การประชุมวิชาการทางวิศวกรรมไฟฟ้า ครั้งที่ 48
The 48th Electrical Engineering Conference (EECON-48)
วันที่ 19-21 พฤศจิกายน 2568 ณ โรงแรมฟูราม่า จังหวัดเชียงใหม่

อัลกอริทึมการรู้จำเสียงสระภาษาไทยโดยใช้ตัวกรองแบบปรับตัวได้

Thai Vowel Recognition Using an Adaptive Filter

วรวุฒิ แจ่มหล้า[1]* ขวัญชัย ทองภูมิพันธ์[1] สัมพันธ์ พรหมพิชัย[1] และ สกล อุดมศิริ[1]

[1]สาขาวิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีปทุมวัน Worawut.jamla@gmail.com*

บทคัดย่อ

บทความนี้นำเสนอการพัฒนาระบบรู้จำเสียงสระในภาษาไทย โดยใช้ Mel-Frequency Cepstral Coefficients (MFCC) ร่วมกับตัวกรองแบบปรับตัวได้ (Adaptive Filter) ด้วย Maximum Likelihood Estimation วิธีการจำแนกประเภทแบบ One-vs-All ระบบได้ทำการทดลองกับชุดข้อมูลเสียงสระเดี่ยวในภาษาไทย 18 เสียง (เสียงสั้น 9 เสียง และเสียงยาว 9 เสียง) คุณลักษณะเด่นของสัญญาณเสียงถูกสกัดโดยใช้สัมประสิทธิ์เซปตรัมบนสเกลเมล (MFCC) ร่วมกับคุณลักษณะเดลต้าความเร็วของเสียงและเดลต้า-เดลต้าความเร่งของเสียง ทำให้ได้เวกเตอร์คุณลักษณะในแต่ละเฟรม จากนั้นจึงสร้างแบบจำลองทางสถิติแบบเกาส์เซียน (Gaussian Model) สำหรับสระแต่ละเสียงเพื่อใช้เป็นตัวกรองแบบปรับตัวได้ ผลการทดสอบประสิทธิภาพของระบบด้วยข้อมูลทดสอบพบว่ามีความแม่นยำโดยรวมสูงถึง 95.11% แสดงให้เห็นถึงศักยภาพของวิธีการที่นำเสนอในการจำแนกเสียงสระภาษาไทยได้อย่างมีประสิทธิภาพ

คำสำคัญ: การรู้จำเสียงพูด, เสียงสระภาษาไทย, สัมประสิทธิ์เซปตรัมบนสเกลเมล, ตัวกรองแบบปรับตัวได้

Abstract

This paper presents the development of a Thai vowel recognition system using Mel-Frequency Cepstral Coefficients (MFCC), an Adaptive Filter, and the One-vs-All classification method. The system was tested on a dataset of 18 Thai monophthong vowels (9 short and 9 long). Features were extracted using MFCCs along with their delta (velocity) and delta-delta (acceleration) counterparts. For each vowel, a statistical Gaussian Model was created using Maximum Likelihood Estimation to serve as an adaptive filter. The performance evaluation with test data showed an overall accuracy of 95.11%, demonstrating the effectiveness of the proposed method for Thai vowel classification.

Keywords: Speech Recognition, Thai Vowels, Mel-Frequency Cepstral Coefficients, Adaptive Filter

