ความก้าวหน้าของการพัฒนาระบบระบุผู้พูดภาษาไทย

Thai Language Speaker Identification System: Development Progress [1]

ชัย วุฒิวิวัฒน์ชัย, สุทัศน์ แซ่ตั้ง และวารินทร์ อัจฉริยะกุลพร
คณะนักวิจัยและพัฒนาระบบระบุผู้พูดสำหรับภาษาไทย [2]
หน่วยปฏิบัติการวิจัยและพัฒนาวิศวกรรมภาษาและซอฟต์แวร์
ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ
สำนักงานพัฒนาวิทยาศาสตร์และเทคโนโลยีแห่งชาติ
539/2 อาคารมหานครยิบซั่ม ชั้น 22 ถนนศรีอยุธยา แขวงพญาไท เขตราชเทวี กรุงเทพฯ 10400

ABSTRACT -- Speaker identification for Thai language project has been initiated by the National Electronics and Computer Technology Center (NECTEC) since 1999. The first objective is to research and develop a text-dependent closed-set speaker identification system in the office environment. The speaking texts for this system are isolated digit utterances 0-9 and their concatenation. This paper gives an overview of the system, explains the route of the past 1-year research history, and some details of the latest identification system, which achieves the best performance of 92.30% for isolated digit "0" and enhances to 98% for 3-concatenated digit.

KEY WORDS -- Speaker Identification, Text Dependent, Closed Set, Thai Language

บทคัดย่อ -- โครงการระบบระบุผู้พูดสำหรับภาษาไทย (Speaker Identification for Thai Language) ของศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ ได้ริเริ่มขึ้นในปีงบประมาณ 2542 โดยเบื้องต้นมุ่งเน้นการวิจัยและพัฒนาระบบระบุผู้พูดที่ใช้กับภาษาไทยแบบกำหนดคำพูดตายตัว (Text dependent) เป็นระบบปิด (Closed set system) และใช้ในสภาพแวดล้อมสำนักงาน (Office environment) คำพูดที่ใช้ในการวิจัยเป็นเสียงตัวเลขโดด 0-9 และตัวเลขโดดต่อกัน บทความฉบับนี้เป็นการสรุปผลการวิจัยโดยนำเสนอภาพรวมของระบบระบุผู้พูด แสดงรายละเอียดของผลงานวิจัยในช่วง 1 ปีที่ผ่านมา รวมทั้งนำเสนอผลงานความก้าวหน้าล่าสุด พร้อมทั้งรายละเอียดของระบบระบุผู้พูดที่ใช้กับผู้พูดจำนวน 50 คน ซึ่งได้ผลอัตราการระบุผู้พูดสูงที่สุด 92.30% เมื่อใช้เสียงตัวเลข 0 และเพิ่มขึ้นเป็น 98% เมื่อใช้เสียงตัวเลขโดดต่อกัน 3 ตัว

คำสำคัญ -- การระบุผู้พูด, กำหนดคำพูดตายตัว, ระบบปิด, ภาษาไทย

1. บทนำ

ในปัจจุบันระบบที่ใช้องค์ประกอบและลักษณะของบุคคลมาระบุตัวบุคคลนั้นๆ (Biometrics personal identification system) เพื่อใช้ในระบบรักษาความปลอดภัย แทนการป้อนรหัสผ่านทางแป้นพิมพ์ (Password) หรือการใช้บัตรแถบแม่เหล็ก (Magnetic card) เป็นที่นิยมมาก เช่น การตรวจสอบลายนิ้วมือ (Fingerprints) การตรวจสอบรูปแบบม่านตา (Retinal patterns) หรือจะเป็นการตรวจสอบใบหน้า (Face recognition) เป็นต้น เหตุผลประการหนึ่งที่ระบบดังกล่าวได้รับความนิยมเพราะยากต่อการปลอมแปลง ในขณะที่การใช้รหัสผ่าน หรือบัตรแถบแม่เหล็กนั้น

[1] บทความนี้ตีพิมพ์ครั้งแรกในเอกสารประกอบการประชุมวิชาการของศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ ปี 2543 หน้า 496-510 และได้รับรางวัลบทความวิชาการดีเด่น

[2] คณะนักวิจัยประกอบด้วย ดร.จุฬารัตน์ ตันประเสริฐ หัวหน้าโครงการ, นายวศิน สินธุภิญโญ, นายเปรมนาถ ดูเบ, นายสุทัศน์ แซ่ตั้ง, นายวารินทร์ อัจฉริยะกุลพร, นายชัย วุฒิวิวัฒน์ชัย และนายศวิต กาสุริยะ

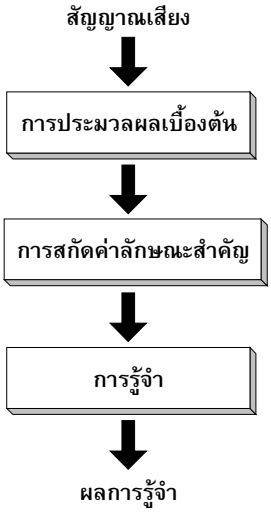
ง่ายต่อการถูกลักขโมย รวมทั้งอาจจะลืมรหัสผ่าน หรือลืมนำบัตรติดตัวมาด้วย

ระบบการรู้จำผู้พูด (Speaker recognition system) ก็เป็นหนึ่งในเทคโนโลยีดังกล่าว ที่ได้รับความสนใจนำมาใช้ในการระบุตัวบุคคล [1,2] นอกเหนือจากระบบระบุตัวบุคคลอื่นๆ การรู้จำผู้พูดสามารถแบ่งออกได้เป็น 2 ประเภทหลักๆ คือ การรับรองผู้พูด (Speaker verification) ซึ่งเป็นการตรวจสอบผู้พูดว่าเป็นบุคคลเดียวกับบุคคลที่กำหนดหรือไม่ และการระบุผู้พูด (Speaker identification) ซึ่งจะทำการตรวจสอบผู้พูดว่าเป็นใคร [3] นอกจากนี้การระบุผู้พูดยังแบ่งได้เป็น 2 อย่างคือ การระบุผู้พูดแบบปิด (Closed-set) เป็นการระบุว่าผู้พูดเป็นบุคคลใดในกลุ่มบุคคลที่กำหนด ในขณะที่การระบุผู้พูดแบบเปิด (Open-set) เป็นการระบุว่าผู้พูดเป็นบุคคลใดในกลุ่มบุคคลที่กำหนด หรือเป็นบุคคลนอกกลุ่ม ระบบการรู้จำผู้พูดยังสามารถแบ่งได้ตามข้อความที่พูดคือ แบบกำหนดคำ หรือประโยคให้พูด (Text dependent) และแบบไม่กำหนดคำ หรือประโยคให้พูด (Text independent) หรือแบ่งตามสถานที่ใช้งาน คือในสภาพแวดล้อมของสำนักงาน (Office environment) และสภาพแวดล้อมทางโทรศัพท์ (Telephone environment)

สำหรับงานวิจัยนี้ เป็นการวิจัยและพัฒนาระบบระบุผู้พูดสำหรับภาษาไทยแบบกำหนดคำพูดตายตัว และเป็นระบบปิด ใช้ในสภาพแวดล้อมสำนักงาน ถือได้ว่าเป็นงานวิจัยในยุคเริ่มแรกของการวิจัยระบบรู้จำผู้พูดสำหรับภาษาไทย และบทความฉบับนี้เป็นบทสรุปความก้าวหน้าของงานวิจัยใน 1 ปีที่ผ่านมา โดยแบ่งหัวข้อดังนี้ หัวข้อที่ 2 จะกล่าวถึงภาพรวมวิธีการระบุผู้พูด โดยให้รายละเอียดคร่าวๆ ของส่วนประกอบต่างๆ ในระบบ หัวข้อที่ 3 จะกล่าวถึงขั้นตอนของงานวิจัยที่ผ่านมาโดยแบ่งแยกเป็นงานวิจัยในส่วนต่างๆ ของระบบ หัวข้อที่ 4 จะกล่าวถึงผลงานล่าสุดที่ให้ผลการระบุผู้พูดสูงที่สุด หัวข้อที่ 5 จะสรุปปัญหา ข้อเสนอแนะและงานวิจัยที่จะดำเนินการต่อไปในอนาคต และสรุปเนื้อหาของบทความนี้ในหัวข้อที่ 6

2. ภาพรวมของระบบระบุผู้พูด

หลักการโดยทั่วไป สำหรับการสร้างระบบระบุผู้พูดได้แสดงไว้ในรูปที่ 1 [3] ประกอบด้วยการประมวลผลเบื้องต้น (Preprocessing) การสกัดค่าลักษณะสำคัญ (Feature extraction) และการรู้จำ (Recognition)



รูปที่ 1. หลักการโดยทั่วไปของระบบระบุผู้พูด

2.1 การประมวลผลเบื้องต้น (Preprocessing)

สัญญาณเสียงที่ผ่านการแปลงสัญญาณเป็นดิจิตอลแล้ว จะนำมาผ่านขั้นตอนการประมวลผลเบื้องต้น ซึ่งประกอบด้วยขั้นตอนต่างๆ ดังนี้

1. การกรองทางความถี่ (Filtering) เป็นขั้นตอนในการกรองสัญญาณในช่วงความถี่ที่ไม่ต้องการออกโดยอาศัยตัวกรองแบบดิจิตอล
2. การตัดหัว-ท้ายเสียง (Endpoint detection) เป็นขั้นตอนในการกำหนดจุดเริ่มต้นและจุดสิ้นสุดของเสียง โดยการแยกส่วนที่เป็นคำพูดออกจากส่วนที่ไม่ใช่คำพูด วิธีในการตัดหัว-ท้ายเสียงมีหลายวิธี เช่น ใช้ค่าระดับพลังงาน (Energy level) ใช้อัตราการตัดศูนย์ (Zero-crossing rate) เป็นต้น
3. การนอร์มอลไลซ์ทางเวลา (Time normalization) เป็นขั้นตอนการเพิ่ม หรือลดขนาดความยาวของสัญญาณในเชิงเวลา เพื่อปรับแต่งขนาดความยาวของสัญญาณให้เหมาะสมตามต้องการ ทั้งนี้จะขึ้นอยู่กับกระบวนการในการรู้จำว่าจำเป็นต้องทำการนอร์มอลไลซ์สัญญาณให้เท่ากันหรือไม่ วิธีในการนอร์มอลไลซ์ทางเวลามีหลายวิธี เช่น การเปลี่ยนอัตราการชักตัวอย่าง (Sampling rate changing) การประมาณค่าในช่วงเชิงเส้น (Linear interpolation) [4] และการเหลื่อมและรวมส่วนย่อยแบบซิงโครไนซ์ (Synchronized overlap-and-add) [5] เป็นต้น

2.2 การสกัดค่าลักษณะสำคัญ (Feature)

การสกัดค่าลักษณะสำคัญ คือการวิเคราะห์หาค่าที่จะใช้แทนสัญญาณเสียง เพื่อนำไปใช้ในขั้นตอนการรู้จำ แบ่งได้เป็น 3 กลุ่มหลัก กลุ่มแรกเป็นค่าลักษณะสำคัญระดับสูง (High level feature) ได้แก่ สำเนียงการพูด รูปแบบในการพูด และความเร็วในการพูด เป็นต้น ในกลุ่มที่สอง จะใช้ค่าลักษณะสำคัญทางฉันทลักษณ์ (Prosodic feature) เช่น ค่าความถี่มูล

ฐาน (Fundamental frequency) ความถี่ฟอร์แมนท์ (Formant frequency) และระดับพลังงาน (Energy profile) เป็นต้น ถึงแม้ว่าค่าลักษณะสำคัญแบบนี้จะมีประสิทธิภาพสูงในการรู้จำ แต่ยากในการสกัดจากสัญญาณ กลุ่มสุดท้ายเรียกว่าค่าลักษณะสำคัญแบบเอนเวโลปของสเปกตรัม (Spectral envelop feature) [6] เป็นกลุ่มที่นิยมใช้กันมาก เนื่องจากค่าลักษณะสำคัญส่วนใหญ่สำหรับการรู้จำเสียงจะรวมอยู่ในข้อมูลเชิงสเปกตรัมนี้ อีกทั้งยังง่ายและสะดวกในการคำนวณหาค่าด้วย ตัวอย่างค่าลักษณะสำคัญแบบนี้ได้แก่ สัมประสิทธิ์การประมาณพันธะเชิงเส้น (Linear prediction coefficients: LPC), สัมประสิทธิ์เซปสตรัม (Cepstral coefficient) และพัฒนาการอีกมากมายจากเซปสตรัมปกติ [7] อาทิเช่น สัมประสิทธิ์เซปสตรัมบนสเกลเมล (Mel frequency cepatral coefficients: MFCC) เซปสตรัมแบบหักลบค่าเฉลี่ย (Cepstral mean substraction: CMS) และเซปสตรัมแบบผ่านตัวกรองภายหลัง (Post filtered cepsturm: PFL) เป็นต้น นอกจากนั้น ยังมีการคำนวณค่าการเปลี่ยนแปลง (Derivative หรือ Delta) ของสัมประสิทธิ์เหล่านี้มาใช้เป็นค่าลักษณะสำคัญเพิ่มเติมได้ด้วย

สำหรับการคำนวณค่าลักษณะสำคัญแบบเอนเวโลปของสเปกตรัมจะมีขั้นตอนดังนี้ [3]

1. การเน้นสัญญาณขั้นต้น (Preemphasis) เป็นขั้นตอนในการบีบอัดสัญญาณเสียงโดยนำสัญญาณเสียงผ่านตัวกรองลำดับหนึ่ง (First-order filter) ซึ่งจะเพิ่มอัตราส่วนสัญญาณต่อสัญญาณรบกวน (Signal to noise ratio)
2. การแบ่งเป็นส่วนย่อย (Frame) เป็นขั้นตอนในการแบ่งสัญญาณเสียงเป็นส่วนย่อย ขนาดความยาวประมาณ 10 – 40 มิลลิวินาที ซึ่งทำให้สัญญาณเสียงมีคุณสมบัติเปลี่ยนแปลงตามเวลาน้อยมาก หรือไม่มีเลย เพื่อให้สามารถสร้างแบบจำลองการกระจายของหน่วยสัญญาณเสียงย่อยทางสถิติได้
3. การลดขอบด้วยฟังก์ชันหน้าต่างสำหรับปรับสัญญาณให้ราบเรียบ (Smoothing window)
4. การสกัดค่าลักษณะสำคัญ (Feature extraction) ในส่วนนี้ จะทำการคำนวณค่าลักษณะสำคัญของสัญญาณเสียงในแต่ละส่วนย่อย ผลลัพธ์อยู่ในรูปแบบของเวกเตอร์ของค่าลักษณะสำคัญ (Feature vector) สำหรับแต่ละส่วนย่อย

2.3 การรู้จำ (Recognition)

ขั้นตอนนี้ประกอบด้วย 2 หน้าที่หลัก คือการนำเวกเตอร์ของค่าลักษณะสำคัญของสัญญาณเสียง ที่อยู่ในชุดอ้างอิงหรือชุดฝึกฝน มาทำการเรียนรู้ เมื่อเรียนรู้แล้วเวกเตอร์ของสัญญาณเสียงที่ต้องการทดสอบการรู้จำ จะถูกนำเข้ามาเทียบเคียงเพื่อรู้จำ ขั้นตอนในการเรียนรู้นั้นขึ้นอยู่กับวิธีในการรู้จำของระบบนั้นๆ บางวิธีก็เพียงแค่เก็บข้อมูลชุดเรียนรู้ไว้เปรียบเทียบกับข้อมูลชุดทอดสอบเท่านั้น เช่น วิธีการรู้จำแบบหาค่าระยะห่างยูคลิเดียน (Euclidean distance) วิธีไดนามิกไทม์วาร์ปปิง (Dynamic time warping: DTW) [6] เป็นต้น ในขณะที่บางวิธี จะนำข้อมูลชุดเรียนรู้ไปแปลงเป็นค่าอ้างอิงที่ต้องการ เช่น โครงข่ายประสาทเทียม (Artificial neural networks: ANN) [8] จะนำข้อมูลชุดเรียนรู้ไปผ่านโครงข่ายที่สร้างขึ้นเพื่อจดจำรูปแบบ และเก็บเป็นค่าน้ำหนัก (Weight) แทน วิธีควอนไตซ์แบบเวกเตอร์ (Vector quantization: VQ) [9] ซึ่งจะแทนเวกเตอร์ทั้งหมดของแต่ละสัญญาณเสียงอ้างอิงด้วยเวกเตอร์จำนวนไม่มาก หรือการใช้แบบจำลองฮิดเดนมาร์คอฟ (Hidden markov model: HMM) [6,9] โดยนำข้อมูลชุดฝึกฝน ไปผ่านแบบจำลองที่สร้างขึ้นเพื่อจดจำรูปแบบ และเก็บค่าทางสถิติและค่าความน่าจะเป็นของแต่ละสถานะไว้ เป็นต้น แต่ทั้งหมดจะมีพื้นฐานอยู่ที่การคำนวณระยะห่างของรูปแบบที่จะรู้จำ และนำค่าระยะห่างที่ได้ไปใช้รู้จำตามแต่ละวิธีนั้นๆ

การเลือกใช้วิธีการรู้จำ ขึ้นอยู่กับข้อกำหนดของงาน เช่น วิธี DTW และ ANN เหมาะสมกับระบบแบบกำหนดคำพูดตายตัว ในขณะที่วิธี VQ และ HMM จะเหมะสมกับระบบงานที่เป็นแบบไม่กำหนดคำพูดมากกว่า [1,9]

3. เส้นทางการวิจัยที่ผ่านมา

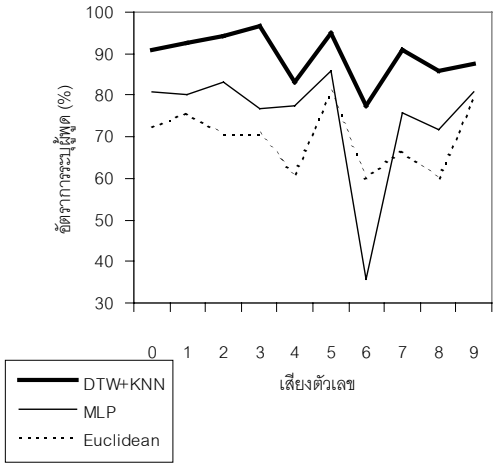
เนื่องจากงานด้านการระบุผู้พูดสำหรับภาษาไทยในประเทศไทยยังไม่มีผลงานวิจัยมากนัก คณะวิจัยจึงเริ่มศึกษาแนวทางจากบทความการระบุผู้พูดจากภาษาต่างประเทศเพื่อเป็นแนวทางในการดำเนินการ งานวิจัยเริ่มแรกควรใช้ระบบระบุผู้พูดแบบกำหนดคำพูดตายตัวเพื่อให้ง่ายต่อการวิจัยและไม่ซับซ้อนเกินไปนัก โดยทดสอบกับเสียงของตัวเลขโดด 0 ถึง 9 ของภาษาไทย ซึ่งคาดว่าจะสามารถประยุกต์นำไปใช้กับระบบรักษาความปลอดภัย ระบุผู้พูด หรือพิสูจน์ผู้พูดได้ เช่น การระบุผู้พูดจากรหัสประจำตัว เป็นต้น

3.1 การทดลองขั้นต้น

ในขั้นเริ่มต้นของงานวิจัย ได้มีความพยายามในการเลือกระบบรู้จำที่จะนำมาใช้ โดยเปรียบเทียบระบบรู้จำอย่างน้อย 3 ระบบ คือวิธีหาระยะห่างแบบยูคลิเดียน, วิธี DTW โดยใช้วิธีตัดสินใจแบบพิจารณาจุดใกล้ K จุด (K-nearest neighbor: K-NN) [2], และการใช้ ANN ชนิดเพอเซปตรอนหลายชั้น (Multilayer perceptron network: MLP) ร่วมกับการเรียนรู้แบบแพร่กระจายย้อนกลับ (Back-propagation) [8]

การทดลองทำการระบุผู้พูดจำนวน 20 คน (ชาย 11 คน หญิง 9 คน) โดยผู้พูดแต่ละคนจะต้องอัดเสียงพูดตัวเลขโดด 0-9 จำนวน 10 ครั้งต่อสัปดาห์ เป็นเวลา 5 สัปดาห์ แบ่งเสียงจาก 3 สัปดาห์แรกเป็นชุดฝึกฝน ที่เหลือเป็นชุดทดสอบ สัญญาณเสียงจะถูกนำมาผ่านผ่านการนอร์มอลไลซ์ทางเวลา

เฉพาะในกรณีของระบบรู้จำแบบยูคลิเดียนและ ANN หลังจากนั้น แบ่งเป็นส่วนย่อยๆ ละ 20 มิลลิวินาที เหลื่อมส่วนย่อยละ 5 มิลลิวินาที และใช้ค่าลักษณะสำคัญแบบ LPC ขนาด 10 อันดับ ผลการทดลอง [10] แสดงดังรูปที่ 2



รูปที่ 2. กราฟเปรียบเทียบผลอัตราการระบุผู้พูดสำหรับวิธี DTW, ANN และระยะห่างแบบยูคลิเดียน

ผลการทดลองชี้ให้เห็นถึงความสามารถของ DTW ซึ่งให้ผลการระบุผู้พูดสูงสุดถึง 96.67% กับเสียงพูดเลข 5 และผลอัตราการระบุผู้พูดเฉลี่ยสำหรับทุกเสียงตัวเลขเท่ากับ 89.42% ในขณะที่ ANN ให้อัตราการระบุผู้พูดสูงสุด 85.83% กับเสียงพูดเลข 3 และอัตราการระบุผู้พูดเฉลี่ย 74.83% ส่วนการใช้ระยะห่างแบบยูคลิเดียนให้ผลต่ำที่สุด โดยมีอัตราการระบุผู้พูดเฉลี่ยเพียง 69.75% เท่านั้น

จากผลการทดลอง ส่วนหนึ่งที่วิเคราะห์ได้คือการใช้วิธีนอร์มอลไลซ์ทางเวลา น่าจะเป็นวิธีที่ลดประสิทธิภาพของการรู้จำได้มาก อย่างไรก็ตาม DTW ได้กลายมาเป็นวิธีที่ถูกนำมาวิจัยและพัฒนาต่อ รวมทั้งการวิจัยวิธีการอื่นๆ ของระบบรู้จำที่หลีกเลี่ยงการนอร์มอลไลซ์ทางเวลา ในขณะเดียวกันการทดลองเพื่อเลือกค่าลักษณะสำคัญที่ให้ประสิทธิภาพสูงขึ้นก็จะทำควบคู่กันไปด้วยกับวิธีการรู้จำแบบต่างๆ เนื่องจากเราไม่สามารถบอกได้ว่า จะใช้ค่าลักษณะสำคัญแบบใดจึงเหมาะสมกับเทคนิคในการเทียบเคียงรูปแบบสัญญาณแต่ละแบบ

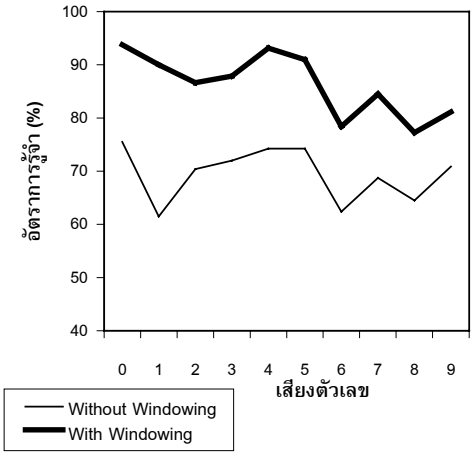
3.2 การวิจัยเกี่ยวกับระบบรู้จำ

หลังจากการทดลองขั้นต้นเกี่ยวกับระบบรู้จำที่จะนำมาใช้ดังได้กล่าวมาแล้ว งานวิจัยก็หันมามุ่งเน้นที่การพัฒนาระบบรู้จำโดยได้ศึกษารายละเอียดของวิธีการรู้จำนั้นๆ เพื่อปรับเปลี่ยนค่าพารามิเตอร์ และขั้นตอนกระบวนการต่างๆ เพื่อให้เหมาะสมกับสัญญาณเสียงภาษาไทย เพื่อเพิ่มผลอัตราการรู้จำให้สูงขึ้น และสามารถรองรับจำนวนผู้พูดได้มากขึ้นโดยไม่มีผลกระทบต่ออัตราการรู้จำ

<u>การทดลองเกี่ยวกับ ANN</u>

ปัญหาสำคัญประการหนึ่งของ ANN แบบ MLP ซึ่งอาศัยการเรียนรู้แบบแพร่กระจายย้อนกลับ คือจะต้องกำหนดจำนวนข้อมูลในชั้นข้อมูลเข้าให้เท่ากันทุกๆ รูปแบบที่จะรู้จำ ซึ่งจำเป็นต้องใช้การนอร์มอลไลซ์ทางเวลามาช่วย ทำให้สูญเสียลักษณะสำคัญบางประเภทที่จำเป็นสำหรับการรู้จำไป อัตราการรู้จำจึงไม่สูงเท่าที่ควร เพื่อไม่ให้ลักษณะสำคัญสูญเสียระหว่างการทำกระบวนการนอร์มอลไลซ์ทางเวลา และเพื่อประหยัดเวลาในการรู้จำ คณะนักวิจัยฯ จึงได้คิดหาวิธีในการส่งข้อมูลลักษณะสำคัญเข้าสู่ชั้นข้อมูลเข้าเพื่อใช้ในการฝึกหัดใหม่ ซึ่งทำให้ไม่มีความจำเป็นในการทำนอร์มอลไลซ์ทางเวลาอีกต่อไป โดยเปลี่ยนให้มีการส่งข้อมูลเข้าแบบหน้าต่าง (Windowing technique) [11] รายละเอียดของกระบวนการจะกล่าวในหัวข้อต่อไป

การทดลองสำหรับระบุผู้พูดจำนวน 20 คนเช่นเดิม โดยใช้ค่าลักษณะสำคัญแบบเซปสตรัมที่คำนวณมาจาก LPC (Linear predictive coding derived cepstrum, LPCC) ขนาด 15 อันดับ ผลการทดสอบแสดงดังรูปที่ 3 ยืนยันว่า วิธีการส่งข้อมูลเข้าแบบหน้าต่างนี้ให้ผลการรู้จำที่สูงกว่าแบบเดิมโดยเฉลี่ยถึง 15% [8] กล่าวคือให้อัตราระบุผู้พูดได้สูงสุดถึง 93.75% สำหรับเสียงตัวเลข 0 และให้อัตราการรู้จำเฉลี่ยสูงถึง 86.37%



รูปที่ 3. กราฟเปรียบเทียบผลอัตราการระบุผู้พูดด้วยวิธี ANN ปกติและแบบใช้เทคนิค Windowing

<u>การทดลองเกี่ยวกับ DTW</u>

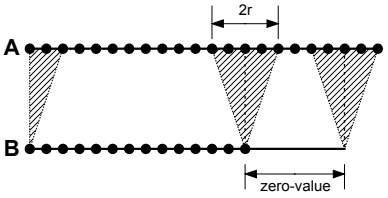
วิธีการรู้จำแบบ DTW เป็นอีกวิธีหนึ่งที่ได้ศึกษาอย่างละเอียด หลังจากได้ศึกษาวิธีการรู้จำแบบ DTW แล้วพบว่าพารามิเตอร์ต่างๆ ใน DTW ได้แก่ การเลือกค่า r หรือกรอบของจุดที่อนุญาตให้มีการจับคู่จุดได้ (Time alignment window) และจำนวนจุดของการก้าวแต่ละครั้งให้เหมาะสม การกำหนดค่า r ไว้คงที่ จะส่งผลกระทบกับการจับคู่ลำดับ 2 ลำดับที่มีความยาวต่างกันมากๆ นอกจากนี้ถ้ามีจำนวนชุดอ้างอิงมากก็จะสูญเสีย

เวลาในการคำนวณระยะห่างมาก และถ้ามีจำนวนผู้พูดมากก็จะเสียเวลาในการคำนวณมากเช่นกัน

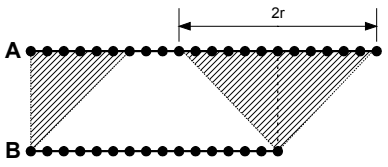
จากปัญหาต่างๆ ดังกล่าว จึงได้มีการทดลอง 3 ส่วน ส่วนแรกได้ทำการทดลองเพื่อปรับค่า r ที่เหมาะสม [12] โดยทำการทดลองระบุผู้พูดจำนวน 20 คน ใช้ค่าลักษณะสำคัญแบบ LPC ขนาด 10 อันดับ ปรากฏว่าค่า r ที่เหมาะสมขึ้นอยู่กับความยาวของเสียงที่ใช้ในการระบุผู้พูด สำหรับเสียงตัวเลขโดด ค่า r เท่ากับ 5 ให้ผลการระบุได้สูง และค่า r ควรจะเพิ่มขึ้นเมื่อใช้เสียงพูดยาวขึ้น

ส่วนที่ 2 เป็นการเสนอขั้นตอนกระบวนการในการกำหนดค่า r รวมทั้งแก้ไขปัญหาความแตกต่างของความยาวของลำดับที่เทียบเคียงกัน ในส่วนนี้ได้เสนอเทคนิคการเทียบเคียงด้วย DTW 3 เทคนิค ดังแสดงในรูปที่ 4 เทคนิคแรกเป็นการกำหนดค่า r คงที่และมีการเพิ่มค่าศูนย์ต่อท้ายสำหรับลำดับที่สั้นกว่า เทคนิคที่ 2 เป็นการกำหนดค่า r ให้เท่ากับความแตกต่างของความยาวของลำดับที่เทียบเคียงกัน เทคนิคสุดท้ายเป็นการยืดจุดที่ควรจะเทียบเคียงกัน ตามสัดส่วนของความยาวแล้วจึงกำหนดค่า r ให้คงที่ค่าหนึ่ง ผลการทดลอง [13] ปรากฏว่าในการระบุผู้พูดจำนวน 50 คน โดยใช้ค่าลักษณะสำคัญแบบ LPCC ขนาด 15 อันดับ เทคนิคที่ 1 และ 3 ให้ผลการระบุผู้พูดเฉลี่ยได้ใกล้เคียงกันกล่าวคือ 84.53% และ 84.29% ตามลำดับ เทคนิคที่ 1 ให้ผลดีกว่าเล็กน้อยและยังใช้จำนวนการคำนวณน้อยกว่าอีกด้วย

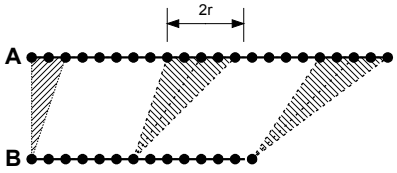
ส่วนสุดท้ายเป็นการทดลองปรับเปลี่ยนจำนวนเสียงอ้างอิงจากเดิมใช้เสียงจาก 3 สัปดาห์แรก (30 เสียง) เป็น 20 และ 10 เสียงโดยคัดแบบคละกันจาก 3 สัปดาห์แรก ผลการทดลอง [13] ปรากฏให้เห็นว่าสำหรับการระบุผู้พูด 50 คน และใช้ค่าลักษณะสำคัญแบบ LPCC ขนาด 15 อันดับ จำนวนเสียงอ้างอิงเท่ากับ 20 ให้ผลการทดลองได้ถึง 84.61% ซึ่งสูงกว่าค่าอื่น แสดงให้เห็นว่าการใช้ชุดอ้างอิงจำนวนมากอาจทำให้ระบบเกิดความสับสนได้มากขึ้น



ก. เทคนิคที่ 1



ข. เทคนิคที่ 2



ค. เทคนิคที่ 3

รูปที่ 4. ขั้นตอนกระบวนการ DTW 3 เทคนิค

การทดลองระบบรู้จำแบบอื่นๆ

นอกเหนือจาก ANN และ DTW แล้ว ยังได้มีความพยายามใช้ระบบรู้จำแบบอื่น ได้แก่ วิธี VQ, วิธี HMM แบบไม่ต่อเนื่อง (Discrete hidden markov model: DHMM) และวิธีแบบจำลองส่วนผสมแบบเกาส์ (Gaussian mixture model: GMM) ซึ่งกำลังอยู่ในขั้นตอนการทดลอง

การทดลองวิธีการตัดสินใจเมื่อใช้เสียงตัวเลขต่อกัน

การทดลองอีกส่วนหนึ่ง คือการเพิ่มความยาวของคำพูดที่ใช้ในระบบระบุผู้พูดโดยใช้เสียงตัวเลขโดดต่อกัน เพื่อเพิ่มประสิทธิภาพของการระบุผู้พูด โดยคัดเลือกเอาค่าลักษณะสำคัญของเสียงตัวเลขโดดที่ให้อัตราการระบุผู้พูดสูงที่สุดอันดับแรกๆ มาต่อกันแล้วจึงเข้าระบบรู้จำวิธีการนี้จะให้ผลอัตราการรู้จำสูงขึ้นกว่าการใช้เสียงตัวเลขโดด อย่างไรก็ตามการใช้ตัวเลขต่อกันนี้ จะทำให้ระบบใช้เวลาในการประมวลผลมากขึ้นโดยเฉพาะอย่างยิ่ง ระบบที่ใช้ DTW ซึ่งต้องกำหนดค่า r ให้กว้างขึ้น ยิ่งเป็นการเพิ่มเวลาในการรู้จำมากขึ้นไปอีก วิธีการหนึ่งสำหรับระบบระบุผู้พูดที่ใช้ DTW และ K-NN ซึ่งได้เสนอไว้ใน [13] คือการรวมเสียงอ้างอิง K เสียงที่ได้จากการระบุผู้พูดด้วยตัวเลขโดดแต่ละตัว เช่น ถ้าสำหรับตัวเลขโดด ใช้กฎการตัดสินใจแบบ 5-NN เมื่อใช้ตัวเลข 3 ตัวต่อกันในระบบระบุผู้พูด จะทำการคำนวณ 5-NN ของตัวเลขโดดแต่ละตัวแล้วจึงนำมารวมกันพิจารณาแบบ 15-NN เป็นต้น

การทดลองระบุผู้พูดจำนวน 50 คนโดยใช้ LPCC ขนาด 15 อันดับ กับตัวเลขต่อกัน 3 ตัวด้วย DTW ผลปรากฏว่าวิธีตัดสินใจแบบใหม่ให้ผลอัตราการระบุผู้พูด 96.10% ในขณะที่วิธีแบบเก่า คือการนำค่าลักษณะสำคัญของแต่ละเลขมาต่อกันก่อน โดยใช้ r = 20 จะให้อัตราการรู้จำ 95.40% ซ้ำยังใช้เวลาในการประมวลผลนานกว่ามาก หลักการเดียวกันนี้สามารถนำไปใช้ได้ในระบบระบุผู้พูดที่ใช้ ANN แบบใหม่ได้เช่นกัน

3.3 การวิจัยเกี่ยวกับค่าลักษณะสำคัญของเสียง

หลังจากได้ทดลองเกี่ยวกับระบบที่ใช้ในการเทียบเคียงพอสมควรแล้ว คณะนักวิจัยได้เริ่มหันมาพิจารณาค่าลักษณะสำคัญที่เหมาะสมสำหรับระบบแต่ละระบบ ค่าลักษณะสำคัญที่เป็นที่นิยม ได้รับการพิสูจน์แล้วว่ามีประสิทธิภาพสูง สำหรับการรู้จำผู้พูดหรือรู้จำเสียงพูดได้ถูกนำมาใช้ในการทดลองเปรียบเทียบ

ในช่วงต้นของงานวิจัย ได้มีการใช้ค่าสัมประสิทธิ์การประมาณพันธะเชิงเส้น (Linear prediction coefficient: LPC) เป็นค่าลักษณะสำคัญ ต่อมาจึงหันมาใช้ค่าสัมประสิทธิ์เซปสตรัม (Cepstral coefficient) ชนิดที่คำนวณมาจาก LPC เรียกว่า LPCC (Linear predictive coding derived cepstrum) โดยมีการทดลองที่แสดงไว้ใน [14] เป็นการทดลองระบุผู้พูดจำนวน 20 คน โดยใช้ค่าลักษณะสำคัญแบบ LPC และ LPCC ขนาด 10 และ 15 อันดับ ใช้ระบบรู้จำแบบ DTW และ K-NN ผลการทดลองดังตารางที่ 1 แสดงให้เห็นอย่างชัดเจนว่า LPCC ให้ผลการระบุผู้พูดสูงกว่า LPC มากทั้งแบบ 10 และ 15 อันดับ สำหรับ LPCC แล้ว ใช้ขนาด 15 อันดับให้ผลดีกว่า 10 ลำดับโดยให้ผลการระบุผู้พูดเฉลี่ยถึง 86.28%

ตารางที่ 1. การทดลองเปรียบเทียบอัตราการรู้จำของระบบที่ใช้ LPC และ LPCC ขนาด 10 และ 15 อันดับ

ตัวเลข	อัตราการระบุผู้พูด (%)			
	LPC		LPCC	
	10	15	10	15
0	84.00	83.75	90.50	91.00
1	71.00	68.75	86.00	89.75
2	64.75	69.50	83.00	87.00
3	74.25	67.75	83.75	85.00
4	77.50	68.50	91.00	93.00
5	70.50	70.75	90.50	91.50
6	69.00	65.25	82.25	85.75
7	60.75	56.75	81.75	84.75
8	48.50	49.50	64.25	70.00
9	70.50	66.50	86.25	85.00
เฉลี่ย	69.08	66.70	83.93	86.28

ในงานวิจัยถัดมา ค่าลักษณะสำคัญอีกหลายชนิดโดยเฉพาะค่าลักษณะสำคัญที่อยู่ในกลุ่มของเซปสตรัมได้ถูกนำขึ้นมาทดลอง อาทิเช่น เซปสตรัมแบบหักลบค่าเฉลี่ย (Cepstral mean substraction: CMS) เซปสตรัมแบบให้น้ำหนักที่ปรับส่วนประกอบได้ (Adaptive component weighted cepstrum: ACW) เซปสตรัมบนสเกลเมล (Mel frequency cepstral coefficient: MFCC) และเซปสตรัมแบบผ่านตัวกรองภายหลัง (Post filtered cepstrum: PFL) ในจำนวนนี้ ค่าลักษณะสำคัญที่ให้ผลการระบุผู้พูดได้สูงได้แก่ MFCC และ PFL จึงมีการทดลองเปรียบเทียบค่าลักษณะสำคัญ 3 ค่าคือ LPCC, MFCC และ PFL กับการระบุผู้พูดจำนวน 50 คน และกำหนดอันดับของค่าลักษณะสำคัญให้คงที่ที่ 15 อันดับ ปรากฏว่าทั้งการทดลองโดยใช้ ANN ที่ป้อนข้อมูลแบบหน้าต่าง และการทดลองโดยใช้ DTW และ K-NN [14] ผลการระบุผู้พูดเฉลี่ยจะสูงที่สุดเมื่อใช้ MFCC ซึ่งสูงกว่า PFL เพียงเล็กน้อยในขณะที่ LPCC ให้ผลต่ำที่สุด รายละเอียดของการคำนวณและผลการทดลองจะได้กล่าวในหัวข้อถัดไป

นอกจากค่าลักษณะสำคัญที่กล่าวมาแล้ว ยังได้มีการวิจัยที่ทดลองใช้ค่าลักษณะสำคัญแบบอื่นๆ อีก อาทิเช่น การประมาณพันธะแบบอิงการรับฟังของมนุษย์ (Perceptual linear predictive: PLP) ซึ่งยังไม่ได้ผลการระบุผู้พูดสูงนัก และการผสมค่าลักษณะสำคัญปกติกับค่าการเปลี่ยนแปลง (Derivative) ซึ่งแม้จะให้ผลดีกว่าแบบปกติก็จริง แต่ต้องใช้ลำดับของค่าลักษณะสำคัญจำนวนมาก เป็นการลดความเร็วของการประมวลผล

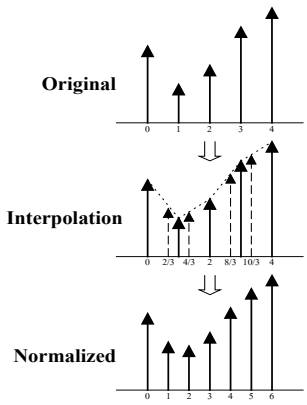
3.4 การทดลองอื่นๆ

นอกจากงานวิจัยในส่วนหลักที่กล่าวมาแล้ว ยังมีการวิจัยในรายละเอียดส่วนอื่นๆ ที่มีความสำคัญควรค่าแก่การพิจารณา เพื่อเพิ่มโอกาสในการพัฒนาผลการระบุผู้พูดให้ดีขึ้น ในที่นี้มีการวิจัยเพิ่มเติม 2 ส่วนดังนี้

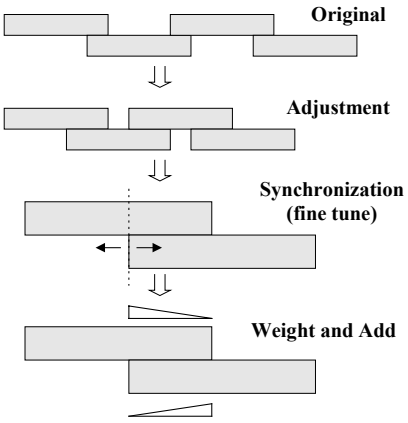
<u>เทคนิคการนอร์มอลไลซ์ทางเวลา</u>

ในขั้นตอนหนึ่งของการวิจัย ได้มีความพยายามปรับปรุงระบบรู้จำที่ใช้ ANN แบบปกติ เนื่องจากสามารถรู้จำได้โดยใช้เวลาประมวลผลไม่มากเมื่อเทียบกับ DTW สำหรับการพัฒนาระบบ ANN ดังที่ได้กล่าวมาแล้ว ส่วนสำคัญที่น่าจะเป็นตัวฉุดความสามารถของการรู้จำลง คือการนอร์มอลไลซ์ทางเวลา ซึ่งเป็นการปรับให้เสียงมีความยาวเท่ากันก่อนผ่านเข้าระบบ เพื่อให้ได้จำนวนข้อมูลที่จะส่งเข้า ANN เท่ากันทั้งหมด

วิธีการนอร์มอลไลซ์ทางเวลามีหลายวิธี ได้แก่ การเปลี่ยนอัตราการชักตัวอย่าง (Sampling rate changing) [4] การประมาณค่าในช่วงเชิงเส้น (Linear interpolation) [4] การเหลื่อมและรวมส่วนย่อยแบบซิงโครไนซ์ (Synchronized overlap-and-add: SOLA) [5] วิธีที่ดีจะต้องทำให้สัญญาณเสียงเพี้ยนไปจากเดิมน้อยที่สุดเท่าที่จะเป็นไปได้ จากการวิจัยที่ผ่านมาได้เปรียบเทียบ 2 วิธีคือ การประมาณค่าในช่วงเชิงเส้นและ SOLA หลักการของการนอร์มอลไลซ์ทางเวลาทั้ง 2 แบบแสดงในรูปที่ 5



ก. การประมาณค่าในช่วงเชิงเส้น



ข. SOLA

รูปที่ 5. การนอร์มอลไลซ์ทางเวลา 2 วิธี.

การประมาณค่าในช่วงเชิงเส้นเป็นวิธีที่ง่าย ทำโดยการเพิ่มหรือลดจำนวนเสียงของสัญญาณตัวอย่างให้มีขนาดตามต้องการ โดยเสียงสัญญาณใหม่จะถูกสร้างขึ้นจากสัญญาณเดิมสองข้างที่อยู่ติดกัน วิธีนี้จะทำให้สัญญาณเสียงเพี้ยนไป (Aliasing) แต่อย่างไรก็ตามระบบการเทียบเคียงสัญญาณเสียงก็ยังสามารถทำการรู้จำสัญญาณเสียงได้ ส่วนวิธี SOLA จะให้ความสำคัญในการปรับสัดส่วนลักษณะสำคัญของเสียงและคุณสมบัติทางเวลาให้มีความคล้ายคลึงกับสัญญาณเสียงต้นแบบมากที่สุด โดยการตัดสัญญาณเสียงเป็นช่วง ๆ แล้วนำมาซ้อนทับกัน ปรับเปลี่ยนระยะทางของสัญญาณที่นำมาซ้อนทับกัน โดยขึ้นอยู่กับสัดส่วนของเวลาที่ต้องการ ให้น้ำหนักความสำคัญของสัญญาณในแต่ละช่วงก่อนที่จะไปรวมกัน

การทดลองกับผู้พูดจำนวน 20 คนโดยใช้ค่าลักษณะสำคัญแบบ LPCC ขนาด 15 อันดับกับเสียงตัวเลขโดด 0-9 ผลการทดลองปรากฏว่าวิธี SOLA ให้อัตราการระบุผู้พูดเฉลี่ยถึง 75.33% [15] ในขณะที่วิธีที่ใช้การประมาณค่าในช่วงเชิงเส้นให้ผลเพียง 67.28 แสดงให้เห็นถึง ความสำคัญของการคงลักษณะดั้งเดิมของเสียงไว้ ดังนั้นถ้าเป็นไปได้ควรหลีกเลี่ยงการนอร์มอลไลซ์ทางเวลา

ผลกระทบของระดับเสียงของคำพูดที่ใช้ระบุผู้พูด

เสียงของภาษาไทยมีความแตกต่างจากเสียงภาษาต่างประเทศอยู่หลายประการ จุดสำคัญจุดหนึ่งคือ ภาษาไทยมีระดับเสียง (Tone) โดยมีห้าระดับคือ สามัญ (Middle) เอก (Low) โท (Falling) ตรี (High) จัตวา (Rising) งานวิจัยอีกส่วนหนึ่งคือการเปรียบเทียบผลของการใช้เสียงพูดในระดับเสียงต่างๆ ในการระบุผู้พูด โดยมีเป้าหมายเพื่อคัดเลือกคำที่เหมาะสมมาใช้ในการระบุผู้พูด

การทดลองระบุผู้พูดจำนวน 9 คน โดยใช้ค่าลักษณะสำคัญแบบ LPC ขนาด 10 อันดับและระบบรู้จำแบบ ANN คำพูดที่ใช้ในการระบุผู้พูดประกอบด้วย 6 ประโยคคือ “เออเออเออเอ”, “เอ่เอ่เอ่เอ่เอ่”, “เอ้เอ้เอ้เอ้เอ้”, “เอ๊เอ๊เอ๊เอ๊เอ๊”, “เอ๋เอ๋เอ๋เอ๋เอ๋” และชุดสุดท้ายเป็นวรรณยุกต์ผสมคือ “เอเอ่เอ้เอ๊เอ๋” ผลการทดลอง [16] พบว่าชุดที่เป็นเสียงวรรณยุกต์ผสมให้ผลการระบุผู้พูดสูงที่สุดคือ 95.56% ส่วนวรรณยุกต์เดี่ยวๆ พบว่าเสียงวรรณยุกต์ในกลุ่มที่มีการเปลี่ยนแปลงในพยางค์คือวรรณยุกต์โทและจัตวาจะให้ผลได้ดีกว่าชุดอื่นๆ

4. ระบบระบุผู้พูดในปัจจุบัน

ณ ปัจจุบันนี้ ผลการทดลองได้ผลดีที่สุดถึง 92.30% สำหรับคำพูดตัวเลขโดด และเพิ่มขึ้นสูงกว่า 98% เมื่อใช้เสียงของตัวเลขต่อกัน ระบบดังกล่าวมีรายละเอียดดังต่อไปนี้

4.1 สัญญาณเสียง

ในงานวิจัย ได้ทำการอัดเก็บเสียงในรูปแบบของสัญญาณดิจิตอล โดยผ่านไมโครโฟนที่ต่อกับคอมพิวเตอร์ผ่านทางการ์ดเสียงปกติ กำหนดให้เก็บเสียงในรูปแบบของไฟล์ WAV อัตราการชักตัวอย่าง (Sampling rate) ที่ 11.025 กิโลเฮิร์ต ตัวอย่างละ 16 บิต และแบบช่องสัญญาณเดียว (Mono)

ทำการอัดเสียงจากผู้พูดจำนวน 50 คน (ชาย 30 คนและหญิง 20 คน) เป็นเวลา 5 สัปดาห์ ในแต่ละสัปดาห์ ผู้พูดแต่ละคนจะต้องพูดเสียงตัวเลขโดด 0-9 เป็นภาษาไทย ตัวเลขละ 10 ครั้ง และเพื่อป้องกันการที่ผู้พูดชินกับการพูดตัวเลขต่อๆ กัน จึงมีการพัฒนาโปรแกรมเฉพาะสำหรับการอัดโปรแกรมจะทำการสุ่มตัวเลขโดด 0-9 ขึ้นแสดงบนจอคอมพิวเตอร์ทีละตัว ผู้พูดจะต้องพูดเสียงตัวเลขที่แสดงเท่านั้น โปรแกรมจะกำหนดช่วงเวลาของการพูดไว้ไม่เกิน 1 วินาทีต่อหนึ่งตัวเลข หากผู้พูดคนใดพูดก่อนช่วงเวลาที่กำหนดหรือพูดช้ากว่าเวลาที่กำหนด โปรแกรมจะสั่งให้พูดใหม่อีกครั้งโดยอัตโนมัติ โดยพิจารณาจากค่าพลังงานของสัญญาณในช่วงเวลา 1 วินาทีดังกล่าว หลังจากนั้นจะแบ่งสัญญาณเสียงออกเป็น 2 กลุ่ม สัญญาณเสียงในสัปดาห์ที่ 1-3 ใช้สำหรับเป็นชุดฝึกฝนหรือชุดอ้างอิง ส่วนสัปดาห์ที่ 4-5 ใช้เป็นชุดทดสอบ

4.2 การประมวลผลขั้นต้น

กระบวนการประมวลผลสัญญาณขั้นต้นประกอบด้วย การกรองสัญญาณ (Filtering) โดยใช้ตัวกรองแบบดิจิตอลชนิดผ่านความถี่สูง กำหนดจุดผ่านความถี่ (Cutoff frequency) ที่ 200 เฮิร์ต เพื่อป้องกันสัญญาณรบกวนความถี่ต่ำที่เกิดจากแหล่งกำเนิดไฟฟ้า ต่อจากนั้นจะผ่านการตัดหัว-ท้าย

เสียง (Endpoint detection) ในที่นี้อาศัยวิธีการตัดโดยพิจารณาจากค่าพลังงานของเสียง [4] โดยไม่มีการนอร์มอลไลซ์ทางเวลา

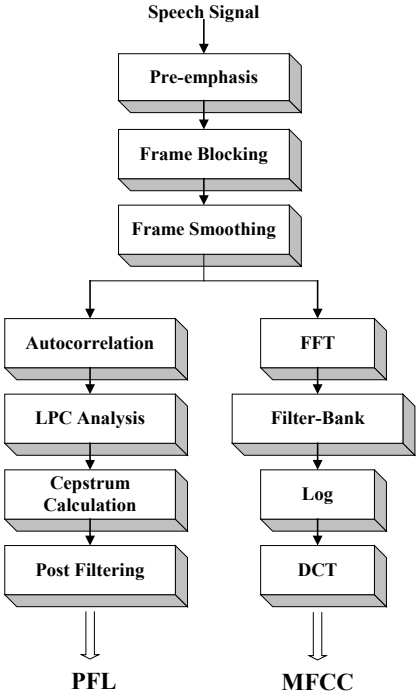
4.3 การสกัดค่าลักษณะสำคัญ

สัญญาณเสียงที่ผ่านการประมวลผลขั้นต้นแล้วจะนำมาผ่านการเน้นสัญญาณเบื้องต้น (Preemphasis) ด้วยตัวกรองอันดับที่ 1 (First order filter) เพื่อเพิ่มค่าอัตราส่วนสัญญาณต่อสัญญาณรบกวน (Signal-to-noise ratio) หลังจากนั้นจะตัดแบ่งสัญญาณเสียงออกเป็นส่วนย่อย (Frame) ขนาดส่วนย่อยละ 20 มิลลิวินาที โดยเหลื่อมส่วนย่อยละ 5 มิลลิวินาที แต่ละส่วนย่อยจะผ่านการปรับให้ราบเรียบ (Smoothing) ด้วยแฮมมิงวินโดว์ (Hamming window) หลังจากนั้นจึงทำการสกัดค่าลักษณะสำคัญ ค่าลักษณะสำคัญที่ให้ผลการระบุผู้พูดสูงที่สุด ณ ปัจจุบันคือ MFCC และ PFL ขนาด 15 อันดับ ซึ่งมีรายละเอียดดังต่อไปนี้

MFCC [6] – ค่าสัมประสิทธิ์เซปสตรัมเป็นค่าลักษณะสำคัญที่นิยมมากทั้งในระบบรู้จำผู้พูดและเสียงพูด โดยพื้นฐานแล้วเซปสตรัมสามารถคำนวณได้จาก การแปลงโคซายน์แบบไม่ต่อเนื่อง (Discrete cosine transformation) ของค่าลอการิธึม (Logarithm) ของสเปกตรัม (Spectrum) ของสัญญาณเสียงแต่ละส่วนย่อย สเปกตรัมของสัญญาณเสียงสามารถหาได้โดยการแปลงฟูริเยร์แบบไม่ต่อเนื่อง (Discrete Fourior transformation) หรือการแปลงฟูริเยร์แบบเร็ว (Fast Fourior transformation) ขั้นตอนดังกล่าวตั้งอยู่บนพื้นฐานแนวคิดที่ว่า สเปกตรัมของสัญญาณเสียงกำเนิดจากส่วนประกอบ 2 ส่วนคือ เอนเวโลปของสเปกตรัม (Spectral envelop) และโครงสร้างรายละเอียดของสเปกตรัม (Spectral fine structure) ทั้ง 2 ส่วนสามารถแยกกันได้ด้วยการใส่ลอการิธึม สัมประสิทธิ์เซปสตรัมเป็นการแทนสัญญาณในส่วนเอนเวโลปของสเปกตรัมเท่านั้น

พัฒนาการหนึ่งของเซปสตรัม คือการผ่านสเปกตรัมของสัญญาณเสียงเข้าไปในกลุ่มของตัวกรอง (Filter bank) ซึ่งกระจายอยู่บนสเกลความถี่แบบไม่สม่ำเสมอ เช่น การกระจายตามสเกลเมล (Mel scale) [6] ซึ่งออกแบบมาให้เหมาะสมกับการรับฟังของหู เป็นต้น ค่าพลังงานของสเปกตรัมของเสียงที่ได้จากตัวกรองแต่ละตัวจะถูกนำมาใช้คำนวณค่าสัมประสิทธิ์เซปสตรัมแทนค่าสเปกตรัมปกติ ค่าสัมประสิทธิ์เซปสตัมที่ได้จากการกระทำเช่นนี้จึงได้ชื่อว่า MFCC

PFL [7,17] – อีกวิธีการหนึ่งของการคำนวณค่าสัมประสิทธิ์เซปสตรัมคือการคำนวณจากค่าสัมประสิทธิ์ LPC วิธีการคำนวณรวมทั้งเหตุผลของการคำนวณแสดงไว้ใน [3,18] หลังจากนั้นมีการเสนอค่าลักษณะสำคัญแบบใหม่ โดยผ่านค่าเซปสตรัมที่ได้เข้าไปยังตัวกรองซึ่งเรียกว่า ตัวกรองภายหลัง (Post filter) ตัวกรองดังกล่าวจะทำการเน้นค่าสเปกตรัมของเสียง ณ บริเวณความถี่ฟอร์แมนท์ (Formant frequency) ซึ่งเป็นการเพิ่มความโดดเด่นของสัญญาณเสียงทั้งในแง่การรู้จำเสียงพูดและรู้จำผู้พูด ภาพรวมของการคำนวณค่าลักษณะสำคัญทั้ง 2 วิธีที่กล่าวมาแสดงไว้ในรูปที่ 6

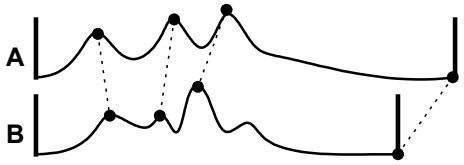


รูปที่ 6. ขั้นตอนการคำนวณค่าลักษณะสำคัญ

4.4 ระบบรู้จำ

ในปัจจุบันระบบรู้จำที่ให้ผลการระบุผู้พูดสูงสุดคือ DTW โดยใช้กฎการตัดสินใจแบบ K-NN รองลงมาคือ ANN โดยใช้วิธีป้อนข้อมูลเป็นช่วงของส่วนย่อย ส่วนต่อไปจะอธิบายหลักการคร่าวๆ ของแต่ละวิธี

DTW และ K-NN [6] – DTW เป็นวิธีการหนึ่งของการโปรแกรมพลวัต (Dynamic programming) ใช้ในการเทียบเคียงเพื่อหาระยะห่างระหว่างลำดับ 2 ชุดซึ่งยาวไม่เท่ากัน กำหนดให้ลำดับ $A = \{a_1, a_2, \ldots, a_I\}$ และ $B = \{b_1, b_2, \ldots, b_J\}$ เป็น 2 ลำดับที่จะเทียบเคียงกัน DTW จะทำการหาจุดเทียบเคียงที่ให้ค่าระยะห่างรวมต่ำที่สุด ดังแสดงในรูปที่ 7



รูปที่ 7. ภาพแสดงวิธีการเทียบเคียงแบบ DTW

โดยอาศัยขั้นตอนกระบวนการดังต่อไปนี้

ขั้นที่ 1: กำหนดค่าเริ่มต้น $D(a_1,b_1) = 2d(a_1,b_1)$ โดยที่ $d(a_i, b_j)$ เป็นค่าระยะห่างระหว่างจุด a_i และ b_j อาจใช้ระยะห่างแบบยูคลิเดียนก็ได้

ขั้นที่ 2: คำนวณแบบวนซ้ำหาจุดเทียบเคียง a_i และ b_j ที่เหมาะสมโดยตั้งอยู่บนพื้นฐานที่ว่า $D(a_i,b_j)$ จะต้องให้ค่าต่ำที่สุด ดังสมการ

$$D(a_i,b_j)=\min\begin{cases} D(a_{i-1},b_j)+d(a_i,b_j) \\ D(a_{i-1},b_{j-1})+2d(a_i,b_j) \\ D(a_i,b_{j-1})+d(a_i,b_j) \end{cases} \qquad (1)$$

ทั้งนี้จะมีข้อกำหนดดังต่อไปนี้

1) $1 \le i \le I, 1 \le j \le J$
2) จุดเทียบเคียงจุดแรกคือ (a_1,b_1) และ (a_I, b_J) เป็นจุดเทียบเคียงจุดสุดท้าย
3) สำหรับแต่ละจุด (a_i, b_j) ที่เทียบเคียงกัน $|i-j| \le r$
4) $0 \le i^{k+1} - i^k \le 1, 0 \le j^{k+1} - j^k \le 1$ โดย k เป็นดัชนีรอบของการเทียบจุด

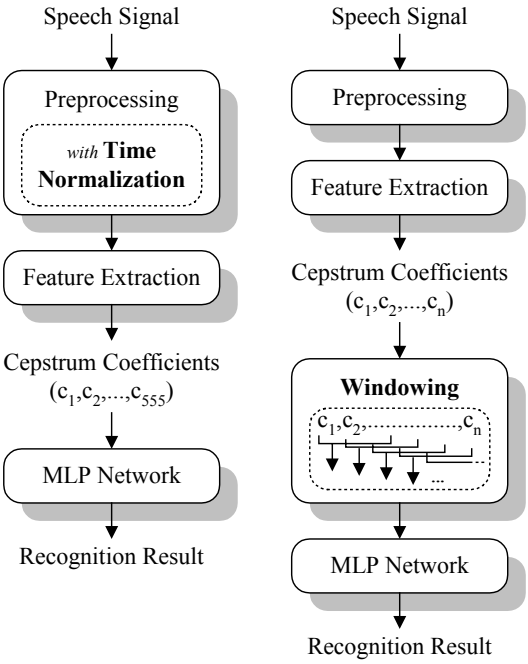
ขั้นที่ 3: ค่าระยะห่างรวมของ 2 ลำดับคือ $\frac{D(a_I,b_J)}{I+J}$

สมการที่ (1) จะตรงตามข้อกำหนดข้อที่ 4 ในตัวเอง กล่าวคืออนุญาตให้มีการขยับจุดที่จะเทียบเคียงได้ทีละ 1 จุดเท่านั้น ค่า r ในข้อกำหนดข้อที่ 3 เป็นค่าที่สำคัญที่ใช้ในการกำหนดความห่างของจุดที่เทียบเคียงกัน เพื่อให้บรรลุตามข้อกำหนดข้อที่ 2 จะมีการเติมเวกเตอร์ศูนย์ต่อท้ายลำดับที่สั้นกว่าดังที่ได้กล่าวมาแล้วในหัวข้อ 3.2 เพื่อให้สามารถจับคู่จุดสุดท้ายได้พอดี

เมื่อได้ค่าระยะห่างระหว่างสัญญาณเสียงที่เข้ามาทดสอบกับสัญญาณเสียงในชุดอ้างอิงแล้ว จะใช้วิธี K-NN ในการตัดสินใจ คือการพิจารณาสัญญาณเสียงอ้างอิง K ตัวที่ให้ค่าระยะห่างต่ำที่สุด ว่าไปตกลงที่สัญญาณเสียงของผู้พูดคนใดมากกว่ากัน ก็จะตอบเป็นผู้พูดคนนั้น ในการทดลองนี้ใช้ 5-NN ในการตัดสินใจ และจะเปลี่ยนเป็น 1-NN เมื่อ 5-NN ไม่สามารถตัดสินใจได้

ANN [8,19] – ANN ที่ใช้เป็นแบบ MLP และวิธีการเรียนรู้แบบแพร่กระจายย้อนกลับ โดยพัฒนาวิธีการป้อนข้อมูลเข้าแบบใหม่คือแบบหน้าต่าง เพื่อหลีกเลี่ยงการนอร์มอลไลซ์ทางเวลา หลักการที่พัฒนาขึ้นนี้เทียบกับวิธีการป้อนข้อมูลแบบเก่าได้แสดงไว้ในรูปที่ 8

ANN แบบเก่าจะมีจำนวนโหนดในชั้นข้อมูลเข้าเท่ากับ 555 โหนดคงที่ (15 ลำดับ * 37 ส่วนย่อยของเสียงที่ผ่านการนอร์มอลไลซ์ทางเวลามา) ส่วนจำนวนโหนดที่ชั้นข้อมูลออกจะเท่ากับจำนวนผู้พูดที่จะระบุและมีเพียง 1 โครงข่ายเท่านั้น แต่โครงสร้างแบบใหม่จะมี 1 โครงข่ายต่อผู้พูด 1 คน มีจำนวนโหนดในชั้นข้อมูลเข้าเพียง 60 โหนด (15 ลำดับ * 4 ส่วนย่อย) คือเลื่อนข้อมูลเข้าทีละ 4 ส่วนย่อยและเหลื่อมครั้งละ 3 ส่วนย่อย มีโหนดในชั้นข้อมูลออกเพียง 2 โหนด ซึ่งทำการรู้จำว่าใช่หรือไม่ใช่ผู้พูดคนนั้น



รูปที่ 8. วิธีการป้อนข้อมูลเข้า ANN

4.5 ผลการระบุผู้พูด

การทดลองระบุผู้พูดจำนวน 50 คนโดยอาศัยค่าลักษณะสำคัญและระบบรู้จำที่ได้กล่าวมาแล้ว แบ่งเป็น 2 ส่วน ส่วนแรกเป็นการระบุผู้พูดโดยใช้เสียงพูดตัวเลขโดด 0-9 ผลการทดลองแสดงไว้ในตารางที่ 2

ตารางที่ 2. ผลการระบุผู้พูดเมื่อใช้เสียงตัวเลขโดด

ตัวเลข	*อัตราการระบุผู้พูด (%)*			
	DTW		*ANN*	
	MFCC	*PFL*	*MFCC*	*PFL*
0	*89.1*	*90.7*	*86.3*	*87.4*
1	*89.4*	*89.4*	*86.3*	*86.4*
2	*87.8*	*86.7*	*80.9*	*82.1*
3	*87.6*	*87.9*	*86.5*	*81.1*
4	*85.7*	*86.9*	*79.4*	*78.6*
5	*92.3*	*89.1*	*90.1*	*82.1*
6	*85.3*	*79.7*	*74.2*	*72.8*
7	*84.2*	*83.6*	*78.2*	*77.0*
8	*79.9*	*75.5*	*77.0*	*72.6*
9	*86.1*	*85.5*	*84.5*	*82.6*
เฉลี่ย	*86.74*	*85.50*	*82.34*	*80.27*

ตารางที่ 3. ผลการระบุผู้พูดเมื่อใช้เสียง 3 ตัวเลขต่อกัน

ระบบ		ตัวเลข	ผลการระบุผู้พูด (%)
DTW	MFCC	"510"	98.70
	PFL	"015"	98.80
ANN	MFCC	"530"	97.30
	PFL	"019"	96.40

เพื่อเพิ่มประสิทธิภาพของการระบุผู้พูด จึงได้ใช้เสียงพูดที่ยาวขึ้นโดยการต่อเสียงตัวเลขโดดเป็น 3 ตัว ตัวเลขโดดที่นำมาต่อกันนั้นจะเลือกมาจากตัวเลขที่ให้ผลการระบุผู้พูดสูงที่สุด 3 อันดับแรกจากการทดลองกับเสียงตัวเลขโดด ผลการระบุผู้พูดรวมทั้งตัวเลขต่อกันที่ใช้ในการทดลองแสดงได้ดังตารางที่ 3

4.6 ซอฟต์แวร์ต้นแบบ

เมื่องานวิจัยมาถึงจุดที่ให้ผลการระบุผู้พูดที่สูง โดยมีอัตราการระบุผู้พูดสูงเกินกว่า 90% คณะนักวิจัยจึงได้พัฒนาซอฟต์แวร์ต้นแบบสำหรับระบุผู้พูด ใช้กับเสียงตัวเลขโดด โดยผู้ใช้สามารถเลือกได้ว่าจะใช้ DTW และ KNN หรือใช้ ANN ในการรู้จำ ทั้งนี้ซอฟต์แวร์จะกำหนดให้ผู้พูดแต่ละคนอัดเสียงตัวเลขที่กำหนดจำนวน 3 ครั้ง ระบบจะนำไปใช้เป็นชุดอ้างอิงสำหรับ DTW หรือนำไปฝึกฝนสำหรับ ANN ก่อนขั้นตอนการทดสอบระบุผู้พูดจริง ผู้สนใจสามารถติดต่อขอชมการทำงานของซอฟต์แวร์ต้นแบบได้ที่ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ

5. อุปสรรคและงานในอนาคต

งานวิจัยและพัฒนาระบบระบุผู้พูดสำหรับภาษาไทยได้ดำเนินการผ่านมา 1 ปีเต็มแล้ว พบอุปสรรคในการดำเนินงานอยู่หลายประการ ได้แก่

1. การรวบรวมข้อมูลจากการอัดเสียง เนื่องจากงานวิจัยมีวัตถุประสงค์ที่ต้องการสร้างระบบที่สามารถรู้จำผู้พูดที่มีประสิทธิภาพสูง ไม่ว่าเวลาจะผ่านไปนานเท่าไร ระบบควรจะสามารถระบุผู้พูดได้ในอัตราความถูกต้องใกล้เคียงเดิม ดังนั้นขบวนการจัดเก็บเสียงจึงจัดเก็บในหลายสัปดาห์ต่อเนื่องกัน ซึ่งทางคณะนักวิจัยได้ขอความร่วมมือจากพนังงานของศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติที่ประจำอยู่ ณ อาคารมหานครยิบซั่มช่วยสละเวลามาอัดเสียงให้ ปัจจุบันรวบรวมได้ 50 คน คณะนักวิจัยคาดหวังว่าจะเก็บเสียงได้อย่างน้อย 100 คนในอนาคต โดยอัดเสียงนักศึกษาฝึกงานและอาจจัดตั้งโครงดารความร่วมมือกับองค์กรต่างๆ เพื่อพัฒนาฐานข้อมูลเสียงภาษาไทย
2. เวลาในการระบุผู้พูดของ DTW ค่อนข้างนานมากเมื่อเทียบกับ ANN แต่เนื่องจากอัตราการรู้จำผู้พูดของ DTW ดีกว่าผลจาก ANN คณะนักวิจัยจึงต้องคิดค้นหาวิธีปรับปรุงเทคนิค DTW ให้สามารถทำงานได้เร็วยิ่งขึ้นหรือพัฒนา ANN ให้สามารถรู้จำได้ถูกต้องมากขึ้น นอกจากนี้คณะนักวิจัยกำลังทดลองเทคนิคการรู้จำผู้พูดวิธีอื่นๆ ด้วย เพราะคาดว่าจะได้ระบบระบุผู้พูดที่ดีมีประสิทธิภาพสูงขึ้นกว่าระบบในปัจจุบัน
3. ฐานความรู้เกี่ยวกับเสียงภาษาไทยค่อนข้างมีจำกัด ส่งผลให้คณะนักวิจัยมีความจำเป็นต้องทดลองในทุกๆ สมมติฐานที่ตั้งขึ้นเอง ซึ่งทำให้ต้องใช้เวลาในการวิจัยมากขึ้น ดังนั้นเมื่อคณะนักวิจัยได้ผลการทดลองจึงได้พยายามจัดทำบทความวิชาการเพื่อเผยแพร่ความรู้อยู่ตลอดเวลา ด้วยความหวังที่ว่าความรู้เหล่านั้นจะได้ช่วยเสริมให้เกิดงานวิจัยทางด้านดาระระบุผู้พูดในประเทศไทยมากขึ้น และลดงานที่ซ้ำซ้อนลงไป

6. บทสรุป

ระบบระบุผู้พูดมีความสำคัญมากกับการเพิ่มประสิทธิภาพของระบบรักษาความปลอกภัย และยังเป็นพื้นฐานที่สำคัญในการพัฒนาระบบรู้จำเสียงพูด (Speech recognition system) สำหรับภาษาไทยอีกด้วย ในปัจจุบันคณะนักวิจัยได้พัฒนาซอฟต์แวร์ต้นแบบระบุผู้พูดสำหรับภาษาไทยที่ให้อัตราความถูกต้องเฉลี่ย 98% กับผู้พูดจำนวน 50 คน คณะนักวิจัยจะพัฒนาระบบนี้ให้สามารถทำงานได้ดีกับผู้พูดจำนวนมากยิ่งขึ้น และจะพัฒนาระบบระบุผู้พูดที่ใช้งานได้กับเสียงพูดผ่านทางสายโทรศัพท์ด้วย นอกจากนี้ทางคณะนักวิจัยกำลังพัฒนาและรวบรวมฐานข้อมูลเสียงพูดตัวเลข 0-9 เพื่อให้นักวิจัยจากที่อื่นๆ สามารถเข้ามาในเว็บไซต์และดึงข้อมูลเพื่อนำไปใช้ทดลองและคิดค้นเทคนิคใหม่ๆ เพื่องานระบุผู้พูดสำหรับภาษาไทยที่ดีขึ้นอีกด้วย

เอกสารอ้างอิง

[1] J. P. Campbell, Jr., "Prolog to Speaker Recognition: A Tutorial", Proceedings of IEEE, Vol. 85, No. 9, pp. 1437-1462, September 1997.

[2] G. R. Doddington, "Speaker Recognition-Identifying People by their Voices", *Proceedings of IEEE*, Vol. 73, No. 11, pp.1651-1664, November 1985.

[3] S. Furui, "Digital Speech Processing, Synthesis, and Recognition", New York and Basel: Marcel Dekker, Inc, 1989.

[4] ชัย วุฒิวิวัฒน์ชัย, "การรู้จำเสียงคำหลายพยางค์แบบไม่ขึ้นกับผู้พูด โดยใช้เทคนิคแบบฟัซซีและนิวรอลเน็ตเวอร์ก", วิทยานิพนธ์วิศวกรรมศาสตรมหาบัณฑิต จุฬาลงกรณ์มหาวิทยาลัย, 2540

[5] S. Roucos and A.M. Wilgus, "High Quality Time Scale Modification for Speech," *IEEE International Conference ASSP*, pp. 493-496, 1985.

[6] L. R. Rabiner and B. -H. Juang, "Fundamentals of Speech Recognition", A. Oppenheim, Series Editor, Englewood Cliffs, NJ: Prentice-Hall, 1993.

[7] R. J. Mammone, X. Zhang, and R. P. Ramachandran, "Robust Speaker Recognition, A Feature-based Approach", *IEEE Signal Processing Magazine*, p. 58-71, September 1996.

[8] L Fausette, "Fundamentals of Neural Networks–Architecture, Algorithm, and Applications", Prentice-Hall, 1994.

[9] K. Yu, J. Mason, and J. Oglesby, "Speaker Recognition using Hidden Markov Models, Dynamic Time Warping and Vector Quantisation*", IEE Proc.-Vis. Image Signal Process*, Vol. 142, No. 5, October 1995.

[10] วศิน สินธุภิญโญ, เปรมนาถ ดูเบ, สุทัศน์ แซ่ตั้ง, วารินทร์ อัจฉริยะกุลพร, ชัย วุฒิวิวัฒน์ชัย และจุฬารัตน์ ตันประเสริฐ, "การระบุผู้พูดด้วย LPC และ DTW สำหรับภาษาไทย", *เอกสารประกอบการประชุมวิชาการ ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ สำนักงานพัฒนาวิทยาศาสตร์และเทคโนโลยีประจำปีงบประมาณ 2542 ณ ศูนย์ประชุมสหประชาชาติ 30 มีนาคม-เมษายน 2542*

[11] S. Sae-Tung and C. Tanprasert, "Feature Windowing based Thai Text-Dependent Speaker Identification using MLP and Backpropagation Algorithm*", Proceedings of the International Symposium on Circuits and Systems*, May 2000.

[12] C. Wutiwiwatchai, V. Achariyakulporn, and C. Tanprasert, "Text-dependent Speaker Identification using LPC and DTW for Thai Language", *1999 IEEE 10th Region Conference (TENCON'99)*, Vol. 1, September 1999.

[13] วารินทร์ อัจฉริยะกุลพร, ชัย วุฒิวิวัฒน์ชัย และ จุฬารัตน์ ตันประเสริฐ, "ระบบระบุผู้พูดภาษาไทยด้วยวิธีไดนามิกส์ไทม์วาร์ปปิง", *กำลังพิจารณาเพื่อตีพิมพ์ใน NECTEC Technical Journal*, Vol.2, No. 7, 2000

[14] C. Wutiwiwatchai and C. Tanprasert, "Thai Text-Dependent Speaker Identification: Features Comparison", *The 4th Symposium on National Language Processing*, May 2000.

[15] C. Wutiwiwatchai, S. Sae-Tung, and C. Tanprasert, "Thai Text-Dependent Speaker Identification by ANN with Two Time Normalization Techniques", *Proceedings of the 1st Workshop on Natural Language Processing and Neural Networks*, pp. 47-52, November 1999.

[16] C. Tanprasert, C. Wutiwiwatchai, and S. Sae-tang, "Text-dependent Speaker Identification Using Neural Network on Distinctive Thai Tone Marks", Proceedings of *International Joint Conference on Neural Networks*, July 1999.

[17] M. S. Zilovic, R. P. Ramachadran, and R. J. Mammone, "Speaker Identification Based on the Use of Robust Cepstral Features Obtained from Pole-Zero Transfer Functions", *IEEE Transactions on Speech and Audio Processing*, Vol.6, No.3, pp.260-267, May 1998.

[18] S. Furui, "Cepstral Analysis Technique for Automatic Speaker Verification", *IEEE Transaction on Acoustic, Speech Signal Processing*, Vol. ASSP-29, pp.254-272, April 1981.

[19] SNNS (Stuttgart Neural Network Simulator) User Manual, Version 4.1, University of Stuttgart, Institute for Parallel and Distributed High Performance Systems (IPVR), Report No. 6/95.

นายชัย วุฒิวิวัฒน์ชัย ตำแหน่งผู้ช่วยนักวิจัย ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ สำเร็จการศึกษา ปริญญาโท (วิศวกรรมศาสตร์มหาบัณฑิต) ปี 2540 จุฬาลงกรณ์มหาวิทยาลัย ประสบการณ์การทำงาน โครงการระบบรู้จำเสียงพูดภาษาไทยซึ่งอยู่ในช่วงเริ่มต้น ผลงานเด่น มีชื่อเป็นผู้แต่งบทความวิชาการที่ได้รับตีพิมพ์ภายในประเทศจำนวน 1 บทความ และระดับนานาชาติ จำนวน 7 บทความ

วารินทร์ อัจฉริยะกุลพร สำเร็จการศึกษาระดับปริญญาตรีสาขาวิทยาการคอมพิวเตอร์ เกียรตินิยมอันดับ 2 จากมหาวิทยาลัยเกษตรศาสตร์ ปี พ.ศ. 2538 และการศึกษาระดับปริญญาโทสาขาวิทยาการคอมพิวเตอร์ จากมหาวิทยาลัยมหิดล ปี พ.ศ. 2541 ปัจจุบันได้ร่วมงานในหน่วยปฏิบัติการวิจัยและพัฒนาวิศวกรรมภาษาและซอฟต์แวร์ โดยรับผิดชอบในหน้าที่การพัฒนาโปรแกรมพัฒนาเว็บเพจ และฐานข้อมูล นอกจากนี้ยังมีส่วนร่วมรับผิดชอบในโครงการแก้ปัญหาคอมพิวเตอร์ปี ค.ศ. 2000 โครงการ Asean-India Digital Archive (AIDA) และโครงการระบบระบุผู้พูดภายใน NECTEC มีความสนใจในด้านการออกแบบระบบงาน การออกแบบโปรแกรมเชิงวัตถุ และงานทางด้านปัญญาประดิษฐ์

สุทัศน์ แซ่ตั้ง สำเร็จการศึกษาระดับปริญญาตรีสาขาระบบสารสนเทศ เกียรตินิยมอันดับ 2 จากสถาบันเทคโนโลยีรัชมงคล ปี พ.ศ. 2536 และการศึกษาระดับปริญญาโทสาขาวิทยาการคอมพิวเตอร์ จากมหาวิทยาลัยมหิดล ปี พ.ศ. 2541 ปัจจุบันได้ร่วมงานในหน่วยปฏิบัติการวิจัยและพัฒนาวิศวกรรมภาษาและซอฟต์แวร์ โดยรับผิดชอบในงานวิจัย และพัฒนาโปรแกรมแปลงรูปภาพเอกสารเป็นข้อความ (Thai OCR) โครงการระบบระบุผู้พูดด้วยเสียง นอกจากนี้ยังมีส่วนร่วมรับผิดชอบในโครงการแก้ปัญหาคอมพิวเตอร์ปี ค.ศ. 2000 มีความสนใจในด้านการประมวลรูปภาพ (Image Procesisng) การรู้จำรูปแบบ (Pattern Recognition) และการออกแบบพัฒนาโปรแกรมเชิงวัตถุ (Object Oriented Systems)