การค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงภาษาไทยด้วยข้อคำถามเสียง

นายภูเบศ โต๊ะลง

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาวิทยาศาสตร์คอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์
คณะคณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2549
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

THAI SPEECH AUDIO RETRIEVAL USING VOICE QUERY

Mr. Phubes Tohlong

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

A Thesis Submitted in Partial Fulfillment of the Requirements

for the Degree of Master of Science Program in Computer Science

Department of Computer Engineering

Faculty of Engineering

Chulalongkorn University

Academic Year 2006

Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์	การค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงภาษาไทยด้วยข้อคำถามเสียง
โดย	นายภูเบศ โต๊ะลง
สาขาวิชา	วิทยาศาสตร์คอมพิวเตอร์
อาจารย์ที่ปรึกษา	อาจารย์ ดร.โชติรัตน์ รัตนามหัทธนะ

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

..... คณบดีคณะวิศวกรรมศาสตร์

(ศาสตราจารย์ ดร.ดิเรก ลาวัณย์ศิริ)

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ

(รองศาสตราจารย์ ดร.ประภาส จงสถิตย์วัฒนา)

..... อาจารย์ที่ปรึกษา

(อาจารย์ ดร.โชติรัตน์ รัตนามหัทธนะ)

..... กรรมการ

(อาจารย์ ดร.อดิวงศ์ สุชาโต)

..... กรรมการ

(อาจารย์ ดร.วิษณุ โคตรจรัส)

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภูเบศ โต๊ะลง : การค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงภาษาไทยด้วยข้อคำถามเสียง. (THAI SPEECH AUDIO RETRIEVAL USING VOICE QUERY) อาจารย์ที่ปรึกษา : อ.ดร.โชติรัตน์ รัตนามหัทธนะ, 66 หน้า.

ปัจจุบันข้อมูลสื่อประสมได้เพิ่มปริมาณขึ้นอย่างรวดเร็วและมีหลายรูปแบบ ทั้งที่อยู่ในรูปแฟ้มข้อมูลเสียง แฟ้มข้อมูลวีดิทัศน์ และแฟ้มข้อมูลภาพ ซึ่งแฟ้มข้อมูลสื่อประสมแต่ละแบบมีวิธีการค้นคืนหลากหลายวิธี งานวิจัยนี้สนใจและเลือกที่จะทำการศึกษาวิธีการที่จะค้นคืนข้อมูลภายในแฟ้มข้อมูลเสียงภาษาไทยขนาดใหญ่ เช่น แฟ้มข้อมูลเสียงสื่อการเรียนการสอนอิเล็กทรอนิกส์ แฟ้มข้อมูลเสียงการอ่านหนังสือ เป็นต้น ปัจจุบันวิธีที่ได้รับความนิยมในการค้นคืนข้อมูลภายในแฟ้มข้อมูลเสียงมักใช้วิธีการสืบค้นด้วยคำหลัก ชื่อเรื่องหรือชื่อผู้แต่ง ซึ่งวิธีการดังกล่าวเป็นการค้นคืนด้วยการพิมพ์ หรือแม้แต่การพูดข้อคำถามเสียงเข้าไปเพื่อค้นหาจากรายการที่มีอยู่ โดยใช้กระบวนการรู้จำคำพูดในการค้นคืนข้อมูลเสียง แต่การใช้กระบวนการรู้จำคำพูดมีข้อจำกัดในเรื่องของเวลาที่ใช้ในการค้นคืน ซึ่งใช้เวลานานในกรณีที่แฟ้มฐานข้อมูลเสียงมีขนาดใหญ่ ดังนั้นงานวิจัยนี้จึงมุ่งเน้นในการค้นคืนข้อมูลเสียงจากแฟ้มข้อมูลเสียงภาษาไทยขนาดใหญ่ ซึ่งเวลาที่ใช้ในการทำงานเป็นเวลาที่ผู้ใช้ยอมรับได้ โดยมีความแม่นยำอยู่ในระดับดี และเนื่องจากภาษาไทยมีการผันวรรณยุกต์ 5 ระดับเสียงต่างกัน คือ สามัญ เอก โท ตรี และจัตวา ผู้เขียนจึงได้เลือกใช้เสียงวรรณยุกต์ในภาษาไทยนี้เข้ามาช่วยในการแยกคำ ซึ่งวรรณยุกต์ในแต่ละพยางค์ของคำก็จะให้ค่าความถี่มูลฐานต่างกัน และสามารถนำเอาคุณลักษณะพิเศษของเสียงในภาษาไทยนี้มาใช้ในการค้นหาคำจากข้อคำถามเสียงโดยใช้วิธีวัดระยะทางแบบไดนามิกไทม์วอร์ปปิง เพื่อช่วยเพิ่มความแม่นยำในการเปรียบเทียบสัญญาณเสียงจากข้อคำถามกับเสียงในแฟ้มฐานข้อมูล จากการทดลอง พบว่าวิธีดังกล่าวสามารถค้นคืนข้อมูลเสียงได้ถูกต้องคิดเป็น 59 เปอร์เซ็นต์

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาควิชา...วิศวกรรมคอมพิวเตอร์... ลายมือชื่อนิสิต...ภูเบศ โต๊ะลง...
สาขาวิชา...วิทยาศาสตร์คอมพิวเตอร์... ลายมือชื่ออาจารย์ที่ปรึกษา...
ปีการศึกษา...2006...

##4871436521 : MAJOR COMPUTER SCIENCE

KEY WORD: AUDIO RETRIEVAL / VOICE QUERY / PITCH DETECTION / DYNAMIC TIME WARPING

PHUBES TOHLONG : THAI SPEECH AUDIO RETRIEVAL USING VOICE QUERY. THESIS ADVISOR : CHOTIRAT RATANAMAHATANA, Ph.D., 66 pp.

Multimedia has increasingly become a prevalent resource in various formats including audio, video, and image archives. Among the varieties of retrieval, this thesis focuses on retrieval of speech audio collections, which include electronic lectures and audio books. Currently, most of audio retrieval systems are based on typed keyword/title/author search or based on voice queries where a speech recognition technique is generally used. However, the main limitation of the speech recognition technique is its slow retrieval time if the audio files are large. Therefore, this research focuses on finding an alternative to speech audio retrieval within the large files with satisfactory retrieval time and accuracy. This work uses Thai tones to help spotting the words because Thai language has 5 different tones, i.e., Low, Middle, High, Falling, and Rising. By exploiting this special property, Fundamental Frequency and Dynamic Time Warping techniques are used to improve performance and to speed up retrieval time. The preliminary experiment result gives a retrieval accuracy of 59%.

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Department Computer Engineering Student's signature

Field of study Computer Science Advisor's signature

Academic year 2006

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงได้ด้วยดี เนื่องมาจากความช่วยเหลืออย่างดียิ่งของท่าน อ.ดร.โชติรัตน์ รัตนามหัทธนะ อาจารย์ที่ปรึกษาวิทยานิพนธ์ ที่ได้สละเวลาให้คำปรึกษาแนะนำแนวทางเกี่ยวกับงานวิจัยอย่างดีตลอดมาจนเสร็จสมบูรณ์ และผู้วิจัยขอกราบขอบพระคุณคณะกรรมการสอบวิทยานิพนธ์ทุกท่านที่ได้ให้คำแนะนำ ข้อคิดเห็น ข้อเสนอแนะ และแนวทางในการพัฒนางานวิจัยนี้

ขอขอบคุณท่านอาจารย์ ดร.อติวงศ์ สุชาโต น.ส.ศิรินาถ ตั้งรวมทรัพย์ นายไพโรจน์ ลีลาภัทรกิจ และน้องๆ ทุกคนที่ให้คำแนะนำและช่วยเหลือในส่วนของการทดลองที่เกี่ยวกับกระบวนการรู้จำคำพูดให้สำเร็จลุล่วงเป็นอย่างดี

ขอขอบคุณพี่ๆ และเพื่อนๆ ทุกคนที่เสียสละเวลามาบันทึกเสียงพูด เพื่อใช้ในงานวิจัยฉบับนี้

ขอขอบคุณ พี่ธุรการภาคฯ ทุกๆ คนที่ช่วยอำนวยความสะดวกในการทำงาน และช่วยตักเตือนแนะนำสิ่งดีๆ เสมอมา

สุดท้ายนี้ ขอกราบขอบพระคุณคุณพ่อคุณแม่ที่ให้โอกาสเราได้เกิด ได้เติบโต ได้เลี้ยงดูเป็นอย่างดี และคอยสนับสนุนในด้านการศึกษาเป็นอย่างดีเสมอมา

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญตาราง

ตาราง หน้า

ตาราง หน้า

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญภาพ

ภาพประกอบ หน้า

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ปัจจุบันข้อมูลสื่อประสมได้เพิ่มปริมาณขึ้นอย่างรวดเร็วและมีหลายรูปแบบทั้งที่อยู่ในรูปแฟ้มข้อมูลเสียง แฟ้มข้อมูลวีดิทัศน์ และแฟ้มข้อมูลภาพ แต่ผู้เขียนมีความสนใจและเลือกที่จะทำการศึกษาวิธีการที่จะค้นหาข้อมูลภายในแฟ้มข้อมูลเสียงภาษาไทย เนื่องจากแรงจูงใจหลักของการทำงานวิจัยนี้เกิดจากการที่ผู้เขียนลงเรียนวิชา ทฤษฎีการคณนา (Theory of Computation) และได้ทำการบันทึกเสียงขณะอาจารย์กำลังสอนไว้ เพื่อเปิดฟังในขณะอ่านหนังสือสอบ เป็นการทบทวนเนื้อหาและความเข้าใจว่าถูกต้องตรงตามที่ท่านอาจารย์ได้สอนหรือไม่ แต่ปรากฏว่าต้องใช้เวลานานในการหาว่าสิ่งที่เราต้องการจะฟังนั้นอยู่ในส่วนใดของแฟ้มข้อมูลเสียง จากสาเหตุดังกล่าวจึงทำให้ผู้เขียนคิดหาวิธีที่จะทำการค้นหาข้อมูลที่อยู่ภายในแฟ้มข้อมูลเสียงภาษาไทยด้วยการใช้ ข้อคำถามเสียง (Voice Query) ซึ่งเป็นการค้นคืนข้อมูลที่สามารถตอบสนองด้วยเวลาที่เหมาะสม ที่ผู้ใช้สามารถรอผลการค้นคืนข้อมูลเสียงได้

จากการศึกษาพบว่าวิธีการค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงขนาดใหญ่นั้น ปัจจุบันมีอยู่หลายวิธีแต่วิธีที่ได้รับความนิยมสูงสุดคือกลวิธีทางด้านกระบวนการรู้จำคำพูด (Speech Recognition) ซึ่งปัจจุบันใช้เวลาในการค้นคืนนาน ดังนั้นผู้เขียน จึงศึกษาวิธีการต่างๆและทำการทดลองเกี่ยวกับการค้นคืนข้อมูลเสียง เช่น การค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงด้วยข้อคำถามเสียงพยางค์เดียว การค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงด้วยข้อคำถามเสียงหลายพยางค์ เพื่อศึกษาเกี่ยวกับการค้นคืนข้อมูลเสียงด้วยการใช้เสียงวรรณยุกต์ในภาษาไทย เนื่องจากภาษาไทยมีการผันวรรณยุกต์ 5 ระดับเสียงต่างกัน และเปรียบเทียบผลการทดลองที่ทำในแต่ละวิธี เพื่อหาวิธีที่ดีที่สุดที่ช่วยลดเวลาในการค้นคืนข้อมูลเสียงและให้ค่าความแม่นยำในระดับดี โดยขั้นตอนและทฤษฎีที่ผู้เขียนสนใจ คือ ทำการบันทึกแฟ้มข้อมูลเสียงและข้อคำถามเสียง ด้วยไมโครโฟนผ่านโปรแกรมบันทึกเสียง จัดเก็บอยู่ในรูปของไฟล์ WAV ทำการลดอัตราการชักตัวอย่าง (Sampling Rate) เหลือ 2000 เฮิรตซ์ และเก็บข้อมูลขนาด 16 บิต แบบช่องสัญญาณเดียว (Mono) หลังจากนั้นนำแฟ้มข้อมูลเสียงที่ทำการบันทึกมาทำการสกัดค่าความถี่มูลฐาน (Fundamental Frequency) ด้วยโปรแกรม Praat และทำการปรับเรียบ (Smoothing) หลังจากทำการปรับเรียบแล้วทำการปรับข้อมูลให้เป็นบรรทัดฐานด้วยค่าคะแนนมาตรฐานซี (Z-Score Normalization) ส่วนวิธีการที่ใช้ในการเปรียบเทียบสัญญาณระหว่างข้อคำถามเสียงกับข้อมูลที่อยู่ในแฟ้มข้อมูลเสียงใช้ วิธีวัดระยะทางแบบไดนามิกไทม์วอร์ปปิง (Dynamic Time Warping หรือ DTW) ในส่วน

ของการแสดงผลลัพธ์ที่ได้จากการค้นหาใช้วิธีการจำแนกประเภทแบบ *K* ลำดับที่ใกล้ที่สุด (*K*-Nearest Neighbor)

ดังที่กล่าวมาข้างต้น เพื่อเพิ่มความเร็วในการค้นคืนข้อมูลเสียงจากแฟ้มข้อมูลเสียงที่มีขนาดใหญ่ ผู้เขียนจึงเสนอวิธีการดังกล่าวในการค้นหาข้อมูลจากแฟ้มข้อมูลเสียงภาษาไทยด้วยข้อคำถามเสียง

1.2 วัตถุประสงค์ของการวิจัย

ศึกษากระบวนการที่ใช้ในการค้นคืนข้อมูลเสียงภาษาไทยจากแฟ้มข้อมูลเสียงขนาดใหญ่ด้วยข้อคำถามเสียง (Voice Query) ซึ่งเวลาที่ใช้ในการทำงานเป็นเวลาที่ผู้ใช้ยอมรับได้

1.3 ขอบเขตของการวิจัย

1. แฟ้มข้อมูลเสียงและข้อคำถามเสียงที่ใช้อยู่ในรูปของเสียงภาษาไทย
2. การค้นคืนข้อมูลเสียงจากแฟ้มข้อมูลเสียงภาษาไทยด้วยข้อคำถามเสียง เป็นการค้นคืนข้อมูลเสียงทีละแฟ้มข้อมูลเสียง
3. ใช้ไมโครโฟนและโปรแกรมบันทึกเสียงในการบันทึกเสียง จัดเก็บในรูปแบบของไฟล์ WAV ซึ่งขนาดไฟล์ที่ทำการบันทึกมีความยาวอย่างน้อย 60 นาที
4. ทำการลดอัตราการชักตัวอย่าง (Sampling Rate) ลงเหลือ 2000 เฮิรตซ์ และเก็บข้อมูลขนาด 16 บิต แบบช่องสัญญาณเดียว (Mono)
5. ข้อคำถามเสียงที่ใช้ในการทดลองเป็นทั้งคำพยางค์เดียว และคำหลายพยางค์
6. แฟ้มข้อมูลเสียงที่ใช้ในการทดลองเป็นเสียงผู้ชายและผู้หญิง
7. ทำการทดลองค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงภาษาไทยด้วยข้อคำถามเสียง กรณีที่ข้อคำถามเสียงกับแฟ้มข้อมูลเสียงเป็นคนพูดคนเดียวกัน (Speaker Dependent)
8. ทำการทดลองค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงภาษาไทยด้วยข้อคำถามเสียง กรณีที่ข้อคำถามเสียงกับแฟ้มข้อมูลเสียงเป็นคนพูดคนละคน (Speaker Independent)
9. ทำการวัดค่าอัตราความถูกต้องการค้นคืน เพื่อประเมินประสิทธิภาพของวิธีการที่ใช้ในการค้นคืนข้อมูล

1.4 ขั้นตอนการวิจัย

1. ศึกษากลวิธีต่างๆ ที่เกี่ยวข้องกับการค้นคืนเสียง
2. ใช้แฟ้มข้อมูลเสียงสื่อการเรียนการสอนอิเล็กทรอนิกส์ แฟ้มข้อมูลเสียงจากอินเทอร์เน็ต และทำการบันทึกเสียงเอง โดยใช้ไมโครโฟนโซนี (Sony) ผ่านโปรแกรม

บันทึกเสียง Sony Sound Forge 8.0 ในการบันทึกเสียงเพื่อทำการทดลอง โดยที่ขนาดไฟล์มีความยาวอย่างน้อย 60 นาที

3. ศึกษากลวิธีการเตรียมข้อมูลเสียง เช่น การหาค่าความถี่มูลฐาน (Fundamental Frequency) การลดอัตราการชักตัวอย่าง (Sampling Rate) การปรับเรียบ (Smoothing) การแปลงข้อมูลให้เป็นบรรทัดฐาน ด้วยค่าคะแนนมาตรฐานซี (Z-Score Normalization)
4. ศึกษากลวิธีการแปลงฟูเรียร์อย่างเร็ว (Fast Fourier Transform)
5. ศึกษาการใช้โปรแกรม MATLAB
6. ศึกษาการสกัดค่าความถี่มูลฐานโดยใช้โปรแกรม Praat
7. ศึกษากลวิธีวิธีการจำแนกประเภทแบบ *K* ลำดับใกล้ที่สุด (*K*-Nearest Neighbor)
8. ศึกษากลวิธีการเปรียบเทียบสัญญาณเสียงระหว่างข้อคำถามเสียงกับข้อมูลเสียงที่อยู่ในแฟ้มข้อมูลเสียงภาษาไทย เช่น วิธีวัดระยะทางแบบไดนามิกไทม์วอร์ปปิง (Dynamic Time Warping: DTW) และวิธีวัดระยะทางแบบยูคลิด (Euclidean Distance)
9. ศึกษากลวิธีการวัดค่าความถูกต้องการค้นคืน เพื่อประเมินประสิทธิภาพของวิธีการที่ใช้ในการค้นคืนข้อมูล
10. สรุปผลการวิจัยและจัดทำวิทยานิพนธ์เป็นรูปเล่ม

1.5 ประโยชน์ที่คาดว่าจะได้รับจากงานวิจัย

1. ช่วยลดเวลาในการค้นคืนข้อมูลเสียงจากแฟ้มข้อมูลเสียง
2. สามารถค้นคืนข้อมูลเสียงได้ตรงตามความต้องการของผู้ใช้
3. สามารถลดภาระในการคัดกรองข้อมูลที่ต้องการค้นหา

1.6 โครงสร้างของวิทยานิพนธ์

เนื้อหาของวิทยานิพนธ์ฉบับนี้ถูกแบ่งออกเป็น 5 บท ดังนี้คือ บทที่ 1 เป็นบทนำ บทที่ 2 กล่าวถึงทฤษฎีและงานวิจัยที่เกี่ยวข้อง เช่น ขั้นตอนการตรวจจับระดับเสียง การเปรียบเทียบสัญญาณเสียง และการวัดประสิทธิภาพของการค้นคืนข้อมูลในแฟ้มข้อมูลเสียงด้วยข้อคำถามเสียง บทที่ 3 กล่าวถึงการดำเนินงานวิจัย โดยอธิบายเป็นขั้นตอนต่างๆทั้งการตรวจจับระดับเสียงและการเปรียบเทียบสัญญาณเสียง ส่วนในบทที่ 4 เป็นการทดลองและผลที่ได้จากการทดลองตามชุดการทดลองต่างๆ และท้ายสุดคือบทที่ 5 เป็นการสรุปผลการทดลองและข้อเสนอแนะของงานวิจัย ซึ่งอาจจะเป็นประโยชน์ต่องานวิจัยอื่นๆ ต่อไปในอนาคต

บทที่ 2

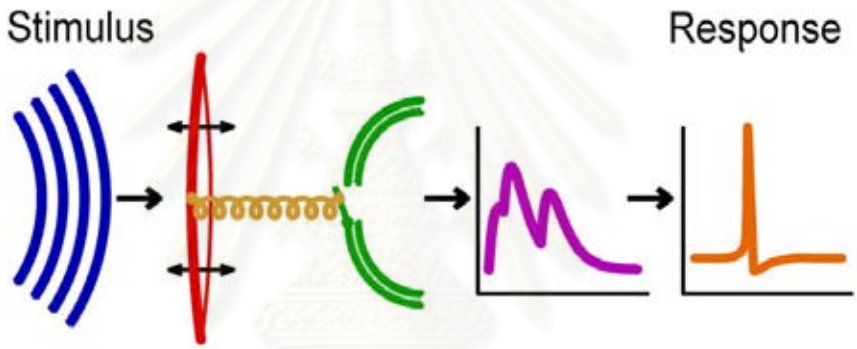
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 ทฤษฎีที่เกี่ยวข้อง

2.1.1 ขั้นตอนวิธีการตรวจจับระดับเสียง (Pitch detection algorithm)

เสียง เกิดจากการสั่นสะเทือนของวัตถุ เมื่อวัตถุสั่นสะเทือน จะทำให้เกิดการอัดตัวและขยายตัวของคลื่นเสียง และถูกส่งผ่านตัวกลาง เช่น อากาศ ไปยังหู แต่เสียงสามารถเดินทางผ่านแก๊ส ของเหลว และของแข็งก็ได้ แต่ไม่สามารถเดินทางผ่านสุญญากาศ เช่น ในอวกาศ ได้

เมื่อการสั่นสะเทือนนั้นมาถึงหูของเรา มันจะถูกแปลงเป็นพัลส์ ซึ่งจะถูกส่งไปยังสมอง ทำให้เรารับรู้และจำแนกเสียงต่างๆ ได้



รูปที่ 2.1 กลไกการได้ยินเสียง (จากซ้ายสุด คือ คลื่นเสียง ถัดมาเป็น กล่องหู หูชั้นในรูปหอยโข่ง (Cochlea) เซลล์รับรู้การได้ยิน สเปกตรัมความถี่ของการตอบสนองการได้ยิน และสุดท้ายคือ อิมพัลส์ประสาท) [1]

เสียงแต่ละเสียงมีความแตกต่างกัน เสียงสูง - เสียงต่ำ เสียงดัง - เสียงเบา หรือคุณภาพของเสียงลักษณะต่างๆ ทั้งนี้ขึ้นอยู่กับแหล่งกำเนิดเสียง และจำนวนรอบต่อวินาทีของการสั่นสะเทือน

