การเปรียบเทียบระบบค้นหาคำหลักสำหรับเสียงพูดภาษาไทย

A Comparison of Keyword Spotting Systems for Thai Speech

โนอาห์ กิจญะวงค์

Noah Kityawong

วิทยานิพนธ์นี้สำหรับการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต

สาขาวิชาวิศวกรรมคอมพิวเตอร์

มหาวิทยาลัยสงขลานครินทร์

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Engineering in Computer Engineering

Prince of Songkla University

2555

ชื่อวิทยานิพนธ์ การเปรียบเทียบระบบค้นหาคำหลักสำหรับเสียงพูดภาษาไทย
ผู้เขียน นายโนอาห์ กิจญะวงค์
สาขาวิชา วิศวกรรมคอมพิวเตอร์

อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

.....
(รองศาสตราจารย์ ดร.มนตรี กาญจนะเดชะ)

อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม

.....
(ดร.อนันท์ ชกสุริวงค์)

คณะกรรมการสอบ

.....ประธานกรรมการ
(ดร.นิคม สุวรรณวร)

.....กรรมการ
(รองศาสตราจารย์ ดร.มนตรี กาญจนะเดชะ)

.....กรรมการ
(ดร.อนันท์ ชกสุริวงค์)

.....กรรมการ
(ดร.ประกาศิต กายะสิทธิ์)

บัณฑิตวิทยาลัย มหาวิทยาลัยสงขลานครินทร์ อนุมัติให้นับวิทยานิพนธ์ฉบับนี้สำหรับการศึกษา ตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมคอมพิวเตอร์

.....
(ศาสตราจารย์ ดร.อมรรัตน์ พงศ์ดารา)
คณบดีบัณฑิตวิทยาลัย

ขอรับรองว่า ผลงานวิจัยนี้เป็นผลมาจาการศึกษาวิจัยของนักศึกษาเอง และขอขอบคุณผู้ที่มีส่วนเกี่ยวข้องทุกท่านไว้ ณ ที่นี้

ลงชื่อ.....
(รองศาสตราจารย์ ดร.มนตรี กาญจนะเดชะ)
อาจารย์ที่ปรึกษาวิทยานิพนธ์

ลงชื่อ.....
(นายโนอาห์ กิจญะวงค์)
นักศึกษา

ข้าพเจ้าขอรับรองว่า ผลงานวิจัยนี้ไม่เคยเป็นส่วนหนึ่งในการอนุมัติปริญญาในระดับใดมาก่อน และไม่ได้ถูกใช้ในการยื่นขออนุมัติปริญญาในขณะนี้

ลงชื่อ.....
(นายโนอาห์ กิจญะวงค์)
นักศึกษา

ชื่อวิทยานิพนธ์ การเปรียบเทียบระบบค้นหาคำหลักสำหรับเสียงพูดภาษาไทย
ผู้เขียน นายโนอาห์ กิจญะวงค์
สาขาวิชา วิศวกรรมคอมพิวเตอร์
ปีการศึกษา 2555

บทคัดย่อ

วิทยานิพนธ์ฉบับนี้เป็นงานวิจัยเพื่อต้องการที่จะศึกษาและเปรียบเทียบระบบค้นหาคำสำคัญบนสื่อเสียงพูดภาษาไทยที่มีระบบพื้นฐานแตกต่างกันสองระบบได้แก่ระบบที่มีพื้นฐานมาจากระบบรู้จำเสียงพูดซึ่งอาศัยหลักการของแบบจำลองฮิดเดนมาร์คอฟและระบบที่มีพื้นฐานมาจากวิธีการจำแนกประเภทโดยใช้เคอร์เนลและขอบเขตการแยกแยะที่มีขนาดกว้างที่สุดซึ่งใช้เทคนิคของซัพพอร์ตเวกเตอร์ แมชชีน และศึกษาและเปรียบเทียบค่าลักษณะเด่นของสัญญาณเสียงพูดที่แตกต่างกันบนระบบค้นหาคำสำคัญบนสื่อเสียงพูดทั้งสองประเภท ไม่ว่าจะเป็นค่าสัมประสิทธิ์ MFCC, ค่าสัมประสิทธิ์การประเมินค่าเชิงเส้น, ค่าสัมประสิทธิ์ DCTC หรือแม้แต่ค่าลักษณะเด่นของเสียงพูดที่ได้มาจากการปรับขนาดของฟีเจอร์เวกเตอร์ หรือค่าลักษณะเด่นที่ได้ปรับแต่งโดยใช้ตัวดำเนินการทางคณิตศาสตร์อย่างง่ายโดยค่าลักษณะเด่นของเสียงพูดทั้งหมดนั้นจะมีทั้งรูปแบบปกติและรูปแบบที่มีการเพิ่มค่าลักษณะพิเศษ (High Order Feature) เพื่อที่จะหาคู่ของระบบค้นหาคำสำคัญและค่าลักษณะเด่นของเสียงพูดที่ได้ผลลัพธ์ที่ดีที่สุดสำหรับเสียงภาษาไทย ซึ่งในการทดลองครั้งนี้ได้ทดสอบบนฐานข้อมูลเสียงภาษาไทย “LOTUS” และผลลัพธ์ที่ได้นั้นจะวัดผลในค่าของ AUC ซึ่งเป็นการเปรียบเทียบระหว่างอัตราค่าบวกที่แท้จริง (การชี้เฉพาะคำสำคัญที่ถูกต้อง) และฟังก์ชั่นของอัตราค่าบวกเท็จ (การชี้เฉพาะคำสำคัญผิดที่) จากการทดลองที่ได้จะเห็นว่า โดยส่วนใหญ่ ระบบค้นหาคำสำคัญที่อยู่บนพื้นฐานของระบบรู้จำเสียงพูดนั้นจะให้ผลลัพธ์ที่ดีกว่าโดยเฉพาะอย่างยิ่งเมื่อใช้ร่วมกับค่าสัมประสิทธิ์ MFCC ที่มีการเพิ่มค่าลักษณะพิเศษที่คำนวณโดยใช้หลักการของ DCTC ซึ่งให้ค่า AUC สูงสุดที่ 0.922 ในส่วนของการปรับแต่งค่าลักษณะเด่นนั้นจะเห็นได้ว่าการใช้ตัวดำเนินการทางคณิตศาสตร์อย่างง่าย โดยเฉพาะอย่างยิ่งการบวกและการลบระหว่างค่าลักษณะเด่น LPC และค่าลักษณะเด่น MFCC ให้ผลลัพธ์ที่ดีขึ้นอย่างมาก เมื่อเปรียบเทียบกับการปรับแต่งค่าลักษณะเด่นโดยใช้วิธีอื่นๆ

Thesis Title A Comparison of Keyword Spotting Systems for Thai Speech

Author Mr. Noah Kityawong

Major Program Computer Engineering

Academic Year 2012

ABSTRACT

This thesis presents a study and comparison of two keyword spotting systems for Thai speech. The first keyword spotting system is based on a HMM-based automatic speech recognizer, and the second is based on the large margin and kernel method which utilizes the support vector machine. This work also compares performance of several speech feature types, including Mel-frequency Cepstral Coefficients (MFCCs), Linear Prediction Coefficients (LPCs) and Discrete Cosine Transform Coefficients (DCTCs). Combinations of these features using a simple mathematic operation and higher order features have also been investigated.

The Thai speech corpus, LOTUS, was used to run experiments. System performance was measured in terms of area under curve (AUC) which is computed from the ratio between the true positive and the false negative.

Experimental results show that the HMM-based keyword spotting system outperformed the other method, especially when it was used with MFCC features that were temporally encoded using the discrete cosine transformation method. The highest AUC obtained was 0.922. In addition, applying addition and subtraction among LPC and MFCC features yielded higher performance improvement when compare to other methods.

กิตติกรรมประกาศ

สำหรับการดำเนินการวิจัยและจัดทำวิทยานิพนธ์นี้ ผู้วิจัยขอขอบพระคุณ รองศาสตราจารย์ ดร.มนตรี กาญจนะเดชะ ประธานกรรมการที่ปรึกษาวิทยานิพนธ์ ที่ได้ให้คำปรึกษา ชี้แนะแนวทางในการทำงาน ทั้งยังให้กำลังใจและการเอาใจใส่กับข้าพเจ้าเป็นอย่างมาก กระตุ้นให้ข้าพเจ้าได้มีความมุ่งมั่นในการทำงานให้สำเร็จลุล่วงไปได้เป็นอย่างดีเรื่อยมา รวมถึงการตรวจและแก้ไขเนื้อหาวิทยานิพนธ์ให้สำเร็จสมบูรณ์

ขอขอบพระคุณ ดร.อนันท์ ชกสุริวงค์ กรรมการที่ปรึกษาวิทยานิพนธ์ ที่ได้ให้คำแนะนำในการปรับปรุงวิทยานิพนธ์ให้สมบูรณ์ยิ่งขึ้น

ขอขอบพระคุณ Mr. Martin Woellmer ที่ได้ให้คำปรึกษาและให้ความช่วยเหลือในการทำวิทยานิพนธ์ครั้งนี้

ขอขอบพระคุณคณาจารย์ และบุคคลากรทุกท่านในภาควิชาวิศวกรรมคอมพิวเตอร์ทุกท่านที่ให้ความช่วยเหลือในระหว่างการทำวิทยานิพนธ์

ขอขอบพระคุณพี่ๆ เพื่อน ๆ และนักศึกษาปริญญาโททุกท่านที่คอยให้คำแนะนำและคอยให้ความช่วยเหลือด้วยดีตลอดมา

สุดท้ายนี้ ขอกราบขอบพระคุณ บิดา มารดาและญาติพี่น้องทุกท่าน ซึ่งเป็นผู้มีพระคุณสูงสุดที่ให้กำลังในและให้การสนับสนุนทุกสิ่งทุกอย่างด้วยดีตลอดมาในชีวิตของข้าพเจ้า

โนอาห์ กิจญะวงค์

สารบัญ

หน้า

รายการรูปภาพ

หน้า

รายการตาราง

หน้า

สารบัญคำศัพท์

การประมวลผลสัญญาณเสียงพูด	Speech Processing
ระบบรู้จำเสียงพูดต่อเนื่องที่ครอบคลุมคำศัพท์จำนวนมาก	Large Vocabulary Continuous Speech Recognition
แบบจำลองฮิดเดนมาร์คอฟ	Hidden Markov Model
แบบจำลองคำหลัก	Keyword Model
แบบจำลองพื้นภูมิ	Background Model
แบบจำลองขยะ	Garbage Model
การถอดรหัสแบบวิเทอบี	Viterbi Decoding
วิธีการจำแนกประเภทโดยใช้เคอร์เนลและขอบเขตการแยกแยะที่มีขนาดกว้างที่สุด	Large Margin and Kernel Method
เคอร์เนล	Kernel
ซัพพอร์ตเวกเตอร์แมชชีน	Support Vector Machine
หน่วยพื้นฐานของเสียงพยางค์	Phoneme
ค่าลักษณะเด่นของข้อมูลเสียงพูด	Speech Feature
ค่าสัมประสิทธิ์ความถี่สเกลเมล	Mel-Frequency Cepstral Coefficient
การรวมค่าลักษณะเด่นของสัญญาณเสียงพูดเข้าด้วยกัน	Speech Feature Combination
ตัวเลขของเมอริท	Figure of Merit
แบบจำลองตัวเติมที่ครอบคลุมทั้งหมด	Universal Filler Model
แบบจำลองของเสียงพยัญชนะและเสียงสระ	Consonant-vowel Model
แบบจำลองของหน่วยเสียงแต่ละตัว	Phoneme-class Model
ไดนามิกไทม์วอร์ปปิง	Dynamic Time Warping
ค่าสัมประสิทธิ์จากการประมาณค่าเชิงเส้น	Linear Prediction Coefficient
ค่าสัมประสิทธิ์เซปสตรัมที่ได้จากการประมาณค่าเชิงเส้น	Linear Prediction Cepstral Coefficient
เพิ่มค่าพลังงาน	Energy
ค่าสัมประสิทธิ์ค่าเฉลี่ยคงที่เท่ากับศูนย์	Zero Mean Static Coefficient
ค่าสัมประสิทธิ์เดลต้า	Delta Coefficient
ค่าสัมประสิทธิ์ความเร่ง	Accuracy Coefficient
การสกัดค่าลักษณะเด่นเสียงพูด	Speech Feature Extraction

การวิเคราะห์การประมาณค่าเชิงเส้น	Linear Prediction Analysis
แบบจำลองการประมาณค่าเชิงเส้น	Linear Prediction Model
การรวมค่าประมาณเชิงเส้น	Linear Combination
ตัวกรองค่าโพล	Pole Filter
ค่าเฉลี่ยกำลังสอง	Mean Square
ลำดับความสัมพันธ์	Autocorrelation Sequence
ค่าสัมประสิทธิ์สะท้อน	Reflection Coefficient
ท่อทางเดินเสียง	Acoustic Tube
ค่าเซปตรัล	Cepstral
การแปลงฟูริเยร์	Fourier Transform
ไดอะโกนอลโควาเรียน	Diagonal Covariance
การวิเคราะห์โดยใช้ชุดตัวกรองความถี่	Filter Bank Analysis
ชุดตัวกรองความถี่	Filter Bank
สเกลเมล	Mel-scale
ความถี่กลาง	Center Frequency
กรอบสัญญาณเสียงพูด	Window of Speech Data
ขนาด	Magnitude
การแปลงฟูริเยย์แบบรวดเร็ว	Fast Fourier Transform
การแปลงโคไซน์ไม่ต่อเนื่อง	Discrete Cosine Transform
ค่าสัมประสิทธิ์การแปลงโคไซน์ไม่ต่อเนื่อง	Discrete Cosine Transform Coefficient
ชุดค่าสัมประสิทธิ์การแปลงโคไซน์ไม่ต่อเนื่อง	Discrete Cosine Series Coefficient
เวกเตอร์พื้นฐาน	Basis Vector
วอร์ปปิ้งฟังก์ชัน	Warping Function
กำหนดค่าที่ต้องการ	Normalize
ค่าเวกเตอร์พื้นฐานของโคไซน์	Cosine Basis Vector
ไบลิเนียวอร์ปปิ้ง	Bilinear Warping
กรอบสัญญาณไคเซอร์	Kaiser Window
ตัวกรองที่ตอบสนองเชิงความถี่	Finite Impulse Response
การวิเคราะห์องค์ประกอบหลัก	Principal Component Analysis
การลดขนาดของข้อมูล	Data Compression

การวิเคราะห์การเปลี่ยนแปลง	Change Analysis
การแปลงเชิงเส้น	Linear Transformation
การวิเคราะห์เฟคเตอร์	Factor Analysis
ไอเกนเวกเตอร์	Eigenvector
เมตริกซ์โควาเรียนซ์	Covariance Matrix
ปริภูมิลักษณะ	Feature Space
ค่าคอร์รีเรชั่น	Correlation
เมตริกซ์ไดอะโกนอล	Diagonal Matrix
ค่าไอเกนเวลู	Eigenvalue
วิธีการของจาโคบี	Jacobi Method
ซูเปอร์เวกเตอร์ของการประมาณรูปร่างผสมของเกาส์	Gaussian Mixture Model –SuperVector
รูปร่างผสมของเกาส์ที่ครอบคลุมทุกแบบจำลอง	Gaussian Mixture Model Universal Background Model
กระบวนการปรับแต่งความเป็นไปได้ภายหลังที่มากที่สุด	Maximum A Posteriori
ตัวกำกับขั้นตอน	State Machine
โปรแกรมแบบพลวัต	Dynamic Programing
แบบจำลองประเภทซ้ายไปขวา	Left-Right Model
แบบจำลองเบคิส	Bakis Model
สถานะซ่อน	Hidden State
การจับคู่รูปแบบ	Pattern Matching
การฝึกฝนแบบจำลอง	Training
การประเมินผล	Evaluation
ลำดับการสังเกตของข้อมูล	Observation Sequence
หลักการของการหาความน่าจะเป็นแบบไปหน้าและย้อนกลับ	Forward-Backward Algorithm
กระบวนการหาความน่าจะเป็นแบบไปหน้า	Forward Algorithm
กระบวนการหาความน่าจะเป็นแบบย้อนกลับ	Backward Algorithm
การจำแนกรูปแบบ	Pattern Recognition
ปริภูมิสมมุติฐาน	Hypothesis Space
กฎการจำแนก	Classifier Rule

ชนิด	Class
สัญลักษณ์แสดงชนิด	Class Label
มิติ Vapnik/Chervonekis	VC-dimension
ปริมาณที่ใช้วัดความสมบูรณ์	Richness
ความเปลี่ยนแปลง	Flexibility
ไฮเปอร์เพลน	Hyperplane
ระยะขอบเขต	Margin
ซัพพอร์ตเวกเตอร์	Support Vector
ค่าความผิดพลาดในการฝึกฝนเป็นศูนย์	Zero Training Error
ระยะระหว่างซัพพอร์ตเวกเตอร์ของทั้ง 2 ชนิดห่างกันมากที่สุด	Maximum Margin
ค่าเหมาะสมที่สุด	Optimization
ความเป็นคู่	Duality
การกำหนดปัญหาที่เราต้องการ	Optimization Problem
ขอบเขตแบบแข็ง	Hard Margin
ขอบเขตแบบอ่อน	Soft Margin
การแลกเปลี่ยน	Trade-off
ปริภูมิขาเข้า	Input Space
คุณลักษณะสำคัญ	Attribute
โพลิโนเมียล	Polynomial
เส้นโค้งคุณลักษณะสมบัติการทำงาน	Receiver Operating Characteristic Curve
พื้นที่ใต้เส้นโค้งค่าคุณลักษณะสมบัติการทำงาน	Area Under the ROC Curve
ความสามารถในการคัดแยก	Classification Performance
ความเที่ยง	Precision
การจดจำ	Precision-recall
ค่าจริงด้านบวก	True Positive Rate
ค่าเท็จด้านบวก	False Positive Rate
เมตริกความสับสน	Confusion Matrix
ค่าจริงบวก	True Positive
ค่าเท็จลบ	False Negative

ค่าเท็จบวก	False Positive
ค่าจริงลบ	True Negative
ค่าบวก	Positive
คลาสบวก	Positive Class
ตัวอย่างด้านบวก	Positive Example
เสียงเงียบ	Silence
หน่วยเสียงเดี่ยว	Monophone
ค่าเฉลี่ย	Mean
ค่าความแปรปรวน	Variance
ฉลาก	Label
การประมาณค่าพารามิเตอร์ใหม่	Parameter Re-estimate
หน่วยเสียงเรียงสาม	Triphone
หน่วยเสียงรอบข้าง	Context-Independent Phone Model
การทำเลื่อนสัญญาณเฟรม	Frame Windowing
เวกเตอร์มูลฐานของการแปลงโคไซน์ไม่ต่อเนื่อง	DCT Basis Vector
การผสมโดยปรับขนาดมิติค่าลักษณะเด่น	Re-dimension of Combine Speech Feature
การรวมโดยใช้ตัวดำเนินการพื้นฐานทางคณิตศาสตร์	Speech Feature Combination Based on Basic Mathematic Operation
การรวมโดยการปรับตำแหน่งการวางข้อมูล	Re-locate Speech Feature combination
สถานะ	State
รูปร่างผสมของเกาส์	Gaussian Mixture
การกำหนดจุดทดสอบ	Threshold
การค้นหาแบบบีม	Beam Search
ค่าลักษณะพิเศษ	High Order Feature
ค่าการเปลี่ยนแปลงสถานะ	Emission
ความน่าจะเป็นในการเปลี่ยนสถานะ	Transition Probability
กรอบสัญญาณ	Frame Window
เบต้า	Beta
การแยกแยะ	Classify
การประเมินค่าเชิงเส้น	Linear Prediction

สารบัญคำย่อ

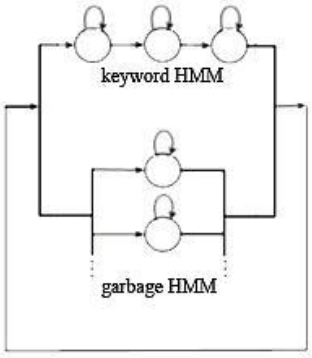
แบบจำลองฮิดเดนมาร์คอฟ	HMM
ระบบรู้จำเสียงพูดต่อเนื่องที่ครอบคลุมคำศัพท์จำนวนมาก	LVCSR
ค่าสัมประสิทธิ์ความถี่สเกลเมล	MFCC
ค่าความดี	FOM
การแปลงโคไซน์ไม่ต่อเนื่อง	DCT
การแปลงฟูริเยย์แบบรวดเร็ว	FFT
ตัวกรองที่ตอบสนองเชิงความถี่	FIR
ชุดค่าสัมประสิทธิ์การแปลงโคไซน์ไม่ต่อเนื่อง	DCSC
การวิเคราะห์องค์ประกอบหลัก	PCA
ซัพพอร์ตเวกเตอร์แมชชีน	SVM
ค่าจริงด้านบวก	TPR
ค่าเท็จด้านบวก	FPR
ค่าสัมประสิทธิ์ที่ได้จากการประมาณค่าเชิงเส้น	LPC
ค่าสัมประสิทธิ์ที่ได้จากค่าสัมประสิทธิ์จากการประมาณค่าเซปตรัลเชิงเส้น	LPCC
ค่าสัมประสิทธิ์ที่ได้จากค่าสัมประสิทธิ์จากการประมาณค่าเซปตรัลเชิงเส้นที่มีการเพิ่มค่าเดลต้า	LPDELCEP
ค่าสัมประสิทธิ์สะท้อนของการประมาณค่าเชิงเส้น	LPREFC
ค่าลักษณะเด่นที่ได้จากผลลัพธ์ของชุดตัวกรองสเกลเมลเชิงเส้น	MELSPEC
ค่าลักษณะเด่นที่ได้จากค่าลอการิทึมของผลลัพธ์ที่ได้จากชุดตัวกรองสเกลเมล	FBANK
ชุดค่าสัมประสิทธิ์การแปลงโคไซน์ไม่ต่อเนื่อง	DCSC
การวิเคราะห์องค์ประกอบหลัก	PCA
กระบวนการปรับแต่งความเป็นไปได้ภายหลังที่มากที่สุด	MAP
เส้นโค้งคุณลักษณะสมบัติการทำงาน	ROC
พื้นที่ใต้เส้นโค้งค่าคุณลักษณะสมบัติการทำงาน	AUC

บทที่ 1
บทนำ

1.1. ความสำคัญและที่มาของการวิจัย

ณ ปัจจุบัน เทคโนโลยีมีบทบาทสำคัญที่ช่วยให้การใช้ชีวิตมีความสะดวกสบายมากยิ่งขึ้น ซึ่งเทคโนโลยีทางด้านการประมวลผลสัญญาณเสียงพูด (Speech Processing) เป็นหนึ่งในเทคโนโลยีเหล่านั้น ผู้ใช้สามารถใช้งานอุปกรณ์ต่างๆ โดยใช้เพียงคำสั่งเสียงเท่านั้น ช่วยประหยัดเวลาและลดขั้นตอนการใช้งานอุปกรณ์ โดยในปัจจุบันได้มีการประยุกต์ใช้งานเทคโนโลยีดังกล่าวร่วมกับระบบอำนวยความสะดวกต่างๆ ไม่ว่าจะเป็น การสั่งงานและควบคุมอุปกรณ์ด้วยเสียงพูด การค้นหาคำที่ต้องการบนเอกสารเสียง การสืบค้นข้อมูลสื่อเสียง เป็นต้น ซึ่งระบบต่างๆ ที่ได้กล่าวมานั้น ระบบค้นหาคำหลักนั้นมีส่วนสำคัญและถูกนำมาประยุกต์ใช้งานมากที่สุด เนื่องจากคุณสมบัติเด่นของระบบค้นหาคำหลักคือ มีความยืดหยุ่นในด้านการใช้งานและใช้เวลาในการประมวลผลน้อย อีกทั้งยังให้ผลลัพธ์การทำงานที่ดียิ่งขึ้นเมื่อนำมาประยุกต์ใช้งานกับกลุ่มคำเฉพาะทางหรือกลุ่มคำที่มีคำศัพท์จำนวนไม่มากนักเช่นการออกคำสั่งเสียงควบทิศทางสำหรับหุ่นยนต์ เป็นต้น

สำหรับการออกแบบโครงสร้างของระบบค้นหาคำหลักบนสื่อเสียงนั้น งานวิจัยส่วนใหญ่จะใช้ระบบรู้จำเสียงพูด (Speech Recognition) เป็นเทคนิคพื้นฐาน [1] ซึ่งมีหลักการทำงานคือแปลงสัญญาณเสียงให้เป็นข้อความอักษรทั้งหมดก่อน แล้วจึงนำเข้าสู่กระบวนการค้นหาคำหลัก (Keyword Text Search) โดยระบบรู้จำเสียงพูดประเภทรู้จำเสียงต่อเนื่องที่ครอบคลุมคำศัพท์จำนวนมาก (Large Vocabulary Continuous Speech Recognition: LVCSR) ซึ่งอาศัยหลักการของแบบจำลองฮิดเดนมาร์คอฟ (Hidden Markov Model: HMM) เป็นระบบรู้จำเสียงพูดที่นิยมใช้มากที่สุด เพราะเป็นเทคนิคที่ง่าย สะดวกรวดเร็ว และเหมาะกับการจำแนกหน่วยเสียง สำหรับการประยุกต์ใช้งาน HMM ทางด้านการค้นหาคำหลักนั้น HMM จะทำการสร้างแบบจำลอง 2 ประเภท คือ แบบจำลองคำหลัก (Keyword Model) และแบบจำลองพื้นภูมิหรือแบบจำลองขยะ (Background Model or Garbage Model) ดังที่แสดงให้เห็นในรูปที่ 1-1



ภาพประกอบที่ 1-1โครงสร้างของ HMM สำหรับการค้นหาคำหลัก

โดยแบบจำลองสำหรับคำหลักนั้นจะผ่านการถอดรหัสแบบวิเทอบี (Viterbi Decoding) คำหลักจะถูกพิจารณาเสมือนคำที่ถูกเปล่งออกมาอย่างเป็นลำดับ ถ้าเป็นเส้นทางที่ดีที่สุด ก็จะสามารถผ่านแบบจำลองคำหลักได้ อย่างไรก็ตาม เมื่อมีการฝึกฝนคำให้แก่ระบบมากขึ้น แบบจำลองพื้นภูมิก็จะมากขึ้น ซึ่งอาจทำให้เกิดแบบจำลองที่มีลักษณะใกล้เคียงกับแบบจำลองของคำหลักที่เราต้องการค้นหา จึงทำให้มีโอกาสทำให้เกิดความผิดพลาดในการหาคำหลักได้

ต่อมาได้มีงานวิจัยที่ได้เสนอวิธีการใหม่ โดยที่ไม่ต้องพึ่งระบบรู้จำเสียงพูดในการค้นหาคำหลัก โดยวิธีดังกล่าวได้ใช้วิธีการจำแนกประเภทโดยใช้เคอร์เนลและขอบเขตการแยกแยะที่มีขนาดกว้างที่สุด (Large Margin and Kernel Method) ในการแยกแยะและค้นหาคำหลัก วิธีการดังกล่าวใช้ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine: SVM) เป็นเครื่องมือในการแยกแยะโดยมุ่งเน้นลักษณะการเรียงลำดับของหน่วยพื้นฐานของเสียงพยางค์ (Phoneme) เป็นปัจจัยสำคัญในการค้นหาคำหลัก ซึ่งสามารถช่วยลดจำนวนข้อมูลที่จะใช้ในการฝึกฝนระบบและสามารถทำงานครอบคลุมกับหน่ยเสียงหรือคำได้หลากหลายขึ้น แต่ต้องแลกกับการใช้ทรัพยากรระหว่างการประมวลผลที่มากกว่า อันเนื่องมากจากลักษณะพื้นฐานการทำงานของ SVM นั้นเอง

นอกจากเทคทิคพื้นฐานที่จะใช้เป็นพื้นฐานของระบบค้นหาคำหลักแล้ว อีกหนึ่งปัจจัยที่มีผลกระทบต่อผลลัพธ์การทำงานทางด้านการประมวลผลสัญญาณเสียงพูดคือข้อมูลที่จะใช้เป็นค่าลักษณะเด่นของสัญญาณเสียงพูด (Speech Feature) ซึ่งข้อมูลดังกล่าวจะแตกต่างกันไปตามกรรมวิธีและหลักการที่ใช้ในการดึงลักษณะเด่นออกมาจากสัญญาณเสียงพูด (Speech Feature Extraction) โดยค่าลักษณะเด่นของสัญญาณเสียงพูดที่เป็นที่นิยมในงานวิจัยหลายๆชิ้นคือค่าสัมประสิทธิ์ความถี่สเกลเมล (Mel-Frequency Cepstral Coefficient: MFCC) ซึ่งผลลัพธ์ส่วนใหญ่อยู่ในระดับที่ดี แต่ยังคงมีหลายๆ งานวิจัยที่พยายามจะพัฒนากรรมวิธีในการดึงค่าลักษณะเด่นของสัญญาณเสียงพูด

เพื่อให้ได้วิธีการดึงค่าลักษณะเด่นของสัญญาณเสียงพูดที่ดียิ่งขึ้น เช่นการรวมค่าลักษณะเด่นของสัญญาณเสียงพูดเข้าด้วยกัน (Speech Feature Combination) [2] [3]

อย่างไรก็ตาม งานวิจัยที่เกี่ยวกับระบบค้นหาคำหลักบนสื่อเสียงพูดหรือวิธีการดึงค่าลักษณะเด่นของเสียงพูด ที่ประยุกต์และใช้งานร่วมกับภาษาไทยเป็นภาษาหลักนั้นยังมีไม่มากนัก งานวิจัยชิ้นนี้จึงต้องการที่จะเปรียบเทียบและหาระบบพื้นฐานสำหรับสร้างระบบค้นหาคำหลักบนสื่อเสียง โดยเปรียบเทียบระหว่าง 2 ระบบซึ่งอยู่ใช้เทคนิคของระบบรู้จำเสียง และเทคนิคการจำแนกประเภทโดยใช้เคอร์เนลและขอบเขตการแยกแยะที่มีขนาดกว้างที่สุดโดยทดสอบร่วมกับค่าลักษณะเด่นของสัญญาณเสียงพูดประเภทต่างๆ เพื่อที่จะหาระบบค้นหาคำหลักบนสื่อเสียงและค่าลักษณะเด่นของสัญญาณเสียงพูดที่เหมาะสมที่สุดสำหรับงานทางด้านการค้นหาคำหลักบนสื่อเสียงภาษาไทย

1.2. งานวิจัยที่เกี่ยวข้อง

ในส่วนนี้จะแบ่งงานวิจัยที่เกี่ยวข้องออกเป็น 2 กลุ่มใหญ่ๆ คือ

1.2.1. งานวิจัยที่เกี่ยวข้องกับระบบค้นหาคำหลักบนสื่อเสียงพูด

สำหรับงานทางด้านการค้นหาคำหลักบนสื่อเสียงภาษาไทยนั้น ยังไม่เป็นที่แพร่หลายนัก อันเนื่องจากงานวิจัยทางด้านการวิเคราะห์สัญญาณเสียงพูดภาษาไทยส่วนใหญ่มุ่งเน้นไปทางด้านการสร้างระบบรู้จำเสียงพูด ซึ่งสามารถนำมาประยุกต์ใช้กับงานวิจัยหลายๆ อย่าง เช่นระบบค้นหาคำหลักบนสื่อเสียงพูด เป็นต้น

ปภาณิน เดชเทวัญดำรง [4] ได้นำเสนองานวิจัยที่เกี่ยวข้องกับการสืบค้นหาข้อมูลข่าวด้วยเสียงสำหรับข่าวภาษาไทย ซึ่งใช้ระบบรู้จำเสียงพูดแบบ LVCSR เป็นพื้นฐานระบบโดยในการวิจัยได้ใช้ฐานข้อมูลเสียงข่าวภาษาไทย (BN-Corpus) ร่วมกับฐานข้อมูลเสียงภาษาไทย LOTUS เฉลิมวุฒิ ไวชน [5] ได้นำเสนอการค้นหาคำหลักบนสื่อเสียงพูดภาษาไทยแบบไม่ขึ้นกับตัวบุคคลโดยการประยุกต์ใช้หลักการของ HMM ซึ่งเป็นแบบจำลองของเสียงในระดับย่อยของพยางค์ใช้วิธีการตรวจสอบหน่วยเริ่มและหน่วยตามของพยางค์ ผลการทดสอบจะวัดอยู่ในค่าของเมอริท (Figure of Merit: FOM) ศิรินาถ ตั้งรวมทรัพย์ [6] ได้เสนอระบบค้นหาคำหลักบนสื่อเสียงที่ใช้ระบบรู้จำเสียงแบบ LVCSR โดยเพิ่มแบบจำลองหน่วยเสียง 3 ชนิด ทั้งรูปแบบของ แบบจำลองตัวเติมที่ครอบคลุมทั้งหมด (Universal Filler Model) แบบจำลองของเสียงพยัญชนะและเสียงสระ (Consonant-vowel model) และแบบจำลองของหน่วยเสียงแต่ละตัว (Phoneme-class Model)

