

การรู้จำภาษาจากเสียงพูดโดยใช้ลักษณะทางสัทวิทยา



นางสาวศิรินุช บุญสุข

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรดุษฎีบัณฑิต
สาขาวิชาวิศวกรรมคอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2556
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

บทคัดย่อและแฟ้มข้อมูลฉบับเต็มของวิทยานิพนธ์ตั้งแต่ปีการศึกษา 2554 ที่ให้บริการในคลังปัญญาจุฬาฯ (CUIR)
เป็นแฟ้มข้อมูลของนิสิตเจ้าของวิทยานิพนธ์ ที่ส่งผ่านทางบัณฑิตวิทยาลัย

The abstract and full text of theses from the academic year 2011 in Chulalongkorn University Intellectual Repository (CUIR)
are the thesis authors' files submitted through the University Graduate School.

SPOKEN LANGUAGE RECOGNITION USING PHONOLOGICAL FEATURES

Miss Sirinoot Boonsuk

จุฬาลงกรณ์มหาวิทยาลัย

CHULALONGKORN UNIVERSITY

A Dissertation Submitted in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy Program in Computer Engineering

Department of Computer Engineering

Faculty of Engineering

Chulalongkorn University

Academic Year 2013

Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์	การรู้จำภาษาจากเสียงพูดโดยใช้ลักษณะทางสัทวิทยา
โดย	นางสาวศิรินุช บุญสุข
สาขาวิชา	วิศวกรรมคอมพิวเตอร์
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	รองศาสตราจารย์ ดร. อติวงศ์ สุชาโต
อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม	ผู้ช่วยศาสตราจารย์ ดร. โปรตปราน บุญยพุกกณะ ดร. ณัฐนันท์ ทัดพิทักษ์กุล

คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้รับวิทยานิพนธ์ฉบับนี้เป็นส่วน
หนึ่งของการศึกษาตามหลักสูตรปริญญาตรีบัณฑิต

.....คณบดีคณะวิศวกรรมศาสตร์
(ศาสตราจารย์ ดร. บัณฑิต เอื้ออาภรณ์)

คณะกรรมการสอบวิทยานิพนธ์

.....ประธานกรรมการ
(ศาสตราจารย์ ดร. ประภาส จงสฤษดิ์วัฒนา)

.....อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
(รองศาสตราจารย์ ดร. อติวงศ์ สุชาโต)

.....อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม
(ผู้ช่วยศาสตราจารย์ ดร. โปรตปราน บุญยพุกกณะ)

.....อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม
(ดร. ณัฐนันท์ ทัดพิทักษ์กุล)

.....กรรมการ
(ศาสตราจารย์ ดร. บุญเสริม กิจสิริกุล)

.....กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร. เศรษฐา ปานงาม)

.....กรรมการภายนอกมหาวิทยาลัย
(ดร. ชัย วุฒิวิวัฒน์ชัย)

ศิรินุช บุญสุข : การรู้จำภาษาจากเสียงพูดโดยใช้ลักษณะทางสัทวิทยา. (SPOKEN LANGUAGE RECOGNITION USING PHONOLOGICAL FEATURES) อ.ที่ปรึกษาวิทยานิพนธ์หลัก: รศ. ดร. อติวงศ์ สุชาโต, อ.ที่ปรึกษาวิทยานิพนธ์ร่วม: ผศ. ดร. โปรตปราณ บุญยพุกกณะ, ดร. ญัฐนันท์ ทัดพิทักษ์กุล, 265 หน้า.

การรู้จำภาษาเสียงพูดได้รับความสนใจในการนำมาใช้ในการรู้จำเสียงพูดที่รองรับหลายภาษาเพื่อเป็นขั้นตอนส่วนต้นเพื่อระบุภาษาของสัญญาณเสียงพูด แนวทางการรู้จำภาษาที่มีส่วนใหญ่อาศัยเทคนิคแบบจำลองทางสถิติร่วมกับค่าลักษณะทางเสียง การเรียงตัวของเสียง และฉันทลักษณ์ ตามการศึกษาความสัมพันธ์ระหว่างค่าพีเอฟ และภาษา ค่าพีเอฟเป็นข้อมูลเกี่ยวกับภาษาเพื่อจับค่าลักษณะทางอะคูสติก และแสดงข้อมูลการเรียงตัวของเสียงจากรูปแบบการเปลี่ยนแปลงของค่าพีเอฟในภาษาต่าง ๆ ระบบการรู้จำภาษาพูดที่เป็นที่ยอมรับในปัจจุบันเป็นการรวมระบบรู้จำย่อยที่ใช้แนวทางต่าง ๆ วิทยานิพนธ์นี้เสนอระบบการรู้จำภาษาพูดที่รวมระบบการรู้จำภาษาพูดย่อย 4 ระบบ ซึ่งประกอบด้วย 1) ระบบการรู้จำภาษาพูดแนวทางที่ใช้การจำลองลำดับหน่วยเสียงด้วยแบบจำลองเวกเตอร์สเปซ 2) ระบบการรู้จำภาษาพูดแนวทางที่ใช้หน่วยเสียงแลตทิซเอสวีเอ็ม 3) ระบบการรู้จำภาษาตามแนวทางการเรียงตัวของเสียงโดยใช้ค่าพีเอฟ และ 4) ระบบการรู้จำภาษาโดยใช้การจำแนกด้วยแบบจำลองลาเทนท์คอนดิชันนอลแรนดอมฟิลด์ร่วมกับค่าพีเอฟ ในระบบการรู้จำภาษาพูดแนวทางการเรียงตัวของเสียง ใช้การจำลองลำดับหน่วยเสียงด้วยซัพพอร์ตเวกเตอร์แมชชีน การถ่วงน้ำหนักของเทอมบนซูปเปอร์เวกเตอร์ของค่าความน่าจะเป็นของแบบจำลองเอ็นแกรมมีความจำเป็นต่อประสิทธิภาพของระบบรู้จำ เพราะการถ่วงน้ำหนักป้องกันฟังก์ชันเคอร์เนลของซัพพอร์ตเวกเตอร์แมชชีนจากการมีอิทธิพลของสมาชิกที่มีค่าความน่าจะเป็นที่มีค่ามาก วิทยานิพนธ์นี้สนใจการเพิ่มประสิทธิภาพการรู้จำภาษาโดยการรวมฟังก์ชันการถ่วงน้ำหนักกับสมาชิกของซูปเปอร์เวกเตอร์เข้าไว้ด้วยกัน การรวมกันของค่าความซ้ำซ้อนของเทอม (อาร์ดี) และค่าลือกของความถี่เทอม (ลือกทีเอฟ) ถูกเสนอเป็นฟังก์ชันการถ่วงน้ำหนักที่มีประสิทธิภาพในการรวมกันระหว่างค่าน้ำหนักแบบท้องถิ่น และค่าถ่วงน้ำหนักแบบครอบคลุม การถ่วงน้ำหนักนี้สามารถลดความถี่หน่วยที่ซ้ำซ้อนที่ปรากฏร่วมกันข้ามภาษาได้ สำหรับระบบการรู้จำภาษาตามแนวทางการเรียงตัวของเสียงโดยใช้ค่าพีเอฟ ใช้สถิติของรูปแบบการปรากฏร่วมกันของค่าพีเอฟบนภาษาที่แตกต่างกัน สำหรับระบบการรู้จำภาษาโดยใช้การจำแนกด้วยแบบจำลองลาเทนท์คอนดิชันนอลแรนดอมฟิลด์ร่วมกับค่าพีเอฟ แบบจำลองลาเทนท์คอนดิชันนอลแรนดอมฟิลด์ร่วมกับค่าพีเอฟถูกใช้เพื่อจับการเปลี่ยนแปลงแบบไดนามิกของลำดับค่าพีเอฟเพื่อสร้างแบบจำลองของภาษา ระบบอ้างอิงที่ถูกจัดสำหรับประเมินระบบการรู้จำภาษาพูดแต่ละแนวทาง และแบบรวมทุกระบบย่อย ผลการทดลองแสดงการทำให้ดีขึ้นเมื่อรวมผลลัพธ์ของระบบย่อยเข้าด้วยกัน การรวมค่าพีเอฟเข้าไว้กับระบบการรู้จำภาษาพูด ทำให้ประสิทธิภาพการรู้จำภาษาดีขึ้น

ภาควิชา วิศวกรรมคอมพิวเตอร์

ลายมือชื่อนิติดี

สาขาวิชา วิศวกรรมคอมพิวเตอร์

ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์หลัก

ปีการศึกษา 2556

ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์ร่วม

ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์ร่วม

5171831321 : MAJOR COMPUTER ENGINEERING

KEYWORDS: SPOKEN LANGUAGE RECOGNITION / PHONOLOGICAL FEATURES / LDCRF

SIRINOOT BOONSUK: SPOKEN LANGUAGE RECOGNITION USING PHONOLOGICAL FEATURES. ADVISOR: ASSOC. PROF. ATIWONG SUCHATO, Ph.D., ASST. PROF. PROADPRAN PUNYABUKKANA, Ph.D., NATTANUN THATPHITHAKKUL, Ph.D., 265 pp.

Spoken language recognition (SLR) has been of increasing interest in multilingual speech recognition as a pre-process for identifying the languages of speech utterances. Most existing SLR approaches apply statistical modeling techniques with acoustic, phonotactic and prosodic features. According to the studies of relationship between phonological features (PFs) and language, this thesis uses PF as the linguistic information to capture acoustic characteristics and to represent phonotactic information from the patterns of PF transition in different languages. The current state-of-the-art system is the fusion of different sub-systems. The proposed SLR system combining four sub-systems: 1) the phone sequence modeling followed by the vector space model (PRVSM), 2) lattice-SVM system, 3) The phonotactic SLR approach using co-occurrence of PFs, and 4) the SLR sub-system based on the latent-dynamic conditional random field (LDCRF) model using PFs. In the phonotactic SLR systems based on the Support Vector Machine (SVM) modeling, term weighting on the supervector of n-gram probabilities is critical to the recognition performance because the weighting prevents the SVM kernel from being dominated by a few large probabilities. This thesis focuses on enhancing the SLR performance by incorporating with term weighting function on the supervector entities. The combination of redundancy of term frequency (rd) and logarithm of term frequency (logtf) is proposed as the effective term weighting functions combining the local and global weighting. It can effectively eliminate the redundancy of unit frequency co-occurrence across languages. For the phonotactic approach using PF, the statistics of co-occurrence of PFs across different languages are captured. For the SLR systems based on LDCRF using PFs, the LDCRF model was employed to capture the dynamics of the PF attribute sequences for constructing language models. Baseline systems were conducted to evaluate the individual and the fused SLR system. The results showed improvements when combining the sub-systems and the results of integrating the PFs into SLR system can achieve better performance.

Department: Computer Engineering Student's Signature

Field of Study: Computer Engineering Advisor's Signature

Academic Year: 2013 Co-Advisor's Signature

Co-Advisor's Signature

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงได้ดีด้วยความอนุเคราะห์อย่างยิ่งจาก รองศาสตราจารย์ ดร.อดิวงค์ สุชาติ อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก และผู้ช่วยศาสตราจารย์ ดร.โปรดปราน บุญยพุกกณะ ดร.ชัย วุฒิวิวัฒน์ชัย และดร.ณัฐนันท์ ทัดพิทักษ์กุล อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม ซึ่งทุกท่านเสียสละเวลาอันมีค่าในการให้คำปรึกษา คำแนะนำ ตลอดจนแก้ไขข้อบกพร่องต่าง ๆ และเสนอแนวทางเกี่ยวกับการทำวิทยานิพนธ์ ให้การสนับสนุนโอกาสที่ดีต่าง ๆ และให้กำลังใจซึ่งเป็นเรื่องสำคัญยิ่งในระหว่างการศึกษาและดำเนินการวิจัย เพื่อทำวิทยานิพนธ์ในครั้งนี้ประสบความสำเร็จได้ด้วยดี ผู้วิจัยขอขอบพระคุณเป็นอย่างสูงไว้ ณ โอกาสนี้

ขอขอบพระคุณ Dr. Li Haizhou Dr. Ma Bin และ Dr. Zhu Donglai จากสถาบัน Institute of Infocomm Research (I2R) ประเทศสิงคโปร์ ซึ่งให้คำแนะนำในการทำวิจัยด้านความรู้จำภาษา และทรัพยากรที่ใช้ในการทำวิจัย

ขอขอบพระคุณศาสตราจารย์ ดร.ประภาส จงสถิตยวัฒนา ศาสตราจารย์ ดร.บุญเสริม กิจสิริกุล และผู้ช่วยศาสตราจารย์ ดร.เศรษฐา ปานงาม กรรมการสอบวิทยานิพนธ์ที่ให้คำแนะนำในการทำวิทยานิพนธ์ฉบับนี้ ตลอดจนแนวคิดในการเรียน การทำวิจัย และอบรมสั่งสอนสิ่งที้นำไปใช้ในชีวิต

ขอขอบคุณทุนสนับสนุนเพื่อการศึกษาและทำวิจัย จากสถาบันบัณฑิตวิทยาศาสตร์และเทคโนโลยีไทย (TGIST) สำนักงานพัฒนาวิทยาศาสตร์และเทคโนโลยีแห่งชาติ (สวทช.) สัญญารับทุนเลขที่ TG-44-09-51-066D

นอกจากนี้ผู้วิจัยขอขอบคุณพี่น้องในห้องปฏิบัติการระบบภาษาพูด และห้องปฏิบัติการเทคโนโลยีช่วยเหลือผู้พิการ ภาควิชาวิศวกรรมคอมพิวเตอร์ จุฬาลงกรณ์มหาวิทยาลัย ที่ให้กำลังใจในการทำวิจัย ขอขอบคุณคุณนิพนธ์ ชินะธิมาตรมงคล ที่สนับสนุนและให้กำลังใจระหว่างทำวิทยานิพนธ์ และช่วยตรวจสอบแก้ไขวิทยานิพนธ์เล่มนี้ เหนือสิ่งอื่นใดขอกราบขอบพระคุณ คุณประสิทธิ์ บุญสุข และคุณโกศล บุญสุข บิดามารดาของผู้วิจัย ตลอดจนครอบครัวที่เข้าใจ ให้กำลังใจ และสนับสนุนการทำวิจัยในครั้งนี้มาโดยตลอดจนประสบความสำเร็จได้ด้วยดี คุณประโยชน์ได้อันพึงมีจากวิทยานิพนธ์นี้ ผู้วิจัยขอมอบ และอุทิศแด่ผู้มีพระคุณทุกๆ ท่านที่มีส่วนร่วมในการให้การอบรมผู้วิจัยตลอดมา

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ง
บทคัดย่อภาษาอังกฤษ.....	จ
กิตติกรรมประกาศ.....	ฉ
สารบัญ.....	ช
สารบัญตาราง.....	ฅ
สารบัญภาพ.....	ด
บทที่ 1 บทนำ.....	1
1.1 ความเป็นมาและความสำคัญของปัญหา.....	1
1.2 วัตถุประสงค์ของการวิจัย.....	3
1.3 ขอบเขตของการวิจัย.....	3
1.4 ลำดับขั้นตอนในการเสนอผลการวิจัย.....	4
1.5 ประโยชน์ที่คาดว่าจะได้รับ.....	4
1.6 ผลงานตีพิมพ์จากวิทยานิพนธ์.....	5
1.7 ลำดับการจัดเรียงเนื้อหาในวิทยานิพนธ์.....	5
บทที่ 2 ทฤษฎีและวรรณกรรมที่เกี่ยวข้อง.....	6
2.1 ทฤษฎีที่เกี่ยวข้องกับภาษา.....	6
2.1.1 ภาษามนุษย์.....	6
2.1.2 ตระกูลภาษา.....	7
2.1.3 ตระกูลภาษาที่ศึกษาในวิทยานิพนธ์.....	8
2.1.4 ลักษณะเฉพาะของภาษา.....	9
2.1.5 ค่าลักษณะสำคัญเกี่ยวกับระบบเสียงในภาษา.....	11
2.2 ทฤษฎีทางสถิติที่ใช้ในการวิเคราะห์ข้อมูล.....	14
2.2.1 การวิเคราะห์ความแปรปรวน.....	14
2.2.2 การพิจารณาค่าลักษณะสำคัญทางสถิติด้วยแผนภาพกล่อง.....	15
2.2.3 การวิเคราะห์ความแปรปรวนพหุคูณ.....	16
2.3 แบบจำลองลาเทนท์ไดนามิคคอนดิชันนอลแรนดอมฟิลด์ (LDCRF).....	17

2.4	ซัพพอร์ตเวกเตอร์แมชชีน	20
2.5	วรรณกรรมที่เกี่ยวข้อง	21
2.5.1	วรรณกรรมที่พิจารณาระบุภาษาด้วยมนุษย์	21
2.5.2	วรรณกรรมที่เกี่ยวข้องในการรู้จำภาษา	21
2.5.2.1	วรรณกรรมการรู้จำภาษาแนวทางใช้ค่าลักษณะทางเสียง	22
2.5.2.2	วรรณกรรมการรู้จำภาษาใช้กฎเกณฑ์การเรียงตัวหน่วยเสียง	24
2.5.2.3	วรรณกรรมการรู้จำภาษาที่รวมจากหลายระบบ	26
2.5.3	วรรณกรรมที่เกี่ยวข้องกับค่าลักษณะเฉพาะค่าพีเอฟ	26
2.5.4	วรรณกรรมที่เกี่ยวกับแบบจำลองคอนดิชันนอลแรนดอมฟิลด์	28
บทที่ 3	ระบบการรู้จำภาษาพูด	29
3.1	ภาพรวมของระบบการรู้จำภาษาพูด	29
3.1.1	กรอบการรู้จำภาษาแบบความน่าจะเป็น (Probabilistic SLR framework)	29
3.1.2	ขั้นตอนการทำงานของกรรรู้จำภาษา	30
3.2	ขั้นตอนการสร้างระบบการรู้จำภาษาพูดที่นำเสนอ	31
3.3	ข้อมูลลักษณะเฉพาะของภาษาจากระบบการรู้จำภาษาพูดที่นำเสนอ	33
3.4	ขั้นตอนการรวมผลลัพธ์และการตัดสินใจรู้จำภาษาจากระบบรู้จำย่อยแบบต่าง ๆ	35
3.4.1	การทำให้คะแนนเป็นบรรทัดฐาน	35
3.4.2	การรวมผลลัพธ์	35
3.4.2.1	การรวมคะแนนแบบเชิงเส้น	36
3.4.2.2	กระบวนการทำงานส่วนหลังด้วยแบบจำลองเกาส์เซียน (Gaussian back end)	37
3.5	ข้อมูลที่ใช้ในการทดลอง	38
3.5.1	ข้อมูลเสียงที่ใช้ในการพัฒนาและทดสอบระบบตรวจหาลักษณะเฉพาะค่าพีเอฟ	38
3.5.2	ข้อมูลเสียงที่ใช้ในการศึกษาการกระจายค่าพีเอฟในแต่ละภาษา	38
3.5.3	ข้อมูลเสียงที่ใช้ในการทดลองระบบการรู้จำภาษาพูด	39
3.5.4	ข้อมูลเสียงที่ใช้ในการทดลองการรู้จำภาษาตามแนวทางการเรียงหน่วยเสียง	40
3.5.4.1	ข้อมูลเสียงที่ใช้ในการฝึกฝนระบบการรู้จำภาษาพูดตามแนวทางการเรียงหน่วยเสียง	40

3.5.4.2 ข้อมูลเสียงที่ใช้ในการทดสอบระบบการรู้จำภาษาพูดตามแนวทางการเรียงหน่วยเสียง	41
3.6 การวัดประสิทธิภาพของการรู้จำภาษา.....	41
3.6.1 การวัดประสิทธิภาพจากความแม่นยำ.....	42
3.6.2 การวัดประสิทธิภาพจากความผิดพลาด	42
3.6.3 การวัดประสิทธิภาพจากความเชื่อมั่น	43
3.6.4 การวัดประสิทธิภาพของการรู้จำภาษาด้วยค่า EER.....	43
3.6.5 การวัดประสิทธิภาพของการรู้จำภาษาด้วยฟังก์ชันค่าการตรวจหา.....	45
3.6.6 การวัดประสิทธิภาพด้วยเส้นโค้ง ROC (Receiver Operating Characteristic curve)	47
3.7 การตรวจสอบแบบไขว้	48
บทที่ 4 การรู้จำภาษาตามแนวทางการเรียงหน่วยเสียง	49
4.1 บทนำ	49
4.2 ระบบการรู้จำภาษาพูดแบบใช้ระบบรู้จำเสียงตามด้วยแบบจำลองเวกเตอร์สเปซ	51
4.2.1 ตัวรู้จำเสียง.....	51
4.2.2 การคำนวณค่าเทอม	52
4.2.3 แบบจำลองเวกเตอร์สเปซ.....	52
4.2.4 ซับพอร์ตเวกเตอร์แมชชีน	53
4.2.5 ส่วนการตัดสินใจระบุภาษา.....	54
4.3 ระบบการรู้จำภาษาพูดแบบใช้หน่วยเสียงแลตทิซตามด้วยแบบจำลองเวกเตอร์สเปซ	55
4.3.1 หน่วยเสียงแลตทิซ.....	56
4.3.2 ตัวรู้จำเสียง.....	57
4.3.3 การคำนวณค่าเทอม	57
4.3.4 การตัดสินใจระบุภาษา	58
4.4 ฟังก์ชันการถ่วงน้ำหนักเทอม	58
4.4.1 ฟังก์ชันการถ่วงน้ำหนักเทอม	59
4.4.2 การทำให้เป็นบรรทัดฐานหรือนอร์มอลไลซ์	62
4.4.3 การประกอบกันของแต่ละฟังก์ชัน	62

4.5 บทสรุป.....	63
บทที่ 5 การรู้จำภาษาโดยอาศัยค่าลักษณะสำคัญเกี่ยวกับระบบเสียงในภาษา.....	64
5.1 ที่มาของการใช้ค่าพีเอฟในระบบการรู้จำภาษาพูด.....	64
5.2 ระบบตรวจหาลักษณะเฉพาะค่าพีเอฟโดยโครงข่ายประสาทเทียมหลายชั้น.....	65
5.2.1 การศึกษาค่าพีเอฟ.....	66
5.2.2 การศึกษาการกระจายค่าพีเอฟในแต่ละภาษา.....	68
5.2.2.1 ผลการวิเคราะห์ความแปรปรวนค่าพีเอฟในแต่ละภาษา.....	71
5.2.2.2 ผลการศึกษาค่าพีเอฟบนเสียง ‘n’.....	72
5.2.2.3 ผลการศึกษาค่าพีเอฟบนเสียง ‘aa’.....	73
5.2.2.4 สรุปผลการศึกษาการกระจายค่าพีเอฟในแต่ละภาษา.....	75
5.2.3 การศึกษาความถี่การปรากฏร่วมกันของค่าพีเอฟในแต่ละภาษา.....	75
5.3 ระบบการรู้จำภาษาพูดโดยใช้การจำแนกด้วยแบบจำลองลาเทนท์ไดนามิกคอนดิชันนอลแรนดอมฟิลด์ร่วมกับค่าพีเอฟ.....	78
5.3.1 การสกัดค่าลักษณะสำคัญทางเสียง.....	78
5.3.1.1 ค่าพีเอฟ PF.....	79
5.3.1.2 ค่าลักษณะสำคัญเอสดีซีที่คำนวณจากค่าพีเอฟ PF.....	79
5.3.2 การจำแนกภาษา.....	80
5.4 ระบบการรู้จำภาษาพูดตามแนวทางการเรียงหน่วยเสียงโดยใช้ค่าพีเอฟ.....	83
5.4.1 ตัวสร้างเหตุการณ์ที่เกิดขึ้นร่วมกันของค่าพีเอฟหรือโคออกเคอร์เรนซ์ค่าพีเอฟ.....	84
5.4.1.1 การสร้างตัวแทนเหตุการณ์ที่เกิดขึ้นร่วมกันของค่าพีเอฟหรือโคออกเคอร์เรนซ์ค่าพีเอฟในแต่ละเฟรม (Co-occurrence of PF within one frame).....	85
5.4.1.2 การหาตัวแทนเหตุการณ์ที่เกิดขึ้นร่วมกันของโคออกเคอร์เรนซ์ค่าพีเอฟในกรอบการพิจารณา (Co-occurrence PF in observing window).....	87
5.4.2 การถ่วงน้ำหนักและทำให้เป็นบรรทัดฐาน.....	91
5.4.3 ตัวจำแนกภาษา.....	91
5.4.4 การรวมตัวจำแนกภาษา.....	92
บทที่ 6 การทดลองและวิธีการวัดผลการทดลอง.....	93
6.1 การสร้างระบบตรวจหาลักษณะเฉพาะค่าพีเอฟ.....	93

6.2 ระบบอ้างอิง	94
6.3 การศึกษาความแปรปรวนของค่าพีเอฟในหน่วยเสียง.....	96
6.3.1 การศึกษาค่าพีเอฟด้วยการวิเคราะห์ความแปรปรวนพหุคูณ	96
6.3.2 การศึกษาค่าพีเอฟด้วยการจัดกลุ่มภาษา	97
6.3.3 การศึกษาค่าพีเอฟด้วยการวิเคราะห์ความแปรปรวน	99
6.3.4 การศึกษาค่าพีเอฟด้วยการวิเคราะห์ความแปรปรวนพหุคูณสำหรับแต่ละคู่ภาษา	106
6.4 การทดลองเพื่อประเมินประสิทธิภาพของการรู้จำภาษา	106
6.4.1 การทดลองที่ 1: ประเมินประสิทธิภาพของระบบตรวจหาลักษณะเฉพาะค่าพีเอฟ... ..	106
6.4.2 การทดลองที่ 2: ประเมินประสิทธิภาพของการรู้จำภาษาตามแนวทางการเรียงหน่วยเสียง.....	106
6.4.3 การทดลองที่ 3: ประเมินประสิทธิภาพของระบบการรู้จำภาษาพูดโดยใช้การจำแนกด้วยแบบจำลองลาเทนท์ไดนามิกคอนดิชันนอลแรนดอมฟิลด์ร่วมกับค่าพีเอฟ	108
6.4.4 การทดลองที่ 4: ประเมินประสิทธิภาพของการรู้จำภาษาตามแนวทางการเรียงหน่วยเสียงโดยใช้ค่าพีเอฟ.....	113
6.4.5 การทดลองที่ 5: ประเมินประสิทธิภาพของระบบการรู้จำภาษาพูดที่น่าเสนอ เมื่อรวมผลลัพธ์และการตัดสินใจรู้จำภาษาจากระบบรู้จำย่อยแนวต่าง ๆ	116
บทที่ 7 ผลการทดลองและวิเคราะห์ผลการทดลอง.....	117
7.1 ประเมินผลระบบตรวจหาลักษณะเฉพาะค่าพีเอฟ	117
7.2 ประเมินผลระบบการรู้จำภาษาพูดตามแนวทางการเรียงหน่วยเสียง	119
7.2.1 ผลการทดลองที่ 2.1: ประเมินประสิทธิภาพของระบบการรู้จำภาษาพูดแบบพ็อาร์วีเอ็สเอ็มเปรียบเทียบกับระบบการรู้จำภาษาพูดแบบแลตทิซ-วีเอ็สเอ็ม	119
7.2.2 ผลการทดลองที่ 2.2: เปรียบเทียบประสิทธิภาพของฟังก์ชันถ่วงน้ำหนักแบบต่าง ๆ	120
7.2.3 ผลการทดลองที่ 2.3: เปรียบเทียบวิธีการทำให้เป็นบรรทัดฐานที่แตกต่างกัน	123
7.2.4 บทอภิปรายการทดลองฟังก์ชันการถ่วงน้ำหนักค่าเทอม	123
7.2.5 บทสรุปการรู้จำภาษาตามแนวทางการเรียงหน่วยเสียง	125
7.2.6 ผลการทดลองที่ 2.4: วัดประสิทธิภาพของระบบการรู้จำภาษาพูดแบบพ็อาร์วีเอ็สเอ็มเปรียบเทียบกับระบบการรู้จำภาษาพูดแบบแลตทิซ-วีเอ็สเอ็ม โดยวัดผลการรู้จำภาษาบน 8 ภาษา.....	125

7.2.7 ผลการทดลองที่ 2.5: วัดประสิทธิภาพของระบบการรู้จำภาษาพูดแบบพ็อดวีเอ็สเอ็มเปรียบเทียบกับระบบการรู้จำภาษาพูดแบบแลตทิซ-วีเอ็สเอ็ม โดยวัดผลการรู้จำภาษาบน 8 ภาษาบนฐานข้อมูลเสียง NIST LRE07 เซตข้อมูลทดสอบ.....	126
7.3 ประเมินประสิทธิภาพของระบบการรู้จำภาษาพูดโดยใช้การจำแนกด้วยแบบจำลองลาเทนท์ไดนามิคคอนดิชันนอลแรนดอมฟิลด์ร่วมกับค่าพีเอฟ.....	126
7.3.1 การทดลองที่ 3.1 ที่วัดประสิทธิภาพของระบบการรู้จำภาษาพูดเมื่อใช้การจำแนกด้วยแบบจำลองลาเทนท์ไดนามิคคอนดิชันนอลแรนดอมฟิลด์ร่วมกับการใช้ค่าลักษณะทางเสียงต่าง ๆ	126
7.3.2 ผลการทดลองที่ 3.2: การประเมินผลกระทบเมื่อลดกลุ่มภาษาเป้าหมายที่มีความคลุมเครือต่อประสิทธิภาพของระบบการรู้จำภาษาพูดที่ใช้ค่าพีเอฟ และระบบการรู้จำภาษาพูดที่ใช้ค่าลักษณะทางเสียงอื่น ๆ.....	137
7.3.3 ผลการทดลองที่ 3.3: วัดประสิทธิภาพของระบบการรู้จำภาษาพูดที่ใช้ตัวจำแนกด้วยแบบจำลองลาเทนท์ไดนามิคคอนดิชันนอลแรนดอมฟิลด์	138
7.3.4 ผลการทดลองที่ 3.4: วัดประสิทธิภาพของระบบการรู้จำภาษาพูดเมื่อลดชนิดค่าพีเอฟที่ใช้ฝึกฝนตัวจำแนกภาษา.....	141
7.3.5 ผลการทดลองที่ 3.5: วัดประสิทธิภาพของระบบการรู้จำภาษาพูดที่นำเสนอเปรียบเทียบกับระบบอ้างอิงโอเวกเตอร์.....	142
7.3.6 ผลการทดลองที่ 3.6: วัดประสิทธิภาพของการรู้จำภาษาของระบบการรู้จำภาษาพูดที่ใช้ตัวจำแนกแบบหนึ่งต่อหนึ่ง และการใช้ตัวจำแนกแบบหนึ่งต่อทั้งหมด	144
7.3.7 บทสรุปการรู้จำภาษาโดยใช้การจำแนกด้วยแบบจำลองลาเทนท์ไดนามิคคอนดิชันนอลแรนดอมฟิลด์ร่วมกับค่าพีเอฟ.....	145
7.4 ประเมินประสิทธิภาพของการรู้จำภาษาตามแนวทางการเรียงหน่วยเสียงโดยใช้ค่าพีเอฟ....	146
7.4.1 ผลการทดลองที่ 4.1: ประเมินผลประสิทธิภาพของการรู้จำภาษาแนวทางการเรียงหน่วยเสียงที่ใช้ค่าพีเอฟโดยใช้การพิจารณาโคออคเคอร์เรนซ์ค่าพีเอฟ.....	146
7.4.2 ผลการทดลองที่ 4.2: ศึกษาผลกระทบของขนาดกรอบการพิจารณาของการปรากฏร่วมกันของโคออคเคอร์เรนซ์ค่าพีเอฟต่อประสิทธิภาพการรู้จำภาษาแนวทางการเรียงหน่วยเสียงโดยใช้ค่าพีเอฟ	147
7.4.3 ผลการทดลองที่ 4.3: ศึกษาผลกระทบของชนิดค่าพีเอฟที่พิจารณาการปรากฏร่วมกันของค่าพีเอฟต่อประสิทธิภาพการรู้จำภาษาแนวทางการเรียงหน่วยเสียงโดยใช้ค่าพีเอฟ	150
7.4.4 ผลการทดลองที่ 4.4: ศึกษาผลกระทบของการเพิ่มจำนวนโคออคเคอร์เรนซ์ค่าพีเอฟต่อประสิทธิภาพการรู้จำภาษาแนวทางการเรียงหน่วยเสียงค่าพีเอฟ.....	150

7.4.5 ผลการทดลองที่ 4.5: เปรียบเทียบประสิทธิภาพการรู้จำภาษาจากระบบที่นำเสนอ กับ ผลการรู้จำภาษาของระบบอ้างอิงตามแนวทางการเรียงหน่วยเสียง	151
7.4.6 บทสรุปการรู้จำภาษาตามแนวทางการเรียงตัวของเสียงโดยใช้ค่าพีเอฟ	157
7.5 ประเมินประสิทธิภาพของระบบการรู้จำภาษาพูดที่นำเสนอ เมื่อรวมผลลัพธ์และการตัดสินใจ รู้จำภาษาจากระบบรู้จำย่อย.....	157
บทที่ 8 สรุปผลการวิจัยและข้อเสนอแนะ	162
8.1 สรุปผลการวิจัย.....	162
8.2 ข้อเสนอแนะ.....	163
ภาคผนวก ก ความถี่ของหน่วยเสียงในฐานข้อมูลเสียง OGI ส่วนเสียงพูดต่อเนื่อง	165
ภาคผนวก ข-1 Box plot ของค่าพีเอฟซึ่งได้จากผลลัพธ์ของโครงข่ายประสาทเทียมหลายชั้นสำหรับ การประมาณค่าที่เหมือนกับค่าพีเอฟเป้าหมายบนหน่วยเสียง ‘n’	171
ภาคผนวก ข-2 Box plot ของค่าพีเอฟซึ่งได้จากผลลัพธ์ของโครงข่ายประสาทเทียมหลายชั้นสำหรับ การประมาณค่าที่เหมือนกับค่าพีเอฟเป้าหมายบนหน่วยเสียง ‘aa’	183
ภาคผนวก ค boxplot และการวิเคราะห์ความแปรปรวนของค่าพีเอฟสำหรับแต่ละภาษา.....	195
ภาคผนวก ง Vowel chart	208
ภาคผนวก จ เปอร์เซ็นต์ของความถี่การปรากฏโคออกเคอเรนซ์ค่าพีเอฟหลัก ร่วมกับค่าพีเอฟชนิดอื่น ในแต่ละภาษา	209
ภาคผนวก ช ผลการศึกษาค่าพีเอฟด้วยการวิเคราะห์ความแปรปรวนพหุคูณ สำหรับแต่ละคู่ภาษา.....	215
ภาคผนวก ซ เส้นโค้ง ROC ของผลการจำแนกภาษาโดยใช้แบบจำลองลาเทนท์ไดนามิคคอนดิชันนอล แรนดอมฟิลด์ร่วมกับค่าพีเอฟ และแบบจำลองคอนดิชันนอลแรนดอมฟิลด์ร่วมกับค่าพีเอฟ สำหรับแต่ละ คู่ภาษา.....	217
ภาคผนวก ฌ เส้นโค้ง ROC ของผลการจำแนกภาษาโดยใช้แบบจำลองลาเทนท์ไดนามิคคอนดิชัน นอลแรนดอมฟิลด์ร่วมกับค่าพีเอฟ และค่าลักษณะสำคัญต่าง ๆ สำหรับแต่ละคู่ภาษา.....	231
ภาคผนวก ฉ-1 คะแนนการตัดสินใจของระบบการรู้จำภาษาแบบแลตทิซ-วีเอสเอ็ม โดยแสดงคะแนน สำหรับแต่ละภาษา	245
ภาคผนวก ฉ-2 คะแนนการตัดสินใจโดยใช้ระบบการรู้จำภาษาพูดแนวทางการเรียงตัวของเสียงโดยใช้ ค่าพีเอฟ สำหรับแต่ละภาษา	249
ภาคผนวก ฉ-3 คะแนนการตัดสินใจโดยใช้การจำแนกด้วยแบบจำลองลาเทนท์ไดนามิคคอนดิชันนอล แรนดอมฟิลด์แบบหนึ่งต่อทั้งหมดร่วมกับค่าพีเอฟ ซึ่งแสดงคะแนนสำหรับแต่ละภาษา	253
รายการอ้างอิง	257

ประวัติผู้เขียนวิทยานิพนธ์265



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

สารบัญตาราง

หน้า

ตารางที่ 2-1 ตารางแสดงความหมายของค่าพีเอฟตามนิยาม SPE.....	12
ตารางที่ 2-2 ตารางแสดงตัวอย่างค่าพีเอฟของคำว่า “khun” ในภาษาไทย.....	13
ตารางที่ 3-1 จำนวนชั่วโมงของเสียงพูดในงานวิจัย แบ่งตามแต่ละภาษา.....	39
ตารางที่ 3-2 จำนวนชั่วโมงของเสียงพูดฐานข้อมูลเสียง CallFriend แบ่งตามแต่ละภาษา.....	40
ตารางที่ 4-1 การประกอบกันของการถ่วงน้ำหนักแบบต่าง ๆ.....	63
ตารางที่ 5-1 ความถี่ของหน่วยเสียง 20 อันดับแรกในฐานข้อมูลเสียง OGI.....	70
ตารางที่ 5-2 ค่า F-ratio และค่า P-value ของค่าพีเอฟแต่ละชนิด บนหน่วยเสียง ‘n’ และ ‘aa’	71
ตารางที่ 6-1 ผลการศึกษาค่าลักษณะสำคัญโดยใช้การวิเคราะห์ความแปรปรวนพหุคูณ.....	97
ตารางที่ 6-2 ค่าลักษณะสำคัญทางเสียงที่ใช้ในการทดลองเพื่อวัดประสิทธิภาพของการรู้จำภาษา เมื่อค่าลักษณะสำคัญทางเสียงชนิดต่าง ๆ	110
ตารางที่ 7-1 ผลความแม่นยำของระบบตรวจหาลักษณะเฉพาะ ค่าพีเอฟแต่ละชนิด	118
ตารางที่ 7-2 ผลความแม่นยำระบบตรวจหาลักษณะเฉพาะค่าพีเอฟ เมื่อใช้จำนวนโหนดซ่อนค่าต่าง ๆ.....	118
ตารางที่ 7-3 ค่าความผิดพลาดการรู้จำภาษา EER ของระบบการรู้จำภาษาพูดแบบพ็อดาร์วีเอ็สเอ็ม และระบบการรู้จำภาษาพูดแบบแลตทิซ-วีเอ็สเอ็ม เมื่อใช้ฟังก์ชันถ่วงน้ำหนักแบบต่าง ๆ และวิธีการทำให้เป็นบรรทัดฐานแบบต่าง ๆ.....	120
ตารางที่ 7-4 ค่าความผิดพลาดการรู้จำภาษา EER ของระบบการรู้จำภาษาพูดแบบพ็อดาร์วีเอ็สเอ็ม และระบบการรู้จำภาษาพูดแบบแลตทิซ-วีเอ็สเอ็ม เมื่อใช้การประกอบกันของฟังก์ชันถ่วง น้ำหนักแบบต่าง ๆ และวิธีการทำให้เป็นบรรทัดฐานแบบต่าง ๆ	122
ตารางที่ 7-5 ค่าความผิดพลาดการรู้จำภาษา EER ของระบบการรู้จำภาษาพูดแบบพ็อดาร์วีเอ็สเอ็ม และระบบการรู้จำภาษาพูดแบบแลตทิซ-วีเอ็สเอ็ม สำหรับข้อมูล NIST LRE07 เซตข้อมูล พัฒนา.....	126
ตารางที่ 7-6 ค่า C_{avg} ของระบบการรู้จำภาษาพูดแบบพ็อดาร์วีเอ็สเอ็ม และระบบการรู้จำภาษาพูด แบบแลตทิซ-วีเอ็สเอ็ม บนฐานข้อมูลเซตทดสอบ 8 ภาษา.....	126
ตารางที่ 7-7 ค่าเฉลี่ยเปอร์เซ็นต์ความผิดพลาดของการจำแนกภาษา เมื่อใช้ค่าลักษณะสำคัญชนิด	

ต่าง ๆ	128
ตารางที่ 7-8 ความแม่นยำของตัวจำแนกภาษาเมื่อใช้ค่าลักษณะสำคัญชนิดต่าง ๆ (PF, SDCPF, PLP, SDCPLP, SDC, และ MFCC).....	131
ตารางที่ 7-9 เปอร์เซนต์ความแม่นยำของการจำแนกของตัวจำแนกภาษา 28 คู่ภาษาเมื่อใช้การจำแนกด้วยแบบจำลองลาเทนท์ไดนามิกคอนดิชันนอลแรนดอมฟีลด์ร่วมกับค่าพีเอฟชนิดต่าง ๆ เมื่อลดค่าพีเอฟที่ละชนิด.....	133
ตารางที่ 7-10 เปอร์เซนต์ความผิดพลาดการจำแนกภาษา เมื่อลดความคลุมเครือของภาษาเป้าหมาย	137
ตารางที่ 7-11 เปอร์เซนต์ความผิดพลาดของการจำแนกภาษา เมื่อใช้แบบจำลองชนิดต่าง ๆ	138
ตารางที่ 7-12 เปอร์เซนต์ความแม่นยำของการจำแนกภาษาเมื่อใช้แบบจำลองชนิดต่าง ๆ ร่วมกับค่าพีเอฟ สำหรับตัวจำแนก 28 คู่ภาษา	140
ตารางที่ 7-13 เปอร์เซนต์ความผิดพลาดของการจำแนกภาษาเมื่อใช้การจำแนกด้วยแบบจำลองลาเทนท์ไดนามิกคอนดิชันนอลแรนดอมฟีลด์ร่วมกับค่าพีเอฟ เมื่อลดค่าพีเอฟหนึ่งชนิด	141
ตารางที่ 7-14 ค่าเฉลี่ยเปอร์เซนต์ความผิดพลาดของการจำแนกภาษาของระบบที่นำเสนอเปรียบเทียบกับระบบการรู้จำภาษาพูดที่ใช้โอเวกเตอร์	143
ตารางที่ 7-15 ค่าความผิดพลาดการรู้จำภาษา EER และค่าความผิดพลาดของการจำแนกภาษาของระบบการรู้จำภาษาพูดที่ใช้ตัวจำแนกแบบหนึ่งต่อหนึ่ง และการใช้ตัวจำแนกแบบหนึ่งต่อทั้งหมด	144
ตารางที่ 7-16 ค่า C_{avg} ของระบบการรู้จำภาษาพูดการจำแนกด้วยแบบจำลองลาเทนท์ไดนามิกคอนดิชันนอลแรนดอมฟีลด์ร่วมกับค่าพีเอฟ บนฐานข้อมูลเซตทดสอบ 8 ภาษา.....	145
ตารางที่ 7-17 ค่าเฉลี่ยเปอร์เซนต์ความแม่นยำของการจำแนกภาษาจาก 28 คู่ภาษาเมื่อพิจารณาโคออร์ดิเนตเรนซ์ค่าพีเอฟจากการจับคู่ค่าพีเอฟ	147
ตารางที่ 7-18 ค่าเฉลี่ยของค่าความแม่นยำของการจำแนกภาษาจาก 28 คู่ภาษา ที่ใช้การปรากฏร่วมกันของการจับคู่ค่าพีเอฟ เมื่อใช้ขนาดกรอบการพิจารณาที่แตกต่างกัน	149
ตารางที่ 7-19 ค่าความแม่นยำการจำแนกภาษาเมื่อใช้พิจารณาการเกิดร่วมกันโคออร์ดิเนตเรนซ์ค่าพีเอฟ 2 โคออร์ดิเนตเรนซ์ค่าพีเอฟ	153

หน้า

ตารางที่ 7-20 ผลการรู้จำภาษาของระบบที่นำเสนอเปรียบเทียบกับ ระบบอ้างอิงที่ 2.1 และ 2.2156	
ตารางที่ 7-21 ค่า C_{avg} ของระบบการรู้จำภาษาพูดการจำแนกด้วยแบบจำลองลาเทนท์ไดนามิก	
คอนดิชันนอลแรนดอมฟิลด์ร่วมกับค่าพีเอฟ บนฐานข้อมูลเซตทดสอบ 8 ภาษา.....	157
ตารางที่ 7-22 ค่า C_{avg} และค่า C_{llr} ของระบบการรู้จำภาษาพูดระบบที่รวมคะแนนจากระบบการ	
รู้จำภาษาพูดแนวทางต่าง ๆ.....	161



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

สารบัญภาพ

หน้า

ภาพที่ 2-1 แผนภูมิรูปวงกลมแสดงเป็นร้อยละตามผู้พูดของภาษามนุษย์ทั่วโลก.....	7
ภาพที่ 2-2 แผนผังตระกูลภาษา ของภาษาของเสียงพูดที่ใช้ในวิทยานิพนธ์	9
ภาพที่ 2-3 แผนผังข้อมูลลักษณะเฉพาะของภาษา	9
ภาพที่ 2-4 ลักษณะของแผนภาพกล่อง	15
ภาพที่ 2-5 แบบจำลองคอนดิชันนอลแรนดอมฟิลด์และแบบจำลองลาเทนท์ไดนามิกคอนดิชันนอล แรนดอมฟิลด์.....	17
ภาพที่ 3-1 ขั้นตอนการทำงานของความรู้จำภาษา.....	30
ภาพที่ 3-2 สถาปัตยกรรมของระบบการรู้จำภาษาพูดที่นำเสนอ.....	31
ภาพที่ 3-3 องค์ประกอบของข้อมูลลักษณะเฉพาะที่ใช้ในระบบการรู้จำภาษาพูดที่เสนอ	33
ภาพที่ 3-4 ความสัมพันธ์ระหว่างเส้นกราฟค่า FAR FRR และจุดตัดค่า EER	44
ภาพที่ 3-5 เส้นโค้ง ROC แสดงความสัมพันธ์ระหว่างอัตราจริงบวกและอัตราเท็จบวก.....	47
ภาพที่ 4-1 องค์ประกอบของระบบการรู้จำภาษาพูดแบบพ็อดาร์วีเอสเอ็ม.....	51
ภาพที่ 4-2 องค์ประกอบของระบบการรู้จำภาษาพูดแบบแลตทิซวีเอสเอ็ม	56
ภาพที่ 4-3 แสดงตัวอย่างของหน่วยเสียงแลตทิซของคำว่า “HELLO”	57
ภาพที่ 5-1 สเปกโตรแกรมและตัวอย่างค่าพีเอฟ.....	67
ภาพที่ 5-2 แผนภาพกล่องค่าพีเอฟ Voice และค่าพีเอฟ Consonantal บนเสียง ‘n’.....	72
ภาพที่ 5-3 แผนภาพกล่องค่าพีเอฟ Voice และค่าพีเอฟ Continianut บนเสียง ‘aa’	74
ภาพที่ 5-4 อัตราของความถี่การปรากฏโคออกเคอเรนซ์ค่าพีเอฟระหว่างภาษาจีนมิน และ ภาษาไทย.....	76
ภาพที่ 5-5 อัตราของความถี่การปรากฏโคออกเคอเรนซ์ค่าพีเอฟระหว่างภาษาอารบิก และ ภาษาไทย.....	77
ภาพที่ 5-6 เปอร์เซ็นต์ของความถี่การปรากฏโคออกเคอเรนซ์ค่าพีเอฟ Vocalic ร่วมกับค่าพีเอฟ ชนิดอื่น ในแต่ละภาษา	77
ภาพที่ 5-7 สถาปัตยกรรมของระบบการรู้จำภาษาพูด โดยใช้การจำแนกร่วมกับค่าพีเอฟ.....	78

หน้า

ภาพที่ 5-8 การคำนวณค่าเอสดีซี	79
ภาพที่ 5-9 การทำงานของระบบการรู้จำภาษาพูดตามแนวทางการเรียงของค่าพีเอฟ	84
ภาพที่ 5-10 เหตุการณ์ที่เกิดขึ้นร่วมกันของค่าพีเอฟสอง ชนิดภายในเฟรมเดียวกัน	86
ภาพที่ 5-11 แสดงการเกิดร่วมกันของสมาชิกโคออคเคอร์เรนซ์ค่าพีเอฟที่อยู่ในกรอบการพิจารณา	88
ภาพที่ 5-12 แสดงกรอบการพิจารณาโคออคเคอร์เรนซ์ค่าพีเอฟที่อยู่ในแต่ละเฟรมภายในขอบเขตที่กำหนด.....	88
ภาพที่ 6-1 ระบบการระบุภาษาที่ใช้ค่าพีเอฟของ Nunez และ Ramon	95
ภาพที่ 6-2 การรวมกลุ่มเป็นแบบลำดับชั้น จากค่าเฉลี่ยของค่าพีเอฟ	99
ภาพที่ 6-3 การรวมกลุ่มเป็นแบบลำดับชั้น จากค่าเฉลี่ยของค่าสัมประสิทธิ์ PLP	99
ภาพที่ 6-4 การรวมกลุ่มเป็นแบบลำดับชั้น จากค่าเฉลี่ยของค่าสัมประสิทธิ์ MFCC	99
ภาพที่ 6-5 แผนภาพสระ	105
ภาพที่ 7-1 (ก) ตัวอย่างหน่วยเสียง 20 อันดับแรกของหน่วยเสียงที่มีความถี่ของการปรากฏบ่อยที่สุดในฐานข้อมูลเสียง ซึ่งเปรียบเทียบกับน้ำหนักจากฟังก์ชันการถ่วงน้ำหนักค่าความถี่เอกสารผกผัน และค่าความซ้ำซ้อนของเทอม และ (ข) หน่วยเสียง 20 อันดับแรกที่มีความแตกต่างกันมากที่สุดระหว่างการถ่วงน้ำหนัก ค่าความถี่เอกสารผกผัน และค่าความซ้ำซ้อนของเทอม	124
ภาพที่ 7-2 เส้นโค้ง ROC ของผลลัพธ์ของการจำแนก 28 คู่ภาษา เมื่อใช้ค่าลักษณะสำคัญชนิดต่าง ๆ.....	129
ภาพที่ 7-3 ค่าความผิดพลาดของการจำแนกภาษาด้วยแบบจำลองลาเทนท์ไดนามิคคอนดิชันนอลแรนดอมฟิลด์ร่วมกับค่าลักษณะสำคัญชนิดต่าง ๆโดยวัดค่าความผิดพลาดจาก 28 คู่ภาษา:	
ภาพที่ 7-4 เส้นโค้ง ROC ของผลการจำแนกภาษาโดยใช้แบบจำลองลาเทนท์ไดนามิคคอนดิชันนอลแรนดอมฟิลด์ร่วมกับค่าพีเอฟ และแบบจำลองคอนดิชันนอลแรนดอมฟิลด์ร่วมกับค่าพีเอฟ	139
ภาพที่ 7-5 แสดงแผนภาพกล่องของค่าความถูกต้องของการจำแนกภาษาจากตัวจำแนกคู่ภาษา 28 คู่ภาษาเมื่อใช้ขนาดกรอบการพิจารณาแตกต่างกัน.....	150
ภาพที่ 7-6 ค่าความผิดพลาดของการจำแนกภาษาของ 28 คู่ภาษาแนวทางการเรียงหน่วยเสียง	

หน้า

เมื่อใช้โคออคเคอเรนซ์ค่าพีเอฟ เปรียบเทียบกับการใช้หน่วยเสียง154

ภาพที่ 7-7 ค่าความผิดพลาดของการจำแนกภาษาของ 28 คู่ภาษาแนวทางการเรียงหน่วยเสียง
 เมื่อใช้โคออคเคอเรนซ์ค่าพีเอฟ เปรียบเทียบกับระบบอ้างอิงที่ 2.1 และระบบอ้างอิงที่
 2.2155



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ในปัจจุบันการติดต่อสื่อสารก้าวหน้าอย่างรวดเร็ว และกว้างขวางมากขึ้น เนื่องจากภาษาพูดของมนุษย์ที่มีจำนวนมาก และต้องการการติดต่อสื่อสารกันระหว่างหลายภาษาเพิ่มขึ้น ทำให้มีการพัฒนาระบบเกี่ยวกับเทคโนโลยีทางด้านเสียงพูดที่สามารถติดต่อสื่อสารระหว่างหลายภาษาได้ และระบบอัตโนมัติที่ให้บริการผู้ใช้งานสามารถตอบสนองทันที จึงมีความต้องการเทคโนโลยีที่สามารถระบุภาษาของเสียงพูดได้โดยอัตโนมัติ ได้แก่ ระบบรู้จำเสียงพูดที่สามารถรองรับกับการรู้จำเสียงพูดของหลายภาษา [1, 2] ระบบบริการข้อมูลซึ่งเป็นระบบการสนทนาที่รองรับภาษาต่าง ๆ และตอบสนองเลือกภาษาสำหรับการสนทนาของผู้ใช้งานได้โดยอัตโนมัติ เช่น ระบบสนทนาที่มีการโอนสายอัตโนมัติ (Call routing) [3] ระบบการเรียกติดต่อหน่วยงานฉุกเฉิน (Emergency call service) [4] ระบบการแปลภาษาอัตโนมัติ (Speech translation) [5] เป็นต้น ซึ่งระบบดังกล่าวต้องการความสามารถในการจำแนกหรือระบุภาษาของเสียงพูดจากผู้ใช้ ทำให้การรู้จำภาษามีบทบาทสำคัญในการพัฒนาระบบเกี่ยวกับเทคโนโลยีทางด้านเสียงพูดมนุษย์ และงานวิจัยด้านการรู้จำภาษาได้รับความสนใจมากขึ้น

การรู้จำภาษาของเสียงพูด (Spoken Language Recognition: SLR) เป็นกระบวนการตัดสินใจระบุภาษาของเสียงพูด การรู้จำภาษาเป็นปัญหาที่ท้าทาย เนื่องจากการรู้จำภาษาจำลองกระบวนการตัดสินใจของมนุษย์ในการระบุภาษาของเสียงพูด ซึ่งมนุษย์สามารถตัดสินใจระบุภาษาของเสียงที่ได้ยินว่าเป็นภาษาใด จากการประมวลผลว่าภาษาของเสียงที่ได้ยินใกล้เคียงกับความรู้ที่เคยรู้มาก่อน อย่างไรก็ตามความสามารถของการระบุภาษาของมนุษย์ลดลงเมื่อมนุษย์ได้ยินเสียงที่ไม่คุ้นเคยมาก่อนทำให้การตัดสินใจระบุภาษาผิดพลาดได้ มนุษย์จึงต้องการความรู้ และเวลาในการเรียนรู้ภาษาใหม่ที่ไม่คุ้นเคยมาก่อนเพื่อใช้ตัดสินใจระบุภาษา ทำให้ระบบการรู้จำภาษาพูดของเสียงพูดแบบอัตโนมัติได้รับความสนใจมากขึ้น ปัญหาในกระบวนการของการรู้จำภาษาสามารถแบ่งตามหลักการพิจารณา ดังนี้

- 1) การพิจารณาตามเซตของภาษา ได้แก่ การรู้จำภาษาภายในเซตแบบปิด (Closed-set) และการรู้จำภาษาภายในเซตแบบเปิด (Open-set) การรู้จำภาษาภายในเซตแบบปิดเป็นการพิจารณาตัดสินใจคำตอบภาษาของเสียงพูดในโดเมนเซตของภาษาพูดที่มีจำนวนจำกัด ขณะที่การรู้จำภาษาภายในเซตแบบเปิดเป็นการพิจารณาตัดสินใจคำตอบภาษาของเสียงพูดที่อยู่นอกขอบเขตของภาษาที่กำหนด ซึ่งการตัดสินใจภาษาที่ไม่อยู่ในเซตภาษาดังกล่าวนี้ เรียกว่าคำตอบของการรู้จำภาษา เป็นภาษาที่ไม่อยู่ในเซตที่พิจารณา (Out-of-set) ซึ่งในวิทยานิพนธ์นี้มีขอบเขตการศึกษาการรู้จำภาษาภายในเซตแบบปิด

- 2) การพิจารณาตามวิธีการตัดสินใจ ได้แก่ การระบุภาษา (Language Identification: LID) และการตรวจหาภาษา (Language detection) การระบุภาษาคือการตัดสินใจระบุคำตอบภาษาของเสียงพูดว่าเป็นภาษาใดจากภาษาภายในเซตของภาษาที่มีจำนวนจำกัด การระบุภาษาต้องการความรู้

ที่มาก่อน (Prior knowledge) ในการตัดสินใจ ขณะที่การตรวจหาภาษา เป็นการตัดสินใจจากการสมมติฐานว่า คำตอบภาษาของเสียงพูดที่พิจารณาเป็นภาษาที่พิจารณาหรือไม่ โดยเปรียบเทียบกับคำตอบภาษาของเสียงพูดเป็นภาษาอื่นที่ไม่ใช่ภาษาที่พิจารณาซึ่งการตรวจหาภาษาสามารถพิจารณาคำตอบจากสองทางเลือกได้

เนื่องจากลักษณะเฉพาะของภาษา (Language characteristic) สามารถพิจารณาได้จากข้อมูลหลายค่า วิทยานิพนธ์นี้สนใจการใช้ความรู้จากหลายองค์ประกอบบนสัญญาณเสียงพูดมาประกอบกันเป็นข้อมูลในการตัดสินใจระบุภาษา โดยเสนอระบบการรู้จำภาษาพูดที่รวมตัวจากระบบการรู้จำภาษาพูดย่อยที่ใช้หลายแนวทาง โดยใช้วิธีแบบจำลองทางสถิติที่แตกต่างกัน และใช้ค่าลักษณะสำคัญที่แตกต่างกัน เพื่อรวบรวมความรู้จากข้อมูลในระดับที่แตกต่างกันมาประกอบการตัดสินใจระบุภาษา ซึ่งข้อมูลที่ใช้ในวิทยานิพนธ์นี้ได้แก่ ค่าลักษณะสำคัญทางเสียง การเรียงกันของหน่วยเสียง (phonotactic) และค่าลักษณะสำคัญเกี่ยวกับระบบเสียงในภาษา (Phonological Features: PF) หรือค่าพีเอฟ เป็นต้น

วิทยานิพนธ์นี้เสนอระบบการรู้จำภาษาพูดที่ใช้ข้อมูลจากกฎเกณฑ์การเรียงกันของหน่วยเสียง โดยมีแนวคิดมาจากการพิจารณาลำดับการเรียงของหน่วยเสียงในแต่ละภาษามีรูปแบบของแบบจำลองทางภาษาที่แตกต่างกัน เช่น การประกอบกันของบางคู่หน่วยเสียงปรากฏในบางภาษา ขณะที่การประกอบกันของบางคู่หน่วยเสียงนั้นไม่ปรากฏในบางภาษาได้ เป็นต้น ดังนั้น ข้อมูลการเรียงกันของหน่วยเสียง สามารถใช้แสดงข้อมูลลักษณะเฉพาะของภาษาได้ ระบบการรู้จำภาษาพูดตามแนวทางการเรียงของหน่วยเสียงที่ใช้ในวิทยานิพนธ์นี้เป็นระบบรู้จำเสียงตามด้วยแบบจำลองเวกเตอร์สเปซ อย่างไรก็ตามอย่างไรก็ตามข้อมูลการเรียงของหน่วยเสียง สามารถเกิดข้อผิดพลาดได้จากขั้นตอนการถอดรหัสหน่วยเสียงของระบบรู้จำเสียง ส่งผลให้หน่วยเสียงที่นำไปสร้างแบบจำลองภาษามีข้อผิดพลาด หรือได้แบบจำลองภาษาที่ไม่เหมาะสมในการบอกความแตกต่างระหว่างภาษา ผู้วิจัยนี้จึงใช้ข้อมูลหน่วยเสียงแลตทิซ แทนหน่วยเสียง เพื่อลดข้อผิดพลาดจากขั้นตอนถอดรหัสหน่วยเสียงของระบบรู้จำเสียง วิทยานิพนธ์นี้จึงใช้ระบบการรู้จำภาษาพูดแบบใช้หน่วยเสียงแลตทิซตามด้วยการใช้แบบจำลองเวกเตอร์สเปซ ซึ่งรายละเอียดของระบบการรู้จำภาษาพูดแนวทางดังกล่าวอธิบายในบทที่ 4 นอกจากนี้วิทยานิพนธ์นี้เสนอการเพิ่มประสิทธิภาพของระบบการรู้จำภาษาพูดตามแนวทางการเรียงของหน่วยเสียง โดยการใช้ฟังก์ชันการถ่วงน้ำหนัก เพื่อให้คะแนนความสามารถค่าลักษณะสำคัญที่ใช้ในการบอกความแตกต่างระหว่างภาษา

นอกจากการศึกษาเพื่อลดข้อผิดพลาดจากระบบรู้จำเสียง วิทยานิพนธ์นี้เสนอการใช้ข้อมูลจากค่าลักษณะสำคัญชนิดอื่น เพื่อเพิ่มประสิทธิภาพของระบบการรู้จำภาษาพูด โดยเสนอการใช้ค่าพีเอฟ แทนการใช้หน่วยเสียง เพื่อแสดงคุณสมบัติเฉพาะของแต่ละภาษา โดยแนวคิดของการใช้ค่าพีเอฟในวิทยานิพนธ์ มาจากหน่วยเสียงเดียวกันในแต่ละภาษามีค่าลักษณะการออกเสียงแตกต่างกัน ดังนั้นจึงเป็นแรงจูงใจให้วิทยานิพนธ์ที่เสนอการใช้ค่าพีเอฟที่ปรากฏบนหน่วยเสียงที่มีค่าแตกต่างกัน เพื่อแสดงข้อมูลลักษณะเฉพาะของภาษา และเพื่อลดช่องว่างระหว่างระดับข้อมูลหน่วยเสียง และค่าลักษณะสำคัญทางเสียง จากระดับเซกเมนต์ (Segment) ของหน่วยเสียง และระดับเฟรม นั่นคือ ค่าพีเอฟเป็นตัวแทนค่าลักษณะการออกเสียงในระดับที่เป็นส่วนประกอบภายในเซกเมนต์ ซึ่งมีขนาดหน่วย

เล็กกว่าหน่วยเสียง และละเอียดกว่าระดับหน่วยเสียง จึงสามารถลดความแตกต่างจากระดับข้อมูลได้นอกจากนี้แต่ละค่าพีเอฟที่แสดงบนระดับเฟรม สามารถอธิบายค่าพารามิเตอร์การออกเสียงของหน่วยเสียง และอธิบายความสัมพันธ์ระหว่างค่าลักษณะการออกเสียงที่ปรากฏร่วมกันของฐานกรณ์ต่าง ๆ ได้อย่างมีความหมายมากกว่าค่าลักษณะสำคัญทางเสียงที่สกัดจากสัญญาณเสียงพูด

วิทยานิพนธ์ที่นำเสนอการใช้การปรากฏร่วมกันของค่าพีเอฟ เพื่อแสดงความสัมพันธ์ของการประกอบกันระหว่างค่าลักษณะการออกเสียง และแสดงลำดับการเรียงกันของค่าลักษณะการออกเสียงที่ปรากฏบนเสียงพูด ซึ่งมีแนวคิดมาจากแบบจำลองภาษา และข้อมูลการเรียงกันของหน่วยเสียงของระบบการรู้จำภาษาพูดตามแนวทางการเรียงกันของหน่วยเสียง นั่นคือ จากลำดับของหน่วยเสียงบนเสียงพูด ทำให้การปรากฏร่วมกันของค่าพีเอฟในแต่ละเฟรมของหน่วยเสียงมีรูปแบบแตกต่างกัน และแบบจำลองภาษาการปรากฏร่วมกันของค่าพีเอฟบนสัญญาณเสียงพูดที่แตกต่างกัน ดังนั้นวิทยานิพนธ์นี้เสนอระบบการรู้จำภาษาพูดที่ใช้ค่าพีเอฟ ซึ่งประกอบด้วย ระบบการรู้จำภาษาพูดโดยใช้การจำแนกด้วยแบบจำลองลาเทนต์คอนดิชันนอลแรนดอมฟิลด์ร่วมกับค่าพีเอฟ และระบบการรู้จำภาษาพูดตามแนวทางการเรียงตัวหน่วยเสียงโดยใช้ค่าพีเอฟ โดยรายละเอียดของแต่ละระบบอธิบายในบทที่ 5

1.2 วัตถุประสงค์ของการวิจัย

- 1) เพื่อเสนอระบบการรู้จำภาษาพูดที่รวมระบบการรู้จำแนวทางต่าง ๆ ที่ใช้ค่าลักษณะสำคัญที่มีความสามารถในการจำแนกภาษาจากข้อมูลในระดับที่แตกต่างกัน
- 2) เพื่อเสนอ และทำการวิเคราะห์ค่าลักษณะสำคัญที่ใช้จำแนกที่ทำให้ประสิทธิภาพของระบบการรู้จำภาษาพูดดีขึ้น
- 3) เพื่อเสนอแนวทางการศึกษา และประเมินค่าลักษณะสำคัญที่ใช้จำแนกที่ใช้ในงานการรู้จำภาษา
- 4) เพื่อเสนอการใช้ค่าลักษณะการออกเสียงเป็นค่าลักษณะสำคัญในระบบการรู้จำภาษาพูด และรวมคะแนนการตัดสินใจจากระบบการรู้จำภาษาพูดที่นำเสนอ

1.3 ขอบเขตของการวิจัย

- 1) ระบบการรู้จำภาษาพูดที่นำเสนอใช้ค่าลักษณะสำคัญมาจากข้อมูลลักษณะสำคัญทางเสียง ข้อมูลการเรียงหน่วยเสียง และข้อมูลการออกเสียงเพื่อบอกความแตกต่างระหว่างภาษา
- 2) การเปรียบเทียบประสิทธิภาพของระบบการรู้จำภาษาพูดเปรียบเทียบกับระบบอ้างอิงทดสอบกับฐานข้อมูลเสียงที่เป็นส่วนหนึ่งของข้อมูลที่ใช้ในการวัดประเมินผลการรู้จำภาษาที่จัดขึ้นด้วยองค์กร NIST ปี คศ. 2007 สำหรับ 8 ภาษา ประกอบด้วย ภาษาอารบิก ภาษาเบงกาลี ภาษารัสเซีย ภาษาไทย ภาษาเยอรมัน ภาษาจีนมิน ภาษาจีนวู และ

ภาษาจีนกวางตุ้ง โดยวัดผลการรู้จำภาษาที่ระบุเป้าหมายของการรู้จำภาษาภายในเขต
ปิด

- 3) ผลลัพธ์ของวิทยานิพนธ์นี้คือการรู้จำภาษาของเสียงพูดว่าภาษาใดเป็นคำตอบของเสียงพูด

1.4 ลำดับขั้นตอนในการเสนอผลการวิจัย

- 4) ขั้นตอนการศึกษาเบื้องต้น
 - 4.1) ศึกษาวรรณกรรมที่เกี่ยวข้องกับการรู้จำภาษา
 - 4.2) ศึกษาวรรณกรรมที่เกี่ยวข้องกับการค่าการออกเสียงในระบบภาษา
 - 4.3) ศึกษาเทคนิคการเรียนรู้ด้วยเครื่องจักร (Machine learning) ที่ใช้สำหรับการจำแนกประเภท
- 5) ขั้นตอนการออกแบบระบบและทำการทดลองออกแบบการทำงานของระบบการรู้จำภาษาพูดแต่ละระบบย่อย
 - 5.2) สำหรับระบบการรู้จำภาษาพูดที่ใช้ค่าพีเอฟ
 - 5.2.1) หาวิธีการสกัดค่าการออกเสียง เพื่อใช้เป็นค่าลักษณะสำคัญหลักของระบบการรู้จำภาษาพูดที่น่าเสนอ
 - 5.2.2) วิเคราะห์และหาตัวแทนของค่าพีเอฟเพื่อแสดงข้อมูลการออกเสียงที่เหมาะสม ที่สามารถบอกความแตกต่างระหว่างภาษาได้
 - 5.3) สร้างแบบจำลองในการจำแนกประเภท
 - 5.4) ออกแบบการทดลอง
 - 5.5) ทดสอบการทำงานและบันทึกผลการทดลอง
 - 5.6) วัดผลความแม่นยำและวิเคราะห์ผลการทดลองของระบบการรู้จำภาษาพูด
 - 5.7) รวมคะแนนของการตัดสินใจจากระบบการรู้จำภาษาพูดแต่ละระบบ
 - 5.8) ทดสอบผลการรู้จำภาษาที่มาจาก การตัดสินใจจากระบบการรู้จำภาษาพูดแต่ละระบบ
- 6) สรุปผลและเรียบเรียงวิทยานิพนธ์

1.5 ประโยชน์ที่คาดว่าจะได้รับ

- 1) ได้แนวทางสำหรับการรู้จำภาษา โดยมาจากระบบการรู้จำภาษาพูดใช้ค่าลักษณะสำคัญต่าง ๆ เช่น การเรียงกันของหน่วยเสียง การใช้ค่าลักษณะการออกเสียง และรวมคำตอบจากระบบภาษาจากแนวทางต่าง ๆ

- 2) เสนอการใช้ค่าลักษณะการออกเสียง เพื่อใช้ในการรู้จำภาษาได้
- 3) เสนอการประเมินค่าลักษณะสำคัญที่เสนอ เพื่อใช้ในการบอกความแตกต่างของภาษา สำหรับการนำไปใช้ในระบบการรู้จำภาษาพูดตามแนวทางที่นำเสนอได้

1.6 ผลงานตีพิมพ์จากวิทยานิพนธ์

ส่วนหนึ่งของวิทยานิพนธ์นี้ได้ตีพิมพ์เป็นบทความทางวิชาการในหัวข้อเรื่อง “A Study of Term Weighting in Phonotactic Approach to Spoken Language Recognition” จัดทำโดย “Sirinoot Boonsuk, Donglai Zhu, Bin Ma, Atiwong Suchato, Proadpran Punyabukkana, Nattanun Thatphithakkul and Chai Wutiwiwatchai” ถูกนำเสนอในงานประชุมวิชาการ “INTERSPEECH” จัดที่ประเทศญี่ปุ่น ในวันที่ 26-30 กันยายน 2010

ส่วนหนึ่งของวิทยานิพนธ์นี้ได้ตีพิมพ์เป็นวารสารวิชาการในหัวข้อเรื่อง “Language Recognition Using Latent Dynamic Conditional Random Field Model with Phonological Features” จัดทำโดย “Sirinoot Boonsuk, Atiwong Suchato, Proadpran Punyabukkana, Chai Wutiwiwatchai and Nattanun Thatphithakkul” ถูกนำเสนอในวารสาร J. Mathematical Problems in Engineering, in Recent Advancements in Signal Processing and Machine Learning Special issues, Hindawi Publishing Corporation, Volume 2014, Article ID 250160, <http://dx.doi.org/10.1155/2014/250160> ตีพิมพ์วันที่ 20 กุมภาพันธ์ 2014 (ISI impact factor 1.383)

1.7 ลำดับการจัดเรียงเนื้อหาในวิทยานิพนธ์

ในวิทยานิพนธ์นี้ได้แบ่งเนื้อหาออกเป็น 8 บทดังต่อไปนี้ บทที่ 1 เป็นบทนำซึ่งกล่าวถึง ความ เป็นมาและความสำคัญของปัญหา วัตถุประสงค์ของการวิจัย ขอบเขตของการวิจัย ลำดับขั้นตอนใน การเสนอผลการวิจัย ประโยชน์ที่คาดว่าจะได้รับและผลงานตีพิมพ์จากวิทยานิพนธ์ บทที่ 2 กล่าวถึง แนวคิดและทฤษฎี ประกอบด้วย ทฤษฎีที่เกี่ยวกับภาษา ทฤษฎีทางสถิติที่ใช้ในการวิเคราะห์ข้อมูล และวรรณกรรมที่เกี่ยวข้องกับการระบุภาษา ในบทที่ 3 กล่าวถึงระบบการรู้จำภาษาพูดที่นำเสนอ ขั้นตอนการดำเนินงานวิจัย ข้อมูลที่ใช้ในการทดลอง และวิธีการวัดผลการทดลอง บทที่ 4 อธิบาย ระบบการรู้จำภาษาพูดตามแนวทางการเรียงหน่วยเสียง บทที่ 5 แสดงระบบการรู้จำภาษาพูดโดย อาศัยค่าลักษณะสำคัญเกี่ยวกับระบบเสียงในภาษา ได้แก่ ระบบการรู้จำภาษาพูดโดยใช้การจำแนก ด้วยแบบจำลองลาเทนต์คอนดิชันนอลแรนดอมฟิลด์ร่วมกับค่าพีเอฟและระบบการรู้จำภาษาพูดตาม แนวทางกฎเกณฑ์ของการเรียงตัวของค่าพีเอฟ บทที่ 6 กล่าวถึงขั้นตอนการทดลอง บทที่ 7 อธิบาย ผลการทดลองและการวิเคราะห์ผลการทดลอง บทที่ 8 บทสรุปผลการวิจัยและข้อเสนอแนะ

บทที่ 2

ทฤษฎีและวรรณกรรมที่เกี่ยวข้อง

ในบทนี้จะกล่าวถึงทฤษฎีที่เกี่ยวข้อง ได้แก่ ทฤษฎีที่เกี่ยวข้องกับภาษา ซึ่งได้อธิบายทฤษฎีที่เกี่ยวข้องกับภาษามนุษย์ ตระกูลภาษา ตาลักักษณะสำคัญเกี่ยวกับระบบเสียงในภาษา เป็นต้น ทฤษฎีทางสถิติที่ใช้ในการวิเคราะห์ข้อมูล เทคนิคการเรียนรู้ด้วยเครื่องจักร ได้แก่ แบบจำลองลาเทนท์ไดนามิค คอนดิชันนอลแรนดอมฟิลด์และซัพพอร์ตเวกเตอร์แมชชีน เป็นต้น นอกจากนี้บทนี้กล่าวถึงวรรณกรรมที่เกี่ยวข้องกับวิทยานิพนธ์นี้ ซึ่งได้แก่ วรรณกรรมที่เกี่ยวข้องกับการรู้จำภาษาตามแนวทางต่าง ๆ ได้แก่ วรรณกรรมความรู้จำภาษาแนวทางใช้คำลักษณะทางเสียง วรรณกรรมความรู้จำภาษาใช้กฎเกณฑ์การเรียงตัวหน่วยเสียง วรรณกรรมความรู้จำภาษาที่รวมจากหลายระบบ วรรณกรรมที่เกี่ยวข้องกับคำลักษณะเฉพาะค่าพีเอฟ และวรรณกรรมที่เกี่ยวข้องกับแบบจำลองคอนดิชันนอลแรนดอมฟิลด์ เป็นต้น

2.1 ทฤษฎีที่เกี่ยวข้องกับภาษา

2.1.1 ภาษามนุษย์

ภาษาคือเซตของสัญลักษณ์ที่มนุษย์ใช้ในการสื่อสาร ทุกภาษาประกอบด้วย ไวยากรณ์ ซึ่งอธิบายคุณสมบัติของสัญลักษณ์และโครงสร้างของความสัมพันธ์ระหว่างสัญลักษณ์ ตามหลักการวิชาที่ศึกษาเกี่ยวกับการแบ่งเชื้อชาติของมนุษย์ (Ethnology) [6] ซึ่งเป็นฐานข้อมูลที่รวบรวมการอธิบายภาษาพูดที่ใช้ทั่วโลก อธิบายว่ามีภาษาพูด 7000 ภาษาที่ใช้ทั่วโลกและมีภาษาจำนวน 330 ภาษาที่มีคนพูดมากกว่าล้านคน ภาษามนุษย์สามารถจำแนกเป็นแผนภูมิแสดงลำดับของตระกูล (Family tree) นักภาษาศาสตร์ได้แบ่งแยกระหว่างภาษาหลักและภาษาท้องถิ่น (Dialect) โดยกล่าวว่า ภาษาท้องถิ่นเป็นภาษาที่สามารถเข้าใจได้ง่าย ไม่เหมือนกับภาษาที่มีความแตกต่างระหว่างกันชัดเจน สังคมและการเมืองสามารถใช้เป็นปัจจัยในการแบ่งแยกระหว่างภาษาหลักและภาษาท้องถิ่นได้ ในทางปฏิบัติ การแบ่งแยกไม่สามารถบอกได้ชัดเจนเพราะเงื่อนไขการบอกความแตกต่างขึ้นอยู่กับประสบการณ์ของผู้พูด ตัวอย่างเช่น ภาษาเยอรมันและภาษาดัตช์ถูกพิจารณาว่าไม่ใช่ภาษาเดียวกัน เนื่องจากใช้ในประเทศที่แตกต่างกัน นั่นคือภาษาเยอรมันนิยมใช้พูดในประเทศเยอรมัน ออสเตรีย และในประเทศสวิตซ์เซอร์แลนด์บางส่วน ขณะที่ภาษาดัตช์เป็นภาษาพูดในประเทศเนเธอร์แลนด์ ถึงแม้ว่าจะมีหน่วยเสียงที่ใช้ร่วมกัน แต่มีบางหน่วยเสียงที่ออกเสียงแตกต่างกัน นอกจากนี้ระบบการเขียนทำให้เกิดความสับสนระหว่างภาษาท้องถิ่นและภาษาหลัก ตัวอย่างเช่น ภาษาฮินดี (Hindi) ใช้ระบบการเขียนแบบ Devanagari และภาษาเออร์ดู (Urdu) ใช้ระบบการเขียนแบบ Perso-Arabic script จึงถูกระบุว่าเป็นภาษาที่แตกต่างกัน [7] โดยภาษามนุษย์มีคุณสมบัติร่วมกัน คือ ทุกภาษาจะมีการแบ่งแยกระหว่างสระและพยัญชนะ ในภาษาส่วนใหญ่มีการแบ่งระหว่างคำนามและคำกริยา อีกทั้งยังมีกฎเกณฑ์ในการเรียงหน่วยเสียง ของทุกภาษาซึ่งมีพยางค์ลงท้ายด้วยสระ แต่บางพยางค์ลงท้ายด้วยพยัญชนะ ในขณะที่เดียวกันทุกภาษามีพยางค์ที่ขึ้นต้นด้วยพยัญชนะแต่ไม่จำเป็นต้องไม่มีพยางค์ที่ขึ้นต้น

แม้ภาษามีส่วนที่เหมือนกันแต่มีคุณสมบัติบางอย่างที่แตกต่างกัน เช่น ในทุกภาษา มีส่วนประธานและส่วนกริยา แต่ตำแหน่งของกริยาที่อาจจะแตกต่างกัน ในภาษาญี่ปุ่นวางกริยาที่ตำแหน่งท้ายประโยค ขณะที่ภาษาอังกฤษ มีตำแหน่งกริยาระหว่างประธานและกรรม [8] เป็นต้น

2.1.2 ตระกูลภาษา

ตามหลักการศึกษาที่ศึกษาเกี่ยวกับการแบ่งเชื้อชาติของมนุษย์ได้สรุปว่า ภาษาสามารถแบ่งออกเป็นกลุ่ม ๆ และแต่ละกลุ่มสัมพันธ์กับอีกกลุ่ม ซึ่งเรียกว่า ตระกูลภาษา (Language family) โดยแบ่งเป็น 136 ตระกูลภาษา ในแต่ละตระกูลภาษาที่มีการสืบทอดมาจากภาษาบรรพบุรุษ (Ancestor language) เดียวกันมีคุณสมบัติทางการศึกษาเกี่ยวกับระบบเสียงในภาษา (Ehonology) คำศัพท์ และไวยากรณ์ที่คล้ายกับ [9] ปัจจัยที่เป็นตัวแปรในการแบ่งตระกูลภาษาที่แตกต่างกัน เช่น จำนวนของผู้พูด ที่ตั้งของภาษา ภาษาท้องถิ่น และความรู้ทางภาษาศาสตร์ เป็นต้น ซึ่งมีตระกูลภาษา 6 ตระกูลที่มีคนพูดจำนวนมาก ได้แก่ ตระกูลภาษาแอฟโฟร-เอเชียติก (Afro-Asiatic) ตระกูลภาษาออสเตรเลเซียน (Australasian) ตระกูลภาษาอินโด-ยูโรเปียน (Indo-European) ตระกูลภาษาไนเจอร์-คองโก (Niger-Congo) ตระกูลภาษาจีน-ทิเบต (Sino-Tibetan) และตระกูลภาษาทราน-นิวไวกินี (Trans-New Guinea) ดังแสดงในภาพที่ 2-1 [10] เป็นแผนภูมิรูปร่างกลมแสดงเป็นร้อยละตามผู้พูดของภาษามนุษย์ทั่วโลก

อย่างไรก็ตาม บางภาษาที่รวมกลุ่มอยู่ในตระกูลภาษาเดียวกัน มีบางคุณสมบัติที่ปรากฏร่วมกันระหว่างภาษาภายในกลุ่ม ดังนั้น ในการศึกษาเรื่องของความแตกต่างของโครงสร้างภาษา จำเป็นต้องมีข้อมูลที่ใช้ในการแยกระหว่างภาษาได้ เรียกว่า คุณสมบัติทางภาษาศาสตร์ (linguistic) ซึ่งจากคุณสมบัติทางภาษาศาสตร์ที่สามารถบอกความแตกต่างระหว่างภาษาภายในกลุ่มได้นี้ ผู้วิจัยจึงมีความสนใจในการใช้คุณสมบัติทางภาษาศาสตร์ดังกล่าวในการศึกษาการรู้จำภาษา

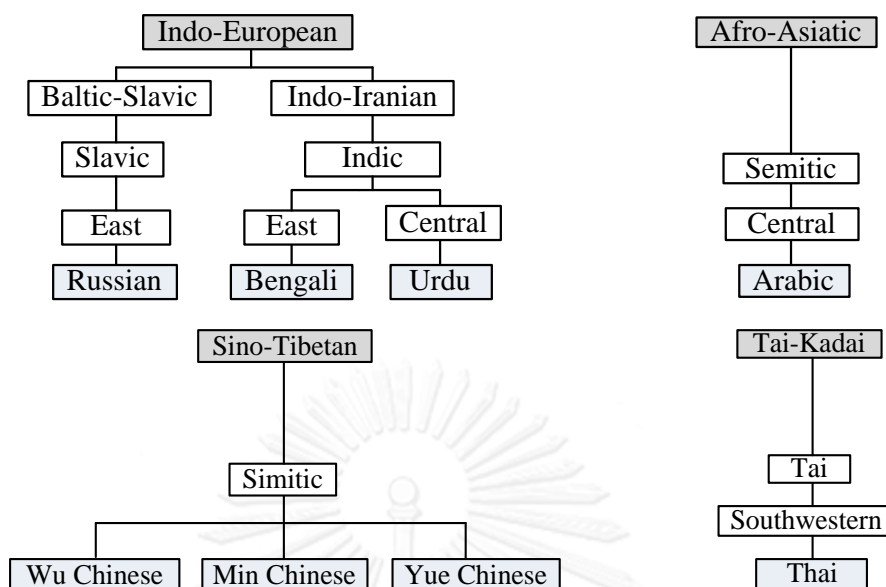
จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ภาพที่ 2-1 แผนภูมิรูปร่างกลมแสดงเป็นร้อยละตามผู้พูดของภาษามนุษย์ทั่วโลก

ในวิทยานิพนธ์นี้ใช้ฐานข้อมูลเสียงส่วนพัฒนา NIST LRE 2007 ในการทดลองการรู้จำภาษา ซึ่งอธิบายในบทที่ 3 โดยภาษาของเสียงพูดในฐานข้อมูลเสียงนี้อยู่ในตระกูลภาษาแอฟโฟร-เอเชียติก ตระกูลภาษาอินโด-ยูโรเปียน ตระกูลภาษาจีน-ทิเบต และตระกูลภาษาไท-กะไดซึ่งเคยเป็นส่วนหนึ่งของตระกูลภาษาจีน-ทิเบตมาก่อน

2.1.3 ตระกูลภาษาที่ศึกษาในวิทยานิพนธ์

ตระกูลของภาษาอธิบายความสัมพันธ์ระหว่างภาษาบรรพบุรุษและภาษาที่สืบทอดตระกูลของภาษา ภาษาของเสียงพูดที่ใช้ในวิทยานิพนธ์นี้มาจากฐานข้อมูลเสียงส่วนพัฒนา NIST LRE 2007 มี 8 ภาษา ซึ่งสามารถบอกความสัมพันธ์ของภาษากับตระกูลภาษา โดยตระกูลภาษาประกอบด้วย 4 ตระกูลภาษาหลัก แสดงดังภาพที่ 2-2 ซึ่งสามารถอธิบายความสัมพันธ์ของแต่ละภาษาและตระกูลภาษาได้ดังนี้ 1) ภาษาอารบิก เป็นภาษาที่อยู่ภายใต้ Semitic โดยเป็นสมาชิกของตระกูลภาษาตระกูลภาษาแอฟโฟร-เอเชียติก ภาษาอารบิกมีการยืมคำมาจากภาษาเบงกาลีและเออร์ดู 2) ภาษารัสเซีย เป็นภาษาที่สืบทอดมาจาก Slavic ซึ่งเป็นกลุ่มภาษาที่สืบทอดมาจาก ตระกูลภาษาอินโด-ยูโรเปียน ซึ่งตระกูลภาษานี้มีสาขาที่เป็นกลุ่มภาษา Indo-aryan ด้วยซึ่งภาษาภายใต้กลุ่มภาษานี้คือ 3) ภาษาเบงกาลี และ 4) ภาษาเออร์ดู ถึงแม้ภาษาเบงกาลีอยู่ภายใต้ภาษาอินดิก (Indic) เหมือนกับภาษาเออร์ดู แต่สองภาษานี้แตกต่างกันเพราะภาษาเบงกาลี เป็นกลุ่มภาษาแถบตะวันออก (East) และภาษาเออร์ดูเป็นกลุ่มภาษาแถบกลาง (Central) เมื่อพิจารณาระดับตระกูลภาษาพบว่า ภาษารัสเซีย เบงกาลี และเออร์ดู อยู่ภายใต้ตระกูลภาษาอินโด-ยูโรเปียน แต่มีระดับกลุ่มภาษาสืบทอดต่างกัน นอกจากนี้มีหลักฐานอื่นที่แสดงว่า คำศัพท์ในภาษาราชการของภาษาเออร์ดู ยืมมาจากภาษาอารบิก ซึ่งอาจจะเป็นหลักฐานที่ทำให้บอกความแตกต่างระหว่าง 2 ภาษานี้ได้ 5) ภาษาไทย เป็นภาษาที่อยู่ในกลุ่มภาษาไท (Tai) อยู่ในตระกูลภาษาไท-กะได ซึ่งเดิมภาษาไทย เคยอยู่ภายใต้ตระกูลภาษาจีน-ทิเบตและตระกูลภาษาจีน-ทิเบต มีกลุ่มภาษาย่อยคือ Sinitic มีภาษาอยู่ภายใต้กลุ่มภาษานี้คือ 6) ภาษาฮก หรือภาษาจีนกวางตุ้ง 7) ภาษาจีน และ 8) ภาษาฮก โดย คนที่พูดภาษาฮก อยู่ในฮ่องกง มาเก๊าและทางใต้ของจีน เช่น จังหวัด Guangdong ส่วนภาษาจีนจีน ใช้ในทางใต้ของ Fujian ซึ่งอยู่ติดกับไต้หวันและเอเชียตะวันออกเฉียงใต้ ภาษาฮก เป็นภาษาพูดในเซี่ยงไฮ้และทางใต้ของ Jiangsu ซึ่งมีการศึกษาทางประวัติศาสตร์พบว่า คนที่พูดภาษาฮกย้ายถิ่นฐานไปพร้อมกับคนที่พูดภาษาจีน เป็นต้น



ภาพที่ 2-2 แผนผังตระกูลภาษา ของภาษาของเสียงพูดที่ใช้ในวิทยานิพนธ์

2.1.4 ลักษณะเฉพาะของภาษา

ข้อมูลลักษณะเฉพาะของภาษา (Language characteristic) สามารถพิจารณาได้จากหลายค่า ภาพที่ 2-3 แสดงข้อมูลลักษณะเฉพาะของภาษา ซึ่งประกอบด้วย ข้อมูลลักษณะทางเสียง (Acoustic) ฉันทลักษณ์ (Prosody) ข้อมูลทางภาษาศาสตร์ ซึ่งข้อมูลแสดงเป็นลำดับชั้น จากภาพแสดงให้เห็นว่าโครงสร้างของประโยคสามารถแบ่งข้อมูลเป็นระดับย่อยได้ จากเสียงพูดหนึ่งข้อความ แสดงระดับวากยสัมพันธ์ (Syntax) แสดงระดับของคำ (Word) โดยในแต่ละคำแสดงข้อมูลฉันทลักษณ์ ในแต่ละคำสามารถแบ่งย่อยเป็นหน่วยเสียง (Phone) ในแต่ละหน่วยเสียงสามารถพิจารณาเป็นข้อมูลลักษณะทางเสียงและข้อมูลคำลักษณะสำคัญเกี่ยวกับระบบเสียงในภาษา (Phonological Feature) ซึ่งอยู่ในระดับเฟรม เป็นต้น จากข้อมูลแสดงลักษณะเฉพาะทางภาษาเหล่านี้มีประโยชน์ต่อการบ่งบอกความแตกต่างระหว่างภาษาและสามารถนำไปใช้ในการรู้จำภาษาต่อไป

Level	utterance																																	
Utterance																																		
Syntax	subject 1									verb 1									subject 2									verb 1						
Word	word 1									word 2									word 3									word 4						
Prosody	~									~									~									~						
Phone	seg1			seg2			seg1			seg2			seg3			seg1			2			seg1												
Acoustic, Phonological feature frame	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34

ภาพที่ 2-3 แผนผังข้อมูลลักษณะเฉพาะของภาษา

การรู้จำภาษาใช้ข้อมูลลักษณะสำคัญหลายค่า ซึ่งข้อมูลแต่ละค่าแสดงลักษณะเฉพาะทางภาษาที่แตกต่างกัน:

- 1) ข้อมูลลักษณะทางเสียง: หน่วยเสียงและค่าลักษณะสำคัญทางเสียง รวมทั้งการปรากฏหน่วยเสียงในแต่ละภาษานั้นมีความแตกต่างกัน ทำให้สามารถบอกความแตกต่างระหว่างภาษาได้ ไม่มีภาษาใดที่มีหน่วยเสียงปรากฏร่วมกัน เหมือนกันทั้งหมด [11] อย่างไรก็ตามถึงแม้ในบางภาษามีหน่วยเสียงที่ใช้ร่วมกันกับภาษาอื่น แต่ความถี่ของการปรากฏของหน่วยเสียงในแต่ละภาษามีค่าแตกต่างกัน
- 2) ข้อมูลการเรียงตัวของหน่วยเสียง (Phonotactic): กฎเกณฑ์การประกอบของเสียงมีความแตกต่างกันระหว่างภาษาตัวอย่างเช่น ภาษาไทยมีพยัญชนะประสมสองตัวปรากฏในพยางค์เดียวกัน ขณะที่ภาษาอังกฤษยอมมีการปรากฏพยัญชนะประสมสามตัว แต่ในภาษาจีนไม่มีพยัญชนะประสม นอกจากนี้เสียงพยัญชนะประสมบางตัว ปรากฏในภาษาหนึ่งบ่อย แต่หน่วยเสียงประสมนั้นอาจจะไม่ปรากฏในอีกภาษาหนึ่ง ดังนั้นความหลากหลายของกฎเกณฑ์การเรียงตัวของหน่วยเสียงจึงแตกต่างกันในแต่ละภาษา ส่งผลให้การกระจายของหน่วยเสียง หรือการกระจายของการปรากฏร่วมกันของหน่วยเสียงมีความแตกต่างกันในแต่ละภาษา
- 3) คำศัพท์ (Vocabulary): คำศัพท์เป็นหน่วยที่เล็กที่มีความหมายและเป็นข้อมูลที่สำคัญในการบ่งบอกลักษณะเฉพาะของภาษา เช่น ในภาษาอังกฤษ แม้ว่าส่วนเสียงนั้นถูกพูดด้วยผู้พูดที่ไม่ใช่เจ้าของภาษา จึงทำให้มีฉันทลักษณ์หรือสำเนียงของการพูดไม่เหมือนเจ้าของภาษา แต่ส่วนเสียงนั้นจะถูกรับรู้เป็นภาษาอังกฤษ เนื่องจาก ส่วนเสียงนั้นประกอบด้วยคำศัพท์ที่อยู่ในภาษาอังกฤษ อย่างไรก็ตาม มีงานวิจัยที่ศึกษาลักษณะภาษา โดยวิเคราะห์โครงสร้างของคำที่สามารถสร้างคำมาจากไวยากรณ์ได้ [12] โดยอธิบายว่าภาษาสร้างมาจากคำ ซึ่งสามารถสร้างได้หลายทาง นั่นคือ คำใหม่สามารถสร้างมาจากกลุ่มไวยากรณ์ที่ต่างกันไป เช่นสร้างคำนามจากคำกริยา ส่งผลให้เกิดความหลากหลายของคำในแต่ละภาษา จึงนำมาบ่งบอกความแตกต่างระหว่างภาษาทางอ้อมได้
- 4) วากยสัมพันธ์: ประโยชน์ของรูปแบบประโยคต่อการบ่งบอกลักษณะเฉพาะของภาษา คือ แม้ในหลายภาษามีคำที่ปรากฏร่วมกัน แต่ตำแหน่งของการปรากฏในแต่ละภาษานั้นแตกต่างกัน หลักกฎเกณฑ์ของการวางคำใด ๆ อยู่ตำแหน่งก่อนหน้าหรือหลังของอีกคำหนึ่ง ส่งผลให้เกิดลักษณะเฉพาะของแต่ละภาษาที่แตกต่างกัน เช่น ภาษาอังกฤษมีโครงสร้างประโยคเป็นแบบ SVO (Subject-verb-object) คือประธาน กริยา และกรรม ซึ่งแตกต่างกับภาษาญี่ปุ่นมีโครงสร้างแบบ SOV (Subject-object-verb) คือประธาน กรรม และกริยา [13] เป็นต้น
- 5) ฉันทลักษณ์: เนื่องจากความรู้ของลักษณะทางเสียงที่เกี่ยวกับภาษาอาจจะไม่เพียงพอสำหรับการระบุภาษา เช่น เมื่อในแต่ละภาษามีคำปรากฏเหมือนกัน แต่เมื่อมีการเน้นเสียงและใช้สำเนียงที่ใช้แตกต่างกัน ซึ่งหมายถึงฉันทลักษณ์สามารถนำไปบอกความแตกต่างระหว่างลักษณะเฉพาะของภาษา (ทางอ้อม) นอกจากนี้ฉันทลักษณ์สามารถ

นำไปบอกความแตกต่างในกลุ่มภาษาที่มีเสียงสูงต่ำ (Tonal language) ซึ่งภาษาที่มีการผันวรรณยุกต์ต่างกันส่งผลให้มีฉันทลักษณ์ที่แตกต่างไปจากกลุ่มภาษาที่ไม่มีการออกเสียงสูงต่ำ เป็นต้น นอกจากนี้ในภาษาที่มีการเน้นเสียง (Stress language) การเน้นเสียงในพยางค์ที่ต่างกันไป (เช่น ภาษาฮังการีมีการเน้นเสียงที่พยางค์เริ่มต้น แต่ในภาษาฝรั่งเศสและภาษาตุรกี มีการเน้นเสียงที่พยางค์ท้ายของเสียงพูด [13]) ซึ่งการเน้นเสียงในพยางค์ที่ต่างกันไปให้มีรูปแบบของฉันทลักษณ์ที่ต่างกันในแต่ละภาษา จึงนำไปใช้ในการระบุภาษาได้

นอกจากนี้ปัจจัยอื่นที่ทำให้ภาษามีความแตกต่างกัน [14, 15] ได้แก่ รูปแบบการพูด ภาษาหลักของผู้พูด ช่องสัญญาณ (Channel) ของสัญญาณเสียงพูด ซึ่งปัจจัยเหล่านี้แม้ไม่สามารถนำมาแบ่งแยกความแตกต่างของลักษณะเฉพาะของภาษาได้โดยตรง แต่มีผลกระทบต่อขั้นตอนการสกัดค่าลักษณะสำคัญในการรู้จำภาษาได้

2.1.5 ค่าลักษณะสำคัญเกี่ยวกับระบบเสียงในภาษา

ในทางภาษาศาสตร์ ค่าลักษณะสำคัญเกี่ยวกับระบบเสียงในภาษา (Phonological features: PF) ซึ่งในวิทยานิพนธ์นี้เรียกว่า ค่าพีเอฟ แสดงพารามิเตอร์การออกเสียงของหน่วยเสียง (Articulation) ค่าพีเอฟใช้วิเคราะห์หน่วยเสียงและอธิบายระบบภาษา ค่าพีเอฟอธิบายท่าทางกลไกการสร้างเสียง ด้วยค่าไบนารี (Binary) ค่าบวก (Positive) แสดงว่ามีลักษณะสำคัญ ค่าลบแสดงว่าไม่มีลักษณะสำคัญนั้นในการออกเสียง ค่าพีเอฟแต่ละตัวนิยามมาจากองค์ประกอบต่าง ๆ คือค่าลักษณะทางเสียง พารามิเตอร์การออกเสียง และนิยามมาจากความสัมพันธ์กันระหว่างค่าลักษณะทางเสียง หรือค่าการออกเสียงที่ต่างกัน [16]

ทฤษฎีการนิยามค่าพีเอฟมีมากมายแตกต่างกัน โดยพิจารณาตามการออกเสียง อะคูสติก ค่าเกี่ยวกับระบบเสียงในภาษา และความสัมพันธ์ระหว่างปัจจัยที่ใช้ในการพิจารณา ซึ่งนิยามที่เป็นที่รู้จักและใช้เป็นมาตรฐานในการอธิบายระบบเสียงในภาษา ได้แก่ ระบบเสียงของภาษาอังกฤษ (Sound Pattern of English: SPE) [17], (Government Phonology: GP) [18] (Multi Valued: MV) และค่าลักษณะสำคัญแบบไฮบริด (Hybrid Features: HF) [19, 20].

ผู้วิจัยเลือกใช้การอธิบายระบบเสียง SPE ซึ่งในวิทยานิพนธ์นี้เรียกว่า แบบเอสพีอี เสนอโดย Chomsky และ Halle [17] อธิบายความสัมพันธ์ระหว่างหน่วยเสียงในภาษาอังกฤษกับลักษณะสำคัญเฉพาะที่ใช้เป็นกลไกการออกเสียงกับหน่วยเสียงโดยแสดงเป็นค่าไบนารี ความสัมพันธ์ระหว่างหน่วยเสียงและลักษณะเฉพาะไม่มีกฎการจับคู่ที่ซ้ำซ้อนกัน การอธิบายระบบเสียงเอสพีอีเป็นที่รู้จักและใช้กันอย่างแพร่หลาย แต่ละหน่วยเสียงสามารถแสดงด้วยค่าพีเอฟ 14 ค่า แสดงดังตารางที่ 2-1 ได้แก่: Vocalic, Consonantal, Nasal, Continuant, Coronal, Anterior, Round, Voice, Strident, Tense, Back, High, Low, และ Silence ค่าพีเอฟ 14 ค่าสามารถพิจารณาเป็น 6 กลุ่มตามค่าลักษณะเฉพาะที่ใช้ในการออกเสียงคือ

- 1) ค่าลักษณะสำคัญกลุ่มหลัก (Major class) ที่ใช้อธิบายการปิดกั้นของกระแสลม (Airstream) ได้แก่ Vocalic Consonantal และ Nasal

- 2) กลุ่มค่าลักษณะสำคัญที่อธิบายการเปล่งเสียง (Manner of articulation) ที่อธิบายการปิดกั้นของช่องอากาศ ได้แก่ Continuant
- 3) กลุ่มค่าลักษณะสำคัญที่อธิบายฐานกรณ์ (Place of articulation) ได้แก่ Coronal Anterior และ Round
- 4) กลุ่มค่าลักษณะสำคัญที่อธิบายแหล่งของเสียง (Source) ได้แก่ Voice Strident Tense
- 5) กลุ่มค่าลักษณะสำคัญที่อธิบายความเป็นเกี่ยวกับตำแหน่งของลิ้น ได้แก่ Back High และ Low ซึ่งมีความสำคัญต่อเสียงสระ
- 6) กลุ่มอื่นได้แก่ เสียงเงียบ *silence*

ตารางที่ 2-2 แสดงตัวอย่างค่าพีเอฟตามนิยาม SPE กับหน่วยเสียงของคำว่า “คุณ” (“khun”) ในภาษาไทย ซึ่งสามารถเขียนในรูปการออกเสียงคือ [k^hun] แสดงในสัญลักษณ์สัทอักษรสากล (International Phonetic Alphabet: IPA) คำนี้สามารถแสดงเป็นชุดลำดับหน่วยเสียงจากการประกอบกันของเสียง /k^h/ /u/ และ /n/ ซึ่งแต่ละเสียงแสดงด้วยเซตของค่าลักษณะเฉพาะของค่าพีเอฟ โดยเซตของค่าพีเอฟได้มาจากกฎการจับคู่ค่าพีเอฟกับหน่วยเสียงด้วย หรือระบบตรวจหาค่าลักษณะเฉพาะของค่าพีเอฟ อธิบายในหัวข้อที่ 5.2

ตารางที่ 2-1 ตารางแสดงความหมายของค่าพีเอฟตามนิยาม SPE

ค่าพีเอฟ	คำอธิบาย
± Vocalic	[+ Vocalic] ค่าบวกแทนเสียงที่ไม่มีการปิดกั้นในช่องทางเดินเสียง แสดงสระที่กลุ่มเสียงสระ (vowel) และเสียงเหลว (liquid) คำนี้แสดงระดับการไม่มีการปิดกั้นระหว่างฐานกรณ์ (articulator)
± Consonantal	[+ Consonantal] แทนเสียงที่สร้างจากการปิดของช่องทางเดินเสียงซึ่งเป็นผลจากการปิดกั้นในกระแสลม (airflow) เสียงสระ และเสียงเลื่อน (glide) มีค่าเป็น [- Consonantal] ซึ่งบ่งบอกว่าไม่มีการปิดกั้นของช่องทางเดินเสียง
± Nasal	[+ Nasal] ค่าบวกแทนเสียงที่ปล่อยให้อากาศผ่านออกทางจมูกได้
± Continuant	เสียงเสียดแทรก (fricative) มีค่า [+ Continuant] ขณะที่เสียงกักมีค่าเป็น [- Continuant]
± Coronal	เสียงจากปุ่มเหงือก (alveolar) และเสียงจากเพดานแข็ง (Palatal) [+ Coronal] สร้างด้วยการปิดกั้นของลิ้นหรือปลายลิ้น เสียงข้างลิ้น (Labial) และเสียงจากเพดานอ่อน (velar) เป็น [- Coronal]
± Anterior	เสียงจากปุ่มเหงือกและเสียงข้างลิ้น เป็นเป็นเสียงที่ออกเสียงด้วยส่วนหน้าของปาก มีค่า [+ Anterior] ส่วนค่าที่เป็น [- Anterior] ได้แก่เสียงจากปุ่มเหงือกและเพดานแข็ง เสียงจากเพดานแข็ง และเสียงจากเพดานอ่อน

ค่าพีเอฟ	คำอธิบาย
± Round	ค่า [+ Round] เป็นเสียงที่สร้างด้วยริมฝีปากที่ม้วน
± Voice	ค่าที่สร้างจากการสั่นของเส้นเสียง (vocal cords) มีค่าเป็น [+ Voice] ขณะที่ค่าเป็น [-Voice] แสดงว่าเสียงนั้นไม่มีการสั่นเป็นรายคาบ
± Strident	เสียงเสียดแทรกและเสียงกักเสียดแทรก (affricates) มีค่าเป็น [+ Strident] เสียงที่สร้างจากการปิดกั้นที่ซับซ้อนบังคับให้กระแสลมด้านพื้นผิวสองด้าน. การสร้างเสียงรบกวนเสียดแทรกที่มีความเข้มสูง ส่วนค่าที่เป็นลบ แสดงเสียงที่สร้างโดยไม่มีการปิดกั้น [-Strident]
± Tense	เสียงสระสร้างจากตำแหน่งของลิ้นที่มีระดับการปิดกั้นมาก โดยฐานกรณใช้กล้ามเนื้อที่ตึงเครียด (Tense) และมากกว่าเสียงสระแบบคลาย [+Tense] ขณะที่เสียงที่เป็นสระแบบคลาย (Lax) มีค่าเป็น [- Tense] ซึ่งโดยปกติเสียง lax ได้แก่ สระเสียงสั้น.
± Back	เสียงสระที่มีตำแหน่งลิ้นในตำแหน่งกลางและหลัง จะมีค่าเป็น [+Back] ขณะที่เสียงสระที่ตำแหน่งลิ้นในตำแหน่งข้างหน้ามีค่าเป็น [-Back]
± High	เสียงสระที่ตำแหน่งลิ้นในตำแหน่งบนมีค่าเป็น [+ High] ขณะที่สระที่ตำแหน่งลิ้นในตำแหน่งกลางและล่างมีค่าเป็น [- High]
± Low	เสียงสระที่ตำแหน่งลิ้นในตำแหน่งกลางและล่างมีค่าเป็น [+Low] ขณะที่เสียงสระที่ตำแหน่งลิ้นในตำแหน่งบนมีค่าเป็น [-Low]
± Silence	เสียงเงียบหรือเสียงรบกวนมีค่าเป็น [+ Silence] ขณะที่เสียงที่เกิดจากการสร้างกลไกการออกเสียงของมนุษย์เป็นค่า [-Silence]

ตารางที่ 2-2 ตารางแสดงตัวอย่างค่าพีเอฟของคำว่า “khun” ในภาษาไทย

IPA	Vocalic	Consonantal	Nasal	Continuant	Coronal	Anterior	Round	Voice	Strident	Tense	Back	High	Low	Silence
[kʰ]	-	+	-	-	-	-	-	-	-	-	+	+	-	-
[u]	+	-	-	+	-	-	+	+	-	+	+	+	-	-
[n]	-	+	+	-	-	+	-	+	-	-	-	-	-	-

2.2 ทฤษฎีทางสถิติที่ใช้ในการวิเคราะห์ข้อมูล

2.2.1 การวิเคราะห์ความแปรปรวน

การวิเคราะห์ความแปรปรวน (Analysis of Variance: ANOVA) เป็นการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรเชิงกลุ่มเพื่อทดสอบความแตกต่างของข้อมูลที่ได้รับปัจจัยที่ต่างระดับกัน ในวิทยานิพนธ์นี้ใช้การวิเคราะห์ความแปรปรวนเพื่อพิจารณาความแปรปรวนในค่าพีเอฟที่ปรากฏในแต่ละภาษา เพื่อเปรียบเทียบความแตกต่างระหว่างค่าเฉลี่ยในค่าพีเอฟระหว่างสองกลุ่ม (นั่นคือระหว่าง กลุ่มภาษาที่ต้องการพิจารณาและกลุ่มที่ไม่ใช่ภาษาที่ต้องการพิจารณา) ซึ่งใช้การวิเคราะห์จากแผนภาพกล่องและตรวจสอบเงื่อนไขว่า ความแตกต่างระหว่างค่าเฉลี่ยระหว่างกลุ่ม ซึ่งประเมินความแตกต่างกันระหว่างกลุ่มว่าแตกต่างกันมากน้อยเท่าไร ซึ่งการตรวจสอบเงื่อนไขการพิจารณาสมมติฐาน โดยที่ μ_k คือ ค่าเฉลี่ยของข้อมูลกลุ่มที่ k (เมื่อ k=1 คือกลุ่มที่เป็นภาษาที่ต้องการพิจารณาและ k=2 คือกลุ่มที่ไม่ใช่ภาษาที่ต้องการพิจารณา)

สมมติฐานที่ $H_0: \mu_1 = \mu_2$ (หมายถึง ค่าเฉลี่ยของกลุ่มภาษาที่ต้องการพิจารณา เท่ากับค่าเฉลี่ยของกลุ่มไม่ใช่ภาษาที่ต้องการพิจารณา)

สมมติฐานที่ $H_1: \mu_1 \neq \mu_2$ (หมายถึง ค่าเฉลี่ยของกลุ่มภาษาที่ต้องการพิจารณา ไม่เท่ากับค่าเฉลี่ยของกลุ่มไม่ใช่ภาษาที่ต้องการพิจารณา)

ถ้าผลการทดสอบสรุปว่า H_0 จริง แสดงว่า ค่าแปรปรวนไม่ต่างกัน นั่นคือค่าเฉลี่ยของข้อมูลทั้งสองกลุ่มไม่แตกต่างกัน ในทางตรงข้ามถ้าผลการทดสอบสรุปว่า H_1 จริง แสดงว่า ค่าแปรปรวนต่างกัน นั่นคือค่าเฉลี่ยของข้อมูลทั้งสองกลุ่มแตกต่างกัน ข้อมูลนี้สามารถใช้อำนาจความแตกต่างระหว่างกลุ่มที่เป็นภาษาที่ต้องการพิจารณาและไม่ใช้ภาษาที่ต้องการพิจารณา ซึ่งการพิจารณาในวิทยานิพนธ์นี้ต้องการหาข้อมูลที่มาเป็นค่าบ่งชี้เพื่อจำแนกกลุ่มภาษาที่ต้องการพิจารณาและกลุ่มที่ไม่ใช่ภาษาที่ต้องการพิจารณาออกจากกัน ดังนั้น ต้องปฏิเสธสมมติฐาน H_0 และยอมรับสมมติฐานที่ H_1

การพิจารณาสมมติฐานใช้ค่าสถิติ F ดังสมการ (2.1) ซึ่งเป็นค่าตรวจสอบ

$$F = \frac{\left[\frac{SS_{boundary}}{a-1} \right]}{\left[\frac{SS_{Error}}{N-a} \right]} \quad (2.1)$$

โดยที่ a คือจำนวนกลุ่ม (ในวิทยานิพนธ์นี้ a มีค่าเท่ากับ 2)

N คือจำนวนข้อมูลที่ใช้วิเคราะห์ทั้งหมด

$$SS_{Error} = \sum_{i=1}^a \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{ij})^2 \quad \text{ซึ่ง } SS_{Error} \text{ คือความแปรผันภายในกลุ่ม}$$

$$SS_{boundary} = \sum_{i=1}^a \sum_{j=1}^{n_i} (\bar{y}_i - \bar{y}_{ij})^2$$

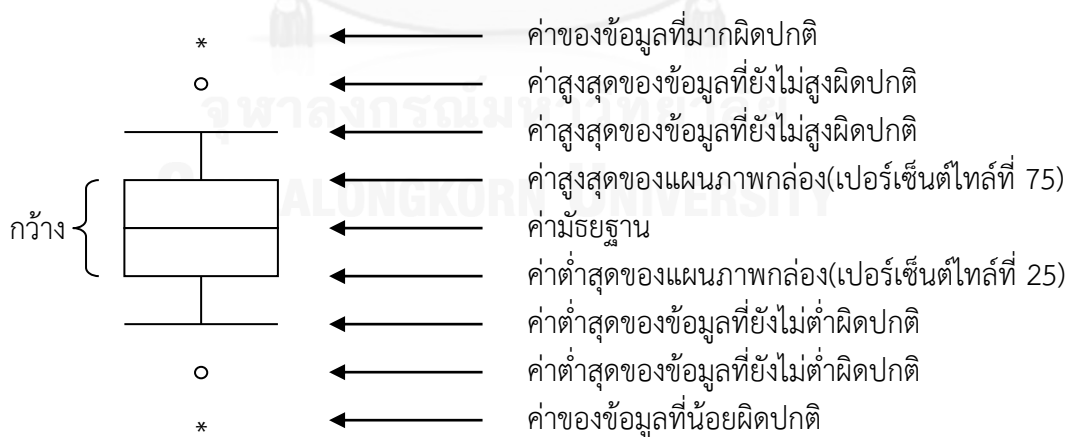
ซึ่ง $SS_{boundary}$ คือความแปรผันระหว่างกลุ่ม

เมื่อ \bar{y}_i คือค่าเฉลี่ยตัวอย่างชุดที่ i ค่า \bar{y}_{ij} คือเฉลี่ยของตัวอย่างทั้งหมด ค่า y_{ij} คือค่าตัวอย่างตัวที่ j ในกลุ่มที่ i และ n_i คือจำนวนข้อมูลของกลุ่มที่ i เมื่อได้ค่าสถิติ F จึงเปรียบเทียบกับ Sig (ค่าทางสถิติเมื่อระดับนัยสำคัญที่กำหนด α คือ 0.01)

2.2.2 การพิจารณาลักษณะสำคัญทางสถิติด้วยแผนภาพกล่อง

แผนภาพกล่อง (Box-and-whisker plot) เป็นแผนภาพที่แสดงรายละเอียดเกี่ยวกับการแจกแจงของข้อมูลเป็นค่าสถิติ เพื่อตรวจสอบการแจกแจงข้อมูล แสดงค่ากลางหรือค่ามัธยฐาน ค่าเปอร์เซ็นต์ไทล์ที่ 25 (Q_1), 75 (Q_3) และให้ค่าข้อมูลที่มีค่าผิดปกติ หมายถึง ค่าที่สูงมากหรือต่ำมาก (Outlier) จากค่ากลางความกว้างของแผนภาพกล่อง แสดง 50% ของข้อมูลอยู่ในแผนภาพกล่อง ดังภาพที่ 2-4

- 1) ค่าสูงสุดของข้อมูลที่ยังไม่สูงผิดปกติคือค่าสูงของข้อมูลชุดนั้น ๆ ที่มีค่าไม่เกิน $Q_3 + 1.5$ เท่าของความกว้างของแผนภาพกล่อง
- 2) ค่าต่ำสุดของข้อมูลที่ยังไม่ต่ำผิดปกติคือต่ำสุดของข้อมูลชุดนั้น ๆ ที่มีค่าไม่ต่ำกว่า $Q_1 - 1.5$ เท่าของความกว้างของแผนภาพกล่อง
- 3) ค่าสูงสุดของข้อมูลที่สูงกว่า $Q_3 + 3$ เท่าของความกว้างของแผนภาพกล่อง จึงเรียกว่า Extreme
- 4) ค่าต่ำสุดของข้อมูลที่น้อยกว่า $Q_1 - 3$ เท่าของความกว้างของแผนภาพกล่อง จึงเรียกว่า Extreme



ภาพที่ 2-4 ลักษณะของแผนภาพกล่อง

2.2.3 การวิเคราะห์ความแปรปรวนพหุคูณ

การวิเคราะห์ความแปรปรวนพหุคูณ (Multivariate Analysis of Variance: MANOVA) เป็นวิธีการวิเคราะห์ความแปรปรวนของข้อมูลระหว่างกลุ่มตั้งแต่ 2 กลุ่มขึ้นไป (ซึ่งเป็นตัวแปรอิสระ) และมีตัวแปรตามเป็นตัวแปรต่อเนื่องหลายตัวแปร หรือมีจำนวนตั้งแต่ 2 ตัวขึ้นไป ซึ่งการทดสอบความแตกต่างใช้การตรวจสอบค่าเฉลี่ย เพื่อวิเคราะห์ว่าเป็นความแตกต่างอันเนื่องมาจากตัวแปรอิสระ หรือเป็นความแตกต่างอันเนื่องมาจากความคลาดเคลื่อน

ซึ่งในวิทยานิพนธ์นี้การวิเคราะห์ความแปรปรวนแบบพหุคูณถูกนำมาใช้เพื่อวิเคราะห์ภาพรวมของความแตกต่างระหว่างกลุ่มซึ่งมีข้อมูลที่ประกอบด้วยตัวแปรหลายค่า ว่ามีผลกระทบต่อความแปรปรวนของข้อมูลแต่ละกลุ่มอย่างไร โดยใช้การวิเคราะห์ความแปรปรวนพหุคูณทางเดียว (One – way MANOVA) ซึ่งเป็นการวิเคราะห์ความแปรปรวนพหุคูณแบบกรณีมีตัวแปรต้นตัวเดียว ซึ่งมีระดับมากกว่า 2 ระดับขึ้นไป ซึ่งค่าสถิติคำนวณจากอัตราส่วนของผลรวมของกำลังสองของสมมติฐานและผลรวมของกำลังสองของความคลาดเคลื่อน ให้ H แทนเมทริกซ์ผลรวมของกำลังสองและการคูณกันของสมมติฐาน และ E แทนเมทริกซ์ผลรวมของกำลังสองและการคูณกันของความคลาดเคลื่อน ดังนั้นค่าสถิติที่มีหลายตัวแปร แทนด้วย $A = HE^{-1}$

อย่างไรก็ตามการประเมินความมีนัยสำคัญของความแตกต่างระหว่างกลุ่มสามารถพิจารณาจากสถิติหลายวิธี เช่น Wilks's Lambda, Hotelling's Trace, Pillai's Trace, Roy's Largest Root ซึ่งวิธีที่เลือกใช้ในวิทยานิพนธ์นี้ ใช้สถิติ Wilk's Lambda ซึ่งเป็นสถิติที่ใช้กันทั่วไป และสถิติ Pillai's Trace ที่มีความแข็งแกร่งในการวิเคราะห์ความแปรปรวนกรณีที่มีจำนวนกลุ่มตัวอยางในแต่ละกลุ่มแตกต่างกันมาก โดยพิจารณาที่ค่า Sig ของสถิติ Wilk's Lambda เปรียบเทียบระดับนัยสำคัญที่กำหนดถ้าค่า Sig มีค่าน้อยกว่าระดับนัยสำคัญ แสดงว่า มีนัยสำคัญทางสถิติ นั่นคือ ค่าเฉลี่ยของตัวแปรตามเมื่อพิจารณาหลายตัวแปรรวมกันทุกตัว มีความแตกต่างกันตามระดับของตัวแปรอิสระ ในทางตรงกันข้ามถ้าค่า Sig มีค่ามากกว่าหรือเท่ากับระดับนัยสำคัญ แสดงว่า การทดสอบนั้นไม่มีนัยสำคัญทางสถิติ

เมื่อ q แทนจำนวนตัวแปรตาม และ λ เป็นค่าไอเกนสำหรับเมทริกซ์ $A = HE^{-1}$ ตัวแปรจำแนกประเภทแต่ละตัว สถิติ Pillai's trace มีค่าเท่ากับผลรวมของอัตราส่วนของความแปรปรวนอธิบายบนฟังก์ชันการจำแนกดังสมการ (2.2)

$$\text{Pillai's trace (V)} = \text{trace}[H(H + E)^{-1}] = \prod_{i=1}^q \frac{\lambda_i}{1 + \lambda_i} \quad (2.2)$$

สถิติ Wilks's Lambda มีค่าเท่ากับผลคูณของอัตราส่วนของความแปรปรวนคลาดเคลื่อนกับความแปรปรวนรวมสำหรับแต่ละตัวแปร สถิติแสดงดังสมการ (2.3)

$$\text{Wilk's lamda}(\Lambda) = \frac{|E|}{|H + E|} = \prod_{i=1}^q \frac{1}{1 + \lambda_i} \quad (2.3)$$

2.3 แบบจำลองลาเทนท์ไดนามิกคอนดิชันนอลแรนดอมฟิลด์ (LDCRF)

แบบจำลองทางสถิติคอนดิชันนอลแรนดอมฟิลด์ (CRF) เป็นแบบจำลองที่ยอมรับกันว่ามีประสิทธิภาพในหลายงานที่แก้ปัญหาที่คำนึงถึงลำดับของข้อมูล ซึ่งใช้การจำลองความน่าจะเป็นแบบมีเงื่อนไข (Conditional probability) $P(y|x)$ แทนการใช้ความน่าจะเป็นแบบร่วม (Joint probability) ระหว่างข้อมูลอินพุตกับฉลากของกลุ่มเป้าหมาย ซึ่งความน่าจะเป็นแบบร่วมใช้กันในแบบจำลองเจนเจอเรทีฟ (Generative) บางเทคนิคมีคุณสมบัติเป็นอิสระต่อกัน จึงไม่สามารถจับความสัมพันธ์ระหว่างคุณสมบัติต่าง ๆ ที่เกี่ยวข้องในข้อมูลได้ ขณะที่แบบจำลองที่เป็นดิสคริมิเนทีฟ (Discriminative) ใช้ค่าความน่าจะเป็นแบบมีเงื่อนไขหาความสัมพันธ์ระหว่างคุณสมบัติต่าง ๆ ของลำดับค่าที่สังเกตที่เป็นอินพุต (Observation sequence) และลำดับฉลากของข้อมูล (Label sequence)

แบบจำลองลาเทนท์ไดนามิกคอนดิชันนอลแรนดอมฟิลด์ (Latent-Dynamic Conditional Random Field: LDCRF) เป็นแบบจำลองคอนดิชันนอลแรนดอมฟิลด์ ที่หาความสัมพันธ์ของคุณสมบัติภายในกลุ่มเป้าหมายและระหว่างกลุ่มเป้าหมาย ที่เกี่ยวกับเซตของสถานะซ่อน (Hidden state) ในแต่ละกลุ่มเป้าหมายที่ระบุตามฉลาก ซึ่งสถานะซ่อนนี้จำลองโครงสร้างย่อยของรูปแบบของแต่ละกลุ่มเป้าหมายและให้ค่าโอกาสที่จะเป็นไปได้ของการรู้จำของทั้งหมด แต่ละสถานะซ่อนสามารถจัดการเหมือนกับแบบจำลองคอนดิชันนอลแรนดอมฟิลด์ ค่าโอกาสที่จะเป็นไปได้ของข้อมูลที่เข้ามาทั้งหมดคือ ผลรวมของค่าโอกาสที่จะเป็นไปได้ของแต่ละส่วนจากสถานะซ่อน ภาพที่ 2-5 แบบจำลองคอนดิชันนอลแรนดอมฟิลด์ (ทางซ้าย) และแบบจำลองลาเทนท์ไดนามิกคอนดิชันนอลแรนดอมฟิลด์ (ทางขวา) [21] เมื่อ x_i แสดงค่าที่สังเกตที่เป็นอินพุตเข้ามาลำดับที่ i เมื่อ h_i แสดงสถานะซ่อนของ x_i และ y_i คือฉลากกลุ่มเป้าหมาย ซึ่งในวิทยานิพนธ์นี้หมายถึงภาษาที่รู้จำ แทนฉลากของค่าที่สังเกตลำดับที่ x_i วงกลมสี่เหลี่ยมแทนตัวแปรที่สังเกตได้และวงกลมสี่เหลี่ยมแทนตัวแปรซ่อนหรือตัวแปรที่สังเกตไม่ได้



แบบจำลองคอนดิชันนอลแรนดอมฟิลด์

แบบจำลองลาเทนท์ไดนามิกคอนดิชันนอล
แรนดอมฟิลด์

ภาพที่ 2-5 แบบจำลองคอนดิชันนอลแรนดอมฟิลด์และแบบจำลองลาเทนท์ไดนามิก
คอนดิชันนอลแรนดอมฟิลด์

แบบจำลองลาเทนท์ไดนามิกคอนดิชันนอลแรนดอมฟิลด์เรียนรู้การจับความสัมพันธ์ระหว่างลำดับข้อมูลอินพุตที่เข้ามา $X = \{x_1, x_2, \dots, x_T\}$ และลำดับของฉลากกลุ่มเป้าหมาย

$Y = \{y_1, y_2, \dots, y_T\}$ โดยแต่ละ y_i คือฉลากกลุ่มเป้าหมายสำหรับข้อมูล x_i ที่เป็นข้อมูลเข้ามาลำดับที่ i

ลำดับข้อมูลอินพุตที่เข้ามาแต่ละเฟรม x_i แสดงเวกเตอร์ค่าลักษณะเฉพาะ $\phi(x_i) \in \mathcal{R}$ กำหนดให้เวกเตอร์ของตัวแปรสถานะซ่อน $h = \{h_1, h_2, \dots, h_n\}$ เป็นตัวแปรที่ใช้ในการจำลองโครงสร้างย่อยของอินพุต ตัวแปร h_i เป็นสมาชิกของเซตตัวแปรสถานะซ่อนทุกตัวซึ่งแทนด้วย H สำหรับแบบจำลองลาเทนท์ไดนามิกคอนดิชันนอลแรนดอมฟิลด์ ค่าโอกาสที่จะเป็นไปได้ของข้อมูลที่เข้ามาทั้งหมดของกลุ่มเป้าหมาย มีค่าเท่ากับ ความน่าจะเป็นเดี่ยวหรือความน่าจะเป็นของเหตุการณ์หนึ่งที่เกิดขึ้น (Marginal probability) ซึ่งหาได้จากผลรวมของความน่าจะเป็นเดี่ยวของสถานะซ่อนของซับเซต H_L แบบจำลองลาเทนท์คอนดิชันนอล คำนวณได้จากสมการ (2.4)

$$P(Y | X, \theta) = \sum_h p(Y | h, X, \theta) p(h | X, \theta) \quad (2.4)$$

เมื่อ X คือการต่อกันของทุกเวกเตอร์ค่าลักษณะเฉพาะ x_i สำหรับลำดับเสียงพูดทั้งหมด และให้ $\theta\{x_i\}$ แทนแบบจำลองที่ฝึกฝน

$$\text{การฝึกฝนแบบจำลองและการอนุมานฟังก์ชันศักยภาพ } P(Y | X, \theta) = \sum_{h: \forall h_i \in H_{y_i}} p(h | X, \theta)$$

(Inference efficient) แบบจำลองถูกจำกัดให้มีเซตที่ไม่เชื่อมกันของสถานะซ่อนที่สอดคล้องกับฉลากกลุ่มเป้าหมาย แต่ละตัวแปร h_i เป็นสมาชิกของเซตตัวแปรสถานะซ่อนที่เป็นไปได้ H_{y_i} สำหรับฉลากกลุ่มเป้าหมายเป้าหมาย y_i เมื่อ H เป็นเซตของสถานะซ่อนที่เป็นไปได้ทั้งหมด ซึ่งเป็นยูเนียน (Union) ของทุกเซต H_{y_i} โดยลำดับของ $h_i \notin H_{y_i}$ มีค่าเป็น 0 เมื่อ $P(Y | h, X, \theta) = 0$ หรือมีค่าเป็น 1 ในกรณีอื่น โดยแบบจำลองสามารถคำนวณได้ดังสมการที่ (2.5)

$$P(Y | X, \theta) = \sum_{h: \forall h_i \in H_{y_i}} p(h | X, \theta) \quad (2.5)$$

เมื่อ $p(h | X, \theta)$ แทนแบบจำลองทางสถิติคอนดิชันนอลแรนดอมฟิลด์ส ซึ่งนิยามด้วยสมการที่ (2.6)

$$p(h | X, \theta) = \frac{1}{Z(X, \theta)} \exp\left(\sum_k \theta_k \cdot F_k(h, X)\right) \quad (2.6)$$

เมื่อ Z คือ การปรับข้อมูลให้เข้ากับบรรทัดฐาน หรือ เรียกอีกอย่างหนึ่งว่าฟังก์ชันพาร์ทิชัน ซึ่งคำนวณได้จากสมการ

$$Z(X, \theta) = \sum_h \exp\left(\sum_k \theta_k \cdot F_k(h, X)\right) \quad (2.7)$$

เวกเตอร์ค่าลักษณะเฉพาะ F_k สามารถนิยามได้สมการที่ (2.8)

$$F_k(h, X) = \sum_{i=1}^T f_k(h_{i-1}, h_i, X, i) \quad (2.8)$$

เมื่อ $f_k(h_{i-1}, h_i, X, i)$ แทนฟังก์ชันคุณสมบัติ (Feature function) หรือ $t_k(h_{i-1}, h_i, X, i)$ แทนฟังก์ชันการเปลี่ยนสถานะ (Transition function) ฟังก์ชันคุณสมบัติ s_k ขึ้นกับตัวแปรซ่อนเดี่ยวในแบบจำลอง ขณะที่ฟังก์ชันการเปลี่ยนสถานะ t_k ขึ้นกับคู่ของตัวแปรซ่อน

ฟังก์ชันคุณสมบัติ ไม่ได้จำลองการขึ้นต่อกันกับลำดับข้อมูลที่เข้ามาทั้งหมด แต่ขึ้นกับซับเซตของลำดับข้อมูลที่เข้ามาก่อนหน้าตัวแปรที่กำลังพิจารณา หรือกล่าวได้อีกอย่างหนึ่งว่า $s_k(h_i, X, i) = s_k(h_i, \tilde{X}_i, i)$ เมื่อ $\tilde{X}_i = [X_{i-m}, \dots, X_i]$ แทนกรอบหน้าต่างพิจารณาที่มีขนาดกรอบ $m+1$ ที่อยู่ก่อนหน้าตัวแปรที่มีข้อมูลที่พิจารณาในเฟรมที่ i ฟังก์ชันคุณสมบัติ $s_k(h_i, X, i)$ คือข้อมูลที่เข้ามาใน \tilde{X}_i ซึ่งนำค่าของสถานะซ่อนที่มีความสัมพันธ์ที่อยู่ติดกัน

กำหนดให้เซตของข้อมูลฝึกฝนประกอบด้วย ลำดับฉลากข้อมูล (X_i, Y_i) เมื่อ i มีค่าตั้งแต่ 1 ถึง n เมื่อฟังก์ชันวัตถุประสงค์ (Objective function) นิยามเหมือนกันกับแบบจำลองคอนดิชันนอลแรนดอมฟิลด์ในงาน [22, 23] ซึ่งนิยามด้วยสมการ (2.9):

$$L(\theta) = \sum_{i=1}^n \log p(Y_i | X_i, \theta) - \frac{1}{2\sigma^2} \|\theta\|^2 \quad (2.9)$$

จากสมการที่ (2.9) พจน์แรกแทนค่าโอกาสที่จะเป็นไปได้แบบล็อกที่มีเงื่อนไขการขึ้นต่อกัน (Conditional log likelihood) ของข้อมูลฝึกฝนและพจน์ที่สองแทนค่าความน่าจะเป็นล่วงหน้าแบบล็อกของเกาส์เซียนที่มีความแปรปรวน σ^2 นั่นคือ $P(\theta) \sim \exp(1/2\sigma^2 \|\theta\|^2)$

ในขั้นตอนการอนุมาน กำหนดให้ X เป็นลำดับของเสียงที่ต้องการทดสอบ ประมาณฉลากหรือคำตอบที่มีความน่าจะเป็นมากที่สุดของลำดับข้อมูล Y^* ซึ่งทำให้แบบจำลองคอนดิชันนอลมีค่ามากที่สุด ดังสมการที่ (2.10) เมื่อพารามิเตอร์ของแบบจำลองที่เหมาะสมที่สุด (Optimal model) แทนด้วย θ^* ได้มาจากตัวอย่างในข้อมูลฝึกฝน

$$Y^* = \arg \max_Y p(Y_i | X_i, \theta^*) \quad (2.10)$$

สมมติให้ฉลากแต่ละกลุ่มสอดคล้องกับ เซตที่ไม่เชื่อมกันของสถานะซ่อน จึงสามารถเขียนสมการที่ (2.10) ในรูปแบบของตัวแปรสถานะซ่อนได้ดังสมการที่ (2.11)

$$Y^* = \arg \max_Y \sum_{h: \forall h_i \in H_{Y_i}} p(h | X, \theta^*) \quad (2.11)$$