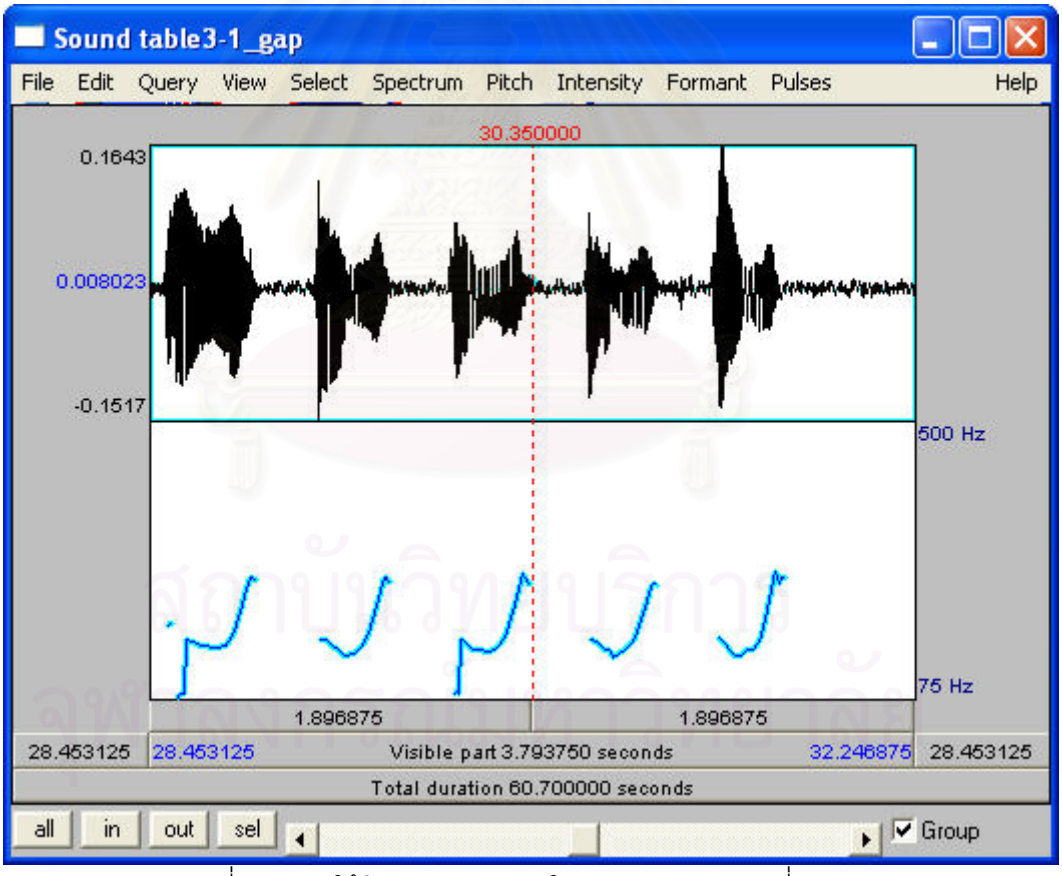
ระดับเสียง (Pitch) [2][3] หมายถึง ระดับความสูง-ต่ำของเสียง ซึ่งเกิดจากค่าความถี่ของการสั่นสะเทือน กล่าวคือ ถ้าเสียงที่มีความถี่สูง ลักษณะการสั่นสะเทือนเร็ว จะส่งผลให้มีระดับเสียงสูง แต่ถ้าหากเสียงมีความถี่ต่ำ ลักษณะการสั่นสะเทือนช้าจะส่งผลให้มีระดับเสียงต่ำ

ขั้นตอนวิธีการตรวจจับระดับเสียง คือ วิธีการประมาณระดับเสียง หรือความถี่มูลฐาน (Fundamental Frequency) ของสัญญาณที่มีลักษณะเป็นคาบหรือลักษณะคล้ายคาบ (Guasiperiodic) โดยส่วนใหญ่สัญญาณเหล่านี้มาจากข้อมูลที่เกี่ยวข้องกับเสียง เช่น เสียงพูดของ

มนุษย์ เสียงดนตรี หรือเสียงธรรมชาติ เป็นต้น ซึ่งขั้นตอนวิธีการตรวจจับระดับเสียงสามารถทำได้ทั้งในโดเมนเวลา (Time Domain) เช่น วิธีสหสัมพันธ์อัตโนมัติ (Autocorrelation Method) และโดเมนความถี่ (Frequency Domain) เช่น การแปลงฟูเรียร์ (Fourier Transformation) [4]

2.1.1.1 **การสกัดค่าความถี่มูลฐาน** (Fundamental Frequency) **ด้วยวิธีสหสัมพันธ์อัตโนมัติ** (Autocorrelation Method) [4]

ในการพูดคุยหรือการสนทนาของมนุษย์เราทุกครั้ง เราจะพบว่ามีความแตกต่างของระดับเสียงชัดเจน โดยเฉพาะในภาษาไทยจะเห็นได้ชัดเจนเพราะเรามีการผันเสียงของวรรณยุกต์ ทำให้เกิดเสียงสูง – ต่ำ และทำให้ความหมายแตกต่างกันออกไป เช่น ปา ป่า ป้า ป๊า ป๋า เป็นต้น จากการที่ภาษาไทยมีการผันวรรณยุกต์ 5 ระดับเสียงต่างกัน ทำให้ค่าความถี่มูลฐานของแต่ละคำต่างกันด้วย ดังนั้นผู้เขียนจึงนำค่าความถี่มูลฐานมาใช้ในการแยกประเภทเสียงวรรณยุกต์ [5] เพื่อใช้ในการคืนคืนข้อมูลเสียง



รูปที่ 2.2 การใช้โปรแกรม Praat ในการสกัดค่าความถี่มูลฐาน

งานวิจัยชิ้นนี้จะสกัดค่าความถี่มูลฐานด้วยด้วยวิธีสหสัมพันธ์อัตโนมัติ (Autocorrelation Method) โดยใช้โปรแกรม Praat [6] ซึ่งเป็นเครื่องมือสำหรับวิเคราะห์ข้อมูลเสียง และสามารถเลือกหน่วยของความถี่ที่ใช้ได้ นอกจากนี้ยังสามารถกำหนดค่าพารามิเตอร์ต่างๆ ที่ใช้ในการ

คำนวณได้ งานวิจัยนี้ผู้เขียนได้กำหนดค่าพารามิเตอร์ดังตารางที่ 2.1 ซึ่งเป็นค่าพารามิเตอร์มาตรฐานและบางตัวเป็นค่าที่ผ่านการทดลองมาแล้ว

ตารางที่ 2.1 ค่าพารามิเตอร์ที่ใช้ในการหาค่าความถี่มูลฐานด้วยโปรแกรม Praat

Parameters	Values
Unit	Hertz
Method	Autocorrelation
Time step	0.01s
Pitch floor	75 Hz
Max. number of candidates	15
Silence threshold	0.03
Voicing threshold	0.45
Octave cost	0.01
Octave-jump cost	0.35
Voiced/unvoiced cost	0.14
Pitch ceiling	600 Hz

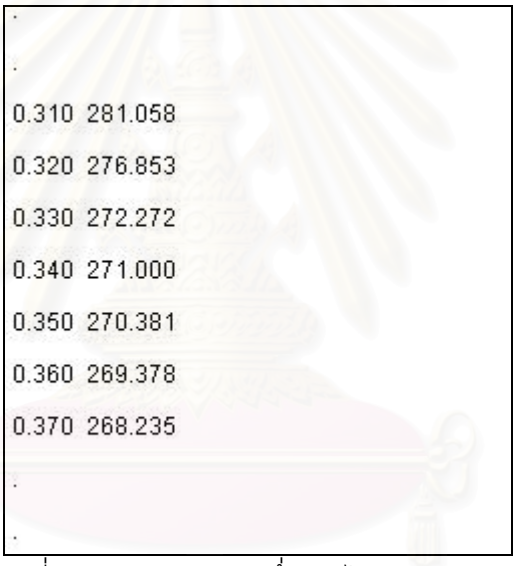
จากตารางที่ 2.1 เป็นตารางค่าพารามิเตอร์ที่ใช้ในการหาค่าความถี่มูลฐานด้วยโปรแกรม Praat เช่น

- Unit หมายถึง หน่วยของค่าความถี่มูลฐาน มีหน่วยเป็น เฮิรตซ์ (Hertz)
- Method หมายถึง วิธีที่ใช้ในการสกัดค่าความถี่มูลฐาน ในงานวิจัยนี้ใช้วิธีสหสัมพันธ์อัตโนมัติ (Autocorrelation Method)
- Time step หมายถึง ช่วงเวลาที่ใช้ในการสกัดค่าความถี่มูลฐาน มีหน่วยเป็นวินาที ค่ามาตรฐาน เท่ากับ 0.0 วินาที ตัวอย่างเช่น ถ้า Time step เท่ากับ 0.01 วินาที และ Pitch floor เท่ากับ 75 เฮิรตซ์ หมายความว่าโปรแกรม Praat จะคำนวณค่าพิทช์ 100 ค่า ใน 1 วินาที ในงานวิจัยนี้ใช้ค่า Time step เท่ากับ 0.01 วินาที
- Pitch floor หมายถึง ขอบเขตล่างที่กำหนดว่าค่าความถี่ที่ต่ำกว่าค่า Pitch floor จะไม่ถูกนำมาคำนวณค่าพิทช์ ค่ามาตรฐาน เท่ากับ 75 Hz

- Pitch ceiling หมายถึง ขอบเขตบนที่กำหนดว่าค่าความถี่ที่สูงกว่าค่า Pitch ceiling จะไม่ถูกนำมาคำนวณค่าพิทช์ ค่ามาตรฐานเท่ากับ 600 Hz

นอกจากพารามิเตอร์ที่กล่าวข้างต้นแล้ว ยังมีพารามิเตอร์อื่นๆ อีก เช่น Max. number of candidates, Silence Threshold, Voicing threshold เป็นต้น ซึ่งในงานวิจัยนี้จะใช้ค่ามาตรฐานของโปรแกรมดังตารางที่ 2.1

การสกัดค่าความถี่มูลฐานของแฟ้มข้อมูลเสียงขนาดใหญ่นั้น จะใช้การเขียนชุดคำสั่ง (Script) เพื่อควบคุมการทำงานของโปรแกรม Praat เมื่อ Praat ทำงานตามชุดคำสั่งนั้น จะได้ไฟล์ชื่อเดียวกับที่เขียนในชุดคำสั่ง ซึ่งในงานวิจัยนี้ คือไฟล์ output.txt ซึ่งเป็นไฟล์ข้อมูลที่ภายในเก็บคู่ลำดับของเวลา (วินาที) และความถี่มูลฐาน (Hz) ดังรูปที่ 2.3



รูปที่ 2.3 ตัวอย่างส่วนหนึ่งของไฟล์ output.txt

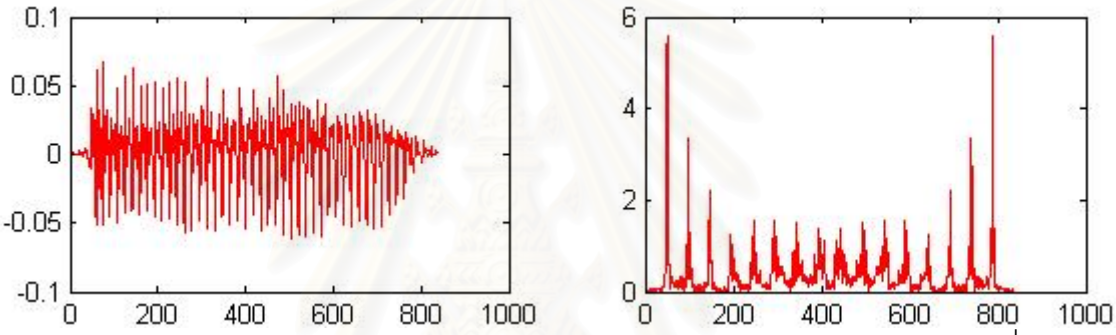
จากรูปที่ 2.3 สามารถอธิบายได้ว่า ณ ช่วงเวลา [0.310 – 0.320) วินาที มีค่าความถี่มูลฐาน เท่ากับ 281.058 เฮิรตซ์ และช่วงเวลา [0.320 – 0.330) วินาที มีค่าความถี่มูลฐาน เท่ากับ 276.583 เฮิรตซ์ เป็นต้น

2.1.1.2 การแปลงฟูเรียร์อย่างเร็ว (Fast Fourier Transform: FFT) [7][8][9]

เสียงเมื่อเดินทางผ่านอากาศมาที่อวัยวะรับการได้ยินของมนุษย์ คือ หู ซึ่งเป็นระบบเปิด ที่มีความไวต่อความดันอากาศมาก มนุษย์จะรู้สึกได้ยินก็ต่อเมื่อ มีการนำสัญญาณนั้นผ่านกระบวนได้ยินในหูชั้นต่างๆ เราสามารถอธิบายเสียงที่ได้ยินในเชิงสมการคณิตศาสตร์ได้สองลักษณะหลัก คือ ฟังก์ชันทางเวลา (Time Domain) และฟังก์ชันทางความถี่ (Frequency Domain)

ฟังก์ชันทางเวลา สามารถอธิบายในลักษณะความดังของเสียง ในช่วงเวลาที่เปลี่ยนแปลงไปเท่านั้น แต่ไม่สามารถบอกได้ในเชิงความถี่ว่ามีความถี่ต่ำหรือสูง ในขณะที่ฟังก์ชันทางความถี่สามารถอธิบายลักษณะของเสียงว่ามีลักษณะสูงหรือต่ำได้ ซึ่งเป็นข้อมูลที่สำคัญในการเข้าถึงธรรมชาติเสียงนั้น

โดยธรรมชาติ หูของมนุษย์จะทำการแปลงสัญญาณเสียงให้อยู่ในรูปของแถบความถี่เสียง นั่นคือ แอมพลิจูด (Amplitude) และความถี่ โดยการแปลงแบบนี้สามารถทำได้ โดยอาศัยหลักการทางคณิตศาสตร์ที่เรียกว่าการแปลงแบบฟูเรียร์ (Fourier Transform) ก็คือการแปลงสัญญาณในโดเมนเวลาให้อยู่ในโดเมนของความถี่นั่นเอง และเสียงที่บันทึกทั่วไปทั่วไปนั้นจะอยู่ในรูปของโดเมนเวเลา (Time Domain) โดยตัวอย่างของการแปลงฟูเรียร์เป็นไปตามรูปที่ 2.4



คำว่า “ตาม” ในโดเมนเวลา คำว่า “ตาม” ในโดเมนความถี่

รูปที่ 2.4 ตัวอย่างการแปลงสัญญาณเสียงจาก โดเมนเวลา (Time Domain) เป็น โดเมนความถี่ (Frequency Domain) ด้วยการแปลงแบบฟูเรียร์อย่างเร็ว

จากรูปที่ 2.4 กราฟด้านซ้ายเป็นกราฟสัญญาณเสียงในโดเมนเวลา ซึ่งเป็นสัญญาณเสียงทั่วไป ในขณะที่กราฟทางด้านขวาคือกราฟที่ได้จากการแปลงสัญญาณเสียงจากโดเมนเวลา เป็นโดเมนความถี่ด้วยการแปลงฟูเรียร์อย่างเร็ว โดยการพลอตกราฟดังกล่าวทำการใส่ค่าสัมบูรณ์ (Absolute) ของค่าที่ได้จากการแปลงฟูเรียร์อย่างเร็ว

การแปลงแบบฟูเรียร์นั้นประกอบด้วย

- การแปลงฟูเรียร์แบบต่อเนื่อง (Fourier Transform) ใช้วิธีการคำนวณโดยการอินทิเกรต
- การแปลงฟูเรียร์แบบไม่ต่อเนื่อง (Discrete Fourier Transform) ใช้วิธีการหาผลบวกแทนการอินทิเกรต

การแปลงฟูเรียร์อย่างเร็ว (Fast Fourier Transform หรือ FFT) นั้นเป็นขั้นตอนวิธีหนึ่งในการทำการแปลงฟูเรียร์แบบไม่ต่อเนื่องที่ใช้เวลาน้อยซึ่งถูกคิดค้นขึ้นโดย James W. Cooley และ John W. Turkey เมื่อปี ค.ศ. 1965 โดยมีขั้นตอนวิธีดังนี้ [9]

เราจะให้

$$W_N = e^{-i2\pi/N} \tag{2.1}$$

ดังนั้นเราจะเขียน สมการการแปลงฟูเรียร์แบบไม่ต่อเนื่อง ($F(s)$) ได้ดังนี้

$$F(s) = \frac{1}{N}\sum_{x=0}^{N-1} f(x)W_{N^{sx}} \tag{2.2}$$

หาก N เป็นเลขคู่ ซึ่ง N คือ จำนวนชุดข้อมูล

$$N = 2M \quad \text{สำหรับ } M \text{ ใดๆ} \tag{2.3}$$

เมื่อแทน N ด้วย $2M$ แล้ว

$$F(s) = \frac{1}{2M}\sum_{x=0}^{2M-1} f(x)W_{2M^{sx}} \tag{2.4}$$

แยกพจน์ที่ M เป็นจำนวนคู่ และจำนวนคี่ แล้วได้ดังนี้

$$F(s) = \frac{1}{2}\left\{\frac{1}{M}\sum_{x=0}^{M-1} f(2x)W_{2M^{s(2x)}} + \frac{1}{M}\sum_{x=0}^{M-1} f(2x+1)W_{2M^{s(2x+1)}}\right\} \tag{2.5}$$

เนื่องจาก $W_{2M}^{2S} = W_M^s$ และ $W_{2M}^{2s+1} = W_M^S W_{2M}^s$ จะได้ว่า

$$F(s) = \frac{1}{2}\left\{\frac{1}{M}\sum_{x=0}^{M-1} f(2x)W_{M^{sx}} + \frac{1}{M}\sum_{x=0}^{M-1} f(2x+1)W_{M^{sx}}W_{2M}^s\right\} \tag{2.6}$$

โดยสมการข้างบนคือสมการแปลงฟูเรียร์ของพจน์ที่เป็นจำนวนคู่ (จะแทนด้วย $F_{even}(s)$) บวกกับค่าคงที่ W_{2M}^s คูณด้วยสมการการแปลงฟูเรียร์ของพจน์ที่เป็นจำนวนคี่ (จะแทนด้วย $F_{odd}(s)$)

นั่นคือ M พจน์แรกของการแปลงฟูเรียร์จะสามารถคำนวณได้ดังนี้

$$F(s) = \frac{1}{2}\left\{F_{even}(s) + F_{odd}(s)W_{2M}^s\right\} \tag{2.7}$$

ทำนองเดียวกัน M พจน์หลังจะสามารถคำนวณได้ดังนี้

$$F(s) = \frac{1}{2}\left\{F_{even}(s) - F_{odd}(s)W_{2M}^s\right\} \tag{2.8}$$

หมายความว่าการแปลงข้อมูล N จุดจะสามารถทำได้โดยแยกเป็นพจน์ที่เป็นจำนวนคู่และจำนวนคี่ โดยคำนวณ $N/2$ พจน์แล้วรวมผลของสมการที่ 2.7 และ 2.8 เข้าด้วยกัน

หาก N เป็นเลขยกกำลังของ 2 ก็ทำซ้ำไปเรื่อยๆ จนได้ถึงค่าฐาน ทำการรวมไปเรื่อยๆ แบบ Recursive สุดท้ายจะได้ผลลัพธ์ของการแปลง ขั้นตอนวิธีแบ่งต่อสู้และเอาชนะ (Divide and Conquer) ที่กล่าวมานี้ใช้เวลา $O(N \log N)$ ซึ่งน้อยกว่าขั้นตอนวิธีแบบธรรมดา ที่ใช้เวลา $O(N^2)$

โดยในขั้นตอนวิธีนี้ สามารถใช้ได้ในกรณีที่ N เป็นเลขยกกำลังของ 2 เท่านั้น

เนื่องจากงานวิจัยนี้มีขอบเขตของการค้นคืนข้อมูลเสียงในลักษณะที่ผู้พูดข้อคำถามเสียงและผู้บันทึกเสียงเป็นคนละคนกัน (Speaker Independent) ได้ แต่จากผลการทดลองที่อยู่ในส่วนของภาคผนวก ก ปรากฏว่าการแปลงฟูเรียร์อย่างเร็วให้ค่าความแม่นยำค่อนข้างต่ำในการค้นคืนข้อมูลเสียงในกรณีที่ผู้พูดข้อคำถามเสียงกับผู้บันทึกเสียงเป็นคนละคนกัน ดังนั้นผู้เขียนจึงเลือกใช้ค่าความถี่มูลฐานในการทดลองการค้นคืนข้อมูลแทนการแปลงฟูเรียร์อย่างเร็ว

2.1.2 **การแปลงข้อมูลให้เป็นบรรทัดฐาน** (Data Normalization)

การแปลงข้อมูลให้เป็นบรรทัดฐาน เป็นการปรับค่าของข้อมูลให้มีขอบเขตอยู่ในช่วงเล็กลง เช่น อยู่ในช่วง -1.0 ถึง 1.0 หรือ ช่วง 0.0 ถึง 1.0 ซึ่งวิธีที่นิยมใช้กันอย่างแพร่หลายเช่น การแปลงตามค่าต่ำสุด-สูงสุด (Min-Max Normalization) การแปลงตามค่าคะแนนมาตรฐานซี (z-score Normalization) และการปรับมาตราทศนิยม (Decimal Scaling) เป็นต้น แต่ในงานวิจัยนี้ผู้เขียนเลือกใช้การแปลงข้อมูลให้เป็นบรรทัดฐานด้วยค่าคะแนนมาตรฐานซี (Z-Score Normalization)

2.1.2.1 **การแปลงตามค่าคะแนนมาตรฐานซี** (Z-Score Normalization) [10]

เป็นการแปลงค่าข้อมูลโดยใช้ค่าเฉลี่ย (Mean) เท่ากับ 0 และ ค่าเบี่ยงเบนมาตรฐาน (Standard Deviation) เท่ากับ 1 ดังสมการ

$$v' = \frac{v - mean(A)}{SD(A)}$$

โดยที่ v คือ ค่าคุณลักษณะเดิม

v' คือ ค่าคุณลักษณะใหม่

$mean(A)$ คือ ค่าเฉลี่ยของคุณลักษณะ A

$SD(A)$ คือ ค่าเบี่ยงเบนมาตรฐานของคุณลักษณะ A

2.1.3 การเปรียบเทียบสัญญาณเสียง

การเปรียบเทียบสัญญาณเสียงระหว่างข้อคำถามเสียงกับข้อมูลเสียงที่อยู่ภายในแฟ้มข้อมูลเสียงมีอยู่หลายวิธี แต่ในงานวิจัยนี้ ผู้เขียนเลือกวิธีการเปรียบเทียบสัญญาณเสียงโดยใช้วิธีวัดระยะทางแบบยูคลิด (Euclidean Distance) และวิธีวัดระยะทางแบบไดนามิกไทม์วอร์ปปิง (Dynamic Time Warping) ในการเปรียบเทียบสัญญาณเสียง เพื่อศึกษาว่าวิธีใดสามารถค้นคืนข้อมูลเสียงภาษาไทยด้วยข้อคำถามเสียงได้ดีที่สุด

วิธีวัดระยะทางแบบยูคลิด วิธีวัดระยะทางแบบไดนามิกไทม์วอร์ปปิง

รูปที่ 2.5 เปรียบเทียบสัญญาณเสียงระหว่างวิธีวัดระยะทางแบบยูคลิดและวิธีวัดระยะทางแบบไดนามิกไทม์วอร์ปปิง (ที่มา : S. Chu, E. Keogh, D. Hart, and M. Pazzani)[11]