1. บทนำ

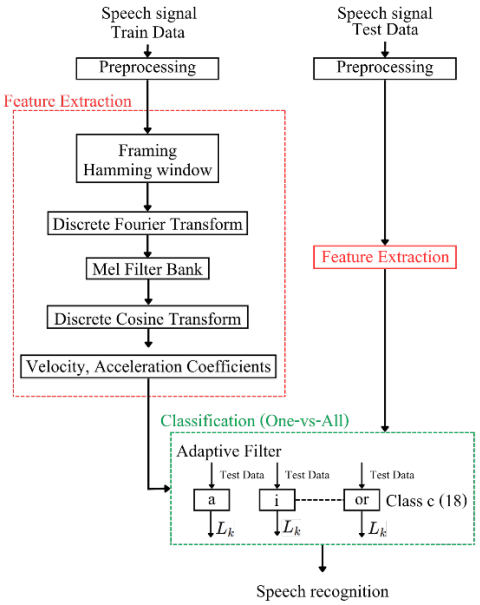
การรู้จำเสียงพูดมีความสำคัญอย่างยิ่งในการประยุกต์ใช้งานหลากหลายรูปแบบ เช่น ระบบสั่งการด้วยเสียง ระบบแปลงเสียงเป็นข้อความ และการวิเคราะห์เสียง สำหรับภาษาไทยซึ่งมีโครงสร้างทางเสียงที่ซับซ้อน โดยเฉพาะเสียงสระที่มีทั้งเสียงสั้นและเสียงยาว งานวิจัยที่ผ่านมาได้มีการศึกษาเทคนิคเพื่อเพิ่มประสิทธิภาพการรู้จำเสียงพูดภาษาไทย โดยงานวิจัย [1] ชี้ให้เห็นว่าฟังก์ชันหน้าต่างแบบ Hamming ให้ผลลัพธ์ที่ดีที่สุดเมื่อใช้ร่วมกับ Support Vector Machine นอกจากนี้ เทคนิคการสกัดคุณลักษณะเด่นของเสียงด้วย Mel-Frequency Cepstral Coefficients (MFCC) [2] ก็ได้รับการยอมรับอย่างกว้างขวางว่ามีประสิทธิภาพสูงจึงนำมาใช้ในงานวิจัยอัลกอริทึมการรู้จำเสียงสระภาษาไทยโดยใช้ตัวกรองแบบปรับตัวได้

งานวิจัยนี้นำเสนอการออกแบบและพัฒนาระบบรู้จำเสียงสระภาษาไทย โดยมีวัตถุประสงค์เพื่อประเมินประสิทธิภาพของตัวกรองแบบปรับตัวได้ (Adaptive Filter) ที่สร้างจากแบบจำลองทางสถิติ Maximum Likelihood Estimation ใช้วิธีการจำแนกประเภทแบบ One-vs-All เปรียบเทียบผลของอัลกอริทึมที่ใช้ Adaptive Filter กับ Support Vector Machine

2. วิธีการดำเนินงาน

อัลกอริทึมการรู้จำเสียงสระภาษาไทยโดยใช้ตัวกรองแบบปรับตัวได้ที่พัฒนาขึ้นแสดงใน รูปที่ 1 กระบวนการเริ่มจากการเตรียมสัญญาณเสียง (Preprocessing) ตามด้วยการสกัดคุณลักษณะ (Feature Extraction) โดยใช้สัมประสิทธิ์ MFCC และอนุพันธ์ (Delta และ Delta-Delta) จากนั้นสร้างตัวกรองแบบปรับตัวได้ (Adaptive Filter) ด้วยแบบจำลองเกาส์เซียนที่ประมาณค่าด้วยวิธี Maximum Likelihood Estimation (MLE) และขั้นตอนสุดท้ายคือการจำแนกประเภท (Classification) ด้วยวิธีวันวีเอสออล (One-vs-All) โดยเลือกเสียงสระจากคลาสที่มีค่า Log-Likelihood สูงที่สุด

*ผู้ประพันธ์บรรณกิจ

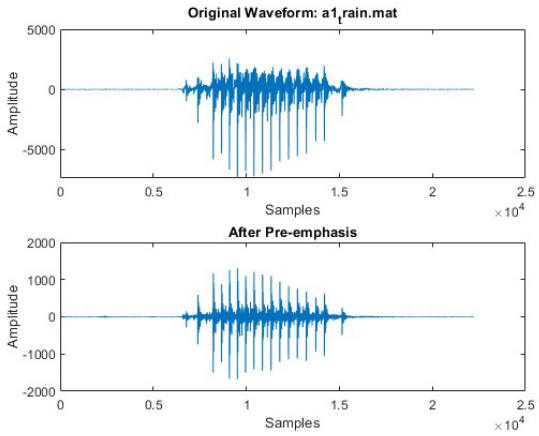


รูปที่ 1 อัลกอริทึมการรู้จำเสียงสระภาษาไทยโดยใช้ตัวกรองแบบปรับตัวได้

2.1 ชุดข้อมูลสัญญาณเสียงและการประมวลผลเบื้องต้น

งานวิจัยนี้ใช้ชุดข้อมูลเสียงสระเดี่ยวในภาษาไทย 18 เสียง แบ่งเป็นสระเสียงสั้น 9 เสียง (อะ, อิ, อึ, อุ, เอะ, แอะ, โอะ, เอาะ, เออะ) และสระเสียงยาว 9 เสียง (อา, อี, อือ, อู, เอ, แอ, โอ, ออ, เออ)

ในขั้นตอนการประมวลผลเบื้องต้น สัญญาณเสียงจะถูกกรองด้วย Pre-emphasis filter เพิ่มความชัดเจนของย่านความถี่สูงแสดงดังรูปที่ 2



รูปที่ 2 สัญญาณเสียงสระเสียง อะ

2.2 การสกัดคุณลักษณะ (Feature Extraction)

คุณลักษณะเด่นของเสียงถูกสกัดโดยใช้วิธี Mel-Frequency Cepstral Coefficients (MFCC) ซึ่งมีขั้นตอนดังนี้

2.2.1 การแบ่งเฟรม (Framing)

แบ่งสัญญาณเสียงออกเป็นเฟรมสั้นๆ ที่ทับซ้อนกัน โดยใช้หน้าต่างแบบแฮมมิ่ง (Hamming window)

2.2.2 การแปลงฟูเรียร์ (DFT)

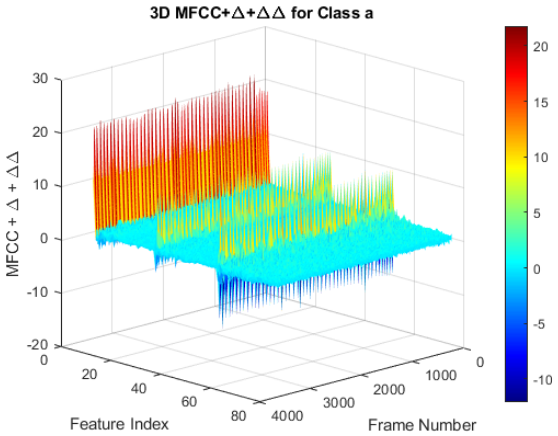
แปลงสัญญาณในแต่ละเฟรมจากโดเมนเวลาไปเป็นโดเมนความถี่ และคำนวณค่า Power Spectrum

2.2.3 Mel Filter Bank

นำผลลัพธ์มาผ่านชุดตัวกรอง Mel-scale filter bank เพื่อเลียนแบบการได้ยินของมนุษย์ [2]

2.2.4 Discrete Cosine Transform (DCT)

แปลงค่าที่ได้เพื่อให้ได้สัมประสิทธิ์ MFCC ซึ่งเป็นคุณลักษณะของเสียงในแต่ละเฟรม นอกจากนี้ ยังมีการคำนวณคุณลักษณะเพิ่มเติมคือ Delta (ความเร็ว) และ Delta-Delta (ความเร่ง) ของ MFCC เพื่อจับการเปลี่ยนแปลงของคุณลักษณะระหว่างเฟรม ทำให้ในแต่ละเฟรมได้เวกเตอร์คุณลักษณะ (Feature Vector) ขนาด 75 มิติ (Base MFCC 25 มิติ, Delta 25 มิติ และ Delta-Delta 25 มิติ) [3] แสดงผลในรูปที่ 3



รูปที่ 3 สกัดคุณลักษณะเฉพาะของสระเสียง อะ

2.3 การสร้าง Adaptive Filter ด้วย MLE

หลังจากการสกัดคุณลักษณะคุณลักษณะ (Feature Extraction) ได้เวกเตอร์ $\mathbf{X}$ ของชุดข้อมูลสำหรับสอน (Training Data) ขั้นตอนต่อไปคือการสร้างตัวกรองแบบปรับตัวได้ (Adaptive Filter) เพื่อรู้จำเสียงสระภาษาไทยแต่ละเสียง C

หลักการที่ใช้คือการประมาณค่าความน่าจะเป็นสูงสุด (Maximum Likelihood Estimation - MLE) ซึ่งเป็นวิธีการทางสถิติ ใช้แบบจำลองเกาส์เซียน (Gaussian Model) เพื่อประมาณค่าพารามิเตอร์ ค่า $\boldsymbol{\mu}_c$ (Mean Vector) ค่าเฉลี่ยของเวกเตอร์คุณลักษณะที่เป็นศูนย์กลางของเสียงสระ c แสดงดังสมการที่ (1) และ $\boldsymbol{\Sigma}_c$ (Covariance Matrix) ค่าที่อธิบายการกระจายตัวของเวกเตอร์คุณลักษณะรอบ ค่าเฉลี่ย แสดงดังสมการที่ (2) แล้วคำนวณความน่าจะเป็น (Likelihood) จากสมการ $p(\mathbf{x} \mid c)$ ซ้ำๆ จนกระทั่งได้ค่าพารามิเตอร์ที่ทำให้ ความน่าจะเป็นรวมของข้อมูลสำหรับสอน ทั้งหมดมีค่าสูงสุด (Maximum) แสดงดังสมการที่ (3) แสดงผลในรูปที่ 4

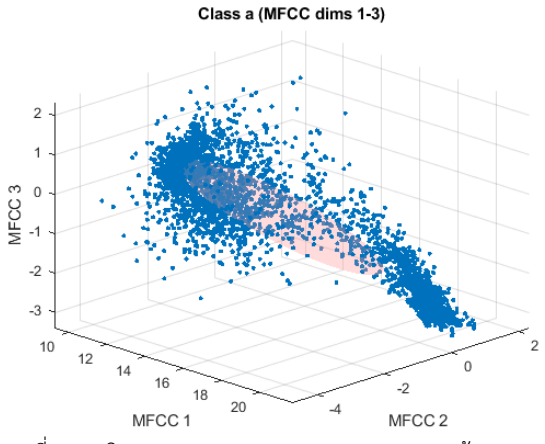
พารามิเตอร์ $\boldsymbol{\mu}_c$ และ $\boldsymbol{\Sigma}_c$ คือ ค่าสัมประสิทธิ์ของฟิลเตอร์ที่ต้องการหาค่าที่เหมาะสมที่สุด [4] [5]

$$\mu_c = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i^{(c)} \tag{1}$$

$$\boldsymbol{\Sigma}_c = \frac{1}{N}\sum_{i=1}^{N}(\mathbf{x}_i^{(c)} - \boldsymbol{\mu}_c)(\mathbf{x}_i^{(c)} - \boldsymbol{\mu}_c)^{\top} \tag{2}$$

$$p(\mathbf{x} \mid c) = \frac{1}{(2\pi)^{d/2} \mid \boldsymbol{\Sigma}_c \mid^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_c)^{\top}\boldsymbol{\Sigma}_c^{-1}(\mathbf{x}-\boldsymbol{\mu}_c)\right) \tag{3}$$

โดยที่ $p(\mathbf{x} \mid c)$ คือความหนาแน่นของความน่าจะเป็นของเวกเตอร์คุณลักษณะ $\mathbf{x}$ ภายใต้คลาส c, $\mathbf{x}$ คือเวกเตอร์คุณลักษณะของเฟรมเสียง (ขนาด d มิติ), c คือคลาสของสระ, d คือจำนวนมิติของเวกเตอร์คุณลักษณะ (ในที่นี้คือ 75), $\boldsymbol{\mu}_c$ คือเวกเตอร์ค่าเฉลี่ยของคลาส c, $\boldsymbol{\Sigma}_c$ คือเมทริกซ์ความแปรปรวนร่วมของคลาส c และ N คือจำนวนเฟรมทั้งหมดของคลาส



รูปที่ 4 การฟิต Gaussian Model (Adaptive Filter) ด้วย MLE

2.4 การจำแนกประเภท

การจำแนกประเภทข้อมูลทดสอบ (Test Data) แบบ One-vs-All

2.4.1การคำนวณ Log-Likelihood

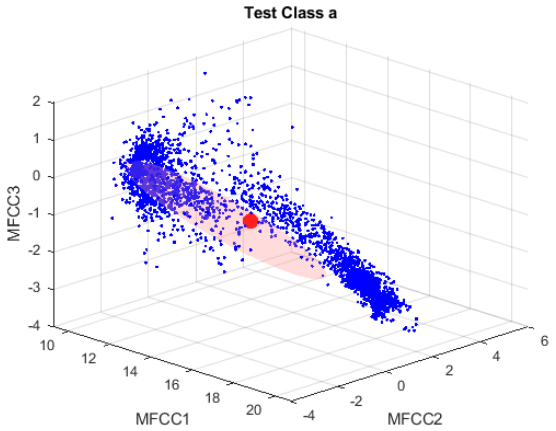
สำหรับไฟล์เสียงที่ต้องการทดสอบ จะมีการคำนวณค่า Sum of Log-Likelihood ของทุกเฟรมเทียบกับแบบจำลองเกาส์เซียนของแต่ละคลาสสระ ดังสมการที่ (4)

$$L_k = \sum_{t=1}^{T}\log(p(\mathbf{x}_t \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)) \tag{4}$$

โดย L_k คือค่า Log-Likelihood รวมของไฟล์ทดสอบภายใต้โมเดลของคลาส k

2.4.2 การตัดสินใจ

ระบบจะทำการเลือกว่าเสียงที่ทดสอบนั้นเป็นสระคลาสใด โดยเลือกคลาสที่ให้ค่า L_k สูงที่สุด แสดงผลในรูปที่ 5

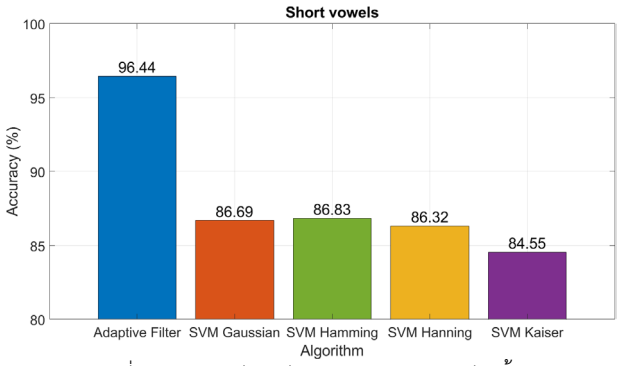


รูปที่ 5 การจำแนกประเภทข้อมูลทดสอบ (Test Data) สระเสียง อะ

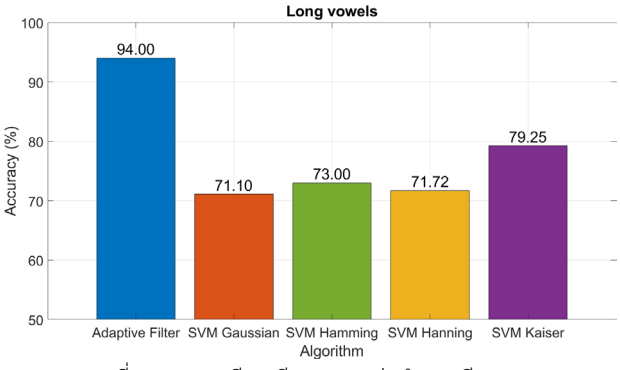
3. ผลการทดลอง

ระบบได้ถูกประเมินประสิทธิภาพโดยใช้ภาษาไทย 18 เสียง แบ่งเป็นสระเสียงสั้น 9 เสียง และสระเสียงยาว 9 เสียง รวมทั้งสิ้น 1,800 คำ โดยแบ่งเป็นข้อมูลสำหรับสอน (Training Data) 900 คำ และข้อมูลสำหรับทดสอบ (Testing Data) 900 คำ

ผลลัพธ์แสดงในกราฟแบ่งเป็นสระเสียงสั้น (Short vowels) รูปที่ 6 และสระเสียงยาว (Long vowels) รูปที่ 7 เปรียบเทียบผลของอัลกอริทึมที่ใช้ Adaptive Filter กับ Support Vector Machine โดยใช้หน้าต่างแบบ Gaussian window, Hamming window, Hanning window, Kaiser window ซึ่งแสดงให้เห็นความแม่นยำของอัลกอริทึม Adaptive Filter สระเสียงสั้น (Short vowels) อยู่ที่ 96.44% และสระเสียงยาว (Long vowels) อยู่ที่ 94%



รูปที่ 6 ผลการเปรียบเทียบความแม่นยำสระเสียงสั้น



รูปที่ 7 ผลการเปรียบเทียบความแม่นยำสระเสียงยาว

4. สรุป

งานวิจัยนี้ได้นำเสนอการพัฒนาระบบรู้จำเสียงสระภาษาไทยโดยใช้ตัวกรองแบบปรับตัวได้ที่สร้างจากแบบจำลองทางสถิติ Maximum Likelihood Estimation ร่วมกับการสกัดคุณลักษณะ MFCC และจำแนกประเภทด้วยวิธี One-vs-All ผลการทดลองแสดงให้เห็นว่าระบบที่พัฒนาขึ้นมีประสิทธิภาพสูงด้วยความแม่นยำ สระเสียงสั้น (Short vowels) อยู่ที่ 96.44% และสระเสียงยาว (Long vowels) อยู่ที่ 94% ซึ่งพิสูจน์ให้เห็นว่าแนวทางนี้เป็นแนวทางที่มีประสิทธิภาพและสามารถนำไปประยุกต์ใช้ในการจำแนกเสียงสระภาษาไทยได้เป็นอย่างดี สำหรับการพัฒนาในอนาคต อาจมีการปรับปรุงโดยการใช้แบบจำลองที่ซับซ้อนขึ้นทดสอบกับชุดข้อมูลที่หลากหลายมากขึ้นเพื่อเพิ่มประสิทธิภาพของระบบ

เอกสารอ้างอิง

[1] วรวุฒิ แจ่มหล้า, สุประวิทย์ เมืองเจริญ, ธีรพงศ์ อรชร, สกล อุดมศิริ, “การเปรียบเทียบประสิทธิภาพฟังก์ชันหน้าต่างในระบบรู้จำเสียงพูดภาษาไทยที่ใช้ซัพพอร์ตเวกเตอร์แมชชีน,” การประชุมวิชาการทางไฟฟ้า ครั้งที่ 43, 28 – 30 ตุลาคม 2563, 161-165

[2] สุประวิทย์ เมืองเจริญ, วรวุฒิ แจ่มหล้า, โสพล ด้วงสูงเนิน, สกล อุดมศิ, “การศึกษาการรู้จำเสียงภาษาไทยโดยใช้เทคนิคการสกัดเสียงหาสัมประสิทธิ์เซปสตรัมบนสเกลเมล,” การประชุมวิชาการทางไฟฟ้า ครั้งที่ 42, 30 ตุลาคม - 1 พฤศจิกายน 2562, 301-304

[3] Steve Young, Gunnar Evermann, Dan Kershaw, Gareth Moore, Julian Odell, Dave Ollason, Dan Povey, Valtcho Valtchev, and Phil Woodland, “The HTK Book,”. Cambridge University, 2002, 63-66

[4] Masahiro IWAHASHI, Sakol UDOMSIRI, “FUNCTIONALLY LAYERED VIDEO CODING FOR RIVER SURVEILLANCE,” *IEEE International Conference on Image Processing*, San Antonio, USA, 2007

[5] Christopher M. Bishop, “Pattern Recognition and Machine Learning,” Springer Nature, 2006, 430-437