ในแบบจำลองลาเทนท์ไดนามิคคอนดิชันนอลแรนดอมฟิลด์ การวัดค่าโอกาสที่เป็นไปได้สำหรับภาษา l เท่ากับ ความน่าจะเป็นเดี่ยว $p(Y_i = l | X, \theta^*)$ ความน่าจะเป็นนี้มีค่าเท่ากับผลรวมของความน่าจะเป็นเดี่ยวของสถานะซ่อนของซับเซต H ดังสมการที่ (2.12) ซึ่งค่าความน่าจะเป็นเดี่ยวได้มาจากการประมาณโดยใช้อัลกอริทึมการส่งผ่านความเชื่อ (Belief propagation):

$$p(Y_i = l | X, \theta^*) = \sum_{h: \forall h_i \in H_i} p(h | X, \theta^*) \quad (2.12)$$

เมื่อ X คือการต่อกันของทุกเวกเตอร์ค่าลักษณะเฉพาะ x_i สำหรับลำดับเสียงพูดทั้งหมด กำหนดให้ลำดับข้อมูลที่เข้ามาแทนด้วย X ซึ่งได้มาจากคุณลักษณะสำคัญทางเสียงนี้ ให้ θ^* แทนพารามิเตอร์ของแบบจำลองที่ได้จากการฝึกฝน ซึ่งในวิทยานิพนธ์นี้ใช้ค่าพีเอฟและค่าลักษณะสำคัญทางเสียงต่าง ๆ เช่น สัมประสิทธิ์ MFCC สัมประสิทธิ์ PLP ซึ่งอธิบายการทดลองอยู่ในบทที่ 6

2.4 ซัพพอร์ตเวกเตอร์แมชชีน

ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine: SVM) เป็นเครื่องมือที่ใช้ในงานการจำแนกรูปแบบ โดยหลักการของซัพพอร์ตเวกเตอร์แมชชีนจะสร้างระนาบการแบ่งข้อมูลออกเป็นกลุ่ม ๆ ฟังก์ชันที่ใช้ในการจำแนกจะสร้างปริภูมิสมมติฐาน (Hypothesis space) เพื่อแบ่งข้อมูล โดยพิจารณาข้อมูลที่เข้ามาที่เรียกว่าค่าสังเกต (observation) ซึ่งแสดงเป็นเวกเตอร์อินพุต x โดยที่ $x \in R^n$ และผลลากกลุ่มของข้อมูลที่เข้ามาซึ่งแทนด้วย y โดยที่ $y \in \{\pm 1\}$ ดังนั้นในแต่ละค่าสังเกตประกอบด้วยคู่ของเวกเตอร์ (x, y) ดังนั้นฟังก์ชันการจำแนกแทนด้วย $h: R^n \rightarrow \{\pm 1\}$ นั่นคือฟังก์ชันการจำแนกแทนด้วย $h(x) = y$

การทำงานของซัพพอร์ตเวกเตอร์แมชชีนซึ่งเป็นตัวจำแนกแบบไบนารี สามารถนิยามได้ดังนี้:

$$f(x) = \sum_{i=1}^N \alpha_i K(x, x_i) + b, \quad (2.13)$$

ให้ x เป็นอินพุตเวกเตอร์และ x_i เป็นเวกเตอร์ซัพพอร์ต โดยที่ $\alpha_i > 0$ และ $K(x, x_i)$ เป็นเคอร์เนลฟังก์ชัน (Kernel function) ที่วัดค่าความเหมือน (Similarity) ระหว่าง x และ x_i

โดย $p(t_i | d_j)$ แทนค่าความน่าจะเป็นของสมาชิก t_i ที่เป็นองค์ประกอบในเวกเตอร์ลักษณะสำคัญของตัวอย่าง d_j ดังนั้นสามารถซูปเปอร์เวกเตอร์ในรูปของเวกเตอร์สเปซได้ดังนี้

$$\Phi(d_j) = [p(t_1 | d_j) \quad p(t_2 | d_j) \quad \dots \quad p(t_M | d_j)] \quad (2.14)$$

เมื่อ M เป็นจำนวนสมาชิกของเวกเตอร์ลักษณะสำคัญ การสร้างเคอร์เนลฟังก์ชันเพื่อวัดความเหมือนระหว่างสองเวกเตอร์ของตัวอย่างระหว่าง d_1 และ d_2 ซึ่งนิยามด้วยการทำอินเนอร์โปรดักต์ (Inner product) ดังสมการ (2.15):

$$K(d_1, d_2) = \langle \Phi(d_1), \Phi(d_2) \rangle = \sum_{i=1}^M p(t_i | d_1) p(t_i | d_2) \quad (2.15)$$

เนื่องจากประสิทธิภาพของการจำแนกขึ้นอยู่กับเคอร์เนลฟังก์ชัน ดังนั้นการเตรียมข้อมูลของตัวอย่างกลุ่มบวกและลบที่นำมาใช้ในการสร้างเคอร์เนลฟังก์ชัน และการเลือกชนิดของเคอร์เนลฟังก์ชันที่เหมาะสมมีความสำคัญต่อความสามารถของการจำแนก

วิทยานิพนธ์นี้ได้นำซัพพอร์ตเวกเตอร์แมชชีนมาใช้ในการจำแนกภาษา โดยใช้ซัพพอร์ตเวกเตอร์แมชชีนเป็นตัวจำแนกแบบไบนารี (Binary classifier) นั่นคือ ให้ทำนายคำตอบภายใน 2

กลุ่มเป้าหมายนั้นคือ 1) พิจารณาคำตอบระหว่างกลุ่มเป้าหมายและกลุ่มที่ไม่ใช่เป้าหมายจากฟังก์ชันของการจำแนกเป็นแบบหนึ่งต่อหนึ่ง (One-Versus-One) ซึ่งฝึกฝนตัวจำแนกจากข้อมูลที่เป็นภาษาเป้าหมายหนึ่งภาษาให้เป็นที่กลุ่มตัวอย่างบวกและข้อมูลที่ไม่ใช่ภาษาเป้าหมายหนึ่งภาษาให้เป็นที่กลุ่มตัวอย่างลบ ซึ่งนำมาในการทดลองสำหรับระบบการรู้จำภาษาพูดตามแนวทางการเรียงของค่าพีเอฟที่นำเสนอในบทที่ 5 หรือ 2) พิจารณาคำตอบระหว่างกลุ่มเป้าหมายและกลุ่มที่ไม่ใช่กลุ่มเป้าหมายจากฟังก์ชันของการจำแนกเป็นแบบหนึ่งต่อทั้งหมด (One-Versus-All) ซึ่งฝึกฝนตัวจำแนกจากข้อมูลที่เป็นภาษาเป้าหมายหนึ่งภาษาให้เป็นที่กลุ่มตัวอย่างบวกและข้อมูลที่ไม่ใช่ภาษาเป้าหมายทุกภาษาให้เป็นที่กลุ่มตัวอย่างลบ ซึ่งนำมาในการทดลองสำหรับระบบการรู้จำภาษาพูดตามแนวทางการเรียงหน่วยเสียงที่นำเสนอในบทที่ 4 เป็นต้น

2.5 วรรณกรรมที่เกี่ยวข้อง

เนื่องจากในเทคโนโลยีทางด้านเสียงพูดมีความก้าวหน้าในระดับที่ต้องการการรู้จำเสียงของเสียงพูดหลายภาษา จึงทำให้การรู้จำภาษาของเสียงพูดได้รับความสนใจ ดังนั้นผู้วิจัยจึงได้ทำการศึกษาวรรณกรรมที่เกี่ยวกับการรู้จำภาษาและวรรณกรรมที่เกี่ยวกับเทคนิคการจำแนกต่าง ๆ นอกจากนี้ผู้วิจัยได้ศึกษาวรรณกรรมที่ใช้คำลักษณะสำคัญที่เกี่ยวกับระบบการออกเสียง เพื่อนำมาเป็นคำลักษณะสำคัญที่ใช้ในการรู้จำภาษา ดังแสดงในหัวข้อถัดไป

2.5.1 วรรณกรรมที่พิจารณาการระบุภาษาด้วยมนุษย์

ระบบการระบุภาษาที่มีความแม่นยำมากที่สุดคือ มนุษย์ [24] เพราะมนุษย์มีความสามารถในการรู้จำเสียงที่ได้ยินได้ และสามารถระบุภาษาได้ทันที มนุษย์ตัดสินใจว่าเสียงพูดที่ได้ยินเป็นภาษาใดโดยวัดเทียบกับประสบการณ์ที่คุ้นเคยมาก่อนและมนุษย์สามารถจับคู่ภาษาที่ไม่เคยรู้จักมาก่อนเข้ากับความรู้ก่อนหน้าได้ นอกจากนี้มีงานวิจัยที่ศึกษาการรับรู้ของมนุษย์ในการระบุภาษา [11] โดยใช้การประกอบกันของหน่วยเสียง หรือการหาตำแหน่งของคำ (Word spotting) และในงานวิจัย [25] ศึกษาว่ามนุษย์ที่มีความคุ้นเคยในภาษามากกว่าหนึ่งภาษามีความสามารถในการระบุภาษาได้ดีกว่าและมีประสิทธิภาพดีขึ้น มีงานวิจัย [26] ที่ศึกษาผลกระทบของความยาวของเสียงพูดต่อการรับรู้ของมนุษย์ และได้เปรียบเทียบประสิทธิภาพของการระบุภาษาของมนุษย์และเครื่องคอมพิวเตอร์ที่ได้รับผลกระทบดังกล่าว จากผลการทดลอง พบว่าเครื่องคอมพิวเตอร์ที่ใช้ความรู้จากเกณฑ์การเรียงของหน่วยเสียง มีความสามารถในการระบุภาษาดีกว่ามนุษย์ เมื่อความยาวของเสียงพูดนั้นสั้น สามารถสรุปได้ว่ามนุษย์ใช้ความรู้ต่าง ๆ และใช้ข้อมูลที่ได้จากเสียงพูดที่มีความยาว ยาวขึ้นมาใช้ในการระบุภาษา

2.5.2 วรรณกรรมที่เกี่ยวข้องในการรู้จำภาษา

ในวรรณกรรม [27] รวบรวมงานวิจัยที่เกี่ยวกับการรู้จำภาษา พบว่า ระบบที่มีการรวบรวมข้อมูลจากหลายแหล่ง ทำให้ประสิทธิภาพการรู้จำภาษาดีขึ้น [28, 29] ซึ่งสามารถแบ่งงานวิจัยการรู้จำภาษาออกเป็น 2 แนวทางหลัก คือ 1) การรู้จำภาษาโดยใช้คำลักษณะทางเสียง (Acoustic

approach) [30, 31] และ 2) การรู้จำภาษาโดยใช้กฎเกณฑ์การเรียงตัวหน่วยเสียง (Phonotactic approach) [28] ซึ่งรายละเอียดของแต่ละแนวทางอธิบายในหัวข้อต่อไป

2.5.2.1 วรรณกรรมการรู้จำภาษาแนวทางใช้ค่าลักษณะทางเสียง

ระบบการรู้จำภาษาพูดแนวทางใช้ค่าลักษณะทางเสียง ใช้ค่าลักษณะสำคัญจากสัญญาณเสียง โดยตรงเป็นค่าเซปตรอล (Cepstral) โดยสกัดค่าลักษณะสำคัญสเปคตรอล (Spectral) จากสัญญาณเสียงพูด จากนั้นสร้างแบบจำลองการกระจายของค่าลักษณะสำคัญโดยตรงด้วยเทคนิคแบบจำลองทางสถิติต่าง ๆ ซึ่งระบบการรู้จำตามแนวทางนี้ไม่ต้องใช้ข้อมูลที่ได้จากการถอดความหน่วยเสียง (Phonetic transcription) ในขั้นตอนการรู้จำภาษานั้น ภาษาของเสียงพูดที่ทดสอบถูกตัดสินใจว่าเป็นภาษาใด จากการพิจารณาค่าความน่าจะเป็นของเสียงพูดเปรียบเทียบกับแต่ละแบบจำลองภาษาที่ได้จากการเรียนรู้ของข้อมูลฝึกฝน

วรรณกรรม [32] ในคศ. 1970 เริ่มมีงานวิจัยการรู้จำภาษาโดยใช้ค่าลักษณะสำคัญทางเสียง ระบบดังกล่าวใช้อัลกอริทึมการจับคู่ต้นแบบของสัญญาณสเปคตรอลขณะทำงาน [33] เสนอระบบการรู้จำภาษาพูดที่ใช้ค่าลักษณะสำคัญทางเสียงโดยใช้แบบจำลองเกาส์เซียนแบบผสม (Gaussian Mixture Models: GMMs) สำหรับในแต่ละภาษา เวกเตอร์ค่าลักษณะสำคัญ (Feature vector) ถูกจำลองด้วยแบบจำลองเกาส์เซียนในขั้นตอนการรู้จำภาษา ซึ่งค่าลักษณะสำคัญทางเสียงที่ใช้ในงานการรู้จำภาษา ได้แก่ สัมประสิทธิ์เซปสตรัมบนสเกลเมล (Mel frequency cepstral coefficients: MFCC) และสัมประสิทธิ์การทำนายเชิงเส้นแบบรับรู้ (Perceptual Linear Prediction: PLP) ถูกนำมาใช้ในการสร้างเวกเตอร์เฉพาะค่าลักษณะสำคัญเพื่อเป็นอินพุตของแบบจำลองเกาส์เซียน [34, 35] ในงานวิจัย [31] เปรียบเทียบผลกระทบของการใช้ค่าลักษณะสำคัญทางเสียงในการรู้จำภาษา พบว่า ระบบการระบุภาษาที่ใช้สัมประสิทธิ์ PLP ในแบบจำลองเกาส์เซียนให้ประสิทธิภาพดีกว่าการใช้ค่าสัมประสิทธิ์ MFCC นอกจากนี้ยังมีงานวิจัยอื่นที่สนใจและนำเสนอค่าลักษณะสำคัญทางเสียงอื่น เพื่อเพิ่มประสิทธิภาพของระบบการระบุภาษา เช่น งานวิจัย [36] เสนอค่า (Revised Perceptual Linear Prediction coefficients: RPLP) ซึ่งเป็นการรวมค่าสัมประสิทธิ์ MFCC และ PLP ซึ่งทำให้ได้ประสิทธิภาพสูงขึ้น ในงานวิจัย [30] เสนอการใช้ (Shifted-Delta Cepstral coefficients: SDC) เป็นการคำนวณค่าการเปลี่ยนแปลงของค่าลักษณะสำคัญที่เกิดขึ้นในหลายเฟรม โดยระบบการรู้จำที่ใช้ค่าลักษณะสำคัญนี้มีประสิทธิภาพสูงอย่างมีนัยสำคัญทำให้ค่า SDC เป็นตัวแทนของค่าลักษณะสำคัญทางเสียงที่ได้รับความนิยม และใช้กันอย่างแพร่หลายในระบบการรู้จำภาษาพูด นอกจากนี้มีงานวิจัยที่เสนอเทคนิคการชดเชยค่าลักษณะสำคัญ (Feature compensation) [37, 38] เพื่อนำไปใช้แก้ปัญหาการไม่ตรงกันของเงื่อนไข (Mismatch condition) ที่เกิดจากความแปรปรวนของผู้พูด ช่องสัญญาณเสียง (Channel) และเสียงรบกวนจากสภาพแวดล้อม เช่น งานวิจัยที่เสนอการใช้ตัวกรอง RASTA (RelAtive SpecTrAl) [28] เพื่อลดผลกระทบที่เกิดจากความแปรปรวนในช่องสัญญาณเสียง ในงานวิจัย [39] เสนอการทานอร์มอลไลซ์ช่องทางเดินเสียง (Vocal Tract Length Normalization: VTLN) งานวิจัยที่เสนอการชดเชยในแบบจำลองด้วยวิธีซัพสเปซ (subspace) ที่เรียกว่าการวิเคราะห์ปัจจัยแฝง (Latent Factor Analysis: LFA) [40] และวิธี (Nuisance Attribute Projection: NAP) [39, 41] เพื่อนำความแปรปรวนที่มาจากแหล่งกำเนิดเสียง

ที่อยู่ในระดับความถี่ต่ำออกไป งานวิจัยแนวทางที่เสนอเทคนิคต่าง ๆ เพื่อลดความแปรปรวนจากช่องสัญญาณเสียง และลดความแปรปรวนจากผู้พูดเหล่านี้เป็นอีกแนวทางที่ได้รับการสนใจและยอมรับในการรู้จำภาษา

นอกจากการเพิ่มพารามิเตอร์ที่ใช้ในการประมาณค่าในแบบจำลองเกาส์เซียน ซึ่งเป็นแบบจำลองแบบเจเนอเรทีฟ (Generative model) และอีกแนวทางที่ทำให้ผลการรู้จำภาษาดีขึ้น คือ การเสนอใช้เทคนิคแบบจำลองดิสคริมีเนทีฟ (Discriminative model) แทนการใช้เทคนิคแบบจำลองแบบเจเนอเรทีฟแบบเดิม [34] ซึ่งงานวิจัยที่เสนอการใช้แบบจำลองดิสคริมีเนทีฟมาเพื่อรู้จำภาษาตามแบบอะคูสติก (Acoustic based) ให้ประสิทธิภาพในการรู้จำภาษาดีขึ้นอย่างเห็นได้ชัด เช่น การใช้ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine: SVM) ด้วยฟังก์ชันเคอร์เนลแบบโพลีโนเมียล (Polynomial kernel) การใช้ซัพพอร์ตเวกเตอร์แมชชีน ร่วมกับฟังก์ชันเคอร์เนลแบบการจำแนกเชิงเส้นของลำดับแบบเจเนอรัลไลซ์ (Generalized Linear Discriminant Sequence kernel: GLDS) [42, 43] การใช้สารสนเทศร่วมค่าสูงสุด (Maximum Mutual Information: MMI) ร่วมกับแบบจำลองเกาส์เซียนแบบผสม (GMM) ซึ่งเรียกว่า GMM-MMI [44, 45] การใช้เวกเตอร์ค่าลักษณะสำคัญที่ได้จากแบบจำลองเกาส์เซียนแบบผสม มาเป็นอินพุทของซัพพอร์ตเวกเตอร์แมชชีน ซึ่งเรียกว่า Hybrid SVM/GMM [46-49] เป็นต้น จากงานวิจัยข้างต้น พบว่าระบบการรู้จำภาษาพูดที่ใช้เทคนิคซัพพอร์ตเวกเตอร์แมชชีนเข้ามาช่วยร่วมด้วย ให้ความถูกต้องในการรู้จำดีกว่าใช้เพียงแบบจำลองเกาส์เซียนและส่งผลให้เป็นแนวทางการวิจัยของงานการรู้จำภาษาตามแนวอะคูสติกที่มีการนำแบบจำลองแบบดิสคริมีเนทีฟมาช่วยร่วมด้วย

ในปีคศ. 2007 เทคนิคการวิเคราะห์ตัวประกอบแบบร่วม (Joint Factor Analysis: JFA) [50] และเทคนิคไอเวกเตอร์ (i-vector) [51] ได้รับความสนใจในการนำมาใช้ในงานการยืนยันเสียงของผู้พูด (Speaker verification) จึงถูกนำมาใช้ในการรู้จำภาษา ซึ่งระบบการรู้จำภาษาพูดที่ใช้เทคนิคดังกล่าวประสบความสำเร็จและมีประสิทธิภาพการรู้จำภาษาที่ดีขึ้น ระบบการรู้จำภาษาพูดตามแนวทางการใช้ค่าลักษณะสำคัญทางเสียงร่วมกับเทคนิคไอเวกเตอร์กลายเป็นที่ระบบการรู้จำภาษาพูดอีกหนึ่งแนวทางที่ได้รับความนิยมร่วมกันในวรรณกรรมการรู้จำภาษาว่ามีประสิทธิภาพที่ดีที่สุด [50, 52]

อย่างไรก็ตามงานวิจัยที่เกี่ยวข้องของการรู้จำภาษาตามแนวทางอะคูสติกที่ผ่านมายังไม่ได้ให้ความสำคัญในการพิจารณาความเป็นลำดับของข้อมูล (Sequential data) ที่อยู่บนค่าลักษณะสำคัญ และนอกจากนี้งานวิจัยที่มีการใช้แบบจำลองแบบดิสคริมีเนทีฟ ต้องใช้ขั้นตอนการแปลงข้อมูล (Feature Transformation) เพื่อให้เป็นข้อมูลที่ใช้เป็นอินพุทของแบบจำลองดิสคริมีเนทีฟ ซึ่งในวิทยานิพนธ์นี้จึงเสนอแนวทางที่เพื่อลดปัญหาในประเด็นในการพิจารณาลำดับของข้อมูล โดยไม่ต้องแปลงค่าลักษณะสำคัญให้อยู่ในสเปซอื่นก่อนที่จะมาใช้ในการจำลอง

2.5.2.2 วรรณกรรมการรู้จำภาษาใช้กฎเกณฑ์การเรียงตัวหน่วยเสียง

วรรณกรรมทางด้านกรรรู้จำภาษาแนวทางการใช้เกณฑ์การเรียงตัวหน่วยเสียง สามารถแบ่งได้เป็น 2 แนวทางหลักคือ วรรณกรรมที่เกี่ยวข้องกับการรู้จำภาษาตามแนวทางการเรียงตัวหน่วยเสียง และวรรณกรรมที่เกี่ยวข้องกับการรู้จำเสียง

1) วรรณกรรมที่เกี่ยวข้องกับการรู้จำภาษาตามแนวทางการเรียงตัวหน่วยเสียง

งานวิจัยการรู้จำภาษาตามแนวทางการใช้เกณฑ์การเรียงตัวหน่วยเสียง เริ่มเมื่อปี 1977 โดย House และ Neuburg [53] ได้เสนอการรู้จำภาษา โดยใช้หน่วยเสียงที่เป็นผลลัพธ์จากตัวรู้จำเสียงมาสร้างแบบจำลองภาษา ต่อมา Hazen [54] เสนอการระบุภาษาโดยอาศัยเซกเมนต์ (Segment) และเสนอกรอบการระบุภาษาอิงตามความน่าจะเป็น นอกจากนี้งานวิจัยของ Hazen เสนอการใช้ตัวรู้จำหน่วยเสียงแบบไม่ขึ้นกับภาษา (Independent phone recognizer) ซึ่งต่างจากงาน [53] งานวิจัยที่เสนอระบบการรู้จำภาษาพูดตามแนวทางการเรียงตัวหน่วยเสียงที่ใช้ตัวรู้จำเสียงตามด้วยแบบจำลองภาษา (Phone Recognizer followed by Language Modeling: PRLM) [55] พบว่ามีประสิทธิภาพการรู้จำภาษาสูงและได้รับการยอมรับและเป็นกรอบการสร้างระบบการรู้จำภาษาพูด ในงานวิจัยนี้ใช้ระบบรู้จำหน่วยเสียงระบบเดียวเพื่อถอดหน่วยเสียงในทุกภาษาและใช้การสร้างแบบจำลองภาษาแยกสำหรับแต่ละภาษาเป้าหมายที่ต้องการรู้จำ แต่แนวทางนี้มีข้อเสียคือ การเลือกตัวรู้จำภาษาที่ไม่เหมาะสมทำให้ไม่ครอบคลุมหน่วยเสียงที่ปรากฏในบางภาษาได้ ต่อมา Zissman [29, 55] นำเสนอระบบการรู้จำภาษาพูดตามแนวทางการเรียงตัวหน่วยเสียงที่ใช้ตัวรู้จำเสียงแบบขนานตามด้วยแบบจำลองภาษา (Parallel Phone Recognizer followed by Language Modeling: PPRLM) งานนี้ใช้ตัวรู้จำหน่วยเสียงหลายตัวเพื่อสร้างหน่วยเสียงที่เป็นผลลัพธ์ที่ได้แตกต่างกันจากตัวรู้จำเสียงหลายตัว แล้วนำผลลัพธ์จากแต่ละตัวรู้จำเสียงมาสร้างแบบจำลองภาษาดังนั้นในขั้นตอนการรู้จำภาษา ภาษาของเสียงพูดถูกพิจารณาจากผลรวมของคะแนนความเหมือนจากแบบจำลองแต่ละภาษาที่มีค่าความเหมือนสูงสุด ข้อดีของการใช้ตัวรู้จำเสียงแบบขนานคือการรู้จำเสียงหลายตัวทำให้ได้ผลลัพธ์การรู้จำเสียงที่มีชุดลำดับหน่วยเสียงที่แตกต่างกัน นำไปสู่การสร้างเกณฑ์การเรียงตัวหน่วยเสียงที่หลากหลายและแตกต่างกัน ส่งผลให้มีข้อมูลในการนำไปสร้างแบบจำลองภาษาเพิ่มขึ้นในแต่ละภาษา แต่แนวทางนี้มีข้อเสียคือ ต้องการข้อมูลในการสร้างตัวรู้จำเสียงหลายตัวเพิ่มขึ้นจากเดิมที่ต้องสร้างตัวรู้จำเสียงเพียงหนึ่งระบบเท่านั้น ต่อมาจึงมีงานวิจัยที่เสนอการใช้ตัวรู้จำเสียงแบบไม่ขึ้นกับภาษาเรียกว่า ตัวรู้จำเสียงแบบยูนิเวอร์ซอล (Universal phone recognizer) [56] ซึ่งเป็นอีกแนวทางที่เสนอการสร้างตัวรู้จำเสียงที่ครอบคลุมหน่วยเสียงในภาษาพูดหลายภาษาและ Ma Bin [57] เสนอการใช้ตัวรู้จำเสียงแบบยูนิเวอร์ซอลด้วยวิธีการสร้างแบบจำลองอะคูสติกจากเซกเมนต์เสียงซึ่งได้มาจากการรวมตัวแทนเสียงอะคูสติกจากหลายภาษา

งานวิจัยของ Ma Bin และงานวิจัยของ Gao, S. [58, 59] นำเสนอการพิจารณารูปแบบของสถิติของหน่วยเสียงจากแบบจำลองเอ็นแกรมแบบลำดับสูง (High order) ได้แก่ ยูนิแกรม ไบแกรม และไตรแกรมเพื่อหารูปแบบของเอ็นแกรมที่แตกต่างกันในแต่ละภาษา งานวิจัย [60] เสนอแนวทางการใช้ตรรกะเชิงแบบอ้างอิงความหมาย (Semantic indexing) เพื่อใช้จัดการเวกเตอร์ลักษณะสำคัญที่ขนาดมิติสูงเนื่องจากประกอบด้วยค่าสถิติแบบจำลองเอ็นแกรมจำนวนมาก Boon Pang [61] ใช้

เทคนิคการสกัดค่าลักษณะสำคัญแบบโกลบอล (Global) เพื่อแสดงค่าสถิติการปรากฏร่วมกันในภาพใหญ่ของเสียง ซึ่งต่อมาในงานวิจัย [62] ที่เสนอการรู้จำภาษาแนวทางการเรียงหน่วยเสียงโดยใช้แบบจำลองเวกเตอร์สเปซ (Vector space model: VSM) ซึ่งเรียกว่า ระบบการรู้จำภาษาพูดแบบใช้ระบบรู้จำเสียงตามด้วยการใช้แบบจำลองเวกเตอร์สเปซ หรือระบบการรู้จำภาษาพูดแบบพรีอาร์วีเอสเอ็ม (Phone recognizer followed by vector space modeling: PRVSM) โดยเวกเตอร์ที่ใช้แสดงค่าสถิติของแบบจำลองเอ็นแกรมของหน่วยเสียงซึ่งเรียกว่า Bag-of-sound [63] พบว่า ประสิทธิภาพของระบบการรู้จำภาษาพูดเมื่อใช้แบบจำลองเวกเตอร์สเปซดีขึ้นเมื่อเทียบกับระบบรู้จำที่ใช้แบบจำลองภาษาแบบเอ็นแกรม แต่ข้อเสียของแบบจำลองเวกเตอร์สเปซ คือเมื่อเวกเตอร์ลักษณะสำคัญมีจำนวนมากจะส่งผลให้การคำนวณค่าความเหมือนระหว่างเวกเตอร์มีผลลัพธ์ที่ไม่ดี เมื่อเพิ่มการพิจารณาแบบจำลองเอ็นแกรมที่ระดับสูงขึ้นทำให้ขนาดของจำนวนเอ็นแกรมที่พิจารณาเพิ่มขึ้นเป็นเอ็กโปเนนเชียล เพื่อลดจำนวนแบบจำลองเอ็นแกรมที่เพิ่มขึ้น Campbell เสนองานวิจัย [64, 65] เสนอการใช้เทคนิคการเลือกคำ (Keyword selection) ที่เลือกจากแบบจำลองเอ็นแกรมที่มีความสามารถในการบอกความแตกต่างได้ดี เป็นตัวแทนไปใช้ในการสร้างลักษณะสำคัญแทน

ในขณะทำงานวิจัย [66, 67] เสนอการใช้หน่วยเสียงที่เป็นระดับย่อยของคำ (sub-word) และเสนอการใช้หน่วยที่คล้ายพยางค์ (Syllable-like) [68] เพื่อเป็นตัวแทนของเสียงที่ได้จากตัวถอดรหัสเสียง

อย่างไรก็ตามผลลัพธ์ที่ได้จากตัวรู้จำเสียงอาจจะมีข้อผิดพลาดจากการถอดรหัสหน่วยเสียง Wang S [69] และ Yang [70] เสนอการใช้หลายสมมติฐานที่ได้จากตัวรู้จำเสียง เพื่อสร้างโอกาสของการปรากฏการเรียงหน่วยเสียงหลายรูปแบบและทำให้ผลการรู้จำภาษาที่มาจากการใช้ชุดลำดับหน่วยเสียงหลายแบบมีประสิทธิภาพดีขึ้น อีกวิธีที่สามารถเพิ่มประสิทธิภาพจากตัวรู้จำเสียงได้คือ การใช้ตัวแทนหน่วยเสียงทั้งหมดที่ได้จากตัวรู้จำเสียง ในงานวิจัยของ Gauvain [71] ได้นำเสนอการรู้จำภาษาตามแนวทางกฎเกณฑ์ของการเรียงหน่วยเสียงแบบใช้หน่วยเสียงแลตทิซแทนการใช้ลำดับหน่วยเสียง ซึ่งเป็นการเสนอแนวทางที่ทำให้ประสิทธิภาพของการรู้จำภาษาตามแนวทางกฎเกณฑ์ของการเรียงหน่วยเสียงดีขึ้นอย่างมีนัยสำคัญ เช่นเดียวกับงานวิจัย [72-75] ที่ใช้หน่วยเสียงแลตทิซในการรู้จำภาษาและมีประสิทธิภาพสูงกว่าการรู้จำภาษาที่ใช้หน่วยเสียง ในการประเมินผลการรู้จำภาษาของ NIST ซึ่งจัดในปี 2009 ระบบการรู้จำภาษาพูดที่นำเสนอโดย Campbell มีประสิทธิภาพการรู้จำภาษาดีกว่าระบบการรู้จำภาษาพูดที่ใช้แนวทางอื่น โดยระบบการรู้จำภาษาพูดนี้ใช้หน่วยเสียงแลตทิซจากตัวรู้จำเสียงและใช้ตัวจำแนกภาษาจากซัพพอร์ตเวกเตอร์แมชชีน (Lattice-SVM) [74] จึงทำให้ระบบการรู้จำภาษาพูดแนวทางการเรียงตัวของหน่วยเสียงแลตทิซเป็นที่ยอมรับ จากความสำเร็จของงานวิจัยที่เกี่ยวกับการรู้จำภาษาตามแนวทางกฎเกณฑ์ของการเรียงหน่วยเสียงจากหน่วยเสียงแลตทิซส่งผลให้เป็นแรงจูงใจในวิทยานิพนธ์นี้เพื่อการพัฒนาการรู้จำภาษาพูดตามแนวทางกฎเกณฑ์ของการเรียงหน่วยเสียงแบบใช้หน่วยเสียงแลตทิซ

2) วรรณกรรมที่เกี่ยวข้องกับการรู้จำเสียง

นอกจากแนวทางการพัฒนาแบบจำลอง อีกแนวทางหนึ่งที่สามารถเพิ่มประสิทธิภาพของการรู้จำภาษาได้คือการใช้ตัวรู้จำเสียงที่มีความแม่นยำสูง ซึ่งตัวรู้จำเสียงที่ใช้เทคนิคของฮิดเด็นมาคอฟ

ร่วมกับโครงข่ายประสาทเทียม (ANN/HMM) [76] นอกจากนี้ตัวรู้จำเสียงที่ใช้โครงข่ายประสาทเทียมแบบโครงสร้างหลายระดับ (Hierarchical structure of ANN) [77] เป็นที่สนใจและให้ผลการรู้จำที่แม่นยำและถูกพัฒนาเป็นตัวรู้จำเสียง BUT ซึ่งใช้เทคนิค ANN/HMM [76] เป็นตัวรู้จำเสียงที่ให้ผลการรู้จำเสียงดีและเป็นที่ยอมรับและถูกนำมาใช้ในระบบการรู้จำภาษาพูดในปี 2007 [37] ซึ่งข้อดีของตัวรู้จำเสียงแบบใช้โครงข่ายประสาทเทียมหลายชั้นคือ หน่วยเสียงที่อยู่รอบข้างมาใช้ในการคำนวณหน่วยเสียงที่กำลังพิจารณา ซึ่งในวิทยานิพนธ์นี้ใช้ตัวรู้จำเสียง BUT เพื่อใช้ถอดรหัสหน่วยเสียงในการสร้างระบบการรู้จำภาษาพูดตามแนวทางกฎเกณฑ์ของการเรียงหน่วยเสียงเช่นกัน

2.5.2.3 วรรณกรรมการรู้จำภาษาที่รวมจากหลายระบบ

เนื่องจากการรู้จำภาษาสามารถพิจารณาจากข้อมูลหลายทาง เช่น ค่าลักษณะทางเสียง (Acoustic) หรือ ข้อมูลการเรียงหน่วยเสียง ข้อมูลฉันทลักษณ์ ดังนั้นเพื่อให้ระบบการรู้จำภาษาพูดมีประสิทธิภาพสูง จึงมีงานวิจัยที่เสนอการรวมค่าลักษณะสำคัญต่าง ๆ หรือคะแนนจากระบบการรู้จำภาษาพูดที่ใช้ข้อมูลแตกต่างกัน ซึ่งกลายเป็นแนวทางของระบบการรู้จำภาษาพูดที่เสนอจากการรวมจากหลายระบบ เช่น Bo Yin [78] เสนอการใช้เวกเตอร์ค่าลักษณะสำคัญที่ได้จากค่าเซปโตรแกรมร่วมกับฉันทลักษณ์เพื่อรู้จำภาษา

ผลการทดลองของงานวิจัย [79, 80] เมื่อรวมระบบหลายระบบการรู้จำภาษาพูดย่อยส่งผลให้การระบุภาษามีความถูกต้องมากขึ้น Singer [29] เสนอการรวมระบบการรู้จำภาษาพูดแนวทางอะคูสติกที่ใช้แบบ GMM/SVM และระบบการรู้จำภาษาพูดแบบใช้ตัวรู้จำเสียงแบบขนาน ตามด้วยการใช้แบบจำลองเวกเตอร์สเปซ หรือพีอาร์วีเอสเอ็ม (PRVSM) โดยใช้ตัวจำแนกส่วนหลัง (Backend classifier) เพื่อตัดสินใจระบุภาษา พบว่าให้ผลการรู้จำมีความถูกต้องเพิ่มขึ้น ซึ่งเหมือนกับงานวิจัยของ Tong Rong [81] ที่เสนอการรวมระบบการรู้จำภาษาพูดจากแนวทางอะคูสติก แนวทางการเรียงตัวหน่วยเสียง (ซึ่งประกอบด้วย ระบบการรู้จำภาษาพูดที่ใช้แบบจำลองภาษาเอ็นแกรมและระบบการรู้จำภาษาพูดที่ใช้แบบจำลองเวกเตอร์สเปซ) และแนวทางการใช้ฉันทลักษณ์ พบว่าเมื่อรวมระบบการรู้จำจากระบบการรู้จำภาษาพูดย่อยแนวทางต่าง ๆ ทำให้ผลการรู้จำมีความถูกต้องเพิ่มขึ้น งานวิจัยของกลุ่มวิจัย ATVS ที่เสนอระบบการรู้จำภาษาพูดที่รวมหลายระบบย่อยเข้าด้วยกันใน ATVS ที่ใช้ในการประเมินผลการรู้จำภาษาของนิซในปี 2009 (NIST LRE 2009) [80] โดยใช้แบบจำลองการรวมแบบแองเคอร์ (Anchor model) [82] เพื่อรวมระบบการรู้จำภาษาพูดที่ใช้ค่าสเปคโตรล เข้ากับระบบการรู้จำภาษาพูดที่ใช้การเรียงตัวหน่วยเสียงแบบพีอาร์วีเอสเอ็ม (PRVSM) ซึ่งทำให้ผลการรู้จำภาษาดีขึ้น ขณะที่งานวิจัยของ Raymond [83] เสนอการรวมระบบการรู้จำภาษาพูดด้วยการใช้การถดถอยโลจิสติกเชิงเส้น (Linear logistic regression) เพื่อรวมระบบการรู้จำภาษาพูดที่ใช้ค่าฉันทลักษณ์และระบบการรู้จำภาษาพูดแนวทางการเรียงหน่วยเสียง ผลจากการรวมระบบการรู้จำภาษาพูดทำให้การระบุภาษาถูกต้องขึ้น

2.5.3 วรรณกรรมที่เกี่ยวข้องกับค่าลักษณะเฉพาะค่าพีเอฟ

ในวิทยานิพนธ์นี้เสนอการนำค่าลักษณะเฉพาะค่าพีเอฟมาใช้ในการรู้จำภาษา ดังนั้นผู้วิจัยจึงศึกษาวรรณกรรมที่เกี่ยวข้องกับค่าลักษณะเฉพาะ โดยแบ่งออกเป็น 2 แนวทางคือ งานวิจัยที่เสนอ

วิธีการสกัดค่าหรือตรวจหาค่าลักษณะเฉพาะ งานวิจัยที่นำเสนอการใช้ค่าลักษณะเฉพาะค่าพีเอฟมาใช้ในการรู้จำภาษา

เนื่องจากข้อจำกัดของการจับคู่ระหว่างการถอดหน่วยเสียงกับค่าพีเอฟ คือต้องการข้อมูลการถอดความของหน่วยเสียงเพื่อใช้แปลงเป็นค่าพีเอฟ จึงทำให้ไม่ยืดหยุ่นต่อการนำไปแปลงค่าลักษณะสำคัญกับภาษาใหม่ ๆ ดังนั้นจึงมีงานวิจัยที่เสนอวิธีการสกัดค่าพีเอฟจากสัญญาณเสียงพูด เช่น การใช้เทคนิคแบบจำลองเกาส์เซียนแบบผสม (GMMs) [84] การใช้โครงข่ายประสาทเทียมแบบหลายชั้น (Multilayer perceptron: MLP) [16] และ [85] งานวิจัยที่เสนอการใช้โครงข่ายประสาทเทียมป้อนกลับแบบหน่วงเวลา (Time-delay recurrent neural networks) [16] เทคนิคฮิดเดินมาคอฟ (Hidden Markov model: HMM) [86] และการใช้ซัพพอร์ตเวกเตอร์แมชชีน (SVM) [87-89]. ด้วยประสิทธิภาพของระบบตรวจหาค่าลักษณะสำคัญที่ใช้โครงข่ายประสาทเทียมแบบหลายชั้น ให้ผลการตรวจหาค่าลักษณะสำคัญได้ดีกว่าวิธีอื่น จึงทำให้ผู้วิจัยสนใจระบบตรวจหาค่าลักษณะสำคัญที่ใช้โครงข่ายประสาทเทียมแบบหลายชั้น นอกจากนี้มีงานวิจัยในการรู้จำภาษาที่เสนอการใช้โครงข่ายประสาทเทียมแบบหลายชั้นในงานการสกัดค่าลักษณะสำคัญ (feature extraction) [90] โดยผลทดลองของงานวิจัยแสดงว่าเมื่อนำโครงข่ายประสาทเทียมแบบหลายชั้นมาใช้ในการสกัดค่าให้ผลการรู้จำดีขึ้น

จากความสำคัญของค่าพีเอฟซึ่งมีผลต่อการบอกความแตกต่างระหว่างภาษา จึงทำให้มีงานวิจัยที่เสนอการใช้ความรู้จากลักษณะการออกเสียงหรือค่าพีเอฟมาช่วยปรับปรุงประสิทธิภาพของการรู้จำภาษา สามารถแบ่งเป็น 2 กลุ่มคือ งานวิจัยการรู้จำภาษาที่ใช้ค่าพีเอฟตามแนวทางการเรียงหน่วยเสียง เช่น งานวิจัย [91, 92] เสนอการใช้ค่าลักษณะสำคัญการออกเสียงเพื่อสร้างแบบจำลองภาษา โดยใช้ชุดลำดับของค่าพีเอฟ แทนชุดลำดับหน่วยเสียง Nunez และ Ramon [92] ใช้ลำดับเสียงแบบไตรแกรมของค่าการออกเสียงแบบ phono-articulatory และพบว่าผลการรู้จำดีกว่าการใช้ไตรแกรมหน่วยเสียง แต่ข้อเสียของงานวิจัยนี้คือ ต้องการข้อมูลการถอดความหน่วยเสียงของแต่ละภาษาในการแปลงหน่วยเสียงกับค่าลักษณะสำคัญ งานวิจัย [92, 93] ศึกษาความสัมพันธ์แบบไดนามิก (dynamic) ระหว่างค่าพีเอฟและค่าความน่าจะเป็นที่แสดงการเปลี่ยนแปลง Parandekar [93, 94] เสนอกรอบแบบความน่าจะเป็นเพื่อจำลองการระบุภาษาจากค่าลักษณะเฉพาะที่มีหลายค่า เป็นต้น

นอกจากนี้อีกกลุ่มได้แก่ งานวิจัยการรู้จำภาษาที่ใช้ค่าพีเอฟตามแนวทางอะคูสติก เช่น งานวิจัย [87] เสนอระบบการรู้จำภาษาพูดที่รวมค่าลักษณะสำคัญเฉพาะ (Distinctive features) ที่แสดงเป็นค่าแบบไบนารี ที่ใช้อธิบายการออกเสียงซึ่งมีลักษณะเหมือนกับค่าพีเอฟ รวมเข้ากับค่าลักษณะสำคัญเอสดีซี (SDC) ซึ่งใช้กันอย่างแพร่หลายในการรู้จำภาษาแนวทางอะคูสติก ผลการทดลองในงานวิจัยนี้แสดงว่า ผลการรู้จำภาษาดีขึ้นเมื่อมีการนำค่าลักษณะสำคัญเฉพาะมาใช้ร่วมด้วย นอกจากนี้มีงานวิจัยที่เสนอการใช้ค่าพีเอฟ ร่วมกับข้อมูลฉันทลักษณ์ [95] เพื่อระบุภาษาและบอกความแตกต่างของภาษาท้องถิ่นได้ [96] ซึ่งให้ผลการระบุภาษาถูกต้องมากขึ้น

2.5.4 วรรณกรรมที่เกี่ยวกับแบบจำลองคอนดิชันนอลแรนดอมฟิลด์

วรรณกรรมที่ใช้แบบจำลองคอนดิชันนอลแรนดอมฟิลด์ (Conditional Random Field: CRF) ในปัญหาการจำแนก ซึ่งทำงานได้ดีกับข้อมูลลำดับ แบบจำลองคอนดิชันนอลแรนดอมฟิลด์เป็นแบบจำลองดิสคริมิเนทีฟ ที่คำนวณจากการรวมเชิงเส้นแบบล็อก (Log linear combination) ของเหตุการณ์ที่เป็นไปได้ทั้งหมด จากค่าลักษณะสำคัญที่เป็นอินพุท ซึ่งเมื่อนำ CRF ไปใช้ในงานวิจัยด้านต่าง ๆ พบว่าให้ประสิทธิภาพที่ดีขึ้นเช่น งานวิจัยด้านรู้จำเสียงพูด [97] การรู้จำหน่วยเสียง [98] การรู้จำคำ (Word recognition) [99] การจำแนกหน่วยเสียง (Phone classification) [100] ซึ่งเมื่อใช้เทคนิคคอนดิชันนอลแรนดอมฟิลด์พบว่าให้ผลการรู้จำดีขึ้นเมื่อเปรียบเทียบกับการใช้ตัวจำแนกวิธีอื่น [101]

อย่างไรก็ตาม มีอีกแนวทางการวิจัยที่ปรับปรุงวิธีการจำลองของคอนดิชันนอลแรนดอมฟิลด์ เช่น การใช้แบบจำลองคอนดิชันนอลแรนดอมฟิลด์ที่มีโครงสร้างแบบลึก (Deep Structure CRF) [102] ทำให้ประสิทธิภาพการจำแนกดีขึ้นเมื่อเทียบกับคอนดิชันนอลแรนดอมฟิลด์ที่มีโครงสร้างแบบปกติ เนื่องจากคอนดิชันนอลแรนดอมฟิลด์สามารถจับพฤติกรรมการเปลี่ยนแปลงภายนอกของค่าลักษณะสำคัญ แต่ยังไม่สามารถจับพฤติกรรมการเปลี่ยนแปลงภายใน ที่เกิดจากลำดับค่าลักษณะสำคัญที่เปลี่ยนจากค่าหนึ่งไปอีกค่าหนึ่งได้ จึงมีงานวิจัยที่เสนอแบบจำลองเพื่อจับรูปแบบการเคลื่อนที่ภายใน (การเปลี่ยนแปลงในโครงสร้างย่อยภายในของค่าลักษณะสำคัญ) คือ แบบจำลองลาเทนท์ไดนามิคคอนดิชันนอลแรนดอมฟิลด์ (Latent-Dynamic Conditional Random Field: LDCRF) ในวรรณกรรม [21, 103, 104] นำแบบจำลองลาเทนท์ไดนามิคคอนดิชันนอลแรนดอมฟิลด์มาใช้ในการรู้จำท่าทางแบบต่อเนื่อง ซึ่งมีผลการทดลองว่า เมื่อแบบจำลองลาเทนท์ไดนามิคคอนดิชันนอลแรนดอมฟิลด์สามารถเรียนรู้โครงสร้างย่อยของลักษณะสำคัญได้ดีกว่าการเรียนรู้การเคลื่อนที่ภายนอกของค่าลักษณะสำคัญ เนื่องจากวรรณกรรมที่ใช้แบบจำลองลาเทนท์ไดนามิคคอนดิชันนอลแรนดอมฟิลด์ให้ผลการจำแนกถูกต้องเพิ่มขึ้น ทำให้ผู้วิจัยสนใจใช้แบบจำลองลาเทนท์ไดนามิคคอนดิชันนอลแรนดอมฟิลด์ในการรู้จำภาษา โดยมีสมมติฐานว่าแบบจำลองลาเทนท์ไดนามิคคอนดิชันนอลแรนดอมฟิลด์สามารถจับรูปแบบการเคลื่อนที่ภายนอกของค่าลักษณะสำคัญที่เปลี่ยนไปตามฉลากของข้อมูลนั้นคือ ภาษา และสามารถหารูปแบบการเคลื่อนที่ภายในแต่ละค่าย่อยของค่าลักษณะสำคัญได้ ทำให้ได้แบบจำลองที่จำแนกภาษาได้ดีขึ้น

บทที่ 3

ระบบการรู้จำภาษาพูด

ในบทนี้จะกล่าวถึงภาพรวมของระบบการรู้จำภาษาพูด ซึ่งประกอบด้วย ขั้นตอนการสร้างระบบการรู้จำภาษาพูดที่นำเสนอ ขั้นตอนการรวมผลลัพธ์ และการตัดสินใจรู้จำภาษาจากระบบรู้จำย่อยแบบต่าง ๆ นอกจากนี้ในบทนี้ยังกล่าวถึงข้อมูลที่ใช้ในการทดลองและการวัดประสิทธิภาพของระบบการรู้จำภาษาพูด ซึ่งรายละเอียดในแต่ละหัวข้อดังนี้

3.1 ภาพรวมของระบบการรู้จำภาษาพูด

3.1.1 กรอบการรู้จำภาษาแบบความน่าจะเป็น (Probabilistic SLR framework)

ปัญหาการรู้จำภาษาสามารถแสดงในรูปของสมการคณิตศาสตร์ กำหนดให้ $L = \{L_1, L_2, \dots, L_n\}$ แทนเซตของภาษาสำหรับการรู้จำจำนวน n ภาษา เมื่อสัญญาณเสียงพูดที่พิจารณาแทนด้วย X กำหนดให้ค่า L^* แทนโอกาสการรู้จำภาษาที่เป็นไปได้ที่มีค่าความน่าจะเป็นสูงสุดแทนด้วย ดังสมการที่ (3.1)

$$L^* = \arg \max_L P(L | X) \quad (3.1)$$

นั่นคือ หาภาษา L ที่มีค่าความน่าจะเป็นการเกิดมากที่สุด L^* เมื่อเกิดค่าที่สังเกต X จากสมการที่ (3.2) พิจารณาด้วยกฎของเบย์ส์ (Bayes rule) จะได้

$$L^* = \arg \max_L P(X | L)P(L) \quad (3.2)$$

เมื่อ $P(X | L)$ แทนความน่าจะเป็นในการพิจารณาภาษา L ของค่าสังเกตจากเสียงพูด X และ $P(L)$ แทนความน่าจะเป็นที่รู้มาก่อนของภาษา L ซึ่งในที่นี้ให้มีความน่าจะเป็นเท่ากันสำหรับทุกภาษา การใช้ความน่าจะเป็นพิจารณาภาษาตามแนวทางการเรียงตัวของหน่วยเสียง ซึ่งค่าความน่าจะเป็นของภาษาที่รู้จำจากลำดับหน่วยเสียง จะเห็นว่าคุณค่า $P(X | L)$ แทนความน่าจะเป็นการเกิดภาษา L เมื่อเกิดค่าสังเกต X ใช้พิจารณาจากผลรวมค่าความน่าจะเป็นดังกล่าวที่สูงที่สุดดังสมการที่ (3.3) จากค่าที่สังเกตที่ได้จากการรู้จำเสียงแทนด้วย $P(X | H, L, \Lambda)$ และค่าความน่าจะเป็นของแบบจำลองภาษา $P(H | L)$

$$L^* \approx \arg \max_L \sum_H P(X | H, L, \Lambda)P(H | L) \quad (3.3)$$

เมื่อสัญญาณเสียงพูด X ถูกถอดรหัสเป็นลำดับหน่วยเสียง H แล้วนำลำดับหน่วยเสียง H ไปพิจารณาค่าความน่าจะเป็นในการเกิดภาษาจากแบบจำลองภาษา ดังนั้นค่า $P(X | H, L, \Lambda)$ หาได้จากโอกาสการเกิดของค่าที่สังเกต X เมื่อเกิดลำดับของหน่วยเสียง H และเกิดกับภาษา L เมื่อ Λ แทนค่าความน่าจะเป็นที่เกิดแบบจำลองอะคูสติกของหน่วยเสียง โดยที่ค่าความน่าจะเป็นของแบบจำลองภาษา $P(H | L)$ สามารถหาได้จากค่าความน่าจะเป็นที่จะเกิดลำดับหน่วยเสียง เมื่อเกิดกับภาษา L ดังสมการ (3.4)

$$P(H | L) = \prod_k P(h_k | h_{k+N-1}, \dots, h_{k-1}, L) \quad (3.4)$$

ค่าผลรวมของความน่าจะเป็นที่เกิดแบบจำลองอะคูสติกของหน่วยเสียง ได้มาจากการประมาณการของตัวรู้จำเสียงที่ให้ผลลัพธ์เป็นลำดับหน่วยเสียงที่ดีที่สุด 1 ลำดับ (1-Best hypothesis) ดังนั้นซึ่งสามารถแทนด้วย $\max_H P(X | H, L, \Lambda)$ จากสมการที่ (3.5) จะได้

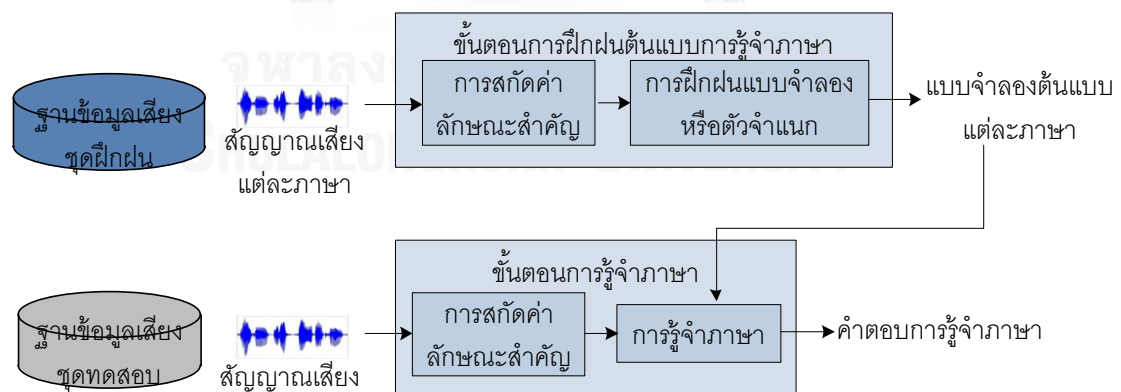
$$L^* \approx \arg \max_L \max_H P(X | H, L, \Lambda) P(H | L) \quad (3.5)$$

เมื่อพิจารณาจากค่าความน่าจะเป็นที่จะเกิดภาษาเมื่อใช้ตัวรู้จำเสียงที่ไม่ขึ้นกับภาษา การพิจารณาค่าความน่าจะเป็นที่เกิดแบบจำลองอะคูสติกของหน่วยเสียง Λ จากผลลัพธ์เป็นลำดับหน่วยเสียงที่ดีที่สุด 1 ลำดับของหนึ่งตัวรู้จำเสียง สามารถหาได้จากสมการที่ (3.6)

$$L^* \approx \arg \max_L \max_H P(X | H, \Lambda) P(H | L) \quad (3.6)$$

3.1.2 ขั้นตอนการทำงานของกรรผู้จำภาษา

ระบบการรู้จำภาษาพูดเป็นการหาคำตอบระบุภาษาของเสียงพูด ซึ่งมีลักษณะของปัญหาในรูปของการรู้จำรูปแบบ (Pattern recognition) ซึ่งขั้นตอนการทำงานของกรรผู้จำรูปแบบสามารถสามารถแบ่งเป็นสองขั้นตอนดังแสดงในภาพที่ 3-1 ได้แก่ ขั้นตอนการสร้างหรือฝึกฝนต้นแบบการรู้จำภาษาและขั้นตอนการรู้จำหรือทดสอบเพื่อตัดสินใจคำตอบของปัญหาที่พิจารณา ในขั้นตอนการสร้างหรือฝึกฝนต้นแบบการรู้จำภาษา นำเสียงพูดแต่ละภาษาจากฐานข้อมูลชุดฝึกฝนมาสกัดค่าลักษณะสำคัญ และนำไปฝึกฝนหรือสร้างต้นแบบการจำแนกภาษาสำหรับแต่ละภาษา ในขั้นตอนการรู้จำภาษา นำเสียงพูดจากฐานข้อมูลชุดทดสอบมาสกัดค่าลักษณะสำคัญ จากนั้นนำค่าลักษณะสำคัญที่ได้เป็นข้อมูลรับเข้าสำหรับการจำแนกภาษา โดยเปรียบเทียบกับแบบจำลองต้นแบบของแต่ละภาษาที่สร้างไว้ ผลลัพธ์ของการจำแนกภาษา ถูกนำมาตัดสินใจเลือกเป็นคำตอบของการรู้จำภาษาและพิจารณาคำตอบของการรู้จำภาษาจากที่ได้จากต้นแบบที่มีค่าความเหมือนที่สุด

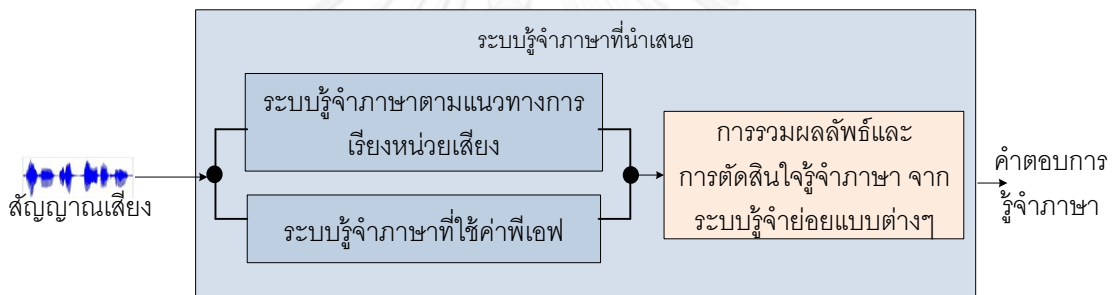


ภาพที่ 3-1 ขั้นตอนการทำงานของกรรผู้จำภาษา

3.2 ขั้นตอนการสร้างระบบการรู้จำภาษาพูดที่นำเสนอ

ระบบการรู้จำภาษาพูดที่มีประสิทธิภาพและเป็นที่ยอมรับในปัจจุบันเป็นระบบการรู้จำภาษาพูดที่รวมระบบการรู้จำภาษาพูดหลายระบบย่อยซึ่งใช้ข้อมูลที่บอกลักษณะเฉพาะของภาษาจากค่าลักษณะสำคัญต่าง ๆ และใช้เทคนิคในการเรียนรู้ข้อมูลเพื่อรู้จำภาษาแตกต่างกัน วิทยานิพนธ์นี้จึงเสนอการระบบการรู้จำภาษาพูดที่มาจากหลายแนวทางและตัดสินใจการรู้จำภาษาจากการใช้ผลการตัดสินใจของระบบการรู้จำภาษาพูดทุกแนวทางร่วมกัน เพื่อใช้ศักยภาพของข้อมูลที่บอกลักษณะเฉพาะของภาษาจากทุกระบบ ช่วยเพิ่มประสิทธิภาพของระบบการรู้จำภาษาพูดที่นำเสนอ

ภาพที่ 3-2 แสดงระบบการรู้จำภาษาพูดที่เสนอในวิทยานิพนธ์นี้ ซึ่งประกอบด้วย ส่วนของระบบการรู้จำภาษาพูดย่อยแต่ละแนวทาง ได้แก่ ระบบการรู้จำภาษาพูดตามแนวทางการเรียงหน่วยเสียง ระบบการรู้จำภาษาพูดที่ใช้ค่าพีเอฟ และการรวมผลลัพธ์สำหรับการใช้รวมผลลัพธ์จากแนวทางต่าง ๆ และตัดสินใจรู้จำภาษา เพื่อตัดสินใจคำตอบการรู้จำภาษาสุดท้ายและเป็นผลลัพธ์ของระบบการรู้จำภาษาพูดที่นำเสนอ



ภาพที่ 3-2 สถาปัตยกรรมของระบบการรู้จำภาษาพูดที่นำเสนอ

รายละเอียดของระบบการรู้จำภาษาพูด แนวทางต่าง ๆ ที่นำเสนอ ดังนี้

- 1) ระบบการรู้จำภาษาพูดตามแนวทางการเรียงหน่วยเสียง ประกอบด้วย ระบบการรู้จำภาษาพูดแบบใช้ระบบรู้จำเสียงตามด้วยการใช้แบบจำลองเวกเตอร์สเปซและระบบการรู้จำภาษาพูดแบบใช้หน่วยเสียงแลตทิซ ร่วมกับใช้แบบจำลองเวกเตอร์สเปซ โดยรายละเอียดของแต่ละระบบแสดงในบทที่ 4
 - 1.1) ระบบการรู้จำภาษาพูดแบบใช้ระบบรู้จำเสียงตามด้วยการใช้แบบจำลองเวกเตอร์สเปซ: ระบบการรู้จำภาษาพูดตามแนวทางนี้ประกอบด้วย ระบบรู้จำเสียง ส่วนการสร้างแบบจำลองเวกเตอร์สเปซ และส่วนจำแนกภาษา เริ่มจากการนำสัญญาณเสียงพูดมาถอดรหัสเป็นลำดับหน่วยเสียงด้วยระบบรู้จำเสียง นำลำดับหน่วยเสียงมาสร้างแบบจำลองภาษาของหน่วยเสียง โดยสร้างเวกเตอร์ค่าลักษณะเฉพาะสำหรับแต่ละภาษา ซึ่งประกอบด้วยค่าความถี่การปรากฏของหน่วยเสียงและค่าความถี่การปรากฏแบบจำลองภาษาของหน่วยเสียงบนสัญญาณเสียงพูด โดยใช้แบบจำลองเวกเตอร์สเปซ นำเวกเตอร์ค่าลักษณะเฉพาะไปใช้ในส่วนการจำแนกภาษา ซึ่งใช้ซัพพอร์ตเวกเตอร์แมชชีนสำหรับการจำแนกภาษา ซัพพอร์ตเวกเตอร์

แมชชีนเป็นการจำแนกแบบดิสคริมิเนทีฟ มีคุณสมบัติในการจำแนกเวกเตอร์ข้อมูลที่มีหลายมิติได้ดี แนวทางนี้ใช้แบบจำลองเวกเตอร์สเปซร่วมกับซับพอร์ตเวกเตอร์แมชชีน แทนการใช้แบบจำลองภาษาเอ็นแกรม

- 1.2) ระบบการรู้จำภาษาพูดแบบใช้หน่วยเสียงแลตทิซ ร่วมกับใช้แบบจำลองเวกเตอร์สเปซ: การทำงานของระบบการรู้จำภาษาพูดแนวทางนี้เหมือนกับระบบการรู้จำภาษาพูดแบบใช้ระบบรู้จำเสียงตามด้วยการใช้แบบจำลองเวกเตอร์สเปซ แต่แนวคิดของระบบการรู้จำภาษาพูดตามแนวทางนี้ใช้ผลลัพธ์จากระบบรู้จำเสียงเป็นหน่วยเสียงแลตทิซ (Phone lattice) แทนลำดับหน่วยเสียงที่เป็นผลลัพธ์ที่ดีที่สุดหนึ่งชุดลำดับ (1-Best phone sequence) จากระบบรู้จำเสียง การใช้ผลลัพธ์เป็นหน่วยเสียงแลตทิซนี้ถูกเสนอเพื่อลดข้อผิดพลาดที่เกิดจากขั้นตอนการถอดรหัสของระบบรู้จำเสียงและเพื่อเพิ่มประสิทธิภาพของระบบการรู้จำภาษาพูดให้มีข้อมูลในการพิจารณาภาษาที่เพิ่มขึ้น

นอกจากนี้วิทยานิพนธ์นี้แนะนำให้เสนอการใช้ฟังก์ชันการถ่วงน้ำหนักเพื่อให้คะแนนการบอกความแตกต่างของข้อมูลในเวกเตอร์ค่าลักษณะเฉพาะไปใช้ในส่วนการจำแนกภาษา เพื่อเพิ่มประสิทธิภาพของการจำแนกภาษาในระบบการรู้จำภาษาพูดตามแนวทางการเรียงหน่วยเสียง ซึ่งรายละเอียดอธิบายในหัวข้อที่ 4.4

- 2) ระบบการรู้จำภาษาพูดที่ใช้ค่าพีเอฟ ประกอบด้วย ระบบการรู้จำภาษาพูดโดยใช้การจำแนกด้วยแบบจำลองลาเทนท์คอนดิชันนอลแรนดอมฟิลด์ร่วมกับค่าพีเอฟและระบบการรู้จำภาษาพูดตามแนวทางการเรียงหน่วยเสียงโดยใช้ค่าพีเอฟ โดยรายละเอียดของแต่ละระบบแสดงในบทที่ 5

- 2.1) ระบบการรู้จำภาษาพูดโดยใช้การจำแนกด้วยแบบจำลองลาเทนท์ไดนามิคคอนดิชันนอลแรนดอมฟิลด์ร่วมกับค่าพีเอฟ: ระบบการรู้จำภาษาพูดตามแนวทางนี้ประกอบด้วย ส่วนการสกัดค่าลักษณะสำคัญและส่วนจำแนกภาษา ซึ่งในวิทยานิพนธ์นี้สนใจการใช้ค่าพีเอฟเพื่อเป็นตัวแทนของค่าลักษณะเฉพาะทางภาษา ในส่วนการสกัดค่าลักษณะสำคัญจึงใช้ตรวจหาค่าลักษณะเฉพาะค่าพีเอฟเพื่อหาตัวแทนเสียงจากสัญญาณเสียงพูดโดยแสดงในรูปของเวกเตอร์ค่าลักษณะเฉพาะ จากนั้นนำเวกเตอร์ค่าลักษณะเฉพาะไปใช้ในส่วนการจำแนกภาษา โดยใช้เทคนิคแบบจำลองลาเทนท์ไดนามิคคอนดิชันนอลแรนดอมฟิลด์เพื่อสำหรับสร้างต้นแบบของการจำแนก ซึ่งเทคนิคนี้สามารถในการเรียนรู้ความสัมพันธ์ระหว่างลำดับค่าที่สังเกตสอดคล้องกับความต้องการของผู้วิจัยที่ต้องการสร้างแบบจำลองของแต่ละภาษาที่มีความสัมพันธ์ระหว่างค่าพีเอฟและลำดับของค่าพีเอฟ

- 2.2) ระบบการรู้จำภาษาพูดตามแนวทางการเรียงหน่วยเสียงโดยใช้ค่าพีเอฟ: ระบบการรู้จำภาษาพูดนี้มีแนวคิดมาจากระบบการรู้จำภาษาพูดแบบใช้ระบบรู้จำเสียงตามด้วยการใช้แบบจำลองเวกเตอร์สเปซ แนวทางนี้ใช้ตัวแทนเสียงอื่นแทนการใช้หน่วยเสียง โดยเสนอการพิจารณาการปรากฏร่วมกันของค่าพีเอฟเป็นตัวแทนเสียง ดังนั้น

ที่ได้จากหน่วยเสียงแลตทิซสามารถแสดงข้อมูลลักษณะทางเสียง เพราะผลลัพธ์ของการรู้จำเสียงได้เป็นกราฟการเชื่อมต่อกันของหน่วยแลตทิซ ซึ่งค่าแต่ละตำแหน่งแสดงค่าความมั่นใจในรูปของคะแนนลักษณะทางเสียง (Acoustic score)

- 2) ข้อมูลค่าพีเอฟ: ค่าพีเอฟถูกใช้แสดงแทนพารามิเตอร์ที่นิยามความสัมพันธ์ของการออกเสียงต่าง ๆ ซึ่งในแต่ละภาษามีหน่วยเสียงที่แตกต่างกันทำให้ค่าการออกเสียงในแต่ละภาษาแตกต่างกัน ส่งผลให้การกระจายของค่าพีเอฟในแต่ละภาษาแตกต่างกัน นอกจากนี้คุณสมบัติของการเปลี่ยนค่าการออกเสียงจากค่าการออกเสียงหนึ่งไปยังอีกค่าหนึ่งในแต่ละภาษามีรูปแบบแตกต่างกัน จึงนำมาสู่การใช้ตัวแทนเสียงด้วยค่าพีเอฟซึ่งคุณสมบัติที่มีลักษณะเฉพาะของแต่ละภาษา ที่แสดงข้อมูลที่อยู่ในระดับเล็กกว่าตัวแทนเสียงจากหน่วยเสียง อย่างไรก็ตามถึงแม้ว่าค่าพีเอฟมีผลกระทบจากตำแหน่งของหน่วยเสียงบนเสียงพูด ในวิทยานิพนธ์นี้ไม่ได้พิจารณาปัจจัยที่มาจากตำแหน่งของหน่วยเสียง แต่ได้พิจารณาความสำคัญของบริบทที่มีผลกระทบต่อค่าพีเอฟ ซึ่งค่าพีเอฟในบริบทที่แตกต่างกันทำให้เกิดความหลากหลายของลำดับค่าพีเอฟชนิดต่าง ๆ ซึ่งวิทยานิพนธ์นี้สนใจพฤติกรรมภายในของการเปลี่ยนแปลงค่าพีเอฟจากการเปลี่ยนค่าการออกเสียงหนึ่งไปยังอีกค่าหนึ่ง ผ่านการเลือกใช้แบบจำลองที่สามารถจับความสัมพันธ์ระหว่างคุณสมบัติต่าง ๆ ในข้อมูลค่าพีเอฟได้ อย่างไรก็ตามถึงแม้ค่าพีเอฟตามนิยาม Chomsky และ Halle ไม่ได้แสดงค่าลักษณะทางเสียงโดยตรง แต่ค่าพีเอฟเป็นผลการจับคู่กับค่าพารามิเตอร์การออกเสียงของแต่ละหน่วยเสียง จึงมีความสัมพันธ์กับค่าอะคูสติกของหน่วยเสียงนั้น ค่าพีเอฟจึงสามารถแสดงค่าลักษณะทางเสียงทางอ้อมได้
- 3) ข้อมูลการเรียงตัวของค่าพีเอฟ: เนื่องจากในแต่ละภาษามีหน่วยเสียงที่แตกต่างกันทำให้ค่าการออกเสียงในแต่ละภาษาแตกต่างกันและกฎเกณฑ์การเรียงตัวของหน่วยเสียงที่แตกต่างกันในแต่ละภาษาทำให้ลำดับหน่วยเสียงในแต่ละภาษาทำให้มีรูปแบบต่างกัน ส่งผลทำให้มีรูปแบบของการเรียงตัวของค่าพีเอฟ นั่นคือ ค่าการออกเสียงที่ปรากฏในแต่ละหน่วยเสียง (สำหรับแต่ละภาษาที่แตกต่างกัน) มีการเปลี่ยนไปยังอีกค่าการออกเสียงที่ปรากฏในหน่วยเสียงที่อยู่ถัดไป จึงทำให้รูปแบบของการปรากฏร่วมกันของค่าพีเอฟแตกต่างกัน ส่งผลให้ความถี่ของการปรากฏร่วมกันของค่าพีเอฟแตกต่างกันสำหรับแต่ละภาษา

จากคุณสมบัติของข้อมูลที่สามารถบอกลักษณะเฉพาะของแต่ละภาษา สนับสนุนให้วิทยานิพนธ์นี้ใช้ข้อมูลข้างต้นมาใช้ในการพัฒนาระบบการรู้จำภาษาพูดที่เสนอ อย่างไรก็ตามถึงแม้ในวิทยานิพนธ์นี้ไม่ได้สกัดค่าลักษณะทางเสียงโดยตรงเพื่อใช้ในการสร้างแบบจำลองสำหรับจำแนกภาษา แต่ข้อมูลของค่าลักษณะทางเสียงได้จากหน่วยเสียงและค่าพีเอฟ นอกจากนี้ในวิทยานิพนธ์นี้ได้ทดลองความสามารถของการจำแนกภาษาเมื่อใช้ข้อมูลที่นำเสนอข้างต้นเปรียบเทียบกับค่าลักษณะสำคัญทางเสียง

3.4 ขั้นตอนการรวมผลลัพธ์และการตัดสินใจรู้จำภาษาจากระบบรู้จำย่อยแบบต่าง ๆ

เนื่องจากระบบการรู้จำภาษาพูดที่มาจากกรรวมหลายระบบนี้ทำให้ประสิทธิภาพการรู้จำภาษาสูงขึ้นและเป็นแนวทางเป็นที่ยอมรับของการพัฒนาระบบการรู้จำภาษาพูดให้มีประสิทธิภาพจากระบบการรู้จำภาษาพูดที่รวมหลายระบบ โดยการรวมระบบการรู้จำภาษาพูดหลายระบบย่อย มาจากการรวมระบบการรู้จำภาษาพูดระบบย่อยที่ไม่มีความสัมพันธ์กัน ซึ่งใช้แนวทางและใช้ค่าลักษณะสำคัญที่แตกต่างกัน หรือมาจากการรวมระบบการรู้จำภาษาพูดที่สัมพันธ์กันโดยใช้แนวทางและใช้ค่าลักษณะสำคัญที่เหมือนกัน นอกจากนี้ด้วยคุณสมบัติของภาษาที่สามารถพิจารณาจากข้อมูลที่แสดงลักษณะเฉพาะของภาษาได้หลายค่า ซึ่งในวิทยานิพนธ์นี้เสนอการใช้ข้อมูลเพื่อใช้ในการบอกความแตกต่างภาษาจากหลายค่า (อธิบายในหัวข้อ 3.3) วิทยานิพนธ์นี้เสนอระบบการรู้จำภาษาพูดที่รวมผลลัพธ์การตัดสินใจจากหลายระบบที่มีแนวทางต่าง ๆ ที่ได้นำเสนอข้างต้น เพื่อนำผลลัพธ์จากระบบแต่ละแนวทางตัดสินใจเป็นคำตอบของการรู้จำภาษาของระบบที่นำเสนอ การรวมคำตอบจากระบบการรู้จำภาษาพูดแต่ละแนวทาง จึงถูกนำมาใช้เพื่อรวมผลลัพธ์ของการใช้ข้อมูลตัดสินใจรู้จำภาษาและช่วยให้ประสิทธิภาพของการรู้จำภาษาดีขึ้น ดังนั้นแนวคิดของรวมคะแนนจากผลการตัดสินใจจากระบบการรู้จำภาษาพูดแต่ละแนวทาง เหมือนกับการพิจารณาคำตอบจากผู้เชี่ยวชาญที่ใช้วิธีและข้อมูลที่แตกต่างกัน เมื่อรวมคะแนนจากระบบแต่ละแนวทางมาตัดสินใจรู้จำภาษาร่วมกัน จึงเป็นผลลัพธ์การตัดสินใจจากคำตอบจากผู้เชี่ยวชาญที่ให้คะแนนร่วมกัน

ขั้นตอนการรวมผลลัพธ์และการตัดสินใจรู้จำภาษาเป็นการรวมค่าโอกาสการเกิด (Likelihood) หรือค่าความน่าจะเป็นของการรู้จำภาษาจากทุกระบบ เพื่อผลลัพธ์สุดท้ายของการตัดสินใจของระบบการรู้จำภาษาพูดที่นำเสนอในวิทยานิพนธ์นี้ ซึ่งมีสมมติฐานคือ คะแนนการตัดสินใจรู้จำภาษาจากแต่ละระบบย่อยซึ่งมีแนวทางที่ไม่มีความสัมพันธ์กัน คะแนนไม่ขึ้นต่อกัน การทำงานในหัวข้อนี้ประกอบด้วย ขั้นตอนการทำให้คะแนนเป็นบรรทัดฐาน (Score normalization) การรวม (Fusion) และการทำให้เป็นค่ามาตรฐาน (Calibration)

3.4.1 การทำให้คะแนนเป็นบรรทัดฐาน

เนื่องจากระบบการรู้จำภาษาพูดย่อยแต่ละแนวทางใช้ค่าลักษณะสำคัญและวิธีการจำแนกที่แตกต่างกัน ทำให้คะแนนที่ได้จากการตัดสินใจรู้จำภาษาจากแต่ละระบบมีค่าแตกต่างกันและมีช่วงคะแนนที่แตกต่างกัน วิทยานิพนธ์นี้จึงนำการทำให้คะแนนเป็นบรรทัดฐาน เพื่อลดความลำเอียงของคะแนนจากแต่ละระบบที่มีค่าแตกต่างกัน และนำคะแนนมารวมกันสำหรับการตัดสินใจรู้จำภาษาเป็นผลลัพธ์สุดท้ายของวิทยานิพนธ์นี้ วิธีการทำให้คะแนนจากแต่ละระบบด้วยเป็นบรรทัดฐานเดียวกันที่นำมาใช้ในหัวข้อนี้ คือ การทำให้เป็นบรรทัดฐานด้วยค่าเฉลี่ย โดยค่าเฉลี่ยของคะแนนจากแต่ละระบบย่อยจะถูกนำมาหารคะแนนของแต่ละระบบเพื่อทำให้เป็นบรรทัดฐานและนำคะแนนมาประกอบกันเป็นเวกเตอร์ของคะแนนการตัดสินใจ

3.4.2 การรวมผลลัพธ์

การรวมผลลัพธ์การตัดสินใจของระบบการรู้จำภาษาพูดที่นำเสนอ เมื่อมีระบบการรู้จำภาษาพูดแต่ละแนวทางย่อย k ระบบ และภาษาเป้าหมายที่ต้องการรู้จำภาษามีจำนวน N ภาษา

กำหนดให้ l แทนภาษาเป้าหมายที่พิจารณาการรู้จำ เวกเตอร์ของคะแนนที่ได้จากระบบย่อยที่ k แทนด้วย s_k คะแนนการตัดสินใจรู้จำภาษาจากแต่ละระบบ N ภาษาจะได้เวกเตอร์ของคะแนนที่ได้จากระบบแต่ละแนวทาง $s_k = [s(k, 1), s(k, 2), \dots, s(k, N)]^T$ ซึ่งถูกใช้พิจารณาคะแนนการตัดสินใจสำหรับการรู้จำภาษาสำหรับแต่ละภาษาเป้าหมายของการรู้จำภาษาของระบบที่นำเสนอ ดังนั้นเวกเตอร์คะแนนจากระบบแต่ละแนวทาง s_k ถูกนำมาประกอบกันเป็นเมทริกซ์คะแนนทั้งระบบซึ่งแทนด้วย $S = [s_1^T, s_2^T, \dots, s_k^T]^T$ โดยที่มีจำนวนของสมาชิกในเมทริกซ์เป็น $k \times N$ ค่า ซึ่งเวกเตอร์คะแนนทั้งระบบนำมาใช้ในการรวมผลลัพธ์ต่อไป

เนื่องจากการรวมผลลัพธ์การตัดสินใจจากระบบย่อยมีหลายวิธี ซึ่งวิธีการรวมคะแนนจากระบบย่อยที่ใช้ในวิทยานิพนธ์นี้ซึ่งเป็นวิธีการรวมคะแนนที่เป็นที่ยอมรับในการระบุผู้พูดและการรู้จำภาษา [105-108] ได้แก่ การรวมคะแนนแบบเชิงเส้น กระบวนการทำงานส่วนหลังด้วยแบบจำลองเกาส์เซียน ดังนี้

3.4.2.1 การรวมคะแนนแบบเชิงเส้น

การรวมคะแนนแบบเชิงเส้นมีสมมติฐานว่า ระบบการรู้จำภาษาพูดแต่ละแนวทางไม่ขึ้นต่อกัน กำหนดให้คะแนน $S(O|\lambda_{k,l})$ เป็นผลลัพธ์ ที่ได้จากแบบจำลอง $\lambda_{k,l}$ ซึ่งแสดงด้วยค่าโอกาสการเกิดของภาษา l ที่ได้จากระบบย่อย k การรวมคะแนนจากระบบย่อยต่าง ๆ คือ ผลรวมค่าล็อกของคะแนนโอกาสการเกิดของระบบย่อยดังสมการที่ (3.7) การรวมผลลัพธ์จากผลบวกค่าล็อกของคะแนนโอกาสการเกิดของระบบย่อยมีข้อดีของคือไม่ต้องการขั้นตอนการฝึกฝน สามารถรวมได้คะแนนโอกาสการเกิดของระบบย่อยได้โดยตรง ซึ่งการรวมคะแนนดังสมการที่ (3.7) เป็นการให้น้ำหนักความสำคัญของคะแนนผลลัพธ์จากแต่ละระบบย่อยเท่า ๆ กัน

$$s_l = \sum_{k=1}^K \log S(O|\lambda_{k,l}) = \sum_{k=1}^K s(k,l) \quad (3.7)$$

อย่างไรก็ตามเนื่องจากความสามารถของการบอกความแตกต่างระหว่างภาษาจากการตัดสินใจของระบบแต่ละแนวทาง มีความสามารถแตกต่างกัน ดังนั้นการถ่วงน้ำหนักคะแนนที่ได้จากแต่ละระบบจึงถูกนำมาใช้ในการรวมคะแนนผลลัพธ์ของแต่ละระบบย่อย ดังสมการที่ (3.8)

$$s_l = \sum_{k=1}^K \alpha_k s(k,l) + \beta_l \quad (3.8)$$

เมื่อ α_k แทนค่าสัมประสิทธิ์การถ่วงน้ำหนักคะแนนที่ได้จากระบบย่อยที่ k และใช้ในการคำนวณคะแนนค่าโอกาสการเกิดของภาษาเป้าหมายทุกภาษาและ β_l แทนเวกเตอร์ค่าระยะห่าง (Offset) สำหรับแต่ละภาษาเป้าหมาย ซึ่งค่าสัมประสิทธิ์การถ่วงน้ำหนักคะแนนและเวกเตอร์ค่าระยะห่างของฟังก์ชันการรวมคะแนนค่าล็อกของโอกาสการเกิด ได้มาจากการหาจุดที่ดีที่สุด (Optimization) ของการคำนวณฟังก์ชันวัตถุประสงค์ (Objective function) ในการค้นหาแบบการค้นหาแบบทั้งหมด (Exhaustive search) การใช้แบบจำลองการถดถอยแบบพหุคูณ (Multiclass logistic regression) ถูกนำมาใช้ในการค้นหาค่าที่เป็นไปได้ทั้งหมดของค่าสัมประสิทธิ์ค่าการถ่วง

น้ำหนักคะแนนของแต่ละระบบและเวกเตอร์ค่าระยะห่างที่ทำให้ค่าฟังก์ชันวัตถุประสงค์มีค่ามากที่สุด ซึ่งคำนวณดังสมการที่ (3.9)

$$(\alpha_1, \alpha_2, \dots, \alpha_K, \beta_l) = \arg \max C'_{lr} \quad (3.9)$$

เมื่อ C'_{lr} แทนค่าฟังก์ชันวัตถุประสงค์ของการวัดประสิทธิภาพของระบบการรู้จำภาษาพูด ซึ่งอธิบายในหัวข้อ 3.7 ในการหาสัมประสิทธิ์ที่ทำให้ค่าฟังก์ชันวัตถุประสงค์ด้วยแบบจำลองการถดถอยที่มีค่ามากที่สุดนี้ ใช้ข้อมูลชุดฝึกฝนในการหาค่าพารามิเตอร์ [105-107] ซึ่งฟังก์ชันวัตถุประสงค์ของแบบจำลองการถดถอย คือค่าล็อกความน่าจะเป็นแบบภายหลังทั้งหมดของข้อมูลชุดฝึกฝน

$$\log \prod_{m \in M} P(L_l, s_l) = \sum_{m=1}^M \sum_{l=1}^N \gamma_{ml} \log P(L_l, s_l) \quad (3.10)$$

เมื่อ M แทนจำนวนข้อมูลในชุดฝึกฝน γ_{ml} แทนค่าถ่วงน้ำหนักของแต่ละภาษา l ซึ่งหาได้จากสัดส่วนของจำนวนข้อมูลฝึกฝนของแต่ละภาษาเทียบกับทั้งหมด หรือในกรณีที่แต่ละภาษามีจำนวนข้อมูลฝึกฝนเท่ากัน ค่าน้ำหนักนี้จะมีค่าให้เป็น 1 และความน่าจะเป็นภายหลัง $P(L_l, s_l)$ คำนวณจากสมการ (3.11)

$$P(L_l | s_l) = \frac{e^{s_l}}{\sum_{i=1}^N e^{s_i}} \quad (3.11)$$

3.4.2.2 กระบวนการทำงานส่วนหลังด้วยแบบจำลองเกาส์เซียน (Guassian back end)

เวกเตอร์คะแนนโอกาสการเกิดที่ได้จากแต่ละระบบนำมารวมกันด้วย กระบวนการทำงานส่วนหลังด้วยแบบจำลองเกาส์เซียน ซึ่งทำหน้าที่ในการหาค่าความน่าจะเป็นภายหลังของคะแนนการตัดสินใจจากระบบย่อยต่าง ๆ เวกเตอร์คะแนนของแต่ละภาษาถูกใช้นำมาฝึกฝนการกระจายตัวแบบปกติที่มีหลายค่าสำหรับแบบจำลองของแต่ละภาษา ในกระบวนการทำงานส่วนหลังใช้วิธีการหาค่าโอกาสการเกิดสูงสุด (Maximum Likelihood: ML) เพื่อหาเซตของค่าพารามิเตอร์ของแบบจำลองเกาส์เซียนสำหรับแต่ละภาษา การกระจายแบบปกติที่มีหลายค่าของภาษา l แทนด้วย $N\{\mu_l, \Sigma_l\}$ ซึ่งประกอบด้วยเวกเตอร์ของค่าเฉลี่ย μ_l และเมทริกซ์ความแปรปรวนร่วม Σ_l คะแนนจากฟังก์ชันตัดสินใจ s_l สำหรับแต่ละภาษา s_l คำนวณได้จากฟังก์ชันการตัดสินใจที่แสดงในรูปของการประมาณจากค่าล็อกโอกาสการเกิดของแบบจำลองเกาส์เซียนของภาษา l ซึ่งในวิทยานิพนธ์นี้เกาส์เซียนทุกภาษาใช้เมทริกซ์ความแปรปรวนร่วมเหมือนกัน ฟังก์ชันการตัดสินใจจึงดำเนินการในรูปการรวมกันแบบเชิงเส้นของเกาส์เซียนทุกภาษานี้ ดังนั้นฟังก์ชันการตัดสินใจจากกระบวนการทำงานส่วนหลังจากแบบจำลองเกาส์เซียน สามารถคำนวณได้ดังสมการ (3.12)

$$s_l = \log N(s | \mu_l, \Sigma) - \beta = (\mu_l^T, \Sigma^{-1})s - \frac{1}{2} \mu_l^T, \Sigma^{-1} \mu_l \quad (3.12)$$

เมื่อ μ_i แทนเวกเตอร์ของค่าเฉลี่ย Σ แทนเมทริกซ์ความแปรปรวนร่วม s แทนเมทริกซ์คะแนน ทั้งระบบ เนื่องจากในระบบการรู้จำภาษาพูดประกอบด้วย 4 ระบบย่อยและแต่ละระบบประกอบด้วย 8 ภาษา ดังนั้นจำนวนมิติของเวกเตอร์คะแนนของทั้งระบบมี 78 มิติ

เนื่องจาก Focal Multi-class [107] เป็นเครื่องมือซึ่งนำไปใช้เพื่อทำให้คะแนนเป็นมาตรฐาน และรวมผลลัพธ์ระบบต่าง ๆ พัฒนาโดย Brummer ถูกนำไปใช้ในระบบการรู้จำภาษาพูดและเป็นที่ยอมรับและประสบความสำเร็จในประเมินผลการรู้จำภาษา NIST LRE ดังนั้นในวิทยานิพนธ์นี้ใช้เครื่องมือ Focal multi-class [107] ซึ่งใช้แบบจำลองการถดถอยแบบพหุคูณ เพื่อดำเนินการกับเวกเตอร์คะแนนที่ได้จากระบบย่อยต่าง ๆ ในขั้นตอนการทำให้เป็นค่ามาตรฐาน การรวมคะแนนและกระบวนการทำงานส่วนหลังด้วยแบบจำลองเกาส์เซียน โดยใช้ข้อมูลชุดฝึกฝนสำหรับขั้นตอนการหาค่าพารามิเตอร์เพื่อการรวมคะแนนและส่วนการทำงานส่วนหลังด้วยแบบจำลองเกาส์เซียน

3.5 ข้อมูลที่ใช้ในการทดลอง

ฐานข้อมูลเสียงที่ใช้ในการทดลองประกอบด้วย ข้อมูลเสียงที่ใช้ในการพัฒนาและทดสอบระบบตรวจหาลักษณะเฉพาะค่าพีเอฟ ข้อมูลเสียงที่ใช้ในการศึกษาการกระจายค่าพีเอฟในแต่ละภาษา ข้อมูลเสียงที่ใช้ในการทดลองระบบการรู้จำภาษาพูดและข้อมูลเสียงที่ใช้ในการพัฒนาและทดสอบระบบการรู้จำภาษาพูดตามแนวทางการเรียงหน่วยเสียง เป็นต้น

3.5.1 ข้อมูลเสียงที่ใช้ในการพัฒนาและทดสอบระบบตรวจหาลักษณะเฉพาะค่าพีเอฟ

ฐานข้อมูลเสียง TIMIT เป็นฐานข้อมูลที่ใช้ในการพัฒนาและทดสอบระบบตรวจหาลักษณะเฉพาะค่าพีเอฟ (ซึ่งขั้นตอนการสร้างระบบอธิบายในหัวข้อ 6.2 ของบทที่ 6) ฐานข้อมูลเสียง TIMIT ประกอบด้วยเสียงพูดที่บันทึกการอ่านประโยคในภาษาอังกฤษจากผู้พูดเพศชายและหญิงจำนวน 630 คนโดยผู้พูดมีสำเนียงภาษาใน 8 ภาษาท้องถิ่นของภาษาอังกฤษแบบอเมริกัน ฐานข้อมูลเสียงนี้มีหน่วยเสียงจำนวน 61 หน่วยเสียง จำนวนเสียงพูดในฐานข้อมูลเสียงนี้มี 6300 ประโยค ซึ่งสามารถแบ่งเป็นข้อมูลเซตฝึกฝน (จำนวน 4620 ประโยคจากผู้พูด 462 คน) และข้อมูลเซตทดสอบ (จำนวน 1680 ประโยคจากผู้พูด 168 คน) โดยข้อมูลเซตฝึกฝน ถูกใช้ในขั้นตอนการสร้างระบบตรวจหาลักษณะเฉพาะค่าพีเอฟและข้อมูลเซตทดสอบใช้ในการทดสอบวัดประสิทธิภาพของการตรวจหาลักษณะเฉพาะค่าพีเอฟ

3.5.2 ข้อมูลเสียงที่ใช้ในการศึกษาการกระจายค่าพีเอฟในแต่ละภาษา

วิทยานิพนธ์นี้ได้ศึกษาการกระจายของค่าพีเอฟบนหน่วยเสียงที่แตกต่างกันในแต่ละภาษา ซึ่งอธิบายวิธีการศึกษาในหัวข้อที่ 5.2.2 ของบทที่ 5 เพื่อศึกษาการกระจายค่าลักษณะสำคัญของแต่ละภาษา จึงต้องการฐานข้อมูลเสียงที่มีหลายภาษาเพื่อนำมาสกัดค่าลักษณะสำคัญจากแต่ละภาษา แต่เนื่องจากฐานข้อมูลเสียง NIST 2007 ที่ใช้ในการรู้จำภาษาที่ใช้ในวิทยานิพนธ์นี้ไม่มีการถอดความหน่วยเสียง ดังนั้นฐานข้อมูลเสียง OGI แบบหลายภาษา (Oregon Graduate Institute Multi-language corpus) [109] ฐานข้อมูลเสียงนี้บันทึกด้วยเสียงพูดผ่านทางโทรศัพท์ ซึ่งมีสโตร์การอ่านเพื่อบันทึกเสียงสองแบบ ได้แก่ บันทึกแบบคำโดดจากคำศัพท์ที่กำหนดและแบบเสียงพูดต่อเนื่องจาก

การพูดในเสียงหัวข้อแบบเปิด (Open topic) เสียงพูดจากส่วนแบบเสียงพูดต่อเนื่องถูกเลือกมาใช้ในการทดลองนี้ เนื่องจากฐานข้อมูลเสียงส่วนนี้มีการถอดความหน่วยเสียงที่ละเอียด แต่ละหน่วยเสียงมีความยาวประมาณ 1 นาที และมีการถอดความหน่วยเสียงสัทลักษณะตามรูปแบบการเข้ารหัสแอสกี (ASCII) แบบ Worldbet [110] ฐานข้อมูลเสียงส่วนนี้ประกอบด้วย 6 ภาษาได้แก่ อังกฤษ (208) เยอรมัน (101) ฮินดี (68) ญี่ปุ่น (64) จีนแบบแมนดาริน (70) และสเปน (108) ซึ่งตัวเลขที่ในวงเล็บแสดงจำนวนของเสียงพูดในแต่ละภาษา

3.5.3 ข้อมูลเสียงที่ใช้ในการทดลองระบบการรู้จำภาษาพูด

ฐานข้อมูลเสียง NIST LRE07 เซตข้อมูลพัฒนา ซึ่งเป็นฐานข้อมูลที่ใช้ในการรู้จำภาษาที่ใช้ในวิทยานิพนธ์นี้ ฐานข้อมูลเสียงนี้เป็นส่วนหนึ่งของข้อมูลที่ใช้ในการพัฒนาระบบการรู้จำภาษาพูด เพื่อวัดประเมินผลการรู้จำภาษา (Language Recognition Evaluation: LRE) ที่จัดขึ้นด้วยองค์กร NIST (National Institute of Standards and Technology: NIST) ปี คศ. 2007 ซึ่งฐานข้อมูลเสียงนี้เป็นเซตข้อมูลพัฒนา (Development data set) ฐานข้อมูลเสียงนี้ประกอบด้วย 8 ภาษา ได้แก่ ภาษาอารบิก (Arabic) แทนด้วย arb, ภาษาเบงกาลี (Bengali) แทนด้วย ben, ภาษาไทย (Thai) แทนด้วย tha, ภาษาเออร์ดู (Urdu) แทนด้วย urd, ภาษารัสเซีย (Russian) แทนด้วย rus, ภาษาจีนกวางตุ้ง (Chinese-Cantonese) แทนด้วย CH.can, ภาษาจีนมิน (Chinese-Min) แทนด้วย CH.min และภาษาจีนวู (Chinese-Wu) แทนด้วย CH.wu เป็นต้น ดังนั้นในการพัฒนาและการทดสอบผลการรู้จำภาษาในวิทยานิพนธ์นี้วัดผลการรู้จำภาษาบน 8 ภาษาฐานข้อมูลเสียงนี้มีจำนวนเสียงพูด 6000 เสียง ประกอบด้วยเสียงพูดที่มีความยาว 3 วินาที 10 วินาที และ 30 วินาที ตารางที่ 3-1 แสดงจำนวนชั่วโมงเสียงพูดของฐานข้อมูลเสียงนี้ที่มีใช้ในวิทยานิพนธ์นี้ซึ่งแบ่งตามแต่ละภาษา

ตารางที่ 3-1 จำนวนชั่วโมงของเสียงพูดในงานวิจัย แบ่งตามแต่ละภาษา

ภาษา	จำนวนชั่วโมง
ภาษาอารบิก	1.45
ภาษาเบงกาลี	1.42
ภาษาจีน (รวมภาษาจีนกวางตุ้ง ภาษาจีนมิน และภาษาจีนวู)	4.20
ภาษาไทย	1.23
ภาษารัสเซีย	1.33
ภาษาเออร์ดู	1.42

3.5.4 ข้อมูลเสียงที่ใช้ในการทดลองการรู้จำภาษาตามแนวทางการเรียงหน่วยเสียง

3.5.4.1 ข้อมูลเสียงที่ใช้ในการฝึกฝนระบบการรู้จำภาษาพูดตามแนวทางการเรียงหน่วยเสียง

ฐานข้อมูลเสียงที่ใช้ในการรู้จำภาษาที่ใช้ในการฝึกฝนของระบบการรู้จำภาษาพูดตามแนวทางการเรียงหน่วยเสียง ประกอบด้วย 3 ส่วน ได้แก่

- 1) ฐานข้อมูลเสียง CallFriend ประกอบด้วย ภาษาอารบิก ภาษาจีน ภาษาจีนแมนดาริน (Mandarin) ภาษาอังกฤษ ภาษาฮินดี (Hindi) ภาษาสเปน ภาษาฟาซี (Farsi) ภาษาเยอรมัน ภาษาญี่ปุ่น ภาษาเกาหลี ภาษาทมิฬ (Tamil) และภาษาเวียดนาม เป็นต้น ตารางที่ 3-2 แสดงจำนวนชั่วโมงของเสียงพูดของฐานข้อมูลเสียง แบ่งตามแต่ละภาษา
- 2) ฐานข้อมูลเสียง OHSU 2005 ซึ่งฐานข้อมูลเสียงนี้จัดเก็บฐานข้อมูลเสียงด้วยองค์กร Oregon Health & Science University (OHSU) ประกอบด้วย ภาษาอังกฤษ ภาษาเยอรมัน ภาษาฮินดี ภาษาญี่ปุ่น ภาษาเกาหลี ภาษาจีน ภาษาสเปน และภาษาทมิฬ เป็นต้น
- 3) ฐานข้อมูลเสียง NIST LRE 2007 เซตข้อมูลพัฒนา ซึ่งอธิบายในข้อ 3.5.3

ตารางที่ 3-2 จำนวนชั่วโมงของเสียงพูดฐานข้อมูลเสียง CallFriend
แบ่งตามแต่ละภาษา

ภาษา	จำนวนชั่วโมง	ภาษา	จำนวนชั่วโมง
ภาษาอารบิก	19.5	ภาษาเยอรมัน	21.6
ภาษาจีน	41.7	ภาษาญี่ปุ่น	19.1
ภาษาอังกฤษ	39.8	ภาษาเกาหลี	18.4
ภาษาฮินดี	19.6	ภาษาทมิฬ	18.4
ภาษาสเปน	43.8	ภาษาเวียดนาม	20.6
ภาษาฟาซี	21.2		

3.5.4.2 ข้อมูลเสียงที่ใช้ในการทดสอบระบบการรู้จำภาษาพูดตามแนวทางการเรียงหน่วยเสียง

การประเมินประสิทธิภาพของระบบการรู้จำภาษาพูดตามแนวทางการเรียงหน่วยเสียง โดยวัดผลการรู้จำภาษาบนฐานข้อมูลเสียง NIST LRE07 เซตข้อมูลทดสอบ ซึ่งฐานข้อมูลเสียงนี้ใช้ในการวัดประเมินประสิทธิภาพของระบบการรู้จำภาษาพูด LRE ที่จัดขึ้นด้วยองค์กร NIST ปี ค.ศ. 2007 การทดสอบผลการรู้จำภาษาวัดภาษาภายในเซตแบบปิดบน 14 ภาษาซึ่งประกอบด้วย ภาษาอารบิก แทนด้วย arb ภาษาเบงกาลี แทนด้วย ben ภาษาจีน แทนด้วย CH ภาษาอังกฤษ แทนด้วย EN ภาษาฟาซิ แทนด้วย far ภาษาเยอรมัน แทนด้วย ger ภาษาฮินดี (Hindustani) แทนด้วย hin ภาษาญี่ปุ่น แทนด้วย jap ภาษาเกาหลี แทนด้วย kor ภาษารัสเซีย แทนด้วย rus ภาษาสเปน แทนด้วย spa ภาษาทมิฬ ภาษาไทย และภาษาเวียดนาม แทนด้วย vie เป็นต้น ฐานข้อมูลเสียงนี้มีจำนวน เสียงพูด 7,530 เสียง ซึ่งประกอบด้วยเสียงพูดที่มีความยาว 3 วินาที 10 วินาที และ 30 วินาที

นอกจากนี้ข้อมูลเสียงนี้ ถูกนำมาเลือกไฟล์เสียงสำหรับ 8 ภาษาตามเซตภาษาที่ใช้ในการรู้จำเท่านั้น ซึ่งแต่ละภาษา มีจำนวนข้อมูลเสียงพูดภาษาละ 240 เสียงรวมทั้งหมดเป็นจำนวนเสียงพูด 1,920 เสียง เพื่อเป็นข้อมูลชุดทดสอบอีกสำหรับชุดประเมินผลระบบการรู้จำภาษาพูดตามแนวทางอื่น เพื่อวัดประสิทธิภาพของการตัดสินใจจากระบบการรู้จำภาษาพูดที่น่าเสนอซึ่งรวมผลลัพธ์จากทุกแนวทาง

3.6 การวัดประสิทธิภาพของการรู้จำภาษา

การรู้จำภาษา สามารถแบ่งออกเป็น 2 ปัญหาซึ่งใช้หลักวิธีการพิจารณาต่างกัน ดังนี้

- 1) การระบุภาษาซึ่งเป็นการตัดสินใจระบุคำตอบภาษาของเสียงพูดว่าเป็นภาษาใดจากภาษาภายในเซตของภาษาที่จำนวนจำกัด เมื่อกำหนดให้ L แทนเซตของภาษาพูดที่มีจำนวน n ภาษาหรือ $L = \{L_1, L_2, \dots, L_n\}$ ดังนั้นผลลัพธ์ของการระบุภาษาเป็นภาษาพูดใด ๆ จากภาษาพูดภายในเซตที่เป็นไปได้จากผลลัพธ์การจำแนก
- 2) การตรวจหาภาษา ซึ่งเป็นการตัดสินใจรู้จำภาษาจากสมมติฐานที่พิจารณาการตัดสินใจแบบสองทางว่า ภาษาของเสียงพูดที่เป็นภาษา L_i หรือไม่ โดยเปรียบเทียบกับคำตอบการตัดสินใจรู้จำภาษาที่เสียงพูดเป็นภาษาอื่นที่ไม่ใช่ภาษา L_i (สำหรับกรณีที่พิจารณาการตัดสินใจมากกว่า 2 ภาษา ใช้การวนซ้ำพิจารณาสมมติฐานสำหรับทุกภาษา โดยแสดงผลการตรวจหาภาษาจาก การรวมคำตอบของสมมติฐานของภาษาอื่น ๆ ที่ไม่ใช่ภาษา L_i)

นอกจากนี้รูปแบบของผลการตัดสินใจคำตอบของการรู้จำภาษา สามารถตอบได้ 2 ลักษณะ ได้แก่ การตัดสินใจแบบถาวร (Hard decision) และการตัดสินใจแบบเปลี่ยนได้ (Soft decision) การตัดสินใจแบบถาวรสำหรับการระบุภาษา คือการตัดสินใจระบุคำตอบภาษาของเสียงพูดว่าเป็นภาษาใดจากเซตกำหนดและการตัดสินใจแบบถาวรสำหรับการตรวจหาภาษา คือการตัดสินใจว่าสมมติฐานของการตัดสินใจแบบสองทางใช่หรือไม่ ขณะที่การตัดสินใจแบบเปลี่ยนได้สำหรับการระบุ

ภาษา คือการกระจายตัวของความน่าจะเป็นของการระบุภาษาสำหรับแต่ละภาษาภายในเซต แทนด้วย $[p(L_1), p(L_2), \dots, p(L_N)]$ เมื่อ $p(L_i)$ แทนความน่าจะเป็น หรือค่าโอกาสการเกิดของแต่ละภาษาและการตัดสินใจแบบเปลี่ยนได้สำหรับการตรวจหาภาษา คือการกระจายตัวของความน่าจะเป็นของสมมติฐานที่พิจารณาการตัดสินใจแบบสองทางว่าภาษาของเสียงพูดที่เป็นภาษา L_i ที่พิจารณาหรือไม่ แทนด้วย $[p(L_i), p(-L_i)]$ เมื่อ $p(-L_i)$ แทนความน่าจะเป็น หรือค่าโอกาสการเกิดของภาษาอื่นที่ไม่ใช่ภาษา L_i เมื่อ $p(-L_i) = 1 - p(L_i) = \sum_{k \neq i} p(L_k)$

การวัดประสิทธิภาพการรู้จำภาษา จึงใช้วิธีการวัดที่แตกต่างกันสำหรับแต่ละปัญหา จึงใช้วิธีการประเมินต่างกัน นั่นคือ สำหรับการวัดประสิทธิภาพของการตัดสินใจแบบถาวรสำหรับการระบุภาษา พิจารณาจากการวัดความแม่นยำ (Accuracy) การวัดอัตราความผิดพลาด (Error rate) - ขณะที่การวัดประสิทธิภาพของการตัดสินใจแบบเปลี่ยนได้สำหรับการตรวจหาภาษา พิจารณาจากการวัดค่าความเชื่อมั่น (Confidence measure) ของการจำแนก โดยค่าความเชื่อมั่นแสดงโอกาสการเกิดผลลัพธ์ของการจำแนกว่าถูกต้องเท่าใด การวัดประสิทธิภาพของการรู้จำภาษาด้วยค่า EER การวัดประสิทธิภาพของการรู้จำภาษาด้วยฟังก์ชันค่าการตรวจหา เป็นต้น ในวิทยานิพนธ์นี้การวัดค่าความเชื่อมั่นคำนวณจาก คะแนนของการตัดสินใจรู้จำภาษาที่ได้จากระบบการรู้จำภาษาพูดแต่ละระบบแนวทาง

3.6.1 การวัดประสิทธิภาพจากความแม่นยำ

การรู้จำภาษาในวิทยานิพนธ์นี้ พิจารณาระบุคำตอบของภาษาของเสียงพูดโดยมีขอบเขตคำตอบภายในเซตปิดบนฐานข้อมูลที่ทดสอบ การวัดประสิทธิภาพจากความแม่นยำของการจำแนกภาษาคำนวณดังสมการ (3.13) ใช้ในการวัดประสิทธิภาพของการระบุภาษาจากผลลัพธ์ของการจำแนกถูกต้องหรือไม่

$$\text{เปอร์เซ็นต์ความแม่นยำ (Acc)} = \frac{\text{จำนวนเสียงพูดที่ระบบทำนายได้ถูกต้อง}}{\text{จำนวนเสียงพูดทั้งหมด}} \times 100 \quad (3.13)$$

3.6.2 การวัดประสิทธิภาพจากความผิดพลาด

การวัดผลประสิทธิภาพของการระบุภาษาโดยพิจารณาผลลัพธ์ของการจำแนกถูกต้องหรือไม่ โดยวัดอัตราความผิดพลาด คำนวณดังสมการ (3.14) ซึ่งการวัดความผิดพลาด นิยามจากอัตราส่วนของความผิดพลาดการจำแนกซึ่งไม่ตรงกับคำตอบจริงและจำนวนเสียงพูดที่ทดสอบทั้งหมด

$$\text{อัตราค่าความผิดพลาด} = \frac{\text{จำนวนของเสียงพูดที่ระบบทำนายผิด}}{\text{จำนวนเสียงพูดทั้งหมดที่ใช้ในการระบุภาษา}} \quad (3.14)$$

นอกจากนี้ในการเปรียบเทียบประสิทธิภาพของระบบเมื่อใช้ค่าลักษณะสำคัญและค่าพารามิเตอร์ที่ต่างกัน การวัดค่าความผิดพลาดที่ลดลงแบบสัมพัทธ์ (Relative resuction of ER) นิยามด้วยสมการที่ (3.15) ถูกนำมาใช้เพื่อคำนวณค่าความผิดพลาดที่ลดลงแบบสัมพัทธ์และเปรียบเทียบประสิทธิภาพการระบุภาษา

$$\text{ค่าความผิดพลาดที่ลดลงแบบสัมพัทธ์} = 100\% - \frac{ER2}{ER1} \times 100\% \quad (3.15)$$

เมื่อ ER1 และ ER2 แทนค่าความผิดพลาดของการจำแนกของระบบที่ 1 และระบบที่ 2 ตามลำดับ

3.6.3 การวัดประสิทธิภาพจากความเชื่อมั่น

ค่าความเชื่อมั่นถูกพิจารณาเปรียบเทียบกับค่าอ้างอิงที่กำหนด (Threshold) ถ้าค่าความเชื่อมั่นต่ำกว่าค่าอ้างอิงที่กำหนด ผลลัพธ์ของคำตอบนั้นจะถูกปฏิเสธ แต่ถ้าค่าความเชื่อมั่นสูงกว่าค่าที่กำหนด ผลลัพธ์ของคำตอบนั้นจะถูกยอมรับ สำหรับการตรวจหาภาษาบนเสียงพูดแต่ละข้อความจากการจำแนกแต่ละระบบ มีกรณีที่เป็นไปได้ของผลลัพธ์การตัดสินใจภาษาของเสียงพูดที่พิจารณาเป็นภาษา L_i หรือไม่ 4 แบบ ดังนี้

- 1) การยอมรับที่ถูกต้อง (Correct acceptance) หรือ ตอบว่า "เสียงพูดเป็นภาษา L_i " ถูกต้อง เกิดขึ้นเมื่อเสียงพูดที่เข้ามาตรงกับภาษา L_i และผลของการตัดสินใจระบุภาษายอมรับว่าเป็นภาษา L_i
- 2) การปฏิเสธที่ถูกต้อง (Correct rejection) หรือ ตอบว่า "เสียงพูดไม่เป็นภาษา L_i " ถูกต้อง เกิดขึ้นเมื่อเสียงพูดที่เข้ามาเป็นภาษาอื่นที่ไม่ใช่ภาษา L_i และผลของการตัดสินใจระบุภาษาปฏิเสธว่าเป็นภาษา L_i
- 3) การปฏิเสธที่ผิดพลาด (False rejection) หรือ ตอบว่า "เสียงพูดไม่เป็นภาษา L_i " ผิดพลาด เกิดขึ้นเมื่อเสียงพูดที่เข้ามาเป็นเสียงพูดภาษา L_i แต่ผลของการตัดสินใจระบุภาษาปฏิเสธว่าเป็นภาษา L_i
- 4) การยอมรับที่ผิดพลาด (False acceptance) หรือ ตอบว่า "เสียงพูดเป็นภาษา L_i " ผิดพลาด เกิดขึ้นเมื่อเสียงพูดที่เข้ามาเป็นภาษาอื่นที่ไม่ใช่ภาษา L_i แต่ผลของการตัดสินใจระบุภาษายอมรับว่าเป็นภาษา L_i

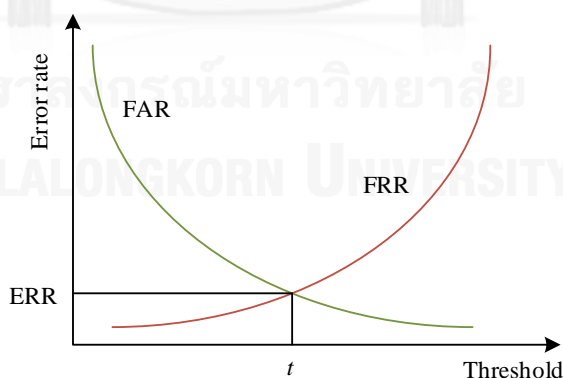
คำตอบของการตัดสินใจจากระบบการรู้จำภาษาพูดย่อยแต่ละแนวทาง ถูกนำมาใช้ในการวัดประสิทธิภาพของระบบ ค่าการยอมรับที่ผิดพลาดทุกภาษาถูกนำมาพิจารณา

3.6.4 การวัดประสิทธิภาพของการรู้จำภาษาด้วยค่า EER

การวัดประสิทธิภาพจากค่าอัตราค่าผิดพลาดระบบที่เท่ากัน (Equal Error Rate: EER) ซึ่งค่า EER คือค่าอัตราความผิดพลาดเมื่อค่าความผิดพลาดจากการยอมรับผิดพลาดและการปฏิเสธผิดพลาดมีค่าเท่ากัน ซึ่งเป็นตำแหน่งที่มีค่าผิดพลาดที่ทำให้ระบบรู้จำมีประสิทธิภาพสูงสุด ค่า EER นี้เป็นวิธีการประเมินที่เป็นที่ยอมรับในการวัดประสิทธิภาพระบบการรู้จำภาษาพูด [111] ความผิดพลาดของระบบการรู้จำภาษาพูด เมื่อมีสมมติฐาน ภาษา L_i เป็นภาษาเป้าหมายที่ใช้ในการพิจารณาคำตอบ การตัดสินใจ สามารถแบ่งได้เป็น 2 กรณีคือ

- 1) อัตราการยอมรับผิดพลาด แทนด้วย FAR เป็นอัตราความผิดพลาดในการตัดสินใจภาษาที่ไม่ใช่ L_i ตอบผิดเป็นภาษา L_i นั่นคือ ความผิดพลาดที่ระบบยอมรับภาษาอื่นที่ไม่ใช่ภาษาที่พิจารณา L_i ตัดสินใจตอบเป็นภาษา L_i ซึ่งคำนวณ FAR จากจำนวนข้อมูลที่ระบบยอมรับภาษาอื่นที่พิจารณาซึ่งเป็นไม่ใช่ภาษา L_i เมื่อคะแนนการตัดสินใจของข้อมูลมีค่ามากกว่าค่าที่กำหนด ทหารด้วยจำนวนข้อมูลทดสอบที่มีคำตอบไม่ใช่ภาษาที่พิจารณา
- 2) อัตราการปฏิเสธผิดพลาด แทนด้วย FRR เป็นอัตราความผิดพลาดในการตัดสินใจภาษาที่เป็นภาษาที่ไม่ใช่ L_i ตอบผิดเป็นภาษาอื่น นั่นคือ ความผิดพลาดที่ระบบปฏิเสธภาษาที่พิจารณาซึ่งเป็นภาษา L_i โดยระบบตัดสินใจตอบเป็นภาษาอื่นที่ไม่ใช่ภาษา L_i ซึ่งคำนวณ FRR จากจำนวนข้อมูลที่ระบบปฏิเสธภาษาที่พิจารณาซึ่งเป็นภาษา L_i เมื่อคะแนนของการตัดสินใจน้อยกว่าค่าที่กำหนด ทหารด้วยจำนวนข้อมูลทดสอบที่มีคำตอบเป็นภาษา

เนื่องจากการเปรียบเทียบประสิทธิภาพของระบบการรู้จำภาษาพูดไม่สามารถโดยใช้ค่าการยอมรับผิดพลาดและการปฏิเสธผิดพลาดได้โดยตรง ค่าผิดพลาดทั้งสองค่ามีการเปลี่ยนแปลงได้ ซึ่งค่าทั้งสองสามารถปรับเปลี่ยนค่าได้ขึ้นอยู่กับวัตถุประสงค์ของการใช้งานระบบ และมุมมองการวัดประสิทธิภาพของระบบการรู้จำภาษาพูด ความสัมพันธ์ของการเปลี่ยนแปลงของค่าทั้งสองคือ ค่าการยอมรับผิดพลาดมีความสัมพันธ์แปรผกผันกับค่าการปฏิเสธผิดพลาด ดังนั้นดังนั้นค่า EER ซึ่งสามารถคำนวณได้จากค่าจุดตัดระหว่างเส้นกราฟของค่า FAR และเส้นกราฟค่า FRR ค่า ภาพที่ 3-4 แสดงเส้นกราฟที่ของค่า FAR (เส้นสีเขียว) และ FRR (เส้นสีแดง) ซึ่งค่า EER ปราบกฏบนแกนค่าความผิดพลาดที่ตำแหน่ง FAR และ FRR ตัดกัน ที่ตำแหน่ง t ค่า EER ของทั้งระบบคือค่าเฉลี่ยของค่า EER แต่ละระบบ



ภาพที่ 3-4 ความสัมพันธ์ระหว่างเส้นกราฟค่า FAR FRR และจุดตัดค่า EER

3.6.5 การวัดประสิทธิภาพของการรู้จำภาษาด้วยฟังก์ชันค่าการตรวจหา

เนื่องจากการพิจารณาเสียงพูดที่ทดสอบแต่ละไฟล์เสียง ประกอบด้วยเสียงทดสอบและภาษาเป้าหมายที่เป็นสมมติฐานของเสียงทดสอบ กำหนดให้ N แทนจำนวนภาษาเป้าหมายทั้งหมดที่พิจารณาและ T แทนจำนวนชุดเสียงที่ทดสอบ ดังนั้นเพื่อวัดประสิทธิภาพของระบบการตรวจหาภาษาที่ตัดสินใจว่า ภาษาเป้าหมายของเสียงพูดที่ทดสอบเป็นภาษาที่ตรงกับสมมติฐานหรือไม่ มีจำนวนกรณีที่วัดผลการตัดสินใจที่ถูกต้องหรือผิดพลาดทั้งหมด $N \times T$ กรณี การวัดประสิทธิภาพนี้จึงเปรียบเทียบการรู้จำภาษาสามารถพิจารณาจากค่าการยอมรับผิดพลาดและการปฏิเสธผิดพลาดร่วมกัน โดยฟังก์ชันค่าการตรวจหา (Detection cost function: DCF)

การวัดประสิทธิภาพของการตรวจหาภาษาซึ่งนิยามตามวิธีการวัดผลของ NIST ที่พิจารณาจากค่าการตรวจหาเฉลี่ย (Average detection cost) แทนด้วย C_{avg} คำนวณดังสมการที่ (3.16)

$$C_{avg} = \frac{1}{N} \sum_{i=1}^N C_{DET}(L_i) \quad (3.16)$$

เมื่อ $C_{DET}(L_i)$ แทนค่าการตรวจหา (Detection Cost) สำหรับซับเซตของชุดเสียงทดสอบ T สำหรับการพิจารณาภาษาเป้าหมาย L_i ซึ่งค่าการตรวจหาคำนวณดังสมการ (3.17) สำหรับแต่ละคู่ภาษา

$$C_{DET}(L_i, L_j) = C_{miss} P_{Target} P_{miss}(L_i) + C_{fa} (1 - P_{Target}) \frac{1}{N-1} P_{fa}(L_i, L_j) \quad (3.17)$$

เมื่อ L_i และ L_j แทนภาษาเป้าหมายและภาษาที่ไม่ใช่ภาษาเป้าหมาย ค่า $P_{miss}(L_i)$ แทนค่าความน่าจะเป็นพลาด (Miss probability) ได้จากค่าการปฏิเสธที่ผิดพลาด ค่า $P_{fa}(L_i, L_j)$ แทนค่าความน่าจะเป็นระบบตอบผิดพลาด (False alarm probability) ได้จากค่าการยอมรับที่ผิดพลาด ซึ่งค่าความน่าจะเป็นคำนวณจากจำนวนของความผิดพลาดนั้นหารด้วยจำนวนของชุดทดสอบและในการวัดประสิทธิภาพการรู้จำภาษากำหนดให้ค่าพารามิเตอร์ C_{miss} มีค่าเป็น 1 ซึ่งมีค่าเท่ากับค่าพารามิเตอร์ C_{fa} และให้ค่าพารามิเตอร์ P_{Target} ซึ่งแทนค่าความน่าจะเป็นที่มีความรู้มาก่อนของภาษาเป้าหมายเป็น 0.5

จากสมการที่ (3.16) ค่าการตรวจหาเฉลี่ย C_{avg} สามารถคำนวณได้ดังสมการ (3.18)

$$C_{avg} = C_{miss} P_{Target} \underbrace{\frac{1}{N} \sum_{i=1}^N P_{miss}(L_i)}_{P_{miss}(\theta_{DET})} + C_{fa} (1 - P_{Target}) \underbrace{\frac{1}{N} \sum_{i=1}^N \left[\frac{1}{N-1} \sum_{j \neq i} P_{fa}(L_i, L_j) \right]}_{P_{fa}(\theta_{DET})} \quad (3.18)$$

เมื่อ $P_{miss}(L_i)$ คำนวณสำหรับแต่ละภาษาเป้าหมายภาษา L_i และค่า $P_{fa}(L_i, L_j)$ คำนวณสำหรับแต่ละคู่การเปรียบเทียบระหว่างภาษาเป้าหมายภาษา L_i และภาษาที่ไม่ใช่ภาษาเป้าหมาย L_j นอกจากนี้สำหรับการวัดประสิทธิภาพของการตัดสินใจแบบถาวรสำหรับการตรวจหาภาษา ซึ่งใช้การคำนวณค่าการตรวจหาเฉลี่ยตามสมการ (3.18) เมื่อคะแนนที่ได้จากการตัดสินใจของระบบ มีค่ามากกว่าค่าอ้างอิงที่กำหนด θ_{DET}

เนื่องจากการวัดประสิทธิภาพของระบบรู้จำขึ้นกับค่าอ้างอิงที่กำหนด θ ซึ่งสามารถเปลี่ยนแปลงตามค่าการดำเนินการที่เป็นไปได้ จึงสามารถแสดงด้วยกราฟซึ่งเป็นการพล็อตระหว่างค่าความน่าจะเป็นพลาด $P_{miss}(\theta)$ และค่าความน่าจะเป็นพลาด $P_{fa}(\theta)$ สำหรับค่าอ้างอิงที่กำหนด θ โดยที่การกำหนดค่าอ้างอิง θ จากการทดสอบระบบการรู้จำเมื่อใช้ข้อมูลชุดฝึกฝนและตั้งค่าอ้างอิงสำหรับจุดดำเนินการที่ให้ค่าการตรวจหาที่มีค่าน้อยที่สุด ซึ่งในการหาค่าอ้างอิงสามารถหาค่าการตรวจหาที่มีค่าน้อยที่สุดจากสมการ (3.11) โดยใช้การวิเคราะห์ความถดถอยแบบพหุคูณ

นอกจากนี้การประเมินประสิทธิภาพของการรู้จำภาษา ซึ่งเป็นการรู้จำรูปแบบหลายกลุ่ม (Multiclass) สามารถใช้ฟังก์ชันวัตถุประสงค์ C_{llr} ดังสมการ (3.19) ซึ่งค่า C_{llr} เป็นฟังก์ชันใช้วัดความดีของคำตอบหรือค่าลือกของโอกาสการเกิดที่ได้จากระบบ เพื่อการหาจุดที่ดีที่สุด (Optimization) ของการหาค่าสัมประสิทธิ์การรวมคะแนนโอกาสการเกิด โดยใช้ข้อมูลชุดฝึกฝน เนื่องจากค่าอ้างอิงที่กำหนด θ การดำเนินการหาสัมประสิทธิ์สำหรับการรวมคะแนนโอกาสการเกิดจึงต้องการหาจุดที่ทำให้มีค่าฟังก์ชันวัตถุประสงค์มีค่าที่ดีที่สุด

$$C_{llr} = -\frac{1}{T \log_2} \sum_L \frac{1}{T_l} \sum_{t=1}^{T_l} \log_2(P'_t) \quad (3.19)$$

เมื่อ T แทนจำนวนชุดเสียงที่ใช้ทดสอบทั้งหมด L แทนเซตของภาษาพูดที่ทดสอบ T_l แทนจำนวนชุดเสียงที่ใช้ทดสอบสำหรับแต่ละภาษา l และ P'_t แทนความน่าจะเป็นภายหลังของระบบทำนายถูกต้องของเสียงทดสอบ t คำนวณจากสมการ (3.20)

$$P'_t = \frac{e^{s_c(t)}}{\sum_{i=1}^N e^{s_i(t)}} \quad (3.20)$$

เมื่อ $s_c(t)$ แทนค่าโอกาสการเกิดสำหรับตัดสินใจภาษาของเสียงทดสอบ t ที่ระบบทำนายถูกต้องและ $s_i(t)$ แทนค่าโอกาสการเกิดสำหรับตัดสินใจภาษาบนเสียงทดสอบ t เมื่อเทียบกับภาษา i สำหรับการพิจารณาค่า C_{llr} ซึ่งมีหน่วยเป็นบิตของข้อมูล ระบบการรู้จำภาษาพูดที่มีค่า C_{llr} ที่มีค่าต่ำแสดงว่ามีประสิทธิภาพดี ตัวอย่างเช่น ค่า C_{llr} มีค่าเท่ากับ 0 หมายถึงระบบการรู้จำมีประสิทธิภาพดีที่สุด ซึ่งได้มาจาก ค่าความน่าจะเป็นภายหลังของระบบทำนายถูกต้องมีค่าเท่ากับ 1 สำหรับเสียงทดสอบทุกเสียง

นอกจากขั้นตอนการทำให้คะแนนเป็นมาตรฐาน (Calibration) และการรวมคะแนนจากระบบรู้จำแต่ละแนวทาง เครื่องมือ Focal multi-class [107] ถูกนำมาใช้ในขั้นตอนการวัดประสิทธิภาพของการรู้จำภาษาด้วยฟังก์ชันค่าการตรวจหา

3.6.6 การวัดประสิทธิภาพด้วยเส้นโค้ง ROC (Receiver Operating Characteristic curve)

การวัดประสิทธิภาพจากเส้นโค้ง ROC ซึ่งแสดงความสัมพันธ์การพล็อตระหว่างค่าอัตราระบบที่ทำนายถูกต้อง (True Positive Rate: TPR) หรือค่าอัตราจริงบวกกับค่าอัตราที่ระบบทำนายผิด (False Positive Rate: FPR) หรือค่าความจดจำโดยค่า TPR และ FPR คำนวณได้จากสมการ (3.21) และ (3.22) ซึ่งค่าอัตราจริงบวกพิจารณาการจำแนกภาษาที่ถูกทดสอบโดยข้อมูลกลุ่มภาษาเป้าหมาย ค่าที่ยิ่งตรงที่มากขึ้นหมายถึงความผิดพลาดที่เกิดขึ้นจากการทดสอบมีน้อย ขณะที่ค่าอัตราเท็จบวกใช้ในการหาภาษาที่ไม่ใช่กลุ่มภาษาเป้าหมายที่พิจารณาที่ทำการทดสอบ เมื่อค่าความจดจำสูง แสดงว่าการระบุภาษาที่ผิดพลาดนั้นมีน้อย

$$TPR = \frac{TP}{(TP+FN)} \quad (3.21)$$

$$FPR = \frac{FP}{(FP+TN)} \quad (3.22)$$

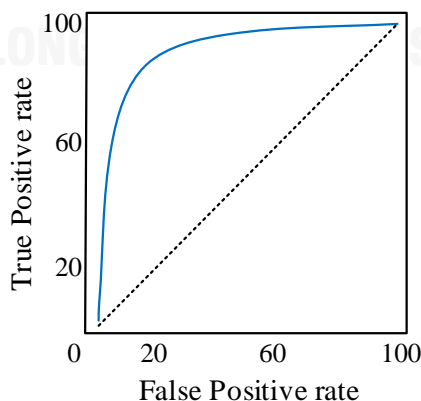
เมื่อ TP (True Positive) คือ จำนวนตัวอย่างที่ทำนายถูกต้องของการเกิดเหตุการณ์

FN (False Positive) คือ จำนวนตัวอย่างที่ทำนายผิดของการเกิดเหตุการณ์

FP (False Negative) คือ จำนวนตัวอย่างที่ทำนายผิดของการไม่เกิดเหตุการณ์

TN (True Negative) คือ จำนวนตัวอย่างที่ทำนายถูกต้องของการไม่เกิดเหตุการณ์

ตัวอย่างเส้นกราฟโค้ง ROC ดังภาพที่ 3-5 ซึ่งแสดงภาพการถ่วงดุลระหว่างอัตราจริงบวกและอัตราเท็จบวก โดยอัตราจริงบวกจะแสดงบนแกน y และอัตราเท็จบวกจะแสดงบนแกน x เส้นกราฟโค้งที่มีคุณสมบัติการทำงานที่ดี จะมีเส้นโค้งใกล้เคียงมุมบนซ้ายมากที่สุด โดยในวิทยานิพนธ์นี้ใช้เส้นโค้ง ROC เป็นเครื่องมือวัดประสิทธิภาพความแม่นยำในการจำแนกภาษา



ภาพที่ 3-5 เส้นโค้ง ROC แสดงความสัมพันธ์ระหว่างอัตราจริงบวกและอัตราเท็จบวก

3.7 การตรวจสอบแบบไขว้

วิทยานิพนธ์นี้ใช้การตรวจแบบไขว้ k ส่วน (Cross validation) เพื่อประเมินความแม่นยำของระบบการรู้จำภาษาพูดและการจำแนกภาษา และเพื่อตรวจสอบค่าความผิดพลาดในการคาดการณ์ของระบบ โดยข้อมูลเสียงจะถูกแบ่งออกเป็น k ส่วนจำนวนเท่า ๆ กัน และทำการคำนวณตรวจสอบจำนวน k รอบ โดยแต่ละรอบข้อมูลซึ่งมีจำนวนเท่ากับ $k-1$ ชุด จะถูกใช้เป็นข้อมูลชุดฝึกฝน และข้อมูลชุดที่เหลืออีกหนึ่งชุดซึ่งได้แก่ชุดที่ k จะถูกเลือกออกมาให้เป็นข้อมูลชุดทดสอบ ค่าความแม่นยำ หรือค่าความผิดพลาดของการตัดสินใจของระบบ จะถูกคำนวณค่าเฉลี่ยจากผลการตัดสินใจจากการตรวจสอบไขว้ k รอบ ในการทดลองนี้ใช้การตรวจแบบไขว้ 5 ส่วนในการประเมินประสิทธิภาพของระบบ



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

บทที่ 4

การรู้จำภาษาตามแนวทางการเรียงหน่วยเสียง

ในบทนี้กล่าวถึงรายละเอียดของระบบการรู้จำภาษาพูดตามแนวทางการเรียงหน่วยเสียง (Phonotactics) ซึ่งประกอบด้วย ส่วนโปรแกรมสองส่วนหลัก คือ ส่วนการสกัดค่าลักษณะสำคัญทางเสียง (Feature extraction) และส่วนแบบจำลองภาษา (Language model) ซึ่งในส่วนของการสกัดค่าลักษณะสำคัญทางเสียง ใช้วิธีการแปลงสัญญาณเสียงให้เป็นหน่วยเสียง ซึ่งในวิทยานิพนธ์นี้หมายถึง ระบบรู้จำเสียง (Phone recognizer) โดยผลลัพธ์ของส่วนโปรแกรมการสกัดค่าลักษณะสำคัญ คือ หน่วยเสียง จากนั้นนำหน่วยเสียงมาสร้างแบบจำลองภาษา เนื้อหาในบทนี้ มีดังต่อไปนี้

- 1) ระบบการรู้จำภาษาพูดโดยใช้ระบบรู้จำเสียงตามด้วยแบบจำลองเวกเตอร์สเปซ (Phone recognizer followed by vector space modeling: PRVSM)
- 2) ระบบการรู้จำภาษาพูดโดยใช้หน่วยเสียงแลตทิซตามด้วยแบบจำลองเวกเตอร์สเปซ (lattice-VSM)

4.1 บทนำ

จากแนวคิดที่ว่าข้อมูลกฎเกณฑ์ของการเรียงหน่วยเสียงในแต่ละภาษามีรูปแบบกฎเกณฑ์แตกต่างกัน จึงสนับสนุนให้มีการใช้ข้อมูลการเรียงหน่วยเสียงมาใช้ในการรู้จำภาษา ระบบการรู้จำภาษาพูดตามแนวทางกฎเกณฑ์ของการเรียงหน่วยเสียงที่ใช้กันอย่างแพร่หลายคือ ระบบการรู้จำภาษาพูดแบบใช้ระบบรู้จำเสียงตามด้วยการใช้แบบจำลองภาษา (Phone recognizer followed by language model: PRLM) ซึ่งวิทยานิพนธ์นี้เรียกระบบการรู้จำภาษาพูดข้างต้นว่า ระบบการรู้จำภาษาพูดแบบฟิวเจอร์แอลเอ็ม แนวคิดของระบบการรู้จำภาษาพูดนี้ ใช้แบบจำลองภาษา ซึ่งการทำงานของระบบการรู้จำภาษาพูด แบ่งเป็น 2 ขั้นตอน ได้แก่ 1) ขั้นตอนการฝึกฝน (Training phase) เสียงจากภาษาเดียวกัน ถูกนำมาสร้างแบบจำลองภาษาแบบการเรียงหน่วยเสียง ซึ่งแบบจำลองของแต่ละภาษาถูกฝึกฝนจากค่าแบบจำลองเอ็นแกรม (N-gram model) จากลำดับหน่วยเสียงที่ปรากฏบนเสียง ค่าแบบจำลองเอ็นแกรมที่ได้จากเสียงในภาษาเดียวกันจะถูกรวมกัน (Interpolate) เพื่อสร้างเป็นตัวแทนของรูปแบบการเรียงหน่วยเสียงของภาษานั้น 2) ขั้นตอนการรู้จำ (Recognition phase) เสียงที่ต้องการรู้จำ ถูกนำมาถอดรหัสหน่วยเสียง แล้วนำหน่วยเสียงที่ได้จากตัวรู้จำเสียงมาสร้างการเรียงของหน่วยเสียง จากนั้นนำไปคำนวณหาคะแนนโอกาสที่จะเป็น (Likelihood) ของแต่ละภาษา คะแนนแต่ละภาษาจะถูกพิจารณาจากค่าคะแนนโอกาสที่คำนวณจากแต่ละภาษา ซึ่งภาษาที่มีค่าสูงสุดถูกพิจารณาเป็นคำตอบของระบบรู้จำ