2.1.3.1 การหาระยะห่างแบบยูคลิด (Euclidean distance) [12]

ฟังก์ชันระยะห่างที่ใช้วัดความคล้ายกันระหว่างวัตถุ 2 ชิ้น มีอยู่หลายวิธี แต่ในที่นี้ผู้เขียนเลือก Minkowski distance ซึ่งเป็นวิธีที่นิยมใช้กันอย่างแพร่หลาย ดังสมการนี้

$$d(i,j) = \sqrt[q]{(|x_{i1} - x_{j1}|^q + |x_{i2} - x_{j2}|^q + ... + |x_{ip} - x_{jp}|^q)}$$

โดยที่ $i = (x_{i1}, x_{i2}, \ldots x_{ip})$ และ $j = (x_{j1}, x_{j2}, \ldots, x_{jp})$ เป็นวัตถุที่มี p มิติ (คุณลักษณะ p ตัว)

q คือ เลขจำนวนเต็มค่าบวก

ถ้า $q = 1$ เราเรียก d ว่า การคำนวณระยะห่างแบบแมนฮัตตัน (Manhattan distance)

$$d(i,j) = |x_{i1} - x_{j1}| + |x_{i2} - x_{j2}| + ... + |x_{ip} - x_{jp}|$$

ถ้า $q = 2$ เราเรียก d ว่า การคำนวณระยะห่างแบบยูคลิด (Euclidean distance)

$$d(i,j) = \sqrt{|x_{i1} - x_{j1}|^2 + |x_{i2} - x_{j2}|^2 + ... + |x_{ip} - x_{jp}|^2}$$

เราสามารถให้ค่าน้ำหนักตามความสำคัญแต่ละตัวแปร ในการใช้สูตรระยะห่างข้างต้น

ในงานวิจัยนี้เลือกใช้ $q = 2$ ซึ่งก็คือ การคำนวณระยะห่างแบบยูคลิด ในการเปรียบเทียบสัญญาณเสียงระหว่างข้อคำถามเสียงกับข้อมูลเสียงในฐานข้อมูล ในการทำการทดลอง

2.1.3.2 วิธีวัดระยะทางแบบไดนามิกไทม์วอร์ปปิง (Dynamic Time Warping: DTW) [13][4]

ในงานวิจัยนี้ใช้วิธีวัดระยะทางแบบไดนามิกไทม์วอร์ปปิง (Dynamic Time Warping) ในการเปรียบเทียบสัญญาณเสียงระหว่างข้อคำถามเสียงกับข้อมูลในแฟ้มข้อมูลเสียง เพื่อหาระยะห่างระหว่างสัญญาณเสียง ซึ่งไดนามิกไทม์วอร์ปปิง (Dynamic Time Warping) เป็นขั้นตอนวิธีที่ใช้หาระยะทาง (Distance Measure) ระหว่างข้อมูล 2 ชุด ซึ่งอนุญาตให้หาระยะทางแบบไม่เป็นเชิงเส้นได้ (Non-linear alignment) วิธีนี้จึงนิยมใช้ในการวัดความคล้ายคลึง (Similarity Measurement) ของข้อมูลประเภทอนุกรมเวลา (Time Series) ที่อาจมีความเร็วต่างกัน และได้มีการนำไปประยุกต์ใช้ในการพัฒนาระบบรู้จำเสียง รวมทั้งการค้นหาเพลงโดยการร้องทำนอง (Query by Humming) อย่างแพร่หลาย เนื่องจากการทำงานของระบบทั้งสองเกี่ยวข้องกับการเปรียบเทียบข้อมูลเสียงที่มาจากการพูดหรือการร้องโดยมนุษย์ ซึ่งอาจมีความเร็วในการพูดหรือการร้องที่แตกต่างกันไปได้

นิยามของไดนามิกไทม์วอร์ปปิง

ให้ข้อมูล $Q = q_1, q_2, \ldots, q_n$ และ $C = c_1, c_2, \ldots, c_m$ โดยจะสามารถนิยามระยะทางไทม์วอร์ปปิงเป็นสมการเวียนเกิดได้ดังต่อไปนี้

$$DTW(Q,C) = \gamma(n,m)$$

$$\gamma(i,j) = D(q_i,c_j) + \min\{\gamma(i-1,j-1), \gamma(i-1,j), \gamma(i,j-1)\}$$

$$D(Q_i,C_j) = (C_i - Q_j)^2$$

โดยที่ γ คือ ระยะทางสะสม

D คือ ฟังก์ชันหาระยะทางระหว่างจุดสองจุด และ

$1 \le i \le n$ และ $1 \le j \le m$

ไดนามิกไทม์วอร์ปปิงใช้กลวิธีกำหนดการพลวัต (Dynamic Programming) ในการคำนวณ ซึ่งใช้เวลาในการทำงานเป็นฟังก์ชันพหุนามตามขนาดข้อมูลขาเข้า (Polynomial Time) ซึ่งยังจัดว่ายังมีความเร็วไม่มากนัก โดยเฉพาะเมื่อเทียบกับวิธีวัดระยะทางแบบยูคลิด (Euclidean

Distance) ซึ่งใช้เวลาการทำงานเป็นแบบเชิงเส้น (Linear Time) ทำให้มีผู้เสนอกลวิธีที่จะช่วยปรับปรุงให้มีความเร็วเพิ่มขึ้นอยู่หลายกลวิธีด้วยกัน

เงื่อนไขบังคับโดยรวม (Global Constraint)

เงื่อนไขบังคับโดยรวมก็ถือเป็นหนึ่งในวิธีปรับปรุงประสิทธิภาพของไดนามิกไทม์วอร์ปปิง โดยวิธีนี้เป็นการบังคับให้กระบวนการไดนามิกไทม์วอร์ปปิงทำงานในขอบเขตที่จำกัดไว้ ซึ่งทำให้ไม่สามารถทำงานเกินขอบเขตที่กำหนดได้ เพื่อป้องกันการเลือกจุดที่ห่างกันมากเกินไปมาพิจารณาหาระยะทาง ซึ่งอาจส่งผลโดยตรงต่อความแม่นยำของกระบวนการนี้

ในงานวิจัยนี้เลือกใช้แถบของซาโก-ชิบะ (Sakoe-chiba Band) [13] ซึ่งมีลักษณะเป็นเส้นขนานสองเส้นบนตารางการคำนวณไดนามิกไทม์วอร์ปปิง โดยหมายความว่าในทุกๆ จุดของชุดข้อมูลที่นำมาพิจารณาหาค่าระยะทาง อนุญาตให้พิจารณาเลือกจุดข้อมูลที่จะนำมาทำการคำนวณระยะทางได้ภายในขอบเขตที่กำหนดเท่านั้น ซึ่งมีขอบเขตเท่ากันทั้งหมดทั้งชุดข้อมูล โดยที่งานวิจัยนี้ใช้ค่าเงื่อนไขบังคับโดยรวมเท่ากับ 3 เปอร์เซ็นต์ เนื่องจากทำการทดลองแล้วให้ค่าความแม่นยำสูงสุด

นิยามของไดนามิกไทม์วอร์ปปิงที่ใช้เงื่อนไขบังคับโดยรวมแบบแถบของซาโก-ชิบะ

ให้ข้อมูล $Q = Q_1, Q_2, \ldots, Q_n$ และ $C = C_1, C_2, \ldots, C_m$ จะสามารถนิยามระยะทางไทม์วอร์ปปิงที่ใช้เงื่อนไขบังคับโดยรวมแบบแถบของซาโก-ชิบะ เป็นสมการเวียนเกิดได้ดังต่อไปนี้

$$cDTW(Q,C,r) = \gamma_r(n,m)$$

$$\gamma_r(i,j) = Dist_r(Q_i, C_j) + \min\{\gamma_r(i,j-1), \gamma_r(i-1,j-1), \gamma_r(i-1,j)\}$$

$$Dist_r(Q,C) = \begin{cases} D(Q_i, C_j) & \text{เมื่อ } |i-j| < r \\ \infty & \text{เมื่อ } |i-j| \geq r \end{cases}$$

โดยที่ γ คือ ระยะทางสะสม

D คือ ฟังก์ชันหาระยะทางระหว่างจุดสองจุด

r คือค่ากำหนดเงื่อนไขบังคับของแถบซาโก-ชิบะ

$1 \leq i \leq n$ และ $1 \leq j \leq m$

จากการเปรียบเทียบความถูกต้องระหว่างวิธีวัดระยะทางแบบไดนามิกไทม์วอร์ปปิงและวิธีวัดระยะทางแบบยูคลิด ดังตารางที่ ก-3 ในส่วนของภาคผนวก ก สามารถสรุปได้ว่าวิธีวัดระยะทางแบบไดนามิกไทม์วอร์ปปิงสามารถค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงภาษาไทยด้วยข้อ

คำถามเสียงได้ถูกต้องกว่าวิธีวัดระยะทางงแบบยูคลิด ดังนั้นผู้เขียนจึงเลือกใช้วิธีวัดระยะทางแบบไดนามิกไทม์วอร์ปปิงในการค้นคืนข้อมูล

2.2 งานวิจัยที่เกี่ยวข้อง

จากการค้นคว้างานวิจัยที่เกี่ยวข้องกับเสียงวรรณยุกต์ในการรู้จำเสียงพูดนั้น ผู้เขียนได้พบงานวิจัยต่างๆ ที่เกี่ยวข้องกับการรู้จำเสียงวรรณยุกต์ ซึ่งมีรายละเอียดดังต่อไปนี้

A. Tungthangthum [15] ทำการศึกษาเกี่ยวกับการรู้จำเสียงวรรณยุกต์ภาษาไทยด้วยแบบจำลองฮิดเดนมาร์คอฟ โดยทำการบันทึกเสียงผู้พูดชายคนเดียว ที่ความถี่การชักตัวอย่าง (Sampling Frequency) 8 kHz งานวิจัยนี้ได้ทำการเลือกเสียงสระมา 10 เสียงสระ เพื่อทำการบันทึกเป็นฐานข้อมูลเสียง ดังตารางที่ 2.2 และทำการแบ่งฐานข้อมูลเสียง ออกเป็น 3 ชุด คือ ชุดข้อมูลสอน (Training Set) ชุดข้อมูลทดสอบชุดที่ 1 (Test Set 1) และชุดข้อมูลทดสอบชุดที่ 2 (Test Set 2) ดังตารางที่ 2.3

ตารางที่ 2.2 แสดงเสียงสระในภาษาไทยที่ใช้ในการทดลอง

Vowel	In Thai
v1	อา
v2	อี
v3	อือ
v4	อู
v5	เอ
v6	แอ
v7	โอ
v8	ออ
v9	เออ
v10	เอีย

ตารางที่ 2.3 แสดงชุดข้อมูลที่ใช้ในการทดลอง

Speech Data	Vowel	Number of Syllables
Training set	v1 to v5	200
Test set 1	v1 to v5	100
Test set 2	v5 to v10	100

งานวิจัยนี้ได้แบ่งวิธีการที่ใช้ในการรู้จำเสียงวรรณยุกต์ภาษาไทยออกเป็น 2 ขั้นตอน คือ ขั้นตอนแรกทำการสกัดค่าความถี่มูลฐาน (Pitch Frequency Detection) ด้วยวิธีสหสัมพันธ์อัตโนมัติ (Autocorrelation) หลักจากนั้นใช้ HMM (Hidden Markov Model) ในการรู้จำเสียงวรรณยุกต์ภาษาไทย ซึ่งพบว่า แม้เสียงวรรณยุกต์จะซ้อนอยู่บนเสียงสระ แต่ข้อมูลทั้งสองไม่ขึ้นแก่กัน นอกจากนี้ยังพบว่าการรู้จำเสียงวรรณยุกต์ตรี มักให้ผลลัพธ์ผิดพลาดเป็นเสียงสามัญ เนื่องจากรูปร่างคอนทัวร์ของความถี่มูลฐานของทั้งสองเสียงมีความใกล้เคียงกัน

จากงานวิจัยนี้ทำให้ทราบว่า เราสามารถใช้ค่าความถี่มูลฐานในการรู้จำเสียงวรรณยุกต์ภาษาไทยได้

A. W. Fu และคณะ [13] ได้ศึกษาเกี่ยวกับวิธีการทำไดนามิกไทม์วอร์ปปิง (Dynamic Time Warping หรือ DTW) ร่วมกับการทำยูนิฟอร์มสเกลลิง (Uniform Scaling หรือ US) เรียกว่า สเกลลิงและไทม์วอร์ปปิง (Scaling and Time Warping) โดยการทำวิธีการทั้งสองไปด้วยกันนั้นจำเป็นในการจัดการปัญหาบางประเภทเพื่อให้ได้ผลลัพธ์ที่ดีและถูกต้อง เช่น ปัญหาในด้านชีวมาตร (Biometrics) การรู้จำลายมือ (Handwriting Recognition) หรือแม้กระทั่งการค้นหาเพลงโดยการร้องทำนอง (Query by Humming)

การค้นหาเพลง โดยการร้องทำนองนั้นจำเป็นต้องรองรับความผิดพลาดจากผู้ใช้ทั้งความเร็วของเสียงร้องทำนองไม่ตรงกับเพลงต้นฉบับ และผู้ใช้อาจมีการร้องโน้ตเพลงขาดหายหรือเกินได้ ซึ่งปัญหาอย่างแรกนั้นสามารถใช้ยูนิฟอร์มสเกลลิงช่วยได้ ส่วนปัญหาอย่างหลังสามารถใช้ไดนามิกไทม์วอร์ปปิงช่วยได้ ดังนั้นวิธีการทั้งสองจึงจำเป็นในการค้นหาเพลงโดยการร้องทำนอง การใช้วิธีใดวิธีหนึ่งอาจทำให้การค้นหาเพลงโดยการร้องทำนองมีความแม่นยำและมีประสิทธิภาพที่ไม่ดีเท่าที่ควร

สำหรับสิ่งที่น่าสนใจในงานวิจัยนี้ คือ การใช้ยูนิฟอร์มสเกลลิงมาช่วยในการแก้ปัญหาความผิดพลาดในการร้องทำนองความเร็วของเสียงร้องไม่ตรงกับเพลงต้นฉบับ ซึ่งผู้เขียนคิดว่าน่าจะนำมาใช้ในการแก้ปัญหาความผิดพลาดกับการค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงภาษาไทยด้วยข้อคำถามเสียงได้ ในเรื่องของความเร็วของการพูดข้อคำถามเสียง

Y. Zhu และ D. Shasha [16] ได้ศึกษาถึงวิธีการที่เพิ่มความเร็วและความแม่นยำในการค้นหาเพลงโดยการร้องทำนองจากวิธีดั้งเดิม โดยจัดเก็บเพลงและแปลงเสียงร้องทำนองให้อยู่ในรูปแบบของอนุกรมเวลา (Time Series) ซึ่งวิธีนี้จะไม่ทำให้เกิดปัญหาจากความผิดพลาดในการแบ่งตัวโน้ต ระบบที่ได้จึงมีความแม่นยำสูงและลดข้อจำกัดของผู้ใช้ที่จะต้องร้องทำนองเพียงแค่เสียง ทา หรือ ดา

นอกจากนี้เมื่อจัดเก็บเพลงในรูปแบบฐานข้อมูลอนุกรมเวลา (Time Series Database) จะทำให้สามารถนำกลวิธีการทำดรรชนีของฐานข้อมูลอนุกรมเวลามาปรับปรุงและใช้ประโยชน์ได้ โดยใช้วิธีการของไดนามิกไทม์วอร์ปปิง เพื่อให้ระบบที่ได้สามารถรองรับความผิดพลาดของผู้ใช้งานในด้านความเร็วของการร้องเพลงที่ไม่คงที่หรือไม่ตรงตามเพลงต้นฉบับ รวมทั้งทำให้ระบบมีความเร็วในการค้นหาเพิ่มมากขึ้นอีกด้วย

ผลการทดสอบระบบในเบื้องต้นชี้ให้เห็นว่าวิธีการนี้มีความเร็วมากกว่าวิธีการแบบดั้งเดิม และมีความถูกต้องอยู่ในระดับที่ดี

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

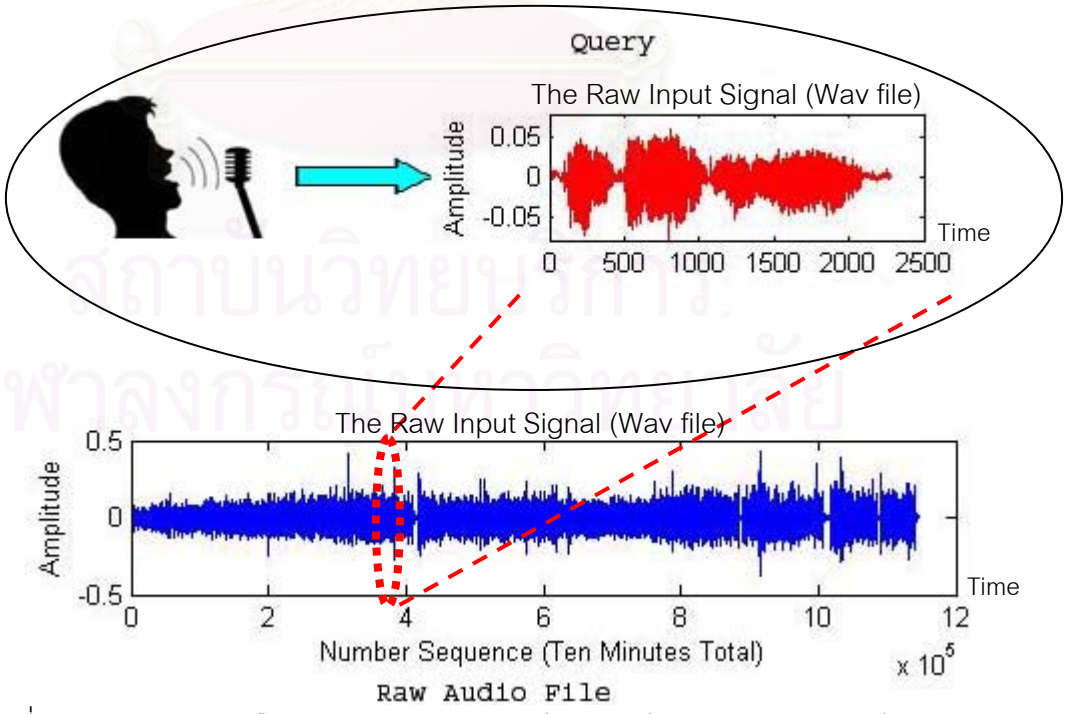
บทที่ 3

ขั้นตอนการดำเนินงานวิจัย

การเพิ่มความเร็วและประสิทธิภาพในการค้นคืนข้อมูลเสียงจากแฟ้มข้อมูลเสียงภาษาไทยด้วยข้อคำถามเสียง ในงานวิจัยนี้มีขั้นตอนการดำเนินงานดังต่อไปนี้

1. บันทึกแฟ้มข้อมูลเสียงภาษาไทยและข้อคำถามเสียง ด้วยไมโครโฟนผ่านโปรแกรมบันทึกเสียง
2. สกัดค่าความถี่มูลฐาน ด้วยโปรแกรม Praat
3. ทำการปรับเรียบ (Smoothing)
4. การปรับข้อมูลให้เป็นบรรทัดฐานด้วยค่าคะแนนมาตรฐานซี
5. ใช้วิธีวัดระยะทางแบบไดนามิกไทม์วอร์ปปิง (Dynamic Time Warping หรือ DTW) ในการเปรียบเทียบสัญญาณเสียงระหว่างข้อคำถามเสียงกับข้อมูลที่อยู่ในแฟ้มข้อมูลเสียงภาษาไทย
6. การแสดงผลลัพธ์ที่ได้จากการค้นคืนใช้วิธีการจำแนกประเภทแบบ *K* ลำดับที่ใกล้ที่สุด (*K*-Nearest Neighbor)

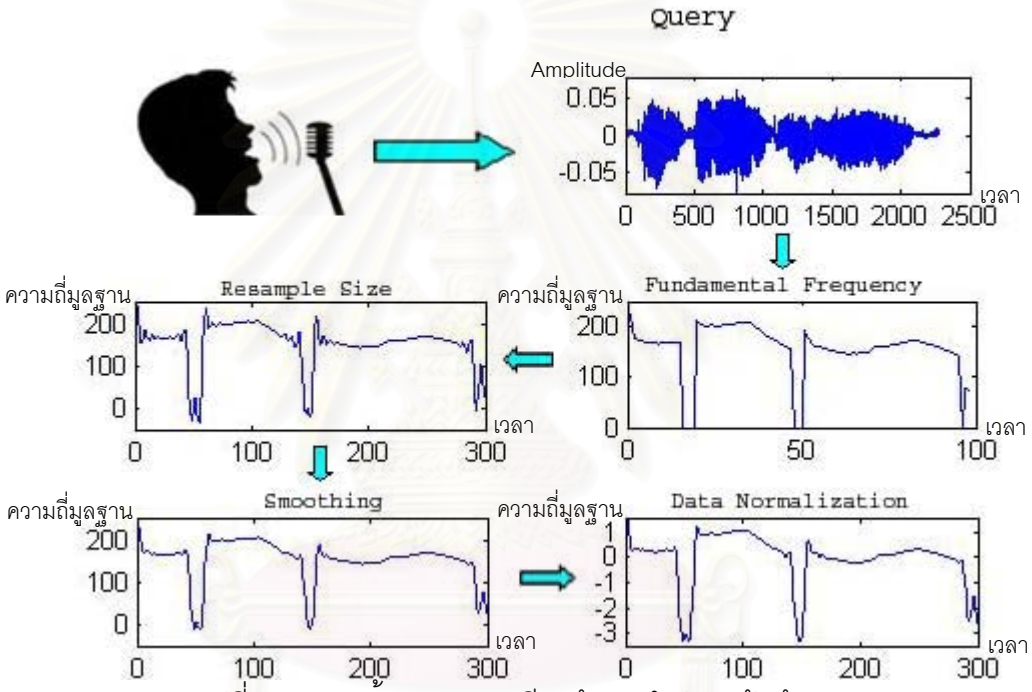
ขั้นตอนการดำเนินงานของงานวิจัยนี้ ซึ่งในที่นี้คือการค้นคืนข้อมูลเสียงจากแฟ้มข้อมูลเสียงภาษาไทยด้วยข้อคำถามเสียง สามารถแสดงดังรูปที่ 3.1 ต่อไปนี้



รูปที่ 3.1 การทำงานค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงภาษาไทยด้วยข้อคำถามเสียง