นอกจากการใช้ระบบรู้จำเสียงพูดเป็นระบบพื้นฐานสำหรับการค้นหาคำหลักบนสื่อเสียงแล้ว การเปรียบเทียบสัญญาณเสียงพูดเพื่อหาข้อมูลเสียงภาษาไทยโดยตรงเพื่อค้นหาคำหลักบนสื่อเสียง

ได้ถูกนำเสนอเช่นเดียวกัน ภูเบศ โต๊ะลง [7] ได้มุ่งเน้นในการค้นคืนข้อมูลเสียงจากแฟ้มข้อมูลเสียงภาษาไทยขนาดใหญ่โดยการเปรียบเทียบสัญญาณเสียงที่เป็นคำถามเทียบกับสื่อเสียงโดยตรง ใช้วิธีวัดระยะทางแบบไดนามิกไทม์วอร์ปปิง (Dynamic Time Warping) และการผันวรรณยุกต์ 5 ระดับของภาษาไทยใช้ร่วมในการเปรียบเทียบ อย่างไรก็ตามวิธีการดังกล่าวยังคงใช้เวลาในการประมวลผลมากกว่าระบบค้นหาคำหลักบนสื่อเสียงที่มีพื้นฐานมากจากระบบรู้จำเสียงพูด

1.2.2. **งานวิจัยที่เกี่ยวข้องกับค่าลักษณะเด่นของเสียงพูด**

งานวิจัยที่เกี่ยวกับค่าลักษณะเด่นของสัญญาณเสียงพูดภาษาไทยนั้น นักวิจัยไทยได้มุ่งเน้นไปในด้านการพัฒนางานวิจัยร่วมกับค่าลักษณะเด่น MFCC ซึ่งเป็นค่าลักษณะเด่นที่เป็นที่นิยมในงานวิเคราะห์สัญญาณเสียงพูด แม้จะเน้นในส่วนของ MFCC แต่ยังคงมีงานวิจัยที่วิเคราะห์การเปรียบเทียบค่าลักษณะเด่นของสัญญาณเสียงพูดภาษาไทย เช่น ภคพงศ์ อมรกุล [8] ได้ทำการทดสอบระบบรู้จำเสียงพูดสำหรับการเน้นเสียง 4 รูปแบบคือ โกรธ ดัง ผลของลอมบาร์ด และปกติ โดยทำการทดสอบบนค่าลักษณะเด่นของสัญญาณเสียงพูดที่แตกต่างกัน ได้แก่ ค่าสัมประสิทธิ์ MFCC, ค่าสัมประสิทธิ์จากการประมาณค่าเชิงเส้น (Linear Prediction Coefficient: LPC) และค่าสัมประสิทธิ์เซปตรัลที่ได้จากการประมาณค่าเชิงเส้น (Linear Prediction Cepstral Coefficient: LPCC) อีกทั้งมีการเปลี่ยนแปลงจำนวนค่าลักษณะเด่นเพิ่มค่าพลังงาน (Energy) เพิ่มค่าสัมประสิทธิ์ค่าเฉลี่ยคงที่เท่ากับศูนย์ (Zero Mean Static Coefficient) และเพิ่มค่าสัมประสิทธิ์เดลต้า (Delta Coefficient) หรือจะเป็นงานของมนตรีและปติมากร [9] ได้สร้างค่าลักษณะเด่นของสัญญาณเสียงพูดที่มีการเพิ่มลักษณะพิเศษโดยใช้ค่าพลวัต (Dynamic Feature) โดยใช้การแปลงโคไซน์ไม่ต่อเนื่อง (Discrete Cosine Transform: DCT) เพื่อนำมาเปรียบเทียบกับค่าลักษณะเด่นที่เพิ่มค่าสัมประสิทธิ์เดลต้าและค่าสัมประสิทธิ์ความเร่ง (Accuracy Coefficient) ซึ่งเป็นค่าลักษณะพิเศษแบบพลวัตที่นิยมใช้กันมาก

1.3. วัตถุประสงค์

1.3.1. เปรียบเทียบระบบค้นหาคำหลักบนสื่อเสียงที่มีพื้นฐานระบบที่แตกต่างกันระหว่างระบบที่อยู่บนพื้นฐานของระบบรู้จำเสียงพูด และระบบที่อยู่บนพื้นฐานของวิธีการจำแนกประเภทโดยใช้เคอร์เนลและขอบเขตการแยกแยะที่มีขนาดกว้างที่สุด

1.3.2. เปรียบเทียบค่าลักษณะเด่นของสัญญาเสียงพูดภาษาไทยแบบต่างๆ

1.3.3. พัฒนาค่าลักษณะเด่นของสัญญาเสียงพูดภาษาไทย

1.3.4. เพื่อหาคู่ที่เหมาะสมระหว่างระบบค้นหาคำหลักบนสื่อเสียงและค่าลักษณะเด่นของสัญญาเสียงพูดที่เหมาะสมที่สุดสำหรับการใช้งานบนภาษาไทย เพื่อที่จะสามารถนำไปประยุกต์ใช้กับงานอื่นๆ ได้

1.4. ขอบเขตงานวิจัย

1.4.1. เสียงพูดที่ใช้ในการค้นหาข้อมูลจะอยู่ในลักษณะของคำโดด

1.4.2. ทำงานบนภาษาไทยเท่านั้น

1.5. ประโยชน์ที่คาดว่าจะได้รับ

ได้ระบบค้นหาคำหลักบนสื่อเสียงและค่าลักษณะเด่นของสัญญาเสียงพูดที่เหมาะสมที่สุดสำหรับการใช้งานบนภาษาไทย เพื่อที่จะสามารถนำไปประยุกต์ใช้กับงานอื่นๆ ได้

บทที่ 2
ทฤษฎีและหลักการ

2.1. การสกัดค่าลักษณะเด่นเสียงพูด (Speech Feature Extraction) [10]

การสกัดค่าลักษณะเด่นของสัญญาณเสียงพูดนั้น เป็นการหาค่าพารามิเตอร์ที่ใช้แทนตัวอย่างสัญญาณเสียง โดยค่าพารามิเตอร์ดังกล่าวจะบอกถึงลักษณะสำคัญของสัญญาณเสียงพูดนั้นๆ ออกมา ซึ่งค่าลักษณะเด่นของสัญญาณเสียงพูดนั้นมีส่วนสำคัญอย่างมากในงานทางด้านการวิเคราะห์สัญญาณเสียง โดยค่าลักษณะเด่นที่เหมาะสมนั้นก็จะให้ผลลัพธ์ที่ถูกต้องและแม่นยำ อีกทั้งในปัจจุบันมีวิธีการสกัดค่าลักษณะเด่นของสัญญาณเสียงพูดหลากหลายรูปแบบ ซึ่งมีหลักการและวิธีการที่แตกต่างกันออกไป

2.1.1. การวิเคราะห์การประมาณค่าเชิงเส้น (Linear Prediction Analysis)

ลักษณะสำคัญของวิธีการประมาณค่าเชิงเส้น คือการแทนสัญญาณเสียงพูดโดยอาศัยสมบัติทางด้านความถี่ด้วยการใช้ค่าพารามิเตอร์จำนวนน้อย แต่กลับสามารถให้ผลลัพธ์ที่แม่นยำและมีประสิทธิภาพ อีกทั้งยังมีวิธีการคำนวณหาค่าพารามิเตอร์ที่ง่ายอีกด้วย

หลักพื้นฐานของแบบจำลองการประมาณค่าเชิงเส้น (Linear Prediction Model) คือการคำนวณเพื่อหาค่าที่จะได้ในอนาคตจากการรวมค่าประมาณเชิงเส้น (Linear Combination) ของค่าสัญญาณที่เกิดขึ้นก่อนหน้าสัญญาณนั้นๆ โดยสมมติให้สัญญาณที่ได้รับความสนใจเกิดจากแหล่งกำเนิดสัญญาณที่ถูกกระตุ้นจากวงจรกรองเชิงเส้น

ในการวิเคราะห์การประมาณค่าเชิงเส้นนั้น ฟังก์ชันสำหรับใช้แทนทางเดินเสียงนั้นจะถูกออกแบบโดยใช้ตัวกรองค่าโพล (Pole Filter) ทั้งหมดโดยจะทำการผ่านฟังก์ชัน

$$H(z) = \frac{1}{\sum_{i=0}^{p} a_i z^{-i}} \tag{2-1}$$

โดย p จะแทนด้วยจำนวนโพลและ $a_0 \equiv 1$ ตัวกรองค่าสัมประสิทธิ์ a_i จะถูกเลือกเพื่อลดผลรวมของค่าความคลาดเคลื่อนที่เกิดจากตัวกรองค่าเฉลี่ยกำลังสอง (Mean Square) ซึ่งได้วิเคราะห์บนช่วงกรอบสัญญาณ (Window) ทั้งช่วง ยกตัวอย่างเช่นการกำหนดให้ช่วงกรอบสัญญาณของ

เสียงพูดให้ $s_n, n=1,N$ ซึ่งเทอมแรกของ $p+1$ ของลำดับความสัมพันธ์ (Autocorrelation Sequence) คำนวณได้จาก

$$r_i = \sum_{j=1}^{N=i} s_j s_{j+1} \tag{2-2}$$

โดยที่ $i=0,p$ ตัวกรองค่าสัมประสิทธิ์จะถูกคำนวณวนซ้ำโดยใช้ชุดสัมประสิทธิ์ตัวช่วย k_i ซึ่งสามารถใช้ในการตีความหมายของค่าข้อมูลเพื่อสร้างค่าสัมประสิทธิ์สะท้อน (Reflection Coefficient) ให้เหมือนกับท่อทางเดินเสียง (Acoustic Tube) ภายในหูของมนุษย์อีกทั้งยังสามารถประมาณค่าข้อมูลที่ผิดพลาด E ซึ่งมีค่าตั้งต้นเท่ากับ r_0 ต่อมากำหนดให้ $k_j^{(i=1)}$ และ $a_j^{(i=1)}$ เป็นค่าสะท้อนและตัวกรองค่าสัมประสิทธิ์สำหรับเป็นตัวกรองลำดับที่ i โดยที่ตัวกรองค่าสัมประสิทธิ์สำหรับเป็นตัวกรองลำดับที่ i สามารถคำนวณได้จาก 3 ขั้นตอนดังต่อไปนี้

ลำดับแรก ทำการตั้งค่าชุดค่าสัมประสิทธิ์สะท้อนซึ่งคำนวณจาก

$$k_j^i = k_j^{i-1} \tag{2-3}$$

สำหรับ $j = IJ-1$ และ

$$k_j^{(i)} = \left\{ r_i + \sum_{j=1}^{i-1} a_j^{(i+1)} r_{i-j} \right\} IE^{(i-1)} \tag{2-4}$$

ต่อมา ทำการคำนวณค่าพลังงานที่มีการเปลี่ยนแปลง

$$E^i = (1-k_i^{(i)} k_j^{(i)}) E^{(i-1)} \tag{2-5}$$

สุดท้าย ทำการคำนวณหาตัวกรองสัมประสิทธิ์ใหม่โดย

$$a_j^{(i)} = a_j^{(i-1)} - k_j^{(i)} a_{i-j}^{(i-1)} \tag{2-6}$$

$$a_i^{(i)} = -k_i^{(i)} \tag{2-7}$$

ซึ่งกระบวนการดังกล่าวจะทำซ้ำไปเรื่อยๆ ตั้งแต่ $i=1$ จนถึงลำดับของตัวกรองที่ต้องการ $i=p$ เพื่อที่จะให้ได้ผลที่ได้จากการแปลงค่าดังกล่าว

อีกทางเลือกหนึ่งสำหรับการหาค่าลักษณะเด่นที่ได้จากค่าสัมประสิทธิ์การประมาณค่าเชิงเส้นนั้นคือ ค่าสัมประสิทธิ์เซปตรัลที่ได้จากการประมาณค่าเชิงเส้นซึ่งเป็นการประมาณค่าเชิงเส้นของข้อมูลเซปตร้า โดยค่าเซปตรัล (Cepstral) ของข้อมูลจะได้จากการแปลงฟูริเยร์ (Fourier Transform) ของค่าลอการิทึมของสเปกตรัม ในกรณีของการประมาณค่าสัมประสิทธิ์เชิงเส้นของข้อมูลเซปตร้า ลักษณะของสเปกตรัมที่จะนำมาใช้นั้นจะต้องมาจากการประมาณค่าเชิงเส้นของข้อมูลเซปตร้าซึ่งแปลงมากจากการแปลงฟูริเยร์ของตัวกรองค่าสัมประสิทธิ์ อย่างไรก็ตามเราสามารถคำนวณหาค่าเซปตรัลโดยใช้วิธีการง่ายๆ แต่ให้ผลลัพธ์ที่มีประสิทธิภาพ โดยการคำนวณแบบวนซ้ำ

$$c_n = a_n + \frac{1}{n}\sum_{i=1}^{n-1}(n \cdot i)a_i c_{n \cdot i} \qquad (2\text{-}8)$$

ข้อดีของค่าสัมประสิทธิ์เซปตรัลที่ได้จากการประมาณค่าเชิงเส้นคือ ค่าเซปตรัลดังกล่าวโดยปกติแล้วจะไม่มีความสัมพันธ์กัน (Uncorrelated) จึงสามารถใช้งานได้ดีกับ HMM แบบไดอะโกนอลโควาเรียน (Diagonal Covariance) อย่างไรก็ตามยังคงมีปัญหาอีกประการหนึ่งคือค่าของเซปตร้าที่อยู่ในอันดับสูงๆจะให้ค่าข้อมูลที่มีขนาดเล็กและจะมีช่วงความแปรปรวนที่กว้างเมื่อไล่ลำดับสัมประสิทธิ์จากลำดับต่ำไปลำดับสูง

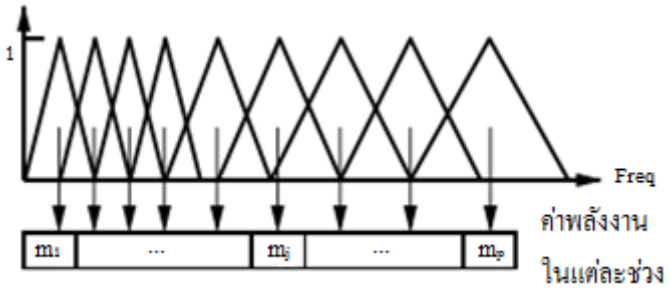
2.1.2. **การวิเคราะห์โดยใช้ชุดตัวกรองความถี่ (Filter Bank Analysis)**

โดยปกติแล้วหูของมนุษย์นั้นมีความสามารถในการได้ยินเสียงคลื่นสัญญาณความถี่ที่ไม่อยู่ในลักษณะของสัญญาณเชิงเส้นได้โดยการเปลี่ยนค่าสัญญาณดังกล่าวให้เป็นค่าสเปกตรัมของสัญญาณเสียง ซึ่งหลักการดังกล่าว ได้ถูกนำไปประยุกต์ใช้ในการสร้างค่าลักษณะเด่นของสัญญาณเสียงพูดโดยการแปลงค่าสัญญาณที่มีลักษณะไม่เป็นเชิงเส้นได้โดยใช้ชุดตัวกรองความถี่ (Filter Bank)

สำหรับชุดตัวกรองความถี่ที่จะนำมาใช้ในงานวิจัยชิ้นนี้ จะใช้การแปลงฟูริเยย์อย่างง่ายในการออกแบบตัวกรองบนพื้นฐานของสเกลเมล (Mel-Scale) ดังที่แสดงในภาพประกอบที่ 2-1 ซึ่งตัวกรองที่เราใช้ในนั้นจะเป็นรูปสามเหลี่ยมและความถี่กลาง (Center Frequency) ของตัวกรองจะกระจายตามสเกลของเมล ซึ่งนิยามไว้ดังสมการที่ 2-9

$$Mel(f) = 2595\log_{10}(1+\frac{f}{700}) \quad (2\text{-}9)$$

ในการสร้างชุดตัวกรองนั้น กรอบสัญญาณเสียงพูด (Window of Speech Data) ซึ่งจะได้ขนาด (Magnitude) ของสัญญาณในช่วงนั้นๆ ออกมากโดยค่าสัมประสิทธิ์ของค่าขนาดนั้นจะถูกกรองออกมาโดยผ่านตัวกรองสามเหลี่ยมในแต่ละช่วง แล้วนำมาหาค่าเฉลี่ยกลางโดยการนำค่าที่ผ่านตัวกรองมานั้นแปลงกับการแปลงฟูริเยย์แบบรวดเร็ว (Fast Fourier Transform: FFT) นำไปคูณกับอัตราการขยายของตัวกรองในแต่ละช่วง สุดท้ายจึงนำค่าที่คำนวณได้มารวมกันแล้วแสดงผลลัพธ์เป็นค่าขนาดของสเปกตรัมของช่วงตัวกรองแต่ละช่วง



ภาพประกอบที่ 2-1ชุดตัวกรองเมลสเกล

2.1.3. **ค่าลักษณะพิเศษ (High Order Feature)**

ค่าลักษณะพิเศษนั้น มีส่วนสำคัญที่จะช่วยให้ค่าลักษณะเด่นของสัญญาณเสียงพูดมีเอกลักษณ์ที่เด่นขึ้น เมื่อนำค่าลักษณะเด่นที่มีการเพิ่มค่าลักษณะพิเศษไปใช้งาน ก็จะส่งผลทำให้ผลลัพธ์ที่ได้มีความถูกต้องเพิ่มมากขึ้น

2.1.3.1. **ค่าพลังงานเสียง (Voice Energy)**

ความแตกต่างระหว่างพลังงานของหน่วยเสียงแต่ละหน่วยนั้น เป็นหนึ่งในคุณสมบัติที่ดีสำหรับการแยกแยะหน่วยเสียงแต่ละหน่วย โดยค่าพลังงานจะถูกคำนวณในรูปแบบของลอการิธึมของสัญญาณพลังงาน ซึ่งสามารถคำนวณได้จากสมการที่ 2-10

$$E_v = \log_{10}\sum_{n=1}^{N} s^2(n) \quad (2\text{-}10)$$

2.1.3.2.**ค่าเปลี่ยนแปลงเซปสตรัมทางเวลา (Temporal Cepstral Derivative)**

ค่าลักษณะเด่นของสัญญาณเสียงพูดที่ได้ จะคำนวณมาจากหน้าต่างของสัญญาณเสียงในแต่ละช่วง ทำให้ขาดลักษณะสำคัญที่เปลี่ยนแปลงระหว่างค่าลักษณะเด่นที่ได้กับเวลา (Dynamic Feature) ซึ่งสามารถหาค่าดังกล่าวได้จากการคำนวณอนุพันธ์อันดับที่หนึ่ง ซึ่งสามารถคำนวณได้จากสมการที่ 2-11 และสามารถหาความเร่ง (Acceleration) ของค่าลักาณะเด่นโดยการนำค่าลักษณะเด่นของสัญญาณเสียงพูดที่ผ่านการหาค่าอนุพันธ์อันดับที่หนึ่งมาทำการหาค่าอนุพันธ์อันดับที่สองได้จากสมการที่ 2-12

$$\Delta^2 c_m(t) \approx \left[\frac{\sum_{k=-M}^{M} c_m(t+k)k}{\sum_{k=-M}^{M} k^2} \right] \tag{2-11}$$

$$\frac{\partial c_m(t)}{\partial t} = \Delta c_m(t) \approx \left[\frac{\sum_{k=-M}^{M} c_m(t+k)k}{\sum_{k=-M}^{M} k^2} \right] \tag{2-12}$$

เมื่อ c_m คือสัมประสิทธิ์ลำดับที่ m ณ เวลาที่ t

M คือค่าคงที่

2.2. ค่าลักษณะเด่นของเสียงบนพื้นฐานของค่าสัมประสิทธิ์การแปลงโคไซน์ไม่ต่อเนื่อง (Speech Feature Based on Discrete Cosine Transform Coefficient) [11]

หลักการของค่าลักษณะเด่นของสัญญาณเสียงพูดชนิดนี้ คือแบ่งสัญญาณเสียงพูดออกเป็นเฟรมย่อยๆ แล้วนำเฟรมเหล่านั้นมาทำการคำนวณฟูริเยย์แบบรวดเร็ว เพื่อให้ได้ค่าสเปกตรัมของเฟรมสัญญาณเสียง แล้วจึงนำค่าสเปกตรัมที่ได้มาทำการแปลงโคไซน์ไม่ต่อเนื่อง เพื่อหาค่าสัมประสิทธิ์โดยไม่จำเป็นที่จะต้องผ่านชุดตัวกรองก่อนเหมือนกับค่าลักษณะเด่น MFCC จากนั้นจึงทำการรวมเฟรมเข้าด้วยกันเป็นบล็อก และทำการแปลงโคไซน์ไม่ต่อเนื่องกับบล็อกอีกครั้งก็จะได้ชุดค่าสัมประสิทธิ์การแปลงโคไซน์ไม่ต่อเนื่อง (Discrete Cosine Series Coefficient: DCSC) ซึ่งใน

การแปลงโคไซน์ไม่ต่อเนื่องในแต่ละครั้งนั้น ก็จะมีการเปลี่ยนแปลงรูปร่างของเวกเตอร์พื้นฐาน (Basis Vector) เล็กน้อยเพื่อให้ค่าที่ได้เหมาะสมกับสัญญาณเสียงพูดมากยิ่งขึ้น

ค่าลักษณะเด่นชนิดนี้จะได้มาจากการแปลงโคไซน์ไม่ต่อเนื่องบนค่าลอการิทึมที่มาจากการแปลงฟูริเยย์ไม่ต่อเนื่องแล้วรวมความถี่เสียงโดยใช้วอร์ปปิ้งฟังก์ชัน (Warping Function) $f = g(f)$

$$DCTC(i) = \int_0^1 X(f')\cos(\pi if')df' \tag{2-13}$$

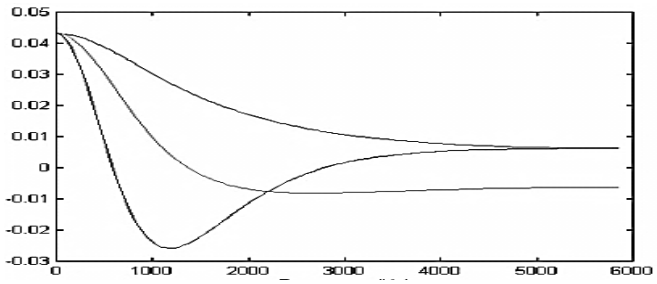
โดยจะเลือกช่วงความถี่ที่ต้องการมากำหนดค่าที่ต้องการ (Normalize) เพื่อใช้ในการเตรียมการแบ่งสัญญาณจาก 0 ถึง 1เมื่อมีการเปลี่ยนค่าตัวแปรบางค่าก็จะสามารถเขียนสมการใหม่ได้เป็น

$$DCTC(i) = \int_0^1 X(f')\Phi_i(f)df \tag{2-14}$$

ซึ่งค่าเวกเตอร์พื้นฐานของโคไซน์ (Cosine Basis Vector) ตั้งต้นจะถูกแปลงเพื่อใช้ในการอธิบายผลกระทบจากการทำวอร์ปปิ้งในรูปแบบของเวกเตอร์พื้นฐาน

$$\Phi_i(f) = \cos[\pi ig(f)]\frac{dg}{df} \tag{2-15}$$

จากรูปที่ 3 ได้แสดงให้เห็นถึงเวกเตอร์พื้นฐาน 3 เวกเตอร์แรก ซึ่งใช้ไบลิเนียวอร์ปปิ้ง (Bilinear Warping) ที่ใช้ค่าสัมประสิทธิ์ที่ 0.45 ซึ่งในทางปฏิบัติแล้ว ค่าสัมประสิทธิ์การแปลงโคไซน์ไม่ต่อเนื่องจะใช้ผลรวมที่สัมพันธ์กับตัวอย่างที่ได้จากการแปลงฟูริเยย์แบบรวดเร็วและ ผลที่ได้จะถูกแปลงเป็นเวกเตอร์พื้นฐานบนช่วงของความถี่ที่ได้ถูกระบุไว้ ด้วยเหตุนี้จึงทำให้ค่าสัมประสิทธิ์การแปลงโคไซน์ไม่ต่อเนื่องมีลักษณะที่เหมือนกับค่าสัมประสิทธิ์เซปตรัลที่นิยมใช้ในงานทางด้านการวิเคราะห์สัญญาณเสียงพูด



ภาพประกอบที่ 2-2 กราฟเวกเตอร์พื้นฐาน 3 เวกเตอร์แรก

จากรูปที่ 2-2 จะเห็นได้ว่าค่าสัมประสิทธิ์การแปลงโคไซน์ไม่ต่อเนื่องสำหรับเฟรมสัญญาณเสียงที่ได้จากการแบ่งสัญญาณเสียง จะถูกเขียนใหม่โดยใช้การขยายตัวของโคไซน์เชิงเวลาตามสมการที่ 2-16

$$DCTC(i,j) = \int_{-1/2}^{1/2} DCTC(i,t')\cos(\pi\ jt')dt' \tag{2-16}$$

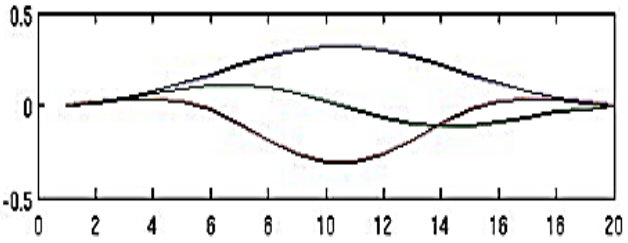
โดยตัวแปร $t' = h(t)$ จะเป็นตัวที่ใช้ในการพิจารณาในการวอร์ปโดยรูปแบบของ t จะขึ้นอยู่กับฟังก์ชัน h ซึ่งถูกเลือกเพื่อใช้ในการเน้นบริเวณส่วนกลางที่สัญญาณที่ได้ทำการแบ่งส่วนไว้ โดยระยะห่างของช่วงเวลาจะถูกำหนดไว้ในช่วง -½ ถึง ½ และเมื่อเปลี่ยนตัวแปรอีกครั้ง เราจะสามารถเขียนสมการใหม่ได้เป็น

$$DCTC(i,j) = \int_{-1/2}^{1/2} DCTC(i,t)\Theta_j(t)dt \tag{2-17}$$

โดยที่เวกเตอร์พื้นฐาน $\Theta_j(t)$ จะถูกแปลงโคไซน์สำหรับการกำหนดระยะห่างของเวลาในการแบ่งสัญญาณเสียง

$$\Theta_j(t) = \cos[\pi\ j\ h(t)]\frac{dh}{dt} \tag{2-18}$$

จากรูปที่ 2-3 จะเป็นการอธิบายถึงเวกเตอร์พื้นฐานของค่าสัมประสิทธิ์การแปลงโคไซน์ไม่ต่อเนื่อง 3 เวกเตอร์แรกที่ใช้กรอบสัญญาณไคเซอร์ (Kaiser Window) ในส่วนของ $\frac{dh}{dt}$ ซึ่งจะนำมาใช้ในการทดลองครั้งนี้



ภาพประกอบที่ 2-3 กราฟของเวกเตอร์ DCTC 3 เวกเตอร์แรกที่ผ่านกรอบสัญญาณไคเซอร์

จากกระบวนการทั้ง 2 อย่างทั้งขั้นตอนการกรองขนาดของค่าลอการิทึมของสเปกตรัมทั้งในด้านความถี่และด้านของเวลา ทำให้ได้ตัวกรอง 2 มิติซึ่งผลลัพธ์ที่ได้นั้นจะสามารถนำไปใช้ในการแยกแยะสัญญาณได้ โดยค่าสัมประสิทธิ์ตัวกรองที่ตอบสนองเชิงความถี่ (Finite Impulse Response: FIR) จะถูกออกแบบให้รับมือกับค่าสเปกตรัมซึ่งมีรูปแบบความถี่ไม่แน่นอนและมีความละเอียดของเวลาที่แตกต่างกัน โดยเฉพาะอย่างยิ่งความถี่ที่มีรายละเอียดใกล้เคียงกันกับช่องว่างของสัญญาณและช่วงเวลาที่มีค่าสัญญาณใกล้เคียงกันตั้งแต่ช่วงกลางของช่วงสัญญาณจนถึงจุดสุดท้ายท้ายของช่วงสัญญาณนั้นๆ

2.3. การปรับปรุงค่าลักษณะเด่นของสัญญาณเสียงพูด โดยการรวมค่าลักษณะเด่นเข้าด้วยกัน (Speech Feature Combination)

2.3.1. การวิเคราะห์องค์ประกอบหลัก (Principal Component Analysis: PCA) [12]

เนื่องจากในการรวมค่าลักษณะเด่นของสัญญาณสียงพูดนั้น บางครั้งได้ค่าลักษณะเด่นใหม่ที่มีขนาดใหญ่ขึ้น ซึ่งในบางครั้งข้อมูลที่ได้เพิ่มขึ้นมามีความแตกต่างกันเพียงเล็กน้อย ทำให้ระบบรับภาระในการทำงานเพิ่มมากขึ้นโดยไม่จำเป็น เพื่อที่จะช่วยให้ระบบสามารถจำแนกประเภทและแปลความหมายข้อมูลสัญญาณเสียงพูดได้อย่างมีประสิทธิภาพ โดย PCA จะทำการรวมข้อมูลที่มีความใกล้เคียงกันเข้าด้วยกันซึ่งสุดท้ายจะได้ขนาดของข้อมูลที่ลดลงแต่ยังคงค่าข้อมูลที่เหมือนต้นฉบับไว้เกือบทั้งหมด

2.3.1.1.**หลักการวิเคราะห์องค์ประกอบหลัก**

การวิเคราะห์องค์ประกอบหลักเป็นเทคนิคการแปลงข้อมูลที่ใช้กันอย่างแพร่หลาย ไม่ว่าจะเป็นงานทางด้านการลดขนาดของข้อมูล (Data Compression) หรืองานทางด้านการวิเคราะห์การเปลี่ยนแปลง (Change Analysis) โดยในวิทยานิพนธ์นี้จะนำการวิเคราะห์องค์ประกอบหลักมาทำการลดขนาดข้อมูลของสัญญาณเสียงพูด

การวิเคราะห์องค์ประกอบหลักนี้เป็นการแปลงเชิงเส้น (Linear Transformation) ที่ใกล้เคียงกันกับการวิเคราะห์เฟคเตอร์ (Factor Analysis) หรือเป็นการรวมน้ำหนักแบบเชิงเส้น ซึ่งหมายความว่ากระบวนการวิเคราะห์องค์ประกอบหลักจะทำการสร้างข้อมูลชุดใหม่ที่คูณสัญญาณเสียงเดิมในแต่ละค่าลักษณะเด่น ด้วยค่าน้ำหนัก และนำผลคูณที่ได้ในแต่ละค่าลักษณะเด่นมารวมกันเป็นสัญญาณเสียงใหม่ ดังสมการที่ 2-19

$$C = w_1 B_1 + w_2 B_2 + \ldots + w_n B_n \tag{2-19}$$

เมื่อค่าน้ำหนัก w_i ที่ใช้ในการแปลงนี้คือค่าอีลีเมนต์ต่างๆ ในแต่ละไอเกนเวกเตอร์ (Eigenvector) ของเมตริกซ์โควาเรียนซ์ (Covariance Matrix) ของสัญญาณเสียงพูดที่เป็นอินพุทเดิม หรือในอีกความหมายคือ กระบวนการวิเคราะห์องค์ประกอบหลักเป็นกระบวนการทางคณิตศาสตร์ที่ใช้สำหรับคำนวณหาชุดข้อมูลของแกนต่างๆ ในปริภูมิลักษณะ (Feature Space) ที่ไม่มีค่าคอร์รีเรชั่น (Correlation) ระหว่างกันและมีการเรียงข้อมูลสูงสุดไว้ที่องค์ประกอบหลักแรกไปจนไปต่ำสุดที่องค์ประกอบหลักสุดท้าย ซึ่งการลดขนาดข้อมูลจะมีขั้นตอนดังต่อไปนี้

2.3.1.1.1.**การคำนวณหาเมตริกซ์โควาเรียนซ์**

ซึ่งมีสูตรในการคำนวณหาดังนี้

$$C_x = \frac{1}{k-1}\sum_{i=1}^{k}(X_i - M)(X_i - M)^T \tag{2-20}$$

เมื่อ C_x คือเมตริกซ์โควาเรียนซ์

X คือการเปลี่ยนแปลง N มิติ

M คือค่าเฉลี่ยเวกเตอร์

k คือ จำนวนค่าลักษณะเด่น

องค์ประกอบหลัก Y_i สามารถเขียนได้ดังนี้คือ

$$Y_i = a_{1i}X_1 + a_{2i}X_2 + \ldots + a_{Ni}X_{N1} \tag{2-21}$$

หรือ

$$Y_i = a_i^T X \tag{2-22}$$

สำหรับการเปลี่ยนแปลงองค์ประกอบทั้งหมดสามารถเขียนได้ดังนี้

$$Y = A^T X \tag{2-23}$$

เมื่อ A คือเมตริกซ์ไอเกนเวกเตอร์

เมตริกซ์โควาเรียนซ์ของ Y ในมิติใหม่เป็นดังนี้

$$C_Y = AC_X A^T \tag{2-24}$$

เมื่อเมตริกซ์ C_X เป็นเมตริกซ์ไดอะโกนอล (Diagonal Matrix) ซึ่งจะประกอบด้วยค่าไอเกนเวลู (Eigenvalue) ต่าง

$$C_x = \begin{bmatrix} \lambda_1 & 0 & 0 & \ldots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ 0 & 0 & \lambda_3 & \ldots & 0 \\ 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \ldots & \lambda_N \end{bmatrix} \tag{2-25}$$

เมื่อ $\lambda_1 > \lambda_2 > \ldots > \lambda_N$

2.3.1.1.2. **การคำนวณหาไอเกนเวกเตอร์และไอเกนเวลู**

ค่าไอเกนเวกเตอร์และไอเกนเวลูขององค์ประกอบหลักต่างๆ คำนวณได้จากโควาเรียนซ์เมตริกซ์ที่ได้จากข้อมูลเดิม โดยใช้วิธีการของจาโคบี (Jacobi Method) จะได้ค่าไอเกนเวลูเท่ากับ

จำนวนค่าลักษณะเด่นของข้อมูลเดิมและแต่ละค่าไอเกนเวลูจะได้ค่าไอเกนเวกเตอร์เท่ากับจำนวนค่าลักษณะเด่นของข้อมูลเดิมเช่นกัน

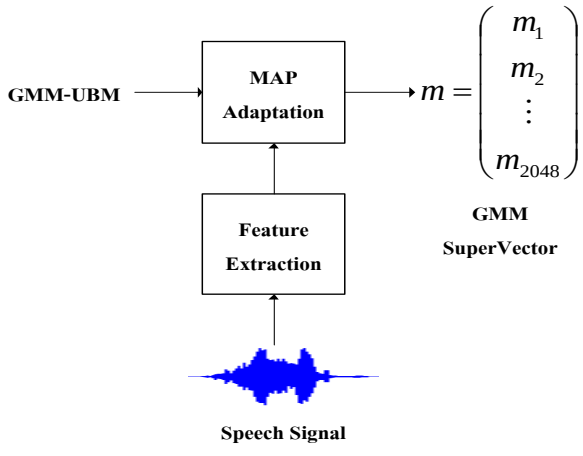
2.3.1.1.3. **วิธีการโปรเจคชั่น**

เมื่อมีการนำค่าไอเกนเวกเตอร์ของแต่ละองค์ประกอบหลักมาคูณกับข้อมูลเดิม (X) จะได้ข้อมูลใหม่ (Y) โดยสูตรการคำนวณข้อมูลใหม่ (Y_i)

$$(Y_i) = [X - \overline{X}] \times \begin{bmatrix} A_i \\ B_i \\ C_i \\ D_i \\ E_i \\ F_i \end{bmatrix} \qquad (2\text{-}26)$$

2.3.2. **ซูเปอร์เวกเตอร์ของการประมาณรูปร่างผสมของเกาส์ (Gaussian Mixture Model - Super Vector) [13]**

แนวคิดของซูเปอร์เวกเตอร์ของการประมาณรูปร่างผสมของเกาส์นั้น คือการแทนค่าฟีเจอร์ของการประมาณรูปร่างผสมของเกาส์ที่ครอบคลุมทุกแบบจำลอง (Gaussian Mixture Model Universal Background Model) ทั้งหมด ผ่านกระบวนการปรับแต่งความเป็นไปได้ภายหลังที่มากที่สุด (MAP Adaptation) เพื่อสร้างข้อมูลเวกเตอร์ขึ้นมาใหม่ โดยค่าข้อมูลในเวกเตอร์แต่ละตัวนั้นจะแทนด้วยค่าเฉลี่ยของฟีเจอร์ต่างๆ ซึ่งสุดท้ายจะได้เวกเตอร์ข้อมูลขนาด $1 \times n$ ดังที่แสดงดังรูปที่ 2-4



ภาพประกอบที่ 2-4 แนวคิดของซูเปอร์เวกเตอร์ของการประมาณรูปร่างผสมของเกาส์

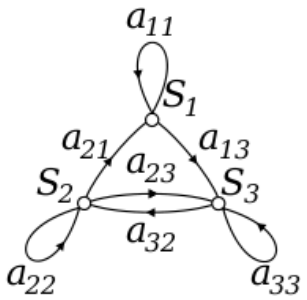
2.4. แบบจำลองฮิดเดนมาร์คอฟ [14]

HMM เป็นแบบจำลองทางสถิติรูปแบบหนึ่ง ที่ได้สร้างบนพื้นฐานของทฤษฎีมาร์คอฟ เป็นหนึ่งในแบบจำลองสำหรับจำแนกสัญญาณเสียงที่ดีที่สุดแบบหนึ่ง HMM นั้นเปรียบเสมือนตัวกำกับขั้นตอน (State Machine) ที่ได้กำหนดลำดับของหน่วยเสียงในการสร้างค่าลักษณะเด่นของสัญญาณเสียงพูดจากสัญญาณเสียงพูด โดยไม่ขึ้นอยู่กับเวลา อีกทั้งลักษณะการทำงานจะอาศัยการโปรแกรมแบบพลวัต (Dynamic Programing) ทำการประมวลผลเป็นไปได้อย่างรวดเร็ว

เหตุที่ทำให้ HMM เป็นที่นิยมนั้น เนื่องมาจาก HMM นั้นเป็นแบบจำลองที่อาศัยโครงสร้างทางคณิตศาสตร์และสามารถเปลี่ยนแปลงทฤษฎีพื้นฐานทำให้สามารถประยุกต์ใช้งานได้อย่างกว้างขวาง และสามารถทำงานได้อย่างมีประสิทธิภาพเมื่อนำไปประยุกต์ใช้งานกับงานที่เหมาะสม

2.4.1. **ประเภทของแบบจำลองฮิดเดนมาร์คอฟ**

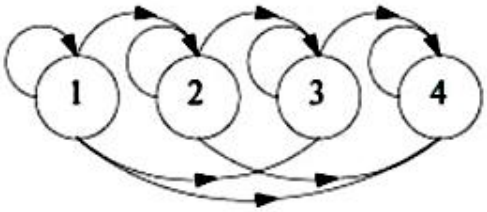
ขั้นตอนวิธีการของ HMM ส่วนใหญ่เป็นการพิจารณาเพียงกรณีพิเศษของแบบจำลองประเภทเออร์กอดิก ซึ่งเป็น HMM ที่ทุกๆ สถานะจะเชื่องโยงกันหมด โดยทุกสถานะของแบบจำลองสามารถเข้าถึงสถานะอื่นๆ ได้ในขั้นตอนเดียว ดังนั้นแบบจำลองแบบดออร์กอดิกจึงมีคุณสมบัติที่ทุกสถานะสามารถเข้าถึงได้จากสถานะอื่นๆ ด้วยขั้นตอนที่จำกัดแน่นอนดังรูปที่ 2-5 ซึ่งเป็นแบบจำลองที่มีจำนวนสถานะ $N = 3$ สถานะและมีคุณสมบัติเฉพาะที่สัมประสิทธิ์ α_{ij} ทั้งหมดมีค่าเป็นบวก



ภาพประกอบที่ 2-5 แบบจำลองที่มีเออร์กอดิก 3 สถานะ

ในการประยุกต์ใช้ HMM กับงานที่มีลักษณะนั้น ก็ยังมีแบบจำลองประเภทอื่นๆ ที่เหมาะสมกับคุณสมบัติที่สังเกตของสัญญาณ ซึ่งถูกจำลองมากกว่าแบบจำลองเออร์กอดิกมาตรฐาน ดังที่แสดงในรูปที่ 2-6 ซึ่งเป็นแบบจำลองประเภทซ้ายไปขวา (Left-Right Model) หรือแบบจำลองเบคิส (Bakis Model) เนื่องจากลำดับสถานะอยู่ภายในที่สัมพันธ์กับแบบจำลองมีคุณสมบัติ ดัชนีของสถานะจะเพิ่มขึ้น หรือมีค่าเท่าเดิม เมื่อระยะเวลาเพิ่มขึ้น เปรียบเสมือนกับสถานะดำเนินจาก

ซ้ายไปขวา ดังนั้นแบบจำลองซ้ายไปขวานี้ จึงมีคุณสมบัติที่เหมาะสมในการจำลองสัญญาณที่เปลี่ยนแปลงไปตามเวลา เช่น สัญญาณเสียงพูดเป็นต้น



ภาพประกอบที่ 2-6 แบบจำลองฮิดเดนมาร์คอฟที่มี 4 สถานะ

คุณสมบัติพื้นฐานของ HMM แบบซ้ายไปขวา สัมประสิทธิ์ของการเปลี่ยนสถานะจะต้องเป็นไปคุณสมบัติ ที่ไม่อนุญาตให้มีการเปลี่ยนแปลงไปยังสถานะที่มีดัชนีต่ำกว่าสถานะปัจจุบันตามสมการที่ 2-27 รวมทั้งค่าความน่าจะเป็นเริ่มต้นจะต้องมีคุณสมบัติเป็นไปตามสมการที่ 2-28

$$\alpha_{ij} = 0,\ j < i \tag{2-27}$$

$$\pi_i = \begin{cases} 0, & i \neq 1 \\ 1, & i = 1 \end{cases} \tag{2-28}$$

เนื่องจากลำดับสถานะจะต้องเริ่มต้นจากสถานะที่ 1 และสิ้นสุดในสถานะที่ N ดังนั้นในแบบจำลองประเภทซ้ายไปขวาจึงต้องเพิ่มเติมเงื่อนไขบังคับให้กับสัมประสิทธิ์ของการเปลี่ยนสถานะ เพื่อไม่ให้เกิดการเปลี่ยนแปลงมากจนเกินไปดังนี้

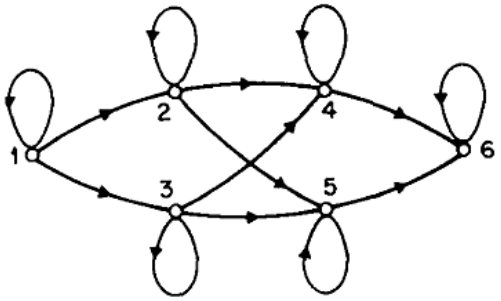
$$\alpha_{ij} = 0,\ j > i + \Delta \tag{2-29}$$

จากรูปที่ 7 จะกำหนดให้ค่า $\Delta = 2$ ซึ่งไม่อนุญาตให้มีการข้ามสถานะเกินกว่า 2 สถานะ จึงได้เมตริกซ์ของการเปลี่ยนแปลงสถานะดังสมการที่ 2-30 และสถานะสุดท้ายของแบบจำลองประเภทซ้าย-ขวาจะมีค่าสัมประสิทธิ์ของการเปลี่ยนแปลงสถานะมีค่าเฉพาะดังสมการที่ 2-31

$$\alpha_{NN} = 1 \tag{2-30}$$

$$\alpha_{Ni} = 0, i < M \quad (2\text{-}31)$$

นอกจากนั้นยังมีรูปแบบของแบบจำลองที่เป็นไปได้อีกมากมาย ตัวอย่างดังรูปที่ 2-7 ซึ่งแสดงถึงการเชื่อมต่อข้ามแบบจำลองประเภทขนานซ้ายไปขวาสองชุด แต่แบบจำลองนี้ยังคงจัดอยู่ในรูปแบบจำลองซ้ายไปขวาเพียงแต่มีความยืดหยุ่นมากขึ้น อย่างไรก็ตามการข้ามสถานะตามเงื่อนไขแบบจำลองนี้ไม่มีผลต่อกระบวนการประมาณค่าใหม่



ภาพประกอบที่ 2-7 แบบจำลองแบบเส้นขนานซ้ายไปขวาที่มี 6 สถานะ

ฮิดเดนมาร์คอฟให้คำจำกัดความของการกระทำหรือเหตุการณ์ที่เป็นไปได้เหล่านั้นว่าเป็นสถานะซ่อน (Hidden State) ซึ่งแต่ละสถานะซ่อนจะมีเส้นโยงถึงกันรวมทั้งมีความน่าจะเป็นกำกับ ฮิดเดนมาร์คอฟได้มีการนำมาประยุกต์ใช้กับงานทางด้านการจับคู่รูปแบบ (Pattern Matching) ที่เปลี่ยนแปลงตามเวลาเช่นการรู้จำเสียงพูด รู้จำท่าทาง ฯลฯ โดยมีองค์ประกอบที่สำคัญดังนี้

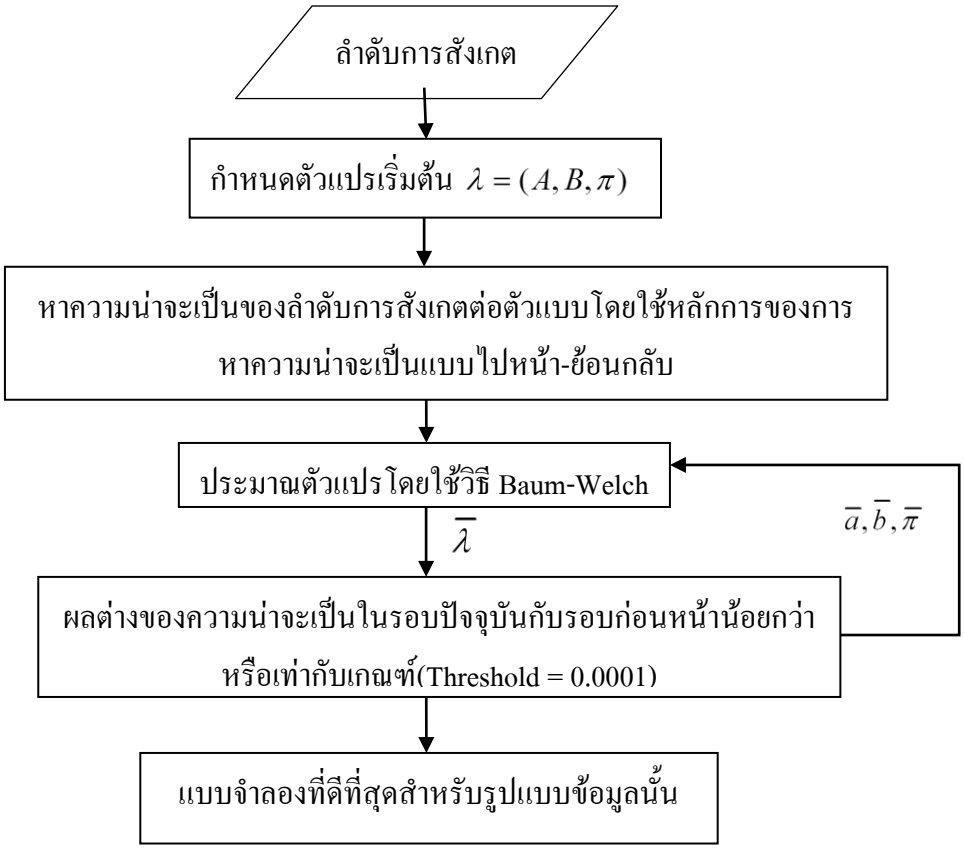
- N คือจำนวนสถานะในตัวแบบสถานะจะเปลี่ยนไปตามเวลา t
- M คือจำนวนของเหตุการณ์ต่อสถานะเหตุการณ์ที่ได้จะสอดคล้องกับอินพุตที่ป้อนให้กับแบบจำลอง
- a_{ij} คือความน่าจะเป็นในการเปลี่ยนสถานะจากสถานะ i ไปสถานะ j
- $b_j(O_t)$ คือค่าความน่าจะเป็นของเหตุการณ์ O ที่เกิดขึ้นในสถานะ j
- π_i คือความน่าจะเป็นที่จะเกิดสถานะแรก

ส่วนประกอบต่างๆ ของแบบจำลองสามารถรวมเข้าด้วยกันแล้วแทนด้วยสมการที่ 2-32ซึ่งแสดงเป็นเซตของตัวแปรที่เสร็จสมบูรณ์ของ HMM ในการจดจำรูปแบบของข้อมูล ซึ่ง HMM สามารถแบ่งการทำงานออกเป็น 2 ขั้นตอนหลักๆ ได้แก่การฝึกฝนแบบจำลอง (Training) และการทดสอบการจดจำของแบบจำลอง (Evaluation)

$$\lambda = (A, B, \pi) \tag{2-32}$$

2.4.2. **การฝึกสอนแบบจำลอง**

การฝึกฝนแบบจำลองให้จดจำรูปแบบข้อมูลเป็นการประมาณตัวแปรของแบบจำลองที่ดีที่สุดสำหรับรูปแบบของข้อมูลในแต่ละรูปแบบซึ่งขั้นตอนการฝึกฝนให้กับแบบจำลองแสดงดังรูปที่ 2-8 โดยอินพุตของแบบจำลองคือลำดับการสังเกตของข้อมูล (Observation Sequence) จากนั้นกำหนดแบบจำลองเริ่มแรก λ ซึ่งประกอบด้วย $\lambda = (A, B, \pi)$ แล้วคำนวณหาค่าความน่าจะเป็นของอินพุตต่อแบบจำลองนั้นโดยใช้หลักการของการหาความน่าจะเป็นแบบไปหน้าและย้อนกลับ (Forward-Backward Algorithm) เมื่อได้ค่าความน่าจะเป็นของอินพุตต่อแบบจำลองจึงเข้าสู่กระบวนการประมาณตัวแปรโดยใช้วิธีของ Baum-Welch เพื่อทำการปรับตัวแบบจนกระทั่งได้แบบจำลองที่ดีที่สุดสำหรับอินพุตนั้น



ภาพประกอบที่ 2-8 ขั้นตอนการฝึกฝนแบบจำลอง

การคำนวณหาค่าความน่าจะเป็นของอินพุตหรือลำดับการสังเกตของแบบจำลอง โดยใช้หลักการหาความน่าจะเป็นแบบไปหน้า-ย้อนกลับมีรายละเอียดดังนี้

ในกระบวนการหาความน่าจะเป็นแบบไปหน้า (Forward Algorithm) มีการกำหนดตัวแปรดังนี้

- $\alpha_t(i)$ เป็นตัวแปรแบบไปหน้า ณ เวลา t ที่สถานะ i
- π_i เป็นค่าความน่าจะเป็น ณ เวลา $t=1$ ที่สถานะ i
- $b_j(O_t)$ เป็นค่าความน่าจะเป็นของข้อมูล O ณ เวลา t ที่สถานะ j
- a_{ij} เป็นค่าความน่าจะเป็นในการเปลี่ยนจากสถานะ i ไปที่สถานะ j
- ขั้นเริ่มต้น

$$\alpha_1(i) = \pi_i b_i(O_1), \quad 1 \le i \le N \tag{2-33}$$

- ขั้นการวนซ้ำ

$$\alpha_{t+1}(j) = [\sum_{i=1}^{N} \alpha_t(i) a_{ij}] \cdot b_j(O_{t+1}), 1 \le t \le T-1, 1 \le j \le N \tag{2-34}$$

- ขั้นสิ้นสุด

$$P(O \mid \lambda) = \sum_{i=1}^{N} \alpha_T(i) \tag{2-35}$$

ส่วนของกระบวนการหาความน่าจะเป็นแบบย้อนกลับ (Backward Algorithm) ได้มีการกำหนดตัวแปรเพิ่มเติมจากกระบวนการหาความน่าจะเป็นแบบไปหน้า ดังนี้

- $\beta_t(i)$ เป็นตัวแปรแบบย้อนกลับ ณ เวลา t ที่สถานะ i
- ขั้นเริ่มต้น

$$\beta_{T(i)} = 1, \quad 1 \le i \le N \tag{2-36}$$

- ขั้นอุปนัย

$$\beta_t(i) = \sum_{j=1}^{N} a_{ij} b_j(O_{t+1}) \beta_{t+1}(j), \quad t = T-1, T-2, \ldots, 1, \quad 1 \le i \le N \tag{2-37}$$

หลังจากกระบวนการหาความน่าจะเป็นแบบไปหน้า-ย้อนกลับ แล้วจะเข้าสู่กระบวนการประมาณค่าพารามิเตอร์โดยใช้วิธี Baum-Welch โดยนำผลลัพธ์ที่ได้จากก่อนหน้านี้ มาใช้ซึ่งมีรายละเอียดดังนี้

- กำหนดให้ ξ_t(i,j) คือความน่าจะเป็นของการอยู่ในสถานะ i ที่เวลา t และสถานะ j ที่เวลา $t+1$

$$\xi_t(i, j) = \frac{\alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)}{\sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)} \tag{2-38}$$

- กำหนดตัวแปร $\gamma_t(i)$ เป็นความน่าจะเป็นของการอยู่ในสถานะ i ที่เวลา t โดยที่ตัวแบบ λ และลำดับของเหตุการณ์ O มีความสัมพันธ์กับ $\xi_t(i, j)$ โดยการบวกกันทุก j

$$\gamma_t(i) = \sum_{j=1}^{N} \xi_t(i, j) \tag{2-39}$$

- การประมาณตัวแปรใหม่ของ A, B และ π

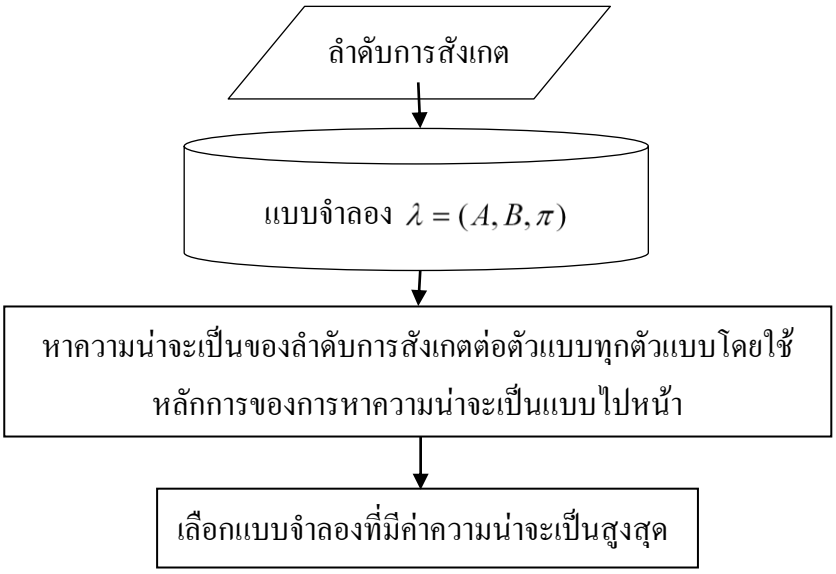
$$\bar{\pi}_i = \gamma_1(i) \tag{40}$$

$$\bar{a}_{ij} = \frac{\sum_{t=1}^{T-1} \xi_t(i, j)}{\sum_{t=1}^{T-1} \gamma_t(i)} \tag{41}$$

$$\bar{b}_j(k) = \frac{\sum_{t=1}^{T} \gamma_t(j)}{\sum_{t=1}^{T} \gamma_t(j)} \tag{42}$$

2.4.3. **การทดสอบการจดจำของแบบจำลอง**

หลังจากที่ได้ทำการฝึกฝนรูปแบบข้อมูลให้กับแบบจำลองแล้วจึงทำการทดสอบการจดจำของแบบจำลองโดยการคำนวณหาความน่าจะเป็นของเหตุการณ์ต่อแบบจำลองโดยใช้หลักการหาความน่าจะเป็นแบบไปหน้ากับทุกๆ แบบจำลอง แบบจำลองใดที่ได้ค่าความน่าจะเป็นสูงสุด ถือว่าเป็นคำตอบของลำดับสังเกตนั้น รายละเอียดของขั้นตอนการทดสอบการจดจำของแบบจำลองอธิบายได้ด้วยภาพประกอบที่ 2-9



ภาพประกอบที่ 2-9 ขั้นตอนการทดสอบการรู้จำของแบบจำลอง

2.5. ซัพพอร์ตเวกเตอร์แมชชีน [15]

SVM เป็นเครื่องมือที่ใช้สำหรับงานทางด้านการจำแนกรูปแบบ (Pattern Recognition) ซึ่งมีแนวคิดคือ การสร้างระนาบหลายมิติที่ใช้ในการแบ่งข้อมูลออกเป็นสองประเภท (Class) โดย SVM จะพยายามที่จะสร้างปริภูมิสมมุติฐาน (Hypothesis Space) ของฟังก์ชันของกฎการจำแนก (Classifier Rule) $h: R^n \to \{\pm 1\}$ โดยกำหนดให้มี l เป็นค่าสังเกต (Observation) แต่ละค่าสังเกตประกอบด้วยคู่ของเวกเตอร์ $\bar{x}_i \in R^n$ และสัญลักษณ์แสดงชนิด (Class Label) $y_i \in \{\pm 1\}$ ดังนั้นจะได้ว่า

$$(\bar{x}_1, y_1), \ldots, (\bar{x}_l, y_l) \in R^n \times \{\pm 1\} \tag{2-43}$$

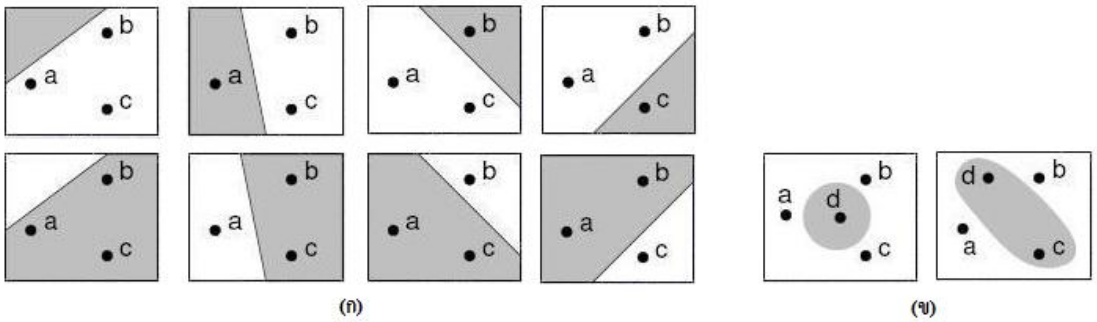
เราจะกล่าวได้ว่า h จำแนกตัวอย่าง $(\bar{x}, y)$ ได้ถูกต้องเมื่อ

$$h(\bar{x}) = y \tag{2-44}$$

โดย $(\bar{x}, y) \sim P(\bar{x}, y)$ ซึ่งได้มาจากการฝึกฝน $(\bar{x}_1, y_1),\ldots(\bar{x}_l, y_l)$

2.5.1. **มิติ Vapnik/Chervonekis (VC-Dimension)**

ปริมาณที่ใช้วัดความสมบูรณ์ (Richness) หรือความเปลี่ยนแปลง (Flexibility) ของชนิดฟังก์ชันส่งผลถึงความจุขอระบบรู้จำเรียกว่ามิติ VC การควบคุมความจุได้ย่อมทำให้ประสิทธิภาพของระบบดีขึ้น ตัวอย่างมิติ VC ของปริภูมิสมมุติฐานในกรณีนี้มี 2 ชนิด จะเท่ากับจำนวนมากที่สุด, d ของตัวอย่างที่สามารถแบ่งออกได้เป็น 2 กลุ่มด้วยวิธี 2^d วิธีใดๆ โดยใช้ปริภูมิสมมติฐานนั้น มิติ VC ของปริภูมิสมมุติฐานใน R^n คือ $N+1$ เช่นใน R^2 จะมีมิติ VC =3 ดังรูปที่ 2-10 จะพบว่ามีเพียง 3 ตัวอย่างเท่านั้นที่เป็นจำนวนมากที่สุดที่สามารถแบ่งออกได้ 2^3=8 วิธี ส่วนกรณีที่ 4 ตัวอย่างนั้นทำไม่ได้



ภาพประกอบที่ 2-10 (ก) มิติ VC ใน R^n กรณี 3ตัวอย่างใน R^n (ข) กรณี 4ตัวอย่างใน R^n

ถ้า $h \in H$ และ H เป็นปริภูมิสมมุติฐานซึ่งมีมิติ VC เท่ากับ d สำหรับทุกๆ การกระจายความน่าจะเป็น D บน $X \times \{-1, 1\}$ ด้วยความน่าจะเป็น $1-\delta$ บนตัวอย่างสุ่มจำนวน l ค่าความน่าผิดพลาดไม่น่าจะเกิน

$$err(h) \le g(l, H, \delta) = \frac{1}{2}(d \log\frac{2el}{d} + \log\frac{2}{\delta}), \quad d \le 1, \quad l > \frac{2}{\delta} \tag{2-45}$$

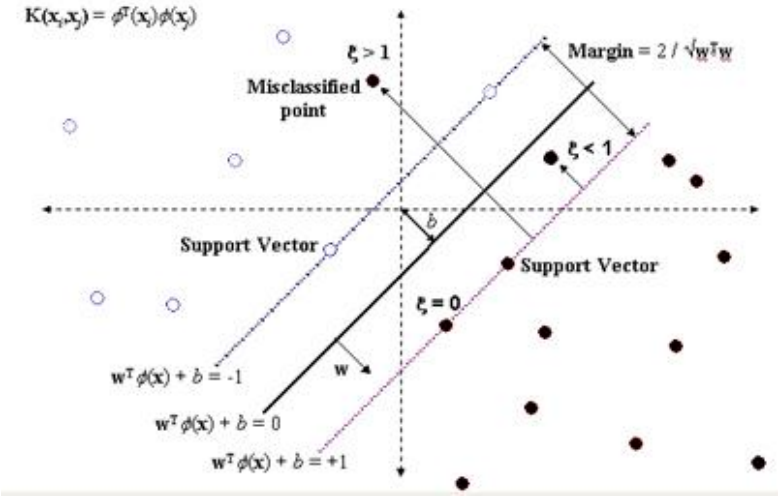
2.5.2. **ซัพพอร์ตเวกเตอร์แมชชีนเชิงเส้น**

SVM เชิงเส้นคือ SVM ที่มีไฮเปอร์เพลน (Hyperplane) เราสมมติให้มีเซตของข้อมูล D ที่ประกอบด้วยตัวอย่างจำนวน 1 ตัวในปริภูมิอันดับที่ n ที่มีสองประเภทคือ $+1, -1$

$$D = \left\{(x_k, y_k) \mid k \in \{1,\ldots,l\},\ x_k \in R^n,\ y_k \in \{+1,-1\}\right\} \tag{2-46}$$

2.5.2.1. **กรณีสามารถแยกกันได้**

เป็นการพิจาณาในกรณีที่ง่ายที่สุดเนื่องจากข้อมูลมีลักษณะแยกกันอย่างสมบูรณ์ซึ่งสามารถจำแนกประเภทชนิดได้ด้วยรูปร่างทางคณิตศาสตร์ (ไฮเปอร์เพลน) ดังภาพประกอบที่ 2-11



ภาพประกอบที่ 2-11 ไฮเปอร์เพลนแบบเชิงเส้นกรณีแยกกันได้

ระนาบหลายมิติในปริภูมิอันดับ n จะถูกกำหนดโดย (w,b) เมื่อ w คือเวกเตอร์ในปริภูมิอันดับ n ที่ตั้งฉากกับระนาบหลายมิติและ b คือค่าคงที่ระนาบหลายมิติ $(w,b)+b$ จะแบ่งข้อมูลได้ก็ต่อเมื่อ

$$\begin{aligned}(w\cdot b_i)+b>0 \text{ ถ้า } y_i=+1\\ (w\cdot b_i)+b<0 \text{ ถ้า } y_i=-1\end{aligned} \tag{2-47}$$

ในกรณีที่ต้องการค่า w และ b ที่ทำให้จุดที่ใกล้ระนาบหลายมิติมากสุดมีระยะ $\frac{1}{|w|}$ แล้วจะได้

$$\begin{aligned}(w\cdot b_i)+b \geq 1 \text{ ถ้า } y_i = +1\\ (w\cdot b_i)+b \leq 1 \text{ ถ้า } y_i = -1\end{aligned} \quad (2\text{-}48)$$

ซึ่งจะเหมือนกับ

$$y_i[(w\cdot b_i)+b] > 1 \quad \forall_i \quad (2\text{-}49)$$

ในการค้นหาระนาบหลายมิติที่ใช้ในการแบ่งข้อมูลที่ดีที่สุด จะต้องทำการค้นหาระนาบหลายมิติที่ระยะห่างระหว่างข้อมูลที่ใช้ในการฝึกฝนกับระนาบที่น้อยที่สุดที่มีค่ามากสุด ดังรูปที่ 2-11 ระยะห่างระหว่างข้อมูลตัวอย่างสองตัวจากประเภทที่แตกต่างกันมีค่าเท่ากับ

$$d(w,b) = \min_{\{x_i / y_i=1\}} \frac{(w\cdot x_i)+b}{|w|} - \max_{\{x_i / y_i=-1\}} \frac{(w\cdot x_i)+b}{|w|} \quad (2\text{-}50)$$

จากสมการที่ 2-48 ค่าที่น้อยที่สุดและมากที่สุดทีเหมาะสมคือ ± 1 ดังนั้นจำเป็นต้องเพิ่มค่าของฟังก์ชัน

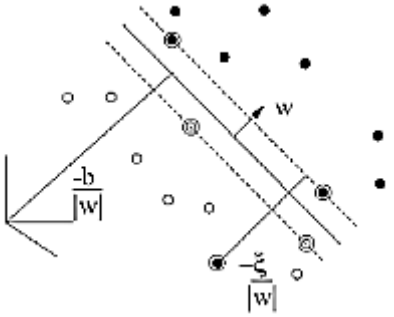
$$d(w,b) = \frac{1}{|w|} - \frac{-1}{|w|} = \frac{2}{|w|} \quad (2\text{-}51)$$

เพื่อให้ได้ค่าสูงที่สุด ซึ่งจะเท่ากับการลดค่า $|w|^2/2$ ให้ต่ำสุดโดยมีเงื่อนไขตามสมการที่ 2-52

$$y_i[(w\cdot x_i)+b] \geq 1 \quad (2\text{-}52)$$

2.5.2.2. **กรณีที่ไม่สามารถแยกออกจากกันได้**

SVM ที่ได้กล่าวไปข้างต้นนั้น สามารถทำงานได้ในการแบ่งส่วนประเภทข้อมูลที่สมบูรณ์เท่านั้น แต่กรณีที่ไม่สามารถแยกจุดตัวอย่างออกจากกันได้ด้วยไฮเปอร์เพลน จะสามารถทำได้เพียงหาไฮเปอร์เพลนที่สามารถแยกจุดตัวอย่างออกจากกันให้ได้มากที่สุดและยอมให้มีจุดตัวอย่างบางส่วนน้อยเพียงบางจุดที่ผิดพลาดดังแสดงในรูปที่ 2-12



ภาพประกอบที่ 2-12 ไฮเปอร์เพลนเชิงเส้น กรณีแยกออกจากกันไม่ได้

สำหรับในกรณีนี้ จะต้องทำการปรับเงื่อนไขใหม่โดยการเพิ่มพจน์ค่าปรับซึ่งประกอบด้วยผลรวมของค่าความคลาดเคลื่อน ζ_i จากขอบเขตกว้างเพิ่มเข้าไปด้วย ซึ่งปัญหาในตอนนี้คือ การลดค่าของ $\frac{|w|^2}{2} + c\sum_{i=1}^{l}\zeta_i$ ให้ต่ำที่สุด โดยการลดค่านั้นจะขึ้นกับเงื่อนไขนั้นคือ

$$y_i[(w \cdot x_i) + b] \geq 1 - \zeta, \quad \zeta_i \geq 0 \quad \forall_i \tag{2-53}$$

พจน์ค่าปรับสำหรับตัวอย่างข้อมูลที่ใช้ในการฝึกฝนที่ทำนายผิดพลาดจะถูกเพิ่มน้ำหนักโดยค่าคงที่ C ซึ่งเป็นค่าถ่วงน้ำหนักสำหรับการให้ความสำคัญระหว่างระยะที่ใช้ในการแยกแยะและค่าความผิดพลาด การเลือกค่า C ที่สูงนั้นจะมีผลทำให้เกิดค่าความคลาดเคลื่อน ζ_i และเพิ่มการคำนวณโดยทำให้การค้นหาวิธีที่จะลดจำนวนตัวอย่างข้อมูลฝึกฝนที่ทำนายผิดพลาดเพิ่มมากขึ้นจะจำเป็นต้องนำวิธีของลากอง (Legong) มาใช้ในการแก้ปัญหาดังกล่าว ด้วยการลดค่าสมการที่ 2-54 ให้มีค่าต่ำที่สุด

$$L(w,b,\alpha) = \sum_{i=1}^{l}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{l}\alpha_i\alpha_j y_i y_j (x_i \cdot x_j), \quad 0 \leq \alpha_i \leq C \; \forall_i, \quad \sum_{i=1}^{l}\alpha_i y_i = 0 \tag{2-54}$$

เมื่อ α^0 ซึ่งเป็นเวกเตอร์ในปริภูมิลำดับที่ 1 แทนค่าต่ำสุดของ $L(w,b,\alpha)$ ถ้า $\alpha_i^0 > 0$ แล้ว x_i เป็นซัพพอร์ตเวกเตอร์ของระนาบหลายมิติที่แบ่งแยกดีที่สุด (w^0, b^0) สามารถเขียนได้ในเทอมของ α^0 และข้อมูลฝึกฝนโดยเฉพาะอย่างยิ่งในเทอมของซัพพอร์ตเวกเตอร์ ดังที่แสดงในสมการที่ 2-55 และ 2-56

$$w^0 = \sum_{i=1}^{l} \alpha_i^0 y_i x_i = \sum_{SupportVector} \alpha_i^0 y_i x_i \qquad (2\text{-}55)$$

$$b^0 = 1 - w^0 x_i, \quad x_i, y_i = 1, \quad 0 < \alpha_i < C \qquad (2\text{-}56)$$

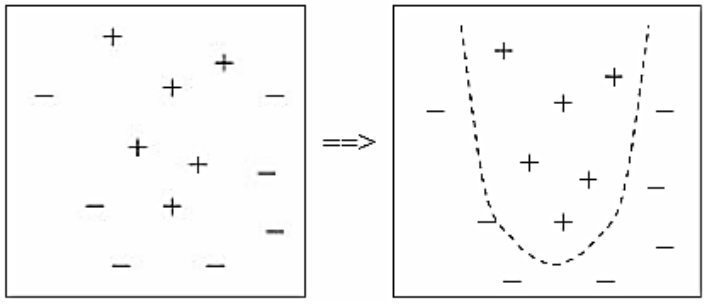
ระนาบหลายมิติที่แบ่งแยกดีที่สุดจะจำแนกจุดต่างๆ ตามเครื่องหมายของผลลัพธ์ของฟังก์ชัน $f(x)$ ดังสมการที่ 2-57

$$f(x) = sign(w^0 \cdot x + b^0) = sign\left(\sum_{SupportVector} \alpha_i^0 y_i (x_i, x) + b^0 \right) \qquad (2\text{-}57)$$

ซึ่งซัพพอร์ตเวกเตอร์ x_i ที่มี $\alpha_i^0 = C$ อาจจะถูกจำแนกผิดพลาดหรือไม่ก็ได้ แต่ x_i ตัวอื่นๆ นั้นจะสามารถจำแนกได้อย่างถูกต้อง

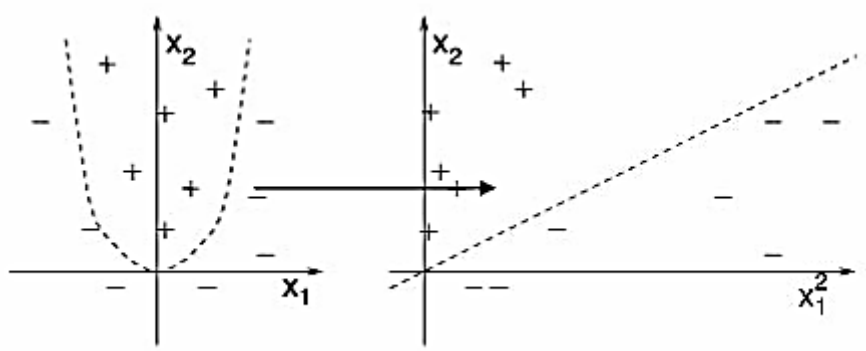
2.5.3. **ซัพพอร์ตเวกเตอร์แมชชีนแบบไม่เป็นเชิงเส้นและฟังก์ชันเคอร์เนล**

ในงานบางประเภทมีลักษณะที่ไม่เป็นเชิงเส้น จึงไม่สามารถใช้ไฮเปอร์เพลนมาจำแนกประเภทได้ ดังที่แสดงในรูปที่ 2-13



ภาพประกอบที่ 2-13 โครงสร้างข้อมูลแบบไม่เป็นเชิงเส้น

เพื่อการหาระนาบหลายมิติเชิงเส้นที่เหมาะสม จึงต้องใช้เทคนิคเข้ามาช่วย นั้นคือการแมป (Map) ข้อมูลตัวอย่างไปยังปริภูมิอันดับสูง โดยใช้ฟังก์ชันการแมป (Mapping Function) Φ แล้วจึงค่อยทำการฝึกฝนและจำแนกข้อมูลดังภาพประกอบที่ 2-14



ภาพประกอบที่ 2-14 การส่งปริภูมิขาเข้า (ซ้าย) ไปสู่ปริภูมิลักษณะ (ขวา)

คุณสมบัติที่ดีอีกประการหนึ่งของ SVM คือ ไม่จำเป็นต้องรู้รูปแบบที่ชัดเจนของ Φ แต่จะต้องนิยามผลคูณภายในมิติปริภูมิอันดับสูงซึ่งเรียกว่าฟังก์ชันเคอร์เนล (Kernel Function) เท่านั้น ดังสมการที่ 2-58

ฟังก์ชันเคอร์เนล: $$K(\vec{x}_i, \vec{x}_j) = \Phi(\vec{x}_i)\Phi(\vec{x}_j) \quad (2\text{-}58)$$

ซึ่งฟังก์ชันการตัดสินใจที่ได้จะเป็นดังสมการที่ 2-59

$$f(x) = sign(\sum_{i=1}^{l} \alpha_i y_i K(x_i, x) + b) \quad (2\text{-}59)$$

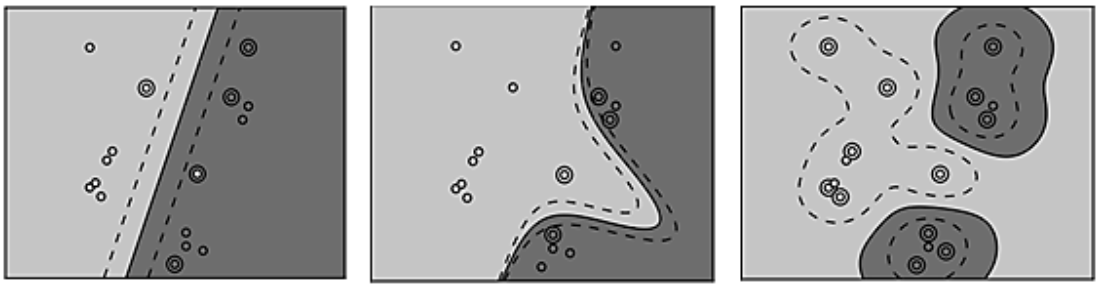
ซึ่งเคอร์เนลมีหลากหลายรูปแบบขึ้นอยู่กับการใช้งาน ดังที่แสดงในรูปที่ 2-15 เช่น

Linear $$K(\vec{x}_i, \vec{x}_j) = \vec{x}_i \cdot \vec{x}_j \quad (2\text{-}60)$$

Polynomial $$K(\vec{x}_i, \vec{x}_j) = [\vec{x}_i \cdot \vec{x}_j + 1]^d \quad (2\text{-}61)$$

Radial Basis Function (RBF)
หรือ Exponential Kernel $$K(\vec{x}_i, \vec{x}_j) = \exp(-\left\|\vec{x}_i \cdot \vec{x}_j\right\|^2 / (2\sigma^2)) \quad (2\text{-}62)$$

Sigmoid $$K(\vec{x}_i, \vec{x}_j) = \tanh[\gamma(\vec{x}_i - \vec{x}_j) + c] \quad (2\text{-}63)$$



ภาพประกอบที่ 2-15 รูปฟังก์ชันแบบ Linear, Polynomial และ RBF ตามลำดับ

อย่างไรก็ตาม ฟังก์ชันที่จะสามารถเป็นเคอร์เนลได้นั้น จะต้องเป็นไปตามทฤษฎีของเมอร์เซอร์ (Mercer's Theorem) ซึ่งกล่าวไว้ว่าสำหรับทุกๆ ฟังก์ชันที่สมมาตร $K(x,y)$ ในปริภูมินำเข้าสามารถแสดงผลคูณภายในของปริภูมิอันดับสูงนั้น ถ้า

$$\iint K(x,y)g(x)g(y)dxdy \geq 0 \tag{2-64}$$

สามารถหาได้สำหรับทุกๆ $g \neq 0$ สำหรับ $\int g^2(u)du \leq \infty$ แล้ว ฟังก์ชันเคอร์เนล K สามารถถูกขยายออกมาในรูปของ Φ_i

$$K(x,y) = \sum_{i=1}^{\infty} \lambda_i \Phi_i(x)\Phi_i(y) \tag{2-65}$$

ด้วย $\lambda_i \geq 0$ ในกรณีนี้ ฟังก์ชันการแมปจากปริภูมินำเข้าไปยังปริภูมิอันดับสูงจะสามารถอธิบายได้เป็น

$$\Phi : x \rightarrow (\sqrt{\lambda_1}\Phi_1(x), \sqrt{\lambda_2}\Phi_2(x), \ldots) \tag{2-66}$$

K สามารถเป็นผลคูณภายในได้ว่า

$$\Phi(x) \cdot \Phi(y) = \sum_{i=1}^{\infty} \lambda_i \Phi_i(x)\Phi_i(y) = K(x,y) \tag{2-67}$$

2.6. การวิเคราะห์เส้นโค้งคุณลักษณะสมบัติการทำงาน และความสัมพันธ์ของพื้นที่ใต้ส่วนโค้ง

คุณลักษณะสมบัติการทำงาน [16]

การวัดประสิทธิภาพการทำงานที่นิยมและใช้อย่างแพร่หลายอีกวิธีหนึ่งคือการวิเคราะห์เส้นโค้งคุณลักษณะสมบัติการทำงาน (Receiver Operating Characteristic Curve: ROC) และความสัมพันธ์ของพื้นที่ใต้ส่วนโค้งของค่าคุณลักษณะสมบัติการทำงาน (Area Under the ROC Curve: AUC) ซึ่งค่าทั้งสองนี้จะใช้ในการวัดความสามารถในการคัดแยก (Classification Performance) โดยเส้นโค้ง ROC นั้นจะแทนเส้นโค้งความเที่ยง (Precision) และการจดจำ (Precision-recall) ซึ่งในการทดลองภาวะถ่วงดุลระหว่างค่าจริงด้านบวก (True Positive Rate: TPR) และค่าเท็จด้านบวก (False Positive Rate: FPR) ซึ่งในระบบค้นหาคำหลักบนสื่อเสียงนั้นได้มีการใช้เพื่อวิเคราะห์ความถูกต้องในการค้นหาคำหลัก ซึ่งในงานวิจัยนี้ได้ใช้เมตริกความสับสน (Confusion Matrix) ในการสรุปจำนวนคำที่พบได้อย่างถูกต้องหรือจำนวนคำที่พบที่ผิดพลาด โดยตารางความสับสนได้แสดงดังในตารางที่ 2-1

ตารางที่ 2-1ตารางแสดงเมตริกความสับสน

		ผลที่ได้จากการกำหนด	
		+	-
ผลที่ได้จากการค้นหาจริง	+	TP (++)	FN (+-)
	-	FP (-+)	TN (--)

โดยข้อมูลในแต่ละช่องจะแทนด้วยจำนวนนับ ซึ่งมีค่าต่างๆ ได้แก่

- ค่าจริงบวก (True Positive) เป็นจำนวนคำที่เราต้องหาหา และระบบสามารถคำค้นหาได้อย่างถูกต้อง
- ค่าเท็จลบ (False Negative) เป็นจำนวนคำที่เราต้องการค้นหา แต่ระบบไม่สามารถค้นหาคำได้อย่างถูกต้อง
- ค่าเท็จบวก (False Positive) เป็นจำนวนคำที่เราไม่ได้ต้องการค้นหาแต่ระบบแจ้งว่าเป็นคำที่เราต้องการค้นหา
- ค่าจริงลบ (True Negative) เป็นจำนวนคำที่เราไม่ต้องการค้นหา และระบบแจ้งว่าเป็นคำดังกล่าวเป็นคำที่เราไม่ต้องการค้นหา

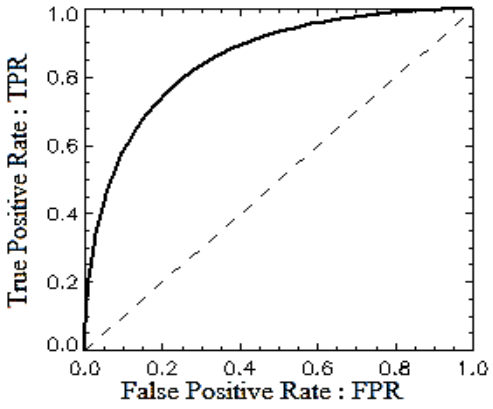
ในการทดลองนี้จะเป็นการวัดค่าความเที่ยงตรงและค่าความจดจำ ซึ่งค่าทั้งสองคำนวณได้จากสมการที่ 2-68

$$TPR = \frac{TP}{(TP + FP)}$$
$$FPR = \frac{TP}{(TP + FN)} \qquad (2\text{-}68)$$

ค่าความเที่ยงตรง จะใช้ในการพิจารณาหาค่าบวก (Positive) ในการแยกคำที่ถูกระบุโดยคำที่ต้องการค้นหาให้เป็นคลาสบวก (Positive Class) ค่าความเที่ยงตรงที่มากขึ้นนั้นหมายถึงความผิดพลาดที่เกิดขึ้นมนการค้นหาคำนั้นมีน้อย

ในส่วนของค่าความจดจำนั้น จะใช้ในการหาคำที่ต้องการที่มีตัวอย่างด้านบวก (Positive Example) ที่ทำการค้นหาได้อย่างถูกต้อง ซึ่งยิ่งค่าความจดจำนั้นมีค่าสูง นั้นแสดงว่าการค้นหาข้อมูลที่ผิดพลาดนั้นมีน้อยเช่นเดียวกัน

สำหรับเส้นโค้งค่าคุณลักษณะสมบัติการทำงานมีลักษณะการแสดงผลเป็นเส้นกราฟ ซึ่งเป็นเส้นกราฟที่แสดงภาวะถ่วงดุลระหว่างค่าอัตราจริงบวกและอัตราเท็จบวก โดยอัตราบวกจริงจะถูกแสดงบนแกน y และอัตราเท็จบวกจะแสดงผลบนแกน x ซึ่งแต่ละจุดจะวาดบนเส้นโค้ง ดังที่แสดงในรูปที่ 2-16 กราฟเส้นโค้งค่าคุณลักษณะสมบัติการทำงานที่ดีนั้นควรจะอยู่จะอยู่ใกล้เคียงมุมบนซ้ายได้มากที่สุด



ภาพประกอบที่ 2-16 ภาพเส้นโค้งค่าคุณลักษณะสมบัติการทำงาน

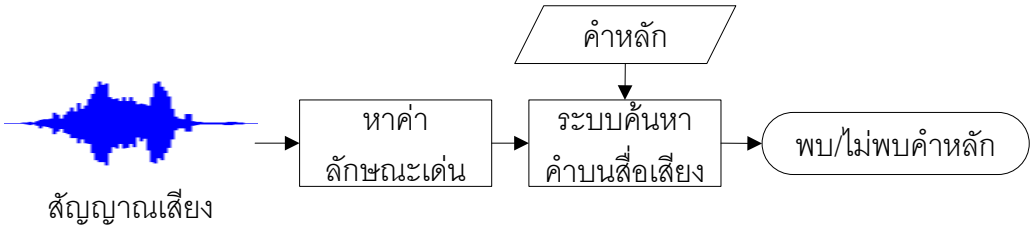
2.7. สรุป

เนื้อหาภายในบทนี้ได้กล่าวถึงระบบที่จะเป็นพื้นฐานของระบบค้นหาคำหลักบนสื่อเสียง ซึ่งเป็นงานวิจัยชิ้นนี้คือ กระบวนการรู้จำเสียงพูดซึ่งใช้ HMM กระบวนการแยกแยะประเภทของ SVM อีกทั้งยังได้กล่าวถึงรวมค่าลักษณะเด่นของเสียงพูดแบบต่างๆ อีก และการวัดค่าประสิทธิภาพการทำงานของระบบที่ใช้ในการทดลองครั้งนี้อีกด้วย

บทที่ 3
การออกแบบและพัฒนาระบบ

งานวิจัยชิ้นนี้ เป็นการทดสอบระบบค้นหาคำหลักบนสื่อเสียงที่มีระบบการทำงานที่ต่างกัน โดยทดสอบบนระบบค้นหาคำหลักบนสื่อเสียงที่ใช้ระบบรู้จำเสียงพูดเป็นพื้นฐาน และระบบค้นหาคำหลักบนสื่อเสียงที่ใช้วิธีการจำแนกประเภทโดยใช้เคอร์เนลและขอบเขตการแยกแยะที่มีขนาดกว้างที่สุดเป็นพื้นฐานร่วมกับค่าลักษณะเด่นของสัญญาณเสียงพูดประเภทต่างๆ เพื่อที่จะหาระบบค้นหาคำหลักและค่าลักษณะเด่นของสัญญาณเสียงพูดที่สามารถทำงานได้อย่างมีประสิทธิภาพมากที่สุดเมื่อนำมาใช้กับเสียงพูดภาษาไทย

ภาพการทำงานโดยรวมของระบบค้นหาคำหลักบนสื่อเสียงนั้นจะมีรูปแบบการทำงานดังที่แสดงในรูปที่ 3-1 คือ จะต้องมีการป้อนข้อมูลสื่อเสียงและคำหลักที่จะใช้ในการค้นหาเข้าสู่ระบบค้นหา ระบบจะทำการประมวลผลและสุดท้ายระบบจะแสดงผลลัพธ์ว่าพบคำหลักอยู่บนข้อมูลสื่อเสียงดังกล่าวหรือไม่

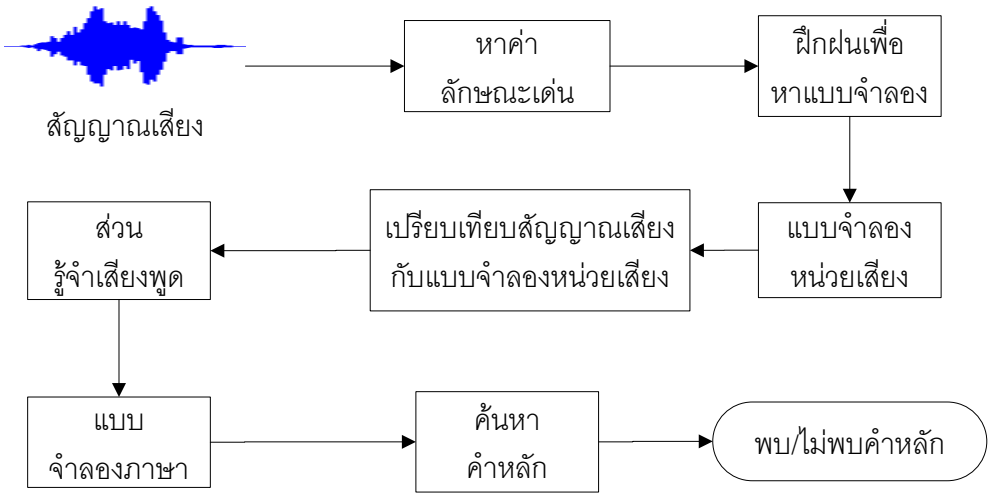


ภาพประกอบที่ 3-1 ภาพการทำงานโดยรวมของระบบค้นหาคำหลักบนสื่อเสียง

3.1. การทำงานของระบบค้นหาคำหลักบนสื่อเสียง

3.1.1. ระบบค้นหาคำหลักบนสื่อเสียงที่ทำงานบนระบบรู้จำเสียงพูด

ภาพการทำงานโดยรวมของระบบค้นหาคำหลักบนสื่อเสียงใช้ระบบรู้จำเสียงพูดเป็นดังภาพประกอบที่ 3-2 โดยมีขั้นตอนดังต่อไปนี้



ภาพประกอบที่ 3-2 ภาพการทำงานโดยรวมของระบบค้นหาคำหลักบนสื่อเสียงที่ใช้เทคนิคของระบบรู้จำเสียงพูด

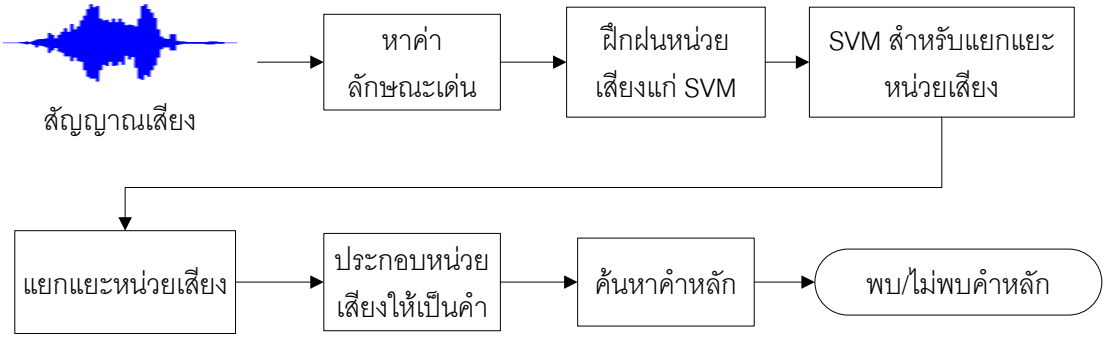
- ทำการวิเคราะห์สัญญาณเสียงพูด ในขั้นตอนนี้ระบบจะทำการเตรียมข้อมูลเพื่อใช้ในการฝึกฝนและทดสอบ โดยจะทำการคำนวณหาค่าลักษณะเด่นของสัญญาณเสียงพูดออกมา
- ในส่วนของการฝึกฝน ระบบจะใช้ค่าลักษณะเด่นของสัญญาณเสียงพูดมาวิเคราะห์เพื่อสร้างแบบจำลอง และแยกแยะรูปแบบของหน่วยเสียงแต่ละหน่วย
- นำแบบจำลองหน่วยเสียงไปประมวลผลอีกครั้งโดยใช้แบบจำลองภาษา เพื่อช่วยให้ผลลัพธ์ถูกต้องตามรูปแบบภาษาที่ได้กำหนดไว้
- หลังจากเสร็จกระบวนการทั้งหมดแล้ว จะได้ผลการรู้จำเสียงในรูปแบบของข้อมูลตัวอักษร จากนั้นนำชุดข้อมูลอักษรดังกล่าวเข้าสู่กระบวนการค้นหาคำหลัก เพื่อตรวจสอบว่าสัญญาณเสียงดังกล่าวมีคำหลักที่ต้องการหาปรากฏหรือไม่

3.1.2. ระบบค้นหาคำหลักบนสื่อเสียงที่ใช้วิธีการจำแนกประเภทโดยใช้เคอร์เนลและขอบเขตการแยกแยะที่มีขนาดกว้างที่สุด

ระบบค้นหาคำหลักบนสื่อเสียงที่ใช้วิธีการดังกล่าว ในการทดลองครั้งนี้ได้ใช้โปรแกรมของ Joseph K. เป็นหลัก ซึ่งโปรแกรมดังกล่าวได้ถูกออกแบบให้ร่วมกับฐานข้อมูลเสียงพูดภาษาอังกฤษ “TIMIT” ซึ่งเมื่อนำมาใช้กับฐานข้อมูลเสียงภาษาไทย “LOTUS” จึงจำเป็นที่จะต้องทำ

การเปลี่ยนรูปแบบข้อมูลในฐานข้อมูลเสียงบางส่วนให้คล้ายคลึงกับ TIMIT เพื่อให้สามารถใช้งานได้

โดยภาพการทำงานโดยรวมของระบบค้นหาคำหลักบนสื่อเสียงที่ใช้วิธีการจำแนกประเภทโดยใช้เคอร์เนลและขอบเขตการแยกแยะที่มีขนาดกว้างที่สุด เป็นดังภาพที่ 3-3 ซึ่งมีขั้นตอนดังต่อไปนี้

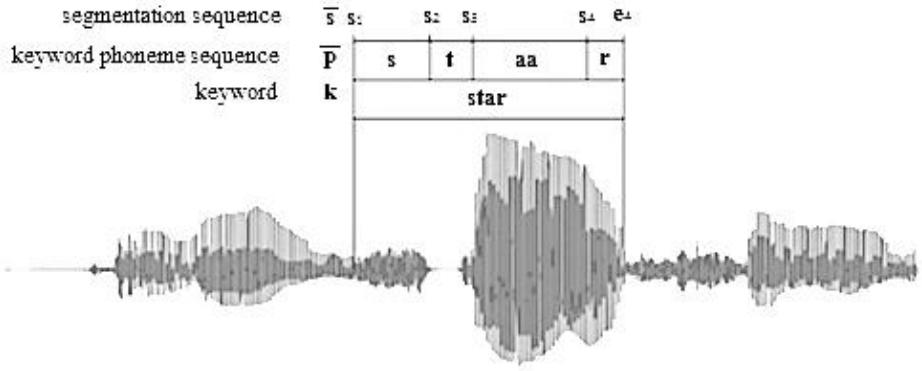


ภาพประกอบที่ 3-3 ภาพการทำงานโดยรวมของระบบค้นหาคำหลักบนสื่อเสียงที่มีพื้นฐานมาจากวิธีการจำแนกประเภทโดยใช้เคอร์เนลและขอบเขตการแยกแยะที่มีขนาดกว้างที่สุด

- ทำการวิเคราะห์สัญญาณเสียงพูด ซึ่งในขั้นตอนนี้ระบบจะทำการเตรียมข้อมูลเพื่อใช้ในการฝึกฝนและทดสอบ โดยจะทำการดึงค่าลักษณะเด่นของสัญญาณเสียงพูดออกมา
- ฝึกฝน SVM ร่วมกับค่าลักษณะเด่นของสัญญาณเสียงพูด เพื่อให้ได้แบบจำลองที่จะสามารถแยกแยะประเภทของหน่วยเสียงได้
- ในส่วนของการค้นหาคำหลัก ระบบจะทำการดึงค่าข้อมูลที่อยู่ในไฟล์เสียงออกมา จากนั้นระบบจะทำการตรวจสอบหน่วยเสียงที่ปรากฏตามลำดับเวลาในไฟล์เสียง โดยจะมีการถอดรหัสค่าหน่วยเสียงออกมาทีละตัว แล้วจึงนำหน่วยเสียงมาประกอบเป็นคำเพื่อเปรียบเทียบกับคำหลักที่ต้องการค้นหา ซึ่งถ้าหากมีคำที่รวมได้และคำหลักเหมือนกัน นั้นหมายความว่ามีคำหลักที่ต้องการปรากฏอยู่บนสื่อเสียง

ระบบค้นหาคำหลักบนสื่อเสียงของ Joseph K. นั้น ใช้ SVM เป็นเครื่องมือหลักในการแยกแยะหน่วยย่อยของเสียงพยางค์ โดยตัวแปรสำคัญในการกำหนดขอบเขตเพื่อแยกแยะได้แก่ คำหลักตาม

พจนานุกรมที่กำหนดไว้ ลำดับของหน่วยพื้นฐานของเสียงในคำหลัก และขอบเขตของช่วงเวลาในหน่วยพื้นฐานของเสียงแต่ละหน่วย ดังรูปที่ 3-4



ภาพประกอบที่ 3-4 แนวคิดของระบบค้นหาคำหลักของ Joseph K.

สัญญาณการพูดเป็นลำดับของลักษณะของเสียง $\bar{x} = (x_1, \ldots, x_T)$ โดยที่ $x_t \in X \subset R^d$ ในช่วง $1 \le t \le T$ และกำหนดให้คำหลัก $k \in K$ โดยที่ K เป็นพจนานุกรมคำในระบบ แต่ละคำหลัก k ประกอบด้วยลำดับของหน่วยของขั้นพื้นฐานของเสียง $\bar{p}^k = (p_1, \ldots, p_L)$ โดยที่ $p_l \in P$ ในช่วง $1 \le l \le L$ และ P เป็นโดเมนของสัญลักษณ์ของหน่วยขั้นพื้นฐานของเสียง ซึ่งกำหนดให้ P^* เป็นเซตของลำดับของ P และกำหนดการจัดวางระหว่างลำดับหน่วยขั้นพื้นฐานของเสียงและสัญญาณการพูด อีกทั้งกำหนดให้ $s_l \in N$ เป็นช่วงเวลาเริ่มต้นของหน่วยขั้นพื้นฐานของเสียง p_l และ ให้ $e_l \in N$ เป็นช่วงเวลาสิ้นสุดของหน่วยขั้นพื้นฐานของเสียง p_l อีกทั้งได้กำหนดช่วงเวลาเริ่มต้นของหน่วยพื้นฐานเสียงถัดไป p_{l+1} ให้มีค่าเหมือนกับช่วงเวลาสิ้นสุดของหน่วยขั้นพื้นฐานของเสียง p_l ซึ่งนั้นหมายถึง $e_l = s_{l+1}$ ในช่วงของ $1 \le l \le L\text{-}1$ ในลำดับของเวลา $\bar{s}^k$ ที่สอดคล้องกับลำดับของหน่วยพื้นฐานของเสียง $\bar{p}^k$ ซึ่งเป็นลำดับของเวลาเริ่มต้นและเวลาสิ้นสุด $\bar{s}^k = (s_1, \ldots, s_L, e_L)$ โดยที่ s_l เป็นเวลาเริ่มต้นของหน่วยพื้นฐานของเสียง p_l และ e_L เป็นเวลาสิ้นสุดของหน่วยพื้นฐานของเสียงลำดับสุดท้าย p_L ตัวอย่างของการกำหนดต่างๆ ได้ให้ไว้ในรูปที่ 3-4 เป้าหมายของ Joseph K. คือการศึกษาเกี่ยวกับระบบค้นหาคำหลักซึ่งกำหนดฟังก์ชัน f เป็นอินพุทคู่ลำดับ $(\bar{x}, \bar{p}^k)$ และค่าความเชื่อมั่นที่คำหลักที่เป้าหมาย k ที่ถูกเปล่งออกมาใน $\bar{x}$ นั้นก็คือ f ที่เป็นฟังก์ชันจาก $X^* \times P^*$ ไปสู่เซต R ค่าของความเชื่อมั่นที่เป็นค่าเอาพุทออกนั้น ได้มาจาก f

สำหรับคู่ลำดับ $(\bar{x}, \bar{p}^k)$ ซึ่งสามารถเปรียบเทียบได้กับ $b \in R$ เพื่อที่จะตัดสินว่า $\bar{p}^k$ ได้ถูกเปล่งเสียงออกมาใน $\bar{x}$ หรือไม่ ซึ่งแสดงให้เห็นในสมการที่ 3-1

$$f(\bar{x}, \bar{p}^k) = \max_{\bar{s}} W \cdot \phi(\bar{x}, \bar{p}^k, \bar{s}) \tag{3-1}$$

สำหรับฟังก์ชันลักษณะ (Feature Function) ที่ใช้ระบุหาคำหลักนั้น Joseph K. ได้เสนอฟังก์ชันลักษณะที่แตกต่างกันเพื่อใช้ในการฝึกฝนและเรียนรู้ของระบบ 7 รูปแบบ ซึ่งฟังก์ชันดังกล่าวนั้นจะถูกนำไปใช้ในการกำหนดฟังก์ชันชี้เฉพาะคำหลัก $f(\bar{x}, \bar{p})$ ดังที่แสดงในสมการที่ 3-1 ซึ่งมีดังต่อไปนี้

ฟังก์ชันลักษณะ 4 ฟังก์ชันแรกดังที่แสดงในสมการที่ 3-2 จะสนใจในค่าการเปลี่ยนแปลง (Transition) ระหว่างหน่วยพื้นฐานของเสียงซึ่งค่าที่ได้จากฟังก์ชันดังกล่าวจะแทนด้วยระยะทางระหว่างเฟรมของหน่วยสัญญาณเสียงก่อนและหลังตามลำดับเวลา $\bar{s}$ ซึ่งค่าระยะทางจะเป็นค่ายูซิเดียน (Euclidean) ระหว่างฟีเจอร์เวกเตอร์ โดยสมมติให้เฟรมที่ x_t และ $x_{t'}$ มีระห่างเป็น $d(x_t, x_{t'})$ โดยตรวจสอบว่า x_t และ $x_{t'}$ เป็นหน่วยพื้นฐานของเสียงหน่วยเดียวกันหรือไม่ ซึ่งถ้าเป็นหน่วยเสียงพื้นฐานเหมือนกันระยะห่างระหว่างเฟรมนั้นก็จะมีขนาดทีสั้นลง ในทางกลับกันถ้าเป็นหน่วยเสียงพื้นฐานที่ต่างกัน ระยะห่างระหว่างเฟรมก็จะมากตามไปด้วย

$$\phi_j(\bar{x}, \bar{p}, \bar{s}) = \frac{1}{|\bar{p}|} \sum_{l=2}^{|\bar{p}|-1} d(x_{-j+s_l}, x_{j+s_l}), \quad j \in \{1,2,3,4\} \tag{3-2}$$

ฟังก์ชันต่อมาจะเป็นฟังก์ชันที่ได้มาจากพื้นฐานของเคอร์เนลที่นำมาใช้ในการแยกแยะหน่วยพื้นฐานของเสียงพูด โดยในแต่ละหน่วยพื้นฐานของเสียง $p \in P$ และเฟรมสัญญาณ $x \in X$ จะเป็นค่าความเชื่อมั่น $g_p(x)$ ที่จะบอกว่ามีหน่วยพื้นฐานของเสียง p อยู่ภายในเฟรม x ซึ่งผลลัพธ์ของฟังก์ชันลักษณะที่ 5 นี้จะเป็นการวัดค่าความเชื่อมั่นสะสมของหน่วยพื้นฐานของสัญญาณเสียงจากลำดับของสัญญาณเสียงพูดและเวลาเริ่มต้น

$$\phi_5(\bar{x}, \bar{p}, \bar{s}) = \frac{1}{|\bar{p}|} \sum_{l=1}^{|\bar{p}|} \frac{1}{s_{l+1} - s_l} \sum_{t=s_l}^{s_{l+1}-1} g_{pl}(x_t) \tag{3-3}$$

ฟังก์ชันลักษณะลำดับที่ 6 จะเป็นฟังก์ชันที่หาค่าคะแนนของลำดับเวลาบนพื้นฐานของความคงทนของหน่วยพื้นฐานของเสียงพูด ซึ่งต่างจากฟังก์ชันลักษณะที่ 5 ซึ่งฟังก์ชันที่ 6 นั้นจะกระทำบนสัญญาณเสียงโดยตรง โดยจะทำการทดสอบเพื่อเปรียบเทียบความยาวของหน่วยพื้นฐานของเสียงแต่ละตัวกับการออกเสียงแต่ละครั้ง ซึ่งกำหนดไว้ในสมการที่ 3-4

$$\phi_6(\bar{x},\bar{p},\bar{s}) = \frac{1}{|\bar{p}|}\sum_{l=1}^{|\bar{p}|}\log N(s_{l+1}-s_l;\hat{\mu}_{pl},\hat{\sigma}_{pl}) \tag{3-4}$$

โดยที่ N คือฟังก์ชันความหนาแน่นของความน่าจะเป็นแบบปกติ (Normal Probability Density) บนค่าเฉลี่ย $\hat{\mu}_p$ และค่าความแปรปรวน $\hat{\sigma}_p$ ซึ่งค่าเฉลี่ยและค่าความแปรปรวนนั้นจะประมาณจากข้อมูลที่นำมาใช้ในการฝึกฝน

ในส่วนของฟังก์ชันสุดท้าย จะเป็นการคาดคะเนอัตราการพูดบนสัญญาณเสียงพูดของผู้พูด ซึ่งโดยปรกติแล้วสัญญาณเสียงพูดจะมีอัตราที่คงที่ ดังนั้นจึงสามารถค้นหาการเปลี่ยนแปลงได้โดยดูจากลำดับของเวลาที่มีการเปลี่ยนแปลงแบบฉับพลัน โดยกำหนดให้ $\hat{\mu}_p$ เป็นความยาวเฉลี่ยของการออกเสียงของหน่วยพื้นฐานของเสียง p^{th} ซึ่งได้มีการกำหนด r_l เป็นอัตราเสียงพูดโดยซึ่ง

$$r_l = \frac{(s_{l+1}-s_l)}{\hat{\mu}_{pl}} \tag{3-5}$$

นั้นคือ r_l จะเป็นอัตราส่วนระหว่างความยาวของหน่วยเสียงพื้นฐาน $\bar{p}$ ที่ได้มีการกำหนดช่วงเวลาตาม $\bar{s}$ บนความยาวเฉลี่ยของหน่วยเสียงพื้นฐาน ซึ่งโดยปกติแล้วอัตราของเสียงพูดจะมีการเปลี่ยนแปลงที่ช้ากว่าช่วงเวลานั้นๆ แต่ในทางปฏิบัติแล้ว อัตราของสัญญาณเสียงในแต่ละครั้งนั้นจะแตกต่างตามผู้พูดและลักษณะการพูดของผู้พูด จึงต้องทำการวัดการเปลี่ยนแปลงในอัตราของสัญญาณเสียง $(r_l - r_{l-1})^2$ ในสัญญาณเสียงนั้นๆ และกำหนดฟังก์ชันลักษณะเพื่อตรวจสอบการเปลี่ยนแปลงของอัตราสัญญาณเสียงดังสมการที่ 3-6

$$\phi_7(\bar{x},\bar{p},\bar{s}) = \frac{1}{|\bar{p}|}\sum_{l=2}^{|\bar{p}|}(r_l - r_{l-1})^2 \tag{3-6}$$

3.2. การสร้างค่าลักษณะเด่นของสัญญาณเสียงพูด

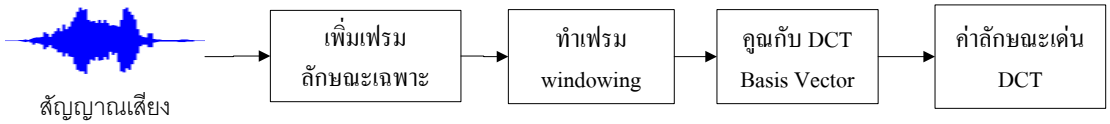
3.2.1. ค่าลักษณะเด่นเสียงพูดที่ได้จากเครื่องมือฮิดเดนมาร์คอฟ (HTK)

ในการหาค่าลักษณะเด่นของสัญญาณเสียงพูดสำหรับการทดลองครั้งนี้จะใช้เครื่องมือ HTK ซึ่งสามารถสร้างชุดค่าลักษณะเด่นของสัญญาณเสียงพูดได้หลากหลายชนิด ซึ่งนอกจากค่าลักษณะเด่นของสัญญาณเสียงพูดปกติแล้ว HTK ยังสามารถเพิ่มค่าลักษณะพิเศษต่างๆ (High Order Feature) ที่ได้จากสัญญาณเสียงพูด เพื่อทำให้ค่าลักษณะเด่นดังกล่าวมีความเป็นเอกลักษณ์มากยิ่งขึ้น ไม่ว่าจะเป็นค่าพลังงานภายในคลื่นสัญญาณเสียง ค่าสัมประสิทธิ์เดลต้า (อนุพันธ์อันดับที่หนึ่ง) ค่าสัมประสิทธิ์ความเร่ง (อนุพันธ์อันดับที่สอง) เป็นต้น ซึ่งการที่ค่าลักษณะเด่นของสัญญาณเสียงพูดมีลักษณะที่มีเอกลักษณ์มากขึ้น ย่อมส่งผลทำให้มีโอกาสที่จะได้ผลลัพธ์ที่ดีขึ้นแต่ในทางกลับกันขนาดของค่าลักษณะเด่นที่ได้จะเพิ่มขึ้นตามไปด้วยซึ่งส่งผลกระทบต่อพื้นที่ในการจัดเก็บและเวลาในการประมวลผลที่จะมากขึ้นตามไปด้วย

สำหรับการทดลองนี้จะใช้ค่าลักษณะเด่นของสัญญาณเสียงพูด 7 ประเภท ได้แก่ค่าลักษณะเด่น MFCC ค่าสัมประสิทธิ์ที่ได้จากการประมาณค่าเชิงเส้น (LPC) ค่าสัมประสิทธิ์ที่ได้จากการประมาณค่าเซปตรัลเชิงเส้น (LPCEPSTRA) ค่าค่าสัมประสิทธิ์ที่ได้จากการประมาณค่าเซปตรัลเชิงเส้นที่มีการเพิ่มค่าเดลต้า (LPDELCEP) ค่าสัมประสิทธิ์การสะท้อนของการประมาณค่าเชิงเส้น (LPREFC) ค่าลักษณะเด่นที่ได้จากผลลัพธ์ของชุดตัวกรองสเกลเมลเชิงเส้น (MELSPEC) และค่าลักษณะเด่นที่ได้จากค่าลอการิทึมของผลลัพธ์ที่ได้จากชุดตัวกรองสเกลเมล (FBANK) และค่าลักษณะพิเศษ 4 รูปแบบคือ ค่าสัมประสิทธิ์เดลต้า (D) ค่าสัมประสิทธิ์ความเร่ง (A) ค่าพลังงาน (E) และค่าสัมประสิทธิ์ค่าเฉลี่ยคงที่เท่ากับศูนย์ (Z) โดยจัดรูปแบบการเพิ่มค่าลักษณะพิเศษได้ 3 รูปแบบคือ ค่าลักษณะเด่นของสัญญาณเสียงพูดปกติ ค่าลักษณะเด่นที่เพิ่มค่าสัมประสิทธิ์เดลต้า ค่าสัมประสิทธิ์ความเร่งค่าพลังงานและค่าพลังงาน (D_A_E) และค่าลักษณะเด่นที่เพิ่มค่าสัมประสิทธิ์เดลต้า ค่าสัมประสิทธิ์ความเร่ง ค่าพลังงานและค่าสัมประสิทธิ์ค่าเฉลี่ยคงที่เท่ากับศูนย์ (D_E_A_Z)

3.2.2. ค่าลักษณะเด่นของเสียงบนพื้นฐานของค่าสัมประสิทธิ์การแปลงโคไซน์ไม่ต่อเนื่อง

ในส่วนของค่าลักษณะเด่นของสัญญาณเสียงพูดโดยใช้ค่าสัมประสิทธิ์การแปลงโคไซน์ไม่ต่อเนื่องที่นำมาใช้ในการทดลองครั้งนี้ได้แสดงภาพรวมของการคำนวณหาค่าลักษณะเด่นชนิดดังที่แสดงในรูปที่ 3-5



ภาพประกอบที่ 3-5 ขั้นตอนการสร้างค่าสัมประสิทธิ์โคไซน์ไม่ต่อเนื่อง

นำค่าลักษณะเด่นของสัญญาณเสียงพูดมาทำการคำนวณจำนวนเฟรมเวกเตอร์ของค่าลักษณะเด่นเสียก่อน เพื่อทำการปรับแต่งโดยกำหนดการเพิ่มจำนวนของเฟรมเวกเตอร์หัว-ท้ายเฟรมเวกเตอร์ของค่าลักษณะเด่น โดยจะเพิ่มเฟรมเวกเตอร์ในส่วนหัวมากกว่าเฟรมในส่วนท้ายในกรณีที่จำนวนเฟรมเวกเตอร์ของค่าลักษณะเด่นเป็นจำนวนคี่ และเพิ่มเฟรมเวกเตอร์ค่าลักษณะเด่นหัวและท้ายเท่ากันในกรณีที่จำนวนเฟรมของค่าลักษณะเด่นเป็นจำนวนคู่ดังที่แสดงในรูปที่ 3-6 เพื่อที่จะให้การทำเลื่อนกรอบสัญญาณเฟรม (Frame Windowing) นั้นครอบคลุมทุกๆ เฟรมเวกเตอร์ของค่าลักษณะเด่น

$$\begin{bmatrix} a_{11} & b_{12} & c_{13} \\ a_{21} & b_{22} & c_{23} \end{bmatrix} \rightarrow \begin{bmatrix} a_{11} & a_{11} & b_{12} & c_{13} & c_{13} \\ a_{21} & a_{21} & b_{22} & c_{23} & c_{23} \end{bmatrix}$$

ภาพประกอบที่ 3-6 การเพิ่มเฟรมของค่าลักษณะเด่น

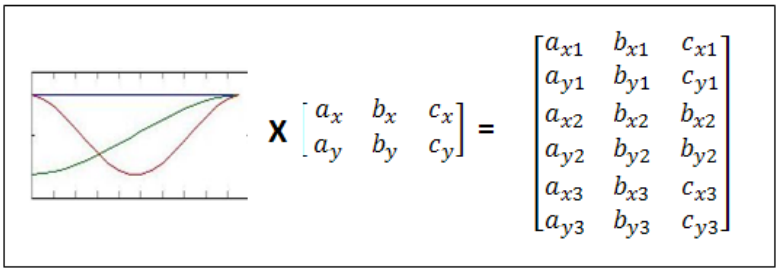
นำค่าลักษณะเด่นที่ได้ทำการปรับแต่งขนาดของเฟรมเวกเตอร์แล้ว มาทำการเลื่อนกรอบสัญญาณเฟรมตามขนาดที่กำหนดเพิ่มให้ได้ เพิ่มให้ได้เฟรมเวกเตอร์ของค่าลักษณะเด่นที่มีเฟรมเป็นเฟรมกึ่งกลางดังรูปที่ 3-7

$$\begin{bmatrix} a_{11} & a_{11} & b_{12} & c_{13} & c_{13} \\ a_{21} & a_{21} & b_{22} & c_{23} & c_{23} \end{bmatrix} \rightarrow \begin{bmatrix} a_x & b_x & c_x \\ a_y & b_y & c_y \end{bmatrix}$$

ภาพประกอบที่ 3-7 การหาค่ากลางจากการเลื่อนกรอบสัญญาณ

หลังจากได้เฟรมเวกเตอร์ของค่าลักษณะเด่นที่เป็นเฟรมกึ่งกลางแล้วโดยในการทดลองครั้งนี้เราจะใช้กรอบสัญญาณชนิดไคเซอร์ เพราะสามารถปรับค่าพารามิเตอร์ความโค้งของกรอบ

สัญญาณได้แล้วนำค่าลักษณะเด่นนั้นมาทำคูณกันกับเวกเตอร์มูลฐานของการแปลงโคไซน์ไม่ต่อเนื่อง (DCT Basis Vector) โดยใช้เวกเตอร์มูลฐานของการแปลงโคไซน์ไม่ต่อเนื่อง 3 ระดับแรก เพื่อเพิ่มค่าสัมประสิทธิ์ให้ค่าลักษณะเด่นของเสียงพูด ดังรูปที่ 3-8



ภาพประกอบที่ 3-8 การสร้างค่าลักษณะเด่นโดยใช้การแปลงโคไซน์ไม่ต่อเนื่อง

3.2.3. ค่าลักษณะเด่นเสียงพูดที่เกิดจากการรวมกันของค่าลักษณะเด่น

ในการทดลองนอกจากจะใช้ค่าลักษณะเด่นที่สร้างโดยมีเครื่องมือและทฤษฎีรองรับแล้ว ยังได้ทำการทดลองกับค่าลักษณะเด่นโดยนำข้อมูลค่าลักษณะเด่นของเสียงต่างชนิดกันมารวมเข้าด้วยกัน ด้วยวิธีง่ายๆ ซึ่งสิ่งที่ต้องคำนึงในการผสานค่าลักษณะเด่นต่างชนิดกันเข้าด้วยกันคือ ขนาดของค่าลักษณะเด่นของค่าลักษณะเด่นที่รวมเข้าด้วยกันจะต้องมีจำนวนค่าสัมประสิทธิ์เท่ากันทุกๆ กรณีสำหรับการทดลองในครั้งนี้ จะมีการรวมค่าลักษณะเด่น 3 รูปแบบใหญ่ๆ คือ

3.2.3.1. การผสมโดยปรับขนาดมิติค่าลักษณะเด่น (Re-dimension of Combine Speech Feature)

ในการผสานโดยการปรับขนาดของค่าลักษณะเด่นนั้น จะมีทั้งการเพิ่มและลดขนาดจากจำนวนเฟรมเวกเตอร์เดิม ซึ่งการทดลองนี้ได้ดำเนินบนกรรมวิธี 2 รูปแบบได้แก่

- **การวิเคราะห์องค์ประกอบหลัก**

การวิเคราะห์องค์ประกอบหลักนั้น สามารถเพิ่มหรือลดขนาดของเฟรมเวกเตอร์ได้ตามต้องการ แต่จะต้องมีจำนวนน้อยกว่าผลรวมของจำนวนเฟรมเวกเตอร์ของทั้ง 2 ค่าลักษณะเด่นรวมกัน โดยการทดลองได้ทำการกำหนดขนาดของค่าสัมประสิทธิ์หลังจากทำการผสานข้อมูลเข้าด้วยกัน 3 ค่า คือ 26, 39 และ 52 และค่าลักษณะเด่นที่นำมาทดลองนั้นเป็นค่าลักษณะเด่นที่เพิ่มค่าสัมประสิทธิ์เดลต้า ค่าสัมประสิทธิ์ความเร่งค่าพลังงานและค่าพลังงาน

- **การใช้ซูเปอร์เวกเตอร์ของการประมาณรูปร่างผสมของเกาส์**

ในการทดลองโดยการใช้ซูเปอร์เวกเตอร์ของการประมาณรูปร่างผสมของเกาส์นั้นจะเน้นการรวมการรวมของค่าลักษณะเด่น MFCC โดยค่าลักษณะเด่นที่ได้จากการใช้ซูเปอร์เวกเตอร์ของการประมาณรูปร่างผสมของเกาส์นั้น จะเหลือเฟรมเวกเตอร์เพียง 1 เฟรมเท่านั้น

3.2.3.2. **การรวมโดยใช้ตัวดำเนินการพื้นฐานทางคณิตศาสตร์ (Speech Feature Combination Based on Basic Mathematic Operation)**

ในการใช้ตัวดำเนินการพื้นฐานทางคณิตศาสตร์ที่จะใช้ในการรวมค่าลักษณะเด่นของสัญญาณเสียงพูดเข้าด้วยกันนั้น จะทดสอบ 6 รูปแบบได้แก่ การบวก การลบ การคูณผลเฉลี่ยของการบวกผลเฉลี่ยของการลบ และผลเฉลี่ยของการคูณ โดยข้อมูลที่นำมาดำเนินการจะเป็นข้อมูลที่อยู่ในตำแหน่งเดียวกันบนเมตริกซ์เวกเตอร์ของค่าลักษณะเด่นนำมากระทำกัน ดังตัวอย่างการบวกที่แสดงในรูปที่ 3-9 และได้ทำการทดสอบบนค่าลักษณะเด่นทั้ง 3 รูปแบบคือ ค่าลักษณะเด่นของสัญญาณเสียงพูดปกติ, แบบ D_A_E และแบบ D_E_A_Z

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \qquad B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$$

$$\text{ฟีเจอร์ใหม่} = \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} \\ a_{21} + b_{21} & a_{22} + b_{22} \end{bmatrix}$$

ภาพประกอบที่ 3-9 การบวกกันแต่ละอีลีเมนต์โดยตรงของค่าลักษณะเด่นทั้ง 2 ชนิด

3.2.3.3. **การรวมโดยการปรับตำแหน่งการวางข้อมูล (Re-locate Speech Feature Combination)**

ในการรวมข้อมูลจัดวางตำแหน่งข้อมูลของค่าลักษณะเด่น 2 ประเภทนั้น ในการทดลองครั้งนี้ได้มีการวางข้อมูล 2 รูปแบบคือการวางตำแหน่งต่อกันหน้า-หลังโดยตรง และการวางตำแหน่งของเฟรมเวกเตอร์แต่ละชนิดเรียงสลับกัน โดยในการทดสอบจะเป็นค่าลักษณะเด่นที่ทำการเพิ่มค่าสัมประสิทธิ์เดลต้า ค่าสัมประสิทธิ์ความเร่ง และค่าพลังงาน ดังที่แสดงในรูปที่ 3-10 และ 3-11

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \quad B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$$

$$\text{ฟีเจอร์ใหม่} = \begin{bmatrix} a_{11} & a_{12} & b_{11} & b_{12} \\ a_{21} & a_{22} & b_{21} & b_{22} \end{bmatrix}$$

ภาพประกอบที่ 3-10 การวางตำแหน่งหน้าหลังต่อกันโดยตรง

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \quad B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$$

$$\text{ฟีเจอร์ใหม่} = \begin{bmatrix} a_{11} & b_{11} & a_{12} & b_{12} \\ a_{21} & b_{21} & a_{22} & b_{22} \end{bmatrix}$$

ภาพประกอบที่ 3-11 การวางตำแหน่งข้อมูลแต่ละเวกเตอร์สลับกัน

3.3 สรุป

เนื้อหาภายในบทนี้กล่าวถึงกระบวนการทดสอบระบบค้นหาคำหลักบนสื่อเสียงภาษาไทย โดยสามารถแบ่งการทดสอบได้เป็น 2 กลุ่มใหญ่ๆ ได้แก่ (1) การทดสอบประสิทธิภาพของระบบค้นหาคำหลักบนสื่อเสียงภาษาไทยที่ใช้พื้นฐานของระบบแตกต่างกัน (2) การทดสอบประสิทธิภาพของระบบค้นหาคำหลักบนสื่อเสียงบนค่าลักษณะเด่นของเสียงที่แตกต่างกัน เพื่อนำไปประยุกต์ใช้กับงานในด้านต่างๆ ได้อย่างเหมาะสม

บทที่ 4
การทดลอง ผลการทดลอง และการวิเคราะห์ผลการทดลอง

4.1. การตั้งค่าตั้งต้นระบบ

4.1.1. **การกำหนดฐานข้อมูลเสียงภาษาไทย LOTUS [17]**

ในการทดลองครั้งนี้ ได้ทดสอบโดยใช้ฐานข้อมูลเสียงภาษาไทย LOTUS ซึ่งมีจำนวนไฟล์เสียง 1,680 ไฟล์เสียง จำแนกได้เป็น 801 ประโยค 2,269 คำ ในการทดสอบกำหนดให้ข้อมูลที่ใช้ในการฝึกฝน (Training Set) และข้อมูลที่ใช้ในการทดสอบระบบ (Testing Set) ทั้งหมด 1,680 ไฟล์เป็นข้อมูลชุดเดียวกันเนื่องจากคำศัพท์ที่ปรากฏในประโยคโดยส่วนใหญ่จะปรากฏน้อยกว่า 3 ประโยค ส่งผลให้ไม่สามารถแบ่งชุดข้อมูลเพื่อใช้ในการฝึกฝนและใช้ในการทดสอบออกจากกันได้ จำนวนคำที่นำมาใช้ในการทดสอบมีทั้งหมด 250 คำจาก 2,269 คำซึ่งได้มาจากการสุ่มเลือกโดยที่คำหลักนั้นจะต้องปรากฏอยู่ในประโยคอย่างน้อย 3 ประโยค

4.1.2. **การกำหนดระบบค้นหาคำหลักบนสื่อเสียง**

ระบบค้นหาคำหลักบนสื่อเสียงที่ใช้ทดลองจะแบ่งการทดสอบออกเป็น 2 ประเภทใหญ่ๆ

- การทดสอบบนระบบที่มีพื้นฐานจากระบบรู้จำเสียงพูด โดยใช้เครื่องมือ HTK เวอร์ชั่น 3.1 ในการสร้างระบบรู้จำเสียงพูดแบบ LVCSR ซึ่งได้กำหนดการประมวลผลการค้นหาแบบบีม (Beam Search) 10 ระดับ เพื่อให้เหมือนการกำหนดจุดทดสอบของระบบที่มีพื้นฐานมากจากวิธีการจำแนกประเภทโดยใช้เคอร์เนลและขอบเขตการแยกแยะที่มีขนาดกว้างที่สุด
- ระบบที่มีพื้นฐานมากจากวิธีการจำแนกประเภทโดยใช้เคอร์เนลและขอบเขตการแยกแยะที่มีขนาดกว้างที่สุดมีการปรับแต่งให้สามารถใช้กับภาษาไทยได้ โดยที่ค่าพารามิเตอร์ต่างๆ ยังคงกำหนดให้เหมือนกับระบบดังเดิม

4.1.3. **การกำหนดค่าลักษณะเด่นของสัญญาณเสียงพูดเพื่อใช้ในการทดสอบ**

ค่าลักษณะเด่นของสัญญาณเสียงพูดที่นำมาใช้ทดสอบนั้น จะมีวิธีหาชุดค่าลักษณะเด่นแบ่งออกเป็น 2 กลุ่มใหญ่ๆ คือ

4.1.3.3. **ชุดค่าลักษณะเด่นโดยใช้เครื่องมือ HTK**

สำหรับชุดค่าลักษณะเด่นของสัญญาณเสียงพูดที่ได้จากเครื่องมือ HTK นั้นมีทั้งค่าลักษณะเด่นของสัญญาณเสียงพูดปกติและค่าลักษณะเด่นของสัญญาณเสียงพูดที่มีการเพิ่มค่าลักษณะพิเศษอีกทั้งยังใช้เป็นค่าลักษณะเด่นพื้นฐานสำหรับการปรับปรุงค่าลักษณะเด่นเพิ่มเติมไม่ว่าจะเป็นการรวมโดยใช้วิเคราะห์องค์ประกอบหลักการรวมโดยใช้ซูเปอร์เวกเตอร์ของการประมาณรูปร่างผสมของเกาส์ การรวมเวกเตอร์ค่าลักษณะเด่นโดยใช้ตัวดำเนินการทางคณิตศาสตร์ หรือการปรับแต่งตำแหน่งเวกเตอร์ของค่าลักษณะเด่น

4.1.3.4. **ชุดค่าลักษณะเด่นของสัญญาณเสียงพูดซึ่งเป็นค่าสัมประสิทธิ์ที่คำนวณจากการแปลงโคไซน์ไม่ต่อเนื่อง**

สำหรับค่าลักษณะเด่นของสัญญาณเสียงพูดซึ่งเป็นค่าสัมประสิทธิ์ที่คำนวณจากการแปลงโคไซน์ไม่ต่อเนื่องที่จะนำมาใช้นั้น ในงานวิจัยชิ้นนี้จะประยุกต์ใช้กับค่าลักษณะเด่นของสัญญาณเสียงพูดที่ได้จาก HTK 4 ชนิดได้แก่ FBANK, MFCC, LPC และ LPDELCEP

4.1.4. **การวัดประสิทธิภาพการทำงานของระบบค้นหาคำหลัก**

ในส่วนของการวัดประสิทธิภาพการทำงานของระบบค้นหาคำหลักบนสื่อเสียง จะใช้ค่าพื้นที่ใต้เส้นโค้ง AUC เป็นตัวชี้วัด ซึ่งความถูกต้องของระบบจะแปรผันตรงกับพื้นที่ใต้กราฟโดยจะเป็นการเปรียบเทียบระหว่างจำนวนไฟล์เสียงที่มีคำหลักปรากฏอยู่ (Positive Sample) และจำนวนไฟล์เสียงที่ไม่มีคำหลักปรากฏอยู่ (Negative Sample) ในอัตราส่วน 3:3 บนคำหลักซึ่งนำมาใช้ในการทดสอบ 250 คำ

4.2. การทดลอง

4.2.1. **การทดลองระบบค้นหาคำหลักบนสื่อเสียงที่มีพื้นฐานมาจากระบบรู้จำเสียงพูด**

ในการทดลองนี้ เป็นการทดลองเพื่อหาประสิทธิภาพการทำงานของระบบค้นหาคำหลักบนสื่อเสียง ซึ่งใช้ระบบรู้จำเสียงพูดที่ทำงานโดยใช้เทคนิคของ HMM เป็นพื้นฐาน ร่วมกับค่าลักษณะเด่นของสัญญาณเสียงพูดรูปแบบต่างๆ เพื่อหาค่าลักษณะเด่นของสัญญาณเสียงพูดที่เหมาะสมที่สุดสำหรับระบบดังกล่าว โดยรายละเอียดของฐานข้อมูลเสียงที่นำมาใช้ในการทดลองค่าลักษณะเด่นของสัญญาณเสียงพูด และการประเมินประสิทธิภาพการทำงานของระบบได้กล่าวไปแล้วในหัวข้อที่ 4.1

4.2.1.1.**ผลการทดลองโดยใช้ค่าลักษณะเด่นจาก HTK**

จากผลการทดลองในตารางที่ 4-1 จะเห็นได้ว่าค่าลักษณะเด่นของสัญญาณเสียงพูดทุกชนิดสามารถให้ค่า AUC ได้มากกว่า 0.5 และให้ผลลัพธ์ที่ดีขึ้นเมื่อเพิ่มค่าลักษณะพิเศษให้แก่ลักษณะเด่นของสัญญาณเสียงพูด สิ่งที่น่าสนใจสำหรับผลการทดลองครั้งนี้คือ ค่าลักษณะเด่น LPDELCEP หรือค่าลักษณะเด่นที่ได้จากการประเมินค่าเชิงเส้นที่มีการเพิ่มเติมค่าสัมประสิทธิ์เดลต้านั้นให้ค่า AUC สูงกว่าค่าลักษณะเด่น MFCC ซึ่งเป็นค่าลักษณะเด่นที่เป็นที่นิยมในงานวิจัยสัญญาณเสียงพูดหลายๆ ชิ้น ไม่ว่าจะมีการเพิ่มค่าลักษณะพิเศษหรือไม่ก็ตาม

ตารางที่ 4-1 ผลลัพธ์ของค่าลักษณะเด่นแต่ละชนิดที่มีการเพิ่มค่าลักษณะพิเศษแบบต่างๆ บนระบบค้นหาคำหลักที่ใช้เทคนิคของระบบรู้จำเสียงพูด

	ประเภทของค่าลักษณะเด่นของสัญญาณเสียงพูด						
ค่าลักษณะเด่น	LPC	LPCEPSTRA	LPREFC	LPDELCEP	FBANK	MELSPEC	MFCC
ค่าลักษณะเด่นปกติ	0.511	0.605	0.575	0.692	0.619	0.578	0.653
ค่าลักษณะเด่น D_A_E	0.55	0.7	0.723	0.698	0.554	0.53	0.689
ค่าลักษณะเด่น D_E_A_Z	0.542	0.717	0.663	0.725	0.551	0.533	0.709

4.2.1.2.**ผลการทดลองที่เกิดจากการรวมค่าลักษณะเด่น**

4.2.1.2.1. **การรวมค่าลักษณะเด่นของสัญญาณเสียงพูดโดยปรับขนาดมิติของค่าลักษณะเด่นโดยการวิเคราะห์องค์ประกอบหลัก**

ในตารางที่ 4-2 จะเห็นได้ว่าผลลัพธ์ที่ได้จากการรวมค่าลักษณะเด่นเข้าด้วยกันสามารถให้ค่า AUC เกินกว่า 0.5 และคู่ลักษณะเด่นของสัญญาณเสียงพูดที่ให้ผลลัพธ์สูงที่สุดคือคู่ของ LPC และ FBANK ที่มีการเพิ่มค่าพิเศษแบบ D_A_E และมีการกำหนดจำนวนค่าสัมประสิทธิ์ของค่าลักษณะเด่นเท่ากับ 52 โดยให้ค่า AUC เท่ากับ 0.635 จะพบว่ามีค่ามากกว่าชุดค่าลักษณะเด่นของสัญญาณเสียงพูดที่ได้จาก HTK ทั้งแบบปกติและแบบที่มีการเพิ่มลักษณะพิเศษ และจะเห็นได้ว่าเมื่อปรับให้มีขนาดของค่าสัมประสิทธิ์ภายในค่าลักษณะเด่นมากขึ้น ค่า AUC ที่ได้จะมีแนวโน้มที่ดีขึ้นเช่นเดียวกัน

ตารางที่ 4-2 ตารางแสดงผลลัพธ์ของการรวมค่าลักษณะเด่นต่างๆ บนวิธีวิเคราะห์องค์ประกอบหลัก โดยกำหนดจำนวนค่าสัมประสิทธิ์ของค่าลักษณะเด่นที่แตกต่างกันบนระบบที่มีพื้นฐานบนระบบรู้จำเสียงพูด

	คู่ของค่าลักษณะเด่นที่นำมารวมกัน					
จำนวนค่าสัมประสิทธิ์ของค่าลักษณะเด่น	MFCC+ LPC	MFCC+ LPDELCEP	MFCC+ FBANK	LPDELCEP+ LPC	LPDELCEP+ FBANK	LPC+ FBANK
PCA=26	0.551	0.538	0.549	0.552	0.588	0.595
PCA=39	0.57	0.572	0.434	0.565	0.596	0.62
PCA=52	0.586	0.583	0.597	0.582	0.61	0.635

4.2.1.2.2. **การรวมค่าลักษณะเด่นของสัญญาณเสียงพูดโดยใช้ซูเปอร์เวกเตอร์ของการประมาณรูปร่างผสมของเกาส์**

ในการทดสอบจะเน้นการรวมค่าลักษณะเด่นของ MFCC เป็นหลักเพื่อดูแนวโน้มของผลลัพธ์ที่ได้จากวิธีการนี้ โดยจะทดสอบร่วมกับค่าลักษณะเด่นของสัญญาณเสียงพูดต่างชนิดกัน ซึ่งจากตารางที่ 4-3 จะเห็นได้ว่าได้ผลลัพธ์ที่ได้ไม่ดีเท่าที่ควร ค่า AUC มีค่าต่ำกว่า 0.5 ในทุกๆ การทดสอบนั้น เนื่องจากการสร้างซูเปอร์เวกเตอร์ของการประมาณรูปร่างผสมของเกาส์แล้ว จำนวนค่าลักษณะเด่นต่อเฟรมในสัญญาณเสียงจะมีค่าเท่ากับ 1 ซึ่งทำให้ไม่สามารถใช้ครอบคลุมค่าลักษณะเด่นของคำหรือหน่วยเสียงในสัญญาณทั้งหมดได้

ตารางที่ 4-3 ตารางแสดงผลลัพธ์การรวมค่าลักษณะต่างๆ ร่วมกับค่าลักษณะ MFCC บนระบบค้นหาคำหลักที่ใช้ระบบรู้จำเสียงพูดร่วมกับซูเปอร์เวกเตอร์ของการประมาณรูปร่างผสมของเกาส์

MFCC+LPC	MFCC+LPDELCEP	MFCC+FBANK
0.493	0.486	0.499

4.2.1.2.3. **การรวมค่าลักษณะเด่นของสัญญาณเสียงพูดโดยใช้ตัวดำเนินการพื้นฐานทางคณิตศาสตร์**

จากการใช้ตัวดำเนินการทางคณิตศาสตร์รวมค่าลักษณะเด่นของสัญญาณเสียงพูดนั้น ผลลัพธ์ส่วนใหญ่จะมีค่า AUC มากกว่า 0.5 แต่จากตารางที่ 4-4 จะพบว่าการใช้ตัวดำเนินการเครื่องหมายลบกับค่าลักษณะเด่นที่มีการเพิ่มค่าลักษณะพิเศษนั้น ไม่สามารถหาผลลัพธ์ได้ เนื่องจากความไม่สมเหตุสมผลในการปรับแต่งค่าการเปลี่ยนแปลงสถานะ (Emission) และความน่าจะเป็นในการเปลี่ยนสถานะ (Transition Probability) ของค่าลักษณะเด่นในแต่ละหน่วยเสียงเมื่อทำการประเมินค่าใหม่ในแต่ละครั้ง ซึ่งขั้นตอนดังกล่าวอยู่ในส่วนของการฝึกฝนระบบ ส่งผลให้ระบบไม่สามารถแยกแยะความแตกต่างของหน่วยเสียงพื้นฐานของแต่ละหน่วยได้ ดังนั้นผลจากการบวกและผลเฉลี่ยของการลบเวกเตอร์ของคู่ลักษณะเด่น MFCC และ LPC ที่ไม่มีการเพิ่มค่าลักษณะพิเศษสามารถให้ค่า AUC ที่มากกว่าการทดสอบโดยใช้ MFCC และ LPC แบบปกติ แสดงว่า การใช้ตัวดำเนินการคณิตศาสตร์กับคู่ของค่าลักษณะเด่นของสัญญาณเสียงพูดที่ไม่เพิ่มลักษณะพิเศษนั้นมีแนวโน้มที่จะทำให้ค่าลักษณะเด่นของสัญญาณเสียงพูดมีคุณลักษณะที่ดีขึ้น โดยเฉพาะผลเฉลี่ยของการลบซึ่งมีแนวโน้มดีที่สุด

ตารางที่ 4-4 ค่า AUC ที่ได้จากการรวมค่าลักษณะเด่นของสัญญาณเสียงพูดด้วยวิธีทางคณิตศาสตร์บนระบบค้นหาคำหลักบนสื่อเสียงที่ใช้ระบบรู้จำเสียงพูดเป็นระบบหลัก

	การดำเนินการทางคณิตศาสตร์					
ค่าลักษณะเด่น	การบวก	การลบ	การคูณ	ผลเฉลี่ยของการบวก	ผลเฉลี่ยของการลบ	ผลเฉลี่ยของการคูณ
MFCC+LPC	0.759	0.517	0.445	0.652	0.789	0.525
MFCC+LPC (D_A_E)	0.679	-	0.594	0.586	-	0.593
MFCC+LPDELCEP (D_A_E)	0.695	-	0.63	0.695	-	0.629
MFCC+LPC (D_E_A_Z)	0.699	-	0.58	0.7	-	0.58

4.2.1.2.4. **การรวมค่าลักษณะเด่นของสัญญาณเสียงพูดโดยการปรับตำแหน่งการวางเวกเตอร์ค่าลักษณะเด่นของสัญญาณเสียงพูด**

จากการทดสอบจะเห็นได้ว่าทั้งการนำเฟรมเวกเตอร์จากค่าลักษณะเด่นของสัญญาณเสียงพูดต่างชนิดกันมาเชื่อมต่อกันโดยตรงและการวางตำแหน่งของเฟรมเวกเตอร์ของค่าลักษณะเด่นของสัญญาณเสียงพูดต่างชนิดมาวางเรียงสลับกันนั้น สามารถให้ค่า AUC สูงกว่า 0.5 แต่ไม่มีผลรวมของค่าลักษณะเด่นคู่ใดที่สามารถให้ค่าผลลัพธ์ที่มีค่ามากกว่าการใช้ค่าลักษณะเด่นที่สร้างจาก HTK ได้ซึ่งวิเคราะห์ได้ว่าวิธีดังกล่าวไม่เหมาะสมสำหรับงานทางด้านการพัฒนาค่าลักษณะเด่นของสัญญาณเสียงพูด ดังที่แสดงในตารางที่ 4-5 และ 4-6

ตารางที่ 4-5 ค่า AUC ที่ได้จากการวางค่าสลับตำแหน่งของเวกเตอร์ระหว่างค่าลักษณะเด่นทั้งสองบนระบบค้นหาคำหลักบนสื่อเสียงที่มีพื้นฐานบนระบบรู้จำเสียงพูด

คู่ลักษณะเด่น	LPC+MFCC	MFCC+LPC	MFCC+ LPDELCEP	LPDELCEP+ MFCC
ค่า AUC	0.609	0.609	0.622	0.622

ตารางที่ 4-6 ค่า AUC ที่ได้จากการเชื่อมต่อกันระหว่างค่าลักษณะเด่นของสัญญาณเสียงพูดทั้งสองโดยตรงบนระบบค้นหาคำหลักที่มีพื้นฐานบนระบบรู้จำเสียงพูด

คู่ ลักษณะเด่น	LPC-FBANK-MFCC	MFCC-FBANK-LPC	FBANK-LPC-MFCC	FBANK-LPC	FBANK-MFCC	LPC-MFCC	LPC-FBANK	MFCC-FBANK	MFCC-LPC
ค่า AUC	0.585	0.542	0.542	0.535	0.538	0.6	0.535	0.599	0.538

4.2.2. การทดลองของระบบค้นหาคำหลักบนสื่อเสียงที่มีพื้นฐานมาวิธีการจำแนกประเภทโดยใช้เคอร์เนลและขอบเขตการแยกแยะที่มีขนาดกว้างที่สุด

ในการทดลองนี้ เป็นการทดลองเพื่อหาประสิทธิภาพการทำงานของระบบค้นหาคำหลักบนสื่อเสียง ซึ่งใช้วิธีการจำแนกประเภทโดยใช้เคอร์เนลและขอบเขตการแยกแยะที่มีขนาดกว้างที่สุดเป็นพื้นฐาน ร่วมกับค่าลักษณะเด่นของสัญญาณเสียงพูดชนิดต่างๆ เพื่อหาค่าลักษณะเด่นของสัญญาณเสียงพูดที่เหมาะสมที่สุดสำหรับระบบดังกล่าวและเปรียบเทียบผลลัพธ์ที่ได้กับระบบค้นหาคำหลักที่มีพื้นฐานจาก HMM โดยการกำหนดค่าในการทดลองไม่ว่าจะเป็นฐานข้อมูลเสียงที่นำมาใช้ในการทดลอง, ค่าลักษณะเด่นของสัญญาณเสียงพูด และการประเมินประสิทธิภาพการทำงานของระบบเหมือนกับการทดสอบบนระบบค้นหาคำหลักที่มีพื้นฐานจาก HMM

4.2.2.1. ผลการทดลองโดยใช้ค่าลักษณะเด่นของสัญญาณเสียงพูดที่ได้จาก HTK

สำหรับผลลัพธ์ของวิธีการจำแนกประเภทโดยใช้เคอร์เนลและขอบเขตการแยกแยะที่มีขนาดกว้างที่สุดโดยใช้ค่าลักษณะเด่นของสัญญาณเสียงพูดที่ได้จาก HTK จะเห็นได้ว่าค่า AUC ส่วนใหญ่จะมีค่ามากกว่า 0.5 เหมือนกับระบบค้นหาคำหลักที่ใช้หลักการของระบบรู้จำเสียงพูด ถึงแม้ว่าไม่มีค่าลักษณะเด่นของสัญญาณเสียงพูดชนิดใดที่สามารถให้ค่า AUC ได้มากกว่า 0.6 ดังที่ได้แสดงในตารางที่ 4-7 โดยชุดค่าลักษณะเด่นที่ให้ค่า AUC สูงสุดคือค่าลักษณะเด่น LPREFC หรือค่าสัมประสิทธิ์สะท้อนเชิงเส้นซึ่งอยู่ในกลุ่มค่าลักษณะเด่นที่คำนวณจากการประเมินค่าเชิงเส้น ซึ่งได้ค่า AUC เท่ากับ 0.595

ตารางที่ 4-7 ผลลัพธ์ของค่าลักษณะเด่นแต่ละชนิดที่มีการเพิ่มค่าลักษณะพิเศษแบบต่างๆ ระบบที่อยู่บนวิธีการจำแนกประเภทโดยใช้เคอร์เนลและขอบเขตการแยกแยะที่มีขนาดกว้างที่สุด

	ประเภทของค่าลักษณะเด่นของสัญญาณเสียงพูด						
ค่าลักษณะเด่น	LPC	LPCEPSTRA	LPREFC	LPDELCEP	FBANK	MELSPEC	MFCC
ค่าลักษณะเด่นปกติ	0.592	0.58	0.56	0.568	0.592	0.521	0.456
ค่าลักษณะเด่น D_A_E	0.545	0.548	0.56	0.558	0.542	0.544	0.561
ค่าลักษณะเด่น D_E_A_Z	0.508	0.56	0.595	0.54	0.544	0.526	0.524

4.2.2.2.**ผลการทดลองที่เกิดจากการรวมค่าลักษณะเด่น**

4.2.2.2.1. **การรวมค่าลักษณะเด่นของสัญญาณเสียงพูดโดยการปรับขนาดมิติเมตริกซ์ของค่าลักษณะเด่นโดยการวิเคราะห์องค์ประกอบหลัก**

จากการรวมกันของค่าลักษณะเด่นของสัญญาณเสียงพูดโดยใช้วิธี PCA สำหรับวิธีการจำแนกประเภทโดยใช้เคอร์เนลและขอบเขตการแยกแยะที่มีขนาดกว้างที่สุดนั้น จะเห็นได้ว่าค่า AUC ที่ได้จะมีค่ามากกว่า 0.5 แต่ว่าไม่มีค่าลักษณะเด่นคู่ใดที่ให้ค่า AUC ที่มากกว่าค่าลักษณะเด่นของสัญญาณเสียงพูดแบบปกติ อีกทั้งยังไม่สามารถสรุปแนวโน้มความสัมพันธ์ระหว่างคู่ของค่าลักษณะเด่นและจำนวนของค่าสัมประสิทธิ์ในค่าลักษณะเด่นเมื่อนำมารวมกันได้ ซึ่งแสดงให้เห็นในตารางที่ 4-8

ตารางที่ 4-8 ตารางแสดงผลลัพธ์ของการรวมค่าลักษณะเด่นต่างๆ บนวิธีวิเคราะห์องค์ประกอบหลัก โดยกำหนดจำนวนค่าสัมประสิทธิ์ของค่าลักษณะเด่นที่แตกต่างกัน บนระบบค้นหาคำหลักบนสื่อเสียงที่ใช้วิธีการจำแนกประเภทโดยใช้เคอร์เนลและขอบเขตการแยกแยะที่มีขนาดกว้างที่สุด

	คู่ของค่าลักษณะเด่นที่นำมารวมกัน					
จำนวนค่าสัมประสิทธิ์ของค่าลักษณะเด่น	MFCC+ LPC	MFCC+ LPDELCEP	MFCC+ FBANK	LPDELCEP+ LPC	LPDELCEP+ FBANK	LPC+ FBANK
PCA=26	0.593	0.513	0.498	0.57	0.599	0.556
PCA=39	0.565	0.556	0.587	0.576	0.53	0.511
PCA=52	0.474	0.582	0.584	0.492	0.549	0.585

4.2.2.2.2. **การรวมค่าลักษณะเด่นของสัญญาณเสียงพูดโดยใช้ซูเปอร์เวกเตอร์ของการประมาณรูปร่างผสมของเกาส์**

สำหรับการใช้ซูเปอร์เวกเตอร์ของการประมาณรูปร่างผสมของเกาส์ของระบบค้นหาคำหลักบนสื่อเสียงที่มีพื้นฐานบนวิธีการจำแนกประเภทโดยใช้เคอร์เนลและขอบเขตการแยกแยะที่มีขนาดกว้างที่สุดนั้น ให้ผลลัพธ์ที่ดีกว่าเมื่อเปรียบเทียบกับระบบที่อยู่บนพื้นฐานของระบบรู้จำเสียงพูด โดยให้ค่า AUC สูงกว่า 0.5 ทุกคู่ลักษณะเด่น แต่อย่างไรก็ตามยังคงให้ค่าที่น้อยกว่า ค่าลักษณะเด่นแบบปกติที่ได้จาก HTK ทำให้วิเคราะห์ได้ว่าการใช้ซูเปอร์เวกเตอร์ของการประมาณ

รูปร่างผสมของเกาส์นั้น เหมาะสำหรับระบบค้นหาคำหลักบนสื่อเสียงที่มีพื้นฐานบนวิธีการจำแนกประเภทโดยใช้เคอร์เนลและขอบเขตการแยกแยะที่มีขนาดกว้างที่สุดมากกว่าระบบค้นหาคำหลักที่ทำงานบนระบบรู้จำเสียงพูด ดังที่แสดงในตารางที่ 4-9

ตารางที่ 4-9 ผลลัพธ์การรวมค่าลักษณะเด่นโดยใช้ซูเปอร์เวกเตอร์ของการประมาณรูปร่างผสมของเกาส์ร่วมกับวิธีการค้นหาขอบเขตที่กว้างที่สุดในการจำแนกประเภทกับการประยุกต์ใช้เคอร์เนลประเภทต่างๆ

MFCC+LPC	MFCC+LPDELCEP	MFCC+FBANK
0.611	0.62	0.528

4.2.2.2.3. **การรวมค่าลักษณะเด่นของสัญญาณเสียงพูดโดยใช้ตัวดำเนินการพื้นฐานทางคณิตศาสตร์**

จากตารางที่ 4-10 จะเห็นได้ว่าได้ผลลัพธ์ส่วนใหญ่ให้ค่า AUC มากกว่า 0.5 แต่ในการรวมค่าลักษณะเด่นของสัญญาณเสียงพูดโดยใช้ตัวดำเนินการเครื่องหมายลบกับคู่ลักษณะของสัญญาณเสียงพูดที่มีการเพิ่มลักษณะพิเศษ D_E_A_Z นั้น จะไม่สามารถนำมาใช้ได้ เพราะระบบไม่สามารถทำการกำหนดขอบเขตระยะเวลาของหน่วยเสียง (Force Alignment) แต่ละหน่วยได้ ทำให้ระบบไม่สามารถคำนวณเวลาที่หน่วยเสียงนั้นๆ เริ่มต้นและสิ้นสุดได้ แต่สำหรับการการบวกกันของลักษณะเด่นของสัญญาณเสียงพูดโดยเฉพาะคู่ของ MFCC และ LPC ที่ไม่มีการเพิ่มค่าพิเศษนั้นสามารถให้ค่า AUC มากกว่า MFCC และ LPC แบบปกติ

ตารางที่ 4-10 ค่า AUC ที่ได้จากการใช้การรวมด้วยวิธีทางคณิตศาสตร์ บนระบบที่มีพื้นฐานของวิธีการจำแนกประเภทโดยใช้เคอร์เนลและขอบเขตการแยกแยะที่มีขนาดกว้างที่สุด

	การดำเนินการทางคณิตศาสตร์					
ค่าลักษณะเด่น	การบวก	การลบ	การคูณ	ผลเฉลี่ยของการบวก	ผลเฉลี่ยของการลบ	ผลเฉลี่ยของการคูณ
MFCC+LPC	0.554	0.609	0.587	0.597	0.64	0.572
MFCC+LPC (D_A_E)	0.501	0.578	0.557	0.572	0.599	0.554
MFCC+LPDELCEP(D_A_E)	0.598	0.513	0.589	0.542	0.57	0.584
MFCC+LPC (D_E_A_Z)	0.568	-	0.554	0.55	-	0.553

4.2.2.2.4. **การรวมค่าลักษณะเด่นโดยการปรับตำแหน่งการวางเวกเตอร์**

จากตารางที่ 4-11 และ 4-12 จะเห็นได้ว่าทั้งการวางตำแหน่งเฟรมเวกเตอร์ของค่าลักษณะเด่นต่อกันและการวางตำแหน่งเฟรมเวกเตอร์ของค่าลักษณะเด่นสลับกันของค่าลักษณะเด่นต่างชนิดกันนั้น ให้ผลลัพธ์ AUC มากกว่า 0.5 แต่ทว่าไม่มีค่าลักษณะเด่นคู่ใดที่ให้ค่า AUC มากกว่าค่าลักษณะเด่นแบบปกติอีกทั้งยังมีบางกรณีที่ไม่สามารถใช้ได้ ทำให้วิเคราะห์ได้ว่าการนำค่าลักษณะเด่นของสัญญาณเสียงพูดมาต่อกันโดยตรงนั้นไม่เหมาะสมสำหรับงานทางด้านการพัฒนาค่าลักษณะเด่นเหมือนกับระบบค้นหาคำหลักที่ใช้ระบบรู้จำเสียงพูดเป็นระบบพื้นฐาน

ตารางที่ 4-11 ค่า AUC ที่เกิดจากการวางค่าสลับตำแหน่งเวกเตอร์ระหว่างค่าลักษณะเด่นทั้งสอง บนระบบที่มีพื้นฐานบนวิธีการจำแนกประเภทโดยใช้เคอร์เนลและ

ขอบเขตการแยกแยะที่มีขนาดกว้างที่สุด

คู่ลักษณะเด่น	LPC+MFCC	MFCC+LPC	MFCC+LPDELCEP	MFCC+LPDELCEP
ค่า AUC	0.526	0.512	0.514	0.6

ตารางที่ 4-12 ค่า AUC ที่เกิดจากการต่อกันระหว่างค่าลักษณะเด่นทั้งสองโดยตรง บนระบบที่มีพื้นฐานบนวิธีการจำแนกประเภทโดยใช้เคอร์เนลและขอบเขตการแยกแยะที่มีขนาดกว้างที่สุด

คู่ลักษณะเด่น	LPC-FBANK-MFCC	MFCC-FBANK-LPC	FBANK-LPC-MFCC	FBANK-LPC	FBANK-MFCC	LPC-MFCC	LPC-FBANK	MFCC-FBANK	MFCC-LPC
ค่า AUC	0.5	0.567	0.468	0.598	0.502	0.528	-	0.46	0.392

4.2.3. ผลการทดลองโดยใช้ค่าลักษณะเด่นของสัญญาณเสียงพูดซึ่งเป็นค่าสัมประสิทธิ์ที่คำนวณจากการแปลงโคไซน์ไม่ต่อเนื่อง

สำหรับการทดสอบค่าลักษณะเด่นสัญญาณเสียงพูดบนพื้นฐานของค่าสัมประสิทธิ์การแปลงโคไซน์ไม่ต่อเนื่องจะเลือกทำการทดสอบบนระบบค้นหาคำหลักบนสื่อเสียงที่มีพื้นฐานมาจากระบบรู้จำเสียงพูดเท่านั้น เนื่องจากผลการทดลองที่ผ่านมาเห็นได้ว่าระบบค้นหาคำหลักบนสื่อเสียงที่มีพื้นฐานมาจากระบบรู้จำเสียงพูดให้ผลลัพธ์ที่ดีกว่าระบบค้นหาคำหลักบนสื่อเสียงที่มีพื้นฐานบนวิธีการจำแนกประเภทโดยใช้เคอร์เนลและขอบเขตการแยกแยะที่มีขนาดกว้างที่สุดอีกทั้งยังใช้เวลาในการประมวลผลน้อยกว่า โดยจะทดสอบบนชุดค่าลักษณะเด่นของสัญญาณเสียงพูด 4 ชนิด ได้แก่ MFCC, LPC, LPDELCEP และ FBANK ที่มีการปรับค่าพารามิเตอร์ของกรอบสัญญาณไคเซอร์และขนาดของกรอบสัญญาณ (Frame Window) เพื่อหาค่าพารามิเตอร์ที่เหมาะสมที่สุด ซึ่งจากตารางที่ 4-13 เห็นได้ว่า

- ในกลุ่มข้อมูลทดลองที่กำหนดค่าเบต้า (Beta) ให้มีค่าเท่ากัน ค่า AUC จะมีแนวโน้มลดลงเมื่อมีการปรับขนาดกรอบสัญญาณให้ใหญ่ขึ้น และการทดลองที่มีการกำหนดขนาดกรอบสัญญาณเหมือนกัน ค่า AUC จะมีแนวโน้มที่ดีขึ้น เมื่อมีค่าเบต้าให้มากขึ้น
- ค่าเบต้าที่มีค่าเท่ากับ 6 และขนาดกรอบสัญญาณที่มีค่าเท่ากับ 3 จะให้ค่าผลลัพธ์ AUC ดีที่สุด เมื่อเปรียบเทียบกับพารามิเตอร์ค่าอื่นๆ โดยที่ค่าลักษณะเด่นของสัญญาณเสียงพูด MFCC ที่ผ่านการแปลงโคไซน์ไม่ต่อเนื่อง จะเป็นค่าลักษณะเด่นของสัญญาณเสียงพูดที่ให้ค่า AUC สูงที่สุด คือ 0.922

ตารางที่ 4-13 ค่า AUC ที่ได้จากลักษณะเด่นของสัญญาณเสียงพูดบนพื้นฐานของค่าสัมประสิทธิ์การแปลงโคไซน์ไม่ต่อเนื่องโดยกำหนดค่าเบต้าเท่ากับ 6 กับกรอบสัญญาณขนาดต่างๆ บนระบบที่มีพื้นฐานบนระบบรู้จำเสียงพูด

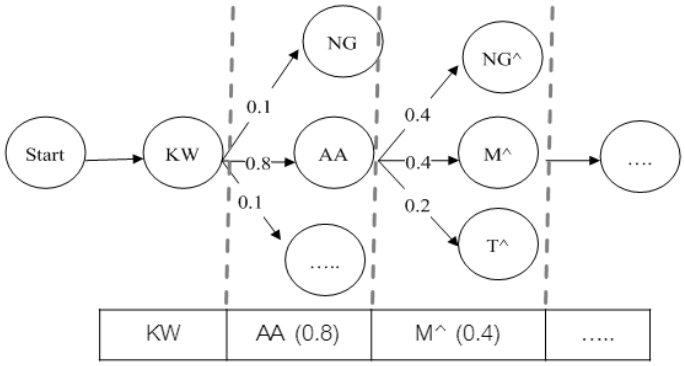
	FBANK	LPC	LPDELCEP	MFCC
w=0 B=6	0.516	0.529	0.532	0.558
w=3 B=6	0.62	0.639	0.762	0.922
w=5 B=6	0.591	0.592	0.743	0.789
w=10 B=6	0.54	0.546	0.587	0.641
w=15 B=6	0.582	0.573	0.724	0.776
w=20 B=6	0.555	0.583	0.583	0.782
w=25 B=6	0.525	0.533	0.538	0.584
w=30 B=6	0.541	0.558	0.565	0.651

*** ตารางเพิ่มเติมในภาคผนวก

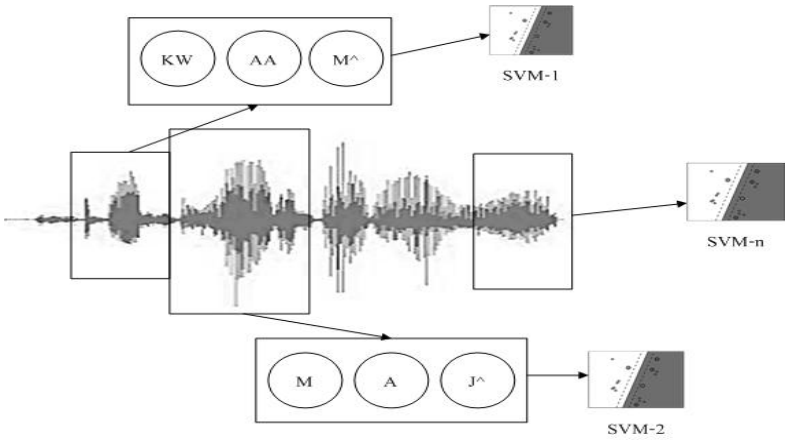
4.3. วิเคราะห์ผลการทดลอง

จากผลการทดลองทั้งหมดจะสรุปได้ว่า ระบบค้นหาคำหลักบนสื่อเสียงที่ใช้ระบบรู้จำเสียงพูดนั้น ให้ผลลัพธ์ดีกว่าระบบค้นหาคำหลักบนสื่อเสียงที่ใช้วิธีการจำแนกประเภทโดยใช้เคอร์เนลและขอบเขตการแยกแยะที่มีขนาดกว้างที่สุด เมื่อทดสอบโดยใช้ข้อมูลเสียงพูดภาษาไทยซึ่งเป็นผลมาจากการรวมคำจากหน่วยเสียงย่อยของแต่ละระบบซึ่งมีพื้นฐานแตกต่างกัน โดยระบบรู้จำเสียงพูดนั้นได้รับการออกแบบให้มีการเรียนรู้และรู้จำหน่วยเสียงแต่ละหน่วย ด้วยเทคนิคของ Viterbi (Viterbi Algorithm) ซึ่งเป็นส่วนหนึ่งของ HMM สำหรับการค้นหาและเชื่อมโยงหน่วยเสียงต่างๆ เพื่อสร้างหรือค้นหาคำหลักโดยใช้วิธีของViterbi นั้น จะเป็นการรวมคำจากหน่วยเสียงผ่านไปตามเส้นทางการเรียงกันของหน่วยเสียงย่อยแต่ละหน่วย โดยในแต่ละเส้นทางนั้นจะมีความน่าจะเป็นคอยควบคุมในการเลือกเส้นทางอยู่ ซึ่งในการเลือกเส้นทางในแต่ละครั้งจะมีโอกาสเลือกเส้นทางที่มีความน่าจะเป็นที่สูงกว่าเพื่อลดความผิดพลาด อีกทั้งยังมีการตรวจสอบเส้นทางและรวมหน่วยเสียงย่อยที่มีความน่าจะเป็นทั้งแบบไปหน้าและถอยกลับ ทำให้การประกอบคำจากหน่วยเสียงย่อยมีความถูกต้องมากยิ่งขึ้นดังที่แสดงในรูปที่ 4-1 ซึ่งแตกต่างจากวิธีการจำแนกประเภทโดยใช้เคอร์เนลและขอบเขตการแยกแยะที่มีขนาดกว้างที่สุด ซึ่งใช้ SVM เป็นพื้นฐานซึ่งเน้นการเรียนรู้และแยกแยะหน่วยเสียงย่อยและทำการสร้างคำจากหน่วยเสียงโดยที่ไม่มีความน่าจะเป็นเข้ามาช่วย

เหมือนระบบค้นหาคำหลักที่ใช้เทคนิคของระบบรู้จำเสียงพูด ดังที่แสดงในรูปที่ 4-2 อีกทั้งค่าความถูกต้องในการแยกแยะ (Classify) ของ SVM จะขึ้นอยู่กับค่าแลกเปลี่ยน (Trade Off) ซึ่งเป็นค่าถ่วงน้ำหนักของค่า 2 ค่าระหว่างระยะที่ใช้ในการแยกแยะและค่าความผิดพลาดซึ่งมีส่วนสำคัญที่จะทำระบบมีโอกาสที่จะแยกแยะหน่วยเสียงและสร้างคำศัพท์ผิดพลาด ซึ่งส่งผลกระทบให้ระบบไม่สามารถหาคำหลักที่ต้องการได้จากเหตุผลดังกล่าวสรุปได้ว่าระบบค้นหาคำหลักบนสื่อเสียงที่ใช้เทคนิคที่มีพื้นฐานมาจาก SVM นั้นให้ผลลัพธ์ที่ด้วยกว่าระบบค้นหาคำหลักที่ทำงานบนระบบรู้จำเสียงพูดซึ่งใช้เทคนิค HMM เป็นพื้นฐาน



ภาพประกอบที่ 4-1 การสร้างคำจากหน่วยเสียงของระบบค้นหาคำหลักที่ใช้เทคนิคของHMM



ภาพประกอบที่ 4-2 การสร้างคำจากหน่วยเสียงของระบบค้นหาคำหลักที่ใช้เทคนิคของSVM

สำหรับผลลัพธ์ที่ได้จากค่าลักษณะเด่นของสัญญาณเสียงพูดที่สร้างจาก HTK นั้นเห็นว่าค่าลักษณะเด่นที่ได้จากเครื่อง HTK เห็นได้ว่าค่าลักษณะเด่นของสัญญาณเสียงพูดที่เป็นค่าสัมประสิทธิ์ซึ่งได้จากการแปลงโคไซน์ไม่ต่อเนื่องของลอการิธึมของสเปกตรัมสัญญาณเสียงพูด

หรือว่าลักษณะเด่นที่มาจากค่าสัมประสิทธิ์ซึ่งมาจากการคำนวณเซปสตรัมได้แก่ LPDELCEP LPCC และ MFCC ให้ผลลัพธ์ที่ดีกว่าค่าลักษณะเด่นชนิดอื่นๆและการทดลองครั้งนี้ปรากฏว่าค่าลักษณะเด่น LPDELCEP และ LPCC ให้ค่า AUC ที่ดีกว่า MFCC ซึ่งให้ผลลัพธ์ที่ดีและใช้อย่างแพร่หลายในงานวิจัยทางด้านการวิเคราะห์เสียงพูด การที่ MFCC ให้ผลลัพธ์ที่ด้อยกว่า LPDELCEP และ LPCC นั้น แสดงว่าค่าลักษณะเด่นดังกล่าวใช้ในงานทางด้านการค้นหาคำหลักได้ดีกว่า

และจากการปรับปรุงค่าลักษณะเด่นของสัญญาณเสียงพูด โดยการรวมค่าลักษณะเด่นของสัญญาณเสียงพูดชนิดต่างๆ เข้าด้วยกัน ไม่ว่าจะเป็นการรวมโดยใช้ตัวดำเนินการพื้นฐานทางคณิตศาสตร์ การรวมค่าลักษณะเด่นโดยการเชื่อมต่อกันโดยตรง การรวมโดยการสลับตำแหน่งของค่าลักษณะเด่นของค่าลักษณะเด่นทั้งสอง และการรวมโดยการปรับจำนวนค่าสัมประสิทธิ์ของการรวมค่าลักษณะเด่นของสัญญาณเสียงพูดต่างชนิดกันจะเห็นได้ว่ามีผลลัพธ์ที่ดีกว่าและผลลัพธ์ที่แย่กว่าค่าลักษณะเด่นแบบปกติที่สร้างจาก HTK แต่การรวมค่าลักษณะเด่นโดยใช้ตัวดำเนินการทางคณิตศาสตร์มีแนวโน้มที่ดีที่สุด โดยเฉพาะอย่างยิ่งการบวกของค่าลักษณะเด่นต่างชนิดกันและผลเฉลี่ยที่ได้จากการลบกันของค่าลักษณะเด่นต่างชนิดกัน ซึ่งให้ค่าผลลัพธ์ที่ใกล้เคียงหรือดีกว่า เมื่อเปรียบเทียบกับค่าลักษณะเด่นของแต่ละตัวก่อนที่จะนำมารวมกัน โดยเฉพาะคู่ของค่าลักษณะเด่น LPC และ MFCC ที่ไม่มีการเพิ่มค่าลักษณะพิเศษเพิ่มเข้าไปให้ผลลัพธ์ที่สูงถึง 0.759 สำหรับการบวก และ 0.789 สำหรับค่าเฉลี่ยที่ได้การลบ บนระบบค้นหาคำบนเสียงพูดที่มีพื้นฐานมาจากระบบรู้จำเสียงพูด ในส่วนของการรวมกันโดยวิธีอื่นๆ นั้น โดยส่วนใหญ่แล้วจะได้ผลลัพธ์ที่แย่กว่าค่าลักษณะเด่นของสัญญาณเสียงพูดแบบปกติ ซึ่งอาจจะสรุปได้ว่าการรวมค่าลักษณะเด่นโดยใช้ตัวดำเนินการทางคณิตศาสตร์มาปรับปรุงค่าลักษณะเด่นของสัญญาณเสียงพูดนั้น มีความน่าจะเป็นที่จะทำให้ค่าลักษณะเด่นมีคุณสมบัติที่ดีขึ้น ซึ่งเป็นวิธีที่ง่ายและไม่ซับซ้อน อย่างไรก็ตามควรทดสอบโดยใช้ฐานข้อมูลเสียงเสียงของภาษาอื่นๆเพิ่มเติมเพื่อเป็นการยืนยันแนวคิดดังกล่าว

ในส่วนของค่าลักษณะเด่นของเสียงพูดซึ่งเป็นค่าสัมประสิทธิ์ที่คำนวณจากการแปลงโคไซน์ไม่ต่อเนื่อง ที่สร้างโดยการนำค่าลักษณะเด่นมาคูณกับเวกเตอร์มูลฐานของการแปลงโคไซน์ไม่ต่อเนื่องเพื่อสร้างค่าลักษณะเด่นออกมาใหม่นั้น ความถูกต้องที่ได้นั้นขึ้นปัจจัย 2 อย่างที่สำคัญคือขนาดของกรอบสัญญาณที่ใช้ในการเลื่อนเพื่อที่จะหาค่าลักษณะเด่นที่เป็นค่ากลางในแต่ละกรอบสัญญาณ และอีกหนึ่งปัจจัยคือค่าพารามิเตอร์เบต้าของกรอบสัญญาณไคเซอร์ซึ่งจากการทดลองเห็นได้ว่า โดยส่วนใหญ่ เมื่อกำหนดค่าเบต้าให้เท่ากัน ค่า AUC จะมีแนวโน้มลดลงเมื่อมีขนาดกรอบสัญญาณที่ใหญ่ขึ้น และโดยส่วนใหญ่ เมื่อกำหนดขนาดกรอบสัญญาณให้เท่ากัน ค่า AUC จะมีแนวโน้มเพิ่มขึ้น เมื่อมีค่าเบต้าที่มากขึ้น และจะเห็นได้ว่าการกำหนดให้ค่าเบต้าเท่ากับ 6 และขนาด

ของกรอบสัญญาณเท่ากับ 3 มีแนวโน้มให้ค่าผลลัพธ์ที่ดีที่สุด โดยเฉพาะอย่างยิ่งเมื่อใช้ร่วมกับค่าลักษณะเด่น MFCC ที่ผ่านการแปลงโคไซน์ไม่ต่อเนื่องซึ่งให้ค่า AUC สูงถึง 0.922

4.4. สรุป

ในบทนี้ได้กล่าวถึงการทดสอบระบบค้นหาคำหลักบนสื่อเสียงที่ใช้ระบบพื้นฐานและค่าลักษณะเด่นของเสียงที่แตกต่างกัน เพื่อเปรียบเทียบและจับคู่ระหว่างระบบพื้นฐานและค่าลักษณะเด่นของเสียงพูดที่สามารถให้ค่าความถูกต้องแม่นยำสูงสุด ซึ่งจะเห็นได้ว่าโดยภาพรวมแล้วระบบค้นหาคำหลักบนสื่อเสียงที่ใช้เทคนิคระบบรู้จำเสียงพูดนั้นจะให้ผลลัพธ์ที่ดีกว่าระบบค้นหาคำหลักบนสื่อเสียงที่ใช้วิธีการจำแนกประเภทโดยใช้เคอร์เนลและขอบเขตการแยกแยะที่มีขนาดกว้างที่สุดเป็นพื้นฐาน ซึ่งระบบค้นหาคำหลักบนสื่อเสียงที่ใช้ระบบรู้จำเสียงพูดเป็นพื้นฐาน ซึ่งใช้ค่าลักษณะเด่นเสียงพูด MFCC ที่ผ่านการแปลงเป็นค่าสัมประสิทธิ์การแปลงโคไซน์ไม่ต่อเนื่องให้ค่า AUC สูงสุดคือ 0.922 สำหรับการรวมค่าลักษณะเด่นของสัญญาณเสียงพูดนั้นช่วยให้ผลลัพธ์ดีขึ้นไม่มากนัก โดยวิธีการรวมค่าลักษณะเด่นของสัญญาณเสียงพูดโดยการรวมกันโดยการบวกกันของค่าลักษณะเด่นต่างชนิดกันและผลเฉลี่ยที่ได้จากการลบกันของค่าลักษณะเด่นต่างชนิดกัน มีแนวโน้มที่จะให้ผลลัพธ์ที่ดีขึ้นมากที่สุด

บทที่ 5
สรุปและข้อเสนอแนะ

5.1. สรุปผลการวิจัย

งานวิจัยนี้ได้ทำการทดลองเปรียบเทียบระบบค้นหาคำหลักบนสื่อเสียงที่มีพื้นฐานระบบแตกต่างกันคือระบบค้นหาคำหลักบนสื่อเสียงที่ใช้ระบบรู้จำเสียงพูด และระบบค้นหาคำหลักที่ใช้วิธีการจำแนกประเภทโดยใช้เคอร์เนลและขอบเขตการแยกแยะที่มีขนาดกว้างที่สุด นอกจากจะเปรียบเทียบระบบหลักทั้งสองแล้ว ในการทดลองครั้งนี้ยังได้มุ่งเน้นการเปรียบเทียบค่าลักษณะเด่นของสัญญาณเสียงพูดที่แตกต่างกัน ไม่ว่าจะเป็นค่าลักษณะเด่นของสัญญาณเสียงพูดทั่วไปที่สร้างจากเครื่องมือของฮิดเดนมาร์คอฟค่าลักษณะเด่นของเสียงบนพื้นฐานของค่าสัมประสิทธิ์การแปลงโคไซน์ไม่ต่อเนื่อง และค่าลักษณะเด่นที่เกิดจากการรวมกันของค่าลักษณะเดิม โดยใช้การวิเคราะห์องค์ประกอบหลัก การเชื่อมค่าลักษณะเด่นอย่างง่าย หรือแม้แต่วิธีการใช้ตัวดำเนินการทางคณิตศาสตร์อย่างง่าย เพื่อที่จะหาคู่ของระบบพื้นฐานและค่าลักษณะเด่นของสัญญาณเสียงพูดที่เหมาะสมที่สุดที่จะนำมาใช้งานกับภาษาไทย โดยสรุปได้ดังนี้

5.1.1. การเปรียบเทียบระหว่างระบบค้นหาคำหลักบนสื่อเสียงที่มีระบบพื้นฐานแตกต่างกัน

ในการเปรียบเทียบระหว่างระบบค้นหาคำหลักบนสื่อเสียงที่มีระบบพื้นฐานแตกต่างกัน ระบบค้นหาคำหลักบนสื่อเสียงที่มีระบบรู้จำเสียงพูด ซึ่งสร้างจากหลักการของ HMM เป็นระบบพื้นฐานนั้น มีแนวโน้มที่จะให้ผลลัพธ์ที่ดีกว่าระบบค้นหาคำหลักบนสื่อเสียงที่ใช้วิธีการจำแนกประเภทโดยใช้เคอร์เนลและขอบเขตการแยกแยะที่มีขนาดกว้างที่สุด ซึ่งใช้เทคนิคของ SVM เป็นเพราะวิธีการสร้างคำจากหน่วยเสียงที่แตกต่างกัน ซึ่งในระบบค้นหาคำหลักที่ใช้เทคนิคของ HMM นั้นจะมีความน่าจะเป็นเข้ามาช่วยในการรวบรวมหน่วยเสียงซึ่งต่างจาก SVM ซึ่งไม่มีตัวช่วยในการรวมหน่วยเสียงเพื่อสร้างคำขึ้นมา อีกทั้งภาษาไทยมีระดับชั้นเสียงหรือวรรณยุกต์ ทำให้คำหนึ่งคำสามารถมีเสียงหลายระดับต่างจากภาษาอังกฤษที่ไม่มีการใช้เสียงวรรณยุกต์ จึงทำให้ผลลัพธ์แตกต่างจากผลการทดลองที่ใช้ภาษาอังกฤษ

สำหรับการนำมาใช้งานจริงนั้นระบบค้นหาคำหลักที่ใช้ระบบรู้จำเสียงพูดเป็นระบบพื้นฐานนั้น จะมีความเหมาะสมมากกว่า เพราะนอกจากจะมีความถูกต้องมากกว่าแล้วยังใช้เวลาใน

การประมวลผลที่น้อยกว่าเมื่อเทียบกับระบบที่อยู่บนพื้นฐานของวิธีการจำแนกประเภทโดยใช้เคอร์เนลและขอบเขตการแยกแยะที่มีขนาดกว้างที่สุด ซึ่งวิธีดังกล่าวความเร็วในการประมวลผลนั้นจะขึ้นอยู่กับจำนวนของหน่วยเสียงที่จะต้องใช้ในการแยกแยะและประเภทของเคอร์เนลที่เลือกมาใช้งาน

5.1.2. **การเปรียบเทียบค่าลักษณะเด่นของสัญญาณเสียงพูดประเภทต่างๆ**

ในการเปรียบเทียบค่าลักษณะเด่นของสัญญาณเสียงพูดประเภทต่างๆ เห็นได้ว่าค่าลักษณะเด่นของสัญญาณเสียงพูดที่คำนวณได้จากการประมาณค่าสัมประสิทธิ์เชิงเส้นให้ผลลัพธ์ที่ดีกว่าค่าลักษณะเด่นของสัญญาณเสียงพูด MFCC ซึ่งได้แก่ LPDELCEP และ LPCEPSTRA ทำให้สรุปได้ว่าค่าลักษณะเด่นของสัญญาณเสียงพูดที่คำนวณได้จากการประมาณค่าสัมประสิทธิ์เชิงเส้น ให้ผลลัพธ์ที่กว่าเมื่อนำมาใช้ในงานทางด้านการค้นหาคำหลักบนสื่อเสียง

ในส่วนการรวมกันของค่าลักษณะเด่นด้วยวิธีต่างๆ ไม่ว่าจะเป็นการปรับแต่งขนาดมิติของเวกเตอร์ค่าลักษณะเด่น ซึ่งใช้ทั้งวิธีการวิเคราะห์องค์ประกอบหลัก และวิธีซูเปอร์เวกเตอร์ของการประมาณรูปร่างผสมของเกาส์ การรวมค่าลักษณะเด่นโดยใช้ตัวดำเนินการทางคณิตศาสตร์อย่างง่าย หรือแม้แต่การรวมโดยการนำค่าลักษณะเด่นมาเชื่อมต่อกันในรูปแบบต่างๆ จะเห็นได้ว่าในผลการทดสอบจะได้ผลลัพธ์ที่ดีกว่าและได้ผลลัพธ์ที่แย่กว่าค่าลักษณะเด่นก่อนที่จะนำมารวมกัน แต่สำหรับการใช้ตัวดำเนินการทางคณิตศาสตร์ในการรวมค่าลักษณะเด่นเข้าด้วยกันมีแนวโน้มที่ดีกว่าเมื่อเปรียบเทียบกับวิธี ซึ่งอาจจะสรุปได้ว่าการรวมค่าลักษณะเด่นโดยใช้ตัวดำเนินการทางคณิตศาสตร์มาปรับแต่งค่าโดยตรงนั้น มีความเป็นไปได้ที่จะทำให้ค่าลักษณะเด่นมีคุณสมบัติที่ดีขึ้นโดยวิธีที่ง่ายและมาซับซ้อน

สำหรับค่าลักษณะเด่นของเสียงบนพื้นฐานของค่าสัมประสิทธิ์การแปลงโคไซน์ไม่ต่อเนื่อง ค่าความถูกต้องที่ได้นั้นขึ้นปัจจัย 2 อย่างที่สำคัญคือขนาดของกรอบสัญญาณที่ใช้ในการเลื่อนเพื่อที่จะหาค่าลักษณะเด่นที่เป็นค่ากลางในแต่ละกรอบสัญญาณ และอีกปัจจัยหนึ่งคือค่าพารามิเตอร์เบต้าของกรอบสัญญาณไคเซอร์ที่ใช้ในการกำหนดความโค้งของกรอบสัญญาณซึ่งจะเห็นได้ว่าการกำหนดให้ค่าเบต้าเท่ากับ 6 และขนาดของกรอบสัญญาณเท่ากับ 3 มีแนวโน้มให้ค่าผลลัพธ์ที่ดีที่สุด โดยเฉพาะอย่างยิ่งเมื่อใช้ร่วมกับ MFCC บนระบบค้นหาคำหลักบนสื่อเสียงที่มีระบบรู้จำเสียงพูดเป็นระบบพื้นฐานซึ่งให้ค่า AUC สูงสุดเมื่อเปรียบเทียบกับค่าลักษณะเด่นทุกประเภทที่ได้กล่าวผ่านมาแล้ว

5.2. ข้อเสนอแนะ

ในงานวิจัยชิ้นนี้ได้มุ่งเน้นที่จะหาระบบค้นหาคำหลักและค่าลักษณะเด่นของสัญญาณเสียงพูดที่สามารถให้ผลลัพธ์ที่ดีที่สุดเมื่อนำมาใช้งานกับภาษาไทย ซึ่งนอกจากการเปรียบเทียบระบบและค่าลักษณะเด่นแล้ว ยังมีการพัฒนาค่าลักษณะเด่นของสัญญาณเสียงพูดโดยการนำค่าลักษณะเด่นต่างชนิดมารวมเข้าด้วยกัน ซึ่งผลการทดลองทั้งหมดนั้นมีพื้นฐานบนเสียงพูดภาษาไทย จึงควรที่จะทำการทดลองเพิ่มเติมบนฐานข้อมูลเสียงภาษาอื่นๆ ทั้งภาษาที่มีเสียงวรรณยุกต์เหมือนกับภาษาไทย เช่นภาษาจีน หรือภาษาเวียดนาม เป็นต้น

บรรณานุกรม

[1] J. Keshet, D. Grangier, S. Bengio, (2009), 'Discriminative Keyword Spotting', *Speech Communication* **51**(4), 317-329.

[2] K.R. Aida-Zade, C. Ardil, S.S. Rustamov,(2006), 'Investigation of Combined use of MFCC andLPC Features in Speech Recognition Systems', World Academy of Science, *Engineering and Technology19*,74-80.

[3] Minghui Liu, Zhongwei Huang, (2009), 'Multi-Feature Fusion Using Multi-GMM Supervector for SVM Speaker Verification', *Image and Signal Processing-CISP'09*. 2nd International Congress on, pp. 1-4, 17-19.

[4] ปภาณิน เดชเทวัญดำรง, (2552), การสืบค้นข้อมูลเสียงสำหรับข่าวภาษาไทย, มหาวิทยาลัยธรรมศาสตร์:กรุงเทพฯ.

[5] เฉลิมวุฒิ ไวชนะ, (2548), การชี้เฉพาะคำหลักเสียงพูดภาษาไทยบนพื้นฐานของการทดสอบหน่วยเริ่มและหน่วยตามของพยางค์, จุฬาลงกรณ์มหาวิทยาลัย: กรุงเทพฯ.

[6] Tangruamsub S., Punyabukkana, P., Suchato, A., (2007), 'Thai Speech Keyword Spotting using Heterogeneous Acoustic Modeling',*in* Research, Innovation and Vision for the Future, 2007 IEEE International Conference on, pp.253-260.

[7] ภูเบศโต๊ะลง, (2549), การค้นคืนข้อมูลจากแฟ้มข้อมูลเสียงภาษาไทยด้วยข้อคำถามเสียง, จุฬาลงกรณ์มหาวิทยาลัย: กรุงเทพฯ.

[8] ภคพงศ์ อมรกุล, (2544), การรู้จำเสียงที่เกิดจากการเน้นด้วยลักษณะสำคัญที่ต่างกัน, มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี: กรุงเทพฯ.

[9] Karnjanadecha M. and Kimsawad P, (2002), 'A Comparison of Front-End Analyses for Thai Speech Recognition', *Proceedings of the International Conference on Spoken Language Processing*, Denver, Colorado, USA.pp. 2141-2144.

[10] Young, S., Jansen, J., Odell, J., Ollasen, D.Woodland, P., (1995), *The HTK Book (Version 3.0)*, Cambridge Research Laboratory, Cambridge, England.

[11] Zahorian S.A., Silsbee P., Xihong Wang, (1997), 'Phone classification with segmental features and a binary-pair partitioned neural network classifier', *ICASSP-97.*, 1997 IEEE International Conference on, **2**, 1011 – 1014.

[12] วีระพล คงนุ่น, (2545), การแบ่งส่วนพื้นที่ของภาพที่ผ่านการวิเคราะห์องค์ประกอบด้วยทฤษฎีกราฟ, สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง: กรุงเทพฯ.

[13] Campbell, W.M.; Sturim, D.E.; Reynolds, D.A., (2006), 'Support vector machines using GMM supervectors for speaker verification', *Signal Processing Letters, IEEE* **13**(5), 308-311.

[14] L.R. Rabiner, (1989), 'A tutorial on hidden markov models and selected application in speech recognition', *Proc. IEEE 77*, pp. 267-293.

[15] ปทุมศิริสงศิริม, (2548), เทคนิคการแตกครึ่งตามสารสนเทศสำหรับซัพพอร์ตเวกเตอร์แมชชีนหลายประเภท, จุฬาลงกรณ์มหาวิทยาลัย: กรุงเทพฯ.

[16] พูนเพิ่ม สุวรรณรัฐภูมิ, (2552), แนวทางการปรับปรุงประสิทธิภาพของการจำแนกข้อมูลด้วยกฎความสัมพันธ์บนฐานข้อมูลที่ไม่สมดุล, บัณฑิตวิทยาลัย, มหาวิทยาลัยเกษตรศาสตร์, กรุงเทพฯ.

[17] Cotsomrong P., Sunpetchniyom T., Kasuriya S., Thatphithakkul N., Wutiwiwatchai C. , (2005), LOTUS: Large vOcabulary Thai contInUous Speech Recognition Corpus., *NAC2005*.

ภาคผนวก

ภาคผนวก ก.

เสียงภาษาไทย

ลักษณะของหน่วยเสียงภาษาไทย

คำในภาษาไทยประกอบด้วยหน่วยที่ย่อยลงมาที่เรียกว่าพยางค์ ซึ่งคำหนึ่งคำนั้นจะมีจำนวนพยางค์มากกว่าหนึ่งพยางค์ก็ได้และทุกพยางค์ไม่จำเป็นต้องมีความหมายในตัวเอง ในส่วนของพยางค์นั้นเกิดจากหน่วยเสียงแต่ละส่วนประกอบกันซึ่งจะประกอบด้วยเสียงพยัญชนะ (Consonant: C) สระ (Vowel: V) และวรรณยุกต์ (Tone: T) ในการออกเสียงภาษาไทยนั้นอาจจะออกเสียงหน่วยเสียงสระเพียงอย่างเดียวก็ได้แต่ไม่สามารถออกเสียงพยัญชนะอย่างเดียวได้การแบ่งเสียงพยัญชนะและสระในทางสัทศาสตร์นั้นจะอาศัยการบังคับลมในการเปล่งเสียงนั้นๆ เป็นเกณฑ์การบังคับลมในการออกเสียงสระช่องทางลมจะต้องเปิดกว้างไม่หรี่ทางลมให้แคบและมีการสั่นของสายเสียงตลอดเวลาส่วนเสียงพยัญชนะเป็นเสียงที่ต้องอาศัยการหรี่ทางลมหรือการปิดกั้นทางลมแบบสนิทด้วยเหตุนี้สระจึงทำหน้าที่เป็นแกนของพยางค์ส่วนพยัญชนะเป็นองค์ประกอบที่มาข้างหน้าหรือข้างหลังสระ

โครงสร้างพยางค์ที่เป็นไปได้ในภาษาไทยมี 4 รูปแบบได้แก่

- CV^T (พยัญชนะต้น + สระ + วรรณยุกต์)
- CCV^T (พยัญชนะควบกล้ำ + สระ+ วรรณยุกต์)
- CV^TC (พยัญชนะต้น + สระ + วรรณยุกต์ +พยัญชนะท้าย)
- CCV^TC (พยัญชนะควบกล้ำ + สระ + วรรณยุกต์ +พยัญชนะท้าย)

เสียงพยัญชนะ

พยัญชนะไทยมีทั้งหมด 44 รูป แต่เมื่อทำการแบ่งตามฐานกรณ์และประเภทของเสียงแล้วจะลดเหลือเพียง 21 เสียง ซึ่งเปรียบเทียบกับตัวอักษรได้ดังตารางภาคผนวกที่ 1

ตารางภาคผนวกที่ 1 เสียงพยัญชนะในภาษาไทยเมื่อเทียบกับพยัญชนะภาษาอังกฤษ

/k-/	ก	/ph-/	พ ภ ผ	/th-/	ท ธ ฒ ฑ ถ ฐ
/kh-/	ข ค ฆ	/f-/	ฝ ฟ	/n-/	น ณ หน-
/ng-/	งหง-	/m-/	ม หม-	/b-/	บ
/c-/	จ	/y-/	ย ญ หย- หญ-อย-	/p-/	ป
/ch-/	ช ฌ ฉ	/r-/	ร หร-	/s-/	ส ศ ษ ซ ทร-
/d-/	ด ฎ	/l-/	ล ฬ หล-	/h-/	ห ฮ
/t-/	ตฏ	/w-/	ว หว-	/z-/	อ

ซึ่งจากหน่วยเสียงทั้ง 21 หน่วยนั้น สามารถนำมาใช้เป็นเสียงพยัญชนะท้ายหรือเสียงตัวสะกดได้เพียง 9 เสียงเท่านั้น ซึ่งได้แสดงไว้ในตารางภาคผนวกที่ 2

ตารางภาคผนวกที่ 2 เสียงพยัญชนะตัวสะกดในภาษาไทยเมื่อเทียบกับพยัญชนะภาษาอังกฤษ

/-z/	ไม่มีเสียงสะกด อ	/-m/	ม
/-b/	บ ป พ ภ ฟ	/-n/	ณ น ล ร ญ ฬ
/-d/	ด ฎ ต ฏ ส ศ ษ ซ จ ช ท ธ ถ ฐ ฑ ฒ	/-ng/	ง
		/-w/	ว
/-k/	ก ข ค ฆ	/-y/	ย

ในภาษาไทย นอกจากเสียงพยัญชนะทั้ง 21 ตัวแล้ว อาจจะมีเสียงควบกล้ำกับเสียงอื่นๆ ได้ โดยมาตรฐานเสียงพยัญชนะที่เกิดขึ้นได้ในตำแหน่งที่สองมี 3 เสียงได้แก่ [-r-], [-l-] และ [-w-] ซึ่งเสียงควบกล้ำในภาษาไทยได้แสดงอยู่ในตารางภาคผนวกที่ 3

ตารางภาคผนวกที่ 3 เสียงพยัญชนะควบกล้ำในภาษาไทย

เมื่อเทียบกับพยัญชนะภาษาอังกฤษ

/pr-/	ปร	/phr-/	พร ผร
/pl-/	ปล	/phl-/	พล ผล
/tr-/	ตร	/thr-/	ทร
/kr-/	กร	/khr-/	ครขร
/kl-/	กล	/khl-/	คล ขล
/kw-/	กว	/khw-/	ควขว

เสียงสระ

สระในภาษาไทยสามารถแบ่งได้ออกเป็น 2 ประเภทใหญ่ได้แก่สระเดี่ยวและสระผสม ซึ่งสระเดี่ยวนั้นจะมีทั้งหมด 18 เสียงซึ่งสามารถแบ่งย่อยออกได้อีกเป็นสระเสียงสั่น 9 เสียงสระเสียงยาว 9 เสียง และสระผสมซึ่งเป็นการนำสระเสียงสั้นและเสียงยาวมาผสมกัน นั้นมีทั้งหมด 6 เสียง ดังที่แสดงไว้ในตารางภาคผนวกที่ 4

ตารางภาคผนวกที่ 4 เสียงสระในภาษาไทยเมื่อเทียบกับพยัญชนะภาษาอังกฤษ

สระเสียงสั้น		สระเสียงยาว		สระเสียงผสม	
/-a/	อะ	/-aa-/	อา	/-ie/	อะ
/-i-/	อิ	/-ii-/	อี	/-va-/	อิ
/-v-/	อึ	/-vv-/	อื	/-ua-/	อึ
/-u-/	อุ	/-uu-/	อู	/iie-/	อา
/-e-/	เอะ	/-ee-/	เอ	/-vva-/	อี
/-x-/	แอะ	/-xx-/	แอ	/-uua-/	อื
/-q-/	เออะ	/-qq-/	เออ		
/-o-/	โอะ	/-oo-/	โอ		
/-@-/	เอาะ	/-@@-/	ออ		

นอกจากเสียงสระทั้ง 24 เสียงที่ได้กล่าวมาข้างต้นแล้ว ภาษาไทยยังมีลักษณะอีกรูปแบบหนึ่งที่เรียกว่าสระอักษร ซึ่งได้แก่สระ อำ ไอ ใอ เอา ซึ่งแม้ว่าจะมีรูปสระเฉพาะ แต่เสียงสระนั้นออกเสียงเหมือนกันสระปกติบวกด้วยพยัญชนะท้ายนั้นเอง ซึ่งได้แก่ /-am/, /-ay/ และ /-aw/

เสียงวรรณยุกต์

เสียงวรรณยุกต์คือระดับเสียงสูงต่ำในคำๆ หนึ่ง ซึ่งเสียงวรรณยุกต์สามารถทำให้คำที่มีเสียงพยัญชนะและเสียงสระเหมือนกันมีความหมายแตกต่างกันได้โดยใช้เสียงวรรณยุกต์ที่แตกต่างกัน โดยเสียงวรรณยุกต์สามารถแบ่งได้เป็น 5 ระดับได้แก่

- เสียงสามัญเริ่มจากเสียงกลางและรักษาระดับให้คงที่ตลอดทั้งพยางค์
- เสียงเอกเริ่มจากระดับเสียงต่ำและรักษาระดับหรือลดระดับเสียงต่ำลงท้ายพยางค์
- เสียงโทเริ่มจากระดับเสียงสูงและลดระดับต่ำลงท้ายพยางค์
- เสียงตรีเริ่มจากระดับเสียงสูงและรักษาระดับหรือเพิ่มระดับเสียงสูงขึ้นท้ายพยางค์
- เสียงจัตวาเริ่มจากระดับเสียงต่ำและเพิ่มระดับเสียงสูงขึ้นท้ายพยางค์

ภาคผนวก ข.

รายละเอียดค่าลักษณะเด่นของสัญญาณเสียงพูดแต่ละประเภทที่สร้างจาก HTK

เนื่องจากการปรับปรุงค่าลักษณะเด่นของสัญญาณเสียงพูด โดยเฉพาะอย่างยิ่งเมื่อนำค่าลักษณะเด่นมารวมเข้าด้วยกัน ในบางครั้งจำเป็นที่จะต้องคำนึงถึงจำนวนค่าสัมประสิทธิ์ของลักษณะเด่นแต่ละชนิด ซึ่งในการทดลองครั้งนี้จำเป็นที่จะต้องทราบจำนวนสัมประสิทธิ์ของลักษณะเด่นแต่ละชนิดก่อน จึงจะนำมาทำการรวมกัน ซึ่งจำนวนสัมประสิทธิ์ที่ได้จาก HTK นั้นแสดงให้เห็นในตารางภาคผนวกที่ 5

ตารางภาคผนวกที่ 5 ตารางแสดงจำนวนค่าสัมประสิทธิ์ของค่าลักษณะเด่นของสัญญาณเสียงพูดแต่ละประเภท

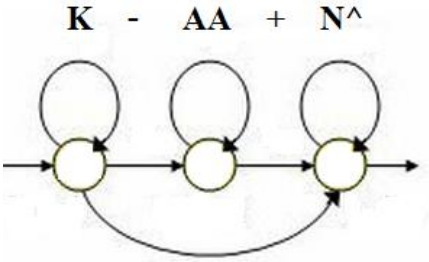
	LPC	LPREFC	LPCEPSTRA	LPDELCEP	MFCC	FBANK	MELSPEC
ค่าลักษณะปกติ	12	12	12	24	12	24	24
D_A_E	39	39	39	39	39	75	75
D_E_A_Z	39	39	39	39	39	75	75

ภาคผนวก ค.

รายละเอียดการทำงานระบบรู้จำเสียงพูดที่ใช้ HTK

ระบบรู้จำเสียงพูดที่นำมาใช้ในการทดลองครั้งนี้ได้ใช้เครื่องมือ HTK เป็นหลัก ซึ่งมีรายละเอียดการทำงานดังต่อไปนี้

- เริ่มจากการกำหนดแบบจำลองหน่วยเสียง เตรียมไฟล์ข้อมูลสำหรับกำกับว่าในแต่ละไฟล์เสียงที่ใช้ในการฝึกฝน ประกอบด้วยหน่วยเสียงใดบ้าง ในแต่ละไฟล์เสียงจะต้องมีหน่วยเสียง “sil” ซึ่งแทนด้วยเสียงเงียบ (Silence) ปิดหัวปิดท้ายเสมอ เพื่อเป็นการกำหนดจุดเริ่มต้นและจุดสิ้นสุดของเสียง
- ทำการสร้างหน่วยเสียงเดี่ยว (Monophone) โดยการทำงานเริ่มต้นจะทำการคำนวณค่าเฉลี่ยของHMM ซึ่งในการทดลองนี้ได้ใช้วิธี Flat-Start นั้นคือทุกหน่วยเสียงจะสร้างมาจากไฟล์ต้นแบบเดียวกัน แล้วนำมาคำนวณค่าโดยใช้คำสั่ง HCompVซึ่งผลที่ได้คือ HMM ของทุกๆ หน่วยเสียงที่ได้มาจากการคำนวณหาค่าเฉลี่ย (Mean) และค่าความแปรปรวน (Variance) จากข้อมูลที่นำมาใช้ในการฝึกฝนทั้งหมด แล้วกำหนดให้ค่าเฉลี่ยและค่าความแปรปรวน ให้มีค่าเท่ากันในทุกๆ แบบจำลอง
- การประมาณค่าพารามิเตอร์ใหม่ (Parameter Re-estimate) ของแต่ละหน่วยเสียงใน HMM ด้วยไฟล์ค่าลักษณะสำคัญทั้งหมดใหม่ ด้วยคำสั่ง HERestโดยจะทำการประมาณค่าใหม่ทั้งหมด 3 ครั้งเพื่อทำการปรับค่าพารามิเตอร์ต่างๆ ในแบบจำลอง หลังจากทำการประเมินแล้วก็จะได้แบบจำลองของหน่วยเสียงเดี่ยวแต่ละหน่วยเสียง
- เพื่อให้ได้ความถูกต้องแม่ยำเพิ่มมากขึ้น จึงใช้หน่วยเสียงเรียงสาม (Triphone) แทนหน่วยเสียงเดี่ยว โดยหน่วยเสียงเดี่ยวนั้นคือ 1HMM จะแทนด้วย 3 หน่วยเสียง ซึ่งจะสนใจหน่วยเสียงรอบข้าง (Context-Independent Phone Model) ซึ่งแสดงดังรูปข้างล่าง



หน่วยเสียงเรียงสามของคำว่า “กาน”

ภาคผนวก ง.

ขั้นตอนการใช้งานระบบค้นหาคำหลักบนสื่อเสียงของ Joseph K.

1. หาค่าลักษณะเด่นและกำหนดฉลาก (Label) ของไฟล์เสียงที่นำไปฝึกฝน โดยใช้คำสั่ง q1-extract_features_and_labels.sh
2. คำนวณหาค่าสถิติทั่วไปของข้อมูลไฟล์เสียงทั้งหมด ไม่ว่าจะเป็นการคำนวณหาค่าระยะห่างของค่าเซปตรัลในข้อมูลเสียงแต่ละไฟล์, คำนวณหาค่าความต่อเนื่องของหน่วยเสียงแต่ละหน่วย, คำนวณหาการวางแนวของสัญญาณเสียง, คำนวณค่าลักษณะเด่นและค่าสถิติของหน่วยเสียงโดยรวมโดยใช้คำสั่ง q2-train_frame_based_phoneme_classifier
3. ทำการฝึกฝนระบบเพื่อในในการแบ่งหน่วยออกจากกันเพื่อคำนวณหาคะแนนของความน่าจะเป็น, คำนวณหาค่าระยะของหน่วยเสียง และหาค่าตั้งต้นของสัญญาณในแต่ละหน่วยเสียง โดยใช้คำสั่ง q3-train_forced_alignment.sh
4. ทำการฝึกฝนในส่วนของการค้นหาคำหลัก โดยจะต้องมีการกำหนดประโยคในการฝึกฝน 2 ลักษณะคือประโยคที่มีคำหลักปรากฏและประโยคที่ไม่มีคำหลักปรากฏ โดยใช้คำสั่ง q4-train_keyword_spotter.sh
5. ทดสอบระบบโดยในการทดสอบระบบจะต้องทำการป้อนคำหลัก ประโยคที่มีคำหลักปรากฏและประโยคที่ไม่มีคำหลักปรากฏ เพื่อคำนวณหาพื้นที่ใต้ส่วนโค้ง ROC (AUC) โดยใช้คำสั่ง q5-decode_keyword_spotter.sh

ภาคผนวก จ.

การค้นหาคำหลักโดยใช้ HTK

เนื่องจาก HTK นั้นได้ถูกออกแบบมาสำหรับงานทางด้านการวิเคราะห์สัญญาณเสียงพูด โดยเฉพาะอย่างยิ่งระบบรู้จำเสียงพูด ซึ่งได้มีการออกเครื่องมือให้ใช้งานสำหรับการค้นหาคำหลัก เช่นเดียวกันคือ HResultซึ่งมีการกำหนดค่าพารามิเตอร์ของคำสั่งดังต่อไปนี้

HResult –w –I ไฟล์อ้างอิงในการกำหนดสลากเสียงพูด (Reference Label File) ไฟล์ที่ได้จากการรู้จำเสียง (Recognition Output File)

ตัวอย่าง

HResults -w -I /config/monoRef.mlfoutput.mlf

เราจำเป็นที่จะต้องกำหนด –w ทุกครั้งใน HResultเพื่อเป็นการกำหนด HResult สำหรับงานทางด้านการค้นหาคำหลัก ซึ่งตัวอย่างผลลัพธของการค้นหาคำหลัก kaan^ จะแสดงให้เห็นดังตารางภาคผนวกที่ 6

ตารางภาคผนวกที่ 6 ผลลัพธ์ของ HResult สำหรับการค้นหาคำหลัก

Figures of Merit

Keyword:	#Hits	#FAs	#Actual	FOM
Kaan^:	8	1	14	30.54
Overall:	8	1	14	30.54

จากตารางดังกล่าวแสดงให้เห็นถึงค่า 3 ค่าที่ได้จากการค้นหาคำหลักคือ จำนวนคำหลักที่ระบบค้นพบ (HIT) จำนวนคำหลักที่ไม่มีในตัวอย่างแต่พบในการทดสอบ (False Alarm: FA) และ FOM ซึ่งเราสามารถนำค่า HIT และค่า FA มาใช้ในการหาค่า ROC ได้

ภาคผนวก ฉ.

คำหลักที่นำใช้ในการทดลอง

สำหรับคำที่ใช้ในการทดลองนั้นมีทั้งหมด 250 คำซึ่งเลือกโดยการสุ่ม ซึ่งแต่ละคำจะปรากฏอยู่ในตัวอย่างประโยคเสียงอย่างน้อย 3 ประโยค

เกณฑ์	กำเริบ	ทอด	รัฐบาล
เกษตรกรรม	กำกับ	ทะเล	รัฐมนตรีว่าการ
เข้ามาใช้	กิตติ	ทำงาน	รับ
เครน	กิติพจน์	ที่	ราชบัณฑิตยสถาน
เครื่องรับวิทยุ	กีรติกร	ที่ตั้ง	ร้านอาหาร
เครื่องสปัตเตอริง	ขณะนี้	นโยบาย	ลอจิก
เคียว	ขยาย	นอกจากนี้	ลักษณะ
เงา	ขอ	นักวิชาการ	ลูกค้า
เจ็ดสิบห้า	ข้อกำหนด	นักศึกษา	วโรตม์สิกขดิตถ์
เจือ	ขอนแก่น	นางสาว	วงจรรวม
เฉพาะ	ข้อสอบ	น้ำหนัก	วงจรรีเลย์
เด็ก	ขั้วโลหะ	นิตตยา	วงจรอิเล็กทรอนิกส์
เดรน	ขาดแคลน	นิวเคลียร์	วรนุช
เดียวกัน	ข้าว	บรรจุ	วิเคราะห์
เตชะดำรงสิน	คณะกรรมการ	บริหาร	วิชา
เต็ม	คณะกรรมการบริหาร	บัด	วิทยาศาสตร์
เตาเอพิแทกซี	คณะรัฐมนตรี	บ้าง	วุฒิพงษ์
เตียวตระกูล	ครั้งละ	บางครั้ง	ศักดาไฟฟ้า
การกระทำ	คลื่น	ประกอบ	ศึกษา
เท่าไหร่	ความเข้ม	ประการ	ศุภเดช
เปลี่ยน	ความเป็นมา	ประทาน	กลไก

เผยแพร่
เพราะ
เมฆ
เย็น
เรื่อง
เลขที่
เลียนแบบ
เวิร์ดโปรเซสเซอร์
เส้นลวดตัวนำ
เห็น
เหมือนกัน
เหลือทิ้ง
เอกสารอ้างอิง
เอง
แชนแนล
แนวตั้ง
แบบ
แปด
แปล
แรงดันไฟฟ้า
แห่ง
แอดเดรส
แอสเซมบลี
โทรทัศน์
โบราณ
โปรแกรม
โฟลชาร์ต
ความจริง
ความรู้
คัมภีร์
คำเชื่อม
คำกำกวม
คำนวณ
งบประมาณ
งวดเงิน
งามเจตนรมณ์
จะ
จะทำ
จัด
จับ
จากนั้น
จาน
จุด
ชาติ
ชาย
ซากถ่านไฟฉาย
ดรัม
ดอกเตอร์
ดัง
ดังกล่าว
ดำเนินงาน
ดิจิตอล
ดิน
ดีซี
ประสาน
ประสานงาน
ปลุกเร้า
ปักดำ
ปากกา
ปีงบประมาณ
ผนัง
ผล
ผลลัพธ์
ผสม
ผู้กระทำ
ผู้ช่วยศาสตราจารย์
ฝนหลวง
ฝรั่งเศส
ฝ่าย
ฝึกสอน
พญาไทย
พร้อม
พระรามหก
พล็อตกราฟ
พลานาร์แมกนิตรอย
พอ
ฟิล์ม
ฟื้นฟู
ภารกิจ
ภาษา
ภาษาศาสตร์
สเปกตรัม
สง่า
รอย
สมเกียรติ
สมุนไพร
สองจุดหนึ่งจุดสาม
สองตุดหนึ่งจุดสี่
อุปสรรค
สัมประสิทธิ์
สายอากาศ
สารตัวนำยิ่งยวด
ตำแหน่ง
ทบทวน
ทบวง
สิ้นเปลือง
สิบเจ็ด
สี่
สี่จุดหนึ่งจุดหนึ่ง
หก
หน้าที่
หนึ่ง
หนึ่งจุดสาม
หนึ่งจุดสี่ศูนย์ศูนย์
หลงลืม
หัวบันทึก
ห้าสิบ
องค์ประกอบ

โมเสด	ดีมอส	มหาวิทยาลัย	อย่างไรก็ตาม
ใคร	ต้นฉบับ	มอเตอร์	อย่างยิ่ง
ใช้เงิน	ต่อไป	มอบหมาย	ออโธรอมบิด
ในกรณีที่	ตอนนี้	มอสเฟสกำลัง	ออกมา
ได้	ต่อๆไป	มารดา	อังกฤษ
ไทย	ตั้งแต่	มิลลิเมตร	อ่านเปรื่อง
ไฟฟ้า	ตันปรียชญา	มือ	อิเล็กทรอนิกส์
ไม่	ตัวตน	ย่อย	อินพุท
ไมครอน	ตัวตั้งเวลา	ยังมี	อุณหภูมิ
ไอเพล็กเซอร์	ตัวนับ	ยาแผนปัจจุบัน	อุดม
กระทรวง	ตี	รถ	อุตสาหกรรม
กรุงเทพ	ถนน	รวดเร็ว	ฮาร์ดแวร์
กฤษณพงศ์	ถ่ายภาพ	ระดับ	
เทคโนโลยีการพลังงาน		สำนักงานปลัดกระทรวง	
สองพันห้าร้อยสามสิบ		รองศาสตราจารย์ดอกเตอร์	

ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ

สถาบันเทคโนโลยีพระจอมเกล้าราชกระบัง

สำนักงานปลัดกระทรวงวิทยาศาสตร์เทคโนโลยีและการพลังงาน

ภาคผนวก ช.

ตารางค่าลักษณะเด่นที่ใช้เมตริกซ์เซปตรัลเชิงเวลา ที่กำหนดค่าเบต้า และขนาดกรอบสัญญาณ ในขนาดต่างๆ บนระบบที่มีพื้นฐานบนระบบรู้จำเสียงพูด

ค่า AUC ที่ได้จากลักษณะเด่นของเสียงเมื่อมีการเพิ่มค่าลักษณะพิเศษ
บนระบบที่มีพื้นฐานบนระบบรู้จำเสียงพูด

	FBANK	LPC	LPDELCEP	MFCC
no high order	0.619	0.511	0.692	0.653
DAE	0.554	0.550	0.698	0.689
DEAZ	0.551	0.542	0.765	0.709

ค่า AUC ที่ได้จากลักษณะเด่นของเสียงบนพื้นฐานของค่าสัมประสิทธิ์การแปลงโคไซน์ไม่ต่อเนื่อง
โดยกำหนดค่าเบต้าเท่ากับ 0 กับกรอบสัญญาณขนาดต่างๆ
บนระบบที่มีพื้นฐานบนระบบรู้จำเสียงพูด

	FBANK	LPC	LPDELCEP	MFCC
w=0 B=0	0.532	0.557	0.562	0.616
w=3 B=0	0.502	0.535	0.661	0.693
w=5 B=0	0.500	0.546	0.634	0.689
w=10 B=0	0.536	0.535	0.581	0.655
w=15 B=0	0.525	0.545	0.545	0.615
w=20 B=0	0.523	0.539	0.535	0.576
w=25 B=0	0.516	0.528	0.531	0.558
w=30 B=0	0.516	0.534	0.527	0.544

ค่า AUC ที่ได้จากลักษณะเด่นของเสียงบนพื้นฐานของค่าสัมประสิทธิ์การแปลงโคไซน์ไม่ต่อเนื่อง
โดยกำหนดค่าเบต้าเท่ากับ 1กับกรอบสัญญาณขนาดต่างๆ
บนระบบที่มีพื้นฐานบนระบบรู้จำเสียงพูด

	FBANK	LPC	LPDELCEP	MFCC
w=0 B=1	0.516	0.550	0.653	0.616
w=3 B=1	0.560	0.550	0.653	0.673
w=5 B=1	0.547	0.548	0.638	0.677
w=10 B=1	0.536	0.566	0.573	0.647
w=15 B=1	0.526	0.561	0.550	0.606
w=20 B=1	0.525	0.543	0.535	0.585
w=25 B=1	0.521	0.531	0.535	0.559
w=30 B=1	0.519	0.530	0.531	0.548

ค่า AUC ที่ได้จากลักษณะเด่นของเสียงบนพื้นฐานของค่าสัมประสิทธิ์การแปลงโคไซน์ไม่ต่อเนื่อง
โดยกำหนดค่าเบต้าเท่ากับ 3 กับกรอบสัญญาณขนาดต่างๆ
บนระบบที่มีพื้นฐานบนระบบรู้จำเสียงพูด

	FBANK	LPC	LPDELCEP	MFCC
w=0 B=3	0.516	0.528	0.531	0.616
w=3 B=3	0.544	0.540	0.636	0.652
w=5 B=3	0.539	0.546	0.613	0.647
w=10 B=3	0.539	0.555	0.585	0.646
w=15 B=3	0.535	0.554	0.548	0.624
w=20 B=3	0.539	0.549	0.539	0.596
w=25 B=3	0.522	0.532	0.533	0.581
w=30 B=3	0.518	0.531	0.528	0.565

ค่า AUC ที่ได้จากลักษณะเด่นของเสียงบนพื้นฐานของค่าสัมประสิทธิ์การแปลงโคไซน์ไม่ต่อเนื่อง
โดยกำหนดค่าเบต้าเท่ากับ 5 กับกรอบสัญญาณขนาดต่างๆ
บนระบบที่มีพื้นฐานบนระบบรู้จำเสียงพูด

	FBANK	LPC	LPDELCEP	MFCC
w=0 B=5	0.516	0.528	0.531	0.558
w=3 B=5	0.546	0.538	0.634	0.650
w=5 B=5	0.545	0.546	0.620	0.637
w=10 B=5	0.540	0.551	0.592	0.640
w=15 B=5	0.532	0.545	0.559	0.622
w=20 B=5	0.527	0.543	0.538	0.595
w=25 B=5	0.526	0.537	0.538	0.580
w=30 B=5	0.522	0.534	0.534	0.570

ค่า AUC ที่ได้จากลักษณะเด่นของเสียงบนพื้นฐานของค่าสัมประสิทธิ์การแปลงโคไซน์ไม่ต่อเนื่อง
โดยกำหนดค่าเบต้าเท่ากับ 8 กับกรอบสัญญาณขนาดต่างๆ
บนระบบที่มีพื้นฐานบนระบบรู้จำเสียงพูด

	FBANK	LPC	LPDELCEP	MFCC
w=0 B=8	0.532	0.557	0.562	0.616
w=3 B=8	0.544	0.537	0.641	0.650
w=5 B=8	0.545	0.541	0.625	0.643
w=10 B=8	0.537	0.544	0.595	0.639
w=15 B=8	0.535	0.551	0.562	0.629
w=20 B=8	0.529	0.547	0.545	0.608
w=25 B=8	0.525	0.534	0.537	0.592
w=30 B=8	0.521	0.529	0.529	0.580

ค่า AUC ที่ได้จากลักษณะเด่นของเสียงบนพื้นฐานของค่าสัมประสิทธิ์การแปลงโคไซน์ไม่ต่อเนื่อง โดยกำหนดค่าเบต้าเท่ากับ 10 กับกรอบสัญญาณขนาดต่างๆ บนระบบที่มีพื้นฐานบนระบบรู้จำเสียงพูด

	FBANK	LPC	LPDELCEP	MFCC
w=0 B=10	0.516	0.528	0.531	0.558
w=3 B=10	0.551	0.539	0.650	0.650
w=5 B=10	0.543	0.543	0.633	0.654
w=10 B=10	0.535	0.545	0.593	0.643
w=15 B=10	0.551	0.539	0.650	0.650
w=20 B=10	0.533	0.548	0.549	0.618
w=25 B=10	0.536	0.539	0.512	0.597
w=30 B=10	0.523	0.532	0.539	0.586

ภาคผนวก ซ.

ผลงานตีพิมพ์เผยแพร่จากวิทยานิพนธ์

1. Noah Kityawong, Montri Karnjanadecha and Anant Choksuriwong, 'A COMPARISON OF THAI KEYWORD SPOTTING SYSTEMS USING DIFFERENT TYPES OF FEATURES', In Proceeding of 4t^{h} International Conference on 2011 International Conference on Signal and Information Processing (ICCEE 2011), Singapore, October 14-16, 2011, pp. 597-602.

A Comparison of Thai Keyword Spotting Systems Using Different Types of Features

NOAH KITYAWONG | ANANT CHOKSURIWONG | MONTRI KARNJANADECHA

Department of Computer Engineering, Faculty of Engineering
Prince of Songkla University, Songkhla, Thailand
noahgundam@gmail.com, ant@coe.psu.ac.th, montri@coe.psu.ac.th

ABSTRACT

Two methods for Thai keyword spotting were studied. The first method was based on a HMM-based LVCSR system and the second was the discriminative keyword spotting technique which based on large margin and kernel method. Each method was evaluated on a Thai keyword spotting task using 19 different feature sets. LOTUS speech corpus was used to conduct all experiments. The objective of this work was to evaluate the performance of both keyword spotting methods on each type of feature set. Experimental results showed that the discriminative keyword spotting approach that utilized DCTC/DCS features outperformed the LVCSR system (AUC = 0.57 versus 0.51). One major drawback of the discriminative keyword spotting method was its computation inefficiency.

KEYWORDS

Thai keyword spotting, Hidden Makov Model, Speech recognition, Feature extraction

1 INTRODUCTION

Keyword spotting is one of the most important topics in speech processing research. Its goal is to determine, from speech utterances, whether or not a specified keyword was spoken. Many algorithms have been proposed to improve the accuracy of this task. Several keyword spotting systems were based on as automatic speech recognition (ASR) system, which is the most interested research area [1] [2] [3]. Among many ASR systems, Hidden Markov Model (HMM) based systems have been the most popular approach. The main function of an ASR system is to translate input speech into text. Thus keyword spotting can be performed by simply locating the keyword in the recognized text. In order to cover a large set of keywords, the ASR system must be able to recognize

a large set of vocabularies. As a result, a large vocabulary speech recognizer (LVCSR) has been used [4]. Note that ASR has a critical drawback when applied to keyword spotting tasks. For example the garbage model which is used to model the non-keyword words must be trained with a large set of vocabularies. This is expensive to train.

Keshet et al. proposed a discriminative keyword spotting technique which was based on large margin and kernel method [5]. The technique used the probability of the keyword to classify between the keyword and the other words which could solve the weak point of the HMM method for keyword spotting task [5]. The experimental results showed that the Area Under Curve (AUC) and Receiver operating characteristic (ROC) curve for the discriminative method were higher than that obtained with the HMM method.

Feature extraction techniques play very important roles for high performance speech recognition systems. This is also true for keyword spotting. There are several well-known speech features, which includes Linear Prediction Coefficient (LPC), Filter Bank (FBANK), Mel frequency cepstral coefficient (MFCC), and Discrete Cosine Transform Coefficients (DCTC) [6]. These features have been always used together with their delta and delta-delta terms. The inclusions of higher order terms have been proved to significantly increase recognition accuracy of ASR systems. In this paper, we evaluated both keyword spotting frameworks with various kind of feature sets to determine the best combination.

Most keyword spotting techniques found in the literatures were proposed to solve keyword spotting problem in major languages such as English and etc. However, the well-studied techniques may not work well for a tonal language such as Thai. The purpose of this research was to find the most suitable keyword spotting system for Thai language.

This paper is organized as follow. Section 2 describes LVCSR keyword spotter and the discriminative keyword spotter. Section 3 briefly explaines the feature extraction methods used in this work. Experimental setup and results are presented in Section 4. Section 5 concludes the paper.

2. Keyword Spotting Systems

Several techniques have been proposed to solve keyword spotting problem. For example the Neural Network (NN) was used in [7], the iterative dynamic programming method was used in [8] and the Gaussian mixture model (GMM) was used in [9]. In this research, we compared 2 keyword spotters: LVCSR and large margin and kernel method.

2.1. LVCSR keyword spotting system

When applied to keyword spotting, ASR is used to decode the most probable string of words from acoustic data. The main idea of this task is get all words from speech signal by speech recognition that most popular based on HMM. That mean correctly and accuracy of keyword spotting system is depend on speech recognizer. However, the word searches are limit by training's word in speech recognizer. To increase number of word in keyword search system, the LVCSR has been proposed. The LVCSR can be improve performance in keyword spotter but for keyword spotting task, we must deal with expensive cost to collect a large amount of labels data for training LVCSR and computation cost implied by large vocabulary decoding.

2.2. Discriminative Keyword spotting system

Discriminative Keyword spotting system was proposed to solve the keyword spotting problem by determining the likelihood of the keyword [10]. Discriminative learning is defined by

$$S = \left\{(\bar{p}_1, \bar{x}_1^+, \bar{x}_1^-, \bar{s}_1), \ldots, (\bar{p}_m, \bar{x}_m^+, \bar{x}_m^-, \bar{s}_m)\right\}$$

Where $\bar{p}$ is phoneme sequence in keyword, $\bar{x}^+$ and $\bar{x}^-$ are utterance sin which keyword is uttered and not. And $\bar{s}$ is alignment of keyword and the utterance with keyword. S will send to discriminative keyword spotting system that will return result in confidence of keyword in speech signal $f(\bar{x}, \bar{p})$. The function that use to calculate keyword confidence prediction score is define by

$$f(\bar{x}, \bar{p}^k) = \max_{\bar{s}} w \cdot \phi(\bar{x}, \bar{p}^k, \bar{s}) \qquad (1)$$

In this system, it has 7 functions for predefined feature function $\{\phi_j\}_{j=1}^n$ that same used in [10] and aim to increase AUC value in keyword spotting system. First four functions concentrate to capture transitions probability between phonemes. Each function means distance between frames of the acoustic signal at both sides of phoneme boundaries. If two frames, x_t and x_t' are derived from the same phoneme then the distance $d(x_t, x_t')$ should be smaller than different one. Next feature is measure the cumulative confidence that denoted by $g_p(x)$ and starts-times of phoneme sequence that utterance in speech signal. The sixth feature function is scoring timing sequences based on phoneme durations that examine the length of each phoneme $\bar{s}$ compared with the typical length required to pronounce in this phoneme. Finally, feature is finding suitable assumptions on speaker's speaking rate. Because less people speak in same rate, this function is calculating

average length in phoneme pronouncing. r_l is denoted for speech rate, but all speakers are different. Speaker rate will be change to $(r_l - r_{l-1})^2$ because speaking rate changes only slowly over time.

3. Speech Features

Feature extraction is the most important step for ASR, speaker identification and keyword spotting. Features are numbers that represent raw speech signal in a compact form and meaningful way. Good features must capture all important information in speech and discard irrelevant information. Several feature computation methods are described in this section.

3.1. Feature extraction methods

Many popular techniques have been used in speech extraction to find speech feature. The speech feature often use in speech research like keyword spotting. In this research use 7 kinds of feature extraction methods that separate in main 3 groups.

3.1.1 Linear Prediction Analysis: One of the most important speech analysis techniques because of its accuracy and computation speed. The basis is the source-filter model where the filter is constrained to be an all-pole linear filter. This amounts to performing a linear prediction of the next sample as a weighted sum of past samples. LPC is support vocal system over short-time interval follow transfer function as in Eq.2

$$H(z) = \frac{A}{1 - \sum_{k=1}^{p} a_k z^{-k}} \qquad (2)$$

Where a_k are prediction coefficients that minimize mean square prediction error.

3.1.2 Filter Bank Analysis: A classical spectral analysis technique which utilizes a bank of overlapping bandpass filters. The filters are spread along the frequency axis. Usually the spreading follows Mel frequency scale. The log energy of the signal in each frequency band is computed. This representation gives a rough approximation of the signal spectral shape while smoothing out the harmonic structure if any. Mel frequency warping is the most adaptation in FBANK that mimics spectral resolution on human ear range. The Mel warping approximations follow Eq. 3

$$mel(f) = 2595\log_{10}(1 + \frac{f}{700}) \qquad (3)$$

3.1.3 Cepstral Feature Analysis: The cepstral coefficients can be derived both from the filter-bank and linear predictive analyses. The cepstrum is defined as the inverse Fourier transform of the logarithm of the Fourier transform. Cepstral coefficients are computed by applying the discrete cosine transform (DCT) on the FBANK or LPC coefficients. MFCCs are the result of applying the DCT on the FBANK coefficients, Eq. 3. LPCCs are the result of applying the DCT on the LPC coefficients, Eq. 4

$$c_i = \sqrt{\frac{2}{N}} \sum_{j=1}^{N} e_j \cos\left(\frac{\pi i(j-0.5)}{N}\right) \qquad (4)$$

From Eq. 4, e_j is energy of the signal in the j^{th} filter channel and *N* is the number of channels in FBANK. $i \in [1, M], (M <= N)$.

$$c_i = -a_i + \frac{1}{i} \sum_{j=1}^{i-1} (i-j) a_j c_{i-j} \qquad (5)$$

From Eq. 5, $i \in [1, M], M <= P$ and P is the prediction order.

3.1.4 Discrete Cosine Transform Coefficient: This feature extraction method was based on applying the DCT directly on the short-time spectrum the speech signal [6]. The DCT basis vectors were modified to give higher resolution at low frequency and less at high frequency, which is similar to spreading the center frequency of the filter banks according to Mel scale in the filter bank analysis. To capture the temporal information into the DCTC features, another DCT was applied on a block consisting of several adjacent frames, resulting in Discrete Cosine Series (DCS) coefficients. This method for temporal encoding was found to be more efficient than the standard delta and delta-delta terms calculations. We refer to these features as DCTC/DCS features in our experiments.

3.2. High order features

It has been shown that higher order terms can improve speech recognition accuracy. Simplest forms of higher order terms are delta (D) and delta-delta or acceleration (A) terms. Other term such as energy (E) has been shown to be beneficial. Also scaling each feature to achieve zero mean (Z) can be useful. The D, A, E, and Z notations are used by Hidden Markov Model Toolkits (HTK), [11]. This work used HTK's HCopy tool was used to compute these parameters.

4. Experimental Setup and Results

The LOTUS corpus [12], a Thai speech database, was used for all experiments. The PD set of this corpus consists of 801 sentences (2,269 words) spoken by 75 speakers. The training set comprised 1680 samples (2,269 words) and the test set comprised 840 samples (250 words).

Performance of the discriminative keyword spotting method and the LVCSR method was compared by using 8 feature extraction methods. The performance was measured using the AUC from averaged ROC.

Tab.1 and Tab.2 show the AUC's of the keyword spotting experiments using the LVCSR method and the discriminative keyword spotting method, respectively.

Tab.1 AUC obtained from the LVCSR system.

	+D+E+A+Z	High order features +D+A+E	None
MFCC	0.46	0.48	0.51
LPC	0.45	0.46	0.42
LPCEPSTRA	0.44	0.46	0.49
LPDELCEP	0.44	0.46	0.50
MELSPEC	0.32	0.18	0.47
FBANK	0.22	0.44	0.45
DCTC	N/A	N/A	0.44
LPREFC	0.48	0.41	0.49

Tab.2 AUC obtained from the discriminative keyword spotting system.

	+D+E+A+Z	High order features +D+A+E	None
MFCC	0.48	0.49	0.42
LPC	0.55	0.53	0.57
LPCEPSTRA	0.53	0.51	0.54
LPDELCEP	0.50	0.05	0.51
MELSPEC	0.50	0.56	0.55
FBANK	N/A	N/A	N/A
DCTC	N/A	N/A	0.57
LPREFC	0.52	0.55	0.53

5. Conclusion

This work compared the performance of 2 Thai keyword spotting systems using various feature sets. The LVCSR-based keyword spotting system worked best with MFCC features without high order terms. Highest AUC was obtained from the keyword spotting system that was based on the large margin and kernel method using the DCTC features. However, the computation inefficiency of the later system was its major drawback.

6. References

[1] [1] Szoke I., Schwarz P., Matejka P., Burget L., Fapso M., Karafiat M., Cernocky J., Comparison of keyword spotting approaches for informal continuous speech. In: Proc. of Joint Workshop on Multimodal Interaction and Related Machine Learning Algorithms, 2005.

[2] [2] Wai Yat Wong, John Robertson, “Application of Phonological Knowledge in Audio Word Spotting”, Proceeding Proceedings of the Eleventh International Florida Artificial Intelligence Research Society Conference AAAI, 1998.

[3] [3] Rose, R., Paul, D., A Hidden Markov Model based keyword recognition system. In: Proc. of International Conference on Audio, Speech and Signal Processing, 1990, PP129-132

[4] [4] Weintraub. M, “Keyword-spotting using SRI's DECIPHER large-vocabulary speech-recognition system” Acoustics, Speech, and Signal Processing, 1993. ICASSP-93., IEEE International Conference on 27-30 April 1993, Volume 2, PP 463.

[5] [5] J. Keshet, D. Grangier, S. Bengio, “Discriminative keyword spotting”, Speech Communication Volume 51, Issue 4, April 2009, PP 317-329.

[6] [6] Zahorian S.A., Silsbee P., Xihong Wang, “Phone classification with segmental features and a binary-pair partitioned neural network classifier”Acoustics, Speech, and Signal Processing, 1997. ICASSP-97., 1997 IEEE International Conference on 21-24 Apr 1997, Vol.2, PP 1011 – 1014.

[7] [7] Hao Ruan, Ravi Sankar, “Applying neural network to robust keyword spotting in speech recognition application” Neural Networks, 1995. Proceedings., IEEE International Conference on Nov/Dec 1995 Volume 5, PP 2882.

[8] [8] Silaghi M., Bourlard H., “A new keyword spotting approach based on iterative dynamic programming”, Acoustics, Speech, and Signal Processing, 2000. ICASSP '00. Proceedings. 2000 IEEE International Conference Volume 3 , PP1831.

[9] [9] Abida K. ; Karray F. ; Jiping Sun, “Comparison of GMM and fuzzy-GMM applied to phoneme classification Signals”, Circuits and Systems (SCS), 2009 3rd International Conference on 6-8 Nov. 2009, PP 1.

[10] [10] Martin Wollmer et. al., “Robust discriminative keyword spotting for emotionally colored spontaneous speech using bidirectional LSTM networks”, Proceeding ICASSP '09 Proceedings of the 2009 IEEE International Conference on Acoustics, Speech and Signal Processing, 2009.

[11] [11] Young, S., Jansen, J., Odell, J., Ollasen, D., Woodland, P., The HTK Book (Version 3.0), Cambridge Research Laboratory, Cambridge, England. 1995.

[12] [12] Cotsomrong P., Sunpetchniyom T., Kasuriya S., Thatphithakkul N., Wutiwiwatchai C. , LOTUS: Large vOcabulary Thai contInUous Speech Recognition Corpus., NAC2005 (in Thai).

ประวัติผู้เขียน

ชื่อ สกุล นายโนอาห์ กิจญะวงค์

รหัสประจำตัวนักศึกษา 5210120080

วุฒิการศึกษา

วุฒิ	ชื่อสถาบัน	ปีที่สำเร็จการศึกษา
วิศวกรรมศาสตรบัณฑิต (วิศวกรรมคอมพิวเตอร์)	มหาวิทยาลัยสงขลานครินทร์	2551

ทุนการศึกษา

ทุนศิษย์ก้นกุฏิ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยสงขลานครินทร์

การตีพิมพ์เผยแพร่ผลงาน

1. Noah Kityawong, Montri Karnjanadecha and Anant Choksuriwong, 'A COMPARISON OF THAI KEYWORD SPOTTING SYSTEMS USING DIFFERENT TYPES OF FEATURES', In Proceeding of 4t^{h} International Conference on 2011 International Conference on Signal and Information Processing (ICCEE 2011), Singapore, October 14-16, 2011, pp. 597-602.