สำนักหอสมุดกลาง พระจอมเกล้าลาดกระบัง

ระบบวิเคราะห์และรู้จำเสียงด้วยฮิดเดนมาร์คอฟโมเดล

SOUND ANALYSIS AND RECOGNITION SYSTEM ON HIDDEN MARKOV MODEL

T117542

ก่อเกียรติ เงินยวง

ชินพัฒน์ มัชฌิมาวิวัฒน์

พัชรพล อยู่สุขสมบูรณ์

เลขหมู่.....
เลขทะเบียน 117542
วัน,เดือน,ปี 5 ส.ค. 2554

b 12313559
i.....

ปริญญานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรบัณฑิต

สาขาวิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

ปีการศึกษา 2553

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ปริญญานิพนธ์ปีการศึกษา 2553

สาขาวิชาวิศวกรรมคอมพิวเตอร์

คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เรื่อง ระบบวิเคราะห์และรู้จำเสียงด้วยฮิดเดนมาร์คอฟโมเดล

SOUND ANALYSIS AND RECOGNITION SYSTEM ON HIDDEN MARKOV MODEL

ผู้จัดทำ

1. นายก่อเกียรติ เงินยวง รหัสนักศึกษา 50010063
2. นายชินพัฒน์ มัชฌิมาวิวัฒน์ รหัสนักศึกษา 50010379
3. นายพัชรพล อยู่สุขสมบูรณ์ รหัสนักศึกษา 50011077

_____อาจารย์ที่ปรึกษา

(อาจารย์สุภกิจ นุตยะสกุล)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ระบบวิเคราะห์และรู้จำเสียงด้วยฮิดเดนมาร์คอฟโมเดล

นายก่อเกียรติ	เงินยวง	50010063
นายชินพัฒน์	มัชฌิมาวิวัฒน์	50010379
นายพัชรพล	อยู่สุขสมบูรณ์	50011077
อาจารย์สุภกิจ	นุตยะสกุล	อาจารย์ที่ปรึกษา

ปีการศึกษา 2553

บทคัดย่อ

ฮิดเดนมาร์คอฟโมเดล (Hidden Markov Model) เป็นโมเดลทางสถิติ ใช้ในระบบรู้จำในรูปแบบต่าง ๆ มากมาย เช่น รู้จำเสียง รู้จำลายมือ รู้จำเครื่องดนตรี รวมถึงการพยากรณ์อนาคตที่อาจเกิดขึ้น เช่น พยากรณ์อากาศและพยากรณ์ตลาดหุ้น เห็นได้ว่าฮิดเดนมาร์คอฟโมเดลเป็นโมเดลที่สามารถพัฒนาใช้ประโยชน์ได้อย่างหลากหลาย

โครงงานนี้นำฮิดเดนมาร์คอฟโมเดลมาใช้ในการรู้จำเสียงพูดภาษาไทย 1 พยางค์ ตัวอย่างเสียงถูกนำมาหาลักษณะเด่นของเสียง โดยใช้ฟาสฟูเรียทรานฟอรม์จากนั้นข้อมูลที่ได้นำมาเรียนรู้โดยใช้ฮิดเดนมาร์คอฟโมเดลเก็บค่าที่ได้ภายในฐานข้อมูล โมเดลเสียงนำมาใช้รู้จำเสียงเพื่อสั่งการโปรแกรมวินโดว์มีเดียเพลย์เยอร์มีคำสั่งจำนวน 8 คำสั่ง การทดสอบพบเปอร์เซนต์ความผิดพลาดเท่ากับ 37.5 และยิ่งจำนวนเสียงที่ใช้รู้จำมีมากขึ้นจะทำให้เปอร์เซนต์ความความผิดพลาดเพิ่มขึ้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

Sound Analysis and Recognition system on Hidden Markov Model

Mr. Korkiat	Ngernyoung	50010063
Mr. Chinnaphat	Matchimaviwat	50010379
Mr. Patcharapon	Usuksomboon	50011077
Mr. Supakit	Nuttayasakul	Advisor

Academic Year 2010

ABSTRACT

Hidden Markov Model (HMM) is a statistical model. It can be use to recognize the patterns in recognition system such as handwriting, musical equipment recognition, bioinformatics recognition. HMM is also used in the weather forecast. Seen that the HMM can develop a wide range of uses.

This project studies the HMM in order to recognize a simple Thai syllable word. The groups of Thai syllable words are classified by the fast Fourier transform. The classified results are then recognized by HMM resulting in sound model which kept to the database. The sound models consist with 8 Thai syllable words, are used to recognize in order to control Windows Media Player. The experimental result shows that error recognition is proximate 37.5 percent. Furthermore, will get more error, if it takes more word to recognize.

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

กิตติกรรมประกาศ

ปริญญานิพนธ์นี้สำเสร็จได้ด้วยดี เพราะได้รับความช่วยเหลือ และร่วมมือจากหลาย ๆ ฝ่าย ด้วยกัน บุคคลแรกที่ต้องกล่าวถึงเพราะเป็นส่วนสำคัญที่ทำให้ปริญญานิพนธ์นี้สำเร็จลงได้ก็คือ อาจารย์ สุภกิจ นุตยะสกุล อาจารย์ที่ปรึกษาปริญญานิพนธ์ ที่มีความดูแลเอาใจใส่ แนะนำ และช่วยเหลือเสมอมา ซึ่งต้องขอขอบพระคุณอาจารย์เป็นอย่างมาก และขอขอบคุณเพื่อน ๆ ทุกคนที่ให้คำปรึกษาการทำงานไม่ว่าจะเป็นด้านทฤษฎี หรือปฏิบัติช่วยให้งานสำเร็จได้ด้วยดี

ขอขอบพระคุณบุคคลสำคัญที่สุดคือ บิดา มารดาที่ทำให้คณะผู้จัดทำมีทุกวันนี้ และเลี้ยงดูมาเป็นอย่างดี พร้อมทั้งให้โอกาสทางการศึกษาอย่างเต็มที่ ทั้งยังให้กำลังใจ เอาใจใส่ดูแลในทุก ๆ ด้าน ข้าพเจ้าขอระลึกในพระคุณอันสุดประมาณ และขอกราบขอบพระคุณมา ณ ที่นี้

ก่อเกียรติ เงินยวง
ชินพัฒน์ มัชฌิมาวิวัฒน์
พัชรพล อยู่สุขสมบูรณ์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ

หน้า

บทคัดย่อภาษาไทย.....I

บทคัดย่อภาษาอังกฤษ.....II

สารบัญ.....IV

สารบัญตาราง.....VI

สารบัญภาพ.....VII

บทที่ 1 บทนำ.....1

1.1 ความเป็นมาของปัญหา.....1

1.2 วัตถุประสงค์ของการศึกษา.....1

1.3 ประโยชน์ที่คาดว่าจะได้รับ.....2

1.4 ขอบเขตของโครงงาน.....2

1.5 ส่วนประกอบของปริญญานิพนธ์.....2

บทที่ 2 ฮิดเดนมาร์คอฟโมเดล.....3

2.1 บทนำ.....3

2.2 รูปแบบที่คาดเดาได้.....3

2.3 รูปแบบที่คาดเดาไม่ได้.....3

2.4 มาร์คอฟโมเดล.....4

2.5 ฮิดเดนมาร์คอฟโมเดล.....5

2.6 ส่วนประกอบของแบบจำลองมาร์คอฟ.....7

2.7 ปัญหาพื้นฐานในฮิดเดนมาร์คอฟโมเดล.....8

2.8 การใช้ประโยชน์ของ เอชเอ็มเอ็ม.....23

บทที่ 3 การรู้จำเสียง.....24

3.1 เสียง.....24

3.2 จิตสวนศาสตร์.....25

3.3 ความถี่ของเสียงพูดมนุษย์.....25

3.4 รายละเอียดไฟล์เวฟ (Wave format).....26

3.5 การออกแบบระบบรู้จำเสียง.....27

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญ (ต่อ)

หน้า

3.6 การหาคุณลักษณะเด่นของเสียง.....28
3.7 ยูเอ็มแอล (UML).....31
บทที่ 4 ผลการทดลองและสรุปผลการทดลอง.....38
4.1 การทดลองใช้งานโปรแกรมรู้จำเสียง.....38
4.2 วิเคราะห์และสรุปผลการทดลอง.....42
4.3 การทดสอบโปรแกรม.....43
บทที่ 5 วิจารณ์ผลการทดลองและแนวทางการพัฒนาต่อ.....45
5.1 การวิจารณ์ผลการทดลอง.....45
5.2 ผลที่ได้รับจากโครงงาน.....45
5.3 ปัญหาที่พบ.....45
5.4 แนวทางแก้ไขปัญหา.....45
5.5 แนวทางในการพัฒนาต่อ.....46
บรรณานุกรม.....47
ภาคผนวก การทำงานของโปรแกรม.....48

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญตาราง

ตาราง **หน้า**

3.1 คำสั่งเสียง 8 คำสั่งที่ใช้ในการสั่งการโปรแกรมวินโดว์มีเดียเพลย์เยอร์.....32
4.1 ผลการทดลองคำสั่งเสียงพยางค์เดียว 5 เสียง ทำการทดสอบเสียงละ 5 ครั้ง.....38
4.2 ผลการทดลองคำสั่งเสียงพยางค์เดียว 8 เสียงที่ใช้ในการสั่งงานโปรแกรม.....38
4.3 ผลการทดลองคำสั่งเสียงพยางค์เดียว 8 เสียงที่ใช้ในการสั่งงานโปรแกรม
ทดสอบเสียงละ 25 ครั้ง กลุ่มตัวอย่าง 5 คน(1 คนเก็บเสียง 5 ครั้งต่อ 1 คำสั่ง).....40
4.4 ผลการทดลองคำสั่งเสียงพยางค์เดียว 8 เสียงที่ใช้ในการสั่งงานโปรแกรม
ทดสอบเสียงละ 50 ครั้ง กลุ่มตัวอย่าง 10 คน(1 คนเก็บเสียง 5 ครั้งต่อ 1 คำสั่ง).....41

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป

รูป หน้า

2.1 ตัวอย่างโมเดลรูปแบบที่คาดเดาได้.....3

2.2 ความน่าจะเป็นในการเปลี่ยนสถานะของมาร์คอฟโมเดล.....4

2.3 ตัวอย่างมาร์คอฟโมเดลของสภาพอากาศ.....4

2.4 เมทริกซ์ใช้แทนความน่าจะเป็นในการเปลี่ยนสถานะ.....5

2.5 ตัวอย่างฮิดเดนมาร์คอฟโมเดลของสภาพอากาศเมื่อให้สภาพอากาศเป็น
ส่วนที่ถูกซ่อนไว้ และส่วนที่เราสังเกตเป็นความชื้นของสาหร่าย.....6

2.6 เมทริกซ์ในฮิดเดนมาร์คอฟโมเดล.....6

2.7 การคำนวณหา P (แห้ง เปียก ชื้น ชื้น | ฝน แดด เมฆ เมฆ).....11

2.8 ฟอร์วาร์ดโพสิเยอร์ที่สถานะเริ่มต้น.....12

2.9 ฟอร์วาร์ดโพสิเยอร์ในวันที่2.....13

2.10 ฟอร์วาร์ดโพสิเยอร์ในขั้นตอนสุดท้าย.....14

2.11 วิเธอร์บิ อัลกอริธึม.....16

2.12 วิเธอร์บิ อัลกอริธึมในตอนเริ่มต้นวันแรก.....17

2.13 วิเธอร์บิ อัลกอริธึมในวันที่2.....18

2.14 วิเธอร์บิ อัลกอริธึมในวันที่2 (2).....19

2.15 วิเธอร์บิ อัลกอริธึมในวันสุดท้าย.....20

2.16 วิเธอร์บิ อัลกอริธึมในขั้นตอนสุดท้าย.....21

2.17 การปรับค่าให้โมเดล.....22

3.1 รายละเอียดการเก็บข้อมูลเวฟไฟล์.....26

3.2 ตัวอย่างการออกแบบระบบรู้จำเสียงอย่างง่าย.....28

3.3 ตัวอย่างข้อมูลเสียงที่ต้องการนำมาวิเคราะห์หาคุณลักษณะเด่น.....29

3.4 ตัวอย่างการแบ่งข้อมูลเสียง เพื่อนำเข้าเอฟเอฟที.....30

3.5 ตัวอย่างข้อมูลที่ได้จากเอฟเอฟที.....30

3.6 ยูสเคสไดอะแกรม.....32

3.7 ซีเควนซ์ไดอะแกรมการสั่งงานคอมพิวเตอร์ด้วยเสียง.....33

3.8 ซีเควนซ์ไดอะแกรมการกำหนดการแสดงผล.....34

3.9 สเตทชาร์ทไดอะแกรมการรู้จำเสียงพูด.....35

3.10 สเตทชาร์ทไดอะแกรมการทำงานของฮิดเดนมาร์คอฟโมเดล.....36

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สารบัญรูป (ต่อ)

รูป	หน้า
3.11 สเตทชาร์ทไดอะแกรมการสั่งงานคอมพิวเตอร์ผ่านเสียง	37
4.1 หน้าต่างการทำงานโปรแกรม WMP SOUND TRINING	43
4.2 หน้าต่างการทำงานโปรแกรม WMP SOUND CONTROL	44
ก.1 หน้าต่างแสดงการทำงานของโปรแกรม WMP SOUND TRINING	48
ก.2 หน้าต่างแสดงการทำงานของโปรแกรม WMP SOUND TRINING	49
ก.3 หน้าต่างแสดงการทำงานของโปรแกรม WMP SOUND TRINING	49
ก.4 หน้าต่างแสดงการทำงานของโปรแกรม WMP SOUND TRINING	50
ก.5 หน้าต่างแสดงการทำงานของโปรแกรม WMP SOUND TRINING	50
ก.6 หน้าต่างแสดงการทำงานของโปรแกรม WMP SOUND TRINING	51
ก.7 หน้าต่างแสดงข้อมูลในดาต้าเบส	51
ก.8 หน้าต่างแสดงการทำงานของโปรแกรม WMP SOUND CONTROL	52

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 1
บทนำ

1.1 ความเป็นมาของปัญหา

เป็นเวลานานมากแล้วที่ได้มีการคิดค้นและพัฒนาคอมพิวเตอร์ให้สามารถประมวลผลเพื่อใช้ในงานต่าง ๆได้ จนในปัจจุบันนี้ขีดความสามารถของเครื่องคอมพิวเตอร์นั้นมากขึ้นมากทำให้มีการใช้คอมพิวเตอร์กันอย่างกว้างขวางทั่วโลก ซึ่งทำให้มีอิทธิพลต่อชีวิตประจำวันของมนุษย์เป็นอย่างยิ่ง และโดยทั่วไปแล้วในการรับคำสั่งจากมนุษย์นั้น คอมพิวเตอร์จะรับทางอุปกรณ์ที่ชื่อ คีย์บอร์ด และเมาส์ โดยในอดีตที่ผ่านมานั้นมนุษย์ได้มีการจินตนาการ ซึ่งอาจเห็นได้ในนิยายวิทยาศาสตร์ว่าการติดต่อของมนุษย์กับเครื่องคอมพิวเตอร์นั้น มนุษย์สามารถสั่งงานโดยเสียงพูดได้โดยตรง ซึ่งเป็นวิธีที่เป็นธรรมชาติมากกว่า ทำให้เกิดการพัฒนาแนวความคิดเหล่านี้ขึ้นเพื่อทำให้คอมพิวเตอร์นั้นสามารถทำความเข้าใจเสียงพูดของมนุษย์ได้

ในอดีตนั้นก็ได้มีการพัฒนาเทคนิควิธีการต่าง ๆ เพื่อที่จะให้มีการติดต่อสื่อสารดังที่ได้กล่าวมาแล้วข้างต้นนั้น จนทำให้เกิดสาขาทางคอมพิวเตอร์ขึ้นมาแขนงหนึ่งก็คือ “การรู้จำเสียงพูด” หรือ “Speech Recognition” แต่เพราะว่าในอดีตความสามารถของเครื่องคอมพิวเตอร์นั้นยังมีประสิทธิภาพที่ต่ำ และเทคนิควิธีการที่ใช้นั้นยังไม่ถูกพัฒนาเพื่อมาทำการจัดการกับเสียงพูดของมนุษย์ซึ่งมีความซับซ้อนมากนัก ทำให้การพัฒนานั้นเป็นไปได้อย่างล่าช้า จนในปัจจุบันนี้ได้มีการคิดค้นเทคนิคต่าง ๆ ขึ้นมากมาย ซึ่งหนึ่งในนั้นก็คือ ฮิดเดนมาร์คอฟโมเดล (Hidden Markov Model)ซึ่งเป็นการสร้างแบบจำลองขึ้นมาเพื่อนำมาใช้ในการเป็นตัวแทนข้อมูลเสียงพูด

ความรู้พื้นฐานของ มาร์คอฟเชน (Markov Chains) นั้นรู้จักกันมาเป็นเวลานานแต่ได้นำมาใช้ในเรื่องของการรู้จำเสียงโดยใช้เพื่อเป็นแบบจำลองที่เป็นตัวแทนเสียงเมื่อไม่นานมานี้เอง ปัญหาหลักอย่างหนึ่งของการสร้างแบบจำลองของเสียงบนพื้นฐานของมาร์คอฟเชน นั้นก็คือการขาดวิธีที่ดี เพื่อที่จะทำการหาค่าตัวแปรของมาร์คอฟโมเดล (Markov Model) ที่ดีที่เหมาะสมกับรูปแบบของข้อมูลเสียง ซึ่งได้มีการนำเสนอวิธีในการหาเมื่อปลายทศวรรษที่ 60 และได้ถูกนำมาใช้กับการประมวลผลสัญญาณ (Signal Processing) เมื่อไม่นานมานี้

1.2 วัตถุประสงค์ของการศึกษา

1) ศึกษาการทำงานและกระบวนการของฮิดเดนมาร์คอฟโมเดล
2) พัฒนาโปรแกรมรู้จำเสียงพูด ใช้สั่งการโปรแกรมวินโดว์มีเดียเพลย์เยอร์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1.3 ประโยชน์ที่คาดว่าจะได้รับ

1) เข้าใจการทำงานของฮิดเดนมาร์คอฟโมเดล
2) ผู้ที่สนใจในการพัฒนาซอฟต์แวร์(software) สามารถนำเอาเอกสารและคู่มือการใช้งานที่ได้จากโครงงานนี้ไปพัฒนา
3) พัฒนาเป็นซอฟต์แวร์ใช้งานบนโมบาย (Mobile) ได้จริง

1.4 ขอบเขตของโครงงาน

1) ศึกษาฮิดเดนมาร์คอฟโมเดลกับปัญหาที่ไม่ซับซ้อน
2) การรับเสียงในห้องเงียบ
3) จำกัดจำนวนเสียงในการรู้จำเสียง
4) โปรแกรมสามารถสั่งการวินโดว์มีเดียเพลย์เยอร์ จำนวน 8 คำสั่ง เล่น หยุด เปิด ปิด ไป กลับ เพิ่ม และลด
5) อุปกรณ์โมบายที่ใช้คือ โน๊ตบุ๊ค
6) การรับเสียงทำโดยไมค์ตัดเสียงรบกวน ยี่ห้อ shure sm-58

1.5 ส่วนประกอบของปริญญานิพนธ์

ในปริญญานิพนธ์ฉบับนี้ประกอบด้วย 5 บท ดังนี้

เนื้อหาในบทที่ 1 กล่าวถึงความเป็นมาของปัญหา วัตถุประสงค์ ประโยชน์ที่คาดว่าจะได้รับ ขอบเขตของโครงงาน และส่วนประกอบต่าง ๆ ของรายงาน

เนื้อหาในบทที่ 2 กล่าวถึงทฤษฎีและหลักการที่เกี่ยวข้องกับฮิดเดนมาร์คอฟโมเดล

เนื้อหาในบทที่ 3 กล่าวถึงการหาลักษณะเด่นของเสียง

เนื้อหาในบทที่ 4 เป็นผลการทดลอง โดยโปรแกรมที่ได้เขียนขึ้นมา

เนื้อหาในบทที่ 5 กล่าวถึงบทสรุปทั้งหมดของโครงงาน การวิจารณ์ผลการทดลอง ผลที่ได้รับจากโครงงาน ปัญหาที่พบ แนวทางในการพัฒนาต่อ และข้อเสนอแนะของโครงงาน

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

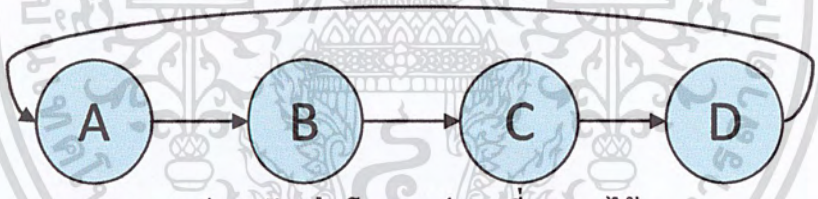
บทที่ 2
ฮิดเดนมาร์คอฟโมเดล

2.1 บทนำ

แบบจำลองมาร์คอฟ (Markov Model) เป็นแบบจำลองทางสถิติซึ่งพัฒนามาเพื่อแบ่งกลุ่มของอนุกรมทางเวลา หรือสัญญาณที่ไม่คงที่ นั่นคือใช้สำหรับจัดกลุ่มของสัญญาณที่ไม่รู้จัก (Unknown signal) ให้ไปอยู่ในกลุ่มใดกลุ่มหนึ่งของสัญญาณ ซึ่งแบบจำลองมาร์คอฟได้ถูกนำมาประยุกต์ใช้ในการรู้จำเสียง รู้จำลายมือ รู้จำเครื่องดนตรี รวมถึงการพยากรณ์อนาคตที่อาจเกิดขึ้น เช่น พยากรณ์อากาศและพยากรณ์ตลาดหุ้น เป็นต้น

2.2 รูปแบบที่คาดเดาได้

รูปแบบที่คาดเดาได้เป็นรูปแบบที่เราสามารถบอกได้ถึงการเปลี่ยนสถานะไปยังสถานะถัดไปได้อย่างแน่นอน เช่น โมเดลในรูป 2.1 แสดงให้เรารู้ว่าเมื่อ สถานะอยู่ที่ A ต่อไปจะต้องเปลี่ยนไปยังสถานะที่อยู่ที่ B อย่างแน่นอน จากนั้นจะเป็น C D และกลับมาที่ A อีกครั้ง



รูป 2.1 ตัวอย่างโมเดลรูปแบบที่คาดเดาได้

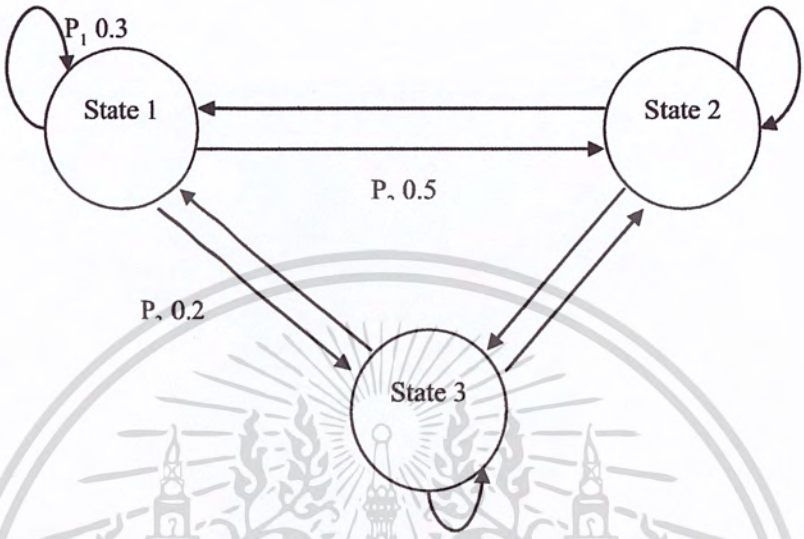
2.3 รูปแบบที่คาดเดาไม่ได้

รูปแบบที่อยู่ในชีวิตจริงส่วนใหญ่ไม่ใช่รูปแบบที่เกิดขึ้นอย่างแน่นอนตายตัว แต่เรายังคงต้องการโมเดลที่จะสามารถให้รูปแบบการเปลี่ยนสถานะแก่เราได้ ซึ่งเราสามารถทำได้โดยเขียนให้สถานะหนึ่ง ขึ้นกับสถานะก่อนหน้านั้น เช่น สภาพอากาศ เราพอจะสามารถคาดการณ์ถึงสภาพอากาศ ณ เวลาปัจจุบันได้ ถ้าหากว่าเรารู้ว่าวันก่อนหน้านั้นมีสภาพอากาศเป็นเช่นใด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

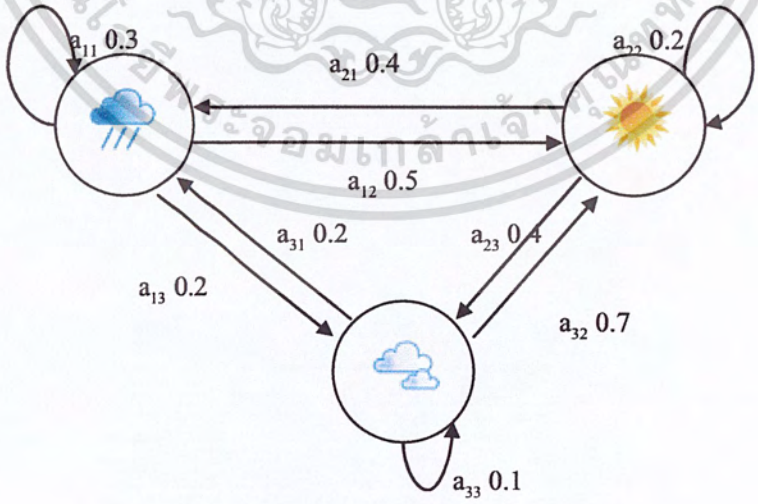
2.4 มาร์คอฟโมเดล

มาร์คอฟโมเดลเป็นโมเดลที่ไม่สามารถรู้การเปลี่ยนสถานะที่แน่นอนได้มีค่าในการเปลี่ยนสถานะเป็นความน่าจะเป็น



รูป 2.2 ความน่าจะเป็นในการเปลี่ยนสถานะของมาร์คอฟโมเดล

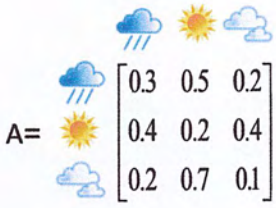
การเปลี่ยนสถานะจากสถานะหนึ่งนั้นทุกเส้นที่ออกจากสถานะนั้นจะต้องมีค่าความน่าจะเป็นรวมกันเป็น1 นั่นคือจากรูป 2.2 $P_1+P_2+P_3=1$ เสมอ และเราสามารถเขียนมาร์คอฟโมเดลของสภาพอากาศได้ดังรูป 2.3 โดยมี 3 สถานะคือ ฝนแดด เมฆ



รูป 2.3 ตัวอย่างมาร์คอฟโมเดลของสภาพอากาศ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

จากรูป 2.3 ความน่าจะเป็นในการเปลี่ยนสถานะของโมเดลที่มี 3 สถานะ สามารถเขียนให้อยู่ในรูปเมทริกซ์ ขนาด 3X3 ได้ ดังรูป 2.4



รูป 2.4 เมทริกซ์ใช้แทนความน่าจะเป็นในการเปลี่ยนสถานะ

จะเห็นว่าค่าของเลขในแต่ละแถวต้องรวมกันได้ 1 ทุกแถว การหาค่าความน่าจะเป็นของลำดับการเปลี่ยนสถานะของสิ่งที่เราสังเกต(Observation) O เมื่อ O = แดด, ฝน, เมฆ

โดยให้ π_{rain} π_{sun} π_{cloud} แทนความน่าจะเป็นในการเริ่มต้นของสถานะฝน เมฆ แดด ตามลำดับ

ค่าเริ่มต้นของสถานะฝนตก $\pi_{rain} = 0.3$ ความน่าจะเป็นที่ จะเกิดเหตุการณ์ O

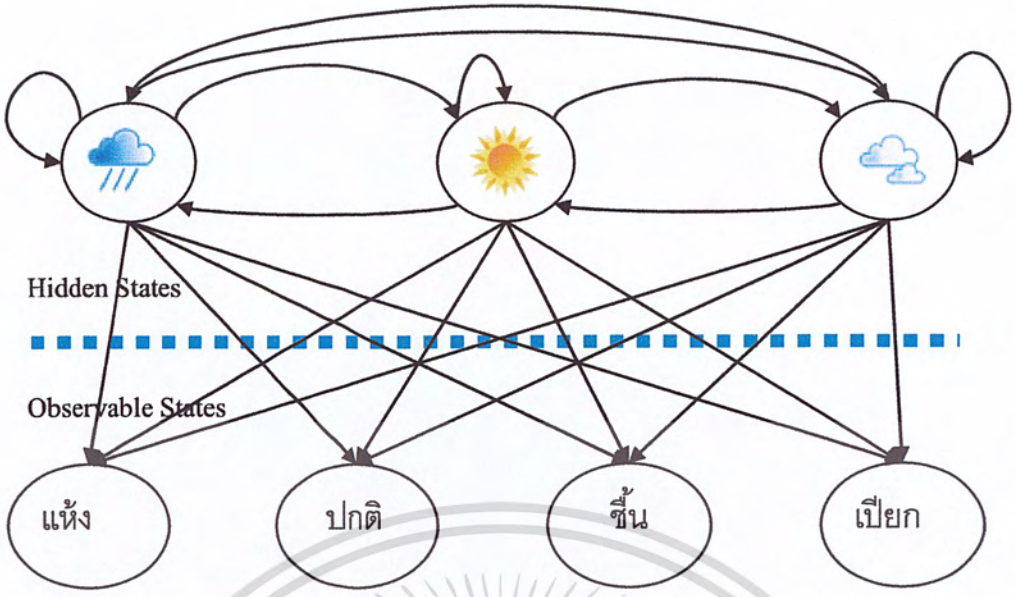
$$P(O) = 0.3 \times 0.4 \times 0.2$$

$$= 0.024$$

2.5 ฮิดเดนมาร์คอฟโมเดล

เพื่อเพิ่มขีดจำกัดในการใช้งานของมาร์คอฟโมเดลจึงเกิดฮิดเดนมาร์คอฟโมเดลขึ้นมา โดยการซ่อนกระบวนการที่เกิดขึ้นไว้ เช่น การทอยเหรียญทีละ1เหรียญ ถ้าเป็นการทอยแบบปกติก็จะเป็นมาร์คอฟโมเดลแต่ถ้าหากมีเหรียญหลายเหรียญซึ่งเป็นเหรียญที่มีโอกาสในการทอยได้หัวหรือก้อยได้ไม่เท่ากันแล้ว ก็จะได้ส่วนที่เราสังเกตเห็นได้เป็นประเภทเหรียญต่าง ๆ แต่ผลที่ได้จากการโยนเหรียญว่าเป็นหัวหรือก้อยจะถูกซ่อนไว้ หรือ เราไม่สามารถเห็นได้ ลักษณะนี้เป็นรูปแบบของ ฮิดเดนมาร์คอฟโมเดล สภาพอากาศ กรณีสภาพอากาศ ปูเสฉวนไม่สามารถดูสภาพอากาศโดยตรงได้ แต่ปูเสฉวนได้ข้อมูลจากแหล่งอื่นแทนเป็นการสังเกตจากความเปียกชื้นจากสาหร่าย จะได้โมเดลสภาพอากาศ ดังรูป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูป 2.5 ตัวอย่างฮิดเดนมาร์คอฟโมเดลของสภาพอากาศเมื่อให้สภาพอากาศเป็นส่วนที่ถูกซ่อนไว้ และส่วนที่เราสังเกตเป็นความชื้นของสาหร่าย

โมเดลของสภาพอากาศความน่าจะเป็นในการเปลี่ยนสถานะจะถูกเก็บลงในเมทริกซ์ A (Transition Matrix) รูป 2.6 ก และ ความน่าจะเป็นในการที่สภาพอากาศจะส่งผลกับสภาพของสาหร่าย จะถูกเก็บลงในเมทริกซ์ B (Confusion Matrix) รูป 2.6 ข

ก) ข)

รูป 2.6 เมทริกซ์ในฮิดเดนมาร์คอฟโมเดล

ก) เมทริกซ์ A

ข) เมทริกซ์ B

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.6 ส่วนประกอบของแบบจำลองมาร์คอฟ

พารามิเตอร์สำคัญที่เกี่ยวข้องในการสร้างแบบจำลองอ้างอิง ที่ต้องรู้จักได้แก่

1) T คือ ความยาวของลำดับข้อมูลที่ได้จากการควอนไตซ์ค่าความถี่มูลฐาน ซึ่งมีขนาดความยาวของลำดับเท่ากับจำนวนเฟรมทั้งหมดในเสียงแต่ละเสียง ซึ่งจะใช้เป็นข้อมูลอินพุทในส่วนของเอชเอ็มเอ็ม (HMM) โดยต่อไปจะเรียกแทนว่า "ลำดับของค่าปรากฏ" (Observation Sequence)

2) N คือ จำนวนสถานะในแบบจำลอง ถ้ากำหนดให้เซตของสถานะเป็น {1, 2, …, N} จะสามารถ แทนสถานะที่เปลี่ยนไปตามเวลา t ด้วย เซตของ $Q = \{q_1, q_2, \ldots, q_N\}$

3) M คือ จำนวนของค่าปรากฏที่สามารถเป็นไปได้ต่อหนึ่งสถานะ แทนสัญลักษณ์ด้วย $V = \{v_1, v_2, \ldots, v_M\}$ ซึ่งจากการจัดระดับของการเปลี่ยนแปลงของความถี่ (ΔF) ออกเป็น 3 ระดับจะได้เซตของค่าปรากฏที่สามารถเป็นไปได้ในแต่ละสถานะมีค่าเป็น V = {-1,0,1}

4) ค่าความน่าจะเป็นในการย้ายสถานะ : $A = \{a_{ij}\}$

โดย a_{ij} แทนการย้ายสถานะจาก i ไป j เมื่อ

$$a_{ij} = P[q_t = j \mid q_{t-1} = i] \; ; 1 \le i, j \le N \qquad (2.1)$$

5) การกระจายความน่าจะเป็นของค่าปรากฏที่สามารถเป็นไปได้ภายในสถานะ : $B = \{b_j(k)\}$

โดยที่

$$b_j(k) = P[v_k \text{ ที่เวลา } t] \; q_j \text{ ที่เวลา } t \; ; 1 \le k \le M \qquad (2.2)$$

เป็นนิยามการกระจายในสถานะ j เมื่อ j = 1,2,…,N

6) ค่าความน่าจะเป็นของการเป็นสถานะเริ่มต้น : $\pi = \{\pi_i\}$ เมื่อ

$$\pi_t = P[q_i \text{ ที่เวลา } t = 1] \; ; 1 \le i \le N \qquad (2.3)$$

จะเห็นเอชเอ็มเอ็มต้องการพารามิเตอร์ของแบบจำลองคือ N, M และ กลุ่มของความน่าจะเป็น A, B, π ดังนั้นในการแสดงเซตของพารามิเตอร์ที่สมบูรณ์ของแบบจำลองอ้างอิง จะแทนด้วยสัญลักษณ์

$$\lambda = (A, B, \pi) \qquad (2.4)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.7 ปัญหาพื้นฐานในฮิดเดนมาร์คอฟโมเดล

1) หาค่าความน่าจะเป็นของลำดับข้อมูล(Observation Sequence)

2) หาเส้นทางของลำดับสถานะ(State Sequence) ที่มีโอกาสเป็นไปได้มากที่สุด

3) การปรับค่าต่าง ๆ ของโมเดล

2.7.1 การแก้ปัญหาที่ 1 ความน่าจะเป็นของลำดับข้อมูลที่เราสังเกต

เราจะทำการคำนวณหาค่าความน่าจะเป็นของลำดับข้อมูล $O = O_1O_2...O_T$ จากโมเดล $Q = q_1q_2...q_T$ โดยเราจะกำหนดให้ลำดับของสถานะคือ λ เราจะเขียนแทนค่านี้ด้วย $P(O|\lambda)$ ให้ q_1 เป็นสถานะตั้งต้น และค่าความน่าจะเป็นของลำดับข้อมูล O โดยที่ลำดับสถานะเป็นดังสมการ คือ

$$P(O|Q,\lambda) = \prod_{t=1}^{T} P(O_t|q_t,\lambda) \tag{2.5}$$

เรากำหนดให้ลำดับข้อมูลเป็นอิสระต่อกันเราจะได้

$$P(O|Q,\lambda) = b_{q1}(O_1) \bullet b_{q2}(O_2) \bullet ... \bullet b_{qT}(O_T) \tag{2.6}$$

ค่าความน่าจะเป็นของชุดลำดับ Q สามารถเขียนได้ดังนี้

$$P(Q|\lambda) = \pi_{q1}a_{q1q2}a_{q2q3}...a_{qT\text{-}1qT} \tag{2.7}$$

เพราะฉะนั้นค่า $P(O|\lambda)$ จะสามารถเขียนได้ดังนี้

$$P(O|\lambda) = \sum_{allQ} P(O|Q,\lambda)P(Q|\lambda) \tag{2.8}$$

$$= \sum_{q1,q2,..qT} \pi_{q1}b_{q1}(O_1)a_{q1q2}b_{q2}(O_2)...a_{qT-1qT}b_{qT}(O_T) \tag{2.9}$$

เนื่องจากว่าถ้าเราใช้คอมพิวเตอร์คำนวณหาค่า $P(O|\lambda)$ โดยวิธีนี้คอมพิวเตอร์จะต้องใช้จำนวนคำสั่งถึง 2T*NT ครั้ง ถ้าสมมติกำหนดให้โมเดลมีสถานะ 5 สถานะ (N = 5) และมีชุดลำดับข้อมูล 100 ข้อมูล (T = 100) เพราะฉะนั้นคอมพิวเตอร์จะต้องให้คำสั่งในการคำนวณ เท่ากับ2 * 100 * 5100 * 1072 ครั้งเลยทีเดียว เราจึงจำเป็นต้องใช้วิธีอื่น ๆ มาช่วยในการลดจำนวนคำสั่งในการคำนวณหาค่า $P(O|\lambda)$ ซึ่งในรายงานนี้ได้เสนอวิธีการลดจำนวนคำสั่งในการคำนวณโดยใช้วิธีฟอร์วาร์ดโพสิเยอร์ (Forward Procedure)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ฟอร์วาร์ดโพสิเยอร์ [2], [3] จะมีตัวแปรเพิ่มขึ้นมาเพื่อจัดรูประบบสมการใหม่ให้เข้าใจได้ง่ายขึ้นและลดรูปสมการด้วย ซึ่งตัวแปรนี้แทนด้วย $\lambda_t(i) = P(O_1O_2...O_t), q_t = S_i | \alpha$ สำหรับวิธีฟอร์วาร์ดโพสิเยอร์นั้นจะประกอบด้วย 3 ส่วนด้วยกัน มีดังนี้

1) Initialization :

$$\alpha_t(i) = \pi_i b_i(O_1), \qquad 1 \le i \le N \tag{2.10}$$

2) Induction :

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^{N} \alpha_t(j) a_{ij}\right] b_j(O_{t+1}), \qquad 1 \le t \le T-1$$
$$1 \le j \le N \tag{2.11}$$

3) Termination :

$$P(O \mid \lambda) = \sum_{i=1}^{N} \alpha_T(i) \tag{2.12}$$

ในข้อ 1) เป็นส่วนของค่าเริ่มต้นของข้อมูลตัวแรก O_1 ของทุก ๆ สถานะในโมเดล และในข้อที่ 2) จะเป็นหัวใจสำคัญของฟอร์วาร์ดโพสิเยอร์ซึ่งเป็นวิธีการคำนวณหาค่า $\alpha_{t(i)}$ ของข้อมูลถัดจาก O_1 ทั้งหมดไปจนถึงตัวสุดท้าย เมื่อการคำนวณดำเนินไปจน $t = T-1$ ก็เป็นอันเสร็จสิ้นของการคำนวณโดยวิธีฟอร์วาร์ดโพสิเยอร์ โดยในข้อที่ 3) จะเป็นสมการที่เสร็จสมบูรณ์ แล้วของค่า $P(O|\lambda)$ ซึ่งเท่ากับผลรวมของค่า $\alpha_{t(i)}$ ของข้อมูลตัวสุดท้ายของทุก ๆ สถานะในโมเดล เราสามารถแสดงลักษณะของการคำนวณโดยวิธี ฟอร์วาร์ดโพสิเยอร์

ถ้าเราใช้วิธีฟอร์วาร์ดโพสิเยอร์ในการคำนวณหาค่า $P(O|\lambda)$ โดยใช้คอมพิวเตอร์จะมีจำนวนคำสั่งในการคำนวณทั้งหมดเป็น N2T เท่านั้น ซึ่งถ้าสมมติให้โมเดลมีสถานะทั้งหมด 5 สถานะ (N = 5) และมีชุดลำดับข้อมูล 100 ข้อมูล (T = 100) เพราะฉะนั้นคอมพิวเตอร์จะต้องให้คำสั่งในการคำนวณ ประมาณ 3000 คำสั่งเท่านั้นเอง ซึ่งเมื่อเทียบกับวิธีตรงแล้วจะแตกต่างกันมาก

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.7.2 ตัวอย่างการแก้ปัญหาที่ 1

สำหรับฮิดเดนมาร์คอฟโมเดลตัวข้อมูลที่เราสังเกตจะไม่ได้อยู่ที่เมทริกซ์ A เหมือนกับมาร์คอฟโมเดลแต่จะอยู่ที่เมทริกซ์ B ซึ่งเมทริกซ์ B ไม่มีการกล่าวถึงการเปลี่ยนสถานะไปมาในการเปลี่ยนแปลงของสิ่งที่เราสังเกตเลย ดังนั้นเราจึงต้องใช้ความสัมพันธ์ของสิ่งที่เราสังเกตกับส่วนที่ซ่อนไว้มาใช้ในการหา โอกาสในการเกิดการลำดับการเปลี่ยนสถานะ

เมื่อกำหนดให้

เมทริกซ์ AและB เป็นดังรูปที่ 2.6

ค่าเริ่มต้นของสถานะฝนตก = 1.0

ค่าเริ่มต้นของสถานะฟ้าโปร่ง = 0

ค่าเริ่มต้นของสถานะเมฆมาก = 0

ความน่าจะเป็นของลำดับข้อมูลของสิ่งที่เราสังเกต O คือ

$$P(O|\lambda)=\sum_{AllQ} P(O|Q,\lambda)P(Q|\lambda) \qquad (2.13)$$

<u>วิธีที่ 1</u> การหาความน่าจะเป็นของการเปลี่ยนสถานะของสิ่งที่สังเกตด้วยวิธีตรง

กำหนดลำดับการเปลี่ยนสถานะของส่วนที่ถูกซ่อนไว้ ลำดับการเปลี่ยนแปลงของส่วนที่เราสังเกต เป็นดังนี้

ให้ Q แทนลำดับการเปลี่ยนของสถานะของสภาพอากาศใน4วัน

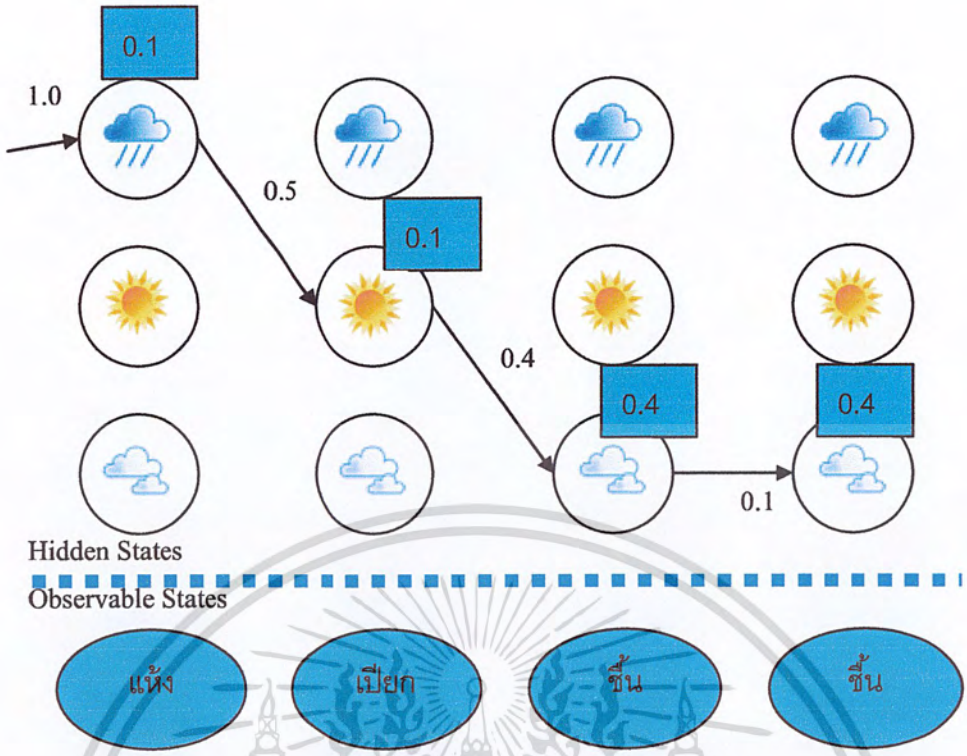
ให้ O แทนลำดับการเปลี่ยนของข้อมูลที่เราสังเกตได้ของสาหร่ายใน4วัน

โดยให้ O = แห้ง เปียก ชื้น ชื้น

จากสมการ 2.13 จะเขียนได้ว่า

$P(O)$= P(แห้ง เปียก ชื้น ชื้น | ฝน,ฝน,ฝน,ฝน) + P(แห้ง เปียก ชื้น ชื้น | ฝน,ฝน,ฝน,แดด) + P(แห้ง เปียก ชื้น ชื้น | ฝน,ฝน,ฝน,เมฆ) + P(แห้ง เปียก ชื้น ชื้น | ฝน,ฝน,แดด,ฝน) .....+ P(แห้ง เปียก ชื้น ชื้น | เมฆ,เมฆ,เมฆ,เมฆ)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูป 2.7 การคำนวณหา *P* (แห้ง เปียก ชื้น ชื้น | ฝน แดด เมฆ เมฆ)

การคำนวณหา *P* (แห้ง เปียก ชื้น ชื้น | ฝน แดด เมฆ เมฆ) เป็นการนำโมเดลมาเขียนในรูปของเวลาที่เปลี่ยนไปโดยที่แต่ละคอลัมน์จะแทนเวลาที่ต่างกัน โดยทางซ้ายเป็นวันแรกและทางขวาสุดเป็นวันสุดท้ายตามลำดับโดยกำหนดลำดับการเปลี่ยนสถานะเพียงรูปแบบเดียวคือ ฝน แดด เมฆ เมฆ

$$P(\text{แห้ง เปียก ชื้น ชื้น} \mid \text{ฝน แดด เมฆ เมฆ})$$

$$= 1.0 \times 0.1 \times 0.5 \times 0.1 \times 0.4 \times 0.4 \times 0.1 \times 0.4$$

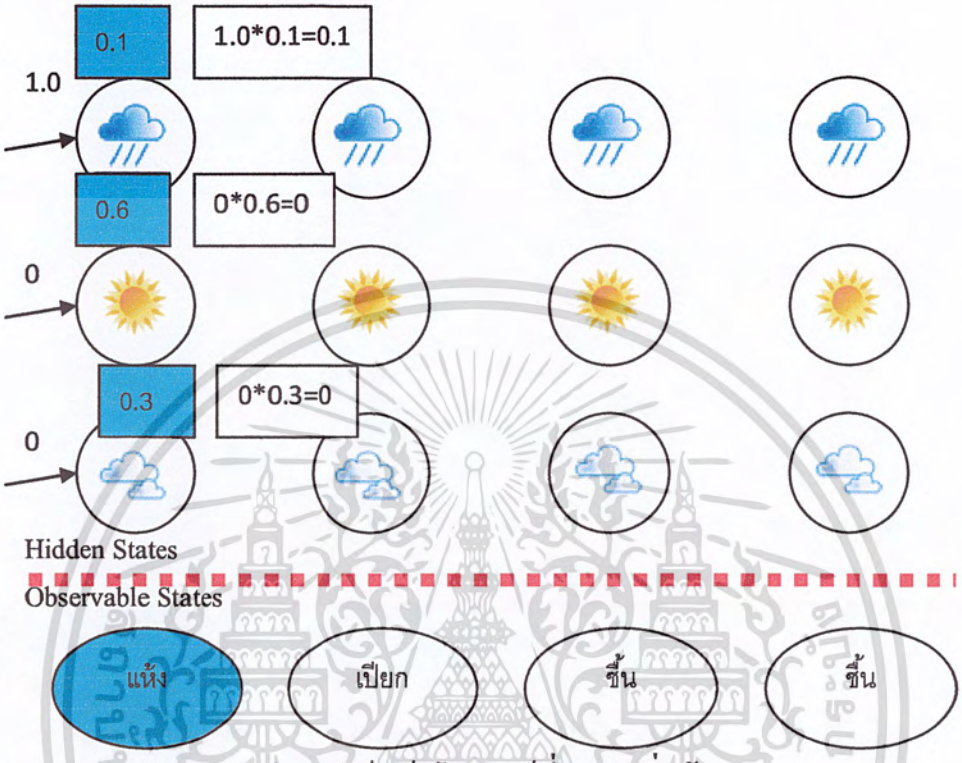
$$= 0.000032$$

จากสมการ 2.13 จะต้องเปลี่ยนลำดับสิ่งที่สังเกต Q จนครบทุกรูปแบบแล้วนำมารวมกันจึงจะได้ค่าความน่าจะเป็นของลำดับสิ่งที่สังเกต O โดยที่ในที่นี้มีลำดับการเปลี่ยนสถานะอยู่ 4 ครั้ง 3สถานะ นั้นคือต้องใช้ลำดับการเปลี่ยนสถานะของสภาพอากาศ ทั้งหมด $3^4 = 81$ รูปแบบ วิธีการนี้เมื่อมีความยาวของลำดับข้อมูลมาก ๆ จะยิ่งใช้เวลานานมาก ในการคำนวณ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

<u>วิธีที่ 2</u> ฟอร์วาร์ดโพสิเยอร์

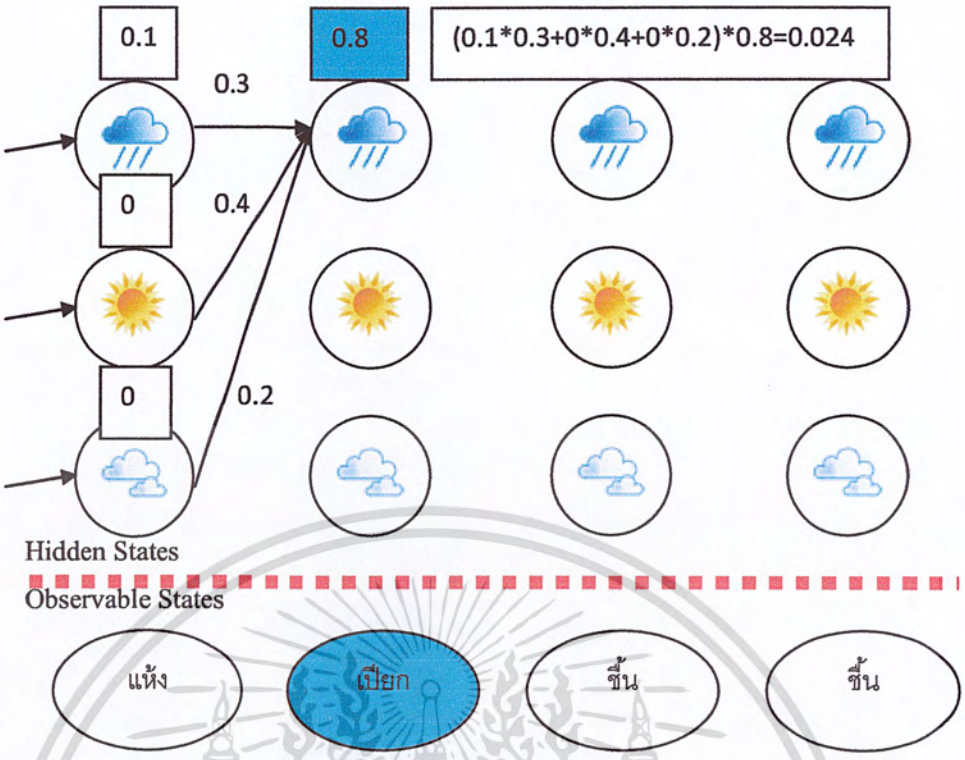
เป็นการหาโอกาสรวมให้แต่ละสถานะโดยที่จะต้องรวมโอกาสทั้งหมดที่สถานะในวันก่อนหน้านั้นจะเปลี่ยนมาเป็นสถานะปัจจุบันทั้งหมด



รูป 2.8 ฟอร์วาร์ดโพสิเยอร์ที่สถานะเริ่มต้น

เมื่อเริ่มต้นโอกาสที่จะเกิดเป็นสถานะนั้น ๆ จะมีค่าความน่าจะเป็นในการเริ่มต้นของสถานะนั้นเพียงอย่างเดียว ดังนั้นจึงนำค่าความน่าจะเป็นในการเริ่มต้นมาคูณกับความน่าจะเป็นของสิ่งที่เราสังเกตเห็นนั่นคือสาหร่ายเปียกในวันแรก จะได้ค่าของความน่าจะเป็นของสถานะในวันแรก

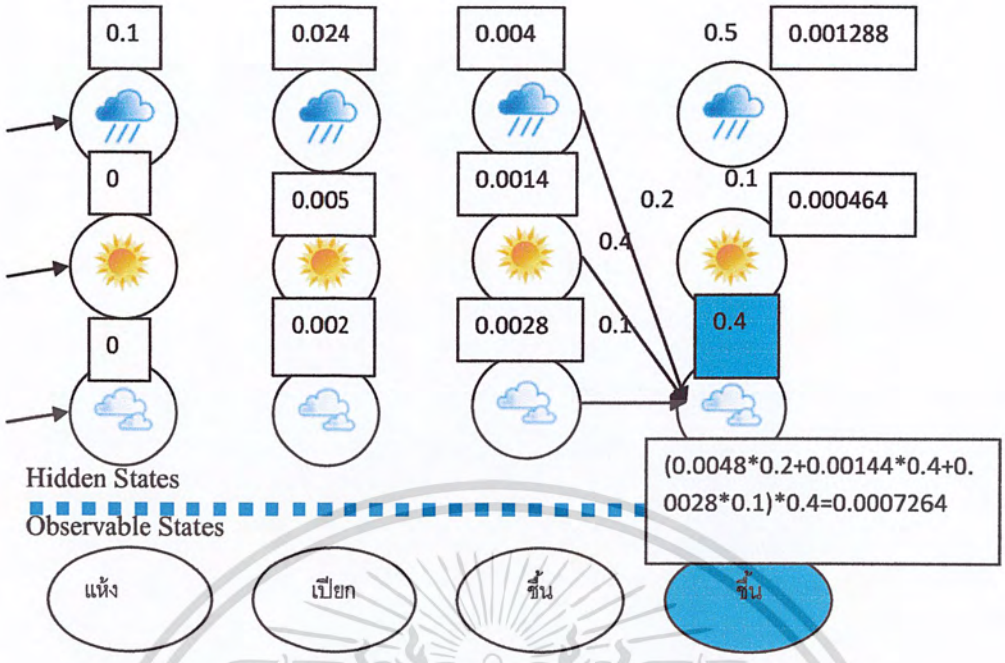
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูป 2.9 ฟอร์วาร์ดโพสิเยอร์ในวันที่2

ในวันที่ 2 นั้นจะต้องนำความน่าจะเป็นของสถานะในวันที่ 1 มาใช้เพื่อหาโอกาสที่แต่ละสถานะในวันที่ 1 ที่จะเปลี่ยนไปอยู่ที่สถานะฝนตกในวันที่ 2 จากนั้นนำค่าที่ได้มาคูณกับค่าความน่าจะเป็นจากสิ่งที่สังเกตได้คือสาหร่ายเปียกคือ 0.8 จะได้ค่าของความน่าจะเป็นรวมในการเกิดฝนในวันที่ 2 จากนั้นทำวิธีเดิมกับทุกสถานะจนครบทุกสถานะ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูป 2.10 ฟอร์วาร์ดโพสิเยอร์ในขั้นตอนสุดท้าย

เมื่อทำฟอร์วาร์ดโพสิเยอร์จนถึงสถานะสุดท้ายแล้วให้นำค่าของสถานะสุดท้ายมารวมกัน จะได้ค่าความน่าจะเป็นของ O โดยที่ O = แห้ง เปียก ชื้น ชื้น จากรูป 2.10 $P(O)$ จะมีค่าเป็น

$$P(O) = 0.001288+0.0004648+0.0007264$$
$$= 0.0024792$$

ซึ่งผลที่ได้จะให้ค่าเท่ากันกับวิธีที่ 1 แต่จะมีความเร็วมากกว่ามากเมื่อข้อมูลมีความยาวมากขึ้น

2.7.3 การแก้ปัญหาที่ 2 หาลำดับการเปลี่ยนสถานะที่เป็นไปได้มากที่สุด ด้วย วิเธอร์บิ อัลกอริธึม (Viterbi Algorithm)

สำหรับปัญหาในข้อที่ 2 นี้ เป็นการเลือกเส้นทางของการเปลี่ยนแปลงสถานะที่ให้ค่าความเป็นไปได้มากที่สุด ในรายงานนี้ได้เสนอวิธีของ วิเธอร์บิ อัลกอริธึมซึ่งเราจะกำหนดให้ชุดลำดับของการเปลี่ยนแปลงสถานะที่ดีที่สุดเป็น $Q = \{q_1 q_2 \dots q_T\}$ และให้ชุดลำดับของข้อมูลเป็น $O = \{O_1 O_2 \dots O_T\}$ และเราต้องการหาค่า $\delta_t(i) = \max_{q1,q2,\dots qt-1} P[q_1 q_2 \dots q_t = i, O_1 O_2 \dots O_t | \lambda]$

ซึ่งค่า $\delta_t(i)$ เป็นค่าที่ดีที่สุดของหนึ่งเส้นทางในลำดับที่ t และ $\varphi_t(j)$ คือเส้นทางที่ให้ค่าที่ดีที่สุด สำหรับวิธีของวิเธอร์บิ อัลกอริธึมจะมีส่วนประกอบที่สำคัญอยู่ 4 ส่วนคือ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

1) Initialization :

$$\delta_t(i) = \pi_i b_i(O_1), \qquad 1 \le i \le N \tag{2.14}$$

$$\varphi_1(i) = 0 \tag{2.15}$$

2) Recursion :

$$\delta_t(j) = \max_{1 \le i \le N} \left[\delta_{t-1}(i) a_{ij}\right] b_j(O_t), \qquad 2 \le t \le T \tag{2.16}$$

$$1 \le j \le N$$

$$\varphi_t(j) = \arg\max_{1 \le i \le N} \left[\delta_{t-1}(i) a_{ij}\right], \qquad 2 \le t \le T \tag{2.17}$$

$$1 \le j \le N$$

3) Termination :

$$p^* = \max_{1 \le i \le N} \left[\delta_T(i)\right] \tag{2.18}$$

$$q_T^* = \arg\max_{1 \le i \le N} \left[\delta_T(i)\right] \tag{2.19}$$

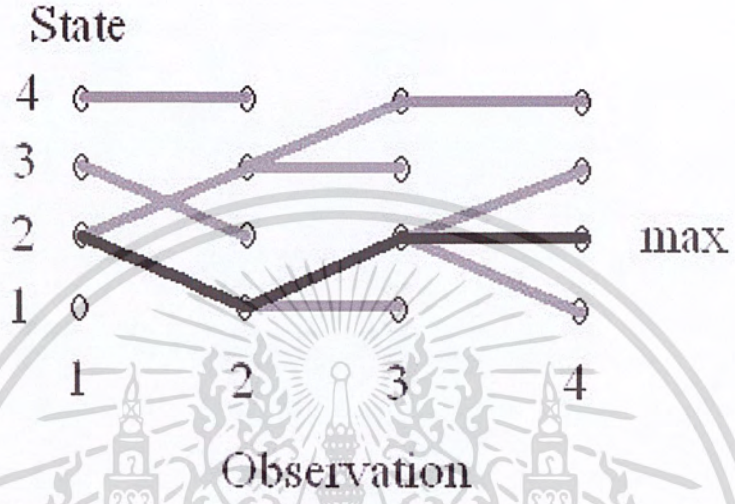
4) Path (state sequence) backtracking :

$$q_t^* = \varphi_{t+1}(q^*_{t+1}), \qquad t = T-1, T-2, \ldots, 1 \tag{2.20}$$

ในส่วนที่ 1 จะเป็นค่าเริ่มต้นของข้อมูลตัวแรกของทุก ๆ สถานะในโมเดล ส่วนเส้นทางเดินนั้นยังไม่มีจึงเป็นศูนย์ สำหรับส่วนที่ 2 ก็เป็นการคำนวณหาค่า $\delta_{T(i)}$ และ $\varphi_{t(i)}$ และในลำดับต่าง ๆ ถัดจากข้อมูลตัวแรกไปจนถึงข้อมูลตัวสุดท้าย ในส่วนที่ 3 ก็จะเป็นส่วนที่สุดท้ายของข้อมูล ค่าที่เราจะเลือกเส้นทางของการเปลี่ยนแปลงสถานะจะเป็นค่า $\delta_{T(i)}$ ตัวสุดท้ายที่มีค่ามากที่สุด และจะเลือกเส้นทางเริ่มต้นจากสถานะนั้นด้วย จากนั้นในส่วนที่ 4 ก็จะเป็นการเลือกเส้นทางเดินทั้งหมด ของชุดลำดับข้อมูลซึ่งจะทำการเลือกย้อนกลับไปและจะมีเพียงเส้นทางเดียวที่ให้เลือกอยู่แล้ว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

สมมติให้โมเดลมีสถานะ 4 สถานะ และจำนวนข้อมูลมี 4 ตัว ดังรูปข้างล่าง จะเห็นว่าเส้นทางที่มีสีเข้มเป็นเส้นทางที่ให้ค่าความเป็นไปได้มากที่สุด เราจะเลือกเส้นทางนี้เป็นเส้นทางการเปลี่ยนแปลงสถานะของโมเดลนี้



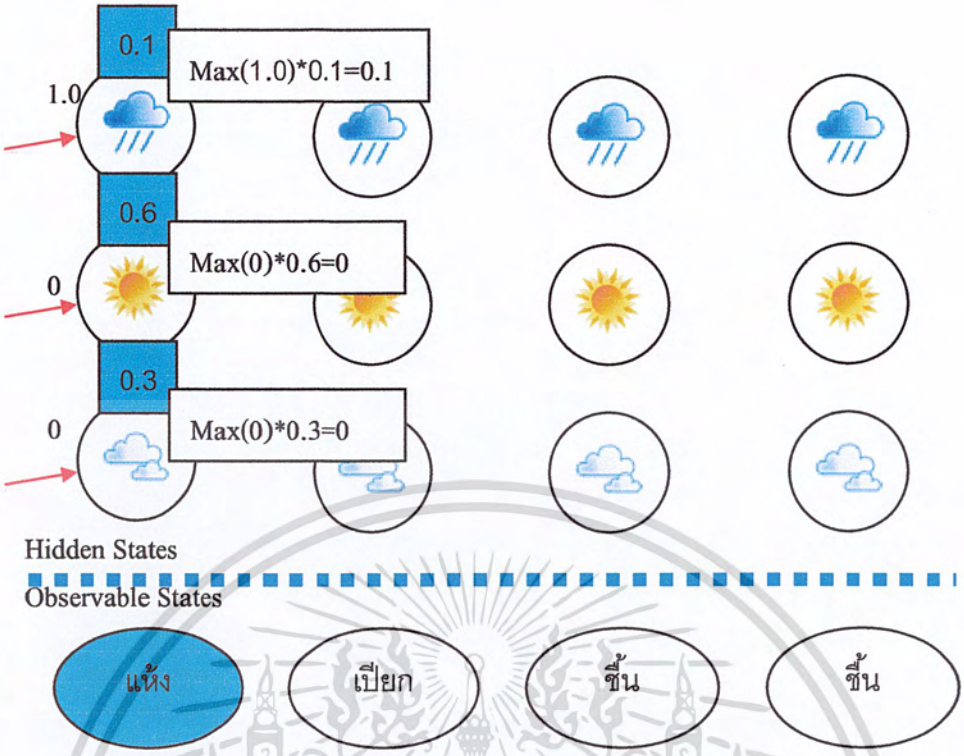
รูป 2.11 วิเธอร์บิ อัลกอริธึม

2.7.4 ตัวอย่างการแก้ปัญหาที่ 2

สิ่งที่เราต้องหาคือลำดับการเปลี่ยนสถานะของส่วนที่ถูกซ่อนไว้ที่เป็นไปได้มากที่สุดเมื่อเรามีข้อมูลลำดับการเปลี่ยนแปลงของส่วนที่เราสังเกต มา 1 ลำดับ เราจะต้องเลือกการเปลี่ยนสถานะระหว่างสถานะในส่วนที่ถูกซ่อนไว้ที่ให้ค่าความเป็นได้มากที่สุด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

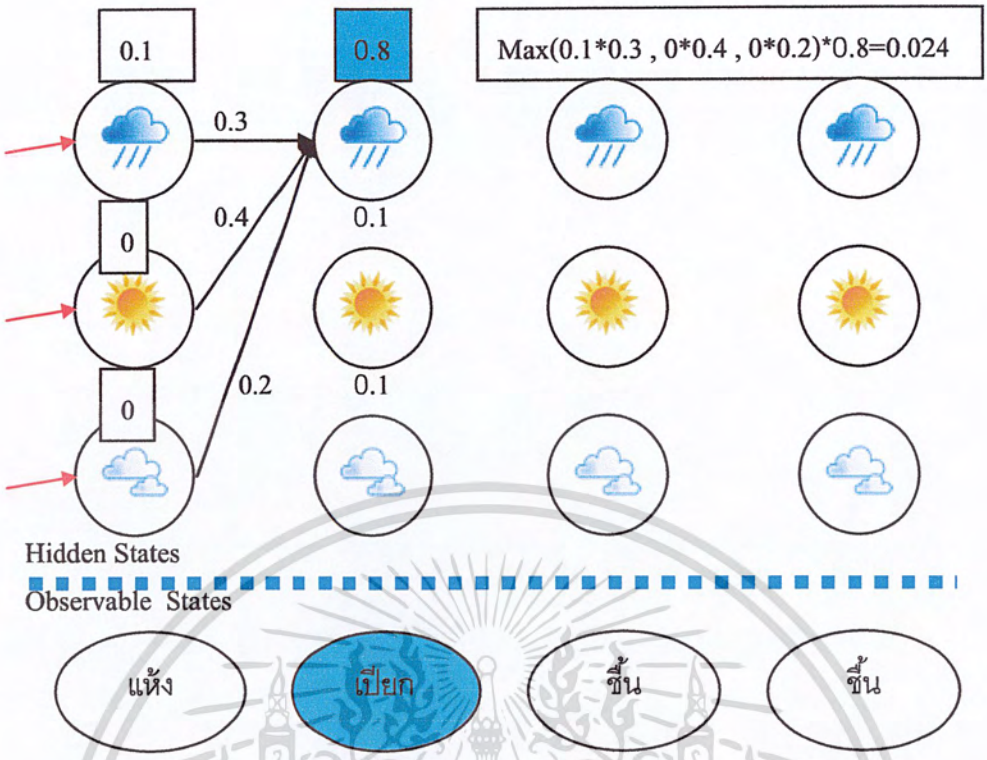
สำนักหอสมุดกลาง พระจอมเกล้าลาดกระบัง



รูป 2.12 วิเธอร์บี อัลกอริธึมในตอนเริ่มต้นวันแรก

จากรูป 2.11 นั้นแต่ละคอลัมน์จะแทนเวลาคนละวันกัน ที่สถานะเริ่มต้นจะมีเพียงค่าเดียว ดังนั้นการหาค่าสูงสุดจึงเท่ากับค่าโอกาสในการเริ่มต้น π ของแต่ละสถานะ โดยที่ช่องสีฟ้าจะเป็นค่าของสถานะในส่วนที่ซ่อนมีผลต่อส่วนที่เราสังเกต ซึ่งจะต้องนำค่าสูงสุดมาคูณ

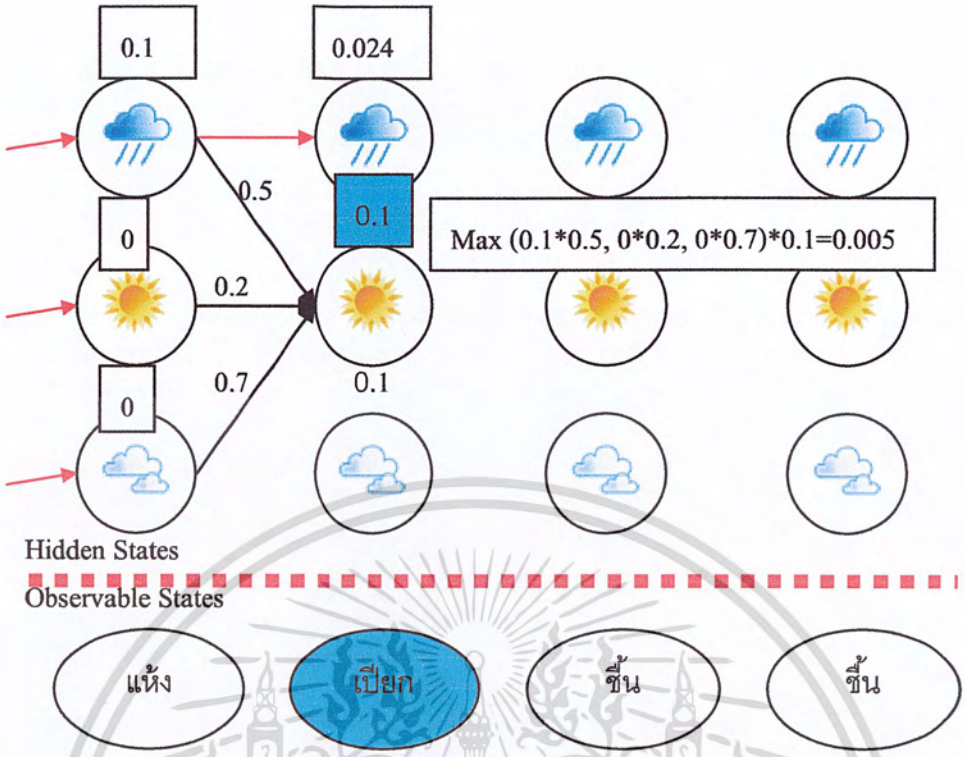
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้
117542



รูป 2.13 วิเธอร์บิ อัลกอริธึมในวันที่2

ในวันที่ 2 ให้เลือกการเปลี่ยนสถานะจากวันที่ 1 ที่ทำให้สถานะฝนในวันที่ 2 นั้นมีค่าความน่าจะเป็นมากที่สุด จะได้ว่าการเปลี่ยนสถานะจากฝนในวันที่ 1 เมื่อนำมาคูณกับการเปลี่ยนจากฝนในวันที่ 1 มาเป็นฝนในวันที่ 2 จะให้ค่าความเป็นมากที่สุด ดังนั้นสำหรับการเปลี่ยนสถานะจากวันที่ 1มาเป็นฝนวันที่ 2 จึงเลือกการเปลี่ยนสถานะจากฝนในวันที่ 1

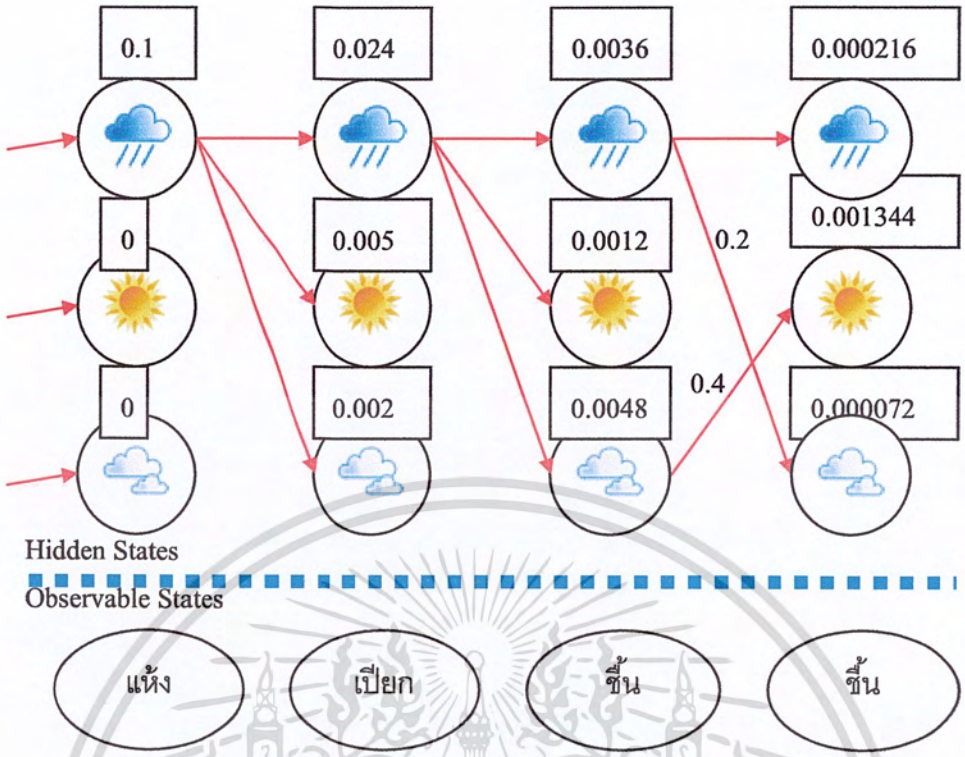
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูป 2.14 วิเธอร์บิ อัลกอริธึมในวันที่2 (2)

ให้เลือกการเปลี่ยนสถานะที่ทำให้สถานะแดดในวันที่ 2 นั้นมีค่าความน่าจะเป็นมากที่สุด ซึ่งจะได้ว่าเมื่อนำค่าการเปลี่ยนสถานะรวมของฝนในวันที่ 1 มาเป็นแดดในวันที่ 2 มีค่ามากที่สุด จึงเลือกการเปลี่ยนสถานะจากฝนในวันที่ 1 นำมาคูณกับการส่งผลของสาหร่ายเปียกในวันที่ 2 จึงจะได้ความน่าจะเป็นที่มากที่สุดในการเกิดแดดในวันที่ 2 ทำเช่นนี้ไปทีละวันจนถึงวันสุดท้าย

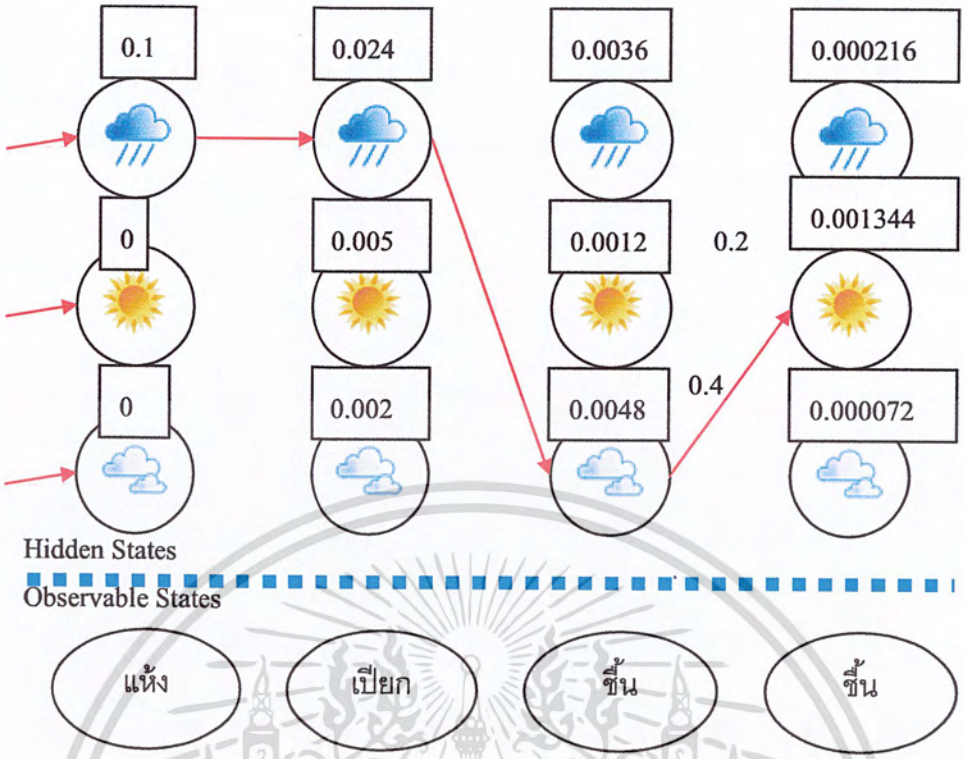
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูป 2.15 วิเธอร์บี อัลกอริธึมในวันสุดท้าย

เมื่อหาจนครบแล้วให้เลือกสถานะในวันสุดท้ายที่มีความน่าจะเป็นมากที่สุดแล้วย้อนกลับไปจะได้เส้นทางการเปลี่ยนสถานะของลำดับการเปลี่ยนสถานะ Q ที่มีโอกาสในการเกิดมากที่สุด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



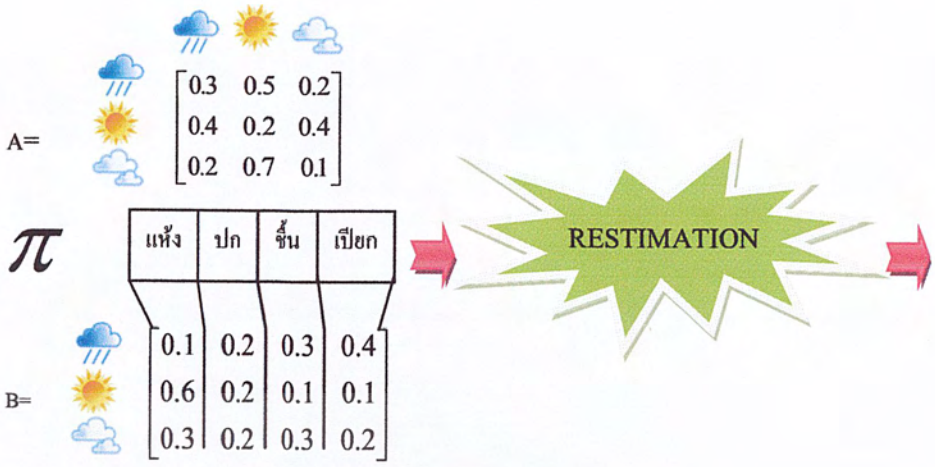
รูป 2.16 วิเธอร์บิ อัลกอริธึมในขั้นตอนสุดท้าย

จะได้การเปลี่ยนลำดับของสถานะที่ถูกซ่อน ที่มีความเป็นไปได้มากที่สุดคือ ฝน ฝน เมฆ แดด สำหรับข้อมูลลำดับการเปลี่ยนสภาพของสาหร่าย แห้ง เปียก ชื้น ชื้น

2.7.5 ปัญหาที่3 การปรับค่าให้โมเดล

การฝึกเพื่อปรับค่าให้กับเมทริกซ์ A, B และค่า π เราจะทำได้ก็ต่อเมื่อมีข้อมูลมากพอ ยิ่งข้อมูลมากค่าต่าง ๆ ของโมเดลจะยิ่งลู่เข้าหาค่าที่ดีที่สุด และเมื่อนำข้อมูลมาทดสอบจะให้ความน่าจะเป็นออกมาสูงสำหรับข้อมูลที่มีความแม่นยำมากขึ้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูป 2.17 การปรับค่าให้โมเดล

$$\bar{a} = \frac{\text{จำนวนครั้งที่มีการเปลี่ยนแปลงสถานะจาก } S_i \text{ ไป } S_j}{\text{จำนวนครั้งที่มีการเปลี่ยนสถานะจาก Si ทั้งหมด}} \quad (2.21)$$

$$\bar{b}_j(K) = \frac{\text{จำนวนครั้งที่ข้อมูลอยู่สถานะ j และข้อมูลเป็น } V_k}{\text{จำนวนครั้งที่อยู่ในสถานะ j ทั้งหมด}} \quad (2.22)$$

$$\bar{\pi} = \frac{\text{จำนวนครั้งที่อยู่ในสถานะ } S_i \text{ ณ ข้อมูลตัวแรกของแต่ละชุด}}{\text{จำนวนชุดข้อมูลทั้งหมด}} \quad (2.23)$$

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2.8 การใช้ประโยชน์ของ เอชเอ็มเอ็ม

สามารถนำไปแก้ปัญหา 3 อย่าง โดยสองปัญหาแรกเป็นการตรวจหารูปแบบ (pattern recognition)

2.8.1 การประเมิน (Evaluation)

โดยใช้อัลกอริธึมฟอร์เวิร์ด (forward algorithm) คำนวณหาความน่าจะเป็น ของลำดับที่สังเกตได้ จากแบบจำลองเอชเอ็มเอ็มจำนวนมาก แล้วเลือกแบบจำลองเอชเอ็มเอ็มที่น่าจะเป็นไปได้มากที่สุด ปัญหาลักษณะนี้พบได้ในการตรวจจับคำพูด (speed recognition) หรือจากแบบจำลองใช้อัลกอริธึมฟอร์เวิร์ด-แบคเวิร์ด (forward-backward algorithm) ในการคำนวณความน่าจะเป็นของลำดับผลลัพธ์ และความน่าจะเป็นของสภาพที่ซ่อนอยู่ที่ให้ลำดับผลลัพธ์นั้น

2.8.2 การอ่านรหัส (Decoding)

เป็นการหาลำดับของเหตุการณ์ (สภาพ) ที่ซ่อนอยู่ ที่ทำให้เกิดลำดับของเหตุการณ์ที่สังเกตเห็น โดยใช้อัลกอริธึมวิเทอร์บี ตัวอย่างเช่น การจัดเรียงเปรียบเทียบลำดับเบสของสายดีเอ็นเอ

2.8.3 การเรียนรู้ (Learning)

เป็นปัญหาที่ยากที่สุดที่เกี่ยวกับเอชเอ็มเอ็มคือการนำลำดับที่สังเกตได้ และรู้ลำดับของสภาพซ่อนอยู่ (hidden states) และหาแบบจำลองเอชเอ็มเอ็มที่เข้าพอดีกับข้อมูลเหล่านั้น นั่นคือหาค่า (π, A, B) ที่น่าจะอธิบายสิ่งที่สังเกตเห็นดีที่สุด การค้นหาพารามิเตอร์ของแบบจำลองเอชเอ็มเอ็มจะใช้ Baum-Welch algorithm

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 3
การรู้จำเสียง

3.1 เสียง

เสียง เป็นคลื่นกลที่เกิดจากการสั่นสะเทือนของวัตถุ เมื่อวัตถุสั่นสะเทือน ก็จะทำให้เกิดการอัดตัวและขยายตัวของคลื่นเสียง และถูกส่งผ่านตัวกลาง เช่น อากาศ ไปยังหู แต่เสียงสามารถเดินทางผ่านสสารในสถานะก๊าซ ของเหลว และของแข็งก็ได้ แต่ไม่สามารถเดินทางผ่านสูญญากาศได้ เมื่อการสั่นสะเทือนนั้นมาถึงหู มันจะถูกแปลงเป็นพัลส์ประสาท ซึ่งจะถูกส่งไปยังสมอง ทำให้เรารับรู้และจำแนกเสียงต่าง ๆ ได้

3.1.1 คุณลักษณะของเสียง

คุณลักษณะเฉพาะของเสียง ได้แก่ ความถี่ ความยาวช่วงคลื่น แอมปลิจูด และความเร็ว เสียงแต่ละเสียงมีความแตกต่างกัน เสียงสูง-เสียงต่ำ, เสียงดัง-เสียงเบา, หรือคุณภาพของเสียง ลักษณะต่าง ๆ ทั้งนี้ขึ้นอยู่กับแหล่งกำเนิดเสียง และจำนวนรอบต่อวินาทีของการสั่นสะเทือน

3.1.1.1 ความถี่

ระดับเสียง (pitch) หมายถึง เสียงสูงเสียงต่ำ สิ่งที่ทำให้เสียงแต่ละเสียงสูงต่ำแตกต่างกันนั้นขึ้นอยู่กับความเร็วในการสั่นสะเทือนของวัตถุ วัตถุที่สั่นเร็วเสียงจะสูงกว่าวัตถุที่สั่นช้า โดยจะมีหน่วยวัดความถี่ของ การสั่นสะเทือนต่อวินาที เช่น 60 รอบต่อวินาที, 2,000 รอบต่อวินาที เป็นต้น และนอกจาก วัตถุที่มีความถี่ในการสั่นสะเทือนมากกว่าจะมีเสียงที่สูงกว่าแล้ว หากความถี่มากขึ้นเท่าตัว ก็จะมีระดับเสียงสูงขึ้นเท่ากับ 1 ออกเตฟ (octave) ภาษาไทยเรียกว่า 1 ช่วงคู่แปด

3.1.1.2 ความยาวช่วงคลื่น

ความยาวช่วงคลื่น (wavelength) หมายถึง ระยะทางระหว่างยอดคลื่นสองยอดที่ติดกันซึ่งเกิดขึ้น ระหว่างการอัดตัวของคลื่นเสียง (คล้ายคลึงกับยอดคลื่นในทะเล) ยิ่งความยาวช่วงคลื่นมีมาก ความถี่ของเสียง (ระดับเสียง) ยิ่งต่ำลง

3.1.1.3 แอมปลิจูด

แอมปลิจูด (amplitude) หมายถึง ความสูงระหว่างยอดคลื่นและท้องคลื่นของคลื่นเสียง ที่แสดงถึงความเข้มของเสียง (Intensity) หรือความดังของเสียง (Loudness) ยิ่งแอมปลิจูดมีค่ามาก ความเข้มหรือความดังของเสียงก็ยิ่งเพิ่มขึ้น

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.2 จิตสวนศาสตร์

จิตสวนศาสตร์ (จิด-ตะ-สะ-วะ-นะ-สาด; ภาษาอังกฤษ: Psychoacoustics) เป็นการศึกษาเกี่ยวกับประสาทสัมผัสของมนุษย์ด้านการได้ยินพื้นฐาน การประยุกต์ใช้งานด้านการประมวลสัญญาณเสียงจำเป็นต้องทราบถึงลักษณะการได้ยินของมนุษย์ เสียงเป็นคลื่นที่เกิดจากการเปลี่ยนแปลงความดันของอากาศ เมื่อกระทบกับอวัยวะภายในหูจะส่งสัญญาณประสาทไปยังสมองเพื่อประมวลเสียงที่ได้มา

3.2.1 ขีดจำกัดของประสาทสัมผัส

อวัยวะของมนุษย์ที่ใช้รับฟังเสียง คือ หู สามารถรับรู้ได้ในช่วงตั้งแต่ความถี่ 20 เฮิรตซ์ ถึง 22000 เฮิรตซ์ เมื่ออายุเพิ่มขึ้นช่วงการได้ยินนี้จะลดลง โดยเฉพาะช่วงความถี่สูง ๆ ความถี่ที่ต่ำมาก ๆ นั้นมนุษย์ไม่สามารถได้ยินแต่สามารถรู้สึกได้ผ่านทางผิวหนัง

การแยกแยะความถี่เสียง 2 ความถี่ของมนุษย์มีลักษณะไม่เชิงเส้นตลอดย่านความถี่ที่มนุษย์สามารถรับรู้ได้ โดยในย่านความถี่เสียงต่ำ ๆ นั้นสามารถแยกแยะความแตกต่างเมื่อความถี่ทั้งคู่แตกต่างกันเพียงไม่กี่เฮิรตซ์ ในขณะที่ในย่านความถี่เสียงสูง ๆ สามารถแยกแยะความแตกต่างของเสียงทั้งสองเมื่อความแตกต่างของความถี่มีค่า เช่น 1000 เฮิรตซ์และในย่านความถี่ 4000 เฮิรตซ์ เป็นต้น

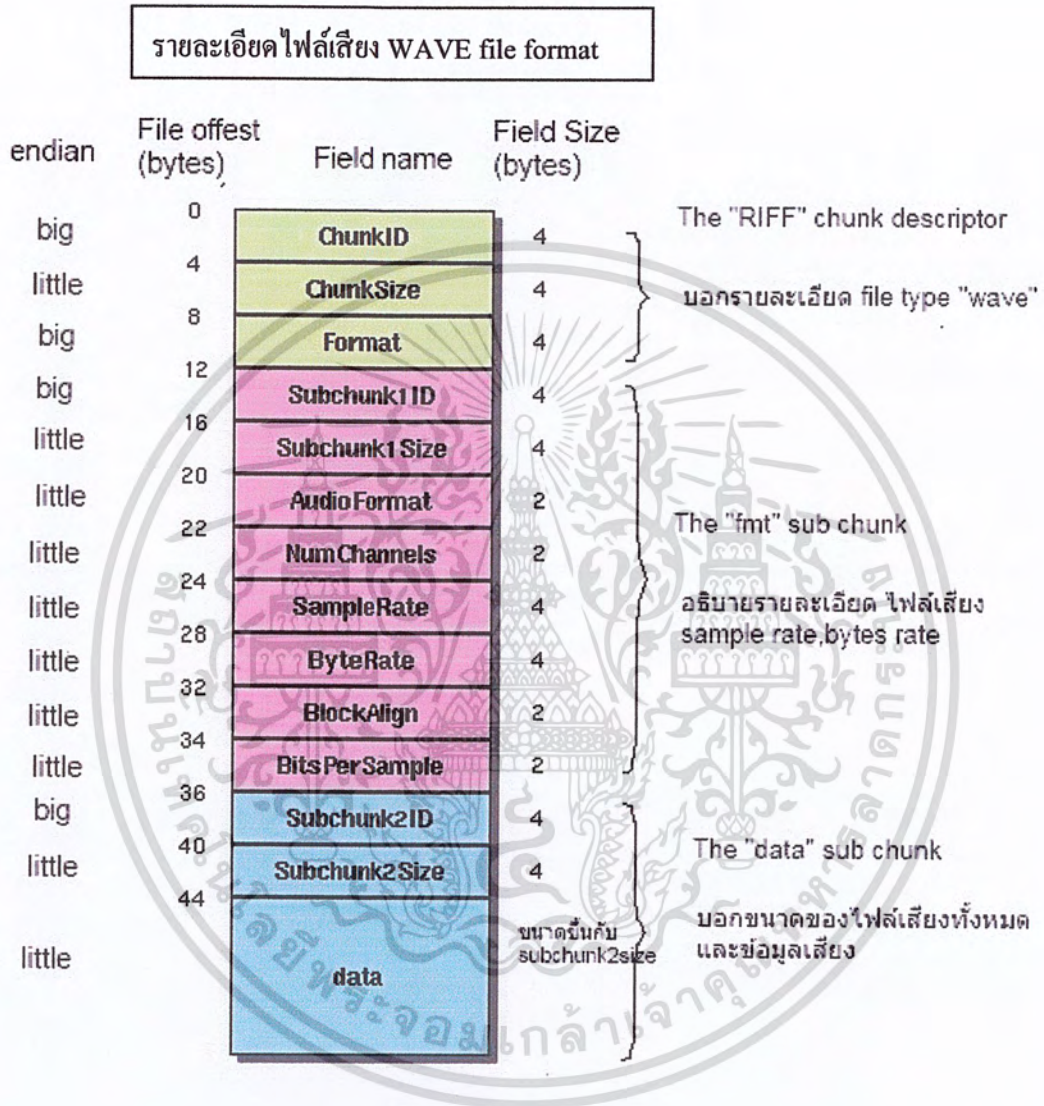
3.3 ความถี่ของเสียงพูดมนุษย์

วีเอฟ (voice frequency) หรือวอยซ์แบน (voice band) เป็นความถี่ช่วงหนึ่ง ในช่วงความถี่เสียงที่มนุษย์ได้ยิน ซึ่งใช้เพื่อในการติดต่อสื่อสารระหว่างบุคคลในการติดต่อสื่อสารทางโทรศัพท์ เสียงที่ใช้จะอยู่ในความถี่ช่วง 300-3400 เฮิรตซ์ ซึ่งเป็นเหตุให้ความถี่ต่ำของคลื่นไฟฟ้า(ความถี่=300-3000 เฮิรตซ์) สามารถแทนความถี่ของเสียงพูดมนุษย์ได้ และค่าแบนด์วิดท์ที่จองไว้สำหรับการสื่อสารทางเดียว อยู่ที่ 4000 เฮิรตซ์ (รวมการ์ดแบนด์ (guard bands) ด้วย) แซมปิ้งเรท= 8000 เฮิรตซ์ ถูกใช้เป็นพื้นฐานในการมอดูเลชั่น (modulation) สำหรับ PSTN(Public switched telephone network)

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.4 รายละเอียดไฟล์เวฟ (Wave format)

เวฟไฟล์ (wave) เป็นรูปแบบหนึ่งของการเก็บข้อมูลเสียงลงสื่อดิจิตอลโดยมีรูปแบบการจัดเก็บดังรูป



รูป 3.1 รายละเอียดการเก็บข้อมูลเวฟไฟล์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.4.1 รูปแบบการเก็บข้อมูลของไฟล์เวฟ

ChunkID	เก็บอักษร "RIFF" ใน แอสกี (ASCII form) (0x52494646 big-endian form).
ChunkSize	36 + SubChunk2Size
Format	เก็บอักษร "WAVE" (0x57415645 big-endian form).
Subchunk1ID	เก็บอักษร "fmt " (0x666d7420 big-endian form).
Subchunk1Size	16 for PCM. บอกขนาด Subchunk1Size
AudioFormat	PCM = 1 (i.e. Linear quantization) ค่าอื่น ๆ นอกจาก 1 แสดงการเข้ารหัสของข้อมูล
NumChannels	Mono = 1, Stereo = 2, etc.
SampleRate	8000, 44100, etc.
ByteRate	== แซมปิ้งเรท * จำนวนช่องสัญญาณ (NumChannels) * จำนวนบิทต่อแซมปิ้ง (BitsPerSample)/8
BlockAlign	== จำนวนช่องสัญญาณ * บิทต่อแซมปิ้ง/8
BitsPerSample	8 บิท = 8, 16 บิท = 16, etc.
Subchunk2ID	เก็บอักษร "data" (0x64617461 big-endian form).
Subchunk2Size	== จำนวนแซมปิ้ง (NumSamples) *จำนวนช่องสัญญาณ * บิทต่อแซมปิ้ง/8
Data	ข้อมูลไฟล์เสียง (.wav)

3.5 การออกแบบระบบรู้จำเสียง

มนุษย์เราสามารถพูดคุยติดต่อสื่อสารกันได้ เสียงที่ใช้มีมากมายหลากหลายเสียง เช่น เสียงหัวเราะ ร้องไห้ ดีใจ และอื่น ๆ มนุษย์เรามีวิธีการแยกความแตกต่าง โดยอาศัยสมองของมนุษย์แปรความหมายเสียงที่ได้ยินเป็นคำพูดหรือภาษา ทำให้เกิดการติดต่อสื่อสารกันได้ แนวคิดนี้เป็นต้นแบบของระบบรู้จำเสียง

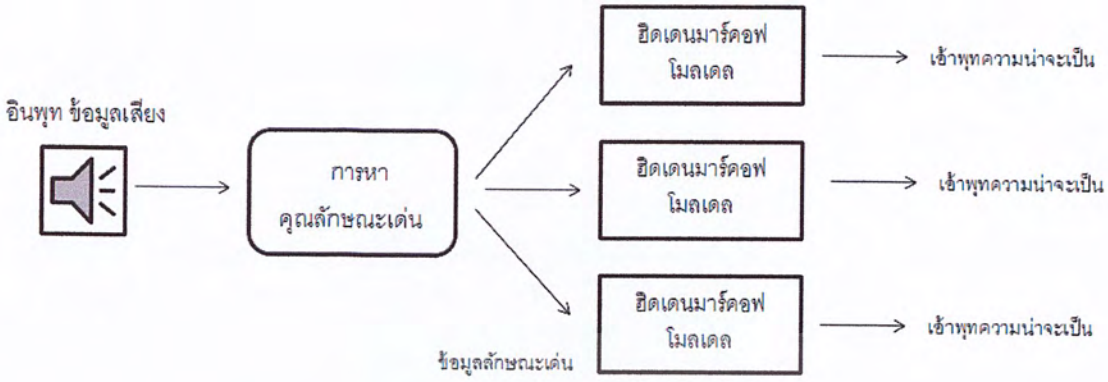
3.5.1 แนวคิดในการออกแบบระบบรู้จำเสียง

1) รับเสียงเข้าระบบ

2) วิเคราะห์หาคุณลักษณะเด่นของเสียง

3) นำคุณลักษณะเด่นของเสียงมาฝึกหรือ หาค่าความน่าจะเป็นจากโมเดลเสียงต่าง ๆ ที่ได้เรียนรู้ไว้แล้ว

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูป 3.2 ตัวอย่างการออกแบบระบบรู้จำเสียงอย่างง่าย

3.6 การหาคุณลักษณะเด่นของเสียง

ขั้นตอนการหาลักษณะเด่นของเสียงในแต่ละจุดออกมาใช้เป็นพารามิเตอร์แก่ฮิดเดนมาร์คอฟโมเดลในการรู้จำเสียง เพื่อนำมาเปรียบเทียบกับตัวอย่างเสียงอื่น ๆ ในเวลาใช้งานจริงในปัจจุบันมีหลายวิธีที่จะหาลักษณะเด่นของเสียง ยกตัวอย่างเช่น

1) ฟาสฟูเรียร์ทรานสฟอร์ม (Fast Fourier Transform)
2) วิธีแอลพีซี (LPC Feature Analysis)
3) เอ็มเอฟซีซี (Mel frequency Cepstral Coefficients (MFCCs))

ในรายงานฉบับนี้จะอธิบายและใช้งานเพียงแค่ฟาสฟูเรียร์ทรานสฟอร์มเพราะเป็นทฤษฎีที่ง่ายต่อการปฏิบัติจริง รวมทั้งมีไลบลารีสำเร็จรูปสำหรับให้ใช้งานได้โดยง่าย ส่วนวิธีอื่น ๆ และรายละเอียดเชิงลึกของฟาสฟูเรียร์ทรานสฟอร์ม นั้นอยู่นอกขอบเขตของการศึกษาของรายงานนี้ และจะไม่นำมาแสดงไว้ในรายงาน

3.6.1 ดีสครีตฟูเรียร์ทรานสฟอร์ม (Discrete Fourier Transform :DFT)

ดีเอฟที (Discrete Fourier Transform) เป็นฟูเรียร์ทรานสฟอร์ม (Fourier Transform) ชนิดหนึ่ง ใช้ในการแปลงฟังก์ชั่นหนึ่งไปสู่อีกฟังก์ชันหนึ่งไปเป็น โดเมนความถี่ (Frequency Domain) ของฟังก์ชั่นเดิม (ซึ่งฟังก์ชั่นเดิมมักเป็นฟังก์ชั่นทางเวลา (Time Domain)) โดยที่ดีเอฟที มีข้อกำหนดว่า ข้อมูลเข้า (Input) จะต้องเป็นข้อมูลไม่ติดต่อกัน (Discrete Input) และมีค่าที่แน่นอน (Finite) ส่วนสมการและรายละเอียดของดีเอฟที นั้นอยู่นอกขอบเขตของการศึกษาของรายงานนี้

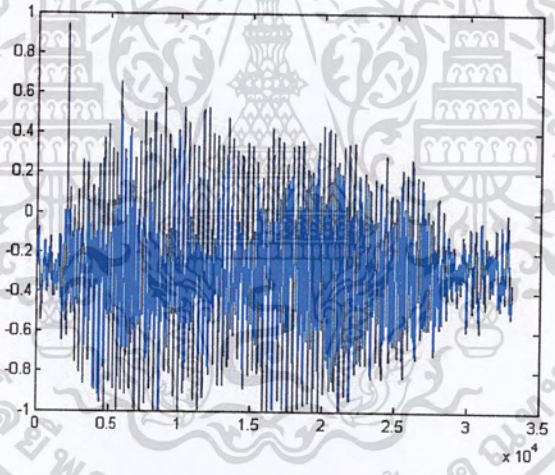
เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.6.2 ฟาส ฟูเรีย ทรานฟอรม์ (Fast Fourier Transform:FFT)

เอฟเอฟทีเป็นอัลกอริทึมที่ใช้ในการคำนวณหาดีเอฟที ให้รวดเร็วยิ่งขึ้น ซึ่งจากเดิมในการหาดีเอฟที ของข้อมูลจำนวน N จุด ต้องใช้ความซับซ้อนมากถึง $O(N^2)$ แต่การใช้เอฟเอฟทีคำนวณจะใช้ความซับซ้อนเพียงแค่ O(N log N) และได้ผลลัพธ์ที่ค่อนข้างใกล้เคียงกับการหาดีเอฟทีซึ่งเอฟเอฟทีเป็นอัลกอริทึมที่มีหลายเทคนิคหลายวิธีมาก ยกตัวอย่างเช่น Cooley–Tukey FFT algorithm, Prime-factor FFT algorithm, Bruun's FFT algorithm, Rader's FFT algorithm, Bluestein's FFT algorithm เป็นต้น แต่ผลลัพธ์ที่ได้จะคล้ายกันคือได้ฟังก์ชั่นผลลัพธ์เป็นโดเมนความถี่

3.6.3 การวิเคราะห์คุณลักษณะเด่นของเสียงโดยใช้เอฟเอฟที

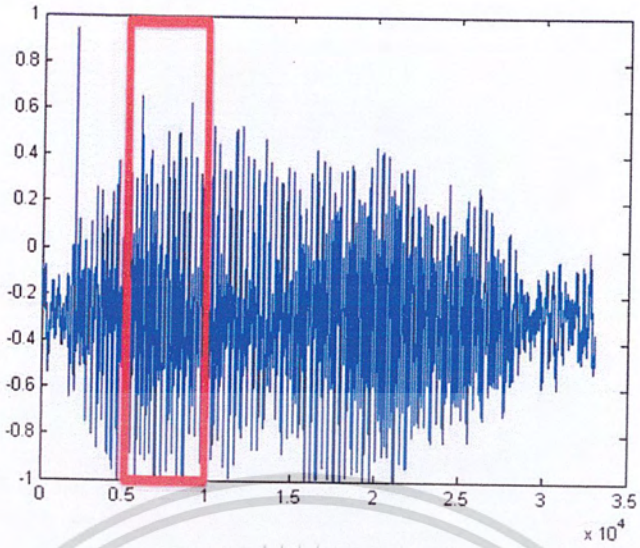
ข้อมูลเสียงที่เราใช้วิเคราะห์หาลักษณะเด่นนั้น เก็บอยู่ในรูปแบบของไฟล์เวฟ ซึ่งเป็นข้อมูลฟังก์ชั่นทางเวลา อีกทั้งมีคุณสมบัติทั้ง ข้อมูลไม่ติดต่อกัน และมีค่าที่แน่นอน อยู่แล้วจึงสามารถนำมาใช้เป็นข้อมูลเข้าของเอฟเอฟทีได้ ตัวอย่างไฟล์เสียง เมื่อเราถอดนำมาเฉพาะค่าข้อมูลของเสียงแล้วจะได้ดังรูป



รูป 3.3 ตัวอย่างข้อมูลเสียงที่ต้องการนำมาวิเคราะห์หาคุณลักษณะเด่น

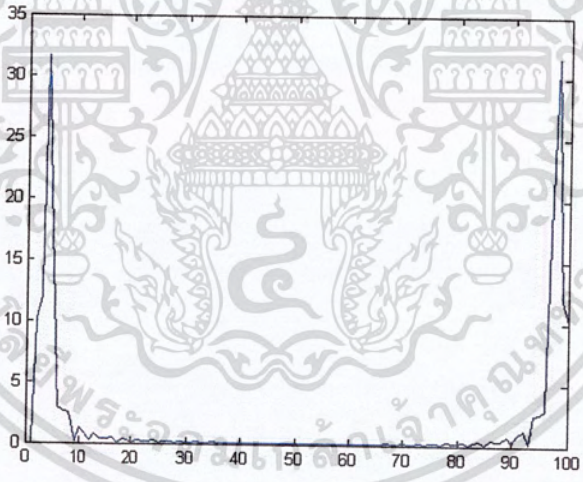
จากนั้นข้อมูลเสียงจะถูกแบ่งเป็นช่วง ๆ แล้วนำเข้าเอฟเอฟที เราจะได้ความถี่ในแต่ละช่วงเวลาออกมา เพื่อนำไปเทรนใน เอชเอ็มเอ็มต่อไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูป 3.4 ตัวอย่างการแบ่งข้อมูลเสียง เพื่อนำเข้าเอฟเอฟที

หลังจากผ่านกระบวนการเอฟเอฟที แล้วข้อมูลที่ได้มีลักษณะดังนี้



รูป 3.5 ตัวอย่างข้อมูลที่ได้จากเอฟเอฟที

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.7 ยูเอ็มแอล (UML)

3.7.1 ความต้องการ

3.7.1.1 ความต้องการที่เป็นฟังก์ชั่น(Functional Requirement)

1) สั่งเปิด-ปิดโปรแกรมวินโดวส์มีเดียเพลเยอร์(Windows Media Player)

2) ควบคุมวินโดวส์มีเดียเพลเยอร์

-เพิ่ม-ลด เสียง

-หยุดการเล่น(Stop)

-หยุดชั่วคราว(Pause)

-เล่น(Play)

-เปิดโปรแกรม

-ปิดเสียง

-เล่นเพลงถัดไป-ก่อนหน้า

3.7.1.2 ความต้องการที่ไม่เป็นฟังก์ชั่น(Non-functional Requirement)

1) แสดงข้อความคำสั่งที่ประมวณผลได้

2) การแสดงสถานะการทำงานของโปรแกรม ขณะนี้กำลังทำอะไรอยู่

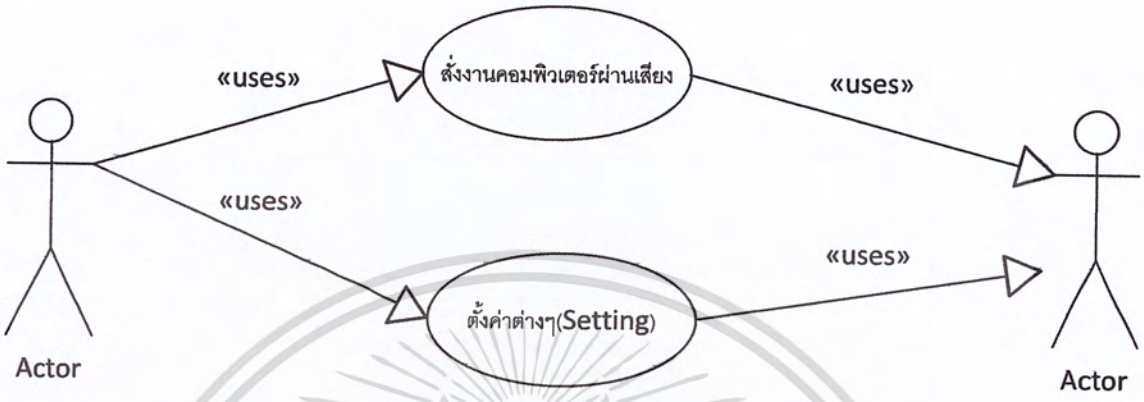
3) ขนาดหน้าต่างโปรแกรม 400*300 พิกเซล

4) เก็บข้อมูลไว้ในดาต้าเบส

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.7.2 ยูสเคสไดอะแกรม(Use Case Diagram)

ผู้ใช้จะใช้งานซอฟต์แวร์ได้สองประเภท คือสั่งงานคอมพิวเตอร์ผ่านเสียงและตั้งค่าต่าง ๆ นำมาเขียนได้ยูสเคสไดอะแกรม ดังภาพ



รูป 3.6 ยูสเคสไดอะแกรม

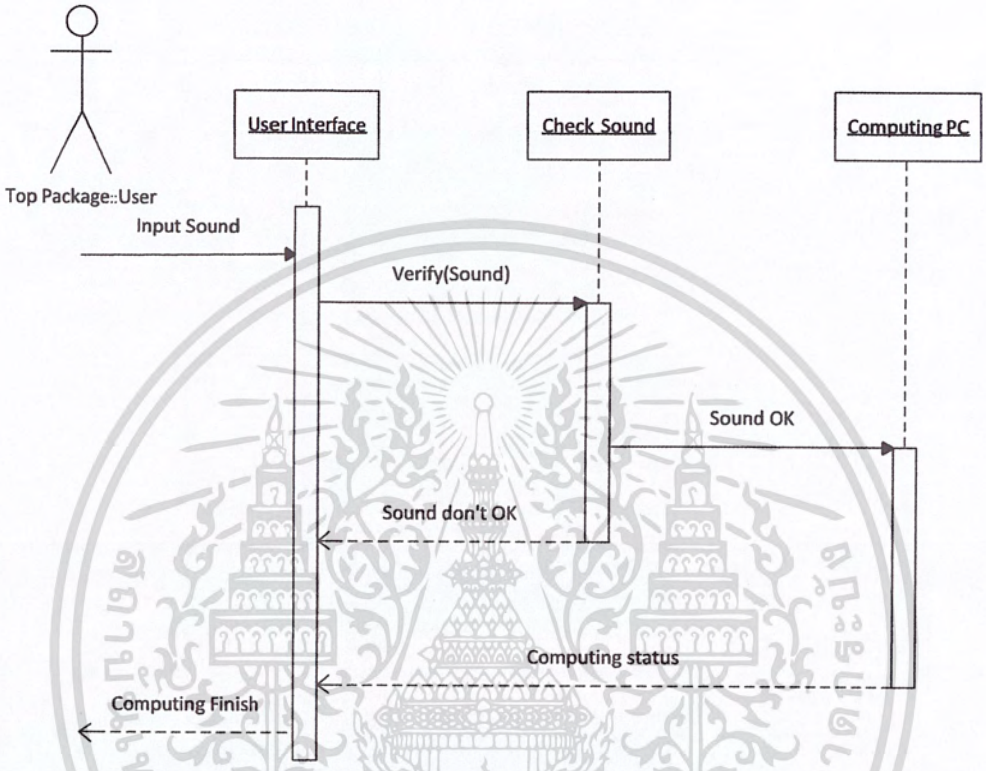
3.1 คำสั่งเสียง 8 คำสั่งที่ใช้ในการสั่งการโปรแกรมวินโดว์มีเดียเพลย์เยอร์

คำสั่งที่	คำสั่งเสียง	รายละเอียดการทำงาน
1	เล่น	สั่งวินโดว์มีเดียเพลย์เยอร์เล่นเพลง
2	หยุด	สั่งวินโดว์มีเดียเพลย์เยอร์หยุดเล่นเพลง
3	ไป	สั่งวินโดว์มีเดียเพลย์เยอร์เล่นเพลงถัดไป
4	กลับ	สั่งวินโดว์มีเดียเพลย์เยอร์เล่นเพลงก่อนหน้า
5	เปิด	สั่งเปิดวินโดว์มีเดียเพลย์เยอร์
6	ปิด	สั่งวินโดว์มีเดียเพลย์เยอร์ปิดเสียง
7	เพิ่ม	สั่งวินโดว์มีเดียเพลย์เยอร์เพิ่มเสียง
8	ลด	สั่งวินโดว์มีเดียเพลย์เยอร์ลดเสียง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

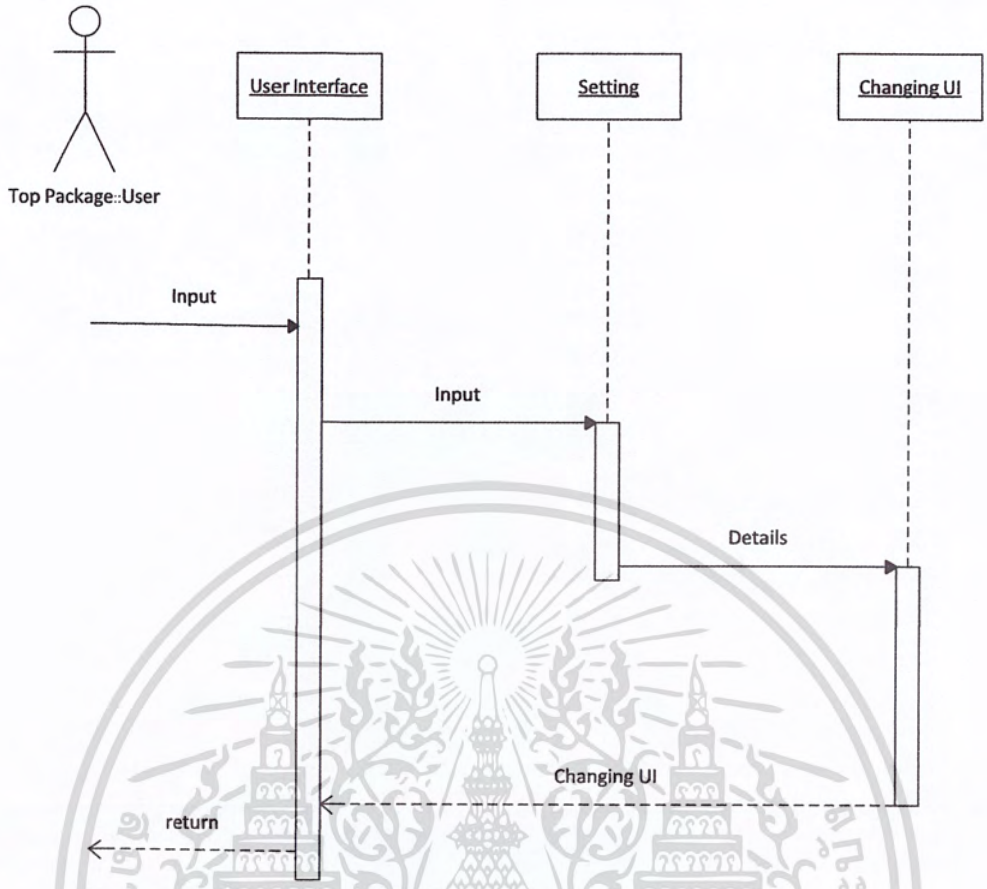
3.7.3 ซีเควนซ์ไดอะแกรม (Sequence Diagram)

จากยูสเคสไดอะแกรมนำมาเขียนเป็นซีเควนซ์ไดอะแกรม ได้ดังนี้



รูป 3.7 ซีเควนซ์ไดอะแกรมการสั่งงานคอมพิวเตอร์ด้วยเสียง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูป 3.8 ซีเควนซ์ไดอะแกรมการกำหนดการแสดงผล

3.7.4 สเตทชาร์ทไดอะแกรม (Statechart Diagram)

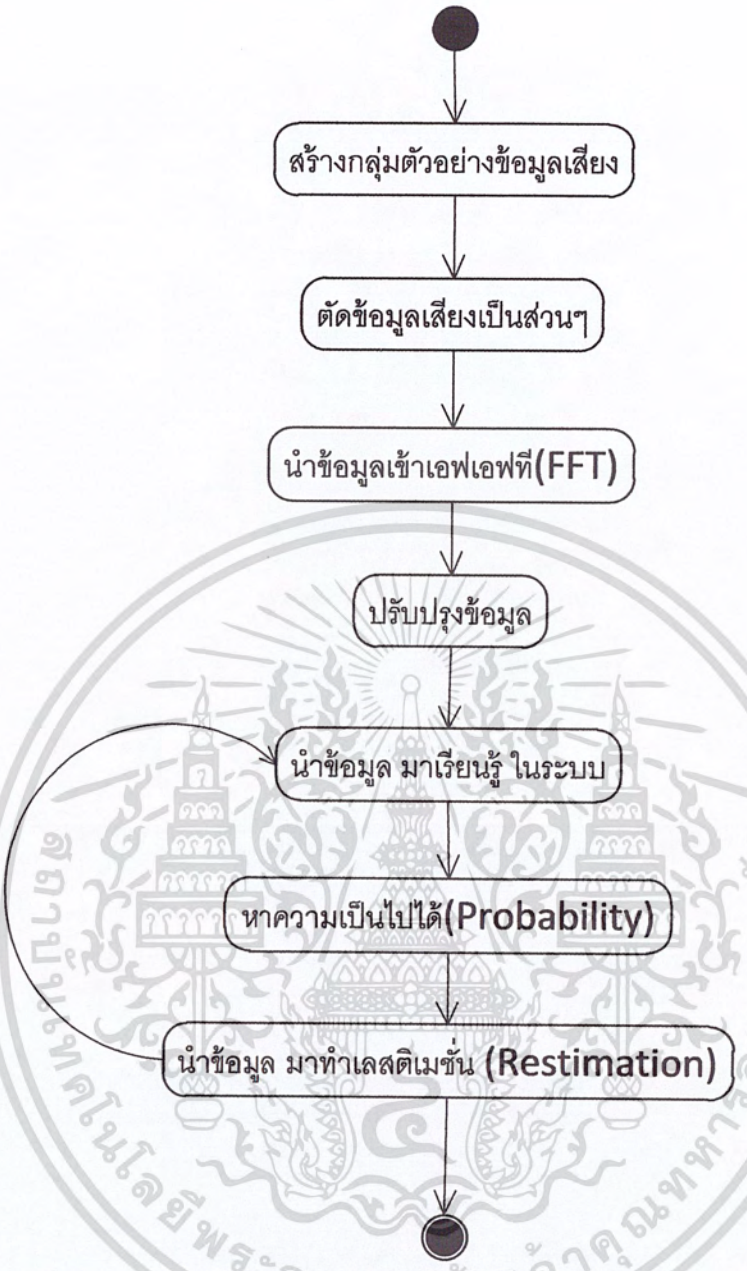
3.7.4.1 การฝึกระบบรู้จำเสียง

ในตอนเริ่มต้น ระบบรู้จำเสียงจะยังไม่มีข้อมูลใด ๆ จำเป็นจะต้องฝึกให้รู้ถึงตัวอย่างเสียงต่าง ๆ ก่อนที่จะนำไปใช้งานได้ แนวคิดในการฝึกระบบรู้จำเสียงมีดังนี้

1) สร้างกลุ่มตัวอย่างข้อมูลเสียงที่ต้องการฝึก (ที่อยู่ในหมวดหมู่เดียวกัน)
2) ตัดข้อมูลเสียงเป็นส่วน ๆ
3) นำข้อมูลแต่ละส่วนเข้าเอฟเอฟที
4) นำข้อมูลมาปรับปรุง (Normalize) ค่าของข้อมูล(0-30) และจำนวนข้อมูล 200 ชุด
5) นำข้อมูลมาเรียนรู้ (Training) เข้า เอชเอ็มเอ็ม โมเดล
6) นำข้อมูลที่ผ่านการเรียนรู้มาหาค่าความเป็นไปได้ (Probability)
7) นำข้อมูลที่ผ่านการเรียนรู้มาทำเลสติเมชั่น (Restimation)

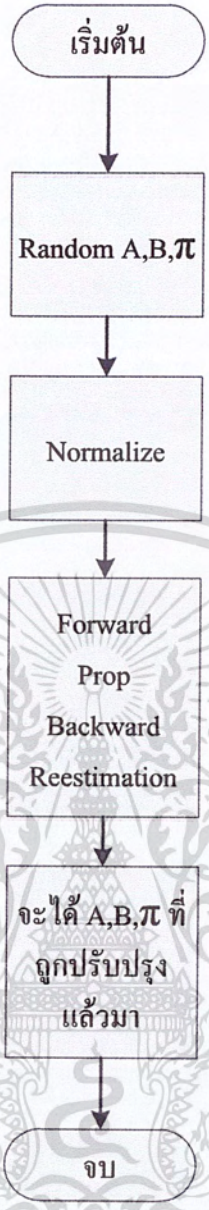
นำมาเขียนเป็น สเตทไดอะแกรม (State Diagram) ได้ดังนี้

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูป 3.9 สเตทชาร์ทไดอะแกรมการรู้จำเสียงพูด

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้



รูป 3.10 สเตทชาร์ทไดอะแกรมการทำงานของฮิดเดนมาร์คอฟโมเดล

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3.7.4.2 การนำระบบรู้จำเสียงไปใช้งาน

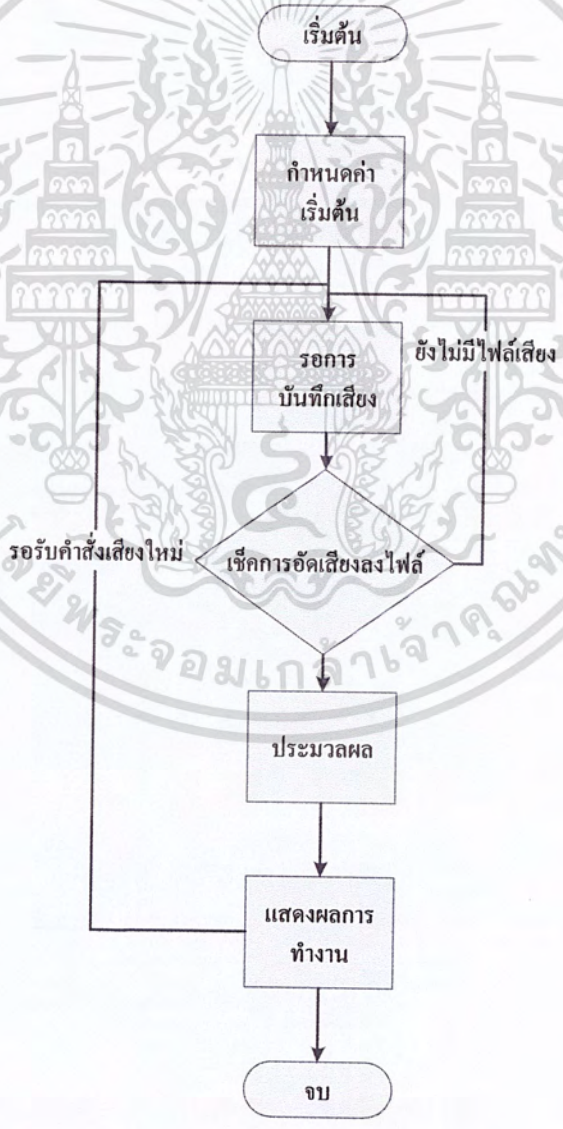
เมื่อระบบรู้จำเสียงผ่านการฝึกมาระดับหนึ่ง จะสามารถนำไปใช้งานได้ แนวคิดการใช้งานระบบรู้จำเสียงมีดังนี้

1) รับข้อมูลเสียงที่ไม่รู้จักหรือไม่ทราบหมวดหมู่เข้ามา

2) วิเคราะห์หาคุณลักษณะเด่นของเสียง

3) หาค่าความน่าจะเป็นจากโมเดลทุก ๆ โมเดลที่มี เลือกโมเดลที่มีความน่าจะเป็นสูงที่สุด

เมื่อข้อมูลเสียงตรงกับลักษณะเด่นของโมเดลใดโมเดลนั้นจะมีค่าความน่าจะเป็นมากกว่าโมเดลอื่น ๆ ถ้าหากค่าโดดเด่นขึ้นมาจากค่าอื่น ๆ (เกณฑ์ในการตัดสินต้องหามาจากการทดลองจริง) จึงเลือกตัดสินว่า ข้อมูลเสียงอินพุทเป็นเสียงประเภทนั้น ๆ



รูป 3.11 สเตทชาร์ทไดอะแกรมการสั่งงานคอมพิวเตอร์ผ่านเสียง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 4

ผลการทดลองและสรุปผลการทดลอง

4.1 การทดลองใช้งานโปรแกรมรู้จำเสียง

การทดลองแบ่งออกเป็น 4 การทดลอง

1) การทดลองที่ 1 ทดลองป้อนไฟล์คำสั่งเสียงจำนวน 5 เสียง ทดสอบเสียงละ 5 ไฟล์ จาก กลุ่มตัวอย่าง 1 คน ได้ผลการทดลองดังตาราง 4.1

ตาราง 4.1 ผลการทดลองคำสั่งเสียงพยางค์เดียว 5 เสียง ทำการทดสอบเสียงละ 5 ครั้ง กลุ่มตัวอย่าง1 คน

ไฟล์เสียง	รู้จำเสียงถูกต้อง(ครั้ง)	รู้จำเสียงผิด(ครั้ง)	เปอร์เซนต์ความผิดพลาด
เล่น	5	0	0
หยุด	2	3	60
ไป	2	3	60
กลับ	3	2	40
ปิด	5	0	0
รวม	17	8	32

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

2) การทดลองที่ 2 ทดลองป้อนไฟล์คำสั่งเสียงจำนวน 8 เสียง ทดสอบเสียงละ 5 ไฟล์ กลุ่มตัวอย่าง 1 คน ได้ผลการทดลองดังตาราง 4.2

ตาราง 4.2 ผลการทดลองคำสั่งเสียงพยางค์เดียว 8 เสียงที่ใช้ในการสั่งงานโปรแกรม ทดสอบเสียงละ 5 ครั้ง กลุ่มตัวอย่าง 1 คน

ไฟล์เสียง	รู้จำเสียงถูกต้อง(ครั้ง)	รู้จำเสียงผิด(ครั้ง)	เปอร์เซนต์ความผิดพลาด
เล่น	5	0	0
หยุด	2	3	60
ไป	3	2	40
กลับ	2	3	60
เปิด	4	1	20
ปิด	5	0	0
เพิ่ม	2	3	60
ลด	2	3	60
รวม	25	15	37.50

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

3) การทดลองที่ 3 ทดลองป้อนไฟล์คำสั่งเสียงจำนวน 8 เสียง ทดสอบเสียงละ 25 ไฟล์ กลุ่มตัวอย่าง 5 คน ได้ผลการทดลองดังตาราง 4.3

ตาราง 4.3 ผลการทดลองคำสั่งเสียงพยางค์เดียว 8 เสียงที่ใช้ในการสั่งงานโปรแกรม ทดสอบเสียงละ 25 ครั้ง กลุ่มตัวอย่าง 5 คน(1 คนเก็บเสียง 5 ครั้งต่อ 1 คำสั่ง)

ไฟล์เสียง	รู้จำเสียงถูกต้อง(ครั้ง)	รู้จำเสียงผิด(ครั้ง)	เปอร์เซนต์ความผิดพลาด
เล่น	18	7	28
หยุด	17	8	32
ไป	10	15	60
กลับ	15	10	40
เปิด	16	9	36
ปิด	11	14	56
เพิ่ม	9	16	64
ลด	14	11	44
รวม	110	90	45

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4) การทดลองที่ 4 ทดลองป้อนไฟล์คำสั่งเสียงจำนวน 8 เสียง ทดสอบเสียงละ 50 ไฟล์ กลุ่มตัวอย่าง 10 คน ได้ผลการทดลองดังตาราง 4.4

ตาราง 4.4 ผลการทดลองคำสั่งเสียงพยางค์เดียว 8 เสียงที่ใช้ในการสั่งงานโปรแกรม ทดสอบเสียงละ 50 ครั้ง กลุ่มตัวอย่าง 10 คน(1 คนเก็บเสียง 5 ครั้งต่อ 1 คำสั่ง)

ไฟล์เสียง	รู้จำเสียงถูกต้อง(ครั้ง)	รู้จำเสียงผิด(ครั้ง)	เปอร์เซนต์ความผิดพลาด
เล่น	37	13	26
หยุด	25	25	50
ไป	30	20	40
กลับ	10	40	80
เปิด	31	19	38
ปิด	45	5	10
เพิ่ม	12	38	76
ลด	17	33	66
รวม	207	193	48.25

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.2 วิเคราะห์และสรุปผลการทดลอง

จากการทดลองคำสั่งเสียง เสียงที่มีคุณภาพดีคือ เสียงเล่น เสียงเปิดและเสียงปิด มีความผิดพลาดน้อยมาก ส่วนคำสั่งเสียงที่มีความผิดพลาดสูงได้แก่ เสียงหยุด เสียงกลับ เสียงเพิ่มและเสียงลด

จากการทดลอง 2 ได้ทำการเพิ่มเสียงคำพยางค์เดียวจากจำนวน 5 คำ เป็น 8 คำ ทำให้ค่าเปอร์เซนต์ความผิดพลาดเพิ่มขึ้นจาก 32 เป็น 37.5 จึงสรุปว่า การเพิ่มจำนวนเสียงที่รู้จำในระบบจะทำให้เปอร์เซนต์ความผิดพลาดเพิ่มขึ้น

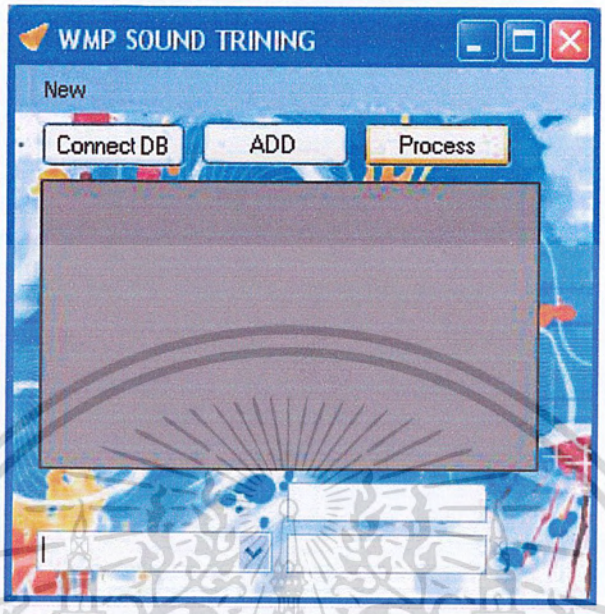
จากการทดลอง 3และ4 เมื่อทำการเพิ่มกลุ่มตัวอย่างเพิ่มขึ้นจาก 1 คนในการทดลอง 2 เป็น 5 คนในการทดลอง 3 และเพิ่มเป็น 10 คนในการทดลอง 4 พบว่าการเพิ่มจำนวนกลุ่มตัวอย่างเพิ่มขึ้น ทำให้เปอร์เซนต์ความผิดพลาดเพิ่มขึ้นเป็น 45 ในการทดลอง 3 และ 48.25 ในการทดลอง 4 ซึ่งค่าที่ได้นี้ถือว่าสูงมากในการรู้จำคำ

สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.3 การทดสอบโปรแกรม

4.3.1 หน้าตาโปรแกรม WMP SOUND TRINING



รูป 4.1 หน้าต่างการทำงานโปรแกรม WMP SOUND TRINING

โปรแกรม WMP SOUND TRINING นี้เราใช้เพื่อเรียนรู้เสียงพูด และจัดเก็บค่าพารามิเตอร์ต่าง ๆ ลงฐานข้อมูล เพื่อใช้ในการสั่งงานคอมพิวเตอร์โดยโปรแกรม WMP SOUND CONTROL ต่อไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

4.3.2 หน้าตาโปรแกรม WMP SOUND CONTROL

รูป 4.2 หน้าต่างการทำงานโปรแกรม WMP SOUND CONTROL

เมื่อเราเทรนเสียงที่ใช้รู้จำโดยโปรแกรม WMP SOUND TRINING แล้ว เราสามารถกดเริ่มการสั่งการโปรแกรม WMP SOUND CONTROL ซึ่งในการสั่งงานโปรแกรม WMP SOUND CONTROL จะใช้คำสั่งเสียง 8 คำสั่ง

-เล่น สั่งวินโดว์มีเดียเพลย์เยอร์เล่นเพลง

-หยุด สั่งวินโดว์มีเดียเพลย์เยอร์หยุดเล่นเพลง

-ไป สั่งวินโดว์มีเดียเพลย์เยอร์เล่นเพลงถัดไป

-กลับ สั่งวินโดว์มีเดียเพลย์เยอร์เล่นเพลงก่อนหน้า

-เปิด สั่งเปิดวินโดว์มีเดียเพลย์เยอร์

-ปิด สั่งวินโดว์มีเดียเพลย์เยอร์ปิดเสียง

-เพิ่ม สั่งวินโดว์มีเดียเพลย์เยอร์เพิ่มเสียง

-ลด สั่งวินโดว์มีเดียเพลย์เยอร์ลดเสียง

เมื่อเราสั่งงานเสียงได้ถูกต้อง คอมพิวเตอร์ก็จะทำงานตามคำสั่งนั้น ๆ

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บทที่ 5
วิจารณ์ผลการทดลองและแนวทางการพัฒนาต่อ

5.1 การวิจารณ์ผลการทดลอง

การทดลองครั้งนี้ทดลองเพื่อทดสอบซอฟท์แวร์ที่ได้พัฒนาขึ้น ว่าได้ผลลัพธ์ถูกต้องตามวัตถุประสงค์ สามารถสั่งงานโปรแกรมวินโดวส์มีเดียเพลเยอร์

ผลการทดลองที่ได้จากกลุ่มตัวอย่างเพียง 1 คนมีเปอร์เซนต์ความผิดพลาดน้อยที่สุด แต่ก็ยังมีเปอร์เซนต์ความผิดพลาดถึง 37.5 ทำให้การสั่งงานด้วยเสียงนั้นยังขาดประสิทธิภาพอยู่มาก เพราะการสั่งงานคอมพิวเตอร์ไม่ควรมีความผิดพลาดเกิดขึ้นเลย อาจทำให้เกิดปัญหากับโปรแกรมที่ทำงานอยู่ ผู้จัดทำจึงได้ใช้เสียงสั่งงานโปรแกรมวินโดวส์มีเดียเพลเยอร์ซึ่งแม้สั่งงานผิดพลาด ก็ไม่ส่งผลกระทบต่อคอมพิวเตอร์ หรือโปรแกรมต่าง ๆ ที่ทำงานอยู่

ผลการทดลองยังมีความผิดพลาดยังมีอยู่มากควรปรับปรุงให้น้อยลง เพื่อความถูกต้องของการรู้จำ

5.2 ผลที่ได้รับจากโครงงาน

1) มีความเข้าใจ และสามารถแก้ไขปัญหาด้วยฮิดเดนมาร์คอฟโมเดล
2) ได้รับทักษะการทำงานร่วมกันเป็นกลุ่ม แบ่งงานตามความถนัดและหน้าที่รับผิดชอบของสมาชิกแต่ละคน
3) ทักษะการพัฒนาซอฟท์แวร์ ด้วยภาษาซี พลัส พลัส (C ++) และมายเอสคิวแอล (MySQL)

5.3 ปัญหาที่พบ

1) ไม่เข้าใจระบบการทำงานของไลบรารีเอฟเอฟทีดับเบิ้ลยู (FFTW) ซึ่งเป็น ไลบรารีฟาสฟูเรียร์ทรานสฟอร์ม
2) เนื่องจากข้อมูลมีขนาดใหญ่ ทำให้ต้องทำการจองพื้นที่บนหน่วยความจำ ซึ่งถ้าการจัดสรรการใช้หน่วยความจำไม่ดี จะทำให้หน่วยความจำไม่พอใช้
3) ภาษา ซี พลัส พลัส ต้องการชนิดตัวแปรคงที่ตายตัว ทำให้เกิดปัญหาการสูญหายของทศนิยมในการคำนวณ

5.4 แนวทางแก้ไขปัญหา

1) ศึกษาและทดลองการใช้งาน ไลบรารีเอฟเอฟทีดับเบิ้ลยู
2) ควรวางแผนการรับข้อมูลและ ลบข้อมูลที่ไม่ได้ใช้แล้วทิ้งไป

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

5.5 แนวทางในการพัฒนาต่อ

1) เพิ่มจำนวนคำในระบบการรู้จำ

2) พัฒนาประสิทธิภาพของระบบรู้จำเสียงให้มีเปอร์เซนต์ความผิดพลาดน้อยลง

3) พัฒนาวิธีหาลักษณะเด่นของเสียงให้มีประสิทธิภาพมากขึ้น

4) เพิ่มคำสั่งใหม่ สำหรับใช้สั่งงานคอมพิวเตอร์

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

บรรณานุกรม

วิกิพีเดีย สารานุกรมเสรี. 2010. [Online].

Available : http://th.wikipedia.org/wiki/เสียง

วิกิพีเดีย สารานุกรมเสรี. 2010. [Online].

Available : http://th.wikipedia.org/wiki/จิตสวนศาสตร์

Wikipedia, the free encyclopedia. 2010. [Online].

Available : http://en.wikipedia.org/wiki/Voice_frequency

Frigo, M. and Johnson, S.G. 2005. The Design and Implementation of FFTW3.

Proc. of the IEEE Vol. 93.

Frigo, M. and Johnson, S.G. 2005. "Fastest Fourier Transform in the West.". [Online].

Available : http://www.fftw.org/

Wikipedia. 2010. Fast Fourier transform. [Online].

Available : http://en.wikipedia.org/wiki/Fast_Fourier_Transform

Wikipedia. 2010. Discrete Fourier transform. [Online].

Available : http://en.wikipedia.org/wiki/Discrete_Fourier_transform

Michael S. Kaplan. 2006. Getting all you can out of a keyboard layout, Part #0. [Online].

Available : http://blogs.msdn.com/b/michkap/archive/2006/03/23/558658.aspx

Armen Hakobyan. 2002. Toggling the Num Lock, Caps Lock, and Scroll Lock keys.. [Online].

Available : http://www.codeproject.com/KB/cpp/togglekeys.aspx

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ภาคผนวก

การทำงานของโปรแกรม

การควบคุมวินโดว์มีเดียเพลย์เยอร์ด้วยคำสั่งเสียงจะใช้โปรแกรมทั้งหมด 2 โปรแกรมด้วยกัน คือ โปรแกรม WMP SOUND TRINING ซึ่งใช้ในการเทรนเสียงและโปรแกรม WMP SOUND CONTROL ที่ใช้ในการรับเสียงและสั่งงานวินโดว์มีเดียเพลย์เยอร์

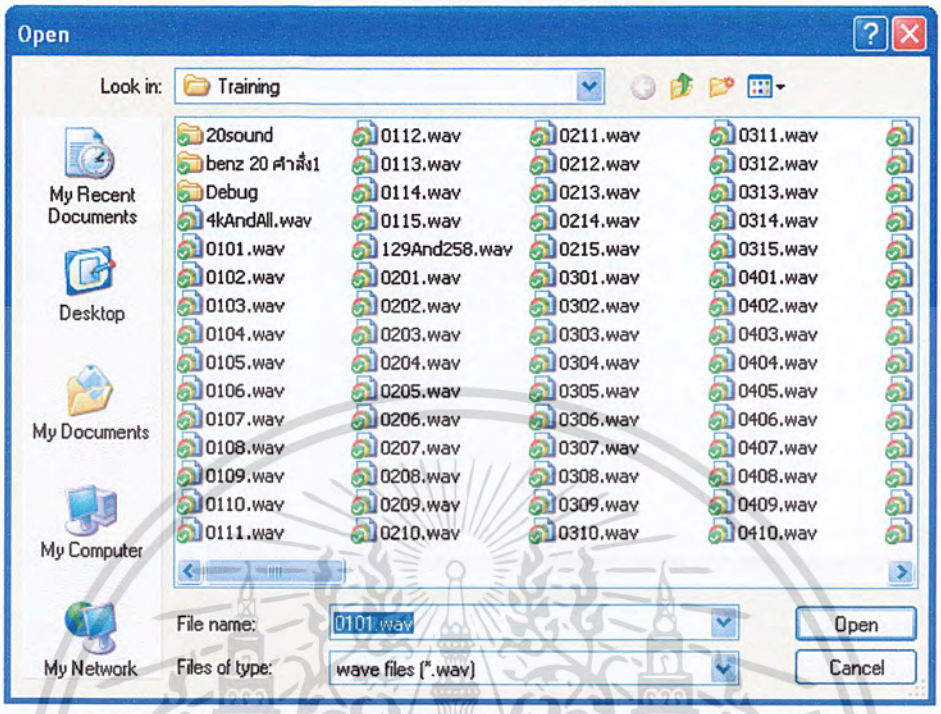
ก.1 การทำงานของโปรแกรม WMP SOUND TRINING

ขั้นที่ 1 เลือก Menu -> Training

รูป ก.1 หน้าต่างแสดงการทำงานของโปรแกรม WMP SOUND TRINING

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ขั้นที่ 2 เลือกไฟล์เสียงที่จะนำมาเทรนโดยไฟล์เสียงที่เลือกจะต้องเป็นคำเดียวกัน



รูป ก.2 หน้าต่างแสดงการทำงานของโปรแกรม WMP SOUND TRINING

ขั้นที่ 3 โปรแกรมจะแสดงการนับไฟล์เสียงให้ดูว่าขณะนี้ใส่ไฟล์เสียงที่ใช้ในการเทรนกี่ไฟล์

รูป ก.3 หน้าต่างแสดงการทำงานของโปรแกรม WMP SOUND TRINING

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ขั้นที่ 4 เลือกคำที่ใช้ในการเทรน

รูป ก.4 หน้าต่างแสดงการทำงานของโปรแกรม WMP SOUND TRINING

ขั้นที่ 5 กด Process

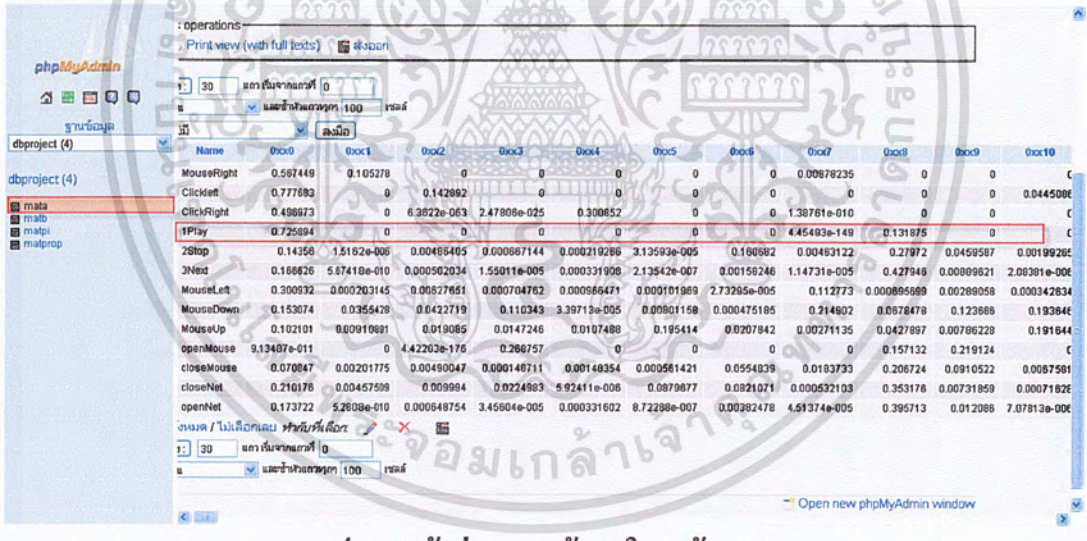
รูป ก.5 หน้าต่างแสดงการทำงานของโปรแกรม WMP SOUND TRINING

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ขั้นที่ 6 รอจนกว่าโปรแกรมจะแสดงตัวเลขในช่อง Process loop

รูป ก.6 หน้าต่างแสดงการทำงานของโปรแกรม WMP SOUND TRINING

ขั้นที่ 7 เมื่อประมวลผลเสร็จ โปรแกรมจะเก็บผลลัพธ์ที่ได้ลงดาต้าเบส



รูป ก.7 หน้าต่างแสดงข้อมูลในดาต้าเบส

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้

ก.2 การทำงานของโปรแกรม WMP SOUND CONTROL

โปรแกรม WMP SOUND CONTROL เป็นโปรแกรมสำหรับสั่งงานวินโดว์มีเดียเพลย์เยอร์ด้วยเสียง วิธีการใช้งานโดยกดปุ่มรัน แล้วสั่งงานด้วยคำสั่งที่มีบอกไว้ในโปรแกรมก็จะสามารถควบคุมโปรแกรมวินโดว์มีเดียเพลย์เยอร์ โดยใช้เสียงได้ทันที

รูป ก.8หน้าต่างแสดงการทำงานของโปรแกรม WMP SOUND CONTROL

เอกสารนี้เป็นเอกสารที่สงวนไว้สำหรับการใช้งานเพื่อการศึกษาเท่านั้น ไม่อนุญาตให้นำไปใช้ประโยชน์ด้านการค้า ไม่ว่ากรณีใดๆ ทั้งสิ้น อีกทั้งห้ามมิให้ดัดแปลงเนื้อหา และต้องอ้างอิงถึงเจ้าของเอกสารทุกครั้งที่มีการนำไปใช้