3.1 การเตรียมข้อมูล

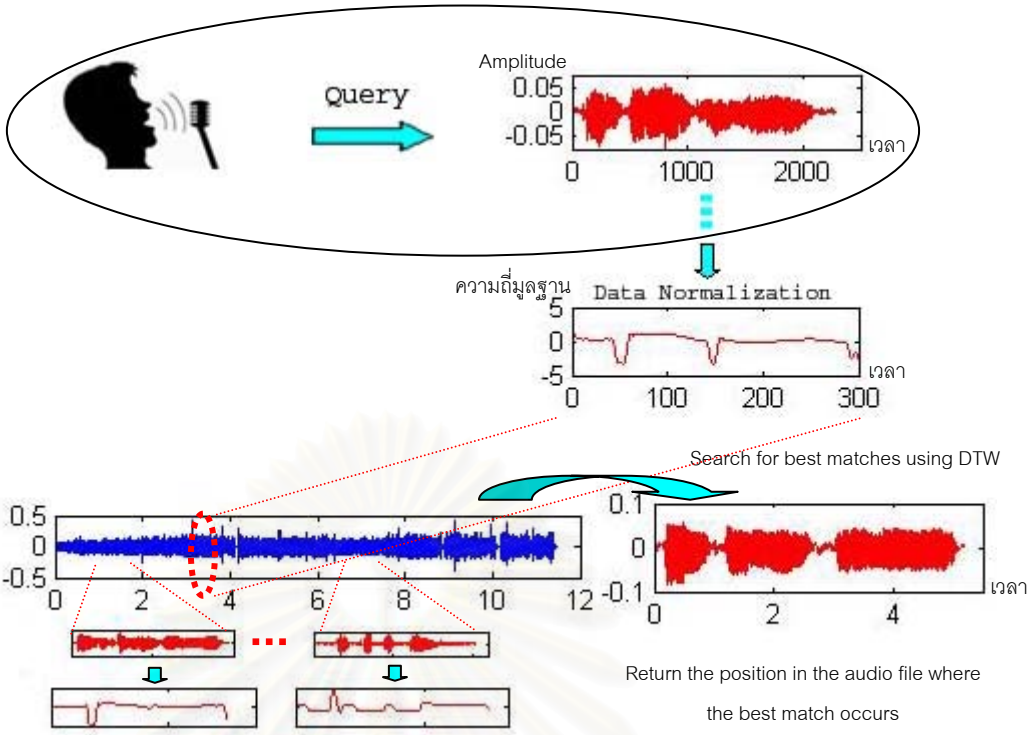
การเก็บข้อมูลเริ่มโดยการบันทึกเสียงจากไมโครโฟน (Microphone) ผ่านโปรแกรมบันทึกเสียง และจัดเก็บในรูปแบบของไฟล์ WAV ที่อัตราการชักตัวอย่าง (Sampling Rate) 2000 เฮิรตซ์ และเก็บข้อมูลขนาด 16 บิต แบบช่องสัญญาณเดียว (Mono) สำหรับขั้นตอนการเตรียมข้อมูลที่ได้จากการบันทึกนั้น จะนำแฟ้มข้อมูลเสียงที่ทำการบันทึกมาสกัดค่าความถี่มูลฐาน (Fundamental Frequency) และนำไปผ่านกระบวนการปรับขนาดตัวอย่าง (Resampling) หลังจากนั้นทำการปรับเรียบ (Smoothing) และการแปลงข้อมูลให้เป็นบรรทัดฐาน (Data Normalization) แสดงดังรูปที่ 3.2



รูปที่ 3.2 แสดงขั้นตอนการเตรียมข้อมูลคำว่า “แฟ้มข้อมูลภายนอก”

3.2 การค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงภาษาไทย

การค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงภาษาไทยนั้น เป็นการเปรียบเทียบสัญญาณเสียงระหว่างข้อคำถามเสียงกับข้อมูลเสียงในแฟ้มข้อมูลเสียงภาษาไทย ซึ่งในงานวิจัยนี้ใช้วิธีวัดระยะทางแบบไดนามิกไทม์วอร์ปปิง (Dynamic Time Warping หรือ DTW) ในส่วนของการแสดงผลลัพธ์ที่ได้จากการค้นคืนใช้วิธีการจำแนกประเภทแบบ *K* ลำดับที่ใกล้ที่สุด (*K*-Nearest Neighbor) ซึ่งในส่วนนี้ใช้วิธีการหาข้อมูลที่ใกล้ที่สุด 25 อันดับแรก การค้นคืนข้อมูลสามารถแสดงดังรูปที่ 3.3



รูปที่ 3.3 แสดงการค้นคืนข้อมูลเสียงจากแฟ้มข้อมูลเสียงภาษาไทย

จากรูปที่ 3.3 สามารถอธิบายขั้นตอนการทำงานได้ดังนี้ เมื่อผู้พูดทำการพูดข้อคำถามเสียงที่ต้องการค้นคืนแล้ว ข้อคำถามเสียงนั้นจะถูกนำไปสกัดค่าความถี่มูลฐาน (Fundamental Frequency) หลังจากนั้นผ่านขั้นตอนการปรับขนาด (Resampling) การปรับเรียบ (Smoothing) และการแปลงข้อมูลให้เป็นบรรทัดฐาน (Data Normalization) เมื่อข้อคำถามเสียงผ่านขั้นตอนการเตรียมข้อมูลเรียบร้อยแล้ว ขั้นตอนต่อไปเป็นการเปรียบเทียบสัญญาณเสียงระหว่างข้อคำถามเสียงกับข้อมูลในแฟ้มข้อมูลเสียงที่ผ่านการเตรียมข้อมูลแล้วเช่นกัน ในขั้นตอนการเปรียบเทียบสัญญาณเสียงนั้นก็คือการค้นคืนข้อมูลเสียงจากแฟ้มข้อมูลเสียงว่ามีคำที่ต้องการค้นหาอยู่หรือไม่

3.3 การหาค่าอัตราความถูกต้อง

เนื่องจากงานวิจัยนี้ เป็นการค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงภาษาไทย ดังนั้นผลลัพธ์ที่ออกมาและถือว่าถูกต้องนั้น จะต้องเป็นคำหรือประโยคเดียวกับข้อคำถามเสียงที่ต้องการค้นหา ซึ่งวิธีที่ใช้วัดอัตราความถูกต้องของงานวิจัยนี้ ใช้ค่าความถูกต้องการค้นคืน ดังสูตรต่อไปนี้

$$\text{ค่าความถูกต้องการค้นคืน} = \frac{\text{ผลลัพธ์การค้นคืนที่ถูกต้องใน } K \text{ รายการแรก}}{\text{จำนวนข้อคำถามเสียงทั้งหมด}}$$

จากสูตรข้างต้น ในงานวิจัยนี้เลือกใช้ผลลัพธ์การค้นคืนและถูกต้องใน $K = 25$ ลำดับแรก เนื่องจากให้ค่าความถูกต้องการค้นคืนอยู่ในระดับที่เหมาะสม ความหมายของ ผลลัพธ์การค้นคืนที่ถูกต้องใน 25 รายการแรก หมายถึง ถ้าหากผลลัพธ์ที่ค้นคืนออกมาและถูกต้องอยู่ใน 25 ลำดับ

แรก จะมีคะแนน เท่ากับ 1 แต่ถ้าผลลัพธ์ที่ถูกต้องไม่อยู่ใน 25 ลำดับแรก จะมีคะแนน เท่ากับ 0 ซึ่งจะถือว่าไม่พบคำนั้นในแฟ้มข้อมูลเสียง

ในส่วนของเวลาที่ใช้ในการค้นคืนข้อมูล เวลาที่ถือว่าผู้ใช้ยอมรับได้นั้น คือ เวลาที่ใช้ในการค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงที่มีความยาวอย่างน้อย 60 นาที จะต้องใช้เวลาในการค้นคืนไม่เกิน 10 นาที จึงจะถือได้ว่าเป็นว่าที่ผู้ใช้สามารถยอมรับได้

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 4

การทดลองและผลการทดลอง

ในบทนี้จะกล่าวถึงวิธีการทดลองและผลการทดลองของงานวิจัยเรื่อง การค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงภาษาไทยด้วยข้อคำถามเสียง ซึ่งเป็นการทดลองหาอัตราความถูกต้องของการค้นคืนข้อมูลเสียงและพิจารณาถึงเวลาที่ใช้ในการทำงาน ซึ่งเป็นเวลาที่ผู้ใช้สามารถยอมรับได้ ในการทดลองแฟ้มข้อมูลเสียงที่ใช้นั้นมาจาก 2 แหล่งคือ

1. การบันทึกด้วยไมโครโฟนผ่านโปรแกรมบันทึกเสียง เช่น แฟ้มข้อมูลเสียงสื่อการเรียนการสอนอิเล็กทรอนิกส์ และแฟ้มข้อมูลเสียงการโอนสายโทรศัพท์ เป็นต้น
2. จากเว็บไซต์ทางอินเทอร์เน็ต

4.1 วิธีการทดลอง

การทดลองการค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงภาษาไทยด้วยข้อคำถามเสียงมีการกำหนดค่าพารามิเตอร์และคุณสมบัติต่างๆ ที่ใช้ในการทดลอง ดังต่อไปนี้

4.1.1 การทดลองค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงภาษาไทยด้วยวิธีวัดระยะทางแบบไดนามิกไทม์วอร์ปปิง โดยใช้ค่าความถี่มูลฐานในการค้นคืน

1. แฟ้มข้อมูลที่ใช้ในการทดลองด้วยวิธีการค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงภาษาไทยด้วยข้อคำถามเสียงมีทั้งหมด 2 ชุด ดังนี้
   - แฟ้มข้อมูลเสียงสื่อการเรียนการสอนอิเล็กทรอนิกส์ เป็นแฟ้มข้อมูลเสียงเกี่ยวกับ การสำรองและกู้คืนข้อมูล และระบบสนับสนุนการตัดสินใจ [17] ความยาว 62.54 นาที มีจำนวนคำทั้งหมด 8,171 คำ และเสียงที่ใช้บันทึกเป็นเสียงผู้ชาย
   - แฟ้มข้อมูลเสียงจากอินเทอร์เน็ต เป็นแฟ้มข้อมูลเสียงเกี่ยวกับธรรมะ เรื่องเหตุใดจึงเกิดเป็นผู้มีรูปงาม จากหนังสือ “เสียดาย…คนตายไม่ได้อ่าน” [18] ความยาว 124.23 นาที มีจำนวนคำทั้งหมด 9,159 คำ โดยเสียงที่ใช้บันทึกเป็นเสียงผู้หญิง บันทึกโดย คุณอลิสา ฉัตรานนท์ จากเว็บไซต์ http://dungtrin.com
2. จำนวนข้อคำถามเสียงที่ใช้ในการทดลองด้วยวิธีการค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงภาษาไทยด้วยข้อคำถามเสียง มีทั้งหมด 2 ชุด ดังต่อไปนี้

   ชุดที่ 1 เป็นข้อคำถามเสียงที่ใช้ในการทดลองวัดค่าความถูกต้องของการค้นคืนข้อมูล กรณีที่ข้อคำถามเสียงกับแฟ้มข้อมูลเสียงเป็น

คนพูดคนเดียวกัน มีจำนวนข้อคำถามเสียง10 ข้อคำถามเสียง โดยเป็นเสียงของผู้พูดคนเดียวกับแฟ้มข้อมูลเสียง ซึ่งมีข้อคำถามเสียงดังตารางที่ 4.1

ตารางที่ 4.1 ข้อคำถามเสียงที่ใช้ในการค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงสื่อการเรียนการสอนอิเล็กทรอนิกส์ กรณีที่ข้อคำถามเสียงกับแฟ้มข้อมูลเสียงเป็นคนพูดคนเดียวกัน

ข้อคำถามเสียงที่	ข้อคำถามเสียง
1	ตัวอย่าง
2	โครงสร้างข้อมูล
3	การกู้คืนข้อมูล
4	การตัดสินใจ
5	การบริหารข้อมูล
6	การสำรองข้อมูล
7	บุคลากร
8	การออกแบบฐานข้อมูล
9	ฐานข้อมูล
10	ร่วมแก้ปัญหา

ชุดที่ 2 เป็นข้อคำถามเสียงที่ใช้ในการทดลองวัดค่าความถูกต้องของการค้นคืนข้อมูล กรณีที่ข้อคำถามเสียงกับแฟ้มข้อมูลเสียงเป็นคนพูดคนละคน มีจำนวนข้อคำถามเสียงทั้งหมด100 ข้อคำถามเสียง โดยมาจากผู้พูด 5 คน แบ่งเป็นเพศชาย 3 คน และเพศหญิง 2 คน ซึ่งผู้พูดแต่ละคน จะทำการเลือกข้อคำถามเสียงจากเอกสารที่ผู้เขียนได้ทำการแจก จำนวน 2 ชุดด้วยกัน ชุดแรกคือเอกสารที่มีเนื้อหาตรงกับแฟ้มข้อมูลเสียงสื่อการเรียนการสอนอิเล็กทรอนิกส์ ส่วนชุดที่สองคือเอกสารที่มีเนื้อหาตรงกับแฟ้มข้อมูลเสียงจากอินเทอร์เน็ต โดยที่ผู้พูดแต่ละท่านจะต้องทำการเลือกข้อคำถามเสียงที่จะค้นหาจากเอกสารแต่ละชุด ชุดละ 10 ข้อคำถามเสียง เมื่อรวมแล้วผู้พูด 1 ท่านจะเลือกข้อคำถามเสียงที่ใช้ในการค้นคืนได้ 20 ข้อคำถามเสียง โดยที่

ข้อคำถามเสียงที่ผู้พูดแต่ละท่านทำการเลือกมีดังตารางที่ 4.2 และ ตารางที่ 4.3

ตารางที่ 4.2 ข้อคำถามเสียงที่ผู้พูดแต่ละคนเลือกใช้ในการค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงสื่อการเรียนการสอนอิเล็กทรอนิกส์

ข้อคำถามเสียงที่	ข้อคำถามเสียง				
	เพศหญิง		เพศชาย		
	คนที่ 1	คนที่ 2	คนที่ 3	คนที่ 4	คนที่ 5
1	DSS	แฟ้มข้อมูล	เทคโนโลยี GDSS	การกู้ฐานข้อมูล	ข้อเสียของการจัดการฐานข้อมูล
2	โครงสร้างข้อมูล	กระบวนการในการตัดสินใจ	โครงสร้างข้อมูล	การตัดสินใจ	ข้อดีของการจัดการฐานข้อมูล
3	การกู้ข้อมูล	การคัดเลือก	การกู้คืนข้อมูล	ความซ้ำซ้อนของข้อมูล	ความสามารถในการเข้าถึงข้อมูล
4	การควบคุม	การสำรอง	การตัดสินใจ	ชุดคำสั่ง	ฐานข้อมูล
5	การตัดสินใจ	การสืบค้นข้อมูล	การบริหารข้อมูล	ฐานแบบจำลอง	ตัวอย่าง
6	การสำรองข้อมูล	ฐานข้อมูล	การสำรองข้อมูล	ตัวอักขระ	บุคลากร
7	Software	ภาษาคำนิยามข้อมูล	การสืบค้นข้อมูล	ผู้บริหารฐานข้อมูล	ประสิทธิผล
8	บุคลากร	ลดความยุ่งยาก	การออกแบบฐานข้อมูล	พจนานุกรม	ประสิทธิภาพ
9	ระบบจัดการฐานข้อมูล	หน้าที่ทางการจัดการ	ฐานข้อมูล	มินิคอมพิวเตอร์	วัตถุประสงค์
10	วิธีการประมวลผล	อุปกรณ์สื่อสาร	ร่วมแก้ปัญหา	ระบบรักษาความปลอดภัย	อุปนัย

ตารางที่ 4.3 ข้อคำถามเสียงที่ผู้พูดแต่ละคนเลือกใช้ในการค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงจากอินเทอร์เน็ต

ข้อคำถามเสียงที่	ผู้พูดข้อคำถามเสียง				
	เพศหญิง		เพศชาย		
	คนที่ 1	คนที่ 2	คนที่ 3	คนที่ 4	คนที่ 5
1	ขาวงามแบบบริสุทธิ์	งามแบบแปลกประหลาด	วิบากกรรม	แปลกประหลาด	ลักษณะขัดแย้ง
2	งามแบบอ่อนหวาน	งามแบบสง่า	โทสะควบคุมจิตใจ	ทำทานด้วยศรัทธา	งามแบบอ่อนหวาน
3	นิยามของความงาม	ฉวีวรรณ	ให้อภัยเป็นทาน	งามแบบสง่า	ทำทานด้วยศรัทธา
4	บทสำรวจตนเอง	ลักษณะขัดแย้ง	งามแบบเร้าความรู้สึกทางเพศ	ทุคติ	พระประธาน
5	ปฏิมากร	ทำทานด้วยศรัทธา	งามแบบแปลกประหลาด	บทสำรวจตนเอง	พระพุทธเจ้า
6	พระฉวี	บทสำรวจตนเอง	งามแบบสง่า	ผู้มีรูปงาม	สรุป
7	มหากุศลกรรม	พระเนตร	ตำราทายมนุษย์	มหากุศลกรรม	รักใคร่เมตตา
8	อาการทางใจและวิธีคิด	พระทนต์	ทำทานด้วยศรัทธา	รักษาศีล	รักษาศีล
9	สรุป	มหากุศลกรรม	รักษาศีล	ลักษณะขัดแย้ง	ศีลธรรม
10	หูบาง	อาการทางใจและวิธีคิด	รูปโฉม	อาการทางใจ	มหากุศลกรรม

3. ข้อคำถามเสียงและแฟ้มข้อมูลเสียงที่ใช้ในการทดลองของวิธีการค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงภาษาไทยด้วยข้อคำถามเสียงนี้ จัดเก็บอยู่ในรูปไฟล์ WAV ทำการลดอัตราการชักตัวอย่าง (Sampling Rate) เหลือ 2000 เฮิรตซ์ และเก็บข้อมูลขนาด 16 บิต แบบช่องสัญญาณเดียว (Mono)
4. ทำการปรับค่าพารามิเตอร์ต่างๆ ที่ใช้ในการค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงภาษาไทยด้วยข้อคำถามเสียง เพื่อให้ได้ค่าอัตราความถูกต้องสูงขึ้น ซึ่งค่าพารามิเตอร์ที่สำคัญ มีดังต่อไปนี้
    - ระยะการเลื่อนกรอบหน้าต่าง (Window Shift) เท่ากับครึ่งหนึ่งของข้อคำถามเสียง มาจากการทดลองเปรียบเทียบระยะการเลื่อนกรอบหน้าต่าง ดังแสดงในภาคผนวก ก ตารางที่ ก-6
    - ช่วงเวลา (Time Step) ที่ใช้ในการสกัดค่าความถี่มูลฐานเท่ากับ 0.01 วินาที มาจากการทดลองการเปรียบเทียบช่วงเวลาที่ใช้ในการสกัดค่าความถี่มูลฐาน ดังแสดงในภาคผนวก ก ตารางที่ ก-9
    - เงื่อนไขบังคับครอบคลุม (Global Constraint) ของวิธีวัดระยะทางแบบไดนามิกไทม์วอร์ปปิง เท่ากับ 3 เปอร์เซ็นต์ เนื่องจากทำการทดลองแล้วให้ค่าความแม่นยำสูงสุด

4.1.2 การทดลองที่ศึกษาถึงการทำงานของกระบวนการรู้จำคำพูด และวิธีการค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงภาษาไทยด้วยวิธีวัดระยะทางแบบไดนามิกไทม์วอร์ปปิง โดยใช้ค่าความถี่มูลฐานในการค้นคืน

1. ทำการทดลองเพื่อศึกษาถึงการทำงานของกระบวนการรู้จำคำพูด (Speech Recognition) และวิธีการค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงภาษาไทยด้วยวิธีวัดระยะทางแบบไดนามิกไทม์วอร์ปปิง โดยใช้ค่าความถี่มูลฐานในการค้นคืน ในงานวิจัยนี้เลือกใช้กระบวนการรู้จำคำพูด ด้วยกลวิธีการหาคำสำคัญ (Keyword Spotting) [19] ในส่วนของข้อมูลเสียงที่ใช้ในการทดลองด้วยวิธีกระบวนการรู้จำคำพูด ได้ทดลองกับกลุ่มเสียงที่ได้มีการเก็บอยู่แล้ว โดยใช้ในงานของการโอนสายโทรศัพท์ภายในองค์กร ซึ่งเลือกใช้เสียงในการโอนสายโทรศัพท์ภายในภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัยในการทดสอบ ในงานโอนสายโทรศัพท์นี้มีคำสำคัญจำนวน 48 คำ เป็นชื่อบุคลากรในภาควิชา และชื่อหน่วยงานในภาควิชา เช่น ธุรการ ผู้ช่วย หัวหน้าภาค เป็นต้น ตัวอย่างเสียงที่ทำการเก็บนั้น ได้ทำการบันทึกที่อัตราการ

ชักตัวอย่าง 8000 เฮิรตซ์ การเก็บข้อมูล 8 บิต แบบช่องสัญญาณเดียว (Mono) และจัดเก็บอยู่ในรูปของไฟล์ WAV ในการทดลองได้แบ่งตัวอย่างเสียงออกเป็น 2 กลุ่ม คือ กลุ่มที่ใช้ในการเรียนรู้ และกลุ่มที่ใช้ในการทดสอบ ซึ่งมีรายละเอียดดังต่อไปนี้

- กลุ่มตัวอย่างเสียงที่ใช้ในการเรียนรู้ ตัวอย่างเสียงที่ใช้ในการเรียนรู้นั้นแบ่งออกเป็น 3 ประเภท คือ
  1. การบันทึกโดยให้กลุ่มตัวอย่างเสียงพูด บทความทั่วไป ในกลุ่มตัวอย่างเสียงนี้ ได้ทำการบันทึกโดยการให้กลุ่มตัวอย่างเสียงทำการพูดบทความที่ได้เตรียมไว้ โดยมีกลุ่มตัวอย่างเสียงจำนวน 59 คน แบ่งเป็นเพศชาย 37 คน และเพศหญิง 22 คน
  2. การบันทึกเสียงโดยให้กลุ่มตัวอย่างเสียงพูด คำสำคัญ ในกลุ่มตัวอย่างเสียงนี้ ได้ทำการบันทึกโดยให้กลุ่มตัวอย่างเสียง พูดแต่คำสำคัญในระบบซึ่งก็คือ ชื่อบุคลากรและหน่วยงานต่างๆ โดยกลุ่มตัวอย่างเสียงนี้มีจำนวน 22 คน แบ่งเป็นเพศชาย 13 คน และเพศหญิง 9 คน
  3. การบันทึกเสียงโดยให้กลุ่มตัวอย่าง พูดกลุ่มคำที่ประกอบด้วยหน่วยเสียงขาดแคลน เนื่องจากในบางหน่วยเสียงที่ประกอบอยู่ในคำพูดทั่วไปค่อนข้างน้อย เช่น /ia/ (หน่วยเสียงสระเอียะ) /@/ (หน่วยเสียงสระเอาะ) ทำให้หน่วยเสียงเหล่านี้มีจำนวนข้อมูลในการเรียนรู้น้อย ดังนั้นจึงได้กำหนดให้กลุ่มตัวอย่างเสียง ทำการบันทึกเสียงกลุ่มคำบางกลุ่ม ที่มีหน่วยเสียงขาดแคลนนี้เป็นส่วนประกอบ โดยทำการบันทึกจากกลุ่มตัวอย่างเสียง 8 คน แบ่งเป็นเพศชาย 6 คน และเพศหญิง 2 คน
- กลุ่มตัวอย่างเสียงที่ใช้ในการทดสอบ สำหรับกลุ่มตัวอย่างเสียงที่ใช้ในการทดสอบนี้ ได้ทำการเก็บบันทึกจากกลุ่มตัวอย่างจำนวน 23 คน จำนวน 250 ประโยค ซึ่งมาจากเพศชาย 20 คน และเพศหญิง 3 คน ในการบันทึกเสียงนั้น ได้กำหนดให้กลุ่มตัวอย่างเสียงเลือกในสิ่งที่จะพูดเอง ภายใต้โดเมนของการขอโอนสายโทรศัพท์

2. ข้อคำถามเสียงที่ใช้ในการทดลองเพื่อศึกษาถึงการทำงานของกระบวนการรู้จำคำพูด ด้วยกลวิธีการหาคำสำคัญ (Keyword Spotting) และวิธีการค้นคืนข้อมูล

จากแฟ้มข้อมูลเสียงการโอนสายโทรศัพท์ด้วยวิธีวัดระยะทางแบบไดนามิกไทม์วอร์ปปิง โดยใช้ค่าความถี่มูลฐานในการค้นคืน ดังตารางที่ 4.4

ตารางที่ 4.4 ข้อคำถามเสียงที่ใช้ในการค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงการโอนสายโทรศัพท์ โดยวิธีกระบวนการรู้จำคำพูด ด้วยกลวิธีการหาคำสำคัญ (Keyword Spotting) และวิธีการค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงการโอนสายโทรศัพท์ด้วยวิธีวัดระยะทางแบบไดนามิกไทม์วอร์ปปิง โดยใช้ค่าความถี่มูลฐานในการค้นคืน

ข้อคำถามเสียงที่	ข้อคำถามเสียง
1	บุญเสริม
2	ผู้ช่วยหัวหน้าภาค
3	บุญชัย
4	เฉลิมเอก
5	หัวหน้าภาค
6	โปรดปราน
7	สมชาย
8	ทวิตีย์
9	ธุรการ
10	อรรถวิทย์

ในงานวิจัยนี้ผู้เขียนได้ทำการทดลองค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงภาษาไทยด้วยวิธีวัดระยะทางแบบไดนามิกไทม์วอร์ปปิง โดยใช้ค่าความถี่มูลฐานในการค้นคืน และทำการทดลองค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงภาษาไทย โดยใช้กระบวนการรู้จำคำพูด ด้วยกลวิธีการหาคำสำคัญ (Keyword Spotting) ซึ่งแบ่งการทดลองออกเป็น 5 ชุดการทดลอง ดังต่อไปนี้

การทดลองชุดที่ 1 การวัดอัตราค่าความถูกต้องของการค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงภาษาไทยด้วยวิธีวัดระยะทางแบบไดนามิกไทม์วอร์ปปิง โดยใช้ค่าความถี่มูลฐานในการค้นคืน กรณีที่ข้อคำถามเสียงกับแฟ้มข้อมูลเสียงเป็นคนพูดคนเดียวกัน

การทดลองชุดที่ 2 การวัดอัตราค่าความถูกต้องของการค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงภาษาไทยด้วยวิธีวัดระยะทางแบบไดนามิกไทม์วอร์ปปิง โดยใช้ค่าความถี่มูลฐานในการค้นคืน กรณีที่ข้อคำถามเสียงกับแฟ้มข้อมูลเสียงเป็นคนพูดคนละคน

การทดลองชุดที่ 3 การวัดอัตราค่าความถูกต้องของการค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงในการโอนสายโทรศัพท์ โดยใช้กระบวนการรู้จำคำพูด ด้วยกลวิธีการหาคำสำคัญ (Keyword Spotting) ซึ่งใช้ค่าสัมประสิทธิ์เมลฟรีเคว็นซีเคปสตรอล 39 มิติ (Mel Frequency Cepstral Coefficient หรือ MFCC) ในการรู้จำ

การทดลองชุดที่ 4 การวัดอัตราค่าความถูกต้องของการค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงในการโอนสายโทรศัพท์ โดยใช้กระบวนการรู้จำคำพูด ด้วยกลวิธีการหาคำสำคัญ (Keyword Spotting) ซึ่งใช้ค่าความถี่มูลฐาน (Fundamental Frequency) ในการรู้จำ

การทดลองชุดที่ 5 การวัดอัตราค่าความถูกต้องของการค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงในการโอนสายโทรศัพท์ด้วยวิธีวัดระยะทางแบบไดนามิกไทม์วอร์ปปิง โดยใช้ค่าความถี่มูลฐานในการค้นคืน

การทดลองชุดที่ 1 และการทดลองชุดที่ 2 เป็นการวัดอัตราค่าความถูกต้องของการค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงสื่อการเรียนการสอนอิเล็กทรอนิกส์และแฟ้มข้อมูลเสียงจากอินเทอร์เน็ตด้วยวิธีวัดระยะทางแบบไดนามิกไทม์วอร์ปปิง โดยใช้ค่าความถี่มูลฐานในการค้นคืน ทั้งกรณีที่ข้อคำถามเสียงกับแฟ้มข้อมูลเสียงเป็นคนพูดคนละคน และกรณีที่ข้อคำถามเสียงกับแฟ้มข้อมูลเสียงเป็นคนพูดคนละคน การทดลองชุดที่ 3 ถึง การทดลองชุดที่ 5 เป็นการวัดอัตราค่าความถูกต้องของการค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงการโอนสายโทรศัพท์ โดยที่การทดลองชุดที่ 3 และการทดลองชุดที่ 4 เป็นการวัดอัตราค่าความถูกต้องโดยใช้กระบวนการรู้จำคำพูด ด้วยกลวิธีการหาคำสำคัญ (Keyword Spotting) ซึ่งใช้ค่าสัมประสิทธิ์เมลฟรีเคว็นซีเคปสตรอล 39 มิติ (Mel Frequency Cepstral Coefficient หรือ MFCC) ในการรู้จำ และใช้ค่าความถี่มูลฐาน (Fundamental Frequency) ในการรู้จำ ตามลำดับ การทดลองชุดที่ 5 เป็นการวัดอัตราค่าความถูกต้องด้วยวิธีวัดระยะทางแบบไดนามิกไทม์วอร์ปปิง โดยใช้ค่าความถี่มูลฐานในการค้นคืน ในส่วนของการทดลองเบื้องต้นอื่นๆ ที่เกี่ยวข้อง เช่น การทดลองค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงภาษาไทยด้วยข้อคำถามเสียงพยางค์เดียว และการทดลองค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงภาษาไทยด้วยข้อคำถามเสียงหลายพยางค์นั้น ได้กล่าวไว้ในส่วนของภาคผนวก ก

4.2 ผลการทดลอง

จากการทดลองดังกล่าวข้างต้น ได้ผลการทดลองดังตารางที่ 4.5 ถึง ตารางที่ 4.9 ดังต่อไปนี้

ตารางที่ 4.5 การทดลองชุดที่ 1 ผลการวัดอัตราค่าความถูกต้องของการค้นคืนข้อมูลใน 25 รายการแรก จากแฟ้มข้อมูลเสียงภาษาไทยด้วยวิธีวัดระยะทางแบบไดนามิกไทม์วอร์ปปิง โดยใช้ค่าความถี่มูลฐานในการค้นคืน กรณีที่ข้อคำถามเสียงกับแฟ้มข้อมูลเสียงเป็นผู้พูดคนเดียวกัน

ข้อคำถามเสียง	ผลการค้นคืนที่ถูกต้อง 25 รายการแรก (0 หรือ 1)
ตัวอย่าง	1
โครงสร้างข้อมูล	1
การกู้คืนข้อมูล	1
การตัดสินใจ	1
การบริหารข้อมูล	1
การสำรองข้อมูล	1
บุคลากร	1
การออกแบบฐานข้อมูล	1
ฐานข้อมูล	1
ร่วมแก้ปัญหา	0
ค่าความถูกต้องการค้นคืน (%)	90.00

จากตารางที่ 4.5 เป็นการคิดค่าความถูกต้องของการค้นคืนข้อมูลใน 25 รายการแรก จากแฟ้มข้อมูลเสียงสื่อการเรียนการสอนอิเล็กทรอนิกส์ด้วยวิธีวัดระยะทางแบบไดนามิกไทม์วอร์ปปิง โดยใช้ค่าความถี่มูลฐานในการค้นคืน กรณีที่ข้อคำถามเสียงกับแฟ้มข้อมูลเสียงเป็นผู้พูดคนเดียวกัน ในระดับของข้อคำถามเสียงที่ใช้ในการทดลอง

ตารางที่ 4.6 การทดลองชุดที่ 2 ผลการวัดอัตราค่าความถูกต้องของการค้นคืนข้อมูลใน 25 รายการแรก จากแฟ้มข้อมูลเสียงภาษาไทยด้วยวิธีวัดระยะทางแบบไดนามิกไทม์วอร์ปปิง โดยใช้ค่าความถี่มูลฐานในการค้นคืน กรณีที่ข้อคำถามเสียงกับแฟ้มข้อมูลเสียงเป็นคนพูดคนละคน

ประเภท	ผู้พูดข้อคำถามเสียง (เพศ)	ค่าความถูกต้อง (%)
แฟ้มข้อมูลเสียงสื่อการเรียนการสอนอิเล็กทรอนิกส์	คนที่ 1 (หญิง)	70.00
	คนที่ 2 (หญิง)	50.00
	คนที่ 3 (ชาย)	90.00
	คนที่ 4 (ชาย)	40
	คนที่ 5 (ชาย)	90
เฉลี่ย		**68**
แฟ้มข้อมูลเสียงจากอินเทอร์เน็ต	คนที่ 1 (หญิง)	50
	คนที่ 2 (หญิง)	60
	คนที่ 3 (ชาย)	30
	คนที่ 4 (ชาย)	60
	คนที่ 5 (ชาย)	50
เฉลี่ย		**50**
เฉลี่ยรวม		**59**

จากตารางที่ 4.6 เป็นการคิดค่าความถูกต้องของการค้นคืนข้อมูลใน 25 รายการแรก จากแฟ้มข้อมูลเสียงสื่อการเรียนการสอนอิเล็กทรอนิกส์และแฟ้มข้อมูลเสียงจากอินเทอร์เน็ต ด้วยวิธีวัดระยะทางแบบไดนามิกไทม์วอร์ปปิง โดยใช้ค่าความถี่มูลฐานในการค้นคืน กรณีที่ข้อคำถามเสียงกับแฟ้มข้อมูลเสียงเป็นผู้พูดคนละคนกัน ในระดับของผู้พูดข้อคำถามเสียง

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 4.7 การทดลองชุดที่ 3 ผลการวัดอัตราค่าความถูกต้องของการค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงในการโอนสายโทรศัพท์ โดยใช้กระบวนการรู้จำคำพูด ด้วยกลวิธีการหาคำสำคัญ (Keyword Spotting) ซึ่งใช้ค่าสัมประสิทธิ์เมลฟรีเคว็นซีเคปสตรอล 39 มิติ (Mel Frequency Cepstral Coefficient หรือ MFCC) ในการรู้จำ

คำที่ใช้ค้นหา	เวลาที่ใช้ในการทำงาน (วินาที)	ค่าความถูกต้อง(%)		
		ค่าความแม่นยำ (Precision)	ค่าความระลึก (Recall)	ค่ามาตรวัด F (F - Measure)
บุญเสริม	2.97	60.00	60.00	60.00
ผู้ช่วยหัวหน้าภาค	3.78	69.23	90.00	78.26
บุญชัย	9.98	14.29	60.00	23.08
เฉลิมเอก	3.22	83.33	55.56	66.67
หัวหน้าภาค	3.12	41.18	82.35	54.90
โปรดปราน	3.14	11.11	10.00	10.53
สมชาย	3.36	50.00	54.55	52.18
ทวิตีย์	3.79	45.45	71.43	55.55
ธุรการ	3.25	50.00	75.00	60.00
อรรถวิทย์	3.35	26.92	77.78	40.00
เฉลี่ย (%)	**4.00**	**45.15**	**63.67**	**50.12**

ตารางที่ 4.8 การทดลองชุดที่ 4 ผลการวัดอัตราค่าความถูกต้องของการค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงในการโอนสายโทรศัพท์ โดยใช้กระบวนการรู้จำคำพูด ด้วยกลวิธีการหาคำสำคัญ (Keyword Spotting) ซึ่งใช้ค่าความถี่มูลฐาน (Fundamental Frequency) ในการรู้จำ

คำที่ใช้ค้นหา	เวลาที่ใช้ในการทำงาน (วินาที)	ค่าความถูกต้อง(%)		
		ค่าความแม่นยำ (Precision)	ค่าความระลึก (Recall)	ค่ามาตรวัด F (F - Measure)
บุญเสริม	1.23	5.00	20.00	8.00
ผู้ช่วยหัวหน้าภาค	1.72	10.00	11.11	10.53

ตารางที่ 4.8 (ต่อ) การทดลองชุดที่ 4 ผลการวัดอัตราค่าความถูกต้องของการค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงในการโอนสายโทรศัพท์ โดยใช้กระบวนการรู้จำคำพูด ด้วยกลวิธีการหาคำสำคัญ (Keyword Spotting) ซึ่งใช้ค่าความถี่มูลฐาน (Fundamental Frequency) ในการรู้จำ

คำที่ใช้ค้นหา	เวลาที่ใช้ในการทำงาน (วินาที)	ค่าความถูกต้อง(%)		
		ค่าความแม่นยำ (Precision)	ค่าความระลึก (Recall)	ค่ามาตรวัด F (F - Measure)
บุญชัย	4.56	3.13	20.00	5.41
เฉลิมเอก	1.67	4.25	22.22	7.14
หัวหน้าภาค	1.43	3.09	17.64	5.26
โปรดปราน	1.44	3.13	10.00	4.77
สมชาย	1.49	2.86	18.18	4.94
ทวิตีย์	1.72	2.22	9.09	3.57
ธุรการ	1.45	2.24	25	4.11
อรรถวิทย์	1.50	2.25	22.22	4.09
เฉลี่ย (%)	**1.82**	**3.82**	**17.55**	**5.78**

ตารางที่ 4.9 การทดลองชุดที่ 5 ผลการวัดอัตราค่าความถูกต้องของการค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงในการโอนสายโทรศัพท์ด้วยวิธีวัดระยะทางแบบไดนามิกไทม์วอร์ปปิง โดยใช้ค่าความถี่มูลฐานในการค้นคืน

ข้อคำถามเสียง	เวลาทำงาน (วินาที)	คะแนน (0 หรือ 1)						
		5 รายการแรก	10 รายการแรก	15 รายการแรก	20 รายการแรก	25 รายการแรก	30 รายการแรก	35 รายการแรก
บุญเสริม	184.38	0	0	1	1	1	1	1
ผู้ช่วยหัวหน้าภาค	113.02	0	0	0	0	1	1	1
บุญชัย	223.22	0	0	0	0	0	0	1
เฉลิมเอก	820.82	0	1	1	1	1	1	1

ตารางที่ 4.9 (ต่อ) การทดลองชุดที่ 5 ผลการวัดอัตราค่าความถูกต้องของการค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงในการโอนสายโทรศัพท์ด้วยวิธีวัดระยะทางแบบไดนามิกไทม์วอร์ปปิง โดยใช้ค่าความถี่มูลฐานในการค้นคืน

ข้อคำถามเสียง	เวลาทำงาน (วินาที)	คะแนน (0 หรือ 1)						
		5 รายการแรก	10 รายการแรก	15 รายการแรก	20 รายการแรก	25 รายการแรก	30 รายการแรก	35 รายการแรก
หัวหน้าภาค	196.80	0	0	0	1	1	1	1
โปรดปราน	187.48	0	0	0	0	0	0	0
สมชาย	241.73	0	0	0	0	0	1	1
ทวิตีย์	225.39	1	1	1	1	1	1	1
ธุรการ	187.88	0	0	0	0	0	0	0
อรรถวิทย์	240.88	0	0	1	1	1	1	1
เฉลี่ย (%)	262.16	10	20	40	50	60	70	80

4.3 วิเคราะห์ผลการทดลอง

จากการทดลอง เมื่อพิจารณาถึงข้อคำถามเสียงที่ใช้ในการทดลองการค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงภาษาไทย ทั้งแฟ้มข้อมูลเสียงสื่อการเรียนการสอนอิเล็กทรอนิกส์และแฟ้มข้อมูลเสียงจากอินเทอร์เน็ต ทำให้พบจุดที่น่าสนใจคือ สามารถแบ่งประเภทของข้อคำถามเสียงที่ใช้ในการค้นคืนออกเป็น 4 ประเภท ดังนี้

1. ข้อคำถามเสียงที่เป็น หัวข้อ ชื่อเรื่อง และ/หรือ อยู่ในบทความ เช่น คำว่า “อุปนัย” ดังรูปที่ 4.1

> **อุปนัย** (Inductive approach) และวิธีนิรนัย (Deductive approach)
>
> 1. วิธี**อุปนัย** การออกแบบฐานข้อมูลด้วยวิธีอุปนัย เป็นการออกแบบฐานข้อมูลจากล่างขึ้นบน (Bottom-up design) ด้วยการเก็บรวบรวมข้อมูลที่มีการใช้งานอยู่แล้วภายในหน่วยงานต่าง ๆ ขององค์กร มาเชื่อมโยงเข้าด้วยกันเพื่อจัดทำเป็นระบบฐานข้อมูลขององค์กร ซึ่งมีข้อจำกัด คือ การนำกรรมวิธีย่อย ๆ ...

รูปที่ 4.1 ตัวอย่างข้อคำถามเสียงที่เป็น หัวข้อ ชื่อเรื่อง และ/หรือ อยู่ในบทความ

2. ข้อคำถามเสียงที่เป็น หัวข้อ ชื่อเรื่อง และ/หรือ คำที่อยู่ต้นย่อหน้า เช่น คำว่า “โครงสร้างข้อมูล” ดังรูปที่ 4.2

> **โครงสร้างข้อมูล** (File Structure)
>
> **โครงสร้างข้อมูล** หมายถึง ลักษณะการจัดแบ่งพิกัดต่าง ๆ ของข้อมูล …

รูปที่ 4.2 ตัวอย่างข้อคำถามเสียงที่เป็น หัวข้อ ชื่อเรื่อง และ/หรือ คำที่อยู่ต้นย่อหน้า

2. ข้อคำถามเสียงที่ปรากฏอยู่ต้นย่อหน้าจุดเดียว เช่น คำว่า “งามแบบแปลกประหลาด” ดังรูปที่ 4.3

> **งามแบบแปลกประหลาด** บางคนมีหลายมุมมองเหลือเกิน บางมุมดูแล้วดี อีกมุมดูแล้วชอบกล เอาไปบอกต่อได้ยากว่างามหรือไม่งามกันแน่ ต้องให้ดูเอาเอง ในอดีตชาติพวกนี้มักทำทานพอประมาณ รักษาศีลพอประมาณ แต่ชอบมีความคิดแหวกแนว พิลึกกึกกือ ไม่ค่อยลงใจสนิทกับทานและศีล …

รูปที่ 4.3 ตัวอย่างข้อคำถามเสียงที่อยู่ต้นย่อหน้า

4 อื่นๆ เช่น คำว่า”ผู้มีรูปงาม” ซึ่งเป็นข้อคำถามเสียงที่ปรากฏอยู่ในบทความ โดยไม่ได้ปรากฏอยู่ในส่วนอื่น เช่น หัวข้อ หรือ ชื่อเรื่อง ดังรูปที่ 4.4

> สำหรับสายตาคนอื่น**ผู้มีรูปงาม**ชนิดแลตะลึง หรือที่เรียกว่า ‘สวยจัด’ กับ ‘หล่อจัด’ นั้น เป็นบุคคลประเภทที่ปลุกเร้าให้เกิดความสับสนวุ่นวายใจ ความสวยหล่อจัดๆสามารถกระตุ้นให้เกิดความคิดหลากหลาย หรืออาจเรียกได้ว่ารบกวนให้คนเห็นกระวนกระวายใจผิดปกติ เพราะในหัวเกิดถ้อยคำพิเศษที่ไม่ค่อยปรากฏนักในการเห็นบุคคลทั่วไป เมื่อคนเราไม่สามารถอธิบายสิ่งที่ตัวเองเห็นออกมาเป็น …

รูปที่ 4.4 ตัวอย่างข้อคำถามเสียงที่ปรากฏอยู่ในบทความ

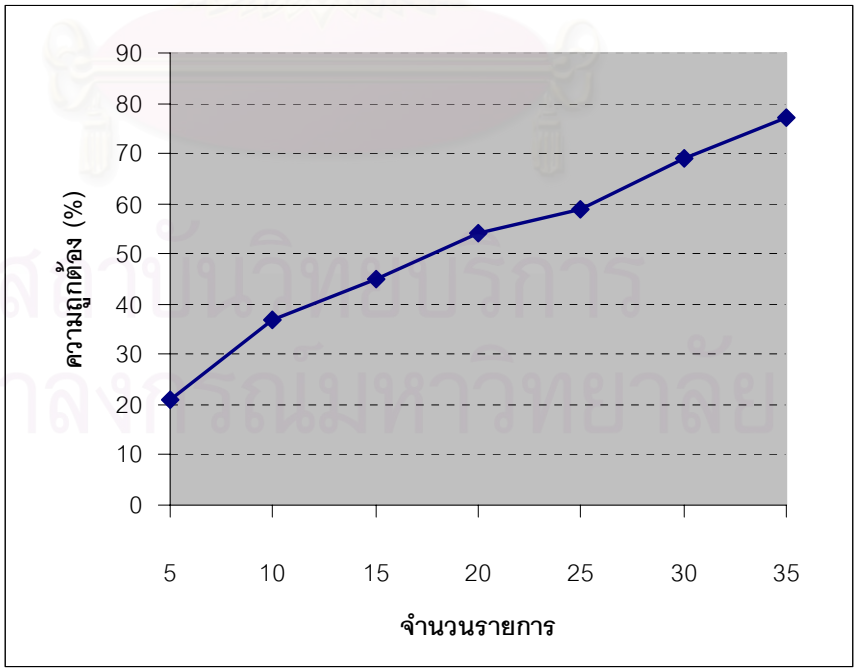
หลังจากทำการแบ่งประเภทของข้อคำถามเสียงออกเป็นแต่ละประเภทแล้ว เมื่อพิจารณาถึงความแม่นยำในการค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงภาษาไทยด้วยข้อคำถามเสียงแต่ละประเภท ปรากฏว่าให้ค่าความถูกต้องดังตารางที่ 4.10

ตารางที่ 4.10 ผลการวัดอัตราความถูกต้องของข้อคำถามเสียงแต่ละประเภท

ประเภทของข้อคำถามเสียง	ค่าความถูกต้อง (%)
ข้อคำถามเสียงที่เป็น หัวข้อ ชื่อเรื่อง และ/หรือ อยู่ในบทความ	69.44
ข้อคำถามเสียงที่เป็น หัวข้อ ชื่อเรื่อง และ/หรือ คำที่อยู่ต้นย่อหน้า	70.00
ข้อคำถามเสียงที่ปรากฏอยู่ต้นย่อหน้าจุดเดียว	53.85
อื่นๆ	48.72

จากตารางที่ 4.10 สรุปได้ว่า ข้อคำถามเสียงลักษณะที่เป็น หัวข้อ ชื่อเรื่อง หรือคำที่อยู่ต้นประโยคจะให้ค่าความแม่นยำการค้นคืนมากสุด ซึ่งเท่ากับ 70 เปอร์เซ็นต์ เนื่องจากข้อคำถามเสียงที่เป็น หัวข้อหรือชื่อเรื่อง นั้น เวลาพูดจะมีช่วงห่างระหว่างคำ ซึ่งทำให้การค้นคืนมีค่าความแม่นยำมากกว่าข้อคำถามเสียงประเภทอื่น

จากตารางที่ 4.5 และตารางที่ 4.6 สรุปได้ว่า อัตราความถูกต้องของการค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงภาษาไทย กรณีที่ข้อคำถามเสียงกับแฟ้มข้อมูลเสียงเป็นผู้พูดคนเดียวกันให้ค่าความแม่นยำสูงกว่า กรณีที่ข้อคำถามเสียงกับแฟ้มข้อมูลเสียงเป็นผู้พูดคนละคน ซึ่งให้ค่าความถูกต้องเท่ากับ 90 เปอร์เซ็นต์ และ 59 เปอร์เซ็นต์ ตามลำดับ นอกจากนี้ยังสรุปได้ว่า การค้นคืนข้อจากแฟ้มข้อมูลสื่อการเรียนการสอนอิเล็กทรอนิกส์ ให้ผลการค้นคืนแม่นยำกว่าแฟ้มข้อมูลเสียงจากอินเทอร์เน็ต ซึ่งให้ค่าความถูกต้องเท่ากับ 68 เปอร์เซ็นต์



รูปที่ 4.5 เปรียบเทียบค่าความถูกต้องในการค้นคืนในแต่ละช่วงของจำนวนรายการค้นคืน

จากรูปที่ 4.5 จะเห็นว่า แนวโน้มของค่าอัตราความถูกต้องมีค่าเพิ่มขึ้นเมื่อจำนวนรายการค้นคืนยิ่งมีค่าเพิ่มขึ้น เช่น ถ้าคิดค่าอัตราความถูกต้องของการค้นคืนที่ 5 รายการแรก จะให้ค่าความถูกต้องเท่ากับ 21 เปอร์เซ็นต์ แต่ถ้าคิดค่าอัตราความถูกต้องของการค้นคืนที่ 25 รายการแรก จะให้ค่าความถูกต้องเท่ากับ 59 เปอร์เซ็นต์ เป็นต้น

จากตารางที่ 4.7 ตารางที่ 4.8 และตารางที่ 4.9 สรุปได้ว่า การค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงในการโอนสายโทรศัพท์ โดยใช้กระบวนการรู้จำคำพูด ด้วยกลวิธีการหาคำสำคัญ (Keyword Spotting) ซึ่งใช้ค่าสัมประสิทธิ์เมลฟรีเคว็นซีเคปสตรอล 39 มิติ (Mel Frequency Cepstral Coefficient หรือ MFCC) ในการรู้จำ ให้ค่าความแม่นยำสูงกว่า การใช้ค่าความถี่มูลฐาน (Fundamental Frequency) เพียงอย่างเดียวในการรู้จำ โดยให้ค่าความถูกต้อง เท่ากับ 50.12 เปอร์เซ็นต์ และ 5.78 เปอร์เซ็นต์ ตามลำดับ นอกจากนี้ยังได้ว่า การค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงในการโอนสายโทรศัพท์ด้วยวิธีวัดระยะทางแบบไดนามิกไทม์วอร์ปปิง โดยใช้ค่าความถี่มูลฐานในการค้นคืน ให้ค่าความถูกต้องของการค้นคืน เท่ากับ 60 เปอร์เซ็นต์ ที่การค้นคืน 25 รายการแรก และเวลาที่ใช้ในการทำงานยังเป็นเวลาที่ผู้ใช้ยอมรับได้ โดยใช้เวลาเฉลี่ยในการค้นคืนเท่ากับ 262.16 วินาที หรือประมาณ 4 นาที

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 5

สรุปผลการวิจัยและข้อเสนอแนะ

งานวิจัยนี้เป็นการศึกษาเกี่ยวกับการค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงภาษาไทยด้วยข้อคำถามเสียง โดยการใช้เสียงวรรณยุกต์ภาษาไทยในการค้นคืน เนื่องจากภาษาไทยมีการผันวรรณยุกต์ 5 ระดับเสียงต่างกัน จึงทำให้ค่าความถี่มูลฐานของแต่ละคำต่างกันด้วย ดังนั้นผู้เขียนจึงนำค่าความถี่มูลฐานมาใช้ในการแยกเสียงวรรณยุกต์

5.1 สรุปผลการวิจัย

งานวิจัยนี้เป็นการค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงภาษาไทยด้วยข้อคำถามเสียง ซึ่งเป็นการค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงภาษาไทยขนาดใหญ่ โดยขนาดแฟ้มข้อมูลเสียงมีความยาวอย่างน้อย 60 นาที และแฟ้มข้อมูลเสียงมีอัตราการชักตัวอย่าง (Sampling Rate) ที่ 2000 เฮิรตซ์ มีการเก็บข้อมูลขนาด 16 บิต แบบช่องสัญญาณเดียว (Mono) ในส่วนของข้อคำถามเสียงที่ใช้ในการค้นคืนข้อมูลเสียงได้ทำการทดลองทั้งที่เป็นคำพยางค์เดียวและคำหลายพยางค์ ซึ่งในการค้นคืนข้อมูลนั้นไม่จำกัดว่าผู้พูดแฟ้มข้อมูลเสียงกับผู้พูดข้อคำถามเสียงจะต้องเป็นบุคคลเดียวกันและไม่จำกัดเพศด้วยเช่นกัน (Speaker Independent) สำหรับวิธีที่ใช้ในการค้นคืนข้อมูลเสียงนั้น ผู้เขียนเลือกใช้ค่าความถี่มูลฐาน (Fundamental Frequency) ในการค้นคืนข้อมูลเสียงและใช้วิธีวัดระยะทางแบบไดนามิกไทม์วอร์ปปิง (Dynamic Time Warping หรือ DTW) ในการเปรียบเทียบสัญญาณเสียงระหว่างข้อคำถามเสียงกับแฟ้มข้อมูลเสียงภาษาไทย ในส่วนของการแสดงผลลัพธ์ที่ได้จากการค้นคืนใช้วิธีการจำแนกประเภทแบบ *K* ลำดับที่ใกล้ที่สุด (*K*-Nearest Neighbor) จากผลการทดลองการค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงสื่อการเรียนการสอนและแฟ้มข้อมูลเสียงจากอินเทอร์เน็ตให้ผลลัพธ์ความถูกต้องเฉลี่ยของการค้นคืน เท่ากับ 59 เปอร์เซ็นต์ ส่วนการค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงในการโอนสายโทรศัพท์ โดยใช้กระบวนการรู้จำคำพูด ด้วยกลวิธีการหาคำสำคัญ (Keyword Spotting) ซึ่งใช้ค่าสัมประสิทธิ์เมลฟรีเคว็นซีเคปสตรอล 39 มิติ (Mel Frequency Cepstral Coefficient หรือ MFCC) ในการรู้จำ ให้ค่าความแม่นยำสูงกว่า การค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงในการโอนสายโทรศัพท์ โดยใช้กระบวนการรู้จำคำพูด ด้วยกลวิธีการหาคำสำคัญ (Keyword Spotting) ซึ่งใช้ค่าความถี่มูลฐาน (Fundamental Frequency) เพียงอย่างเดียวในการรู้จำ โดยให้ค่าความถูกต้อง เท่ากับ 50.12 เปอร์เซ็นต์ และ 5.78 เปอร์เซ็นต์ ตามลำดับ ส่วนการค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงในการโอนสายโทรศัพท์ด้วยวิธีวัดระยะทางแบบไดนามิกไทม์วอร์ปปิง โดยใช้ค่าความถี่มูลฐานในการค้นคืน ให้ค่าความถูกต้องของการค้นคืน เท่ากับ 60 เปอร์เซ็นต์ ที่การค้นคืน 25 รายการแรก ในด้านของเวลาที่ใช้ในการทำงานนั้นแม้ว่าไม่อาจจะสามารถระบุเป็นเชิงปริมาณว่าวิธีใดมีความเร็วมากน้อยกว่ากันเพียงไร เนื่องจากผลลัพธ์จากการ

ค้นคืนในแต่ละวิธีมีการจัดลำดับที่ต่างกัน แต่สามารถสรุปการทำงานโดยรวมได้ว่า วิธีการค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงภาษาไทยด้วยวิธีวัดระยะทางแบบไดนามิกไทม์วอร์ปปิงใช้เวลาในการทำงานอยู่ในระดับที่ผู้ใช้ยอมรับได้ เมื่อเทียบกับกระบวนการรู้จำคำพูด ด้วยกลวิธีการหาคำสำคัญ เนื่องจากไม่ต้องสูญเสียเวลาในการเรียนรู้และมีความยืดหยุ่นในเรื่องของการเลือกข้อคำถามเสียงที่จะค้นคืนมากกว่า เพราะในวิธีกระบวนการรู้จำคำพูด ด้วยกลวิธีการหาคำสำคัญนั้น ถ้าต้องการหาคำสำคัญที่ไม่ได้อยู่ในโดเมนที่สนใจอาจต้องทำการเรียนรู้โมเดลที่ใช้ในการรู้จำใหม่

5.2 **ข้อเสนอแนะ**

1. ค่าความถูกต้องจะเพิ่มมากขึ้น ถ้าผู้พูดข้อคำถามเสียงและแฟ้มข้อมูลเสียงที่ใช้ในการทดลองพูดชัดเจนและมีสำเนียงการพูดที่ดี
2. ขั้นตอนการเลื่อนกรอบหน้าต่าง (Window Shift) ในการเปรียบเทียบสัญญาณเสียงระหว่างข้อคำถามเสียงกับข้อมูลในแฟ้มข้อมูลเสียง ควรเลื่อนในระยะที่สั้นลง ซึ่งจะทำให้ผลการค้นคืนข้อมูลเสียงดีขึ้น เนื่องจากในการพูดของแต่ละคนจะมีลักษณะและทำนองของการพูดที่ต่างกัน
3 การเว้นวรรคการพูดในแฟ้มข้อมูลเสียงและข้อคำถามเสียง ควรมีความคล้ายกัน ซึ่งจะส่งผลให้การค้นคืนข้อมูลมีความถูกต้องมากขึ้น ดังตัวอย่างเช่น
   - การบันทึกแฟ้มข้อมูลเสียง ผู้พูดแฟ้มข้อมูลเสียงพูดว่า “ระบบสนับสนุน การตัดสินใจ”
   - ผู้พูดข้อคำถามเสียงพูดว่า “ระบบสนับสนุนการตัดสินใจ”

   จากตัวอย่างข้างต้นจะเห็นว่า การพูดข้อคำถามเสียงกับข้อมูลที่อยู่ในแฟ้มข้อมูลเสียงพูดเว้นวรรคต่างกัน ซึ่งจะส่งผลให้การค้นคืนให้ค่าความถูกต้องลดน้อยลง

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

รายการอ้างอิง

[1] เสียง, Available from: http://th.wikipedia.org/wiki/%E0%B9%80%E0%B8%AA%E0%B8%B5%E0%B8%A2%E0%B8%87 [2007,February 25].

[2] เสียง, Available from: http://tcmc.nisit.kps.ku.ac.th/tcmc/modules.php?op=modload&name=News&file=article&sid=92&mode=thread&order=0&thold=0 [2007,February 25].

[3] เสียง, Available from: http://classroom.psu.ac.th/users/wkomson/data/western-musuc/Chapter2/chap2-1.htm [2007,February 25].

[4] บัณฑิต จิตคงชื่น. การใช้ข้อมูลเสียงวรรณยุกต์ในการรู้จำเสียงพูดภาษาไทย. โครงงานทางวิศวกรรมระดับปริญญาตรี ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย, 2549.

[5] Boersma and D. Weenink, Praat doing phonetics by computer (Version 4.5.14) [Computer software], Available from: http://www.fon.hum.uva.nl/praat/ [2007,February 25].

[6] การแปลงฟูเรียร์อย่างเร็ว, Available from: www.kmitl.ac.th/~kchsomsa/somsak/crse_dsp/chap_6pn.pdf [2007,February 25].

[7] สมบูรณ์ แซ่เล้า และธนนท์ สุทธิกุล. การพัฒนาระบบประเมินเสียงร้องเพลง. โครงงานทางวิศวกรรมระดับปริญญาตรี ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย, 2549.

[8] W. A. Barret. Introduction to Digital Signal and Image Processing, Available from: http://iul.cs.byu.edu/450/F2000/ [2007,February 25].

[9] S. Chu, E. Keogh, D. Hart, and M. Pazzani. Iterative deepening dynamic time warping for time series. In: Proceedings of 2nd SIAM international conference on data mining, 2002.

[10] L. Tan, M. Karnjanadecha, T. Khaorapapong, and P. Tandayya. A Study of Thai Tone Classification. In: Proceedings of 4th Information Engineering Postgraduate Workshop 2004, pp. 24-27, Phuket, Thailand, Jan. 22-23, 2004.

[11] A. W. Fu, E. Keogh, L. Y. H. Lau, and C. A. Ratanamahatana. Scaling and Time Warping in Time Series Querying. In Proceedings of 31st International conference on Very Large Data Bases (VLDB), Trondheim, Norway, 2005.

[12] Euclidean Distance, Available from: http://en.wikipedia.org/wiki/Euclidean_distance [2007,February 25].

[13] พงศกร ธีรภาพวงศ์ และไวยณ์วุฒิ เอื้อจงประสิทธิ์. การค้นหาเพลงโดยการร้องทำนอง. โครงงานทางวิศวกรรมระดับปริญญาตรี ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย, 2549.

[14] จันทิมา พลพินิจ, วีรยา อมรพงษ์กุล, อัษฎาภรณ์ พิมพ์สมบูรณ์ และอุมาภรณ์ สายแสงจันทร์. การเปรียบเทียบเทคนิคการตัดคำที่แตกต่างสำหรับการย่อข้อความภาษาไทย. The Proc of NCSEC, 1999.

[15] A. Tungthangthum. Tone Recognition for Thai. IEEE Trans. Speech Audio Processing, pp 157-160, 1998.

[16] Y. Zhu and D. Shasha. Warping Indexes with Envelope Transforms for Query by Humming. In Proceedings of the ACM SIGMOD, International Conference on management of Data, pp. 181 – 192, 2003.

[17] ระบบสนับสนุนการตัดสินใจ (Decision Support Systems), Available from: http://www.sirikitdam.egat.com/WEB_MIS/107/index.html [2007,February 25].

[18] ดังตฤน. เสียดาย… คนตายไม่ได้อ่าน, Available from: http://multimedia.dungtrin.net/siadai.html [2007,February 25].

[19] S. Tangruamsub, P. Punyabukkana, and A. Suchato. Thai Speech Keyword Spotting using Heterogeneous Acoustic Modeling. 5th International Conference on Research, Innovation & Vision for the Future Information & Communication Technologies, IEEE RIVF'07, March 5-9, 2007.

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก ก

การทดลองที่เกี่ยวข้อง

งานวิจัยนี้ ทำการศึกษาวิธีการที่ช่วยลดเวลาในการค้นคืนข้อมูลเสียงภายในแฟ้มข้อมูลเสียงขนาดใหญ่ ด้วยข้อคำถามเสียง ซึ่งมีด้วยกันหลายวิธี และมีประสิทธิภาพที่ต่างกัน ดังนั้นผู้เขียนจึงจำเป็นต้องทดสอบและวัดประสิทธิภาพในการทำงาน เพื่อให้ได้มาซึ่งวิธีการที่เหมาะสมและมีประสิทธิภาพมากที่สุด ซึ่งวิธีการที่ใช้ในการวัดประสิทธิภาพมีดังต่อไปนี้

1 ค่าความแม่นยำ (P) [14]

คือการหาอัตราส่วนของการค้นพบคำที่ต้องการค้นหาได้และตรงกับคำที่ต้องการค้นหาจากจำนวนคำที่ค้นหาได้ทั้งหมด ดังสมการ

$$P = \frac{C}{A_R}$$

โดยที่ P คือ ค่าความแม่นยำ (Precision Value)

C คือ จำนวนคำที่ค้นหาได้และตรงกับคำที่ต้องการค้นหา

A_Rคือ จำนวนคำที่ค้นหาได้ทั้งหมด

ตัวอย่างการหาค่าความแม่นยำของการค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงด้วยข้อคำถามเสียง เช่น จำนวนคำที่ต้องการค้นหาทั้งหมดมี 5 คำ ผลจากการค้นคืนของคำที่ต้องการหาทั้งหมดค้นคืนออกมา 8 รายการ ดังนั้นค่าความแม่นยำ มีค่าเท่ากับ $\frac{5}{8}$ หรือเท่ากับ 0.625

2 ค่าความระลึก (R) [14]

คือ การหาอัตราส่วนของการค้นพบคำที่ต้องการค้นหาได้และตรงกับคำที่ต้องการค้นหาจากจำนวนคำที่ต้องการค้นหาทั้งหมด ดังสมการ

$$R = \frac{C}{A_C}$$

โดยที่ R คือ ค่าความระลึก (Recall Value)

C คือ จำนวนคำที่ค้นหาได้และตรงกับคำที่ต้องการค้นหา

A_Cคือ จำนวนคำที่ต้องการค้นหาทั้งหมด

ตัวอย่างการหาค่าความระลึกของการค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงด้วยข้อคำถามเสียง เช่น จำนวนคำที่ต้องการค้นหาทั้งหมดมี 5 คำ ผลจากการค้นคืนของคำที่ต้องการหาทั้งหมด

ค้นคืนออกมา 8 รายการ เพราะฉะนั้นจากสมการจะได้ ค่าความระลึก มีค่าเท่ากับ $\frac{5}{5}$ หรือเท่ากับ 1

3 ค่ามาตรวัด F (The F Measure) [14]

เป็นการแสดงความสัมพันธ์ระหว่างค่าความแม่นยำ (*P*) และ ค่าความระลึก (*R*) มีสมการดังนี้

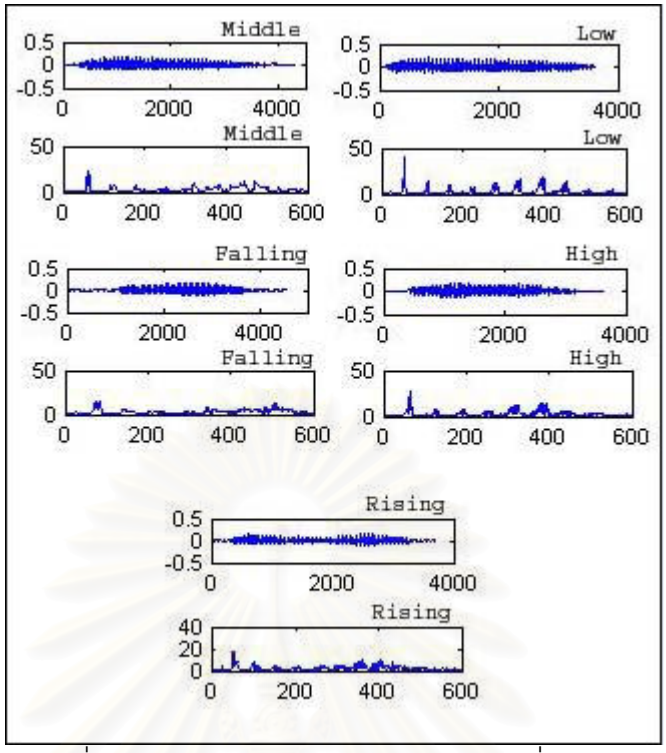
$$F\_Measure = \frac{(2 \times P \times R)}{P + R}$$

เนื่องจากการวัดประสิทธิภาพของการค้นคืนจะต้องดูทั้งค่าความแม่นยำและค่าความระลึกซึ่งทำให้ไม่สะดวก ดังนั้นงานวิจัยนี้จึงใช้ค่ามาตรวัด F ในการวัดประสิทธิภาพของการค้นคืน ตัวอย่างการหาค่ามาตรวัด F ของการค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงด้วยข้อคำถามเสียง เช่น จำนวนคำที่ต้องการค้นหาทั้งหมดมี 5 คำ ผลจากการค้นคืนของคำที่ต้องการหาทั้งหมดค้นคืนออกมา 8 รายการ ดังนั้นค่ามาตรวัด F มีค่าเท่ากับ $\frac{(2 \times 0.625 \times 1)}{0.625 + 1}$ หรือเท่ากับ 0.77

ในการศึกษาวิธีการที่ช่วยลดเวลาในการค้นคืนข้อมูลภายในแฟ้มข้อมูลเสียงขนาดใหญ่ด้วยข้อคำถามเสียง ซึ่งมีด้วยกันหลายวิธี และมีประสิทธิภาพที่ต่างกัน ดังนั้นผู้เขียนจึงทำการทดลองในหลายรูปแบบดังต่อไปนี้

ก.1 การค้นคืนข้อมูลเสียงจากแฟ้มข้อมูลเสียงภาษาไทยด้วยวิธีวัดระยะทางแบบไดนามิกไทม์วอร์ปปิง โดยวิธีการแปลงฟูเรียร์อย่างเร็ว

การออกเสียงในภาษาต่างๆ มีโครงสร้างที่คล้ายกันซึ่งประกอบไปด้วยความถี่มูลฐาน และความถี่ที่แสดงคุณสมบัติของเสียง หน่วยของเสียงที่มีขนาดเล็กที่สุดคือ หน่วยเสียง (Phoneme) ซึ่งประกอบด้วย เสียงของพยัญชนะ และเสียงของสระ ในภาษาไทยเราใช้ความถี่มูลฐานเป็นตัวกำหนดระดับของสัญญาณเสียง (Tone) ซึ่งเสียงภาษาไทยมีด้วยกัน 5 ระดับ คือ สามัญ (Middle) เอก (Low) โท (Falling) ตรี (High) และจัตวา (Rising) ซึ่งเมื่อนำเสียงไปผ่านการแปลงให้อยู่ในโดเมนความถี่ (Frequency domain) ด้วยวิธีการแปลงฟูเรียร์อย่างเร็วจะได้ผลลัพธ์ดังรูปที่ ก-1 ต่อไปนี้



รูปที่ ก-1 สัญญาณเสียงที่อยู่ในรูปไฟล์ WAV และสัญญาณเสียงที่ผ่านการแปลงให้อยู่ในโดเมนความถี่ด้วยวิธีการแปลงฟูเรียร์อย่างเร็ว ของสัญญาณเสียง 5 ระดับ

จากรูปที่ ก-1 จะเห็นว่า สัญญาณเสียงที่ผ่านการแปลงให้อยู่ในโดเมนความถี่ ของทั้ง 5 ระดับเสียง คือคำว่า ดา ด่า ด้า ด๊า ด๋า มีความแตกต่างกัน โดยการพลอตกราฟดังกล่าวทำการใส่ค่าสัมบูรณ์ (Absolute) ของค่าที่ได้จากการแปลงฟูเรียร์อย่างเร็ว ดังนั้นผู้เขียนจึงทำการศึกษาและทดลองเกี่ยวกับผลของระดับสัญญาณเสียงที่มีต่อการค้นคืนข้อมูลเสียง จากการศึกษาผู้เขียนได้ทำการเปรียบเทียบการค้นคืนข้อมูลเสียงระหว่างแฟ้มข้อมูลเสียงกับข้อคำถามเสียง ด้วยวิธีวัดระยะทางแบบไดนามิกไทม์วอร์ปปิงและวิธีการหาระยะห่างแบบยูคลิด โดยข้อคำถามเสียงกับแฟ้มข้อมูลเสียงเป็นเสียงบุคคลเดียวกัน (Speaker Dependent) ซึ่งทำการทดลองดังนี้

ก.1.1 การทดลองจำแนกประเภทเสียงวรรณยุกต์ลักษณะที่ 1

เป็นการทดลองค้นคืนข้อมูลเสียงจากแฟ้มข้อมูลเสียงที่ป็นคำพยางค์เดียวและมีช่วงห่างระหว่างคำประมาณ 0.5 - 1 วินาที เช่น ดา ด่า ด้า ด๊า ด๋า เป็นต้น

จากผลการทดลองพบว่าวิธีวัดระยะทางแบบไดนามิกไทม์วอร์ปปิงสามารถแยกประเภทเสียงวรรณยุกต์ลักษณะที่ 1 ได้ดีกว่าวิธีวัดระยะทางแบบยูคลิด ซึ่งได้ค่าความถูกต้องเท่ากับ 98.18 เปอร์เซ็นต์ โดยสามารถแสดงผลลัพธ์แยกตามประเภทวรรณยุกต์ได้ดังตารางที่ ก-1

ตารางที่ ก-1 ผลการจำแนกประเภทเสียงวรรณยุกต์ลักษณะที่ 1

เสียงวรรณยุกต์	ความถูกต้อง (%)					
	การหาระยะห่างแบบยูคลิด			วิธีวัดระยะทางแบบไดนามิกไทม์วอร์ปปิง		
	ค่าความแม่นยำ	ค่าความระลึก	ค่ามาตรวัด F	ค่าความแม่นยำ	ค่าความระลึก	ค่ามาตรวัด F
สามัญ	83.33	100	90.91	100	100	100
เอก	83.33	100	90.91	100	100	100
โท	62.50	100	76.92	83.33	100	90.91
ตรี	100	100	100	100	100	100
จัตวา	100	100	100	100	100	100
เฉลี่ย	**85.83**	**100**	**91.75**	**96.67**	**100**	**98.18**

ก.1.2 การทดลองจำแนกประเภทเสียงวรรณยุกต์ลักษณะที่ 2

เป็นการทดลองค้นคืนข้อมูลเสียงจากแฟ้มข้อมูลเสียงที่เป็นคำพยางค์เดียว แต่ไม่มีช่วงห่างระหว่างคำ เช่น ดาด่าด้าด๊าด๋า เป็นต้น

จากผลการทดลองพบว่าทั้งวิธีวัดระยะทางแบบไดนามิกไทม์วอร์ปปิงและวิธีวัดระยะทางแบบยูคลิดสามารถแยกประเภทเสียงวรรณยุกต์ลักษณะที่ 2 ได้ใกล้เคียงกัน โดยสามารถแสดงผลลัพธ์แยกตามประเภทวรรณยุกต์ได้ดังตารางที่ ก-2

ตารางที่ ก-2 ผลการจำแนกประเภทเสียงวรรณยุกต์ลักษณะที่ 2

เสียงวรรณยุกต์	ความถูกต้อง (%)					
	การหาระยะห่างแบบยูคลิด			วิธีวัดระยะทางแบบไดนามิกไทม์วอร์ปปิง		
	ค่าความแม่นยำ	ค่าความระลึก	ค่ามาตรวัด F	ค่าความแม่นยำ	ค่าความระลึก	ค่ามาตรวัด F
สามัญ	100	100	100	100	100	100
เอก	83.33	100	90.91	100	100	100
โท	50	100	66.66	62.50	100	76.92
ตรี	83.33	100	90.91	62.50	100	76.92
จัตวา	83.33	100	90.91	62.50	100	76.92
เฉลี่ย	**80.00**	**100**	**87.88**	**77.50**	**100**	**86.15**

ก.1.3 การทดลองจำแนกประเภทเสียงวรรณยุกต์ลักษณะที่ 3

เป็นการทดลองค้นคืนข้อมูลเสียงจากแฟ้มข้อมูลเสียงที่เป็นคำผสมมากกว่า 1 พยางค์ และมีช่วงห่างระหว่างคำประมาณ 0.5 – 1 วินาที เช่น ต้องการ ถูกตัด อัตโนมัติ เป็นต้น ซึ่งคำที่ใช้เป็นข้อคำถามเสียงคือ คำว่า ต้องการ ซึ่งเป็นการผสมระหว่างเสียงวรรณยุกต์โทกับสามัญ และเป็นคำที่ให้ค่าความถูกต้องมากสุด

ตารางที่ ก-3 เปรียบเทียบความถูกต้องระหว่างวิธีวัดระยะทางแบบไดนามิกไทม์วอร์ปปิงและวิธีวัดระยะทางแบบยูคลิด

วิธี	ความถูกต้อง (%)		
	ค่าความแม่นยำ (Precision)	ค่าความระลึก (Recall)	F-Measure
ยูคลิคลิด	68.42	81.25	74.29
ไดนามิกไทม์วอร์ปปิง	80.00	75.00	77.42

จากตารางที่ ก-3 สามารถสรุปได้ว่า วิธีวัดระยะทางแบบไดนามิกไทม์วอร์ปปิงสามารถค้นคืนข้อมูลเสียงจากแฟ้มข้อมูลเสียงภาษาไทยด้วยข้อคำถามเสียงได้ถูกต้องกว่าวิธีการหาระยะห่างแบบยูคลิด ซึ่งให้ค่าความถูกต้องเท่ากับ 77.42 เปอร์เซ็นต์ จากผลการทดลองที่กล่าวมา ผู้เขียนจึงเลือกใช้วิธีวัดระยะทางแบบไดนามิกไทม์วอร์ปปิงในการทำการทดลองต่อไป

ก.1.4 การเก็บข้อมูล

ปัจจุบันแฟ้มข้อมูลเสียงทำการเก็บข้อมูลอยู่ที่ 16 บิต แต่เนื่องจากงานวิจัยชิ้นนี้มุ่งเน้นที่ความเร็วในการค้นคืนข้อมูลเสียง ผู้เขียนจึงตั้งข้อสันนิษฐานว่า การเก็บข้อมูลที่น้อยกว่า 16 บิต จะทำให้การค้นคืนข้อมูลเสียงเร็วยิ่งขึ้นและความถูกต้องยังอยู่ในระดับดี ดังนั้นผู้เขียนจึงได้ทำการลดการเก็บข้อมูลเหลือ 8 บิต ในการทดลองผู้เขียนเลือกใช้คำหลายพยางค์ซึ่งเป็นการผสมเสียงข้ามระดับเสียง ซึ่งได้ผ่านการทดลองมาแล้วว่าให้ผลลัพธ์ในการค้นหาดีกว่าคำพยางค์เดียวเป็นข้อคำถามเสียง คำที่ผู้เขียนเลือกใช้คือคำว่า ต้องการ

ตารางที่ ก-4 เปรียบเทียบความถูกต้องผลการลดขนาดการเก็บข้อมูลจาก 16 บิต เหลือ 8 บิต

การเก็บข้อมูล (บิต)	ความถูกต้อง (%)		
	ค่าความแม่นยำ (Precision)	ค่าระลึก (Recall)	F-Measure
8	75.00	75.00	75.00
16	80.00	75.00	77.42

จากตารางที่ ก-4 พบว่าการเก็บข้อมูลขนาด 16 บิต สามารถค้นคืนข้อมูลเสียงจากแฟ้มข้อมูลเสียงได้ถูกต้องมากกว่าการเก็บข้อมูลขนาด 8 บิต ซึ่งให้ค่าความถูกต้องเท่ากับ 77.42 เปอร์เซ็นต์ โดยที่เวลาที่ใช้ในการค้นคืนไม่ต่างกันมากนัก จากผลการทดลองที่กล่าวมาผู้เขียนจึงเลือกใช้การเก็บข้อมูลขนาด 16 บิต ในการทำการทดลองต่อไป

ก.1.5 การแปลงข้อมูลให้เป็นบรรทัดฐานด้วย ค่าคะแนนมาตรฐานซี (Z-Score Normalization)

การทดลองนี้ เป็นการเปรียบเทียบผลของการทำการแปลงข้อมูลให้เป็นบรรทัดฐานด้วยค่าคะแนนมาตรฐานซี (Z-Score Normalization) ว่ามีผลต่อการค้นคืนเสียงจากแฟ้มข้อมูลเสียงภาษาไทยด้วยข้อคำถามเสียงหรือไม่ ในการทดลองผู้เขียนเลือกใช้คำหลายพยางค์ซึ่งเป็นการผสมเสียงข้ามระดับสียง ซึ่งได้ผ่านการทดลองมาแล้วว่าให้ผลลัพธ์ในการค้นหาดีกว่าคำพยางค์เดียว เป็นข้อคำถามเสียง คำที่ผู้เขียนเลือกใช้คือคำว่า ต้องการ

ตารางที่ ก-5 เปรียบเทียบผลของการแปลงข้อมูลกับไม่มีการแปลงข้อมูลให้เป็นบรรทัดฐานด้วยค่าคะแนนมาตรฐานซี

วิธี	ความถูกต้อง (%)		
	ค่าความแม่นยำ (Precision)	ค่าระลึก (Recall)	F-Measure
ทำการแปลงข้อมูลให้เป็นบรรทัดฐานด้วยค่าคะแนนมาตรฐานซี	80.00	75.00	77.42
ไม่มีการแปลงข้อมูลให้เป็นบรรทัดฐานด้วยค่าคะแนนมาตรฐานซี	62.50	62.50	62.50

จากตารางที่ ก-5 พบว่าเมื่อทำการแปลงข้อมูลให้เป็นบรรทัดฐานด้วยค่าคะแนนมาตรฐานซี ทำให้สามารถค้นคืนข้อมูลเสียงจากแฟ้มข้อมูลเสียงได้ถูกต้องมากขึ้น ซึ่งให้ค่าความถูกต้องเท่ากับ 77.42 เปอร์เซ็นต์ จากผลการทดลองที่กล่าวมาผู้เขียนจึงเลือกทำการแปลงข้อมูลให้เป็นบรรทัดฐานด้วยค่าคะแนนมาตรฐานซี ในการทำการทดลองต่อไป

ก.1.6 การเลื่อนกรอบหน้าต่าง

การทดลองนี้เป็นการทดลองเกี่ยวกับการหาระยะของการเลื่อนกรอบหน้าต่างที่เหมาะสม โดยวัตถุประสงค์หลักเพื่อที่จะหาระยะที่ดีที่สุด ที่ทำให้ผลการค้นคืนข้อมูลเสียงจากแฟ้มข้อมูลเสียงภาษาไทยด้วยข้อคำถามเสียงมีค่าความถูกต้องแม่นยำสูงสุด ในการทดลองผู้เขียนเลือกใช้คำ

หลายพยางค์ซึ่งเป็นการผสมเสียงข้ามระดับเสียง ซึ่งได้ผ่านการทดลองมาแล้วว่าให้ผลลัพธ์ในการค้นหาดีกว่าคำพยางค์เดียวเป็นข้อคำถามเสียง คำที่ผู้เขียนเลือกใช้คือคำว่า ต้องการ

ตารางที่ ก-6 เปรียบเทียบระยะการเลื่อนกรอบหน้าต่าง

ระยะการเลื่อนกรอบหน้าต่าง (Window Shift)	ความถูกต้อง (%)		
	ค่าความแม่นยำ (Precision)	ค่าระลึก (Recall)	F-Measure
1 ใน10 ของข้อคำถามเสียง	62.50	62.50	62.50
ครึ่งหนึ่งของข้อคำถามเสียง	80.00	75.00	77.42
เท่ากับขนาดของข้อคำถามเสียง	76.47	72.50	74.43

จากตารางที่ ก-6 พบว่าระยะการเลื่อนกรอบหน้าต่างเป็นครึ่งหนึ่งของข้อคำถามเสียง ทำให้สามารถค้นคืนข้อมูลเสียงจากแฟ้มข้อมูลเสียงได้ถูกต้องมากที่สุด ซึ่งมีค่าความถูกต้องเท่ากับ 77.42 เปอร์เซ็นต์ จากผลการทดลองที่กล่าวมาผู้เขียนจึงเลือกระยะการเลื่อนกรอบหน้าต่างเป็นครึ่งหนึ่งของข้อคำถามเสียง ในการทำการทดลองต่อไป จากผลการทดลองที่ผ่านมา ปรากฏว่าการแปลงฟูเรียร์อย่างเร็วให้ค่าความแม่นยำในระดับดีในการค้นคืนข้อมูลเสียงในกรณีที่ผู้พูดข้อคำถามเสียงกับผู้บันทึกเสียงเป็นคนเดียวกัน แต่ให้ผลไม่ดีในการค้นคืนข้อมูลเสียงในกรณีที่ผู้พูดข้อคำถามเสียงกับผู้บันทึกเสียงเป็นคนละคนกัน ดังนั้นผู้เขียนจึงทดลองใช้ค่าความถี่มูลฐานในการทดลองการค้นคืนข้อมูลแทนการแปลงฟูเรียร์อย่างเร็ว

ก.2 การค้นคืนข้อมูลเสียงจากแฟ้มข้อมูลเสียงภาษาไทยด้วยวิธีวัดระยะทางแบบไดนามิกไทม์วอร์ปปิง โดยใช้ค่าความถี่มูลฐาน (Fundamental Frequency)

ในภาษาไทยนั้นมีวรรณยุกต์ด้วยกัน 5 ระดับเสียง ซึ่งคำที่มีเสียงวรรณยุกต์ต่างกัน จะมีลักษณะของความถี่มูลฐานต่างกันด้วย ดังนั้นผู้เขียนจึงนำค่าความถี่มูลฐานมาใช้ในการแยกประเภทเสียงวรรณยุกต์

จากเหตุผลที่กล่าวมาข้างต้น ผู้เขียนจึงทำการศึกษาและทดลองเกี่ยวกับผลของค่าความถี่มูลฐานที่มีต่อการค้นคืนข้อมูลเสียง จากการศึกษาผู้เขียนได้ทำการทดลองค้นคืนข้อมูลเสียงจากแฟ้มข้อมูลเสียงด้วยวิธีวัดระยะทางแบบไดนามิกไทม์วอร์ปปิง ซึ่งทำการทดลองดังนี้

ก.2.1 การทดลองจำแนกประเภทเสียงวรรณยุกต์ลักษณะที่ 1

เป็นการทดลองค้นคืนข้อมูลเสียงจากแฟ้มข้อมูลเสียงที่เป็นคำพยางค์เดียวและมีช่วงห่างระหว่างคำประมาณ 0.5 - 1 วินาที เช่นคำว่า ตาม อ่าน ก้าม ป๊า จ๋า เป็นต้น ในส่วนของข้อคำถามเสียงที่ใช้ในการค้นคืนข้อมูลเสียงจะเป็นเสียงของบุคคลอื่น (Speaker Independent)

ตารางที่ ก-7 ผลการจำแนกประเภทเสียงวรรณยุกต์ลักษณะที่ 1

เสียงวรรณยุกต์	ความถูกต้อง (%)		
	ค่าความแม่นยำ (Precision)	ค่าความระลึก (Recall)	ค่ามาตรวัด F (F-Measure)
สามัญ	40.00	100	57.14
เอก	28.57	100	44.44
โท	28.57	100	44.44
ตรี	18.18	100	30.77
จัตวา	8.70	100	16.01
เฉลี่ย	**24.80**	**100**	**38.56**

จากตารางที่ ก-7 พบว่าสามารถค้นคืนข้อมูลเสียงได้ครบทุกคำ ที่เป็นคำเดียวกับข้อคำถามเสียงนั้น ซึ่งได้ค่าความถูกต้องเฉลี่ยเท่ากับ 38.56 เปอร์เซ็นต์ นอกจากนี้ยังพบว่าคำอื่นที่ค้นคืนออกมาด้วยนั้น เป็นคำที่มีระดับสัญญานเสียง (tone) เดียวกับข้อคำถามเสียง

ก.2.2 การทดลองจำแนกประเภทเสียงวรรณยุกต์ลักษณะที่ 2

เป็นการทดลองค้นคืนข้อมูลเสียงจากแฟ้มข้อมูลเสียงที่เป็นคำผสมมากกว่า 1 พยางค์ และมีช่วงห่างระหว่างคำประมาณ 0.5 – 1 วินาที เช่น จะเห็นได้ว่า ได้ว่า อัตโนมัติ จุดมุ่งหมาย ซอฟต์แวร์ ธรรมชาติ เป็นต้น ซึ่งคำที่ใช้เป็นข้อคำถามเสียงคือ คำว่า อัตโนมัติ ซึ่งเป็นเสียงของผู้พูดคนละคนกับเสียงในแฟ้มข้อมูลเสียง (Speaker Independent)

ตารางที่ ก-8 ผลการจำแนกประเภทเสียงวรรณยุกต์ลักษณะที่ 2

คำที่ต้องการค้นหา	ความถูกต้อง (%)		
	ค่าความแม่นยำ (Precision)	ค่าความระลึก (Recall)	ค่ามาตรวัด F (F-Measure)
อัตโนมัติ	100	100	100

จากตารางที่ ก-8 พบว่าสามารถค้นคืนข้อมูลเสียงได้ครบทุกคำ ที่เป็นคำเดียวกับข้อคำถามเสียงนั้น ซึ่งได้ค่าความถูกต้องเท่ากับ 100 เปอร์เซ็นต์

ก.2.3 การทดลองจำแนกประเภทเสียงวรรณยุกต์ลักษณะที่ 3

การทดลองนี้ ทำเพื่อหาค่าช่วงเวลา (Time Step) ที่เหมาะสมที่สุด เพื่อให้การค้นคืนข้อมูลเสียงจากแฟ้มข้อมูลเสียงด้วยข้อคำถามเสียงมีความถูกต้องมากที่สุด ในการทดลองนี้แฟ้มข้อมูลเสียงเป็นการบันทึกเสียงอย่างต่อเนื่อง ความยาวประมาณ 3 นาที โดยคำที่ใช้เป็นข้อคำถามเสียง คือ คำว่า อัตโนมัติ ซึ่งเป็นเสียงของผู้พูดคนละคนกับเสียงในแฟ้มข้อมูลเสียง (Speaker Independent)

ตารางที่ ก-9 การเปรียบเทียบช่วงเวลา (Time Step) ที่ใช้ในการสกัดค่าความถี่มูลฐาน

ข้อคำถามเสียง	ความถูกต้อง (%)					
	ช่วงเวลา (Time Step) เท่ากับ 0.01s			ช่วงเวลา (Time Step) เท่ากับ 0.05s		
	ค่าความแม่นยำ	ค่าความระลึก	ค่ามาตรวัด F	ค่าความแม่นยำ	ค่าความระลึก	ค่ามาตรวัด F
อัตโนมัติ	22.22	100	36.36	18.18	100	30.76

จากตารางที่ ก-9 พบว่าค่าช่วงเวลาทั้งสองค่าสามารถค้นคืนข้อมูลเสียงได้ครบทุกคำ ที่เป็นคำเดียวกับข้อคำถามเสียงนั้น แต่ที่ช่วงเวลา 0.01 วินาที ให้ค่าความถูกต้องมากกว่าช่วงเวลา 0.05 วินาที ซึ่งได้ค่าความถูกต้องเท่ากับ 36.36 เปอร์เซ็นต์

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก ข

การแจกแจงอัตราความถูกต้องการค้นคืนข้อมูลในการทดลองชุดที่ 2

การแจกแจงอัตราความถูกต้องการค้นคืนข้อมูลในการทดลองชุดที่ 2 ประกอบด้วยตารางที่ ข-1 ถึง ตารางที่ ข-10 ซึ่งตารางที่ ข-1 ถึง ตารางที่ ข-5 เป็นการคิดอัตราค่าความถูกต้องของการค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงสื่อการเรียนการสอนอิเล็กทรอนิกส์ ของผู้พูดคนที่ 1 ถึง ผู้พูดคนที่ 5 ส่วนตารางที่ ข-6 ถึง ตารางที่ ข-10 เป็นการค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงจากอินเทอร์เน็ตของผู้พูดคนที่ 1 ถึง ผู้พูดคนที่ 5 โดยในการค้นคืนนั้นจะทำการค้นคืนที่ K = 5 ถึง 35 รายการแรก

ตารางที่ ข-1 อัตราค่าความถูกต้องของการค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงสื่อการเรียนการสอนอิเล็กทรอนิกส์ ของผู้พูดคนที่ 1 ซึ่งเป็นเพศหญิง

ข้อคำถามเสียง	จำนวนข้อคำถามเสียงในฐานข้อมูลทั้งหมด	คะแนน (0 หรือ 1)						
		5 รายการแรก	10 รายการแรก	15 รายการแรก	20 รายการแรก	25 รายการแรก	30 รายการแรก	35 รายการแรก
DSS	112	1	1	1	1	1	1	1
โครงสร้างข้อมูล	4	1	1	1	1	1	1	1
การกู้ข้อมูล	3	0	1	1	1	1	1	1
การควบคุม	3	0	0	1	1	1	1	1
การตัดสินใจ	109	1	1	1	1	1	1	1
การสำรองข้อมูล	5	0	0	0	0	0	1	1
Software	8	1	1	1	1	1	1	1
บุคลากร	15	1	1	1	1	1	1	1
ระบบจัดการฐานข้อมูล	2	0	0	0	0	0	0	0
วิธีการประมวลผล	1	0	0	0	0	0	0	0
เฉลี่ย (%)		50	60	70	70	70	80	80

ตารางที่ ข-2 อัตราค่าความถูกต้องของการค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงสื่อการเรียนการสอนอิเล็กทรอนิกส์ ของผู้พูดคนที่ 2 ซึ่งเป็นเพศหญิง

ข้อคำถามเสียง	จำนวนข้อคำถามเสียงในฐานข้อมูลทั้งหมด	คะแนน (0 หรือ 1)						
		5 รายการแรก	10 รายการแรก	15 รายการแรก	20 รายการแรก	25 รายการแรก	30 รายการแรก	35 รายการแรก
แฟ้มข้อมูล	9	0	0	0	0	0	1	1
กระบวนการในการตัดสินใจ	1	1	1	1	1	1	1	1
การคัดเลือก	3	0	0	0	0	0	0	0
การสำรอง	8	1	1	1	1	1	1	1
การสืบค้นข้อมูล	1	0	0	0	0	0	0	0
ฐานข้อมูล	82	0	1	1	1	1	1	1
ภาษาคำนิยามข้อมูล	1	0	0	0	0	0	0	0
ลดความยุ่งยาก	1	1	1	1	1	1	1	1
หน้าที่ทางการจัดการ	1	0	0	0	1	1	1	1
อุปกรณ์สื่อสาร	2	0	0	0	0	0	1	1
เฉลี่ย (%)		**30**	**40**	**40**	**50**	**50**	**70**	**70**

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ ข-3 อัตราค่าความถูกต้องของการค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงสื่อการเรียนการสอนอิเล็กทรอนิกส์ ของผู้พูดคนที่ 3 ซึ่งเป็นเพศชาย

ข้อคำถามเสียง	จำนวนข้อคำถามเสียงในฐานข้อมูลทั้งหมด	คะแนน (0 หรือ 1)						
		5 รายการแรก	10 รายการแรก	15 รายการแรก	20 รายการแรก	25 รายการแรก	30 รายการแรก	35 รายการแรก
เทคโนโลยี GDSS	1	0	0	1	1	1	1	1
โครงสร้างข้อมูล	4	1	1	1	1	1	1	1
การกู้คืนข้อมูล	5	1	1	1	1	1	1	1
การตัดสินใจ	109	1	1	1	1	1	1	1
การบริหารข้อมูล	2	0	0	0	1	1	1	1
การสำรองข้อมูล	5	1	1	1	1	1	1	1
การสืบค้นข้อมูล	1	0	1	1	1	1	1	1
การออกแบบฐานข้อมูล	9	0	1	1	1	1	1	1
ฐานข้อมูล	82	1	1	1	1	1	1	1
ร่วมแก้ปัญหา	1	0	0	0	0	0	0	0
เฉลี่ย (%)		**50**	**70**	**80**	**90**	**90**	**90**	**90**

ตารางที่ ข-4 อัตราค่าความถูกต้องของการค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงสื่อการเรียนการสอนอิเล็กทรอนิกส์ ของผู้พูดคนที่ 4 ซึ่งเป็นเพศชาย

ข้อคำถามเสียง	จำนวนข้อคำถามเสียงในฐานข้อมูลทั้งหมด	คะแนน (0 หรือ 1)						
		5 รายการแรก	10 รายการแรก	15 รายการแรก	20 รายการแรก	25 รายการแรก	30 รายการแรก	35 รายการแรก
การกู้ฐานข้อมูล	1	0	0	0	0	0	0	0
การตัดสินใจ	109	1	1	1	1	1	1	1
ความซ้ำซ้อนของข้อมูล	2	0	0	0	0	0	0	1
ชุดคำสั่ง	14	1	1	1	1	1	1	1
ฐานแบบจำลอง	4	0	0	0	0	0	1	1
ตัวอักขระ	3	0	0	1	1	1	1	1
ผู้บริหารฐานข้อมูล	3	0	0	0	0	0	1	1
พจนานุกรม	3	1	1	1	1	1	1	1
มินิคอมพิวเตอร์	2	0	0	0	0	0	0	0
ระบบรักษาความปลอดภัย	2	0	0	0	0	0	0	0
เฉลี่ย (%)		**30**	**30**	**40**	**40**	**40**	**60**	**70**

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ ข-5 อัตราค่าความถูกต้องของการค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงสื่อการเรียนการสอนอิเล็กทรอนิกส์ ของผู้พูดคนที่ 5 ซึ่งเป็นเพศชาย

ข้อคำถามเสียง	จำนวนข้อคำถามเสียงในฐานข้อมูลทั้งหมด	คะแนน (0 หรือ 1)						
		5 รายการแรก	10 รายการแรก	15 รายการแรก	20 รายการแรก	25 รายการแรก	30 รายการแรก	35 รายการแรก
ข้อเสียของการจัดการฐานข้อมูล	1	0	0	0	0	1	1	1
ข้อดีของการจัดการฐานข้อมูล	1	0	0	0	1	1	1	1
ความสามารถในการเข้าถึงข้อมูล	2	0	1	1	1	1	1	1
ฐานข้อมูล	82	1	1	1	1	1	1	1
ตัวอย่าง	6	0	1	1	1	1	1	1
บุคลากร	6	1	1	1	1	1	1	1
ประสิทธิผล	4	0	0	0	0	1	1	1
วัตถุประสงค์	5	0	0	0	0	0	1	1
อุปนัย	3	0	0	1	1	1	1	1
เฉลี่ย (%)		30	50	60	70	90	100	100

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ ข-6 อัตราค่าความถูกต้องของการค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงจากอินเทอร์เน็ต ของผู้พูดคนที่ 1 ซึ่งเป็นเพศหญิง

ข้อคำถามเสียง	จำนวนข้อคำถามเสียงในฐานข้อมูลทั้งหมด	คะแนน (0 หรือ 1)						
		5 รายการแรก	10 รายการแรก	15 รายการแรก	20 รายการแรก	25 รายการแรก	30 รายการแรก	35 รายการแรก
ขาวงามแบบบริสุทธิ์	1	0	1	1	1	1	1	1
งามแบบอ่อนหวาน	1	0	0	0	0	0	0	1
นิยามของความงาม	1	1	1	1	1	1	1	1
บทสำรวจตนเอง	1	0	0	0	0	0	0	1
ปฏิมากร (ประติมากร)	2	1	1	1	1	1	1	1
พระฉวี	2	0	0	0	0	0	0	1
มหากุศลกรรม	1	1	1	1	1	1	1	1
อาการทางใจและวิธีคิด	3	0	0	0	0	0	1	1
สรุป	3	1	1	1	1	1	1	1
หูบาง	1	0	0	0	0	0	0	0
เฉลี่ย (%)		**40**	**50**	**50**	**50**	**50**	**60**	**90**

ตารางที่ ข-7 อัตราค่าความถูกต้องของการค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงจากอินเทอร์เน็ต ของผู้พูดคนที่ 2 ซึ่งเป็นเพศหญิง

ข้อคำถามเสียง	จำนวนข้อคำถามเสียงในฐานข้อมูลทั้งหมด	คะแนน (0 หรือ 1)						
		5 รายการแรก	10 รายการแรก	15 รายการแรก	20 รายการแรก	25 รายการแรก	30 รายการแรก	35 รายการแรก
งามแบบแปลกประหลาด	1	0	1	1	1	1	1	1
งามแบบสง่า	1	0	0	0	1	1	1	1
ฉวีวรรณ	1	0	0	0	0	0	0	0
ลักษณะขัดแย้ง	1	0	0	0	0	0	0	1
ทำทานด้วยศรัทธา	4	0	0	0	0	1	1	1
บทสำรวจตนเอง	1	0	0	0	0	0	0	0
พระเนตร	3	0	0	0	0	0	0	0
พระทนต์	4	0	1	1	1	1	1	1
มหากุศลกรรม	1	0	0	1	1	1	1	1
อาการทางใจและวิธีคิด	3	0	0	0	1	1	1	1
เฉลี่ย (%)		**0**	**20**	**30**	**50**	**60**	**60**	**70**

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ ข-8 อัตราค่าความถูกต้องของการค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงจากอินเทอร์เน็ต ของผู้พูดคนที่ 3 ซึ่งเป็นเพศชาย

ข้อคำถามเสียง	จำนวนข้อคำถามเสียงในฐานข้อมูลทั้งหมด	คะแนน (0 หรือ 1)						
		5 รายการแรก	10 รายการแรก	15 รายการแรก	20 รายการแรก	25 รายการแรก	30 รายการแรก	35 รายการแรก
วิบากกรรม	1	0	0	0	0	0	0	0
โทสะควบคุมจิตใจ	1	0	0	0	0	0	0	0
ให้อภัยเป็นทาน	2	0	0	0	0	0	0	0
งามแบบเร้าความรู้สึกทางเพศ	1	0	0	0	0	0	0	0
งามแบบแปลกประหลาด	1	0	0	0	0	0	0	0
งามแบบสง่า	1	0	0	1	1	1	1	1
ตำราทายมนุษย์	1	0	0	0	0	0	0	0
ทำทานด้วยศรัทธา	4	0	0	0	1	1	1	1
รักษาศีล	11	0	1	1	1	1	1	1
รูปโฉม	8	0	0	0	0	0	1	1
เฉลี่ย (%)		**0**	**10**	**20**	**30**	**30**	**40**	**40**

ตารางที่ ข-9 อัตราค่าความถูกต้องของการค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงจากอินเทอร์เน็ต ของผู้พูดคนที่ 4 ซึ่งเป็นเพศชาย

ข้อคำถามเสียง	จำนวนข้อคำถามเสียงในฐานข้อมูลทั้งหมด	คะแนน (0 หรือ 1)						
		5 รายการแรก	10 รายการแรก	15 รายการแรก	20 รายการแรก	25 รายการแรก	30 รายการแรก	35 รายการแรก
แปลกประหลาด	2	0	0	0	1	1	1	1
ทำทานด้วยศรัทธา	4	0	0	0	1	1	1	1
งามแบบสง่า	1	0	0	0	0	0	1	1
ทุคติ	2	0	0	0	0	0	0	1
บทสำรวจตนเอง	1	0	0	0	0	0	0	0
ผู้มีรูปงาม	4	0	0	1	1	1	1	1
มหากุศลกรรม	1	0	0	0	1	1	1	1
รักษาศีล	11	1	1	1	1	1	1	1
ลักษณะขัดแย้ง	1	0	0	0	0	0	0	1
อาการทางใจ	5	0	0	1	1	1	1	1
เฉลี่ย (%)		**10**	**10**	**30**	**60**	**60**	**70**	**90**

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ ข-10 อัตราค่าความถูกต้องของการค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงจากอินเทอร์เน็ต ของผู้พูดคนที่ 5 ซึ่งเป็นเพศชาย

ข้อคำถามเสียง	จำนวนข้อคำถามเสียงในฐานข้อมูลทั้งหมด	คะแนน (0 หรือ 1)						
		5 รายการแรก	10 รายการแรก	15 รายการแรก	20 รายการแรก	25 รายการแรก	30 รายการแรก	35 รายการแรก
ลักษณะขัดแย้ง	1	0	0	0	0	0	0	1
งามแบบอ่อนหวาน	1	0	0	0	0	0	0	0
ทำทานด้วยศรัทธา	4	0	0	0	0	0	1	1
พระประธาน	1	0	0	0	0	0	0	0
พระพุทธเจ้า	6	0	0	0	0	1	1	1
สรุป	3	0	1	1	1	1	1	1
รักใคร่เมตตา	1	0	0	0	0	0	0	0
รักษาศีล	11	1	1	1	1	1	1	1
ศีลธรรม	2	0	1	1	1	1	1	1
มหากุศลกรรม	1	0	0	0	0	1	1	1
เฉลี่ย (%)		**10**	**30**	**30**	**30**	**50**	**60**	**70**

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก ค

ผลงานตีพิมพ์

งานประชุมวิชาการนานาชาติห้องสมุดดิจิทัลอาเซียน ครั้งที่ 9 (9th International Conference on Asian Digital Libraries (ICADL)) ระหว่างวันที่ 27 - 30 พฤศจิกายน 2549 ณ มหาวิทยาลัยเกียวโต เมืองเกียวโต ประเทศญี่ปุ่น ในบทความเรื่อง Speech Audio Retrieval Using Voice Query ตีพิมพ์ในวารสาร Lecture Notes In Computer Science (LNCS) Vol.4312: 494 – 497, Springer-Verlag Berlin Heidelberg, 2006

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Speech Audio Retrieval Using Voice Query

Chotirat Ann Ratanamahatana and Phubes Tohlong

Dept. of Computer Engineering, Chulalongkorn University, Bangkok 10330 Thailand
ann@cp.eng.chula.ac.th, Tohlong@hotmail.com

Abstract. Multimedia data has increasingly become a prevalent resource in Digital Library system; this includes audio, video, and image archives. However, each type of these data may need specific tools to help facilitate effective and efficient retrieval tasks. In this paper, we focus on retrieval of speech audio collection, which includes audio books, speech recordings, interviews, and lectures. Currently, most of the audio retrieval systems are based on keyword/title/author search typed into the system by users. The system then searches for particular keywords and gives a list of entire audio files that potentially are relevant to the query. Nonetheless, browsing audio content for particular section of the audios without knowing the actual content is yet a very difficult task. Moreover, since audio transcription or keyword annotation is very labor intensive and becomes infeasible for large data, we introduce here a preliminary framework that locates subsections of the audio that correspond to the voice query made by a user. We demonstrate a utility of our approach on query retrieval tasks in various types of audio recordings. We also show that this simple framework can potentially help retrieve and locate the voice query within the audio accurately and efficiently.

Keywords: Audio retrieval, time series, query by example, voice search.

1 Introduction

Speech processing has established itself in research communities since 1950s. However, it still has many unsolved problems and remains a very challenging and active area of research nowadays due to its exceptionally complex nature of the problem itself. Many speech processing techniques have been proposed for speech audio retrieval [1][6][7]. In spite of this, none of them have really solved a problem of searching the actual content within large retrieved audio files. Instead, the search processes usually are text-based or voice query, searching for entire audio files according to provided titles, authors, and keywords [1][3]. Some system may need manual transcription of the speech/audio into text before searching can be performed. It would be very helpful and more convenient if we can search *any* part of the speech audio using our voice as a query without having to do the transcription. This paper proposes a preliminary alternative to textual annotations, which is based on time series features extracted from the raw speech data.

S. Sugimoto et al. (Eds.): ICADL 2006, LNCS 4312, pp. 494 – 497, 2006.
© Springer-Verlag Berlin Heidelberg 2006

1.1 Motivation

Our motivation started from an attempt to search the recorded lectures archived in the digital library collection. At this point, we can generally search for audio files as a whole, based on keywords, titles, and authors. However, if the retrieved audio file is very long, it is still extremely hard to browse or locate the exact content within the audio, when we are interested only in some parts of the recording; it is more likely that a user would like to only hear subsections within the audio file instead of having to listen to the whole audio from the beginning.

This work is based on a query-by-example (QBE) technique, where users provide voice/speech examples of the word or phrase they seek. Some may argue that this query-by-example approach has a major limitation when the users really want to search for semantic concepts rather than the "exact" word or phrase; rather, query by keyword approaches may be more appropriate. Since these types of research have been a research of interest within speech communities for years and still have not been considered a completely solved problem, we are taking this opportunity to explore an alternative in approaching the problem without using the full speech processing techniques. We would like to be able to search *inside* each audio file to locate the exact content that we want based on a given voice query.

2 Time Series Representation for Voice Searching

More complex analysis of the speech audio cannot be achieved by looking at the raw audio plots alone. To get some information about the frequency distribution, harmonics, and others, some signal processing such as Fourier analysis is needed. In this work, we propose a simple approach to approximately represent audio features using time series representation, an approach recently used in query by humming system [4][5]. Note that by looking at the raw audio plot (.WAV file), we can extract several features, such as volume from the amplitude and the "timbre" of the voice. However, these characteristics are irrelevant to the task of differentiating one word from another. Instead, we propose to simply use the frequencies information as approximations of words in the audio. Note that the effectiveness and accuracy of this approach essentially depend on the nature of the spoken languages themselves as well. In this work, we test our method in Thai language, a tonal language with 5 tones.

We start off with acquiring the Thai digital audio recording in WAV file format. In our experiment, all recordings are originally recorded at the sampling rate of 22,050 Hz, but we decide to downsample the data to only 2,000 Hz (16 bits mono) to significantly speed up the search process and make sure that we do not lose too much of important features during the reduction and calculation. In speech community, such sampling rate is considered unacceptably low; however, our proposed work has one big difference in the algorithm in that we process the speech in *word level*, instead of *phoneme level* as typically being done in speech processing. This in turn allows us to easily process a 1-hour audio which almost seems unfeasible if we were to employ a traditional automatic speech recognition process.

We then preprocess the data by transforming a raw audio into a frequency domain using Fast Fourier Transformation (FFT), which gives the frequency distribution information about the spoken word or subsequence of the recording. This is a time series to be later used in similarity search in our framework. In addition, to further

remove noise and outlier, we also apply some smoothing and z-score normalization to all datasets in our work before utilizing a Dynamic Time Warping distance measure (DTW) to locate the *K*-nearest neighbor query word within the given recording.

The algorithm is simply a subsequence matching using a sliding window of the size of the query window. Starting from the beginning of the recording until the end, it looks for the one with best match using a similarity measure. To simplify the implementation, a Euclidean distance metric could be used. However, we believe that a more sophisticated similarity measures, such as Dynamic Time warping [2][8], could significantly improve the accuracy of the result since it could gracefully resolve the problem of discrepancies or minor time variation in the time series, where we could intuitively map the time series query to the appropriate section of the recording.

3 Experimental Evaluation

We have put together a collection of various audio recordings for our experiment; some are audio books, and some are real lectures with both male and female speakers. Each one is approximately 45 to 60 minutes in length, with word content ranging from 6,000 to 9,000 words. We have chosen some words from each recording and exclusively removed those occurrences from the recording to avoid getting an exact match during the search. To evaluate the retrieval's effectiveness, we calculate the Precision, Recall, as well as the F-Measure to compare results among various parameter settings and approaches.

3.1 Experiment Results and Discussions

At this preliminary stage of our work, the evaluation process must be done manually. After the query words are selected, we have to actually listen to the whole recording and mark all the actual occurrences of each word within the recording, since there is no transcription available. The main contribution of our work is an ability to perform a voice search within a large audio file, where speech processing community may still have difficulties with. We demonstrate our utility by querying a word in an hour-long audio then measure the retrieval effectiveness both by looking at the precision/recall as well as the running time. Up to this point, we have demonstrated that Dynamic Time warping distance measure always outperforms the classic Euclidean distance metric in terms of the accuracy but with the price of higher time complexity.

In addition, we also consider another approach using Mel Frequency Cepstral Coefficients or MFCC that is regularly employed in speech processing to see if its superiority still holds for voice search in the word level. We first compare its time complexity with the Euclidean and Dynamic Time Warping distance measures. With exactly the same parameters and settings, MFCC measure is running 30 times slower

Table 1. Comparison of results between FFT with DTW and MFCC measures, showing that DTW gives more accurate results

Approach	Precision	Recall	F-Measure
FFT with DTW	80%	75%	77.42%
MFCC	61.11%	68.75%	64.71%

than Euclidean distance and about 5 times slower than the Dynamic Time warping. The retrieval's effectiveness between the two approaches is shown in Table 1.

Since the MFCC's running time is larger and its F-Measure is much lower, FFT with DTW distance measure is then employed in our experiments. With speaker-dependent experiment, as expected, we get much worse results; there are many more query words that were left undetected, as well as a lot more false alarms. Ideally, we would like to minimize the number of False Negatives as much as possible, with an acceptable number of False Positives. Looking closely, we found that the results are affected across genders as well. We look at the Fourier analysis of the same word spoken by different speakers and discover that they approximately have the similar shape but relatively shifted along the frequency axis. That means the structure of the word spoken are quite similar across the speakers, but the overall speaking frequency for each person differs and can be thought of as a frequency offset.

4 Conclusions and Future Work

In this preliminary work, we have proposed a simple approach to approximately represent speech audio features using time series representation, then to locate a voice query within the audio recordings. We have demonstrated the utility of our approach on query retrieval tasks for audio recordings in Thai language, i.e., to locate a voice query within the lecture recordings. From the experiment results, we have demonstrated that this simple framework can potentially help retrieve and locate the audio according to voice query inputs, especially in the speaker-dependent situation. Since the pitch discrepancies among speakers pose a limitation in our current framework, we need to look more closely into these features and see if any normalization among various speakers could be attained. Together with a Dynamic Time Warping distance measure as well as some lowerbounding and dimensionality reduction techniques, this could potentially resolve the problem and to help speed up the overall search process.

Referenes

[1] Franz, A. & Milch, B. (2002). Searching the Web by Voice. In Proceedings of COLING.
[2] Kruskall, J. B. & Liberman, M. (1983). The symmetric time warping algorithm: From continuous to discrete. In Time Warps, String Edits and Macromolecules.
[3] Klabbhankao, B. (2000). Online Information Retrieval Using Genetic Algorithms. NECTEC Technical Journal Vol 2, No.7. March-June.
[4] Zhu, Y., Shasha, D., & Zhao, X. (2003). Query by Humming – in Action with its Technology Revealed. ACM SIGMOD, June 9-12.
[5] Zhu, Y. & Shasha, D. (2003). Warping Indexes with Envelope Transforms for Query by Humming. ACM SIGMOD, June 9-12.
[6] Hazen, T.J., Saenko, K., La, C.-H., & Glass, J.R. (2004). A Segment-Based Audio-Visual Speech Recognizer: Data Collection, Development, and Initial Experiments. Proc. ICMI.
[7] Gutkin, A. & King, S. (2004). Structural Representation of Speech for Phonetic Classification. In Proc. 17th International Conference on Pattern Recognition (ICPR), volume 3, pages 438-441, Cambridge, UK, August 2004. IEEE Computer Society Press.
[8] Ratanamahatana, C.A. & Keogh, E. (2005). Three Myths about Dynamic Time Warping Data Mining. SIAM International Conference on Data Mining (SDM).

ประวัติผู้เขียนวิทยานิพนธ์

นายภูเบศ โต๊ะลง เกิดเมื่อวันที่ 19 มกราคม พ.ศ. 2526 ที่จังหวัดฉะเชิงเทรา สำเร็จการศึกษาหลักสูตรวิทยาศาสตรบัณฑิต (วท.บ.) เกียรตินิยมอันดับหนึ่ง (เหรียญเงิน) สาขาวิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยหัวเฉียวเฉลิมพระเกียรติ เมื่อปีการศึกษา 2547 และเข้าศึกษาต่อหลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาศาสตร์คอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย เมื่อปีการศึกษา 2548 ขณะศึกษาได้มีโอกาสไปเสนอผลงานเรื่อง Speech Audio Retrieval Using Voice Query ในงานประชุมวิชาการนานาชาติห้องสมุดดิจิทัลอาเซียน ครั้งที่ 9 (9th International Conference on Asian Digital Libraries (ICADL)) ณ มหาวิทยาลัยเกียวโต เมืองเกียวโต ประเทศญี่ปุ่น ปัจจุบันทำงานอยู่ที่ บริษัท ไอโซเน็ต จำกัด

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย