บ                | /b/             | /p/   |
| ป                | /p/             |       |
| พ, ภา, ผ         | /ph/            |       |
| ฟ, ฝ             | /f/             |       |
| ม                | /m/             | /m/   |
| ร                | /r/             |       |
| ล, ฬ             | /l/             |       |
| ว                | /w/             | /w/   |
| ห, ฮ             | /h/             | -     |
| อ                | /ʔ/             | -     |

ที่มา: Mittrapiyanurak et al., 2000

### 2.1.2.3. พยัญชนะควบกล้ำ (Cluster Consonants)

ภาษาไทยมาตรฐานมีหน่วยเสียงพยัญชนะควบกล้ำ 12 หน่วยเสียง ได้แก่ /pl / , /phl / , /kl / , /khl / , /pr / , /phr / , /kr / , /khr / , /tr / , /thr / , /kw / และ /khw / ซึ่งมีรายละเอียดดังนี้

### ตารางที่ 2.3

สัญลักษณ์เสียงของพยัญชนะควบกล้ำ

| Thai Letter | Phonetic Symbol of Consonant Cluster |       | English Letter | Phonetic Symbol of Consonant Cluster |       |
|-------------|--------------------------------------|-------|----------------|--------------------------------------|-------|
|             | Initial                              | Final |                | Initial                              | Final |
| ปร-         | /pr/                                 | -     | br-            | /br/                                 | -     |
| ปล-         | /pl/                                 | -     | bl-            | /bl/                                 | -     |
| ตร-         | /tr/                                 | -     | fr-            | /fr/                                 | -     |
| กร-         | /kr/                                 | -     | fl-            | /fl/                                 | -     |
| กล-         | /kl/                                 | -     | dr-            | /dr/                                 | -     |
| กว-         | /kw/                                 | -     | f-             | -                                    | /f/   |
| พร-, ผร-    | /phr/                                | -     | l-             | -                                    | /l/   |
| พล-, ผล-    | /phl/                                | -     | s-             | -                                    | /s/   |
| ทร-         | /thr/                                | -     | ch-            | -                                    | /ch/  |
| คร-, ขร-    | /khr/                                | -     | -              | -                                    | -     |
| คล-, ขล-    | /khl/                                | -     | -              | -                                    | -     |
| คว-         | /khw/                                | -     | -              | -                                    | -     |

ที่มา : Mittrapiyanurak et al., 2000

#### 2.1.2.4. หน่วยเสียงวรรณยุกต์

ภาษาไทยมาตรฐานมีหน่วยเสียงวรรณยุกต์ทั้งหมด 5 หน่วยเสียง ซึ่งสามารถแบ่งออกได้เป็นวรรณยุกต์ระดับ (Level Tone) และวรรณยุกต์เปลี่ยน (Contour Tone) ดังนี้

วรรณยุกต์ระดับ (Level Tone) มี 3 หน่วยเสียง คือ

- (1) หน่วยเสียงวรรณยุกต์ต่ำ (Low Tone) หรือที่เรียกทั่วไปว่า เสียงเอก
- (2) หน่วยเสียงวรรณยุกต์สามัญ (Middle Tone) หรือที่เรียกทั่วไปว่า เสียงสามัญ
- (3) หน่วยเสียงวรรณยุกต์ตรี (High Tone) หรือที่เรียกทั่วไปว่า เสียงตรี

วรรณยุกต์เปลี่ยน (Contour Tone) มี 2 หน่วยเสียง คือ

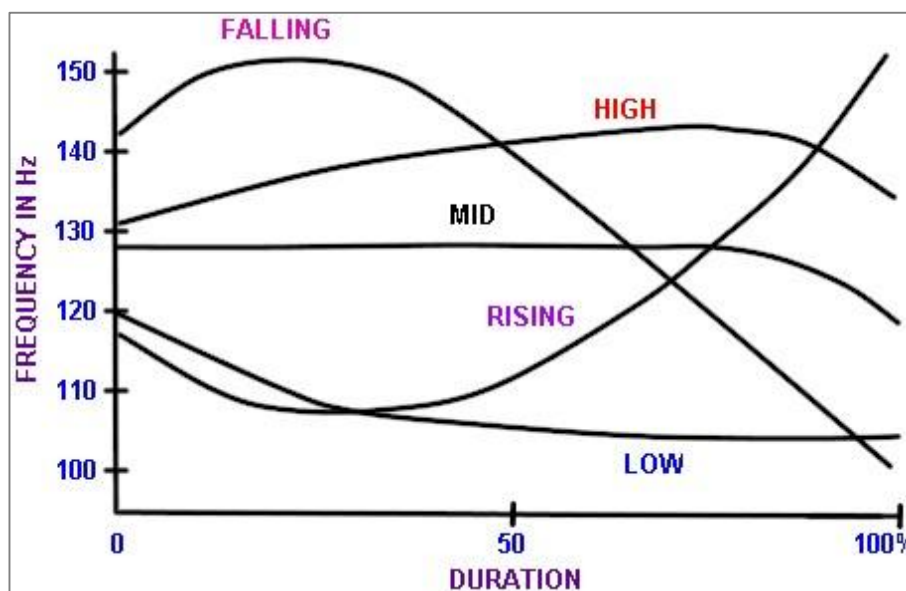
- (4) หน่วยเสียงวรรณยุกต์โท (Falling Tone) หรือที่เรียกทั่วไปว่า เสียงโท
- (5) หน่วยเสียงวรรณยุกต์จัตวา (Rising Tone) หรือที่เรียกทั่วไปว่า เสียงจัตวา

เสียงสูงเสียงต่ำมี 5 ระดับ ที่แสดงในรายการกับตัวเลขภาษาอาหรับ (0-4) [0, เสียงสามัญ], [1, ระดับต่ำ -], ระดับกลางที่เสียงต่ำลง [-, 2], ระดับสูง-[-, 3], และสูงขึ้น [-, 4]

ข้อมูลวรรณยุกต์ (Tone) ของพยางค์เป็นส่วนที่มีลักษณะเฉพาะในภาษาไทย เนื่องจากภาษาไทยเป็นภาษาที่มีระดับเสียงสูงต่ำที่แสดงความหมายของคำ (Tonal Language)

ภาพที่ 2.1

ระดับเสียงสูงต่ำในคำภาษาไทย



ที่มา : Jackson Gandour (1976)

ภาษาไทยมาตรฐานมีหน่วยเสียงวรรณยุกต์ทั้งหมด 5 หน่วยเสียง

- (1) หน่วยเสียงวรรณยุกต์สามัญ (Middle Tone) ป่า
- (2) หน่วยเสียงวรรณยุกต์ต่ำ (Low Tone) ป่า
- (3) หน่วยเสียงวรรณยุกต์โท (Falling Tone) ป่า
- (4) หน่วยเสียงวรรณยุกต์ตรี (High Tone) ป่า
- (5) หน่วยเสียงวรรณยุกต์จัตวา (Rising Tone) ป่า

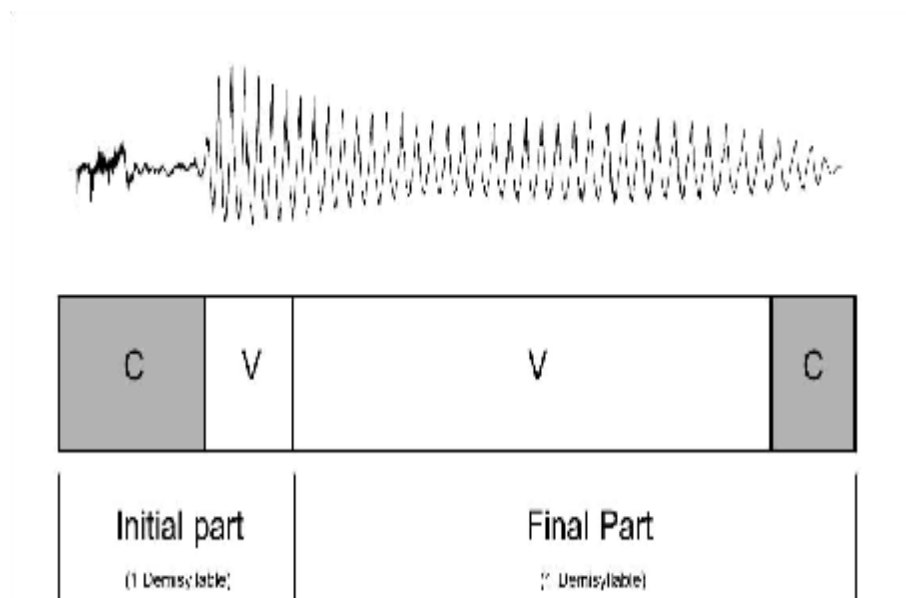
### 2.1.3 เสียงครึ่งพยางค์ (Demisyllable)

เสียงครึ่งพยางค์ คือหน่วยเริ่มต้นพยางค์และหน่วยครึ่งหลังของพยางค์ บนแนวความคิดนี้รูปแบบของคลื่นเสียงถูกสร้างโดยการต่อส่วนของพยางค์ รูปแบบของคลื่นพยางค์ถูกสร้างจากการเริ่มต้นของหน่วยพยางค์และหน่วยของครึ่งพยางค์สุดท้าย หน่วยทั้งสองคือส่วนจากพยางค์ ของเสียงสระคงที่ โดยทั่วไปแล้ว พยางค์ไทย มีโครงสร้างคือ “C (C) VC ” เพราะฉะนั้น พยางค์ถูกแบ่งเป็นส่วนสองส่วนคือ “C (C)V” และ “VC ” หน่วยเริ่มต้นบรรจุเสียงพยัญชนะเดียว หรือ เชื่อมเสียงพยัญชนะซ้อนกับเสียงสระ หน่วยสุดท้ายเชื่อมเสียงสระกับเสียงพยัญชนะสุดท้ายและ ลักษณะเกี่ยวกับเสียงสูงเสียงต่ำที่บอกถึงลักษณะเสียงของพยางค์

เสียงพยางค์ไทยสามารถแสดงลักษณะโดยทั่วไปโดยแบ่งเป็นสี่กลุ่ม: เสียงพยัญชนะเริ่มต้น, เสียงสระ, เสียงพยัญชนะสุดท้ายและเสียงสูงเสียงต่ำ เมื่อพิจารณาถึงโครงสร้างพยางค์ไทยจะถูกแบ่งเป็น 4 แบบแผน

- 2.1.3.1. แบบ CV เช่น ( ป่า [ pa ;0], รี [ ri ;1] )
- 2.1.3.2. แบบ CCV เช่น ( ปลา [ pla ;0], ครู [ khru ;0] )
- 2.1.3.3. แบบ CVC เช่น ( ปาด [ pa ; t1 ], กาก [ ka ; k1 ] )
- 2.1.3.4. แบบ CCVC เช่น ( ปราบ [ pra ; t1 ], กวาด [ kwa ; t1 ] )

ภาพที่ 2.2  
แสดงโครงสร้างหน่วยเสียงครึ่งพยางค์



ที่มา : Mittrapiyanurak et al., 2000

#### 2.1.4 โครงสร้างพยางค์

ภาษาไทยมาตรฐานเป็นภาษาที่ได้รับอิทธิพลจากภาษาต่างประเทศหลายประเทศ จึงได้รับเอาศัพท์จากภาษาต่างประเทศมาใช้โดยอาศัยหลักและวิธีทางด้านการกำหนดศัพท์ ดังนั้น โครงสร้างพยางค์จึงมี 3 แบบ

##### 2.1.4.1. โครงสร้างคำพยางค์เดียว (Monosyllabic Structure)

ภาษาไทยมาตรฐานเป็นภาษาจัดอยู่ในตระกูลคำโดด (Isolating Language) จำนวน คำในโครงสร้างจึงมีเป็นจำนวนมาก แต่สามารถจัดรูปแบบได้ 4 แบบดังนี้

- (1) CV เช่น กา มา ตา หมา หา เสีย
- (2) CVC เช่น การ ด้าน ศาล วัด สัตว์ หมายถึง เสีย
- (3) CCV เช่น ตู๋ ครู พลี พร่า คว่า
- (4) CCVC เช่น กราบ คลาน เปลี่ยน เกวียน ตราบ ตรม



#### 2.1.4.2. โครงสร้างคำสองพยางค์ (Disyllabic Structure)

เป็นโครงสร้างคำที่ประกอบด้วยคำ 2 พยางค์ แบ่งได้ 14 รูปแบบ ดังนี้

- (1) CV - CV เช่น กากี สีดามาลี ดารา พานา ภาษา
- (2) CV - CVC เช่น สามารถ สบาย สภาพ
- (3) CV - CCV เช่น ชาตรี สตรี
- (4) CV - CCVC เช่น โสโครก
- (5) CCV - CV เช่น ปรีดา กริธา ปรานี
- (6) CCV - CVC เช่น ปราศรัย ปราโมทย์ ปราสาท ปรางค์ ทรายสาร ทรายดวง  
ปรารภ
- (7) CVC - CV เช่น กานดา บรรดา นุภาพ
- (8) CVC - CVC เช่น สังคม กำนาน สันดาน
- (9) CVC - CCV เช่น นิทรา ศาสตรา จันทรา ดนตรี มาตรา ไมตรี
- (10) CCVC - CV เช่น กระทุ้ง กระจู้ ปราบดา กระเช้า
- (11) CCVC - CVC เช่น ตระการ ตระกูล ตรุษจีน
- (12) CCV - CCVC เช่น ปริเปรม ผลิผลาม
- (13) CCVC - CCVC เช่น ปริกปรี ทรายทราย ฝนฝน
- (14) CVC - CCVC เช่น มะกูด ตะกูด สงกรานต์

#### 2.1.4.3. โครงสร้างคำมากพยางค์ (Polysyllabic Structure)

เนื่องจากภาษาไทยมาตรฐานได้มีการเปลี่ยนแปลงจากตระกูลคำโดด กลายเป็นคำ  
มากพยางค์ ด้วยเหตุผลหลายประการ เช่นยืมคำจากภาษาอื่นมาใช้ ดังนั้นโครงสร้างคำมาก  
พยางค์จึงแบ่งได้เป็น 13 รูปแบบดังนี้

- (1) CV - CV - CV เช่น โสเภณี สาวิกา เสวนา สารภี
- (2) CV - CV - CVC เช่น โทรศัพท์ โทรศัพท์ โทรเลข โสดาบัน สรณะ
- (3) CVC - CV - CV เช่น บุษบา ดาษดา ศาสตรา
- (4) CV - CVC - CV เช่น สุวรรณี
- (5) CVC - CV - CVC เช่น ปัจจุบัน ประชาชน ประเทศชาติ สามเณร แสนยาวกร
- (6) CV - CV - CV - CV เช่น อเสวนา
- (7) CCVC - CV - CV เช่น ทฤษฎี กฤษฎา ปรารภนา
- (8) CV - CV - CV - CVC เช่น สาธารณะ สารานีเยะ

- (9) CV - CV - CV - CCVC เช่น สารานุกรม
- (10) CCVC - CV - CVC เช่น ปราศจาก
- (11) CV - CV - CV - CV - CVC เช่น สารานุกรม
- (12) CV - CV - CVC - CV - CVC เช่น สารวิริกธาตุ
- (13) CV - CV - CCVC - CVC เช่น สารประโยชน์

#### ข้อสังเกต

- (1) ไม่มีหน่วยเสียงพยัญชนะต้น เช่น VC
- (2) โครงสร้างคำหนึ่งพยางค์ส่วนมากเป็นคำศัพท์ร่วมตระกูล (Cognate forms)
- (3) โครงสร้างคำสองพยางค์และคำมากพยางค์เป็นคำมาจากภาษาต่างประเทศ โดยเฉพาะภาษาบาลีและสันสกฤต
- (4) โครงสร้างคำมากพยางค์มีจำนวนคำไม่มาก และส่วนมากเป็นคำที่มีการสร้างคำตามระเบียบของภาษาคำประสม (Compound words) หรือที่เรียกว่า คำสมาสและคำสนธิ

## 2.1.5 ภาษา (Language)

### 2.1.5.1. ภาษาวรรณยุกต์ (Tonal Language)

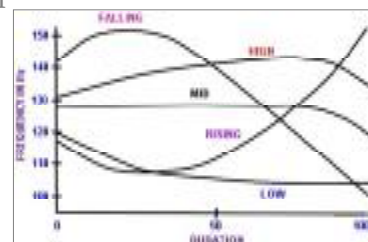
ภาษาที่มีระดับเสียงสูงต่ำที่แสดงความหมายของคำ เช่น ภาษาไทย จีน ลาว

ภาพที่ 2.3

ภาษาวรรณยุกต์

## Tonal language

- |                |     |       |
|----------------|-----|-------|
| • Middle Tone  | ปา  | /pa:/ |
| • Low Tone     | ป๋า | /pà:/ |
| • Falling Tone | ป๊า | /pâ:/ |
| • High Tone    | ป๊า | /pá:/ |
| • Rising Tone  | ป๋า | /pǎ:/ |



### 2.1.5.2. ภาษาทำนองเสียง (Intonation Language)

ภาษาที่มีระดับเสียงสูงต่ำที่แสดงความหมายของประโยค เช่น ภาษาอังกฤษ ฝรั่งเศส เยอรมัน

ภาพที่ 2.4

ภาษาทำนองเสียง

## Intonation language

*He found it on the street?*

[hi faʊnd it | an ɔ̃ / stɪt ɪ ]

*Yes, he found it on the street.*

[ \ jɛs ɪ | hi faʊnd it | an ɔ̃ \ stɪt ɪ ]

## 2.2 ทฤษฎีระบบสังเคราะห์เสียงพูด

ระบบสังเคราะห์เสียงพูดจากข้อความ เป็นระบบช่วยแปลงข้อความให้เป็นสัญญาณเสียงพูด ซึ่งคนสามารถเข้าใจได้ ซึ่งระบบนี้ส่วนใหญ่จะถูกพัฒนาสำหรับภาษาหลักๆ เช่น ภาษาอังกฤษ, จีน, ญี่ปุ่น และภาษาทางยุโรป เป็นต้น สำหรับภาษาไทยยังมีเป็นส่วนน้อย และขาดการพัฒนาที่ต่อเนื่อง มีหลายโปรแกรมที่ใช้ประโยชน์จากระบบนี้ เช่น โปรแกรมช่วยอ่านสำหรับผู้พิการทางสายตา โปรแกรมอ่านจดหมายอิเล็กทรอนิกส์ผ่านทางโทรศัพท์ เป็นต้น จุดหมายของระบบนี้คือ การที่ระบบสามารถสังเคราะห์เสียงธรรมชาติออกมาได้

ในอดีตมีงานวิจัยเกี่ยวกับระบบสังเคราะห์เสียงพูดซึ่งมีอยู่ 3 ส่วนหลักๆ ส่วนแรกคือ ส่วนประมวลผลข้อความจะทำหน้าที่แปลงข้อความเป็นตัวเขียนแสดงเสียงอ่าน ส่วนที่สองคือส่วนพจนานุกรมเสียงส่วนนี้จะทำการจับคู่ตัวแปรแสดงเสียงอ่านกับเสียงสังเคราะห์ ส่วนที่สามจะทำการสังเคราะห์เสียงโดยนำเสียงที่อยู่ในรูปคลื่นเสียงมาต่อกัน (Luksaneeyanawin et al., 1992) และมีอีกงานเป็นการพัฒนาระบบง่าย ๆ แต่เป็นระบบที่ใช้ได้ ภายในประกอบด้วย ส่วนแปลงข้อความเป็นหน่วยแสดงการออกเสียงโดยใช้กฎการสนทนาและพจนานุกรมขนาดเล็กสำหรับคำยกเว้น และอีกส่วนเป็นส่วนสังเคราะห์เสียงโดยเชื่อมเสียงที่อยู่ในรูปของคลื่นเสียงจากฐานข้อมูลเสียงครึ่งพยางค์ (Taisertavattanakul และ Kanawaree, 1995)

งานวิจัยที่เน้นลงไปเฉพาะจุดได้เสนอข้อคิดเห็นเกี่ยวกับโครงสร้างของเสียงในการสังเคราะห์เสียงพูดภาษาไทย ซึ่งในงานนี้คลื่นเสียงของพยางค์ถูกสร้างขึ้นจากการนำเอาคลื่นเสียงครึ่งพยางค์มาเชื่อมกัน (Kiat-arpakul, Fakcharoenphol และ Keretho, 1995) และอีกงานได้เสนอเทคนิคเพื่อแปลงรูปแบบของเสียงแต่ละพยางค์ด้วย PSOLA (Pitch Synchronous Overlap-Add) ซึ่ง จะทำการสังเคราะห์ F0 Contour อีกครั้ง เทคนิคนี้จะได้เปรียบในเรื่องของการลดจำนวนการสังเคราะห์สัญญาณเสียงได้ประมาณ 5 เท่า เทคนิคนี้จะเก็บเฉพาะเสียงกลางของพยางค์และจะสังเคราะห์เสียงอื่นจากสัญญาณเสียงนี้ (Luksaneeyanawin, S., 1995)

มีการนำเสนอบทความหัวข้อสำคัญในการวิจัยและพัฒนา ระบบสังเคราะห์เสียงพูดจากข้อความภาษาไทย ประกอบด้วยการวิเคราะห์ข้อความ, การสังเคราะห์สัทสัมพันธ์ และการสังเคราะห์สัญญาณเสียงพูด ในหัวข้อการวิเคราะห์ข้อความจะกล่าวถึงปัญหาที่สำคัญในการประมวลผลข้อความภาษาไทย และรายละเอียดของส่วนประกอบภายในซึ่งประกอบด้วย 3 ส่วน ได้แก่ การตัดประโยค การหาขอบเขตวลีเพื่อหยุดเว้นวรรคการอ่าน และการแปลงรูปเขียนเป็นรูปเสียงอ่านในหัวข้อการสังเคราะห์สัทสัมพันธ์จะกล่าวถึงกฎในการกำหนดช่วงเวลาของพยางค์ และ

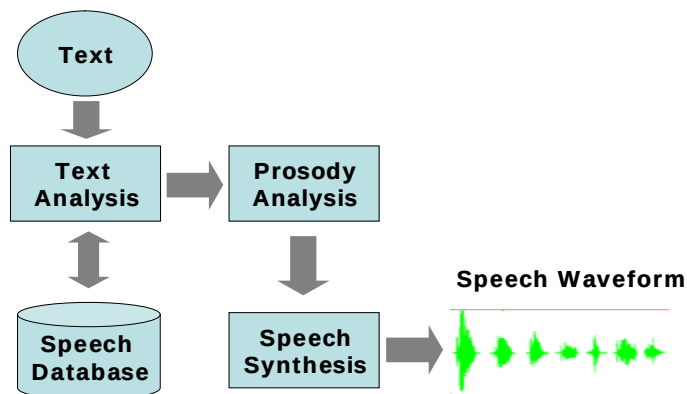
F0 Contour ซึ่งจะทำให้สามารถสังเคราะห์เสียงที่มีความสัมพันธ์ในระดับเหนือหน่วยเสียงได้ ส่วนหัวข้อการสังเคราะห์สัญญาณเสียงพูดกล่าวถึงโครงสร้างหน่วยเสียงแบบครึ่งพยางค์และอัลกอริทึมทางการประมวลผลสัญญาณในการปรับสัญญาณที่รอยต่อให้ต่อเนื่องและปรับสัญญาณให้มีลักษณะสัมพันธ์ตามที่ได้กำหนดมา (Mittrapiyanurak et al., 2000)

โดยสรุปการสังเคราะห์เสียงจากข้อความ ประกอบด้วยส่วนสำคัญ 3 ส่วน คือ ส่วนการวิเคราะห์ข้อความ (Text Analysis) , ส่วนการวิเคราะห์ลักษณะสัมพันธ์ (Prosody Analysis) และส่วนการสังเคราะห์สัญญาณเสียงพูด (Speech Synthesis)

ภาพที่ 2.5

ขั้นตอนการสังเคราะห์เสียง

### ขั้นตอนการสังเคราะห์เสียง



#### 2.2.1 การวิเคราะห์ข้อความ (Text Analysis)

ส่วนนี้จะมีหน้าที่วิเคราะห์ข้อความอินพุตจะกล่าวถึงปัญหาในการประมวลผลข้อความเพื่อแปลงเป็นข้อมูล เสียงอ่าน (Phoneme) ของคำนั้น และส่งต่อให้ส่วนของการ สังเคราะห์เสียง (Speech Synthesis) ต่อไป นอกจากนี้ส่วนนี้ยังทำหน้าที่อย่างอื่น ได้แก่

2.2.1.1. การตัดประโยค โดยการแบ่งประโยคจากข้อความที่ยาว (Sentence Breaking)

2.2.1.2. การทำข้อความให้อยู่ในรูปปกติ (Text Normalization) ได้แก่ การแปลง ตัวเลข, คำย่อ และเครื่องหมายอื่นๆ ที่ไม่ใช่ข้อ ความให้ กลายเป็นข้อความ

2.2.1.3. การหาขอบเขตของวลีของการอ่านในประโยค เพื่อหยุดเว้นวรรคการอ่าน

2.2.1.4. การแปลงรูปเขียนเป็นรูปเสียงอ่าน

## 2.2.2 การวิเคราะห์สัทสัมพันธ์ (Prosody Analysis)

ส่วนนี้ทำหน้าที่ในการวิเคราะห์และสังเคราะห์ข้อมูล สัทสัมพันธ์ (Prosody) ของ ประโยคใดๆ จากข้อมูลเสียงอ่าน และข้อความ จะกล่าวถึงกฎในการกำหนดช่วงเวลาของพยางค์ และ F0 Contour ซึ่งจะทำให้สามารถสังเคราะห์เสียงที่มีความสัมพันธ์ในระดับเหนือหน่วยเสียงได้ ข้อมูลสัทสัมพันธ์ ที่วิเคราะห์ ออกมาได้ในระบบโดยทั่วไปได้แก่

### 2.2.2.1. Segment Duration

คือความยาวของเสียงย่อยที่ต้องการสังเคราะห์ ค่านี้จะ มีผลต่อ จังหวะของเสียงที่ทำการ สังเคราะห์ เช่น ถ้ากำหนด ให้ค่า ความ ยาว ของเสียงย่อยที่ต้องการสังเคราะห์มี ขนาดสั้น เสียงที่ ทำ การสังเคราะห์ก็จะเหมือนกับการพูดเร็ว

### 2.2.2.2. Pitch Contour

คือ ค่าความสัมพันธ์ของความถี่มูลฐานกับเวลา ค่านี้จะมีผลต่อเสียงสูงต่ำ (Intonation) ของ ประโยคนั้นๆ

## 2.2.3 การสังเคราะห์สัญญาณเสียงพูด (Speech Synthesis)

ส่วนนี้ทำหน้าที่ในการสร้างสัญญาณคลื่นเสียง จากข้อมูลเสียง อ่าน (Phonetic Transcription) และข้อมูลสัทสัมพันธ์ (Prosody Transcription) จากข้อ 1 และ 2 และส่งออกสู่ ลำโพง เพื่อให้เราได้ยินเสียงพูดประโยคนั้นๆ โดยทั่วไป ส่วนนี้สามารถแบ่งตามเทคนิควิธีการ สังเคราะห์เสียง ได้ 3 ประเภท คือ

### 2.2.3.1. Formant Synthesis

เทคนิคการสังเคราะห์วิธีแบบนี้ ข้อมูลเสียงอ่านใดๆ จะถูกกำหนด ไว้อยู่ในรูป ของ ความถี่ฟอร์แมนท์ต่างๆ (F1,F2,F3) ของเสียงนั้นๆ เมื่อต้องการ สังเคราะห์ เสียงใดๆ ก็นำข้อมูล เหล่านี้มาทำการสังเคราะห์ให้ เป็นสัญญาณ เสียง ซึ่งวิธีการนี้จะมีข้อดี ที่สามารถควบคุมค่าความ เปลี่ยนแปลงของ ความถี่ฟอร์แมนท์ (Formant Transition) ที่บริเวณรอย ต่อระหว่าง เสียงได้ง่าย แต่มีข้อเสีย คือ การจะแทนเสียงใดๆด้วยค่าฟอร์แมนท์ทำได้ยาก จะต้องมีความรู้ในการสังเคราะห์ เสียงใดๆจำนวนมาก และเสียงที่ สังเคราะห์ออก มาได้จะไม่ค่อยเป็นธรรมชาติ ตัวอย่างของระบบ แบบนี้ได้แก่ MITALK (Allen et al., 1987) และ DECTALK (Klatt, 1987)

### 2.2.3.2. Articulation Synthesis

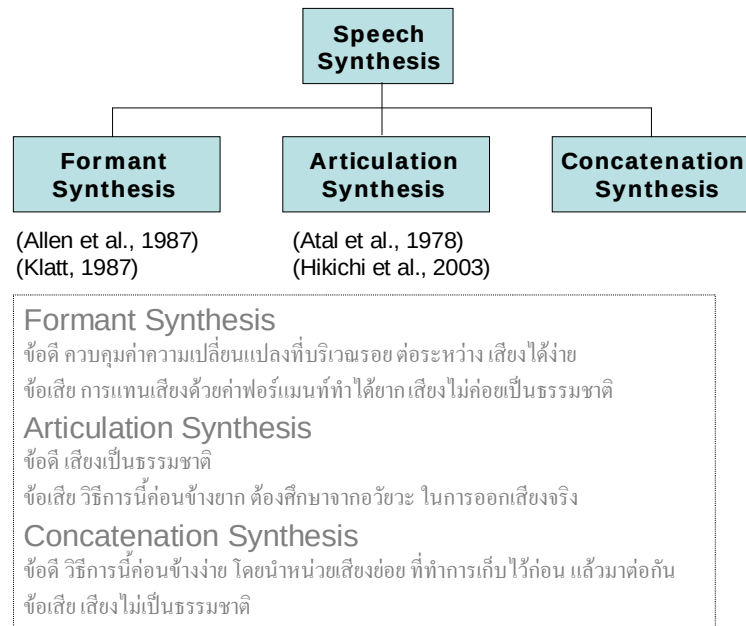
สำหรับวิธีนี้ข้อมูลเสียงที่ต้องการสังเคราะห์จะ โมเดล อยู่ในรูปของค่า พารามิเตอร์ของ โครงสร้าง ทางกายภาพของการเคลื่อนไหวของ อวัยวะ ในช่อง ปากที่ทำให้เกิดเสียงต่างๆวิธีการนี้ ค่อนข้างยาก ในแง่การโมเดล เสียงต่างๆ ซึ่งจะต้องศึกษาจากอวัยวะ ในการออกเสียงจริงๆ (Atal et al., 1978) และ (Hikichi et al., 2003)

### 2.2.3.3. Concatenation Synthesis

เทคนิควิธีการนี้เป็นวิธีที่ได้รับความนิยมมากที่สุดในปัจจุบัน โดยเสียง ที่ทำการ สังเคราะห์ขึ้น นี้เกิดจากการนำหน่วยเสียงย่อย ที่ทำการเก็บไว้ก่อน แล้วมาต่อกันเป็น เสียงพูดที่ ต้องการ โดยทั่วไปหน่วยเสียงย่อยที่ทำเก็บ ไว้จะอยู่ระดับต่ำกว่าคำ เช่น หน่วยของเสียงพยางค์ หน่วยของเสียง ครึ่งพยางค์ (Demisyllable) หน่วยของเสียงคู่เสียง (Diphone) เป็นต้น ซึ่งในงานวิจัยนี้ได้เลือกใช้เทคนิควิธีการนี้ โดยใช้หน่วยของ เสียงครึ่งพยางค์ เป็นพื้นฐาน



ภาพที่ 2.6  
จำแนกประเภทการสังเคราะห์เสียง



ในปัจจุบันนี้เป้าหมายของการพัฒนาระบบสังเคราะห์เสียงนี้ได้แก่ เสียงที่สังเคราะห์ออกมาสามารถฟังแล้วเข้าใจได้ (Intelligibility) เสียงที่สังเคราะห์ออกมาเป็นธรรมชาติใกล้เคียงเสียงมนุษย์ (Naturalness) ซึ่งระบบที่พัฒนากันโดยทั่วไปสามารถพัฒนาให้ถึง เป้าหมาย Intelligibility แต่ก็ยังขาดในเรื่องของ Naturalness

ในการที่จะทำให้เสียงมีความเป็นธรรมชาตินั้นต้องอาศัยฐานข้อมูลเสียงขนาดใหญ่ ฐานข้อมูลเสียงพูดเป็นส่วนประกอบสำคัญที่จะทำให้ระบบรู้จำเสียงพูดมีประสิทธิภาพและให้ผลการรู้จำถูกต้องสูงสุดแม้ว่าในปัจจุบันได้มีการคิดค้นนำเสนอวิธีมากมายในการพัฒนาระบบรู้จำเสียงพูดให้ดีขึ้น แต่ระบบรู้จำที่ดียังคงต้องอาศัยฐานข้อมูลเสียงขนาดใหญ่ เพื่อใช้เป็นข้อมูลพื้นฐานในการวิเคราะห์และฝึกฝนระบบ

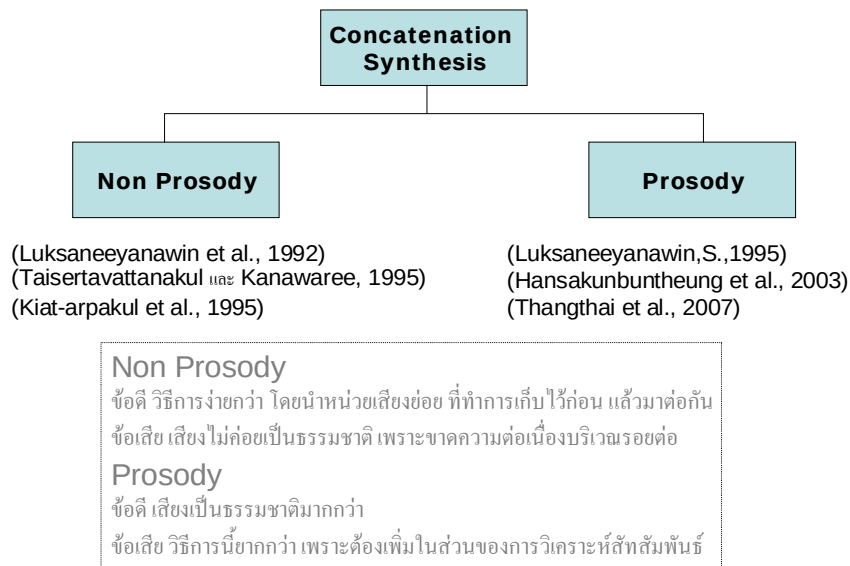
งานวิจัยได้มีการพัฒนาฐานข้อมูลเสียงเพื่อใช้ในการสังเคราะห์สัญญาณเสียงพูด การเขียนแบบแสดงเสียงอ่านและข้อมูลเกี่ยวกับเสียงเพื่อที่จะนำไปพัฒนารูปแบบสัทสัมพันธ์การพูดภาษาไทยต่อไปในอนาคต ในงานนี้แบบแสดงเสียงอ่านประกอบด้วย รูปอ่าน, สัญลักษณ์แสดงเสียง, ขอบเขตของเสียงพูด, ตำแหน่งของพยางค์, เสียงพูด, เสียงสูงต่ำ และข้อมูลเกี่ยวกับเสียงได้แก่

ค่าพลังงาน,ระดับเสียง,การแยกช่วงของเสียง และเส้นเสียง ซึ่งในฐานะข้อมูลเสียงนี้ประกอบด้วยคำ 5,200 ประโยค ซึ่งครอบคลุมหมดทุกเสียงทั้งคำไทย และคำที่ยืมมาจากภาษาต่างประเทศ (Hansakunbuntheung et al., 2003)

การพัฒนาทางด้านสัทสัมพันธ์ จะทำให้การสังเคราะห์เสียงที่มีความสัมพันธ์ในระดับเหนือหน่วยเสียง ได้มีการวิเคราะห์ปัจจัยควบคุมความต่อเนื่องพยางค์ไทย และใช้สถิติเทคนิคเชิงเส้นเกี่ยวกับหน่วยควบคุม การวิเคราะห์ได้แสดงทั้งระดับพยางค์ และระดับวลี และการสังเกตหน่วยควบคุมความต่อเนื่องระดับพยางค์ ผลกระทบของระดับเสียงทั้งห้าและโครงสร้างพยางค์ เพื่อวิเคราะห์สถิติผลกระทบโครงสร้างพยางค์ ได้มีการประยุกต์ทฤษฎี Quantification กับปัจจัยเกี่ยวกับภาษาทั้งสอง (Hansakunbuntheung et al., 2003)

สำหรับระบบการสังเคราะห์เสียงที่มีการสร้างขึ้นมาใช้งานจริงนั้นในปัจจุบันมีการพัฒนาระบบการสังเคราะห์เสียงพูดภาษาไทย “วาจา” (Vaja) ที่มีคุณภาพเสียงเป็นธรรมชาติมากกว่าขึ้น โดยมีส่วนที่ได้รับการพัฒนา 3 ส่วนคือ 1. ในส่วนประมวลผลข้อความ(Text Analysis) มีการเปลี่ยนขั้นตอนการหาขอบเขตของประโยค (Sentence Boundary Detection) มาเป็นการหาขอบเขตของวลี (Phrase Break Detection) ทำให้มีจังหวะในการพูดที่เป็นธรรมชาติและฟังง่ายขึ้น และพัฒนา ส่วนการแปลงรูปเขียนเป็นรูปอ่าน (Grapheme-to-Phoneme Conversion) ให้มีประสิทธิภาพสูงขึ้น 2. เพิ่มส่วนการวิเคราะห์ระยะเวลาของหน่วยเสียง (Phoneme Duration Modeling) Time Domain Pitch Synchronous Overlap-Add (TD-PSOLA) เพื่อปรับความยาวของสัญญาณเสียงให้ใกล้เคียงกับเสียงพูดที่เป็นธรรมชาติมากขึ้น 3. ใช้เทคนิคการคัดเลือกหน่วยเสียงแบบครึ่งเสียง (Half Phone) เพื่อแก้ปัญหาในกรณีที่การคัดเลือกหน่วยเสียง (Unit Selection) ไม่พบหน่วยเสียงแบบคู่เสียง(Di-Phone) อยู่ในฐานะข้อมูล (Corpus) (Thangthai et al., 2007)

ภาพที่ 2.7  
การสังเคราะห์เสียงแบบเชื่อมต่อหน่วยเสียง



## 2.3 ทฤษฎีการโมเดลทำนองเสียง

ทำนองเสียง หมายถึงระดับสูง ต่ำของเสียงเป็นสิ่งที่ใช้บอกความหมายของประโยค แบ่งออกเป็น 2 รูปแบบคือ

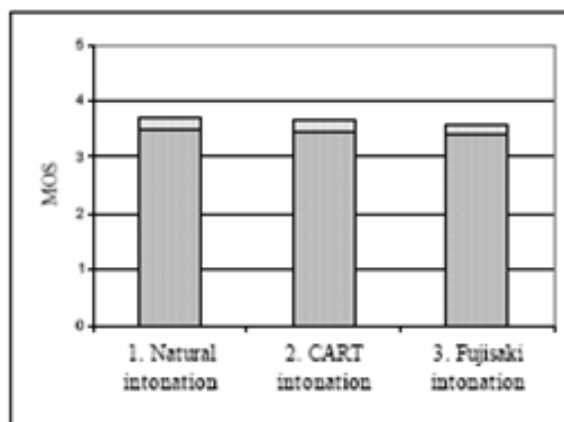
ทำนองเสียงตก เป็นลักษณะของประโยคบอกเล่า

ทำนองเสียงขึ้น เป็นลักษณะของประโยคคำถาม และประโยคคำสั่งบางชนิด

มีงานวิจัยที่นำเสนอเทคนิคในการออกแบบทำนองเสียงในต่างประเทศโดยใช้การเรียนรู้ของเครื่องจักร (Machine Learning) เข้ามาช่วยในการสร้างแบบจำลอง โดยใช้วิธี Classification and Regression Trees (CART), Fujisaki Model ทำการเปรียบเทียบประสิทธิภาพการทำนายค่าความถี่มูลฐาน(F0 Contour) กับเสียงพูดธรรมชาติของมนุษย์ ปรากฏว่าวิธี Classification and Regression Trees (CART) ให้ผลลัพธ์ออกมาใกล้เคียงเสียงพูดของมนุษย์มากกว่า Fujisaki Model

### ภาพที่ 2.8

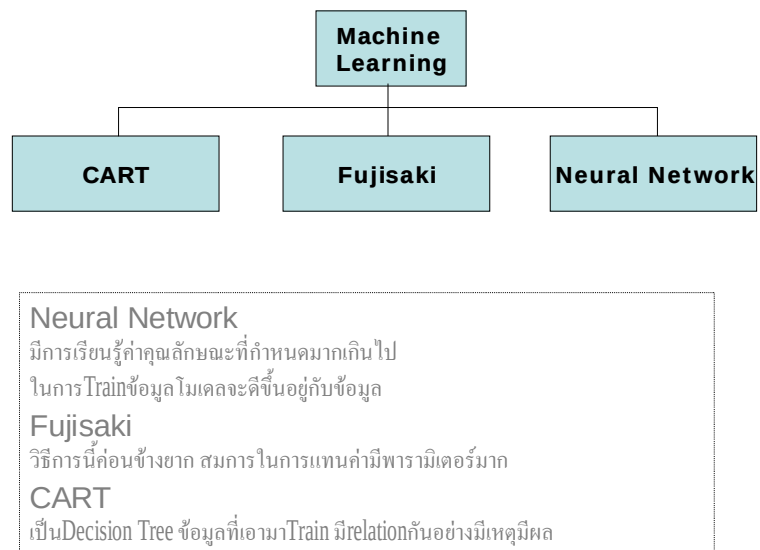
เปรียบเทียบประสิทธิภาพการทํานายค่าความถี่มูลฐาน



*Figure 4. Listening test results comparing 20 English names using Fujisaki intonation, CART based intonation and natural intonation.*

สำหรับงานวิจัยของไทยก็มียงานวิจัยการทํานายค่าความถี่มูลฐาน(F0 Contour) เป็นการทดลองโดยใช้เทคนิค **Fujisaki Model** ซึ่งให้ผลการทํานายถูกต้องสูง แต่วิธีการนี้ค่อนข้างยาก สมการในการแทนค่ามีพารามิเตอร์มาก

ภาพที่2.9  
เปรียบเทียบเทคนิคการเรียนรู้ของเครื่องจักร



ดังนั้นบทความฉบับนี้ จะนำเสนอเทคนิคในการสังเคราะห์เสียงให้ได้คล้ายมนุษย์ โดยจะใช้หลักการเรียนรู้ของเครื่องจักร จะใช้ทฤษฎีต้นไม้ตัดสินใจ (Decision Tree) โดยจะเรียนรู้จากคลังข้อมูลเสียงชื่อว่า "TSynC1" ซึ่งถูกพัฒนาโดย ศูนย์เทคโนโลยีอิเล็กทรอนิกส์ (NECTEC)

## 2.4 งานวิจัยที่เกี่ยวข้อง

ตารางที่ 2.4  
เปรียบเทียบงานวิจัยที่เกี่ยวข้อง

| ชื่องานวิจัย                                                                                                | เทคนิคที่ใช้                                                                                                                                                                                | ข้อดี                                | ข้อเสีย                                                                                           |
|-------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------|---------------------------------------------------------------------------------------------------|
| A Thai text-to-speech system<br>(Luksaneeyanawin et al., 1992)                                              | ระบบสังเคราะห์เสียง มี 3 ส่วน คือ ประมวลผลข้อความ พจนานุกรมเสียง และ สังเคราะห์เสียง                                                                                                        | ทำได้ง่าย                            | ยังขาดความเป็นธรรมชาติ การสังเคราะห์เสียงโดยนำเสียงที่อยู่ ในรูปคลื่นเสียงมาต่อกัน                |
| Thai speech synthesizer<br>(Taisertavattanakul และ Kanawaree, 1995)                                         | พัฒนาระบบง่ายๆ ที่ใช้ได้ ภายในประกอบ ด้วย ส่วนแปลงข้อความ เป็นหน่วยแสดง การออกเสียงโดยใช้ กฎการสนทนาและ พจนานุกรมขนาดเล็ก สำหรับคำยกเว้น และ ส่วนสังเคราะห์เสียง โดยเชื่อมเสียงครึ่ง พยางค์ | ทำได้ง่าย และสามารถให้ระบบ ได้ดีขึ้น | ยังขาดความเป็น ธรรมชาติ การสังเคราะห์เสียง โดยนำเสียงครึ่งพยางค์ ที่อยู่ ในรูปคลื่นเสียงมา ต่อกัน |
| A combined phoneme-based and demisyllable approach for Thai speech synthesis<br>(Kiat-arpakul et al., 1995) | เสนอโครงสร้างของ เสียงในการสังเคราะห์ เสียงพูดภาษาไทย ซึ่ง คลื่นเสียงของพยางค์ ถูก ส ร้าง จา ก การ นำเอาคลื่นเสียงครึ่ง พยางค์มาเชื่อมกัน                                                   | ทำได้ง่าย และสามารถให้ระบบ ได้ดีขึ้น | ยังขาดความเป็น ธรรมชาติ การสังเคราะห์เสียง โดยนำเสียงครึ่งพยางค์ ที่อยู่ ในรูปคลื่นเสียงมา ต่อกัน |

## ตารางที่ 2.4 (ต่อ)

| ชื่องานวิจัย                                                                                   | เทคนิคที่ใช้                                                                                                                                                                                        | ข้อดี                                                             | ข้อเสีย                                                                                                             |
|------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------|
| Tone transformation<br>(Luksaneeyanawin,S.,1995)                                               | แปลงรูปแบบเสียง<br>แต่ละพยางค์ด้วย<br>PSOLA ซึ่งจะทำการ<br>สังเคราะห์ F0<br>Contour อีกครั้ง<br>เทคนิคนี้จะเก็บ<br>เฉพาะเสียงกลาง<br>ของพยางค์และจะ<br>สังเคราะห์เสียงอื่น<br>จากสัญญาณเสียง        | สามารถลดจำนวน<br>การสังเคราะห์<br>สัญญาณเสียงได้<br>ประมาณ 5 เท่า | ทำได้ยาก                                                                                                            |
| Issues in Thai Text-to-Speech Synthesis: The NECTEC Approach<br>(Mittrapiyanurak et al., 2000) | บทความหัวข้อ<br>สำคัญในการวิจัย<br>และพัฒนาระบบ<br>สังเคราะห์เสียงพูด<br>จากข้อความภาษาไทย<br>ประกอบด้วยการวิเคราะห์<br>ข้อความ, การสังเคราะห์<br>สัทสัมพันธ์<br>และการสังเคราะห์<br>สัญญาณเสียงพูด | สรุปประเด็นปัญหา<br>สำคัญในการทำวิจัย<br>การสังเคราะห์พูด         | ยังขาดความเป็น<br>ธรรมชาติ<br>การสังเคราะห์<br>เสียงโดยนำเสียง<br>ครึ่งพยางค์ที่อยู่<br>ในรูปคลื่นเสียงมา<br>ต่อกัน |

ตารางที่ 2.4 (ต่อ)

| ชื่องานวิจัย                                                                                          | เทคนิคที่ใช้                                                                                                      | ข้อดี                                       | ข้อเสีย                                          |
|-------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------|---------------------------------------------|--------------------------------------------------|
| Thai Tagged Speech Corpus for Speech Synthesis (Hansakunbuntheung et al., 2003)                       | พัฒนาฐานข้อมูลเสียง เพื่อใช้ในการสังเคราะห์สัญญาณเสียงพูด มีการเขียนแบบแสดงเสียงอ่านและข้อมูลเกี่ยวกับเสียง       | พัฒนารูปแบบ สัทสัมพันธ์การพูด ภาษาไทย       | ใช้เวลา กำลังแรงงาน และผู้เชี่ยวชาญในการพัฒนามาก |
| Analysis and Modeling of Syllable Duration for Thai Speech Synthesis (Hansakunbuntheung et al., 2003) | การวิเคราะห์ปัจจัยควบคุมความต่อเนื่อง พยางค์ไทย และใช้สถิติเทคนิคเชิงเส้นเกี่ยวกับหน่วยควบคุม                     | พัฒนารูปแบบ สัทสัมพันธ์การพูด ภาษาไทย       | ใช้ผู้เชี่ยวชาญในการพัฒนา                        |
| Recent Progress of "VAJA" Thai Text-to-Speech Synthesis (Thangthai et al., 2007)                      | ส่วนประสมผลข้อความมีการหาขอบเขตของวลี มีการเพิ่มส่วนของการวิเคราะห์ระยะเวลาของหน่วยเสียง และการคัดเลือกหน่วยเสียง | จังหวะในการพูดที่เป็นธรรมชาติและฟังง่ายขึ้น | เสียงสูงต่ำในระดับประโยคยังไม่มีดี               |



### บทที่ 3

#### วิธีการดำเนินการวิจัย

ในบทนี้กล่าวถึงวิธีดำเนินการวิจัยเพื่อให้ได้ตามวัตถุประสงค์ที่ระบุไว้ในบทนำ ซึ่งประกอบไปด้วย ข้อมูลที่ใช้ในการวิจัย เครื่องมือในการเรียนรู้ระบบ ขั้นตอนในการดำเนินการระยะเวลาในการดำเนินการวิจัย โดยมีรายละเอียดดังนี้

#### 3.1 ข้อมูลที่ใช้ในการวิจัย

งานวิจัยนี้เป็นงานวิจัยมีศาสตร์ทางด้านเสียงโดยผู้วิจัยมีเป้าหมายที่จะพัฒนาคุณภาพของเสียงที่สังเคราะห์ออกมาให้มีความชัดเจนใกล้เคียงเสียงมนุษย์ ดังนั้นผู้วิจัยจึงทำการทดลองกับคลังข้อมูลเสียงขนาดใหญ่ (Corpus) ในการทดลองดังนี้

##### Thai Tagged Speech Corpus for Speech Synthesis (TSynC1)

เป็นคลังข้อมูลเสียงขนาดใหญ่ที่จะจัดเก็บฐานข้อมูลเสียง เพื่อใช้ในการสังเคราะห์สัญญาณเสียงพูด มีการเขียนแบบแสดงเสียงอ่านและข้อมูลเกี่ยวกับเสียงเพื่อที่จะนำไปพัฒนารูปแบบสัทสัมพันธ์การพูดภาษาไทยต่อไป โดยแบบแสดงเสียงอ่านประกอบด้วย รูปอ่าน, สัญลักษณ์แสดงเสียง, ขอบเขตของเสียงพูด, ตำแหน่งของพยางค์, เสียงพูด, เสียงสูงต่ำ และข้อมูลเกี่ยวกับเสียงได้แก่ ค่าพลังงาน, ระดับเสียง, การแยกช่วงของเสียง และเส้นเสียง ซึ่งในฐานข้อมูลเสียงนี้ประกอบด้วยประโยค 5,200 ประโยค ซึ่งครอบคลุมหมดทุกเสียงทั้งคำไทย และคำที่ยืมมาจากภาษาต่างประเทศ ซึ่งถูกพัฒนาโดย ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ (Hansakunbuntheung et al., 2003)

#### 3.2 เครื่องมือในการเรียนรู้ระบบ

งานวิจัยนี้เสนอหลักการปัญญาประดิษฐ์ เรื่องการเรียนรู้ของเครื่องจักร (Machine Learning) เข้ามาช่วยในการสร้างแบบจำลองโดยจะใช้วิธีการของต้นไม้ตัดสินใจ (Decision Tree) มาทำการทดลอง

### 3.2.1 การเรียนรู้ต้นไม้ตัดสินใจ (Decision Tree Learning)

ต้นไม้ตัดสินใจ (Decision Tree) เป็นโมเดลทางคณิตศาสตร์ที่ใช้กันอย่างกว้างขวาง (Quinlan, 1993)

โดยเป็นวิธีการเรียนรู้แบบมีผู้สอน (Supervised learning) การประเมินค่า ทนต่อข้อมูลรบกวน และเป็นวิธีในการประมาณผลลัพธ์ที่ได้ อัลกอริทึมในการสร้างต้นไม้ตัดสินใจมีหลายวิธี ได้แก่ Classification and Regression Trees (CART) (Breiman et al., 1984) Induction of Decision Trees (ID3) (Quinlan, 1986, pp 81-106) และ C4.5 (Quinlan, 1993)

#### 3.2.1.1. การแทนต้นไม้ตัดสินใจ (Decision Tree Representation)

ในการจำแนกของต้นไม้ตัดสินใจจะแสดงข้อมูลโดยเรียงจากรากของต้นไม้ตัดสินใจไป ยังแต่ละส่วนของกิ่งแต่ละส่วนของโหนดภายใน (Inner Node) และไปสิ้นสุดที่โหนดใบ (Leaf Node) ดังนั้น การทำนายผลลัพธ์ จะพิจารณาโดยเริ่มต้นจากรากของต้นไม้ตัดสินใจแล้วทำการ ย้ายไปตามกิ่งของต้นไม้เรื่อยๆ จนถึงโหนดสุดท้าย โดยการแทนค่าในต้นไม้มีลักษณะของคู่ (ค่าคุณลักษณะ, ค่าของค่าคุณลักษณะ) ดังนี้

- (1) โหนดภายใน (Inner Node) สัมพันธ์กับ ค่าคุณลักษณะ (Attribute)
- (2) กิ่งของต้นไม้ สัมพันธ์กับ ค่าของค่าคุณลักษณะ (Attribute Value)
- (3) โหนดใบ (Leaf Node) สัมพันธ์กับ ค่าคลาสของต้นไม้ตัดสินใจ (Class)

#### 3.2.1.2. ต้นไม้ตัดสินใจ (Decision Tree)

การจำแนกข้อมูลโดยใช้ต้นไม้การตัดสินใจ เป็นการเลือกตัดสินใจแบบโครงสร้างต้นไม้ โดยปัญหาคือราก และวิธีการหรือเงื่อนไขต่างๆ คือ ใบที่จะแตกแขนงไปเรื่อยๆ จนพบคำตอบ โดยมีกระบวนการทำงานหลักๆสามารถทำได้ 2 ขั้นตอน

- (1) สร้างต้นไม้การตัดสินใจจาก Training Data และทดสอบความถูกต้องด้วย Testing Data
- (2) ใช้ต้นไม้การตัดสินใจในการแยกกลุ่มของข้อมูล

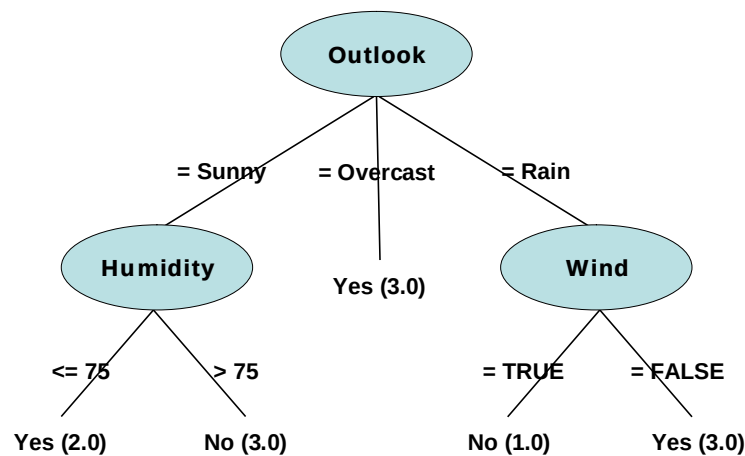
#### 3.2.1.3. วิธีสร้างต้นไม้ตัดสินใจ (Decision Tree)

- (1) เลือก Attribute ที่ดีที่สุดสำหรับการตัดสินใจ
- (2) กำกับ Attribute ที่เลือกได้กับโหนดปัจจุบัน
- (3) สำหรับค่า Attribute แต่ละช่วง สร้างโหนดลูกของโหนดปัจจุบัน
- (4) จัดเรียงข้อมูลฝึกฝน และจำแนกไปยังโหนดที่สร้างขึ้นในข้อ (3)

- (5) ถ้าข้อมูลจำแนกได้สมบูรณ์ให้หยุดทำงาน ถ้ายังไม่สมบูรณ์ให้กลับไปทำข้อ (1)

ภาพที่ 3.1

โครงสร้างต้นไม้เพื่อคาดเดาว่าจะเล่นกีฬาเทนนิสหรือไม่



ตัวอย่างโครงสร้างต้นไม้เพื่อคาดเดาว่าจะเล่นกีฬาเทนนิสหรือไม่

อัลกอริทึมต้นไม้ตัดสินใจ (Decision Tree) มีหลายวิธี แต่ในการทดลองนี้ จะเลือกใช้วิธี Classification and Regression Trees (CART) (Breiman et al., 1984) และ C4.5 (Quinlan, 1993) มาทำการทดลอง

#### 3.2.1.4. Classification and Regression Trees (CART)

เป็นเทคนิคการตัดสินใจด้วยแผนภูมิต้นไม้ที่ได้กำหนดคลาสของด้าเซตแล้ว โดยมีกลุ่มของกฎที่สามารถนำไปใช้กับด้าเซตใหม่เพื่อการคาดเดาว่าข้อมูลจะออกมาในลักษณะใด ซึ่งการแบ่งข้อมูลนั้น จะแบ่งข้อมูลออกเป็น 2 ทาง จึงใช้ข้อมูลในการวิเคราะห์น้อย

## ภาพที่3.2

การทดลองโดยใช้เครื่องมือต้นไม้ตัดสินใจ

## CART

```

pb.desc
(
  (out play yes no)
  (outlook sunny overcast rainy)
  (temperat float)
  (humidity float)
  (windy TRUE FALSE)
)

pb.res

Dataset of 12 vectors of 5 parameters from: pb.data
Dataset of 4 vectors of 5 parameters from: pb.test

(((play 0) (yes 0.666667) (no 0.333333) yes))
  pla yes
  play 0 0 0 0 [0/0] 100.000
  yes 0 1 0 1 [1/1] 100.000
  no 0 3 0 3 [0/3] 0.000
    0 4 0
total 4 correct 1.000 25.000%

pb.data
no sunny 85 85 FALSE
no sunny 80 90 TRUE
yes overcast 83 86 FALSE
yes rainy 70 96 FALSE
yes rainy 68 80 FALSE
no rainy 65 70 TRUE
yes overcast 64 65 TRUE
no sunny 72 95 FALSE
yes sunny 69 70 FALSE
yes rainy 75 80 FALSE
yes sunny 75 70 TRUE
yes overcast 72 90 TRUE

pb.test
yes overcast 81 75 FALSE
no rainy 71 91 TRUE
no sunny 71 91 TRUE
no sunny 71 91 TRUE

```

## 3.2.2 ข่ายงานประสาทเทียม (Neural Network)

ข่ายงานประสาทเทียม (Neural Network) ในการทดลองนี้ จะเลือกใช้วิธีข่ายงานประสาทเทียมแบบแพร่กระจายย้อนกลับหลายชั้น (Multilayer Backpropagation Neural Network) เป็นข่ายงานประสาทเทียม (Neural Network) ที่ใช้อัลกอริทึมแพร่กระจายย้อนกลับ (Backpropagation) (Rumelhart และ McClelland, 1986) คุณสมบัติพิเศษที่เพิ่มขึ้นมากก็คือสามารถสร้างพื้นผิวการตัดสินใจแบบไม่เป็นเชิงเส้นได้ (Nonlinear Decision Surface) ที่แบ่งแยกตัวอย่างได้ดีกว่าพื้นผิวการตัดสินใจแบบเชิงเส้น (Linearly Decision Surface) อัลกอริทึมแพร่กระจายย้อนกลับใช้กฎการเรียนรู้คือ กฎเดลต้า (Delta Rule) ซึ่งมีข้อดีตรงที่การเรียนรู้จะเข้าสู่ระนาบหลายมิติที่ให้ค่าผิดพลาดน้อยที่สุด โดยใช้หลักการเคลื่อนลงตามความชัน (Gradient Descent) กฎเดลต้าจะหาเวกเตอร์น้ำหนักที่ให้ค่าผิดพลาดของตัวอย่างสอนน้อยที่สุดโดยการหาอนุพันธ์ทางคณิตศาสตร์ ดังนั้นจึงต้องใช้ฟังก์ชันกระตุ้นที่สามารถหาอนุพันธ์ได้ อาทิ ฟังก์ชันเชิงเส้น (Linear Function) หรือฟังก์ชันซิกมอยด์ (Sigmoid Function)

### 3.3 ขั้นตอนในการดำเนินการ

ผู้วิจัยได้กำหนดวิธีการดำเนินการวิจัยให้สอดคล้องกับวัตถุประสงค์และขอบเขตของการวิจัยตามขั้นตอนดังต่อไปนี้

- 3.3.1 ทบทวนทฤษฎีและงานวิจัยที่เกี่ยวข้อง
- 3.3.2 ทำการออกแบบทำนองเสียงภาษาไทย ที่เหมาะกับการสังเคราะห์เสียง
- 3.3.3 ศึกษาเครื่องมือ วิธี และอัลกอริทึมที่เกี่ยวข้อง
- 3.3.4 พัฒนาโปรแกรมที่ใช้ในการทดลอง
- 3.3.5 ทดสอบและวัดประสิทธิภาพการสังเคราะห์เสียง
- 3.3.6 สรุปรวบรวมผลการวิจัย พร้อมทั้งจัดทำเอกสารที่เกี่ยวข้องกับวิทยานิพนธ์

### 3.4 ระยะเวลาในการดำเนินการวิจัย

ตารางที่ 3.1

แผนการดำเนินงาน

| ขั้นตอนการดำเนินงาน                   | เม.ย. | พ.ค. | มิ.ย. | ก.ค. | ส.ค. | ก.ย. | ต.ค. | พ.ย. | ธ.ค. | ม.ค. | ก.พ. | มี.ค. |
|---------------------------------------|-------|------|-------|------|------|------|------|------|------|------|------|-------|
| 1. เสนอหัวข้องานวิจัย                 | ■     | ■    |       |      |      |      |      |      |      |      |      |       |
| 2. ทบทวนทฤษฎีและงานวิจัยที่เกี่ยวข้อง |       | ■    | ■     | ■    | ■    | ■    | ■    |      |      |      |      |       |
| 3. จัดหาเครื่องมือ วิธี และอัลกอริทึม |       |      | ■     | ■    | ■    | ■    | ■    | ■    |      |      |      |       |
| 4. ออกแบบทำนองเสียงภาษาไทย            |       |      |       | ■    | ■    | ■    | ■    | ■    | ■    |      |      |       |
| 5. พัฒนาโปรแกรมที่ใช้ในการทดลอง       |       |      |       |      |      |      |      | ■    | ■    | ■    | ■    | ■     |
| 6. ทดสอบและวัดประสิทธิภาพ             |       |      |       |      |      |      |      |      | ■    | ■    | ■    | ■     |
| 7. จัดทำเอกสาร + นำเสนอผลงาน          |       |      |       |      |      |      | ■    | ■    | ■    | ■    | ■    | ■     |

## บทที่ 4

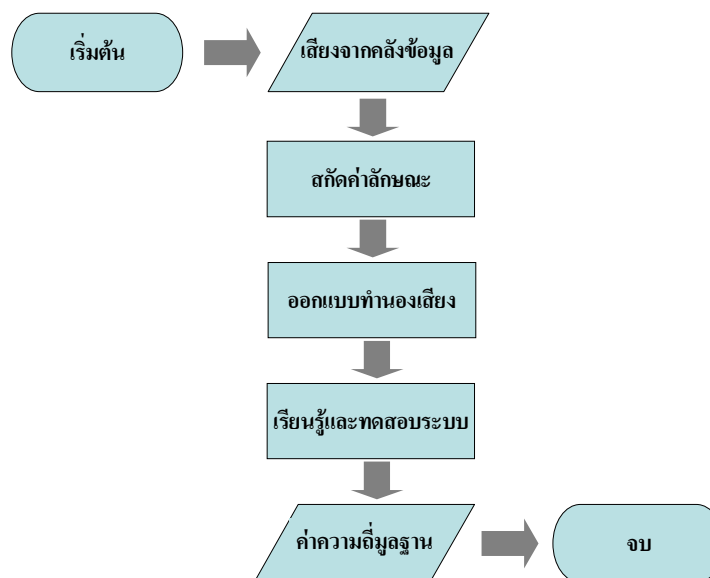
### การทดลองและผลการทดลอง

การทดลองจะใช้ประโยคจากคลังข้อมูลเสียงขนาดใหญ่ (Corpus) ชื่อว่า "TSynC1" เพื่อใช้ในการสังเคราะห์สัญญาณเสียงพูด ซึ่งในฐานข้อมูลเสียงนี้ประกอบด้วยประโยค 5,200 ประโยค โดยจะแบ่งเป็นข้อมูลที่ใช้ในการทดลองคือ ข้อมูลใช้ในการฝึกฝน (Training Data) ในสัดส่วนร้อยละ 90 ของเสียงทั้งหมด และ ข้อมูลใช้ในการทดสอบ (Testing Data) ในสัดส่วนร้อยละ 10 ของเสียงทั้งหมด ในส่วนของการสังเคราะห์ทำนองเสียง จะสังเคราะห์ทำนองเสียงทีละประโยค โดยพิจารณาเอาเสียงของแต่ละหน่วยเสียงย่อยมาเชื่อมต่อกัน ตามโมเดลทำนองเสียงที่ได้กำหนดไว้ ซึ่งโมเดลทำนองเสียงได้มาจากการฝึกฝนจากคลังข้อมูลเสียงขนาดใหญ่ โดยใช้เทคนิคทางปัญญาประดิษฐ์ ที่เรียกว่า เทคนิคต้นไม้ตัดสินใจ (Decision Tree) และเทคนิคข่ายงานประสาทเทียม (Neural Network) ในการพิจารณา เพื่อสร้างโมเดลทำนองเสียง

### 4.1 วิธีการทดลอง

ผู้วิจัยได้กำหนดขั้นตอนในการทดลองใช้สอดคล้องกับวัตถุประสงค์และขอบเขตของการวิจัยตามขั้นตอนดังต่อไปนี้

ภาพที่ 4.1  
แผนภาพขั้นตอนการทดลอง



#### 4.1.1 ข้อมูลนำเข้า

ขั้นตอนนี้จะทำการศึกษาข้อมูลที่ใช้ในการทดลอง โดยข้อมูลที่ใช้ในการฝึกฝนและทดสอบระบบ นำมาจากคลังข้อมูลเสียงขนาดใหญ่ (Corpus) ชื่อว่า "TSynC1" ซึ่งเป็นคลังข้อมูลเสียงภาษาไทยสำหรับงานวิจัยด้านการสังเคราะห์เสียง ซึ่งถูกพัฒนาโดย ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ โดยตัวอย่างข้อมูลในคลังข้อมูล ระบุไว้ในภาคผนวก ก

ในการทดลองจะแบ่งข้อมูลเป็น 2 ชุด คือชุดแรก ข้อมูลใช้ในการฝึกฝนระบบ (Training Data) 90% และ ชุดที่สอง ข้อมูลใช้ในการทดสอบระบบ (Testing Data) 10%

#### 4.1.2 ขั้นตอนการทดลอง

การสังเคราะห์เสียงพูดที่มีอยู่จะนำเอาเสียงในหน่วยย่อยๆ เช่น คำ พยางค์ และ หน่วยย่อยของเสียง (Phoneme) มาเรียงต่อกัน ซึ่งทำให้เสียงที่สังเคราะห์ได้ไม่เป็นธรรมชาติเพราะขาดความต่อเนื่องระหว่างรอยต่อของหน่วยย่อยนั้น

การออกแบบทำนองเสียงภาษาไทย งานวิจัยนี้เสนอหลักการบัญญัติประติสฐ์ เรื่องการเรียนรู้ของเครื่องจักร (Machine Learning) เข้ามาช่วยในการสร้างแบบจำลองของทำนองเสียงภาษาไทย ในแบบต่างๆ โดยพิจารณาในระดับหน่วยย่อยของเสียง (Phoneme)

ผู้วิจัยได้ทำการออกแบบวิธีการสกัดลักษณะ (Feather Extraction) จากสัญญาณเสียงพูดที่เหมาะสมต่อประโยคทำนองเสียงภาษาไทย โดยพิจารณาข้อมูลทั้งในส่วนภายในหน่วยย่อยของเสียง (Phoneme) และระหว่างหน่วยย่อยของเสียง (Phoneme) งานวิจัยนี้ได้ลักษณะสำคัญ (Feature) ออกเป็น 12 พารามิเตอร์ (Parameters) เพื่อนำมาใช้ในการฝึกฝนและทดสอบระบบ

#### 4.1.2.1. การสกัดค่าลักษณะสำคัญ (Feature Extraction)

การทดลองนี้จะทำนายค่าความถี่มูลฐานของเสียงพูดภาษาไทย ซึ่งจะมีการกำหนดค่าลักษณะสำคัญ (Feature) ให้สอดคล้องกับลักษณะประโยคเสียงพูดภาษาไทย โดยในการทดลองได้มีการแบ่งกลุ่มที่พิจารณาลักษณะข้อมูลเป็น 3 ระดับ ตามลักษณะของเสียง ดังนี้

- (1) ระดับเริ่มต้น(First Level) จะใช้ข้อมูลทางด้านหน่วยเสียง (Phoneme)
- (2) ระดับที่สอง(Second Level) จะใช้ข้อมูลการแบ่งระดับข้อความ
- (3) ระดับที่สาม(Third Level) จะใช้ข้อมูลวรรณยุกต์ (Tone) ของพยางค์

งานวิจัยนี้ได้แบ่งลักษณะสำคัญ (Feature) ออกเป็น 12 พารามิเตอร์ (Parameters) เพื่อนำมาใช้ในการฝึกฝนและทดสอบระบบ ได้แก่

- (1) PhType คือ ชนิดของหน่วยเสียง (Phoneme)
- (2) Voiced คือ เสียงก้อง (Voiced) เสียงไม่ก้อง (Unvoiced)
- (3) Ph คือ หน่วยเสียง (Phoneme)
- (4) PhL คือ หน่วยเสียง (Phoneme) ซ้าย
- (5) PhR คือ หน่วยเสียง (Phoneme) ขวา
- (6) PosPh คือ ตำแหน่งของหน่วยเสียง (Phoneme) ในพยางค์
- (7) VowType คือ ชนิดของสระ (Vowel)
- (8) PosSyl คือ ตำแหน่งของพยางค์ (Syllable) ในคำ
- (9) PosWrd คือ ตำแหน่งของคำ (Word) ในวลี (Phrase)
- (10) T คือ วรรณยุกต์ (Tone) ของพยางค์
- (11) TL คือ วรรณยุกต์ (Tone) ของพยางค์ซ้าย
- (12) TR คือ วรรณยุกต์ (Tone) ของพยางค์ขวา



#### 4.1.2.2. การออกแบบทำนองเสียงภาษาไทย

งานวิจัยนี้เสนอหลักการปัญหาประดิษฐ์ เรื่องการเรียนรู้ของเครื่องจักร (Machine Learning) เข้ามาช่วยในการสร้างแบบจำลอง โดยพิจารณาในระดับหน่วยย่อยของเสียง (Phoneme) สำหรับเครื่องมือที่นำมาใช้ในการฝึกฝนและทดสอบระบบ ในวิทยานิพนธ์นี้ใช้การฝึกฝนแบบมีผู้สอน (Supervised Learning) โดยใช้เทคนิคทางด้านปัญหาประดิษฐ์ คือ เทคนิคต้นไม้ตัดสินใจ (Decision Tree) ชื่อว่า Classification and Regression Trees (CART) มาช่วยในการเรียนรู้โดยแบ่งการทำงานเป็น 2 ชุด

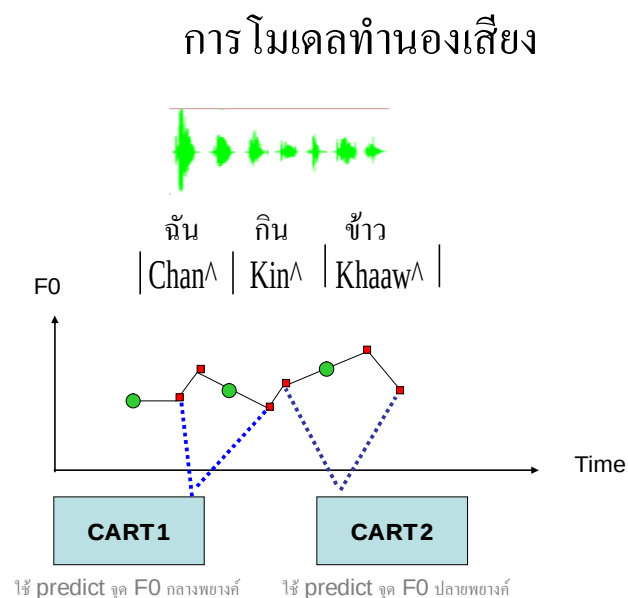
(1) CART 1 ใช้ทำนาย (Predict) จุด F0 กลางพยางค์

(2) CART 2 ใช้ทำนาย (Predict) จุด F0 ปลายพยางค์

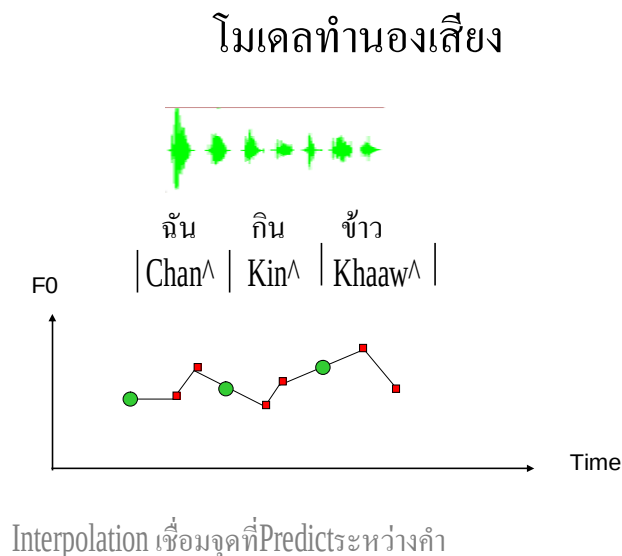
หลังจากนั้นจะทำการเชื่อมจุดที่ทำนาย (Predict) ระหว่างคำ (Interpolation)

ภาพที่ 4.2

การทำนายค่าความถี่มูลฐาน



ภาพที่ 4.3  
การเชื่อมจุดที่ทำนายค่าความถี่มูลฐาน



#### 4.1.2.3. การเรียนรู้และทดสอบระบบ

ในการทดลองจะแบ่งข้อมูลเป็น 2 ชุด คือชุดแรก ข้อมูลใช้ในการฝึกฝนระบบ (Training Data) ในสัดส่วนร้อยละ 90 ของเสียงทั้งหมด และ ชุดที่สอง ข้อมูลใช้ในการทดสอบระบบ (Testing Data) ในสัดส่วนร้อยละ 10 ของเสียงทั้งหมด

สำหรับเครื่องมือที่นำมาใช้ในการฝึกฝนและทดสอบระบบ ในวิทยานิพนธ์นี้ใช้การฝึกฝนแบบมีผู้สอน (Supervised Learning) โดยใช้เทคนิคทางด้านปัญญาประดิษฐ์ คือ เทคนิคต้นไม้ตัดสินใจ (Decision Tree) ชื่อว่า Classification and Regression Trees (CART) มาช่วยในการเรียนรู้ ซึ่งจะใช้โปรแกรม “Wagon” โดยตัวอย่างการใช้โปรแกรมที่ใช้ในการเรียนรู้ระบบ ระบุไว้ในภาคผนวก ข

ผู้วิจัยได้ทำการออกแบบการทำงานของระบบ โดยแบ่งการทำงานเป็น 2 ขั้นตอน คือ

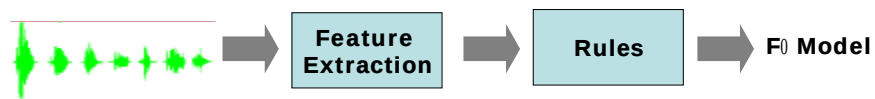
- (1) ขั้นตอนการเรียนรู้ระบบ (Training)
- (2) ขั้นตอนการทดสอบระบบ (Testing)

ภาพที่ 4.4  
ขั้นตอนการเรียนรู้และทดสอบระบบ

#### Training

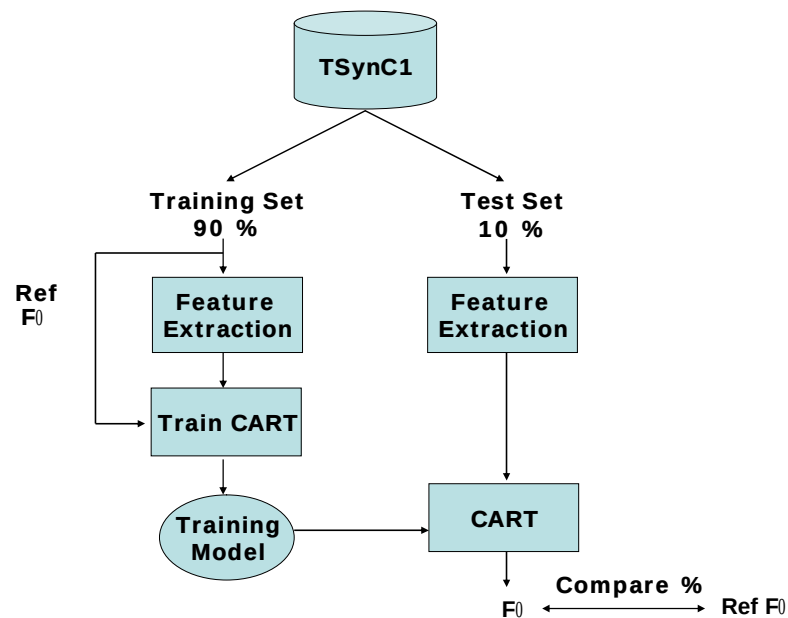


#### Testing



เมื่อฝึกฝนและทดสอบระบบแล้ว ก็จะทำการวิเคราะห์ผลและแก้ไขข้อบกพร่องของระบบต่อไปเพื่อพัฒนาการเรียนรู้ระบบให้มีประสิทธิภาพมากขึ้น

ภาพที่ 4.5  
การทำงานของระบบ



การทดลองนี้ได้ทำการออกแบบการประเมินผลลัพธ์โดยการวิเคราะห์สหสัมพันธ์ (Correlation Analysis) ระหว่างค่าความถี่มูลฐานของทำนองเสียงที่สังเคราะห์ออกมา กับค่าความถี่มูลฐานของทำนองเสียงของเสียงต้นฉบับ

#### 4.1.3 การประเมินผลลัพธ์

ในวิทยานิพนธ์นี้ จะใช้ระบบการสังเคราะห์เสียงพูดภาษาไทย “วาจา” (Vaja) เป็นเครื่องมือในการสังเคราะห์เสียง ซึ่งถูกพัฒนาโดย ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ โดยผลการวิจัยจะเป็นการเปรียบเทียบระหว่างงานวิจัยที่เสนอ กับ “วาจา” (Vaja)

การวิจัยนี้ได้ทำการออกแบบการประเมินผลการทดลองโดยการวิเคราะห์สหสัมพันธ์ (Correlation Analysis) และค่าความผิดพลาดกำลังสองเฉลี่ย (Root Mean Square Error, RMSE) ระหว่างทำนองเสียงที่สังเคราะห์ออกมา กับทำนองเสียงต้นฉบับ

## 4.2 การคำนวณค่าผลการทดลอง

ในการวัดผลงานวิจัยนี้ใช้การคำนวณค่าเพื่อแสดงถึงประสิทธิภาพ ได้แก่ การประเมินผลลัพธ์โดยการวิเคราะห์สหสัมพันธ์ (Correlation Analysis) และค่าความผิดพลาดกำลังสองเฉลี่ย (Root Mean Square Error, RMSE)

### 4.2.1 การวิเคราะห์สหสัมพันธ์ (Correlation Analysis)

เป็นค่าความถูกต้องระหว่างค่าความถี่มูลฐานของทำนองเสียงที่ทำนายออกมากับทำนองเสียงของเสียงต้นฉบับ

$$Correlation = \frac{n}{N} \quad (4.1)$$

โดยที่

$n$  คือ จำนวนข้อมูลทั้งหมดที่ทำนายถูกต้อง

$N$  คือ จำนวนข้อมูลทั้งหมด

### 4.2.2 ค่าความผิดพลาดกำลังสองเฉลี่ย (Root Mean Square Error, RMSE)

นำมาใช้ในการพิจารณาค่าความผิดพลาดระหว่างค่าความถี่มูลฐานของทำนองเสียงที่ทำนายออกมากับทำนองเสียงของเสียงต้นฉบับ

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (Fo_{ref_i} - Fo_{predict_i})^2}{N}} \quad (4.2)$$

โดยที่

$Fo_{ref_i} - Fo_{predict_i}$  คือ ค่าความผิดพลาดจากผลต่างของค่าค่าความถี่มูลฐานที่แท้จริง กับค่าความถี่มูลฐานที่ได้จากการทำนายโดยเทคนิคต้นไม้ตัดสินใจ และเทคนิคข่ายงานประสาทเทียม

$Fo_{ref_i}$  คือ ค่าความถี่มูลฐานที่แท้จริงที่ได้จากคลังข้อมูลเสียง

$Fo_{predict_i}$  คือ ค่าความถี่มูลฐานที่ได้จากการทำนายโดยเทคนิคต้นไม้ตัดสินใจ และเทคนิคข่ายงานประสาทเทียม

$N$  คือ จำนวนข้อมูลทั้งหมด

### 4.3 ผลการทดลอง

การทดลองนี้ได้ทำการประเมินผลลัพธ์โดยการวิเคราะห์สหสัมพันธ์ (Correlation Analysis) และค่าความผิดพลาดกำลังสองเฉลี่ย (Root Mean Square Error, RMSE) ระหว่าง ทำนองเสียงที่สังเคราะห์ออกมากับทำนองเสียงของเสียงต้นฉบับ ซึ่งได้ผลดังนี้

#### ตารางที่ 4.1

เปรียบเทียบผลการทดลองระหว่างค่าที่ทำนาย 2 จุดในพยางค์

| F0 range | RMSE  |       | Correlation |       |
|----------|-------|-------|-------------|-------|
|          | Cart1 | Cart2 | Cart1       | Cart2 |
| 20       | 48.15 | 47.67 | 0.18        | 0.18  |
| 50       | 50.54 | 50.49 | 0.40        | 0.40  |
| 100      | 57.94 | 57.98 | 0.67        | 0.67  |
| 150      | 78.73 | 77.65 | 0.73        | 0.73  |

เปรียบเทียบผลการทดลองระหว่างค่าความถี่มูลฐานที่ทำนาย 2 จุดในพยางค์ ระหว่าง ค่าความถี่กลางพยางค์ และ ค่าความถี่ปลายพยางค์ โดยการวิเคราะห์สหสัมพันธ์ (Correlation

Analysis) และค่าความผิดพลาดกำลังสองเฉลี่ย (Root Mean Square Error, RMSE) ระหว่างค่าที่ทำนายได้กับค่าจริงที่ใช้ในการทดลอง ให้ผลลัพธ์การทำนายทั้ง 2 จุดใกล้เคียงกัน

จากตารางที่ 4.1 การทดลองได้ทำการแบ่งช่วงความถี่มูลฐาน (F0 range) ออกเป็น 4 รูปแบบ คือ ช่วงความถี่ 20 , 50 , 100 และ 150 ผลจากการทดลอง แสดงให้เห็นว่าช่วงความถี่มาก ๆ ผลลัพธ์ของการวิเคราะห์สหสัมพันธ์ (Correlation Analysis) ยิ่งให้ค่าความถูกต้องสูงและให้ค่าความผิดพลาดกำลังสองเฉลี่ย (Root Mean Square Error, RMSE) ต่ำ แต่ถ้าค่าช่วงความถี่แคบ ๆ ผลลัพธ์ของการวิเคราะห์สหสัมพันธ์ (Correlation Analysis) จะให้ค่าความถูกต้องต่ำและให้ค่าความผิดพลาดกำลังสองเฉลี่ย (Root Mean Square Error, RMSE) สูง

จากผลของการแบ่งช่วงความถี่มูลฐาน (F0 range) ที่ใช้ในการทดลอง แสดงให้เห็นว่าการแบ่งช่วงความถี่มูลฐาน (F0 range) มีผลต่อการทำนายค่าความถี่มูลฐาน (F0 Prediction)

สำหรับการประเมินผลการทดลองโดยการวิเคราะห์สหสัมพันธ์ (Correlation Analysis) เปรียบเทียบระหว่างเสียงที่สังเคราะห์ได้จากระบบสังเคราะห์เสียงพูดวาจา (VAJA) กับเสียงของเสียงต้นฉบับ และ ค่าที่ทำนายได้ (Proposed Values) คือ เสียงที่สังเคราะห์ออกมากับเสียงของเสียงต้นฉบับ ได้ผลลัพธ์ดังนี้

#### ตารางที่ 4.2

ผลการทดลองระหว่างระบบสังเคราะห์เสียงพูดวาจาและค่าที่ทำนายได้

| F0 range | RMSE   |          | Correlation |          |
|----------|--------|----------|-------------|----------|
|          | VAJA   | Proposed | VAJA        | Proposed |
| 20       | 243.63 | 48.15    | 0.0074      | 0.18     |
| 50       | 243.05 | 50.54    | 0.0296      | 0.40     |
| 100      | 250.92 | 57.94    | 0.0667      | 0.67     |
| 150      | 228.40 | 78.73    | 0.0815      | 0.73     |

เปรียบเทียบผลการทดลองระหว่างเสียงที่สังเคราะห์ได้จากระบบสังเคราะห์เสียงพูดวาจา (VAJA) กับเสียงของเสียงต้นฉบับ และ ค่าที่ทำนายได้ (Proposed Values) ผลการทดลองค่าที่ทำนายได้ (Proposed Values) ให้ค่าสหสัมพันธ์ความถูกต้อง (Correlation Analysis) สูงกว่า

เสียงที่สังเคราะห์ได้จากระบบสังเคราะห์เสียงพูดวาจา (VAJA) และให้ค่าความผิดพลาดกำลังสองเฉลี่ย (Root Mean Square Error, RMSE) ต่ำกว่า

จากผลการทดลองแสดงให้เห็นว่าการวิจัยในส่วนสัทสัมพันธ์โดยใช้เครื่องมือในการเรียนรู้ต้นไม้ตัดสินใจ (Decision Tree) ให้ผลการทำนายดีกว่า ระบบสังเคราะห์เสียงในรูปแบบเดิมมาก

การทดลองได้ทำการเปรียบเทียบประสิทธิภาพเครื่องมือที่ใช้ในการทำวิจัย คือ เครื่องมือในการเรียนรู้ต้นไม้ตัดสินใจ (Decision Tree) กับ เครื่องมือการเรียนรู้ข่ายงานประสาทเทียมแบบแพร่กระจายย้อนกลับหลายชั้น (Multilayer Backpropagation Neural Network) ซึ่งให้ผลการทดลองดังนี้

### ตารางที่ 4.3

เปรียบเทียบผลการทดลองในแต่ละวิธีการเรียนรู้

| Method | RMSE   | Correlation |
|--------|--------|-------------|
| CART   | 57.935 | 0.6691      |
| C4.5   | 57.212 | 0.6736      |
| NN     | 60.008 | 0.6535      |

เปรียบเทียบผลการทดลองระหว่างเครื่องมือในการเรียนรู้ต้นไม้ตัดสินใจ (Decision Tree) กับ เครื่องมือการเรียนรู้ข่ายงานประสาทเทียมแบบแพร่กระจายย้อนกลับหลายชั้น (Multilayer Backpropagation Neural Network) โดยการวิเคราะห์สหสัมพันธ์ (Correlation Analysis) และค่าความผิดพลาดกำลังสองเฉลี่ย (Root Mean Square Error, RMSE) ผลการทดลองโดยใช้เครื่องมือในการเรียนรู้ต้นไม้ตัดสินใจ (Decision Tree) ให้ผลการวิเคราะห์สหสัมพันธ์ (Correlation Analysis) สูงกว่า และค่าความผิดพลาดกำลังสองเฉลี่ย (Root Mean Square Error, RMSE) ต่ำกว่าเครื่องมือการเรียนรู้ข่ายงานประสาทเทียมแบบแพร่กระจายย้อนกลับหลายชั้น (Multilayer Backpropagation Neural Network)



## บทที่ 5

### สรุปผลการวิจัยและข้อเสนอแนะ

งานวิจัยนี้นำเสนอแนวทางการประยุกต์ใช้คลังข้อมูลเสียง และเครื่องมือปัญญาประดิษฐ์ มาช่วยในการทำนายค่าความถี่มูลฐาน และต้องการทดสอบสมมติฐาน 2 ข้อ ได้แก่ ข้อมูลความถี่มูลฐานจากคลังข้อมูลเสียง สามารถใช้ทำนายค่าความถี่มูลฐานสำหรับการสังเคราะห์เสียงพูดภาษาไทย และเทคนิคทางด้านปัญญาประดิษฐ์ที่ใช้โดยการเรียนรู้ต้นไม้ตัดสินใจ (Decision Tree) ในการทำนายค่าความถี่มูลฐานมีความถูกต้องมากกว่าเทคนิคข่ายงานข่ายงานประสาทเทียมแบบแพร่กระจายย้อนกลับหลายชั้น (Multilayer Backpropagation Neural Network) ซึ่งผลการวิเคราะห์ที่ได้จากบทที่แล้ว สามารถสรุปผลได้ดังนี้

#### 5.1 สรุปผลการวิจัย

จากตารางที่ 4.1 การทดลองได้ทำการแบ่งช่วงความถี่มูลฐาน (F0 range) ออกเป็น 4 รูปแบบ คือ ช่วงความถี่ 20 , 50 , 100 และ 150 ผลจากการทดลอง แสดงให้เห็นว่าช่วงความถี่มาก ๆ ผลลัพธ์ของการวิเคราะห์สหสัมพันธ์ (Correlation Analysis) ยิ่งให้ค่าความถูกต้องสูงและให้ค่าความผิดพลาดกำลังสองเฉลี่ย (Root Mean Square Error, RMSE) ต่ำ แต่ถ้าวัดช่วงความถี่แคบ ๆ ผลลัพธ์ของการวิเคราะห์สหสัมพันธ์ (Correlation Analysis) จะให้ค่าความถูกต้องต่ำและให้ค่าความผิดพลาดกำลังสองเฉลี่ย (Root Mean Square Error, RMSE) สูง

จากผลของการแบ่งช่วงความถี่มูลฐาน (F0 range) ที่ใช้ในการทดลอง แสดงให้เห็นว่าการแบ่งช่วงความถี่มูลฐาน (F0 range) มีผลต่อการทำนายค่าความถี่มูลฐาน (F0 Prediction) เนื่องจากว่าถ้ามีการแบ่งช่วงความถี่มูลฐาน (F0 range) กว้างขึ้น จะทำให้ความน่าจะเป็นในการทำนายผลอย่างถูกต้องมีสูงกว่ เช่น ถ้ามีช่วงความถี่รวมทั้งหมด ตั้งแต่ 1 ถึง 600

เมื่อมีการแบ่งช่วงความถี่มูลฐาน (F0 range) เป็น 20 จะทำให้ความน่าจะเป็นในการทำนายผลอย่างถูกต้อง เท่ากับ  $20/600 = 0.034$

เมื่อมีการแบ่งช่วงความถี่มูลฐาน (F0 range) เป็น 50 จะทำให้ความน่าจะเป็นในการทำนายผลอย่างถูกต้อง เท่ากับ  $50/600 = 0.084$

เมื่อมีการแบ่งช่วงความถี่มูลฐาน (F0 range) เป็น 100 จะทำให้ความน่าจะเป็นในการทำนายผลอย่างถูกต้อง เท่ากับ  $100/600 = 0.167$

เมื่อมีการแบ่งช่วงความถี่มูลฐาน (F0 range) เป็น 20 จะทำให้ความน่าจะเป็นในการทำนายผลอย่างถูกต้อง เท่ากับ  $150/600 = 0.250$

จะเห็นได้ว่าความน่าจะเป็นในการทำนายผลอย่างถูกต้อง ในการแบ่งช่วงความถี่มูลฐาน (F0 range) กว้างขึ้น จะทำให้ความน่าจะเป็นในการทำนายผลอย่างถูกต้องมีสูงกว่า

ดังนั้นจะเห็นได้ว่าผลลัพธ์ของการทดลอง จากตารางที่ 4.1 โดยการทดลองได้ทำการแบ่งช่วงความถี่มูลฐาน (F0 range) ออกเป็น 4 รูปแบบ คือ ช่วงความถี่ 20 , 50 , 100 และ 150 มีความสอดคล้องกับค่าความน่าจะเป็นที่ควรจะได้

การทดลองมีการใช้เทคนิคต้นไม้ตัดสินใจ (Decision Tree) โดยเลือกใช้วิธี Classification and Regression Trees (CART) และ C4.5 มาทำการทดลอง ผลการทดลองโดยใช้วิธี Classification and Regression Trees (CART) ให้ผลการวิเคราะห์สหสัมพันธ์ (Correlation Analysis) ต่ำกว่า และค่าความผิดพลาดกำลังสองเฉลี่ย (Root Mean Square Error, RMSE) สูงกว่าวิธี C4.5 ไม่มากนัก

ดังนั้นจะเห็นได้ว่าผลลัพธ์ของการทดลอง จากตารางที่ 4.3 วิธี Classification and Regression Trees (CART) และ C4.5 ให้ผลการวิเคราะห์สหสัมพันธ์ (Correlation Analysis) และค่าความผิดพลาดกำลังสองเฉลี่ย (Root Mean Square Error, RMSE) ไม่ต่างกันมากนัก

จากตารางที่ 4.3 ได้ทำการเปรียบเทียบผลการทดลองระหว่างเครื่องมือในการเรียนรู้ต้นไม้ตัดสินใจ (Decision Tree) กับ เครื่องมือการเรียนรู้ข่ายงานประสาทเทียมแบบแพร่กระจายย้อนกลับหลายชั้น (Multilayer Backpropagation Neural Network) โดยการวิเคราะห์สหสัมพันธ์ (Correlation Analysis) และค่าความผิดพลาดกำลังสองเฉลี่ย (Root Mean Square Error, RMSE) ผลการทดลองโดยใช้เครื่องมือในการเรียนรู้ต้นไม้ตัดสินใจ (Decision Tree) ให้ผลการวิเคราะห์สหสัมพันธ์ (Correlation Analysis) สูงกว่า และค่าความผิดพลาดกำลังสองเฉลี่ย (Root Mean Square Error, RMSE) ต่ำกว่าเครื่องมือการเรียนรู้ข่ายงานประสาทเทียมแบบแพร่กระจายย้อนกลับหลายชั้น (Multilayer Backpropagation Neural Network)

จากผลการทดลองแสดงให้เห็นการใช้เทคนิคการเรียนรู้ต้นไม้ตัดสินใจ (Decision Tree) จะสามารถจำแนกลักษณะเด่นและทำนายค่าความถี่มูลฐานได้ดีกว่าเทคนิคข่ายงานประสาทเทียม (Neural Network) เนื่องจากว่าเทคนิคการเรียนรู้ต้นไม้ตัดสินใจ (Decision Tree) เป็นการเรียนรู้อย่างเป็นเหตุเป็นผล ในการจำแนกลักษณะเด่นของกลุ่มข้อมูลที่นำมาเรียนรู้และทดสอบ แต่เครื่องมือการเรียนรู้ข่ายงานประสาทเทียมแบบแพร่กระจายย้อนกลับหลายชั้น (Multilayer Backpropagation Neural Network) กลับให้ผลลัพธ์ของการทำนายความถูกต้องต่ำ

กว่า เนื่องจากการเรียนรู้แบบเทคนิคข่ายงานประสาทเทียม (Neural Network) จะทำให้การเรียนรู้ของระบบมีมากขึ้นไป เมื่อนำมาใช้กับฐานข้อมูลที่มีจำกัด จึงทำให้ผลลัพธ์ของการทำนายค่าความถูกต้องออกมาต่ำกว่า

### 5.1.1 สรุปผลการทดสอบสมมติฐานวิจัยข้อที่ 1

ข้อมูลความถี่มูลฐานจากคลังข้อมูลเสียง สามารถใช้ทำนายค่าความถี่มูลฐานสำหรับการสังเคราะห์เสียงพูดภาษาไทย

จากการเปรียบเทียบผลการทดลองที่ได้ในบทที่ 4 ระหว่างเสียงที่สังเคราะห์ได้จากระบบสังเคราะห์เสียงพูดวาจา (VAJA) กับเสียงของเสียงต้นฉบับ และ ค่าที่ทำนายได้ (Proposed Values) ผลการทดลองค่าที่ทำนายได้ (Proposed Values) ให้ค่าสหสัมพันธ์ความถูกต้อง (Correlation Analysis) สูงกว่า เสียงที่สังเคราะห์ได้จากระบบสังเคราะห์เสียงพูดวาจา (VAJA) และให้ค่าความผิดพลาดกำลังสองเฉลี่ย (Root Mean Square Error, RMSE) ต่ำกว่า

จากผลการทดลองแสดงให้เห็นว่าการวิจัยในส่วนสหสัมพันธ์โดยใช้คลังข้อมูลเสียงมาช่วยในการเรียนรู้ระบบ สามารถช่วยให้การทำนายค่าความถี่มูลฐานของระบบดีขึ้น

### 5.1.2 สรุปผลการทดสอบสมมติฐานวิจัยข้อที่ 2

เทคนิคทางด้านปัญญาประดิษฐ์ที่ใช้โดยการเรียนรู้ต้นไม้ตัดสินใจ (Decision Tree) ในการทำนายค่าความถี่มูลฐานมีความถูกต้องมากกว่าเทคนิคข่ายงานข่ายงานประสาทเทียมแบบแพร่กระจายย้อนกลับหลายชั้น (Multilayer Backpropagation Neural Network)

จากการเปรียบเทียบผลการทดลองที่ได้ในบทที่ 4 เปรียบเทียบผลการทดลองระหว่างเครื่องมือในการเรียนรู้ต้นไม้ตัดสินใจ (Decision Tree) กับ เครื่องมือการเรียนรู้ข่ายงานประสาทเทียมแบบแพร่กระจายย้อนกลับหลายชั้น (Multilayer Backpropagation Neural Network) โดยการวิเคราะห์สหสัมพันธ์ (Correlation Analysis) และค่าความผิดพลาดกำลังสองเฉลี่ย (Root Mean Square Error, RMSE) ผลการทดลองโดยใช้เครื่องมือในการเรียนรู้ต้นไม้ตัดสินใจ (Decision Tree) ให้ผลการวิเคราะห์สหสัมพันธ์ (Correlation Analysis) สูงกว่า และค่าความผิดพลาดกำลังสองเฉลี่ย (Root Mean Square Error, RMSE) ต่ำกว่าเครื่องมือการเรียนรู้ข่ายงานประสาทเทียมแบบแพร่กระจายย้อนกลับหลายชั้น (Multilayer Backpropagation Neural Network)

จากผลการทดลองแสดงให้เห็นว่าเครื่องมือในการเรียนรู้สามารถช่วยในการทำนายค่าความถี่มูลฐานได้ค่าความถูกต้องสูง เหมาะกับในกรณีที่ฐานข้อมูลในคลังข้อมูลมีจำกัด เพราะจะช่วยให้การสกัดค่าลักษณะสำคัญของข้อมูล แต่จะเห็นได้ว่าจากผลการทดลองซึ่งทดลองกับคลังข้อมูลที่มีขนาดข้อมูลจำกัด จากการใช้เทคนิคการเรียนรู้ต้นไม้ตัดสินใจ (Decision Tree) จะสามารถจำแนกลักษณะเด่นและทำนายค่าความถี่มูลฐานได้ดีกว่าเทคนิคข่ายงานประสาทเทียม (Neural Network) เนื่องจากว่าเทคนิคการเรียนรู้ต้นไม้ตัดสินใจ (Decision Tree) เป็นการเรียนรู้อย่างเป็นเหตุเป็นผล ในการจำแนกลักษณะเด่นของกลุ่มข้อมูลที่นำมาเรียนรู้และทดสอบ แต่เครื่องมือการเรียนรู้ข่ายงานประสาทเทียมแบบแพร่กระจายย้อนกลับหลายชั้น (Multilayer Backpropagation Neural Network) กลับให้ผลลัพธ์ของการทำนายความถูกต้องต่ำกว่า เนื่องจากการเรียนรู้แบบเทคนิคข่ายงานประสาทเทียม (Neural Network) จะทำให้การเรียนรู้ของระบบมีมากเกินไป เมื่อนำมาใช้กับฐานข้อมูลที่มีจำกัด จึงทำให้ผลลัพธ์ของการทำนายค่าความถูกต้องออกมาต่ำกว่า

## 5.2 ข้อเสนอแนะ

ส่วนนี้นำเสนอข้อเสนอแนะ เพื่อเป็นแนวทางในการศึกษาและพัฒนางานวิจัยที่เกี่ยวข้องต่อไปในอนาคต

ในการทำนายค่ายังสามารถปรับปรุงคุณภาพในการทำนายได้อีก ดังนั้นควรปรับปรุงวิธีการดังลักษณะสำคัญ

การเลือกใช้เครื่องมือในการเรียนรู้ควรได้มีการทดลองกับคลังข้อมูลขนาดใหญ่มากกว่านี้ ในการเปรียบเทียบผลการทดลองระหว่างเทคนิคการเรียนรู้ต้นไม้ตัดสินใจ (Decision Tree) และเทคนิคข่ายงานประสาทเทียม (Neural Network) เนื่องจากการเรียนรู้มากๆของเทคนิคข่ายงานประสาทเทียม (Neural Network) อาจทำให้กระบวนการเรียนรู้และทดสอบของระบบมีความถูกต้องมากขึ้น หรืออาจมีผลการทำนายแม่นยำมากกว่าเทคนิคการเรียนรู้ต้นไม้ตัดสินใจ (Decision Tree)

จากการทดลองสามารถทำนายค่าความถี่มูลฐานได้ และมีค่าความถูกต้องสูง น่าจะนำมาปรับใช้กับระบบการสังเคราะห์เสียงจริงในปัจจุบัน

ภาคผนวก

## ภาคผนวก ก

### ตัวอย่างข้อมูลในคลังข้อมูลเสียง

Thai Tagged Speech Corpus for Speech Synthesis (TSynC1) เป็นคลังข้อมูลเสียงขนาดใหญ่ที่จะจัดเก็บฐานข้อมูลเสียง เพื่อใช้ในการสังเคราะห์สัญญาณเสียงพูด มีการเขียนแบบแสดงเสียงอ่านและข้อมูลเกี่ยวกับเสียงเพื่อที่จะนำไปพัฒนารูปแบบสัทสัมพันธ์การพูดภาษาไทยต่อไป โดยแบบแสดงเสียงอ่านประกอบด้วย รูปอ่าน, สัญลักษณ์แสดงเสียง, ขอบเขตของเสียงพูด, ตำแหน่งของพยางค์, เสียงพูด, เสียงสูงต่ำ และข้อมูลเกี่ยวกับเสียงได้แก่ ค่าพลังงาน, ระดับเสียง, การแยกช่วงของเสียง และเส้นเสียง ซึ่งในฐานข้อมูลเสียงนี้ประกอบด้วยประโยค 5,200 ประโยค ซึ่งครอบคลุมหมดทุกเสียงทั้งคำไทย และคำที่ยืมมาจากภาษาต่างประเทศ ซึ่งถูกพัฒนาโดย ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ

#### ภาพที่ ก.1

แสดงขอบเขตข้อมูลในคลังข้อมูล “TSynC1”

```
Directory of E:\TSynC1
[.]          [..]          [lab_phASRT] [f0_pitch]    [f0_sm]      [lab_ph]
[lab_syl]    [wav]          [lab_phASR]
              0 File(s)      0 bytes
              9 Dir(s)  26,382,237,696 bytes free
```

ในคลังข้อมูล “TSynC1” ประกอบด้วยข้อมูลดังนี้

|            |                                          |
|------------|------------------------------------------|
| f0_pitch   | บอก Pitch Mark                           |
| f0_sm      | บอก F0 แต่ละ Window                      |
| lab_ph     | บอกขอบเขตของ Phoneme                     |
| lab_phASR  | บอกขอบเขตของ Phoneme                     |
| lab_phASRT | บอกขอบเขตของ Phoneme และ Tone เสียงกำกับ |
| lab_syl    | บอกขอบเขตของ Syllable                    |
| wav        | แสดงเสียงอ่าน                            |

f0\_pitch

File type = "ooTextFile short"  
"PointProcess"

0  
8.33125  
1346  
0.1848441722972973  
0.18775459381044488  
0.19120254994742375  
0.19461199874686716  
0.19784503682624707  
0.2017622720459655  
0.2052383275092215  
0.20868331096196868  
0.21207980039016233  
0.21544736470809556  
0.2188102057478772  
0.22218516810597802  
0.2255744700984103  
0.22894807797413008  
0.2324375  
0.23575820386234936  
0.2391584465005223  
0.2425144380221488  
0.24588131284277878  
0.24920031283280084  
0.25260614600107356



f0\_sm

File type = "ooTextFile short"  
"PitchTier"

0  
8.33125  
1873  
0.0034147412026750284  
292.8479614257817  
0.006829482405350043  
292.8479614257817  
0.010244223608025058  
292.8479614257817  
0.013658964810700072  
292.8479614257817  
0.017073706013375087  
292.8479614257817  
0.0204884472160501  
292.8479614257817  
0.023903188418725116  
292.8479614257817  
0.02731792962140013  
292.8479614257817  
0.030732670824075145  
292.8479614257817  
0.03414741202675016  
292.8479614257817  
0.037562153229425174

lab\_ph

0 1576250 sil  
1576250 2257500 pr  
2257500 2677500 a  
2677500 3286250 th  
3286250 4382500 ee  
4382500 4786250 t^  
4786250 5378750 th  
5378750 6418750 a  
6418750 7342500 j^  
7342500 7837500 d  
7837500 8202500 aa  
8202500 8558750 j^  
8558750 9213750 m  
9213750 9955000 ii  
9955000 10432500 k  
10432500 11683750 aa  
11683750 12451250 n^  
12451250 13163750 pr  
13163750 13516250 a  
13516250 14315000 p^  
14315000 15235000 pl  
15235000 16130000 iia  
16130000 17035000 n^  
17035000 17822500 khr  
17822500 18392500 oo  
18392500 19197500 ng^  
19197500 20090000 s

lab\_phASR

0 1576250 sil  
1576250 2257500 pr  
2257500 2677500 a  
2677500 3286250 th  
3286250 4382500 ee  
4382500 4786250 t^  
4786250 5378750 th  
5378750 6418750 a  
6418750 7342500 j^  
7342500 7837500 d  
7837500 8202500 aa  
8202500 8558750 j^  
8558750 9213750 m  
9213750 9955000 ii  
9955000 10432500 k  
10432500 11683750 aa  
11683750 12451250 n^  
12451250 13163750 pr  
13163750 13516250 a  
13516250 14315000 p^  
14315000 15235000 pl  
15235000 16130000 iia  
16130000 17035000 n^  
17035000 17822500 khr  
17822500 18392500 oo  
18392500 19197500 ng^  
19197500 20090000 s

lab\_phASRT

0 1576250 sil  
1576250 2257500 pr  
2257500 2677500 a1  
2677500 3286250 th  
3286250 4382500 ee2  
4382500 4786250 t^  
4786250 5378750 th  
5378750 6418750 a0  
6418750 7342500 j^  
7342500 7837500 d  
7837500 8202500 aa2  
8202500 8558750 j^  
8558750 9213750 m  
9213750 9955000 ii0  
9955000 10432500 k  
10432500 11683750 aa0  
11683750 12451250 n^  
12451250 13163750 pr  
13163750 13516250 a1  
13516250 14315000 p^  
14315000 15235000 pl  
15235000 16130000 iia1  
16130000 17035000 n^  
17035000 17822500 khr  
17822500 18392500 oo0  
18392500 19197500 ng^  
19197500 20090000 s

lab\_syl

|          |          |             |
|----------|----------|-------------|
| 0        | 1576250  | \$-\$-\$-\$ |
| 1576250  | 2677500  | pr-a-z^1    |
| 2677500  | 4786250  | th-ee-t^2   |
| 4786250  | 7342500  | th-a-j^0    |
| 7342500  | 8558750  | d-aa-j^2    |
| 8558750  | 9955000  | m-ii-z^0    |
| 9955000  | 12451250 | k-aa-n^0    |
| 12451250 | 14315000 | pr-a-p^1    |
| 14315000 | 17035000 | pl-ia-n^1   |
| 17035000 | 19197500 | kh-oo-ng^0  |
| 19197500 | 22078750 | s-aa-ng^2   |
| 22078750 | 23295000 | n-a-j^0     |
| 23295000 | 25460000 | k-aa-n^0    |
| 25460000 | 26191250 | ph-a-t^3    |
| 26191250 | 27031250 | th-a-z^3    |
| 27031250 | 28906250 | n-aa-z^0    |
| 28906250 | 30201250 | s-ee-t^1    |
| 30201250 | 31263750 | th-a-z^1    |
| 31263750 | 32856250 | k-i-t^1     |
| 32856250 | 35327500 | kh-@@-ng^4  |
| 35327500 | 36635000 | pr-a-z^1    |
| 36635000 | 40897500 | th-ee-t^2   |
| 40897500 | 43631250 | \$-\$-\$-\$ |
| 43631250 | 46076250 | c-aa-k^1    |
| 46076250 | 47152500 | pr-a-z^1    |
| 47152500 | 49188750 | th-ee-t^2   |
| 49188750 | 50101250 | k-a-z^1     |

|          |          |                        |
|----------|----------|------------------------|
| 50101250 | 51643750 | s-ee-t <sup>-1</sup>   |
| 51643750 | 52686250 | tr-a-z <sup>-1</sup>   |
| 52686250 | 56706250 | k-a-m <sup>-0</sup>    |
| 56706250 | 57916250 | p-a-j <sup>-0</sup>    |
| 57916250 | 60275000 | s-uu-z <sup>-1</sup>   |
| 60275000 | 62525000 | khw-aa-m <sup>-0</sup> |
| 62525000 | 64553750 | p-e-n <sup>-0</sup>    |
| 64553750 | 65521250 | pr-a-z <sup>-1</sup>   |
| 65521250 | 67385000 | th-ee-t <sup>-2</sup>  |
| 67385000 | 68140000 | z-u-t <sup>-1</sup>    |
| 68140000 | 70488750 | s-aa-z <sup>-4</sup>   |
| 70488750 | 71187500 | h-a-z <sup>-1</sup>    |
| 71187500 | 74218750 | k-a-m <sup>-0</sup>    |
| 74218750 | 76166250 | m-aa-k <sup>-2</sup>   |
| 76166250 | 78532500 | j-i-ng <sup>-2</sup>   |
| 78532500 | 82708750 | kh-v-n <sup>-2</sup>   |
| 82708750 | 83300000 | \$-\$-\$-\$            |

## ผนวก ข

### ตัวอย่างการใช้โปรแกรมที่ใช้ในการเรียนรู้ระบบ

การทดลองงานวิจัยจะใช้เทคนิคต้นไม้ตัดสินใจ(Decision Tree)มาช่วยในการเรียนรู้ระบบ โดยจะใช้โปรแกรม“Wagon”ซึ่งสามารถเรียนรู้ฝึกฝนระบบ และทดสอบระบบได้

ภาพที่ ข.1

แสดงขอบเขตข้อมูลในโปรแกรม“Wagon”

```
Directory of E:\WAGON
[.]          [..]          pb.data          pb.desc
pb.test      train.bat     wagon.exe       wagon_test.exe
pb.res              7 File(s)    1,279,042 bytes
                  2 Dir(s)  26,385,448,960 bytes free
```

ในโปรแกรม“Wagon” ประกอบด้วยข้อมูลดังนี้

|                |                             |
|----------------|-----------------------------|
| pb.desc        | อธิบาย Feature ของชุดข้อมูล |
| pb.data        | ข้อมูลสำหรับ Train          |
| pb.test        | ข้อมูลสำหรับ Test           |
| pb.res         | ส่วนแสดงผลลัพธ์             |
| train.bat      | ใช้สำหรับ Train             |
| wagon.exe      | โปรแกรม “Wagon”             |
| wagon_test.exe | ใช้สำหรับ Test              |

ภาพที่ ข.2

แสดงรายละเอียดการใช้งานโปรแกรม“Wagon”

```

E:\WAGON>wagon -h
Usage:
wagon <options>
Wagon CART building program <defaults in {}>
-desc <ifile>      Field description file
-data <ifile>      Datafile, one vector per line
-stop <int> <50>   Minimum number of examples for leaf nodes
-test <ifile>      Datafile to test tree on
-frs <float> <10>  Float range split, number of partitions to
                  split a float feature range into
-distmatrix <ifile>
                  A distance matrix for clustering
-dlist            Build a decision list (rather than tree)
-dtree           Build a decision tree (rather than list) default
-output <ofile>   File to save output tree in
-v              Print version number and exit
-quiet          No questions printed during building
-verbose        Lost of information printing during build
The following are *provisional* and may not work
-stepwise       Incrementally find best features
-swlimit <float> <0.0>
                  Percentage necessary improvement for stepwise
-balance <float> For derived stop size, if dataset at node, divided
                  by balance is greater than stop it is used as stop
                  if balance is 0 (default) always use stop as is.
-held_out <int>   Percent to hold out for pruning
-heap <int> <210000>
                  Set size of Lisp heap, should not normally need
                  to be changed from its default, only with *very*
                  large description files (> 1M)
-noprune        No (same class) pruning required

```

ภาพที่ ข.3

แสดงการทดสอบโปรแกรม

```

E:\WAGON>wagon_test -h
Usage:
wagon_test <options>
program to test CART trees on data
-desc <ifile>      Field description file
-data <ifile>      Datafile, one vector per line
-tree <ifile>      File containing CART tree
-predict          Predict for each vector returning distribution
-predict_val      Predict for each vector returning just value
-heap <int> <210000>
                  Set size of Lisp heap, should not normally need
                  to be changed from its default
-o <ofile>        File to save output in

```

ภาพที่ ข.4

แสดงการฝึกฝนโปรแกรม

```

E:\WAGON>train -h
E:\WAGON>wagon -desc ph.desc -data ph.data -test ph.test -quiet -heap 2000000 1
>ph.res

```



ตัวอย่างการใช้โปรแกรมที่ใช้ในการเรียนรู้ระบบ

ภาพที่ ข.5

ตัวอย่างการอธิบายFeatureของชุดข้อมูล

```
pb.desc
(
  (out play yes no)
  (outlook sunny overcast rainy)
  (temperat float)
  (humidity float)
  (windy TRUE FALSE)
)
```

ภาพที่ ข.6

ตัวอย่างข้อมูลสำหรับTrain

```
pb.data

no      sunny      85      85      FALSE
no      sunny      80      90      TRUE
yes     overcast   83      86      FALSE
yes     rainy      70      96      FALSE
yes     rainy      68      80      FALSE
no      rainy      65      70      TRUE
yes     overcast   64      65      TRUE
no      sunny      72      95      FALSE
yes     sunny      69      70      FALSE
yes     rainy      75      80      FALSE
yes     sunny      75      70      TRUE
yes     overcast   72      90      TRUE
```

ภาพที่ ข.7  
ตัวอย่างข้อมูลสำหรับTest

```
pb.test

yes    overcast    81    75    FALSE
no     rainy       71    91    TRUE
no     sunny       71    91    TRUE
no     sunny       71    91    TRUE
```

ภาพที่ ข.8  
ตัวอย่างส่วนแสดงผลลัพธ์

```
pb.res

Dataset of 12 vectors of 5 parameters from: pb.data
Dataset of 4 vectors of 5 parameters from: pb.test

(((play 0) (yes 0.666667) (no 0.333333) yes))
  pla yes
  play 0 0 0 0 [0/0] 100.000
  yes 0 1 0 1 [1/1] 100.000
  no 0 3 0 3 [0/3] 0.000
    0 4 0
total 4 correct 1.000 25.000%
```

- Allen, J., Hunnicutt, S. and Klatt, D.H., 1987, FROM TEXT TO SPEECH: THE MITALK SYSTEM, Cambridge Univ. Press, Cambridge UK (1987). (K) Complete description of MITalk system.
- Atal, B. S., Chang, J. J., Mathews, M. V. and Turkey, J. W., 1978, "Inversion of Articulatory-to-acoustic Transformation in the Vocal Tract by a Computer sorting Technique", J. Acoust. Soc. Am., 63, pp. 1535-1555.
- Breiman, L., Frieman, J.H., Olshen, R. A., and Stone, P.J., 1984, "Classification and Regression Tree.", Wadsworth International Groups, CA.
- Hansakunbuntheung, C., Tesprasit, V., and Sornlertlamvanich, V., 2003, "Thai Tagged Speech Corpus for Speech Synthesis", The Oriental COCOSDA 2003, Singapore, October, 1-3, 2003.
- Hansakunbuntheung, C., Tesprasit, V., Siricharoenchai, R. and Sagisaka, Y., 2003-9, "Analysis and Modeling of Syllable Duration for Thai Speech Synthesis", Proc. European Conference on Speech Communication and Technology, Geneva-Switzerland, pp.93-96.
- Hikichi, T., Osaka, N. and Itakura, F., 2003, "Sho-So-In: New synthesis method for addition of articulations based on a sho-type physical model", Proceedings of the 2003 International Computer Music Conference, 333-336.
- Kiat-arpakul, R., Fakcharoenphol, J. and Keretho, S., 1995, "A combined phoneme-based and demisyllable approach for Thai speech synthesis", Proceedings of the 2nd Symposium on Natural Language Processing SNLP'95, pp 361-369.

บรรณานุกรม (ต่อ)

- Klatt, D.H., 1987, "How KLATTALK became DECtalk: An academic's experiences in the business world," Speech Tech '87, New York, April 1987.
- Luksaneeyanawin, S. et., al., 1992, "A Thai text-to-speech system", Proceeding of 4th NECTEC Conference, pp.65-78.
- Luksaneeyanawin, S., 1995, "Tone transformation", Proceedings of the 2nd Symposium on Natural Language Processing SNLP'95, pp. 345-360.
- Mittrapiyanurak, P., Hansakunbuntheung, C., Tesprasit, V. and Sornlertlumvanich, V., 2000, "Issues in Thai Text-to-Speech Synthesis: The NECTEC Approach", NECTEC Technical Journal Thailand, vol. 2, No. 7, pp. 36-47, June 2000.
- Quinlan, J.R., 1986, "Induction of Decision Tree", In Readings in Machine Learning, Morgan Kaufmann Publishers In., San Francisco, CA, 1993.
- Quinlan, J.R., 1993, C4.5: programs for machine learning, Morgan Kaufmann Publishers In., San Francisco, CA, 1993.
- Taisertavattanakul, S. and Kanawaree, W., 1995, "Thai speech synthesizer", Unpublished senior project report, Department of Computer Engineering, Chulalongkorn University.
- Thangthai, A., Saichum, S., Rugchatjaroen, A., Thatphithakkul, N. and Wutiwiwatchai, C., 2007, "Recent Progress of "VAJA" Thai Text-to-Speech Synthesis", NSTDA Annual Conference (NAC 2007), March 28-30, 2007, Pathumthani, Thailand.

## ประวัติการศึกษา

|                 |                                                                                                                                                                                                                                                                                                                                    |
|-----------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| ชื่อ (ไทย)      | นางสาว เปรมวดี อุดมเรืองเกียรติ                                                                                                                                                                                                                                                                                                    |
| ชื่อ (อังกฤษ)   | Ms. Premvadee Udomruangkiat                                                                                                                                                                                                                                                                                                        |
| วุฒิการศึกษา    | บริหารธุรกิจบัณฑิต (บธ.บ.) สาขาการจัดการการผลิต<br>คณะบริหารธุรกิจ มหาวิทยาลัยเกษตรศาสตร์                                                                                                                                                                                                                                          |
| ผลงานทางวิชาการ | การทำนายค่าความถี่มูลฐานสำหรับการสังเคราะห์เสียงพูดภาษาไทย<br>(F0 Prediction for Thai Speech Synthesis)<br>เสนอผลงานและเข้าร่วมการประชุมทางวิชาการ<br>The 2008 International Conference on Embedded Systems and<br>Intelligent Technology (ICESIT 2008)<br>February 27-29, 2008, Grand Mercure Fortune Hotel, Bangkok,<br>Thailand |