ระบบการรู้จำภาษาพูดตามแนวทางกฎเกณฑ์ของการเรียงหน่วยเสียงในวิทยานิพนธ์นี้ พิจารณาการใช้ซัพพอร์ตเวกเตอร์แมชชีนเพื่อจำแนกรูปแบบของตัวแทนเสียงที่ปรากฏในแต่ละภาษา ซึ่งเป็นวิธีการจำแนกแบบดิสคริมิเนทีฟ (Discriminative classification) แทนวิธีการเปรียบเทียบจากแบบจำลองภาษาเอ็นแกรมที่เป็นการจำแนกแบบเจเนอเรทีฟ (Generative classification) เพื่อเพิ่มประสิทธิภาพของความสามารถของการจำแนกภาษา นอกจากนี้ตัวแทนเสียงที่ใช้ในการพิจารณา

รูปแบบของตัวแทนเสียงในระบบที่นำเสนอได้ใช้หน่วยเสียงแบบแลตทิซ (Phone lattice) แทนการใช้หนึ่งชุดลำดับหน่วยเสียงที่ได้จากขั้นตอนการถอดรหัสหน่วยเสียง เพื่อลดข้อผิดพลาดที่เกิดจากระบบการรู้จำเสียง

วิทยานิพนธ์นี้เสนอระบบการรู้จำภาษาพูดตามแนวทางการเรียงหน่วยเสียงซึ่งแบ่งเป็น 2 ระบบ ซึ่งทั้งสองระบบเป็นระบบการรู้จำภาษาพูดที่มีประสิทธิภาพสูงและได้รับการยอมรับ (State-of-the-art) ในการรู้จำภาษาตามแนวทางการเรียงหน่วยเสียง ได้แก่ 1) ระบบการรู้จำภาษาพูดแบบใช้ระบบรู้จำเสียงตามด้วยการใช้แบบจำลองเวกเตอร์สเปซ (Phone recognizer followed by vector space modeling: PRVSM) ซึ่งเรียกชื่อในวิทยานิพนธ์นี้ว่า ระบบการรู้จำภาษาพูดแบบพ็ออาร์วีเอสเอ็มและ 2) ระบบการรู้จำภาษาพูดแบบใช้หน่วยเสียงแลตทิซตามด้วยการใช้แบบจำลองเวกเตอร์สเปซ (Lattice-VSM) ซึ่งเรียกชื่อในวิทยานิพนธ์นี้ว่า ระบบการรู้จำภาษาพูดแบบแลตทิซ-วีเอสเอ็ม

ระบบการรู้จำภาษาพูดแบบพ็ออาร์วีเอสเอ็ม ประกอบด้วย ระบบรู้จำเสียงแบบขึ้นกับภาษา (Language-dependent phone recognizer) และแบบจำลองเวกเตอร์สเปซ แบบจำลองเวกเตอร์สเปซได้รับความสนใจมาใช้ในการใช้งานระบบการรู้จำภาษาพูดจากงาน [63] ซึ่งได้รับแนวคิดมาจากการใช้แบบจำลองเวกเตอร์สเปซที่ประสบความสำเร็จในการสืบค้นข้อมูล (Information retrieval) ระบบการรู้จำภาษาพูดแบบพ็ออาร์วีเอสเอ็มมีประสิทธิภาพสูงและได้รับการยอมรับเพื่อใช้เป็นระบบการรู้จำภาษาพูดตามแนวทางกฎเกณฑ์ของการเรียงหน่วยเสียง ความแตกต่างระหว่างระบบการรู้จำภาษาพูดแบบพ็ออาร์วีเอสเอ็มและรู้จำภาษาแบบพ็ออาร์แอลเอ็มคือ ระบบการรู้จำภาษาพูดแบบพ็ออาร์วีเอสเอ็มใช้แบบจำลองเวกเตอร์สเปซเพื่อสร้างตัวแทนของตัวแบบของลำดับหน่วยเสียงที่ปรากฏบนสัญญาณเสียง ในขณะที่ระบบการรู้จำภาษาพูดแบบพ็ออาร์แอลเอ็มใช้แบบจำลองภาษาโดยใช้เอ็นแกรมของหน่วยเสียงเป็นวิธีการสร้างตัวแบบของลำดับหน่วยเสียง ข้อดีของการใช้แบบจำลองเวกเตอร์สเปซคือ รูปแบบของลำดับหน่วยเสียงสามารถสร้างให้อยู่ในเวกเตอร์ที่มีขนาดมิติสูง ซึ่งใช้เวกเตอร์ดังกล่าวมาใช้เป็นอินพุตของตัวจำแนกซัพพอร์ตเวกเตอร์สเปซแมชชีน (Support vector machine: SVM)

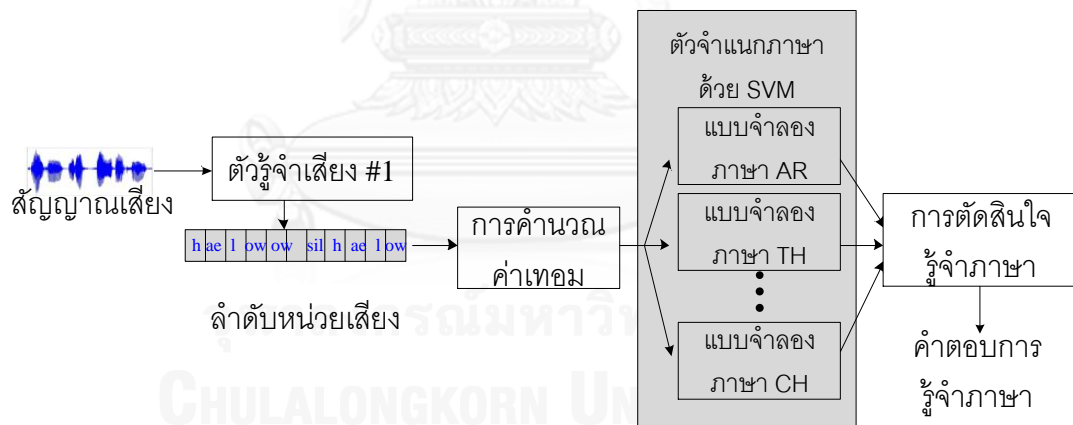
ระบบการรู้จำภาษาพูดแบบแลตทิซ-วีเอสเอ็ม คล้ายกับระบบการรู้จำภาษาพูดแบบพ็ออาร์วีเอสเอ็ม นั่นคือเป็นระบบการรู้จำภาษาพูดที่ประกอบด้วยตัวรู้จำเสียง เพื่อใช้ในการถอดรหัสหน่วยเสียงและแบบจำลองเวกเตอร์สเปซ เพื่อมาสร้างตัวแบบของการแบบจำลองภาษา อย่างไรก็ตาม ความแตกต่างของทั้งสองระบบคือ ระบบการรู้จำภาษาพูดแบบแลตทิซ-วีเอสเอ็ม ใช้ผลลัพธ์ที่ได้จากการรู้จำเสียงเป็นหน่วยเสียงแลตทิซแทนลำดับหน่วยเสียง (Phone sequence) ซึ่งช่วยลดข้อผิดพลาดของการตัดแต่ง (Pruning) เป็นขั้นตอนหนึ่งของการถอดรหัสเสียง ข้อผิดพลาดจากผลลัพธ์ของการรู้จำเสียงส่งผลต่อลำดับของการเรียงตัวของหน่วยเสียงในการรู้จำภาษา ดังนั้นเพื่อป้องกันการเกิดข้อผิดพลาดนี้หน่วยเสียงแลตทิซจึงถูกเสนอมาใช้ในการงานวิจัยนี้

รายละเอียดของระบบการรู้จำภาษาพูดแบบพ็ออาร์วีเอสเอ็มและระบบการรู้จำภาษาพูดแบบแลตทิซ-วีเอสเอ็มอธิบายในหัวข้อถัดไป

4.2 ระบบการรู้จำภาษาพูดแบบใช้ระบบรู้จำเสียงตามด้วยแบบจำลองเวกเตอร์สเปซ

ระบบการรู้จำภาษาพูดแบบพ็อดาร์วีเอ็สเอ็มใช้ข้อมูลกฎเกณฑ์ของการเรียงหน่วยเสียงมาคำนวณค่าสถิติของการปรากฏของแบบจำลองภาษาของหน่วยเสียง จากนั้นนำมาสร้างเป็นเวกเตอร์ค่าลักษณะสำคัญ เพื่อใช้สร้างแบบจำลองภาษา สมมติฐานของแนวทางนี้คือ ภาษาของเสียงพูดสามารถพิจารณาได้จากความถี่ของข้อมูลการเรียงหน่วยเสียง นั่นคือ ภาษาที่มีความถี่การปรากฏของแบบจำลองภาษาของหน่วยเสียงที่ มาก หรือพบลำดับหน่วยเสียงบ่อยครั้ง สามารถแยกแยะจากภาษาที่มีความถี่ของการปรากฏของแบบจำลองภาษาของหน่วยเสียงที่น้อย

ภาพที่ 4-1 แสดงส่วนประกอบหลักของระบบการรู้จำภาษาพูดแบบพ็อดาร์วีเอ็สเอ็ม ซึ่งประกอบด้วย ตัวรู้จำเสียง ส่วนการคำนวณค่าเทอมและแบบจำลองเวกเตอร์สเปซและส่วนการตัดสินใจ โดยในระบบการรู้จำภาษาพูดแบบพ็อดาร์วีเอ็สเอ็ม เสียงพูดจากภาษาเป้าหมายเดียวกันถูกถอดรหัสหน่วยเสียง ด้วยตัวรู้จำเสียง แล้วคำนวณความถี่ของหน่วยเสียงที่ปรากฏบนเสียงพูด จากนั้นนำไปสร้างเวกเตอร์ค่าลักษณะสำคัญ ในส่วนการคำนวณค่าเทอม จากนั้นนำเวกเตอร์ไปใช้ในขั้นตอนการสร้างตัวจำแนกภาษา โดยใช้เวกเตอร์ที่ได้จากข้อมูลเซตฝึกฝนของภาษาเป้าหมายและเวกเตอร์ที่ได้จากภาษาที่ไม่ใช่ภาษาเป้าหมาย นำมาฝึกฝนตัวจำแนกภาษา ซึ่งตัวจำแนกใช้เทคนิคซัพพอร์ตเวกเตอร์สเปซแมชชีน สำหรับขั้นตอนการรู้จำ เสียงที่ต้องการรู้จำนำมาถอดรหัสเป็นหน่วยเสียงและคำนวณความถี่หน่วยเสียงที่ปรากฏบนเสียงพูดนั้นแล้วประกอบเป็นเวกเตอร์ จากนั้นคำนวณค่าคะแนนโอกาสที่คำนวณจากแต่ละภาษาแล้วเลือกตัดสินใจระบุภาษา



ภาพที่ 4-1 องค์ประกอบของระบบการรู้จำภาษาพูดแบบพ็อดาร์วีเอ็สเอ็ม

4.2.1 ตัวรู้จำเสียง

เสียงพูดถูกถอดรหัสเสียงเป็นหน่วยเสียงและลำดับหน่วยเสียง ซึ่งใช้เป็นอินพุทของส่วนการคำนวณค่าเทอมและแบบจำลองเวกเตอร์สเปซ สำหรับตัวรู้จำเสียง (Phone recognizer) ที่ใช้ในวิทยานิพนธ์นี้ ใช้ตัวรู้จำเสียงของ BUT ซึ่งเป็นตัวรู้จำเสียงที่ใช้แนวทางของการใช้เทคนิค ANN/HMM [76] ซึ่งเป็นตัวรู้จำเสียงที่มีประสิทธิภาพและเป็นที่ยอมรับในงานการรู้จำเสียง ตัวรู้จำเสียงนี้ถูกฝึกฝนด้วยฐานข้อมูลเสียงภาษาอังกฤษ ซึ่งมีประกอบด้วย 61 หน่วยเสียง ฐานข้อมูลเสียงที่ใช้ฝึกฝนตัวรู้จำเสียง มาจากผู้พูด 1000 คน มีความยาว 10 ชั่วโมง ผลลัพธ์ที่ได้จากตัวรู้จำเสียงเป็นหน่วยเสียงที่เป็น

ลำดับหน่วยเสียงจากการถอดรหัสที่ดีที่สุด 1 ลำดับ ซึ่งสามารถแสดง ผลการถอดรหัสของเสียงเป็นลำดับหน่วยเสียง $S = \{w_1, w_2, w_3, \dots, w_n\}$ เมื่อ w_i แทน หน่วยเสียงลำดับที่ i ในชุดหน่วยเสียง

นอกจากการพิจารณาจากความแม่นยำของตัวรู้จำเสียง (ซึ่งตัวรู้จำเสียง BUT สูงและเป็นที่ยอมรับ) ความหลากหลายของหน่วยเสียงเป็นอีกหนึ่งตัวช่วยในการเพิ่มประสิทธิภาพในการรู้จำภาษา เนื่องจากหน่วยเสียงมีความสำคัญต่อภาษา ตัวรู้จำเสียงที่ฝึกฝนมาจากฐานข้อมูลเสียงที่ประกอบด้วยจำนวนหน่วยเสียงที่มากและความหลากหลายของทางเสียง จึงถูกพิจารณาเลือกเป็นตัวรู้จำเสียง ดังนั้นตัวรู้จำเสียง BUT ที่ฝึกฝนด้วยภาษาอังกฤษ (ประกอบด้วย 61 หน่วยเสียง) ถูกเลือกในงานวิทยานิพนธ์นี้ เนื่องจากมีจำนวนหน่วยเสียงมากกว่าตัวรู้จำเสียง BUT ที่ฝึกฝนด้วยภาษาอื่น

4.2.2 การคำนวณค่าเทอม

หัวข้อนี้อธิบายการสร้างข้อมูลการเรียงหน่วยเสียง โดยความถี่ของหน่วยเสียงที่เป็นผลลัพธ์จากตัวรู้จำเสียงถูกคำนวณและสร้างแบบจำลองเวกเตอร์สเปซ ซึ่งเรียกเวกเตอร์ที่ประกอบด้วยความถี่ของทุกหน่วยเสียงที่ปรากฏบนเสียงพูด ว่า ซุปเปอร์เวกเตอร์ (Supervector) มีค่าถ่วงน้ำหนักที่คำนวณจากฟังก์ชันถ่วงน้ำหนักค่าเทอม (Term-weighting function) ซึ่งฟังก์ชันนี้มีหน้าที่ให้คะแนนและถ่วงน้ำหนักของความสามารถในการบอกความแตกต่างระหว่างภาษา ของแต่ละหน่วยเสียง รายละเอียดของฟังก์ชันการถ่วงน้ำหนักเทอมอธิบายในหัวข้อ 4.4

4.2.3 แบบจำลองเวกเตอร์สเปซ

ลำดับหน่วยเสียงที่เป็นผลลัพธ์จากตัวรู้จำเสียง ถูกคำนวณความถี่ของหน่วยเสียงและประกอบกันเป็นเวกเตอร์ลักษณะสำคัญ แนวคิดของแบบจำลองเวกเตอร์สเปซ (Vector space modeling paradigm) ถูกใช้และประสบความสำเร็จในการสืบค้นข้อมูล (Information retrieval) นั่นคือ การใช้เวกเตอร์เพื่อเป็นตัวแทนเอกสารข้อความ (Text document) โดยมีจำนวนของการปรากฏของคำในเอกสาร เป็นสมาชิกองค์ประกอบของเวกเตอร์ ดังนั้น ในการรู้จำภาษา ใช้เวกเตอร์เพื่อแสดงค่าสถิติของหน่วยเสียงที่ปรากฏบนเสียงพูด

ข้อมูลการเรียงหน่วยเสียงได้รับจากแบบจำลองภาษาของหน่วยเสียงที่มีลำดับสูง ให้ผลลัพธ์ของการจำแนกภาษาที่มีประสิทธิภาพสูง [64] โดยแบบจำลองภาษาของหน่วยเสียงที่มีลำดับสูง ส่งผลให้รูปแบบของข้อมูลการเรียงหน่วยเสียงมีจำนวนมากขึ้นและให้ข้อมูลสำหรับเงื่อนไขของคำศัพท์เพิ่มขึ้น ในวิทยานิพนธ์นี้ ใช้แบบจำลองภาษาของหน่วยเสียง ในระดับแบบยูนิแกรม (Unigram) แบบไบแกรม (Bigram) และแบบไตรแกรม (Trigram) เพื่อแสดงการปรากฏของหน่วยเสียงในแต่ละภาษา ตัวอย่างเช่น ลำดับหน่วยเสียงที่ได้จากตัวรู้จำเสียง คือ $S = \{w_1, w_2, w_3, \dots, w_n\}$ เมื่อ w_i แทนหน่วยเสียงที่ i ของลำดับหน่วยเสียง แบบจำลองภาษาของหน่วยเสียงระดับไบแกรม สามารถแสดงเป็น $\#w_1 w_2 w_2 w_3 w_3 w_4$ และ $w_4 \#$ แบบจำลองภาษาของหน่วยเสียงระดับไตรแกรมสามารถแสดงเป็น $\#w_1 w_2, w_1 w_2 w_3, w_2 w_3 w_4$ และ $w_3 w_4 \#$, เมื่อ $\#$ แทนที่ด้วยบริบทว่าง

ลำดับหน่วยเสียงที่ได้จากตัวรู้จำเสียง ถูกแปลงให้อยู่ในรูปของเวกเตอร์ V ดังสมการที่ (4.1) แสดงความถี่ของการปรากฏของเอ็นแกรม

$$V = [v_1 \quad v_2 \quad \dots \quad v_n] \quad (4.1)$$

เมื่อ v_n แสดงเวกเตอร์ของค่าสถิติเอ็นแกรมลำดับที่ n ของหน่วยเสียง ซึ่งมีสมาชิกแต่ละตัวคือ เวกเตอร์ v ประกอบด้วยสมาชิกของค่าสถิติเอ็นแกรมของหน่วยเสียง ดังสมการ (4.2) ที่แสดงเวกเตอร์ของไบแกรม v_2 ซึ่งประกอบด้วย ความถี่ของการปรากฏของไบแกรมของหน่วยเสียง

$$v_2 = [P_{(1,2)} \quad P_{(2,3)} \quad P_{(i-1,i)}] \quad (4.2)$$

กำหนดให้ w_i แทนหน่วยเสียงที่ i ของลำดับหน่วยเสียงที่ได้จากการถอดรหัสเสียงและแต่ละองค์ประกอบ $P_{(i-1,i)}$ แทนค่าความน่าจะเป็นแบบร่วม (Joint probability) ของไบแกรมของหน่วยเสียง $w_{i-1}w_i$ ในลำดับหน่วยเสียง เมื่อให้ N แทนจำนวนของหน่วยเสียงที่ได้จากตัวรู้จำเสียง ดังนั้นค่าความน่าจะเป็นแบบร่วม สามารถคำนวณได้จาก:

$$P_{(i-1,i)} = \frac{\text{count}(w_{i-1}, w_i | S)}{\sum_{j=1}^N \text{count}(w_{j-1}, w_j | S)} \quad (4.3)$$

และเวกเตอร์ของไตรแกรม v_3 มีสมาชิกประกอบด้วย $v_3 = [P_{(1,2,3)} \quad P_{(2,3,4)} \quad P_{(i-2,i-1,i)}]$ ค่าโดยสมาชิกแต่ละตัวของเวกเตอร์ $P_{(i-2,i-1,i)}$ สามารถคำนวณได้จากความน่าจะเป็นแบบร่วมของไตรแกรมของหน่วยเสียงที่ปรากฏบนเสียงพูด

ตัวอย่างเช่น ผลลัพธ์ของตัวรู้จำเสียงมี k หน่วยเสียง เวกเตอร์ค่าลักษณะสำคัญ V ประกอบด้วยรูปแบบของหน่วยเสียงและแบบจำลองภาษาของหน่วยเสียงที่ปรากฏทั้งหมดในภาษาดังนั้นขนาดมิติของเวกเตอร์ค่าลักษณะสำคัญสามารถเขียนอยู่ในรูปของ k หน่วยเสียง รวมไบแกรม $k \times k$ และไตรแกรมของหน่วยเสียง $k \times k \times k$ ที่ปรากฏบนเสียงพูด ในงานวิจัยนี้จำนวนหน่วยเสียงที่เป็นผลลัพธ์ของตัวรู้จำเสียง BUT มีจำนวน 61 หน่วยเสียง ดังนั้นขนาดมิติของซูปเปอร์เวกเตอร์ที่ใช้ในระบบการรู้จำภาษาพูดตามแนวทางนี้ มีค่าเท่ากับ $61 + (61 \times 61) + (61 \times 61 \times 61) = 230,673$ ดังนั้นผลลัพธ์ของขั้นตอนนี้คือ เวกเตอร์ค่าลักษณะสำคัญซึ่งประกอบด้วยแบบจำลองภาษาของหน่วยเสียงแบบยูนิแกรม แบบไบแกรม และแบบไตรแกรม โดยเรียกเวกเตอร์ค่าลักษณะสำคัญนี้ว่า ซูปเปอร์เวกเตอร์ ซึ่งใช้เป็นอินพุตในการฝึกฝนตัวจำแนกภาษาในหัวข้อ 4.2.4

4.2.4 ซัพพอร์ตเวกเตอร์แมชชีน

ซัพพอร์ตเวกเตอร์แมชชีนถูกใช้มาเป็นตัวจำแนกภาษาไบนารี (Binary classifier) นั่นคือ ให้ทำนายคำตอบภายใน 2 กลุ่มเป้าหมาย โดยพิจารณาคำตอบระหว่างกลุ่มเป้าหมายและกลุ่มที่ไม่ใช่เป้าหมาย โดยแบบจำลองของการจำแนกเป็นแบบหนึ่งต่อต่อทั้งหมด (One-versus-all) ในขั้นตอนการฝึกฝนตัวจำแนกภาษา ใช้เวกเตอร์ค่าลักษณะสำคัญที่ได้จากเสียงของภาษาเป้าหมายหนึ่งภาษาให้เป็นที่กลุ่มตัวอย่างบวกและเวกเตอร์ค่าลักษณะสำคัญที่ได้จากเสียงพูดที่ไม่ใช่ภาษาเป้าหมายทุกภาษาเป็นที่กลุ่มตัวอย่างลบ

ซัพพอร์ตเวกเตอร์แมชชีนเป็นตัวจำแนกแบบไบนารี (binary classifier) สามารถนิยามได้ดังนี้:

$$f(x) = \sum_{i=1}^N \alpha_i K(x, x_i) + b, \quad (4.4)$$

ให้ x เป็นอินพุทเวกเตอร์และ x_i เป็นเวกเตอร์ซัพพอร์ต โดยที่ $\alpha_i > 0$ และ $K(x, x_i)$ เป็นเคอร์เนลฟังก์ชัน (kernel function) ที่วัดค่าความเหมือน (similarity) ระหว่าง x และ x_i

โดย $p(t_i | d_j)$ แทนค่าความน่าจะเป็นของสมาชิก t_i ที่เป็นองค์ประกอบในเวกเตอร์ ลักษณะสำคัญของเสียงพูด d_j ดังนั้นสามารถซัพพอร์ตเวกเตอร์ในรูปของเวกเตอร์สเปซได้ดังนี้:

$$\Phi(d_j) = [p(t_1 | d_j) \quad p(t_2 | d_j) \quad \dots \quad p(t_M | d_j)] \quad (4.5)$$

เมื่อ M เป็นจำนวนสมาชิกของเวกเตอร์ลักษณะสำคัญ และเคอร์เนลฟังก์ชันเพื่อวัดความเหมือนระหว่างสองเวกเตอร์ของเสียงพูดระหว่าง d_1 และ d_2 นิยามด้วยการทำอินเนอร์โปรดักต์ (inner product) ดังสมการ (4.6):

$$K(d_1, d_2) = \langle \Phi(d_1), \Phi(d_2) \rangle = \sum_{i=1}^M p(t_i | d_1) p(t_i | d_2) \quad (4.6)$$

เนื่องจากประสิทธิภาพของการจำแนกขึ้นอยู่กับเคอร์เนลฟังก์ชันซึ่งได้มาจากค่าของข้อมูลของตัวอย่างกลุ่มบวกและลบที่ใช้ในการสร้างเคอร์เนลฟังก์ชัน เพื่อเพิ่มประสิทธิภาพของการจำแนกของเคอร์เนลฟังก์ชัน เวกเตอร์ลักษณะสำคัญจึงถูกนำไปคำนวณด้วยฟังก์ชันถ่วงน้ำหนักค่าเทอม $w(t_i)$ เพื่อให้คะแนนความสามารถของการบอกความแตกต่างของภาษาของสมาชิกในเวกเตอร์ ซึ่งค่า $w(t_i)$ ถูกคำนวณแตกต่างกันขึ้นกับความสามารถในการจำแนกของสมาชิกแต่ละตัวในเวกเตอร์ลักษณะสำคัญ ดังนั้นเคอร์เนลฟังก์ชันเมื่อใช้ฟังก์ชันถ่วงน้ำหนักค่าเทอมแสดงดังสมการ (4.7)

$$\tilde{K}(d_1, d_2) = \sum_{i=1}^M w(t_i)^2 p(t_i | d_1) p(t_i | d_2) \quad (4.7)$$

ในวิทยานิพนธ์นี้ใช้ฟังก์ชันการจำแนกจากเคอร์เนลฟังก์ชันแบบเชิงเส้นด้วยเครื่องมือ SVM^{light} [112]

4.2.5 ส่วนการตัดสินใจระบุภาษา

ในการตัดสินใจของการระบุภาษาของเสียงพูด ใช้เอาต์พุตของตัวจำแนกภาษาเพื่อสร้างเป็นเวกเตอร์แบบดิสคริมิเนทีฟ (Discriminative vector) ซึ่งมีขนาดมิติ เท่ากับจำนวนของตัวจำแนกภาษา ผู้วิจัยใช้การทดสอบสมมติฐาน (Hypothesis test) ในการตัดสินใจของการระบุภาษา ระบุตรวจหาภาษาสร้างจากแบบจำลองเกาส์เซียน โดยมีแบบจำลองเกาส์เซียนสำหรับภาษาเป้าหมาย (แทนกลุ่มบวก) และสำหรับแบบจำลองสำหรับภาษาที่ไม่ใช่ภาษาเป้าหมาย (แทนกลุ่มลบ) โดยมี I^+ แทนแบบจำลองกลุ่มบวกซึ่งได้มาจากการฝึกฝนจากเวกเตอร์แบบดิสคริมิเนทีฟของภาษาเป้าหมาย และมี I^- แทนแบบจำลองกลุ่มลบซึ่งได้มาจากการฝึกฝนจากเวกเตอร์แบบดิสคริมิเนทีฟของภาษาที่ไม่ใช่เป้าหมาย เราพิจารณาการตัดสินใจระบุภาษาจากค่าอัตราส่วนของค่าล็อกโอกาสการเกิด (Log-likelihood) λ ระหว่างภาษาเป้าหมายและภาษาที่ไม่ใช่ภาษาเป้าหมาย คะแนนความเชื่อมั่นของการรู้จำภาษาของเสียงพูด O พิจารณาโดยใช้การทดสอบสมมติฐานจากแบบจำลองเบย์เซียน

(Bayesian) ให้สมมติฐาน H_0 แทนเสียงพูด O ที่ตัดสินใจระบุภาษาเป็นภาษาเป้าหมาย I^+ และให้สมมติฐาน H_1 แทนเสียงพูด O ที่ตัดสินใจระบุภาษาเป็นภาษาที่ไม่ใช่เป้าหมาย I^- สมมติฐานเป็นจริง เมื่ออัตราส่วนของโอกาสการเกิดแบบล็อก λ ของเสียงพูด O (คำนวณความน่าจะเป็นจากเวกเตอร์แบบดิสคริมีเนทีฟของเสียงพูด O เทียบกับแบบจำลองภาษาเป้าหมาย I^+ กับความน่าจะเป็นจากเวกเตอร์แบบดิสคริมีเนทีฟของเสียงพูด O เทียบกับแบบจำลองภาษาที่ไม่ใช่ภาษาเป้าหมาย I^- มีค่ามากกว่าระดับเกณฑ์ที่กำหนด โดยการคำนวณอัตราส่วนของโอกาสการเกิดแบบล็อก λ ของเสียงพูด O แสดงดังสมการที่ (4.8)

$$\lambda = \log \frac{P(O|I^+)}{P(O|I^-)} = \log P(O|I^+) - \log P(O|I^-) \quad (4.8)$$

4.3 ระบบการรู้จำภาษาพูดแบบใช้หน่วยเสียงแลตทิซตามด้วยแบบจำลองเวกเตอร์สเปซ

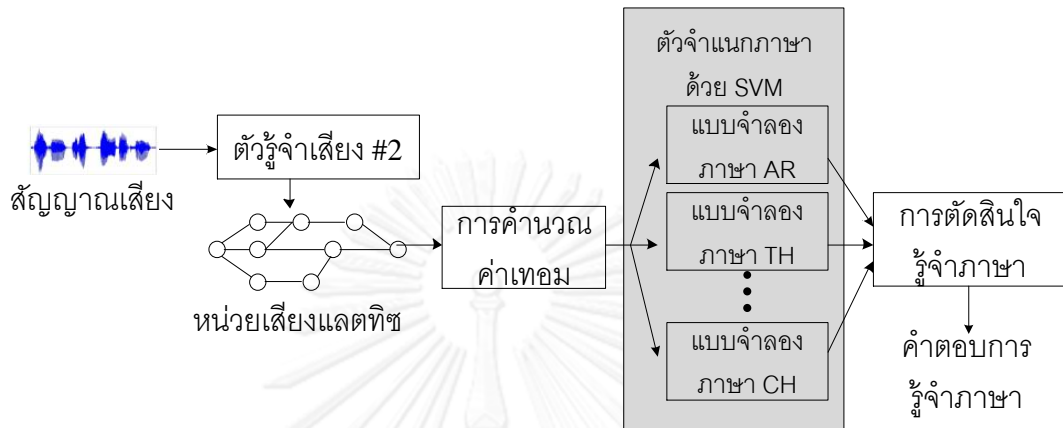
แนวคิดของระบบการรู้จำภาษาพูดแบบแลตทิซ-วีเอ็สเอ็ม คล้ายกับระบบการรู้จำภาษาพูดแบบพ็อรวีเอ็สเอ็ม นั่นคือใช้ความถี่ของการปรากฏของหน่วยเสียงเพื่อเป็นข้อมูลของกฎของการเรียงหน่วยเสียงที่มีความแตกต่างกันระหว่างภาษา ความแตกต่างกันระหว่างสองระบบ คือ ระบบการรู้จำภาษาพูดแบบแลตทิซ-วีเอ็สเอ็ม ใช้หน่วยเสียงแลตทิซ แทนลำดับหน่วยเสียง ดังนั้นค่าลักษณะสำคัญของ การเรียงหน่วยเสียงของแนวทางนี้ได้มาจากผลรวมทั้งหมดของความน่าจะเป็นจากแบบจำลองเอ็นแกรมของหน่วยเสียงแลตทิซที่อยู่บนเสียงพูด

สมมติฐานของแนวทางนี้คือ ลำดับหน่วยเสียงที่เป็นผลลัพธ์จากการตัดแต่งกิ่งในขั้นตอนการถอดรหัสเสียงของตัวรู้จำเสียงอาจจะเกิดข้อผิดพลาดจากการตัดแต่งกิ่งได้ (นั่นคือ ชุดลำดับหน่วยเสียงที่มีคะแนนความมั่นใจต่ำกว่าค่าเกณฑ์ที่กำหนดจะถูกลบทิ้งไป ทำให้การตัดแต่งชุดสายของลำดับหน่วยเสียงตัดหรือลบชุดลำดับที่มีโอกาสจะเกิดเป็นชุดลำดับหน่วยเสียงที่ถูกต้องของการรู้จำเสียงได้ ทำให้ผลลัพธ์ที่ได้ของการถอดรหัสหน่วยเสียงเป็นชุดสายของลำดับหน่วยเสียงของผลการรู้จำเสียงที่ดีที่สุด 1 ชุด อาจจะเกิดข้อผิดพลาดจากการตัดแต่งกิ่งได้) จากผลลัพธ์ที่ให้ลำดับหน่วยเสียงที่ผิดพลาดส่งผลให้นำหน่วยเสียงไปสร้างต้นแบบของการเรียงตัวหน่วยเสียงที่ผิดพลาดได้ ดังนั้นเพื่อลดข้อผิดพลาดดังกล่าว การใช้หน่วยเสียงแลตทิซทั้งหมด (ที่รวมชุดสายของลำดับหน่วยเสียงไว้ทั้งหมด) จึงถูกเสนอมาใช้เป็นตัวแทนของผลลัพธ์ของตัวรู้จำเสียงในระบบการรู้จำภาษาพูด

ภาพที่ 4-2 แสดงองค์ประกอบของระบบการรู้จำภาษาพูดแบบแลตทิซวีเอ็สเอ็ม ประกอบด้วย ตัวรู้จำเสียงที่ให้ผลลัพธ์ที่ได้จากตัวรู้จำเสียงเป็นหน่วยเสียงแลตทิซ ส่วนการคำนวณค่าเทอม และส่วนการตัดสินใจ ซึ่งแตกต่างกับระบบการรู้จำภาษาพูดแบบพ็อรวีเอ็สเอ็มดังแสดงในภาพที่ 4-1

ระบบการรู้จำภาษาพูด ประกอบด้วย ตัวรู้จำเสียง การคำนวณค่าเทอม และตัวจำแนกภาษา โดยใช้ซัพพอร์ตเวกเตอร์แมชชีน เสียงพูดที่เป็นอินพุตถูกแปลงหน่วยเสียงแลตทิซ แล้วนำไปคำนวณความถี่ของหน่วยเสียงแลตทิซที่ปรากฏบนเสียงพูด แล้วประกอบกันเป็นเวกเตอร์ ซึ่งเรียกว่า ซุปเปอร์เวกเตอร์ (ประกอบด้วย ความน่าจะเป็นของแบบจำลองภาษาหน่วยเสียงแลตทิซ) ในขั้นตอนการ

ฝึกฝน ซุปเปอร์เวกเตอร์คำนวณจากเสียงพูดของภาษาเป้าหมายและเสียงพูดจากภาษาที่ไม่ใช่ภาษาเป้าหมาย ถูกใช้มาสร้างตัวจำแนกภาษาจากซัพพอร์ตเวกเตอร์แมชชีน แล้วในขั้นตอนการรู้จำภาษา ซุปเปอร์เวกเตอร์ของเสียงทดสอบถูกนำมาเปรียบเทียบกับตัวจำแนกภาษาที่ได้จากการขั้นตอนเรียนรู้

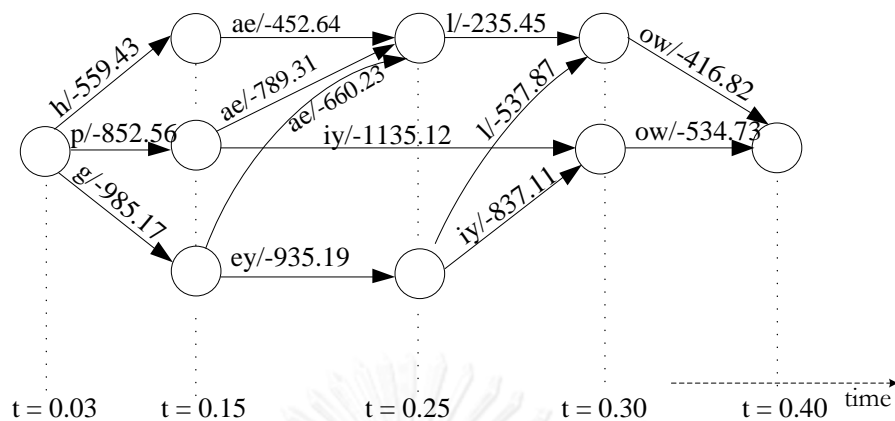


ภาพที่ 4-2 องค์ประกอบของระบบการรู้จำภาษาพูดแบบแลตทิซวีเอสเอ็ม

4.3.1 หน่วยเสียงแลตทิซ

เนื่องจากการรู้จำเสียงซึ่งให้ผลลัพธ์แบบชุดลำดับหน่วยเสียงที่ดีที่สุด 1 ลำดับ (1-Best phone sequence) เป็นผลลัพธ์ที่มีข้อผิดพลาดจากขั้นตอนการถอดรหัสหน่วยเสียง งานวิจัยด้านการรู้จำเสียงจึงเสนอผลลัพธ์ทางเลือกของขั้นตอนการถอดรหัสเสียง โดยให้ผลลัพธ์ที่ได้ของการถอดรหัสเสียงเป็นหน่วยเสียงแลตทิซ ซึ่งผลลัพธ์ดังกล่าวสามารถลดข้อผิดพลาดที่เกิดในการหาผลลัพธ์ที่ดีที่สุดจากการถอดรหัสหน่วยเสียงได้

หน่วยเสียงแลตทิซ คือกราฟที่มีทิศทางแต่ละโหนด (Node) มีความสัมพันธ์กับเฟรมของสัญญาณเสียงพูด มีเส้นเชื่อม (Edge) ระหว่างโหนดที่สัมพันธ์กับคะแนนลักษณะทางเสียง (Acoustic score) เพื่อสะท้อนความมั่นใจของโอกาสการเกิดหน่วยเสียงแต่ละหน่วยเสียงที่แตกต่างกัน หน่วยเสียงแลตทิซที่เป็นผลลัพธ์การรู้จำเสียงได้รวมวิถี (Path) ของชุดลำดับหน่วยเสียงหลายชุดสายและแต่ละวิถีของชุดสายลำดับหน่วยเสียง มีค่าความน่าจะเป็นของคะแนนทางเสียงที่แสดงบนแต่ละเส้นเชื่อมระหว่างหน่วยเสียง ตลอดบนทุกชุดสายลำดับหน่วยเสียง ภาพที่ 4-3 แสดงตัวอย่างของหน่วยเสียงแลตทิซของคำว่า "HELLO" ซึ่งถอดความเป็นหน่วยเสียง "h ae l ow" จากภาพแสดงว่า หน่วยเสียงเดียวกันสามารถสร้างวิถีของชุดสายลำดับหน่วยเสียงได้หลายวิถี ในแต่ละวิถีที่มีเส้นเชื่อมระหว่างหน่วยเสียง หน่วยเสียงที่อยู่บนในแต่ละวิถีไม่ขึ้นต่อกัน หน่วยเสียงที่อยู่ในตำแหน่งที่แตกต่างกัน (ถึงแม้จะเป็นหน่วยเสียงเดียวกัน) มีค่าความน่าจะเป็นของคะแนนทางเสียงของโอกาสการเกิดหน่วยเสียงแต่ละหน่วยเสียงแตกต่างกัน ผลรวมของค่าโอกาสการเกิดหน่วยเสียงจากทุกหน่วยเสียงบนแต่ละวิถีแสดงคะแนนของการให้ผลลัพธ์ของการถอดรหัสหน่วยเสียง



ภาพที่ 4-3 แสดงตัวอย่างของหน่วยเสียงแลตทิซของคำว่า “HELLO”

ข้อมูลเกณฑ์การเรียงหน่วยเสียงแบบใช้หน่วยเสียงแลตทิซสามารถคำนวณได้จากผลรวมของคะแนนโอกาสการเกิดหน่วยเสียงและการเกิดหน่วยเสียงร่วมกัน (นั่นคือ คะแนนทางเสียงที่ปรากฏบนแต่ละเส้นเชื่อม จากแต่ละวิถีของชุดสายลำดับหน่วยเสียง) ที่แสดงบนทุกวิถีของสายลำดับหน่วยเสียงแลตทิซ

4.3.2 ตัวรู้จำเสียง

ผู้วิจัยใช้ตัวรู้จำเสียงแบบเดียวกับที่ใช้ในการรู้จำภาษาแบบพ็อดาร์วีเอสเอ็ม (ตามหัวข้อ 4.2.1) เพื่อสร้างผลลัพธ์ของการรู้จำเสียงเป็นหน่วยเสียงแลตทิซ โดยใช้เครื่องมือ HTK HMM toolkit [113] เพื่อการสร้างผลลัพธ์จากตัวรู้จำเสียงให้เป็นหน่วยเสียงแลตทิซ แล้วนำหน่วยเสียงแลตทิซไปคำนวณค่าความถี่คาดหวัง (Expected count) ในส่วนโปรแกรมการคำนวณค่าเทอมต่อไป

4.3.3 การคำนวณค่าเทอม

ขั้นตอนนี้เป็นการสร้างแบบจำลองเวกเตอร์สเปซ เพื่อเตรียมเวกเตอร์เป็นอินพุทของตัวจำแนกภาษา ผู้วิจัยสร้างเวกเตอร์โดยใช้วิธีคล้ายกับแนวทางระบบการรู้จำภาษาพูดแบบพ็อดาร์วีเอสเอ็ม เนื่องจากผลลัพธ์ของตัวรู้จำเสียงเป็นหน่วยเสียงแลตทิซ จึงใช้ค่าความถี่คาดหวังของหน่วยเสียงแลตทิซเป็นตัวแทนของข้อมูลการเรียงหน่วยเสียง เพื่อใช้เป็นสมาชิกของเวกเตอร์ลักษณะสำคัญแทนการใช้ค่าความถี่ของหน่วยเสียง

การคำนวณค่าความถี่คาดหวังของหน่วยเสียง มาจากผลรวมของความน่าจะเป็นของแบบจำลองเอ็นแกรมของหน่วยเสียงแลตทิซทั้งหมดที่ปรากฏบนเสียงพูด ซึ่งได้จากโอกาสการเกิดหน่วยเสียงบนแต่ละวิถีของชุดลำดับหน่วยเสียงที่ปรากฏบนเสียงพูด กำหนดให้ลำดับหน่วยเสียงที่ได้จากตัวรู้จำเสียง ในแต่ละวิถีของชุดสายลำดับหน่วยเสียงเป็น $S = \{w_1, w_2, w_3, \dots, w_n\}$ เมื่อ w_i แทนหน่วยเสียงลำดับที่ i ของวิถีของชุดลำดับหน่วยเสียง การคำนวณค่าความถี่คาดหวังสามารถแสดงอยู่ในรูปของฟังก์ชันความถี่ (count function) ของลำดับหน่วยเสียงแทนด้วย $C(w_1, \dots, w_n)$ ซึ่งฟังก์ชันค่าความถี่คาดหวังของหน่วยเสียงแลตทิซดังสมการที่ (4.9):

$$E[C(w_1, \dots, w_n) | X, \theta] \approx \sum_{w(e_i)=w_i} P(e_1, \dots, e_N | L) \quad (4.9)$$

เมื่อ X แทนส่วนเสียงพูด, θ แทนเซตของแบบจำลองทางเสียง, L แทนหน่วยเสียงแลตทิซ และค่าความน่าจะเป็นแบบภายหลัง (Posterior probability) คำนวณได้จากอัลกอริทึมแบบไปข้างหน้าและถอยหลัง (Forward-backward algorithm) [64] ดังนี้:

$$P(e_1, \dots, e_N | L) = \alpha(e_1) \beta(e_N) \prod_i \xi(e_i) \quad (4.10)$$

เมื่อ $\alpha(e)$ แทนค่าโอกาสการเกิดแบบไปข้างหน้าของโหนดเริ่มต้นของเส้นเชื่อม e และมี $\beta(e)$ แทนค่าโอกาสการเกิดแบบถอยหลังของโหนดสิ้นสุดของเส้นเชื่อม e มี $\xi(e)$ แทนความน่าจะเป็นแบบภายหลังของเส้นเชื่อม e การประมาณความน่าจะเป็นของแบบจำลองทางภาษาสามารถวนซ้ำเพื่อคำนวณใหม่ได้จากคะแนนความน่าจะเป็นภายหลังโดยใช้อัลกอริทึมการคาดหมายสูงสุด (Expectation maximization: EM)

ในการเตรียมเวกเตอร์ ค่าความถี่คาดหวังของหน่วยเสียงแลตทิซที่คำนวณได้ จะนำมาสร้างเป็นสมาชิกของเวกเตอร์ เวกเตอร์ที่ได้จะนำมาคำนวณด้วยฟังก์ชันการถ่วงน้ำหนักเทอม เพื่อนำไปใช้เป็นข้อมูลรับเข้าของตัวจำแนกภาษา รายละเอียดของตัวจำแนกได้อธิบายในหัวข้อ 4.2.4 จากนั้นเอาที่พูดของตัวจำแนกภาษาซึ่งเป็นคะแนนความมั่นใจของการระบุภาษานำไปใช้เตรียมเวกเตอร์แบบดิสคริมิเนทีฟ เพื่อใช้ในการตัดสินใจของการระบุภาษา

4.3.4 การตัดสินใจระบุภาษา

การตัดสินใจของการระบุภาษา ของระบบการรู้จำภาษาพูดแบบแลตทิซใช้วิธีเดียวกับระบบการรู้จำภาษาพูดแบบพ็อดาร์วีเอสเอ็ม โดยอธิบายรายละเอียดของการทำงานในหัวข้อ 4.2.5 โดยอินพุตของการตัดสินใจระบุภาษาใช้เวกเตอร์แบบดิสคริมิเนทีฟ ที่ได้จากผลลัพธ์ของตัวจำแนกภาษา จากนั้นนำมาคำนวณค่าอัตราส่วนของโอกาสการเกิดแบบล็อก (Log-likelihood) ระหว่างภาษาเป้าหมายและภาษาที่ไม่ใช่ภาษาเป้าหมายแล้วตัดสินใจคำตอบของการรู้จำภาษา จากคะแนนโอกาสการเกิดที่มีค่าสูง

4.4 ฟังก์ชันการถ่วงน้ำหนักเทอม

หัวข้อนี้อธิบายฟังก์ชันการถ่วงน้ำหนัก ซึ่งมีหน้าที่ในการคัดเลือก หรือให้ความสำคัญกับเทอมหรือตัวแทนของหน่วยเสียงที่อยู่ในเวกเตอร์ของเสียงพูด ซึ่งแต่ละเทอมมีความสำคัญและแสดงลักษณะเฉพาะของภาษาแตกต่างกัน เวกเตอร์ของเสียงพูดนี้เป็นผลลัพธ์มาจากขั้นตอนการสร้างแบบจำลองเวกเตอร์สเปซ โดยเวกเตอร์แสดงความถี่ของลำดับหน่วยเสียง (จากหัวข้อที่ 4.2.3) และเวกเตอร์แสดงความถี่คาดหวัง (จากหัวข้อ 4.3.3) เวกเตอร์จะถูกนำมาคำนวณด้วยฟังก์ชันการถ่วงน้ำหนัก ก่อนนำไปใช้เป็นอินพุตของตัวจำแนกภาษาด้วยซอฟต์แวร์เวกเตอร์แมชชีน

ฟังก์ชันถ่วงน้ำหนักจะให้คะแนนความสามารถของการบอกความแตกต่างของภาษา ในแต่ละเทอมหรือหน่วยเสียง ซึ่งสมาชิกของเวกเตอร์ คือ เทอม (Term) แทนลำดับหน่วยเสียงที่พิจารณา

หรือ หน่วยเสียงแลตทิซที่ปรากฏบนเสียงพูด ดังนั้น ค่าของแต่ละเทอม แสดงค่าความถี่ของเทอม หรือความถี่คาดหวังของเทอมที่ปรากฏบนเสียงพูด โดยแต่ละเทอมมีค่าที่แตกต่างกัน เนื่องจากในงานวิจัยนี้ใช้ตัวจำแนกซัพพอร์ตเวกเตอร์แมชชีนซึ่งความสามารถของการจำแนกระหว่างสองกลุ่ม ขึ้นกับใช้คอร์เนลฟังก์ชัน ดังนั้นเพื่อสร้างฟังก์ชันคอร์เนลที่มีความสามารถในการจำแนกได้ดี อินพุทที่ใช้สร้างคอร์เนลฟังก์ชันจึงมีผลกระทบต่อความสามารถในการแบ่งแยกระหว่างกลุ่ม ตัวอย่างเช่น เมื่อคอร์เนลฟังก์ชันสร้างจากเวกเตอร์ที่ประกอบด้วยค่าความน่าจะเป็นของเทอมหนึ่งสูงมากและค่าความน่าจะเป็นของอีกเทอมมีค่าต่ำ ซึ่งค่าของสมาชิกแตกต่างกันมาก สมาชิกของเวกเตอร์ที่มีค่าความน่าจะเป็นสูงมีอิทธิพลต่อการสร้างคอร์เนลฟังก์ชัน ทำให้ความสามารถของการแบ่งระหว่างกลุ่มจากฟังก์ชันของระนาบของฟังก์ชันสร้างจากสมาชิกของเวกเตอร์ที่มีค่าน้อยมีความสามารถด้อยลง ดังนั้นเพื่อป้องกันการมีอิทธิพลของค่าเทอมต่อการสร้างคอร์เนลฟังก์ชัน การนำฟังก์ชันการถ่วงน้ำหนักเทอมมาใช้ร่วมกับเวกเตอร์จึงมีความสำคัญ เพื่อทำให้เป็นบรรทัดฐานหรือการนอร์มอลไลซ์ (Normalize) ค่าของเวกเตอร์ และให้คะแนนความสามารถในการจำแนกของแต่ละเทอมที่เป็นสมาชิกของเวกเตอร์

เนื่องจากฟังก์ชันการถ่วงน้ำหนักเทอมสามารถหาได้จากการวิเคราะห์ความหมายแอบแฝง (Latent Semantic Analysis: LSI) โดยพิจารณาจากการปรากฏร่วมของคำต่าง ๆ ซึ่งฟังก์ชันไม่ได้ทำเพียงการนอร์มอลไลซ์ค่า แต่ทำการรวมแบบจำลองเชิงความหมายเทอมในกรอบยาว (Long-term semantic) หรือพิจารณาตามภาพใหญ่ รวมกับแบบจำลองของเทอมในกรอบสั้นหรือพิจารณาแบบภาพเล็ก

ในหัวข้อนี้ผู้วิจัยได้ศึกษาฟังก์ชันการถ่วงน้ำหนักเทอมที่ถูกเสนอในงานการสืบค้นข้อมูล [114] เปรียบเทียบประสิทธิภาพของฟังก์ชันการถ่วงน้ำหนักเทอมแบบต่าง ๆ นอกจากนี้ในหัวข้อนี้ผู้วิจัยเสนอฟังก์ชันการถ่วงน้ำหนักเทอม และเปรียบเทียบประสิทธิภาพของ ระบบการรู้จำภาษาพูดตามแนวทางกฎเกณฑ์ของการเรียงหน่วยเสียง (ได้แก่ ระบบการรู้จำภาษาพูดแบบฟิอาร์วีเอ็สเอ็มและระบบการรู้จำภาษาพูดแบบแลตทิซเอ็สเอ็ม) เมื่อใช้ฟังก์ชันการถ่วงน้ำหนักเทอมที่แตกต่างกัน

4.4.1 ฟังก์ชันการถ่วงน้ำหนักเทอม

งานวิจัย [62, 115] กล่าวว่า การถ่วงน้ำหนักเทอมมีความสำคัญต่อประสิทธิภาพของการจำแนก เนื่องจากมีการทำนอร์มอลไลซ์ซูปเปอร์เวกเตอร์ของค่าความน่าจะเป็นของแบบจำลองเอ็นแกรม ทำให้ฟังก์ชันคอร์เนลที่ได้ไม่ได้รับผลอิทธิพลจากองค์ประกอบของเวกเตอร์ที่มีค่าความน่าจะเป็นที่มีค่ามาก ซึ่งองค์ประกอบเหล่านั้นมีจำนวนน้อย

ผู้วิจัยได้ศึกษาฟังก์ชันการถ่วงน้ำหนักเทอมที่ใช้แพร่หลายในงานการสืบค้นข้อมูล ซึ่งประกอบด้วย: ค่าความถี่เทอม (Term frequency: tf), ความถี่เอกสารผกผัน (Inverse document frequency: idf), ค่าความเกี่ยวข้องกันของเทอม (Term relevance: tr), ค่าสถิติไคสแควร์ (Chi-square statistic: chi), ค่าความซ้ำซ้อนของเทอม (Redundancy of term frequency: rd) และการประกอบกันของแต่ละฟังก์ชัน โดยฟังก์ชันเหล่านี้ได้มาจากการวิเคราะห์ความหมายแอบแฝง ซึ่ง

เป็นการรวมคะแนนจากแบบจำลองความหมายเทอมแบบยาว ร่วมกับคะแนนจากแบบจำลองของเทอมแบบสั้น [114]

- 1) ค่าความถี่เทอม (Term frequency: tf) : ความถี่เทอมสามารถแสดงน้ำหนักแบบท้องถิ่น (local) เนื่องจากค่านี้มาจากการความถี่เทอม t_i ภายในเสียงพูด d_j ดังนั้นค่าความถี่เทอมสามารถแทนด้วย $tf(t_i, d_j)$ โดย $tf(t_i, d_j)$ คือความถี่ของหน่วยเสียงหรือชุดลำดับหน่วยเสียงที่ปรากฏในเสียงพูด สามารถอธิบายค่าลักษณะเฉพาะของภาษา โดยค่าความถี่ของเทอมสะท้อนความสำคัญของเทอมนั้น นอกจากนี้ค่าความถี่เทอมสามารถแสดงในรูปแบบต่าง ๆ เช่น ล็อกการิทึม (logarithm) ของความถี่เทอมแทนด้วย $\log tf(t_i, d_j)$ แสดงดังสมการที่ (4.11) ฟังก์ชันล็อกการิทึมทำให้ไม่มีค่าน้ำหนักที่มีค่าสูงมาก เมื่อเปรียบเทียบกับน้ำหนักของเทอมที่ปรากฏบนเสียงพูดที่มีความถี่การปรากฏครั้งหนึ่งของเสียงพูด ขณะที่ค่าความถี่แบบผกผัน (Inverse term frequency: itf) แทนด้วย $ITF(t_i, d_j)$ ซึ่งเป็นค่าที่ได้จากความหลากหลายของค่าความถี่เทอม สามารถคำนวณค่านี้ตามสมการที่ (4.12)

$$\log tf(t_i, d_j) = \log(tf(t_i, d_j) + 1) \quad (4.11)$$

$$ITF(t_i, d_j) = 1 - 1 / (1 + tf(t_i, d_j)) \quad (4.12)$$

- 2) ความถี่เอกสารผกผัน (Inverse document frequency: idf): ความถี่ของจำนวนเสียงพูดที่มีเทอมนั้นปรากฏ เปรียบเทียบกับจำนวนเสียงพูดในฐานข้อมูลเสียงทั้งหมด โดยค่าความถี่เอกสารผกผันแทนด้วย idf ใช้เพื่อลดความสำคัญของเทอมที่พบบ่อยในเสียงพูดหลายเสียงพูดในฐานข้อมูลและใช้เพิ่มความสำคัญของเทอมที่ไม่ค่อยพบในฐานข้อมูล ค่านี้สามารถนิยามเป็นตัวแทนของน้ำหนักแบบครอบคลุม (Global) เพื่อสะท้อนความเฉพาะเจาะจงของเทอม t_i ต่อจำนวนเสียงพูดทั้งหมดที่อยู่ในฐานข้อมูลเสียง ซึ่งค่าความถี่เอกสารผกผันสามารถคำนวณได้ดังสมการที่ (4.13) โดยค่า N เป็นจำนวนเสียงพูดทั้งหมดที่อยู่ในฐานข้อมูลและค่า $f(t_i)$ คือจำนวนเสียงพูดที่มีเทอม t_i ปรากฏในเสียงพูด สำหรับงานการรู้จำภาษา เทอมที่ปรากฏในเสียงพูดน้อยครั้ง แสดงว่ามีความสำคัญมากกว่า เทอมที่ปรากฏในทุกเสียงพูดที่อยู่ในฐานข้อมูล

$$idf(t_i) = \log\left(\frac{N}{f(t_i)}\right) \quad (4.13)$$

เนื่องจากค่าน้ำหนัก idf สะท้อนความสำคัญของเทอมเมื่อเปรียบเทียบในฐานข้อมูลเสียงทั้งหมด ผู้วิจัยจึงเสนอ ค่าถ่วงน้ำหนักที่พิจารณาเทอมเปรียบเทียบภายในกลุ่มเดียวกัน (Intra-class) แทนการพิจารณาเทอมเปรียบเทียบกับฐานข้อมูลเสียงทั้งหมดแทนด้วย idf_cate เพื่อพิจารณาค่าลักษณะของเทอมที่ปรากฏในกลุ่มเดียวกัน ดังสมการที่ (4.14) เมื่อพิจารณาเทอมในภาษาเดียวกัน ให้ค่า c_k แทนภาษาเดียวกัน ค่า

N_{c_k} แทนจำนวนหน่วยเสียงในภาษาที่พิจารณา c_k และค่า $f(t_i, c_k)$ แทนจำนวนความถี่ของเทอมในภาษา c_k

$$idf_cate(t_i, c_k) = \log\left(\frac{N_{c_k}}{f(t_i, c_k)}\right) \quad (4.14)$$

- 3) ค่าความเกี่ยวข้องกันของเทอม (Term relevance: tr): ค่าน้ำหนักค่าความเกี่ยวข้องกันของเทอมแสดงการถ่วงน้ำหนักระหว่างกลุ่ม (Inter-class) โดยแสดงสัดส่วนของจำนวนเสียงพูดที่มีเทอมปรากฏกับจำนวนเสียงพูดที่ไม่มีความเกี่ยวข้องกันกับเทอม (นั่นคือจำนวนเสียงพูดที่ไม่มีเทอมบนเสียงพูด) คำนวณดังสมการ (4.15) มีค่า N เป็นจำนวนเสียงพูดทั้งหมดที่อยู่ในฐานข้อมูลเสียง และค่า $f(t_i)$ คือจำนวนเสียงพูดที่มีเทอม t_i ปรากฏ

$$tr(t_i) = \log\left(\frac{N - f(t_i)}{f(t_i)}\right) \quad (4.15)$$

- 4) ค่าสถิติไคสแควร์ (Chi-square statistic: chi): ค่าถ่วงน้ำหนักแบบสถิติไคสแควร์พิจารณาทั้งการปรากฏของเทอมในเสียงพูดของแต่ละภาษาและเทอมที่ไม่ปรากฏในเสียงพูดของแต่ละภาษา ซึ่งสามารถคำนวณได้ดังสมการที่ (4.16) โดยที่ $a = f(t_i, c_k)$ และ $b = f(t_i, \tilde{c}_k)$ แทนความถี่ของเทอม t_i ที่ปรากฏในกลุ่มภาษาเป้าหมาย c_k และในกลุ่มที่ไม่ใช่ภาษาเป้าหมาย \tilde{c}_k ตามลำดับ กำหนดให้ \tilde{t}_i แทนด้วยเทอมความถี่ของเทอมอื่น ๆ ที่ไม่ใช่เทอม t_i ดังนั้นค่า $c = f(\tilde{t}_i, c_k)$ และ $d = f(\tilde{t}_i, \tilde{c}_k)$ แทนความถี่ของเทอม \tilde{t}_i ที่ปรากฏในภาษาเป้าหมาย c_k และในภาษาที่ไม่ใช่เป้าหมาย \tilde{c}_k ตามลำดับ โดยที่ N คือจำนวนความถี่ทั้งหมดที่เทอม t_i ปรากฏในเสียงพูดที่อยู่ในฐานข้อมูลเสียงทั้งหมด

$$chi(t_i) = \frac{N(ad - cb)^2}{[(a + c)(b + d)(a + b)(c + d)]} \quad (4.16)$$

ข้อเสียของค่าถ่วงน้ำหนักนี้ในการรู้จำภาษาคือ เป็นการถ่วงน้ำหนักที่ไม่เหมาะสมในกรณีที่จำนวนของเสียงพูดของภาษาที่ไม่ใช่ภาษาเป้าหมาย มีจำนวนมากกว่า เสียงพูดของภาษาเป้าหมาย

- 5) ค่าความซ้ำซ้อนของเทอม (Redundancy of term frequency: rd): การถ่วงน้ำหนักแบบค่าความซ้ำซ้อนของเทอม สามารถแสดงได้ดังสมการที่ (4.17) โดยที่ N เป็นจำนวนเสียงพูดในฐานข้อมูลเสียง มีค่า $tf(t_i, d_j)$ แสดงค่าความถี่ของเทอม t_i ปรากฏภายในเสียงพูด d_j และค่า $tfsum(t_i)$ แทนผลรวมความถี่ของเทอม t_i ที่ปรากฏในทุกเสียงพูดที่อยู่ในฐานข้อมูลเสียง เมื่อ j แทนลำดับของเสียงพูดในฐานข้อมูลเสียง และ i แทนลำดับเทอมที่พิจารณา ซึ่งคำนวณดังสมการที่ (4.18)

$$rd(t_i) = \log N + \sum_{i=1}^N \frac{tf(t_i, d_j)}{tfsun(t_i)} \log \left(\frac{tf(t_i, d_j)}{tfsun(t_i)} \right) \quad (4.17)$$

$$tfsun(t_i) = \sum_{j=1}^N tf(t_i, d_j) \quad (4.18)$$

ข้อดีของการถ่วงน้ำหนักแบบนี้คือ สามารถวัดการกระจายของเทอมที่ปรากฏในแต่ละเสียงพูด และคำนวณค่าความถี่เทอมในแต่ละเสียงพูดโดยตรง ผู้วิจัยเสนอการถ่วงน้ำหนักนี้เพื่อแก้ข้อเสียของการถ่วงน้ำหนักวิธีแบบ *idf* (ซึ่งนับเพียงจำนวนเสียงพูดที่มีเทอม t_i ปรากฏบนเสียงพูดเท่านั้น แต่ไม่ได้พิจารณาค่าความถี่เทอมบนแต่ละเสียงพูด) ถ้าค่าถ่วงน้ำหนักมาก แสดงว่าเทอมมีความสำคัญน้อย ผู้วิจัยใช้ค่าถ่วงน้ำหนักนี้เพื่อศึกษาผลกระทบของความซ้ำซ้อนของเทอม (หรือหน่วยเสียงที่พิจารณา) ในเสียงพูดสำหรับแต่ละภาษา

4.4.2 การทำให้เป็นบรรทัดฐานหรือนอร์มอลไลซ์

ความยาวของเสียงพูดส่งผลต่อความหลากหลายของแบบจำลองของตัวจำแนก ความยาวของเสียงพูดสามารถจัดการได้ด้วยการทำนอร์มอลไลซ์ (Nomalization) ฟังก์ชันนอร์มอลไลซ์ด้วยวิธีผลรวม (Summation) ซึ่งแทนด้วย $w_k(t_i).normS'$ และฟังก์ชันนอร์มอลไลซ์ด้วยวิธีระยะห่างแบบยูคลิเดียน (Euclidean distance) $w_k(t_i).normE'$ สามารถนิยามได้ดังสมการที่ (4.19) และ (4.20) ตามลำดับ

$$w_k(t_i).normS' = \frac{w_k(t_i)}{\sum_{i=1}^T w_k(t_i)} \quad (4.19)$$

$$w_k(t_i).normE' = \frac{w_k(t_i)}{\sqrt{\sum_{i=1}^T (w_k(t_i))^2}} \quad (4.20)$$

4.4.3 การประกอบกันของแต่ละฟังก์ชัน

ฟังก์ชันการถ่วงน้ำหนักที่ดี ควรประกอบกันระหว่างค่าถ่วงน้ำหนักแบบท้องถิ่น (ซึ่งแสดงลักษณะเฉพาะเจาะจงของเทอมที่ปรากฏในเสียงพูด) และค่าถ่วงน้ำหนักแบบครอบคลุม (ซึ่งแสดงความสามารถในการจำแนกภาษาระหว่างภาษา และเปรียบเทียบความสำคัญของเทอมที่พิจารณาเทียบกับเทอมอื่น) ด้วยความหลากหลายของการประกอบกันของฟังก์ชันการถ่วงน้ำหนัก ฟังก์ชันถ่วงน้ำหนักที่ประกอบกันจากค่าฟังก์ชันย่อย ๆ จึงถูกนำมาใช้ในการรู้จำภาษาและเปรียบเทียบแต่ละฟังก์ชัน ซึ่งตารางที่ 4-1 ได้รวบรวมฟังก์ชันถ่วงน้ำหนักแบบต่าง ๆ นั่นคือ สามอันดับแรกเป็นการประกอบกันของความถี่เทอม tf ร่วมกับ idf รวมทั้งความหลากหลายของการรวมของฟังก์ชัน ได้แก่ $tf.idf$, $\log tf.idf$ และ $ITF.idf$ ซึ่งเป็นการประกอบกันของค่าถ่วงน้ำหนักแบบท้องถิ่นและค่าถ่วงน้ำหนักแบบครอบคลุมทั้งหมด นอกจากนี้การประกอบกันของฟังก์ชันอื่นที่ไม่ได้เป็นการค่าถ่วงน้ำหนักแบบท้องถิ่นและค่าถ่วงน้ำหนักแบบครอบคลุมทั้งหมด เช่น ฟังก์ชัน $\log tf.ITF$ (อันดับที่ 4 ซึ่ง

แสดงค่าการถ่วงน้ำหนักแบบท้องถิ่น) ฟังก์ชัน $tf.idf_cate$, ฟังก์ชัน $tf.tr$ และ ฟังก์ชัน $tf.chi$ (อันดับที่ 5, 6, และ 7 ตามลำดับ ซึ่งแสดงค่าการถ่วงน้ำหนักแบบครอบคลุมทั้งหมด) และสองอันดับสุดท้ายของตารางแสดงการประกอบกันของฟังก์ชัน tf หรือ $logtf$ ร่วมกับ rd

ตารางที่ 4-1 การประกอบกันของการถ่วงน้ำหนักแบบต่าง ๆ

ลำดับที่	การประกอบกันของค่าถ่วงน้ำหนัก	คำอธิบาย
1	$tf.idf$	$w_{i,j} = tf(t_i, d_j) \bullet idf(t_i)$
2	$logtf.idf$	$w_{i,j} = \log(tf(t_i, d_j) + 1) \bullet idf(t_i)$
3	$ITF.idf$	$w_{i,j} = ITF(t_i, d_j) \bullet idf(t_i)$
4	$logtf.ITF$	$w_{i,j} = \log(tf(t_i, d_j) + 1) \bullet ITF(t_i, d_j)$
5	$tf.idf_cate$	$w_{i,j,k} = tf(t_i, d_j) \bullet idf_cate(t_i, c_k)$
6	$tf.tr$	$w_{i,j} = tf(t_i, d_j) \bullet tr(t_i)$
7	$tf.chi$	$w_{i,j} = tf(t_i, d_j) \bullet chi(t_i)$
8	$tf.rd$	$w_{i,j} = tf(t_i, d_j) \bullet rd(t_i)$
9	$logtf.rd$	$w_{i,j} = \log(tf(t_i, d_j) + 1) \bullet rd(t_i)$

4.5 บทสรุป

ในบทนี้ได้เสนอระบบการรู้จำภาษาพูดตามแนวทางกฎเกณฑ์ของการเรียงหน่วยเสียง ซึ่งใช้ร่วมกับระบบการรู้จำภาษาพูดแบบพ็อดาร์วีเอ็สเอ็มและระบบการรู้จำภาษาพูดแบบแลตทิซวีเอ็สเอ็ม ความแตกต่างระหว่างสองระบบคือ การใช้หน่วยเสียง (สำหรับระบบการรู้จำภาษาพูดแบบพ็อดาร์วีเอ็สเอ็ม) และใช้หน่วยเสียงแบบแลตทิซ (สำหรับระบบการรู้จำภาษาพูดแบบแลตทิซวีเอ็สเอ็ม) เพื่อเป็นตัวแทนของเทอมที่ใช้พิจารณาการจำแนกภาษา นอกจากนี้ผู้วิจัยได้ศึกษาการใช้ฟังก์ชันถ่วงน้ำหนักแบบต่าง ๆ เพื่อใช้เพิ่มความสามารถของการจำแนกภาษาและเสนอการประกอบกันของฟังก์ชันถ่วงน้ำหนัก และเพื่อเปรียบเทียบประสิทธิภาพของระบบการรู้จำภาษาพูดเมื่อใช้ฟังก์ชันถ่วงน้ำหนักแบบต่าง ๆ ในบทที่ 6 ต่อไป

บทที่ 5

การรู้จำภาษาโดยอาศัยค่าลักษณะสำคัญเกี่ยวกับระบบเสียงในภาษา

ในบทนี้เสนอระบบการรู้จำภาษาพูดโดยอาศัยค่าลักษณะสำคัญเกี่ยวกับระบบเสียงในภาษา (Phonological Features: PF) ซึ่งในวิทยานิพนธ์นี้เรียกว่า ค่าพีเอฟ เนื้อหาในบทนี้ประกอบด้วย หัวข้อที่ 5.1 ที่มาของการใช้ค่าพีเอฟในระบบการรู้จำภาษาพูด หัวข้อที่ 5.2 ระบบตรวจหา ลักษณะเฉพาะค่าพีเอฟโดยใช้โครงข่ายประสาทเทียมแบบเพอร์เซพตรอนหลายชั้นและการศึกษาการ กระจายของค่าพีเอฟในแต่ละภาษา หัวข้อที่ 5.3 และ 5.4 อธิบายระบบการรู้จำภาษาพูดที่นำเสนอใน วิทยานิพนธ์นี้ ซึ่งแบ่งเป็น 2 แนวทาง ได้แก่

- 1) ระบบการรู้จำภาษาพูดโดยใช้การจำแนกด้วยแบบจำลองลาเทนท์ไดนามิกคอนดิชันนอล แรนดอมฟิลด์ร่วมกับค่าพีเอฟ
- 2) ระบบการรู้จำภาษาพูดตามแนวทางกฎเกณฑ์ของการเรียงตัวของค่าพีเอฟ ซึ่ง รายละเอียดแสดงตามลำดับดังนี้

5.1 ที่มาของการใช้ค่าพีเอฟในระบบการรู้จำภาษาพูด

วิทยานิพนธ์นี้เสนอการใช้ค่าพีเอฟแทนการใช้ค่าลักษณะสำคัญทางเสียงค่าอื่น ๆ ที่ใช้กันใน วรรณกรรมงานการรู้จำภาษา เนื่องจากค่าพีเอฟสามารถอธิบายค่าพารามิเตอร์การออกเสียงต่าง ๆ ซึ่งเป็นคุณสมบัติพื้นฐานของหน่วยเสียงในแต่ละภาษา ดังนั้นค่าพีเอฟจึงถูกนำมาใช้เพื่อแสดง คุณสมบัติค่าการออกเสียงในแต่ละภาษาที่มีลักษณะที่แตกต่างกัน นอกจากนี้เพื่ออธิบายคุณสมบัติ ของค่าพีเอฟที่นำไปใช้ในการรู้จำภาษา จึงอธิบายที่มาของการนำค่าพีเอฟไปใช้ในการรู้จำภาษาแต่ละ แนวทาง อธิบายดังนี้

- 1) ที่มาของการนำค่าพีเอฟไปใช้ในระบบการรู้จำภาษาพูดโดยใช้การจำแนกด้วย แบบจำลองลาเทนท์ไดนามิกคอนดิชันนอลแรนดอมฟิลด์ร่วมกับค่าพีเอฟ มาจาก การศึกษาความแตกต่างของการกระจายค่าพีเอฟในแต่ละภาษา และมาจากรวมกรรม การรู้จำภาษา แนวทางใช้ค่าลักษณะทางเสียงหรือการรู้จำภาษาตามแนวทางอะคูสติกที่ มีตัวระบุภาษา สร้างขึ้นโดยใช้แบบจำลองของการกระจายค่าลักษณะสำคัญทางเสียงใน แต่ละภาษา

จากการศึกษาความแตกต่างของการกระจายของค่าพีเอฟในแต่ละภาษา ผู้วิจัยจึง ต้องการสร้างตัวจำแนกที่ใช้แบบจำลองที่มาจากค่าพีเอฟแทนค่าลักษณะสำคัญทางเสียง ค่าอื่นในการสร้างตัวจำแนกเพื่องานการรู้จำภาษา แต่เนื่องจากคุณสมบัติของการเปลี่ยน สถานะ (Transition) ของค่าการออกเสียงหนึ่งไปยังอีกค่าหนึ่งในแต่ละภาษามีรูปแบบ แตกต่างกัน นั่นคือ การเปลี่ยนแปลงค่าพีเอฟบนหน่วยเสียงในแต่ละภาษามี ลักษณะเฉพาะแตกต่างกัน ผู้วิจัยจึงสนใจนำแบบจำลองลาเทนท์ไดนามิกคอนดิชันนอล แรนดอมฟิลด์มาใช้ในการจำแนก เพื่อพิจารณาจับความสัมพันธ์ระหว่างคุณสมบัติต่าง ๆ ในข้อมูลค่าพีเอฟด้วย

- 2) การนำค่าพีเอฟไปใช้ในแนวทางการรู้จำภาษาด้วยแนวทางการเรียงตัวของค่าพีเอฟ มาจากคุณสมบัติของการเรียงตัวของหน่วยเสียงในแต่ละภาษามีรูปแบบการเรียงตัวของหน่วยเสียงที่แตกต่างกันทำให้รูปแบบการเรียงตัวของค่าพีเอฟแตกต่างกันในแต่ละภาษา และมาจากความแปรปรวนของค่าพีเอฟในแต่ละหน่วยเสียงที่แตกต่างกันระหว่างภาษา

เนื่องจากเกณฑ์การเรียงหน่วยเสียงมีความเฉพาะเจาะจงกับภาษา นั่นคือ ลำดับหน่วยเสียงที่ปรากฏในแต่ละภาษามีรูปแบบแตกต่างกัน เช่น บางภาษาอนุญาตให้มีเสียงของพยัญชนะประสมภายในพยางค์เดียวกัน เช่น ภาษาไทยมีเสียงพยัญชนะควบกล้ำ ซึ่งเป็นพยัญชนะประสม 2 ตัวที่ประกอบด้วย เสียงกัก (Stop consonant) ปรากฏร่วมกับเสียง ร /r/ เสียง ล /l/ หรือเสียง ว /w/ เสียงพยัญชนะควบกล้ำรูปแบบนี้ไม่ปรากฏในบางภาษา ขณะที่ภาษาญี่ปุ่นมีเสียงพยัญชนะประสม 2 ตัว ที่เกิดจากเสียงพยัญชนะที่ปรากฏร่วมกับเสียง ย /j/ แต่ในภาษาอังกฤษยอมให้มีเสียงพยัญชนะประสม 3 ตัว เช่น /spl/ ที่เกิดจากการปรากฏร่วมกันของเสียง /s/ /p/ และ /l/ แต่ในขณะที่บางภาษา เช่น ภาษาจีนไม่มีเสียงพยัญชนะประสม แต่อนุญาตให้มีเสียงสระประสมได้ ตัวอย่างของเสียงสระประสม ได้แก่ เสียงสระประสม /aw/ ซึ่งเกิดจากการเปลี่ยนแปลงของพารามิเตอร์การออกเสียงในช่องทางเดินเสียง ที่เปลี่ยนจากการออกเสียงสระ /a/ ไปเป็นเสียงเลื่อน /w/ ซึ่งการประกอบกันของเสียงสระประสมนี้ เกิดขึ้นในภาษาไทย ในขณะที่ไม่สามารถมีสระประสมนี้ในภาษารัสเซีย เพราะภาษารัสเซียไม่อนุญาตให้มีเสียงสระประสมที่ลงท้ายด้วยเสียง /w/ เป็นต้น จากกฎเกณฑ์การเรียงตัวของหน่วยเสียงที่แตกต่างกันในแต่ละภาษา ทำให้มีรูปแบบการเปลี่ยนแปลงค่าพีเอฟที่แตกต่างกัน

นอกจากนี้เนื่องจากความแตกต่างกันของการแปรปรวนของค่าพีเอฟบนหน่วยเสียงในแต่ละภาษา เมื่อหาความสัมพันธ์ของค่าพีเอฟชนิดต่าง ๆ บนหน่วยเสียงเดียวกันในแต่ละภาษา จึงมีรูปแบบของการประกอบกันของค่าพีเอฟชนิดต่าง ๆ ที่แตกต่างกัน เมื่อพิจารณาการเรียงตัวของค่าพีเอฟบนทั้งเสียงพูดที่ประกอบด้วยหน่วยเสียงต่าง ๆ เปลี่ยนแปลงจากหน่วยเสียงหนึ่งไปยังอีกหน่วยเสียง ทำให้การเรียงตัวของค่าพีเอฟชนิดต่าง ๆ ในแต่ละภาษามีรูปแบบแตกต่างกัน จึงเป็นที่มาของแนวคิดในการหารูปแบบเหตุการณ์การเกิดขึ้นร่วมกันของค่าพีเอฟ เพื่อใช้ในระบบการรู้จำภาษาพูดตามแนวทางที่สองของบทนี้ โดยพิจารณาความถี่ของการปรากฏของรูปแบบเหตุการณ์การเกิดขึ้นร่วมกันของค่าพีเอฟที่มีความถี่การปรากฏที่ไม่เท่ากันในแต่ละภาษา เพื่อจำแนกความแตกต่างกันระหว่างภาษาได้

5.2 ระบบตรวจหาลักษณะเฉพาะค่าพีเอฟโดยโครงข่ายประสาทเทียมหลายชั้น

วิทยานิพนธ์นี้นำเสนอระบบตรวจหาลักษณะเฉพาะค่าพีเอฟด้วยโครงข่ายประสาทเทียมหลายชั้น (Multilayer Perceptron: MLP) เพื่อสกัดค่าพีเอฟจากสัญญาณเสียงพูด ซึ่งขั้นตอนการสร้างของระบบตรวจหาลักษณะเฉพาะค่าพีเอฟ ประกอบด้วย 3 ขั้นตอน ดังนี้

ขั้นตอนที่ 1) การเตรียมข้อมูลสำหรับฝึกฝนระบบตรวจหาลักษณะเฉพาะค่าพีเอฟ รวมขั้นตอนการสกัดค่าลักษณะสำคัญทางเสียงจากเสียงพูดในฐานข้อมูลเสียงส่วนที่ใช้สำหรับการฝึกฝนซึ่งมีฉลากหน่วยเสียง และการเตรียมฉลากค่าพีเอฟซึ่งได้จากการแปลงฉลากหน่วยเสียงที่กำกับสัญญาณเสียงพูดเป็นค่าพีเอฟ ฉลากค่าพีเอฟที่ได้จากการถอดความหน่วยเสียงประกอบด้วยค่าพีเอฟ 13 ชนิด ซึ่งนิยามตาม SPE (อธิบายไว้ในบทที่ 2) โดยไม่รวมค่าพีเอฟเสียงเงียบ

ขั้นตอนที่ 2) การฝึกฝนระบบตรวจหาลักษณะเฉพาะค่าพีเอฟ เนื่องจากค่าพีเอฟที่ใช้ในวิทยานิพนธ์นี้มี 13 ชนิด เพื่อสร้างระบบตรวจหาลักษณะเฉพาะสำหรับค่าพีเอฟแต่ละชนิด ในขั้นตอนนี้จึงสร้างระบบตรวจหาค่าพีเอฟแต่ละชนิดย่อย โดยนำค่าลักษณะสำคัญทางเสียงและฉลากค่าพีเอฟแต่ละชนิดมาฝึกฝนระบบตรวจหาลักษณะเฉพาะค่าพีเอฟแต่ละชนิดโดยใช้โครงข่ายประสาทเทียมแบบเพอร์เซพตรอนหลายชั้น หรือเรียกอีกอย่างหนึ่งว่า การฝึกฝนแต่ละระบบตรวจหาย่อยโดยแยกกันที่ละชนิดของค่าพีเอฟ (รายละเอียดของขั้นตอนการฝึกฝนระบบตรวจหาลักษณะเฉพาะค่าพีเอฟอธิบายในบทที่ 6)

ขั้นตอนที่ 3) การวัดผลการจำแนกของระบบตรวจหาเฉพาะค่าพีเอฟ วัตถุประสงค์ของขั้นตอนนี้คือวัดประสิทธิภาพของระบบตรวจหาค่าลักษณะเฉพาะค่าพีเอฟก่อนที่จะนำไปใช้ในการสกัดค่าพีเอฟในงานการรู้จำภาษา โดยขั้นตอนนี้ทดสอบคือ นำเสียงพูดในฐานข้อมูลเสียงส่วนที่ใช้สำหรับการทดสอบมาสกัดค่าลักษณะสำคัญทางเสียง แล้วนำมาทดสอบกับระบบตรวจหาลักษณะเฉพาะค่าพีเอฟที่ได้จากขั้นตอนนี้ฝึกฝน เมื่อได้คำตอบของการจำแนกค่าพีเอฟในแต่ละเฟรมนำมาเปรียบเทียบกับคำตอบจริงที่เตรียมไว้ และวัดความแม่นยำของการตรวจหา

เมื่อได้ระบบตรวจหาเฉพาะค่าพีเอฟสำหรับแต่ละชนิดที่มีประสิทธิภาพ จึงนำไปใช้ในการสกัดค่าพีเอฟเพื่อใช้ในการรู้จำภาษาต่อไป ผลลัพธ์ของระบบตรวจหาเฉพาะค่าพีเอฟสำหรับค่าพีเอฟแต่ละชนิด นำมาประกอบกันเป็นเวกเตอร์ลักษณะเฉพาะของค่าพีเอฟ เพื่อเป็นเวกเตอร์ค่าลักษณะเฉพาะค่าพีเอฟ 13 ชนิดซึ่งผลลัพธ์สุดท้ายของระบบตรวจหาลักษณะเฉพาะค่าพีเอฟ และใช้เวกเตอร์ค่าพีเอฟดังกล่าวเป็นอินพุตของระบบการรู้จำภาษาพูดที่นำเสนอต่อไป ดังนั้นบนแต่ละเฟรมของสัญญาณเสียงพูด เวกเตอร์ค่าลักษณะเฉพาะค่าพีเอฟนี้ประกอบด้วยค่าพีเอฟ 13 ชนิด แต่ละค่าพีเอฟแสดงลักษณะเฉพาะการออกเสียงต่าง ๆ และมีความสัมพันธ์กันระหว่างชนิดค่าพีเอฟ นอกจากนี้ นำผลลัพธ์ของโครงข่ายประสาทเทียมมาแปลงเป็นค่าความน่าจะเป็นภายหลังค่าพีเอฟในแต่ละเฟรม ซึ่งมีค่าอยู่ระหว่าง 0 ถึง 1 ค่าผลลัพธ์ที่ใกล้เคียงค่า 1 แสดงว่าค่านั้นมีคุณสมบัติใกล้เคียงกับค่าพีเอฟเป้าหมาย (Target PF) และมีการปรากฏค่าพีเอฟชนิดนั้น และค่าผลลัพธ์ที่ใกล้เคียงค่า 0 แสดงว่าค่านั้นมีคุณสมบัติไม่ใกล้เคียงพีเอฟเป้าหมาย (Non-target PF) และไม่มีการปรากฏค่าพีเอฟชนิดนั้น

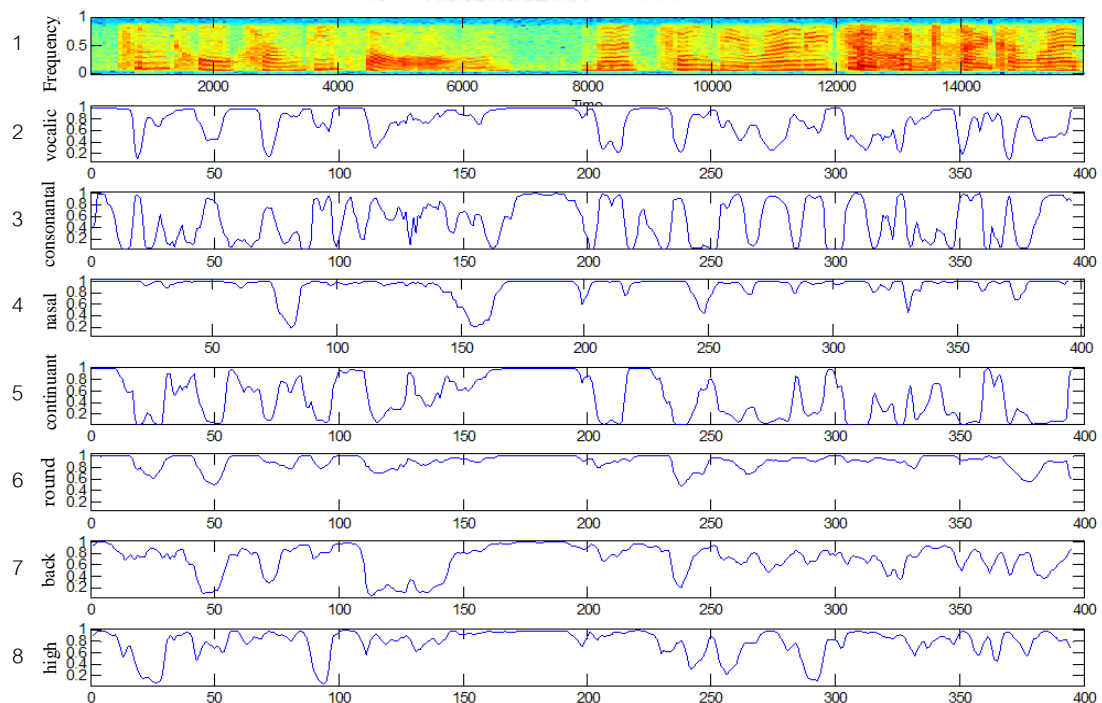
5.2.1 การศึกษาค่าพีเอฟ

เนื่องจากเราไม่สามารถวิเคราะห์หน่วยเสียงต่าง ๆ หรือค่าลักษณะสำคัญทางเสียงที่อยู่บนสัญญาณเสียงจากรูปแบบคลื่นในโดเมนทางเวลาได้ ดังนั้นเราจึงพิจารณาสัญญาณเสียงพูดในรูปแบบสเปกโตรแกรม ซึ่งสเปกโตรแกรมเป็นตัวแทนของสัญญาณเสียงที่แสดงออกมาในรูปของกราฟิกและเป็นเครื่องมือที่บอกข้อมูลเกี่ยวกับหน่วยเสียงจากรูปแบบคลื่นในโดเมนความถี่

(Frequency domain) พร้อมทั้งแสดงผลลัพธ์ของระบบตรวจหาเฉพาะค่าพีเอฟที่สอดคล้องกับสเปกโตรแกรมของสัญญาณเสียงพูด

ภาพที่ 5-1 แสดงสเปกโตรแกรมและค่าพีเอฟของเสียงพูดภาษาไทยข้อความว่า “ฟังโทรครั้งที่สองและโชคดีมากได้เจอคนไทยครั้งแรก” ซึ่งสามารถแสดงการถอดข้อความหน่วยเสียงด้วย /pvnŋ[^] toz[^] maz[^] kraang[^] thiiz[^] s@[^]ng[^] laez[^] chok[^] diiz[^] mak[^] dhayz[^] jherz[^] kon[^] thiiz[^] kraang[^] raek[^]/ ภาพย่อยที่ 1 แสดงสเปกโตรแกรมของเสียงพูดและภาพย่อยที่ 2 ถึง 8 ของภาพที่ 5-1 แสดงค่าพีเอฟชนิดต่าง ๆ ที่ได้มาจากระบบตรวจหาค่าลักษณะค่าพีเอฟ

จากการสังเกตภาพย่อยที่ 2 ถึง 8 พบว่าในแต่ละภาพย่อยสัมพันธ์กันและมีจังหวะของการเปลี่ยนแปลงค่าพีเอฟในเวลาใกล้เคียงกัน ค่าพีเอฟบางชนิดแสดงการเปลี่ยนแปลงไปในทิศทางเหมือนกัน ถึงแม้ว่าค่าพีเอฟแต่ละค่าได้มาจากระบบตรวจหาค่าพีเอฟที่ฝึกฝนแยกกันสำหรับแต่ละค่าพีเอฟ เช่น การเปลี่ยนแปลงของค่าพีเอฟเปลี่ยนแปลงเหมือนกันและเป็นไปในทิศทางเดียวกันในภาพย่อยที่ 7 และ 8 ที่แสดงค่าพีเอฟ Back และ High ซึ่งเป็นค่าคุณสมบัติแสดงตำแหน่งของลิ้นที่สัมพันธ์กับการออกเสียงสระ ภาพย่อยที่ 7 และ 8 มีการเปลี่ยนแปลงค่าพีเอฟเป็นไปในทิศทางเดียวกัน เมื่อจังหวะการเปลี่ยนแปลงเข้าใกล้ค่า 1 หรือการลดลงเข้าใกล้ค่า 0 เกิดขึ้นในเวลาใกล้เคียงกัน นอกจากนี้เมื่อพิจารณาการเปลี่ยนแปลงค่าพีเอฟในภาพย่อยที่ 6 ที่แสดงค่าพีเอฟ Round บ่งบอกรูปร่างการห่อปากซึ่งเป็นคุณสมบัติของการออกเสียงสระ พบว่ามีลักษณะการเปลี่ยนแปลงคล้ายกัน ภาพย่อยที่ 7 ที่แสดงค่าพีเอฟที่เกี่ยวข้องกับคุณสมบัติของสระ เช่นกัน เป็นต้น



ภาพที่ 5-1 สเปกโตรแกรมและตัวอย่างค่าพีเอฟ

อย่างไรก็ตามจากการพิจารณาการเปลี่ยนแปลงของค่าพีเอฟทุกชนิดสามารถสรุปได้ว่า ค่าพีเอฟทุกค่าไม่ได้เกิดการเปลี่ยนแปลงทันที (Instantaneously) และไม่ได้เกิดการเปลี่ยนแปลงค่าพร้อมกัน (Synchronous) ทุกค่า นอกจากนี้ค่าพีเอฟบางค่าค่อย ๆ เปลี่ยนแปลงที่ตำแหน่งขอบเขตหน่วยเสียง แต่ค่าพีเอฟค่าอื่นเกิดการเปลี่ยนแปลงในอีกสองเฟรมถัดไป ซึ่งแสดงให้เห็นว่าค่าพีเอฟที่ค่อย ๆ เปลี่ยนแปลงจากหน่วยเสียงหนึ่งไปยังอีกเสียงหนึ่ง นอกจากนี้ค่าพีเอฟที่ได้จากระบบตรวจหาค่าลักษณะเฉพาะค่าพีเอฟเป็นค่าความน่าจะเป็นของค่าการออกเสียงแต่ละชนิด ไม่ใช่ค่าเลขจำนวนเต็ม 0 หรือ 1 ทำให้คุณสมบัติของค่าพีเอฟที่มีลักษณะค่อย ๆ เปลี่ยนแปลง และการเปลี่ยนแปลงของค่าพีเอฟที่จะค่อย ๆ เปลี่ยนแปลงที่ตำแหน่งตำแหน่งขอบเขตหน่วยเสียงจากหน่วยเสียงหนึ่งไปเป็นอีกหน่วยเสียงหนึ่งบนสัญญาณเสียงพูด ไปแสดงพฤติกรรมของการเปลี่ยนแปลงของค่าพีเอฟที่แตกต่างกันในแต่ละภาษา เพื่อแทนค่าพีเอฟที่มีค่าเป็น 0 หรือ 1 ที่ได้จากการแปลงค่าโดยใช้ผลจากค่าพีเอฟที่แปลงจากฉลากหน่วยเสียงโดยตรง

5.2.2 การศึกษาการกระจายค่าพีเอฟในแต่ละภาษา

สมมติฐานของการทดลองนี้ คือ หน่วยเสียงเดียวกันมีความแปรปรวนของค่าพีเอฟแตกต่างกันในแต่ละภาษา ดังนั้นในการทดลองผู้วิจัยต้องการศึกษาความแปรปรวนของค่าพีเอฟบนหน่วยเสียงในภาษาต่าง ๆ โดยใช้ฐานข้อมูลเสียง OGI ส่วนเสียงพูดต่อเนื่อง (อธิบายในหัวข้อ 3.5.2) ในการทดลองเพื่อหาความแปรปรวนของค่าพีเอฟบนเสียงพูดในแต่ละภาษา

ขั้นตอนการศึกษาความแปรปรวนค่าพีเอฟ มีดังนี้

- 1) การเลือกหน่วยเสียง: ขั้นตอนนี้ทำการเลือกตัวแทนของหน่วยเสียงสำหรับเป็นกรณีศึกษาการกระจายของค่าพีเอฟ โดยมีเงื่อนไขของการเลือกหน่วยเสียงสำหรับตัวแทนหน่วยเสียงที่เป็นเสียงสระและตัวแทนของเสียงพยัญชนะ พิจารณาจากหน่วยเสียงที่มีความถี่การปรากฏของหน่วยเสียงสูงสุดในฐานข้อมูลเสียงแสดงดังภาคผนวก ก ตารางที่ 5-1 แสดงความถี่การปรากฏของหน่วยเสียง 20 อันดับแรกที่ปรากฏในฐานข้อมูลเสียง OGI เรียงลำดับจากความถี่ของการปรากฏหน่วยเสียงจากความถี่สูงไปต่ำ จากตารางพบว่า หน่วยเสียง 'ก' และหน่วยเสียง 'aa' เป็นหน่วยเสียง 2 อันดับแรกที่มีความถี่สูงสุด ผู้วิจัยจึงใช้หน่วยเสียง 'ก' เป็นตัวแทนตัวแทนของเสียงพยัญชนะ และใช้หน่วยเสียง 'aa' เป็นตัวแทนของเสียงสระ ซึ่งหน่วยเสียง 'ก' แทนเสียง น ในภาษาไทย และหน่วยเสียง 'aa' แทนเสียง อะ ในภาษาไทย
- 2) การสกัดค่าพีเอฟ: ค่าพีเอฟที่ใช้ในการศึกษาความแตกต่างของความแปรปรวนค่าพีเอฟได้มาจากผลลัพธ์ของระบบตรวจหาลักษณะเฉพาะค่าพีเอฟ (โดยใช้ค่าจากชั้นผลลัพธ์ของโครงข่ายประสาทเทียมหลายชั้นจากโหนดผลลัพธ์การประมาณค่าความน่าจะเป็นที่เหมือนกับค่าพีเอฟเป้าหมาย) ค่าพีเอฟที่พิจารณาการวิเคราะห์ความแปรปรวน ถูกสกัดมาจากหน่วยเสียง 'ก' และ 'aa' ในแต่ละภาษาจากเสียงพูดในฐานข้อมูลเสียง OGI

การวิเคราะห์: ขั้นตอนการวิเคราะห์ความแปรปรวนของค่าพีเอฟในหน่วยเสียง ผู้วิจัยใช้การวิเคราะห์ความแปรปรวนด้วย ANOVA เพื่อทดสอบว่าค่าพีเอฟในแต่ละภาษาแตกต่างกัน

อย่างมีนัยสำคัญหรือไม่ โดยพิจารณาจากค่า F-ratios และ P-values มีค่า 0.01 เป็นระดับนัยสำคัญ สำหรับการทดสอบความแตกต่าง ผู้วิจัยพิจารณาการวิเคราะห์ความแปรปรวนค่าพีเอฟแบบทั้งหมดในทุกภาษาและพิจารณาการวิเคราะห์ความแปรปรวนค่าพีเอฟแบบที่ละคู่ภาษา (Language pairwise) ซึ่งมีจำนวนคู่ภาษาทั้งหมด 15 คู่ภาษา ดังนั้นจำนวนกรณีที่ศึกษาความแปรปรวนค่าพีเอฟจาก 13 ค่าลักษณะเฉพาะเมื่อพิจารณาแบบที่ละคู่ภาษามีทั้งหมด 195 กรณีศึกษา นอกจากนี้ผู้วิจัยยังพิจารณาค่าเฉลี่ยของค่าพีเอฟในภาษาต่าง ๆ และใช้แผนภาพกล่องเพื่อดูความแตกต่างของค่าพีเอฟในแต่ละภาษา



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ตารางที่ 5-1 ความถี่ของหน่วยเสียง 20 อันดับแรกในฐานข้อมูลเสียง OGI

หน่วยเสียง	อังกฤษ	เยอรมัน	สเปน	อินเดีย	จีนแมนดาริน	ญี่ปุ่น	จำนวนหน่วยเสียงทั้งหมดในทุกภาษา	อัตราส่วนของหน่วยเสียงต่อทั้งหมด
sil	9802	4268	5475	3485	2100	3663	28793	9.54
n	5586	4324	2706	958	1372	2434	17380	5.76
aa	2216	951	5285	1963	103	4473	14991	4.97
s	4832	2013	3387	818	128	929	12107	4.01
tcl	4633	2484	2210	443		1981	11751	3.89
iy	3780	453	3171	924	8	2435	10771	3.57
t	3771	1943	2313	206	862	1631	10726	3.55
ey	1717		5206	1148	1	2549	10621	3.52
ih	6265	2467		938			9670	3.2
eh	2928	3499	1339	1339	7		9112	3.02
ah	6620			2488			9108	3.02
m	2950	1531	2008	1153	355	1110	9107	3.02
k	2608	542	1479	1169	906	1877	8581	2.84
ow			3866	637	3	3772	8278	2.74
kcl	2885	547	1466	1292		1739	7929	2.63
l	3351	1379	1720	722	346	12	7530	2.5
r	3391	114	2078	98	131	1052	6864	2.27
d	2351	1815	692	87			4945	1.64
ay	3610	990	72	109	3	11	4795	1.59
อื่น ๆ	42727	19719	9578	11405	8069	7195	98693	32.69
รวม	106221	44771	48576	27897	12294	33200	301752	100

5.2.2.1 ผลการวิเคราะห์ความแปรปรวนค่าพืเอฟในแต่ละภาษา

ผลการทดลองของการศึกษาการกระจายค่าพืเอฟในแต่ละภาษา (โดยใช้ค่าจากชั้นผลลัพธ์ของโครงข่ายประสาทเทียมหลายชั้นจากโหนดผลลัพธ์การประมาณค่าความน่าจะเป็นที่เหมือนกับค่าพืเอฟเป้าหมาย) พบว่าค่าเฉลี่ยของค่าพืเอฟแตกต่างกันอย่างมีนัยสำคัญเมื่อวิเคราะห์ความแปรปรวนของค่าพืเอฟแบบทั้งหมดรวมทุกค่าพืเอฟ ตารางที่ 5-2 แสดงผลการวิเคราะห์ความแปรปรวนของค่าพืเอฟแต่ละชนิดบนหน่วยเสียง 'n' และ 'aa'

ตารางที่ 5-2 ค่า F-ratio และค่า P-value ของค่าพืเอฟแต่ละชนิดบนหน่วยเสียง 'n' และ 'aa'

ค่าพืเอฟ	หน่วยเสียง n		หน่วยเสียง aa	
	F-ratio	P-value	F-ratio	P-value
Vocalic	420.54	0	580.39	0
Voice	964.05	0	824.53	0
Continuant	564.85	0	425.63	0
Consonantal	325.79	0	147.00	2.05×10^{-48}
Coronal	113.13	4.69×10^{-119}	270.92	0
Nasal	323.32	0	14.27	5.71×10^{-14}
Round	316.93	0	439.64	0
High	635.21	0	404.03	0
Low	150.78	5.13×10^{-159}	201.79	0
Back	646.36	0	73.05	6.69×10^{-76}
Tense	452.29	0	464.71	0
Anterior	107.45	0	208.41	0
Strident	64.26	5.23×10^{-67}	61.87	4.08×10^{-64}

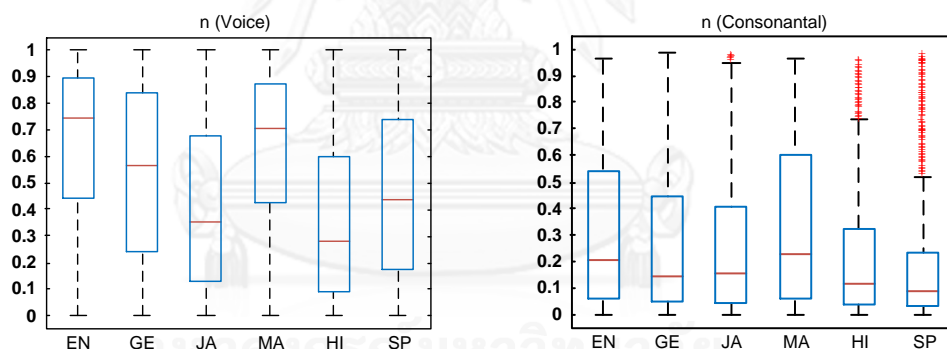
จากตารางที่ 5-2 พบว่าค่า P-value ของการวิเคราะห์ค่าแปรปรวนแบบทั้งหมดในทุกภาษามีค่าน้อยกว่า 0.01 หมายความว่าค่าเฉลี่ยของค่าพืเอฟมีความแตกต่างกันอย่างมีนัยสำคัญระหว่างภาษาทั้งในเสียง 'n' และเสียง 'aa' ค่าพืเอฟแต่ละชนิดสามารถบอกความแตกต่างระหว่างภาษาได้จากการพิจารณาค่า F-ratio พบว่า

- 1) ค่าลักษณะเฉพาะของการออกเสียง Voice มีค่าเฉลี่ยของค่าพืเอฟแตกต่างระหว่างภาษามากกว่าค่าลักษณะเฉพาะพืเอฟค่าอื่น
- 2) ค่าลักษณะเฉพาะของการออกเสียงที่เกี่ยวข้องกับการปิดกั้นของช่องทางเดินเสียง ได้แก่ ค่าพืเอฟ Vocalic ค่าพืเอฟ Continuant มีค่าเฉลี่ยของค่าพืเอฟแตกต่างระหว่างภาษามากกว่าค่าพืเอฟ Consonantal

- 3) ค่าลักษณะเฉพาะค่าพีเอฟ Nasal ในหน่วยเสียง 'n' มีค่าเฉลี่ยของค่าพีเอฟแตกต่างระหว่างภาษา มากกว่าค่าลักษณะเฉพาะค่าพีเอฟ Nasal ในหน่วยเสียง 'aa' เนื่องจากหน่วยเสียง 'n' เป็นเสียงพยัญชนะในกลุ่มนาสิก จึงแสดงคุณสมบัติของค่าการออกเสียงของนาสิกเด่นชัดกว่าค่าพีเอฟที่ได้จากหน่วยเสียง 'aa' ที่ไม่ใช่เสียงกลุ่มนาสิก
- 4) เมื่อเปรียบเทียบค่าลักษณะเฉพาะในกลุ่มที่เกี่ยวข้องกับการออกเสียงสระ ระหว่างค่าพีเอฟ High ค่าพีเอฟ Back ค่าพีเอฟ Low และค่าพีเอฟ Round พบว่า ค่าลักษณะเฉพาะค่าพีเอฟ High มีค่าเฉลี่ยของค่าพีเอฟแตกต่างระหว่างภาษามากกว่าค่าพีเอฟอื่น

5.2.2.2 ผลการศึกษาค่าพีเอฟบนเสียง 'n'

จากตารางที่ 5-2 แสดงการวิเคราะห์ความแปรปรวนค่าพีเอฟแบบทั้งหมดในทุกภาษา ค่า P-value ของค่าพีเอฟบนเสียง 'n' มีค่าน้อยกว่า 0.01 หมายความว่า ค่าเฉลี่ยของค่าพีเอฟใน 6 ภาษา มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ภาพที่ 5-2 แสดงแผนภาพกล่องของค่าพีเอฟ Voice (รูปทางซ้าย) และค่าพีเอฟของ Consonantal (รูปทางขวา) บนเสียง 'n' พบว่า ค่าเฉลี่ยของค่าพีเอฟ Voice บนเสียง 'n' มีค่ามากกว่า 0.5 ในภาษาอังกฤษ และภาษาจีนแมนดารินเมื่อเปรียบเทียบกับภาษาอื่นซึ่งมีค่าเฉลี่ยต่ำกว่า 0.5 เช่น ในภาษาฮินดีมีค่าเฉลี่ยกว่า 0.3 และค่าเฉลี่ยของค่าพีเอฟ Consonantal บนเสียง 'n' มีค่าเฉลี่ยต่ำกว่า 0.3 สำหรับทุกภาษา



(a) ค่าพีเอฟ Voice บนเสียง 'n'

(b) ค่าพีเอฟ Consonantal บนเสียง 'n'

ภาพที่ 5-2 แผนภาพกล่องค่าพีเอฟ Voice และค่าพีเอฟ Consonantal บนเสียง 'n'

เมื่อพิจารณาแผนภาพกล่องของค่าพีเอฟ Voice พบว่าสามารถบอกความแตกต่างระหว่างภาษาได้ เพราะค่าเฉลี่ยของค่าพีเอฟ Voice ในแต่ละภาษามีความแตกต่างกันอย่างชัดเจน และเมื่อพิจารณาค่าส่วนเบี่ยงเบนมาตรฐานของ 6 ภาษาในภาคผนวก ข-1 พบว่ามีค่ามากเมื่อเทียบกับค่าเฉลี่ย เมื่อเปรียบเทียบแผนภาพกล่องค่าพีเอฟ Voice และ ค่าพีเอฟ Consonantal พบว่าค่าพีเอฟ Consonantal ในแต่ละภาษามีการเหลื่อมซ้อนทับกัน (Overlapping) ระหว่างภาษา แสดงถึงค่าเฉลี่ยของค่าพีเอฟ Consonantal ของแต่ละภาษามีค่าใกล้เคียง ขณะที่แผนภาพกล่องของค่าพีเอฟ Voice มีการเหลื่อมซ้อนทับกันของกล่องน้อยกว่า นอกจากนี้ข้อสังเกตที่ได้จากการพิจารณาแผนภาพกล่องค่าพีเอฟชนิดอื่นจากภาคผนวก ข-1 พบว่าความแตกต่างระหว่างค่าเฉลี่ยของค่าพีเอฟ Low

และ Tense มีความแตกต่างกันอย่างชัดเจนระหว่าง 6 ภาษา แต่เมื่อพิจารณาแผนภาพกล่อง Low และ Tense มีข้อมูลที่อยู่นอกขอบเขตกล่องจำนวนมาก เมื่อเทียบกับค่าพีเอฟ Voice และ Consonantal พบว่า ค่าพีเอฟ Low และค่าพีเอฟ Tense มีความสามารถในการแยกได้แยกว่า

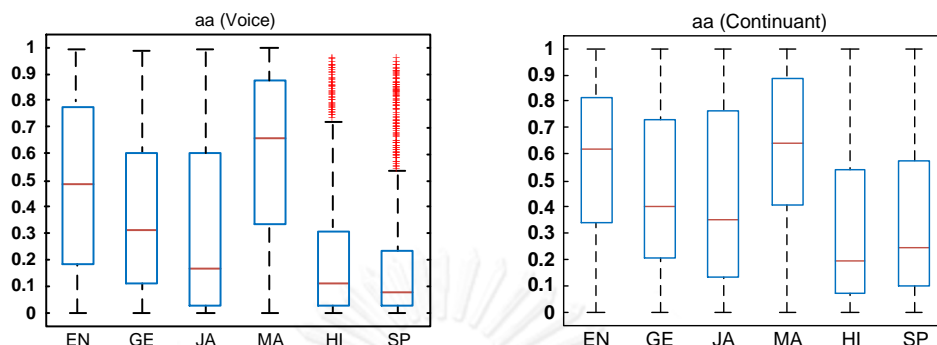
จากภาคผนวก ข-1 แสดงผลการวิเคราะห์ความแปรปรวนค่าพีเอฟโดยพิจารณาแบบที่ละคู่ ภาษา ค่า P-value ของพีเอฟ Voice ค่าพีเอฟ Consonantal ค่าพีเอฟ Low และค่าพีเอฟ Tense บนเสียง 'n' มีค่าน้อยกว่า 0.01 หมายความว่า ค่าเฉลี่ยของพีเอฟบนเสียง 'n' ระหว่าง 6 ภาษามีความแตกต่างกันอย่างมีนัยสำคัญ พบว่า 93.84% (183 กรณีจาก 195 กรณีศึกษา) ของการวิเคราะห์ความแปรปรวนค่าพีเอฟโดยพิจารณาแบบที่ละคู่ภาษามีค่า P-value น้อยกว่า 0.01 แสดงว่าเมื่อพิจารณาที่ละคู่ภาษา ค่าเฉลี่ยของค่าพีเอฟบนเสียง 'n' มีความแตกต่างกันอย่างมีนัยสำคัญ ดังนั้นค่าพีเอฟจึงเป็นค่าที่เหมาะสมในการใช้เป็นค่าลักษณะสำคัญที่บอกความแตกต่างระหว่างสองภาษาได้

นอกจากนี้ผลการวิเคราะห์ความแปรปรวนที่พิจารณาแบบที่ละคู่ภาษาที่มีค่า P-value ของมากกว่า 0.01 ซึ่งค่าเฉลี่ยของค่าพีเอฟไม่มีความแตกต่างกันอย่างมีนัยสำคัญ ได้แก่ ค่าพีเอฟ Vocalic ระหว่างภาษาอังกฤษคู่กับภาษาจีนแมนดารินมีค่า P-value 0.4 ค่าพีเอฟ Continuant ระหว่างภาษาญี่ปุ่นคู่กับภาษาอินดีมีค่า P-value 0.84 ค่าพีเอฟ Coronal ระหว่างภาษาอังกฤษคู่กับภาษาจีนแมนดารินมีค่า P-value 0.74 และระหว่างภาษาจีนแมนดารินคู่กับภาษาสเปนซึ่งค่า P-value มีค่า 0.03 ค่าพีเอฟ Round ระหว่างภาษาอังกฤษคู่กับภาษาจีนแมนดารินซึ่งค่า P-value มีค่า 0.04 ค่าพีเอฟ High ระหว่างภาษาเยอรมันคู่กับภาษาสเปนมีค่า P-value 0.35 ค่าพีเอฟ Back ระหว่างภาษาอังกฤษคู่กับภาษาจีนแมนดารินมีค่า P-value 0.41 ค่าพีเอฟ Strident ระหว่างภาษาอังกฤษคู่กับภาษาเยอรมันมีค่า P-value 0.65 และระหว่างภาษาจีนแมนดารินคู่กับภาษาญี่ปุ่นมีค่า P-value 0.44 ค่าพีเอฟ Nasal ระหว่างภาษาเยอรมันคู่กับภาษาญี่ปุ่นมีค่า P-value 0.13 ระหว่างภาษาเยอรมันคู่กับภาษาสเปนมีค่า P-value 0.34 และระหว่างภาษาญี่ปุ่นคู่กับภาษาสเปนมีค่า P-value 0.55 เป็นต้น

5.2.2.3 ผลการศึกษาค่าพีเอฟบนเสียง 'aa'

ผลการวิเคราะห์ความแปรปรวนของค่าพีเอฟแบบทั้งหมดในทุกภาษา แสดงดังตารางที่ 5-1 พบว่าค่า P-value ของค่าพีเอฟบนเสียง 'aa' มีค่าน้อยกว่า 0.01 แสดงว่าการวิเคราะห์ความแปรปรวนของค่าพีเอฟแบบทั้งหมดใน 6 ภาษามีความแตกต่างกันอย่างมีนัยสำคัญ เมื่อพิจารณาแยกตามแต่ละชนิดค่าพีเอฟ ผลการวิเคราะห์ความแปรปรวนของค่าพีเอฟเสียง 'aa' โดยพิจารณาแบบที่ละคู่ภาษา พบว่าค่าพีเอฟ Voice ค่าพีเอฟ Consonantal ค่าพีเอฟ Low และค่าพีเอฟ Tense บนเสียง 'aa' ใน 6 ภาษา มีค่า P-value น้อยกว่า 0.01 ซึ่งหมายความว่าค่าเฉลี่ยของค่าพีเอฟนั้นมีความแตกต่างกันอย่างมีนัยสำคัญ ภาพที่ 5-3 แสดงแผนภาพกล่องของค่าพีเอฟ Voice และค่าพีเอฟ Continuant บนเสียง 'aa' ระหว่าง 6 ภาษา พบว่า ค่าพีเอฟ Voice บนเสียง 'aa' มีค่าเฉลี่ยมากกว่า 0.5 ในจีนแมนดารินเท่านั้น และค่าพีเอฟ Continuant บนเสียง 'aa' มีค่าเฉลี่ยมากกว่า 0.5 สำหรับภาษาจีนและภาษาอังกฤษ นอกจากนี้เมื่อเปรียบเทียบแผนภาพกล่องของค่าพีเอฟ Voice บนเสียง 'aa' และบนเสียง 'n' (ภาพที่ 5-2) แสดงให้เห็นว่าการกระจายของค่าพีเอฟที่ปรากฏบนแต่ละหน่วยเสียงมีความแตกต่างกัน สำหรับแผนภาพกล่องของค่าพีเอฟอื่น ๆ บนหน่วยเสียง 'aa' แสดง

ในภาคผนวก ข-2 จากการพิจารณาแผนภาพกล่องสามารถสรุปว่า ความแปรปรวนของค่าพีเอฟบนหน่วยเสียงแต่ละภาษามีความแตกต่างกัน



(a) ค่าพีเอฟ Voice บนเสียง 'aa'

(b) ค่าพีเอฟ Continuant บนเสียง 'aa'

ภาพที่ 5-3 แผนภาพกล่องค่าพีเอฟ Voice และค่าพีเอฟ Continuant บนเสียง 'aa'

ผลการพิจารณารวิเคราะห์ความแปรปรวนของค่าพีเอฟสำหรับการสังเกตแบบที่ละคู่ภาษา และแผนภาพกล่องของค่าพีเอฟของเสียง 'aa' จากภาคผนวก ข-2 พบว่า 175 กรณีจาก 195 กรณีศึกษาหรือร้อยละ 89.74 ของกรณีศึกษาที่พิจารณาแบบที่ละคู่ภาษาที่มีค่า P-value น้อยกว่า 0.01 ซึ่งหมายถึงเมื่อพิจารณาความแปรปรวนระหว่างสองภาษา ค่าเฉลี่ยของพีเอฟแต่ละชนิดบนเสียง 'aa' มีความแตกต่างกันอย่างมีนัยสำคัญ ได้แก่ เมื่อพิจารณาค่าพีเอฟ Voice ค่าพีเอฟ Continuant และค่าพีเอฟ Low พบว่าทุกคู่ภาษามีค่าเฉลี่ยของค่าพีเอฟแตกต่างกันอย่างมีนัยสำคัญ โดยร้อยละ 93.33 ของกรณีศึกษาที่พิจารณาแบบคู่ภาษาของค่าพีเอฟ Vocalic และค่าพีเอฟ Tense ที่ค่าเฉลี่ยค่าพีเอฟแตกต่างกันอย่างมีนัยสำคัญซึ่งมี 2 กรณีที่ค่า P-value มากกว่า 0.01 ได้แก่ การพิจารณาค่าพีเอฟ Vocalic ระหว่างภาษาญี่ปุ่นคู่กับภาษาฮินดีซึ่งค่า P-value มีค่า 0.018 และพิจารณาค่าพีเอฟ Tense ระหว่างภาษาเยอรมันคู่กับภาษาญี่ปุ่น ซึ่งค่า P-value มีค่า 0.93

นอกจากนี้ผลการวิเคราะห์ความแปรปรวนที่พิจารณาแบบที่ละคู่ภาษาที่มีค่า P-value ของมากกว่า 0.01 โดยค่าเฉลี่ยของค่าพีเอฟไม่มีความแตกต่างกันอย่างมีนัยสำคัญ ได้แก่ ค่าพีเอฟ Round ของคู่ภาษาอังกฤษกับภาษาเยอรมันมีค่า P-value 0.10 ระหว่างคู่ภาษาอังกฤษกับภาษาสเปน มีค่า P-value 0.55 ระหว่างคู่ภาษาเยอรมันกับภาษาสเปนมีค่า P-value 0.16 ระหว่างภาษาญี่ปุ่นกับภาษาสเปนมีค่า P-value เป็น 0.07 ค่าพีเอฟ Consonantal ระหว่างภาษาอังกฤษกับภาษาญี่ปุ่นมีค่า P-value 0.22 ระหว่างภาษาจีนแมนดารินคู่กับภาษาเยอรมันมีค่า P-value 0.21 ระหว่างภาษาสเปนคู่กับภาษาเยอรมันมีค่า P-value 0.10 และระหว่างภาษาจีนแมนดารินคู่กับภาษาสเปนมีค่า P-value 0.61 ค่าพีเอฟ Coronal ระหว่างภาษาจีนแมนดารินคู่กับภาษาสเปนซึ่งมีค่า P-value 0.81 ค่าพีเอฟ Back ระหว่างภาษาฮินดีคู่กับภาษาสเปนมีค่า P-value 0.22 ค่าพีเอฟ Strident ระหว่างภาษาอังกฤษคู่กับภาษาเยอรมันมีค่า P-value 0.08 ค่าพีเอฟ Nasal ระหว่างภาษาอังกฤษคู่กับภาษาเยอรมันมีค่า P-value 0.53 ระหว่างภาษาอังกฤษคู่กับภาษาญี่ปุ่นมีค่า P-value 0.75 ระหว่างภาษาอังกฤษคู่กับภาษาสเปนมีค่า P-value 0.68 ระหว่างภาษาเยอรมันคู่กับภาษาญี่ปุ่นมีค่า P-value 0.80 ระหว่างภาษาเยอรมันคู่กับภาษาสเปนมีค่า P-value 0.32 ระหว่างภาษาญี่ปุ่นคู่กับ

ภาษาสเปนมีค่า P-value 0.50 และระหว่างภาษาฮินดีคู่กับภาษาสเปนมีค่า P-value 0.02 นอกจากนี้ยังสามารถพิจารณาจากค่า F-ratio ของค่าพีเอฟในแต่ละคู่ภาษาซึ่งค่า F-ratio ที่มีค่ามากสะท้อนคุณสมบัติของค่าพีเอฟนั้นในการนำไปบอกความแตกต่างระหว่างภาษาได้ดี และ ค่า F-ratio ต่ำ สะท้อนว่าค่าเฉลี่ยของค่าพีเอฟไม่แตกต่างกันระหว่างกลุ่ม (ภาษา) เช่น พบในกรณีวิเคราะห์ความแปรปรวนของบางค่าพีเอฟที่มีค่า P-value สูงกว่าระดับความมั่นใจ เป็นต้น

5.2.2.4 สรุปผลการศึกษาระยะการกระจายค่าพีเอฟในแต่ละภาษา

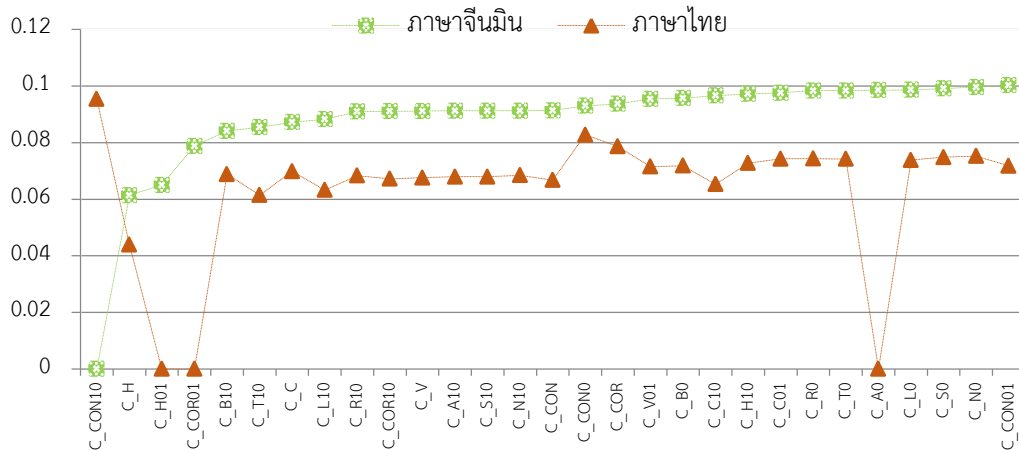
จากผลการศึกษาระยะการกระจายค่าพีเอฟบนเสียง 'n' และเสียง 'aa' ใน 6 ภาษา พบว่าค่า P-value ของการพิจารณาความแปรปรวนของค่าพีเอฟใน 6 ภาษามีค่าน้อยกว่า 0.01 และจากแผนภาพกล่องของค่าพีเอฟจากเสียง 'n' และเสียงอะ 'aa' ระหว่าง 6 ภาษา พบว่าค่าพีเอฟบนหน่วยเสียงเดียวกันในภาษาที่ต่างกัน มีการกระจายค่าพีเอฟที่ต่างกัน ถึงแม้แผนภาพกล่องในบางคู่ภาษามีการเหลื่อมซ้อนทับกันของกล่องค่าพีเอฟ นอกจากนี้ผลวิเคราะห์ความแปรปรวนที่พิจารณาแบบทีละคู่ภาษาพบว่าร้อยละ 90 ของกรณีศึกษามีค่าเฉลี่ยของค่าพีเอฟแตกต่างกันอย่างมีนัยสำคัญ ทำให้ผู้วิจัยเลือกใช้ค่าพีเอฟมาเป็นค่าลักษณะสำคัญเพื่อบอกความแตกต่างระหว่างภาษา ถึงแม้ว่าเมื่อพิจารณาค่าพีเอฟบางชนิดที่ค่าเฉลี่ยของค่าพีเอฟไม่แตกต่างกันอย่างมีนัยสำคัญในบางคู่ภาษา แต่ใช้คุณสมบัติของค่าพีเอฟอื่นที่มีค่าเฉลี่ยแตกต่างกันมาเพิ่มความสามารถในการบอกความแตกต่างระหว่างคู่ภาษานั้นแทนได้

5.2.3 การศึกษาความถี่การปรากฏร่วมกันของค่าพีเอฟในแต่ละภาษา

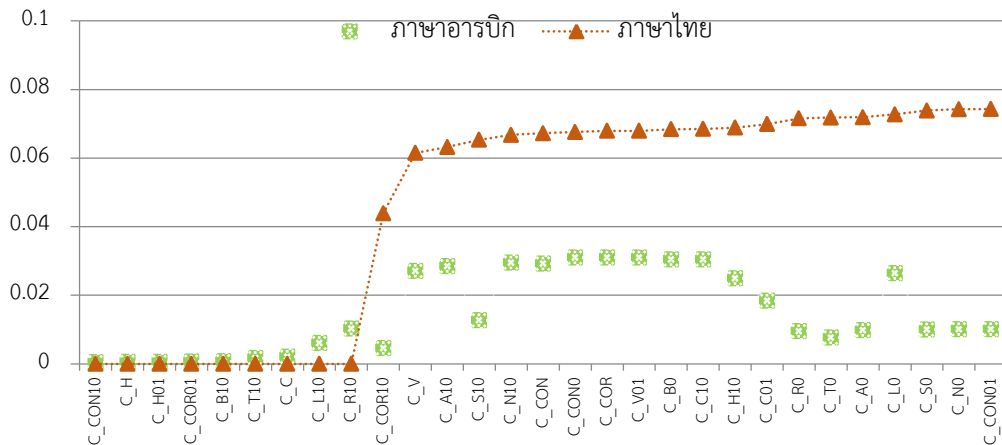
สมมติฐานของการทดลองนี้คือ ความถี่ของการปรากฏร่วมกันของค่าพีเอฟแต่ละชนิดในแต่ละภาษามีค่าแตกต่างกัน การทดลองนี้พิจารณาความถี่ของค่าพีเอฟแต่ละชนิดและการเกิดร่วมกันของค่าพีเอฟในแต่ละภาษา จากฐานข้อมูลเสียงที่ใช้ในการรู้จำภาษา เนื่องจากในงานวิจัยนี้เสนอการใช้ความถี่ของการปรากฏของค่าพีเอฟ และการปรากฏร่วมกันของค่าพีเอฟ เพื่อพิจารณาว่าความถี่ของการปรากฏของพีเอฟในแต่ละภาษาที่ใช้ในงานการรู้จำภาษาแตกต่างกันหรือไม่ สมมติฐานของการทดลองนี้คือ ความถี่ของการปรากฏร่วมกันของค่าพีเอฟแต่ละชนิดในแต่ละภาษามีค่าแตกต่างกัน โดยพิจารณาความถี่ของการเกิดร่วมกันของค่าพีเอฟในแต่ละภาษาจากข้อมูลเสียงชุดฝึกฝนของฐานข้อมูลเสียงสำหรับการรู้จำภาษา (ตามหัวข้อที่ 3.5.3) ซึ่งแนวคิดของการทดลองนี้ใช้แนวคิดของการรู้จำภาษาตามแนวทางการเรียงตัวหน่วยเสียง ที่ใช้ความถี่ของการเรียงตัวของหน่วยเสียง หรือความถี่ของแบบจำลองภาษาของหน่วยเสียงในการบอกความแตกต่างระหว่างภาษา ซึ่งในวิทยานิพนธ์นี้สนใจการใช้ค่าพีเอฟแทนหน่วยเสียงในแต่ละภาษา โดยเสนอระบบการรู้จำภาษาพูดตามแนวทางการเรียงหน่วยเสียงโดยใช้ค่าพีเอฟ การทดลองนี้จึงศึกษารูปแบบการปรากฏร่วมกันของค่าพีเอฟและเปรียบเทียบความถี่ของแต่ละรูปแบบระหว่างแต่ละภาษา

ผลการทดลองการเปรียบเทียบอัตราของความถี่การปรากฏโคออคเคอเรนซ์ค่าพีเอฟในแต่ละภาษาต่อความถี่ของแต่ละรูปแบบนั้นในทุกภาษา ดังแสดงในภาพที่ 5-4 ที่เปรียบเทียบระหว่างภาษาไทย และภาษาจีนมิน พบว่าค่าอัตราความถี่การปรากฏร่วมกันของค่าพีเอฟของแต่ละรูปแบบในภาษาไทย และภาษาจีนมินมีรูปแบบคล้ายกัน หมายถึงตัวอย่างของการปรากฏร่วมกันของค่าพีเอฟที่

ปรากฏในภาษาไทย จะปรากฏในภาษาจีนมินเช่นเดียวกัน โดยความถี่ของตัวอย่างของการปรากฏร่วมกันของค่าพีเอฟในภาษาจีนมินมีค่าสูงกว่าภาษาไทย ขณะที่ในภาพที่ 5-5 เปรียบเทียบระหว่างภาษาไทย และภาษาอารบิก พบว่าค่าอัตราความถี่ของตัวอย่างของการปรากฏร่วมกันของค่าพีเอฟในภาษาไทยมีค่าสูงกว่าภาษาอารบิก ซึ่งความถี่การปรากฏของรูปแบบนั้นนำไปใช้บอกความแตกต่างระหว่างภาษาจีนมิน และภาษาไทยได้ หรือบอกความแตกต่างระหว่างภาษาไทย และภาษาอารบิกตามลำดับ นอกจากนี้ภาพที่ 5-6 และแสดงเปอร์เซ็นต์ของความถี่การปรากฏโคออคเคอเรนซ์ค่าพีเอฟ Vocalic ร่วมกับค่าพีเอฟอื่น ๆ ในแต่ละภาษา จากภาพในภาคผนวก จ แสดงเปอร์เซ็นต์ของความถี่การปรากฏโคออคเคอเรนซ์ค่าพีเอฟต่าง ๆ ในแต่ละภาษา พบว่าตัวอย่างของความถี่การปรากฏโคออคเคอเรนซ์ค่าพีเอฟ Vocalic แบบรวมทุกรูปแบบ ในแต่ละภาษามีค่าความถี่แตกต่างกัน เช่นจากภาพที่ 5-6 เปอร์เซ็นต์ของความถี่การปรากฏโคออคเคอเรนซ์ค่าพีเอฟ Vocalic คู่กับค่าพีเอฟ Strident (ซึ่งแทนด้วย Vocalic-S) ในภาษาเออร์ดูมีความถี่สูงเมื่อเปรียบเทียบกับภาษาอื่น ในขณะที่ในภาษาเบงกาลี ภาษาจีนมิน ภาษาจีนวู ภาษาจีนกวางตุ้ง และภาษาไทย มีความถี่ต่ำมาก เปอร์เซ็นต์ของความถี่การปรากฏโคออคเคอเรนซ์ค่าพีเอฟ Vocalic คู่กับค่าพีเอฟ Consonant (ซึ่งแทนด้วย Vocalic-C) ในภาษาเออร์ดูมีความถี่ต่ำกว่าภาษาอื่น ประมาณ 7% ซึ่งเท่ากับภาษาจีนกวางตุ้ง และภาษาไทย ขณะที่ความถี่ของโคออคเคอเรนซ์ค่าพีเอฟมีความถี่สูงในภาษาอารบิก ภาษาเบงกาลี ภาษารัสเซีย ภาษาจีนวู และภาษาจีนมิน เป็นต้น จากผลการทดลองนี้เป็นแรงจูงใจในการนำความถี่ของการปรากฏโคออคเคอเรนซ์ค่าพีเอฟไปใช้เป็นค่าลักษณะสำคัญในการบอกความแตกต่างระหว่างภาษา สำหรับการรู้จำภาษาตามแนวทางการเรียงตัวหน่วยเสียง ที่ใช้ความถี่ของการเรียงตัวของหน่วยเสียงโดยใช้ค่าพีเอฟต่อไป

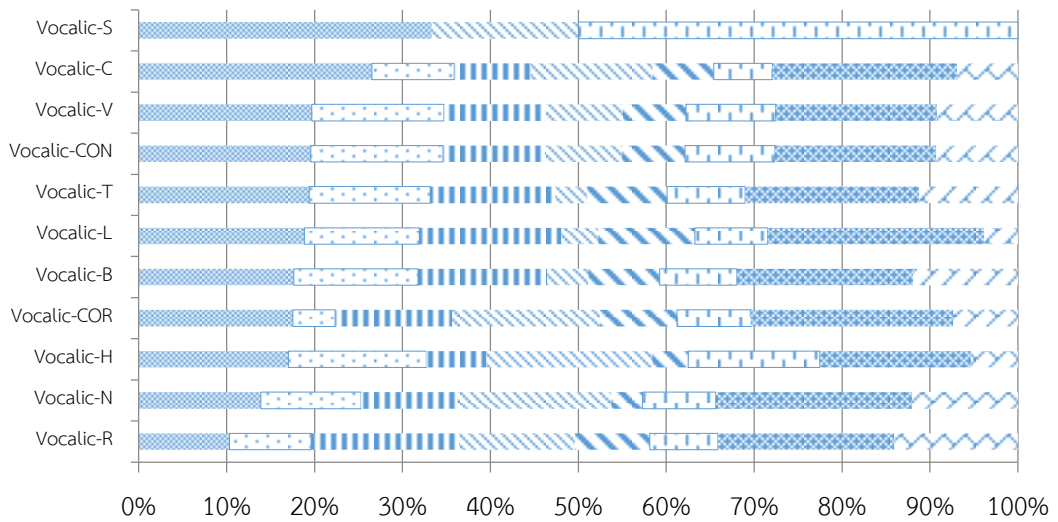


ภาพที่ 5-4 อัตราของความถี่การปรากฏโคออคเคอเรนซ์ค่าพีเอฟระหว่างภาษาจีนมิน และภาษาไทย



ภาพที่ 5-5 อัตราของความถี่การปรากฏโคออกเคอเรนซ์ค่าพีเอระหว่างภาษาอารบิกและภาษาไทย

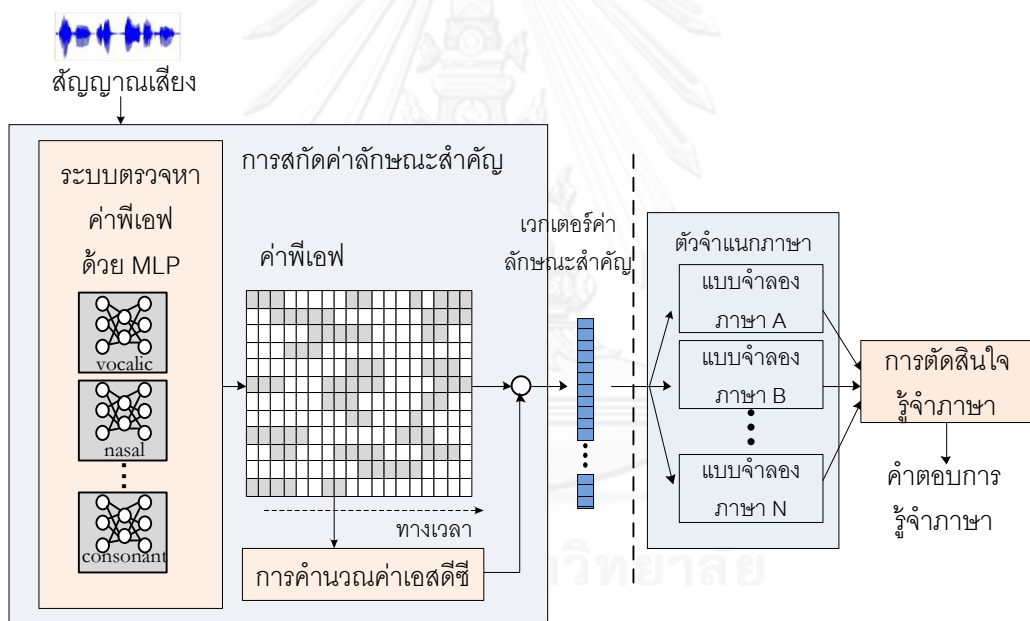
■ อารบิก ■ เบงกาลี ■ จีนมิน ■ รัสเซีย ■ ไทย ■ เออร์ดู ■ จีนวู <math>\lt;/math> จีนกวางตุ้ง



ภาพที่ 5-6 เปอร์เซนต์ของความถี่การปรากฏโคออกเคอเรนซ์ค่าพีเอ Vocalic ร่วมกับค่าพีเอชนิดอื่น ในแต่ละภาษา

5.3 ระบบการรู้จำภาษาพูดโดยใช้การจำแนกด้วยแบบจำลองลาเทนท์ไดนามิกคอนดิชันนอล แรนดอมฟิลต์ร่วมกับค่าพีเอฟ

วิทยานิพนธ์นี้เสนอระบบการรู้จำภาษาพูดโดยใช้การจำแนกด้วยแบบจำลองลาเทนท์ไดนามิกคอนดิชันนอลแรนดอมฟิลต์ร่วมกับค่าพีเอฟ ซึ่งใช้ค่าพีเอฟเป็นค่าลักษณะสำคัญในการบอกความแตกต่างของภาษา ภาพที่ 5-7 แสดงสถาปัตยกรรมของระบบการรู้จำภาษาพูดโดยใช้การจำแนกร่วมกับค่าพีเอฟ ซึ่งประกอบด้วย 2 ส่วนโปรแกรมหลัก ได้แก่ การสกัดค่าลักษณะสำคัญและการจำแนกภาษา ในการสกัดค่าลักษณะสำคัญ นำสัญญาณเสียงพูดมาแปลงเป็นเวกเตอร์ค่าลักษณะเฉพาะของค่าพีเอฟที่สกัดด้วยระบบตรวจหาค่าพีเอฟ ผลลัพธ์ที่ได้คือ เวกเตอร์ของค่าพีเอฟซึ่งนำไปใช้เป็นอินพุทของตัวจำแนกภาษา ตัวจำแนกภาษาถูกฝึกฝนด้วยเวกเตอร์ค่าพีเอฟของแต่ละภาษาเพื่อใช้ในการตัดสินใจระบุภาษาของเสียงที่ต้องการรู้จำภาษา นอกจากนี้ในขั้นตอนการสกัดค่าลักษณะสำคัญ สามารถสร้างผลลัพธ์อื่นแทนเวกเตอร์ของค่าพีเอฟ นั่นคือ เวกเตอร์ค่าลักษณะเฉพาะของค่าเอสดีซีของค่าพีเอฟ โดยคำนวณค่าการเปลี่ยนแปลงไดนามิกจากค่าพีเอฟ



ภาพที่ 5-7 สถาปัตยกรรมของระบบการรู้จำภาษาพูด
โดยใช้การจำแนกร่วมกับค่าพีเอฟ

5.3.1 การสกัดค่าลักษณะสำคัญทางเสียง

การสกัดค่าลักษณะสำคัญทางเสียงทำหน้าที่ หาตัวแทนที่แสดงลักษณะเฉพาะของภาษาจากสัญญาณเสียงพูด ตัวแทนเสียงที่ใช้ในการรู้จำเสียงตามแนวทางนี้คือ ค่าพีเอฟและการเปลี่ยนแปลงของค่าพีเอฟ ซึ่งคำนวณการเปลี่ยนแปลงค่าพีเอฟแบบค่าเอสดีซี (SDC) จากนั้นเปรียบเทียบความสามารถการจำแนกภาษาของค่าลักษณะสำคัญที่นำเสนอ กับค่าลักษณะสำคัญทางเสียงค่าอื่นที่ใช้กันทั่วไปในงานการรู้จำภาษา ซึ่งได้แก่ สัมประสิทธิ์ MFCC และสัมประสิทธิ์ PLP

5.3.1.1 ค่าพีเอฟ PF

ค่าพีเอฟที่ใช้ในการทดลองตามหัวข้อนี้ได้มาจากระบบตรวจหาลักษณะเฉพาะค่าพีเอฟ ซึ่งอธิบายในหัวข้อ 5.2 ประกอบเป็นเวกเตอร์ค่าลักษณะสำคัญที่มีขนาดมิติ 26 ค่า

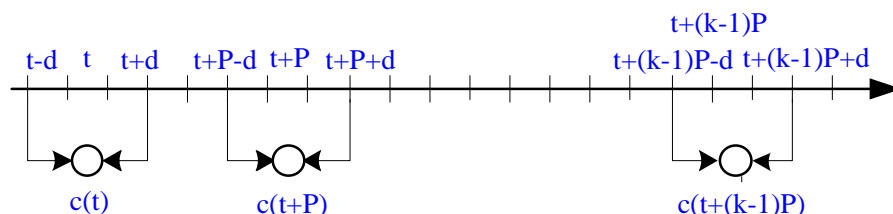
5.3.1.2 ค่าลักษณะสำคัญเอสดีซีที่คำนวณจากค่าพีเอฟ PF

วิทยานิพนธ์นี้ได้นำวิธีการคำนวณค่าการเปลี่ยนแปลงแบบเลื่อนของสัมประสิทธิ์หรือเอสดีซี (Shifted Delta Coefficient: SDC) มาคำนวณค่าการเปลี่ยนแปลงของค่าลักษณะสำคัญทางเสียงที่เกิดขึ้นบนเสียงพูด โดยใช้กรอบการพิจารณาเพื่อคำนวณค่าการเปลี่ยนแปลงของค่าสัมประสิทธิ์ที่อยู่ในเฟรมถัดไป นอกจากนี้ในงานนี้ได้นำวิธีการคำนวณค่าเอสดีซีมาประยุกต์ใช้กับค่าพีเอฟ เพื่อหาค่าการเปลี่ยนแปลงแบบเลื่อนของค่าพีเอฟโดยกรอบการพิจารณาคำนวณค่าการเปลี่ยนแปลงค่าพีเอฟที่เกิดขึ้นในเฟรมถัดไป และใช้การเปลี่ยนแปลงของค่าพีเอฟที่เกิดขึ้นในช่วงกรอบพิจารณานี้แสดงคุณลักษณะของภาษาที่มีรูปแบบค่าการเปลี่ยนแปลงที่แตกต่างกัน การนำวิธีการคำนวณค่าเอสดีซีมาใช้ในวิทยานิพนธ์นี้เนื่องมาจากวรรณกรรมการรู้จำภาษา [37, 116] ที่ได้มีการเสนอค่าเอสดีซี เป็นค่าลักษณะสำคัญให้ผลการรู้จำภาษาแม่นยำมากขึ้นเมื่อเปรียบเทียบกับการใช้ค่าลักษณะสำคัญทางเสียงอื่น โดยค่าเอสดีซีแสดงการเปลี่ยนแปลงแบบไดนามิกของค่าเซปโตรอลโดยใช้กรอบการพิจารณาทางเวลาในช่วงยาวได้ดี

การคำนวณค่าการเปลี่ยนแปลงแบบเลื่อนของสัมประสิทธิ์โดยใช้ค่าพีเอฟแทนสัมประสิทธิ์หรือ การคำนวณค่าการเปลี่ยนแปลงแบบเลื่อนของค่าพีเอฟดังสมการที่ (5.1) และ ภาพที่ 5-8 [43, 52] แสดงการคำนวณค่าเอสดีซีซึ่งมีพารามิเตอร์ที่เกี่ยวข้องในการคำนวณค่าเอสดีซี 4 ค่า คือ $N-d-P-k$ และให้ผลลัพธ์ของการคำนวณนี้เป็นค่าเอสดีซีพีเอฟ (SDCPF)

$$\Delta c(t+iP) = c(t+iP+d) - c(t+iP-d) \quad (5.1)$$

เมื่อค่าพารามิเตอร์ N แทนมิติของค่าสัมประสิทธิ์ที่คำนวณในแต่ละเฟรม พารามิเตอร์ d แทนช่วงระยะเวลาก่อนและหลังที่ต้องการคำนวณ พารามิเตอร์ P แทนจำนวนเฟรมที่ต้องการเลื่อนกรอบการคำนวณไปยังกรอบพิจารณาถัดไป พารามิเตอร์ k แทนจำนวนบล็อกที่คำนวณที่ต้องการนำค่าสัมประสิทธิ์การเปลี่ยนแปลงมาประกอบกันเพื่อสร้างเวกเตอร์ลักษณะเฉพาะสุดท้าย เวกเตอร์ลักษณะเฉพาะค่าเอสดีซีที่เฟรม t คือการนำค่าสัมประสิทธิ์การเปลี่ยนแปลง $\Delta c(t+iP)$ มาต่อกัน



ภาพที่ 5-8 การคำนวณค่าเอสดีซี

ในวิทยานิพนธ์นี้ค่าพารามิเตอร์ที่ใช้ในการคำนวณเอสดีซีคือ 7-1-3-7 อิงตามงานวิจัยที่ศึกษาค่าพารามิเตอร์เอสดีซีที่เหมาะสมที่ให้ผลการรู้จำภาษาดีที่สุด [43, 52]

5.3.2 การจำแนกภาษา

ค่าพีเอฟและค่าเอสดีซีพีเอฟที่ได้จากการขั้นตอนการสกัดค่าลักษณะสำคัญจากหัวข้อที่ 5.3.1 นำมาใช้เป็นข้อมูลของการจำแนกภาษา ในขั้นตอนนี้สร้างต้นแบบของการบอกความแตกต่างระหว่างภาษาเพื่อนำผลลัพธ์ของการจำแนกมาตัดสินใจรู้จำภาษาต่อไป การวัดประสิทธิภาพของผลลัพธ์ที่ได้จากการจำแนกภาษาแสดงเป็น 2 แบบคือ 1) ค่าตอบการตัดสินใจเป็นผลการจำแนกภาษาของเสียงพูดว่าเป็นภาษาที่พิจารณาแบบใช่หรือไม่ ซึ่งใช้ในการวัดความแม่นยำของการจำแนกภาษา เพื่อเปรียบเทียบประสิทธิภาพของการใช้เทคนิคในการจำแนกแบบต่าง ๆ 2) อัตราส่วนความน่าจะเป็นของการตัดสินใจเป็นผลการจำแนกภาษาของเสียงพูดระหว่างคู่ภาษาที่พิจารณา ซึ่งใช้ในการนำคะแนนมาเป็นผลลัพธ์ของการตอบของระบบการรู้จำภาษาพูดแนวทางนี้สำหรับนำไปใช้ในขั้นตอนการรวมผลลัพธ์ต่อไป

ปัญหาการรู้จำภาษาสามารถเขียนในรูปแบบของสมการคณิตศาสตร์ $L = \{L_1, L_2, \dots, L_n\}$ เมื่อ L แทนเซตของภาษาพูดที่ต้องการรู้จำจำนวน n ภาษา เมื่อมีสัญญาณเสียงพูดเป็นข้อมูลอินพุทลำดับที่ X ค่าโอกาสที่จะเป็นไปได้ของข้อมูลที่ต้องการรู้จำแทนด้วย L^* สามารถแสดงได้ดังนี้

$$L^* = \arg \max_L P(L | X) \quad (5.2)$$

โดยปัญหาการรู้จำภาษา คล้ายกับปัญหาการจำแนกหลายกลุ่ม (multiclass) สามารถแสดงเป็นปัญหาการจำแนกแบบทวิภาคหรือไบนารี ในวิทยานิพนธ์นี้พิจารณาการรู้จำภาษา จากปัญหาการจำแนกแบบไบนารี ซึ่งมาจากแบบจำลองที่หนึ่งต่อหนึ่ง สามารถแทนเซตของภาษาที่ต้องการรู้จำด้วย $L = \{l_1, l_2\}$ เมื่อ l_1 แทนด้วยกลุ่มภาษาเป้าหมายหนึ่งภาษาและ l_2 แทนด้วยกลุ่มภาษาที่ไม่ใช่ภาษาเป้าหมายหนึ่งภาษา

วิทยานิพนธ์นี้ใช้แบบจำลองลาเทนท์ไดนามิคคอนดิชันนอลแรนดอมฟิลด์สร้างตัวจำแนกภาษา ดังนั้น แบบจำลองลาเทนท์ไดนามิคคอนดิชันนอลแรนดอมฟิลด์เรียนรู้การจับความสัมพันธ์ระหว่าง ลำดับข้อมูลอินพุทที่เข้ามา $X = \{x_1, x_2, \dots, x_T\}$ และลำดับของฉลากกลุ่มเป้าหมาย $Y = \{y_1, y_2, \dots, y_T\}$ โดยแต่ละ y_i คือฉลากกลุ่มเป้าหมายสำหรับข้อมูล x_i ที่เป็นข้อมูลเข้ามา ลำดับที่ i และสมาชิกของเซต L คือฉลากกลุ่มภาษาเป้าหมาย ตัวอย่างเช่น เมื่อ $Y = \{Chinese, Thai, \dots, Russian\}$ แทนเซตของภาษาที่ต้องการรู้จำที่ประกอบด้วยภาษาจีน ภาษาไทย ภาษารัสเซีย เป็นต้น เนื่องจากสัญญาณเสียงพูดที่ใช้ในงานการรู้จำภาษาหนึ่งเสียงพูดมีคำตอบฉลากของภาษาที่ใช้ในการรู้จำภาษาหนึ่งคำตอบ ดังนั้นกำหนดให้ฉลากของข้อมูลเป็น $y_1 = y_2 = y_T$ สำหรับฉลากของเป้าหมายของข้อมูลที่เข้ามาเป็นกลุ่มภาษาเป้าหมายเดียวกัน ลำดับข้อมูลอินพุทที่เข้ามาแต่ละเฟรม x_i แสดงเวกเตอร์ค่าลักษณะเฉพาะ $\phi(x_i) \in \mathcal{R}$ กำหนดให้เวกเตอร์ของตัวแปรสถานะซ่อน $h = \{h_1, h_2, \dots, h_n\}$ เป็นตัวแปรที่ใช้ในการจำลองโครงสร้างย่อยของสัญญาณเสียงพูด ตัวแปร h_i เป็นสมาชิกของเซตตัวแปรสถานะซ่อนทุกตัวซึ่งแทนด้วย H สำหรับแบบจำลองลาเทนท์ไดนามิคคอนดิชันนอลแรนดอมฟิลด์ ค่าโอกาสที่จะเป็นไปได้ของข้อมูลที่เข้ามาทั้งหมดของภาษา มีค่าเท่ากับ ความน่าจะเป็นเดียวหรือความน่าจะเป็นของเหตุการณ์หนึ่งที่เกิดขึ้น (Marginal

probability) ซึ่งหาได้จากผลรวมของความน่าจะเป็นเดียวของสถานะซ่อนของซบเซต H_L แบบจำลองลาเทนท์คอนดิชันนอล คำนวณได้จากสมการ (5.3) เมื่อ X คือการต่อกันของทุกเวกเตอร์ ค่าลักษณะเฉพาะ x_i สำหรับลำดับเสียงพูดทั้งหมดและให้ $\theta\{x_i\}$ แทนแบบจำลองที่ฝึกฝน

$$P(Y | X, \theta) = \sum_h p(Y | h, X, \theta) p(h | X, \theta) \quad (5.3)$$

จากการอนุมานฟังก์ชันศักยภาพดังสมการ (5.4) แบบจำลองถูกจำกัดให้มีเซตที่ไม่เชื่อมกันของสถานะซ่อนที่สอดคล้องกับฉลากกลุ่มเป้าหมาย

$$P(Y | X, \theta) = \sum_{h: \forall h_i \in H_{y_i}} p(h | X, \theta) \quad (5.4)$$

เมื่อ $p(h | X, \theta)$ แทนแบบจำลองทางสถิติคอนดิชันนอลแรนดอมฟิลด์ส ซึ่งนิยามด้วยสมการที่ (5.5)

$$P(h | X, \theta) = \frac{1}{Z(X, \theta)} \exp\left(\sum_k \theta_k \bullet F_k(h, X)\right) \quad (5.5)$$

และมีเวกเตอร์ค่าลักษณะเฉพาะ F_k สามารถนิยามได้สมการที่ (5.6)

$$F_k(h, X) = \sum_{i=1}^T f_k(h_{i-1}, h_i, X, i) \quad (5.6)$$

เมื่อ $f_k(h_{i-1}, h_i, X, i)$ แทนฟังก์ชันคุณสมบัติ (Feature function) ซึ่งเป็น $t_k(h_{i-1}, h_i, X, i)$ แทนฟังก์ชันการเปลี่ยนสถานะ (Transition function) ที่จำลองการขึ้นต่อกันทางเวลาระหว่างค่าลักษณะสำคัญ หรือเป็นฟังก์ชันสถานะ $s_k(h_i, x, i)$ ที่พิจารณาความสัมพันธ์ระหว่างค่าลักษณะสำคัญ ซึ่งฟังก์ชันสถานะ s_k ขึ้นกับตัวแปรซ่อนเดี่ยวในแบบจำลอง ขณะที่ฟังก์ชันการเปลี่ยนสถานะ t_k ขึ้นกับคู่ของตัวแปรซ่อน

กำหนดให้ฟังก์ชันคุณสมบัติ $s_\phi(h_i, X, i)$ ในสมการที่ (5.6) เป็นฟังก์ชันสถานะบนสัญญาณเสียงพูด X ตำแหน่งที่ i ซึ่งมีฉลากกำกับสถานะกับภาษา ϕ ดังสมการที่ (5.7) เมื่อ x_i เป็นค่าที่สังเกตที่ตำแหน่งที่ i ซึ่งในวิทยานิพนธ์นี้ นิยามฟังก์ชันสถานะจากฉลากภาษาของเสียงพูดและค่าพีเอฟบนสัญญาณเสียงพูด x_i ซึ่งได้จากผลลัพธ์ของระบบตรวจหาค่าพีเอฟ จึงแทนฟังก์ชันด้วยค่าผลลัพธ์จากระบบตรวจหาค่าพีเอฟ ตัวอย่างเช่น ฟังก์ชันของความสัมพันธ์ระหว่างภาษากับผลลัพธ์ของระบบตรวจหาค่าพีเอฟสำหรับค่าการออกเสียง Vocalic บนสัญญาณเสียงพูดที่ตำแหน่ง x_i แทนด้วย $MLP_{vocalic}(x_i)$ ดังสมการที่ (5.8):

$$s_\phi(h_i, X, i) = \begin{cases} x_i & \text{if } h_i = \phi \\ 0 & \text{otherwise} \end{cases} \quad (5.7)$$

$$s_{\phi, vocalic}(h_i, X, i) = \begin{cases} MLP_{vocalic}(x_i), & \text{if } h_i = \phi \\ 0 & ; \text{otherwise} \end{cases} \quad (5.8)$$

สำหรับฟังก์ชันการเปลี่ยนสถานะ t_k ดังสมการที่ (5.9) ถูกนิยามด้วย $|H| \times |H|$ สำหรับแต่ละคู่ของโหนดสถานะซ่อน (h', h'') ซึ่งมีค่าน้ำหนักที่สัมพันธ์กับฟังก์ชันการเปลี่ยนสถานะสำหรับโหนดสถานะซ่อนเป็นซับเซตของ H_{y_i} จะถูกนำมาใช้ในการจำลองรูปแบบของโครงสร้างย่อยภายในแบบจำลอง ค่าน้ำหนักสำหรับแต่ละฟังก์ชันการเปลี่ยนสถานะ จากซับเซตที่แตกต่างกันจะนำมาใช้ในการจำลองรูปแบบประหวาดภาษา

$$t_{h',h''}(h_{i-1}, h_i, X, i) = \begin{cases} 1, & \text{if } h_{i-1} = h' \text{ and } h_i = h'' \\ 0 & ; \text{otherwise} \end{cases} \quad (5.9)$$

สำหรับขั้นตอนการฝึกฝนแบบจำลอง ข้อมูลบนสัญญาณเสียงพูดถูกระบุฉลากของภาษาเป้าหมายในระดับเฟรมเพื่อนำมาใช้สำหรับการฝึกฝนแบบจำลอง ค่าพารามิเตอร์จำนวนของสถานะซ่อนสำหรับแบบจำลองลาเทนท์ไดนามิคคอนดิชันนอลแรนดอมฟิลด์ในขั้นตอนการฝึกฝนและการทดสอบโดยมีจำนวนโหนด 13 สถานะ ขนาดหน้าต่างมีค่าเป็น 3 เฟรม จำนวนของฟังก์ชันสถานะ s_k มีค่าเท่ากับความยาวของเวกเตอร์ค่าลักษณะเฉพาะคูณด้วยจำนวนโหนดของสถานะซ่อน $|H|$ ซึ่งในกรณีนี้จำนวนของการจำแนกภาษาที่ใช้ค่าพีเอฟเป็นข้อมูลอินพุท ในแต่ละเฟรมเวกเตอร์ค่าลักษณะเฉพาะมี 13 มิติและทำให้แบบจำลองมีจำนวนโหนดสถานะซ่อน 3 โหนดสำหรับแต่ละฉลาก (แต่ละภาษา) จะได้จำนวนโหนดสถานะซ่อนทั้งหมด 6 โหนด (สำหรับการสร้างการจำแนกแบบระหว่าง 2 กลุ่ม) นั่นคือ จำนวนฟังก์ชันสถานะ s_k ทั้งหมด เท่ากับ $13 \times 6 = 78$ และในการสร้างแบบจำลองให้มีความสม่ำเสมอ (regularization) ใช้เทคนิค BFGS [117] ทำการหาค่าที่ดีที่สุดของฟังก์ชันล๊อคโอกาสการเกิด โดยให้วนซ้ำเพื่อให้ค่าพารามิเตอร์จำนวนรอบไม่เกิน 300 รอบแบบจำลองถูกฝึกฝนด้วยฟังก์ชันวัตถุประสงค์ดังสมการ (2.9) นอกจากนี้ค่าลักษณะสำคัญเอสดีซีที่คำนวณจากค่าพีเอฟ SDCPF ถูกนำมาใช้ร่วมกับแบบจำลองลาเทนท์ไดนามิคคอนดิชันนอลแรนดอมฟิลด์ ซึ่งการนิยามฟังก์ชันคุณสมบัติเหมือนกับการใช้ผลลัพธ์ที่ได้จากระบบตรวจหาค่าพีเอฟ

ในขั้นตอนการอนุมานกำหนดให้ X เป็นลำดับของเสียงที่ต้องการทดสอบ ประเมินฉลากหรือคำตอบที่มีความน่าจะเป็นมากที่สุดของลำดับข้อมูล Y^* ซึ่งทำให้แบบจำลองคอนดิชันนอลมีค่ามากที่สุด ดังสมการที่ (5.10) เมื่อพารามิเตอร์ของแบบจำลองที่เหมาะสมที่สุด (Optimal model) แทนด้วย θ^* ได้มาจากตัวอย่างในข้อมูลฝึกฝน

$$Y^* = \arg \max_Y p(Y_i | X_i, \theta^*) \quad (5.10)$$

สมมติให้ฉลากแต่ละกลุ่มสอดคล้องกับ เซตที่ไม่เชื่อมกันของสถานะซ่อน จึงสามารถเขียนสมการที่ (8) ในรูปแบบของตัวแปรสถานะซ่อนได้ดังสมการที่ (5.11)

$$Y^* = \arg \max_Y \sum_{h: \forall h_i \in H_{y_i}} p(h | X, \theta^*) \quad (5.11)$$

ในแบบจำลองลาเทนท์ไดนามิคคอนดิชันนอลแรนดอมฟิลด์ การวัดค่าโอกาสที่เป็นไปได้สำหรับภาษา l เท่ากับ ความน่าจะเป็นเดี่ยว $p(Y_i = l | X, \theta^*)$ ความน่าจะเป็นนี้มีค่าเท่ากับผลรวม

ของความน่าจะเป็นเดียวของสถานะซ่อนของซบเซต H ดังสมการที่ (5.12) ซึ่งค่าความน่าจะเป็นเดียวได้มาจากการประมาณโดยใช้อัลกอริทึมการแพร่กระจายความเชื่อ

$$p(Y_i = l | X, \theta^*) = \sum_{h: \forall h_i \in H_i} p(h | X, \theta^*) \quad (5.12)$$

เมื่อ X คือลำดับของเวกเตอร์ค่าลักษณะเฉพาะ x_i สำหรับลำดับข้อมูลที่เข้ามาจากเสียงพูดทั้งหมดฝึกฝน โดยในระบบการรู้จำภาษาพูดแนวทางนี้คือเวกเตอร์ค่าลักษณะเฉพาะของค่าพีเอฟ และ θ^* แทนพารามิเตอร์ของแบบจำลองที่ได้จากการ

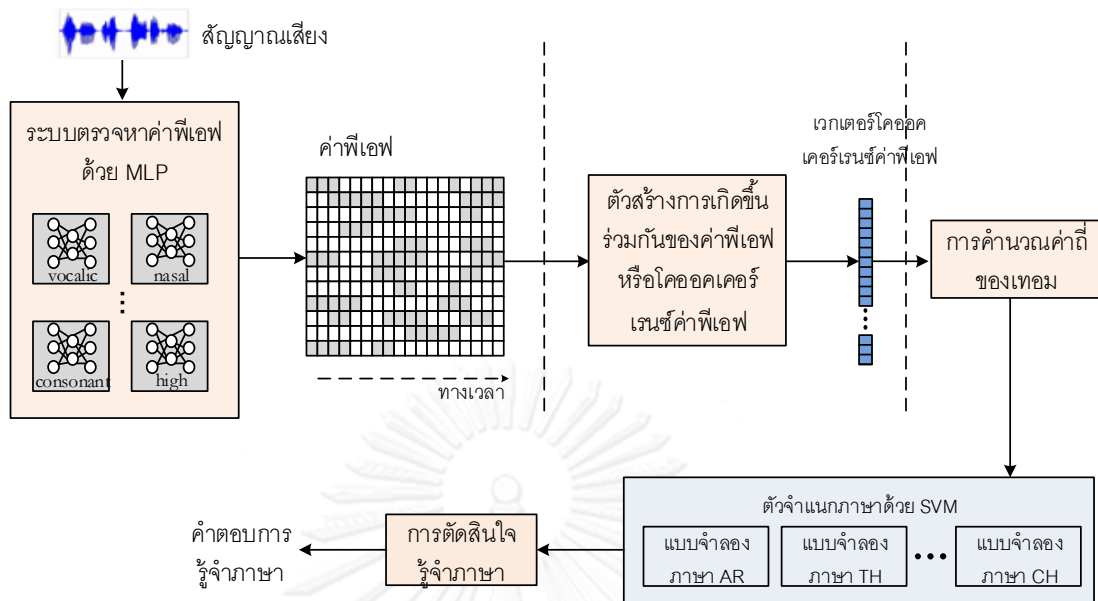
นอกจากนี้เวกเตอร์ค่าคุณลักษณะสำคัญทางเสียงจากค่าสัมประสิทธิ์ MFCC และค่าสัมประสิทธิ์ PLP ถูกนำมาใช้เป็นข้อมูลอินพุทของตัวจำแนกเช่นกัน เพื่อเปรียบเทียบประสิทธิภาพของค่าพีเอฟที่ได้นำเสนอ โดยรายละเอียดอธิบายในการทดลองหัวข้อ 6.4.2

5.4 ระบบการรู้จำภาษาพูดตามแนวทางการเรียงหน่วยเสียงโดยใช้ค่าพีเอฟ

ในหัวข้อนี้เสนอระบบการรู้จำภาษาพูดตามแนวทางการเรียงหน่วยเสียงโดยใช้ค่าพีเอฟ ซึ่งมีแนวคิดมาจากการพิจารณาความสัมพันธ์ระหว่างค่าพีเอฟภายในเฟรมเดียวกันมีรูปแบบการประกอบกันของค่าพีเอฟที่แตกต่างกัน หรือแนวคิดจากการประกอบกันของค่าพีเอฟในแต่ละหน่วยเสียงมีรูปแบบแตกต่างกัน นอกจากนี้มีการเปลี่ยนแปลงของการออกเสียงในหน่วยเสียงหนึ่งไปเป็นอีกหน่วยเสียงหนึ่งในแต่ละภาษาทำให้เป็นหน่วยแทนหน่วยเสียงรูปแบบแตกต่างกัน วิทยานิพนธ์นี้เสนอรูปแบบของเหตุการณ์ที่เกิดขึ้นร่วมกันของค่าพีเอฟ หรือโคออคเคอร์เรนซ์ค่าพีเอฟ (Co-occurrence PF) เพื่อแสดงการจับคู่ค่าพีเอฟสองชนิดและพิจารณารูปแบบของการปรากฏร่วมกันของค่าพีเอฟ เช่น การปรากฏร่วมกันของค่าพีเอฟ Nasal ที่มีค่าเป็น 1 ร่วมกับการปรากฏร่วมกันของค่าพีเอฟ Strident ที่มีค่าเป็น 0 เป็นต้น

ภาพที่ 5-9 แสดงระบบการรู้จำภาษาพูดตามแนวทางกฎการเรียงของค่าพีเอฟ โดยใช้ค่าความถี่ของโคออคเคอร์เรนซ์ค่าพีเอฟ ซึ่งมีส่วนประกอบการทำงาน 5 ส่วน ประกอบด้วย 1) ระบบตรวจหาลักษณะเฉพาะค่าพีเอฟ 2) ตัวสร้างเหตุการณ์ที่เกิดขึ้นร่วมกันของค่าพีเอฟหรือโคออคเคอร์เรนซ์ค่าพีเอฟ 3) การคำนวณค่าความถี่เหตุการณ์ 4) จำแนกภาษา และ 5) การตัดสินใจ

การทำงานแบ่งเป็น 2 ขั้นตอนคือ ขั้นตอนการสร้างต้นแบบของตัวจำแนกภาษาและขั้นตอนการรู้จำภาษา โดยขั้นตอนการสร้างต้นแบบของตัวจำแนกภาษา นำเสียงพูดจากฐานข้อมูลส่วนฝึกฝนสกัดค่าพีเอฟด้วยระบบตรวจหาลักษณะเฉพาะค่าพีเอฟ จากนั้นนำมาสร้างรูปแบบของเหตุการณ์ที่เกิดขึ้นร่วมกันของค่าพีเอฟ แล้วหาความถี่ของเหตุการณ์ที่เกิดขึ้นร่วมกันของค่าพีเอฟนั้น แล้วนำความถี่มาประกอบกันเป็นเวกเตอร์ลักษณะเฉพาะ แล้วนำเวกเตอร์ไปใช้ฝึกฝนตัวจำแนกภาษาเพื่อสร้างต้นแบบของการจำแนกภาษา สำหรับขั้นตอนการรู้จำภาษา นำเสียงพูดที่ต้องการรู้จำมาแปลงเป็นค่าพีเอฟและสร้างเวกเตอร์ลักษณะเฉพาะของความถี่ของเหตุการณ์ที่เกิดขึ้นร่วมกันของค่าพีเอฟ จากนั้นนำเวกเตอร์มาเป็นอินพุทของตัวจำแนกภาษาและนำคำตอบของการจำแนกภาษาที่ได้จากแต่ละคู่ภาษา มาพิจารณาเพื่อตัดสินใจระบุภาษาของเสียงทดสอบเป็นขั้นตอนสุดท้าย



ภาพที่ 5-9 การทำงานของระบบการรู้จำภาษาพูดตามแนวทางการเรียงของค่าพยางค์

5.4.1 ตัวสร้างเหตุการณ์ที่เกิดขึ้นร่วมกันของค่าพยางค์หรือโคออดเคอร์เรนซ์ค่าพยางค์

โคออดเคอร์เรนซ์ค่าพยางค์ถูกใช้เป็นตัวแทนที่แสดงความสัมพันธ์ของเหตุการณ์ที่เกิดขึ้นร่วมกันระหว่างค่าการออกเสียงภายในแต่ละเฟรมและเป็นตัวแทนในสร้างข้อมูลการเรียงของโคออดเคอร์เรนซ์ค่าพยางค์ เพื่อแสดงการเปลี่ยนแปลงของค่าการออกเสียงที่เกิดขึ้นผ่านค่าพยางค์จากเฟรมหนึ่งไปอีกเฟรมหนึ่ง ซึ่งใช้ในการเตรียมข้อมูลการเรียงของค่าการออกเสียงต่อไป

การสร้างโคออดเคอร์เรนซ์ค่าพยางค์ เริ่มต้นจากการนำผลลัพธ์ที่ได้มาจากระบบการตรวจหาลักษณะเฉพาะค่าพยางค์มาทำการควอนไทซ์ (Quantize) เป็นค่า 1 เมื่อค่าผลลัพธ์มีค่าใกล้เคียงกับค่าพยางค์เป้าหมายและมีค่าเป็น 0 เมื่อค่าผลลัพธ์มีค่าไม่ใกล้เคียงกับพยางค์เป้าหมาย ซึ่งผลลัพธ์ที่ได้จากระบบตรวจหาลักษณะเฉพาะค่าพยางค์ ใช้แสดงค่าพยางค์แต่ละชนิด โดยแต่ละระบบตรวจหาลักษณะเฉพาะค่าพยางค์ จะให้ผลลัพธ์ออกมา 2 ค่า คือ ค่าที่บอกความเหมือนของการออกเสียง และค่าที่บอกว่าไม่เหมือนการออกเสียงนั้น ดังนั้นในแต่ละเฟรมจะมีทั้งหมด 26 ค่าและค่าพยางค์แต่ละค่า จะแสดงลักษณะการออกเสียงที่มีความสัมพันธ์กัน

จากเวกเตอร์ลักษณะเฉพาะที่ประกอบด้วยค่าพยางค์ชนิดต่าง ๆ หลายค่าเรียงกันแบบขนานที่แสดงค่าการออกเสียงชนิดต่าง ๆ ทำให้การหารูปแบบการเปลี่ยนแปลงของการออกเสียงทำได้ยาก เพื่อแก้ปัญหาจึงแปลงเวกเตอร์ที่มีหลายมิติเป็นหน่วยหรือตัวแทนที่สามารถแสดงความสัมพันธ์ระหว่างค่าการออกเสียงในแต่ละเฟรมได้ ดังนั้นเป้าหมายของขั้นตอนนี้ คือการสร้างรูปแบบของการประกอบกันของค่าพยางค์แบบต่าง ๆ ซึ่งการปรากฏร่วมกันของค่าพยางค์มีรูปแบบความสัมพันธ์ของการออกเสียงที่ต่างกัน เช่น รูปแบบของการปรากฏของค่าพยางค์ชนิดหนึ่ง ร่วมกับการไม่ปรากฏของค่าพยางค์อีกชนิดหนึ่ง รูปแบบที่หลากหลายนี้อาจสะท้อนความแตกต่างระหว่างภาษาได้

เมื่อได้รูปแบบของการเกิดร่วมกันของค่าพีเอฟ ที่ใช้เป็นตัวแทนเหตุการณ์การเกิดร่วมกันของค่าพีเอฟในแต่ละเฟรมของเสียงพูด (ซึ่งใช้แทนหน่วยเสียงหรือค่าลักษณะทางเสียงต่าง ๆ) จึงทำการหาตัวแทนที่แสดงการเปลี่ยนแปลงของค่าการออกเสียงจากหน่วยเสียงหนึ่งเป็นอีกหน่วยเสียงหนึ่ง โดยใช้การพิจารณาตัวแทนการเกิดร่วมกันของค่าพีเอฟที่เกิดขึ้นพร้อมกันในหลายเฟรม ซึ่งใช้เป็นข้อมูลการเรียงของลำดับค่าพีเอฟที่สามารถนำไปใช้บอกความแตกต่างระหว่างภาษาได้ หัวข้อ 5.4.1.1 อธิบายการสร้างตัวแทนการเกิดร่วมกันของค่าพีเอฟในแต่ละเฟรมและหัวข้อ 5.4.1.2 อธิบายการหาตัวแทนการเกิดร่วมกันของโคออคเคอร์เรนซ์ค่าพีเอฟในรอบพิจารณา เพื่อใช้แทนลำดับข้อมูลการเรียงกันของค่าพีเอฟ

5.4.1.1 การสร้างตัวแทนเหตุการณ์ที่เกิดขึ้นร่วมกันของค่าพีเอฟหรือโคออคเคอร์เรนซ์ค่าพีเอฟในแต่ละเฟรม (Co-occurrence of PF within one frame)

โคออคเคอร์เรนซ์ค่าพีเอฟถูกใช้แสดงความสัมพันธ์ระหว่างค่าพีเอฟ โดยพิจารณาการจับคู่ค่าพีเอฟสองชนิดภายในเฟรมเดียวกัน กำหนดให้แต่ละค่าพีเอฟที่เป็นผลลัพธ์ของระบบการตรวจหา ลักษณะเฉพาะค่าพีเอฟ แทนด้วย s และ $s \in S$ เมื่อ $S = \{s_1, s_2, \dots, s_{13}\}$ เป็นเซตของค่าพีเอฟ ทั้งหมด 13 ค่า ดังนั้นค่า $PF_t(s_i)$ แสดงค่าพีเอฟ s_i และ $PF_t(s_j)$ แสดงค่าพีเอฟ s_j เมื่อ $s_i, s_j \in S$ และ $1 \leq i \leq 13$ และ $1 \leq j \leq 13$ ค่าพีเอฟแต่ละค่าที่ได้จากระบบตรวจหา ลักษณะเฉพาะค่าพีเอฟถูกควอนไทซ์ให้เป็นค่า 1 เมื่อค่าความน่าจะเป็นภายหลังจากที่ได้จากระบบตรวจหามีค่ามากกว่า 0.5 แสดงว่ามีค่าการออกเสียงชนิดนั้น และเป็น 0 เมื่อค่าความน่าจะเป็นที่ได้มีค่าน้อยกว่า 0.5 ซึ่งแสดงว่าไม่มีค่าการออกเสียงชนิดนั้น กำหนดให้โคออคเคอร์เรนซ์ค่าพีเอฟแทนด้วย $COPE_t(s_i, s_j)$ เพื่อแสดงการจับคู่ที่ละสองค่าพีเอฟ $COPE_t(s_i, s_j) = (PF_t(s_i), PF_t(s_j))$ โดยพิจารณาโคออคเคอร์เรนซ์ค่าพีเอฟ s_i เกิดขึ้นร่วมกับกับค่าพีเอฟ s_j ในเฟรม t มีซึ่งมีรูปแบบของเหตุการณ์เกิดขึ้นทั้งหมด 4 แบบ ดังนี้

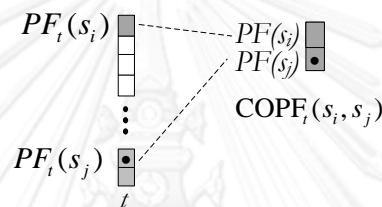
1) $COPE_t(s_i, s_j) = (1, 1)$ แสดงค่าเหตุการณ์ที่ค่าพีเอฟ s_i เกิดขึ้นร่วมกับกับค่าพีเอฟ s_j ในเฟรม t โดยที่ค่าพีเอฟทั้งสองชนิด มีค่าเป็น 1 นั่นคือมีค่าการออกเสียงของค่าพีเอฟเป้าหมายทั้งสองชนิด

2) $COPE_t(s_i, s_j) = (1, 0)$ แสดงค่าเหตุการณ์ที่ค่าพีเอฟ s_i เกิดขึ้นร่วมกับกับค่าพีเอฟ s_j ในเฟรม t โดยที่ค่าพีเอฟ s_i มีค่าเป็น 1 (นั่นคือมีค่าการออกเสียงของค่าพีเอฟเป้าหมาย) และค่าพีเอฟ s_j มีค่าเป็น 0 (นั่นคือ ไม่มีค่าการออกเสียงของค่าพีเอฟเป้าหมาย)

3) $COPE_t(s_i, s_j) = (0, 1)$ แสดงค่าเหตุการณ์ที่ค่าพีเอฟ s_i เกิดขึ้นร่วมกับกับค่าพีเอฟ s_j ในเฟรม t โดยที่ค่าพีเอฟ s_i มีค่าเป็น 0 (นั่นคือไม่มีค่าการออกเสียงของค่าพีเอฟเป้าหมาย) และค่าพีเอฟ s_j มีค่าเป็น 1 (นั่นคือ มีค่าการออกเสียงของค่าพีเอฟเป้าหมาย)

4) $\text{COPF}_t(s_i, s_j) = (0, 0)$ แสดงค่าเหตุการณ์ที่ค่าพีเอฟ s_i เกิดขึ้นร่วมกับค่าพีเอฟ s_j ในเฟรม t โดยที่ค่าพีเอฟทั้งสองชนิด มีค่าเป็น 0 นั่นคือ ไม่มีมีค่าการออกเสียงของค่าพีเอฟเป้าหมายทั้งสองชนิด

ภาพที่ 5-10 แสดงตัวอย่างเหตุการณ์ที่เกิดขึ้นร่วมกันของค่าพีเอฟสองชนิดภายในเฟรมเดียวกัน ซึ่งพิจารณา ณ เฟรม t ให้กล่องสี่เหลี่ยมแสดงว่าค่าพีเอฟมีค่าเป็น 1 และกล่องสีขาวแสดงว่าค่าพีเอฟมีค่าเป็น 0 จุดภายในกล่องแสดงชนิดค่าพีเอฟที่ถูกกำหนดให้เป็นค่าพีเอฟหลักของการพิจารณาการปรากฏร่วมของค่าพีเอฟ รูปทางขวามือของภาพแสดงเหตุการณ์ที่ค่าพีเอฟ s_j ที่ถูกกำหนดให้เป็นค่าพีเอฟหลักของการพิจารณาการเกิดร่วมกับค่าพีเอฟอื่น ๆ ซึ่งในตัวอย่างคือค่าพีเอฟ s_i ดังนั้นโคออร์ดิเนตค่าพีเอฟ ของตัวอย่างตรงกับกรณี 1) คือ $\text{COPF}_t(s_i, s_j) = (1, 1)$ ที่มีค่าพีเอฟ s_j เกิดขึ้นร่วมกับค่าพีเอฟ s_i ในเฟรม t และทั้งสองค่าพีเอฟมีค่าเป็น 1



ภาพที่ 5-10 เหตุการณ์ที่เกิดขึ้นร่วมกันของค่าพีเอฟสองชนิดภายในเฟรมเดียวกัน

เพื่อป้องกันไม่ให้เกิดการสร้างกฎที่ไม่ยืดหยุ่น เช่น การกำหนดค่าพีเอฟชนิดใดบ้างที่สามารถพิจารณาการปรากฏร่วมกันและเพื่อให้ยืดหยุ่นต่อการนำกฎไปใช้ในภาษาอื่น ๆ ในวิทยานิพนธ์นี้ใช้โคออร์ดิเนตค่าพีเอฟที่สร้างจากกฎของเหตุการณ์การเกิดขึ้นร่วมกันของค่าพีเอฟโดยจับคู่ค่าพีเอฟที่ละสองชนิดทุกแบบ โดยจับคู่ค่าไบนารีของค่าพีเอฟทั้งสองชนิด

เนื่องจากการจับคู่ระหว่างค่าพีเอฟหลายชนิดทำให้เกิดกฎที่ไม่ยืดหยุ่น เช่น การสร้างกฎจากการปรากฏร่วมกันระหว่างค่าไบนารีของค่าพีเอฟสามชนิด หรือ การปรากฏร่วมกันของค่าไบนารีของค่าพีเอฟทั้ง 13 ชนิด เป็นต้น ทำให้ได้จำนวนของแต่ละรูปแบบน้อย ส่งผลให้การจำแนกแต่ละภาษาที่ใช้ข้อมูลที่เฉพาะเจาะจงมีความลำเอียงกับข้อมูลที่ฝึกฝนเกินไป (รูปแบบที่น้อยแสดงถึงตัวแทนเหตุการณ์การเกิดขึ้นร่วมกันของค่าพีเอฟที่เฉพาะเจาะจงเกินไป หรือเป็นข้อมูลที่ไม่มีลักษณะเป็นนัยทั่วไป) จึงไม่เหมาะสมในการนำไปสร้างตัวจำแนกกับข้อมูลทดสอบต่อไปได้ ดังนั้นการสร้างกฎจากการจับคู่ของค่าพีเอฟสองชนิด จึงเป็นการเพิ่มความยืดหยุ่นของการสร้างกฎ ลดความลำเอียงของกฎที่เกิดจากการจับคู่ค่าพีเอฟหลายชนิดและทำให้ได้ความหลากหลายของรูปแบบการเกิดร่วมกันของค่าพีเอฟ

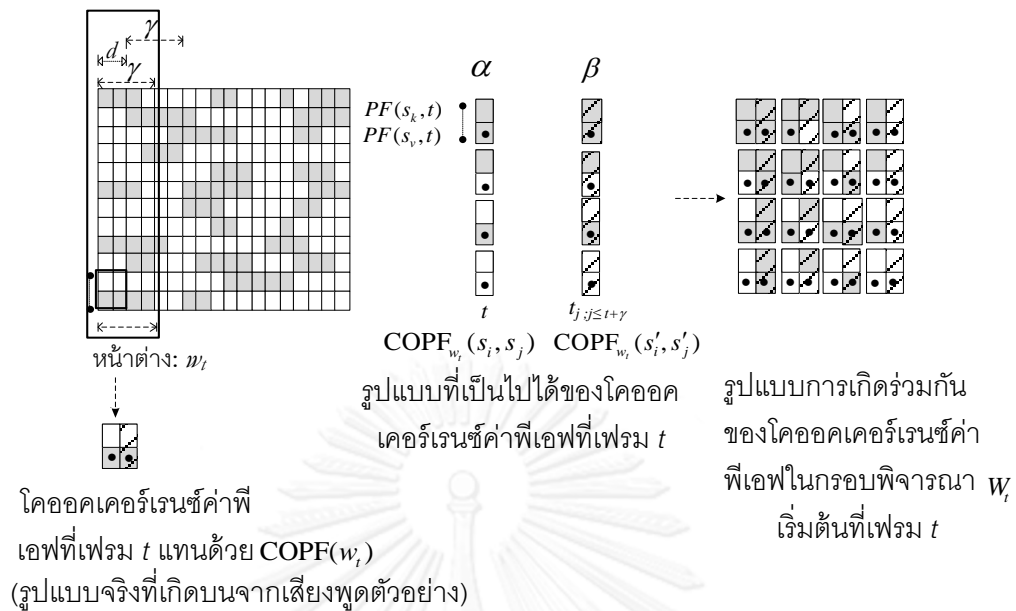
ดังนั้นเพื่อหาตัวแทนของเหตุการณ์ที่เกิดขึ้นร่วมกันของค่าพีเอฟ ควรพิจารณาการเกิดร่วมกันในแบบต่าง ๆ โดยเปลี่ยนจำนวนของชนิดค่าพีเอฟที่ต้องพิจารณาร่วมกัน แต่เนื่องจากในทางปฏิบัติการหาแบบการเกิดร่วมกันของค่าพีเอฟที่พิจารณาจากจำนวนค่าพีเอฟที่เกิดร่วมกันให้ครบทุกรูปแบบต้องใช้ทรัพยากรในการประมวลผลจำนวนมาก ผู้วิจัยพิจารณาเหตุการณ์ที่เกิดขึ้นร่วมกัน

ของระหว่างคำพืชีพสองชนิด โดยเริ่มทดลองสร้างกฎของการจับคู่ระหว่างคำพืชีพ 2 ชนิดด้วยการประกอบกันของคำไบนารี โดยจับคู่การประกอบกันของคำพืชีพที่ละสองชนิดไปจนครบการจับคู่ชนิดของคำพืชีพ

5.4.1.2 การหาตัวแทนเหตุการณ์ที่เกิดขึ้นร่วมกันของโคออกเคอร์เรนซ์คำพืชีพในกรอบการพิจารณา (Co-occurrence PF in observing window)

ในหัวข้อนี้กล่าวถึงการสร้างข้อมูลการเรียงของลำดับคำพืชีพที่แสดงการเปลี่ยนแปลงการออกเสียงบนหลายเฟรมในกรอบเวลาบนเสียงพูด ผู้วิจัยได้นำแนวคิดของการประมาณค่าด้วยไบแกรม (Bigram) ซึ่งเป็นการประมาณค่าความน่าจะเป็นของสายลำดับสิ่งที่พิจารณาที่เกิดขึ้นร่วมกันว่ามีค่าเท่ากับผลคูณของความน่าจะเป็นที่จะพบสมาชิกของสิ่งที่พิจารณา เช่น พิจารณาหน่วยเสียงทีละ 2 ตัวที่ติดกันในสายลำดับ เป็นต้น แนวคิดไบแกรมนำมาใช้ในการพิจารณาการเรียงกันของโคออกเคอร์เรนซ์คำพืชีพที่อยู่บนเสียงพูด เพื่อเป็นค่าในการบอกความแตกต่างระหว่างภาษา โดยหารูปแบบของสมาชิกของการเกิดร่วมกันของคำพืชีพที่อยู่ติดกันสองเฟรม นั่นคือ ใช้กรอบการพิจารณาเพื่อกำหนดขอบเขตของการหาสมาชิกเหตุการณ์ที่ต้องการพิจารณาร่วมกัน ซึ่งกรอบการพิจารณานี้จะอนุญาตให้มีระยะห่างระหว่างสมาชิกที่อยู่ในกรอบได้ มีสมาชิก $COPE_t(s_i, s_j)$ แทนเหตุการณ์ที่เกิดขึ้นร่วมกันของคำพืชีพที่เฟรม t และมีสมาชิก $COPE_{t+1}(s_i, s_j)$ แทนเหตุการณ์ที่เกิดขึ้นร่วมกันของคำพืชีพที่อยู่ในเฟรมถัดไปจากเฟรม t โดยสมาชิกทั้งสองนั้นอยู่ในกรอบการพิจารณาเดียวกัน และสมาชิกที่อยู่ในเฟรมถัดไปจากเฟรม t ไม่จำเป็นต้องเป็นเฟรมที่อยู่ติดกับเฟรม t แต่มีเงื่อนไขว่าต้องอยู่ในขอบเขตกรอบการพิจารณาที่กำหนด ความกว้างระยะห่างระหว่างสมาชิกมีผลต่อรูปแบบเหตุการณ์การเกิดร่วมกันของคำพืชีพ ดังนั้นผู้วิจัยจึงได้ทดลองผลกระทบของความกว้างของระยะห่างระหว่างสมาชิกซึ่งอธิบายในหัวข้อที่ 6.4.4

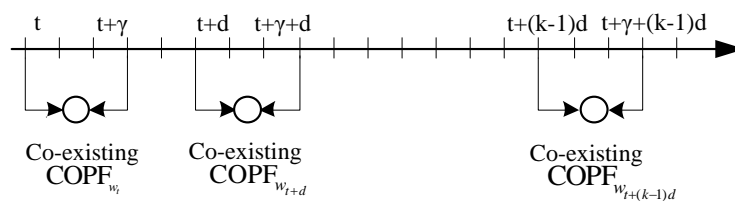
กำหนดให้ $co-existing_{w_t}(\alpha, \beta)$ แทนการเกิดร่วมกันของสมาชิก α และ β ที่เกิดขึ้นในกรอบการพิจารณาซึ่งเริ่มต้นที่เฟรม t มีสมาชิก α และ β เป็นโคออกเคอร์เรนซ์คำพืชีพในเฟรม t และเฟรมที่อยู่ถัดไปจากเฟรม t ตามลำดับ เช่น กำหนดให้การเกิดร่วมกันของโคออกเคอร์เรนซ์คำพืชีพ $co-existing_{w_t}(\alpha, \beta) = (COPE_{w_t}(s_i, s_j), COPE_{w_{t+1}}(s'_i, s'_j))$ แสดงการเกิดร่วมกันของสมาชิก α และ β เมื่อสมาชิก α เป็นโคออกเคอร์เรนซ์คำพืชีพ $COPE_{w_t}(s_i, s_j)$ บนเฟรม t และสมาชิก β เป็นโคออกเคอร์เรนซ์คำพืชีพ $COPE_{w_t}(s'_i, s'_j)$ บนเฟรม $t+1$ โดยที่ทั้งสองเฟรมอยู่ในกรอบการพิจารณา w_t ภาพที่ 5-11 แสดงการเกิดร่วมกันของสมาชิกที่อยู่ในกรอบการพิจารณา กล่องที่มีจุดและกล่องที่เป็นช่องว่างที่อยู่ติดกันในแนวตั้ง แสดงโคออกเคอร์เรนซ์คำพืชีพที่อยู่ในเฟรม t ซึ่งในภาพเป็นตัวอักษรของโคออกเคอร์เรนซ์คำพืชีพของคำพืชีพ s_i และคำพืชีพ s_j เส้นแถบเอียงแสดง โคออกเคอร์เรนซ์คำพืชีพที่อยู่ในเฟรมถัดไปจากเฟรม t และรูปทางขวา แสดงรูปแบบที่เกิดขึ้นได้ทั้งหมด 16 แบบของการเกิดร่วมกันของโคออกเคอร์เรนซ์คำพืชีพ



ภาพที่ 5-11 แสดงการเกิดร่วมกันของสมาชิกโคออกเคอร์เรนซ์ค่าพีเอฟที่อยู่ในกรอบการพิจารณา

กำหนดให้ กรอบการพิจารณา w ใช้พิจารณาโคออกเคอร์เรนซ์ค่าพีเอฟที่อยู่ในแต่ละเฟรมภายในกรอบหน้าตาต่าง โดยมีพารามิเตอร์ γ แทนขนาดของกรอบการพิจารณา ซึ่งความกว้างของกรอบหน้าตาต่าง w สามารถคำนวณได้ดังสมการที่ (5.13) ซึ่งมีขอบเขตการพิจารณาโคออกเคอร์เรนซ์ค่าพีเอฟตั้งแต่เฟรม t ถึงเฟรม $t + \gamma$ และภาพที่ 5-12 แสดงกรอบการพิจารณาโคออกเคอร์เรนซ์ค่าพีเอฟที่อยู่ในแต่ละเฟรมภายในขอบเขตที่กำหนด มีค่าพารามิเตอร์ γ แทนขนาดของกรอบการพิจารณา ค่าพารามิเตอร์ d แทนขนาดของการเลื่อนของกรอบการพิจารณา และค่าพารามิเตอร์ k แทนจำนวนกรอบการพิจารณาที่เลื่อนไปคำนวณการเกิดร่วมกันของโคออกเคอร์เรนซ์ค่าพีเอฟที่อยู่ตำแหน่งสุดท้ายของเสียงพูด เมื่อใช้กรอบการพิจารณารูปแบบการเกิดร่วมกันของโคออกเคอร์เรนซ์ค่าพีเอฟจนสิ้นสุดความยาวของเสียงพูดแล้ว จึงหาความถี่ของเหตุการณ์การเกิดร่วมกันของโคออกเคอร์เรนซ์ค่าพีเอฟได้

$$w_t = \{t \mid 0 \leq t \leq t + \gamma\}; \gamma \geq 1 \tag{5.13}$$



ภาพที่ 5-12 แสดงกรอบการพิจารณาโคออกเคอร์เรนซ์ค่าพีเอฟที่อยู่ในแต่ละเฟรมภายในขอบเขตที่กำหนด

ผลลัพธ์ของการใช้กรอบการพิจารณาทำให้ได้รูปแบบการเกิดร่วมกันของโคออคเคอร์เรนซ์ค่าพีเอฟที่อยู่ภายในกรอบหน้าต่าง แล้วนำมาหาความถี่ของรูปแบบของการเกิดร่วมกัน จากตัวอย่าง ให้พิจารณาเหตุการณ์การเกิดร่วมกันของ α และ β ภายในกรอบหน้าต่าง ดังนั้นความถี่ของเหตุการณ์เกิดร่วมกันของโคออคเคอร์เรนซ์ค่าพีเอฟสามารถคำนวณได้จากสมการที่ (5.14) ให้ $f(\text{co-existing}_{w_t}(\alpha, \beta))$ แทนความถี่ของเหตุการณ์เกิดร่วมกันของสมาชิก α และ β ที่ปรากฏร่วมกันในกรอบการพิจารณา และมีความถี่ของสมาชิก α และ β เป็น $f(\alpha)$ และ $f(\beta)$ ตามลำดับ ดังนั้นความถี่ของเหตุการณ์เกิดร่วมกันของสมาชิก α และ β เป็น 1 เมื่อทั้งสองเหตุการณ์เกิดขึ้นพร้อมกัน นั่นคือ ความถี่ของโคออคเคอร์เรนซ์ค่าพีเอฟของสมาชิก α และ β มากกว่า 1 เป็นต้น

$$f(\text{co-existing}_{w_t}(\alpha, \beta)) = \begin{cases} 1; & \sum_{r=t}^{t+\gamma} f(\alpha) \geq 1 \text{ and } \sum_{r=t+1}^{t+\gamma} f(\beta) \geq 1 \\ 0; & \text{otherwise} \end{cases} \quad (5.14)$$

where $r \in w, t \leq r \leq r + \gamma$

ตัวอย่างเช่น $f(\text{co-existing}_{w_t}(\text{COPF}_{w_t}(s_i, s_j), \text{COPF}_{w_t}(s_i, s_o)))$ แทนความถี่ของการเกิดร่วมกันของเหตุการณ์ที่มีสมาชิก $\text{COPF}_{w_t}(s_i, s_j)$ ปรากฏร่วมกับสมาชิก $\text{COPF}_{w_t}(s_i, s_o)$ ที่อยู่ในกรอบการพิจารณา w_t ซึ่งคำนวณดังสมการ (5.15) เมื่อกำหนดให้ $f(\text{COPF}_{w_t}(s_i, s_j))$ และ $f(\text{COPF}_{w_t}(s_i, s_o))$ แทนความถี่ของการเกิดร่วมกันของโคออคเคอร์เรนซ์ค่าพีเอฟ $\text{COPF}_{w_t}(s_i, s_j)$ และ $\text{COPF}_{w_t}(s_i, s_o)$ ในกรอบการพิจารณา w_t

$$f(\text{co-existing}_{w_t}(\text{COPF}_{w_t}(s_i, s_j), \text{COPF}_{w_t}(s_i, s_o))) = \begin{cases} 1 & \sum_{r=t}^{t+\gamma} f(\text{COPF}_{w_t}(s_i, s_j)) \geq 1 \\ & \text{and } \sum_{r=t+1}^{t+\gamma} f(\text{COPF}_{w_t}(s_i, s_o)) \geq 1 \\ 0 & \text{otherwise} \end{cases} \quad (5.15)$$

where $r \in w, t \leq r \leq t + \gamma$

เมื่อหาความถี่ของการเกิดร่วมกันของโคออคเคอร์เรนซ์ค่าพีเอฟทุกเหตุการณ์บนเสียงพูดแล้ว นำแต่ละค่ามาประกอบกันเป็นเวกเตอร์ค่าลักษณะเฉพาะของความถี่ของการเกิดร่วมกันของเหตุการณ์ เพื่อสร้างเป็นซูเปอร์เวกเตอร์ (Supervector) สำหรับฝึกฝนตัวจำแนกภาษาต่อไป

จากการจับคู่ที่ละสองค่าพีเอฟจากค่าพีเอฟ 13 ชนิด ทำให้ได้รูปแบบโคออคเคอร์เรนซ์ค่าพีเอฟ $C(13,2)$ เท่ากับ 78 แบบ โดยจำนวนรูปแบบของเหตุการณ์ที่เกิดร่วมกันระหว่างโคออคเคอร์เรนซ์ค่าพีเอฟที่ปรากฏในสองเฟรม ที่เกิดจากโคออคเคอร์เรนซ์ค่าพีเอฟสมาชิกที่หนึ่งในเฟรม t และสมาชิกที่สองอยู่ในเฟรมถัดไปจากเฟรม t มีการเกิดร่วมกันของค่าไบนารีทั้งหมด 16 แบบ ดังนั้นโอกาสเกิดเหตุการณ์ทั้งหมดเป็น $78 \times 78 \times 16 = 97,344$ แบบ ซึ่งจำนวนของการจับคู่ของเหตุการณ์การเกิดร่วมกันจากการจับคู่ค่าพีเอฟ 13 ค่า มีจำนวนมาก ทำให้มิติของเวกเตอร์ค่าลักษณะเฉพาะมีมาก ส่งผลให้เกิดการใช้เวลาในการคำนวณและประมวลผลอย่างมาก เพื่อลดปัญหาจำนวนมิติของเวกเตอร์ที่มีจำนวนมากนี้ ผู้วิจัยจึงเสนอการใช้กฎการจับคู่ค่าพีเอฟที่ละสองชนิดโดย

กำหนดค่าพีเอพหลักของการพิจารณาการเกิดร่วมกับค่าพีเอพอื่น ๆ ที่ละชนิดและสังเกตผลกระทบของค่าพีเอพแต่ละชนิดว่ามีผลกระทบต่อความสามารถในการจำแนกภาษาหรือไม่

การพิจารณาการเกิดร่วมกันของโคออกเคอร์เรนซ์ค่าพีเอพ โดยกำหนดให้พิจารณาทีละค่าพีเอพ s_i ที่เกิดขึ้นร่วมกับค่าพีเอพชนิดอื่น จากนั้นคำนวณความถี่ของเหตุการณ์เกิดร่วมกันของโคออกเคอร์เรนซ์ค่าพีเอพที่มีค่าพีเอพ s_i เกิดขึ้นร่วมกับค่าพีเอพอื่น ๆ เช่น การพิจารณาโคออกเคอร์เรนซ์ค่าพีเอพ Vocalic เกิดร่วมกับค่าพีเอพอื่น ๆ ได้แก่ ค่าพีเอพ Vocalic คู่กับค่าพีเอพ Consonantal, ค่าพีเอพ Vocalic เกิดคู่กับค่าพีเอพ Nasal และจับคู่ค่าพีเอพ Vocalic ร่วมกับชนิดอื่นไปจนครบทุกชนิดของค่าพีเอพ ดังนั้นจำนวนโคออกเคอร์เรนซ์ค่าพีเอพที่มีค่าพีเอพ Vocalic เกิดในทั้งสองเฟรมเกิดร่วมกับค่าพีเอพชนิดอื่น ๆ มีทั้งหมด 12 แบบ และจำนวนรูปแบบการเกิดร่วมกันที่มีโคออกเคอร์เรนซ์ค่าพีเอพ Vocalic คู่กับชนิดอื่น (ในเฟรมแรก) ปรากฏร่วมกับโคออกเคอร์เรนซ์ค่าพีเอพ Vocalic ร่วมกับค่าอื่น (ในเฟรมถัดไป) มีทั้งหมด $12 \times 12 \times 16 = 2,304$ แบบ

เวกเตอร์ค่าความถี่ของการเกิดร่วมกันของโคออกเคอร์เรนซ์ค่าพีเอพประกอบด้วย 2,304 มิติ (ที่ได้จากหัวข้อ 5.4.1.2) ถูกนำมาทำให้เป็นบรรทัดฐาน (หรือนอร์มอลไลซ์) เพื่อป้องกันไม่ให้เกิดความลำเอียงที่มาจากความยาวของแต่ละส่วนเสียงพูดที่ไม่เท่ากัน การทำนอร์มอลไลซ์ความถี่ของแต่ละพจน์ในเวกเตอร์นั้นเป็นการลดข้อแตกต่างของแต่ละพจน์ที่เกิดขึ้นบนสัญญาณเสียงพูด รายละเอียดของการทำให้เป็นบรรทัดฐานได้อธิบายในหัวข้อ 5.4.2

นอกจากนี้ผู้วิจัยได้ทำการทดลองการเปรียบเทียบประสิทธิภาพของการจำแนกภาษา เมื่อใช้การเกิดร่วมกันของโคออกเคอร์เรนซ์ค่าพีเอพที่นำเสนอในหัวข้อ 6.4.4 และผู้วิจัยได้ขยายการทดลองเพื่อศึกษาผลกระทบของจำนวนค่าพีเอพที่นำมาประกอบกัน โดยเพิ่มจำนวนค่าพีเอพที่พิจารณาการเกิดร่วมกันของสองโคออกเคอร์เรนซ์ค่าพีเอพภายในแต่ละเฟรม เช่น กำหนดให้พิจารณาสมาชิกภายในเฟรม t ประกอบด้วย โคออกเคอร์เรนซ์ค่าพีเอพ Vocalic เกิดคู่กับค่าพีเอพชนิดอื่น (ประกอบด้วย ค่าพีเอพ Vocalic เกิดคู่กับค่าพีเอพ Consonantal, ค่าพีเอพ Vocalic เกิดคู่กับค่าพีเอพ Nasal และจับคู่ค่าพีเอพ Vocalic กับชนิดอื่นจนครบทุกชนิด) และโคออกเคอร์เรนซ์ค่าพีเอพ Nasal เกิดคู่กับค่าพีเอพชนิดอื่น (ประกอบด้วยค่าพีเอพ Nasal เกิดคู่กับค่าพีเอพ Consonantal ค่าพีเอพ Nasal เกิดคู่กับค่าพีเอพ Voice และจับคู่ค่าพีเอพ Nasal คู่กับชนิดอื่นจนครบทุกชนิด) และสมาชิกที่เหมือนกัน ที่ปรากฏในเฟรมถัดจากเฟรม t เพื่อหาการเกิดร่วมกันของสมาชิกทั้งสองตัวภายในกรอบการพิจารณาเดียวกัน เป็นต้น ซึ่งการทดลองเพิ่มจำนวนโคออกเคอร์เรนซ์ค่าพีเอพที่พิจารณานี้ ทำให้ได้รูปแบบที่พิจารณาการเกิดร่วมกันของการออกเสียงหลายชนิดที่ปรากฏในแต่ละเฟรมเพิ่มขึ้น อย่างไรก็ตามผู้วิจัยต้องการศึกษาผลกระทบของการเพิ่มขึ้นของรูปแบบโคออกเคอร์เรนซ์ค่าพีเอพที่พิจารณา แต่ไม่ได้มุ่งทดลองเพื่อหาจำนวนค่าพีเอพที่เหมาะสมในการพิจารณารูปแบบของการเกิดร่วมกันของค่าพีเอพ จึงไม่ได้ทดลองการเพิ่มจำนวนที่พิจารณาการเกิดร่วมกันของค่าพีเอพให้ครบทุกกรณี

5.4.2 การถ่วงน้ำหนักและทำให้เป็นบรรทัดฐาน

ในการจำแนกภาษาเหตุการณ์ที่ปรากฏในเสียงพูดน้อยครั้งมีความสำคัญมากกว่าเหตุการณ์ที่ปรากฏทั่วไปในทุกเสียงพูด การถ่วงน้ำหนักแต่ละเหตุการณ์ให้มีความสำคัญแตกต่างกันจึงสามารถเพิ่มประสิทธิภาพของฟังก์ชันการจำแนกให้สามารถบอกความแตกต่างระหว่างภาษาได้ดีขึ้น

เมื่อได้เวกเตอร์ค่าความถี่การเกิดร่วมกันของโคออกเคอร์เรนซ์ค่าพีเอฟจากหัวข้อ 5.4.1.2 พบว่าค่าความถี่ของการเกิดร่วมกันของโคออกเคอร์เรนซ์ค่าพีเอฟบางเหตุการณ์มีค่าสูงกว่าบางเหตุการณ์ นั่นคือ ค่าของสมาชิกในเวกเตอร์มีค่าแตกต่างกันมาก ทำให้สมาชิกแต่ละตัวในเวกเตอร์มีอิทธิพลในการสร้างระนาบการแบ่งข้อมูลออกเป็นกลุ่มของเคอร์เนลฟังก์ชันของซัพพอร์ตเวกเตอร์แมชชีน ทำให้ระนาบของฟังก์ชันการจำแนกของปริภูมิสมมติฐานที่ได้มีคุณสมบัติแตกต่างจากฟังก์ชันการจำแนกที่ได้จากสมาชิกอื่นในเวกเตอร์ ส่งผลต่อความสามารถในการจำแนกข้อมูลได้ ดังนั้นการถ่วงน้ำหนักของสมาชิกแต่ละตัวในเวกเตอร์ค่าความถี่ของการเกิดร่วมกันของโคออกเคอร์เรนซ์ค่าพีเอฟจึงได้มาไว้ในหัวข้อนี้

ฟังก์ชันการถ่วงน้ำหนัก ที่ใช้กับเวกเตอร์ค่าความถี่ของการเกิดร่วมกันของโคออกเคอร์เรนซ์ค่าพีเอฟคือ ค่าความถี่เทอม (tf) ร่วมกับค่าความถี่เอกสารผกผัน (idf) ซึ่งรายละเอียดของฟังก์ชันการถ่วงน้ำหนักอธิบายในหัวข้อ 4.4.1 โดยค่าความถี่เทอมใช้แสดงค่าน้ำหนักแบบท้องถิ่นซึ่งมี $tf(e_i, d_j)$ แทนค่าความถี่ของเหตุการณ์การเกิดร่วมกันของโคออกเคอร์เรนซ์ค่าพีเอฟแทน e_i บนเสียงพูด d_j ค่าความถี่เทอมของสมาชิกในเวกเตอร์ที่มีค่ามากอธิบายลักษณะเฉพาะและความสำคัญของสมาชิกนั้นในเสียงพูดนั้น ขณะที่ค่าความถี่เอกสารผกผันแสดงค่าน้ำหนักแบบครอบคลุมซึ่งแทนด้วย $idf(e_i)$ โดยค่านี้อธิบายความสำคัญของสมาชิกนั้นสะท้อนจำนวนเสียงพูดทั้งหมดที่อยู่ในฐานข้อมูลเสียง นั่นคือสามารถคำนวณได้จาก $idf(e_i) = \log(N / f(e_i))$ เมื่อ N แทนจำนวนเสียงพูดทั้งหมดทุกภาษาและ $f(e_i)$ แทนจำนวนเสียงพูดที่มีเหตุการณ์นั้นปรากฏ โดยค่าน้ำหนักความถี่เอกสารผกผันที่มีค่าต่ำแสดงว่าสมาชิกนั้นมีความสำคัญมาก

นอกจากนี้เวกเตอร์ค่าความถี่ของการเกิดร่วมกันของโคออกเคอร์เรนซ์ค่าพีเอฟถูกนำมาทำให้เป็นบรรทัดฐาน เพื่อป้องกันไม่ให้เกิดความลำเอียงที่มาจากความยาวของแต่ละส่วนเสียงพูดที่ไม่เท่ากัน เป็นการลดข้อแตกต่างของความสามารถของการจำแนกของสมาชิกแต่ละพจน์ในเวกเตอร์

5.4.3 ตัวจำแนกภาษา

ระบบการรู้จำภาษาพูดที่นำเสนอนี้ได้นำซัพพอร์ตเวกเตอร์แมชชีนมาใช้ในรูปแบบจำลองของการจำแนกภาษาแบบหนึ่งต่อหนึ่ง โดยทำการจำแนกภาษาแบบทีละคู่ภาษาจึงมีตัวจำแนกทั้งหมด $C(8, 2)$ เท่ากับ 28 ตัว ในการฝึกฝนตัวจำแนกนั้นจะใช้เวกเตอร์ค่าความถี่ของการเกิดร่วมกันของโคออกเคอร์เรนซ์ค่าพีเอฟเป็นข้อมูลสำหรับการสร้างระนาบการแบ่งกลุ่มของฟังก์ชันเคอร์เนล ซึ่งฝึกฝนจากข้อมูลที่เป็นภาษาเป้าหมายหนึ่งภาษาให้เป็นที่กลุ่มตัวอย่างบวก และข้อมูลที่ไม่ใช่ภาษาเป้าหมายหนึ่งภาษาให้เป็นที่กลุ่มตัวอย่างลบ

ซัพพอร์ตเวกเตอร์แมชชีนสร้างระนาบแบ่งกลุ่มข้อมูลด้วยสมการที่ (5.16) โดยใช้ฟังก์ชันเคอร์เนล แบบเชิงเส้น $K(x, x_i) = x^T x_i + C$ ซึ่งฟังก์ชันเคอร์เนลเป็นระบบสมการเชิงเส้นมีค่าเท่ากับการคูณกันของสองเวกเตอร์ x และ x_i และมีค่าคงที่ C จากสมการมีฟังก์ชันค่าน้ำหนัก $w = \sum_{i=1}^n \alpha_i x$ ซึ่งเท่ากับค่าผลรวมของการคูณของน้ำหนักและเวกเตอร์

$$f(x) = \sum_{i=1}^N \alpha_i x^T x_i + b, \quad (5.16)$$

ในการสร้างตัวจำแนกภาษาด้วยซัพพอร์ตเวกเตอร์แมชชีน ผู้วิจัยใช้ SVM^{light} [112] เป็นเครื่องมือในการสร้างตัวจำแนก ผู้วิจัยได้ศึกษาความสามารถของการจำแนกซัพพอร์ตเวกเตอร์แมชชีน ที่มีฟังก์ชันเคอร์เนลแบบต่าง ๆ ได้แก่ ฟังก์ชันเคอร์เนลแบบเชิงเส้นและเรเดียลเบสิสฟังก์ชัน (Radial Basis Function: rbf) ร่วมกับพารามิเตอร์ค่า g ที่ต่างกัน เช่น 0.001, 0.01, 0.1, 1 และ 10 เป็นต้น ผลการจำแนกของซัพพอร์ตเวกเตอร์แมชชีนพบว่าประสิทธิภาพการจำแนกซัพพอร์ตเวกเตอร์แมชชีน ที่มีฟังก์ชันเคอร์เนลแบบเชิงเส้นให้ค่าความถูกต้องของการจำแนกสูงกว่า ดังนั้นในการทดลองนี้จึงเลือกใช้ฟังก์ชันเคอร์เนลแบบเชิงเส้น เพื่อสร้างการจำแนกภาษาด้วยซัพพอร์ตเวกเตอร์แมชชีน

5.4.4 การรวมตัวจำแนกภาษา

ผลลัพธ์ที่ได้จากการจำแนกภาษาจากซัพพอร์ตเวกเตอร์แมชชีน นำไปใช้ในการรวมคะแนนเพื่อระบุภาษาของระบบแนวทางนี้ โดยใช้การทำงานส่วนหลังด้วยแบบจำลองเกาส์เซียนตามที่อธิบายในหัวข้อ 3.4.2

บทที่ 6

การทดลองและวิธีการวัดผลการทดลอง

ในบทนี้จะกล่าวถึงวิธีการทดลองแบบต่าง ๆ เพื่อวัดผลของการรู้จำภาษา โดยทดสอบกับชุดข้อมูลทดสอบที่ได้จากการตรวจสอบแบบไขว้ ซึ่งหัวข้อ 6.1. อธิบายการสร้างระบบตรวจหาลักษณะเฉพาะค่าพีเอฟ 6.2 อธิบายระบบอ้างอิงซึ่งประกอบด้วย ระบบการรู้จำภาษาพูดตามแนวทางอะคูสติกแบบใช้ไอเวกเตอร์ ระบบการรู้จำภาษาพูดตามแนวทางการเรียงตัวหน่วยเสียง และระบบการรู้จำภาษาพูดที่ใช้ค่าพีเอฟ 6.3 การศึกษาความแปรปรวนของค่า PF ในหน่วยเสียง และ 6.4 การทดลองเพื่อประเมินประสิทธิภาพของระบบการรู้จำภาษาพูด

6.1 การสร้างระบบตรวจหาลักษณะเฉพาะค่าพีเอฟ

ผู้วิจัยใช้ฐานข้อมูลเสียง TIMIT (อธิบายในบทที่ 3) เพื่อฝึกฝนระบบตรวจหาลักษณะเฉพาะค่าพีเอฟ โดยเตรียมข้อมูลอินพุตส่วนฉลากของข้อมูลฝึกฝน จากการแปลงฉลากหน่วยเสียงที่กำกับสัญญาณเสียงพูดให้เป็นฉลากของค่าลักษณะเฉพาะค่าพีเอฟ 13 ค่าที่นิยามตาม SPE และเตรียมข้อมูลส่วนค่าลักษณะสำคัญ ในรูปของเวกเตอร์ค่าลักษณะสำคัญที่มี 39 ค่า ประกอบด้วยค่าสัมประสิทธิ์ MFCC 12 ค่า ค่าพลังงานเสียง 1 ค่า และ 26 ค่าจากค่า MFCC และค่าพลังงานเสียงที่คำนวณการเปลี่ยนแปลงเซปสตรัมทางเวลาจากสมการอนุพันธ์อันดับหนึ่ง (Delta) และอนุพันธ์อันดับสอง (Delta-delta) MFCC โดยในขั้นตอนการสกัดค่าลักษณะสำคัญนี้ใช้กรอบการพิจารณา (Window) 25 มิลลิวินาที และมีการซ้อนทับของเฟรม (Frame shift) 10 มิลลิวินาที เวกเตอร์จากการสกัดค่าลักษณะสำคัญจากสัญญาณเสียงพูด นำไปใช้เป็นฝึกฝนระบบตรวจหาลักษณะเฉพาะค่าพีเอฟด้วยโครงข่ายประสาทเทียมหลายชั้น โดยใช้วิธีการเรียนรู้แบบแพร่กระจายกลับ ลักษณะของโครงข่ายประสาทเทียมหลายชั้น ประกอบด้วย 3 ชั้น คือ ชั้นข้อมูลเข้ามี 39 โหนด โดยนำเวกเตอร์ที่ได้จากการสกัดค่าลักษณะสำคัญเป็นอินพุต ชั้นที่ 2 เป็นชั้นซ่อน ซึ่งการกำหนดจำนวนของชั้นซ่อนและจำนวนโหนดของแต่ละชั้นซ่อนควรให้เหมาะสมกับข้อมูล ซึ่งได้ทำการทดลองเพื่อหาค่าพารามิเตอร์ที่เหมาะสม และชั้นที่ 3 เป็นชั้นผลลัพธ์ ซึ่งกำหนดให้เป็นผลลัพธ์ของการประมาณค่าความน่าจะเป็นของค่าพีเอฟจำนวน 2 ค่าคือค่า 1 แทนค่าที่เหมือนกับค่าพีเอฟเป้าหมายและค่า 0 แทนค่าที่ไม่เหมือนกับค่าพีเอฟเป้าหมาย สำหรับการทดลองเพื่อหาจำนวนโหนดซ่อนที่เหมาะสมพิจารณาจากจำนวนโหนดซ่อนที่ให้ความแม่นยำของการตรวจหาค่าพีเอฟที่มีค่าสูงที่สุด โดยปรับค่าพารามิเตอร์ให้มีจำนวนโหนดซ่อนเป็น 100 150 200 250 และ 300 สำหรับการฝึกฝนแบบโครงข่ายประสาทเทียม และเปรียบเทียบความแม่นยำของการตรวจหาค่าพีเอฟเมื่อมีค่าพารามิเตอร์ที่แตกต่างกัน สำหรับการประเมินประสิทธิภาพของระบบการตรวจหาค่าพีเอฟ วัดความแม่นยำของการตรวจหา โดยทดสอบจากการสกัดเวกเตอร์ค่าลักษณะสำคัญ 39 ค่า ดังข้างต้น จากเสียงพูดในชุดข้อมูลทดสอบ และนำมาเป็นอินพุตของการทดสอบระบบตรวจหาค่าพีเอฟแต่ละชนิด จากนั้นวัดประสิทธิภาพของระบบตรวจหาค่าพีเอฟแต่ละชนิด และเปรียบเทียบค่าเฉลี่ยของความแม่นยำ เมื่อได้ระบบตรวจหาค่าพีเอฟที่มีความแม่นยำสูงจึงนำมาเป็นระบบตรวจหาค่าพีเอฟเพื่อสกัดค่าลักษณะสำคัญเพื่อนำไปใช้เป็นอินพุตของระบบการรู้จำภาษาพูดที่ใช้ค่าพีเอฟต่อไป ผลลัพธ์ที่ได้จากระบบ

ตรวจหาค่าพีเอฟแต่ละชนิด ใช้แทนค่าพีเอฟแต่ละชนิด และนำมาประกอบกันเป็นเวกเตอร์ค่าลักษณะสำคัญพีเอฟ ซึ่งสำหรับแต่ละเฟรมของเวกเตอร์ ประกอบด้วยค่าความน่าจะเป็นที่เหมือนค่าพีเอฟเป้าหมาย และค่าที่ไม่เหมือนค่าพีเอฟเป้าหมาย ของแต่ละค่าพีเอฟ 13 ชนิดรวมเป็นเวกเตอร์ที่มีจำนวนมิติ 26 ค่า

6.2 ระบบอ้างอิง

ระบบอ้างอิง (Baseline Systems) ที่ใช้สำหรับการเปรียบเทียบประสิทธิภาพของระบบการรู้จำภาษาพูดที่นำเสนอ เพื่อนำไปสู่การพัฒนาการรู้จำภาษาพูดที่มีประสิทธิภาพดีขึ้น จึงเปรียบเทียบการประเมินประสิทธิภาพระบบการรู้จำภาษาพูดที่นำเสนอกับระบบอ้างอิง ประกอบด้วย 3 แบบ ดังนี้

ระบบอ้างอิงที่ 1 สำหรับระบบการรู้จำภาษาพูดตามแนวทางเกณฑ์การเรียงหน่วยเสียง ซึ่งแบ่งเป็น 2 ส่วน ดังนี้

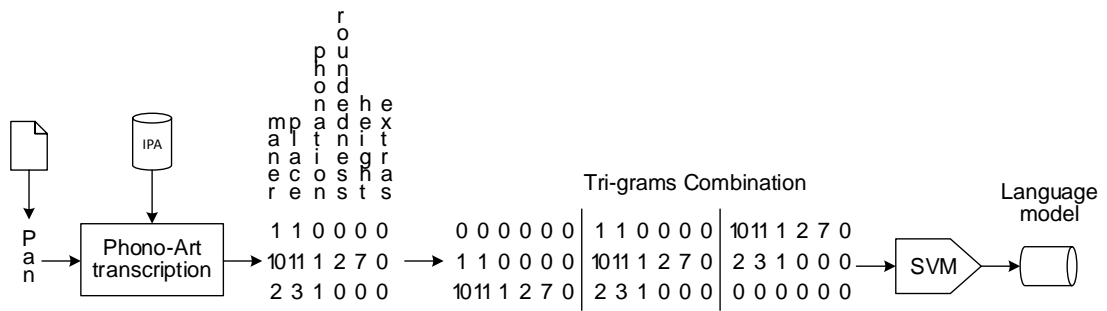
ระบบอ้างอิงที่ 1.1 คือ ระบบการรู้จำภาษาพูดแบบพ็อดาร์วีเอสเอ็ม ระบบอ้างอิงนี้ใช้เพื่อประเมินประสิทธิภาพของระบบการรู้จำภาษาพูดตามแนวทางการเรียงตัวของหน่วยเสียงที่ใช้แบบจำลองเวกเตอร์สเปซ โดยระบบอ้างอิงที่ 1.1 นี้ใช้เปรียบเทียบกับระบบการรู้จำภาษาพูดแบบแลตทิซ-วีเอสเอ็ม เพื่อพิจารณาผลกระทบจากการใช้หน่วยเสียงแตกต่างกัน

ระบบอ้างอิงที่ 1.2 คือ ระบบการรู้จำภาษาพูดแบบพ็อดาร์วีเอสเอ็ม หรือระบบการรู้จำภาษาพูดแบบแลตทิซ-วีเอสเอ็ม ที่ใช้ฟังก์ชันถ่วงน้ำหนักที่มาจากการประกอบกันของค่าความถี่เทอม และค่าความถี่เอกสารผกผัน เนื่องจากฟังก์ชันถ่วงน้ำหนักที่มาจากการประกอบกันของค่าความถี่เทอม และค่าความถี่เอกสารผกผัน เป็นฟังก์ชันการถ่วงน้ำหนักที่ยอมรับในงานสืบค้นเอกสาร ระบบอ้างอิงนี้ถูกใช้เพื่อพิจารณาความสามารถการใช้ฟังก์ชันถ่วงน้ำหนัก เมื่อใช้ร่วมกับแบบจำลองเวกเตอร์สเปซในระบบการรู้จำภาษาพูดตามแนวทางเกณฑ์การเรียงหน่วยเสียง

ระบบอ้างอิงที่ 2 สำหรับเปรียบเทียบประสิทธิภาพของระบบการรู้จำภาษาพูดตามการเรียงค่าพีเอฟ แบ่งเป็น 2 ส่วน ดังนี้

ระบบอ้างอิงที่ 2.1 คือ ระบบการรู้จำภาษาพูดแบบพ็อดาร์วีเอสเอ็ม ซึ่งเป็นที่ยอมรับในระบบการรู้จำภาษาพูดตามแนวทางการเรียงหน่วยเสียง ระบบอ้างอิงนี้ ถูกใช้เพื่อพิจารณาผลกระทบของข้อมูลที่ได้จากการเรียงลำดับเสียงที่แตกต่างกัน เมื่อตัวแทนเสียงต่างกัน โดยระบบการรู้จำภาษาพูดแบบพ็อดาร์วีเอสเอ็ม ใช้หน่วยเสียง ถูกเปรียบเทียบกับระบบการรู้จำภาษาพูดที่เสนอที่ใช้การปรากฏกันของค่าพีเอฟ

ระบบอ้างอิงที่ 2.2 คือ ระบบการรู้จำภาษาพูดของ Nunez และ Ramon [92] ซึ่งเป็นวรรณกรรมล่าสุดทางด้าน phonotactic ที่ใช้ค่าพีเอฟ ระบบอ้างอิงนี้ใช้ลำดับเสียงแบบไตรแกรมของค่าการออกเสียงที่ปรากฏบนหน่วยเสียง (แบบ phono-articulatory) เพื่อสร้างแบบจำลองภาษาขั้นตอนการทำงานของระบบการรู้จำภาษาพูดแสดงดังภาพที่ 6-1



ภาพที่ 6-1 ระบบการระบุภาษาที่ใช้ค่าพีเอฟของ Nunez และ Ramon

ระบบอ้างอิงที่ 3 คือระบบการรู้จำภาษาพูดตามแนวทางอะคูสติกแบบใช้ไอเวกเตอร์ (i-vector) เพื่อเปรียบเทียบประสิทธิภาพของระบบการรู้จำภาษาพูดโดยใช้การจำแนกด้วยแบบจำลองลาเทนท์ไดนามิคคอนดิชันนอลแรนดอมฟิลด์ร่วมกับค่าพีเอฟ

เนื่องจากระบบการรู้จำภาษาพูดโดยใช้การจำแนกด้วยแบบจำลองลาเทนท์ไดนามิคคอนดิชันนอลแรนดอมฟิลด์ร่วมกับค่าพีเอฟ ที่นำเสนอเป็นการสร้างแบบจำลองเพื่อการจำแนกภาษาจากค่าลักษณะสำคัญภาษาโดยตรง ซึ่งมีแนวทางคล้ายกับระบบการรู้จำภาษาพูดแบบอะคูสติก เพื่อประเมินประสิทธิภาพของระบบการรู้จำภาษาพูดที่นำเสนอ จึงเลือกระบบการรู้จำภาษาพูดตามแนวทางอะคูสติกที่เป็นที่ยอมรับในการรู้จำภาษาในปัจจุบัน ได้แก่ ระบบการรู้จำภาษาพูดตามแนวทางอะคูสติกแบบใช้การวิเคราะห์ปัจจัยร่วมกัน (Joint factor analysis: JFA) และระบบการรู้จำภาษาพูดตามแนวทางอะคูสติกแบบใช้ไอเวกเตอร์ เป็นต้น

ระบบอ้างอิงที่ 3 ใช้ระบบการรู้จำภาษาพูดตามแนวทางอะคูสติกแบบใช้ไอเวกเตอร์จากค่าสัมประสิทธิ์สเปกตรอล โดยวิธีไอเวกเตอร์มีแนวคิดมาจากการประสบความสำเร็จของการวิเคราะห์ปัจจัยร่วมกันหรือ JFA ในการรู้จำภาษา และการระบุเสียงผู้พูด (Speaker recognition) ซึ่งสร้างแบบจำลองของภาษา และความแปรปรวนของช่องสัญญาณ (Channel variability) แยกออกจากกัน แต่แนวคิดของไอเวกเตอร์ใช้ซัพสเปซความแปรปรวนแบบทั้งหมด (Total variability subspace) ซึ่งตรงกันข้ามกับ JFA โดยจำลองความแปรปรวนทั้งหมดให้อยู่ในซัพสเปซที่มีจำนวนมิติน้อย (Low dimension subspace) กำหนดให้ซัพสเปซความแปรปรวนแบบทั้งหมดแทนการรวมความแปรปรวนภาษา และความแปรปรวนช่องสัญญาณ แนวคิดของซัพสเปซความแปรปรวนแบบทั้งหมดเป็นการปรับค่าพารามิเตอร์จากแบบจำลองภาวะพื้นหลังแบบสากลหรือ ยูบีเอ็ม (Universal Background Model: UBM) เป็นไอเกนซ็วอยซ์ (Eigenvoice) โดยที่แบบจำลองยูบีเอ็มฝึกฝนจากตัวอย่างของทุกภาษาที่ใช้ในการทดลองนี้ ซัพสเปซความแปรปรวนแบบทั้งหมดนี้เป็นซูปเปอร์เวกเตอร์ที่ขึ้นกับภาษา และขึ้นกับช่องสัญญาณแทนด้วย M ซึ่งได้จากการต่อกันของเวกเตอร์ค่าเฉลี่ยของแบบจำลองเกาส์เซียนแบบผสม (GMM) ซูปเปอร์เวกเตอร์ M สามารถคำนวณจากสมการที่ (6.1)

$$M = m + Tw \tag{6.1}$$

เมื่อ m เป็นซูปเปอร์เวกเตอร์ UBM (นั่นคือ ส่วนประกอบที่ไม่ขึ้นกับภาษา และไม่ขึ้นกับช่องสัญญาณ) T คือ เมทริกซ์สี่เหลี่ยมผืนผ้า (Rectangular matrix) ซึ่งเรียกว่า ความแปรปรวนแบบทั้งหมด และ w เป็นไอเวกเตอร์ ซึ่งได้จากเวกเตอร์แบบสุ่มที่มีการกระจายแบบปกติ $N(0,1)$ โดยไอเวกเตอร์ w ได้มาจากสัญญาณเสียงพูด

สำหรับขั้นตอนการรู้จำภาษาแบบใช้ไอเวกเตอร์ หรือการจำแนกระหว่างไอเวกเตอร์ พิจารณาจากค่าล็อกของโอกาสการเกิดที่ใช้บอกคะแนนความเหมือนระหว่างไอเวกเตอร์ของเสียงทดสอบ และไอเวกเตอร์ของกลุ่มที่ฝึกฝน วิธีการคำนวณคะแนนความเหมือนกันระหว่างไอเวกเตอร์ w_1 และ w_2 มีหลายวิธีโดยการวิเคราะห์จำแนกเชิงเส้นด้วยความน่าจะเป็น (Probabilistic Linear Discriminant Analysis: pLDA) [118] เป็นวิธีการให้คะแนนแบบจำลองไอเวกเตอร์ที่ได้รับการยอมรับวิธีหนึ่งในการระบุเสียงผู้พูด จากค่าล็อกของโอกาสการเกิดจากวิธีการวิเคราะห์จำแนกเชิงเส้นด้วยความน่าจะเป็นหรือ pLDA คำนวณระยะห่างระหว่างไอเวกเตอร์สองอัน ซึ่งสามารถคำนวณได้ดังสมการ (6.2)

$$\begin{aligned} score(w_1, w_2) = & \log N \left(\begin{bmatrix} w_1 \\ w_2 \end{bmatrix}; \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{total} & B \\ B & \Sigma_{total} \end{bmatrix} \right) \\ & - \log N \left(\begin{bmatrix} w_1 \\ w_2 \end{bmatrix}; \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{tot} & 0 \\ 0 & \Sigma_{total} \end{bmatrix} \right) \end{aligned} \quad (6.2)$$

เมื่อ w_1 และ w_2 เป็นไอเวกเตอร์ 2 เวกเตอร์ $N(\cdot)$ เป็นฟังก์ชันความหนาแน่นของเกาส์เซียนที่มีการแจกแจงแบบปกติ และ Σ_{total} เป็นเมทริกซ์ความแปรปรวนร่วม (Covariance matrix) แบบทั้งหมดจากไอเวกเตอร์ที่มาจากชุดฝึกฝน และ B แทนเมทริกซ์ความแปรปรวนร่วมระหว่างกลุ่มจากไอเวกเตอร์ที่มาจากชุดฝึกฝน ไอเวกเตอร์ภายในแต่ละกลุ่ม ซึ่งหมายถึงแต่ละภาษา ถูกหาค่าเฉลี่ยและแสดงด้วยไอเวกเตอร์หนึ่งค่า ขั้นตอนการรู้จำภาษาของระบบอ้างอิงตามวิธีการวิเคราะห์จำแนกเชิงเส้นด้วยความน่าจะเป็นนี้ใช้สำหรับจำแนกไอเวกเตอร์ในวิทยานิพนธ์นี้

การสกัดค่าลักษณะสำคัญที่ใช้ในระบบการรู้จำภาษาพูดแบบใช้ไอเวกเตอร์ใช้ค่าพารามิเตอร์เหมือนในวรรณกรรม [119] ซึ่งเวกเตอร์ลักษณะเฉพาะมีมิติ 56 ค่า ได้มาจากการประกอบกันของสัมประสิทธิ์ MFCC 7 ค่า ร่วมกับค่าเอสดีซีที่คำนวณจากค่าสัมประสิทธิ์ MFCC โดยใช้ค่าพารามิเตอร์ 7-1-3-7 ในการคำนวณค่าเอสดีซี เวกเตอร์ลักษณะเฉพาะถูกนำมาใช้เป็นเวกเตอร์อินพุตของการสร้างค่าไอเวกเตอร์ โดยใช้ ALIZE [120] ซึ่งแต่ละแบบจำลองเกาส์เซียนที่มีหลายมิติ มีจำนวนองค์ประกอบเกาส์เซียนผสม 2048 ส่วนประกอบ

6.3 การศึกษาความแปรปรวนของค่าพีเอฟในหน่วยเสียง

6.3.1 การศึกษาค่าพีเอฟด้วยการวิเคราะห์ความแปรปรวนพหุคูณ

การวิเคราะห์ความแปรปรวนพหุคูณใช้พิจารณาความแปรปรวนค่าลักษณะสำคัญที่มีข้อมูลหลายมิติ เพื่อเปรียบเทียบความแตกต่างระหว่างค่าเฉลี่ยของค่าลักษณะสำคัญระหว่าง 8 ภาษา โดย

วัดที่ระดับความเชื่อมั่น 0.05 ซึ่งค่า P-value ที่มากกว่าระดับความเชื่อมั่น แสดงว่าไม่มีความแตกต่างระหว่างค่าเฉลี่ยในค่าลักษณะสำคัญระหว่างสองกลุ่มภาษาอย่างมีนัยสำคัญ

ตารางที่ 6-1 แสดงผลการศึกษาค่าลักษณะสำคัญด้วยการวิเคราะห์ความแปรปรวนพหุคูณ ซึ่งเรียงลำดับค่า F-ratio เรียงลำดับจากสูงไปหาต่ำพบว่า ผลการวิเคราะห์ความแปรปรวนพหุคูณของค่าพีเอฟ และค่าสัมประสิทธิ์ MFCC มีค่า P-value เป็น 0 ซึ่งแสดงว่ามีอย่างน้อย 1 คู่ภาษาที่มีความแตกต่างระหว่างสองภาษา ซึ่งแตกต่างกับการวิเคราะห์ในค่าสัมประสิทธิ์ PLP ที่มีค่า P-value เท่ากับ 1 และเมื่อเปรียบเทียบค่า F-ratio ของค่าพีเอฟมีค่า 1425.43 ซึ่งมากกว่าค่าสัมประสิทธิ์ MFCC ที่มีค่า 22.63 หมายความว่าในค่าพีเอฟมีความแตกต่างระหว่างค่าเฉลี่ยสองภาษามากกว่าค่าสัมประสิทธิ์ MFCC

ตารางที่ 6-1 ผลการศึกษาค่าลักษณะสำคัญโดยใช้การวิเคราะห์ความแปรปรวนพหุคูณ

ค่าลักษณะสำคัญ	F-Ratio	P-Value
PF	1425.43	0
MFCC	22.63	0
PLP	0.11	1

6.3.2 การศึกษาค่าพีเอฟด้วยการจัดกลุ่มภาษา

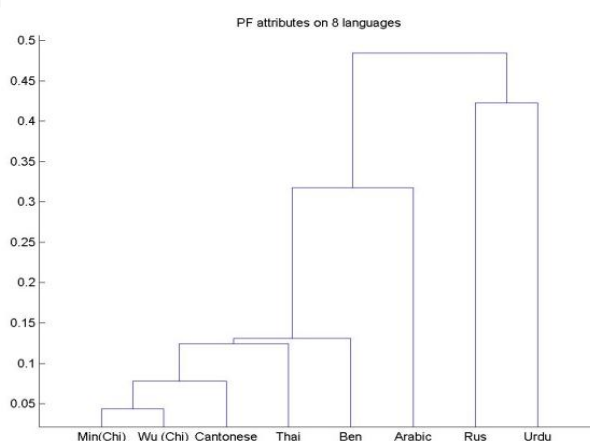
หัวข้อนี้พิจารณาความแตกต่างในค่าลักษณะสำคัญระหว่างภาษา โดยแสดงความสัมพันธ์ระหว่างกลุ่มด้วยเส้นเชื่อม กลุ่มที่อยู่ภายใต้เส้นเชื่อมมีระดับเดียวกันหรือใกล้เคียงกัน มีความสัมพันธ์ระหว่างกลุ่มมากกว่ากลุ่มที่อยู่ในเส้นเชื่อมระดับต่างกัน ผลลัพธ์ของหัวข้อนี้คือ เดนโตแกรมที่แสดงความสัมพันธ์ด้วยเส้นเชื่อมต่อของการรวมภาษา หรือกลุ่มการรวมของภาษา

โดยหัวข้อนี้วัดความไม่เหมือนระหว่างภาษาโดยใช้ฟังก์ชันของ MATLAB ชื่อ Manovacluster ซึ่งใช้ค่าเฉลี่ยในค่าลักษณะสำคัญของแต่ละภาษาที่ซึ่งได้จากการวิเคราะห์ความแปรปรวนพหุคูณในหัวข้อ 6.3.1 มาจัดกลุ่ม โดยอัลกอริทึมการจัดกลุ่ม (Cluster) คำนวณเมตริกซ์ของความเหมือนระหว่างแต่ละคู่ภาษา แล้วจัดรวมกลุ่มเป็นแบบลำดับขั้น (Hierarchical cluster) ให้ภาษาที่มีค่าเฉลี่ยในค่าลักษณะสำคัญใกล้เคียงกันมีเส้นเชื่อมต่อกัน ระยะห่างสั้นที่สุดระหว่างคู่ภาษา คือภาษาที่มีความเหมือนกันสูงที่สุด หรือการรวมกลุ่มกันสองภาษาที่มีความแตกต่างกันน้อยที่สุด ขั้นตอนการจัดรวมกลุ่ม จะรวมกลุ่มที่มีความแตกต่างกันน้อยที่สุดก่อน แล้วรวมกลุ่มที่มีความแตกต่างกันน้อยในลำดับถัดไป ไปเรื่อยๆ จนครบทุกกลุ่ม ความสูงของเส้นเชื่อมระหว่างกลุ่มแสดงระยะห่างระหว่างสองภาษา หรือระหว่างสองกลุ่มของภาษาที่เชื่อมต่อกัน ดังนั้นระยะห่างของกลุ่มที่มีความสูงมากแสดงความไม่เหมือนกันระหว่างภาษามากกว่า ภาษาที่รวมกันภายในกลุ่มเดียวกัน

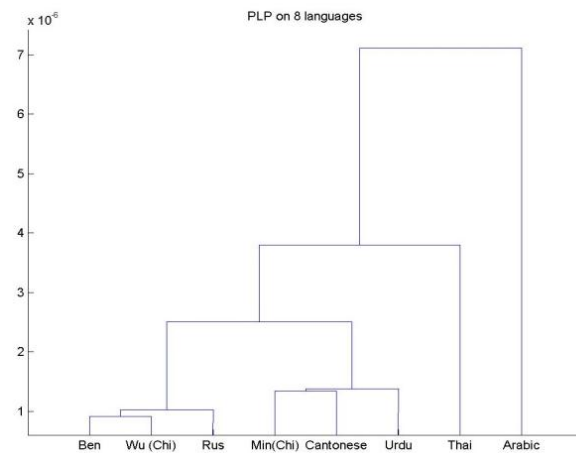
ภาพที่ 6-2 ภาพที่ 6-3 และภาพที่ 6-4 แสดงเดนโตแกรมของการรวมกลุ่มภาษาแบบลำดับขั้น โดยพิจารณาค่าเฉลี่ยในค่าลักษณะสำคัญค่าพีเอฟ ค่าสัมประสิทธิ์ PLP และค่าสัมประสิทธิ์ MFCC ระหว่าง 8 ภาษา ตามลำดับ ภาพที่ 6-2 แสดงความแตกต่างค่าเฉลี่ยของค่าพีเอฟในกลุ่มภาษา โดย

การรวมกันของค่าพีเอฟเริ่มจากการรวมกลุ่มภายในกลุ่มภาษาจีน ซึ่งมีเส้นเชื่อมระหว่างภาษาจีนมิน และภาษาจีนวู เพื่อรวมกลุ่มภาษาที่ไม่เหมือนกันน้อยที่สุด และรวมภาษาจีนกวางตุ้งในระดับชั้นของการรวมกลุ่มที่สูงขึ้นไป และทำการรวมกลุ่มภาษาที่ต่างกันในระดับชั้นสูงขึ้นถัดไป จากผลลัพธ์การรวมกลุ่มภาษาจีนที่อยู่ในระดับชั้นเดียวกันและใกล้กันทั้งสามภาษานี้ แสดงถึงค่าเฉลี่ยของค่าพีเอฟไม่มีความแตกต่างกันระหว่างภาษา เมื่อเปรียบเทียบกับภาษาอื่น ซึ่งสอดคล้องกับทฤษฎีของตระกูลภาษาทั้งสามภาษาอยู่ในกลุ่มตระกูลภาษาเดียวกันคือ ตระกูลภาษาจีน-ทิเบต นอกจากนี้ต้นไม้หรือเส้นเชื่อมของการรวมของกลุ่มภาษาจีน ถูกรวมเข้ากับภาษาไทย ภาษาเบงกาลี ภาษาอารบิก และการรวมกลุ่มของภาษารัสเซียและภาษาเออร์ดู ตามลำดับ ซึ่งลำดับของการรวมภาษาแสดงความไม่เหมือนกัน ที่แตกต่างกันขึ้นระหว่างภาษา ตามลำดับ นอกจากนี้การรวมกลุ่มของภาษารัสเซีย และภาษาเออร์ดู แสดงว่ามีความเหมือนกันระหว่างภาษารัสเซีย และภาษาเออร์ดูมากกว่าภาษาอื่นๆ นั่นคือมีความแตกต่างกับกลุ่มที่รวมภาษามาก่อนหน้า ซึ่งตรงกับทฤษฎีของตระกูลภาษาที่ภาษารัสเซีย และภาษาเออร์ดูเป็นภาษาที่อยู่ในตระกูลภาษาเดียวกันคือ ตระกูลภาษาอินโด-ยุโรปเปียน อย่างไรก็ตาม ผลลัพธ์ที่ได้จากการวิเคราะห์ด้วยค่าเฉลี่ยของค่าพีเอฟ แสดงความสัมพันธ์ระหว่างบางคู่ภาษา ไม่ตรงกับทฤษฎีของตระกูลภาษา เช่น ภาษาเบงกาลี ไม่รวมกลุ่มกับภาษารัสเซียและภาษาเออร์ดู เป็นต้น ซึ่งอาจจะเกิดจากขั้นตอนการสกัดค่าลักษณะสำคัญ และจำนวนข้อมูลที่ศึกษา

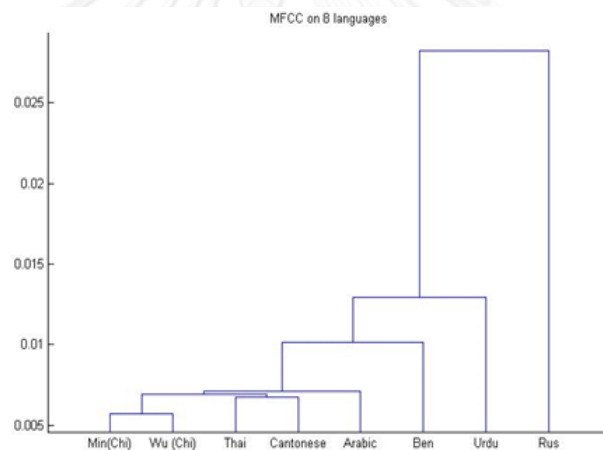
เมื่อเปรียบเทียบภาพที่ 6-2 ซึ่งพิจารณาใช้ค่าเฉลี่ยของค่าพีเอฟ พบว่าความสัมพันธ์กันระหว่างภาษา และระยะของการรวมกลุ่มภาษาแตกต่างกันกับภาพที่ 6-3 และภาพที่ 6-4 ซึ่งวิเคราะห์โดยใช้ค่าเฉลี่ยของค่าสัมประสิทธิ์ MFCC และค่าสัมประสิทธิ์ PLP เช่น ค่าเฉลี่ยของค่าสัมประสิทธิ์ PLP แสดงความไม่เหมือนกันของการรวมกลุ่มภาษาเบงกาลี และภาษาจีนวู และรวมกลุ่มกับภาษารัสเซีย ในระดับชั้นถัดไป ซึ่งเมื่อพิจารณาความสัมพันธ์ระหว่างภาษาระหว่างภาษาจีนวู และภาษาเบงกาลี ไม่สอดคล้องทฤษฎีตระกูลภาษา ซึ่งความสัมพันธ์ระหว่างภาษาในภาพที่ 6-3 และภาพที่ 6-4 ที่พิจารณาเมื่อใช้ค่าเฉลี่ยของค่าสัมประสิทธิ์ PLP และค่าสัมประสิทธิ์ MFCC พบว่าเส้นเชื่อมระหว่างการรวมกลุ่มภาษาที่แสดงการรวมภาษาในระดับชั้นเดียวกัน ไม่สอดคล้องตามทฤษฎีของตระกูลภาษา ที่ความสัมพันธ์ของภาษาภายในตระกูลภาษาเดียวกันควรมีความเหมือนกันมากกว่าภาษาที่อยู่ในตระกูลภาษาอื่น นอกจากนี้เมื่อวิเคราะห์ความสูงของเส้นเชื่อมระหว่างภาษาด้วยค่าพีเอฟมีระยะห่างมากกว่าค่าสัมประสิทธิ์ PLP และค่าสัมประสิทธิ์ MFCC



ภาพที่ 6-2 การรวมกลุ่มเป็นแบบลำดับขั้น จากค่าเฉลี่ยของค่าพีเอฟ



ภาพที่ 6-3 การรวมกลุ่มเป็นแบบลำดับขั้น จากค่าเฉลี่ยของค่าสัมประสิทธิ์ PLP



ภาพที่ 6-4 การรวมกลุ่มเป็นแบบลำดับขั้น จากค่าเฉลี่ยของค่าสัมประสิทธิ์ MFCC

6.3.3 การศึกษาค่าพีเอฟด้วยการวิเคราะห์ความแปรปรวน

หัวข้อนี้เปรียบเทียบความแตกต่างระหว่างค่าเฉลี่ยในค่าพีเอฟแต่ละชนิดระหว่าง 8 ภาษา ด้วยการวิเคราะห์ความแปรปรวน (ANOVA) ของค่าพีเอฟ จากค่า F-ratio และค่า P-value โดยพิจารณาค่าพีเอฟทีละชนิด โดยวัดความแตกต่างระหว่างกลุ่มที่ระดับความเชื่อมั่น 0.01 ซึ่งค่า P-value ที่มากกว่าระดับความเชื่อมั่น แสดงว่า ไม่มีความแตกต่างระหว่างค่าเฉลี่ยในค่าลักษณะสำคัญระหว่างสองกลุ่มภาษาอย่างมีนัยสำคัญ นอกจากนี้ทำการทดสอบการวิเคราะห์ความแปรปรวนค่าพีเอฟทีละคู่ภาษา เพื่อพิจารณาความแตกต่างของค่าเฉลี่ยในค่าพีเอฟระหว่างคู่ภาษา เช่น เปรียบเทียบภาษา A กับภาษา B เปรียบเทียบภาษา A กับภาษา C และจับคู่ไปเรื่อยๆจนครบทุกภาษา เป็นต้น นอกจากนี้นำแผนภาพกล่องมาใช้อธิบายการกระจายตัวของค่าพีเอฟแต่ละชนิด ซึ่งแสดงในภาคผนวก และนำแผนภาพเสียงสระ (Vowel chart) มาประกอบการศึกษาความสัมพันธ์ระหว่างภาษาเพื่ออธิบายพฤติกรรมของค่าพีเอฟบางชนิด เมื่ออธิบายด้วยระบบการออกเสียงกลุ่มสระในแต่ละภาษา

ผลการวิเคราะห์ความแปรปรวนค่าพีเอฟแต่ละชนิดเมื่อพิจารณาที่ละคู่ภาษาพบว่า จากกรณีคู่ภาษาทั้งหมด 364 คู่ภาษา ที่พิจารณาค่าเฉลี่ยในค่าพีเอฟแต่ละชนิด มี 296 กรณีหรือ 81% ที่มีความแตกต่างกันอย่างมีนัยสำคัญในค่าเฉลี่ยของค่าพีเอฟระหว่างสองภาษา ซึ่งผลการวิเคราะห์ค่าพีเอฟแต่ละชนิด แสดงดังนี้

1) ค่าพีเอฟ Vocalic

ผลการวิเคราะห์ความแปรปรวนค่าพีเอฟใน 8 ภาษา พบว่าค่า P-value มีค่าเท่ากับ 0 ซึ่งหมายความว่า เมื่อเปรียบเทียบค่าเฉลี่ยในค่าพีเอฟ Vocalic มีอย่างน้อย 1 คู่ภาษาที่มีความแตกต่างค่าเฉลี่ยในค่าพีเอฟ Vocalic ระหว่างคู่ภาษา และเมื่อวิเคราะห์ความแตกต่างของค่าเฉลี่ยค่าพีเอฟ Vocalic ในแต่ละคู่ภาษาพบว่ามี 22 คู่ภาษาจากทั้งหมด 28 คู่ภาษาหรือ 78.57% ของคู่ภาษาที่มีค่า P-value น้อยกว่า 0.01 ซึ่งมีความแตกต่างของค่าเฉลี่ยค่าพีเอฟ Vocalic ระหว่างภาษา

เมื่อวิเคราะห์แผนภาพกล่องพบว่า ภาษาที่มีความแตกต่างของค่าพีเอฟ Vocalic ที่แตกต่างจากภาษาอื่นได้แก่ ภาษาอารบิก ภาษาไทย และภาษาจีนวู ขณะที่จากการวิเคราะห์ใน 6 คู่ภาษาที่มีค่าเฉลี่ยค่าพีเอฟ Vocalic ไม่แตกต่างกันอย่างมีนัยสำคัญ พบว่า ภาษาจีนมินและภาษารัสเซียมีค่าเฉลี่ยเท่ากัน แผนภาพกล่องของภาษาจีนมิน ซ้อนทับกับกล่องของภาษาเบงกาลี เออร์ดู และภาษาจีนกวางตุ้ง การกระจายของค่าพีเอฟ Vocalic ของภาษาเออร์ดู ไม่สามารถแยกแยะจากภาษาเบงกาลี และภาษารัสเซีย เป็นต้น

2) ค่าพีเอฟ Consonantal

ผลการวิเคราะห์ความแปรปรวนค่าพีเอฟ Consonantal ใน 8 ภาษา พบว่าเมื่อค่า P-value เท่ากับ 0 จะมีความแตกต่างกันของค่าเฉลี่ยในค่าพีเอฟ Consonantal อย่างน้อย 1 คู่ภาษา และเมื่อวิเคราะห์ค่าความแปรปรวนค่าพีเอฟ Consonantal ที่ละคู่ภาษา มี 78% จาก 28 คู่ภาษาที่มีค่า P-value น้อยกว่า 0.01 โดยมี 6 คู่ภาษาที่มีค่า P-value มากกว่า 0.01 ได้แก่ ระหว่างคู่ภาษาอารบิกกับภาษาจีนมิน ระหว่างคู่ภาษาเบงกาลีกับภาษาจีนมิน ระหว่างคู่ภาษาไทยกับภาษาจีนมิน ระหว่างคู่ภาษาจีนกวางตุ้งกับภาษาจีนมิน ระหว่างคู่ภาษารัสเซียกับภาษาอารบิก และระหว่างคู่ภาษารัสเซียกับภาษาจีนวู หมายความว่าค่าเฉลี่ยของค่าพีเอฟ Consonantal ของภาษาจีนมินไม่แตกต่างอย่างมีนัยสำคัญกับภาษาอารบิก เบงกาลี ไทย และภาษาจีนกวางตุ้ง และเช่นเดียวกับค่าเฉลี่ยของค่าพีเอฟ Consonantal ของภาษารัสเซียไม่แตกต่างอย่างมีนัยสำคัญกับภาษาจีนวู เป็นต้น ขณะที่การวิเคราะห์ความแปรปรวนแบบที่ละคู่ภาษา ระหว่างคู่ภาษาเออร์ดูกับภาษาอื่น มีค่า P-value ต่ำกว่า 0.01 ซึ่งหมายความว่าค่าเฉลี่ยของค่าพีเอฟ Consonantal ของภาษาเออร์ดู แตกต่างกับภาษาอื่น

3) ค่าพีเอฟ High

ผลการวิเคราะห์ความแปรปรวนค่าพีเอฟ High ใน 8 ภาษา พบว่าค่า P-value มีค่าเท่ากับ 0 นั่นคือมีความแตกต่างกันในค่าเฉลี่ยของค่าพีเอฟ High อย่างน้อย 1 คู่ภาษา จากแผนภาพกล่องของค่าพีเอฟ High (แสดงในภาคผนวก ค) พบว่า มีข้อมูลที่อยู่นอกขอบเขตของกล่องจำนวนมาก เมื่อวิเคราะห์ค่าความแปรปรวนค่าพีเอฟ High แบบที่ละคู่ภาษา พบว่า 75% ของ 28 คู่ภาษามีค่า P-value น้อยกว่า 0.01 ค่าเฉลี่ยในค่าพีเอฟ High ของภาษาเออร์ดู ภาษาจีนมิน ภาษาจีนวู แตกต่าง

จากภาษาอื่นอย่างมีนัยสำคัญทางสถิติ ซึ่งหมายความว่าค่าพีเอฟ High สามารถใช้บอกความแตกต่างระหว่างภาษาข้างต้น กับภาษาอื่นได้ ขณะที่ค่าเฉลี่ยในค่าพีเอฟ High ของภาษารัสเซีย ไม่แตกต่างกับภาษาอารบิก ภาษาเบงกาลี ภาษาไทย และภาษาจีนกวางตุ้ง ซึ่งแสดงว่าค่าพีเอฟ High ไม่มีความแตกต่างระหว่างภาษาดังกล่าวอย่างมีนัยสำคัญ นอกจากนี้จากการสังเกตแผนภาพกล่องพบว่า การกระจายของข้อมูลค่าพีเอฟ High ของภาษาเบงกาลี ซ้อนทับกับภาษาอารบิก และภาษาไทย ซึ่งหมายความว่าค่าพีเอฟ High ของภาษาเบงกาลีไม่สามารถแยกจากภาษาอารบิก และภาษาไทยได้ เช่นเดียวกับภาษาอารบิกไม่สามารถแยกออกจากภาษาไทย เป็นต้น

4) ค่าพีเอฟ Back

ผลการวิเคราะห์ความแปรปรวนค่าพีเอฟ Back ใน 8 ภาษา พบว่า ค่า P-value มีค่าเท่ากับ 0 นั่นคือ มีความแตกต่างกันในค่าเฉลี่ยของค่าพีเอฟ Back อย่างน้อย 1 คู่ภาษา และเมื่อพิจารณาทีละคู่ภาษาพบว่า มี 6 คู่ภาษาจาก 28 คู่ภาษาที่มีค่า P-value มากกว่า 0.01 ได้แก่ ระหว่างคู่ภาษาไทย กับภาษาจีนมิน ภาษาอารบิก และภาษาจีนกวางตุ้ง และระหว่างคู่ภาษาเบงกาลีกับภาษาเออร์ดู ภาษาจีนมิน และภาษารัสเซีย

เนื่องจากค่าพีเอฟ Back เป็นค่าที่อธิบายการออกเสียงตำแหน่งลิ้น ซึ่งเป็นส่วนหนึ่งของค่าลักษณะสำคัญของการออกเสียงสระ ดังนั้นผู้วิจัยจึงใช้แผนภาพสระ เพื่อวิเคราะห์ความสัมพันธ์ของสระที่ลิ้นอยู่ตำแหน่งด้านหลัง (Back vowel) ที่ปรากฏในแต่ละภาษา จากสมมติฐาน สระที่ลิ้นอยู่ตำแหน่งหลัง ซึ่งแทนด้วย สระหลัง มีความสัมพันธ์กับค่าพีเอฟ Back และสระหลังที่พบบ่อยในภาษาหนึ่งและไม่ปรากฏบ่อยในภาษาอื่น สระนั้นสามารถใช้แยกความแตกต่างระหว่างภาษาได้ ซึ่งส่งผลให้ค่าพีเอฟ Back ที่สกัดมาจากสระในภาษาหนึ่งนั้น มีค่าแตกต่างจากค่าพีเอฟที่สกัดจากอีกภาษาได้ในทางตรงกันข้ามหากสระหลังที่พบบ่อยในหลายภาษา จะทำให้ไม่สามารถนำสระนั้นมาช่วยบอกความแตกต่างระหว่างภาษาได้ ตัวอย่างเช่น จากการวิเคราะห์แผนภาพสระดังภาพที่ 6-5 พบว่ามีสระหลัง 5 ตัว ที่ปรากฏในภาษาไทย และปรากฏร่วมกันในภาษาจีนมิน เช่น ซึ่งจากสระที่เหมือนกันปรากฏในภาษาไทย และภาษาจีนมินนั้น อาจจะไม่สามารถแยกกระหว่างสองภาษาได้ดีนักด้วยค่าพีเอฟ Back นั่นคือ สระ [u] ซึ่งเป็นสระหลังแบบสระปิด (close vowel) ตำแหน่งลิ้นยกสูง ปากไม่ห่อ, สระ [u] ซึ่งเป็นสระหลังแบบสระปิด ตำแหน่งลิ้นยกสูง ปากห่อ, สระ [ɤ] ซึ่งเป็นสระหลังแบบสระปิด ตำแหน่งลิ้นยกระดับกลาง ปากไม่ห่อ, สระ [o] ซึ่งเป็นสระหลังแบบสระปิด ตำแหน่งลิ้นยกระดับกลาง ปากห่อ และสระ [ɔ] ซึ่งเป็นสระหลังแบบสระเปิด (open vowel) ตำแหน่งลิ้นยกระดับกลาง ปากห่อ นอกจากนี้จากแผนภาพสระพบว่า สระหลังที่ปรากฏในภาษาเบงกาลี จะปรากฏในภาษาเออร์ดูเช่นกัน

5) ค่าพีเอฟ Low

ผลการวิเคราะห์ความแปรปรวนค่าพีเอฟ Low ใน 8 ภาษา มีค่า P-value เท่ากับ 0 ซึ่งค่าเฉลี่ยในค่าพีเอฟ Low อย่างน้อย 1 คู่ภาษาที่มีความแตกต่างกับภาษาอื่น จากการพิจารณาทีละคู่ภาษาพบว่า 75% ของคู่ภาษาจาก 28 คู่ภาษาที่มีค่า P-value น้อยกว่า 0.01 มีค่าเฉลี่ยในค่าพีเอฟแตกต่างกันอย่างมีนัยสำคัญทางสถิติเมื่อเปรียบเทียบระหว่างสองภาษา และมี 7 คู่ภาษาจาก 28 คู่ภาษาที่มีค่า P-value มากกว่า 0.01 ได้แก่ ระหว่างคู่เบงกาลี กับภาษาอารบิก ภาษาจีนมิน และ

ภาษาเออร์ดู และระหว่างคู่ภาษาอารบิกกับภาษาไทย ภาษาจีนมิน และภาษาเออร์ดู และระหว่างภาษารัสเซียกับภาษาจีนกวางตุ้ง

เนื่องจากค่าพีเอฟ Low เป็นค่าที่อธิบายการออกเสียงตำแหน่งลิ้น และเป็นส่วนหนึ่งของค่าการออกเสียงสระ ผู้วิจัยจึงใช้แผนภาพสระพิจารณาสระที่ปรากฏในแต่ละภาษา เมื่อพิจารณาแผนภาพสระดังภาพที่ 6-5 และแผนภาพสระในภาคผนวก ง เพื่อวิเคราะห์ความสัมพันธ์ของสระที่ลิ้นอยู่ตำแหน่งด้านล่าง (Low vowel) พบว่า สระต่ำที่ปรากฏในภาษาหนึ่งจะปรากฏในอีกภาษาหนึ่ง เช่น สระ [a] ซึ่งเป็นสระเปิด ลิ้นอยู่ตำแหน่งหน้า และตำแหน่งลิ้นต่ำ ปากไม่ห่อ สระ [æ] ซึ่งเป็นสระเปิด ลิ้นอยู่ใกล้ตำแหน่งหน้า และตำแหน่งลิ้นต่ำ ปากไม่ห่อ มีการปรากฏของสระต่ำข้างต้นในภาษาเบงกาลีและภาษาเออร์ดู

ในทางทฤษฎีการวิเคราะห์จากแผนภาพสระ สำหรับพิจารณาสระต่ำที่ปรากฏในภาษาหนึ่งและไม่ปรากฏในภาษาอื่น แต่เมื่อพิจารณาสระต่ำ [ɑ] ซึ่งเป็นสระหลังแบบสระเปิด ปากไม่ห่อ ที่ปรากฏในภาษาอารบิก แต่ไม่ปรากฏในภาษาเบงกาลี ซึ่งไม่สอดคล้องกับการวิเคราะห์ค่าพีเอฟ Low จากฐานข้อมูลเสียงที่ได้จากค่าพีเอฟที่สกัดจากเสียงในแต่ละภาษา ซึ่งสาเหตุที่ไม่สอดคล้องกันเกิดจากค่าพีเอฟที่สกัดจากเสียงในแต่ละภาษาอาจจะไม่ครอบคลุมตัวอย่างทั้งหมดของค่าการออกเสียงในแต่ละภาษาได้

6) ค่าพีเอฟ Anterior

ผลการวิเคราะห์ความแปรปรวนค่าพีเอฟ Anterior ใน 8 ภาษา พบว่า ค่า P-value มีค่าเท่ากับ 0 มีค่าเฉลี่ยในค่าพีเอฟ Anterior อย่างน้อย 1 คู่ภาษาที่มีความแตกต่างกันในค่าเฉลี่ย จากการพิจารณาทีละคู่ภาษาพบว่า 89% ของคู่ภาษาทั้งหมด 28 คู่ภาษาที่ค่าเฉลี่ยในค่าพีเอฟ Anterior แตกต่างกันอย่างมีนัยสำคัญทางสถิติ และมี 3 คู่ภาษาจาก 28 คู่ภาษาที่มีค่า P-value มากกว่า 0.01 ได้แก่ ระหว่างคู่เบงกาลี กับภาษาอารบิก ระหว่างคู่ภาษาอารบิกกับภาษาจีนมิน และระหว่างคู่ภาษาอารบิกกับภาษาไทย ที่ค่าเฉลี่ยของค่าพีเอฟ Anterior ไม่แตกต่างกันอย่างมีนัยสำคัญ

7) ค่าพีเอฟ Coronal

ผลการวิเคราะห์ความแปรปรวนค่าพีเอฟ Coronal ใน 8 ภาษา พบว่า ค่า P-value มีค่าเท่ากับ 0 และจากการพิจารณาทีละคู่ภาษาพบว่า 78% ของคู่ภาษาทั้งหมด 28 คู่ภาษาที่ค่าเฉลี่ยในค่าพีเอฟ Coronal แตกต่างกันอย่างมีนัยสำคัญทางสถิติเมื่อเปรียบเทียบระหว่างสองภาษา และมี 6 คู่ภาษาจาก 28 คู่ภาษาที่มีค่า P-value มากกว่า 0.01 ที่ค่าการทดสอบความแตกต่างของค่าเฉลี่ยของค่าพีเอฟ Coronal ไม่แตกต่างกันอย่างมีนัยสำคัญ ได้แก่ ระหว่างคู่ภาษาอารบิกกับภาษารัสเซีย ระหว่างคู่ภาษาอารบิกกับภาษาไทย ระหว่างคู่ภาษาเบงกาลี กับภาษาจีนมิน ระหว่างคู่ภาษาเบงกาลีกับภาษาจีนกวางตุ้ง ระหว่างคู่ภาษารัสเซียกับภาษาไทย และระหว่างคู่ภาษาเออร์ดูกับภาษาจีนวู เป็นต้น

8) ค่าพีเอฟ Round

ผลการวิเคราะห์ความแปรปรวนค่าพีเอฟ Round ใน 8 ภาษา พบว่า ค่า P-value มีค่าเท่ากับ 0 และจากการพิจารณาทีละคู่ภาษาพบว่า 89% ของคู่ภาษาทั้งหมด 28 คู่ภาษาที่ค่าเฉลี่ยใน

ค่าพีเอฟ Round แตกต่างกันอย่างมีนัยสำคัญทางสถิติ และมีเพียง 4 คู่ภาษา ที่มีค่าการทดสอบความแตกต่างของค่าเฉลี่ยของค่าพีเอฟ Round ไม่แตกต่างกันอย่างมีนัยสำคัญ ได้แก่ ระหว่างคู่ภาษาอารบิกกับภาษารัสเซีย ระหว่างคู่ภาษาจีนมินกับภาษาจีนกวางตุ้ง ระหว่างคู่ภาษาจีนมินกับภาษาจีนวู ระหว่างคู่จีนมินกับภาษาไทย ซึ่งทั้งสี่คู่ภาษาไม่มีความแตกต่างกันอย่างมีนัยสำคัญ สอดคล้องตามทฤษฎีตระกูลภาษา ที่บอกว่าภาษาจีนมิน จีนวู และจีนกวางตุ้งอยู่ในตระกูลภาษาเดียวกัน และมีความใกล้เคียงตระกูลภาษาของภาษาไทย

จากภาคผนวก ค แสดงแผนภาพกล่องของค่าพีเอฟ Round โดยพบว่าค่าของภาษาจีนมิน ใกล้เคียงกับภาษาจีนกวางตุ้ง และภาษาจีนวู การกระจายของค่าพีเอฟ Round และค่าเฉลี่ยทั้งสามภาษามีตำแหน่งใกล้เคียงกัน อีกทั้งความแตกต่างของค่าเฉลี่ยค่าพีเอฟ Round ของภาษาจีนมิน และภาษาไทยไม่แตกต่างกันอย่างมีนัยสำคัญ ทำให้การทดสอบความแตกต่างของค่าเฉลี่ยค่าพีเอฟ Round ในสองภาษานี้ไม่แตกต่างกัน และจากการพิจารณาค่าเฉลี่ยของค่าพีเอฟ Round ของภาษาจีนมิน ใกล้เคียงกับภาษาจีนกวางตุ้ง และภาษาจีนวู พบว่าผลการวิเคราะห์ค่าเฉลี่ยในค่าพีเอฟ Round ในภาษาจีนที่ไม่แตกต่างกันอย่างมีนัยสำคัญ ขณะที่ค่าเฉลี่ยในค่าพีเอฟ Round ระหว่างภาษาจีนกับภาษาอื่น มีความแตกต่างกันอย่างมีนัยสำคัญ มีสาเหตุมาจากมีสระบางตัวที่ปรากฏร่วมกันในภาษาจีน แต่ไม่ปรากฏในภาษาอื่น ส่งผลกระทบต่อการสกัดค่าพีเอฟ ที่นำมาวิเคราะห์ในแต่ละภาษา

เนื่องจากค่าพีเอฟ Round เป็นค่าที่อธิบายการออกเสียงสระ จากการพิจารณาแผนภาพสระที่ปรากฏในแต่ละภาษา พบว่าสระที่ปรากฏร่วมกันในภาษาจีน เช่น สระ [y] ซึ่งเป็นสระปิด ลื่นอยู่ตำแหน่งหน้า ปากห่อ, สระ [œ] ซึ่งเป็นสระเปิด ลื่นอยู่ใกล้ตำแหน่งหน้า และ ปากห่อ เป็นต้น จากตัวอย่างของสระ ที่มีค่าการออกเสียงปากห่อที่ปรากฏร่วมกันในกลุ่มภาษาจีน แต่ไม่ปรากฏในภาษาอื่น แสดงให้เห็นว่า ค่าพีเอฟ Round ที่อธิบายการออกเสียงปากห่อ สามารถจำแนกความแตกต่างระหว่างภาษาจีน กับภาษาอื่นได้ และสามารถบอกความแตกต่างระหว่างภาษาจีนได้น้อย

9) ค่าพีเอฟ Tense

ผลการวิเคราะห์ความแปรปรวนค่าพีเอฟ Tense ใน 8 ภาษา พบว่า ค่า P-value มีค่าเท่ากับ 0 และจากการพิจารณาทีละคู่ภาษาพบว่าร้อยละ 89 ของคู่ภาษา 28 คู่ภาษาที่ค่าเฉลี่ยในค่าพีเอฟ Tense แตกต่างกันอย่างมีนัยสำคัญทางสถิติเมื่อเปรียบเทียบระหว่างสองภาษา และมีเพียง 3 คู่ภาษา ที่มีค่าการทดสอบความแตกต่างของค่าเฉลี่ยของค่าพีเอฟ Tense ไม่แตกต่างกันอย่างมีนัยสำคัญ ได้แก่ ระหว่างคู่ภาษาอารบิกกับภาษาเออร์ดู ระหว่างคู่ภาษาจีนมินกับภาษาเออร์ดู และระหว่างคู่ภาษาจีนมินกับภาษาจีนกวางตุ้ง ถึงแม้ค่าพีเอฟ Tense จะมีความสำคัญในการอธิบายการออกเสียงสระ แต่การวิเคราะห์ด้วยแผนภาพสระไม่แสดงลักษณะของการออกเสียงเกร็ง (tense) ของเสียงสระจึงไม่สามารถใช้เพื่อพิจารณาความแตกต่างของค่าพีเอฟ Tense ได้

10) ค่าพีเอฟ Voice

ผลการวิเคราะห์ความแปรปรวนค่าพีเอฟ Voice ใน 8 ภาษา พบว่าค่า P-value มีค่าเท่ากับ 0 และจากการพิจารณาทีละคู่ภาษาพบว่าร้อยละ 78 ของคู่ภาษาทั้งหมด 28 คู่ภาษาที่ค่าเฉลี่ยในค่าพี

เอฟ Voice แตกต่างกันอย่างมีนัยสำคัญทางสถิติเมื่อเปรียบเทียบระหว่างสองภาษา และมี 6 คู่ภาษา ที่มีค่าการทดสอบความแตกต่างของค่าเฉลี่ยของค่าพีเอฟ Voice ไม่แตกต่างกันอย่างมีนัยสำคัญ ได้แก่ ระหว่างคู่ภาษาอารบิกกับภาษาเบงกาลี ภาษาจีนมิน ภาษารัสเซีย ภาษาไทย ภาษาเออร์ดุ และ ภาษาจีนกวางตุ้ง ซึ่งเมื่อพิจารณาจากแผนภาพกล่องของค่าพีเอฟ Voice ของภาษาอารบิกพบว่า ค่าเฉลี่ยของค่าพีเอฟ Voice ในภาษาอารบิกไม่แตกต่างจาก 6 ภาษาข้างต้น

11) ค่าพีเอฟ Continuant

ผลการวิเคราะห์ความแปรปรวนค่าพีเอฟ Continuant ใน 8 ภาษา พบว่า ค่า P-value มีค่าเท่ากับ 0 และจากการพิจารณาทีละคู่ภาษาพบว่า ร้อยละ 85 ของคู่ภาษาทั้งหมด 28 คู่ภาษาที่ ค่าเฉลี่ยในค่าพีเอฟ Continuant ที่แตกต่างกันอย่างมีนัยสำคัญทางสถิติเมื่อเปรียบเทียบระหว่างสอง ภาษา และมี 4 คู่ภาษา ที่มีค่าการทดสอบความแตกต่างของค่าเฉลี่ยของค่าพีเอฟ Continuant ไม่ แตกต่างกันอย่างมีนัยสำคัญ ได้แก่ ระหว่างคู่ภาษาอารบิกกับภาษาเบงกาลี ภาษารัสเซีย ภาษาไทย และภาษาจีนกวางตุ้ง ซึ่งจากผลการวิเคราะห์ความแปรปรวนค่าพีเอฟ Continuant แบบทีละคู่ภาษา พบว่าการทดสอบความแตกต่างของค่าเฉลี่ยของค่าพีเอฟระหว่างภาษา เหมือนกับค่าพีเอฟ Voice แสดงว่าค่าเฉลี่ยของค่าพีเอฟ Continuant และ Voice ในภาษาอารบิกนี้ ไม่แตกต่างกับ 4 ภาษา ข้างต้น ได้แก่ ภาษาเบงกาลี ภาษารัสเซีย ภาษาไทย และภาษาจีนกวางตุ้ง แต่ผลลัพธ์การทดสอบที่ ทำให้ค่าพีเอฟ Continuant และค่าพีเอฟ Voice ไม่เหมือนกัน คือ ค่าเฉลี่ยของค่าพีเอฟ Continuant ระหว่างภาษาอารบิก กับภาษาเออร์ดุ และภาษาจีนมินแตกต่างกันอย่างมีนัยสำคัญ

12) ค่าพีเอฟ Nasal

ผลการวิเคราะห์ความแปรปรวนค่าพีเอฟ Nasal ใน 8 ภาษา พบว่าค่า P-value มีค่าเท่ากับ 0 และจากการพิจารณาทีละคู่ภาษาพบว่า ร้อยละ 85 ของคู่ภาษาทั้งหมด 28 คู่ภาษาที่ค่าเฉลี่ยในค่า พีเอฟ Nasal แตกต่างกันอย่างมีนัยสำคัญทางสถิติเมื่อเปรียบเทียบระหว่างสองภาษา และมี 4 คู่ ภาษา ที่มีค่าการทดสอบความแตกต่างของค่าเฉลี่ยของค่าพีเอฟ Nasal ไม่แตกต่างกันอย่างมี นัยสำคัญ ได้แก่ ระหว่างคู่ภาษาอารบิกกับภาษารัสเซีย ภาษาไทย และภาษาเออร์ดุ และระหว่างคู่ ภาษาเบงกาลีกับภาษาจีนกวางตุ้ง เป็นต้น

ถึงแม้ว่าภาษาพูดมนุษย์ส่วนส่วนใหญ่มีการปรากฏของหน่วยเสียงนาสิก ซึ่งสามารถอธิบาย การออกเสียงด้วยค่าพีเอฟ Nasal ได้ ส่งผลให้ค่าพีเอฟ Nasal อาจจะไม่มีความสำคัญในการบอก ความแตกต่างระหว่างภาษาได้ จากการพิจารณาแผนภาพกล่องของค่าพีเอฟ Nasal ของแต่ละภาษา พบว่า ช่วงกล่องของค่าพีเอฟ Nasal ในแต่ละภาษาซ้อนทับกัน และมีจำนวนข้อมูลที่อยู่นอกขอบเขต กล่องเป็นจำนวนมาก อย่างไรก็ตาม ตามทฤษฎีของภาษาศาสตร์ เสียงสระนาสิก ซึ่งเป็นสระที่อยู่ติด กับพยัญชนะเสียงนาสิกและเกิดจากการหย่อนเพดานอ่อนในปาก แล้วปล่อยอากาศออกทางจมูก ได้รับอิทธิพลของค่าพีเอฟ Nasal เป็นเสียงที่พบได้ในหลายภาษา เช่น ภาษาเบงกาลี ภาษาเออร์ดุ ภาษาจีนมิน ภาษาจีนวู เป็นต้น เสียงสระนาสิกที่ปรากฏในบางภาษา และไม่ปรากฏในภาษาอื่น จะ สามารถนำค่าลักษณะสำคัญที่ได้บอกความแตกต่างระหว่างภาษาได้ อีกหนึ่งตัวอย่างที่บอกความ แตกต่างระหว่างภาษาได้ คือ หน่วยเสียงซึ่งเป็นพยัญชนะนาสิกที่ปรากฏในบางภาษา เช่น หน่วยเสียง [ŋ] ซึ่งเป็นเสียงนาสิก เพดานอ่อน แทนด้วยเสียง /ng/ แทนด้วยเสียง ง ในภาษาไทย ปรากฏใน

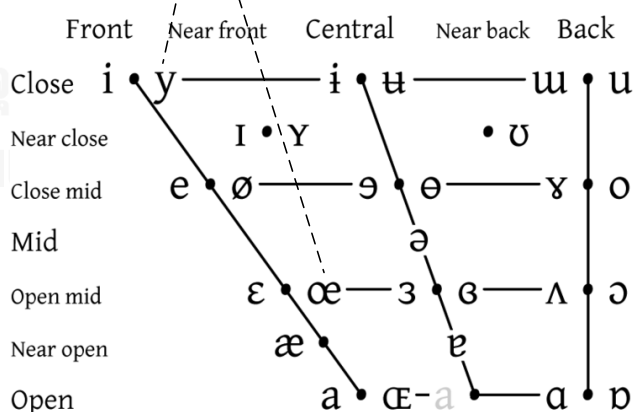
ภาษาจีนมิน ภาษาจีนกวางตุ้ง ภาษาจีนวู ภาษาเออร์ดู่ ภาษาไทย และภาษารัสเซีย ในขณะที่ไม่ปรากฏในหลายภาษา เป็นต้น จากความถี่ของการปรากฏของการออกเสียงนาสิกในแต่ละภาษาที่แตกต่างกันนี้ จึงนำไปสู่การบอกความแตกต่างระหว่างภาษาได้

13) ค่าพีเอฟ Strident

ผลการวิเคราะห์ความแปรปรวนค่าพีเอฟ Strident ใน 8 ภาษา พบว่า ค่า P-value มีค่าเท่ากับ 0 และจากการพิจารณาทีละคู่ภาษาพบว่าร้อยละ 75 ของคู่ภาษาทั้งหมด 28 คู่ภาษาที่ค่าเฉลี่ยในค่าพีเอฟ Strident แตกต่างกันอย่างมีนัยสำคัญทางสถิติเมื่อเปรียบเทียบระหว่างสองภาษา และมี 7 คู่ภาษา ที่ค่าเฉลี่ยของค่าพีเอฟ Strident ไม่แตกต่างกันอย่างมีนัยสำคัญ ได้แก่ ระหว่างคู่ภาษาอารบิกกับภาษาไทย ภาษาเบงกาลี ภาษารัสเซีย ภาษาจีนมิน และภาษาจีนกวางตุ้ง ระหว่างคู่ภาษาเออร์ดู่กับภาษาจีนวู และระหว่างคู่ภาษาจีนมินกับภาษาจีนกวางตุ้ง ซึ่งเป็นที่น่าสังเกตว่า คู่ภาษาที่มีค่าเฉลี่ยของค่าพีเอฟ Strident ไม่แตกต่างกับภาษาอื่นมาก ได้แก่ ระหว่างภาษาไทยกับภาษาอื่น ซึ่งเมื่อวิเคราะห์ความเป็นไปได้ของความไม่แตกต่างนี้ อาจเกิดจาก มีหน่วยเสียงเสียดแทรก (Fricative) 3 หน่วยเสียง ประกอบด้วย /f/, /s/ และ /h/ จากทั้งหมด 20 หน่วยเสียงเสียดแทรก ที่ปรากฏในภาษาไทย และ 2 หน่วยเสียงได้แก่ เสียง /s/ และ /f/ เป็นหน่วยเสียงที่ปรากฏทั่วไปในภาษาอื่นๆ จึงทำให้การใช้ค่าการออกเสียงที่มาจากค่าพีเอฟ Strident ที่สกัดจากข้อมูลเสียงในภาษาไทย ไม่แตกต่างกับค่าพีเอฟ Strident ที่ได้จากภาษาอื่น คุณสมบัติของค่าพีเอฟ Strident ซึ่งเป็นคำอธิบายการออกเสียงเสียดแทรกหรือเสียงรบกวน และแยกความแตกต่างของหน่วยเสียง [f, v, s, z, ʃ, ʒ] ออกจากเสียงเสียดแทรกอื่นๆได้ แต่มั่นยากต่อเมื่อใช้ค่าพีเอฟ Strident เพื่อบอกความแตกต่างระหว่างภาษาไทยกับภาษาอื่นๆ โดยใช้ค่าพีเอฟ Strident ที่ได้จากหน่วยเสียงเสียดแทรก

CH.min, CH.can, CH.wuu

VOWELS



Vowels at right & left of bullets are rounded & unrounded.

ภาพที่ 6-5 แผนภาพสระ

6.3.4 การศึกษาค่าพีเอฟด้วยการวิเคราะห์ความแปรปรวนพหุคูณสำหรับแต่ละคู่ภาษา

การทดลองนี้ต้องการศึกษาความสัมพันธ์ระหว่างค่าพีเอฟทุกชนิดในแต่ละภาษา โดยพิจารณาความแปรปรวนพหุคูณของกลุ่มค่าพีเอฟ ซึ่งประกอบด้วยค่าพีเอฟ 13 ชนิด สำหรับแต่ละคู่ภาษา โดยทดสอบบนค่าสถิติแลมด้า Wilk และสถิติ Pillai's trace (ดังแสดงในภาคผนวก ข) พบว่าค่า P-value มีค่าเท่ากับ 0 สำหรับทุกคู่ภาษา ซึ่งต่ำกว่าระดับความเชื่อมั่น ซึ่งสอดคล้องกับผลการวิเคราะห์ความแปรปรวน ตามหัวข้อที่ 6.3.3 ที่พิจารณาความแตกต่างของค่าเฉลี่ยพีเอฟแต่ละชนิดระหว่างภาษา จึงสรุปได้ว่า ค่าพีเอฟมีความแตกต่างระหว่างค่าเฉลี่ยในค่าพีเอฟระหว่างคู่ภาษา 28 ภาษา จากผลการศึกษาค่าพีเอฟด้วยการวิเคราะห์ความแปรปรวนนี้ นำไปสู่การนำค่าพีเอฟไปใช้ในการบอกความแตกต่างระหว่างภาษาต่อไป

6.4 การทดลองเพื่อประเมินประสิทธิภาพของการรู้จำภาษา

วิทยานิพนธ์นี้เสนอระบบการรู้จำภาษาพูดตามแนวทางต่าง ๆ ดังนั้นเพื่อประเมินประสิทธิภาพของแต่ละระบบการรู้จำ ผู้วิจัยจึงออกแบบการทดลองเพื่อทดสอบปัจจัยต่าง ๆ และหาค่าพารามิเตอร์ที่เหมาะสมในการสร้างต้นแบบระบบการรู้จำของแต่ละแนวทาง เช่น การหาตัวแทนของการเกิดร่วมกันของค่าพีเอฟ เพื่อใช้ในระบบการรู้จำภาษาพูดแนวทางการเรียงหน่วยเสียงที่ใช้ค่าพีเอฟ การเปรียบเทียบความสามารถของค่าลักษณะสำคัญต่าง ๆ เมื่อใช้ในการจำแนกภาษาของระบบการรู้จำภาษาพูดโดยใช้การจำแนกด้วยแบบจำลองลาเทนท์ไดนามิคคอนดิชันนอลแรนดอมฟิลด์ร่วมกับค่าพีเอฟ การวัดประสิทธิภาพของระบบตรวจหาลักษณะเฉพาะค่าพีเอฟ เพื่อนำค่าพีเอฟมาใช้เป็นอินพุทของระบบการรู้จำภาษาพูด เป็นต้น ซึ่งรายละเอียดของการทดลองแสดงดังต่อไปนี้

6.4.1 การทดลองที่ 1: ประเมินประสิทธิภาพของระบบตรวจหาลักษณะเฉพาะค่าพีเอฟ

การทดลองนี้ประเมินประสิทธิภาพของระบบตรวจหาลักษณะเฉพาะค่าพีเอฟ ก่อนที่จะนำระบบตรวจหาลักษณะเฉพาะค่าพีเอฟมาใช้สกัดค่าพีเอฟเพื่อใช้ในระบบการรู้จำภาษาพูด โดยใช้ข้อมูลเซตทดสอบของฐานข้อมูลเสียง TIMIT ทดสอบกับระบบการตรวจหาลักษณะเฉพาะค่าพีเอฟสำหรับค่าพีเอฟแต่ละชนิด โดยนำข้อมูลทดสอบมาทำการทดสอบ และวัดผลการทำนายของแต่ละเสียงพูดที่ได้จากระบบตรวจหาลักษณะเฉพาะค่าพีเอฟแต่ละชนิดมาเปรียบเทียบกับคำตอบจริงที่ได้จากฉลากค่าพีเอฟ จากขั้นตอนการเตรียมข้อมูล โดยถ้าผลลัพธ์ของการทำนายตรงกับคำตอบของค่าพีเอฟนั้นจึงถูกพิจารณาว่าเป็นคำตอบที่ระบบทำนายได้ถูกต้อง และวัดความแม่นยำของระบบตรวจหาลักษณะเฉพาะค่าพีเอฟ นอกจากนี้การวัดประสิทธิภาพถูกเปรียบเทียบกับประสิทธิภาพของระบบตรวจหาค่าพีเอฟที่ใช้นิยามตาม SPE 13 ค่าเช่นเดียวกัน ตามวรรณกรรม [16] ซึ่งมีค่าเปอร์เซ็นต์ความแม่นยำเฉลี่ย 95.8

6.4.2 การทดลองที่ 2: ประเมินประสิทธิภาพของการรู้จำภาษาตามแนวทางการเรียงหน่วยเสียง

บทที่ 4 เสนอระบบการรู้จำภาษาพูดตามแนวทางการเรียงหน่วยเสียง ได้แก่ ระบบการรู้จำภาษาพูดแบบใช้ระบบรู้จำเสียงตามด้วยแบบจำลองเวกเตอร์สเปซ และระบบการรู้จำภาษาพูดแบบใช้หน่วยเสียงแลตทิซร่วมกับแบบจำลองเวกเตอร์สเปซ การทดลองที่ 2 จึงเสนอการทดลองเพื่อประเมิน

ประสิทธิภาพของระบบการรู้จำภาษาพูดดังกล่าว การทดลองนี้ศึกษาผลกระทบของฟังก์ชันค่าถ่วงน้ำหนักแบบต่าง ๆ และผลกระทบของวิธีการทำให้เป็นบรรทัดฐานที่แตกต่างกันต่อประสิทธิภาพของการรู้จำภาษาตามแนวทางการเรียงหน่วยเสียง

การทดลองที่ 2.1 ถึง 2.3 ใช้ฐานข้อมูลเสียงซึ่งอธิบายในหัวข้อ 3.5.4.1 เพื่อพัฒนาระบบการรู้จำภาษาพูดตามแนวทางการเรียงต่อกันของหน่วยเสียง และวัดประสิทธิภาพของระบบการรู้จำภาษาพูด โดยทดสอบการรู้จำภาษาภายในเซตแบบปิด บนฐานข้อมูลเสียง NIST LRE07 เซตข้อมูลทดสอบซึ่งประกอบด้วย 14 ภาษา โดยฐานข้อมูลเสียงนี้ที่อธิบายในหัวข้อ 3.5.4.2 ขณะที่การทดลองที่ 2.4 พัฒนาระบบการรู้จำภาษาพูด และทดสอบประสิทธิภาพการรู้จำภาษาโดยใช้ฐานข้อมูลเสียงตามหัวข้อ 3.5.3 ซึ่งเป็นข้อมูลเสียงที่ใช้ในการรู้จำภาษาใน 8 ภาษาซึ่งใช้การตรวจสอบแบบไขว้โดยแบ่งข้อมูลออกเป็น 5 เซตเพื่อจำแนกข้อมูลสำหรับฝึกฝนระบบ และทดสอบระบบ ในการทดลองที่ 2.5 ทดสอบประสิทธิภาพการรู้จำภาษาโดยใช้ฐานข้อมูลเสียง NIST LRE07 เซตข้อมูลทดสอบตามหัวข้อ 3.5.4.2 โดยไฟล์ทดสอบสำหรับเป็นข้อมูลเสียงที่ใช้ในการรู้จำภาษามี 8 ภาษาเท่านั้น

การทดลองที่ 2.1: ประเมินประสิทธิภาพของระบบการรู้จำภาษาพูดแบบพ็อดาร์วีเอ็สเอ็มเปรียบเทียบกับระบบการรู้จำภาษาพูดแบบแลตทิซ-วีเอ็สเอ็ม

สมมติฐานของการทดลองคือ ลำดับหน่วยเสียงซึ่งเป็นผลลัพธ์ของการรู้จำเสียง มีข้อผิดพลาดจากขั้นตอนการถอดรหัสเสียง ขณะที่หน่วยเสียงแลตทิซคือ การใช้ผลลัพธ์ของการถอดรหัสเสียงทั้งหมดทำให้สามารถลดความผิดพลาดที่ได้จากการรู้จำเสียงได้

การทดลองนี้เปรียบเทียบผลการรู้จำภาษาตามแนวทางการเรียงหน่วยเสียงเมื่อใช้ข้อมูลจากการรู้จำเสียงที่มีหน่วยเสียงแตกต่างกัน ได้แก่ ระบบการรู้จำภาษาพูดแบบพ็อดาร์วีเอ็สเอ็มซึ่งใช้ผลลัพธ์ที่ได้จากการรู้จำเสียงเป็นหน่วยเสียง และหาความถี่การปรากฏของหน่วยเสียง หรือแบบจำลองภาษาของหน่วยเสียงที่ปรากฏบนเสียงพูด ขณะที่ระบบการรู้จำภาษาพูดแบบแลตทิซ-วีเอ็สเอ็มซึ่งใช้ผลลัพธ์ที่ได้จากการรู้จำเสียงเป็นหน่วยเสียงแลตทิซ และคำนวณค่าความถี่คาดหวังของหน่วยเสียงแลตทิซจากผลรวมของความถี่ทั้งหมดที่ปรากฏบนเสียงพูด แทนความถี่การปรากฏของแบบจำลองภาษาของหน่วยเสียง

การทดลองที่ 2.2: เปรียบเทียบประสิทธิภาพของฟังก์ชันถ่วงน้ำหนักแบบต่าง ๆ

การทดลองนี้เปรียบเทียบผลการรู้จำภาษาตามแนวทางการเรียงหน่วยเสียงเมื่อใช้ฟังก์ชันถ่วงน้ำหนักแบบต่าง ๆ ซึ่งประกอบด้วย: ค่าความถี่เทอม (tf), ความถี่เอกสารผกผัน (idf), ค่าความเกี่ยวข้องกันของเทอม (tr), ค่าสถิติไคสแควร์ (chi) และค่าความซ้ำซ้อนของเทอม (rd) รวมทั้งการประกอบกันระหว่างค่าน้ำหนักแบบท้องถิ่นและค่าน้ำหนักแบบครอบคลุม การรู้จำภาษาเมื่อใช้ฟังก์ชันถ่วงน้ำหนักต่าง ๆ ถูกเปรียบเทียบกับระบบอ้างอิงที่ 1.2 คือ ระบบการรู้จำภาษาพูดตามแนวทางเกณฑ์การเรียงหน่วยเสียงเมื่อใช้ฟังก์ชันถ่วงน้ำหนักที่มาจากการประกอบกันของค่าความถี่เทอม และค่าความถี่เอกสารผกผัน

การทดลองที่ 2.3: เปรียบเทียบประสิทธิภาพของวิธีการทำให้เป็นบรรทัดฐานที่แตกต่างกัน

สมมติฐานของการทดลองนี้มาจากความยาวของเสียงพูดในการรู้จำภาษาแตกต่างกัน ทำให้ค่าความถี่การปรากฏของเทอมที่เป็นสมาชิกในเวกเตอร์อินพุทของแต่ละเสียงพูด มีค่าแตกต่างกันสำหรับเสียงพูดที่มีความยาวแตกต่างกัน เพื่อให้เวกเตอร์อินพุทของการจำแนกมีความสามารถของการบอกความแตกต่างภาษาของทุกเสียงพูดเท่ากัน จึงใช้การทำให้เป็นบรรทัดฐาน

การทดลองนี้เปรียบเทียบผลการรู้จำภาษาตามแนวทางการเรียงหน่วยเสียงเมื่อใช้วิธีการทำให้เป็นบรรทัดฐานที่ต่างกัน ได้แก่ ฟังก์ชันนอร์มอลไลซ์ด้วยวิธีผลรวม และด้วยวิธีระยะห่างแบบยูคลิเดียน โดยวัดผลการรู้จำภาษาของระบบการรู้จำภาษาพูดแบบพ็อดาร์วีเอสเอ็ม และระบบการรู้จำภาษาพูดแบบแลตทิซ-วีเอสเอ็ม

การทดลองที่ 2.4: วัดประสิทธิภาพของระบบการรู้จำภาษาพูดแบบพ็อดาร์วีเอสเอ็ม เปรียบเทียบกับระบบการรู้จำภาษาพูดแบบแลตทิซ-วีเอสเอ็ม โดยวัดผลการรู้จำภาษาบน 8 ภาษา

สำหรับการทดลองนี้วัดผลการรู้จำภาษาตามแนวทางการเรียงหน่วยเสียงของระบบการรู้จำภาษาพูดแบบพ็อดาร์วีเอสเอ็ม และระบบการรู้จำภาษาพูดแบบแลตทิซ-วีเอสเอ็ม โดยใช้ฟังก์ชันถ่วงน้ำหนัก และใช้วิธีการทำให้เป็นบรรทัดฐานที่ให้ประสิทธิภาพการรู้จำที่ดีที่สุด ซึ่งได้จากการเปรียบเทียบการทดลองที่ 2.2 และการทดลองที่ 2.3 ตามลำดับ โดยใช้ฐานข้อมูลเสียง NIST LRE07 เซตข้อมูลพัฒนา ที่วัดผลการรู้จำภาษาบน 8 ภาษาซึ่งอธิบายตามหัวข้อ 3.5.3

การทดลองที่ 2.5: วัดประสิทธิภาพของระบบการรู้จำภาษาพูดแบบพ็อดาร์วีเอสเอ็ม เปรียบเทียบกับระบบการรู้จำภาษาพูดแบบแลตทิซ-วีเอสเอ็ม โดยพัฒนาจาก NIST LRE07 เซตข้อมูลพัฒนา และวัดผลการรู้จำภาษาบน 8 ภาษาบนฐานข้อมูลเสียง NIST LRE07 ตามหัวข้อ 3.5.4.2

นอกจากนี้ นำคะแนนการตัดสินใจที่ได้จากระบบการรู้จำภาษาพูดแบบพ็อดาร์วีเอสเอ็ม และระบบการรู้จำภาษาพูดแบบแลตทิซ-วีเอสเอ็ม สำหรับการรู้จำภาษา 8 ภาษาไปเป็นคะแนนสำหรับขั้นตอนการรวมผลลัพธ์และตัดสินใจการรู้จำภาษาของระบบที่นำเสนอต่อไป

6.4.3 การทดลองที่ 3: ประเมินประสิทธิภาพของระบบการรู้จำภาษาพูดโดยใช้การจำแนกด้วยแบบจำลองลาเทนท์ไดนามิกคอนดิชันนอลแรนดอมฟิลด์ร่วมกับค่าพีเอฟ

ในหัวข้อที่ 5.3 เสนอระบบการรู้จำภาษาพูดโดยใช้การจำแนกด้วยแบบจำลองลาเทนท์ไดนามิกคอนดิชันนอลแรนดอมฟิลด์ร่วมกับค่าพีเอฟ การทดลองที่ 3 นี้จึงเสนอเพื่อประเมินประสิทธิภาพของระบบการรู้จำภาษาพูดดังกล่าว โดยหาค่าพารามิเตอร์ที่เหมาะสม ทดสอบค่าลักษณะสำคัญต่างๆ และเทคนิคที่ใช้ในการจำแนกภาษา รายละเอียดของการทดลองแสดงดังต่อไปนี้

การทดลองที่ 3.1: วัดประสิทธิภาพของระบบการรู้จำภาษาพูดเมื่อใช้ค่าพีเอฟ เปรียบเทียบกับการใช้ค่าลักษณะทางเสียงอื่นๆ

การทดลองนี้วัดประสิทธิภาพของการรู้จำภาษาเปรียบเทียบผลของการใช้ค่าพีเอฟกับการใช้ค่าลักษณะสำคัญทางเสียงชนิดอื่น ซึ่งใช้กันในงานการรู้จำภาษา และการรู้จำเสียง ตารางที่ 6-2 แสดงค่าลักษณะสำคัญที่ใช้ในการฝึกฝน และทดสอบตัวจำแนกภาษา และวัดความแม่นยำของตัวจำแนกภาษาเมื่อใช้ค่าพีเอฟ เปรียบเทียบกับการใช้ค่าลักษณะสำคัญทางเสียงอื่นๆ โดยพิจารณา

คำตอบจากผลลัพธ์ของการจำแนกแบบลำดับ (Sequence classification) ที่นำผลลัพธ์ของทุกเฟรมบนเสียงพูดทั้งหมดมาตัดสินใจคำตอบ

ตัวจำแนกภาษาสำหรับ 8 ภาษามีจำนวน $C(8, 2)$ หรือ 28 คู่ภาษา ตัวจำแนกแต่ละตัวฝึกฝนด้วยค่าลักษณะสำคัญทางเสียงที่แตกต่างกัน และวัดความแม่นยำจากค่าเฉลี่ยจากตัวจำแนกภาษาทั้งหมด



ตารางที่ 6-2 ค่าลักษณะสำคัญทางเสียงที่ใช้ในการทดลองเพื่อวัดประสิทธิภาพของการรู้จำภาษาเมื่อใช้ค่าลักษณะสำคัญทางเสียงชนิดต่าง ๆ

ค่าลักษณะสำคัญ	คำอธิบาย	จำนวนมิติ
PF	ค่าพีเอฟ	26
SDCPF	ค่าลักษณะสำคัญเอสดีซีซีที่คำนวณจากค่าพีเอฟ	182
PLP	ค่าพีแอลพี 7 ค่า และค่าพลังงาน	8
MFCC	(ค่าสัมประสิทธิ์ MFCC 12 ค่าและค่าพลังงาน) $+\Delta + \Delta\Delta$	39
SDCPLP	ค่าพีแอลพี 7 ค่า และค่าพลังงาน เพื่อคำนวณค่าลักษณะสำคัญเอสดีซีซีที่คำนวณจากค่า PLP	56
SDC+PLP	ค่าลักษณะสำคัญเอสดีซีซีที่คำนวณจากค่า PLP 56 ค่า ต่อกับค่า PLP 7 ค่า	64

ค่าลักษณะสำคัญต่าง ๆ ที่แสดงจากตารางที่ 6-2 มีรายละเอียดดังนี้

- 1) กลุ่มค่าลักษณะสำคัญเซปตรอล (Cepstral feature) การสกัดค่าลักษณะสำคัญเซปตรอลประกอบด้วย ค่าสัมประสิทธิ์ MFCC และ PLP ซึ่งแต่ละเฟรมถูกวิเคราะห์ด้วยรอบหน้าต่างที่มีขนาด 25 มิลลิวินาที มีช่วงเหลื่อมซ้อนทับกัน 10 มิลลิวินาที เวกเตอร์ลักษณะสำคัญของสัมประสิทธิ์ MFCC ประกอบด้วย ค่าสัมประสิทธิ์ MFCC 12 ค่า ค่าพลังงาน และค่าสัมประสิทธิ์ MFCC และค่าพลังงานเสียงที่คำนวณการเปลี่ยนแปลงเซปสตรัมทางเวลาจากสมการอนุพันธ์อันดับหนึ่ง และอนุพันธ์อันดับสอง ส่วนเวกเตอร์ลักษณะสำคัญของค่าสัมประสิทธิ์ PLP ประกอบด้วย ค่าสัมประสิทธิ์ PLP 7 ค่าและค่าพลังงาน
- 2) กลุ่มค่าลักษณะสำคัญเอสดีซีซี (SDC feature): จากหัวข้อ 5.3 การคำนวณเอสดีซีซีถูกนำไปใช้คำนวณการเปลี่ยนแปลงค่าบนค่าพีเอฟ และในหัวข้อนี้การคำนวณเอสดีซีซีถูกนำมาใช้ในการคำนวณเอสดีซีซีเพื่อหาค่าเปลี่ยนแปลงของค่าสัมประสิทธิ์ PLP ซึ่งผลลัพธ์ของการคำนวณนี้คือ ค่าเอสดีซีซีพีแอลพี แทนด้วย SDCPLP นอกจากนี้นำค่าเอสดีซีซีพีแอลพี ที่คำนวณได้มาประกอบกันกับค่าสัมประสิทธิ์ PLP แทนด้วย SDC+PLP ซึ่งทั้งสองค่าลักษณะสำคัญที่ได้จากหัวข้อนี้ นำไปใช้ในการทดสอบเพื่อเปรียบเทียบประสิทธิภาพของการรู้จำภาษาเมื่อใช้ค่าลักษณะสำคัญทางเสียงชนิดต่าง ๆ โดยค่าพารามิเตอร์ที่ใช้ในการคำนวณเอสดีซีซีในการทดลองนี้คือ 7-1-3-7 อิงตามงานวิจัยที่ศึกษาค่าพารามิเตอร์เอสดีซีซีที่เหมาะสมที่ให้ผลการรู้จำภาษาดีที่สุที่สุด [52]

การทดลองที่ 3.2: ศึกษาผลกระทบเมื่อลดกลุ่มภาษาเป้าหมายที่มีความคลุมเครือต่อประสิทธิภาพของระบบการรู้จำภาษาพูดที่ใช้ค่าพีเอฟ และระบบการรู้จำภาษาพูดที่ใช้ค่าลักษณะทางเสียงอื่น ๆ

การทดลองนี้ประเมินผลความสามารถของตัวจำแนกเมื่อลดความคลุมเครือระหว่างภาษานั้นคือ ในวิทยานิพนธ์นี้มีภาษาจีน 3 ภาษาที่เป็นภาษาเป้าหมาย ได้แก่ ภาษาจีนมิน ภาษาจีนวู และภาษาจีนกวางตุ้ง เพื่อลดความคลุมเครือระหว่างภาษาจีนจึงลดภาษาที่ทำให้เกิดความสับสนในการรู้จำ โดยพิจารณาภาษาเป้าหมายให้เหลือ 6 ภาษา โดยไม่รวมภาษาจีนมิน และภาษาจีนวู ดังนั้นตัวจำแนกภาษาสำหรับแต่ละคู่ภาษามีจำนวน 15 คู่ภาษาซึ่งมาจากการเลือกจับคู่ $C(5, 2)$ ด้วยเหตุนี้จึงวัดความสามารถของการจำแนกภาษา เมื่อลดความคลุมเครือที่ทำให้สับสนในการจำแนกเปรียบเทียบกับผลการทดลองที่ 5.1 ที่ให้ผลการจำแนกที่ไม่ได้มีการลดความคลุมเครือระหว่างภาษา ดังนั้นจึงใช้ค่าลักษณะสำคัญ และค่าพารามิเตอร์ที่ใช้ในการฝึกฝนตัวจำแนกเหมือนกับการทดลองที่ 5.1 จากนั้นจึงวัดความสามารถของการบอกความแตกต่างระหว่างภาษาเมื่อลดภาษาที่ทำให้คลุมเครือออกและดูผลกระทบที่เกิดจากความสับสนระหว่างภาษา โดยวัดผลการจำแนกภาษาเมื่อใช้ค่าพีเอฟ และค่าลักษณะสำคัญทางเสียงอื่น ๆ

การทดลองที่ 3.3: เปรียบเทียบประสิทธิภาพของระบบการรู้จำภาษาพูดที่ใช้ตัวจำแนกด้วยแบบจำลองลาเทนท์ไดนามิคคอนดิชันนอลแรนดอมฟิลด์ กับแบบจำลองชนิดอื่น

การทดลองที่ 3.3.1: การทดลองนี้เปรียบเทียบประสิทธิภาพของตัวจำแนกด้วยแบบจำลองลาเทนท์ไดนามิคคอนดิชันนอลแรนดอมฟิลด์ซึ่งเป็นแบบจำลองดิสคริมิทีฟ กับตัวจำแนกที่ใช้แบบจำลองเกาส์เซียนซึ่งเป็นแบบเจนเนอเรทีฟ โดยใช้แบบจำลองเกาส์เซียนแบบผสม สำหรับแต่ละภาษา ฝึกฝนจากค่าพีเอฟที่สกัดจากเสียงพูดของแต่ละภาษา ร่วมกับอัลกอริทึมการคาดหมายสูงสุด ซึ่งมีพารามิเตอร์กำหนดค่าเริ่มต้นจากการวนซ้ำด้วยวิธีการหาค่าโอกาสการเกิดสูงสุด 10 รอบ ในการฝึกฝนแบบจำลองเกาส์เซียนแบบผสมที่มีจำนวนเกาส์เซียนผสม 256 ส่วนประกอบ นอกจากนี้ในการทดลองนี้ได้ฝึกฝนแบบจำลองเกาส์เซียนด้วยค่าพีแอลพีเป็นค่าลักษณะสำคัญ เพื่อเปรียบเทียบความสามารถของการจำแนกภาษาเมื่อใช้แบบจำลองเกาส์เซียนที่มาจากค่าลักษณะสำคัญทางเสียงต่างกัน

การทดลองที่ 3.3.2: การทดลองนี้เปรียบเทียบตัวจำแนกแบบจำลองลาเทนท์ไดนามิคคอนดิชันนอลแรนดอมฟิลด์ กับตัวจำแนกที่ใช้แบบจำลองคอนดิชันนอลแรนดอมฟิลด์ซึ่งแทนด้วย CRF ซึ่งเป็นแบบจำลองดิสคริมิเนทีฟอีกหนึ่งวิธีที่นิยมใช้ในการแก้ปัญหาแบบลำดับ ตัวจำแนกภาษาที่ใช้แบบจำลอง CRF ถูกฝึกฝนด้วยพารามิเตอร์เดียวกันกับพารามิเตอร์ที่ใช้ในการฝึกฝนตัวจำแนกภาษา LDCRF และใช้ค่าพีเอฟเป็นค่าลักษณะสำคัญทางเสียงเพื่อฝึกฝนตัวจำแนกภาษา

การทดลองที่ 3.4: วัดประสิทธิภาพของระบบการรู้จำภาษาพูดเมื่อลดชนิดค่าพีเอฟที่ใช้ฝึกฝนตัวจำแนกภาษา

การทดลองนี้วิเคราะห์ผลกระทบความสามารถในการบอกความแตกต่างระหว่างภาษา เมื่อลดค่าพีเอฟแต่ละชนิดจากเวกเตอร์ค่าลักษณะค่าพีเอฟ และวัดประสิทธิภาพของตัวจำแนกเมื่อทำการลดค่าพีเอฟเปรียบเทียบกับตัวจำแนกที่ไม่มีการลดค่าพีเอฟ

ในการทดลองที่ 3.1 เวกเตอร์ค่าลักษณะเฉพาะของค่าพีเอฟประกอบด้วยค่าพีเอฟ 13 ชนิด เวกเตอร์ค่าลักษณะเฉพาะของค่าพีเอฟที่พิจารณาสำหรับการทดลองนี้ ประกอบด้วยค่าพีเอฟ 12 ชนิด ซึ่งเกิดจากรีดค่าพีเอฟที่ละชนิดออกจากเวกเตอร์ค่าลักษณะเฉพาะ และนำมาเป็นอินพุทของตัวจำแนก เพื่อพิจารณาผลกระทบเมื่อขาดค่าพีเอฟแต่ละชนิด โดยใช้ค่าพารามิเตอร์สำหรับการฝึกฝนตัวจำแนกที่เหมือนกับการทดลองที่ 3.1 เพื่อควบคุมตัวแปรที่เกี่ยวข้องกับการเพิ่มประสิทธิภาพของตัวจำแนก และสามารถเปรียบเทียบผลการจำแนกแบบจำลองลาเทนท์ไดนามิคคอนดิชันนอลแรนดอมฟิลด์ ที่ใช้เวกเตอร์ค่าลักษณะเฉพาะค่าพีเอฟที่ลดชนิดค่าพีเอฟ กับการจำแนกที่ไม่มีการลดค่าพีเอฟ

การทดลองที่ 3.5: วัดประสิทธิภาพของระบบการรู้จำภาษาพูดเปรียบเทียบกับระบบอ้างอิงโอเวกเตอร์ (i-vector)

การทดลองประเมินประสิทธิภาพของระบบการรู้จำภาษาพูดที่นำเสนอ เปรียบเทียบกับระบบอ้างอิงที่ 3 ซึ่งเป็นระบบการรู้จำภาษาพูดตามแนวอะคูสติกที่ใช้โอเวกเตอร์ [119] ที่เป็นที่ยอมรับ เพื่อประเมินประสิทธิภาพของระบบที่นำเสนอ จากการวัดคะแนนความเหมือนของโอเวกเตอร์ด้วยวิธี pLDA

การทดลองที่ 3.6: วัดประสิทธิภาพของการรู้จำภาษาของระบบการรู้จำภาษาพูดที่ใช้ตัวจำแนกแบบหนึ่งต่อหนึ่ง และตัวจำแนกแบบหนึ่งต่อทั้งหมด

การทดลองที่ 3.1 ถึง 3.5 เป็นการวัดประสิทธิภาพจากค่าความผิดพลาดของตัวจำแนกภาษาสำหรับแต่ละคู่ภาษา อย่างไรก็ตามคำตอบการรู้จำภาษาเป็นการตัดสินใจระบุภาษาแบบปัญหาที่มีลักษณะเป็นหลายกลุ่ม ดังนั้นจากผลการจำแนกสำหรับแต่ละคู่ภาษาจากตัวจำแนกสำหรับแต่ละคู่ภาษา หรือแบบหนึ่งต่อหนึ่ง จึงต้องนำมาตัดสินใจระบุภาษาแบบปัญหาการจำแนกหลายกลุ่ม โดยพิจารณาคะแนนการโหวตร่วมกันของคำตอบการตัดสินใจแบบถาวรส่วนใหญ่ของตัวจำแนกสำหรับแต่ละคู่ภาษา ดังสมการที่ (6.3) เมื่อ s_{ij} แทนผลลัพธ์การตัดสินใจแบบถาวรจากตัวจำแนกภาษาแบบหนึ่งต่อหนึ่งโดย s_{ij} มีค่าเป็น 1 เมื่อคะแนนการตัดสินใจจากตัวจำแนกภาษาสำหรับคู่ภาษา (i, j) หรือแทนด้วย r_{ij} มีค่ามากกว่าคะแนนการตัดสินใจจากตัวจำแนกภาษาสำหรับคู่ภาษา (j, i) หรือแทนด้วย r_{ji}

$$Class = \arg \max_{i=1..L} \sum_{1 \leq j \neq i \leq m} s_{ij} \quad (6.3)$$

เพื่อประเมินประสิทธิภาพของการรวมคำตอบจากคะแนนโหวตของการตัดสินใจจากตัวจำแนกภาษาแบบหนึ่งต่อหนึ่ง จึงเปรียบเทียบกับคำตอบของการระบุภาษาเป็นคำตอบการตัดสินใจแบบถาวรจากตัวจำแนกแบบหนึ่งต่อทั้งหมด หรือภาษาเป้าหมายหนึ่งภาษาต่อภาษาที่ไม่ใช่ภาษาเป้าหมายทั้งหมด

นอกจากนี้วิธีการวัดประสิทธิภาพของการรู้จำภาษาสามารถพิจารณาจากการตัดสินใจแบบเปลี่ยนได้ โดยพิจารณาจากคะแนนความน่าจะเป็นของการรู้จำภาษาที่ระบบให้คำตอบ ซึ่งแตกต่างจากการวัดประสิทธิภาพของการทดลองที่ 3.1 ถึง 3.5 เป็นการประเมินจากผลลัพธ์ของการตัดสินใจแบบถาวร คำตอบของการระบุภาษาพิจารณาจากคะแนนของการตัดสินใจแบบเปลี่ยนได้ จากการตัดสินใจการระบุภาษาจากตัวจำแนกภาษาแบบหนึ่งต่อหนึ่งทุกตัว ซึ่งคำนวณจากผลคูณของความน่าจะเป็นของตัวจำแนกภาษาสำหรับแต่ละคู่ภาษา ดังสมการที่ (6.4) เมื่อค่า $P(X|\lambda_i)$ แทนค่าความน่าจะเป็นของการระบุภาษาของข้อมูล X สำหรับแบบจำลองจากตัวจำแนกภาษา λ_i และ L แทนจำนวนของแบบจำลองสำหรับแต่ละตัวจำแนก ซึ่งมีตัวจำแนกภาษาทั้งหมดสำหรับแต่ละคู่ภาษา 28 ตัว

$$P_i = \prod_{i=1}^L P(X|\lambda_i) \quad (6.4)$$

การทดลองนี้เปรียบเทียบค่าความผิดพลาดของการจำแนกภาษา และค่าความผิดพลาดการรู้จำภาษา EER จากคะแนนของการตัดสินใจแบบเปลี่ยนได้จากทั้งสองวิธี และผลลัพธ์ที่ได้ของวิธีที่ให้ค่าความผิดพลาดในการรู้จำภาษา EER ค่าน้อยกว่าจะถูกนำไปใช้ในการรวมคะแนนการตัดสินใจของระบบที่นำเสนอ

6.4.4 การทดลองที่ 4: ประเมินประสิทธิภาพของการรู้จำภาษาตามแนวทางการเรียงหน่วยเสียงโดยใช้ค่าพีเอฟ

การทดลองนี้วัดประสิทธิภาพของการรู้จำภาษาตามแนวทางการเรียงตัวของเสียงโดยใช้ค่าพีเอฟ แบ่งการทดลองเป็น 5 แบบเพื่อศึกษาผลกระทบของค่าพารามิเตอร์ที่มีผลต่อการรู้จำภาษา ดังนี้

การทดลองที่ 4.1: ประเมินประสิทธิภาพของการรู้จำภาษาตามแนวทางการเรียงหน่วยเสียงที่ใช้ค่าพีเอฟ โดยใช้การพิจารณาโคออคเคอร์เรนซ์ค่าพีเอฟ

การทดลองนี้วัดประสิทธิภาพของการรู้จำภาษาตามแนวทางการเรียงหน่วยเสียงที่ใช้ค่าพีเอฟ โดยพิจารณาการปรากฏร่วมกันที่นำเสนอ และหาตัวแทนที่เหมาะสมเพื่อแสดงความสัมพันธ์ของค่าพีเอฟที่แตกต่างกันในแต่ละภาษา ซึ่งใช้จำนวนการปรากฏร่วมกันของค่าพีเอฟที่เหมาะสมในระบบการรู้จำภาษาพูดแตกต่างกัน โดยพิจารณาความถี่ของการปรากฏของตัวแทนในกรอบการพิจารณาขนาด 2 เฟรม หรือแบบ win2 ดังนี้

การทดลองที่ 4.1.1: เสนอโคออคเคอร์เรนซ์ค่าพีเอฟ เพื่อเป็นตัวแทนพิจารณาการปรากฏร่วมกันของค่าพีเอฟสองชนิด โดยกำหนดให้พิจารณาหนึ่งค่าพีเอฟหลักปรากฏร่วมกับค่าพีเอฟอื่นๆ เช่น จับคู่ค่าพีเอฟ Vocalic เป็นค่าพิจารณาหลักร่วมกับค่าพีเอฟชนิดอื่นๆ เป็นต้น ซึ่งรูปแบบโคออคเคอร์เรนซ์ค่าพีเอฟมาจากการจับคู่ที่ละสองค่าพีเอฟเมื่อกำหนดค่าพีเอฟหลักหนึ่งค่า ทำให้ซูเปอร์เวกเตอร์ที่มีจำนวนมิติ 2,304 มิติ

การทดลองที่ 4.1.2: เสนอโคออคเคอร์เรนซ์ค่าพีเอฟ เพื่อเป็นตัวแทนพิจารณาการใช้ค่าพีเอฟที่ทุกคู่ของเหตุการณ์ที่เกิดขึ้นร่วมกันของค่าพีเอฟสองค่า โดยให้พิจารณาจับคู่การปรากฏร่วมกัน

ของค่าพีเอฟสองค่าทุกกรณี ซึ่งการเลือกค่าพีเอฟทีละสองค่าจากค่าพีเอฟทั้งหมด 13 ชนิด ทำให้ ซุปเปอร์เวกเตอร์มีจำนวนมิติขนาด 97,344 มิติ

เนื่องจากระบบการรู้จำภาษาพูดที่เสนอใช้ความถี่ของการปรากฏร่วมกันของเหตุการณ์ของ สมาชิกสองตัว เป็นโคออคเคอร์เรนซ์ค่าพีเอฟสองตัว ซุปเปอร์เวกเตอร์ที่ได้จากการจับคู่กันทุกคู่ของ โคออคเคอร์เรนซ์ค่าพีเอฟตามการทดลองที่ 4.1.2 ทำให้มีสมาชิกจำนวนมาก ขณะที่ซุปเปอร์เวกเตอร์ ตามตามการทดลองที่ 4.1.1 มีขนาดมิติน้อยกว่า ผู้วิจัยจึงสนใจตัวแทนจากการปรากฏร่วมกันของค่า พีเอฟจากการพิจารณาทีละค่าพีเอฟหลักเนื่องจากมีจำนวนมิติของเวกเตอร์ที่น้อยกว่า ว่ามีผลกระทบต่อประสิทธิภาพของการจำแนกภาษาหรือไม่

การทดลองที่ 4.2: ศึกษาผลกระทบของขนาดกรอบการพิจารณาของการปรากฏร่วมกันของ โคออคเคอร์เรนซ์ค่าพีเอฟต่อประสิทธิภาพการรู้จำภาษาแนวทางการเรียงหน่วยเสียงโดยใช้ค่าพีเอฟ

แนวคิดของการทดลองนี้มาจาก ขนาดของกรอบการพิจารณา และช่วงการเลื่อนกรอบการ พิจารณา (Frame shift period) มีผลกระทบต่อรูปแบบของความถี่การปรากฏร่วมกันของโคออค เคอร์เรนซ์ค่าพีเอฟในแต่ละกรอบ ซึ่งนำไปสู่การบอกลักษณะเฉพาะของแต่ละภาษา นอกจากนี้ขนาด กรอบการพิจารณาสามารถทำให้เกิดรูปแบบที่ซ้ำซ้อนกันซึ่งส่งผลให้ความสามารถในการบอกความ แตกต่างระหว่างภาษาลดลงได้ ดังนั้นการทดลองนี้วัดผลกระทบของขนาดกรอบการพิจารณาเพื่อหา การปรากฏร่วมกันของโคออคเคอร์เรนซ์ค่าพีเอฟสองตัว เมื่อใช้ขนาดกรอบที่แตกต่างกัน ซึ่งขนาด กรอบการพิจารณาที่ใช้ในการทดลองนี้ ได้แก่ i) ขนาด 2 เฟรมเลื่อนกรอบพิจารณาทีละ 1 เฟรม (เหลื่อมซ้อนทับกันร้อยละ 50), ii) ขนาด 3 เฟรมเลื่อนกรอบพิจารณาทีละ 2 เฟรม (เหลื่อมซ้อนทับ กันร้อยละ 66), iii) ขนาด 4 เฟรมเลื่อนกรอบพิจารณาทีละ 3 เฟรม (เหลื่อมซ้อนทับกันร้อยละ 75), iv) ขนาด 5 เฟรมเลื่อนกรอบพิจารณาทีละ 3 เฟรม (เหลื่อมซ้อนทับกันร้อยละ 60) เป็นต้น ขนาด กรอบพิจารณา 3 ถึง 5 เฟรมอนุญาตให้มีช่องว่างระหว่างเฟรมของการเกิดโคออคเคอร์เรนซ์ค่าพีเอฟ ร่วมกัน ขณะที่กรณี i) ที่มีรูปแบบการพิจารณาเหมือนแบบไบแกรมโดยไม่มีช่องว่างระหว่างเฟรม ของการปรากฏร่วมกันของโคออคเคอร์เรนซ์ ผลลัพธ์ของการใช้กรอบการพิจารณาขนาดต่าง ๆ ที่ให้ ประสิทธิภาพดีที่สุด จะนำไปใช้ในการทดลองต่อไป

การทดลองที่ 4.3: ศึกษาผลกระทบของชนิดค่าพีเอฟที่พิจารณาการปรากฏร่วมกันของค่าพี เอฟต่อประสิทธิภาพการรู้จำภาษาแนวทางการเรียงหน่วยเสียงโดยใช้ค่าพีเอฟ

จากการทดลองที่ 4.1 และ 4.2 กำหนดให้พิจารณาหนึ่งค่าพีเอฟหลักปรากฏร่วมกับค่าพีเอฟ อื่นๆ ในการทดลองนี้ศึกษาผลกระทบของการพิจารณาโคออคเคอร์เรนซ์ค่าพีเอฟ เมื่อใช้ชนิดค่าพีเอฟ ที่แตกต่างกันเพื่อกำหนดเป็นค่าพีเอฟหลักปรากฏร่วมกับค่าพีเอฟอื่นๆ เช่น COPF(Vocalic, Consonantal), COPF(Vocalic, Nasal) เป็นต้น การพิจารณาจากชนิดของค่าพีเอฟที่แตกต่างกันทำ ให้ระบบรู้จำสามารถบอกความแตกต่างระหว่างภาษาได้ดีขึ้นหรือไม่

การทดลองที่ 4.4: ศึกษาผลกระทบของการเพิ่มจำนวนโคออคเคอร์เรนซ์ค่าพีเอฟต่อ ประสิทธิภาพการรู้จำภาษาแนวทางการเรียงหน่วยเสียงค่าพีเอฟ

แนวคิดของการทดลองนี้มาจากข้อมูลที่เพิ่มขึ้นจากการเกิดร่วมกันของโคออกเคอร์เรนซ์ค่าพีเอฟหลายชนิดในแต่ละเฟรม ทำให้ได้รูปแบบการพิจารณาการเกิดร่วมกันของการออกเสียงหลายชนิดในแต่ละเฟรมเพิ่มมากขึ้น ซึ่งส่งผลให้ระบบรู้จำสามารถจำแนกภาษาได้ดีขึ้น

จากการทดลองที่ 4.1 ถึง 4.3 ใช้โคออกเคอร์เรนซ์ค่าพีเอฟ 1 แบบซึ่งใช้เป็นตัวแทนการพิจารณาการปรากฏร่วมกันของค่าพีเอฟ 2 ชนิดในแต่ละเฟรม ในการทดลองนี้เพิ่มจำนวนค่าพีเอฟที่พิจารณาให้มากขึ้น ซึ่งแทนการปรากฏร่วมกันของค่าพีเอฟ 3 หรือ 4 ชนิดในแต่ละเฟรม โดยการพิจารณาการเกิดร่วมกันโคออกเคอร์เรนซ์ค่าพีเอฟ 2 โคออกเคอร์เรนซ์ค่าพีเอฟในแต่ละเฟรม เช่น โคออกเคอร์เรนซ์ค่าพีเอฟ Vocalic ร่วมกับ โคออกเคอร์เรนซ์ค่าพีเอฟ Nasal ซึ่งแทนรูปแบบที่พิจารณาด้วย COPF(Vocalic, *) และ COPF(Nasal, *) เมื่อ * แทนค่าพีเอฟอื่นๆที่ปรากฏร่วมกับค่าพีเอฟหลักที่กำหนด แต่ละการปรากฏร่วมกันของค่าพีเอฟมีจำนวนรูปแบบ 24 แบบ ทำให้รูปแบบของการเกิดร่วมกันของโคออกเคอร์เรนซ์ค่าพีเอฟเป็น $24 \times 24 \times 16 = 9216$ แบบ และมีการเลือกรูปแบบการเกิดกันของ 2 ค่าพีเอฟ 78 แบบ

การทดลองที่ 4.5: เปรียบเทียบประสิทธิภาพการรู้จำภาษาจากระบบที่นำเสนอกับผลการรู้จำภาษาของระบบอ้างอิงตามแนวทางการเรียงหน่วยเสียง

การทดลองนี้วัดประสิทธิภาพของระบบที่นำเสนอระบบการรู้จำภาษาพูดที่ใช้ค่าพีเอฟ ที่มีค่าพารามิเตอร์ที่ให้ประสิทธิภาพดีที่สุด ซึ่งได้จากผลลัพธ์ของการทดลองที่ 4.1 ถึง 4.4 โดยใช้ความถี่การปรากฏของโคออกเคอร์เรนซ์ค่าพีเอฟในการจำแนกภาษา เปรียบเทียบกับระบบอ้างอิงที่ 2.1 ซึ่งเป็นระบบการรู้จำภาษาพูดแบบพ็อดาร์วีเอ็สเอ็มใช้เกณฑ์การเรียงหน่วยเสียง และระบบอ้างอิงที่ 2.2 ระบบการรู้จำภาษาพูดที่ใช้ค่าพีเอฟของ Nunez และ Ramon [92] ซึ่งเป็นงานวิจัยล่าสุดที่ใช้ค่าพีเอฟ โดยระบบอ้างอิงที่ 2.1 และระบบอ้างอิงที่ 2.2 พัฒนารูปแบบตัวจำแนกภาษาแบบหนึ่งต่อหนึ่ง และแบบหนึ่งต่อทั้งหมด

เนื่องจากในการทดลองที่ 4.1 ถึง 4.4 เป็นการประเมินผลการจำแนกภาษาเมื่อใช้ตัวจำแนกภาษาสำหรับแต่ละคู่ภาษา 28 คู่ภาษา เพื่อวิเคราะห์ความสามารถของการระบุภาษาสำหรับแบบคู่ภาษา ซึ่งคะแนนของตัวจำแนกสะท้อนความสามารถของการรู้จำภาษาระหว่างคู่ภาษาได้ อย่างไรก็ตามเพื่อนำคะแนนของระบบรู้จำแนวทางนี้ไปใช้เพื่อตัดสินร่วมกับระบบแนวทางอื่นจึงต้องการคะแนนการตัดสินใจแบบเปลี่ยน จากความน่าจะเป็นการเกิดแต่ละภาษาที่พิจารณา และจากผลการประเมินความแม่นยำของการจำแนกภาษาที่ตัดสินใจจากโหวตร่วมกันของคำตอบการตัดสินใจแบบถาวรส่วนใหญ่จากตัวจำแนกภาษาแบบหนึ่งต่อหนึ่งให้ผลไม่แตกต่างกับความแม่นยำของการจำแนกภาษาที่ตัดสินใจจากการใช้ตัวจำแนกแบบหนึ่งต่อทั้งหมด ดังแสดงในการทดลองที่ 3.6 ดังนั้นเพื่อประเมินความผิดพลาดของการรู้จำภาษา จึงใช้คะแนนแบบตัดสินใจเปลี่ยนได้จากคะแนนที่ได้จากการตัดสินใจของตัวจำแนกสำหรับแต่ละภาษาจากการใช้ตัวจำแนกแบบหนึ่งต่อทั้งหมด โดยในการทดลองนี้สร้างตัวจำแนกภาษาแบบหนึ่งภาษาต่อภาษาอื่นที่ไม่ใช่ภาษาที่พิจารณาทั้งหมด รวมทั้งสิ้น 8 ตัวจำแนกและวัดผลการรู้จำภาษา ที่พิจารณาผลลัพธ์ของระบบการรู้จำภาษาที่ตัดสินใจของตัวจำแนกภาษาแบบหนึ่งต่อทั้งหมดบนเสียงพูดทดสอบ

6.4.5 การทดลองที่ 5: ประเมินประสิทธิภาพของระบบการรู้จำภาษาพูดที่นำเสนอ เมื่อรวมผลลัพธ์และการตัดสินใจรู้จำภาษาจากระบบรู้จำย่อยแนวต่าง ๆ

การทดลองนี้ประเมินประสิทธิภาพของการรวมผลลัพธ์การรู้จำภาษาจากระบบการรู้จำภาษาตามแนวทางต่าง ๆ โดยใช้วิธีการรวมคะแนนแบบเชิงเส้น และกระบวนการทำงานส่วนหลังด้วยแบบจำลองเกาส์เซียน โดยเวกเตอร์ของคะแนนที่ได้จากแต่ละระบบย่อยถูกนำมาใช้พิจารณาคะแนนการตัดสินใจสำหรับการรู้จำภาษาของระบบที่นำเสนอ ซึ่งระบบการรู้จำภาษาพูดแนวต่าง ๆ ประกอบด้วย 4 ระบบย่อย

- 1) ระบบการรู้จำภาษาพูดตามแนวทางการเรียงหน่วยเสียง ที่มาจากระบบการรู้จำภาษาพูดแบบพ็อรวีเอ็สเอ็มซึ่งได้จากผลการทดลองที่ 2.4
- 2) ระบบการรู้จำภาษาพูดตามแนวทางการเรียงหน่วยเสียง ที่มาจากระบบการรู้จำภาษาพูดแบบแลตทิซ-วีเอ็สเอ็มซึ่งได้จากผลการทดลองที่ 2.4
- 3) ระบบการรู้จำภาษาพูดที่ใช้ค่าพีเอฟ ที่มาจากระบบการรู้จำภาษาพูดโดยใช้การจำแนกด้วยแบบจำลองลาเทนท์ไดนามิกคอนดิชันนอลแรนดอมฟิลด์ร่วมกับค่าพีเอฟ ซึ่งได้จากผลการทดลองที่ 3.6
- 4) ระบบการรู้จำภาษาพูดที่ใช้ค่าพีเอฟ ที่มาจากระบบการรู้จำภาษาพูดตามแนวทางการเรียงหน่วยเสียงโดยใช้ค่าพีเอฟซึ่งได้จากผลการทดลองที่ 4.5

โดยวัดประสิทธิภาพของการรู้จำภาษาที่ได้จากการตัดสินใจรวมคะแนนระบบการรู้จำภาษาพูดจาก 2 ระบบ 3 ระบบ และ 4 ระบบ

สำหรับการวัดประสิทธิภาพของการตัดสินใจที่ได้ในขั้นตอนสุดท้าย แบ่งการทดสอบออกเป็น 2 ส่วนเพื่อวัดผลการรู้จำบนเซตข้อมูลทดสอบคนละเซตกัน ได้แก่

1) ข้อมูลทดสอบแบบที่ 1: ชุดข้อมูลทดสอบที่ได้จากฐานข้อมูลเสียงตามหัวข้อ 3.5.3 ซึ่งเป็นข้อมูลเสียงที่ใช้ในการรู้จำภาษาใน 8 ภาษาซึ่งข้อมูลถูกแบ่งออกเป็น 5 ส่วนสำหรับการตรวจสอบแบบไขว้ ดังนั้นเพื่อหาค่าพารามิเตอร์สำหรับการรวมคะแนนแบบเชิงเส้น และการรวมโดยใช้การทำงานส่วนหลังด้วยแบบจำลองเกาส์เซียน ระบบการรู้จำภาษาพูดที่ใช้ประเมินผลสร้างมาจากข้อมูลฝึกฝนจากแต่ละส่วนไขว้ และนำมาตัดสินใจภาษา คะแนนของการตัดสินใจที่ได้จากการทดสอบโดยใช้ข้อมูลเสียงชุดฝึกฝนของแต่ละส่วนไขว้จึงนำมาใช้ในหาค่าพารามิเตอร์สำหรับขั้นตอนการรวมคะแนนของแต่ละวิธีข้างต้น และใช้ข้อมูลทดสอบแบบที่ 1 เพื่อประเมินผลการรู้จำภาษาสำหรับแต่ละส่วนไขว้

2) ข้อมูลทดสอบแบบที่ 2: ชุดทดสอบที่ใช้ข้อมูลเสียงพูดบนฐานข้อมูลเสียง NIST LRE07 เซตข้อมูลทดสอบตามหัวข้อ 3.5.4.2 ซึ่งไฟล์ทดสอบสำหรับเป็นข้อมูลเสียงที่ใช้ในการรู้จำภาษาใน 8 ภาษาเท่านั้น ดังนั้นการหาค่าพารามิเตอร์สำหรับการรวมคะแนนแบบเชิงเส้น และการรวมโดยใช้การทำงานส่วนหลังด้วยแบบจำลองเกาส์เซียนได้มาจากคะแนนของการตัดสินใจที่ได้จากการทดสอบโดยใช้ข้อมูลเสียงชุดฝึกฝน และใช้ข้อมูลส่วนทดสอบแบบที่ 2 มาประเมินผลการรู้จำภาษา

บทที่ 7

ผลการทดลองและวิเคราะห์ผลการทดลอง

เนื้อหาในบทนี้แสดงผลการวิเคราะห์ประสิทธิภาพระบบการรู้จำภาษาพูดแนวทางต่าง ๆ ซึ่งอธิบายในบทที่ 4 และ 5 และวิธีการวัดผลการทดลองอธิบายในบทที่ 6 โดยแบ่งตามหัวข้อดังนี้

- 1) ประเมินประสิทธิภาพของระบบตรวจหาลักษณะเฉพาะค่าพีเอฟ
- 2) ประเมินประสิทธิภาพของระบบการรู้จำภาษาพูดตามแนวทางการเรียงหน่วยเสียง
- 3) ประเมินประสิทธิภาพของระบบการรู้จำภาษาพูดโดยใช้การจำแนกด้วยแบบจำลองลาเทนท์ไดนามิกคอนดิชันนอลแรนดอมฟิลด์ร่วมกับค่าพีเอฟ
- 4) ประเมินประสิทธิภาพของการรู้จำภาษาแบบเรียงหน่วยเสียงโดยใช้ค่าพีเอฟ
- 5) ประเมินประสิทธิภาพของระบบการรู้จำภาษาพูดจากแนวทางที่นำเสนอที่รวมคะแนนจากแต่ละแนวทาง

7.1 ประเมินผลระบบตรวจหาลักษณะเฉพาะค่าพีเอฟ

ผลการทดลองที่ 1 ซึ่งประเมินประสิทธิภาพของระบบการตรวจหาลักษณะเฉพาะค่าพีเอฟแต่ละชนิด เพื่อที่จะนำไปใช้ในใช้ในการหาค่าพีเอฟในระบบการรู้จำภาษาพูด โดยวัดความแม่นยำของระบบตรวจหาลักษณะเฉพาะค่าพีเอฟสำหรับแต่ละชนิด โดยการทดลองนี้วัดผลความแม่นยำของระบบตรวจหาบนข้อมูลเซตฝึกฝน และบนข้อมูลเซตทดสอบ เพื่อวัดความแม่นยำของระบบตรวจหา และแสดงว่าระบบตรวจหาค่าพีเอฟไม่มีความเฉพาะเจาะจงหรือไม่ลำเอียงที่เกิดจากข้อมูลเซตฝึกฝน จึงเปรียบเทียบประสิทธิภาพของระบบตรวจหาที่วัดบนข้อมูลเซตฝึกฝน และบนข้อมูลเซตทดสอบ ดังตารางที่ 7-1 แสดงความแม่นยำของระบบตรวจหาลักษณะเฉพาะค่าพีเอฟสำหรับแต่ละชนิด พบว่าความแม่นยำของระบบตรวจหาลักษณะเฉพาะค่าพีเอฟที่วัดบนเซตข้อมูลฝึกฝนมีค่าเฉลี่ยของเปอร์เซ็นต์ความแม่นยำ 88.17 และเมื่อวัดบนเซตทดสอบ มีค่าเปอร์เซ็นต์ความแม่นยำเฉลี่ย 87.95 โดยมีช่วงของค่าเปอร์เซ็นต์ความแม่นยำของระบบตรวจหาระหว่าง 81 ถึง 98 ระบบตรวจหาค่าพีเอฟนาสิก (Nasal) ให้ค่าเปอร์เซ็นต์ความแม่นยำ 95.8 สูงกว่าระบบตรวจหาค่าพีเอฟชนิดอื่น ซึ่งผลการวัดความแม่นยำเมื่อวัดบนข้อมูลฝึกฝน และวัดบนเซตทดสอบไม่แตกต่างกันอย่างมีนัยสำคัญ

ตารางที่ 7-1 ผลความแม่นยำของระบบตรวจหาลักษณะเฉพาะ
ค่าพีเอฟแต่ละชนิด

ค่าพีเอฟ	เปอร์เซ็นต์ความแม่นยำ		ค่าพีเอฟ	เปอร์เซ็นต์ความแม่นยำ	
	ข้อมูลฝึกฝน	ข้อมูลทดสอบ		ข้อมูลฝึกฝน	ข้อมูลทดสอบ
High	81.9	81.9	Continuant	88.8	88.7
Back	82.0	82.6	Low	88.9	88.8
Coronal	83.9	82.8	Voice	90.6	90.3
Vocalic	83.3	82.8	Round	91.7	91.7
Anterior	83.5	83.3	Strident	95.1	95.0
Tense	83.9	84.3	Nasal	96.0	95.8
Consonantal	86.0	84.8	Silence	98.8	98.5
ค่าเฉลี่ย	88.17	87.95			

สำหรับการสร้างระบบตรวจหาค่าพีเอฟที่ใช้แบบโครงข่ายประสาทเทียมหลายชั้น การกำหนดจำนวนโหนดซ่อนที่เหมาะสมมีผลต่อประสิทธิภาพของความแม่นยำของระบบตรวจหาค่าพีเอฟตารางที่ 7-2 แสดงความแม่นยำของระบบตรวจหาค่าพีเอฟเมื่อใช้จำนวนโหนดซ่อนเป็น 100 150 200 250 และ 300 พบว่าการปรับค่าพารามิเตอร์จำนวนโหนดซ่อนที่มี 250 โหนดให้ความแม่นยำของการตรวจหาค่าพีเอฟที่มีค่าสูงที่สุด ซึ่งค่าเฉลี่ยของเปอร์เซ็นต์ความแม่นยำของการตรวจหาค่าพีเอฟของระบบการตรวจหาค่าพีเอฟของทุกชนิดเท่ากับ 87.9 ซึ่งเท่ากับกับผลการทดลองที่ใช้จำนวนโหนด 300 จึงเลือกจำนวนโหนด 250 โหนดเป็นค่าพารามิเตอร์สำหรับการสร้างระบบตรวจหาค่าพีเอฟต่อไป

ตารางที่ 7-2 ผลความแม่นยำระบบตรวจหาลักษณะเฉพาะค่าพีเอฟ เมื่อใช้จำนวน
โหนดซ่อนค่าต่าง ๆ

ข้อมูลฝึกฝน	เปอร์เซ็นต์ความแม่นยำ
100	82.2
150	85.6
200	87.6
250	87.9
300	87.9

7.2 ประเมินผลระบบการรู้จำภาษาพูดตามแนวทางการเรียงหน่วยเสียง

การทดลองที่ 2 ประเมินผลระบบการรู้จำภาษาพูดแบบใช้ระบบรู้จำเสียงตามด้วยการใช้แบบจำลองเวกเตอร์สเปซ ซึ่งแสดงผลการทดลองดังนี้

7.2.1 ผลการทดลองที่ 2.1: ประเมินประสิทธิภาพของระบบการรู้จำภาษาพูดแบบพ็อดาร์วีเอสเอ็มเปรียบเทียบกับระบบการรู้จำภาษาพูดแบบแลตทิซ-วีเอสเอ็ม

การทดลองนี้เปรียบเทียบประสิทธิภาพของระบบรู้จำเสียงตามด้วยการใช้แบบจำลองเวกเตอร์สเปซ ตารางที่ 7-3 แสดงค่าความผิดพลาดการรู้จำภาษา EER ของระบบการรู้จำภาษาพูดแบบพ็อดาร์วีเอสเอ็ม และระบบการรู้จำภาษาพูดแบบแลตทิซ-วีเอสเอ็มเมื่อใช้ฟังก์ชันถ่วงน้ำหนักแบบต่าง ๆ และวิธีการทำให้เป็นบรรทัดฐานแบบต่าง ๆ พบว่าผลการรู้จำภาษาส่วนใหญ่ของระบบการรู้จำภาษาพูดแบบแลตทิซ-วีเอสเอ็มให้ผลดีกว่าผลการรู้จำภาษาจากระบบการรู้จำภาษาพูดแบบพ็อดาร์วีเอสเอ็ม นั่นคือ ค่าความผิดพลาดการรู้จำภาษา EER จากระบบการรู้จำภาษาพูดแบบพ็อดาร์วีเอสเอ็มมีค่าสูงกว่าระบบการรู้จำภาษาพูดแบบแลตทิซ-วีเอสเอ็ม ยกเว้นเมื่อใช้ฟังก์ชันถ่วงน้ำหนักแบบค่าความถี่โทม และค่าสถิติโคสแควร์ ที่ให้ค่าความผิดพลาดการรู้จำภาษาของระบบการรู้จำภาษาพูดแบบแลตทิซ-วีเอสเอ็มไม่แตกต่างกับระบบการรู้จำภาษาพูดแบบพ็อดาร์วีเอสเอ็ม

ระบบการรู้จำภาษาพูดแบบแลตทิซ-วีเอสเอ็มมีผลการรู้จำภาษาดีกว่าระบบการรู้จำภาษาพูดแบบพ็อดาร์วีเอสเอ็ม เนื่องจากระบบการรู้จำภาษาพูดแบบแลตทิซ-วีเอสเอ็มใช้ผลลัพธ์จากการรู้จำเสียงเป็นหน่วยเสียงแลตทิซ ขณะที่ระบบการรู้จำภาษาพูดแบบพ็อดาร์วีเอสเอ็มใช้ผลลัพธ์ที่ได้จากการรู้จำเสียงเป็นหน่วยเสียงเพื่อเป็นอินพุตของแบบจำลองเวกเตอร์สเปซ ซึ่งการใช้หน่วยเสียงแลตทิซสามารถลดข้อผิดพลาดของขั้นตอนการตัดแต่งกิ่งในการรู้จำเสียง และรักษาโอกาสที่จะได้ผลลัพธ์ของลำดับหน่วยเสียงที่เป็นคำตอบที่ถูกต้องของการรู้จำเสียงพูด โดยหน่วยเสียงแลตทิซรวบรวมชุดลำดับหน่วยเสียงที่เป็นผลลัพธ์ที่เป็นไปได้ทุกลำดับ หรือการสร้างแบบจำลองภาษาจากผลลัพธ์การรู้จำเสียงแบบลำดับหน่วยเสียงทำให้ได้แบบจำลองภาษาที่ไม่ถูกต้องหรือไม่ครอบคลุมแบบจำลองภาษาที่เกิดขึ้น ส่งผลให้เกิดข้อผิดพลาดต่อการสร้างตัวจำแนกแต่ละภาษาได้ นอกจากนี้ข้อดีของการใช้หน่วยเสียงแลตทิซแทนหน่วยเสียงในระบบการรู้จำภาษาพูด คือ ข้อมูลของลำดับหน่วยเสียงที่มากขึ้นในหน่วยเสียงแลตทิซ สามารถสร้างรูปแบบแบบจำลองภาษาได้มากขึ้น ทำให้เมื่อพิจารณาความถี่การปรากฏของผลลัพธ์นี้ทำให้ได้ข้อมูลในการบอกความแตกต่างระหว่างภาษาเพิ่มขึ้น

อย่างไรก็ตามเมื่อเปรียบเทียบประสิทธิภาพระหว่างระบบการรู้จำภาษาพูดแบบพ็อดาร์วีเอสเอ็ม กับระบบการรู้จำภาษาพูดแบบแลตทิซ-วีเอสเอ็ม จากตารางที่ 7-4พบว่าระบบการรู้จำภาษาพูดแบบแลตทิซ-วีเอสเอ็มให้ค่าความผิดพลาดการรู้จำ EER ลดลงประมาณ 44.46% เมื่อเปรียบเทียบกับตามระบบอ้างอิงที่ 1.2 และระบบการรู้จำภาษาพูดแบบแลตทิซ-วีเอสเอ็มให้ค่าความผิดพลาดการรู้จำ EER ลดลงประมาณ 45.93% เมื่อใช้ระบบการรู้จำภาษาพูดแบบใช้ฟังก์ชันถ่วงน้ำหนักค่าล็อกการิทึมค่าความถี่โทมและความซ้ำซ้อนของโทม

ตารางที่ 7-3 ค่าความผิดพลาดการรู้จำภาษา EER ของระบบการรู้จำภาษาพูดแบบพีอาร์วีเอสเอ็ม และระบบการรู้จำภาษาพูดแบบแลตทิซ-วีเอสเอ็ม เมื่อใช้ฟังก์ชันถ่วงน้ำหนักแบบต่าง ๆ และวิธีการทำให้เป็นบรรทัดฐานแบบต่าง ๆ

ลำดับที่	ฟังก์ชันถ่วงน้ำหนัก	แลตทิซ-วีเอสเอ็ม		พีอาร์วีเอสเอ็ม	
		วิธีทำบรรทัดฐาน		วิธีทำบรรทัดฐาน	
		การรวม	ยูคลิเดียน	การรวม	ยูคลิเดียน
1	tf	5.83	5.98	5.50	5.32
2	logtf	3.12	3.27	5.30	5.35
3	ITF	2.80	2.77	5.32	5.28
4	idf	3.04	2.94	4.76	4.76
5	tr	3.14	3.05	5.85	5.68
6	chi	17.69	18.53	15.23	15.28
7	rd	2.85	2.69	4.81	4.71

7.2.2 ผลการทดลองที่ 2.2: เปรียบเทียบประสิทธิภาพของฟังก์ชันถ่วงน้ำหนักแบบต่าง ๆ

ผลการทดลองแสดงผลการรู้จำภาษาเมื่อใช้ฟังก์ชันถ่วงน้ำหนักแบบต่าง ๆ และวิธีการทำให้เป็นบรรทัดฐานแบบต่าง ๆ แสดงดังตารางที่ 7-3 พบว่า เมื่อเปรียบเทียบฟังก์ชันค่าความถี่เทอม (tf) กับฟังก์ชันถ่วงน้ำหนักที่มีความหลากหลายของความถี่เทอมแบบต่าง ๆ เช่น ค่าความถี่แบบผกผัน (ITF) และ ล็อกการิทึมของความถี่เทอม (logtf) ที่แสดงในข้อ 1 ถึงข้อ 3 ของตารางที่ 7-3 พบว่า ฟังก์ชันถ่วงน้ำหนักแบบค่าความถี่แบบผกผัน (ITF) และล็อกการิทึมของความถี่เทอม ให้ผลการรู้จำดีกว่าฟังก์ชันถ่วงน้ำหนักแบบค่าความถี่เทอม

ผลการทดลองของฟังก์ชันถ่วงน้ำหนักที่เป็นตัวแทนของน้ำหนักแบบครอบคลุม พบว่าฟังก์ชันความซ้ำซ้อนของเทอม (rd) ดังข้อ 7 ให้ประสิทธิภาพดีที่สุด เนื่องจากค่าถ่วงน้ำหนักความซ้ำซ้อนของเทอมแสดงความเจาะจงของเทอมที่อยู่บนเสียงพูดในภาษาที่พิจารณาเทียบกับเสียงที่อยู่ในฐานข้อมูลเสียงทั้งหมด ซึ่งค่าถ่วงน้ำหนักความซ้ำซ้อนดีกว่าฟังก์ชันความถี่เอกสารผกผันดังข้อ 4 เป็นอีกหนึ่งฟังก์ชันที่ใช้ตัวแทนของน้ำหนักแบบครอบคลุม ขณะที่ค่าสถิติไคสแควร์ (chi) สามารถบอกความแตกต่างระหว่างภาษาได้ไม่ดี ซึ่งค่าสถิติไคสแควร์ให้ค่าความผิดพลาดการรู้จำแตกต่างจากฟังก์ชันถ่วงน้ำหนักแบบอื่นและมีความผิดพลาดสูงที่สุด

เมื่อพิจารณาผลการรู้จำภาษาจากระบบการรู้จำภาษาพูดแบบพีอาร์วีเอสเอ็ม พบว่าฟังก์ชันความซ้ำซ้อนของเทอมให้ค่าความผิดพลาดต่ำที่สุด และผลการรู้จำเมื่อใช้ฟังก์ชันถ่วงน้ำหนักแบบต่าง ๆ ไม่แตกต่างกันมากนักซึ่งค่าความผิดพลาดอยู่ในช่วง 4.71 ถึง 5.85 ยกเว้นผลการรู้จำจากการใช้

ฟังก์ชันค่าสถิติไคสแควร์ (chi) ซึ่งค่าแตกต่างจากฟังก์ชันถ่วงน้ำหนักแบบอื่น และให้ค่าความผิดพลาดการรู้จำ EER เป็น 15.23 ซึ่งเป็นค่าที่สูงที่สุด เพราะค่าสถิติไคสแควร์เบี่ยงเบนผิดไปซึ่งเป็นผลมาจากความถี่การปรากฏของเทอมในภาษาที่ไม่ใช่ภาษาเป้าหมายมีค่าสูงมาก ทำให้มีความแตกต่างกันมากระหว่างความถี่การปรากฏของเทอมในภาษาที่ไม่ใช่ภาษาเป้าหมายและในภาษาเป้าหมาย ขณะที่ค่าความเกี่ยวข้องกันของเทอม (tr) ให้ค่าความผิดพลาดการรู้จำ EER เป็น 5.85 ซึ่งค่าความผิดพลาดสูงเป็นอันดับสองรองจากค่าสถิติไคสแควร์

เมื่อพิจารณาผลการรู้จำภาษาของระบบการรู้จำภาษาพูดแบบแลตทิซ-วีเอสเอ็ม พบว่า ผลการรู้จำจากการใช้ฟังก์ชันค่าสถิติไคสแควร์ให้ค่าความผิดพลาดการรู้จำ EER เป็น 17.69 ซึ่งเป็นค่าความผิดพลาดที่สูงที่สุด ซึ่งมีเหตุผลเดียวกับระบบการรู้จำภาษาพูดแบบพ็อดาร์วีเอสเอ็มที่เกิดจากความแตกต่างกันระหว่างความถี่การปรากฏของเทอมในภาษาเป้าหมายและในภาษาที่ไม่ใช่ภาษาเป้าหมายที่ทำให้การคำนวณค่าสถิติไคสแควร์เบี่ยงเบนผิดไป และค่าความผิดพลาดนี้สูงกว่าค่าสถิติไคสแควร์ที่ได้จากผลการรู้จำภาษาจากระบบการรู้จำภาษาพูดแบบพ็อดาร์วีเอสเอ็ม และความถี่ผลการรู้จำจากการใช้ฟังก์ชันค่าความถี่เทอม ให้ค่าความผิดพลาดการรู้จำ EER เป็น 5.83 ซึ่งสูงรองลงมาเป็นอันดับสอง สำหรับผลการรู้จำภาษาของระบบการรู้จำภาษาพูดแบบแลตทิซ-วีเอสเอ็มเมื่อใช้ฟังก์ชันถ่วงน้ำหนักแบบอื่น ๆ ให้ค่าความผิดพลาดไม่แตกต่างกันนักซึ่งอยู่ในช่วงระหว่าง EER เป็น 2.69 ถึง 3.27 ซึ่งฟังก์ชันความซ้ำซ้อนของเทอมให้ประสิทธิภาพดีที่สุด

ตารางที่ 7-4 แสดงค่าความผิดพลาดของการรู้จำภาษาจากระบบการรู้จำภาษาพูดแบบพ็อดาร์วีเอสเอ็ม และระบบการรู้จำภาษาพูดแบบแลตทิซ-วีเอสเอ็มเมื่อใช้การประกอบกันของฟังก์ชันถ่วงน้ำหนักแบบต่าง ๆ และวิธีการทำให้เป็นบรรทัดฐานแบบต่าง ๆ พบว่า ผลการทดลองของการประกอบกันของฟังก์ชันถ่วงน้ำหนักระหว่างฟังก์ชันถ่วงน้ำหนักแบบท้องถิ่น และค่าถ่วงน้ำหนักแบบครอบคลุม แสดงดังข้อ 8, 9, 10, 11, 16 และ 17 ให้ผลการรู้จำภาษาดีกว่าการทดลองของการประกอบกันของฟังก์ชันถ่วงน้ำหนักระหว่างค่าถ่วงน้ำหนักแบบท้องถิ่น เช่น การประกอบกันของค่าถ่วงน้ำหนักค่าความถี่เทอมและค่าถ่วงน้ำหนักที่พิจารณาเทอมเปรียบเทียบภายในกลุ่มเดียวกัน แทนด้วย *idf_cate* ดังข้อ 13 หรือการประกอบกันของค่าถ่วงน้ำหนักค่าความถี่เทอม และค่าความเกี่ยวข้องกันของเทอมที่ใช้แสดงการถ่วงน้ำหนักระหว่างกลุ่มดังข้อ 14

จากตารางที่ 7-3 ที่แสดงว่าฟังก์ชันความซ้ำซ้อนของเทอมให้ผลการรู้จำภาษาที่ดีกว่าฟังก์ชันค่าถ่วงน้ำหนักแบบครอบคลุมประเภทอื่น สอดคล้องกับผลการรู้จำภาษาเมื่อใช้การประกอบกันของฟังก์ชันถ่วงน้ำหนักดังตารางที่ 7-4 คือผลการทดลองของการประกอบกันของฟังก์ชันถ่วงน้ำหนักค่าถ่วงน้ำหนักความซ้ำซ้อนซึ่งใช้เป็นฟังก์ชันค่าถ่วงน้ำหนักแบบครอบคลุม ร่วมกับค่าฟังก์ชันค่าถ่วงน้ำหนักแบบท้องถิ่นต่าง ๆ ให้ค่าความผิดพลาดการรู้จำภาษาต่ำที่สุด เช่น การประกอบกันของฟังก์ชันค่าความถี่เทอมและค่าความซ้ำซ้อนของเทอม ซึ่งแทนด้วย *tf.rd* ดังข้อ 16 และการประกอบกันของฟังก์ชันถ่วงน้ำหนักค่าถ่วงน้ำหนักค่าความถี่เทอมและความซ้ำซ้อนของเทอม ซึ่งแทนด้วย *logtf.rd* ดังข้อ 17 ให้ค่าความผิดพลาด EER 2.43 และ 2.46 ตามลำดับ ซึ่งเป็นค่าความผิดพลาดที่ต่ำที่สุด

จากผลการทดลองสรุปได้ว่าฟังก์ชันถ่วงน้ำหนักมีผลทำให้ผลการรู้จำภาษาตามแนวการใช้เกณฑ์การเรียงหน่วยเสียงดีขึ้น และผลการรู้จำภาษาเมื่อใช้ฟังก์ชันความซ้ำซ้อนของเทอม เป็นค่าถ่วงน้ำหนักค่าเทอมในการให้คะแนนความสำคัญของเทอมเพื่อบอกความแตกต่างระหว่างภาษาได้ดีกว่าฟังก์ชันถ่วงน้ำหนักแบบอื่น ๆ นอกจากนี้เมื่อเปรียบเทียบผลการรู้จำภาษาเมื่อใช้การประกอบกันของฟังก์ชันค่าความถี่เทอมและค่าความซ้ำซ้อนของเทอม ให้ค่าความผิดพลาดการรู้จำ EER ต่ำกว่าระบบอ้างอิงที่ 1.2 ซึ่งเป็นระบบการรู้จำภาษาพูดตามแนวทางเกณฑ์การเรียงหน่วยเสียงเมื่อใช้ฟังก์ชันถ่วงน้ำหนักที่มาจากการประกอบกันของฟังก์ชันค่าความถี่เทอมและค่าความถี่เอกสารพจนานุกรม ทั้งในระบบการรู้จำภาษาพูดแบบพ็อดาร์วีเอสเอ็ม และระบบการรู้จำภาษาพูดแบบแลตทิซ-วีเอสเอ็ม เปอร์เซ็นต์ความผิดพลาดการรู้จำภาษา EER แบบสัมพัทธ์ของระบบการรู้จำภาษาพูดแบบพ็อดาร์วีเอสเอ็ม และระบบการรู้จำภาษาพูดแบบแลตทิซ-วีเอสเอ็มที่ใช้ฟังก์ชันถ่วงน้ำหนักค่าถ่วงน้ำหนักค่าความถี่เทอมและความซ้ำซ้อนของเทอม ลดลงเมื่อเปรียบเทียบกับระบบอ้างอิงที่ 1.2 มีค่าเป็น 1.30% และ 3.90% ตามลำดับ

ตารางที่ 7-4 ค่าความผิดพลาดการรู้จำภาษา EER ของระบบการรู้จำภาษาพูดแบบพ็อดาร์วีเอสเอ็ม และระบบการรู้จำภาษาพูดแบบแลตทิซ-วีเอสเอ็ม เมื่อใช้การประกอบกันของฟังก์ชันถ่วงน้ำหนักแบบต่าง ๆ และวิธีการทำให้เป็นบรรทัดฐานแบบต่าง ๆ

ลำดับที่	การประกอบกันของฟังก์ชัน	แลตทิซ-วีเอสเอ็ม		พ็อดาร์วีเอสเอ็ม	
		วิธีทำบรรทัดฐาน		วิธีทำบรรทัดฐาน	
		การรวม	ยูคลิเดียน	การรวม	ยูคลิเดียน
8	tf.idf. (ไม่ทำ Norm)	3.11		4.42	
9	tf.idf (ระบบอ้างอิงที่ 1.2)	2.56	2.49	4.61	4.50
10	logtf.idf	2.68	2.69	4.39	4.52
11	ITF.idf	2.62	2.73	4.38	4.70
12	logtf.ITF	4.31	4.34	10.80	11.68
13	tf.idf_cate	2.59	2.59	4.39	4.41
14	tf.tr	8.43	9.46	10.97	11.65
15	tf.chi	7.93	8.93	13.71	12.04
16	tf.rd	2.49	2.46	4.45	4.64
17	logtf.rd	2.46	2.43	4.55	4.68

7.2.3 ผลการทดลองที่ 2.3: เปรียบเทียบวิธีการทำให้เป็นบรรทัดฐานที่ต่างกัน

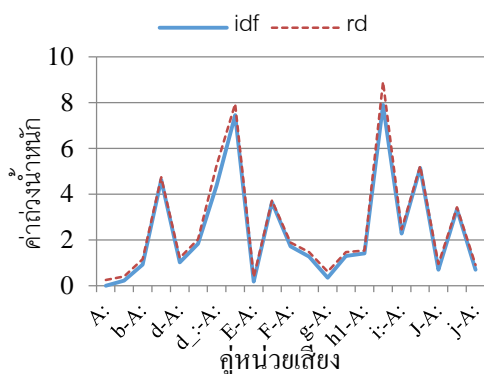
ประสิทธิภาพของระบบการรู้จำภาษาพูดแบบพ็อดาร์วีเอสเอ็ม และระบบการรู้จำภาษาพูดแบบแลตทิซ-วีเอสเอ็มเมื่อใช้ฟังก์ชันถ่วงน้ำหนักแบบต่าง ๆ และเมื่อใช้การประกอบกันของฟังก์ชันถ่วงน้ำหนักแบบต่าง ๆ โดยใช้ร่วมกับวิธีการทำให้เป็นบรรทัดฐานที่ต่างกัน ดังตารางที่ 7-3 และตารางที่ 7-4 พบว่าค่าความผิดพลาดการรู้จำภาษาเมื่อใช้วิธีการทำให้เป็นบรรทัดฐานด้วยวิธีการรวมและวิธีวัดระยะห่างแบบยูคลิเดียนไม่มีความแตกต่างกันอย่างมีนัยสำคัญ และเมื่อค่าความผิดพลาดการรู้จำภาษาของระบบรู้จำภาษาที่ใช้วิธีการทำให้เป็นบรรทัดฐานมีค่าความผิดพลาดการรู้จำภาษา EER ต่ำกว่าการไม่ใช้การทำให้เป็นบรรทัดฐาน ซึ่งแสดงถึงผลการทดลองเมื่อเปรียบเทียบระหว่างข้อ 8 ซึ่งเป็นการรู้จำภาษาเมื่อไม่ใช้การทำให้เป็นบรรทัดฐาน และข้อ 9 ซึ่งเป็นการรู้จำภาษาเมื่อใช้วิธีการทำให้เป็นบรรทัดฐานแบบรวม

เมื่อพิจารณาผลการทดลอง โดยใช้ฟังก์ชันถ่วงน้ำหนัก ร่วมกับการทำให้เป็นบรรทัดฐานแบบต่าง ๆ พบว่า ประสิทธิภาพของระบบการรู้จำภาษาพูดแบบพ็อดาร์วีเอสเอ็มเมื่อใช้การทำให้เป็นบรรทัดฐานด้วยวิธีวัดระยะห่างแบบยูคลิเดียน และระบบการรู้จำภาษาพูดแบบแลตทิซ-วีเอสเอ็ม มีผลการทดลองเป็นแนวโน้มทิศทางเดียวกัน คือ ระบบการรู้จำภาษาพูดที่ใช้วิธีการทำให้เป็นบรรทัดฐานด้วยวิธีวัดระยะห่างแบบยูคลิเดียน ให้ค่าความผิดพลาดการรู้จำภาษาต่ำกว่าการทำให้เป็นบรรทัดฐานด้วยวิธีการรวม

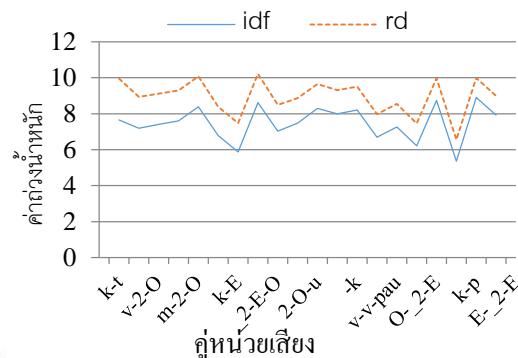
อย่างไรก็ตามเมื่อเปรียบเทียบผลการทดลองของวิธีการทำให้เป็นบรรทัดฐาน พบว่าผลการทดลองส่วนใหญ่สำหรับระบบการรู้จำภาษาพูดแบบแลตทิซ-วีเอสเอ็มเมื่อใช้ฟังก์ชันถ่วงน้ำหนัก ร่วมกับการทำให้เป็นบรรทัดฐานด้วยวิธีวัดระยะห่างแบบยูคลิเดียน มีผลทำให้ค่าความผิดพลาดการรู้จำภาษาต่ำกว่าการทำให้เป็นบรรทัดฐานด้วยวิธีการรวม มากกว่า ผลการทดลองของการทำให้เป็นบรรทัดฐานในระบบการรู้จำภาษาพูดแบบพ็อดาร์วีเอสเอ็ม จากผลการทดลองสรุปได้ว่าการทำให้เป็นบรรทัดฐานด้วยวิธีวัดระยะห่างแบบยูคลิเดียนเหมาะสมสำหรับระบบการรู้จำภาษาพูดแบบแลตทิซวีเอสเอ็ม ซึ่งใช้ค่าเทอมจากหน่วยเสียงแลตทิซ เนื่องจากค่าน้ำหนักที่ได้จากการทำบรรทัดฐานด้วยวิธีวัดระยะห่างแบบยูคลิเดียนมีค่าน้อยกว่า เมื่อนำไปใช้ร่วมกับค่าน้ำหนักของเทอมจึงทำให้แสดงค่าเทอมเด่นชัดขึ้น ซึ่งเพิ่มน้ำหนักให้กับค่าเทอมที่ได้จากหน่วยเสียงแลตทิซได้

7.2.4 บทอภิปรายการทดลองฟังก์ชันการถ่วงน้ำหนักค่าเทอม

จากผลการทดลองที่ 2.2 ที่ศึกษาผลการรู้จำภาษาเมื่อใช้ฟังก์ชันถ่วงน้ำหนักแบบต่าง ๆ พบว่าค่าถ่วงน้ำหนักความซ้ำซ้อนของเทอมให้ผลการรู้จำที่ดีที่สุด เพื่อเปรียบเทียบความสามารถของการใช้ถ่วงน้ำหนักแบบโกลบอล ผู้วิจัยพิจารณาผลกระทบของคุณสมบัติของการถ่วงน้ำหนักแบบโกลบอลโดยวิเคราะห์คะแนนของค่าถ่วงน้ำหนักระหว่างแบบค่าความถี่เอกสารผกผัน และค่าความซ้ำซ้อนของเทอม เพื่อเปรียบเทียบค่าที่ปรากฏในแต่ละหน่วยเสียง



(ก)



(ข)

ภาพที่ 7-1 (ก) ตัวอย่างหน่วยเสียง 20 อันดับแรกของหน่วยเสียงที่มีความถี่ของการปรากฏบ่อยที่สุดในฐานข้อมูลเสียง ซึ่งเปรียบเทียบน้ำหนักจากฟังก์ชันการถ่วงน้ำหนัก ค่าความถี่เอกสารผกผัน และค่าความซ้ำซ้อนของเทอม และ (ข) หน่วยเสียง 20 อันดับแรกที่มีความแตกต่างกันมากที่สุดระหว่างการถ่วงน้ำหนัก ค่าความถี่เอกสารผกผัน และค่าความซ้ำซ้อนของเทอม

จากการวิเคราะห์หน่วยเสียงในภาษาอังกฤษอินิยามไว้ตาม HU-BUT ที่ปรากฏในฐานข้อมูลเสียงในการทดลองนี้ พบว่าหน่วยเสียงที่มีความถี่ของการปรากฏบ่อยที่สุดในแต่ละภาษามีแนวโน้มเดียวกัน มีหน่วยเสียง'A:' เป็นหน่วยเสียงเดี่ยว และไบแกรมของหน่วยเสียง'A:' มีความถี่ของการปรากฏบ่อยที่สุดในฐานข้อมูลเสียง ซึ่งตัวอย่างหน่วยเสียง 20 อันดับแรกของหน่วยเสียงที่มีความถี่ของการปรากฏมากที่สุดในฐานข้อมูลเสียง แสดงดังภาพที่ 7-1 (ก) แผนภาพด้านซ้าย โดยได้แสดงเปรียบเทียบค่าน้ำหนักจากฟังก์ชันการถ่วงน้ำหนักค่าความถี่เอกสารผกผันแทนด้วย idf และค่าความซ้ำซ้อนของเทอมแทนด้วย rd จากการสังเกตพบว่าตัวอย่างหน่วยเสียง 20 อันดับแรกที่มีแนวโน้มของค่าน้ำหนักแบบค่าความถี่เอกสารผกผัน และค่าความซ้ำซ้อนของเทอมไม่แตกต่างกันนั้น ฟังก์ชันการถ่วงน้ำหนักทั้งสองแบบมีค่าความต่างน้อยมาก ซึ่งค่าน้ำหนักจากฟังก์ชันการถ่วงน้ำหนักค่าความซ้ำซ้อนของเทอมมีค่ามากกว่าจากฟังก์ชันการถ่วงน้ำหนักค่าความถี่เอกสารผกผัน อย่างไรก็ตามค่าน้ำหนักสำหรับแต่ละไบแกรมมีค่าแตกต่างกันเล็กน้อย เช่น ไบแกรม'd_-A:' และ'i-A:' เป็นต้น ขณะที่ภาพที่ 7-1 (ข) หน้าต่างทางขวา แสดงตัวอย่างหน่วยเสียง 20 อันดับแรกที่มีความแตกต่างกันมากที่สุดระหว่างค่าน้ำหนักจากฟังก์ชันการถ่วงน้ำหนักค่าความถี่เอกสารผกผัน และค่าความซ้ำซ้อนของเทอม เช่น ไบแกรม'k-t' เป็นต้น ซึ่งน้ำหนักจากค่าน้ำหนักจากค่าความซ้ำซ้อนของเทอม มีค่ามากกว่าค่าน้ำหนักจากค่าความถี่เอกสารผกผันอย่างชัดเจน จากการวิเคราะห์พบว่า ค่าความซ้ำซ้อนของเทอมมีค่ามากกว่าค่าความถี่เอกสารผกผัน สรุปได้ว่าค่าน้ำหนักค่าความซ้ำซ้อนของเทอมเป็นค่าที่เหมาะสมเพื่อถ่วงน้ำหนักในการจำแนกไบแกรมเหล่านี้ได้ดีกว่าค่าความถี่เอกสารผกผัน เพราะมีค่ามากกว่าและสามารถเพิ่มน้ำหนักของเทอมที่พิจารณาให้มีความสำคัญในการจำแนกภาษาได้ดีกว่า ค่าน้ำหนักที่มากกว่าของค่าความซ้ำซ้อนของเทอมนี้ทำให้เทอมที่มีความสำคัญยังคงปรากฏในเสียงพูด และเทอมที่ไม่มีความสำคัญถูกตัดทิ้งหรือลดคะแนนไปจากเสียงพูดซึ่งส่งผลให้การจำแนกมีประสิทธิภาพที่ดีขึ้น

7.2.5 บทสรุปการรู้จำภาษาตามแนวทางการเรียงหน่วยเสียง

การศึกษาฟังก์ชันถ่วงน้ำหนักที่ใช้ในระบบการรู้จำภาษาพูดตามแนวทางกฎการเรียงหน่วยเสียงที่ใช้ลำดับหน่วยเสียง และที่ใช้หน่วยเสียงแลตทิซ จากผลการทดลองพบว่าฟังก์ชันค่าถ่วงน้ำหนักแบบค่าความซ้ำซ้อนของเทอม ทำให้ความสามารถในการจำแนกภาษาดีกว่าฟังก์ชันค่าถ่วงน้ำหนักแบบอื่น เนื่องจากสามารถลดความซ้ำซ้อนของความสัมพันธ์การปรากฏของหน่วยเสียงหรือเทอมที่ปรากฏในภาษาต่าง ๆ ได้ ดังนั้นการประกอบกันของฟังก์ชันถ่วงน้ำหนักระหว่างค่าน้ำหนักแบบท้องถิ่นใด ๆ ร่วมกับค่าความซ้ำซ้อนของเทอมที่ใช้เป็นตัวแทนของฟังก์ชันถ่วงน้ำหนักจากค่าแบบครอบคลุมทำให้ผลการจำแนกภาษาดีขึ้น และการประกอบกันของฟังก์ชันถ่วงน้ำหนักค่าล็อกการตีพิมพ์ของค่าความถี่เทอมร่วมกับค่าความซ้ำซ้อนของเทอมเป็นฟังก์ชันถ่วงน้ำหนักที่ทำให้ตัวจำแนกมีประสิทธิภาพที่ดีที่สุด

7.2.6 ผลการทดลองที่ 2.4: วัดประสิทธิภาพของระบบการรู้จำภาษาพูดแบบพ็อดาร์วีเอสเอ็ม เปรียบเทียบกับระบบการรู้จำภาษาพูดแบบแลตทิซ-วีเอสเอ็ม โดยวัดผลการรู้จำภาษาบน 8 ภาษา

จากผลการทดลองที่ 2.2 และการทดลองที่ 2.3 พบว่าระบบการรู้จำภาษาพูดตามแนวทางการเรียงหน่วยเสียงของระบบการรู้จำภาษาพูดแบบพ็อดาร์วีเอสเอ็ม และระบบการรู้จำภาษาพูดแบบแลตทิซ-วีเอสเอ็ม ที่ใช้ฟังก์ชันค่าถ่วงน้ำหนักแบบค่าความซ้ำซ้อนของเทอม และใช้วิธีการทำให้เป็นบรรทัดฐานแบบวัดระยะห่างแบบยูคลิเดียนให้ประสิทธิภาพการรู้จำที่ดีที่สุด จึงนำต้นแบบของฟังก์ชันการจำแนกดังกล่าวมาสร้างฟังก์ชันการจำแนกสำหรับวัดผลการรู้จำภาษา 8 ภาษา โดยใช้ฐานข้อมูลเสียง NIST LRE07 เซตข้อมูลพัฒนา และทำตรวจสอบไขว้โดยแบ่งข้อมูลเป็น 5 ส่วน

ภาคผนวก ก-1 แสดงแผนภาพกล่องของคะแนนที่ได้จากคะแนนการตัดสินใจของระบบการรู้จำภาษาแบบแลตทิซ-วีเอสเอ็ม เมื่อใช้ข้อมูลชุดฝึกฝนจากแต่ละภาษา พบว่ากล่องของคะแนนที่ได้จากการจำแนกด้วยซัพพอร์ตเวกเตอร์แมชชีนจากข้อมูลฝึกฝนสำหรับภาษาเป้าหมาย มีค่าเฉลี่ยสูงกว่า 0 และไม่ซ้อนทับกับกล่องของคะแนนที่ได้การจำแนกด้วยซัพพอร์ตเวกเตอร์แมชชีนเมื่อใช้ข้อมูลฝึกฝนของภาษาอื่นที่มีค่าเฉลี่ยต่ำกว่า -1

ตารางที่ 7-5 แสดงผลการรู้จำภาษาตามแนวทางการเรียงหน่วยเสียงของระบบการรู้จำภาษาพูดแบบพ็อดาร์วีเอสเอ็ม และระบบการรู้จำภาษาพูดแบบแลตทิซ-วีเอสเอ็ม โดยใช้ฟังก์ชันค่าถ่วงน้ำหนักแบบค่าความซ้ำซ้อนของเทอม และใช้วิธีการทำให้เป็นบรรทัดฐานแบบยูคลิเดียน ซึ่งให้ประสิทธิภาพการรู้จำที่ดีที่สุด พบว่าระบบการรู้จำภาษาพูดแบบแลตทิซ-วีเอสเอ็มให้ค่าความผิดพลาดการรู้จำภาษา EER มีค่า 5.70 ซึ่งต่ำกว่าระบบการรู้จำภาษาพูดแบบพ็อดาร์วีเอสเอ็มที่มีค่า 8.61 อย่างไรก็ตามค่าความผิดพลาดการรู้จำภาษา EER ที่ได้จากการทดลองที่ 2.4 มีค่าสูงกว่าผลการทดลองที่ 2.3 ดังตารางที่ 7-4 เนื่องจากระบบการรู้จำภาษาที่พัฒนา และทดสอบในการทดลองนี้ใช้ฐานข้อมูลเสียงที่มีจำนวนข้อมูลเสียงในแต่ละภาษาน้อยกว่า ข้อมูลเสียงที่ใช้ในการทดลองที่ 2.3 ทำให้ประสิทธิภาพของการรู้จำภาษา ที่ใช้ตัวจำแนกภาษาด้วยซัพพอร์ตเวกเตอร์แมชชีนมีความสามารถในการจำแนกตัวอย่างระหว่างภาษาต่อยกกว่า ส่งผลให้มีค่าความแม่นยำในการตัดสินใจจำแนกภาษาน้อยกว่า

ตารางที่ 7-5 ค่าความผิดพลาดการรู้จำภาษา EER ของระบบการรู้จำภาษาพูดแบบพ็อาร์วีเอสเอ็ม และระบบการรู้จำภาษาพูดแบบแลตทิซ-วีเอสเอ็ม สำหรับข้อมูล NIST LRE07 เซตข้อมูลพัฒนา

ระบบการรู้จำภาษาพูด	ค่าความผิดพลาดการรู้จำภาษา EER
พ็อาร์วีเอสเอ็ม	8.61
แลตทิซ-วีเอสเอ็ม	5.70

7.2.7 ผลการทดลองที่ 2.5: วัดประสิทธิภาพของระบบการรู้จำภาษาพูดแบบพ็อาร์วีเอสเอ็ม เปรียบเทียบกับระบบการรู้จำภาษาพูดแบบแลตทิซ-วีเอสเอ็ม โดยวัดผลการรู้จำภาษาบน 8 ภาษานับฐานข้อมูลเสียง NIST LRE07 เซตข้อมูลทดสอบ

การทดลองนี้ทำเพื่อประเมินผลการรู้จำภาษาพูดกับฐานข้อมูลชุดประเมินผล จึงใช้ระบบการรู้จำภาษาพูดตามแนวทางการเรียงหน่วยเสียงของระบบการรู้จำภาษาพูดแบบพ็อาร์วีเอสเอ็ม และระบบการรู้จำภาษาพูดแบบแลตทิซ-วีเอสเอ็มที่ได้จากผลการทดลองที่ 2.4 ซึ่งพัฒนาสำหรับการจำแนกภาษา 8 ภาษา โดยพัฒนาจาก NIST LRE07 เซตข้อมูลพัฒนา แล้วนำมาทดสอบการจำแนกภาษานับฐานข้อมูลเสียง NIST LRE07 เซตข้อมูลทดสอบที่เลือกมา 8 ภาษา ตามหัวข้อ 3.5.4.2 พบว่าระบบการรู้จำภาษาพูดแบบแลตทิซ-วีเอสเอ็มให้ค่า C_{avg} มีค่า 0.558 ซึ่งต่ำกว่าระบบการรู้จำภาษาพูดแบบพ็อาร์วีเอสเอ็ม ดังแสดงในตารางที่ 7-6

ตารางที่ 7-6 ค่า C_{avg} ของระบบการรู้จำภาษาพูดแบบพ็อาร์วีเอสเอ็ม และระบบการรู้จำภาษาพูดแบบแลตทิซ-วีเอสเอ็ม บนฐานข้อมูลเซตทดสอบ 8 ภาษา

ระบบการรู้จำภาษาพูด	ค่า $100 \times C_{avg}$
พ็อาร์วีเอสเอ็ม	9.52
แลตทิซ-วีเอสเอ็ม	5.58

7.3 ประเมินประสิทธิภาพของระบบการรู้จำภาษาพูดโดยใช้การจำแนกด้วยแบบจำลองลาเทนท์ไดนามิคคอนดิชันนอลแรนดอมฟิลด์ร่วมกับค่าพีเอฟ

7.3.1 การทดลองที่ 3.1 ที่วัดประสิทธิภาพของระบบการรู้จำภาษาพูดเมื่อใช้การจำแนกด้วยแบบจำลองลาเทนท์ไดนามิคคอนดิชันนอลแรนดอมฟิลด์ร่วมกับการใช้ค่าลักษณะทางเสียงต่าง ๆ

ผลการทดลองที่ 3.1 ที่วัดประสิทธิภาพของระบบการรู้จำภาษาพูดเมื่อใช้การจำแนกด้วยแบบจำลองลาเทนท์ไดนามิคคอนดิชันนอลแรนดอมฟิลด์ร่วมกับการใช้ค่าลักษณะทางเสียงต่าง ๆ ได้แก่ ค่าพีเอฟ, ค่าเอสดีซีพีเอฟ, ค่าสัมประสิทธิ์ MFCC, ค่าสัมประสิทธิ์พีแอลพี, ค่าเอสดีซีพีแอลพี,

และการประกอบร่วมกันของค่าเอสดีซีพีแอลพีและค่าสัมประสิทธิ์พีแอลพี เป็นต้น ตารางที่ 7-7 แสดงค่าเฉลี่ยเปอร์เซ็นต์ความผิดพลาดของการจำแนกภาษาจากตัวจำแนก 28 คู่ภาษา เมื่อใช้ค่าลักษณะสำคัญที่แตกต่างกัน พบว่าค่าเฉลี่ยเปอร์เซ็นต์ความผิดพลาดเมื่อใช้ตัวจำแนกที่มีค่าลักษณะสำคัญพีเอฟมีค่า 17.46 ซึ่งมีค่าต่ำกว่าตัวจำแนกที่ใช้ค่าลักษณะสำคัญทางเสียงชนิดอื่น ค่าเฉลี่ยเปอร์เซ็นต์ความผิดพลาดของการจำแนกเมื่อใช้ค่าลักษณะสำคัญที่แตกต่างกัน แสดงเรียงลำดับจากน้อยไปหามาก จะได้ $PF < PLP < SDCPF < SDC+PLP < SDCPLP < MFCC$ ซึ่งผลการทดลองพบว่าเมื่อใช้ค่าการเปลี่ยนแปลงแบบเลื่อนของสัมประสิทธิ์ หรือค่านวนค่าเอสดีซี (SDC) เพื่อค่าการเปลี่ยนแปลงของค่าพีเอฟ (ค่า SDCPF) และค่าการเปลี่ยนแปลงของค่าพีแอลพี (ค่า SDCPLP) ให้ความผิดพลาดของการจำแนกเพิ่มขึ้น 17.13% และ 16.48% เมื่อเปรียบเทียบกับกรณีไม่ใช้ค่านวนค่าเอสดีซีบนค่าพีเอฟ และค่าพีแอลพี



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ตารางที่ 7-7 ค่าเฉลี่ยเปอร์เซ็นต์ความผิดพลาดของการจำแนกภาษา เมื่อใช้ค่าลักษณะสำคัญชนิดต่าง ๆ

ค่าลักษณะสำคัญ	เปอร์เซ็นต์ความผิดพลาด
PF	17.46
SDCPF	21.07
MFCC	25.48
PLP	19.51
SDCPLP	23.36
SDC+PLP	21.15

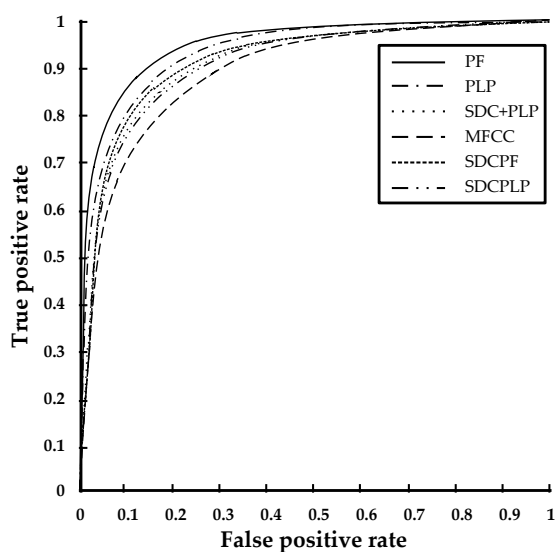
ตัวหนาแสดงค่าผิดพลาดที่น้อยที่สุด

เมื่อพิจารณาค่าลักษณะสำคัญของพีแอลพีในแบบต่าง ๆ ได้แก่ ค่าสัมประสิทธิ์พีแอลพี ค่าเอสดีซีพีแอลพี (SDCPLP) และค่าการประกอบร่วมกันของเอสดีซีพีแอลพีและค่าสัมประสิทธิ์พีแอลพี (SDC+PLP) พบว่าประสิทธิภาพของการจำแนกภาษาเมื่อใช้ค่าสัมประสิทธิ์พีแอลพีให้ผลดีกว่าการใช้ค่าเอสดีซีพีแอลพี (SDCPLP) และค่าการประกอบกันของเอสดีซีพีแอลพีและค่าสัมประสิทธิ์พีแอลพี (SDC+PLP) แต่ประสิทธิภาพของการใช้ค่าลักษณะสำคัญจากค่าพีแอลพีแบบต่าง ๆ นี้ไม่มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติ เนื่องจากค่าลักษณะสำคัญทั้งสามค่ามีความสัมพันธ์กันสูง โดยทุกค่ามีค่าสัมประสิทธิ์พีแอลพีเป็นค่าลักษณะสำคัญหลักร่วมกันทั้งหมด

สำหรับผลการทดลองที่ใช้การคำนวณค่าเอสดีซีเพื่อหาค่าการเปลี่ยนแปลงของค่าลักษณะสำคัญต่าง ๆ พบว่าค่าลักษณะสำคัญที่ได้จากการคำนวณเอสดีซีให้ความผิดพลาดของการจำแนกภาษาเพิ่มขึ้นแต่ไม่แตกต่างกันอย่างมีนัยสำคัญเมื่อเปรียบเทียบกับค่าลักษณะสำคัญที่ไม่ใช้การคำนวณเอสดีซี ซึ่งผลการทดลองของการใช้ค่าเอสดีซีในวิทยานิพนธ์นี้แตกต่างไปจากรวรรณกรรมอื่นที่ได้นำการคำนวณค่าเอสดีซีไปใช้หาค่าลักษณะสำคัญเพื่อใช้ในการรู้จำภาษาที่ให้ผลการรู้จำดีขึ้น และผลการทดลองของค่าลักษณะสำคัญที่ได้จากการคำนวณค่าเอสดีซี และค่าลักษณะสำคัญต่าง ๆ ไม่แตกต่างกันอย่างมีนัยสำคัญ เนื่องจากเวกเตอร์ค่าลักษณะสำคัญที่ได้มีส่วนข้อมูลที่ซ้ำกันและมีความสัมพันธ์กัน คือข้อมูลจากค่าเอสดีซี และค่าลักษณะสำคัญที่ใช้คำนวณเอสดีซี และเมื่อนำเวกเตอร์ที่มีข้อมูลที่เหมือนกันไปใช้ร่วมกับการจำแนกด้วยแบบจำลองลาเทนท์ไดนามิคคอนดิชันนอลแรนดอมฟิลด์ซึ่งมีความสามารถเรียนรู้ระหว่างข้อมูลที่อินพุตที่มีความสัมพันธ์กัน ทำให้นำไปสร้างแบบจำลองของการจำแนกที่สนใจคุณสมบัติของความสัมพันธ์ระหว่างข้อมูลได้

ภาพที่ 7-2 แสดงเส้นโค้ง ROC ซึ่งได้จากผลลัพธ์ของการจำแนก 28 คู่ภาษาพบว่า เส้นโค้ง ROC ของผลการจำแนกเมื่อใช้แบบจำลองลาเทนท์ไดนามิคคอนดิชันนอลแรนดอมฟิลด์ร่วมกับค่าพี

เอฟสูงกว่าเส้นของค่าลักษณะสำคัญชนิดอื่น ๆ แสดงว่าค่าพีเอฟให้ผลการจำแนกภาษาดีกว่าค่าลักษณะสำคัญชนิดต่าง ๆ และเมื่อเปรียบเทียบระหว่างการใช้ค่าพีเอฟ และค่าเอสดีซีร่วมกับค่าพีเอฟ พบว่า ค่าพีเอฟให้ผลการรู้จำแม่นยำมากกว่า และเมื่อใช้การคำนวณค่าเอสดีซีมาใช้กับค่าพีเอฟ ไม่ทำให้ผลการจำแนกแม่นยำเพิ่มขึ้น และภาคผนวก ฅ เส้นโค้ง ROC ซึ่งได้จากผลลัพธ์ของแต่ละตัวจำแนกด้วยแบบจำลองลาเทนท์ไดนามิคคอนดิชันนอลแรนดอมฟิลด์ร่วมกับค่าพีเอฟ และค่าลักษณะสำคัญต่าง ๆ สำหรับแต่ละคู่ภาษา



ภาพที่ 7-2 เส้นโค้ง ROC ของผลลัพธ์ของการจำแนก 28 คู่ภาษา เมื่อใช้ค่าลักษณะสำคัญชนิดต่าง ๆ

ผลการทดลองที่ 3.1 ซึ่งเปรียบเทียบประสิทธิภาพของตัวจำแนกภาษา 28 คู่ภาษาเมื่อใช้ค่าลักษณะสำคัญชนิดต่าง ๆ ได้แก่ PF, SDCPF, MFCC, PLP, SDCPLP, และ SDC+PLP แสดงดังตารางที่ 7-8 พบว่า ความสามารถของการจำแนกภาษาของแต่ละคู่ภาษาส่วนใหญ่มีแนวโน้มไปในทิศทางเดียวกัน นั่นคือ ตัวจำแนกภาษาที่ใช้ค่าพีเอฟ มีค่าความแม่นยำของการจำแนกดีกว่าการใช้ค่าลักษณะสำคัญชนิดอื่น และตัวจำแนกที่ใช้ค่าสัมประสิทธิ์ MFCC ให้ผลความแม่นยำของการจำแนกที่ต่ำกว่าค่าลักษณะสำคัญชนิดอื่น ๆ จากผลการทดลองพบว่า มีตัวจำแนกภาษาจำนวน 22 คู่ภาษาที่ใช้ค่าพีเอฟให้ผลการจำแนกดีกว่าการใช้ค่าสัมประสิทธิ์ค่าพีแอลพี และจากจำนวนนี้มี 3 คู่ภาษาที่ให้ผลการจำแนกแตกต่างกันอย่างมีนัยสำคัญ ได้แก่ ระหว่างคู่ภาษารัสเซียกับภาษาจีนกวางตุ้ง ระหว่างคู่ภาษาไทยกับภาษาจีนงู และระหว่างคู่ภาษาเออร์ดูกับภาษาจีนงู ในทางตรงกันข้ามผลการจำแนกเมื่อใช้ค่าพีเอฟให้ผลการจำแนกแย่กว่าการใช้ค่าสัมประสิทธิ์ PLP มีจำนวน 6 คู่ภาษา แต่ไม่แตกต่างกันอย่างมีนัยสำคัญ ได้แก่ ระหว่างคู่ภาษาจีนมินกับภาษาจีนงู ระหว่างคู่ภาษาจีนมินกับภาษาเออร์ดู ระหว่างคู่ภาษาอารบิกกับภาษาเออร์ดู ระหว่างคู่ภาษาไทยกับภาษาเออร์ดู ระหว่างคู่ภาษาจีนมินกับภาษารัสเซีย และระหว่างคู่ภาษาอารบิกกับภาษารัสเซีย ขณะที่เปรียบเทียบการใช้ค่าพีเอฟกับค่าเอสดีซีพีแอลพี พบว่าค่าพีเอฟให้ความแม่นยำของการจำแนกดีกว่าค่าเอสดีซีพีแอลพี และค่าแตกต่างกันอย่างมีนัยสำคัญมีจำนวน 16 คู่ภาษา ภาพที่ 7-3 แสดงค่าความผิดพลาดของการจำแนกภาษา

สำหรับ 28 คู่ภาษาที่ใช้แบบจำลองลาเทนท์ไดนามิคคอนดิชันนอลแรนดอมฟิลด์ร่วมกับค่าลักษณะสำคัญต่าง ๆ ซึ่งกราฟแห่งของความผิดพลาดของตัวจำแนกเปรียบเทียบการใช้ค่าลักษณะสำคัญแต่ละชนิด พบว่าผลการจำแนกส่วนใหญ่เมื่อใช้ค่าพีเอฟให้ความผิดพลาดการจำแนกต่ำกว่าตัวจำแนกที่ใช้ค่าลักษณะสำคัญชนิดอื่น ๆ

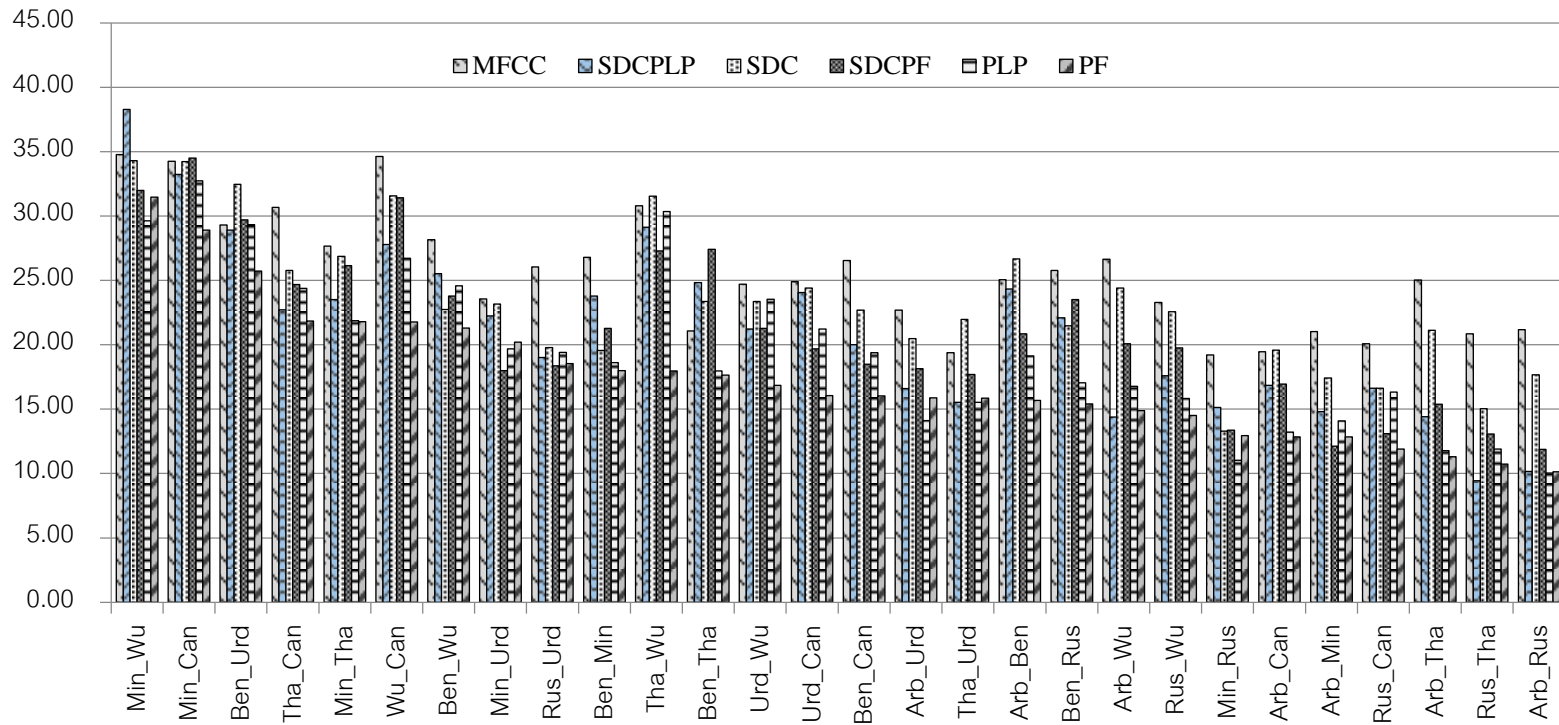
เมื่อเปรียบเทียบการใช้ค่าพีเอฟกับค่าเอสดีซีพีแอลพีร่วมกับค่าพีแอลพี (SDCPLP +PLP) พบว่า ตัวจำแนกที่ใช้ค่าพีเอฟให้ความแม่นยำการจำแนกสูงกว่าการใช้ค่าเอสดีซีพีแอลพีร่วมกับค่าพีแอลพี มีจำนวน 25 คู่ภาษา ซึ่งในจำนวนนี้มีจำนวน 10 คู่ภาษาที่ให้ความแตกต่างกันอย่างมีนัยสำคัญ ได้แก่ ระหว่างคู่ภาษาจีนมินกับภาษาจีนวู ระหว่างคู่ภาษาจีนมินกับภาษาเออร์ดู ระหว่างคู่ภาษาอารบิกกับภาษาเบงกาลี ระหว่างคู่ภาษาเบงกาลีกับภาษาจีนมิน ระหว่างคู่ภาษาเบงกาลีกับภาษารัสเซีย ระหว่างคู่ภาษาเบงกาลีกับภาษาไทย ระหว่างคู่ภาษาเบงกาลีกับภาษาจีนวู ระหว่างคู่ภาษาจีนมินกับภาษาจีนวู ระหว่างคู่ภาษารัสเซียกับภาษาจีนกวางตุ้ง ระหว่างคู่ภาษาไทยกับภาษาจีนวู ระหว่างคู่ภาษาเออร์ดูกับภาษาจีนวู ระหว่างคู่ภาษาเออร์ดูกับภาษาจีนกวางตุ้ง และ ระหว่างคู่ภาษาจีนวูกับภาษาจีนกวางตุ้ง ขณะที่ตัวจำแนกที่ใช้ค่าพีเอฟให้ความแม่นยำการจำแนกแย่งเมื่อเปรียบเทียบกับ การใช้ค่าเอสดีซีพีแอลพีมีจำนวน 3 คู่ภาษา ได้แก่ ระหว่างคู่ภาษารัสเซียกับภาษาจีนกวางตุ้ง ระหว่างคู่ภาษาไทยกับภาษาจีนวู และระหว่างคู่ภาษาเออร์ดูกับภาษาจีนวู เมื่อเปรียบเทียบค่าพีเอฟกับค่าเอสดีซีพีเอฟ (SDCPF) พบว่า ค่าพีเอฟให้ผลการจำแนกดีกว่า จำนวน 27 คู่ภาษา และมีจำนวน 7 คู่ภาษาที่ผลการจำแนกของการใช้ค่าพีเอฟสูงกว่าค่าเอสดีซีพีเอฟ ซึ่งให้ผลแตกต่างกันอย่างมีนัยสำคัญ แต่มีจำนวน 1 คู่ภาษาที่ผลการจำแนกของการใช้ค่าพีเอฟต่ำกว่าค่าเอสดีซีพีเอฟอย่างมีนัยสำคัญ ได้แก่ ระหว่างคู่ภาษาจีนมิน กับภาษาเออร์ดู ขณะที่ตัวจำแนกที่ใช้ค่าพีเอฟให้ความแม่นยำของการจำแนกดีกว่าการใช้ค่าสัมประสิทธิ์ MFCC จำนวน 24 คู่ภาษาและแตกต่างกันอย่างมีนัยสำคัญ

ตารางที่ 7-8 ความแม่นยำของตัวจำแนกภาษาเมื่อใช้ค่าลักษณะสำคัญชนิดต่าง ๆ
(PF, SDCPF, PLP, SDCPLP, SDC, และ MFCC)

แต่ละคู่ภาษา	PF	SDCPF	PLP	SDC+PLP	SDCPLP	MFCC
CH.min-CH.wuu	68.53±2.88	68.01±4.35	70.60±4.39	61.80±2.17*	65.60±6.54	65.22±4.06
CH.min-CH.can	71.09±3.75	65.49±6.41	67.00±2.65	66.80±2.95	65.60±3.58*	65.74±3.32*
ben-urd	74.28±6.55	70.29±3.78	70.80±6.38	71.20±2.49	67.40±4.67	70.70± 4.20
tha-CH.can	78.15±3.81	75.33±3.16	75.80±4.09	77.40±4.34	74.00±4.47	69.32±5.94*
CH.min-tha	78.20±6.91	73.86±5.08	78.20±3.42	76.40±5.22	73.00±2.24	72.33±2.86
CH.wuu-CH.can	78.23±2.99	68.59±5.09*	73.20±3.96	72.20±2.95*	68.60±2.30*	65.38± 3.69
ben-CH.wuu	78.70±3.45	76.22±1.55	75.40±4.51	74.20±1.79*	77.40±4.56	71.84± 3.50
CH.min-urd	79.80±1.55	82.01±5.46	80.60±3.78	78.00±4.90	76.80±3.90	76.44± 2.51
rus-urd	81.46±2.49	81.63±3.50	80.40±3.51	81.00±3.54	80.40±3.58	73.95± 4.58
ben-CH.min	82.00±1.57	78.72±4.13	81.40±1.95	76.20±2.77*	80.40±2.70	73.19± 3.64
tha-CH.wuu	82.03±1.79	72.72±3.45*	69.60±5.59*	70.60±6.88*	68.40±3.78*	69.20± 2.37
ben-tha	82.36±1.18	72.57±2.13*	82.00±2.00	75.20±4.55*	76.80±4.76*	78.93± 1.13
urd-CH.wuu	83.15±2.60	78.73±1.77*	76.40±4.16*	78.80±5.31	76.80±2.86*	75.29± 3.30
urd-CH.can	83.95±2.92	80.32±2.27	78.60±4.39	76.00±4.47*	75.60±6.31*	75.10± 3.82
ben-CH.can	83.96±2.10	81.52±4.60	80.60±2.61	80.00±3.39	77.40±3.78*	73.45± 3.29
arb-urd	84.12±1.68	81.86±3.59	85.80±3.56	83.40±1.82	79.60±2.30*	77.30± 1.68
tha-urd	84.14±1.52	82.31±2.67	84.40±2.41	84.40±1.34	78.20±3.19*	80.61±4.72
arb-ben	84.33±3.04	79.16±4.65	80.80±3.42	75.80±1.92*	73.20±0.84*	74.94± 2.42
ben-rus	84.60±5.13	76.49±5.33*	83.00±3.74	77.80±1.10*	78.60±3.58	74.23± 2.08
arb-CH.wuu	85.12±2.09	79.91±2.08*	83.00±3.54	85.60±1.52	75.80±2.68*	73.36± 2.01
rus-CH.wuu	85.49±2.95	80.25±4.31	84.40±2.07	82.60±3.78	77.60±2.70*	76.71± 4.04
CH.min-rus	87.06±3.68	86.62±3.50	89.00±2.45	84.80±3.03	86.60±4.83	80.78± 3.11
arb-CH.min	87.16±1.50	87.88±2.84	86.00±2.92	85.20±2.05	82.40±1.52*	80.54± 2.96
arb-CH.can	87.16±2.29	83.06±3.19*	87.00±2.83	83.40±7.06	80.20±3.11*	78.98± 3.91
rus-CH.can	88.10±2.54	86.91±1.31	83.80±3.27*	83.20±3.83*	83.60±4.56	79.92± 3.43
arb-tha	88.70±2.80	84.62±2.42*	88.20±2.49	85.40±3.36	79.00±2.35*	74.97± 1.76
rus-tha	89.26±2.43	86.93±2.33	88.00±2.83	90.60±2.07	85.00±3.54	79.16± 3.66
arb-rus	89.87±2.78	88.12±1.51	90.00±2.00	89.80±2.17	82.20±3.90*	78.82± 4.99

- ช่องที่แสดงเน้นสีเทาแสดงค่าความแม่นยำที่สูงที่สุดสำหรับแต่ละคู่ภาษา. เครื่องหมาย * แสดงการจำแนกของคู่ภาษาเมื่อใช้ค่าพีเอฟให้ค่าความแม่นยำกันอย่างมีนัยสำคัญทั้ง 5 ชุดทดสอบไขว้ เทียบกับค่าลักษณะสำคัญที่พิจารณา โดยการทดสอบความแตกต่าง t-test ที่ระดับความเชื่อมั่น $\alpha = 0.05$

- ค่าพีเอฟแม่นยำสูงกว่าค่าพีแอลพีอย่างมีนัยสำคัญ 3 คู่ภาษา ค่าพีเอฟแม่นยำสูงกว่าค่า SDCPLP อย่างมีนัยสำคัญ 16 คู่ภาษา ค่าพีเอฟแม่นยำสูงกว่าค่า SDCPLP+PLP อย่างมีนัยสำคัญ 10 คู่ภาษา ค่าพีเอฟแม่นยำสูงกว่าค่า SDCPF อย่างมีนัยสำคัญ 8 คู่ภาษา และค่าพีเอฟแม่นยำสูงกว่าค่า MFCC อย่างมีนัยสำคัญ 24 คู่ภาษา และผลการจำแนกโดยใช้ค่าพีเอฟกับค่าลักษณะสำคัญอื่น ๆ โดยทดสอบความแตกต่างจาก 28 คู่ภาษา ความแม่นยำการจำแนกโดยใช้ค่าพีเอฟสูงกว่าค่าสัมประสิทธิ์ MFCC, ค่า SDC, ค่า SDC+PLP อย่างมีนัยสำคัญ และความแม่นยำการจำแนกโดยใช้ค่า PLP สูงกว่า SDC อย่างมีนัยสำคัญ



ภาพที่ 7-3 ค่าความผิดพลาดของการจำแนกภาษาด้วยแบบจำลองลาเทนท์ไดนามิคคอนดิชันนอลแรนดอมฟีลด์ร่วมกับค่าลักษณะสำคัญชนิดต่าง ๆ โดยวัดค่าความผิดพลาดจาก 28 คู่ภาษา

ตารางที่ 7-9 เปอร์เซ็นต์ความแม่นยำของการจำแนกของตัวจำแนกภาษา 28 คู่ภาษาเมื่อใช้การจำแนกด้วยแบบจำลองลาเทนท์ไดนามิคคอนดิชันนอลแรนดอมฟิลด์ร่วมกับค่าพีเอฟชนิดต่าง ๆ เมื่อลดค่าพีเอฟทีละชนิด

คู่ภาษา	Strident	Nasal	Continuant	Voiced	Tense	Round	Coronal
CH.min-CH.wuu	68.03±0.04	68.38±0.02	68.25±0.06	69.78±0.01	69.55±0.04	69.37±0.04	69.03±0.04
CH.min-CH.can	70.37±0.01	68.38±0.04	69.79±0.02	68.60±0.05	69.44±0.03	71.00±0.04	70.71±0.01
ben-urd	75.94±0.03	76.98±0.02	77.34±0.01	78.17±0.04	77.33±0.03	75.14±0.04	73.45±0.04
tha-CH.can	80.36±0.04	77.23±0.02	77.17±0.03	78.90±0.03	78.83±0.03	79.73±0.05	79.23±0.05
CH.min-tha	75.22±0.06	75.58±0.06	77.95±0.03	77.55±0.05	75.81±0.01	76.22±0.02	74.11±0.04
CH.wuu-CH.can	78.36±0.02	77.60±0.03	76.03±0.03	77.57±0.03	78.39±0.04	75.35±0.04	77.80±0.04
ben-CH.wuu	77.40±0.04	78.39±0.04	78.02±0.04	77.43±0.03	79.51±0.05	78.03±0.02	80.57±0.02
CH.min-urd	80.16±0.02	79.99±0.03	82.30±0.02	78.76±0.05	82.08±0.02	80.25±0.04	80.75±0.03
rus-urd	82.77±0.04	79.58±0.05	82.51±0.04	83.73±0.04	81.99±0.02	81.01±0.05	80.61±0.03
ben-CH.min	81.61±0.02	82.38±0.03	80.39±0.02	82.31±0.02	81.93±0.02	80.21±0.03	81.60±0.02
tha-CH.wuu	78.45±0.02*	80.16±0.04	79.95±0.06	79.12±0.03	79.68±0.04	77.36±0.01*	78.84±0.03
ben-tha	83.26±0.04	81.25±0.04	80.85±0.05	83.46±0.02	81.46±0.04	82.12±0.06	83.05±0.04
urd-CH.wuu	80.53±0.03	84.43±0.02	78.04±0.03*,**	82.47±0.03	84.60±0.03	82.82±0.03	83.83±0.03
urd-CH.can	83.31±0.04	82.99±0.03	82.99±0.02	81.28±0.04	83.42±0.02	80.44±0.05	82.44±0.03

คู่ภาษา	Strident	Nasal	Continuant	Voiced	Tense	Round	Coronal
ben-CH.can	85.71±0.03	83.15±0.02	82.71±0.02**	85.31±0.00	82.99±0.05	83.46±0.03	82.41±0.03
arb-urd	85.53±0.01	86.18±0.01*	85.88±0.01	86.50±0.02	85.47±0.03	84.73±0.02	83.73±0.02
tha-urd	85.13±0.04	83.06±0.03	86.44±0.02	85.94±0.03	84.37±0.02	84.56±0.04	84.99±0.03
arb-ben	80.29±0.00*	82.06±0.01	81.67±0.02	81.56±0.02	81.85±0.05	82.21±0.04	81.80±0.02
ben-rus	82.03±0.03	81.66±0.02	82.52±0.04	82.76±0.03	80.14±0.05	81.60±0.03	81.56±0.04
arb-CH.wuu	86.39±0.02	85.96±0.01	87.28±0.02	85.73±0.01	85.68±0.02	85.25±0.02	86.56±0.01
rus-CH.wuu	82.75±0.02	83.74±0.02	83.47±0.02	83.53±0.02	83.91±0.03	82.55±0.05	81.66±0.03
CH.min-rus	86.79±0.01	86.05±0.02	85.67±0.04	85.52±0.03	84.49±0.04	86.07±0.02	84.75±0.02
arb-CH.can	86.49±0.03	85.92±0.03	87.03±0.02	83.87±0.03	86.23±0.02	85.85±0.01	87.34±0.04
arb-CH.min	85.25±0.01	86.23±0.01	85.32±0.01*	85.32±0.02	87.85±0.03	86.05±0.03	85.79±0.02
rus-CH.can	87.26±0.03	88.56±0.02	87.85±0.03	88.58±0.03	88.31±0.04	87.71±0.04	88.09±0.01
arb-tha	88.21±0.03	88.51±0.02	88.51±0.03	89.78±0.03	87.72±0.04	89.58±0.02	90.75±0.02
rus-tha	88.89±0.03	88.46±0.03	89.74±0.01	88.21±0.04	88.54±0.02	89.59±0.03	88.98±0.04
arb-rus	89.40±0.01	88.25±0.03	87.37±0.02	87.89±0.02	89.95±0.03	89.22±0.04	87.31±0.04

คู่ภาษา	Anterior	Low	Back	High	consonant	Vocalic
CH.min-CH.wuu	66.57±0.05	67.84±0.02	65.53±0.03	69.50±0.02	70.13±0.02	69.22±0.05
CH.min-CH.can	70.73±0.03	69.63±0.05	69.69±0.03	68.45±0.05	68.48±0.03	71.00±0.01
ben-urd	70.96±0.06	75.02±0.04	72.62±0.06	76.25±0.03	73.80±0.02	76.51±0.05
tha-CH.can	80.53±0.03	80.35±0.02	76.39±0.05	78.44±0.03	78.44±0.03	79.61±0.03
CH.min-tha	74.27±0.05	77.40±0.06	75.70±0.04	77.36±0.02	77.42±0.05	78.69±0.04
CH.wuu-CH.can	75.76±0.03	77.76±0.01	76.14±0.03	77.28±0.03	77.90±0.04	74.54±0.04
ben-CH.wuu	79.23±0.06	79.36±0.04	78.35±0.04	78.72±0.04	80.15±0.05	78.30±0.04
CH.min-urd	79.52±0.03	81.23±0.04	81.70±0.05	80.46±0.02	80.86±0.03	80.68±0.02
rus-urd	80.69±0.06	82.66±0.04	82.05±0.03	83.83±0.02	79.61±0.05	83.64±0.01
ben-CH.min	82.19±0.01	80.99±0.03	82.31±0.02	82.11±0.02	79.55±0.02*	81.82±0.02
tha-CH.wuu	81.56±0.05	77.76±0.02*	78.98±0.03	80.10±0.02	78.71±0.03	79.54±0.04
ben-tha	83.34±0.03	84.42±0.04	82.34±0.05	82.72±0.05	82.59±0.05	83.35±0.04
urd-CH.wuu	84.35±0.04	80.93±0.04	81.84±0.02	82.40±0.03	80.93±0.02	81.22±0.04
urd-CH.can	83.41±0.03	81.27±0.05	83.55±0.02	82.14±0.02	80.97±0.04	80.15±0.04
ben-CH.can	84.90±0.02	83.46±0.04	82.87±0.03	84.27±0.04	85.68±0.03	82.99±0.01

คู่ภาษา	Anterior	Low	Back	High	consonant	Vocalic
arb-urd	84.18±0.02	84.88±0.01	85.86±0.02	84.72±0.01	85.87±0.02	84.34±0.03
tha-urd	85.51±0.02	85.26±0.02	85.88±0.03	85.03±0.02	83.92±0.02	84.58±0.02
arb-ben	82.98±0.03	82.51±0.03	80.58±0.04	82.47±0.03	80.57±0.01*	80.14±0.03
ben-rus	81.86±0.06	81.45±0.02	81.58±0.02	80.77±0.03	82.50±0.03	81.59±0.02
arb-CH.wuu	85.24±0.01	86.77±0.03	84.78±0.03	87.30±0.02	85.14±0.03	85.67±0.02
rus-CH.wuu	82.44±0.02	83±0.03	82.82±0.01	83.06±0.03	81.01±0.03	83.40±0.03
CH.min-rus	85.65±0.02	80.85±0.02	85.36±0.03	85.39±0.04	84.88±0.02	86.24±0.02
arb-CH.can	86.99±0.02	80.87±0.03	84.40±0.03	86.32±0.04	86.61±0.03	86.62±0.02
arb-CH.min	86.74±0.01	85±0.02	85.19±0.04	86.87±0.02	84.38±0.02*	85.11±0.02
rus-CH.can	86.71±0.02	88.22±0.04	87.46±0.02	86.89±0.03	86.34±0.03	88.20±0.03
arb-tha	90.07±0.01	89.58±0.03	88.22±0.03	89.87±0.03	89.78±0.03	88.22±0.02
rus-tha	90.13±0.03	88.52±0.03	88.44±0.03	91.11±0.02	87.61±0.03	89.85±0.02
arb-rus	88.92±0.03	87.85±0.03	88.47±0.01	89.18±0.02	90.38±0.03	88.13±0.04

7.3.2 ผลการทดลองที่ 3.2: การประเมินผลกระทบเมื่อลดกลุ่มภาษาเป้าหมายที่มีความคลุมเครือต่อประสิทธิภาพของระบบการรู้จำภาษาพูดที่ใช้ค่าพีเอฟ และระบบการรู้จำภาษาพูดที่ใช้ค่าลักษณะทางเสียงอื่น ๆ

ตารางที่ 7-10 แสดงค่าเฉลี่ยเปอร์เซ็นต์ความผิดพลาดการจำแนกภาษาเมื่อลดความคลุมเครือของภาษาเป้าหมาย พบว่าผลการจำแนกภาษาเมื่อใช้ตัวจำแนกภาษา 15 คู่ภาษา (เมื่อลดความคลุมเครือในภาษาจีนโดยไม่รวมการพิจารณาภาษาจีนวู และภาษาจีนมิน) ทำให้ค่าเฉลี่ยเปอร์เซ็นต์ความผิดพลาดลดลง เมื่อเปรียบเทียบกับผลการจำแนกภาษาเมื่อใช้ตัวจำแนก 28 คู่ภาษา (ที่ไม่ได้ลดความคลุมเครือในภาษาเป้าหมาย) ซึ่งค่าเฉลี่ยเปอร์เซ็นต์ความผิดพลาดของตัวจำแนกภาษา 15 คู่ภาษา เมื่อใช้ค่าพีเอฟต่ำกว่า ตัวจำแนกที่ใช้ค่าลักษณะสำคัญทางเสียงชนิดอื่น และมีค่าเปอร์เซ็นต์ค่าผิดพลาดลดลงเมื่อเปรียบเทียบกับผลการทดลองที่ใช้ตัวจำแนกภาษา 28 คู่ภาษาเป็น 10.08%, 8.58%, 6.33%, 10.87%, 6.09% และ 9.99% สำหรับการใช้ค่า PF, SDCPF, MFCC, PLP, SDCPLP และ SDC+PLP เป็นค่าลักษณะสำคัญ ตามลำดับ โดยค่าเปอร์เซ็นต์ผิดพลาดลดประมาณ 2 เปอร์เซ็นต์ สำหรับทุกค่าลักษณะสำคัญ จากผลการทดลองสรุปได้ว่า การจำแนกภาษาหรือการรู้จำภาษาที่มีภาษาเป้าหมายที่มีความคลุมเครือ ทำให้ความแม่นยำของการจำแนกภาษาไม่ดี และเพื่อปรับปรุงประสิทธิภาพระบบการรู้จำภาษาพูดควรหาวิธีลดความคลุมเครือระหว่างภาษาเป้าหมายก่อนทำการรู้จำภาษา อย่างไรก็ตามถึงแม้ไม่สามารถลดความคลุมเครือของการตัดสินใจจำแนกภาษาได้จริง แต่วัตถุประสงค์ของการทดลองนี้คือ การทดลองเพื่อหาแนวทางการปรับปรุงระบบให้ผลการจำแนกภาษาดีขึ้นต่อไป

ตารางที่ 7-10 เปอร์เซ็นต์ความผิดพลาดการจำแนกภาษาเมื่อลดความคลุมเครือของภาษาเป้าหมาย

ค่าลักษณะสำคัญ	ตัวจำแนกภาษา (28 คู่ภาษา)	เมื่อลดภาษาจีน (15 คู่ภาษา)	เปอร์เซ็นต์ค่าผิดพลาดลดลง
PF	17.46	15.70	10.08
SDCPF	21.07	19.26	8.58
MFCC	25.48	23.87	6.33
PLP	19.51	17.39	10.87
SDCPLP	23.36	21.94	6.09
SDC+PLP	21.15	19.04	9.99

ตัวหนาแสดงค่าผิดพลาดที่น้อยที่สุด

7.3.3 ผลการทดลองที่ 3.3: วัดประสิทธิภาพของระบบการรู้จำภาษาพูดที่ใช้ตัวจำแนกด้วยแบบจำลองลาเทนท์ไดนามิกคอนดิชันนอลแรนดอมฟิลด์

ตารางที่ 7-11 แสดงความผิดพลาดของตัวจำแนกภาษาเมื่อใช้แบบจำลองชนิดต่าง ๆ ร่วมกับค่าพีเอฟ ได้แก่ แบบจำลองลาเทนท์ไดนามิกคอนดิชันนอลแรนดอมฟิลด์ แบบจำลองคอนดิชันนอลแรนดอมฟิลด์ และแบบจำลองเกาส์เซียน พบว่าการจำแนกด้วยแบบจำลองลาเทนท์ไดนามิกคอนดิชันนอลแรนดอมฟิลด์ให้ค่าเปอร์เซ็นต์ความผิดพลาดการจำแนกภาษา 17.46 ซึ่งต่ำกว่าค่าเปอร์เซ็นต์ความผิดพลาดการจำแนกภาษาเมื่อใช้แบบจำลองคอนดิชันนอลแรนดอมฟิลด์ และแบบจำลองเกาส์เซียน

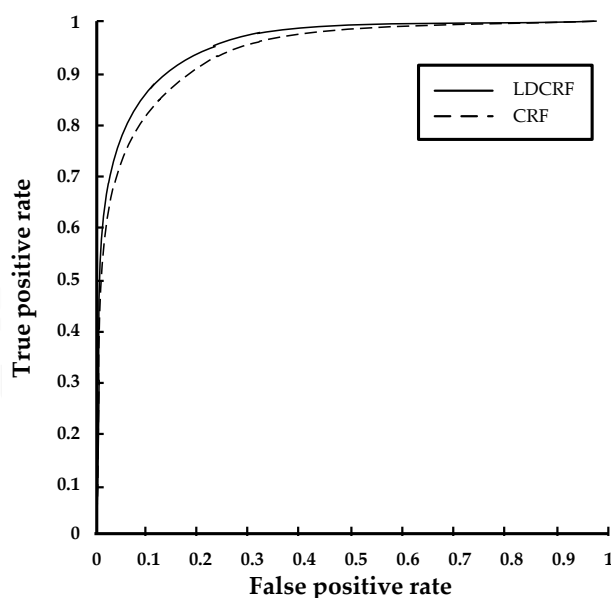
ตารางที่ 7-11 เปอร์เซนต์ความผิดพลาดของการจำแนกภาษาเมื่อใช้แบบจำลองชนิดต่าง ๆ

แบบจำลอง	เปอร์เซนต์ความผิดพลาดการจำแนก
LDCRF	17.46 *
CRF	19.11
GMM (PF)	30.63
GMM (PLP)	52.28

* แสดงค่าผิดพลาดที่น้อยที่สุด

นอกจากนี้จากตารางที่ 7-11 เปรียบเทียบประสิทธิภาพของตัวจำแนกภาษาแบบจำลองเกาส์เซียน เมื่อใช้ค่าลักษณะสำคัญจากค่าพีเอฟ และค่าพีแอลพี ตามการทดลองที่ 3.3.1 พบว่า เปอร์เซนต์ความผิดพลาดของการจำแนกภาษาด้วยแบบจำลองเกาส์เซียนเมื่อใช้ค่าพีเอฟ ให้เปอร์เซนต์ความผิดพลาดของการจำแนกภาษาต่ำกว่าการใช้ค่าพีแอลพี ซึ่งเมื่อพิจารณาผลการทดลองในตารางที่ 7-12 ที่แสดงเปอร์เซนต์ความแม่นยำของการจำแนกภาษาจากตัวจำแนก 28 คู่ภาษา เมื่อใช้แบบจำลองชนิดต่าง ๆ ร่วมกับค่าพีเอฟ พบว่า มีตัวจำแนกภาษา 22 คู่ภาษาที่ใช้แบบจำลองลาเทนท์ไดนามิกคอนดิชันนอลแรนดอมฟิลด์ร่วมกับค่าพีเอฟให้ประสิทธิภาพดีกว่าตัวจำแนกแบบจำลองเกาส์เซียนร่วมกับค่าพีเอฟ และมีตัวจำแนก 6 คู่ภาษาที่ใช้แบบจำลองเกาส์เซียนให้ความแม่นยำสูงกว่าตัวจำแนกภาษาที่ใช้แบบจำลองลาเทนท์ไดนามิกคอนดิชันนอลแรนดอมฟิลด์ ได้แก่ ระหว่างคู่ภาษาอารบิกและภาษาจีนดู ระหว่างคู่ภาษารัสเซียและภาษาจีนดู ระหว่างคู่ภาษาเบงกาลีและภาษาเออร์ดู ระหว่างคู่ภาษาเบงกาลีและภาษาจีนดู ระหว่างคู่ภาษาเบงกาลีและภาษาจีนกวางตุ้ง และระหว่างคู่ภาษาอารบิกและภาษาจีนมิน ขณะที่ผลการจำแนกภาษาของ 28 คู่ภาษาเมื่อตัวจำแนกที่ใช้แบบจำลองลาเทนท์ไดนามิกคอนดิชันนอลแรนดอมฟิลด์ร่วมกับค่าพีเอฟให้ค่าความแม่นยำสูงกว่าตัวจำแนกที่ใช้แบบจำลองเกาส์เซียนร่วมกับค่าพีแอลพี และมีความแตกต่างกันอย่างมีนัยสำคัญ

ผลการทดลองที่ 3.3.2 ซึ่งเปรียบเทียบประสิทธิภาพของตัวจำแนกภาษาด้วยแบบจำลองลาเทนท์ไดนามิกคอนดิชันนอลแรนดอมฟิลด์ กับตัวจำแนกที่ใช้แบบจำลองคอนดิชันนอลแรนดอมฟิลด์ พบว่า การจำแนกด้วยแบบจำลองลาเทนท์ไดนามิกคอนดิชันนอลแรนดอมฟิลด์ให้ค่าเปอร์เซ็นต์ความผิดพลาดการจำแนกต่ำกว่าการใช้แบบจำลองคอนดิชันนอลแรนดอมฟิลด์ แต่ไม่แตกต่างกันอย่างมีนัยสำคัญ ภาพที่ 7-4 แสดงเส้นโค้ง ROC ของผลการจำแนกภาษาโดยใช้แบบจำลองลาเทนท์ไดนามิกคอนดิชันนอลแรนดอมฟิลด์ร่วมกับค่าพีเอฟ และแบบจำลองคอนดิชันนอลแรนดอมฟิลด์ร่วมกับค่าพีเอฟจากตัวจำแนก 28 คู่ภาษา พบว่าเส้นโค้งของแบบจำลองลาเทนท์ไดนามิกคอนดิชันนอลแรนดอมฟิลด์ร่วมกับค่าพีเอฟ สูงกว่าเส้นโค้งของแบบจำลองคอนดิชันนอลแรนดอมฟิลด์เล็กน้อย แสดงว่าความสามารถของการจำแนกภาษาเมื่อใช้แบบจำลองลาเทนท์ไดนามิกคอนดิชันนอลแรนดอมฟิลด์ไม่แตกต่างกับการใช้แบบจำลองคอนดิชันนอลแรนดอมฟิลด์อย่างมีนัยสำคัญ จากตารางที่ 7-12 ที่แสดงเปอร์เซ็นต์ความแม่นยำของการจำแนกภาษาเมื่อใช้แบบจำลองชนิดต่าง ๆ ร่วมกับค่าพีเอฟของตัวจำแนก 28 คู่ภาษา พบว่ามีตัวจำแนก 5 คู่ภาษา ที่ใช้แบบจำลองคอนดิชันนอลแรนดอมฟิลด์ให้ค่าเปอร์เซ็นต์ความแม่นยำของการจำแนกสูงกว่า ซึ่งมีตัวจำแนก 2 คู่ภาษา ที่ค่าเปอร์เซ็นต์ความแม่นยำการจำแนกด้วยแบบจำลองลาเทนท์ไดนามิกคอนดิชันนอลแรนดอมฟิลด์สูงกว่าการใช้แบบจำลองคอนดิชันนอลแรนดอมฟิลด์โดยแตกต่างกันอย่างมีนัยสำคัญ ได้แก่ ระหว่างคู่ภาษาจีนมินกับภาษาจิงกวางตุง และระหว่างคู่ภาษาเออร์ดูกกับภาษาจิงวู และภาคผนวก ซ เส้นโค้ง ROC ซึ่งได้จากผลลัพธ์ของแต่ละตัวจำแนกด้วยแบบจำลองลาเทนท์ไดนามิกคอนดิชันนอลแรนดอมฟิลด์ร่วมกับค่าพีเอฟ และแบบจำลองคอนดิชันนอลแรนดอมฟิลด์ร่วมกับค่าพีเอฟสำหรับแต่ละคู่ภาษา



ภาพที่ 7-4 เส้นโค้ง ROC ของผลการจำแนกภาษาโดยใช้แบบจำลองลาเทนท์ไดนามิกคอนดิชันนอลแรนดอมฟิลด์ร่วมกับค่าพีเอฟ และแบบจำลองคอนดิชันนอลแรนดอมฟิลด์ร่วมกับค่าพีเอฟ

ตารางที่ 7-12 เปอร์เซ็นต์ความแม่นยำของการจำแนกภาษาเมื่อใช้แบบจำลองชนิด
ต่าง ๆ ร่วมกับค่าพีเอฟ สำหรับตัวจำแนก 28 คู่ภาษา

คู่ภาษา	LDCRF(PF)	CRF(PF)	GMM_PF	GMM_PLP
CH.min-CH.wuu	68.53 ± 2.88	65.62 ± 2.77	58.55 ± 22.82	42.70 ± 0.19
CH.min-CH.can	71.09 ± 3.75	64.75 ± 1.46*	54.79 ± 13.75	53.12 ± 0.22
ben-urd	74.28 ± 6.55	69.69 ± 1.49	83.92 ± 20.36	53.51 ± 3.14
tha-CH.can	78.15 ± 3.81	76.93 ± 2.35	52.85 ± 3.52	46.61 ± 3.54
CH.min-tha	78.20 ± 6.91	74.40 ± 1.65	49.83 ± 8.82	57.90 ± 4.72
CH.wuu-CH.can	78.23 ± 2.99	75.17 ± 1.33	47.10 ± 9.13	69.37 ± 11.32
ben-CH.wuu	78.70 ± 3.45	76.70 ± 3.63	82.72 ± 20.79	51.30 ± 4.19
CH.min-urd	79.80 ± 1.55	79.60 ± 3.07	60.80 ± 18.16	52.28 ± 0.11
rus-urd	81.46 ± 2.49	80.77 ± 3.38	49.44 ± 11.66	44.80 ± 0.13
ben-CH.min	82.00 ± 1.57	80.43 ± 2.91	93.05 ± 7.00	47.27 ± 0.10
tha-CH.wuu	82.03 ± 1.79	80.82 ± 4.01	54.85 ± 9.19	45.89 ± 15.92
ben-tha	82.36 ± 1.18	81.82 ± 4.92	80.41 ± 29.93	44.71 ± 14.61
urd-CH.wuu	83.15 ± 2.60	79.30 ± 1.86*	60.61 ± 8.01	56.41 ± 7.10
urd-CH.can	83.95 ± 2.92	80.47 ± 2.24	50.32 ± 9.24	52.69 ± 3.57
ben-CH.can	83.96 ± 2.10	83.68 ± 3.19	90.11 ± 13.92	50.37 ± 6.80
arb-urd	84.12 ± 1.68	84.47 ± 1.71	68.74 ± 19.03	50.62 ± 8.71
tha-urd	84.14 ± 1.52	85.67 ± 3.32	58.76 ± 17.73	42.96 ± 0.14
arb-ben	84.33 ± 3.04	81.08 ± 2.51	65.04 ± 17.28	53.66 ± 10.00
ben-rus	84.60 ± 5.13	82.67 ± 2.65	75.53 ± 15.99	59.43 ± 1.28
arb-CH.wuu	85.12 ± 2.09	84.70 ± 1.99	91.63 ± 6.89	41.64 ± 22.46
rus-CH.wuu	85.49 ± 2.95	81.68 ± 3.74	90.64 ± 11.28	35.78 ± 0.77
CH.min-rus	87.06 ± 3.68	85.21 ± 2.22	86.07 ± 20.43	38.52 ± 42.44
arb-CH.min	87.16 ± 1.50	87.07 ± 2.21	68.84 ± 23.75	42.61 ± 3.84
arb-CH.can	87.16 ± 2.29	85.90 ± 2.78	88.77 ± 12.01	44.38 ± 0.22
rus-CH.can	88.10 ± 2.54	87.51 ± 2.56	64.21 ± 18.59	45.69 ± 0.25
arb-tha	88.70 ± 2.80	89.19 ± 1.90	82.98 ± 14.61	38.76 ± 4.94
rus-tha	89.26 ± 2.43	89.39 ± 3.39	63.18 ± 27.62	42.75 ± 12.55
arb-rus	89.87 ± 2.78	90.35 ± 2.89	68.70 ± 11.37	30.50 ± 9.95

7.3.4 ผลการทดลองที่ 3.4: วัดประสิทธิภาพของระบบการรู้จำภาษาพูดเมื่อลดชนิดค่าพีเอฟที่ใช้ฝึกฝนตัวจำแนกภาษา

จากการทดลองที่ 3.4 ที่ศึกษาผลกระทบของค่าพีเอฟแต่ละชนิดในเวกเตอร์ค่าพีเอฟ เมื่อลดค่าพีเอฟที่ละชนิดจากเวกเตอร์ค่าลักษณะเฉพาะ และวัดประสิทธิภาพของตัวจำแนกพบว่า ผลการจำแนกด้วยแบบจำลองลาเทนท์ไดนามิคคอนดิชันนอลแรนดอมฟิลด์ ร่วมกับค่าพีเอฟเมื่อลดค่าพีเอฟหนึ่งชนิดจากเวกเตอร์ค่าลักษณะเฉพาะ มีเปอร์เซ็นต์ความผิดพลาดของการจำแนกภาษาเพิ่มขึ้นเมื่อเปรียบเทียบกับผลการจำแนกเมื่อใช้เวกเตอร์ค่าลักษณะเฉพาะที่ไม่ทำการลดค่าพีเอฟ อย่างไรก็ตาม ค่าเฉลี่ยของความผิดพลาดเพิ่มขึ้นนั้นไม่แตกต่างกันอย่างมีนัยสำคัญ ดังตารางที่ 7-13 โดยที่ผลการจำแนกภาษาเมื่อลดค่าพีเอฟ Back ออกจากเวกเตอร์ค่าลักษณะเฉพาะ ให้ค่าเปอร์เซ็นต์ความผิดพลาดการจำแนกเพิ่มขึ้นน้อยที่สุด ซึ่งตรงกันข้ามกับผลการจำแนกภาษาเมื่อลดค่าพีเอฟ High ที่ทำให้ค่าเปอร์เซ็นต์ความผิดพลาดของการจำแนกภาษาเพิ่มขึ้นมากที่สุด

ตารางที่ 7-13 เปอร์เซ็นต์ความผิดพลาดของการจำแนกภาษาเมื่อใช้การจำแนกด้วยแบบจำลองลาเทนท์ไดนามิคคอนดิชันนอลแรนดอมฟิลด์ร่วมกับค่าพีเอฟ เมื่อลดค่าพีเอฟหนึ่งชนิด

ค่าพีเอฟที่ลดจากเวกเตอร์	เปอร์เซ็นต์ความผิดพลาด	ค่าพีเอฟที่ลดจากเวกเตอร์	เปอร์เซ็นต์ความผิดพลาด
ไม่มีการลดค่าพีเอฟ	17.46 *	Vocalic	18.10
High	17.79	Continuant	18.12
Tense	17.80	Coronal	18.15
Voiced	17.87	Nasal	18.17
Strident	18.00	Round	18.31
Anterior	18.02	Consonantal	18.44
Low	18.05	Back	18.60

* แสดงค่าผิดพลาดที่น้อยที่สุด

จากตารางที่ 7-9 แสดงเปอร์เซ็นต์ความแม่นยำของการจำแนกเมื่อใช้การจำแนกด้วยแบบจำลองลาเทนท์ไดนามิคคอนดิชันนอลแรนดอมฟิลด์ร่วมกับค่าพีเอฟ จากตัวจำแนกภาษา 28 คู่ภาษา ซึ่งค่าพีเอฟที่แสดงในแต่ละคอลัมน์ของตารางแทนชนิดของค่าพีเอฟที่นำออกจากเวกเตอร์ค่าลักษณะเฉพาะค่าพีเอฟ พบว่าค่าความแม่นยำของการจำแนกเมื่อลดค่าพีเอฟไปหนึ่งชนิดออกจากเวกเตอร์ค่าลักษณะเฉพาะ เปรียบเทียบกับการไม่ลดค่าพีเอฟ มีตัวจำแนก 11 คู่ภาษาที่มีค่าความแม่นยำของการจำแนกแตกต่างกันอย่างมีนัยสำคัญ โดยมีตัวจำแนกภาษาจำนวน 1 คู่ภาษาที่มีความ

แม่นยำของการจำแนกเพิ่มขึ้นอย่างมีนัยสำคัญ เมื่อลดค่าพีเอฟหนึ่งชนิด นั่นคือระหว่างคู่ภาษาอารบิกคู่กับภาษาเออร์ดู ในทางตรงกันข้ามมีตัวจำแนกภาษา 10 คู่ภาษาที่มีค่าความแม่นยำลดลงเมื่อลดค่าพีเอฟหนึ่งชนิด ได้แก่ ระหว่างคู่ภาษาไทยกับภาษาจีนว เมื่อลดค่าพีเอฟ Strident ค่าพีเอฟ Round และค่าพีเอฟ Low คู่ภาษาเออร์ดูกับภาษาจีนว เมื่อลดค่าพีเอฟ Continuant คู่ภาษาเบงกาลีกับภาษาจีนกวางตุ้ง เมื่อลดค่าพีเอฟ Continuant คู่ภาษาอารบิกกับภาษาจีนมิน เมื่อลดค่าพีเอฟ Continuant หรือ Consonantal คู่ภาษาอารบิกกับภาษาเบงกาลี เมื่อลดค่าพีเอฟ Strident หรือค่าพีเอฟ Consonantal และคู่ภาษาเบงกาลีกับภาษาจีนมินเมื่อลดค่าพีเอฟ Consonantal เป็นต้น

จากการเปรียบเทียบผลการทดลองของการจำแนกภาษาเมื่อใช้เวกเตอร์ค่าลักษณะเฉพาะที่ไม่มีค่าพีเอฟ (หรือใช้ค่าพีเอฟทุกชนิด) และเมื่อใช้เวกเตอร์ที่ลดค่าพีเอฟหนึ่งชนิด พบว่า ค่าความแม่นยำของการจำแนกภาษาที่ลดลงเพียงเล็กน้อย และไม่แตกต่างกันอย่างมีนัยสำคัญ เนื่องจากการใช้เวกเตอร์ค่าลักษณะเฉพาะที่ลดค่าพีเอฟหนึ่งชนิด ยังคงสามารถแสดงข้อมูลลักษณะการออกเสียงจากค่าพีเอฟอื่น ๆ ที่ไม่ถูกลดไปช่วยในการบอกความแตกต่างระหว่างภาษาได้ นั่นคือ ค่าพีเอฟชนิดอื่นที่อยู่ในเวกเตอร์สามารถชดเชยข้อมูลการออกเสียงของค่าพีเอฟชนิดที่ลดได้ ทำให้ตัวจำแนกยังคงความสามารถในการจำแนกความแตกต่างระหว่างภาษาได้ เช่น การใช้ข้อมูลลักษณะการออกเสียงในกลุ่มที่เกี่ยวข้องกับเสียงสระ เมื่อลดค่าพีเอฟ High ตัวจำแนกภาษายังสามารถใช้ข้อมูลจากค่าพีเอฟ Low ค่าพีเอฟ Back และค่าพีเอฟ Round ในการบอกค่าลักษณะการออกเสียงสระได้ เป็นต้น จึงทำให้เมื่อตัวจำแนกภาษาที่ใช้เวกเตอร์ค่าลักษณะเฉพาะที่ไม่รวมค่าพีเอฟ High แล้ว ยังคงสามารถจำแนกภาษาได้ดี อีกหนึ่งตัวอย่างได้แก่ ตัวจำแนกภาษาที่ใช้เวกเตอร์ค่าลักษณะเฉพาะที่ไม่รวมค่าพีเอฟ Voice ยังคงมีความสามารถในการบอกความแตกต่างระหว่างภาษาได้ดี ซึ่งอาจจะมีเหตุผลมาจากค่าพีเอฟ Voice เป็นค่าลักษณะสำคัญที่อธิบายแหล่งของเสียง มีคุณสมบัติคล้ายกันกับค่าพีเอฟ Vocalic และ Consonantal ซึ่งเป็นกลุ่มค่าลักษณะสำคัญที่อธิบายการปิดกั้นของกระแสลม ดังนั้นเมื่อใช้เวกเตอร์ค่าลักษณะเฉพาะที่ไม่รวมค่าพีเอฟ Voice เป็นค่าลักษณะสำคัญของตัวจำแนกภาษา ยังคงสามารถใช้ข้อมูลลักษณะการออกเสียงจากค่าพีเอฟอื่นแทนได้ ดังนั้นจากผลการทดลองนี้สรุปได้ว่าการลดค่าพีเอฟใดออกจากเวกเตอร์ค่าลักษณะเฉพาะจะไม่ส่งผลกระทบต่อความสามารถของการจำแนกให้ลดลงอย่างชัดเจน แต่มีข้อที่น่าสนใจสำหรับงานวิจัยต่อไป เช่น การศึกษาผลกระทบการใช้เวกเตอร์ค่าลักษณะเฉพาะที่ลดค่าพีเอฟบางชนิดต่อความสามารถของการจำแนกบางคู่ภาษาสามารถทำให้ประสิทธิภาพของตัวจำแนกภาษาดีขึ้นเมื่อมีการลดค่าพีเอฟ เป็นต้น

7.3.5 ผลการทดลองที่ 3.5: วัดประสิทธิภาพของระบบการรู้จำภาษาพูดที่นำเสนอเปรียบเทียบกับระบบอ้างอิงไอเวกเตอร์

ตารางที่ 7-14 แสดงค่าเฉลี่ยของเปอร์เซ็นต์ความผิดพลาดของการจำแนกภาษาของระบบที่นำเสนอ ซึ่งใช้การจำแนกด้วยแบบจำลองลาเทนท์ไดนามิกคอนดิชันนอลแรนดอมฟิลด์ร่วมกับค่าพีเอฟ เปรียบเทียบกับระบบการรู้จำภาษาพูดที่ใช้ไอเวกเตอร์ซึ่งใช้เป็นระบบอ้างอิงพบว่า ระบบที่นำเสนอให้ค่าเฉลี่ยของเปอร์เซ็นต์ความผิดพลาดของการจำแนกภาษามีค่า 17.46 ขณะที่ระบบอ้างอิงค่าเฉลี่ยของเปอร์เซ็นต์ความผิดพลาดของการจำแนกภาษามีค่า 19.94 ซึ่งสะท้อนผลการรู้จำภาษา

สำหรับแต่ละคู่ภาษาจากระบบที่นำเสนอให้ค่าเฉลี่ยของเปอร์เซ็นต์ความผิดพลาดของการจำแนกภาษาไม่แตกต่างจากระบบอ้างอิงแบบไอเวกเตอร์ เมื่อพิจารณาผลการจำแนกแต่ละคู่ภาษาพบว่า มีตัวจำแนก 21 คู่ภาษาที่ใช้ค่าพีเอฟให้ค่าความแม่นยำการจำแนกสูงกว่าระบบอ้างอิงที่ใช้ไอเวกเตอร์ และ 2 คู่ภาษาที่แตกต่างกันอย่างมีนัยสำคัญได้แก่ ระหว่างคู่ภาษาจีนกวางตุ้งกับภาษารัสเซีย และระหว่างคู่ภาษารัสเซียกับภาษาจีนวู เป็นต้น ในทางตรงกันข้าม มีตัวจำแนก 7 คู่ภาษาที่ใช้ค่าพีเอฟให้ค่าความแม่นยำการจำแนกต่ำกว่า ระบบที่ใช้ไอเวกเตอร์ แต่ไม่แตกต่างกันอย่างมีนัยสำคัญได้แก่ ระหว่างคู่ภาษาเบงกอลกับภาษารัสเซีย ระหว่างคู่ภาษาจีนมินกับภาษารัสเซีย ระหว่างคู่ภาษาไทยกับภาษารัสเซีย ระหว่างคู่ภาษาไทยกับภาษาจีนวู ระหว่างคู่ภาษาเออร์ดุกับภาษาจีนวู และระหว่างคู่ภาษาไทยกับภาษาเออร์ดุ จากการพิจารณาพบว่าค่าความแม่นยำการจำแนกของระบบที่นำเสนอไม่แตกต่างกันอย่างมีนัยสำคัญกับระบบอ้างอิง อย่างไรก็ตามประสิทธิภาพของระบบอ้างอิงยังไม่ดี ซึ่งผู้วิจัยควรทำการทดลองเพื่อปรับปรุงประสิทธิภาพให้เพิ่มขึ้น ถึงแม้ในการทดลองจะใช้ ALIZE ซึ่งได้รับความนิยมในการนำมาใช้มาสกัดไอเวกเตอร์ และได้ทดลองปรับค่าพารามิเตอร์เพื่อให้ได้ประสิทธิภาพ อย่างไรก็ตามวิทยานิพนธ์นี้ไม่ได้ให้ความสำคัญในการทดลองในขั้นตอนการปรับค่าพารามิเตอร์ที่เหมาะสมเนื่องจากในขั้นตอนนี้ต้องใช้เวลามากในการทดลองเพื่อปรับค่าพารามิเตอร์ จึงเสนอแนะสำหรับงานวิจัยต่อไปควรปรับค่าพารามิเตอร์ให้ประสิทธิภาพดีขึ้น นอกจากนี้ระบบที่นำเสนอมีค่าความแม่นยำสูงกว่าระบบอ้างอิง มีสาเหตุมาจาก 1) ข้อดีการใช้ไอเวกเตอร์เมื่อใช้กับฐานข้อมูลเสียงที่มีจำนวนมากเพียงพอ ซึ่งฐานข้อมูลมาจากทุกภาษาสำหรับฝึกฝนซูปเปอร์เวกเตอร์แบบจำลองยูบีเอ็ม และสำหรับปรับค่าพารามิเตอร์แบบจำลองยูบีเอ็มเป็นเวกเตอร์ที่ขึ้นกับภาษา และขึ้นกับช่องสัญญาณ แต่ในวิทยานิพนธ์นี้ใช้ฐานข้อมูลเสียงที่มีขนาดเล็ก ซึ่งอาจจะไม่เหมาะสมกับการใช้ไอเวกเตอร์ 2) ในการทดลองนี้ใช้วิธีการวัดความเหมือนระหว่างไอเวกเตอร์ของเสียงทดสอบกับแบบจำลองต้นแบบ ด้วยวิธีการวิเคราะห์จำแนกเชิงเส้นด้วยความน่าจะเป็น หรือแบบ pLDA ซึ่งเป็นแบบจำลองแบบเจนเนอเรทีฟ ซึ่งมีวิธีอื่น ๆ เพื่อวัดความเหมือนหรือการจำแนกบนสเปซของไอเวกเตอร์ ที่เป็นแบบจำลองแบบดิสคริมิเนทีฟเพื่อมาสร้างต้นแบบและจำแนกไอเวกเตอร์ได้ดีกว่าการใช้วิธีการวัดแบบเจนเนอเรทีฟ เช่น การใช้ซัพพอร์ตเวกเตอร์แมชชีน การวิเคราะห์การถดถอยโลจิสติก (Logistic Regression) เป็นต้น นอกจากนี้ผู้วิจัยมีข้อเสนอแนะให้ลองนำกระบวนการค้น (Paradigm) แบบไอเวกเตอร์ไปใช้ร่วมกับค่าพีเอฟที่นำเสนอ ในงานการรู้จำภาษา

ตารางที่ 7-14 ค่าเฉลี่ยเปอร์เซ็นต์ความผิดพลาดของการจำแนกภาษาของระบบที่
นำเสนอเปรียบเทียบกับระบบการรู้จำภาษาพูดที่ใช้ไอเวกเตอร์

แนวทาง	ค่าผิดพลาดของการจำแนก ภาษา (%)
LDCRF (PF)	17.46 *
i-vector	19.94

* แสดงค่าผิดพลาดที่น้อยที่สุด

7.3.6 ผลการทดลองที่ 3.6: วัดประสิทธิภาพของการรู้จำภาษาของระบบการรู้จำภาษาพูดที่ใช้ตัว จำแนกแบบหนึ่งต่อหนึ่ง และการใช้ตัวจำแนกแบบหนึ่งต่อทั้งหมด

เนื่องจากการวัดประสิทธิภาพของการทดลองที่ 3.1 ถึง 3.5 ดังตารางที่ 7-7 ถึงตารางที่ 7-14 เป็นค่าความผิดพลาดของตัวจำแนกที่ได้จากตัวจำแนกภาษาสำหรับแต่ละคู่ภาษา เพื่อประเมินประสิทธิภาพของการตัดสินใจการหาคำตอบของการระบุภาษาจากคะแนนโหวตของตัวจำแนกภาษาแบบหนึ่งต่อหนึ่ง จึงเปรียบเทียบกับผลการหาคำตอบของการระบุภาษาเป็นคำตอบการตัดสินใจแบบถาวรจากตัวจำแนกแบบหนึ่งต่อทั้งหมด ตารางที่ 7-15 แสดงความผิดพลาดของระบบการรู้จำภาษาพูดที่ใช้การจำแนกด้วยแบบจำลองลาเทนท์ไดนามิกคอนดิชันนอลแรนดอมฟิลด์ร่วมกับค่าพีเอฟ เมื่อใช้โครงสร้างของตัวจำแนกแบบต่าง ๆ พบว่าค่าเฉลี่ยเปอร์เซ็นต์ผิดพลาดของการจำแนกภาษาของระบบการรู้จำภาษาพูดที่ใช้คะแนนโหวตของตัวจำแนกแบบหนึ่งต่อหนึ่งให้ค่า 19.89 ซึ่งมีค่ามากกว่าระบบการรู้จำภาษาพูดที่ใช้ตัวจำแนกแบบหนึ่งต่อทั้งหมด ค่าความผิดพลาดของการตัดสินใจแบบถาวรที่ได้จากการโหวตของตัวจำแนกแบบหนึ่งต่อหนึ่งนี้มีค่าความผิดพลาดมากกว่า เนื่องจากความผิดพลาดของการตัดสินใจระบุภาษาจากตัวจำแนก 28 คู่ภาษา ที่โหวตจากตัวจำแนกที่เป็นตัวจำแนกที่ไม่ใช่คู่ภาษาที่ตรงกับข้อมูลทดสอบหรือเป็นภาษาที่พิจารณา ทำให้คำตอบที่ได้จากตัวจำแนกแบบหนึ่งต่อหนึ่งนี้จำเป็นต้องให้คำตอบของข้อมูลทดสอบ ดังนั้นคำตอบที่ได้จากตัวจำแนกดังกล่าวส่งผลให้มีความผิดพลาดได้ ทำให้ผลคะแนนโหวตของคำตอบส่วนใหญ่ผิดพลาดได้

ขณะที่การพิจารณาค่าความผิดพลาดการรู้จำภาษา EER ดังตารางที่ 7-15 พบว่าระบบการรู้จำภาษาพูดที่ใช้ LDCRF (PF) แบบหนึ่งต่อทั้งหมด มีค่า 8.49 ซึ่งต่ำกว่าระบบการรู้จำภาษาพูดที่ใช้ LDCRF (PF) แบบหนึ่งต่อหนึ่ง ดังนั้นเพื่อนำคะแนนการตัดสินใจระบุภาษาจากระบบการรู้จำภาษาพูดโดยใช้การจำแนกด้วยแบบจำลองลาเทนท์ไดนามิกคอนดิชันนอลแรนดอมฟิลด์ร่วมกับค่าพีเอฟไปรวมเข้ากับระบบรู้จำภาษาพูดแนวทางอื่น จึงใช้ระบบการรู้จำภาษาที่ใช้ LDCRF (PF) แบบหนึ่งต่อทั้งหมดสำหรับการตัดสินใจให้คะแนนการรู้จำภาษา และถูกนำไปใช้ในการรวมคะแนนการตัดสินใจของระบบที่นำเสนอต่อไป ในภาคผนวก ญ-3 แสดงแผนภาพกล่องของคะแนนที่ได้จากตัวจำแนกด้วยแบบจำลองลาเทนท์ไดนามิกคอนดิชันนอลแรนดอมฟิลด์แบบหนึ่งต่อทั้งหมดใช้ร่วมกับค่าพีเอฟ เมื่อใช้ข้อมูลชุดฝึกฝนจากแต่ละภาษา พบว่ากล่องของคะแนนของระบบการรู้จำภาษาจากข้อมูลฝึกฝนสำหรับภาษาเป้าหมาย การกระจายตัว มีค่าเฉลี่ยสูงกว่า และไม่ซ้อนทับกับคะแนนที่ได้จากข้อมูลฝึกฝนจากภาษาอื่นซึ่งส่วนใหญ่มีค่าเฉลี่ยต่ำกว่า 0.1

ตารางที่ 7-15 ค่าความผิดพลาดการรู้จำภาษา EER และค่าความผิดพลาดของการ
จำแนกภาษาของระบบการรู้จำภาษาพูดที่ใช้ตัวจำแนกแบบหนึ่งต่อหนึ่ง และการใช้
ตัวจำแนกแบบหนึ่งต่อทั้งหมด

แนวทาง	ค่าเฉลี่ยความผิดพลาดของการ จำแนกภาษา (%)	ค่าความผิดพลาดการ รู้จำภาษา EER
ระบบการรู้จำภาษาพูดใช้ LDCRF	19.89	11.98

(PF) แบบหนึ่งต่อหนึ่ง		
ระบบการรู้จำภาษาพูดใช้ LDCRF (PF) แบบหนึ่งต่อทั้งหมด	13.24 *	8.49 *

* แสดงค่าผิดพลาดที่น้อยที่สุด

นอกจากนี้ผลการวัดประสิทธิภาพของระบบการรู้จำภาษาพูดโดยใช้การจำแนกด้วยแบบจำลองลาเทนท์ไดนามิคคอนดิชันนอลแรนดอมฟีลด์ร่วมกับค่าพีเอฟ ซึ่งตัวจำแนกเป็นแบบหนึ่งต่อทั้งหมด โดยทดสอบการจำแนกภาษาบนฐานข้อมูลเสียง NIST LRE07 เซตข้อมูลทดสอบ แสดงดังตารางที่ 7-15 พบว่าระบบการรู้จำภาษาพูดการจำแนกด้วยแบบจำลองลาเทนท์ไดนามิคคอนดิชันนอลแรนดอมฟีลด์ร่วมกับค่าพีเอฟให้ค่า C_{avg} มีค่า 0.0861 ดังแสดงในตารางที่ 7-16 ซึ่งนำผลลัพธ์ของระบบนี้ไปเปรียบเทียบกับระบบที่นำเสนอเมื่อรวมคะแนนจากการรู้จำภาษาจากแนวทางนี้ต่อไป

ตารางที่ 7-16 ค่า C_{avg} ของระบบการรู้จำภาษาพูดการจำแนกด้วยแบบจำลองลาเทนท์ไดนามิคคอนดิชันนอลแรนดอมฟีลด์ร่วมกับค่าพีเอฟ บนฐานข้อมูลเซตทดสอบ 8 ภาษา

ระบบการรู้จำภาษาพูด	ค่า $100 \times C_{avg}$
ระบบการรู้จำภาษาพูดใช้ LDCRF (PF)	8.61

7.3.7 บทสรุปการรู้จำภาษาโดยใช้การจำแนกด้วยแบบจำลองลาเทนท์ไดนามิคคอนดิชันนอลแรนดอมฟีลด์ร่วมกับค่าพีเอฟ

จากผลการทดลองของการรู้จำภาษาโดยใช้การจำแนกด้วยแบบจำลองลาเทนท์ไดนามิคคอนดิชันนอลแรนดอมฟีลด์ร่วมกับค่าพีเอฟ พบว่าเปอร์เซ็นต์ความแม่นยำการจำแนกภาษาของระบบการรู้จำภาษาพูดที่เสนอดีกว่าระบบการรู้จำภาษาพูดตามแนวทางอะคูสติกแบบใช้ไอเวกเตอร์ ซึ่งเป็นระบบอ้างอิงที่ 3 ที่ใช้เปรียบเทียบประสิทธิภาพของระบบที่นำเสนอสำหรับแต่ละแนวทาง เมื่อเปรียบเทียบความสามารถของการจำแนกของค่าพีเอฟกับค่าลักษณะสำคัญทางเสียงอื่น เช่น ค่าสัมประสิทธิ์ MFCC ค่าสัมประสิทธิ์ PLP และค่าเอสดีซีพี พบว่าตัวจำแนกภาษาที่ใช้ค่าพีเอฟ ซึ่งผลการทดลองสอดคล้องกับการทดลองของระบบการรู้จำภาษาพูดที่ใช้การจำแนกภาษาด้วยแบบจำลองเกาส์เซียน (ดังการทดลองที่ 3.3) ซึ่งมีเปอร์เซ็นต์ความผิดพลาดของการจำแนกภาษาเมื่อใช้ค่าพีเอฟต่ำกว่าการใช้ค่าสัมประสิทธิ์ PLP มีค่าความแม่นยำของการจำแนกดีกว่าการใช้ค่าลักษณะสำคัญชนิดอื่น นอกจากนี้เมื่อใช้การจำแนกด้วยแบบจำลองลาเทนท์ไดนามิคคอนดิชันนอลแรนดอมฟีลด์ เพื่อเรียนรู้รูปแบบของลำดับแบบไดนามิคของค่าพีเอฟ พบว่าค่าเปอร์เซ็นต์ความผิดพลาดการจำแนกภาษาเมื่อใช้การจำแนกด้วยแบบจำลองลาเทนท์ไดนามิคคอนดิชันนอลแรนดอมฟีลด์นั้นให้ค่าต่ำกว่าการใช้แบบจำลองคอนดิชันนอลแรนดอมฟีลด์ และแบบจำลองเกาส์เซียน

7.4 ประเมินประสิทธิภาพของการรู้จำภาษาตามแนวทางการเรียงหน่วยเสียงโดยใช้คำพืดอ

ผลการทดลองที่ 4 เพื่อวัดประสิทธิภาพการรู้จำภาษาแบบเรียงหน่วยเสียงโดยพิจารณาโคออกเคอร์เรนซ์คำพืดอแทนหน่วยเสียง ซึ่งวิธีการทดลองของแต่ละการทดลองอธิบายในหัวข้อ 6.4.4 มีผลการทดลองดังต่อไปนี้

7.4.1 ผลการทดลองที่ 4.1: ประเมินผลประสิทธิภาพของการรู้จำภาษาแนวทางการเรียงหน่วยเสียงที่ใช้คำพืดอโดยใช้การพิจารณาโคออกเคอร์เรนซ์คำพืดอ

การทดลองนี้หาตัวแทนของเหตุการณ์ที่เกิดขึ้นร่วมกันของคำพืดอที่เหมาะสมในการจำแนกภาษา ซึ่งในวิทยานิพนธ์ใช้การพิจารณาโคออกเคอร์เรนซ์คำพืดอแทนการปรากฏร่วมกันของคำพืดอตารางที่ 7-17 แสดงค่าเฉลี่ยเปอร์เซ็นต์ความแม่นยำของการจำแนกภาษาจาก 28 คู่ภาษาเมื่อพิจารณาโคออกเคอร์เรนซ์คำพืดอจากการจับคู่คำพืดอสองชนิด และเมื่อพิจารณาโคออกเคอร์เรนซ์คำพืดอจากการจับคู่คำพืดอจากทุกชนิดคำพืดอ โดยใช้การพิจารณาที่มีกรอบการพิจารณาขนาด 2 เฟรม

ผลการทดลองที่ 4.1.1 แสดงดังตารางที่ 7-17 พบว่าค่าเฉลี่ยเปอร์เซ็นต์ความแม่นยำของการจำแนกภาษาเมื่อพิจารณาโคออกเคอร์เรนซ์คำพืดอจากการจับคู่คำพืดอสองชนิด โดยกำหนดการพิจารณาหนึ่งคำพืดอหลักปรากฏร่วมกับคำพืดออื่น ๆ มีค่าเท่ากับ 71.50 ซึ่งโคออกเคอร์เรนซ์คำพืดอจากการพิจารณาคำพืดอ Vocalic เป็นคำพืดอหลักร่วมกับคำพืดออื่น ๆ ให้ค่าเฉลี่ยเปอร์เซ็นต์ความแม่นยำของการจำแนกภาษาสูงที่สุดคือ 78.19 ขณะที่เมื่อใช้โคออกเคอร์เรนซ์คำพืดอจากการพิจารณาคำพืดอ Anterior เป็นคำพืดอหลักร่วมกับคำพืดออื่น ๆ ให้ค่าเฉลี่ยเปอร์เซ็นต์ความแม่นยำของการจำแนกภาษาต่ำที่สุดคือ 69.56 โดยแต่ละแถวของตารางที่ 7-17 แสดงค่าเฉลี่ยเปอร์เซ็นต์ความแม่นยำของการจำแนกเมื่อพิจารณาคำพืดอชนิดนั้น ๆ เป็นคำพืดอหลักปรากฏร่วมกับคำพืดออื่น ๆ ในการหาโคออกเคอร์เรนซ์คำพืดอ

ผลการทดลองที่ 4.1.2 พบว่าค่าเฉลี่ยเปอร์เซ็นต์ความแม่นยำของการจำแนกภาษาเมื่อพิจารณาโคออกเคอร์เรนซ์คำพืดอจากการจับคู่คำพืดอจากทุกชนิดคำพืดอ ให้ค่าความแม่นยำ 60.84 ซึ่งมีค่าต่ำกว่าการพิจารณาโคออกเคอร์เรนซ์คำพืดอจากการจับคู่คำพืดอสองชนิดที่พิจารณาหนึ่งคำพืดอหลักปรากฏร่วมกับคำพืดออื่น ๆ โดยค่าเฉลี่ยเปอร์เซ็นต์ความแม่นยำของการจำแนกภาษาเมื่อพิจารณาโคออกเคอร์เรนซ์คำพืดอจากการจับคู่คำพืดอสองชนิด โดยกำหนดการพิจารณาหนึ่งคำพืดอหลักปรากฏร่วมกับคำพืดออื่น ๆ มีค่าเฉลี่ย 71.50 ซึ่งความแม่นยำของการจำแนกเมื่อใช้โคออกเคอร์เรนซ์คำพืดอจากการจับคู่คำพืดอสองชนิด โดยกำหนดการพิจารณาหนึ่งคำพืดอหลักปรากฏร่วมกับคำพืดออื่น ๆ ให้ผลการจำแนกเพิ่มขึ้นประมาณร้อยละ 17.52

ตารางที่ 7-17 ค่าเฉลี่ยเปอร์เซ็นต์ความแม่นยำของการจำแนกภาษาจาก 28 คู่ภาษา
เมื่อพิจารณาโคออคเคอร์เรนซ์ค่าพีเอฟจากการจับคู่ค่าพีเอฟ

ชนิดค่าพีเอฟที่กำหนดให้พิจารณา เป็นค่าพีเอฟหลักในโคออคเคอร์เรนซ์ ค่าพีเอฟ (การทดลองที่ 4.1.1)	ค่าความแม่นยำ
Vocalic	78.19
Continuant	72.12
Consonantal	72.00
Voice	71.39
High	71.36
Strident	70.92
Nasal	70.88
Back	70.84
Coronal	70.78
Low	70.65
Round	70.62
Tense	70.23
Anterior	69.56
ค่าเฉลี่ย	71.50
ค่าพีเอฟการจับคู่ค่าพีเอฟจากทุกชนิด ค่าพีเอฟ ในการหาโคออคเคอร์เรนซ์ ค่าพีเอฟ (การทดลองที่ 4.1.2)	60.84

7.4.2 ผลการทดลองที่ 4.2: ศึกษาผลกระทบของขนาดกรอบการพิจารณาของการปรากฏร่วมกัน ของโคออคเคอร์เรนซ์ค่าพีเอฟต่อประสิทธิภาพการรู้จำภาษาแนวทางการเรียงหน่วยเสียงโดยใช้ ค่าพีเอฟ

เนื่องจากในวิทยานิพนธ์นี้เสนอการใช้ขนาดกรอบหน้าต่างที่พิจารณาการเกิดเหตุการณ์
ร่วมกันของสมาชิกสองตัว ซึ่งเป็นเหตุการณ์ที่เกิดขึ้นร่วมกันของสองค่าพีเอฟ ดังนั้นในการทดลองนี้
เปรียบเทียบความสามารถของการจำแนกภาษาเมื่อใช้ขนาดกรอบการพิจารณา และช่วงการเลื่อน
กรอบพิจารณาที่แตกต่างกัน ความถูกต้องของการจำแนกเมื่อใช้ขนาดกรอบการพิจารณาที่แตกต่างกัน

กัน แสดงดังตารางที่ 7-18 พบว่า ตัวจำแนกภาษาที่ใช้กรอบการพิจารณาการเกิดร่วมกันของพีเอฟที่มีขนาด 3 เฟรม (แทนด้วย win3) ให้ค่าเฉลี่ยความแม่นยำการจำแนกเป็นร้อยละ 72.40 ซึ่งมีค่าสูงสุด แต่อย่างไรก็ตามความถูกต้องของการจำแนกเมื่อใช้กรอบการพิจารณาขนาดที่ต่างกัน ให้ค่าความถูกต้องของการจำแนกไม่แตกต่างกันอย่างมีนัยสำคัญ

การทดลองนี้ใช้กรอบการพิจารณาเพื่อหารูปแบบการเกิดร่วมกันของค่าพีเอฟมีขนาด 2 เฟรม 3 เฟรม 4 เฟรม และ 5 เฟรม ซึ่งกรอบการพิจารณาขนาด 3 ถึง 5 เฟรมนี้ ใช้พิจารณาเหตุการณ์ของโคออกเคอเรนซ์ที่เกิดขึ้นร่วมกันสองเฟรมภายในกรอบหน้าต่างเดียวกัน และยอมให้มีช่องว่างเกิดขึ้นระหว่างสองเหตุการณ์ได้ ผลการจำแนกเมื่อใช้กรอบการพิจารณาขนาด 3 เฟรม 4 เฟรม และ 5 เฟรม พบว่า ผลการทดลองส่วนใหญ่ให้ค่าความแม่นยำสูงกว่าการใช้กรอบการพิจารณาขนาด 2 เฟรม ซึ่งอาจจะมีสาเหตุมาจากการเกิดร่วมกันของสองเหตุการณ์โดยยอมให้มีช่องว่างเกิดขึ้นระหว่างเหตุการณ์ได้ ทำให้ได้ข้อมูลที่มีรูปแบบที่มีความยืดหยุ่น และทำให้หลีกเลี่ยงสิ่งรบกวนที่เกิดขึ้นในข้อมูลฝึกฝนได้

ภาพที่ 7-5 แสดงแผนภาพกล่องของค่าความแม่นยำของการจำแนกภาษาจากตัวจำแนกคู่ภาษา 28 คู่เมื่อใช้ขนาดกรอบการพิจารณาแตกต่างกัน แต่ละกล่องแสดงค่าความแม่นยำของตัวจำแนกภาษาที่ใช้กรอบการพิจารณาหารูปแบบเหตุการณ์ที่เกิดขึ้นร่วมกันของโคออกเคอเรนซ์ค่าพีเอฟ เมื่อมีขนาดกรอบการพิจารณาแตกต่างกัน กล่องของกรอบการพิจารณาที่มีขนาด 3 เฟรมอยู่ตำแหน่งสูงกว่ากล่องของขนาดกรอบการพิจารณาค่าอื่น แต่ไม่แตกต่างกันอย่างสังเกตเห็นชัด และค่าเฉลี่ยของผลการจำแนกของกรอบหน้าต่างที่มีขนาด 3 เฟรมที่สูงกว่าขนาดกรอบพิจารณาค่าอื่นนี้ สะท้อนว่าผลการจำแนกเมื่อใช้กรอบหน้าต่างที่มีขนาด 3 เฟรมให้ผลดีที่สุด ผู้วิจัยจึงเลือกกรอบหน้าต่างที่มีขนาด 3 เฟรมเป็นขนาดกรอบการพิจารณาของในการทดลองที่ 4.3 ถึง 4.5 ต่อไป

ตารางที่ 7-18 ค่าเฉลี่ยของค่าความแม่นยำของการจำแนกภาษาจาก 28 คู่ภาษา ที่ใช้การปรากฏร่วมกันของการจับคู่ค่าพีเอฟ เมื่อใช้ขนาดกรอบการพิจารณาที่แตกต่างกัน

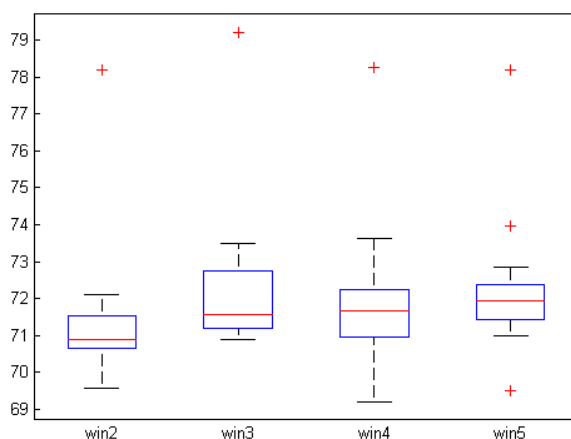
ค่าพีเอฟที่กำหนดเป็นค่าพีเอฟหลักปรากฏร่วมกับค่าพีเอฟอื่น	ขนาดกรอบการพิจารณา			
	win2	win3	win4	win5
Vocalic	78.19	79.19	78.24	78.17
Continuant	72.12	72.66	72.65	72.84
Consonantal	72.00	73.50	73.63	73.96
Voice	71.39	72.10	72.09	72.24
High	71.36	71.98	71.71	71.98
Strident	70.92	71.55	71.04	71.48
Nasal	70.88	70.89	70.72	70.98
Back	70.84	71.22	71.65	71.95
Coronal	70.78	71.38	71.11	71.54
Low	70.65	71.48	71.72	72.04
Round	70.62	71.13	71.02	71.47
Tense	70.23	71.13	70.8	71.28
Anterior	69.56	73.00	69.21	69.49
ค่าเฉลี่ย	71.50	72.40	71.97	72.26

win2 แทนกรอบการพิจารณาที่มีขนาด 2 เฟรม

win3 แทนกรอบการพิจารณาที่มีขนาด 3 เฟรมและช่วงการเลื่อนกรอบทีละ 2 เฟรม

win4 แทนกรอบการพิจารณาที่มีขนาด 4 เฟรมและช่วงการเลื่อนกรอบทีละ 3 เฟรม

win5 แทนกรอบการพิจารณาที่มีขนาด 5 เฟรมและช่วงการเลื่อนกรอบทีละ 3 เฟรม



ภาพที่ 7-5 แสดงแผนภาพกล่องของค่าความถูกต้องของการจำแนกภาษาจากตัว
จำแนกคู่ภาษา 28 คู่ภาษาเมื่อใช้ขนาดกรอบการพิจารณาแตกต่างกัน

7.4.3 ผลการทดลองที่ 4.3: ศึกษาผลกระทบของชนิดคำพืเอฟที่พิจารณาการปรากฏร่วมกันของ คำพืเอฟต่อประสิทธิภาพการรู้จำภาษาแนวทางการเรียงหน่วยเสียงโดยใช้คำพืเอฟ

ผลการทดลองที่ 4.3 เพื่อเปรียบเทียบชนิดของคำพืเอฟที่กำหนดเป็นคำพืเอฟหลักในการพิจารณาเหตุการณ์ที่เกิดขึ้นร่วมกันของคำพืเอฟสองคำ แสดงดังตารางที่ 7-18 พบว่าค่าความแม่นยำของการจำแนกภาษาเท่ากับ 79.19 เมื่อใช้คำพืเอฟ Vocalic เป็นคำพืเอฟหลักที่ใช้ในการพิจารณา รูปแบบเหตุการณ์ที่เกิดขึ้นร่วมกัน ซึ่งมีค่าความแม่นยำสูงกว่าการใช้คำพืเอฟชนิดอื่น

ค่าความแม่นยำของการจำแนกแต่ละคู่ภาษาที่ใช้คำพืเอฟ Vocalic เป็นคำพืเอฟหลักที่ใช้ในการพิจารณา รูปแบบเหตุการณ์ที่เกิดขึ้นร่วมกันคำพืเอฟ มีตัวจำแนกเพียง 3 คู่ภาษาที่ให้ค่าเปอร์เซ็นต์ความแม่นยำการจำแนกต่ำกว่า 70 ได้แก่ คู่ภาษาจีนกวางตุ้งกับภาษาจีนมิน มีค่าเปอร์เซ็นต์ความแม่นยำ 63.82 คู่ภาษาจีนกวางตุ้งกับภาษาจีนจูมีค่าเปอร์เซ็นต์ความแม่นยำ 63.86 และคู่ภาษาจีนกวางตุ้งกับภาษาไทยมีค่าเปอร์เซ็นต์ความแม่นยำ 64.25 ตัวจำแนกทั้งสามคู่ภาษานี้จำแนกภาษาได้ไม่ดีนัก เนื่องจากทั้งสามอยู่ในตระกูลภาษาเดียวกันทำให้มีค่าลักษณะการออกเสียง และมีลักษณะเฉพาะภาษาร่วมกัน ทำให้ผลการแยกแยะระหว่างภาษาที่คุณสมบัติคล้ายกันจำแนกไม่ได้ นอกจากนี้จากการเปรียบเทียบของการใช้คำโคออกเคอร์เรนซ์คำพืเอฟชนิดต่าง ๆ ที่แตกต่างกัน ทำให้สามารถปรับปรุงผลการจำแนกที่ดีขึ้นสำหรับแต่ละคู่ภาษา โดยการเลือกคำพืเอฟที่มีคุณสมบัติลักษณะเฉพาะของภาษา และนำมาทำให้ประสิทธิภาพของการจำแนกภาษาดีขึ้นต่อไป

7.4.4 ผลการทดลองที่ 4.4: ศึกษาผลกระทบของการเพิ่มจำนวนโคออกเคอร์เรนซ์คำพืเอฟต่อ ประสิทธิภาพการรู้จำภาษาแนวทางการเรียงหน่วยเสียงคำพืเอฟ

ในการทดลองที่ 4.4 ศึกษาผลกระทบของจำนวนคำพืเอฟที่นำมาประกอบกันโดยเพิ่มจำนวนคำพืเอฟที่พิจารณาการเกิดร่วมกันโคออกเคอร์เรนซ์คำพืเอฟ 2 โคออกเคอร์เรนซ์คำพืเอฟภายในแต่

ละเฟรม ซึ่งการเพิ่มจำนวนโคออกเคอร์เรนซ์ค่าพีเอฟทำให้ได้รูปแบบที่พิจารณาการเกิดร่วมกันของ การออกเสียงหลายชนิดที่ปรากฏในแต่ละเฟรมเพิ่มขึ้น ตารางที่ 7-19 แสดงค่าเปอร์เซ็นต์ความ แม่นยำของการจำแนกภาษาเมื่อใช้การพิจารณาการเกิดร่วมกันโคออกเคอร์เรนซ์ค่าพีเอฟ 2 โคออก เคอร์เรนซ์ค่าพีเอฟในแต่ละเฟรม พบว่า เมื่อพิจารณาการเกิดร่วมกันโคออกเคอร์เรนซ์ค่าพีเอฟ 2 โค ออกเคอร์เรนซ์ค่าพีเอฟ ที่มีสมาชิกหนึ่งตัวมาจากโคออกเคอร์เรนซ์ค่าพีเอฟ Continuant ที่กำหนด เป็นค่าพีเอฟหลักที่ใช้ในการพิจารณารูปแบบเหตุการณ์ที่เกิดขึ้นร่วมกัน หรือแทนด้วย COPF(Continuant, *) ให้ค่าเฉลี่ยเปอร์เซ็นต์ความแม่นยำของการจำแนกเป็น 76.57 ซึ่งสูงกว่าการ ปรากฏร่วมกันของโคออกเคอร์เรนซ์ค่าพีเอฟ 2 โคออกเคอร์เรนซ์ค่าพีเอฟที่มีสมาชิกแบบอื่น

นอกจากนี้เมื่อเปรียบเทียบผลการพิจารณาในแต่ละเฟรมเมื่อเพิ่มการเกิดร่วมกันโคออกเคอร์ เรนซ์ค่าพีเอฟ 2 โคออกเคอร์เรนซ์ค่าพีเอฟ เทียบกับผลการทดลองของการพิจารณาโคออกเคอร์เรนซ์ ค่าพีเอฟตามแบบการทดลองที่ 4.4.1 โดยเลือกผลการทดลองจากตารางที่ 7-18 ที่ใช้โคออกเคอร์ เรนซ์ค่าพีเอฟที่ใช้ค่าพีเอฟ Vocalic กำหนดเป็นค่าพีเอฟหลักเพราะให้ค่าเปอร์เซ็นต์ความแม่นยำการ จำแนกสูงสุดคือ 79.19 พบว่ามี 4 กรณีของการเพิ่มการเกิดร่วมกันโคออกเคอร์เรนซ์ค่าพีเอฟ 2 โค ออกเคอร์เรนซ์ค่าพีเอฟที่ให้ค่าเปอร์เซ็นต์ความแม่นยำการจำแนกสูงกว่าการพิจารณาโคออกเคอร์ เรนซ์ค่าพีเอฟตามแบบการทดลองที่ 4.4.1 เมื่อใช้โคออกเคอร์เรนซ์ค่าพีเอฟที่มาจาก Vocalic ซึ่ง 4 กรณี ได้แก่ โคออกเคอร์เรนซ์ค่าพีเอฟ Consonantal-* , ร่วมกับโคออกเคอร์เรนซ์ค่าพีเอฟ Voice-* ซึ่งให้ค่าเปอร์เซ็นต์ความแม่นยำ 79.74 , โคออกเคอร์เรนซ์ค่าพีเอฟ Vocalic -* , ร่วมกับโคออกเคอร์ เรนซ์ค่าพีเอฟ Continuant -* ซึ่งให้ค่าเปอร์เซ็นต์ความแม่นยำ 79.55, โคออกเคอร์เรนซ์ค่าพีเอฟ Voice -* , ร่วมกับโคออกเคอร์เรนซ์ค่าพีเอฟ Continuant -* ซึ่งให้ค่าเปอร์เซ็นต์ความแม่นยำ 79.55, และ โคออกเคอร์เรนซ์ค่าพีเอฟ Vocalic-* , ร่วมกับโคออกเคอร์เรนซ์ค่าพีเอฟ Voice-* ซึ่งให้ ค่าเปอร์เซ็นต์ความแม่นยำ 79.53 เป็นต้น ซึ่งผลการทดลองที่ทำให้การพิจารณาจากการเกิดร่วมกัน ของโคออกเคอร์เรนซ์ค่าพีเอฟ 2 โคออกเคอร์เรนซ์ค่าพีเอฟทำให้รูปแบบของการออกเสียงเพิ่มขึ้น ส่งผลให้สามารถนำไปใช้ในการบอกลักษณะเฉพาะของภาษาได้เพิ่มขึ้น และเป็นที่น่าสังเกตว่า จาก 4 กรณีนี้มาจากการพิจารณาค่าพีเอฟที่มีความสัมพันธ์ในกลุ่มการออกเสียงเดียวกันจึงอาจทำให้เพิ่ม ข้อมูลการออกเสียงในการยืนยันลักษณะเฉพาะของแต่ละภาษาได้เพิ่มขึ้น

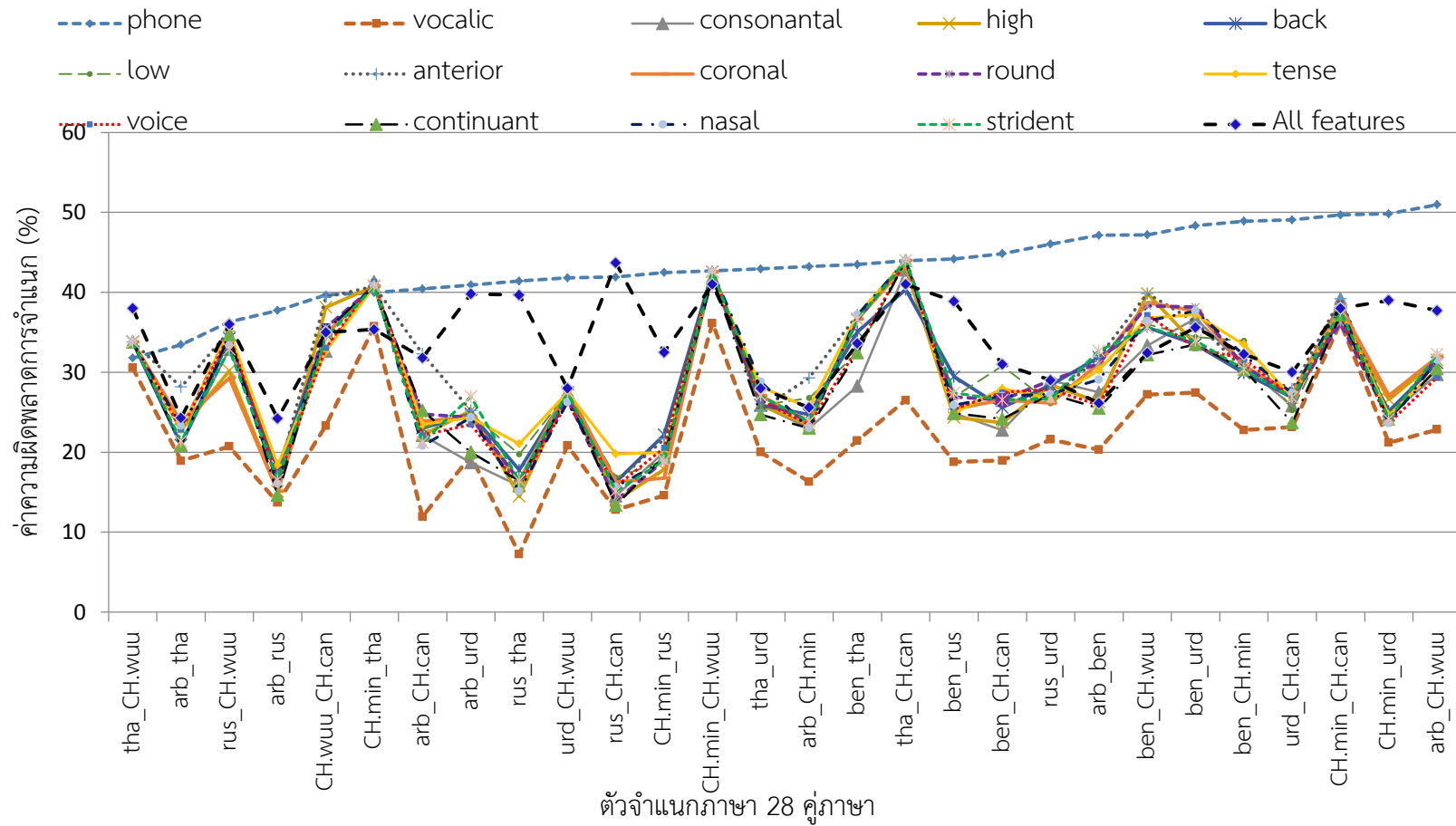
7.4.5 ผลการทดลองที่ 4.5: เปรียบเทียบประสิทธิภาพการรู้จำภาษาจากระบบที่นำเสนอกับผล การรู้จำภาษาของระบบอ้างอิงตามแนวทางการเรียงหน่วยเสียง

ภาพที่ 7-6 แสดงค่าความผิดพลาดของการจำแนกภาษาของ 28 คู่ภาษาของระบบการรู้จำ ภาษาพูดตามแนวทางการเรียงหน่วยเสียง เมื่อพิจารณาจากโคออกเคอร์เรนซ์ค่าพีเอฟตามแบบการ ทดลองที่ 4.4.1 เปรียบเทียบกับการพิจารณาจากหน่วยเสียงพบว่า แนวโน้มค่าผิดพลาดของการ จำแนกมีรูปแบบเดียวกัน ความผิดพลาดของการจำแนกจากโคออกเคอร์เรนซ์ค่าพีเอฟต่ำกว่าการใช้ หน่วยเสียง และค่าความผิดพลาดของการจำแนกเมื่อใช้โคออกเคอร์เรนซ์ค่าพีเอฟที่มาจาก การพิจารณาจากโคออกเคอร์เรนซ์ค่าพีเอฟ Vocalic กำหนดเป็นค่าพีเอฟหลักที่ใช้ในการพิจารณา รูปแบบเหตุการณ์ที่เกิดขึ้นร่วมกัน ให้ค่าความผิดพลาดต่ำที่สุดสำหรับการจำแนกภาษาของ 28 คู่ ภาษา

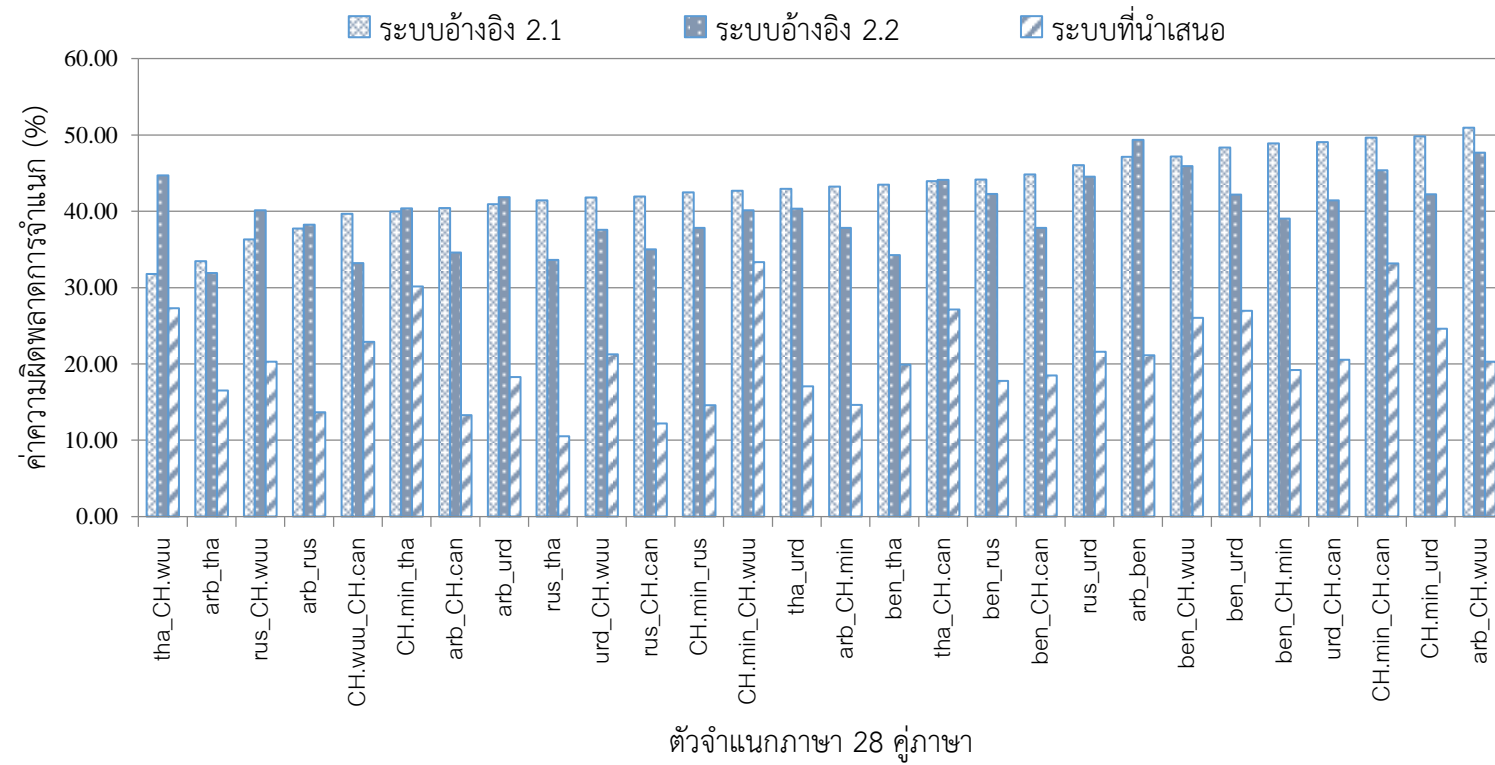
จากผลการทดลองที่ 4.1 ถึง 4.4 พบว่า พารามิเตอร์และผลการทดลองที่ให้ค่าความแม่นยำของการจำแนกที่ดีที่สุดคือ คือการใช้โคออคเคอเรนซ์ค่าพีเอฟ Vocalic เพื่อกำหนดเป็นค่าพีเอฟหลักในการพิจารณาเหตุการณ์ที่เกิดขึ้นร่วมกันของค่าพีเอฟสองค่า และเมื่อใช้ขนาดกรอบการพิจารณาเพื่อหาความถี่ของโคออคเคอเรนซ์ค่าพีเอฟขนาด 3 เฟรมให้ค่าความแม่นยำการจำแนกที่ดีที่สุด จึงเลือกนำมาใช้เป็นค่าพารามิเตอร์ของระบบที่นำเสนอ เพื่อเปรียบเทียบประสิทธิภาพของระบบที่นำเสนอกับระบบอ้างอิงที่ 2.1 และระบบอ้างอิงที่ 2.2 ซึ่งภาพที่ 7-7 แสดงค่าความผิดพลาดของการจำแนกภาษาของ 28 คู่ภาษาแนวทางการเรียงหน่วยเสียงเมื่อใช้โคออคเคอเรนซ์ค่าพีเอฟ เปรียบเทียบกับระบบการรู้จำภาษาพูดแบบพ็อดาร์วีเอส ซึ่งเป็นระบบอ้างอิงที่ 2.1 และ ระบบการรู้จำภาษาพูดที่ใช้ค่าพีเอฟ [92] ซึ่งเป็นระบบอ้างอิงที่ 2.2 พบว่าค่าความผิดพลาดของการจำแนกภาษาจากระบบที่นำเสนอ มีค่าต่ำกว่าระบบอ้างอิงที่ 2.1 และ 2.2 ในตัวจำแนกทั้ง 28 คู่ภาษา

ตารางที่ 7-19 ค่าความแม่นยำการจำแนกภาษาเมื่อใช้พิจารณาการเกิดร่วมกันโคออคเคอร์เรนซ์ค่าพีเอฟ 2 โคออคเคอร์เรนซ์ค่าพีเอฟ

	Vocalic	Continuant	consonant	Voice	High	Strident	Nasal	Back	Coronal	Low	Round	Tense	Anterior
Vocalic		79.55	72.18	79.53	76.33	77.47	78.02	73.78	75.87	75.93	77.19	74.45	77.89
Continuant	79.55		74.62	79.55	75.27	76.90	76.11	74.38	78.67	74.72	77.67	74.77	76.60
consonant	72.18	74.62		79.74	74.38	74.12	74.09	73.02	73.72	74.2	74.14	74.71	75.18
Voice	79.53	79.55	79.74		75.1	77.35	76.84	75.39	75.57	74.79	73.67	75.02	74.43
High	76.33	75.27	74.38	75.1		74.76	77.26	73.88	74.99	75.81	76.42	77.23	76.88
Strident	77.47	76.9	74.12	77.35	74.76		74.47	74.37	75.73	75.4	75.03	74.58	74.64
Nasal	78.02	76.11	74.09	76.84	77.26	74.47		74.01	75.93	74.96	75.06	74.11	75.03
Back	73.78	74.38	73.02	75.39	73.88	74.37	74.01		73.43	75.71	73.76	74.15	76.15
Coronal	75.87	78.67	73.72	75.57	74.99	75.73	75.93	73.43		74.89	74.52	74.7	74.3
Low	75.93	74.72	74.2	74.79	75.81	75.4	74.96	75.71	74.89		75.12	72.46	73.41
Round	77.19	77.67	74.14	73.67	76.42	75.03	75.06	73.76	74.52	75.12		75.47	74.68
Tense	74.45	74.77	74.71	75.02	77.23	74.58	74.11	74.15	74.7	72.46	75.47		74.09
Anterior	77.89	76.6	75.18	74.43	76.88	74.64	75.03	76.15	74.3	73.41	74.68	74.09	
ค่าเฉลี่ย	76.52	76.57	74.51	76.41	75.69	75.4	75.49	74.34	75.19	74.78	75.23	74.65	75.27



ภาพที่ 7-6 ค่าความผิดพลาดของการจำแนกภาษาของ 28 คู่ภาษาแนวทางการเรียงหน่วยเสียง
เมื่อใช้โคออร์ดิเนตค่าพิตช์ เปรียบเทียบกับการใช้หน่วยเสียง



ภาพที่ 7-7 ค่าความผิดพลาดของการจำแนกภาษาของ 28 คู่ภาษาแนวทางการเรียงหน่วยเสียงเมื่อใช้โคออคเคอเรนซ์ค่าพีเอฟ
เปรียบเทียบกับระบบอ้างอิงที่ 2.1 และระบบอ้างอิงที่ 2.2

ภาคผนวก ญ-2 คะแนนการตัดสินใจโดยใช้ระบบการรู้จำภาษาพูดแนวทางการเรียงตัวของเสียงโดยใช้ค่าพีเอฟ เมื่อใช้ข้อมูลชุดฝึกฝนจากแต่ละภาษา พบว่ากล่องของคะแนนที่ได้จากการจำแนกด้วยซัพพอร์ตเวกเตอร์แมชชีนแบบหนึ่งต่อทั้งหมดที่ใช้ข้อมูลฝึกฝนสำหรับภาษาเป้าหมาย มีค่าเฉลี่ยสูงกว่า 0 และไม่ซ้อนทับกับกล่องของคะแนนที่ได้การจำแนกด้วยซัพพอร์ตเวกเตอร์แมชชีนเมื่อใช้ข้อมูลฝึกฝนของภาษาอื่นที่มีค่าเฉลี่ยต่ำกว่า -1

ตารางที่ 7-20 แสดงค่าเฉลี่ยค่าความแม่นยำของการจำแนกภาษาที่ใช้ และค่าความผิดพลาด EER ของระบบที่นำเสนอเปรียบเทียบกับระบบอ้างอิงที่ 2.1 และ 2.2 พบว่า ระบบที่นำเสนอที่ใช้โคออคเคอเรนซ์ค่าพีเอฟมีค่าความผิดพลาดการรู้จำภาษา EER 6.80 ซึ่งต่ำกว่าระบบอ้างอิงที่ 2.1 และระบบอ้างอิง 2.2 ที่มีค่าความผิดพลาด EER เป็น 8.61 และ 8.82 ตามลำดับ นอกจากนี้เมื่อพิจารณาตามขนาดเวกเตอร์ที่ใช้ในแบบจำลองเวกเตอร์สเปซเพื่อเป็นอินพุตให้กับตัวจำแนกซัพพอร์ตเวกเตอร์แมชชีนระหว่างเวกเตอร์ของค่าความถี่ของการเกิดร่วมกันของโคออคเคอเรนซ์ค่าพีเอฟ กับเวกเตอร์ความถี่ของการปรากฏของแบบจำลองภาษาหน่วยเสียงจากระบบอ้างอิงที่ 2.1 และเวกเตอร์ของการประกอบกันของไตรแกรมของค่าการออกเสียงที่ปรากฏบนหน่วยเสียง พบว่าขนาดมิติของเวกเตอร์ที่นำเสนอมีค่าน้อยกว่า ซึ่งข้อเสียของขนาดมิติของเวกเตอร์ที่มากขึ้นนี้ทำให้ต้องการการคำนวณเพื่อประมวลผลที่เพิ่มขึ้นของจำแนกภาษา

ตารางที่ 7-20 ผลการรู้จำภาษาของระบบที่นำเสนอเปรียบเทียบกับระบบอ้างอิงที่ 2.1 และ 2.2

ระบบการรู้จำภาษาพูด	ค่าความผิดพลาด ERR
ระบบอ้างอิงที่ 2.1 (ระบบการรู้จำภาษาพูดแบบพ็อดริวีเอ็สเอ็ม)	8.61
ระบบอ้างอิงที่ 2.2 (งานวิจัย Nunez และ Ramon [92])	8.82
ระบบที่นำเสนอ (โคออคเคอเรนซ์ค่าพีเอฟ Vocalic)	6.80 *

* แสดงค่าผิดพลาดที่น้อยที่สุด

นอกจากนี้เพื่อวัดระบบการรู้จำภาษาพูดแนวทางการเรียงหน่วยเสียงโดยใช้ค่าพีเอฟ จากการปรากฏร่วมกันโคออคเคอเรนซ์ค่าพีเอฟ Vocalic โดยทดสอบการจำแนกภาษาบนฐานข้อมูลเสียง NIST LRE07 เซตข้อมูลทดสอบที่เลือกมา 8 ภาษา พบว่าระบบการรู้จำภาษาพูดแนวทางการเรียงหน่วยเสียงโดยใช้ค่าพีเอฟให้ค่า C_{avg} มีค่า 0.0605 ดังแสดงในตารางที่ 7-21

ตารางที่ 7-21 ค่า C_{avg} ของระบบการรู้จำภาษาพูดการจำแนกด้วยแบบจำลองลาเทนท์ไดนามิกคอนดิชันนอลแรนดอมฟิลด์ร่วมกับค่าพีเอฟ บนฐานข้อมูลเซตทดสอบ 8 ภาษา

ระบบการรู้จำภาษาพูด	ค่า $100 \times C_{avg}$
ระบบการรู้จำภาษาพูดใช้เรียงหน่วยเสียงโดยใช้ค่าพีเอฟ (Co-PF)	6.05

7.4.6 บทสรุปการรู้จำภาษาตามแนวทางการเรียงตัวของเสียงโดยใช้ค่าพีเอฟ

จากผลการทดลองของการรู้จำภาษาตามแนวทางการเรียงหน่วยเสียง ซึ่งทดสอบกับฐานข้อมูลเสียง NIST LRE07 พบว่าระบบที่นำเสนอที่ใช้เวกเตอร์ของค่าความถี่ของการเกิดร่วมกันของโคออคเคอร์เรนซ์ค่าพีเอฟให้ผลการรู้จำภาษา และเปอร์เซ็นต์ความแม่นยำการจำแนกภาษาระบบการรู้จำภาษาพูดที่เสนอดีกว่าระบบอ้างอิงที่ 2.1 ที่เป็นระบบการรู้จำภาษาพูดแบบฟิอาร์วีเอสเอ็ม และระบบอ้างอิงที่ 2.2 ซึ่งเป็นระบบการรู้จำภาษาพูดที่ใช้ค่าการออกเสียงที่ปรากฏบนโปรแกรมของหน่วยเสียง นอกจากนี้ข้อดีของการใช้โคออคเคอร์เรนซ์ค่าพีเอฟซึ่งมีขนาดของเวกเตอร์มีจำนวนมิติที่น้อยกว่าระบบอ้างอิง ทำให้การสิ้นเปลืองทรัพยากรในการคำนวณเพื่อสร้างแบบจำลองภาษา และเพื่อจำแนกภาษาน้อยกว่าระบบอ้างอิง

7.5 ประเมินประสิทธิภาพของระบบการรู้จำภาษาพูดที่นำเสนอ เมื่อรวมผลลัพธ์และการตัดสินใจรู้จำภาษาจากระบบรู้จำย่อย

การทดลองเพื่อวัดประสิทธิภาพระบบการรู้จำภาษาพูดที่นำเสนอ เมื่อรวมผลลัพธ์การรู้จำภาษาจากระบบรู้จำย่อยตามแนวทางต่าง ๆ โดยใช้คะแนนที่ได้จากข้อมูล ซึ่งแสดงตามการรวมแต่ละแบบ และมีค่าน้ำหนักการถ่วงแต่ละระบบที่ได้จากการคำนวณด้วยข้อมูลชุดฝึกฝนดังนี้

การรวมแบบที่ 1 PRVSM + lattice-VSM แทนการรวมระบบการรู้จำภาษาพูดตามแนวทางการเรียงหน่วยเสียง ได้แก่ ระบบการรู้จำภาษาพูดฟิอาร์วีเอสเอ็ม + ระบบการรู้จำภาษาพูดแลตทิซ-วีเอสเอ็ม ซึ่งมีค่าน้ำหนักของการรวมของระบบการรู้จำภาษาพูดฟิอาร์วีเอสเอ็มเป็น 12.48 และค่าน้ำหนักการรวมของระบบการรู้จำภาษาพูดแลตทิซ-วีเอสเอ็มเป็น 56.59

การรวมแบบที่ 2 LDCRF (PF) + Co-PF แทนการรวมระบบการรู้จำภาษาพูดที่ใช้ค่าพีเอฟ ได้แก่ ระบบการรู้จำภาษาพูดโดยใช้การจำแนกด้วยแบบจำลองลาเทนท์ไดนามิกคอนดิชันนอลแรนดอมฟิลด์ร่วมกับค่าพีเอฟ + ระบบการรู้จำภาษาพูดตามแนวทางการเรียงตัวของเสียงโดยใช้ค่าพีเอฟ ซึ่งมีค่าน้ำหนักของการรวมของระบบการรู้จำภาษาพูดโดยใช้การจำแนกด้วยแบบจำลองลาเทนท์ไดนามิกคอนดิชันนอลแรนดอมฟิลด์ร่วมกับค่าพีเอฟเป็น 7.18 และค่าน้ำหนักการรวมของระบบการรู้จำภาษาพูดตามแนวทางการเรียงตัวของเสียงโดยใช้ค่าพีเอฟเป็น 53.26

การรวมแบบที่ 3 PRVSM + Co-PF แทนการรวมระบบการรู้จำภาษาพูดฟิอาร์วีเอสเอ็ม + ระบบการรู้จำภาษาพูดตามแนวทางการเรียงตัวหน่วยเสียงโดยใช้ค่าพีเอฟ ซึ่งมีค่าน้ำหนักของการรวม

ทางการเรียงตัวของเสียงโดยใช้ค่าพีเอฟเป็น 12.91 และค่าน้ำหนักการรวมของระบบการรู้จำภาษาพูดโดยใช้การจำแนกด้วยแบบจำลองลาเทนท์ไดนามิคคอนดิชันนอลแรนดอมฟิลด์ร่วมกับค่าพีเอฟเป็น 3.95

ผลการประเมินประสิทธิภาพระบบการรู้จำภาษาพูดที่นำเสนอเมื่อรวมผลลัพธ์การรู้จำภาษาจากระบบรู้จำย่อยตามแนวทางต่าง ๆ และตัดสินใจการรู้จำภาษาเพื่อเป็นคำตอบการรู้จำภาษา โดยประเมินบนฐานข้อมูลเสียง NIST LRE07 เซตข้อมูลทดสอบที่เลือกมา 8 ภาษา ดังตารางที่ 7-22 แสดงค่า C_{avg} ค่า C_{llr} และค่า $multiC_{llr}$ ของระบบการรู้จำภาษาพูดระบบที่รวมคะแนนจากแนวทางต่าง ๆ

เนื่องจากระบบการรู้จำภาษาพูดแลตทิซ-วีเอ็สเอ็ม ให้ค่าประสิทธิภาพที่ดีกว่าแนวทางอื่น ๆ จึงถูกนำมาใช้เป็นคะแนนการตัดสินใจพื้นฐาน ในการรวมกับคะแนนจากแนวทางอื่น เช่น วิธีการรวมแบบที่ 1 แบบที่ 4 และแบบที่ 5 จากตารางที่ 7-22 ระบบการรู้จำภาษาพูดที่ได้จากการรวมแบบที่ 1 มีค่า C_{avg} 0.0473 และค่า $multiC_{llr}$ 0.663 ซึ่งให้ผลการรู้จำภาษาจากการรวมคะแนนของสองระบบที่ดีที่สุด มีค่า C_{avg} ลดลงเมื่อเปรียบเทียบกับตารางที่ 7-6 โดยเปอร์เซ็นต์ค่าลดลงเมื่อเปรียบเทียบกับระบบการรู้จำภาษาพูดแบบพาร์วีเอ็สเอ็ม และแบบแลตทิซ-วีเอ็สเอ็มคิดเป็น 45.06% และ 17.01% ตามลำดับ ขณะที่การรวมแบบที่ 4 จากระบบการรู้จำภาษาพูดแลตทิซ-วีเอ็สเอ็ม + ระบบการรู้จำภาษาพูดตามแนวทางการเรียงตัวหน่วยเสียงโดยใช้ค่าพีเอฟ มีค่า C_{avg} 0.043 และค่า $multiC_{llr}$ 0.615 ขณะที่การรวมแบบที่ 2 เป็นการพิจารณาประสิทธิภาพของระบบการรู้จำภาษาพูดที่ใช้ค่าพีเอฟ พบว่ามีค่า C_{avg} เป็น 0.0481 และค่า $multiC_{llr}$ เป็น 0.785 ไม่ให้ประสิทธิภาพสูงนัก

การรวมคะแนนระบบรู้จำภาษาพูดแบบที่ 5 มีค่า C_{avg} 0.0686 และมีค่า $multiC_{llr}$ เป็น 0.880 ซึ่งค่าน้อยกว่าการรวมวิธีแบบอื่น เนื่องจากว่าผลการรู้จำภาษาของระบบการรู้จำภาษาพูดพาร์วีเอ็สเอ็ม และระบบการรู้จำภาษาพูดโดยใช้การจำแนกด้วยแบบจำลองลาเทนท์ไดนามิคคอนดิชันนอลแรนดอมฟิลด์ร่วมกับค่าพีเอฟมีค่าไม่ดีเมื่อเปรียบเทียบกับระบบการรู้จำภาษาพูดแลตทิซ-วีเอ็สเอ็ม

นอกจากนี้จากประสิทธิภาพที่ได้จากการรวมของระบบย่อยสองระบบ นำมาใช้เป็นคะแนนการรวมของแนวทางต่าง ๆ ต่อไป เช่น วิธีการรวมแบบที่ 7 และแบบที่ 8 มีพื้นฐานจากการรวมแบบที่ 1 ซึ่งเป็นการรวมคะแนนของระบบการรู้จำภาษาพูดตามแนวทางการเรียงหน่วยเสียง การรวมแบบที่ 7 มีค่า C_{avg} 0.0441 และค่า $multiC_{llr}$ 0.612 และการรวมแบบที่ 8 มีค่า C_{avg} 0.0451 และค่า $multiC_{llr}$ 0.625 เมื่อเปรียบเทียบผลการรวมแบบที่ 1 และการรวมคะแนนแบบที่ 8 พบว่าประสิทธิภาพของการรู้จำภาษาลดลง สะท้อนว่าการนำคะแนนการตัดสินใจของหลายแนวทางมาช่วยกันตัดสินใจรู้จำภาษา อาจจะไม่ทำให้ประสิทธิภาพการรู้จำภาษาดีขึ้นได้ ขณะที่เมื่อเปรียบเทียบการรวมผลลัพธ์การตัดสินใจจากทุกแนวทางมีค่า C_{avg} 0.0440 และค่า $multiC_{llr}$ 0.605 ซึ่งมีค่า

ไม่แตกต่างกับการรวมแบบที่ 4 และแบบที่ 7 ที่ใช้ผลการตัดสินใจที่ได้จากระบบการรู้จำภาษาพูด แลตทิซ-วีเอ็สเอ็ม + ระบบการรู้จำภาษาพูดตามแนวทางการเรียงตัวหน่วยเสียงโดยใช้ค่าพีเอฟ ซึ่งเมื่อวิเคราะห์ค่าน้ำหนักที่ได้จากการการหาจากข้อมูลชุดฝึกฝน พบว่าค่าน้ำหนักที่ใช้ในการรวมคะแนนมีค่าสอดคล้องกันกับประสิทธิภาพที่ได้จากการรวม อย่างไรก็ตามการรวมคะแนนจากระบบแนวทางอื่นๆ เพิ่มขึ้นไม่ได้ช่วยให้ผลลัพธ์การตัดสินใจรู้จำภาษาดีขึ้นเสมอ



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

ตารางที่ 7-22 ค่า C_{avg} และค่า C_{llr} ของระบบการรู้จำภาษาพูดระบบที่รวมคะแนน
จากระบบการรู้จำภาษาพูดแนวทางต่าง ๆ

การรวมระบบการรู้จำภาษาพูด	ค่า $100 * C_{avg}$	ค่า C_{llr}	ค่า $multiC_{llr}$
การรวมแบบที่ 1 PRVSM + lattice-VSM	4.73	0.187	0.663
การรวมแบบที่ 2 LDCRF (PF) + Co-PF	4.81	0.272	0.785
การรวมแบบที่ 3 PRVSM + Co-PF	4.83	0.277	0.796
การรวมแบบที่ 4 lattice-VSM + Co-PF	4.43	0.166	0.615
การรวมแบบที่ 5 PRVSM + LDCRF (PF)	6.86	0.839	0.880
การรวมแบบที่ 6 lattice-VSM + LDCRF (PF)	4.74	0.186	0.665
การรวมแบบที่ 7 PRVSM + lattice-VSM + Co-PF	4.41	0.162	0.612
การรวมแบบที่ 8 PRVSM + lattice-VSM + LDCRF (PF)	4.51	0.167	0.625
การรวมแบบที่ 9 PRVSM + lattice-VSM + Co-PF + LDCRF (PF)	4.40	0.160	0.605

บทที่ 8

สรุปผลการวิจัยและข้อเสนอแนะ

8.1 สรุปผลการวิจัย

วิทยานิพนธ์นี้นำเสนอระบบการรู้จำภาษาพูดที่รวมแนวทางต่าง ๆ ที่ใช้ข้อมูลเพื่อแสดงข้อมูลลักษณะเฉพาะของภาษาแตกต่างกัน ประกอบด้วยระบบรู้จำภาษาพูดย่อย 4 ระบบ ซึ่งประกอบด้วย 1) ระบบการรู้จำภาษาพูดแนวทางที่ใช้การจำลองลำดับหน่วยเสียงด้วยแบบจำลองเวกเตอร์สเปซ 2) ระบบการรู้จำภาษาพูดแนวทางที่ใช้หน่วยเสียงแลตทิซเอสวีเอ็ม 3) ระบบการรู้จำภาษาพูดตามแนวทางการเรียงตัวของเสียงโดยใช้ค่าพีเอฟ และ 4) ระบบการรู้จำภาษาพูดโดยใช้การจำแนกด้วยแบบจำลองลาเทนท์คอนดิชันนอลแรนดอมฟิลด์ร่วมกับค่าพีเอฟ โดยแต่ละระบบการรู้จำภาษาพูดแนวทางได้อภิปรายถึงค่าลักษณะสำคัญ ตลอดจนเทคนิคที่ใช้ในการสร้างต้นแบบของการจำแนก การระบุ และการรู้จำภาษา และได้สรุปการวิจัยโดยย่อในแต่ละบท

วิทยานิพนธ์นี้เสนอระบบการรู้จำภาษาพูดจากค่าลักษณะสำคัญเกี่ยวกับระบบเสียงในภาษาหรือค่าพีเอฟ และค่าลักษณะสำคัญทางเสียงและคุณลักษณะการออกเสียง ซึ่งในวิทยานิพนธ์นี้ได้ศึกษาลักษณะและสมบัติของค่าลักษณะสำคัญเกี่ยวกับระบบเสียงในภาษา จากผลวิเคราะห์สรุปได้ว่าค่าลักษณะสำคัญเกี่ยวกับระบบเสียงในภาษาที่ปรากฏในแต่ละภาษามีค่าแตกต่างกัน มีความถี่ของการปรากฏแตกต่างกัน และมีรูปแบบที่แสดงความสัมพันธ์ระหว่างค่าแตกต่างกัน ควรเลือกคุณสมบัติที่แตกต่างกันสำหรับบอกความแตกต่างของแต่ละคู่ภาษา ซึ่งขึ้นอยู่กับความสัมพันธ์ระหว่างภาษาว่าอยู่ในตระกูลภาษาเดียวกันหรือไม่ จากผลการศึกษาความสัมพันธ์ระหว่างค่าพีเอฟต่าง ๆ และความสัมพันธ์ของค่ากับแสดงความสัมพันธ์กันระหว่างภาษาที่อยู่ในตระกูลภาษาเดียวกัน พบว่าความสัมพันธ์ของภาษาภายในตระกูลภาษาเดียวกันเมื่อวิเคราะห์ด้วยค่าพีเอฟสอดคล้องกับทฤษฎีตระกูลภาษา ซึ่งแตกต่างกับการวิเคราะห์ด้วยค่าลักษณะสำคัญทางเสียงแบบอื่นๆ นอกจากนี้ในขั้นตอนการวิเคราะห์ความแปรปรวนของค่าลักษณะสำคัญเกี่ยวกับระบบเสียงในภาษา สรุปได้ว่าค่าเฉลี่ยของลักษณะสำคัญเกี่ยวกับระบบเสียงในภาษาแต่ละชนิด สำหรับแต่ละภาษามีค่าแตกต่างกัน และเหมาะสมกับการใช้ค่าดังกล่าวมาบอกความแตกต่างระหว่างภาษาได้ ความสัมพันธ์ของการปรากฏของค่าลักษณะสำคัญเกี่ยวกับระบบเสียงในภาษามีรูปแบบที่แตกต่างกัน วิทยานิพนธ์นี้เสนอการใช้การปรากฏร่วมกันของค่าพีเอฟสองชนิดเพื่อมาเป็นตัวแทนของหน่วยเสียง และหาความสัมพันธ์ของการปรากฏของรูปแบบดังกล่าวในแต่ละภาษา นอกจากนี้จากสมมติฐานจากการเรียงตัวของหน่วยเสียงแตกต่างกันในแต่ละภาษา ส่งผลให้การหาตัวแทนเสียงจากการปรากฏของค่าพีเอฟที่บริเวณรอยต่อระหว่างหน่วยเสียง ว่าเกิดการเปลี่ยนแปลงของค่าลักษณะสำคัญเกี่ยวกับระบบเสียงในภาษามีรูปแบบแตกต่างกันหรือไม่ และนำรูปแบบการเปลี่ยนแปลง สะท้อนให้ค่าลักษณะสำคัญที่สกัดได้จากสัญญาณเสียงพูดที่มีค่าแตกต่างกันระหว่างภาษา

นอกจากนี้จากคุณสมบัติของภาษาสามารถแสดงได้จากข้อมูลลักษณะเฉพาะของภาษาหลายแบบ ซึ่งในวิทยานิพนธ์นี้ศึกษาประสิทธิภาพของระบบการรู้จำภาษาพูดที่ใช้แนวทางการเรียงตัวของหน่วยเสียงที่ใช้หน่วยเสียงที่แตกต่างกัน สามารถสรุปได้ว่า การเลือกใช้หน่วยเสียงมีผลกระทบต่อ

ประสิทธิภาพของระบบการรู้จำภาษา เนื่องจากข้อมูลจากการเลือกหน่วยเสียงดังกล่าวมีค่าความน่าจะเป็นของการเกิดหน่วยเสียงที่แตกต่างกัน และส่งผลให้เกิดรูปแบบของข้อมูลการเรียงตัวของหน่วยเสียงที่ใช้ในการบอกความแตกต่างระหว่างภาษาได้แตกต่างกัน จากผลการทดลองพบว่า ระบบการรู้จำภาษาพูดแบบแลตทิซ-วีเอ็สเอ็มซึ่งใช้แบบจำลองภาษาที่ได้มาจากค่าความคาดหวังที่คำนวณจากความน่าจะเป็นภายหลังของหน่วยเสียงแลตทิซ มาแสดงเป็นค่าสถิติของข้อมูลการเรียงตัวของหน่วยเสียง ให้ผลการรู้จำภาษาดีกว่าระบบการรู้จำภาษาพูดแบบพ็ออาร์วีเอ็สเอ็มซึ่งใช้ข้อมูลการเรียงตัวของหน่วยเสียงจากค่าความถี่ของการปรากฏหน่วยเสียงบนสัญญาณเสียงพูด จากผลการทดลองของระบบรู้จำภาษาพูดแบบแลตทิซ-วีเอ็สเอ็ม มีค่าความผิดพลาดการรู้จำภาษาต่ำกว่าระบบการรู้จำภาษาพูดแบบพ็ออาร์วีเอ็สเอ็ม คิดเป็น 44.46% นอกจากนี้วิทยานิพนธ์นี้ได้ศึกษาผลกระทบของฟังก์ชันถ่วงน้ำหนักว่ามีผลต่อความสามารถของฟังก์ชันเคอร์เนลของซัพพอร์ตเวกเตอร์แมชชีน ซึ่งใช้ในเป็นฟังก์ชันการจำแนกระหว่างภาษา โดยวิทยานิพนธ์นี้เสนอฟังก์ชันถ่วงน้ำหนักจากค่าถ่วงน้ำหนักความซ้ำซ้อน เพื่อแสดงความสามารถของการใช้ถ่วงน้ำหนักแบบโกลบอล ร่วมกับค่าความถี่เทอมและเปรียบเทียบกับฟังก์ชันการถ่วงน้ำหนักแบบอื่นๆที่มาจากค่าความถี่เทอมร่วมกับค่าถ่วงน้ำหนักแบบโกลบอลแบบอื่น เช่น ค่าความถี่เอกสารผกผัน เป็นต้น จากผลการทดลองของระบบการรู้จำภาษาพูดแบบพ็ออาร์วีเอ็สเอ็ม และระบบการรู้จำภาษาพูดแบบแลตทิซ-วีเอ็สเอ็มพบว่า ค่าถ่วงน้ำหนักความซ้ำซ้อนมีความสามารถของแต่ละเทอมการถ่วงน้ำหนักแบบโกลบอลได้ดีเนื่องจากค่าถ่วงน้ำหนักความซ้ำซ้อนแสดงความเจาะจงของเทอมที่อยู่บนเสียงพูดในภาษาที่พิจารณาเทียบกับเสียงที่อยู่ในฐานข้อมูลเสียงทั้งหมด ซึ่งความเจาะจงของเทอมนั้นบอกการกระจายของเทอมที่ปรากฏในแต่ละเสียงพูด ดีกว่าการพิจารณาจากค่าความถี่ของจำนวนเสียงพูดที่อยู่ในฐานข้อมูลที่มีเทอมนั้น โดยจากผลการทดลองพบว่าระบบการรู้จำภาษาพูดแบบพ็ออาร์วีเอ็สเอ็ม และระบบรู้จำภาษาพูดแบบแลตทิซ-วีเอ็สเอ็ม และ เมื่อใช้ฟังก์ชันการถ่วงน้ำหนักที่นำเสนอให้ค่าความผิดพลาดการรู้จำภาษาลดลงคิดเป็น 1.30% และ 3.90% ตามลำดับ เมื่อเปรียบเทียบกับการใช้ฟังก์ชันการถ่วงน้ำหนักที่เป็นระบบอ้างอิง

ระบบรู้จำภาษาพูดที่รวมผลการตัดสินใจของระบบรู้จำภาษาแนวทางต่าง ๆ ให้ประสิทธิภาพการรู้จำภาษาดีขึ้น การรวมระบบการรู้จำภาษาพูดแลตทิซ-วีเอ็สเอ็ม และระบบการรู้จำภาษาพูดแบบพ็ออาร์วีเอ็สเอ็มให้ประสิทธิภาพการรู้จำภาษาที่ดีที่สุดสำหรับการรวมแบบสองระบบ อย่างไรก็ตามการรวมผลลัพธ์จากบางระบบอาจทำให้ประสิทธิภาพการรู้จำภาษาลดลงได้

8.2 ข้อเสนอแนะ

การปรับปรุงประสิทธิภาพของระบบการรู้จำภาษาพูด สามารถปรับปรุงประสิทธิภาพของระบบรู้จำภาษาพูดแต่ละแนวทางให้สามารถรู้จำภาษา และปรับปรุงขั้นตอนการรวมผลลัพธ์ของการตัดสินใจแต่ละระบบ โดยการปรับปรุงระบบรู้จำภาษาพูดแต่ละแนวทาง สามารถปรับปรุงในหลายขั้นตอน ได้แก่ ขั้นตอนการสกัดค่าลักษณะสำคัญทางเสียงที่สามารถบอกความแตกต่างระหว่างภาษา ขั้นตอนการวิเคราะห์ค่าลักษณะสำคัญ และศึกษาค่าลักษณะสำคัญทางเสียงอื่น ๆ ที่มีคุณสมบัติแสดงลักษณะเฉพาะของภาษา

นอกจากนี้เนื่องจากปัญหาการรู้จำภาษาพูด ซึ่งมีความสัมพันธ์กันระหว่างภาษา ในการปรับปรุงขั้นตอนการจำแนกภาษา นอกจากเลือกเทคนิคการจำแนกภาษาวิธีอื่นๆแล้ว การเลือกเทคนิคสำหรับการจำแนกกลุ่มภาษา และกลุ่มที่ไม่ใช่ภาษาเป้าหมาย สามารถทำปัญหาให้ง่ายขึ้นโดยแบ่งกลุ่มภาษาที่มีความสัมพันธ์กันเพื่อจำแนกภาษาที่ละกลุ่ม โดยใช้การรวมกลุ่ม หรือแยกกลุ่มของภาษา จากค่าลักษณะสำคัญทางเสียง ทำให้ได้กลุ่มที่เกิดขึ้นมีสมาชิกทุกตัวมีสมบัติคล้ายกัน และนำประยุกต์ใช้ในการสร้างแบบจำลองของการจำแนกภาษาที่ใช้ค่าลักษณะสำคัญทางเสียงของกลุ่มภาษาที่มีความสัมพันธ์กัน เป็นต้น ซึ่งในวิทยานิพนธ์นี้ได้ศึกษาผลกระทบของการคลุ่มเครือของภาษาต่อการระบุคำตอบของภาษาเป้าหมาย เมื่อคำตอบของการระบุภาษาไม่มีความคลุมเครือทำให้ประสิทธิภาพของการจำแนกภาษาดีขึ้น

สำหรับการปรับปรุงขั้นตอนการรวมผลลัพธ์การตัดสินใจจากแต่ละเพื่อตัดสินใจระบุภาษา ได้แก่ ขั้นตอนการรวมผลลัพธ์ที่ได้จากตัวจำแนกภาษาแบบหนึ่งต่อหนึ่ง สำหรับแต่ละคู่ภาษา ให้เป็นผลลัพธ์ของคำตอบในรูปของปัญหาที่มีลักษณะแบบหลายกลุ่ม ซึ่งใช้วิธีการรวมคะแนนจากตัวจำแนกแบบหนึ่งต่อหนึ่งด้วยวิธีอื่น เช่น วิธีการพิจารณาจากค่าความน่าจะเป็น การเลือกความน่าจะเป็นสูงสุดสองอันดับจากตัวจำแนกภาษาแบบหนึ่งต่อทั้งหมด และค่อยพิจารณาผลการตัดสินใจจากความน่าจะเป็นที่ได้จากตัวจำแนกภาษาแบบหนึ่งต่อหนึ่ง หรือใช้วิธีการรวมผลลัพธ์เป็นข้อมูลที่มีลักษณะแบบหลายกลุ่ม จากตัวจำแนกแบบหนึ่งต่อหนึ่ง ด้วยวิธีทฤษฎีรหัสสำหรับแก้ไขข้อผิดพลาด (Error-correcting-code) เป็นต้น

ภาคผนวก ก ความถี่ของหน่วยเสียงในฐานข้อมูลเสียง OGI ส่วนเสียงพูดต่อเนื่อง

ตารางที่ ก-1 ความถี่ของหน่วยเสียงในฐานข้อมูลเสียง OGI ส่วนเสียงพูดต่อเนื่อง

หน่วยเสียงใน OGI	ภาษาอังกฤษ	ภาษาเยอรมัน	ภาษาสเปน	ภาษาฮินดี	ภาษาจีนแมนดาริน	ภาษาญี่ปุ่น	จำนวนหน่วยเสียงในทศภาษา	ร้อยละของหน่วยเสียงต่อจำนวนหน่วยเสียงทั้งหมด
sil	9802	4268	5475	3485	2100	3663	28793	9.54
n	5586	4324	2706	958	1372	2434	17380	5.76
aa	2216	951	5285	1963	103	4473	14991	4.97
s	4832	2013	3387	818	128	929	12107	4.01
tcl	4633	2484	2210	443		1981	11751	3.89
iy	3780	453	3171	924	8	2435	10771	3.57
t	3771	1943	2313	206	862	1631	10726	3.55
ey	1717		5206	1148	1	2549	10621	3.52
ih	6265	2467		938			9670	3.2
eh	2928	3499	1339	1339	7		9112	3.02
ah	6620			2488			9108	3.02
m	2950	1531	2008	1153	355	1110	9107	3.02
k	2608	542	1479	1169	906	1877	8581	2.84
ow			3866	637	3	3772	8278	2.74
kcl	2885	547	1466	1292		1739	7929	2.63
l	3351	1379	1720	722	346	12	7530	2.5
r	3391	114	2078	98	131	1052	6864	2.27
d	2351	1815	692	87			4945	1.64

หน่วยเสียงใน OGI	ภาษาอังกฤษ	ภาษาเยอรมัน	ภาษาสเปน	ภาษาฮินดี	ภาษาจีนแมนดาริน	ภาษาญี่ปุ่น	จำนวนหน่วยเสียงในทศภาษา	ร้อยละของหน่วยเสียง ต่อจำนวนหน่วยเสียงทั้งหมด
ay	3610	990	72	109	3	11	4795	1.59
dcl	2713	1305	411	313			4742	1.57
w	2341		262	385	1077	410	4475	1.48
ax	3224	625	136	194	9	10	4198	1.39
p	1638	249	972	416	568	90	3933	1.3
uw	1056		1363	167	2	1333	3921	1.3
b	1581	924	333	506		244	3588	1.19
y	719	207	140	462	1461	551	3540	1.17
pcl	1696	266	876	392		78	3308	1.1
bcl	1438	764	222	603		213	3240	1.07
ae	2436	266	4	499			3205	1.06
f	1688	999	246	138	114		3185	1.06
g	945	914	166	290		488	2803	0.93
sh	631	710	43	181	461	682	2708	0.9
ng	996	238	311	70	949	12	2576	0.85
v	1692	840			11		2543	0.84
z	1914	321		98		1	2334	0.77
dh	2281			3			2284	0.76
dx	1233		957				2190	0.73
gcl	721	734	117	258		316	2146	0.71

หน่วยเสียงใน OGI	ภาษาอังกฤษ	ภาษาเยอรมัน	ภาษาสเปน	ภาษาฮินดี	ภาษาจีนแมนดาริน	ภาษาญี่ปุ่น	จำนวนหน่วยเสียงในทุกภาษา	ร้อยละของหน่วยเสียง ต่อจำนวนหน่วยเสียงทั้งหมด
non	659	389		1096			2144	0.71
uh	423	1023		516			1962	0.65
hh	1079		7		530	331	1947	0.65
a		1929					1929	0.64
ix	1473	65	234	85	2		1859	0.62
er	1556						1556	0.52
ao	1031			489	7		1527	0.51
ts		476			808	163	1447	0.48
jh	567	9	50	362		245	1233	0.41
ddcl				328		880	1208	0.4
dd				370		788	1158	0.38
h		555		599			1154	0.38
ch	494	25	183	187		247	1136	0.38
cx		1066					1066	0.35
rd				1061			1061	0.35
aw	630	362		13	6		1011	0.34
x		979					979	0.32
c					863		863	0.29
oh		850					850	0.28
axr	765						765	0.25

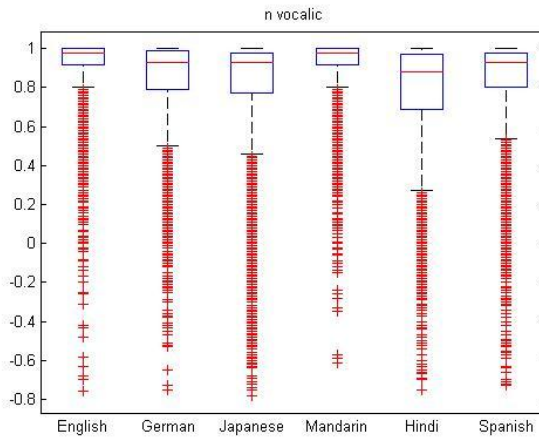
หน่วยเสียงใน OGI	ภาษาอังกฤษ	ภาษาเยอรมัน	ภาษาสเปน	ภาษาฮินดี	ภาษาจีนแมนดาริน	ภาษาญี่ปุ่น	จำนวนหน่วยเสียงในทศภาษา	ร้อยละของหน่วยเสียง ต่อจำนวนหน่วยเสียงทั้งหมด
bx			814				814	0.27
dt				765			765	0.25
nx	706						706	0.23
th	617			16			633	0.21
shr					603		603	0.2
ea		539					539	0.18
ee		478					478	0.16
hs			445				445	0.15
jhcl	393	5		28			426	0.14
chcl	349	21		1		1	372	0.12
en	365						365	0.12
el	338						338	0.11
hx			329				329	0.11
tsr					322		322	0.11
gx			314				314	0.1
oy	122	179					301	0.1
oo		276					276	0.09
ihw		252					252	0.08
ia		248					248	0.08
kh				122	119		241	0.08

หน่วยเสียงใน OGI	ภาษาอังกฤษ	ภาษาเยอรมัน	ภาษาสเปน	ภาษาฮินดี	ภาษาจีนแมนดาริน	ภาษาญี่ปุ่น	จำนวนหน่วยเสียงในทศภาษา	ร้อยละของหน่วยเสียง ต่อจำนวนหน่วยเสียงทั้งหมด
ny			201				201	0.07
uu		193			1		194	0.06
ly			162				162	0.05
dth				150			150	0.05
chl			147				147	0.05
oa		147					147	0.05
ehw		131					131	0.04
rr		7	123				130	0.04
dz						111	111	0.04
chh				103			103	0.03
ox					94		94	0.03
ux	89						89	0.03
bh				81			81	0.03
ph				17	59		76	0.03
iyw		75					75	0.02
eyw		66					66	0.02
em	51						51	0.02
zh	43	3					46	0.02
gh				18			18	0.01
jhh				17			17	0.01

หน่วยเสียงใน OGI	ภาษาอังกฤษ	ภาษาเยอรมัน	ภาษาสเปน	ภาษาฮินดี	ภาษาจีนแมนดาริน	ภาษาญี่ปุ่น	จำนวนหน่วยเสียงในทศภาษา	ร้อยละของหน่วยเสียง ต่อจำนวนหน่วยเสียงทั้งหมด
ew	16						16	0.01
eng	11						11	0
tx			10				10	0
pf		7					7	0
@	5						5	0
ddh				5			5	0
oe					3		3	0
ax0	2						2	0
done		1					1	0
in		1					1	0
nr				1			1	0
tscl						1	1	0
ทั้งหมด	106221	44771	48576	27897	12294	33200	301752	100

ภาคผนวก ข-1 Box plot ของค่าพีเอฟซึ่งได้จากผลลัพธ์ของโครงข่ายประสาทเทียมหลายชั้น
สำหรับการประมาณค่าที่เหมือนกับค่าพีเอฟเป้าหมายบนหน่วยเสียง 'n'

ค่าพีเอฟ Vocalic บนเสียง 'n'



F = 420.54, P-value = 0

	อังกฤษ	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
ค่าเฉลี่ย	0.92	0.84	0.81	0.92	0.78	0.83
ค่าเบี่ยงเบนมาตรฐาน	0.15	0.22	0.29	0.15	0.28	0.25

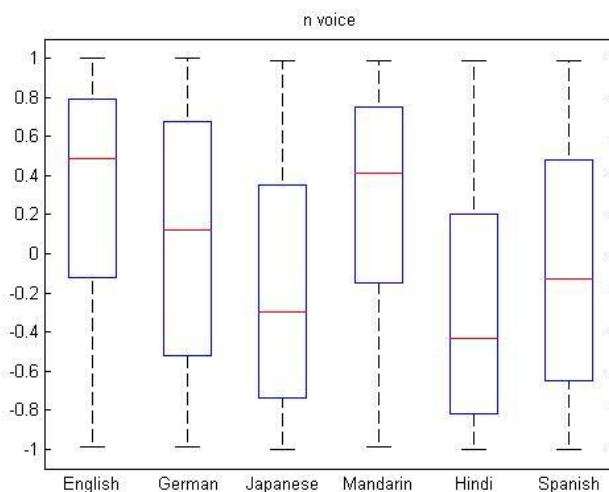
Pairwise F-ratio

	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
อังกฤษ	517.62	727.86	0.69	1291.7	568.68
เยอรมัน		63.43	563.71	228.94	8.12
ญี่ปุ่น			770.34	35.40	26.04
จีนแมนดาริน				1358.3	611.54
ฮินดี					137.87

Pairwise P-Value

	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
อังกฤษ	3.2×10^{-112}	1.1×10^{-155}	0.4	9.38×10^{-269}	7.63×10^{-123}
เยอรมัน		1.8×10^{-15}	8.22×10^{-122}	3.0×10^{-51}	0.0044
ญี่ปุ่น			2.21×10^{-164}	2.75×10^{-9}	3.39×10^{-7}
จีนแมนดาริน				8.6×10^{-282}	9.94×10^{-132}
ฮินดี					1.15×10^{-31}

ค่าพีเอฟ Voice บนเสียง 'n'



F = 964.05, P-value = 0

	อังกฤษ	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
ค่าเฉลี่ย	0.31	0.07	-0.18	0.26	-0.28	-0.07
ค่าเบี่ยงเบนมาตรฐาน	0.57	0.65	0.60	0.56	0.59	0.60

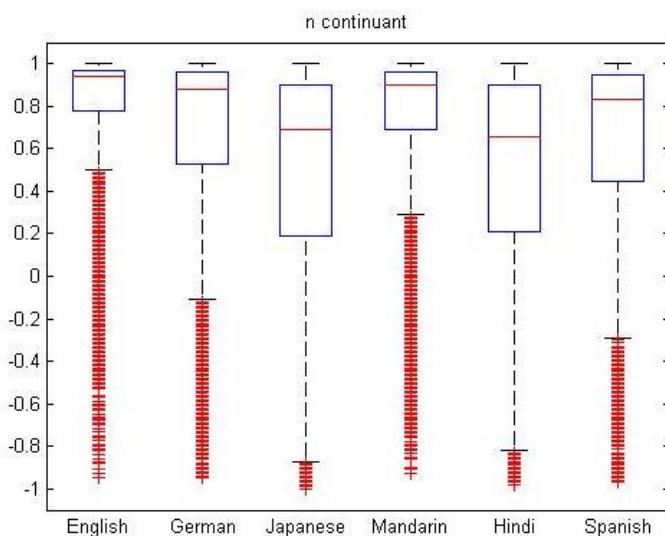
Pairwise F-ratio

	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
อังกฤษ	481.18	2103.60	19.52	3150.7	1303.6
เยอรมัน		468.57	318.33	952.36	158.72
ญี่ปุ่น			1734.6	85.23	86.16
จีนแมนดาริน				2689.6	1017.2
ฮินดี					346.56

Pairwise P-Value

	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
อังกฤษ	1.3×10^{-104}	0	1×10^{-5}	0	4.36×10^{-271}
เยอรมัน		5.8×10^{-102}	2.7×10^{-70}	2.6×10^{-201}	3.64×10^{-36}
ญี่ปุ่น			0	3.1×10^{-20}	1.94×10^{-20}
จีนแมนดาริน				0	2.48×10^{-214}
ฮินดี					2.80×10^{-76}

ค่าพีเอฟ Continuant บนเสียง 'ก'



F = 564.85, P-value = 0

	อังกฤษ	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
ค่าเฉลี่ย	0.80	0.68	0.48	0.75	0.48	0.48
ค่าเบี่ยงเบนมาตรฐาน	0.31	0.40	0.51	0.36	0.50	0.43

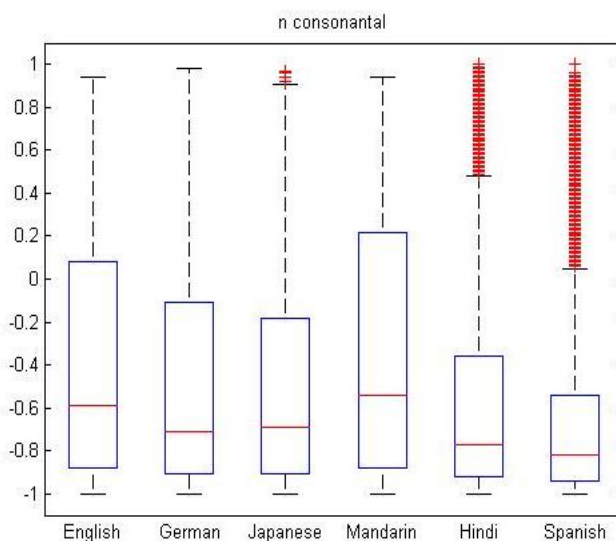
Pairwise F-ratio

	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
อังกฤษ	310.14	1624.6	71.85	1674.3	563.10
เยอรมัน		535.41	87.02	544.24	36.15
ญี่ปุ่น			1049.7	0.03	300.49
จีนแมนดาริน				1076.0	238.41
ฮินดี					303.24

Pairwise P-Value

	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
อังกฤษ	1.47×10^{-68}	0	2.58×10^{-17}	0	1.10×10^{-121}
เยอรมัน		6.33×10^{-116}	1.25×10^{-20}	9.19×10^{-118}	1.87×10^{-9}
ญี่ปุ่น			7.73×10^{-221}	0.84	1.65×10^{-66}
จีนแมนดาริน				4.34×10^{-226}	2.83×10^{-53}
ฮินดี					4.30×10^{-67}

ค่าพีเอพ Consonant บนเสียง 'n'



F = 325.79, P-value = 0

	อังกฤษ	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
ค่าเฉลี่ย	-0.37	-0.45	-0.48	-0.31	-0.57	-0.65
ค่าเบี่ยงเบนมาตรฐาน	0.58	0.57	0.53	0.61	0.46	0.41

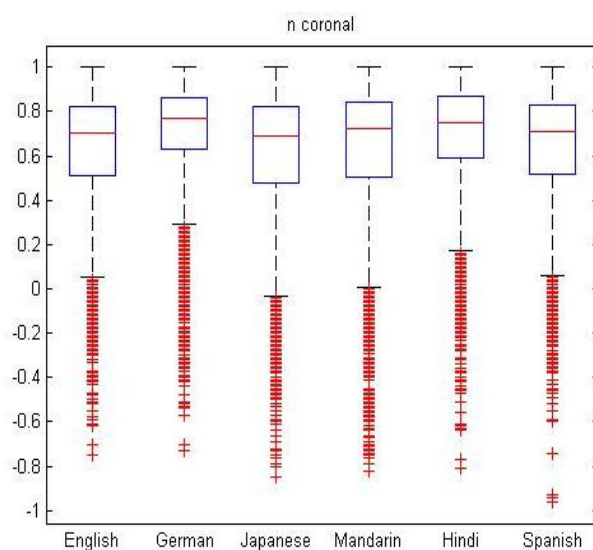
Pairwise F-ratio

	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
อังกฤษ	55.07	115.93	23.90	429.53	937.03
เยอรมัน		9.52	148.06	162.82	503.33
ญี่ปุ่น			243.09	99.53	402.88
จีนแมนดาริน				648.64	1238.5
ฮินดี					106.53

Pairwise P-Value

	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
อังกฤษ	1.23×10^{-13}	6.52×10^{-27}	1.02×10^{-6}	8.94×10^{-94}	3.24×10^{-198}
เยอรมัน		0.0020	7.26×10^{-34}	4.74×10^{-37}	3.10×10^{-109}
ญี่ปุ่น			2.82×10^{-54}	2.37×10^{-23}	3.60×10^{-88}
จีนแมนดาริน				2.13×10^{-139}	2.58×10^{-258}
ฮินดี					7.14×10^{-25}

ค่าพีเอฟ Coronal บนเสียง 'ก'



F = 113.13, P-value = 4.69×10^{-119}

	อังกฤษ	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
ค่าเฉลี่ย	0.63	0.70	0.61	0.63	0.69	0.64
ค่าเบี่ยงเบนมาตรฐาน	0.26	0.23	0.28	0.29	0.25	0.26

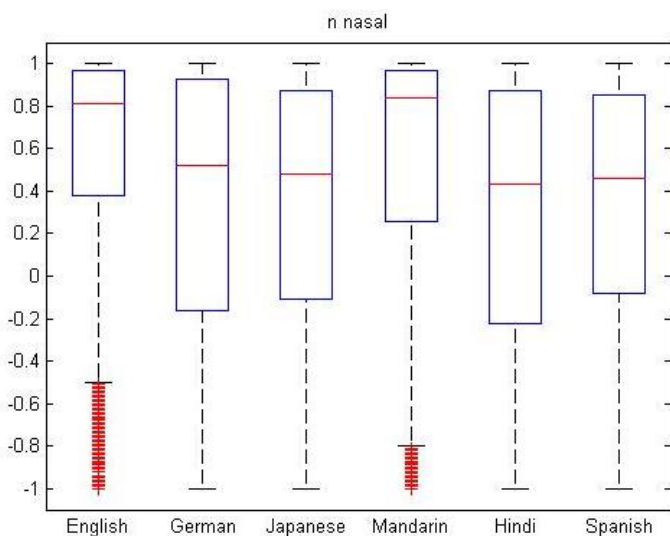
Pairwise F-ratio

	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
อังกฤษ	251.52	13.86	0.11	168.42	6.98
เยอรมัน		363.98	210.21	6.68	170.00
ญี่ปุ่น			14.71	264.11	39.13
จีนแมนดาริน				140.45	4.67
ฮินดี					104.88

Pairwise P-Value

	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
อังกฤษ	4.5×10^{-56}	0.0020	0.74	2.96×10^{-38}	0.0082
เยอรมัน		5.8×10^{-80}	3.09×10^{-47}	0.0097	1.34×10^{-38}
ญี่ปุ่น			0.0001	9.24×10^{-59}	4.09×10^{-10}
จีนแมนดาริน				3.21×10^{-32}	0.0306
ฮินดี					1.63×10^{-24}

ค่าพีเอฟ Nasal บนเสียง 'ก'



F = 323.32, P-value = 0

	อังกฤษ	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
ค่าเฉลี่ย	0.60	0.35	0.34	0.57	0.30	0.34
ค่าเบี่ยงเบนมาตรฐาน	0.48	0.60	0.56	0.53	0.59	0.55

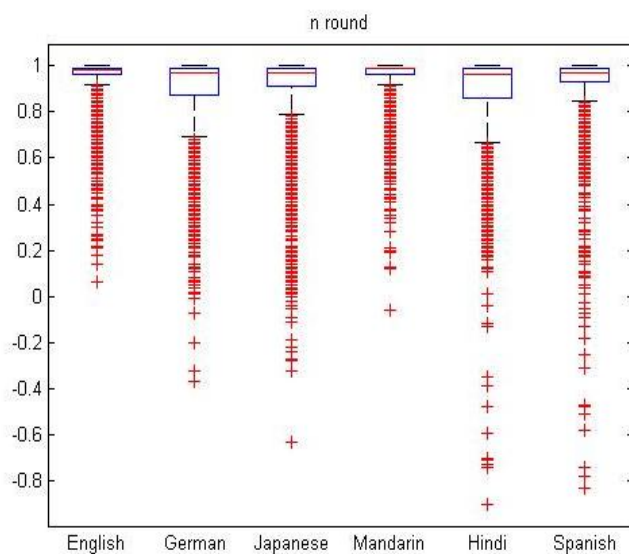
Pairwise F-ratio

	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
อังกฤษ	589.47	716.52	9.49	890.13	703.81
เยอรมัน		2.22	422.53	23.49	0.8754
ญี่ปุ่น			519.34	12.12	0.3511
จีนแมนดาริน				668.22	505.79
ฮินดี					16.84

Pairwise P-Value

	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
อังกฤษ	3.7×10^{-127}	2.31×10^{-153}	0.0021	9.72×10^{-189}	9.41×10^{-151}
เยอรมัน		0.1356	2.64×10^{-92}	1.26×10^{-6}	0.3495
ญี่ปุ่น			1.41×10^{-112}	5.0×10^{-4}	0.5535
จีนแมนดาริน				1.96×10^{-143}	9.48×10^{-110}
ฮินดี					4.09×10^{-5}

ค่าพีเอฟ Round บนเสียง 'น'



F = 316.93, P-value = 0

	อังกฤษ	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
ค่าเฉลี่ย	0.96	0.90	0.91	0.96	0.89	0.93
ค่าเบี่ยงเบนมาตรฐาน	0.08	0.14	0.15	0.07	0.16	0.11

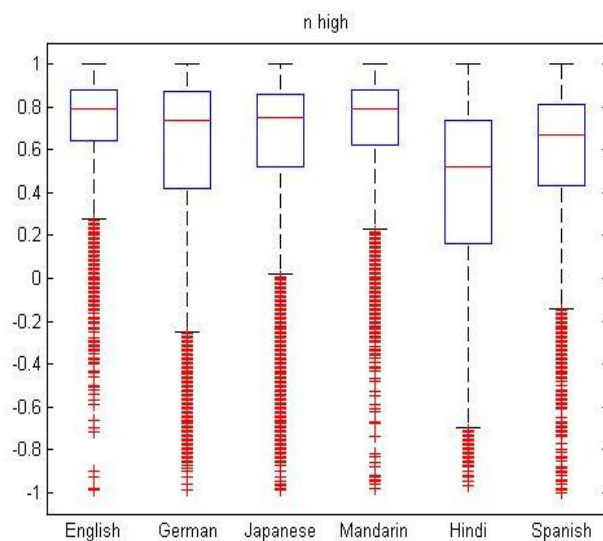
Pairwise F-ratio

	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
อังกฤษ	703.19	389.65	4.14	777.16	150.82
เยอรมัน		19.31	818.07	12.77	204.72
ญี่ปุ่น			462.59	56.84	78.42
จีนแมนดาริน				881.31	206.08
ฮินดี					285.43

Pairwise P-Value

	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
อังกฤษ	1.26×10^{-150}	2.21×10^{-85}	0.0419	8.96×10^{-166}	1.84×10^{-34}
เยอรมัน		1.1×10^{-5}	4.07×10^{-174}	3.52×10^{-4}	4.64×10^{-46}
ญี่ปุ่น			1.03×10^{-100}	5.1×10^{-14}	9.45×10^{-19}
จีนแมนดาริน				5.92×10^{-143}	2.37×10^{-46}
ฮินดี					2.64×10^{-63}

ค่าพีเอฟ High บนเสียง 'ก'



F = 635.21, P-value = 0

	อังกฤษ	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
ค่าเฉลี่ย	0.72	0.58	0.61	0.70	0.41	0.57
ค่าเบี่ยงเบนมาตรฐาน	0.24	0.39	0.37	0.26	0.42	0.32

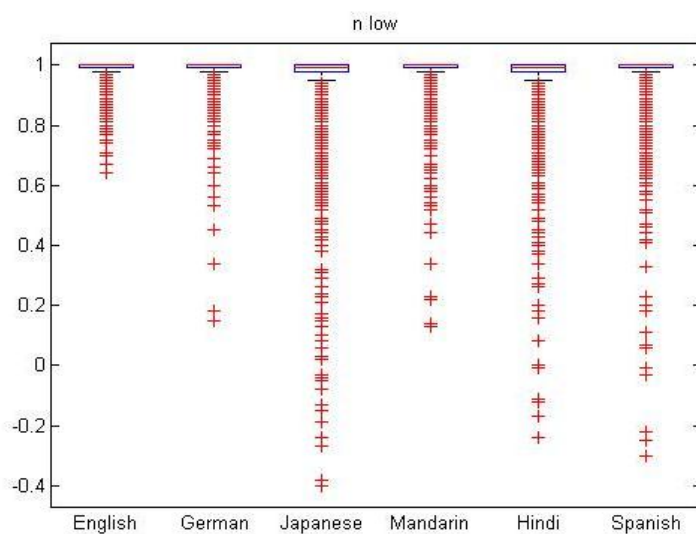
Pairwise F-ratio

	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
อังกฤษ	486.67	308.11	10.35	2435.6	685.67
เยอรมัน		18.43	361.44	582.63	0.8636
ญี่ปุ่น			210.92	825.89	31.82
จีนแมนดาริน				2100.6	507.34
ฮินดี					633.98

Pairwise P-Value

	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
อังกฤษ	9.41×10^{-106}	3.98×10^{-68}	0.0013	0	5.02×10^{-147}
เยอรมัน		1.76×10^{-5}	1.99×10^{-79}	9.69×10^{-126}	0.3527
ญี่ปุ่น			2.17×10^{-47}	1.05×10^{-175}	1.72×10^{-8}
จีนแมนดาริน				0	4.50×10^{-110}
ฮินดี					2.27×10^{-136}

ค่าพีเอฟ Low บนเสียง 'n'



$F = 150.78$, $P\text{-value} = 5.13 \times 10^{-159}$

	อังกฤษ	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
ค่าเฉลี่ย	0.99	0.99	0.97	0.99	0.98	0.98
ค่าเบี่ยงเบนมาตรฐาน	0.02	0.03	0.09	0.04	0.06	0.06

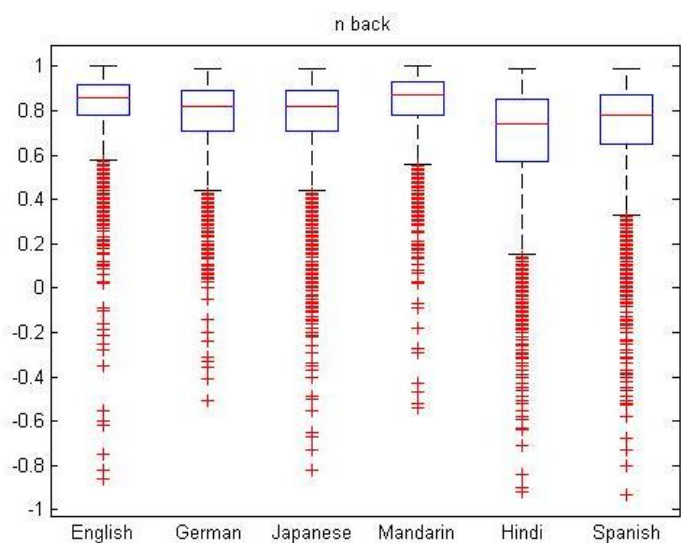
Pairwise F-ratio

	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
อังกฤษ	10.79	10.88	10.88	220.56	51.40
เยอรมัน		368.82	30.10	257.22	76.81
ญี่ปุ่น			234.45	28.88	127.55
จีนแมนดาริน				126.65	16.42
ฮินดี					44.80

Pairwise P-Value

	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
อังกฤษ	9.7×10^{-4}	0.001	9.7×10^{-4}	1.87×10^{-49}	7.95×10^{-13}
เยอรมัน		5.49×10^{-81}	4.18×10^{-8}	2.72×10^{-57}	2.13×10^{-18}
ญี่ปุ่น			1.99×10^{-52}	7.83×10^{-8}	1.98×10^{-70}
จีนแมนดาริน				3.10×10^{-29}	5.08×10^{-5}
ฮินดี					2.27×10^{-11}

ค่าพีเอฟ Back บนเสียง 'ก'



F = 646.36, P-value = 0

	อังกฤษ	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
ค่าเฉลี่ย	0.83	0.78	0.77	0.83	0.67	0.72
ค่าเบี่ยงเบนมาตรฐาน	0.14	0.16	0.18	0.14	0.24	0.20

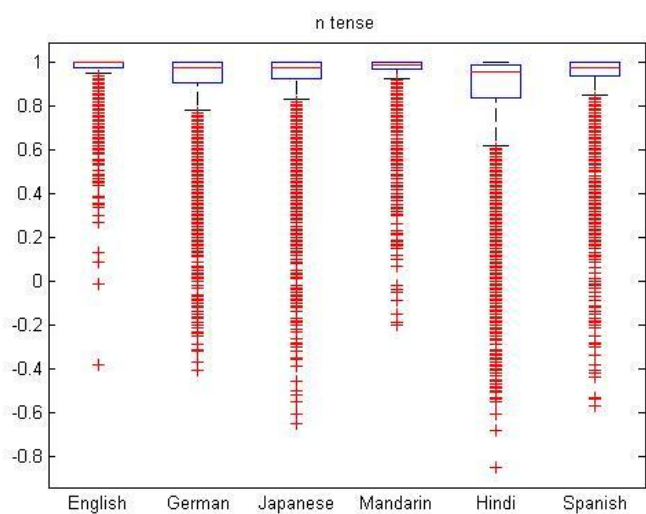
Pairwise F-ratio

	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
อังกฤษ	276.44	343.80	0.65	1799.4	969.36
เยอรมัน		10.92	307.91	827.79	259.52
ญี่ปุ่น			375.49	602.48	146.12
จีนแมนดาริน				1869.4	1024.9
ฮินดี					174.42

Pairwise P-Value

	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
อังกฤษ	2.16×10^{-61}	1.07×10^{-75}	0.4189	0	1.01×10^{-204}
เยอรมัน		9.49×10^{-4}	4.38×10^{-68}	4.28×10^{-176}	8.78×10^{-58}
ญี่ปุ่น			2.14×10^{-82}	7.47×10^{-13}	1.90×10^{-33}
จีนแมนดาริน				0	7.34×10^{-216}
ฮินดี					1.50×10^{-39}

ค่าพีเอฟ Tense บนเสียง 'n'



F = 452.29, P-value = 0

	อังกฤษ	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
ค่าเฉลี่ย	0.98	0.91	0.91	0.96	0.85	0.93
ค่าเบี่ยงเบนมาตรฐาน	0.06	0.18	0.18	0.09	0.26	0.14

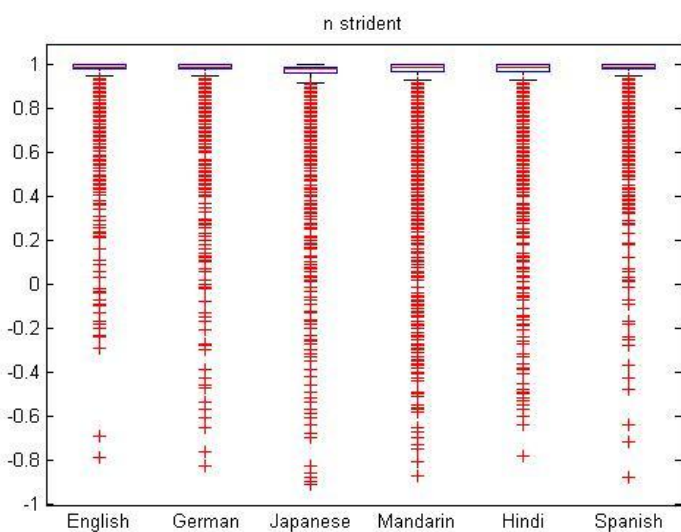
Pairwise F-ratio

	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
อังกฤษ	775.77	590.11	57.37	1335.4	427.26
เยอรมัน		6.9417	495.57	206.57	72.49
ญี่ปุ่น			356.82	271.01	31.64
จีนแมนดาริน				1060.1	207.24
ฮินดี					472.47

Pairwise P-Value

	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
อังกฤษ	1.71×10^{-165}	2.72×10^{-127}	3.86×10^{-14}	2.61×10^{-277}	2.68×10^{-93}
เยอรมัน		0.0084	1.28×10^{-107}	1.86×10^{-46}	1.86×10^{-17}
ญี่ปุ่น			1.89×10^{-78}	3.11×10^{-60}	1.89×10^{-8}
จีนแมนดาริน				6.65×10^{-223}	1.33×10^{-46}
ฮินดี					8.78×10^{-103}

ค่าพีเอฟ Strident บนเสียง 'ก'



$F = 64.26, P\text{-value} = 5.23 \times 10^{-67}$

	อังกฤษ	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
ค่าเฉลี่ย	0.97	0.97	0.94	0.94	0.95	0.95
ค่าเบี่ยงเบนมาตรฐาน	0.10	0.11	0.15	0.18	0.14	0.10

Pairwise F-ratio

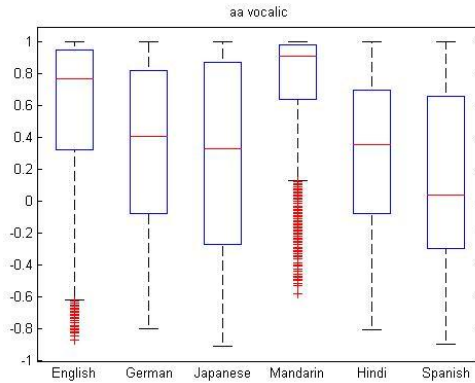
	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
อังกฤษ	0.1972	162.75	114.49	76.01	12.51
เยอรมัน		161.45	115.30	77.06	14.39
ญี่ปุ่น			0.58	15.60	97.99
จีนแมนดาริน				8.26	66.15
ฮินดี					32.94

Pairwise P-Value

	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
อังกฤษ	0.657	4.92×10^{-37}	1.33×10^{-26}	3.2×10^{-18}	4.05×10^{-4}
เยอรมัน		9.4×10^{-37}	8.93×10^{-27}	1.87×10^{-18}	1.48×10^{-4}
ญี่ปุ่น			0.4444	7.85×10^{-5}	5.14×10^{-23}
จีนแมนดาริน				0.004	4.56×10^{-16}
ฮินดี					9.71×10^{-9}

ภาคผนวก ข-2 Box plot ของค่าพีเอฟซึ่งได้จากผลลัพธ์ของโครงข่ายประสาทเทียมหลายชั้น
สำหรับการประมาณค่าที่เหมือนกับค่าพีเอฟเป้าหมายบนหน่วยเสียง 'aa'

ค่าพีเอฟ Vocalic บนเสียง 'aa'



F = 580.39, P-value = 0

	อังกฤษ	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
ค่าเฉลี่ย	0.92	0.84	0.81	0.92	0.78	0.14
ค่าเบี่ยงเบนมาตรฐาน	0.48	0.49	0.59	0.30	0.46	0.53

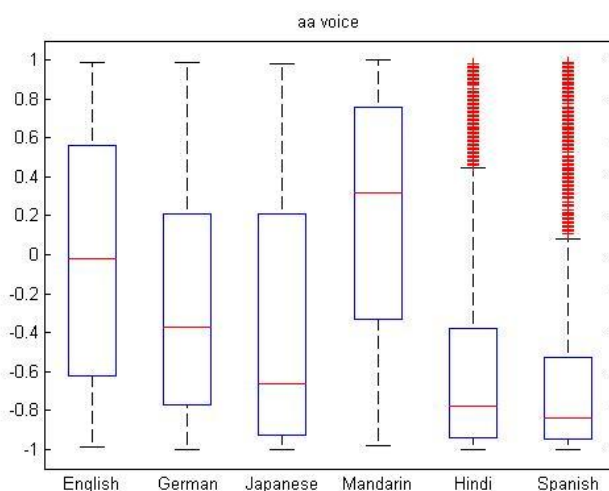
Pairwise F-ratio

	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
อังกฤษ	259.91	412.02	332.66	414.59	901.63
เยอรมัน		32.53	1387.5	14.67	209.43
ญี่ปุ่น			1522.0	5.58	58.79
จีนแมนดาริน				1862.2	2725.8
ฮินดี					123.49

Pairwise P-Value

	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
อังกฤษ	3.80×10^{-57}	2.45×10^{-88}	3.58×10^{-72}	7.42×10^{-89}	2.52×10^{-183}
เยอรมัน		1.23×10^{-8}	1.73×10^{-270}	1.29×10^{-4}	1.34×10^{-46}
ญี่ปุ่น			1.64×10^{-293}	0.0182	2.05×10^{-14}
จีนแมนดาริน				0	0
ฮินดี					2.20×10^{-28}

ค่าพีเอฟ Voice บนเสียง 'aa'



F = 824.53, P-value = 0

	อังกฤษ	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
ค่าเฉลี่ย	-0.02	-0.24	-0.36	0.20	-0.57	-0.62
ค่าเบี่ยงเบนมาตรฐาน	0.6241	0.5887	0.6594	0.5994	0.4987	0.48

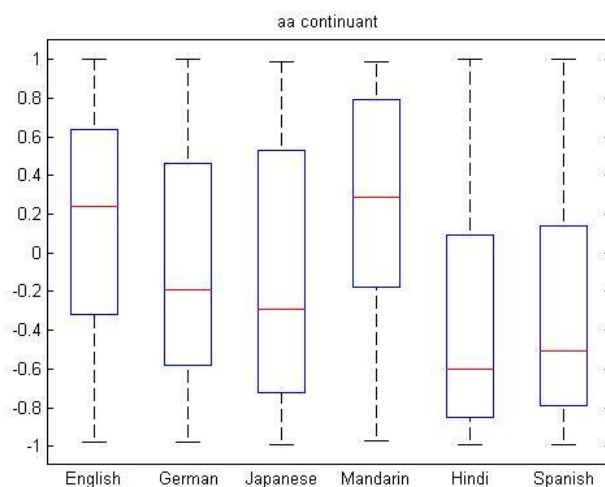
Pairwise F-ratio

	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
อังกฤษ	172.66	364.88	187.88	1275.1	1565.5
เยอรมัน		47.13	758.26	500.00	688.26
ญี่ปุ่น			1074.1	182.98	290.51
จีนแมนดาริน				2683.0	3127.5
ฮินดี					15.95

Pairwise P-Value

	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
อังกฤษ	7.6×10^{-39}	9.19×10^{-79}	4.63×10^{-42}	6.9×10^{-251}	7.20×10^{-301}
เยอรมัน		7.36×10^{-12}	2.62×10^{-156}	5.42×10^{-106}	6.96×10^{-143}
ญี่ปุ่น			5.20×10^{-215}	5.0×10^{-41}	1.71×10^{-63}
จีนแมนดาริน				0	0
ฮินดี					6.56×10^{-5}

ค่าพีเอฟ Continuant บนเสียง 'aa'



F =425.63, P-value = 0

	อังกฤษ	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
ค่าเฉลี่ย	0.15	-0.08	-0.12	0.26	-0.35	-0.28
ค่าเบี่ยงเบนมาตรฐาน	0.5783	0.5899	0.6542	0.5414	0.6048	0.62

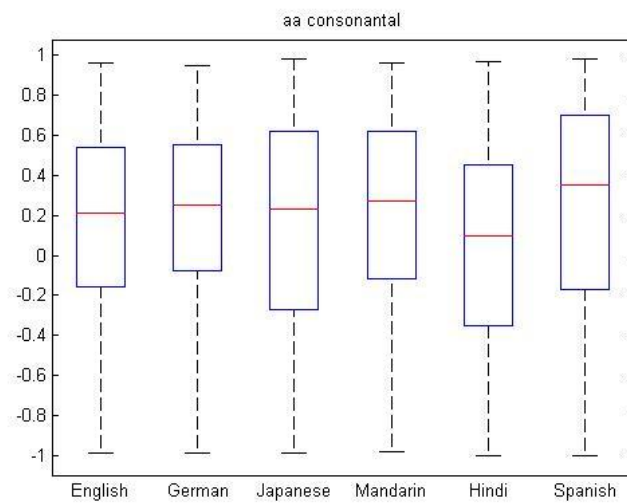
Pairwise F-ratio

	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
อังกฤษ	203.75	264.12	50.73	955.53	714.06
เยอรมัน		7.41	474.08	277.52	161.71
ญี่ปุ่น			545.46	171.73	88.70
จีนแมนดาริน				1506.7	1180.9
ฮินดี					13.39

Pairwise P-Value

	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
อังกฤษ	2.10×10^{-45}	5.07×10^{-58}	1.19×10^{-1}	2.53×10^{-193}	7.52×10^{-148}
เยอรมัน		0.0065	8.08×10^{-101}	8.38×10^{-61}	1.58×10^{-36}
ญี่ปุ่น			5.22×10^{-115}	1.19×10^{-38}	6.61×10^{-21}
จีนแมนดาริน				6.39×10^{-291}	3.17×10^{-234}
ฮินดี					2.55×10^{-4}

ค่าพีเอพ Consonantal บนเสียง 'aa'



F =47, P-value = 2.0534e-048

	อังกฤษ	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
ค่าเฉลี่ย	0.16	0.20	0.14	0.21	0.04	0.22
ค่าเบี่ยงเบนมาตรฐาน	0.47	0.44	0.55	0.48	0.52	0.56

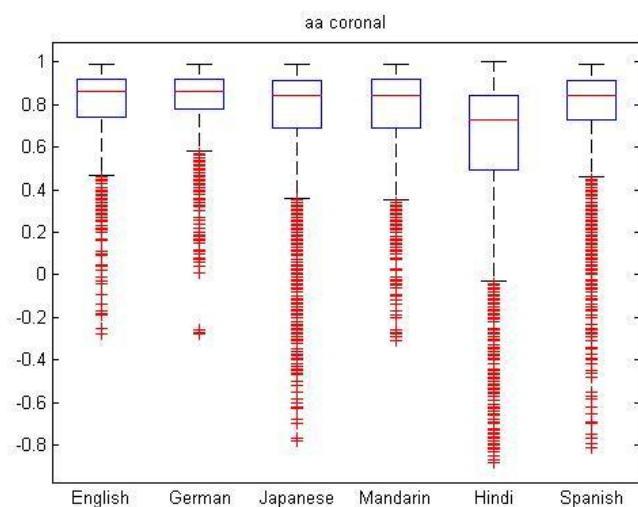
Pairwise F-ratio

	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
อังกฤษ	10.83	1.49	18.96	72.80	20.21
เยอรมัน		18.24	1.51	141.68	2.6951
ญี่ปุ่น			27.35	45.41	28.3279
จีนแมนดาริน				159.11	0.2516
ฮินดี					147.12

Pairwise P-Value

	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
อังกฤษ	0.0010	0.2212	1.35×10^{-5}	1.84×10^{-17}	7.06×10^{-6}
เยอรมัน		1.97×10^{-5}	0.2180	2.88×10^{-32}	0.1007
ญี่ปุ่น			1.76×10^{-7}	1.75×10^{-11}	1.06×10^{-7}
จีนแมนดาริน				5.65×10^{-36}	0.6160
ฮินดี					1.99×10^{-33}

คำพยางค์ Coronal บนเสียง 'aa'



F =270.92, P-value = 3.7570e-279

	อังกฤษ	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
ค่าเฉลี่ย	0.80	0.82	0.75	0.77	0.61	0.76
ค่าเบี่ยงเบนมาตรฐาน	0.16	0.13	0.26	0.21	0.36	0.23

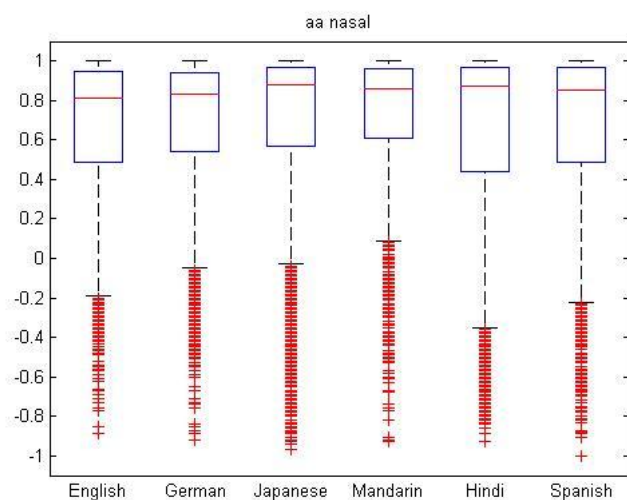
Pairwise F-ratio

	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
อังกฤษ	24.98	86.38	52.15	648.63	41.1848
เยอรมัน		181.69	147.58	849.75	117.0859
ญี่ปุ่น			8.33	261.80	8.6882
จีนแมนดาริน				387.39	0.0546
ฮินดี					367.9180

Pairwise P-Value

	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
อังกฤษ	5.94×10^{-7}	2.10×10^{-20}	5.82×10^{-13}	3.24×10^{-135}	1.50×10^{-10}
เยอรมัน		9.39×10^{-41}	1.59×10^{-33}	1.27×10^{-173}	5.19×10^{-27}
ญี่ปุ่น			0.0039	1.53×10^{-57}	0.0032
จีนแมนดาริน				2.40×10^{-83}	0.8152
ฮินดี					2.21×10^{-79}

ค่าพีเอฟ Nasal บนเสียง 'aa'



F = 14.27, P-value = 5.7106e-014

	อังกฤษ	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
ค่าเฉลี่ย	0.67	0.67	0.67	0.72	0.63	0.66
ค่าเบี่ยงเบนมาตรฐาน	0.36	0.37	0.45	0.34	0.46	0.41

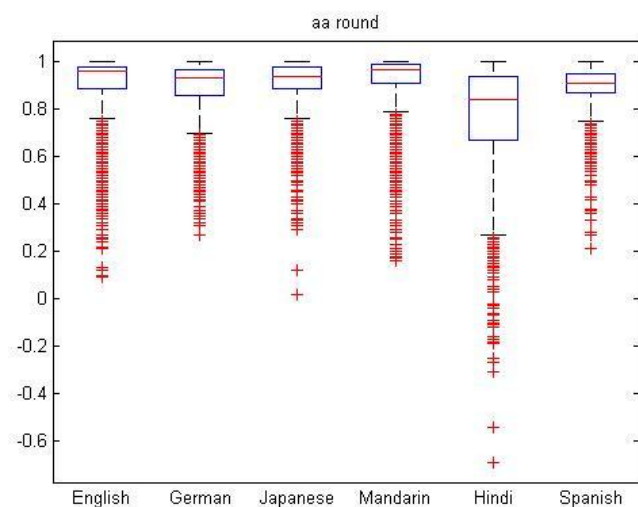
Pairwise F-ratio

	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
อังกฤษ	0.3909	0.0908	38.45	7.27	0.1700
เยอรมัน		0.0583	30.08	10.37	0.9801
ญี่ปุ่น			26.22	7.48	0.4468
จีนแมนดาริน				65.84	37.85
ฮินดี					4.7605

Pairwise P-Value

	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
อังกฤษ	0.5319	0.7521	6.0×10^{-10}	0.0070	0.6801
เยอรมัน		0.8092	4.33×10^{-8}	0.0013	0.3222
ญี่ปุ่น			3.15×10^{-7}	0.0063	0.5039
จีนแมนดาริน				5.98×10^{-16}	8.17×10^{-10}
ฮินดี					0.0292

คำพืเอฟ Round บนเสียง 'aa'



F =439.64, P-value =0

	อังกฤษ	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
ค่าเฉลี่ย	0.90	0.90	0.92	0.92	0.77	0.89
ค่าเบี่ยงเบนมาตรฐาน	0.14	0.11	0.1	0.12	0.22	0.08

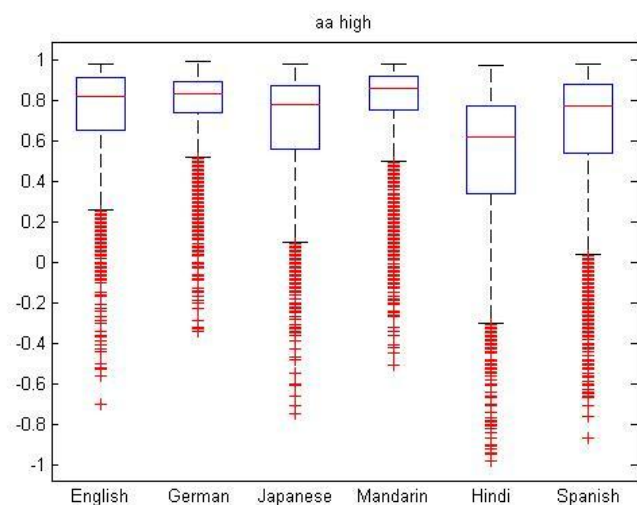
Pairwise F-ratio

	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
อังกฤษ	2.67	19.29	30.93	648.33	0.3561
เยอรมัน		48.70	63.93	658.92	1.9389
ญี่ปุ่น			3.13	940.60	42.71
จีนแมนดาริน				933.35	58.18
ฮินดี					770.97

Pairwise P-Value

	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
อังกฤษ	0.1021	1.13×10^{-5}	2.79×10^{-8}	3.7×10^{-135}	0.5507
เยอรมัน		3.32×10^{-12}	1.56×10^{-15}	3.26×10^{-137}	0.1638
ญี่ปุ่น			0.0767	1.45×10^{-190}	6.92×10^{-11}
จีนแมนดาริน				3.22×10^{-189}	2.80×10^{-14}
ฮินดี					9.95×10^{-159}

คำพืเอฟ High บนเสียง 'aa'



F =404.03, P-value = 0

	อังกฤษ	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
ค่าเฉลี่ย	0.74	0.78	0.68	0.80	0.51	0.65
ค่าเบี่ยงเบนมาตรฐาน	0.2420	0.1829	0.2639	0.2014	0.3587	0.32

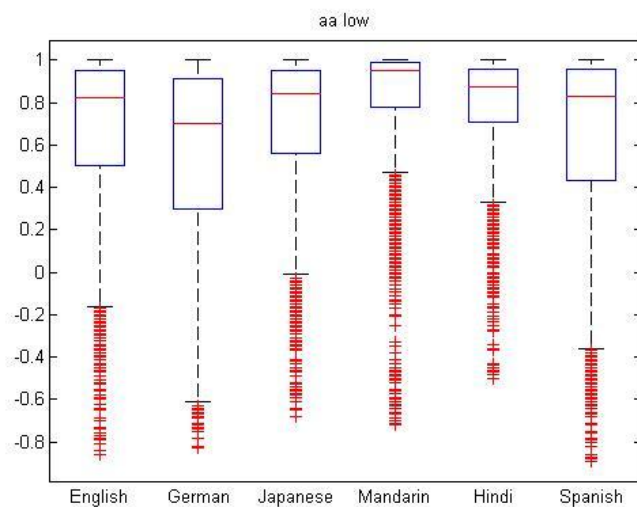
Pairwise F-ratio

	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
อังกฤษ	35.69	73.03	81.26	755.38	140.99
เยอรมัน		230.27	14.22	1158.7	314.72
ญี่ปุ่น			315.65	393.29	17.66
จีนแมนดาริน				1282.4	399.24
ฮินดี					212.40

Pairwise P-Value

	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
อังกฤษ	2.44×10^{-9}	1.63×10^{-17}	2.69×10^{-19}	9.28×10^{-156}	4.03×10^{-32}
เยอรมัน		5.77×10^{-51}	1.64×10^{-4}	3.01×10^{-230}	1.73×10^{-68}
ญี่ปุ่น			1.11×10^{-68}	1.52×10^{-84}	2.68×10^{-5}
จีนแมนดาริน				3.60×10^{-252}	9.49×10^{-86}
ฮินดี					3.20×10^{-47}

คำพืเอฟ Low บนเสียง 'aa'



F = 2 201.7954.8562, P-value = 2.2661e-209

	อังกฤษ	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
ค่าเฉลี่ย	0.67	0.54	0.70	0.81	0.78	0.63
ค่าเบี่ยงเบนมาตรฐาน	0.3715	0.4416	0.3343	0.2973	0.2572	0.43

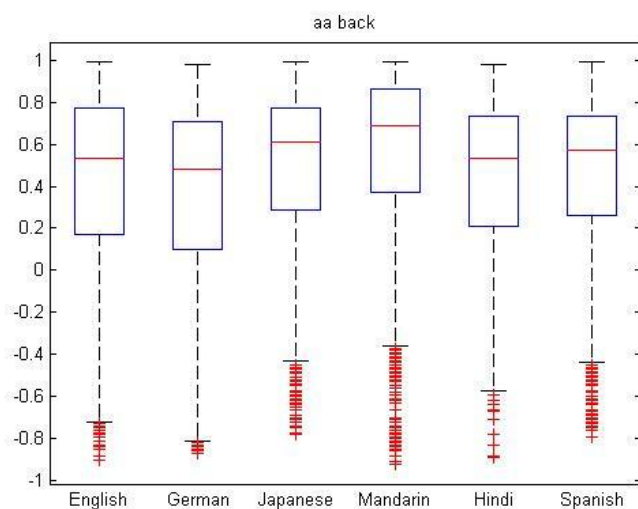
Pairwise F-ratio

	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
อังกฤษ	127.89	11.08	250.73	170.19	10.90
เยอรมัน		218.66	697.59	590.74	55.38
ญี่ปุ่น			172.15	100.59	41.75
จีนแมนดาริน				17.39	317.30
ฮินดี					235.23

Pairwise P-Value

	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
อังกฤษ	2.52×10^{-29}	8.74×10^{-4}	3.08×10^{-55}	2.52×10^{-38}	9.65×10^{-4}
เยอรมัน		1.55×10^{-48}	1.10×10^{-144}	6.37×10^{-124}	1.14×10^{-13}
ญี่ปุ่น			9.75×10^{-39}	1.81×10^{-23}	1.12×10^{-10}
จีนแมนดาริน				3.08×10^{-5}	5.12×10^{-69}
ฮินดี					5.31×10^{-52}

ค่าพีเอฟ Back บนเสียง 'aa'



F = 73.05 , P-value = 6.6908e-076

	อังกฤษ	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
ค่าเฉลี่ย	0.42	0.36	0.50	0.55	0.44	0.45
ค่าเบี่ยงเบนมาตรฐาน	0.4243	0.4267	0.3644	0.4063	0.3654	0.39

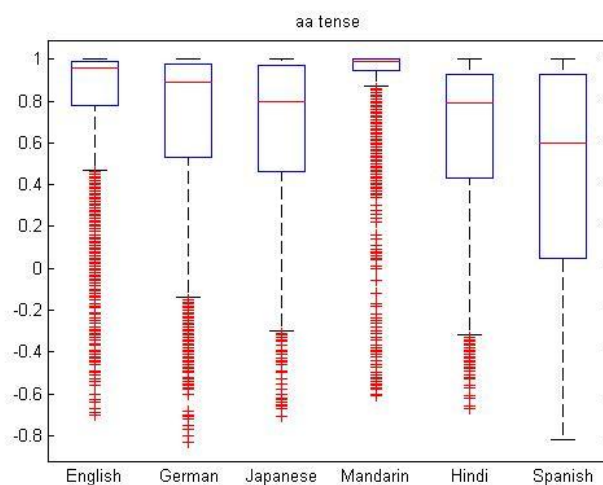
Pairwise F-ratio

	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
อังกฤษ	23.39	55.37	138.69	3.14	8.05
เยอรมัน		158.86	278.28	48.28	61.24
ญี่ปุ่น			25.50	37.70	21.78
จีนแมนดาริน				117.57	86.15
ฮินดี					1.47

Pairwise P-Value

	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
อังกฤษ	1.35×10^{-6}	1.15×10^{-13}	1.25×10^{-31}	0.0762	0.0046
เยอรมัน		6.40×10^{-36}	5.83×10^{-61}	4.12×10^{-12}	6.01×10^{-15}
ญี่ปุ่น			4.56×10^{-7}	8.81×10^{-10}	3.12×10^{-6}
จีนแมนดาริน				4.08×10^{-27}	2.35×10^{-20}
ฮินดี					0.2244

ค่าพีเอฟ Tense บนเสียง 'aa'



F = 464.71 , P-value = 0

	อังกฤษ	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
ค่าเฉลี่ย	0.81	0.67	0.67	0.92	0.63	0.45
ค่าเบี่ยงเบนมาตรฐาน	0.3117	0.4377	0.3542	0.2129	0.3783	0.50

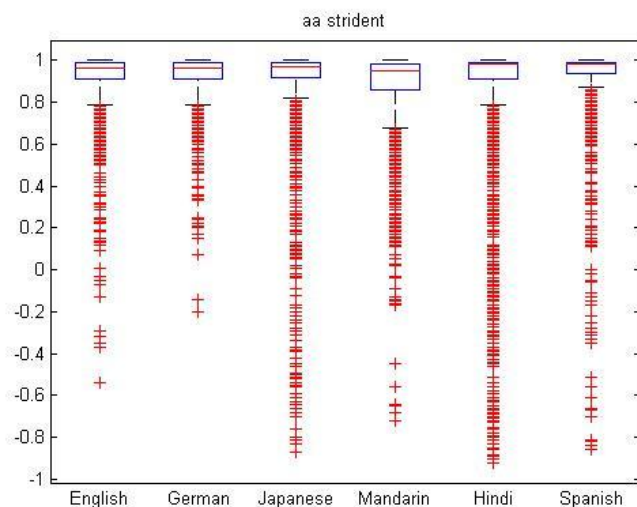
Pairwise F-ratio

	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
อังกฤษ	177.33	232.95	204.05	368.22	955.09
เยอรมัน		0.0065	664.62	15.22	279.45
ญี่ปุ่น			928.73	18.09	325.48
จีนแมนดาริน				1.16	1877.4
ฮินดี					199.89

Pairwise P-Value

	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
อังกฤษ	7.81×10^{-40}	1.59×10^{-51}	1.81×10^{-45}	1.91×10^{-79}	3.06×10^{-193}
เยอรมัน		0.9357	2.57×10^{-138}	1.01×10^{-4}	3.33×10^{-61}
ญี่ปุ่น			2.13×10^{-5}	2.13×10^{-5}	1.06×10^{-70}
จีนแมนดาริน				6.78×10^{-23}	0
ฮินดี					1.36×10^{-44}

ค่าพีเอฟ Strident บนเสียง 'aa'



F = 61.87 , P-value = 4.0872e-064

	อังกฤษ	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
ค่าเฉลี่ย	0.92	0.93	0.90	0.88	0.86	0.93
ค่าเบี่ยงเบนมาตรฐาน	0.1271	0.1055	0.2305	0.1857	0.3294	0.16

Pairwise F-ratio

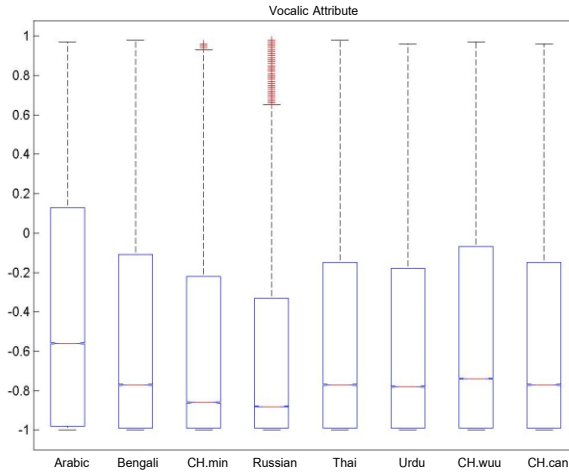
	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
อังกฤษ	2.92	22.20	114.6728	93.36	7.48
เยอรมัน		36.08	158.87	114.05	2.21
ญี่ปุ่น			15.60	29.15	40.51
จีนแมนดาริน				7.01	141.72
ฮินดี					116.18

Pairwise P-Value

	เยอรมัน	ญี่ปุ่น	จีนแมนดาริน	ฮินดี	สเปน
อังกฤษ	0.0875	2.51×10^{-6}	1.70×10^{-26}	6.54×10^{-22}	0.0062
เยอรมัน		2.01×10^{-9}	6.35×10^{-36}	2.32×10^{-26}	0.1372
ญี่ปุ่น			7.9×10^{-5}	6.96×10^{-8}	2.11×10^{-10}
จีนแมนดาริน				0.0081	2.82×10^{-32}
ฮินดี					8.13×10^{-27}

ภาคผนวก ค boxplot และการวิเคราะห์ความแปรปรวนของค่าพีเอฟสำหรับแต่ละภาษา

ค่าพีเอฟ Vocalic



F = 3706.52, P-value = 0

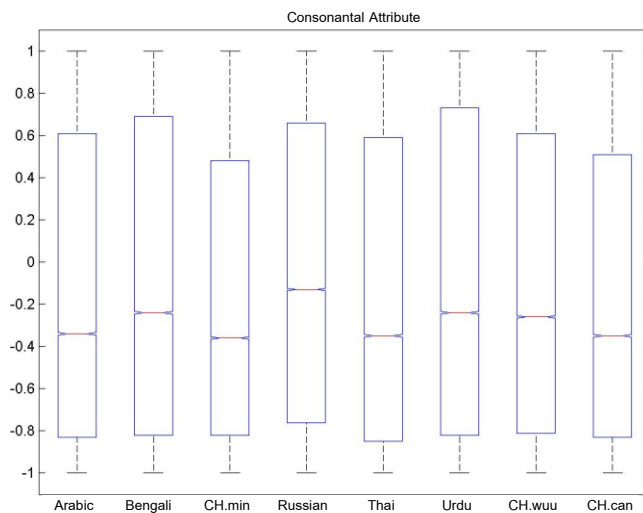
Pairwise F-ratio

	Bengali	CH.min	Russian	Thai	Urdu	CH.wuu	CH.can
Arabic	12.98	49.81	15.76	8.93	45.03	340.12	30.87
Bengali		3.18	27.14	151.99	5.14	106.61	11.70
CH.min			2.32	78.04	2.07	197.27	0.0559
Russian				153.85	4.0759	119.71	10.31
Thai					88.32	11.86	110.08
Urdu						189.08	0.0537
CH.wuu							155.64

Pairwise P-Value

	Bengali	CH.min	Russian	Thai	Urdu	CH.wuu	CH.can
Arabic	3.13×10^{-4}	1.68×10^{-12}	7.18×10^{-5}	0.0028	1.94×10^{-11}	6.56×10^{-76}	2.75×10^{-8}
Bengali		0.0745	1.88×10^{-7}	6.45×10^{-35}	0.0234	5.45×10^{-25}	6.22×10^{-4}
CH.min			0.1274	1.01×10^{-18}	0.1502	8.45×10^{-45}	0.8132
Russian				2.54×10^{-35}	0.0435	7.40×10^{-28}	0.0013
Thai					5.57×10^{-21}	5.73×10^{-4}	9.49×10^{-27}
Urdu						5.17×10^{-43}	0.8168
CH.wuu							1.03×10^{-35}

ค่าพีเอฟ Consonantal



F = 830.29, P-value = 0

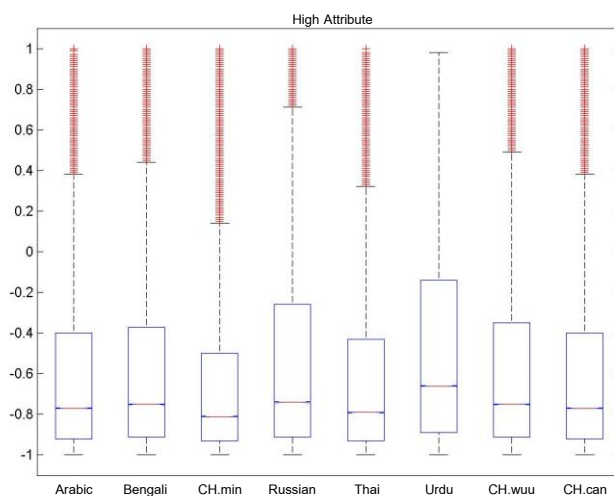
Pairwise F-ratio

	Bengali	CH.min	Russian	Thai	Urdu	CH.wuu	CH.can
Arabic	16.15	6.0855	5.91	13.49	26.18	7.98	28.90
Bengali		3.49	8.86	17.62	20.48	11.31	34.70
CH.min			12.90	5.25	124.92	10.18	0.4270
Russian				6.85	38.81	3.1031	18.59
Thai					24.56	9.02	30.97
Urdu						34.07	11.73
CH.wuu							22.22

Pairwise P-Value

	Bengali	CH.min	Russian	Thai	Urdu	CH.wuu	CH.can
Arabic	5.83×10^{-5}	0.0136	0.0150	2.39×10^{-4}	3.10×10^{-11}	0.0047	7.58×10^{-8}
Bengali		0.0617	0.0029	2.69×10^{-5}	6.01×10^{-6}	7.69×10^{-4}	1.39×10^{-8}
CH.min			3.27×10^{-4}	0.0218	5.34×10^{-29}	0.0014	0.5134
Russian				0.0088	4.66×10^{-10}	0.0781	1.61×10^{-5}
Thai					7.18×10^{-7}	0.0027	2.61×10^{-8}
Urdu						5.30×10^{-9}	6.14×10^{-4}
CH.wuu							2.42×10^{-6}

ค่าพีเอฟ High



F = 3214.64, P-value = 0

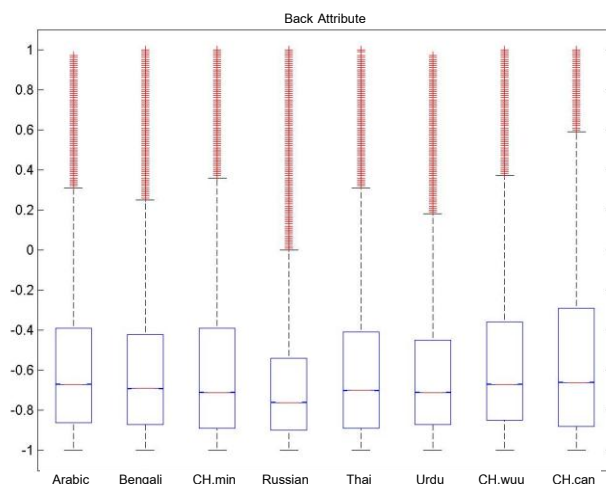
Pairwise F-ratio

	Bengali	CH.min	Russian	Thai	Urdu	CH.wuu	CH.can
Arabic	0.0065	42.81	1.0434	3.1659	130.38	35.70	16.03
Bengali		13.03	2.77	1.12	78.52	78.61	0.2374
CH.min			18.63	28.64	226.21	7.94	59.14
Russian				0.0020	85.93	54.08	4.18
Thai					106.44	49.94	8.22
Urdu						8.81	154.76
CH.wuu							24.04

Pairwise P-Value

	Bengali	CH.min	Russian	Thai	Urdu	CH.wuu	CH.can
Arabic	0.9357	6.03×10^{-11}	0.3070	0.0752	3.41×10^{-30}	2.29×10^{-9}	6.21×10^{-5}
Bengali		3.05×10^{-4}	0.0955	0.2889	7.94×10^{-19}	7.57×10^{-19}	6.22×10^{-4}
CH.min			1.58×10^{-5}	8.67×10^{-8}	4.14×10^{-51}	0.0048	1.46×10^{-14}
Russian				0.9640	1.87×10^{-20}	1.92×10^{-13}	0.0408
Thai					5.94×10^{-25}	1.58×10^{-12}	0.0041
Urdu						0.0030	1.60×10^{-35}
CH.wuu							9.40×10^{-7}

ค่าพีเอฟ Back



F = 3041.22, P-value = 0

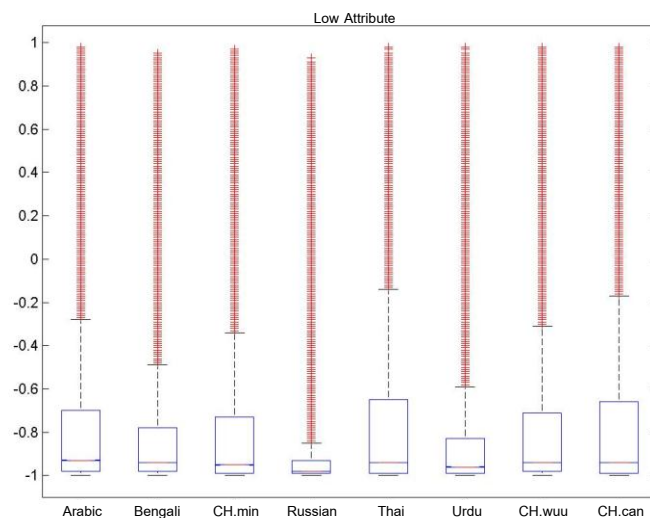
Pairwise F-ratio

	Bengali	CH.min	Russian	Thai	Urdu	CH.wuu	CH.can
Arabic	34.02	115.51	75.40	3.19	87.50	456.18	151.57
Bengali		6.02	1.12	47.26	0.0401	172.02	18.22
CH.min			2.32	1.5507	32.44	303.31	80.22
Russian				26.74	7.39	266.40	44.36
Thai					14.07	79.95	0.3707
Urdu						276.38	55.30
CH.wuu							353.08

Pairwise P-Value

	Bengali	CH.min	Russian	Thai	Urdu	CH.wuu	CH.can
Arabic	5.44×10^{-9}	6.13×10^{-27}	3.85×10^{-18}	0.0737	8.43×10^{-21}	3.82×10^{-101}	7.97×10^{-35}
Bengali		0.0141	0.2884	6.22×10^{-12}	0.8413	2.74×10^{-39}	1.95×10^{-5}
CH.min			1.30×10^{-6}	0.2130	1.22×10^8	6.72×10^{-68}	3.36×10^{-19}
Russian				2.32×10^{-7}	0.0065	7.30×10^{-60}	2.73×10^{-11}
Thai					1.75×10^{-4}	3.84×10^{-19}	0.5426
Urdu						4.90×10^{-62}	1.03×10^{-13}
CH.wuu							9.98×10^{-79}

ค่าพีเอฟ Low



F = 6597.28, P-value = 0

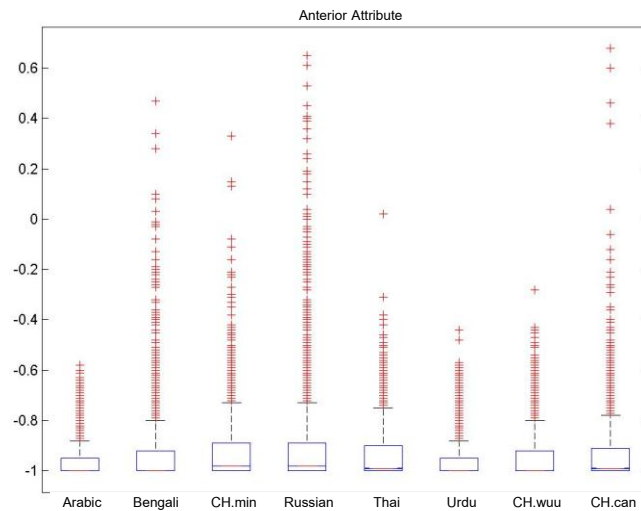
Pairwise F-ratio

	Bengali	CH.min	Russian	Thai	Urdu	CH.wuu	CH.can
Arabic	1.48	142.69	44.54	0.6885	100.24	318.70	39.37
Bengali		0.4136	66.91	157.21	6.03	49.99	25.11
CH.min			77.56	0.7625	140.51	379.07	62.92
Russian				43.54	41.23	254.26	3.79
Thai					3.5773	46.03	18.90
Urdu						329.99	32.70
CH.wuu							188.38

Pairwise P-Value

	Bengali	CH.min	Russian	Thai	Urdu	CH.wuu	CH.can
Arabic	0.2224	6.96×10^{-33}	2.49×10^{-11}	0.4067	1.35×10^{-23}	3.00×10^{-71}	3.50×10^{-10}
Bengali		0.5202	2.84×10^{-16}	4.68×10^{-36}	0.0140	1.54×10^{-12}	5.40×10^{-7}
CH.min			1.29×10^{-18}	0.3826	2.08×10^{-32}	2.22×10^{-84}	2.14×10^{-15}
Russian				4.14×10^{-11}	1.35×10^{-10}	3.21×10^{-57}	0.0514
Thai					0.0586	1.16×10^{-11}	1.37×10^{-5}
Urdu						1.05×10^{-73}	1.07×10^{-8}
CH.wuu							7.38×10^{-43}

ค่าพีเอฟ Anterior



F = 4919.50, P-value = 0

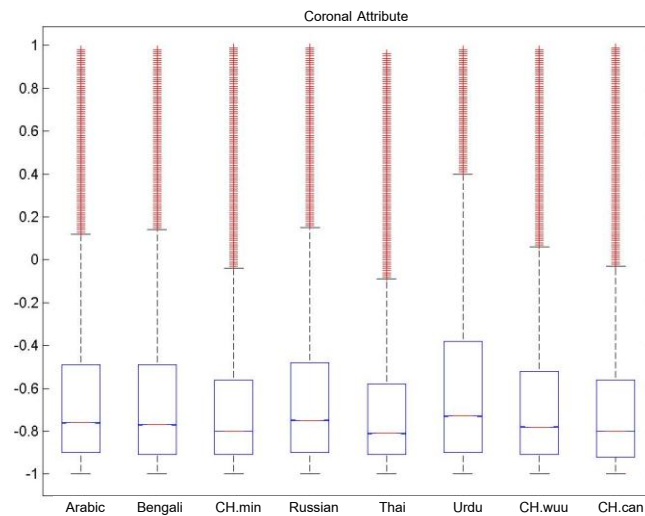
Pairwise F-ratio

	Bengali	CH.min	Russian	Thai	Urdu	CH.wuu	CH.can
Arabic	0.4902	2.0599	10.71	0.9732	37.50	635.27	7.74
Bengali		45.53	125.22	85.51	219.01	248.19	35.13
CH.min			101.64	65.65	172.07	247.14	23.71
Russian				154.07	308.39	118.49	87.75
Thai					252.10	185.39	53.89
Urdu						96.72	162.32
CH.wuu							317.22

Pairwise P-Value

	Bengali	CH.min	Russian	Thai	Urdu	CH.wuu	CH.can
Arabic	0.4838	0.1512	0.0011	0.3239	9.15×10^{-10}	4.90×10^{-140}	0.0054
Bengali		1.50×10^{-11}	4.60×10^{-29}	2.31×10^{-20}	1.53×10^{-49}	6.76×10^{-56}	3.08×10^{-9}
CH.min			6.68×10^{-24}	5.40×10^{-16}	2.66×10^{-39}	1.14×10^{-55}	1.11×10^{-6}
Russian				2.27×10^{-35}	5.27×10^{-69}	1.36×10^{-27}	7.45×10^{-21}
Thai					9.47×10^{-57}	3.30×10^{-42}	2.12×10^{-13}
Urdu						8.01×10^{-23}	3.59×10^{-37}
CH.wuu							6.30×10^{-71}

ค่าพีเอฟ Coronal



F = 3414.93, P-value = 0

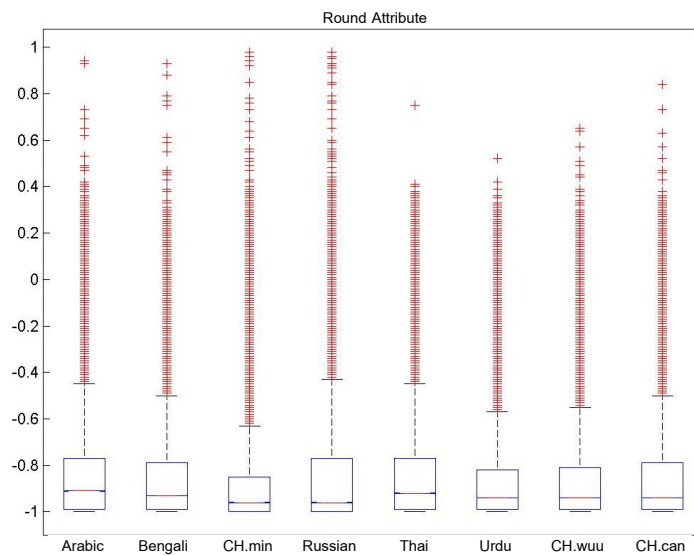
Pairwise F-ratio

	Bengali	CH.min	Russian	Thai	Urdu	CH.wuu	CH.can
Arabic	11.21	14.67	0.1979	0.3476	290.48	297.59	11.17
Bengali		0.2590	13.06	14.84	174.59	446.56	0.8226
CH.min			11.79	15.53	407.46	225.94	43.92
Russian				0.0551	274.98	300.11	7.87
Thai					315.28	357.97	11.52
Urdu						0.8506	381.96
CH.wuu							229.40

Pairwise P-Value

	Bengali	CH.min	Russian	Thai	Urdu	CH.wuu	CH.can
Arabic	8.11×10^{-4}	1.27×10^{-4}	0.6564	0.5555	4.15×10^{-65}	1.18×10^{-66}	8.28×10^{-4}
Bengali		0.6108	3.01×10^{-4}	1.16×10^{-4}	7.51×10^{-40}	4.71×10^{-99}	0.3644
CH.min			5.92×10^{-4}	8.10×10^{-5}	1.49×10^{-90}	4.74×10^{-51}	3.41×10^{-11}
Russian				0.8144	9.88×10^{-62}	3.33×10^{-67}	0.0030
Thai					1.67×10^{-70}	8.61×10^{-80}	6.86×10^{-4}
Urdu						0.3564	5.22×10^{-37}
CH.wuu							8.39×10^{-52}

ค่าพีเอฟ Round



F = 748.32, P-value = 0

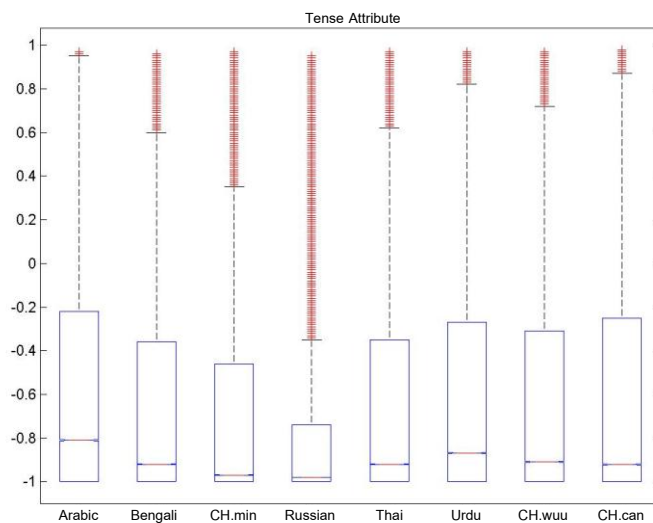
Pairwise F-ratio

	Bengali	CH.min	Russian	Thai	Urdu	CH.wuu	CH.can
Arabic	254.67	69.37	183.86	122.66	329.96	113.29	128.15
Bengali		33.38	125.86	73.97	248.37	66.36	78.93
CH.min			17.48	1.78	64.35	0.66	2.92
Russian				46.69	177.06	40.77	50.73
Thai					116.20	13.26	19.91
Urdu						107.26	121.75
CH.wuu							15.87

Pairwise P-Value

	Bengali	CH.min	Russian	Thai	Urdu	CH.wuu	CH.can
Arabic	2.61×10^{-57}	8.18×10^{-17}	7.12×10^{-42}	1.66×10^{-28}	1.06×10^{-73}	1.87×10^{-26}	1.05×10^{-29}
Bengali		7.57×10^{-9}	3.33×10^{-29}	7.96×10^{-18}	6.16×10^{-56}	3.76×10^{-16}	6.45×10^{-19}
CH.min			2.89×10^{-5}	0.1815	1.04×10^{-15}	0.4149	0.0874
Russian				8.29×10^{-12}	2.17×10^{-40}	1.71×10^{-67}	1.06×10^{-12}
Thai					4.32×10^{-27}	2.69×10^{-4}	8.08×10^{-6}
Urdu						3.94×10^{-25}	2.63×10^{-28}
CH.wuu							6.77×10^{-5}

ค่าพีเอฟ Tense



F = 5237.42, P-value = 0

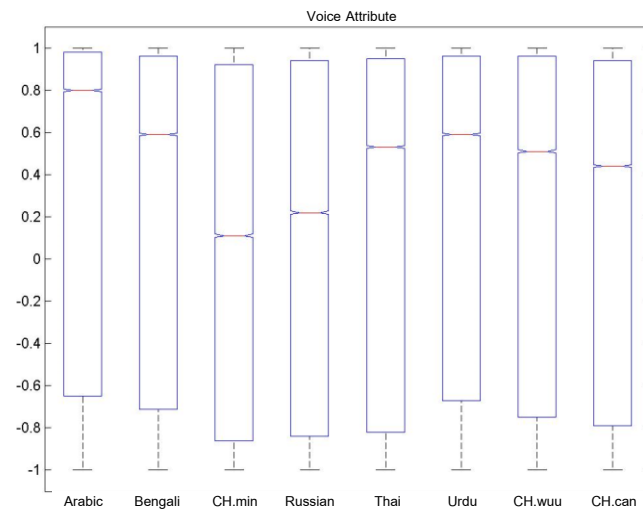
Pairwise F-ratio

	Bengali	CH.min	Russian	Thai	Urdu	CH.wuu	CH.can
Arabic	44.16	62.03	6.91	66.40	0.9232	691.26	13.67
Bengali		27.32	160.41	461.64	200.25	196.71	277.68
CH.min			19.21	45.87	0.2164	781.06	5.34
Russian				196.71	33.21	649.84	76.34
Thai					237.24	141.42	317.59
Urdu						362.66	117.72
CH.wuu							254.28

Pairwise P-Value

	Bengali	CH.min	Russian	Thai	Urdu	CH.wuu	CH.can
Arabic	3.02×10^{-11}	3.37×10^{-15}	0.0085	3.68×10^{-18}	0.3366	1.27×10^{-153}	2.17×10^{-4}
Bengali		1.73×10^{-7}	9.39×10^{-37}	2.50×10^{-102}	1.90×10^{-45}	1.12×10^{-44}	2.54×10^{-62}
CH.min			1.17×10^{-5}	1.26×10^{-11}	0.6418	1.14×10^{-171}	0.0207
Russian				1.12×10^{-44}	8.26×10^{-9}	3.37×10^{-143}	2.39×10^{-18}
Thai					1.64×10^{-53}	1.31×10^{-32}	5.23×10^{-71}
Urdu						8.21×10^{-81}	2.01×10^{-27}
CH.wuu							3.18×10^{-57}

ค่าพีเอฟ Voice



F = 2683.94, P-value = 0

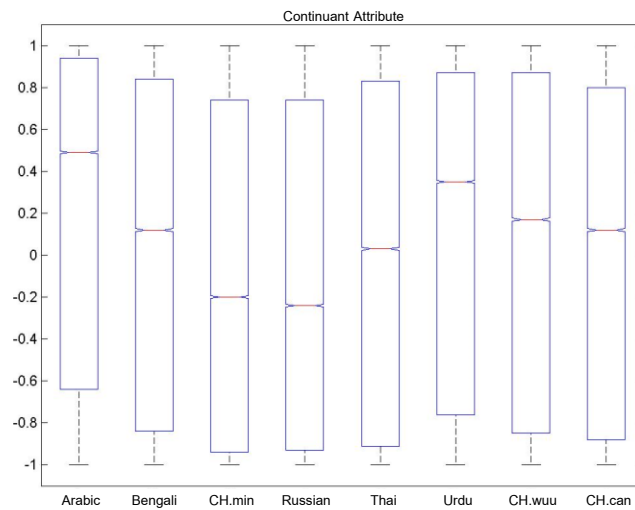
Pairwise F-ratio

	Bengali	CH.min	Russian	Thai	Urdu	CH.wuu	CH.can
Arabic	0.7047	0.5993	0.8108	1.94	0.5339	279.77	2.30
Bengali		84.05	81.26	128.80	88.37	45.46	72.21
CH.min			54.44	93.12	58.62	68.64	46.49
Russian				90.05	56.34	70.37	44.41
Thai					97.19	37.21	80.55
Urdu						69.96	48.66
CH.wuu							82.31

Pairwise P-Value

	Bengali	CH.min	Russian	Thai	Urdu	CH.wuu	CH.can
Arabic	0.4012	0.4388	0.3679	0.1603	0.4625	8.93×10^{-63}	0.1291
Bengali		4.83×10^{-20}	1.97×10^{-19}	7.57×10^{-30}	5.45×10^{-21}	1.55×10^{-11}	1.94×10^{-17}
CH.min			1.60×10^{-13}	4.94×10^{-22}	1.91×10^{-14}	1.18×10^{-16}	9.21×10^{-12}
Russian				2.35×10^{-21}	6.09×10^{-14}	4.91×10^{-17}	2.65×10^{-11}
Thai					6.33×10^{-23}	1.06×10^{-9}	2.83×10^{-19}
Urdu						6.05×10^{-17}	3.04×10^{-12}
CH.wuu							1.16×10^{-19}

ค่าพีเอฟ Contingent



F = 830.29, P-value = 0

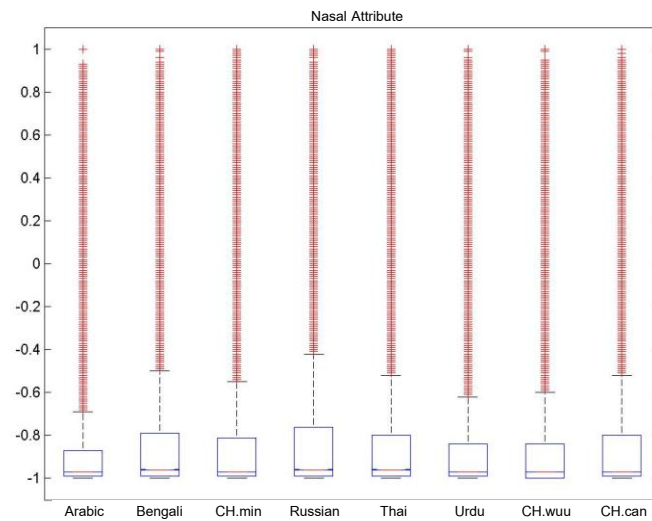
Pairwise F-ratio

	Bengali	CH.min	Russian	Thai	Urdu	CH.wuu	CH.can
Arabic	4.77	16.65	0.2686	1.0515	43.25	691.54	3.5056
Bengali		7.99	41.24	63.10	183.44	373.79	26.18
CH.min			20.12	35.87	131.55	447.34	9.89
Russian				91.29	226.40	308.54	45.53
Thai					275.06	254.54	68.37
Urdu						114.16	192.36
CH.wuu							362.72

Pairwise P-Value

	Bengali	CH.min	Russian	Thai	Urdu	CH.wuu	CH.can
Arabic	0.0288	4.49×10^{-5}	0.6043	0.3051	4.81×10^{-11}	3.00×10^{-152}	0.0621
Bengali		0.0047	1.34×10^{-10}	1.96×10^{-15}	8.82×10^{-42}	3.12×10^{-83}	3.09×10^{-7}
CH.min			7.26×10^{-6}	2.10×10^{-9}	1.89×10^{-30}	3.19×10^{-99}	0.0017
Russian				1.24×10^{-21}	3.78×10^{-51}	4.87×10^{-69}	1.50×10^{-11}
Thai					9.36×10^{-62}	2.79×10^{-57}	1.35×10^{-16}
Urdu						1.21×10^{-26}	9.99×10^{-44}
CH.wuu							7.95×10^{-81}

ค่าพีเอฟ Nasal



F = 970.02, P-value = 0

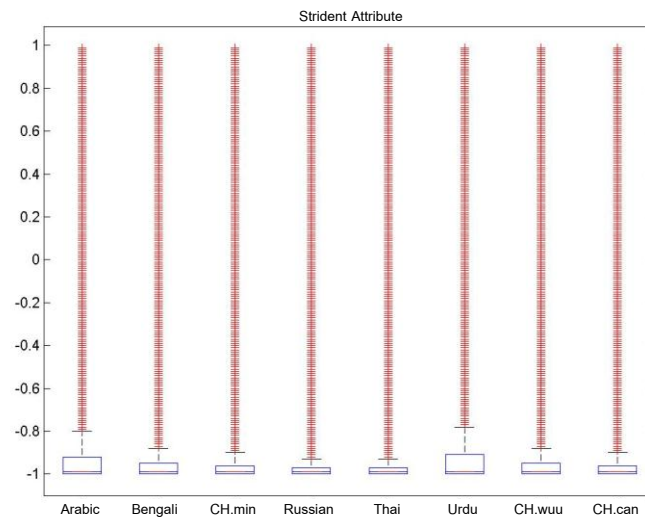
Pairwise F-ratio

	Bengali	CH.min	Russian	Thai	Urdu	CH.wuu	CH.can
Arabic	79.12	145.81	0.0717	5.35	8.01	189.92	21.44
Bengali		288.97	34.44	11.78	10.67	350.75	1.78
CH.min			77.56	40.86	40.68	470.02	18.96
Russian				18.58	23.53	102.59	41.54
Thai					53.53	58.13	77.75
Urdu						59.28	92.24
CH.wuu							30.42

Pairwise P-Value

	Bengali	CH.min	Russian	Thai	Urdu	CH.wuu	CH.can
Arabic	5.86×10^{-19}	1.45×10^{-33}	0.7889	0.0206	0.0046	3.39×10^{-43}	3.64×10^{-6}
Bengali		8.90×10^{-65}	4.38×10^{-9}	5.96×10^{-4}	0.0011	3.19×10^{-78}	0.1817
CH.min			1.28×10^{-18}	1.63×10^{-10}	1.79×10^{-10}	3.77×10^{-104}	1.34×10^{-5}
Russian				1.62×10^{-5}	1.22×10^{-6}	4.15×10^{-24}	1.15×10^{-10}
Thai					2.55×10^{-13}	2.45×10^{-14}	1.17×10^{-18}
Urdu						1.17×10^{-18}	7.71×10^{-22}
CH.wuu							3.46×10^{-8}

ค่าพีเอฟ Strident



F = 2695.36, P-value = 0

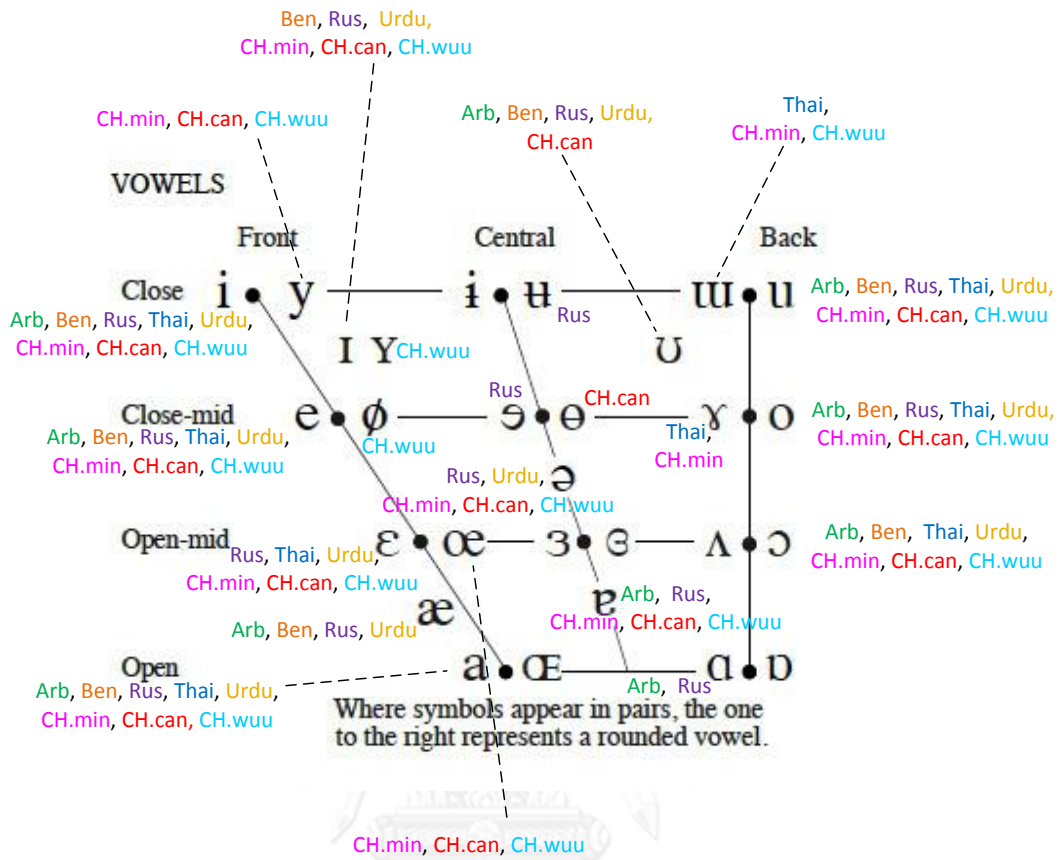
Pairwise F-ratio

	Bengali	CH.min	Russian	Thai	Urdu	CH.wuu	CH.can
Arabic	31.24	27.58	36.73	2.77	242.86	126.76	11.79
Bengali		23.50	32.80	0.6714	251.83	174.00	8.09
CH.min			29.35	0.0012	258.00	219.80	5.38
Russian				1.83	270.87	169.22	11.14
Thai					177.61	429.67	4.28
Urdu						0.3235	193.68
CH.wuu							252.43

Pairwise P-Value

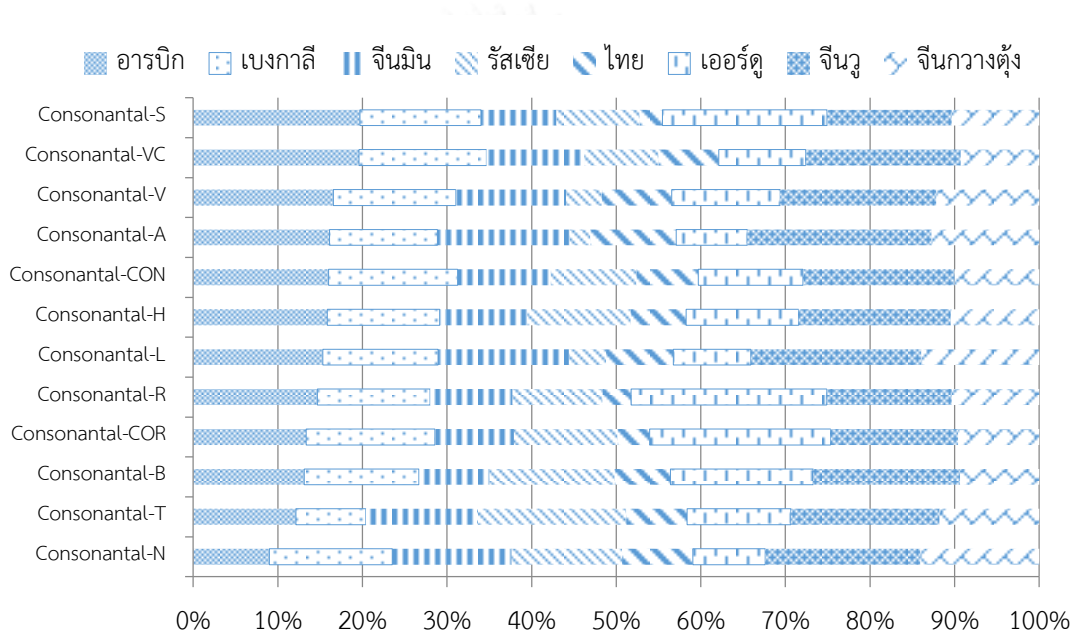
	Bengali	CH.min	Russian	Thai	Urdu	CH.wuu	CH.can
Arabic	2.27×10^{-8}	1.50×10^{-7}	1.35×10^{-9}	0.0958	9.80×10^{-55}	2.11×10^{-29}	5.92×10^{-4}
Bengali		1.24×10^{-6}	1.01×10^{-8}	0.4114	1.08×10^{-56}	1.01×10^{-39}	0.0044
CH.min			6.02×10^{-8}	0.9729	4.93×10^{-58}	1.03×10^{-49}	0.0204
Russian				0.1757	7.76×10^{-61}	1.11×10^{-38}	8.41×10^{-4}
Thai					1.65×10^{-40}	2.21×10^{-95}	0.0384
Urdu						0.5695	5.14×10^{-44}
CH.wuu							8.02×10^{-57}

ภาคผนวก ง Vowel chart

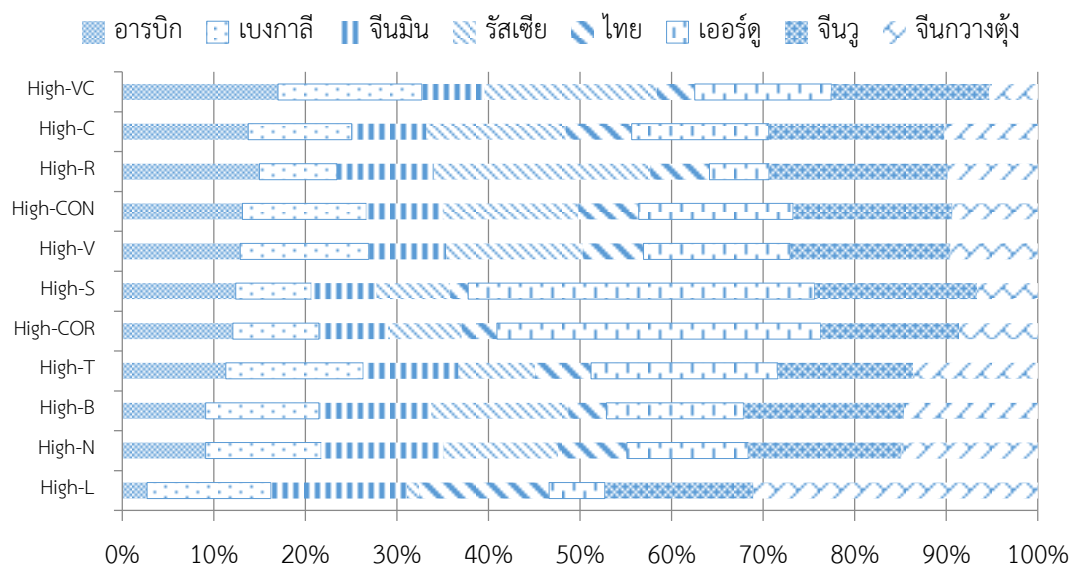


ภาคผนวก จ เปอร์เซ็นต์ของความถี่การปรากฏโคออกเคอเรนซ์ค่าพืเอฟหลัก ร่วมกับค่าพืเอฟชนิดอื่นในแต่ละภาษา

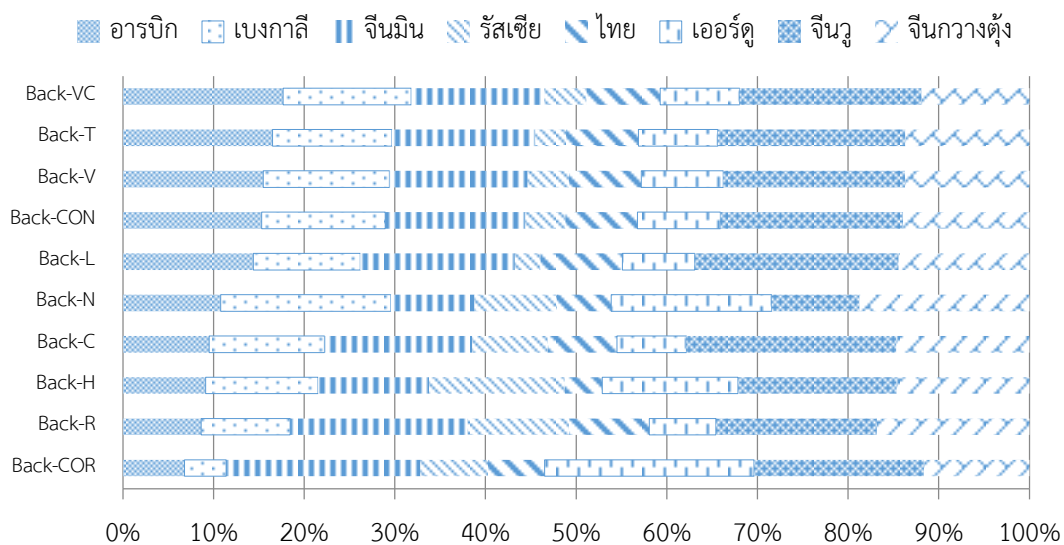
โดยที่ค่าพืเอฟอื่นแสดงด้วยตัวย่อของค่าพืเอฟแต่ละชนิด S แทนค่าพืเอฟ Strident, VC แทนค่าพืเอฟ Vocie, V แทนค่าพืเอฟ Vocalic, A แทนค่าพืเอฟ Anterior, C แทนค่าพืเอฟ Consonant, CON แทนค่าพืเอฟ Continuant, H แทนค่าพืเอฟ High, L แทนค่าพืเอฟ Low, R แทนค่าพืเอฟ Round, COR แทนค่าพืเอฟ Coronal, B แทนค่าพืเอฟ Back, T แทนค่าพืเอฟ Tense, N แทนค่าพืเอฟ Nasal ดังนี้



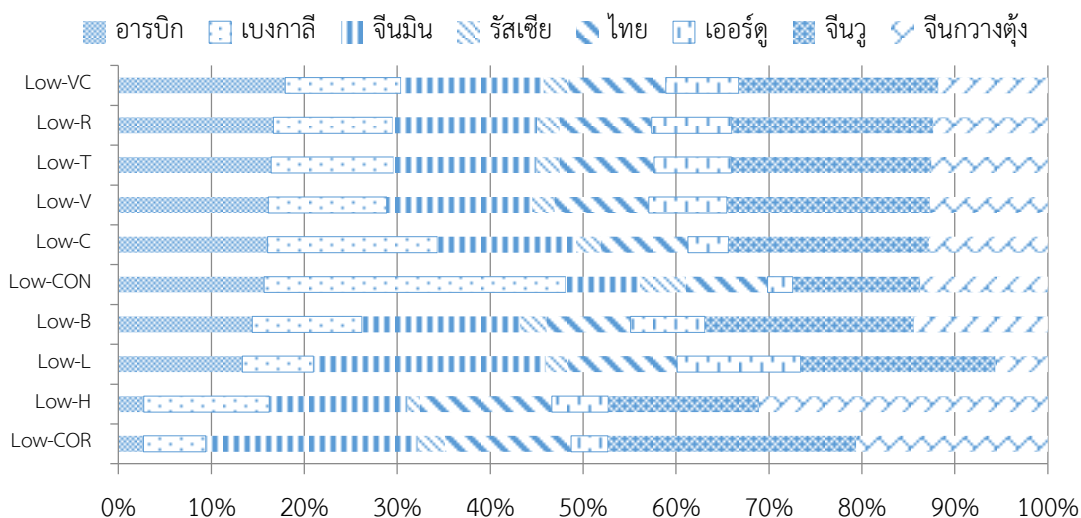
ภาพที่ จ-1 เปอร์เซ็นต์ของความถี่การปรากฏโคออกเคอเรนซ์ค่าพืเอฟหลัก Consonantal ร่วมกับค่าพืเอฟชนิดอื่น ในแต่ละภาษา



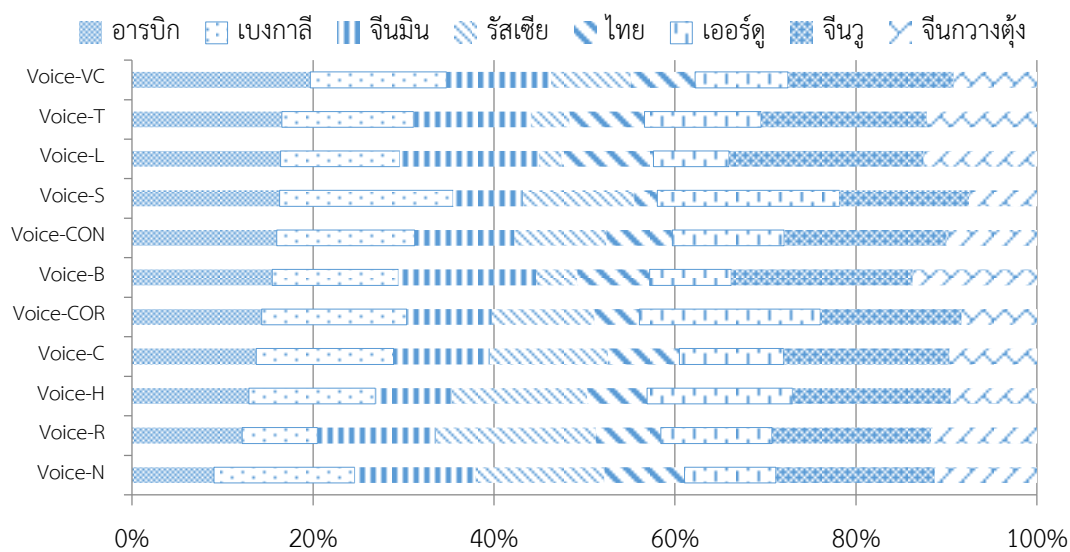
ภาพที่ จ-2 เปอร์เซนต์ของความถี่การปรากฏโคออร์ดิเนตค่าพีเอฟหลัก High ร่วมกับค่าพีเอฟชนิดอื่น ในแต่ละภาษา



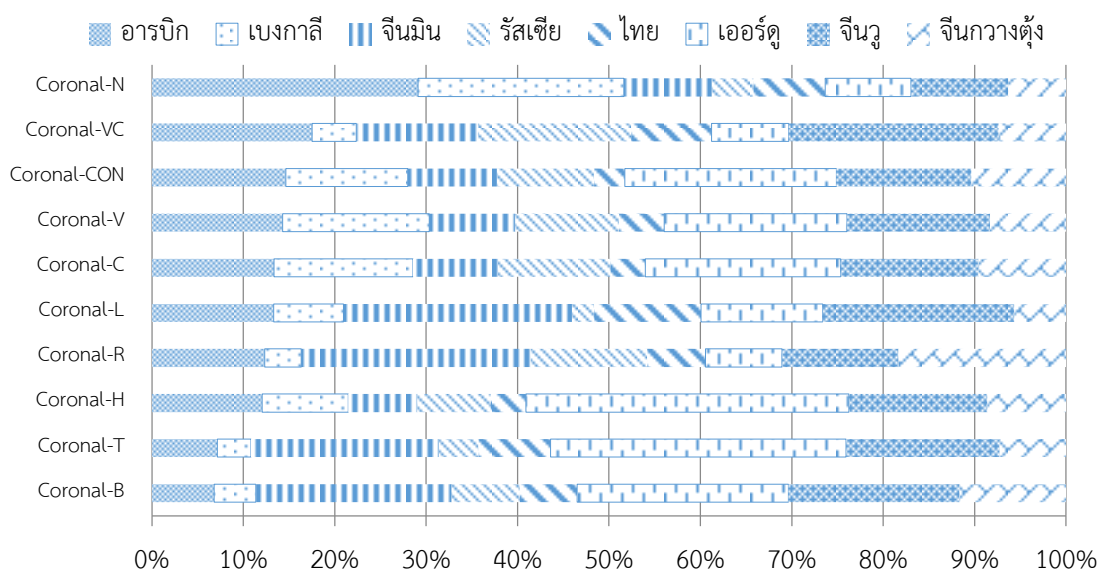
ภาพที่ จ-3 เปอร์เซนต์ของความถี่การปรากฏโคออร์ดิเนตค่าพีเอฟหลัก Back ร่วมกับค่าพีเอฟชนิดอื่น ในแต่ละภาษา



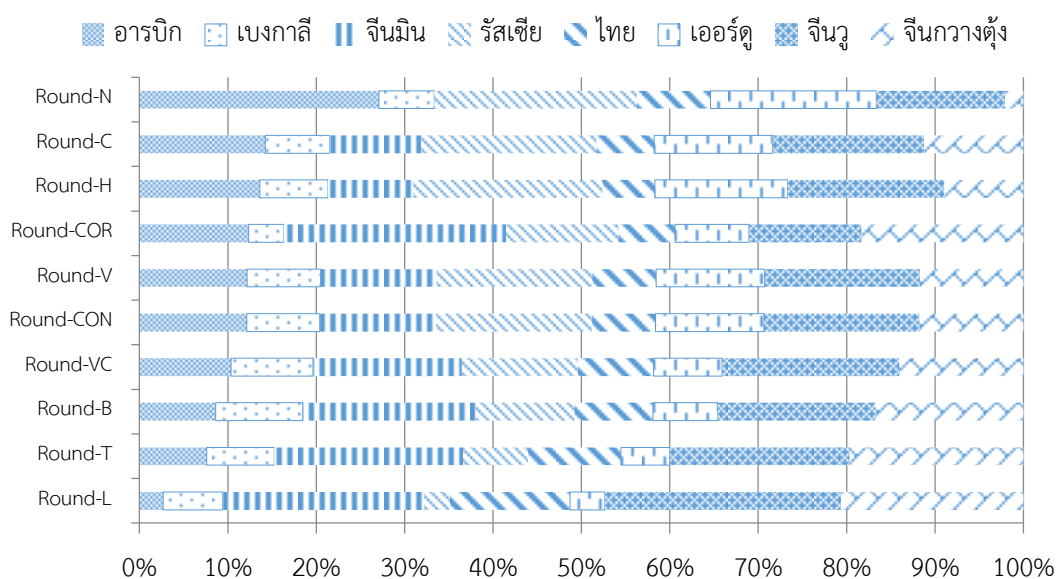
ภาพที่ จ-4 เปอร์เซ็นต์ของความถี่การปรากฏโคออคเคอเรนซ์ค่าพีเอฟหลัก Low ร่วมกับค่าพีเอฟชนิดอื่น ในแต่ละภาษา



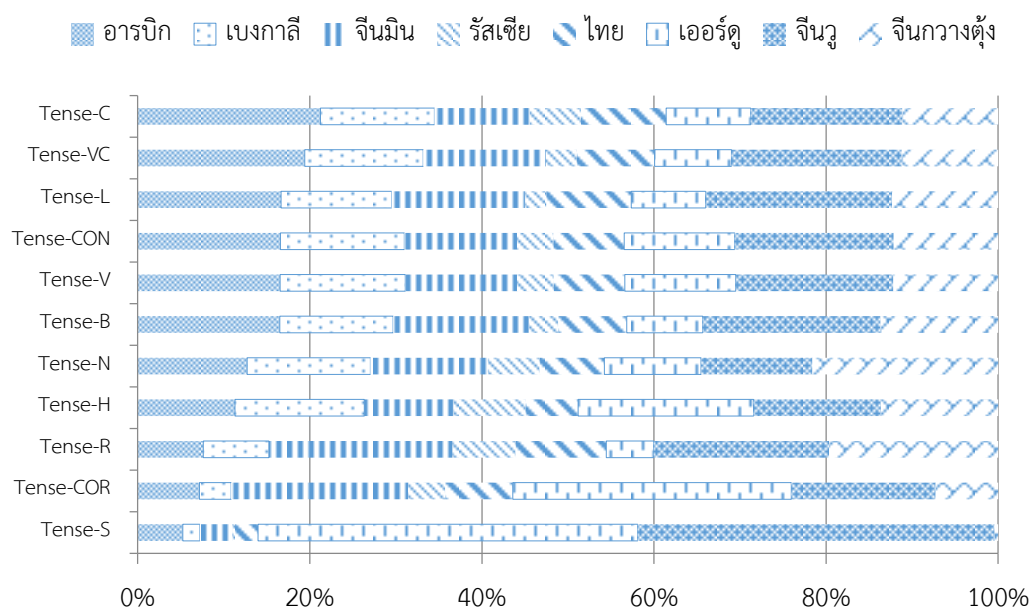
ภาพที่ จ-5 เปอร์เซ็นต์ของความถี่การปรากฏโคออคเคอเรนซ์ค่าพีเอฟหลัก Voice ร่วมกับค่าพีเอฟชนิดอื่น ในแต่ละภาษา



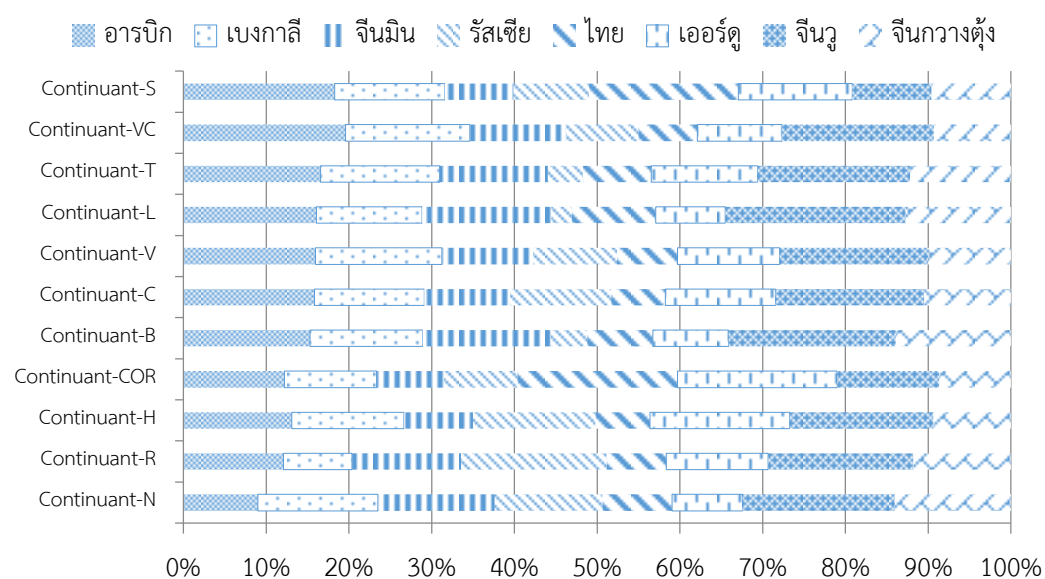
ภาพที่ จ-6 เปอร์เซนต์ของความถี่การปรากฏโคออกเคอเรนซ์ค่าพีเอฟ Coronal ร่วมกับค่าพีเอฟชนิดอื่น ในแต่ละภาษา



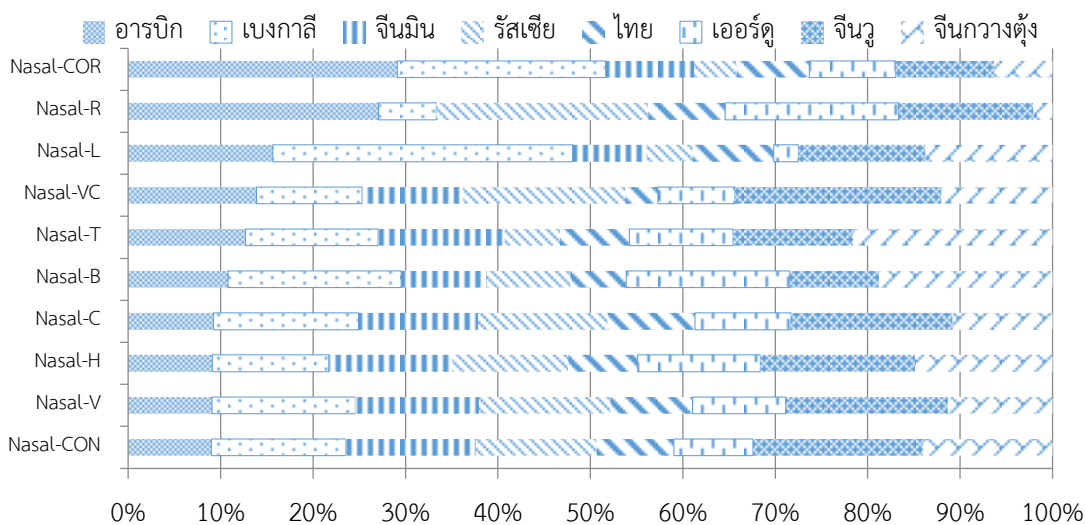
ภาพที่ จ-7 เปอร์เซนต์ของความถี่การปรากฏโคออกเคอเรนซ์ค่าพีเอฟ Round ร่วมกับค่าพีเอฟชนิดอื่น ในแต่ละภาษา



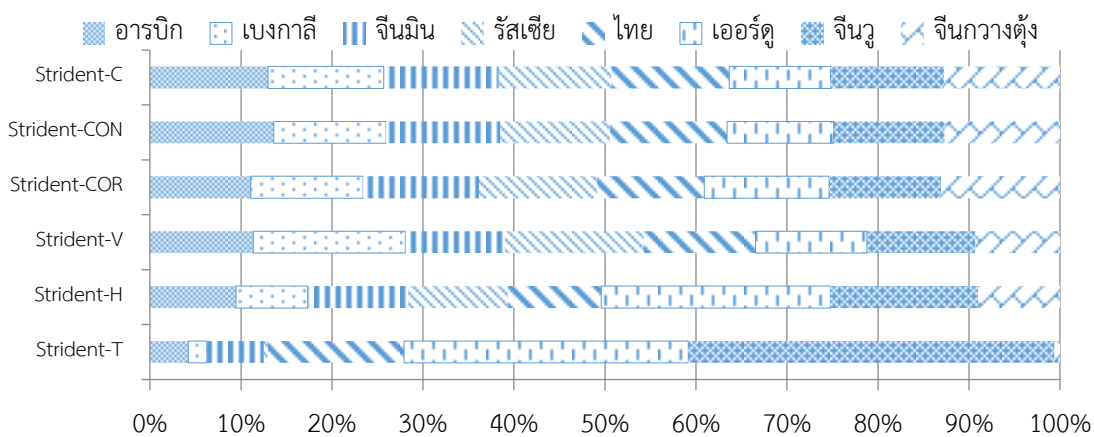
ภาพที่ จ-8 เปอร์เซ็นต์ของความถี่การปรากฏโคออกเคอเรนซ์ค่าพีเอฟ Tense ร่วมกับค่าพีเอฟชนิดอื่น ในแต่ละภาษา



ภาพที่ จ-9 เปอร์เซ็นต์ของความถี่การปรากฏโคออกเคอเรนซ์ค่าพีเอฟ Continuant ร่วมกับค่าพีเอฟชนิดอื่น ในแต่ละภาษา



ภาพที่ จ-10 เปอร์เซนต์ของความถี่การปรากฏโคออกเคอเรนซ์ค่าพีเอฟ Nasal ร่วมกับค่าพีเอฟชนิดอื่น ในแต่ละภาษา



ภาพที่ จ-11 เปอร์เซนต์ของความถี่การปรากฏโคออกเคอเรนซ์ค่าพีเอฟ Strident ร่วมกับค่าพีเอฟชนิดอื่น ในแต่ละภาษา

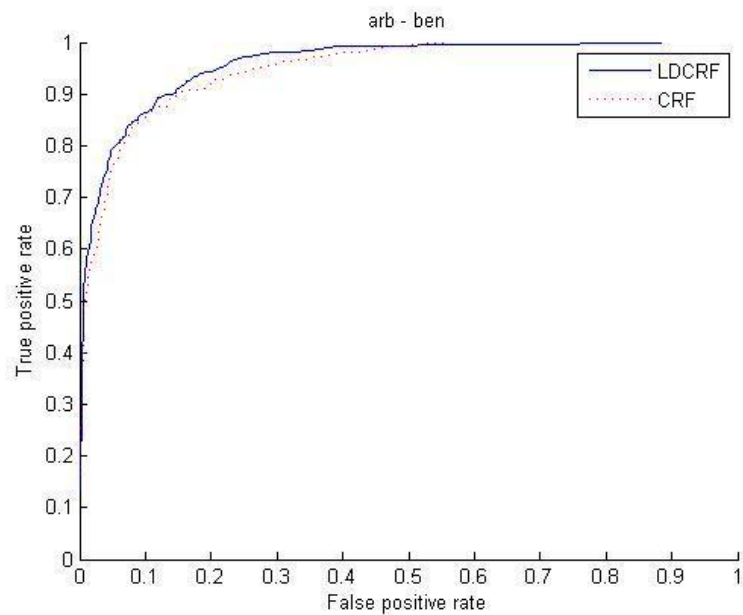
ภาคผนวก ข ผลการศึกษาค่าพีเอฟด้วยการวิเคราะห์ความแปรปรวนพหุคูณ สำหรับแต่ละคู่ภาษา

ตารางที่ ข-1 ผลการศึกษาค่าพีเอฟด้วยการวิเคราะห์ความแปรปรวนพหุคูณ
สำหรับแต่ละคู่ภาษา

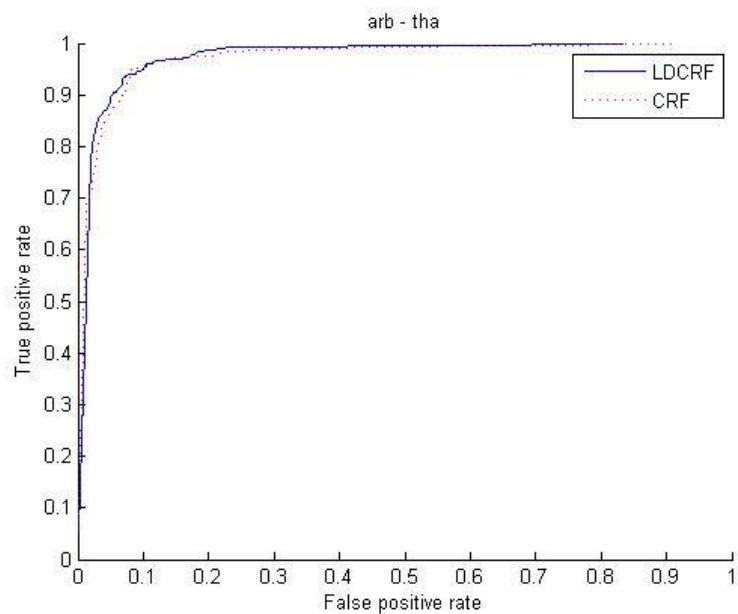
คู่ภาษา	ค่าสถิติ	df1 (chisqdf)	df2(dfW)	F	P	นัยสำคัญ	Lamda
arb-ben	0.032	13	623921	1536.37	0	S	0.969
arb-rus	0.135	13	552406	5720.11	0	S	0.8814
arb-tha	0.069	13	453451	2420.29	0	S	0.9351
arb-urd	0.073	13	543556	3045.22	0	S	0.9321
arb-CH.wu	0.049	13	700311	2640.49	0	S	0.9533
arb-CH.can	0.064	13	508841	2503.9	0	S	0.9399
ben-CH.min	0.031	13	604401	1460.92	0	S	0.9695
ben-rus	0.069	13	585781	3108.41	0	S	0.9355
ben-tha	0.03	13	486826	1114.08	0	S	0.9711
ben-urd	0.041	13	576931	1816.65	0	S	0.9607
ben-wuu	0.016	13	733686	882.43	0	S	0.9846
ben-CH.can	0.028	13	542216	1151.44	0	S	0.9731
CH.min-rus	0.098	13	532886	4009.63	0	S	0.9109
CH.min-tha	0.022	13	433931	746.82	0	S	0.9781
CH.min-urd	0.089	13	524036	3578.88	0	S	0.9185
CH.min-CH.wu	0.011	13	680791	553.86	0	S	0.9895
CH.min-CH.can	0.015	13	489321	572.46	0	S	0.985
rus-tha	0.121	13	415311	3854.31	0	S	0.8923
rus-urd	0.107	13	505416	4175.81	0	S	0.903
rus-CH.wu	0.068	13	662171	3473.09	0	S	0.9362
rus-CH.can	0.121	13	470701	4384.01	0	S	0.892

คู่ภาษา	ค่าสถิติ	df1 (chisqdf)	df2(dfW)	F	P	นัยสำคัญ	Lamda
tha-urd	0.094	13	406461	2927.85	0	S	0.9144
tha-CH.wu	0.016	13	563216	713.85	0	S	0.9838
tha-CH.can	0.031	13	371746	892.21	0	S	0.9697
urd-CH.wu	0.064	13	653321	3226.02	0	S	0.9397
urd-CH.can	0.055	13	461851	1948.07	0	S	0.948
CH.wu-CH.can	0.017	13	618606	799.3	0	S	0.9835

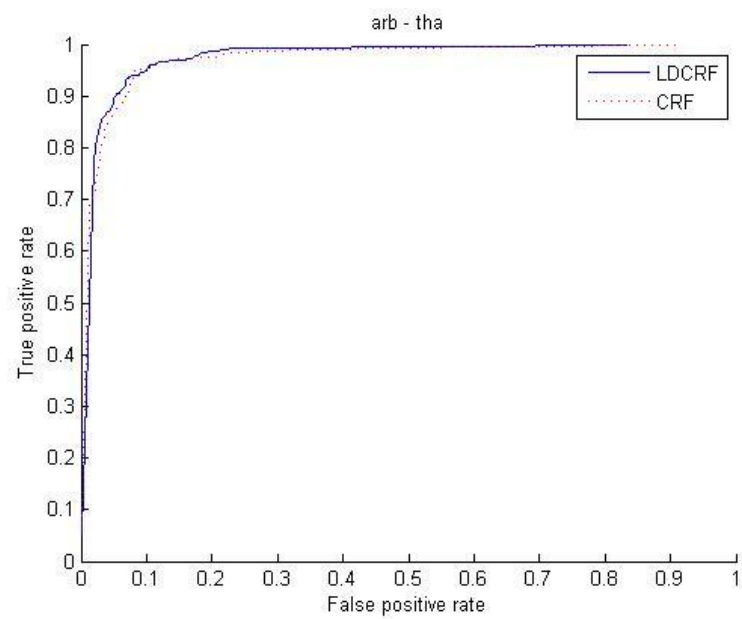
ภาคผนวก ข เส้นโค้ง ROC ของผลการจำแนกภาษาโดยใช้แบบจำลองลาเทนท์ไดนามิกคอนดิชันนอลแรนดอมฟิลด์ร่วมกับค่าพีเอฟ และแบบจำลองคอนดิชันนอลแรนดอมฟิลด์ร่วมกับค่าพีเอฟ สำหรับแต่ละคู่ภาษา



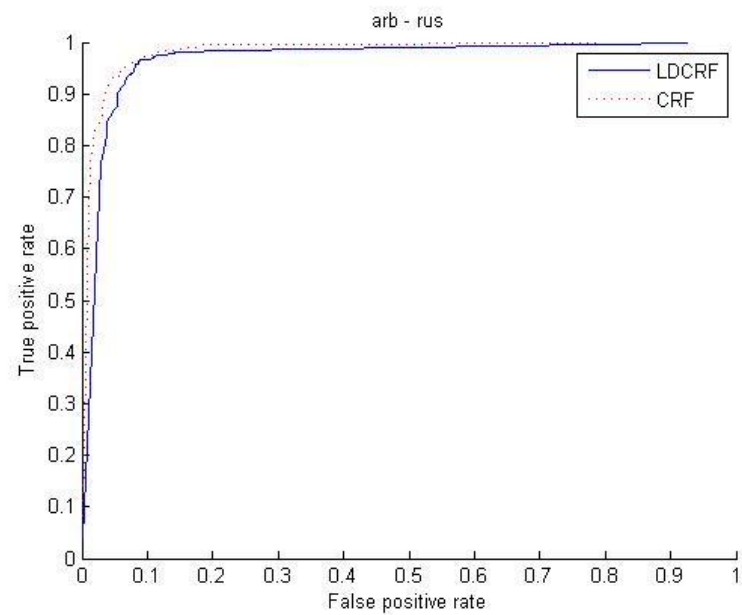
ภาพที่ ข-1 เส้นโค้ง ROC ของผลการจำแนกภาษาของแบบจำลอง LDCRF และ CRF สำหรับคู่ภาษาอารบิก และภาษาเบงกาลี



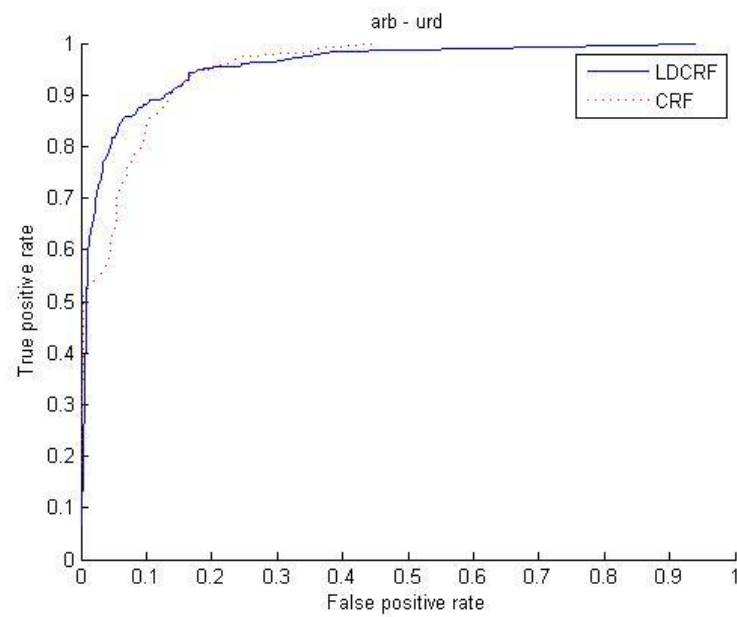
ภาพที่ ข-2 เส้นโค้ง ROC ของผลการจำแนกภาษาของแบบจำลอง LDCRF และ CRF สำหรับคู่ภาษาอารบิก และภาษาจีนมิน



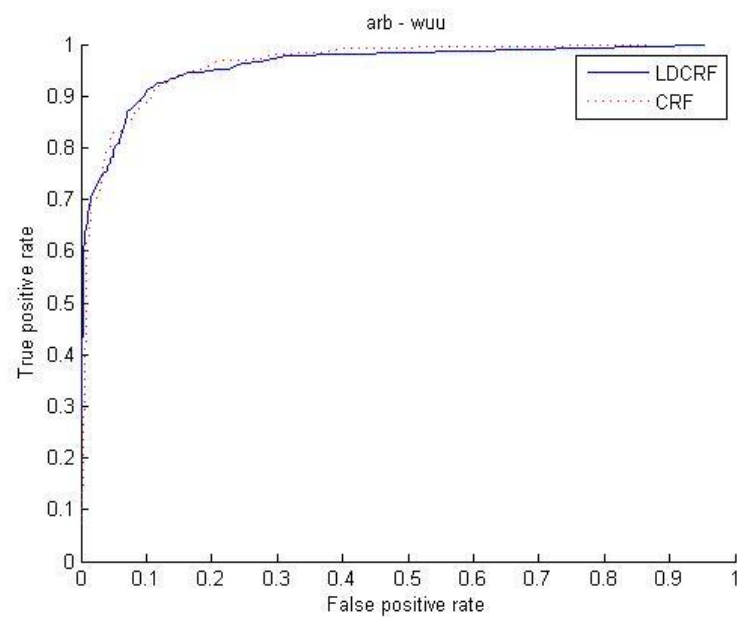
ภาพที่ ข-3 เส้นโค้ง ROC ของผลการจำแนกภาษาของแบบจำลอง LDCRF และ CRF สำหรับคู่ภาษาอารบิก และภาษาไทย



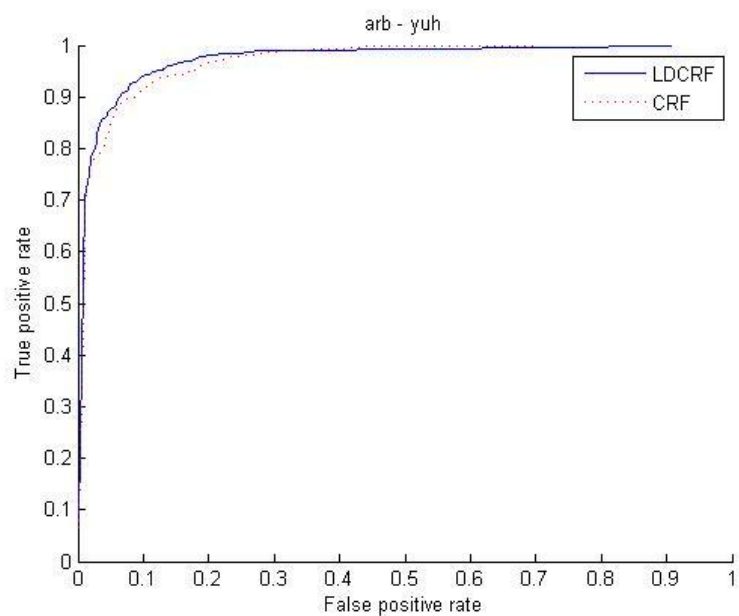
ภาพที่ ข-4 เส้นโค้ง ROC ของผลการจำแนกภาษาของแบบจำลอง LDCRF และ CRF สำหรับคู่ภาษาอารบิก และภาษารัสเซีย



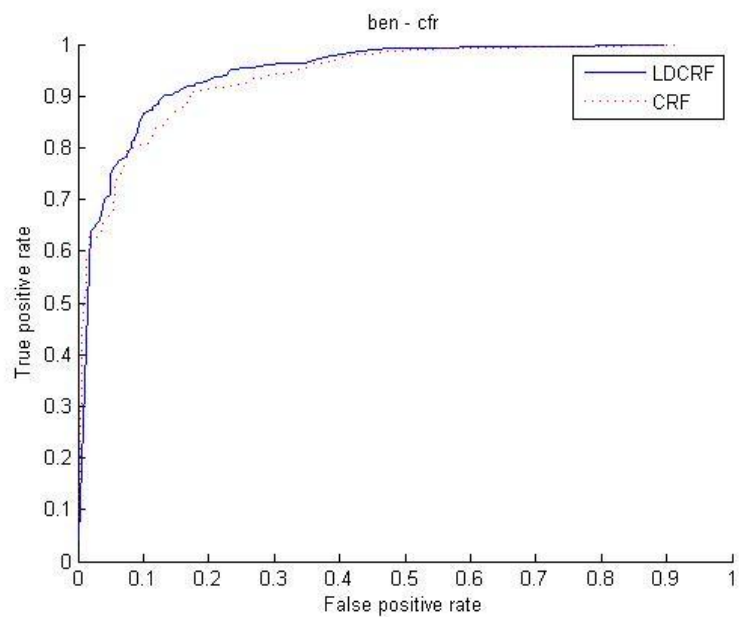
ภาพที่ ข-5 เส้นโค้ง ROC ของผลการจำแนกภาษาของแบบจำลอง LDCRF และ CRF สำหรับคู่ภาษาอารบิก และภาษาเออร์ดู



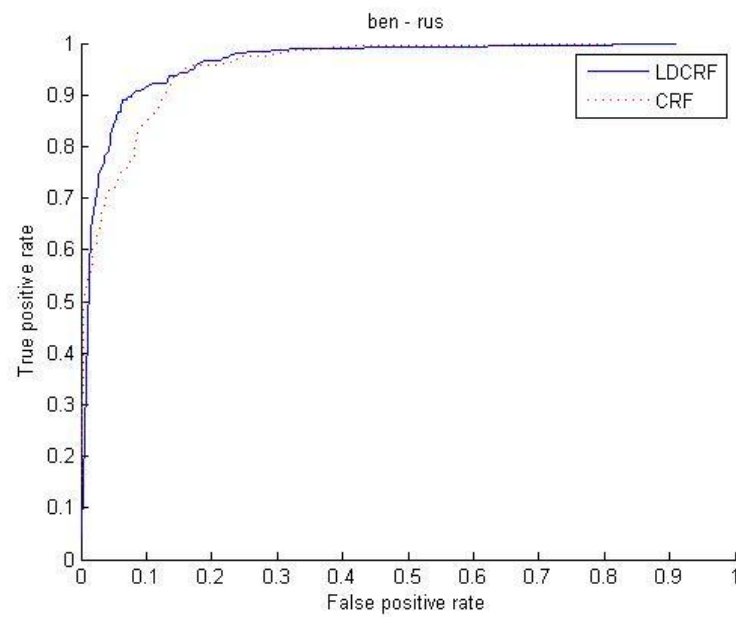
ภาพที่ ข-6 เส้นโค้ง ROC ของผลการจำแนกภาษาของแบบจำลอง LDCRF และ CRF สำหรับคู่ภาษาอารบิก และภาษาจิ้นงู



ภาพที่ ข-7 เส้นโค้ง ROC ของผลการจำแนกภาษาของแบบจำลอง LDCRF และ CRF สำหรับคู่ภาษาอารบิก และภาษาจีนกวางตุ้ง

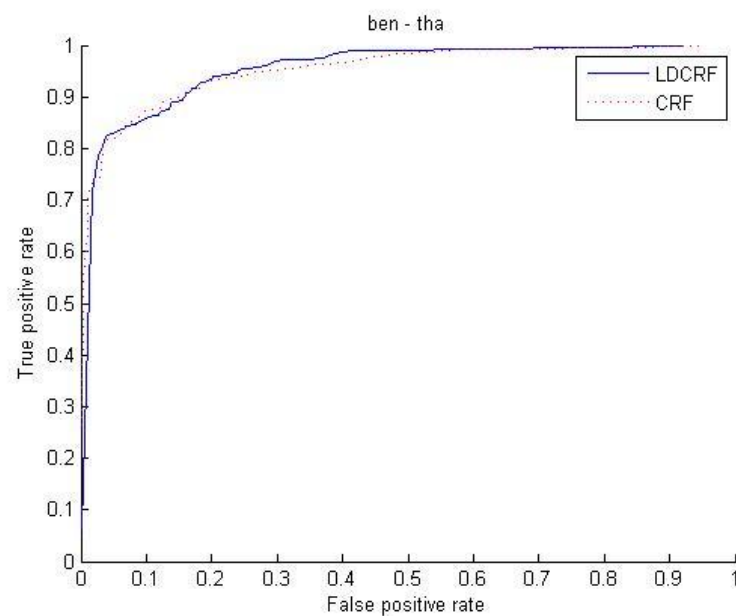


ภาพที่ ข-8 เส้นโค้ง ROC ของผลการจำแนกภาษาของแบบจำลอง LDCRF และ CRF สำหรับคู่ภาษาเบงกาลี และภาษาจีนมิน

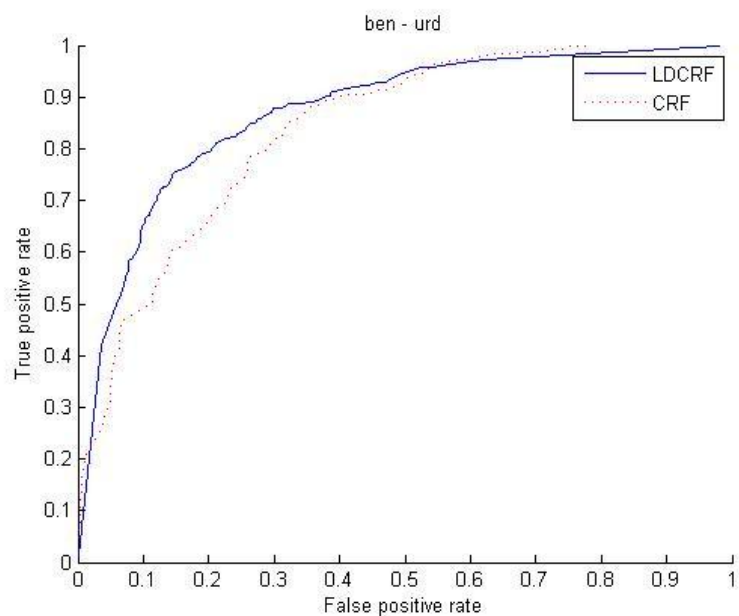


ภาพที่ ข-9 เส้นโค้ง ROC ของผลการจำแนกภาษาของแบบจำลอง LDCRF และ CRF
สำหรับ

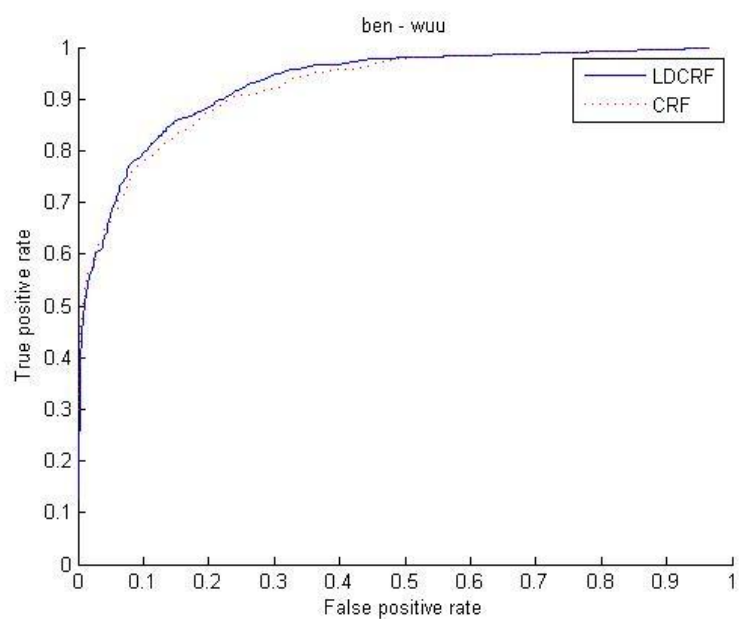
คู่ภาษาเบงกาลี และภาษารัสเซีย



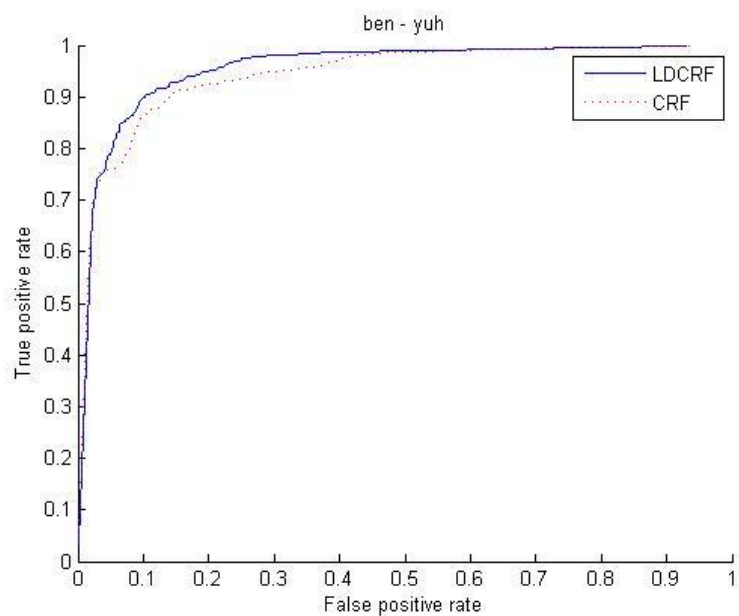
ภาพที่ ข-10 เส้นโค้ง ROC ของผลการจำแนกภาษาของแบบจำลอง LDCRF และ CRF
สำหรับคู่ภาษาเบงกาลี และภาษาไทย



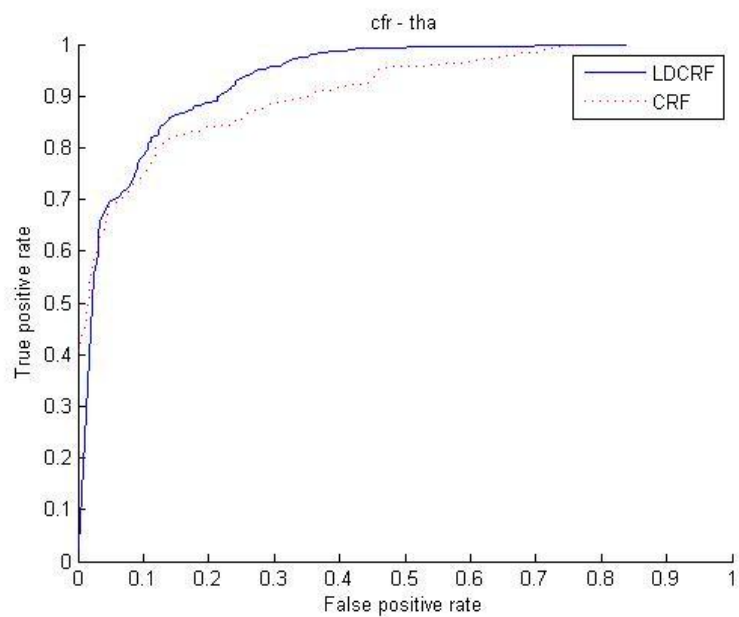
ภาพที่ ข-11 เส้นโค้ง ROC ของผลการจำแนกภาษาของแบบจำลอง LDCRF และ CRF สำหรับคู่ภาษาเบงกาลี และภาษาเอร์ดู



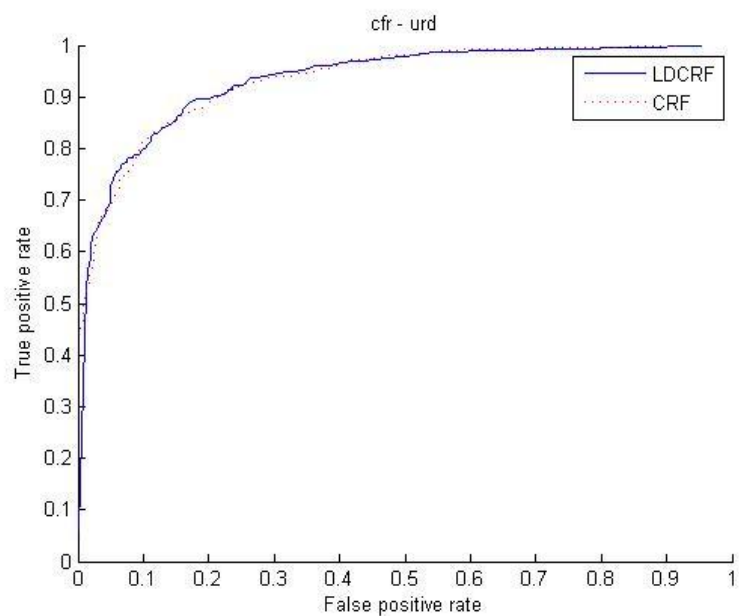
ภาพที่ ข-12 เส้นโค้ง ROC ของผลการจำแนกภาษาของแบบจำลอง LDCRF และ CRF สำหรับคู่ภาษาเบงกาลี และภาษาจื๋นงู



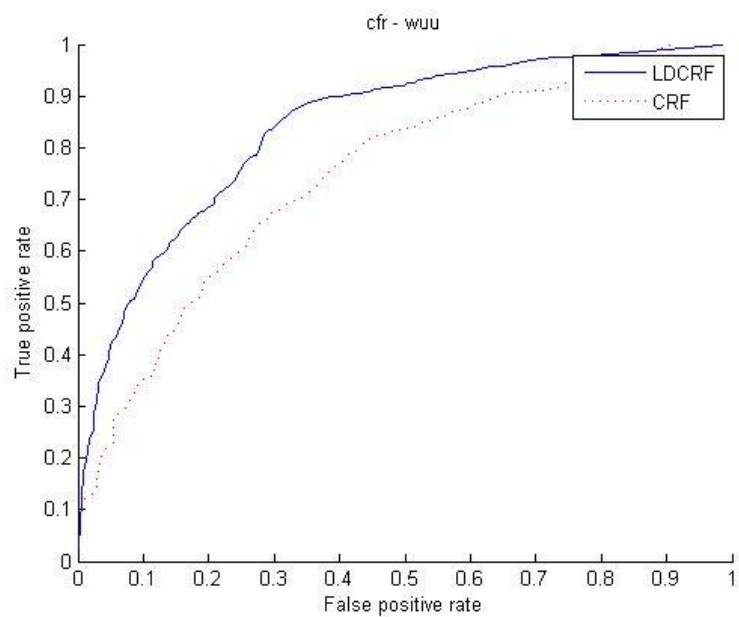
ภาพที่ ข-13 เส้นโค้ง ROC ของผลการจำแนกภาษาของแบบจำลอง LDCRF และ CRF สำหรับคู่ภาษาเบงกาลี และภาษาจันกวางตุ้ง



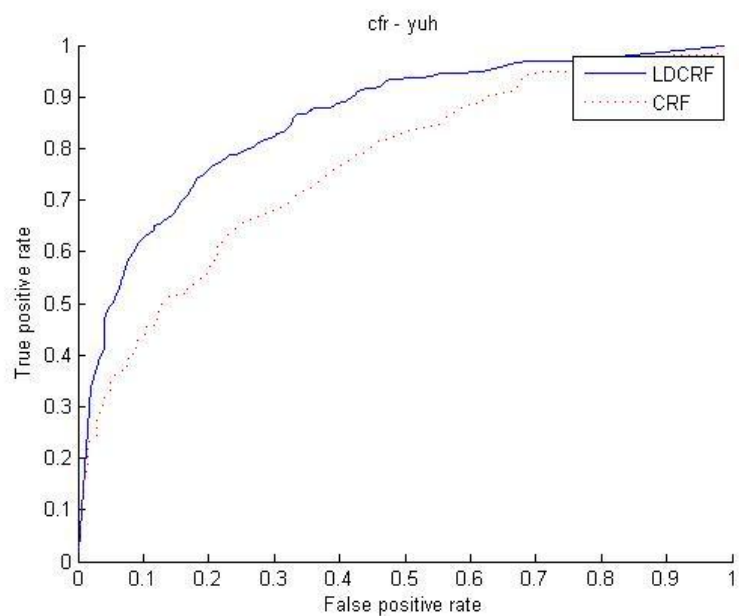
ภาพที่ ข-14 เส้นโค้ง ROC ของผลการจำแนกภาษาของแบบจำลอง LDCRF และ CRF สำหรับคู่ภาษาจันมิน และภาษาไทย



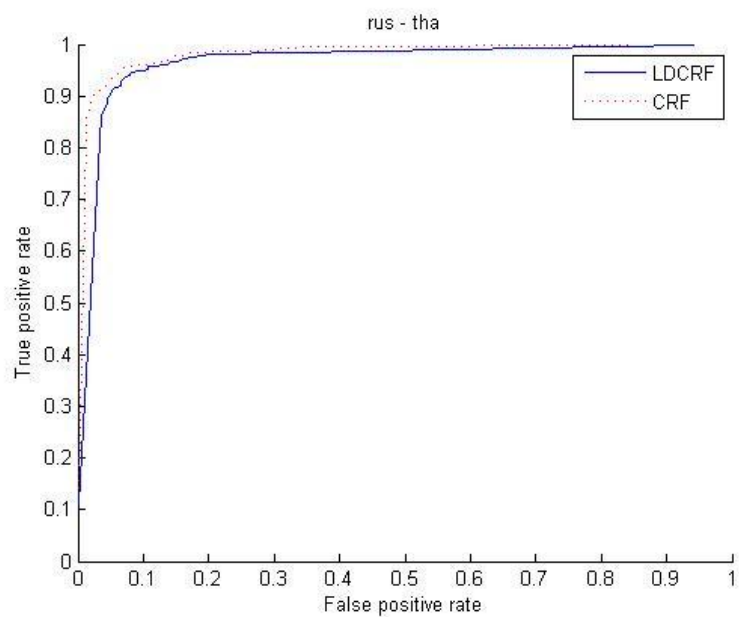
ภาพที่ ข-15 เส้นโค้ง ROC ของผลการจำแนกภาษาของแบบจำลอง LDCRF และ CRF สำหรับคู่ภาษาภาษาจีนมิน และภาษาเออร์ดู



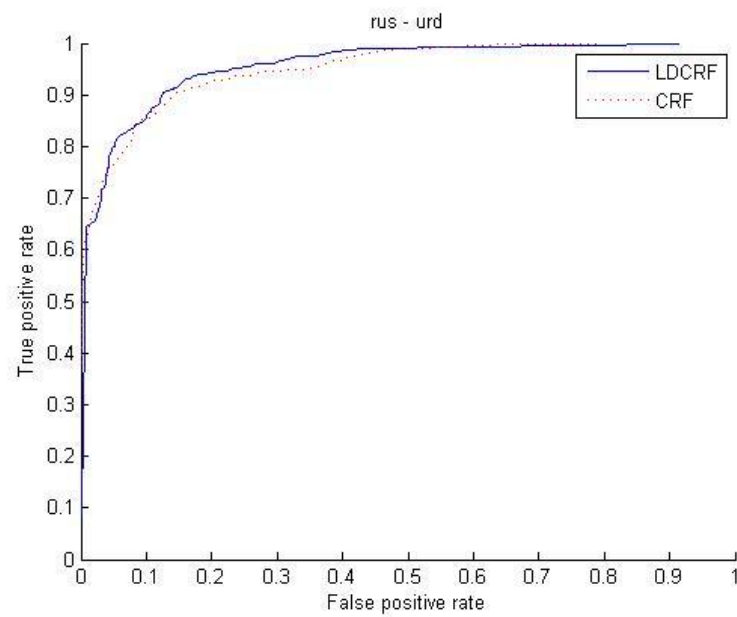
ภาพที่ ข-16 เส้นโค้ง ROC ของผลการจำแนกภาษาของแบบจำลอง LDCRF และ CRF สำหรับคู่ภาษาจีนมิน และภาษาจีนวู



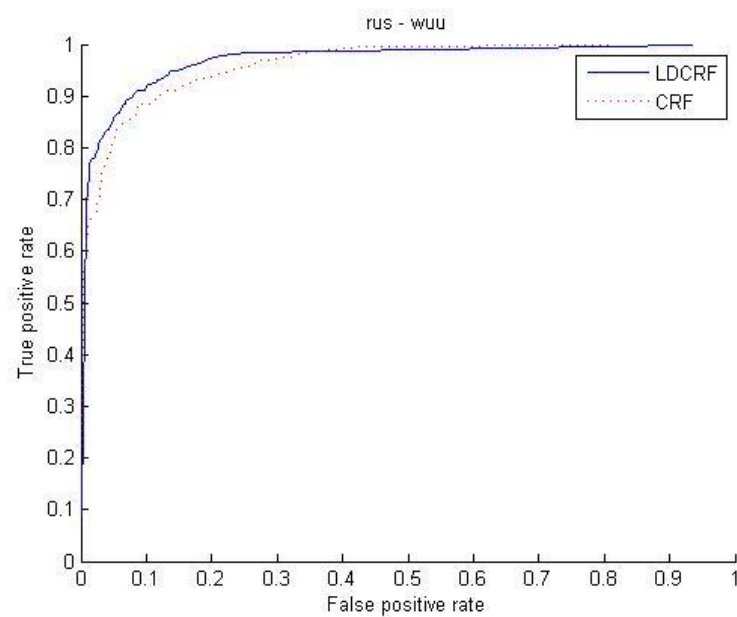
ภาพที่ ข-17 เส้นโค้ง ROC ของผลการจำแนกภาษาของแบบจำลอง LDCRF และ CRF สำหรับคู่ภาษาจีนมิน และภาษาจีนกวางตุ้ง



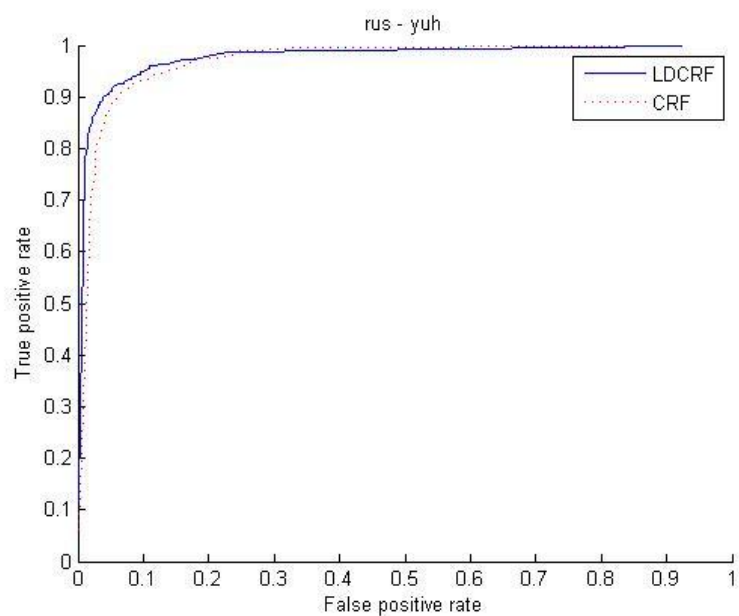
ภาพที่ ข-18 เส้นโค้ง ROC ของผลการจำแนกภาษาของแบบจำลอง LDCRF และ CRF สำหรับคู่ภาษาจีนมิน และภาษาจีนกวางตุ้ง



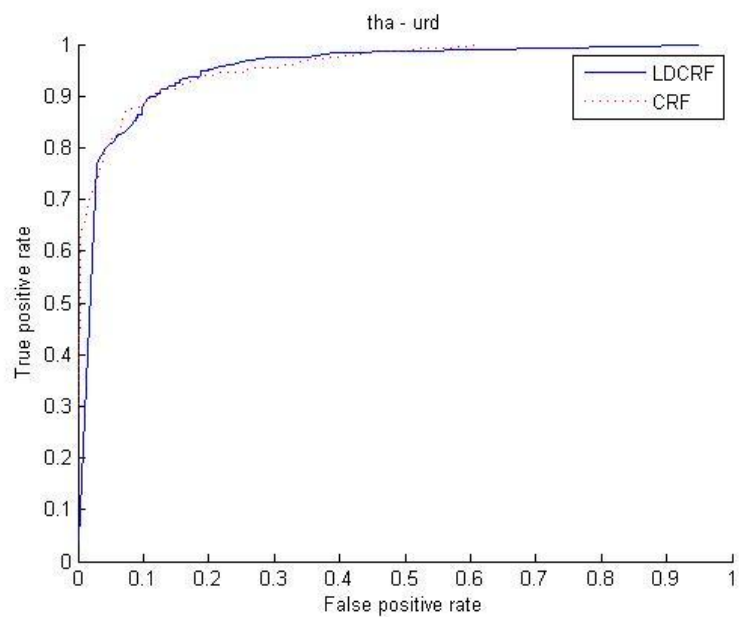
ภาพที่ ข-19 เส้นโค้ง ROC ของผลการจำแนกภาษาของแบบจำลอง LDCRF และ CRF สำหรับคู่ภาษารัสเซีย และภาษาเออร์ดู



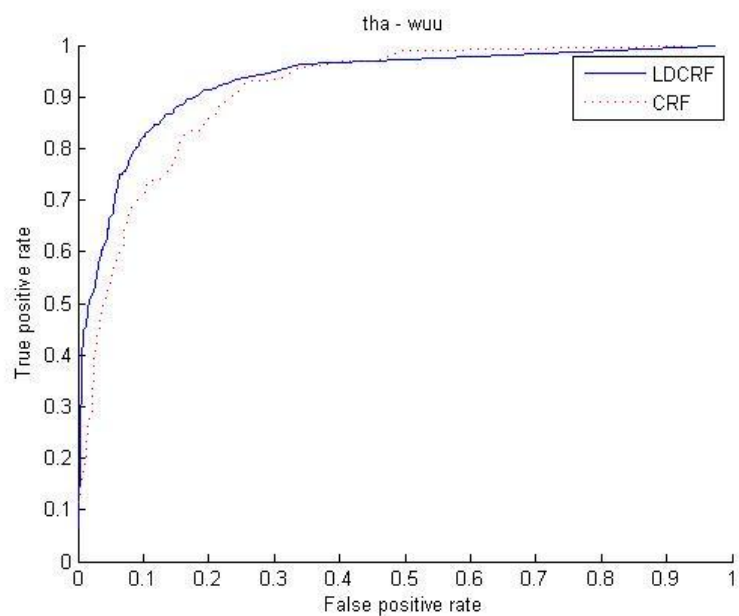
ภาพที่ ข-20 เส้นโค้ง ROC ของผลการจำแนกภาษาของแบบจำลอง LDCRF และ CRF สำหรับคู่ภาษารัสเซีย และภาษาจื๋นงู



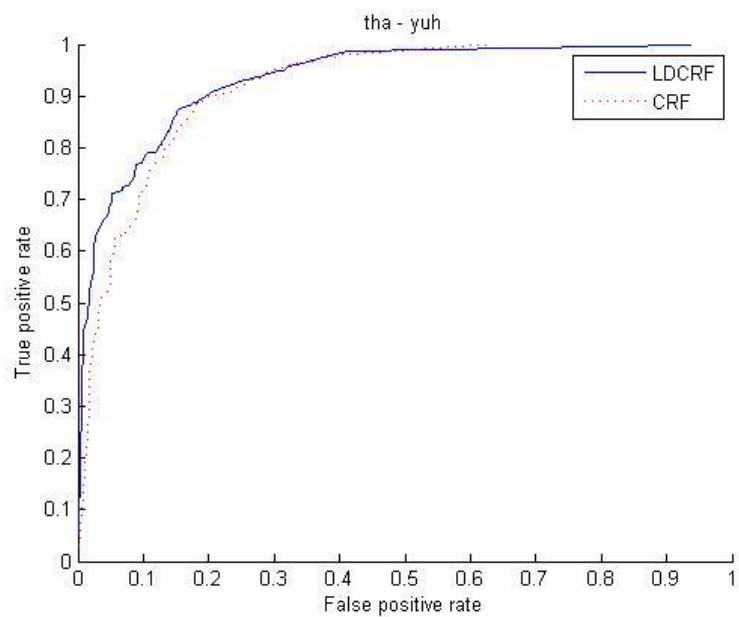
ภาพที่ ข-21 เส้นโค้ง ROC ของผลการจำแนกภาษาของแบบจำลอง LDCRF และ CRF สำหรับคู่ภาษารัสเซีย และภาษาจีกวางต้ง



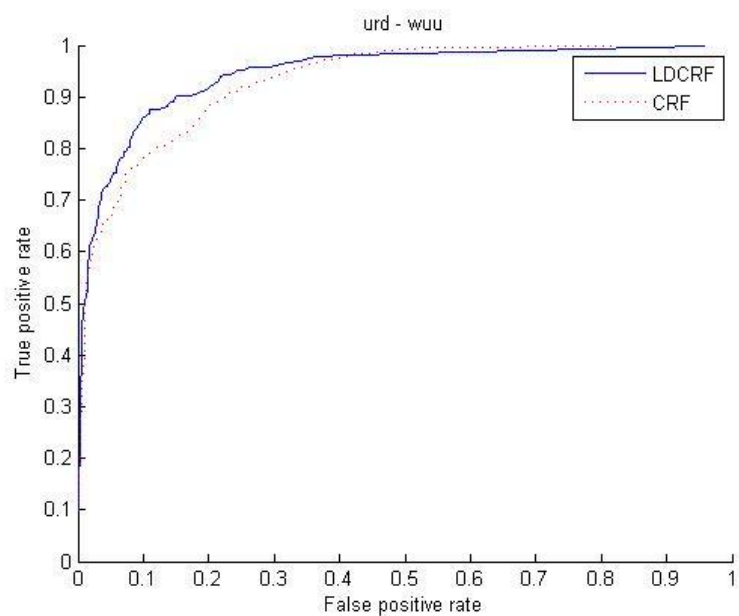
ภาพที่ ข-22 เส้นโค้ง ROC ของผลการจำแนกภาษาของแบบจำลอง LDCRF และ CRF สำหรับคู่ภาษาไทย และภาษาเออร์ดุ



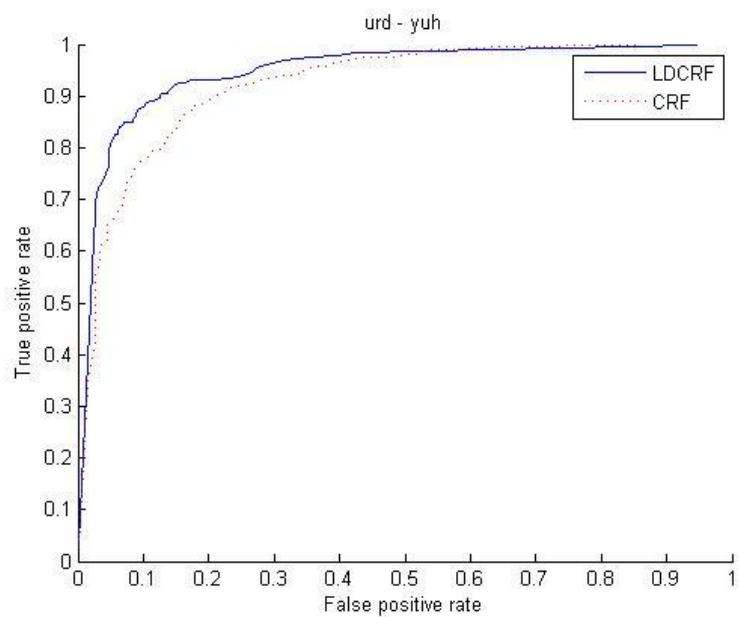
ภาพที่ ข-23 เส้นโค้ง ROC ของผลการจำแนกภาษาของแบบจำลอง LDCRF และ CRF สำหรับคู่ภาษาไทย และภาษาจีน



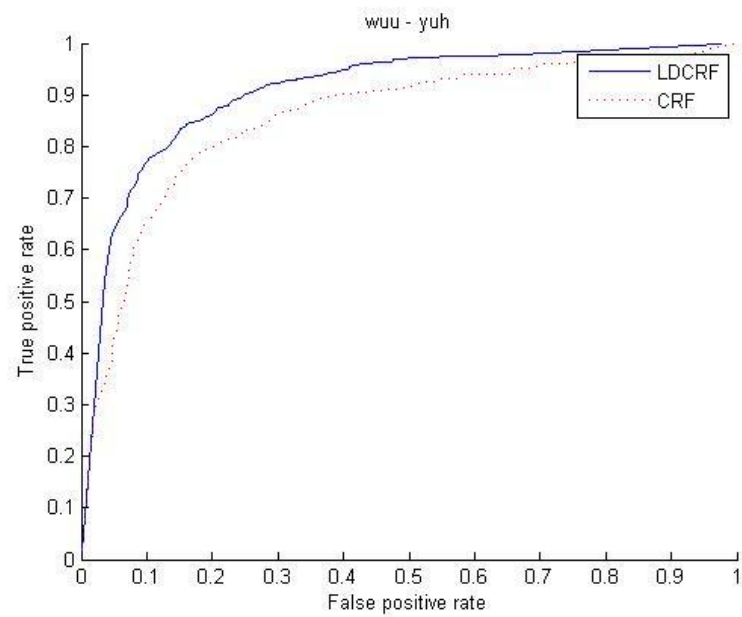
ภาพที่ ข-24 เส้นโค้ง ROC ของผลการจำแนกภาษาของแบบจำลอง LDCRF และ CRF สำหรับคู่ภาษาไทย และภาษาจีนกวางตุ้ง



ภาพที่ ข-25 เส้นโค้ง ROC ของผลการจำแนกภาษาของแบบจำลอง LDCRF และ CRF สำหรับคู่ภาษาเออร์ดุ และภาษาจิ้นวู

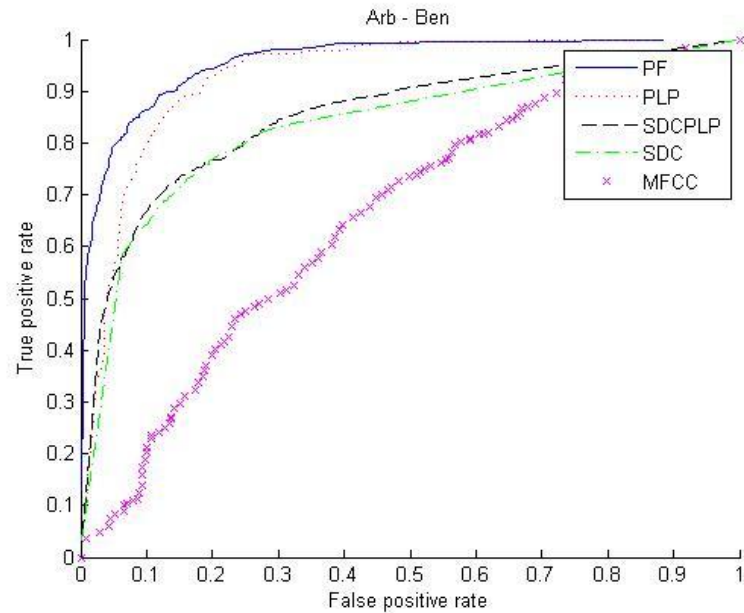


ภาพที่ ข-26 เส้นโค้ง ROC ของผลการจำแนกภาษาของแบบจำลอง LDCRF และ CRF สำหรับคู่ภาษาเออร์ดุ และภาษาจิ้นกวางตุ้ง

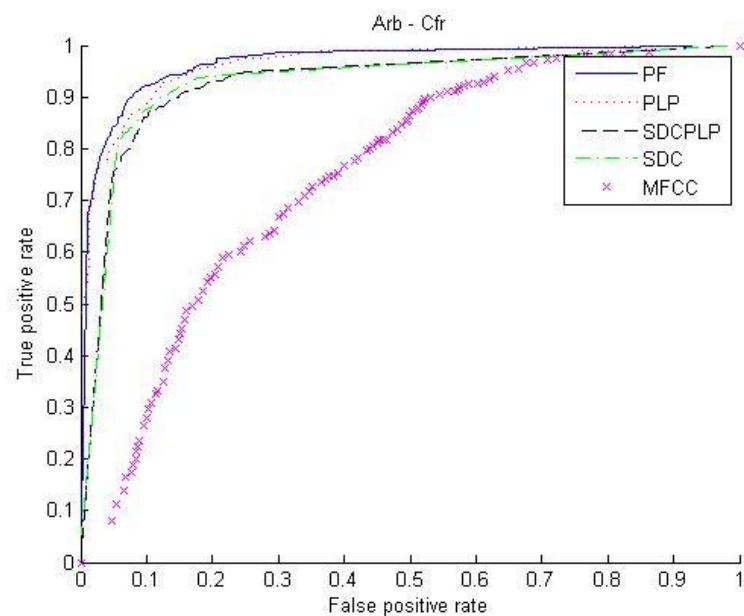


ภาพที่ ช-27 เส้นโค้ง ROC ของผลการจำแนกภาษาของแบบจำลอง LDCRF และ CRF สำหรับคู่ภาษาจีนวู และภาษาจีนกวางตุ้ง

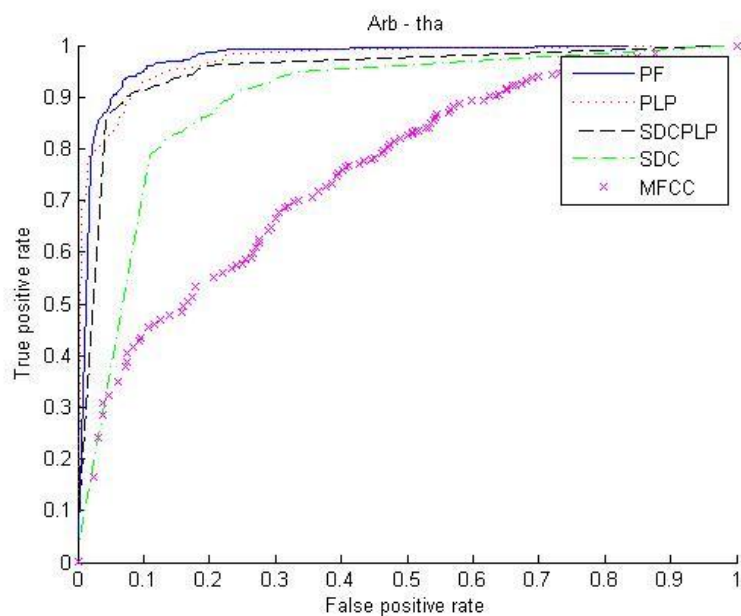
ภาคผนวก ฅ เส้นโค้ง ROC ของผลการจำแนกภาษาโดยใช้แบบจำลองลาเทนท์ไดนามิกคอนดิชันนอลแรนดอมฟิลด์ร่วมกับค่าพีเอฟ และค่าลักษณะสำคัญต่าง ๆ สำหรับแต่ละคู่ภาษา



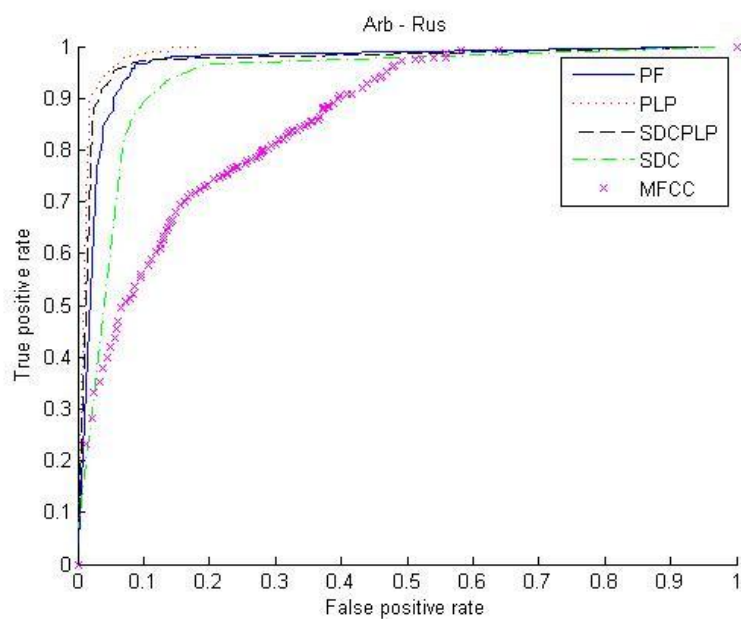
ภาพที่ ฅ-1 เส้นโค้ง ROC ของผลการจำแนกภาษาของแบบจำลอง LDCRF เมื่อใช้ค่าลักษณะสำคัญชนิดต่าง ๆ สำหรับคู่ภาษาอารบิก และภาษาเบงกาลี



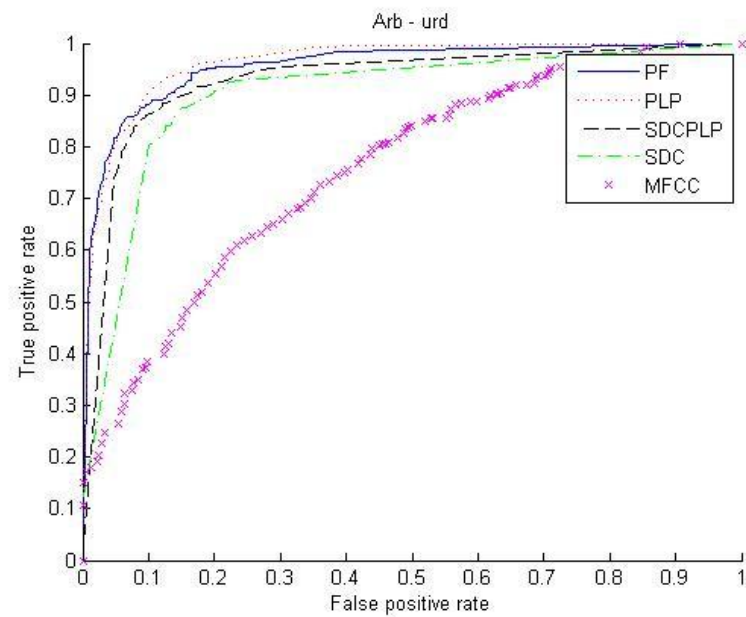
ภาพที่ ฅ-2 เส้นโค้ง ROC ของผลการจำแนกภาษาของแบบจำลอง LDCRF เมื่อใช้ค่าลักษณะสำคัญชนิดต่าง ๆ สำหรับคู่ภาษาอารบิก และภาษาจีนมิน



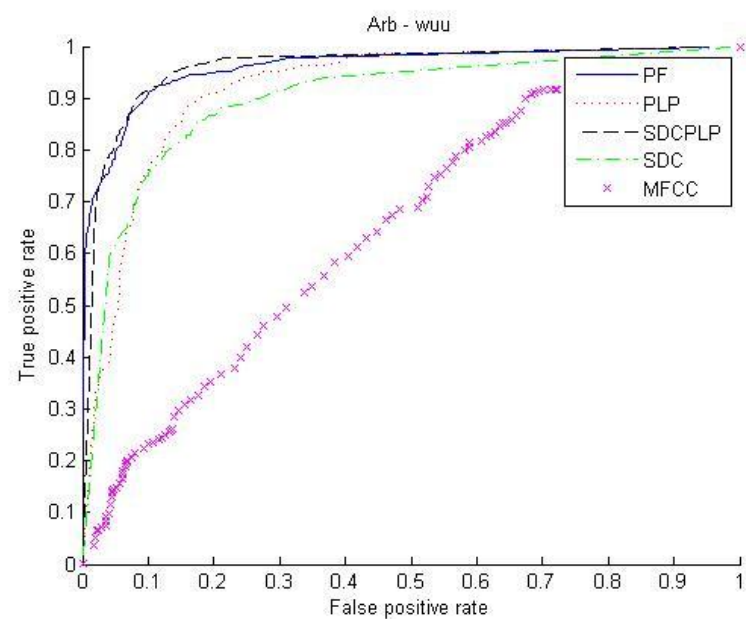
ภาพที่ ๓-3 เส้นโค้ง ROC ของผลการจำแนกภาษาของแบบจำลอง LDCRF เมื่อใช้ค่าลักษณะสำคัญชนิดต่าง ๆ สำหรับคู่ภาษาอารบิก และภาษาไทย



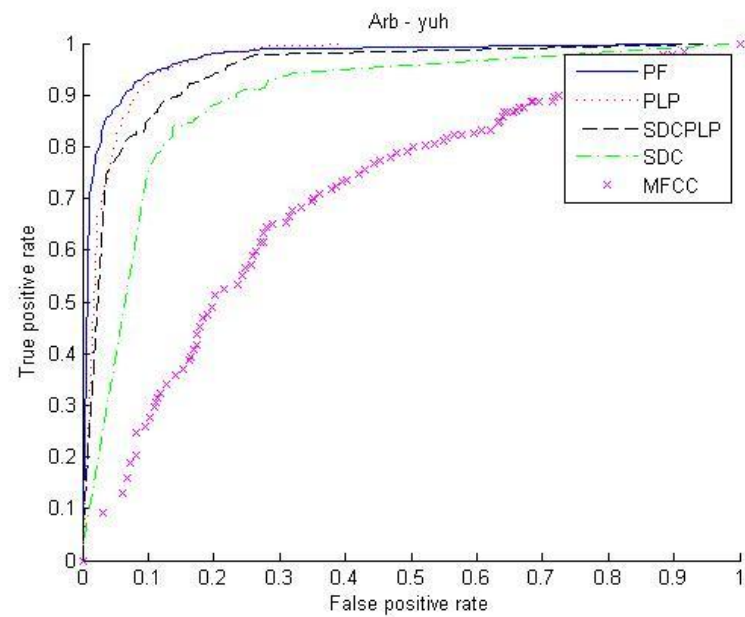
ภาพที่ ๓-4 เส้นโค้ง ROC ของผลการจำแนกภาษาของแบบจำลอง LDCRF เมื่อใช้ค่าลักษณะสำคัญชนิดต่าง ๆ สำหรับคู่ภาษาอารบิก และภาษารัสเซีย



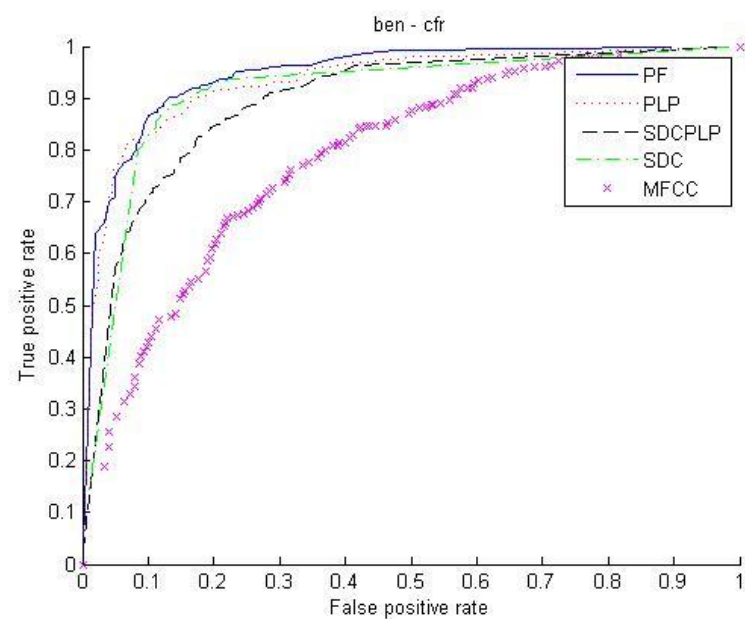
ภาพที่ ฅ-5 เส้นโค้ง ROC ของผลการจำแนกภาษาของแบบจำลอง LDCRF เมื่อใช้ค่าลักษณะสำคัญชนิดต่าง ๆ สำหรับคู่ภาษาอารบิก และภาษาเออร์ดู



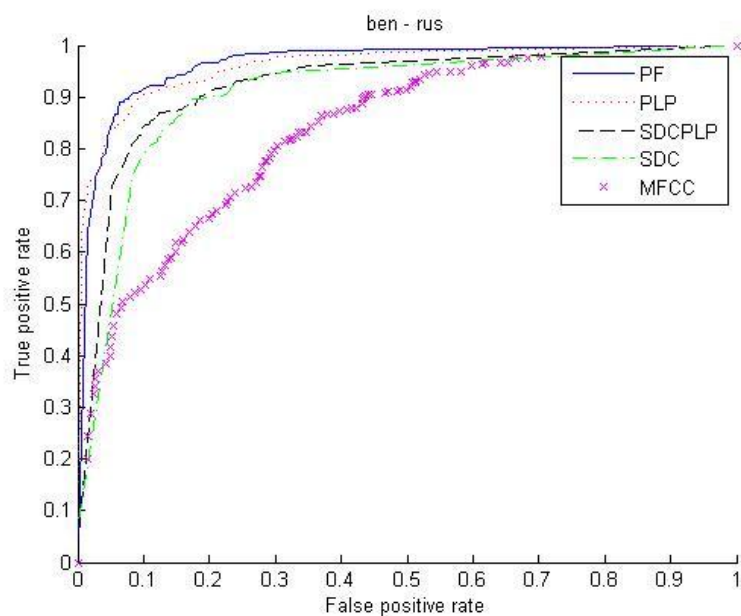
ภาพที่ ฅ-6 เส้นโค้ง ROC ของผลการจำแนกภาษาของแบบจำลอง LDCRF เมื่อใช้ค่าลักษณะสำคัญชนิดต่าง ๆ สำหรับคู่ภาษาอารบิก และภาษาจีนวู



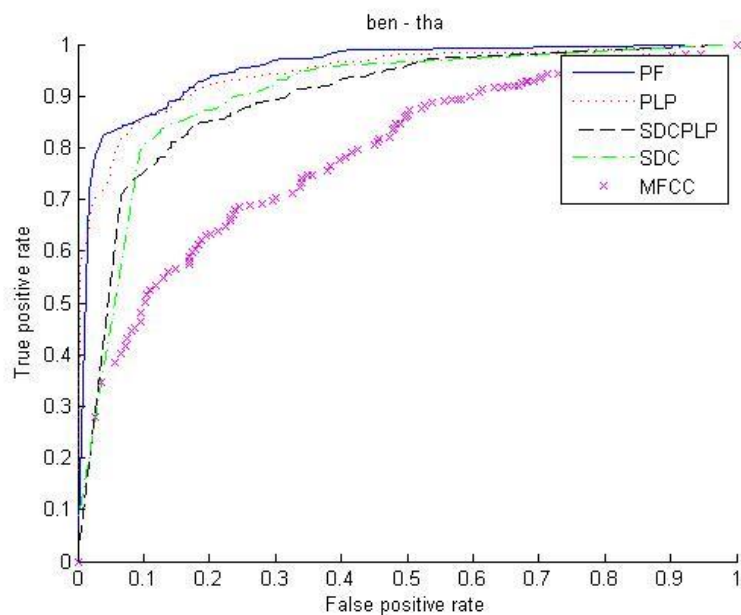
ภาพที่ ๗-7 เส้นโค้ง ROC ของผลการจำแนกภาษาของแบบจำลอง LDCRF เมื่อใช้ค่าลักษณะสำคัญชนิดต่าง ๆ สำหรับคู่ภาษาอารบิก และภาษาจีนกวางตุ้ง



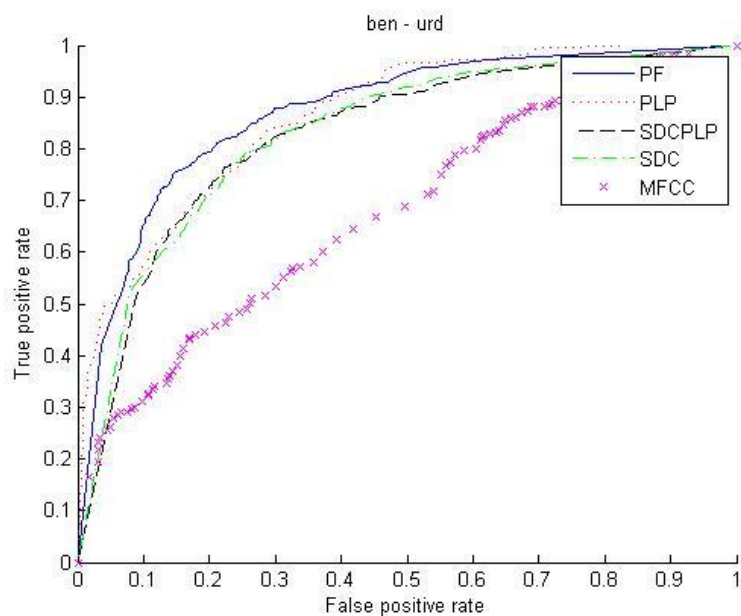
ภาพที่ ๗-8 เส้นโค้ง ROC ของผลการจำแนกภาษาของแบบจำลอง LDCRF เมื่อใช้ค่าลักษณะสำคัญชนิดต่าง ๆ สำหรับคู่ภาษาเบงกาลี และภาษาจีนมิน



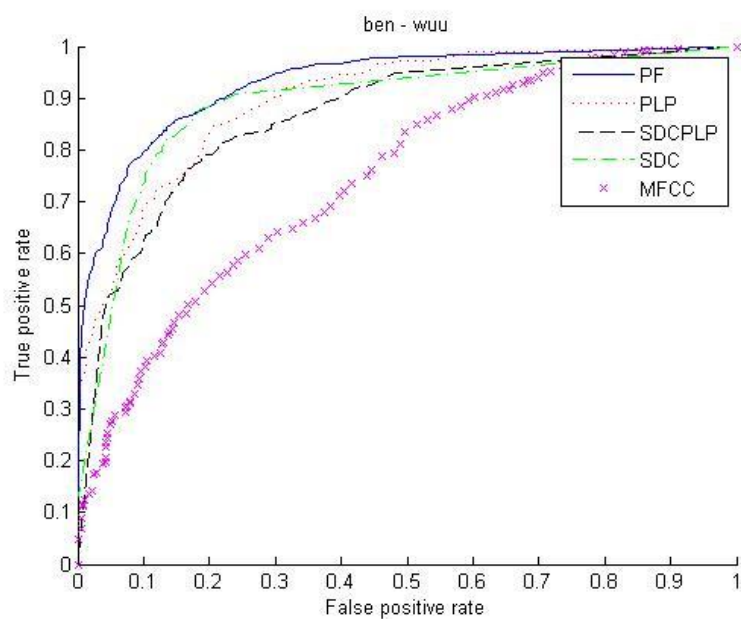
ภาพที่ ๘-9 เส้นโค้ง ROC ของผลการจำแนกภาษาของแบบจำลอง LDCRF เมื่อใช้ค่าลักษณะสำคัญชนิดต่าง ๆ สำหรับคู่ภาษาเบงกาลี และภาษารัสเซีย



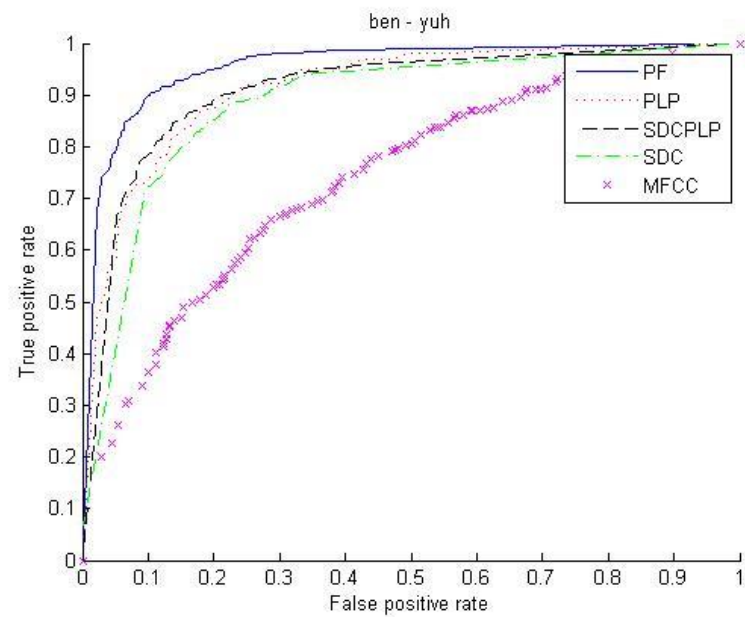
ภาพที่ ๘-10 เส้นโค้ง ROC ของผลการจำแนกภาษาของแบบจำลอง LDCRF เมื่อใช้ค่าลักษณะสำคัญชนิดต่าง ๆ สำหรับคู่ภาษาเบงกาลี และภาษาไทย



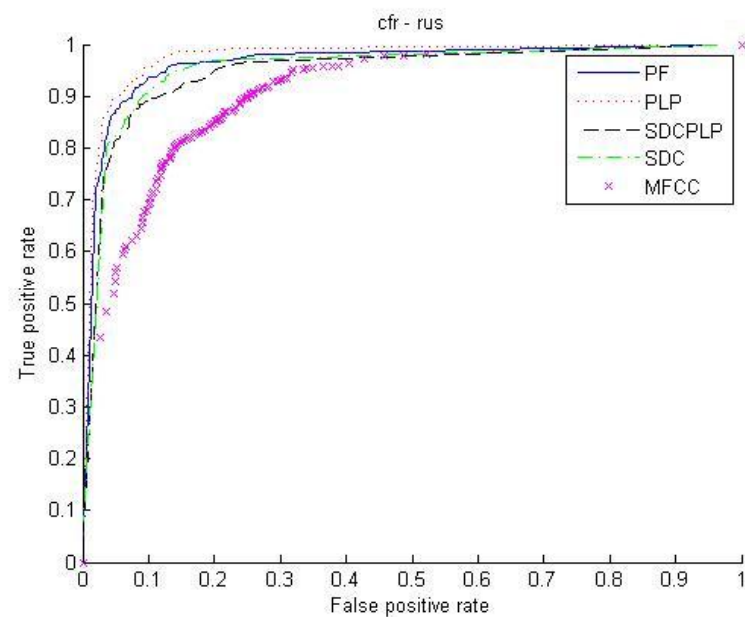
ภาพที่ ๑๑-11 เส้นโค้ง ROC ของผลการจำแนกภาษาของแบบจำลอง LDCRF เมื่อใช้ค่าลักษณะสำคัญชนิดต่าง ๆ สำหรับคู่ภาษาเบงกาลี และภาษาเอร์ดู



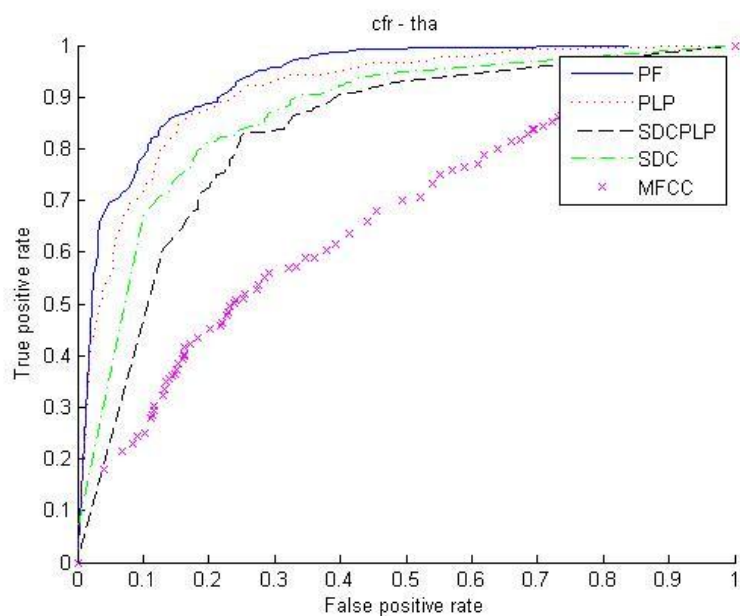
ภาพที่ ๑๑-12 เส้นโค้ง ROC ของผลการจำแนกภาษาของแบบจำลอง LDCRF เมื่อใช้ค่าลักษณะสำคัญชนิดต่าง ๆ สำหรับคู่ภาษาเบงกาลี และภาษาจื๋นงู



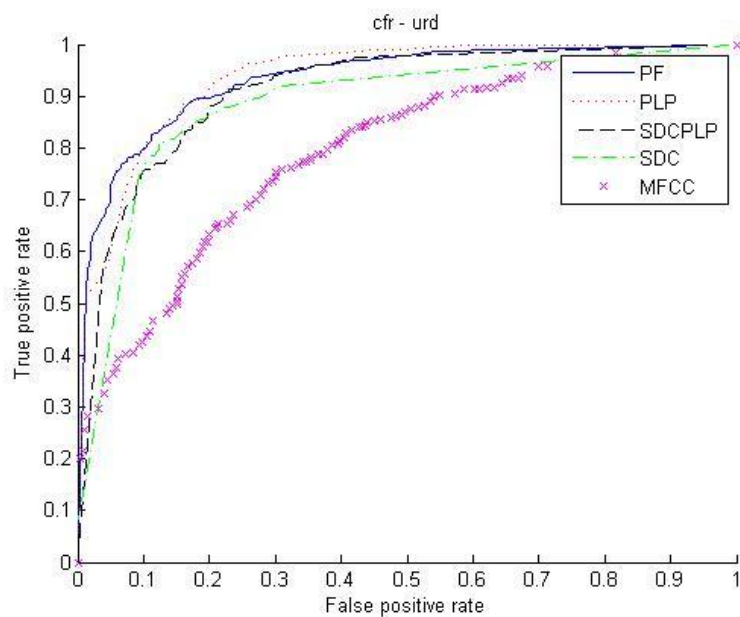
ภาพที่ ฅ-13 เส้นโค้ง ROC ของผลการจำแนกภาษาของแบบจำลอง LDCRF เมื่อใช้ค่าลักษณะสำคัญชนิดต่าง ๆ สำหรับคู่ภาษาเบงกาลี และภาษาจันทวาทั้ง



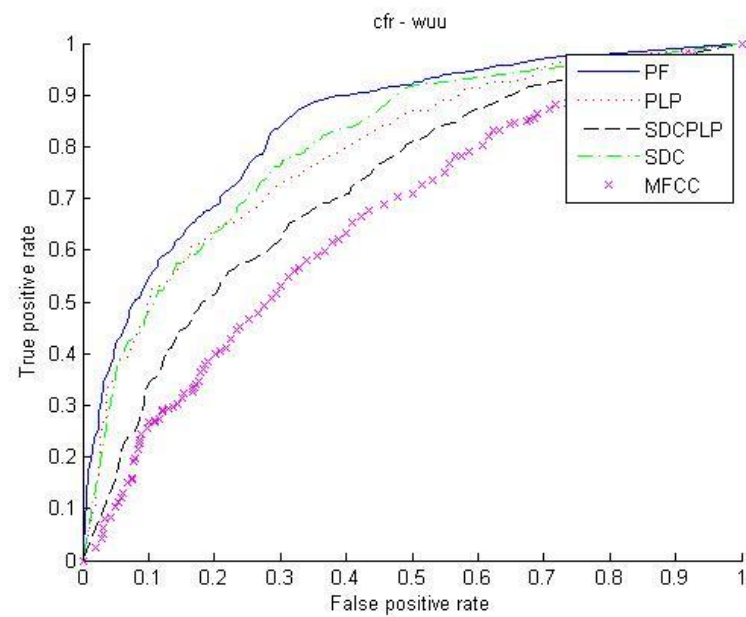
ภาพที่ ฅ-14 เส้นโค้ง ROC ของผลการจำแนกภาษาของแบบจำลอง LDCRF เมื่อใช้ค่าลักษณะสำคัญชนิดต่าง ๆ สำหรับคู่ภาษาจันทมิน และภาษาไทย



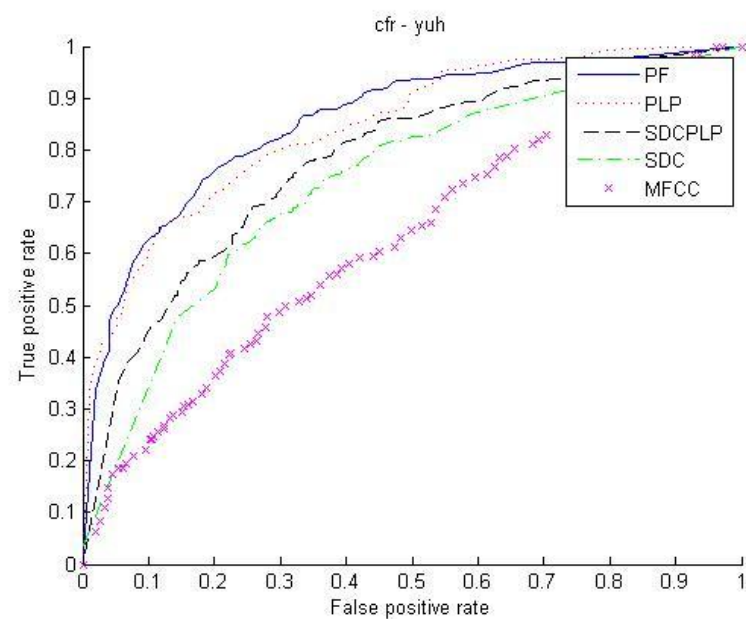
ภาพที่ ๑๕-15 เส้นโค้ง ROC ของผลการจำแนกภาษาของแบบจำลอง LDCRF เมื่อใช้ค่าลักษณะสำคัญชนิดต่าง ๆ สำหรับคู่ภาษาจีนมิน และภาษาไทย



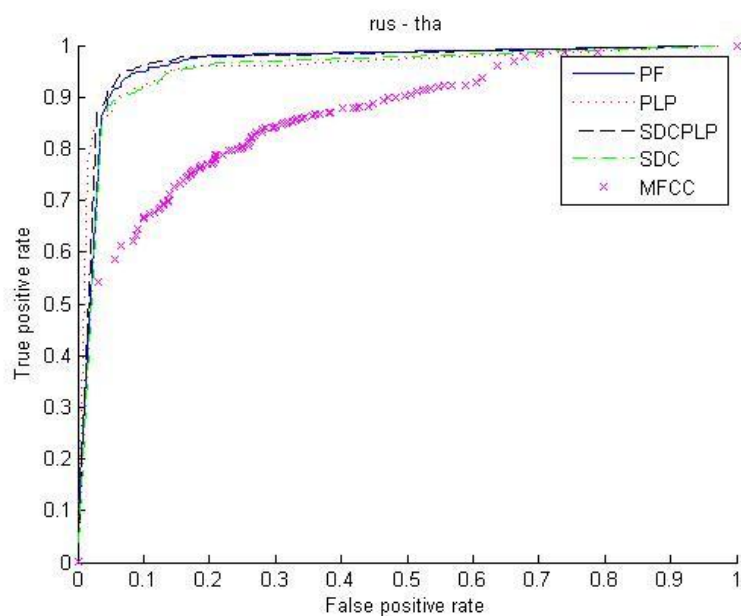
ภาพที่ ๑๕-16 เส้นโค้ง ROC ของผลการจำแนกภาษาของแบบจำลอง LDCRF เมื่อใช้ค่าลักษณะสำคัญชนิดต่าง ๆ สำหรับคู่ภาษาภาษาจีนมิน และภาษาเออร์ดู



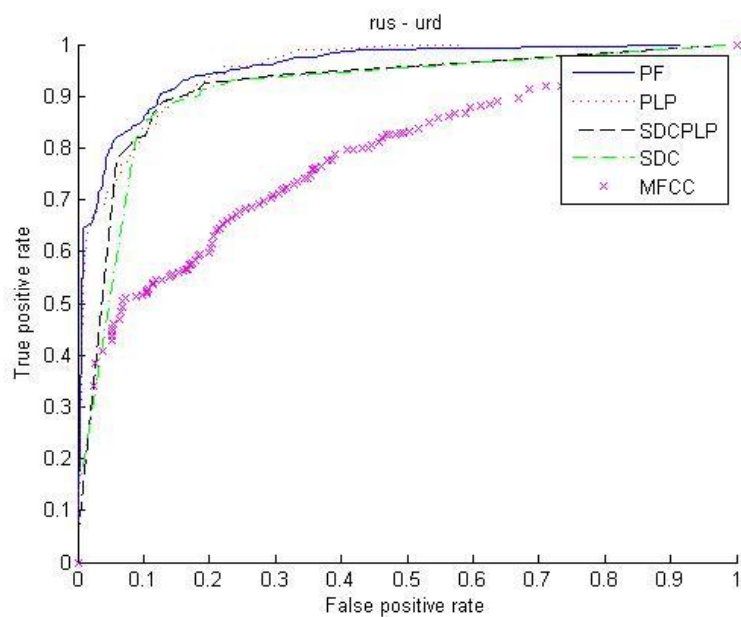
ภาพที่ ๑๗-17 เส้นโค้ง ROC ของผลการจำแนกภาษาของแบบจำลอง LDCRF เมื่อใช้ค่าลักษณะสำคัญชนิดต่าง ๆ สำหรับคู่ภาษาจีนมิน และภาษาจีนวู



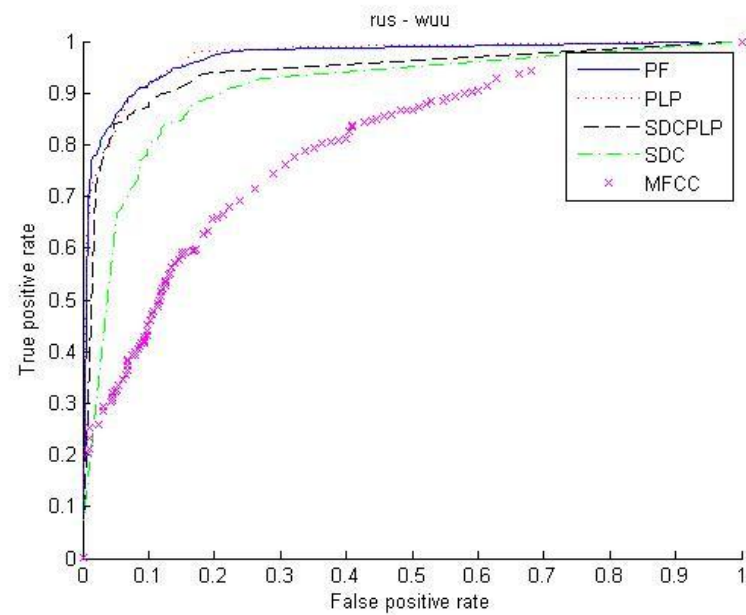
ภาพที่ ๑๗-18 เส้นโค้ง ROC ของผลการจำแนกภาษาของแบบจำลอง LDCRF เมื่อใช้ค่าลักษณะสำคัญชนิดต่าง ๆ สำหรับคู่ภาษาจีนมิน และภาษาจีนกวางตุ้ง



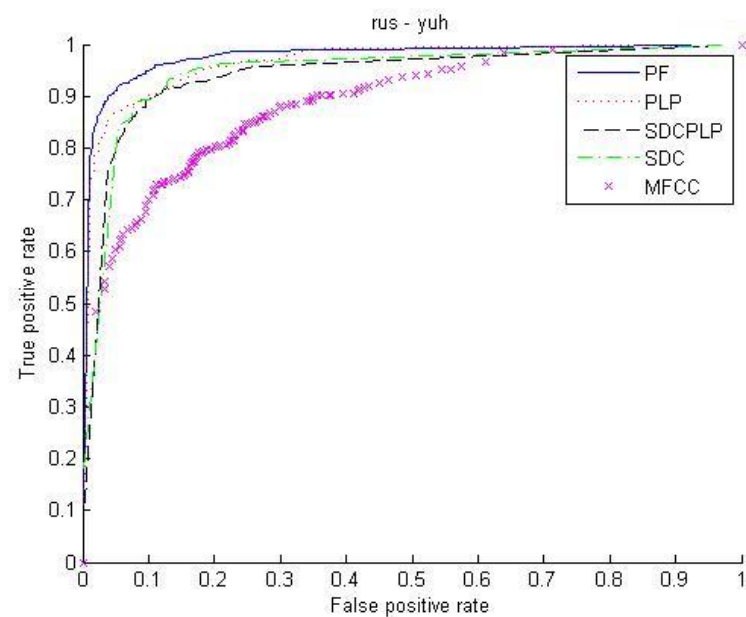
ภาพที่ ๑๙-19 เส้นโค้ง ROC ของผลการจำแนกภาษาของแบบจำลอง LDCRF เมื่อใช้ค่าลักษณะสำคัญชนิดต่าง ๆ สำหรับคู่ภาษาจีนมิน และภาษาจีนกวางตุ้ง



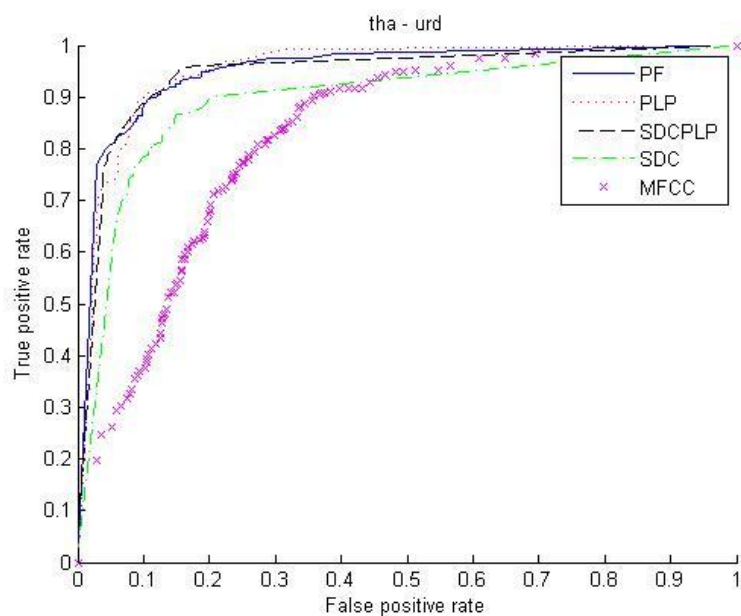
ภาพที่ ๑๙-20 เส้นโค้ง ROC ของผลการจำแนกภาษาของแบบจำลอง LDCRF เมื่อใช้ค่าลักษณะสำคัญชนิดต่าง ๆ สำหรับคู่ภาษารัสเซีย และภาษาเออร์ดุ



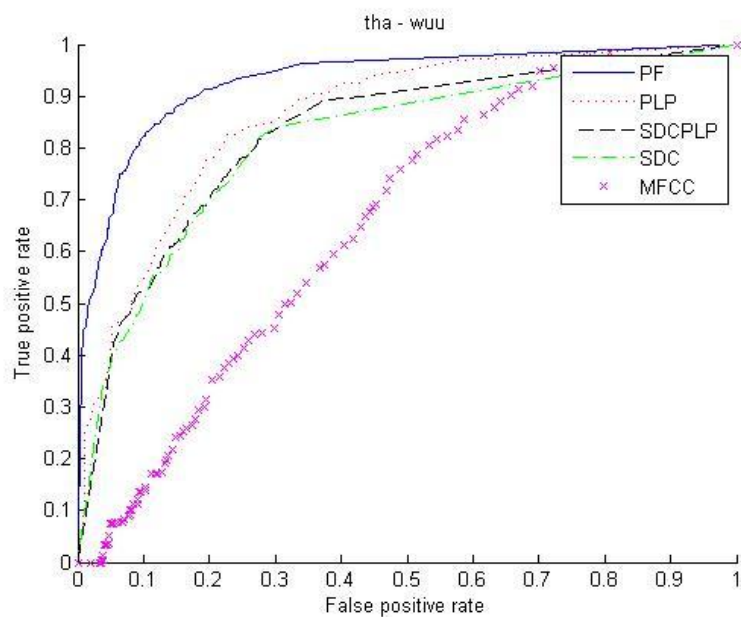
ภาพที่ ๘-21 เส้นโค้ง ROC ของผลการจำแนกภาษาของแบบจำลอง LDCRF เมื่อใช้ค่าลักษณะสำคัญชนิดต่าง ๆ สำหรับคู่ภาษารัสเซีย และภาษาจีนวู



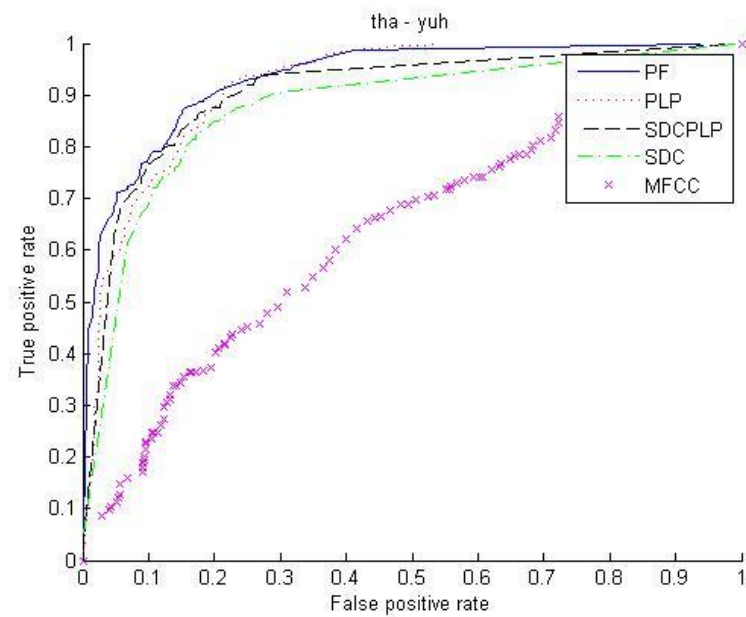
ภาพที่ ๘-22 เส้นโค้ง ROC ของผลการจำแนกภาษาของแบบจำลอง LDCRF เมื่อใช้ค่าลักษณะสำคัญชนิดต่าง ๆ สำหรับคู่ภาษารัสเซีย และภาษาจีนกวางตุ้ง



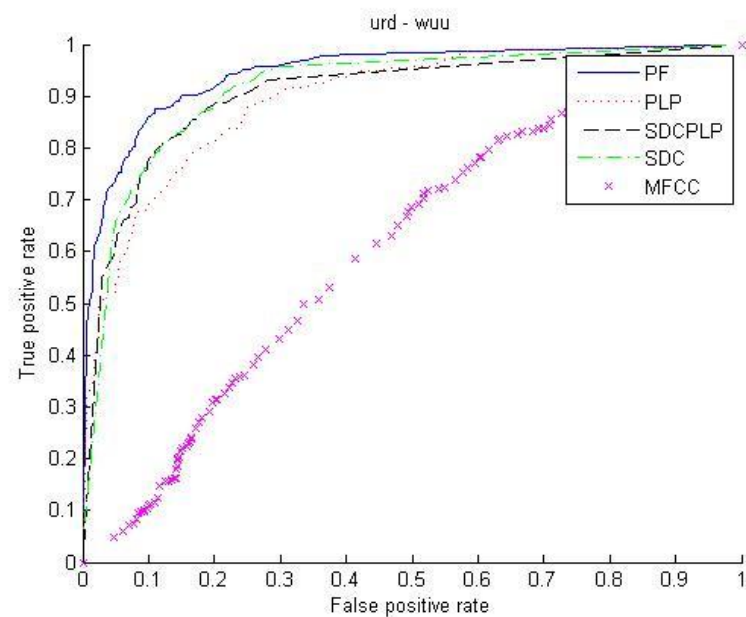
ภาพที่ ๒๓ เส้นโค้ง ROC ของผลการจำแนกภาษาของแบบจำลอง LDCRF เมื่อใช้ค่าลักษณะสำคัญชนิดต่าง ๆ สำหรับคู่ภาษาไทย และภาษาเออร์ดุ



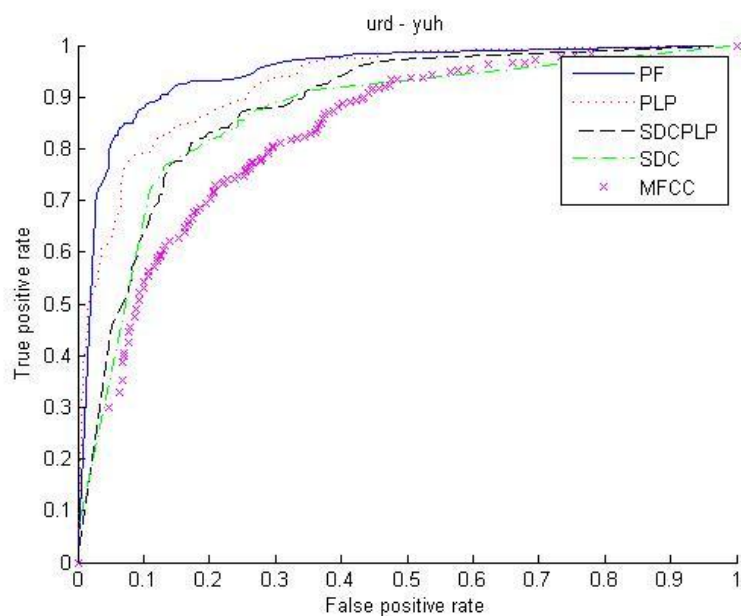
ภาพที่ ๒๔ เส้นโค้ง ROC ของผลการจำแนกภาษาของแบบจำลอง LDCRF เมื่อใช้ค่าลักษณะสำคัญชนิดต่าง ๆ สำหรับคู่ภาษาไทย และภาษาจิ้นวู



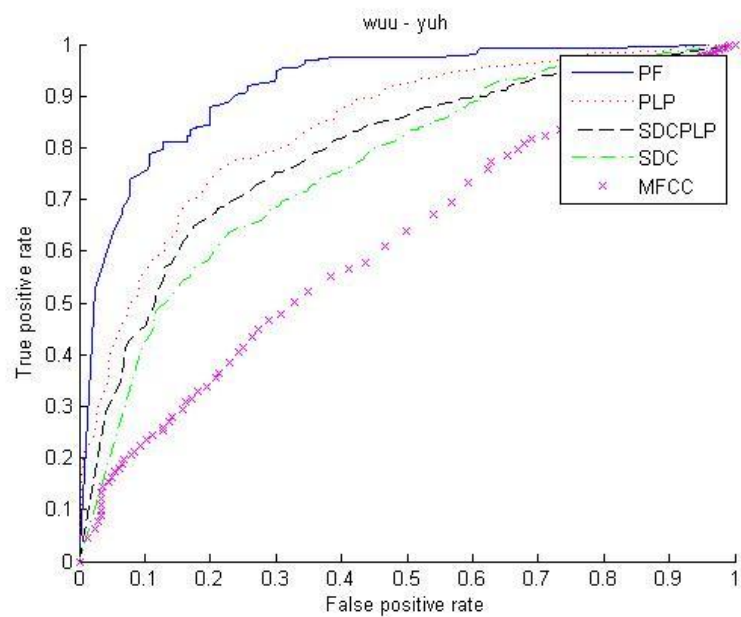
ภาพที่ ๘-25 เส้นโค้ง ROC ของผลการจำแนกภาษาของแบบจำลอง LDCRF เมื่อใช้ค่าลักษณะสำคัญชนิดต่าง ๆ สำหรับคู่ภาษาไทย และภาษาจีนกวางตุ้ง



ภาพที่ ๘-26 เส้นโค้ง ROC ของผลการจำแนกภาษาของแบบจำลอง LDCRF เมื่อใช้ค่าลักษณะสำคัญชนิดต่าง ๆ สำหรับคู่ภาษาเอร์ดู และภาษาจีนวู

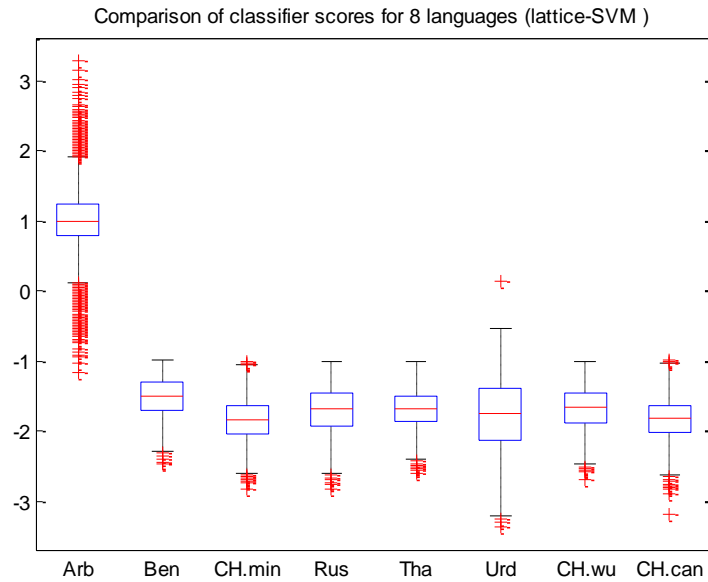


ภาพที่ ๓-27 เส้นโค้ง ROC ของผลการจำแนกภาษาของแบบจำลอง LDCRF เมื่อใช้ค่าลักษณะสำคัญชนิดต่าง ๆ สำหรับคู่ภาษาเออร์ดุ และภาษาจีนกวางตุ้ง

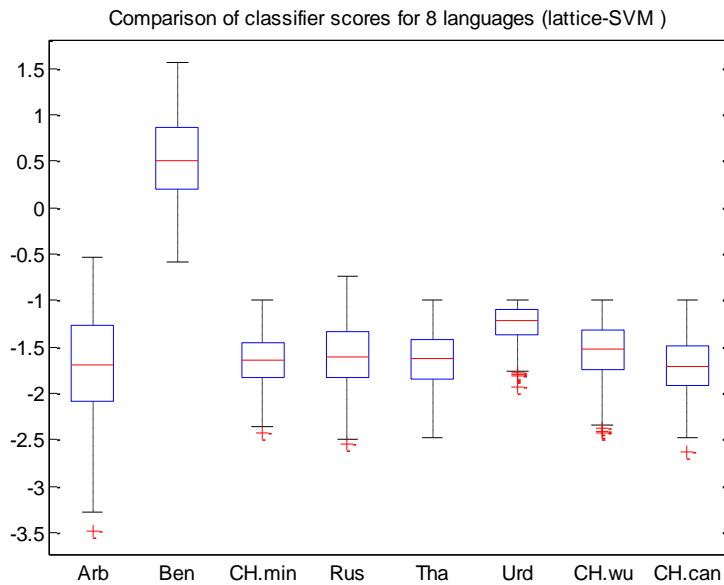


ภาพที่ ๓-28 เส้นโค้ง ROC ของผลการจำแนกภาษาของแบบจำลอง LDCRF เมื่อใช้ค่าลักษณะสำคัญชนิดต่าง ๆ สำหรับคู่ภาษาจีนนง และภาษาจีนกวางตุ้ง

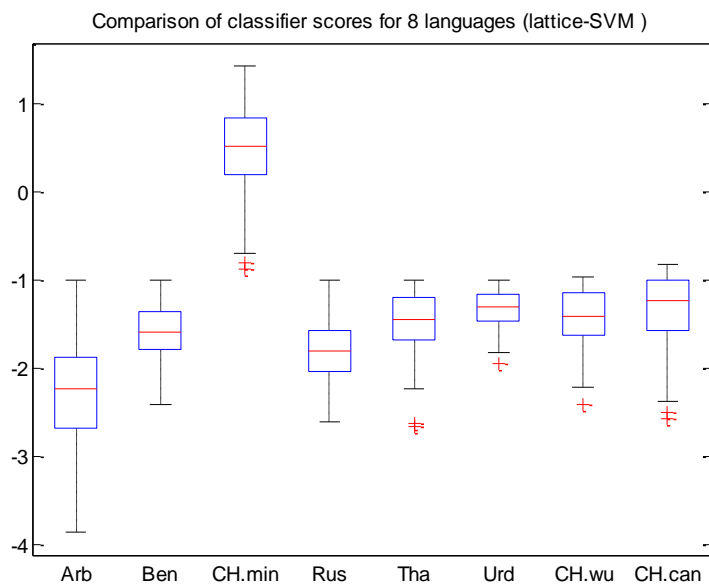
ภาคผนวก ญ-1 คะแนนการตัดสินใจของระบบการรู้จำภาษาแบบแลตทิซ-วีเอสเอ็ม โดยแสดง
คะแนนสำหรับแต่ละภาษา



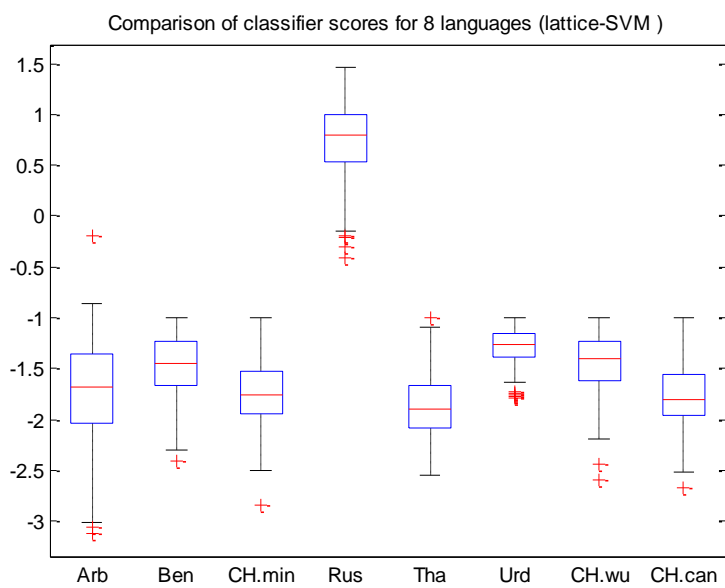
ภาพที่ ญ-1 คะแนนของระบบการรู้จำภาษาแบบแลตทิซ-วีเอสเอ็มจากข้อมูลฝึกฝน
ภาษาอารบิก



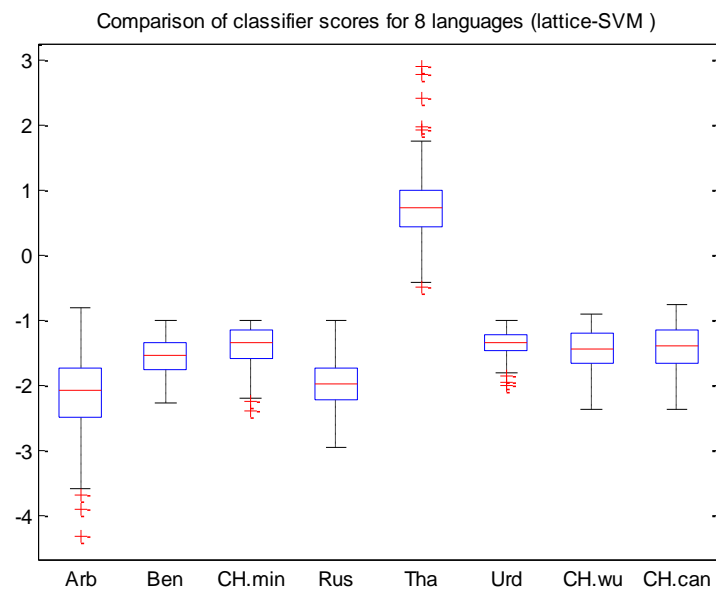
ภาพที่ ญ-2 คะแนนของระบบการรู้จำภาษาแบบแลตทิซ-วีเอสเอ็มจากข้อมูลฝึกฝน
ภาษาเบงกาลี



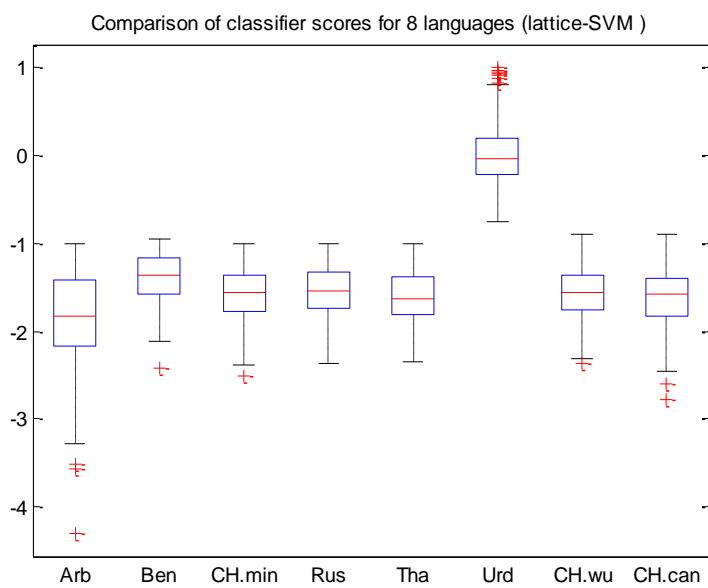
ภาพที่ ๓-๓ คะแนนของระบบการรู้จำภาษาแบบแลตทิซ-วีเอสเอ็มจากข้อมูลฝึกฝน
ภาษาจีนมิน



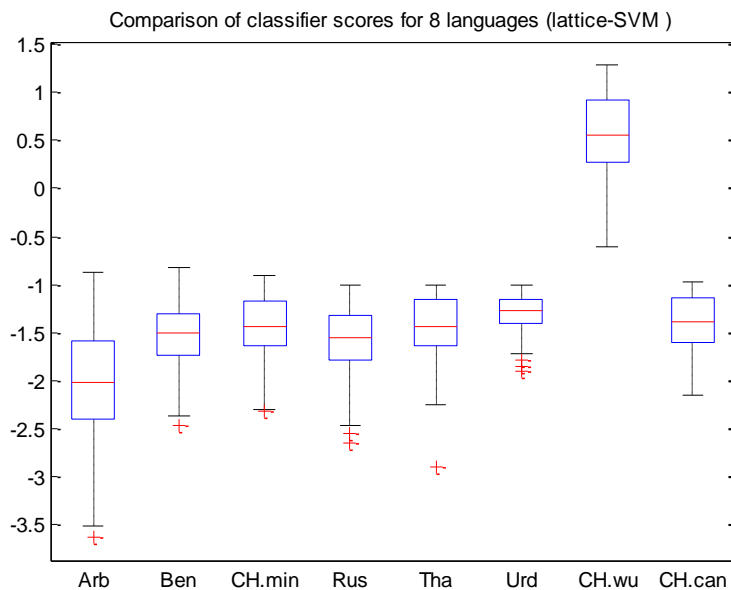
ภาพที่ ๓-๔ คะแนนของระบบการรู้จำภาษาแบบแลตทิซ-วีเอสเอ็มจากข้อมูลฝึกฝน
ภาษารัสเซีย



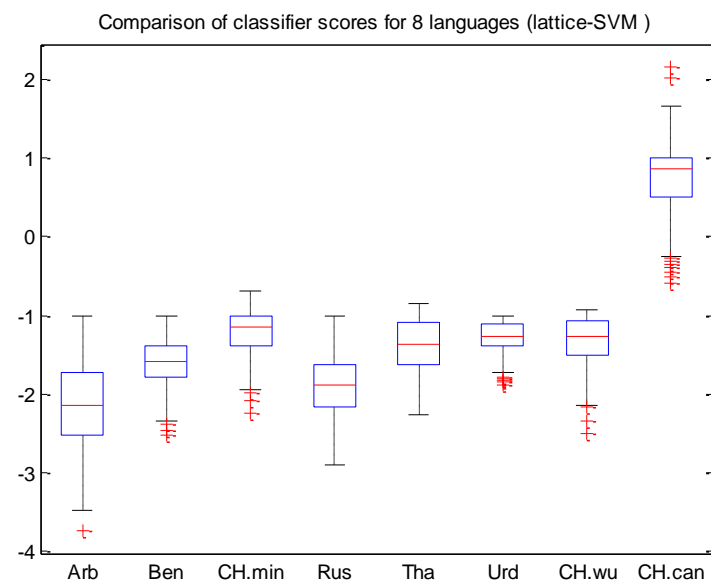
ภาพที่ ๕-5 คะแนนของระบบการรู้จำภาษาแบบแลตทิซ-วีเอสเอ็มจากข้อมูลฝึกฝนภาษาไทย



ภาพที่ ๕-6 คะแนนของระบบการรู้จำภาษาแบบแลตทิซ-วีเอสเอ็มจากข้อมูลฝึกฝนภาษาเออร์ดีดู

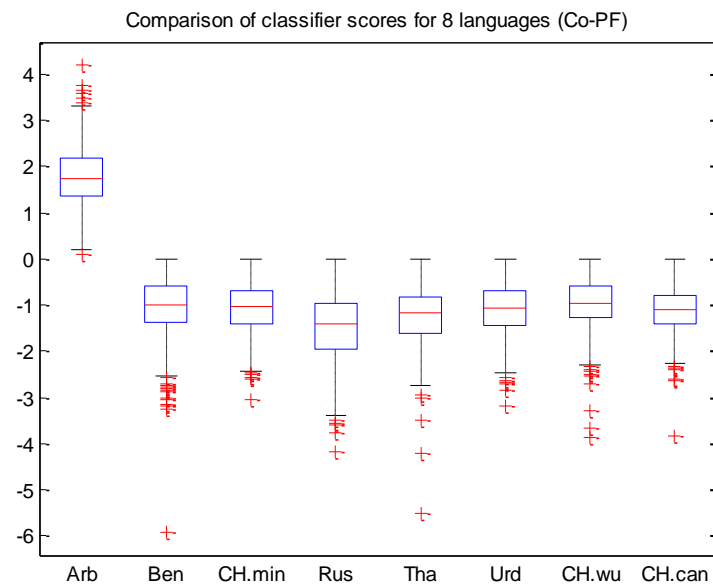


ภาพที่ ๗-7 คะแนนของระบบการรู้จำภาษาแบบแลตทิซ-วีเอสเอ็มจากข้อมูลฝึกฝน ภาษาจีนวู

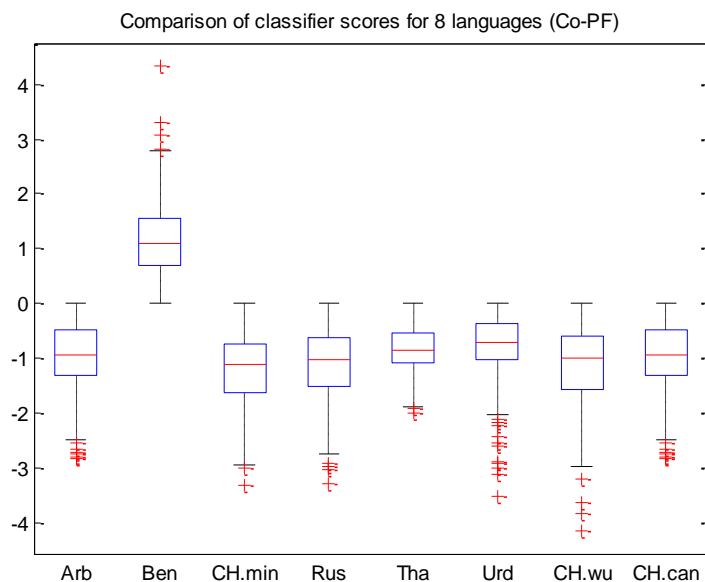


ภาพที่ ๗-8 คะแนนของระบบการรู้จำภาษาแบบแลตทิซ-วีเอสเอ็มจากข้อมูลฝึกฝน ภาษาจีนกวางตุ้ง

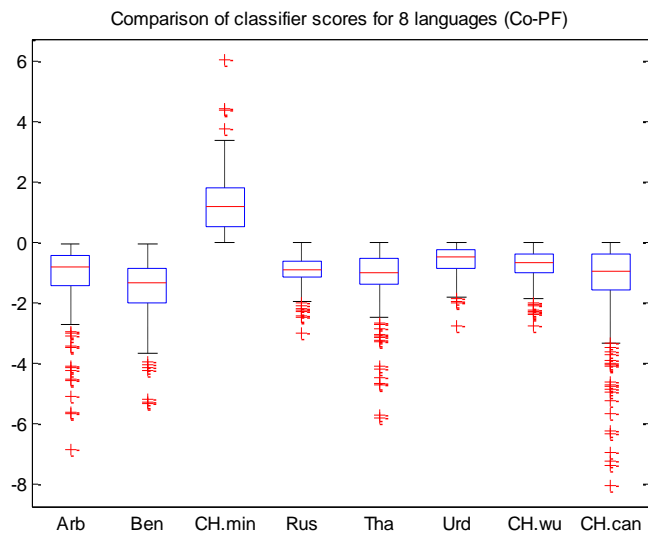
ภาคผนวก ญ-2 คะแนนการตัดสินใจโดยใช้ระบบการรู้จำภาษาพูดแนวทางการเรียงตัวของเสียง
โดยใช้ค่าพีเอฟ สำหรับแต่ละภาษา



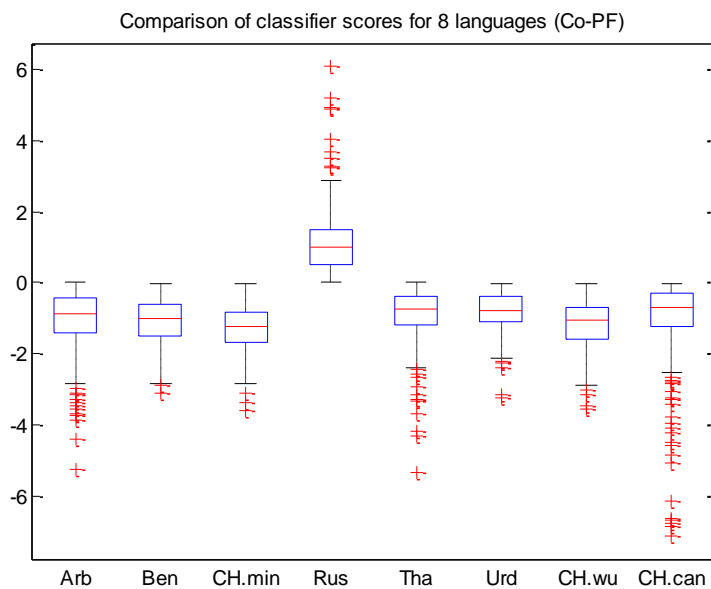
ภาพที่ ญ-9 คะแนนของระบบการรู้จำภาษาพูดแนวทางการเรียงตัวของเสียงโดยใช้ค่า
พีเอฟ จากข้อมูลฝึกฝนภาษาอารบิก



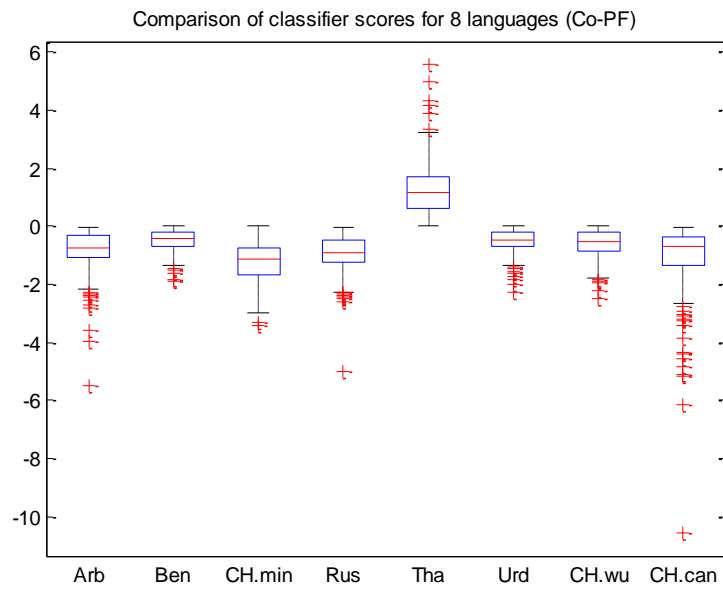
ภาพที่ ญ-10 คะแนนของระบบการรู้จำภาษาพูดแนวทางการเรียงตัวของเสียงโดยใช้ค่า
พีเอฟ จากข้อมูลฝึกฝนภาษาเบงกาลี



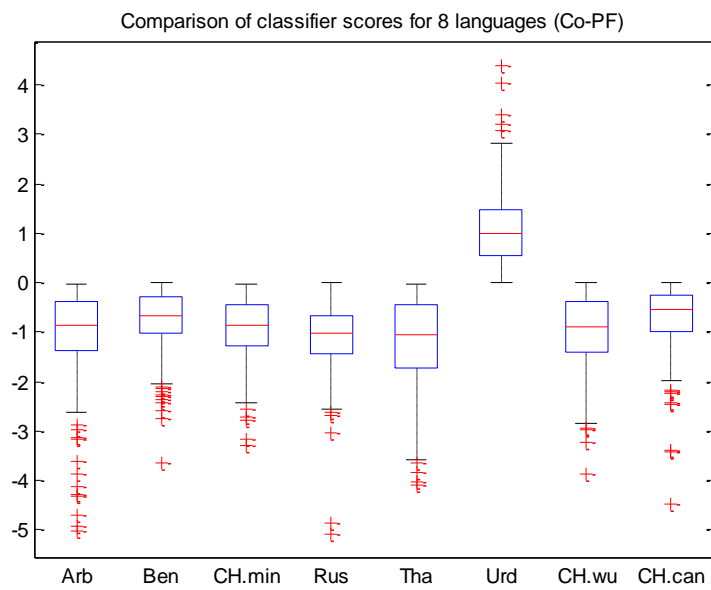
ภาพที่ ๑๑-11 คะแนนของระบบการรู้จำภาษาพูดแนวทางการเรียงตัวของเสียงโดยใช้ค่าพีเอฟ จากข้อมูลฝึกฝนภาษาจีนมิน



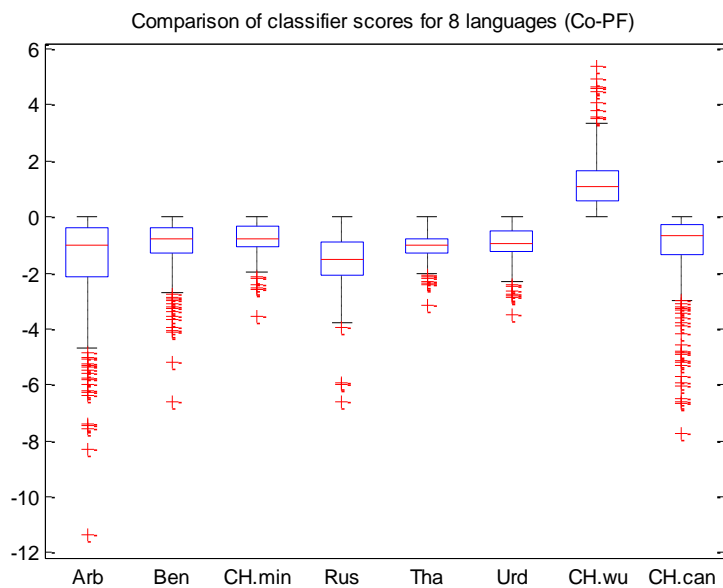
ภาพที่ ๑๑-12 คะแนนของระบบการรู้จำภาษาพูดแนวทางการเรียงตัวของเสียงโดยใช้ค่าพีเอฟ จากข้อมูลฝึกฝนภาษารัสเซีย



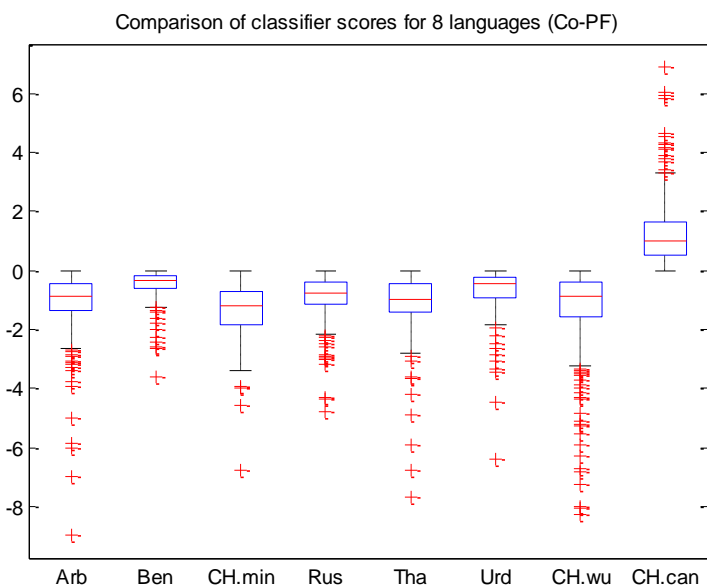
ภาพที่ ๑๓-13 คะแนนของระบบการรู้จำภาษาพูดแนวทางการเรียงตัวของเสียงโดยใช้ค่าพีเอฟ จากข้อมูลฝึกฝนภาษาไทย



ภาพที่ ๑๓-14 คะแนนของระบบการรู้จำภาษาพูดแนวทางการเรียงตัวของเสียงโดยใช้ค่าพีเอฟ จากข้อมูลฝึกฝนภาษาเออร์ดู

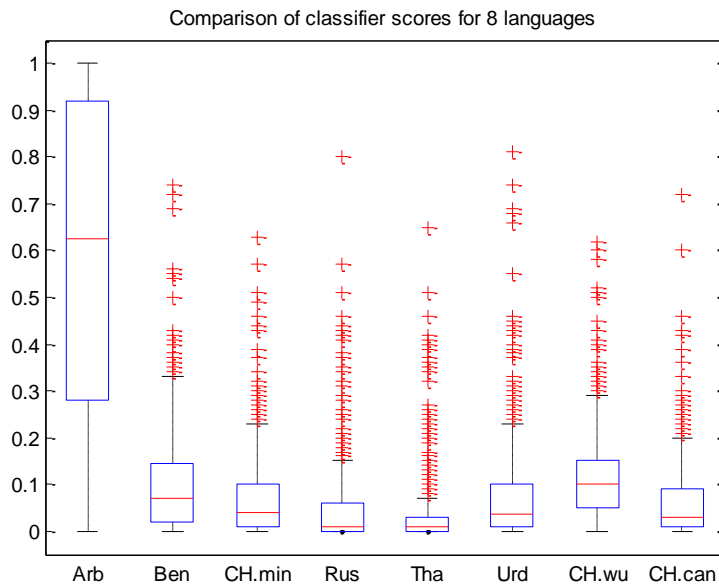


ภาพที่ ๑๕-15 คะแนนของระบบการรู้จำภาษาพูดแนวทางการเรียงตัวของเสียงโดยใช้ค่าพีเอฟ จากข้อมูลฝึกฝนภาษาจีนวู

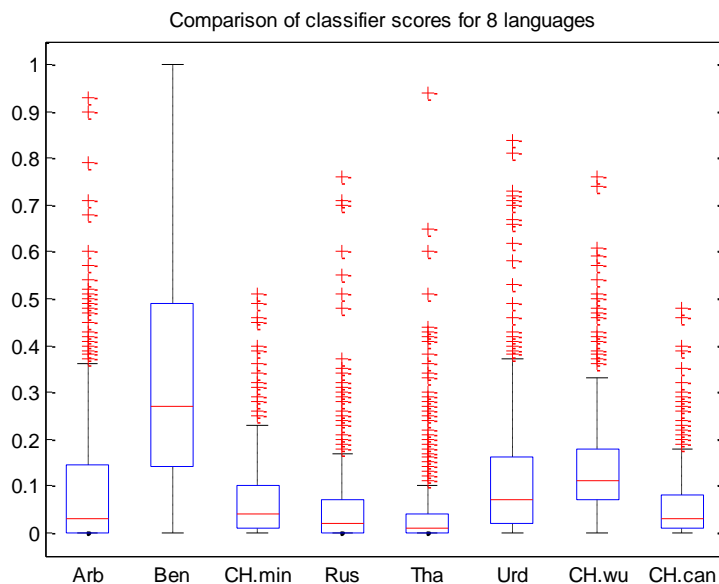


ภาพที่ ๑๕-16 คะแนนของระบบการรู้จำภาษาพูดแนวทางการเรียงตัวของเสียงโดยใช้ค่าพีเอฟ จากข้อมูลฝึกฝนภาษาจีนกวางตุ้ง

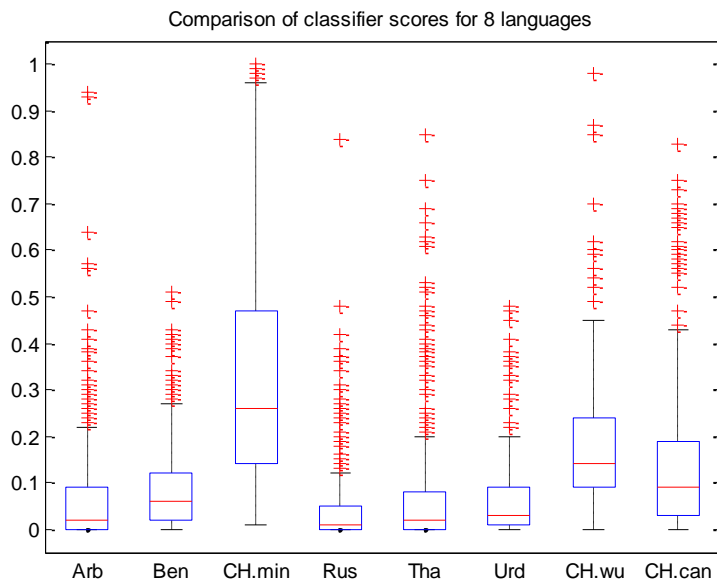
ภาคผนวก ญ-3 คะแนนการตัดสินใจโดยใช้การจำแนกด้วยแบบจำลองลาเทนท์ไดนามิกคอนดิชันนอลแรนดอมฟิลด์แบบหนึ่งต่อทั้งหมดร่วมกับค่าพีเอฟ ซึ่งแสดงคะแนนสำหรับแต่ละภาษา



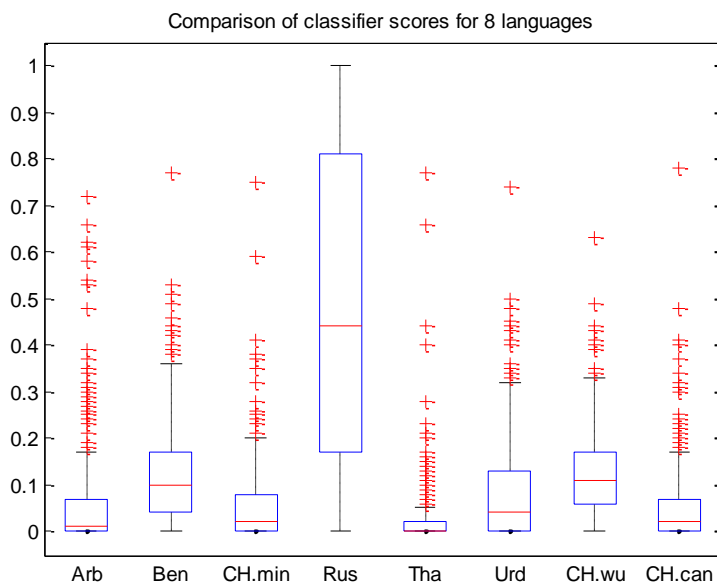
ภาพที่ ญ-17 คะแนนของระบบการรู้จำภาษาที่ใช้จำแนกด้วยแบบจำลองลาเทนท์ไดนามิกคอนดิชันนอลแรนดอมฟิลด์ร่วมกับค่าพีเอฟ จากข้อมูลฝึกฝนภาษาอารบิก



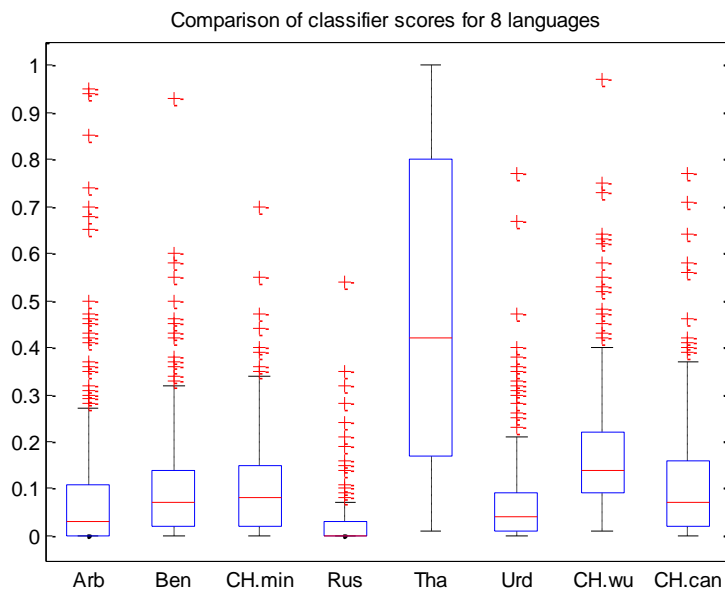
ภาพที่ ญ-18 คะแนนของระบบการรู้จำภาษาที่ใช้จำแนกด้วยแบบจำลองลาเทนท์ไดนามิกคอนดิชันนอลแรนดอมฟิลด์ร่วมกับค่าพีเอฟ จากข้อมูลฝึกฝนภาษาเบงกาลี



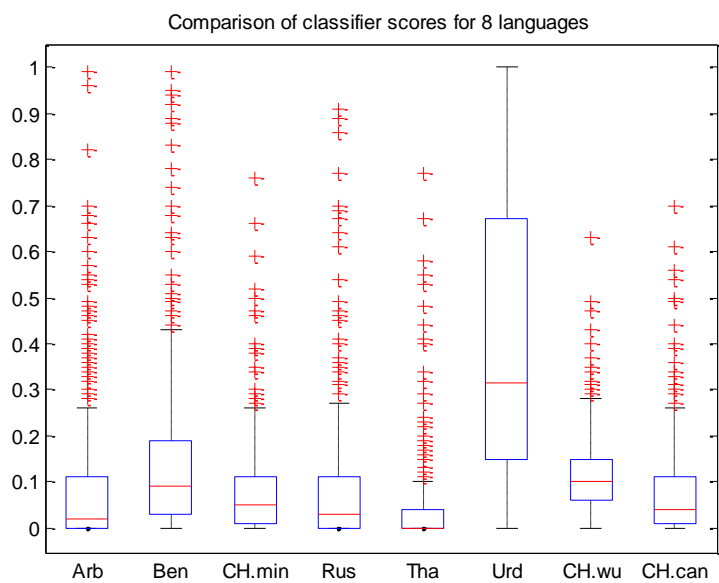
ภาพที่ ๑๙-19 คะแนนของระบบการรู้จำภาษาที่ใช้จำแนกด้วยแบบจำลองลาเทนท์ไดนามิกคอนดิชันนอลแรนดอมฟิลด์ร่วมกับค่าพีเอช จากข้อมูลฝึกฝนภาษาจีนมิน



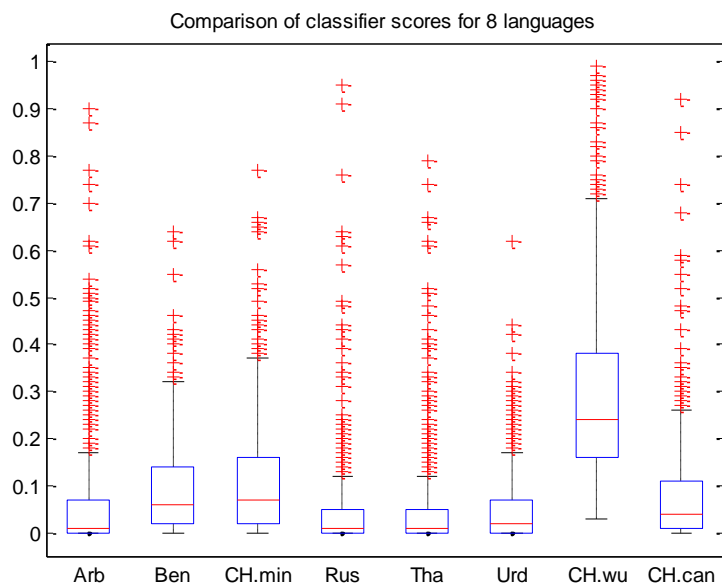
ภาพที่ ๑๙-20 คะแนนของระบบการรู้จำภาษาที่ใช้จำแนกด้วยแบบจำลองลาเทนท์ไดนามิกคอนดิชันนอลแรนดอมฟิลด์ร่วมกับค่าพีเอช จากข้อมูลฝึกฝนภาษารัสเซีย



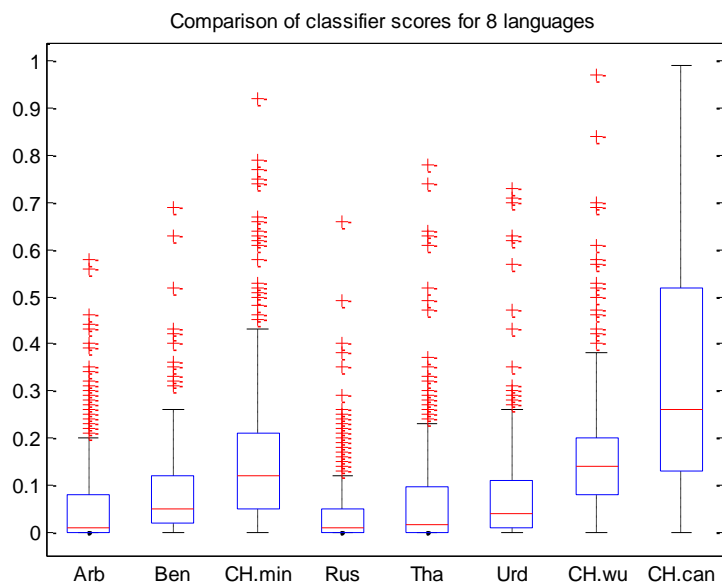
ภาพที่ ๒๑-21 คะแนนของระบบการรู้จำภาษาที่ใช้จำแนกด้วยแบบจำลองลาเทนท์ไดนามิกคอนดิชันนอลแรนดอมฟิลด์ร่วมกับค่าพีเอฟ จากข้อมูลฝึกฝนภาษาไทย



ภาพที่ ๒๑-22 คะแนนของระบบการรู้จำภาษาที่ใช้จำแนกด้วยแบบจำลองลาเทนท์ไดนามิกคอนดิชันนอลแรนดอมฟิลด์ร่วมกับค่าพีเอฟ จากข้อมูลฝึกฝนภาษาเออร์ดู



ภาพที่ ๓-23 คะแนนของระบบการรู้จำภาษาที่ใช้จำแนกด้วยแบบจำลองลาเทนท์ไดนามิกคอนดิชันนอลแรนดอมฟิลด์ร่วมกับค่าพีเอฟ จากข้อมูลฝึกฝนภาษาจีนวู



ภาพที่ ๓-24 คะแนนของระบบการรู้จำภาษาที่ใช้จำแนกด้วยแบบจำลองลาเทนท์ไดนามิกคอนดิชันนอลแรนดอมฟิลด์ร่วมกับค่าพีเอฟ จากข้อมูลฝึกฝนภาษาจีนกวางตุ้ง

รายการอ้างอิง

1. Ma, B., et al. *Multilingual speech recognition with language identification*. in *ICSLP-2002*. 2002.
2. Fuliang Weng, H.B., Leonardo Neumeyer, and Andreas Stolcke. *A Study of Multilingual Speech Recognition*. in *European Conf. on Speech Communication and Technology*. 1997.
3. Zue, V.W. and J.R. Glass. *Conversational interfaces: Advances and challenges*. in *Proceedings of the IEEE*. 2000.
4. *Language Line Services*. Available from: <http://www.language.com>.
5. Waibel, A., et al. *Multilinguality in speech and spoken language systems*. in *Proceedings of the IEEE*. 2000.
6. Gordon, R.G., B.F. Grimes, and S.I.o. Linguistics, *Ethnologue: Languages of the world*. 15 ed. 2005: SIL International.
7. Masica, C., *The Indo-Aryan Languages*. 1991: Cambridge University Press.
8. Eifring, H. and R. Theil, *Linguistics for Students of Asian and African Languages*. 2005.
9. Fennell, B.A., *A history of English: A sociolinguistic approach*. Blackwell Textbooks in linguistics. Vol. 17. 2001: Wiley-Blackwell.
10. *Language families* http://en.wikipedia.org/wiki/List_of_language_families.
11. Muthusamy, Y.K., *A segmental approach to automatic language identification*. 1993, Oregon Graduate Institute of Science & Technology.
12. Crystal, D., *A dictionary of linguistics and phonetics*. 2003: Wiley-Blackwell.
13. Benesty, J., *Springer handbook of speech processing*. 2008: Springer Verlag.
14. Byrne, W., et al. *Pronunciation modelling using a hand-labelled corpus for conversational speech recognition*. in *ICASSP*. 1998. IEEE.
15. Fosler-Lussier, E. and N. Morgan. *Effects of speaking rate and word frequency on conversational pronunciations*. in *MPV*. 1998.
16. King, S. and P. Taylor, *Detection of phonological features in continuous speech using neural networks*. *Computer Speech & Language*, 2000. **14**(4): p. 333-353.
17. Chomsky, N. and M. Halle, *The sound pattern of English*. 1968.
18. Harris, J., *English Sound Structure*. 1994: Blackwell, Oxford.
19. Scharenborg, O., V. Wan, and R.K. Moore, *Towards capturing fine phonetic variation in speech using articulatory features*. *Speech Communication*, 2007. **49**(10): p. 811-826.

20. Frankel, J., et al. *Articulatory feature classifiers trained on 2000 h of telephone speech*. in *Interspeech*. 2007.
21. Morency, L.P., A. Quattoni, and T. Darrell. *Latent-dynamic discriminative models for continuous gesture recognition*. in *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*. 2007. IEEE.
22. Lafferty, J., A. McCallum, and F.C.N. Pereira, *Conditional random fields: Probabilistic models for segmenting and labeling sequence data*. 2001.
23. Kumar, S. and M. Hebert. *Discriminative random fields: A discriminative framework for contextual interaction in classification*. in *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*. 2003. IEEE.
24. Muthusamy, Y.K., N. Jain, and R.A. Cole. *Perceptual benchmarks for automatic language identification*. in *ICASSP*. 1994. IEEE.
25. Maddieson, I. and I. Vasilescu. *Factors in human language identification*. in *International Conference on Spoken Language Processing*. 2002.
26. Adda-Decker, M., et al. *Phonetic knowledge, phonotactics and perceptual validation for automatic language identification*. in *ICPhS*. 2003.
27. Muthusamy, Y.K., E. Barnard, and R.A. Cole, *Reviewing automatic language identification*. *Signal Processing Magazine, IEEE*, 1994. **11**(4): p. 33-41.
28. Zissman, M.A., *Comparison of four approaches to automatic language identification of telephone speech*. *IEEE Transactions on Speech and Audio Processing*, 1996. **4**(1): p. 31-44.
29. Singer, E., et al. *Acoustic, phonetic and discriminative approaches to automatic language recognition*. in *EUROSPEECH-2003*. 2003.
30. Torres-Carrasquillo, P.A., et al. *Approaches to language identification using Gaussian mixture models and shifted delta cepstral features*. in *Proc. ICSLP*. 2002.
31. Wong, E., et al. *Language identification using efficient Gaussian mixture model analysis*. in *Australian International Conference on Speech Science and Technology*. 2000.
32. Leonard, R.G., *Automatic language identification*, in *RADC/Texas Instruments, Inc., Dallas. Tech Rep.* . 1974.
33. Zissman, M.A. *Automatic language identification using Gaussian Mixture and Hidden Markov models*. in *ICASSP*. 1993.
34. Torres-Carrasquillo, P.A., D.A. Reynolds, and J.R. Deller. *Language identification using Gaussian mixture model tokenization*. in *ICASSP*. 2002. IEEE.

35. Wong, K.E., *Automatic spoken language identification utilizing acoustic and phonetic speech information*. 2004: PhD thesis, Queensland University of Technology.
36. Kumar, P., et al., *Spoken Language Identification Using Hybrid Feature Extraction Methods*. arXiv preprint arXiv:1003.5623, 2010.
37. Matejka, P., et al., *BUT language recognition system for NIST 2007 evaluations*. Interspeech, 2008: p. 739-742.
38. Campbell, W., et al. *A comparison of subspace feature-domain methods for language recognition*. in *Proc. interspeech*. 2008.
39. W. M. Campbell, D.E.S., D. A. Reynolds, and A. Solomonoff. *SVM based speaker verification using a GMM supervector kernel and NAP variability compensation*. in *ICASSP*. 2006.
40. Dumouchel, P.K.a.P. *Experiments in speaker verification using factor analysis likelihood ratios*. in *Odyssey04*. 2004.
41. Alex Solomonoff, Carl Quillen, and a.W.M. Campbell. *Channel compensation for SVM speaker recognition*. in *The Speaker and Language Recognition Workshop*. 2004.
42. Campbell, W.M., et al. *Language recognition with support vector machines*. in *ODYSSEY04-The Speaker and Language Recognition Workshop*. 2004.
43. Campbell, W.M., et al., *Support vector machines for speaker and language recognition*. *Computer Speech & Language*, 2006. **20**(2): p. 210-229.
44. Burget, L., P. Matejka, and J. Cernocky. *Discriminative training techniques for acoustic language identification*. in *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*. 2006. IEEE.
45. Torres-Carrasquillo, P.A., et al. *Eigen-channel compensation and discriminatively trained Gaussian mixture models for dialect and accent recognition*. in *Proc. Interspeech*. 2008.
46. Campbell, W.M., D.E. Sturim, and D.A. Reynolds, *Support vector machines using GMM supervectors for speaker verification*. *Signal Processing Letters, IEEE*, 2006. **13**(5): p. 308-311.
47. Castaldo, F., et al. *Acoustic language identification using fast discriminative training*. in *Proc. Interspeech*. 2007.
48. Campbell, W. *A covariance kernel for SVM language recognition*. in *ICASSP*. 2008. IEEE.

49. You, C.H., H. Li, and K.A. Lee. *A GMM-supervector approach to language recognition with adaptive relevance factor*. in *18th Europ. Signal Process. Conf.* 2010.
50. Kenny, P., *Joint Factor Analysis versus Eigenchannel in Speaker Recognition*. IEEE Transaction on Audio Speech and Language Processing, 2007. **15**: p. 1435-1447.
51. Dehak, N., et al., *Front-End Factor Analysis for Speaker Verification*. IEEE Transaction on Audio Speech and Language Processing, 2011. **19**: p. 788-798.
52. Martínez González David, B.L., Ferrer Luciana, Scheffer Nicolas. *Ivector-Based Prosodic System For Language Identification*. in *International Conference on Acoustics Speec (ICASSP)*. 2012. Kyoto, Japan.
53. House, A.S. and E.P. Neuburg, *Toward automatic identification of the language of an utterance. I. Preliminary methodological considerations*. The Journal of the Acoustical Society of America, 1977. **62**: p. 708.
54. Hazen, T., *Automatic language identification using a segment-based approach*. 1993, Massachusetts Institute of Technology.
55. Zissman, M.A., *Comparison of four approaches to automatic language identification of telephone speech*. Speech and Audio Processing, IEEE Transactions on, 1996. **4**(1): p. 31.
56. Lee, C.H., F.K. Soong, and B.H. Juang. *A segment model based approach to speech recognition*. in *ICASSP*. 1988. IEEE.
57. Ma, B., H. Li, and C.H. Lee. *An acoustic segment modeling approach to automatic language identification*. in *INTERSPEECH*. 2005.
58. Bin, M. and L. Haizhou. *SPOKEN LANGUAGE IDENTIFICATION USING BAG-OF-SOUNDS*. in *ICCC*. 2005.
59. Gao, S., et al. *A text-categorization approach to spoken language identification*. in *INTERSPEECH*. 2005.
60. Bellegarda, J.R. *Exploiting latent semantic information in statistical language modeling*. in *Proceedings of the IEEE*. 2000.
61. Boon Pang, L., L. Haizhou, and M. Bin. *Using Local and Global Phonotactic Features in Chinese Dialect Identification*. in *ICASSP*. 2005.
62. Li, H., B. Ma, and C.H. Lee, *A vector space modeling approach to spoken language identification*. Audio, Speech, and Language Processing, IEEE Transactions on, 2007. **15**(1): p. 271-284.
63. Li, H. and B. Ma. *A phonotactic language model for spoken language identification*. in *ACL*. 2005. Association for Computational Linguistics.

64. Richardson, F. and W. Campbell. *Language recognition with discriminative keyword selection*. 2008. IEEE.
65. Campbell, W.M., Richardson, F. *Discriminative keyword selection using support vector machines*. in *Neural Information Processing Systems*. 2008. Vancouver, BC Canada.
66. Jayram, A.K.V.S., V. Ramasubramanian, and T. Sreenivas. *Language identification using parallel sub-word recognition*. in *ICASSP*. 2003. IEEE.
67. Ramasubramanian, V., A.K.V.S. Jayram, and T. Sreenivas. *Language identification using parallel phone recognition*. in *WSLP*. 2003.
68. Nagarajan, T. and H.A. Murthy. *Language identification using parallel syllable-like unit recognition*. in *ICASSP*. 2004. IEEE.
69. Wang, S., J. Liu, and R. Liu. *Language identification using discriminative weighted language models*. in *ISCSLP*. 2004. IEEE.
70. Yang, X. and M. Siu. *N-Best Tokenization in a GMM-SVM Language Identification System*. in *ICASSP*. 2007. IEEE.
71. Gauvain, J.L., A. Messaoudi, and H. Schwenk. *Language recognition using phone lattices*. in *ICSLP*. 2004.
72. Shen, W., et al. *Experiments with lattice-based PPRLM language identification*. in *ODYSSEY*. 2006. IEEE.
73. Zhu, D. and M. Adda-Decker. *Language identification using lattice-based phonotactic and syllabotactic approaches*. in *The Speaker and Language Recognition Workshop*. 2006. IEEE.
74. Campbell, W.M., F. Richardson, and D. Reynolds. *Language recognition with word lattices and support vector machines*. in *ICASSP*. 2007. IEEE.
75. Montero-Asenjo, A., et al. *On the use of high-level information in speaker and language recognition*. in *The Spanish Network of Speech Technologies Workshop*. 2006.
76. Schwarz, P., P. Matejka, and J. Cernocky. *Hierarchical structures of neural networks for phoneme recognition*. in *ICASSP*. 2006. IEEE.
77. P. Schwarz, P.M., and J. Cernocky. *Towards lower error rates in phoneme recognition*. in *TSD 2004, Brno*, . 2004. Czech Republic.
78. Yin, B., E. Ambikairajah, and F. Chen. *Combining cepstral and prosodic features in language identification*. in *18th International Conference on Pattern Recognition*. 2006.
79. Wong, E. and S. Sridharan. *Fusion of output scores on language identification system*. in *Multilingual Speech and Language*. 2003.

80. Gonzalez-Dominguez, J., et al., *Multilevel and Session Variability Compensated Language Recognition: ATVS-UAM Systems at NIST LRE 2009*. Selected Topics in Signal Processing, IEEE Journal of, 2010. **4**(6): p. 1084-1093.
81. Tong, R., et al. *Integrating acoustic, prosodic and phonotactic features for spoken language identification*. in *ICASSP*. 2006. IEEE.
82. Lopez-Moreno, I., et al. *Anchor-model fusion for language recognition*. in *INTERSPEECH*. 2008.
83. Ng, R.W.M., et al. *Prosodic attribute model for spoken language identification*. in *ICASSP*. 2010. IEEE.
84. Stuker, S., et al. *Multilingual articulatory features*. in *ICASSP*. 2003. IEEE.
85. Kirchhoff, K., *Robust speech recognition using articulatory information*. 1999, Universität Bielefeld.
86. Li, J. and C.H. Lee. *On designing and evaluating speech event detectors*. in *Proc. Interspeech*. 2005.
87. Harwath, D. and M. Hasegawa-Johnson, *Phonetic Landmark Detection for Automatic Language Identification*. Urbana, 2010. **51**: p. 61801.
88. Chaudhari, U. and M. Picheny. *Articulatory feature detection with Support Vector Machines for integration into ASR and phone recognition*. in *ASRU*. 2009.
89. S. Kanokphara , J. Macek, and J. Carson-Berndsen. *Comparative study: HMM and SVM for automatic articulatory feature extraction*. in *IEA/AIE'06 Proceedings of the 19th international conference on Advances in Applied Artificial Intelligence: industrial, Engineering and Other Applications of Applied Intelligent Systems* 2006.
90. Haipeng Wang, C.-C.L., Tan Lee, Bin Ma, Haizhou Li, *Shifted-Delta MLP Features for Spoken Language Recognition*. IEEE Signal Process. Lett. 20, 2013: p. 15-18.
91. Kanokphara, S. and J. Carson-Berndsen. *Articulatory-Acoustic-Feature-based Automatic Language Identification*. in *Multilingual Speech and Language Processing*. 2006.
92. Núñez Cuadra, O. and J.R. Calvo de Lara, *A New Phono-Articulatory Feature Representation for Language Identification in a Discriminative Framework*. *Computación y Sistemas*, 2011. **15**(1).
93. Parandekar, S. and K. Kirchhoff. *Multi-stream language identification using data-driven dependency selection*. in *Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03). 2003 IEEE International Conference on*. 2003. IEEE.
94. Parandekar, S.P., *Feature-Based Language Identification Using Data-driven model selection*. 2003, University of Washington.

95. Sangwan, A., M. Mehrabani, and J.H.L. Hansen. *Language identification using a combined articulatory prosody framework*. in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*. 2011. IEEE.
96. Sangwan, A. and J.H.L. Hansen, *Automatic analysis of Mandarin accented English using phonological features*. *Speech Communication*, 2012. **54**(1): p. 40-54.
97. Abdel-Haleem, Y.H., *Conditional random fields for continuous speech recognition*. 2006, University of Sheffield.
98. Yu, D. and L. Deng. *Deep-structured hidden conditional random fields for phonetic recognition*. in *Interspeech*. 2010.
99. Morris, J. and E. Fosler-Lussier, *Crandem: Conditional random fields for word recognition*. *Interspeech*, Brighton, England, 2009.
100. Gunawardana, A., et al. *Hidden conditional random fields for phone classification*. in *Proc. Interspeech*. 2005.
101. Morris, J. and E. Fosler-Lussier, *Conditional random fields for integrating local discriminative classifiers*. *Audio, Speech, and Language Processing, IEEE Transactions on*, 2008. **16**(3): p. 617-628.
102. Yu, D., et al. *Language recognition using deep-structured conditional random fields*. in *ICASSP*. 2010. IEEE.
103. Quattoni, A.J., *Object recognition with latent conditional random fields*. 2005, Massachusetts Institute of Technology.
104. Deufemia, V., M. Risi, and G. Tortora. *Sketched Symbol Recognition with a Latent-Dynamic Conditional Model*. in *Pattern Recognition (ICPR), 2010 20th International Conference on*. 2010. IEEE.
105. Leeuwen, N.B.a.D.v. *On calibration of language recognition scores*. in *Speaker and Language Recognition Workshop, 2006. IEEE Odyssey 2006*. San Juan.
106. Brummer, N., *Measuring, refining and calibrating speaker and language information extracted from speech*,. 2010, University of Stellenbosch.
107. Brummer, N. *Focal multiclass tools for evaluation, calibration and fusion of, and decision-making with, multi-class statistical pattern recognition scores*. June 2007; Available from: <http://sites.google.com/site/nikobrummer/focalmulticlass>.
108. BenZeghiba, M.F., J.-L. Gauvain, and a.L. Lamel. *Language Score Calibration using Adapted Gaussian Back* in *INTERSPEECH*. 2009.
109. K., M.Y., C.R. A., and a.B.T. Oshika. *The OGI Multi-language Telephone Speech Corpus*. in *the International Conference on Spoken Language Processing*. 1992. Banff, Alberta, Canada.

110. Hieronymus, J.L., *ASCII Phonetic Symbols for the World's Languages: Worldbet*. 1993: preprint.
111. Mark Przybocki and A. Martin. *NIST Speaker Recognition Evaluation Chronicles*. in *IEEE Odyssey 2006: Speaker and Language Recognition Workshop*, . 2006. Toledo, Spain.
112. T., J., *Making large-Scale SVM Learning Practical*. *Advances in Kernel Methods - Support Vector Learning*. Support Vector Learning, B. Schölkopf and C. Burges and A. Smola (ed.), MIT-Press, 1999.
113. Young, S., et al., *The HTK book*. 2006: Cambridge.
114. Joachims, T., *Learning to Classify Text Using Support Vector Machines*. Kluwer Academic Publishers, 2002.
115. W. M. Campbell, J.P.C., D. A. Reynolds, D. A. Jones, and T. R. Leek, *Phonetic speaker recognition with support vector machines*. *Advances in Neural Information Processing* 15. 2003.
116. Allen, F., E. Ambikairajah, and J. Epps. *Language Identification Using Warping and the Shifted Delta Cepstrum*. in *IEEE Int. Workshop on Multimedia Signal Processing (Shanghai, China)*. 2005.
117. McCallum, A. *Efficiently inducing features of conditional random fields*. in *Nineteenth Conference on Uncertainty in Artificial Intelligence UAI03*. 2003.
118. Burget, L., et al. *Discriminantly trained probabilistic linear discriminant analysis for speaker verification*. in *ICASSP*. 2011.
119. Martínez G. David, O.P.B.L., Ferrer Luciana, Scheffer Nicolas. *Language Recognition in iVector Space*. in *Interspeech2011*. 2011. Florence, Italy.
120. Larcher, A., et al. *ALIZE 3.0 - Open Source Toolkit for State-of-the-Art Speaker Recognition*. in *Interspeech*. 2013.

ประวัติผู้เขียนวิทยานิพนธ์

นางสาวศิรินุช บุญสุข เกิดเมื่อวันที่ 23 สิงหาคม พ.ศ.2523 สำเร็จการศึกษาระดับปริญญาตรีหลักสูตรวิศวกรรมศาสตรบัณฑิต สาขาวิชาวิศวกรรมคอมพิวเตอร์ จากภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์มหาวิทยาลัยเกษตรศาสตร์ ในปี 2546 สำเร็จการศึกษาระดับปริญญาโทในหลักสูตรวิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมคอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ในปี 2550 และเข้าศึกษาต่อระดับปริญญาเอกในหลักสูตรวิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมคอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY