รายงานผลการวิจัย

เรื่อง

การศึกษาระบบรู้จำเสียงพูดอัตโนมัติ

The Study of The Speech Recognition

ส่วนวิจัยและพัฒนาอุปกรณ์พิเศษ กองเทคโนโลยีและศูนย์ข้อมูลการตรวจสอบ

กรมสอบสวนคดีพิเศษ กระทรวงยุติธรรม

ปีงบประมาณ พ.ศ. 2561

คำนำ

งานวิจัย เรื่อง “การศึกษาระบบรู้จำเสียงพูดอัตโนมัติ (The Study of The Speech Recognition)” เป็นงานวิจัยที่ศึกษาเกี่ยวกับแนวทางในการพัฒนาระบบรู้จำเสียงพูดอัตโนมัติ โดยได้นำแนวคิดเรื่องระบบรู้จำเสียงพูดอัตโนมัติมาพัฒนาเพื่อแปลสัญญาณเสียงเป็นข้อความทันที ซึ่งเป็นการช่วยลดระยะเวลาในการในการทำงานของเจ้าหน้าที่คดีพิเศษและพนักงานสอบสวนคดีพิเศษ ทั้งนี้ ตัวชี้วัดตามแผนยุทธศาสตร์ กรมสอบสวนคดีพิเศษ ปี พ.ศ. 2560-2564 ประจำปีงบประมาณ พ.ศ. 2561 ตัวชี้วัดที่ 2.6 ระดับความสำเร็จของการดำเนินการเกี่ยวกับการวิจัยหรือผลงานทางวิชาการเพื่อการป้องกันการกระทำความผิดที่มีลักษณะเป็นคดีพิเศษที่มีประสิทธิภาพ

โดยในส่วนของการวิเคราะห์และออกแบบระบบ ผู้วิจัยได้เก็บข้อมูลด้วยแบบสอบถามปลายเปิด ซึ่งเป็นการรวบรวมข้อมูลการทำงาน ปัญหา อุปสรรค และความต้องการต่อการพัฒนาระบบรู้จำเสียงพูดของผู้ใช้งาน จากนั้น นำข้อมูลที่ได้มาวิเคราะห์และออกแบบระบบรู้จำเสียงพูดอัตโนมัติให้สอดคล้องกับความต้องการของผู้ใช้งานโดยเริ่มจากในขั้นตอนของการสอบสวนพนักงานสอบสวนใช้คอมพิวเตอร์ที่มีไมโครโฟน ทำการเลือกชนิดของเอกสารที่ต้องการใช้งาน ไม่ว่าจะเป็นเอกสารต้นแบบ เอกสารที่ไม่มีต้นแบบ ที่สามารถเรียกได้อีกอย่างว่าเอกสารเปล่า จากนั้นพนักงานสอบสวนสั่งการระบบผ่านไมโครโฟน โดยจะมีไมโครโฟนประจำตัวของแต่ละคน จากนั้นข้อมูลเสียงของการสอบสวนหรือสนทนาทำการประมวลผล เพื่อตรวจสอบคุณลักษณะของเสียงที่ไม่พึงประสงค์ออกไป เช่นเสียงรบกวน เป็นต้น จากนั้นระบบทำการส่งข้อมูลเสียงแยกตามไมโครโฟนของแต่ละคนไปยังโปรแกรมพาที (PARTY) ที่พัฒนาโดยศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ (NECTEC) เพื่อประมวลผลให้ออกมาเป็นข้อความตัวหนังสืออัตโนมัติตามเสียงพูด

เมื่อได้ตัวหนังสือตามที่พนักงานสอบสวนได้พูดหรือสั่งการแล้ว ระบบจะทำการจับคู่ข้อความกับตำแหน่งของข้อความในไฟล์เอกสารต้นแบบ โดยจะเป็นการเติมข้อความแบบอัตโนมัติ ซึ่งหลักการจะคล้ายคลึงกับระบบสั่งการด้วยเสียงอัตโนมัติ จากนั้นระบบทำการจัดเก็บข้อมูลเสียงและข้อความจากการสอบสวนหรือสนทนา เพื่อใช้ประโยชน์ในการสืบค้นได้ในอนาคต

งานวิจัยนี้ได้ดำเนินการโดยมีระเบียบวิธีวิจัยที่ถูกต้องตามหลักวิชาการว่าด้วยการศึกษาระเบียบวิธีวิจัย พิจารณาตามขอบเขต บทบาทหน้าที่ และภารกิจหลัก ตามเจตนารมณ์ของการจัดตั้งกรมสอบสวนคดีพิเศษ ทั้ง 4 ด้าน ดังนี้

1. ดำเนินการป้องกัน ปราบปราม และควบคุมอาชญากรรมพิเศษ
2. พัฒนาระบบ รูปแบบ วิธีการและมาตรการในการป้องกัน ปรามปราม และควบคุมอาชญากรรม
3. พัฒนาบุคลากรในการเสริมสร้างศักยภาพในด้านความรู้ ความสามารถ คุณธรรมจริยธรรม และขวัญกำลังใจ
4. ประสานความร่วมมือในการป้องกันปราบปรามและควบคุมอาชญากรรมพิเศษจากหน่วยงานภาครัฐและภาคประชาชน ทั้งภายในและต่างประเทศ

ส่วนวิจัยและพัฒนาอุปกรณ์พิเศษ

กันยายน 2561

สารบัญ

หน้า

คำนำ 1
สารบัญ 2
สารบัญภาพ 3
บทที่ 1 บทนำ 5
1.1 ความเป็นมา 5
1.2 วัตถุประสงค์ 6
1.3 เป้าหมาย 6
1.4 ขอบเขตการดำเนินการ 6
1.5 กรอบแนวคิด 6
1.6 ระเบียบวิธีวิจัย 6
1.7 ขั้นตอนและระยะเวลาดำเนินงาน 7
1.8 ประโยชน์ที่คาดว่าจะได้รับ 7
บทที่ 2 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง 8
2.1 หลักการทำงานพื้นฐาน 8
2.2 เครื่องมือการแปลงเสียงพูดเป็นข้อความ 9
2.3 Voice Type 9
2.4 Siri 10
2.5 Windows Speech Recognition 10
2.6 Dragon 11
2.7 Speech to Text 11
2.8 พาที (PARTY) 11
2.9 แนวทางในการพัฒนาระบบรู้จำเสียงพูด 13
บทที่ 3 การวิเคราะห์และออกแบบระบบ 22
3.1 การวิเคราะห์และศึกษาสภาพปัญหา 22
3.2 การวิเคราะห์และออกแบบระบบ 25
บทที่ 4 บทสรุป 29
4.1 สรุปผลการดำเนินงาน 29
4.2 ปัญหาและอุปสรรค 29
4.3 แนวทางการดำเนินการต่อไป 29

สารบัญภาพ

ภาพที่ หน้า
2.1 ส่วนประกอบหลักของระบบรู้จำเสียงพูด (Speech Recognition) 8
2.2 หลักการโดยทั่วไปของระบบระบุคำพูด 13
2.3 ฟังก์ชันกรอบชนิด Hamming Windows 14
2.4 การกรองความถี่สัญญาณเสียงพูดแบบ low-pass filter 14
2.5 การตัดหัว-ท้ายเสียง ของสัญญาณเสียงพูดที่ต่อเนื่อง 15
2.6 การสุ่มสัญญาณ (signal sampling) และการควอนไทซ์ (quantization)................................ 15
2.7 ตัวอย่างเสียงพูด.. 16
2.8 แผนภาพแสดง HMM แบบ 5 สถานะ ที่มีการเปลี่ยนแปลงสถานะแบบซ้ายไปขวา 19
2.9 แบบจำลองทางคณิตศาสตร์ของสมองมนุษย์ 20
2.10 โครงข่ายประสาทเทียมแบบหลายชั้น ที่ใช้การสอนแบบอัลกอริทึมการแพร่ย้อนกลับ (Backpropagation Neural Network).. 20
2.11 ภาพรวมขั้นตอนการใช้งาน 21
3.1 การใช้งานกับเอกสารต้นแบบ (template).. 23
3.2 ใช้งานกับเอกสารที่ไม่มีต้นแบบ (template) 24
3.3 สามารถนำไฟล์เสียงเข้าสู่ระบบและแปลงเป็นข้อความ 24
3.4 ไมโครโฟนลักษณะต่าง 24
3.5 แผนภาพบริบท (Context Diagram): ระบบรู้จำเสียงพูดอัตโนมัติเพื่องานสอบสวน 25
3.6 Data Flow Diagram Level 1 : ระบบรู้จำเสียงพูดอัตโนมัติเพื่องานสอบสวน 26

สารบัญตาราง

ตารางที่ หน้า

1 คำอธิบายการประมวลผลของโปรเซสที่ 1.0 : เลือกเอกสารต้นแบบ..27
2 คำอธิบายการประมวลผลของโปรเซสที่ 2.0 : การรู้จำเสียงพูด(Speech Recognition)...............27
3 คำอธิบายการประมวลผลของโปรเซสที่ 3.0 : การรู้จำเสียงพูด (Speech Recognition)..............28

บทที่ 1
บทนำ

1.1 ความเป็นมา

กรมสอบสวนคดีพิเศษมีพันธกิจเกี่ยวกับการป้องกัน การปราบปราม การสืบสวนและการสอบสวนคดีความผิดทางอาญาที่ต้องดำเนินการสืบสวนและสอบสวนโดยใช้วิธีการพิเศษตามกฎหมายว่าด้วยการสอบสวนคดีพิเศษ โดยในปัจจุบันการเก็บข้อมูลการสอบสวนบุคคลต่างๆ อยู่ในรูปแบบของข้อความที่ต้องถูกถอดออกมาจากเสียงในขั้นตอนการสอบสวน ซึ่งเป็นกระบวนการที่ซับซ้อนและใช้เวลามาก ดังนั้นคณะผู้วิจัยจึงมีแนวคิดที่จะนำเอาระบบรู้จำเสียงพูดโดยอัตโนมัติ (Automatic Speech Recognition) เข้ามาเพื่อแก้ปัญหาดังกล่าว โดยหลักการทำงานคือ เมื่อผู้ใช้พูดใส่ไมโครโฟน ระบบรู้จำเสียงพูดอัตโนมัติ (Automatic Speech Recognition) จะแปลงสัญญาณเสียงเป็นข้อความที่พร้อมนำไปใช้งานต่อได้ทันที

ในการดำเนินชีวิตประจำวันของมนุษย์นั้นมีการติดต่อสื่อสารกับคอมพิวเตอร์อยู่ตลอด อาจจะโดยการติดต่อด้วยมือผ่านอุปกรณ์นำเข้า (Input Devices) ที่ปัจจุบันได้มีการพัฒนาขึ้นมาหลายชนิด เช่น คีย์บอร์ด หรือ เมาส์ หรือ ปุ่มที่มีอยู่หลากหลายบนแป้นโทรศัพท์ แต่การติดต่อที่เป็นธรรมชาติมากที่สุดคือ การใช้เสียงพูด การติดต่อกับคอมพิวเตอร์ด้วยเสียงพูดนั้นช่วยอำนวยความสะดวกในการใช้งานคอมพิวเตอร์ให้กับผู้ใช้ต่าง ๆ ได้ เช่น ผู้สูงอายุที่ไม่มีความชำนาญในการใช้อุปกรณ์นำเข้าที่มีอยู่มากมายหลายชนิด และผู้ที่มีความทุพพลภาพทางร่างกาย ได้แก่ ผู้พิการทางมือและผู้พิการทางสายตา

ระบบรู้จำเสียงพูดอัตโนมัติ (Automatic Speech Recognition) เป็นเทคโนโลยีที่สามารถนำไปใช้ได้อย่างหลากหลาย อย่างเช่น ในอุตสาหกรรมดูแลสุขภาพ (Health Care) ผู้ที่ใช้ประโยชน์จากเทคโนโลยีนี้คือ ฝ่ายธุรการ และหมอ พยาบาล เภสัชกรที่ไม่ถนัดการพิมพ์, หรือแม้กระทั่งทางการทหาร ก็สามารถนำเทคโนโลยีนี้ไปใช้เพื่อสั่งการระบบนักบินอัตโนมัติ (Autopilot), ติดตั้งความถี่คลื่นวิทยุ หรือควบคุมการบิน (flight display) เป็นต้น นอกจากนี้ เทคโนโลยีนี้ยังสามารถนำไปใช้ประโยชน์อย่างอื่นได้อีก เช่น การแปลอัตโนมัติ, การสั่งการรถยนต์, การโทรสนเทศ (Telematics), การรายงานในศาล (Court reporting หรือ Real-time Voice Writing), คอมพิวเตอร์แฮนด์ฟรี, โทรศัพท์มือถือ, หุ่นยนต์, ระบบตอบรับอัตโนมัติ (Interactive Voice Response) และการควบคุมการจราจรทางอากาศ เป็นต้น

อีกงานที่สำคัญอย่างหนึ่งของการใช้งานระบบรู้จำเสียงพูด (Speech Recognition) เทคโนโลยีคือแปลงเสียงจากคำพูดของเรากลายเป็นตัวหนังสือได้อย่างแม่นยำ เพียงแค่พูดภาษานั้นชัดเจนก็สามารถแปลงเป็นตัวเขียนได้ ซึ่งจะเป็นการลดเวลาการทำงานของหน่วยงานหลาย ๆ หน่วยงานได้โดยไม่จำเป็นต้องไปนั่งฟังและพิมพ์ตาม ซึ่งเป็นการสิ้นเปลืองเวลาและทำงานซ้ำซ้อนหลายขั้นตอน

ส่วนวิจัยและพัฒนาอุปกรณ์พิเศษ กองเทคโนโลยีและศูนย์ข้อมูลการตรวจสอบ มีหน้าที่รับผิดชอบงานด้านวิจัยและพัฒนาอุปกรณ์พิเศษของกรมสอบสวนคดีพิเศษ ได้เล็งเห็นถึงความจำเป็นในการใช้งานระบบรู้จำเสียงพูดอัตโนมัติเพื่องานสอบสวนคดีพิเศษ จึงได้ดำเนินการวิจัย

เรื่อง “การศึกษาระบบรู้จำเสียงพูดอัตโนมัติ” โดยคาดหวังว่าการวิจัยเรื่องนี้จะเป็นแนวทางที่นำไปสู่การพัฒนางานทางด้านการสืบสวนและสอบสวนคดีพิเศษต่อไป

1.2 วัตถุประสงค์

เพื่อศึกษาองค์ความรู้ ข้อมูลที่เกี่ยวข้อง และการเก็บข้อมูลโดยแบบสอบถามของเจ้าหน้าที่คดีพิเศษและพนักงานสอบสวนคดีพิเศษ เพื่อนำมาวิเคราะห์และออกแบบระบบรู้จำเสียงพูดอัตโนมัติ

1.3 เป้าหมาย

ได้แนวทางและรูปแบบสำหรับการพัฒนาระบบรู้จำเสียงพูดอัตโนมัติสำหรับงานทางด้านการสืบสวนและสอบสวนคดีพิเศษ

1.4 ขอบเขตการดำเนินการ

1) ศึกษาการจัดเก็บข้อมูลเสียงเพื่อสามารถนำไปใช้ในการแปลงเป็นตัวหนังสือภายหลังได้

2) วิเคราะห์และออกแบบระบบจัดเก็บเสียงพูดหรือสนทนา

1.5 กรอบแนวคิด

อำนวยความสะดวกสำหรับเจ้าหน้าที่คดีพิเศษและพนักงานสอบสวนคดีพิเศษของกรมสอบสวนคดีพิเศษ เพื่อลดเวลาในการบันทึกข้อมูลการสอบสวนจากเดิมที่ใช้การพิมพ์สัมผัสตามเสียงพูด โดยมีการปรับเปลี่ยนมาเป็นระบบพิมพ์ตัวหนังสืออัตโนมัติตามเสียงพูดในเวลาเดียวกันและเป็นแบบเรียลไทม์ (real-time) และยังสามารถจัดเก็บลงในระบบฐานข้อมูลโดยอัตโนมัติเพื่อใช้ประโยชน์ในการสืบค้นภายหลังได้

1.6 ระเบียบวิธีวิจัย

1. ศึกษาความเป็นมาและความสำคัญของปัญหา หลักการและเหตุผล วัตถุประสงค์ของการศึกษาวิจัย ขอบเขตการศึกษา วิธีการศึกษาและวิเคราะห์ ผลประโยชน์ที่คาดว่าจะได้รับหรือเกี่ยวกับการกระทำความผิดที่มีลักษณะเป็นคดีพิเศษหรือเรื่องที่จะศึกษาวิจัย

2. ทบทวน กรอบแนวคิด ทฤษฎี วรรณกรรม กฎหมาย ที่เกี่ยวข้อง

3. กำหนดกรอบแนวคิดในการวิจัย และแบบการเก็บข้อมูล

4. เก็บข้อมูล และรวบรวมข้อมูลที่ทำการวิจัย โดยใช้แบบสอบถามสำรวจปัญหาและความต้องการของผู้ใช้งาน

5. วิเคราะห์ข้อมูล ตามระเบียบวิธีวิจัย และจัดทำรายงานผลการวิจัย

1.7 ขั้นตอนและระยะเวลาดำเนินงาน

ลำดับ	ขั้นตอนการดำเนินงานหลัก	ระยะเวลาการดำเนินงาน (ปี พ.ศ.2559)							
		ส.ค.				ก.ย.			
		1	2	3	4	1	2	3	4
1.	ศึกษาทฤษฎีที่เกี่ยวของ	———	———	———					
2.	รวบรวมแบบสอบถามและความต้องการของผู้ใช้งานเพื่อออกแบบระบบรู้จำเสียงพูดอัตโนมัติ		———	———	———				
3.	ออกแบบระบบรู้จำเสียงพูดอัตโนมัติ			———	———	———			
4.	จัดกิจกรรมการนำเสนอร่างผลงานวิจัยและรับฟังความเห็นที่มีต่อร่างผลงานวิจัยเรื่อง “การศึกษาระบบรู้จำเสียงพูดอัตโนมัติ”						———		
5.	วิเคราะห์ผลการทดลองฯ						———	———	———
6.	จัดทำเอกสารายงานฯ	———	———	———	———	———	———	———	———

1.8 ประโยชน์ที่คาดว่าจะได้รับ

กรมสอบสวนคดีพิเศษได้แนวทางการพัฒนาระบบรู้จำเสียงพูดอัตโนมัติสำหรับงานด้านการสืบสวนและสอบสวนคดีพิเศษ ให้อยู่ในรูปแบบของตัวหนังสือตามเสียงพูดได้จากขั้นตอนกระบวนการสอบสวนได้

บทที่ 2
ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 หลักการทำงานพื้นฐาน

เทคโนโลยีการแปลงเสียงพูดเป็นข้อความ เป็นสร้างข้อความตามคำบอก (Dictation) โดยทำความเข้าใจความหมายของลำดับข้อมูลที่สกัดออกมาจากสัญญาณเสียงพูดแปลงเป็นข้อความ โดยอาศัยพื้นฐานของเทคโนโลยีการรู้จำเสียงพูด (Speech Recognition) โดยระบบรู้จำเสียงพูด มีส่วนประกอบหลัก 5 ส่วน ดังรูปที่ 2.1 คือ

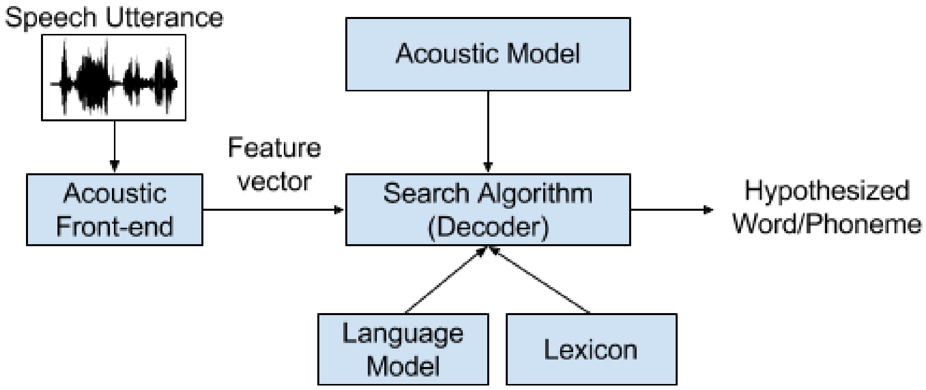
1) ส่วนรับเสียง (Acoustic Front-end) เป็นส่วนที่ทำหน้าที่เกี่ยวกับการประมวลผลสัญญาณ (signal processing) และสกัดคุณลักษณะ (feature extraction)

2) โมเดลเสียง (Acoustic Model) เป็นส่วนที่จัดเก็บความรู้สำหรับระบบรู้จำเสียงพูด

3) โมเดลภาษา (Language Model) เป็นส่วนที่รวบรวมเงื่อนไขของลำดับคำในภาษา

4) ตัวตรวจจับ (Decoder) เป็นส่วนที่ค้นหาความคล้ายของลำดับคำจากเสียงที่ได้รับ

5) คลังศัพท์ (Lexicon) เป็นส่วนที่จัดเก็บคำศัพท์ทั้งหมด



ภาพที่ 2.1 ส่วนประกอบหลักของระบบรู้จำเสียงพูด (Speech Recognition)

บทบาทของเทคโนโลยีการรู้จำเสียงพูดที่สำคัญในปัจจุบัน คือ เป็นตัวเชื่อมประสานกับผู้ใช้งาน (User Interface) ซึ่งอำนวยความสะดวกในการติดต่อระหว่างมนุษย์กับคอมพิวเตอร์ ขณะที่มือไม่ว่าง ต้องการความคล่องตัว สายตามไม่ว่าง ไม่ต้องการใช้คีย์บอร์ด ทัศนวิสัยไม่ดี มีข้อจำกัดด้านร่างกาย ฯลฯ

ทั้งนี้ ความท้าทายและทิศทางการพัฒนาเทคโนโลยีดังกล่าว มีดังต่อไปนี้

1) ความทนทาน (Robustness) เพื่อให้ได้ความถูกต้องของการรู้จำที่ไม่ลดลง เมื่อข้อมูลที่ส่งมาเกิดความผิดเพี้ยนหรือหายไปเนื่องจากสิ่งรบกวนต่างๆ

2) การเรียนรู้และปรับปรุงตัวเองโดยอัตโนมัติ (Automatic Training and Adaptation) เพื่อให้ระบบสามารถเรียนรู้และปรับปรุงตัวเองให้เข้ากับการทำงานในลักษณะต่างๆ ได้อย่างรวดเร็ว ประหยัด และเรียบง่าย

3) การรู้จำเสียงพูดที่เป็นธรรมชาติ (Spontaneous Speech) เพื่อให้ระบบสามารถรู้จำสำเนียงการพูด (Prosody) จังหวะการพูด อารมณ์ และพฤติกรรมการพูดรูปแบบแบบต่างๆ

4) การสนทนา (Dialogue Models) เพื่อให้ระบบสามารถเข้าใจบทสนทนาของผู้ใช้

5) การสร้างภาษาโต้ตอบ (Natural Language Response Generation) เพื่อให้ระบบสามารถสร้างภาษาโต้ตอบกับผู้ใช้ โดยภาษาที่สร้างขึ้นต้องสอดคล้องและเหมาะสมกับเรื่องที่กำลังสนทนา

6) การสังเคราะห์และสร้างเสียงพูด (Speech Synthesis and Generation) เพื่อให้ระบบสามารถสังเคราะห์เสียงพูด และสนทนาโต้ตอบกับผู้ใช้

7) ระบบหลายภาษา (Multilingual Systems) เพื่อการเข้าถึงข้อมูลข้ามภาษา และการแปลภาษาแบบทันกาลจากเสียงพูด

8) ระบบแบบผสมผสาน (Multimodal Systems) เป็นการนำข้อมูลด้านอื่นที่นอกเหนือจากข้อมูลทางภาษาและเสียงพูด เช่น สีหน้า ฝีปาก ท่าทาง และลายมือ เข้ามาใช้เพื่อเพิ่มความถูกต้องของการรู้จำและความเข้าใจในภาษา

2.2 เครื่องมือการแปลงเสียงพูดเป็นข้อความ

ปัจจุบันเทคโนโลยีการแปลงเสียงพูดเป็นข้อความ (speech to text) ได้มีการพัฒนาให้มีความถูกต้องแม่นยำเพิ่มมากขึ้น ระยะเวลาการตอบสนองลดน้อยลง อีกทั้งการใช้งานมีความสะดวกสามารถทำงานผ่านอุปกรณ์ต่างๆ ไม่ว่าจะเป็นเครื่องคอมพิวเตอร์ แท็บเล็ต หรือสมาร์ทโฟน ซึ่งมีผลิตภัณฑ์ให้เลือกใช้หลายหลาย ทั้งนี้ แต่ละผลิตภัณฑ์จะมีข้อดีและข้อจำกัดการใช้งานแตกต่างกัน ขึ้นอยู่กับระบบปฏิบัติการหรือแพลตฟอร์ม ภาษาที่รองรับ และการปรับแต่งผลิตภัณฑ์ให้เป็นไปตามความต้องการ

ทั้งนี้ ได้ดำเนินการศึกษาผลิตภัณฑ์ที่สำคัญ จำนวน 6 ผลิตภัณฑ์ ประกอบด้วย

1) Voice Typing ของ Google
2) Siri ของ Apple
3) Windows Speech Recognition ของ Microsoft
4) Dragon ของ Nuance Communications
5) Speech to Text API ของ IBM
6) พาที (PARTY) ของ NECTEC

2.3 Voice Type

Voice Typing เป็นฟังก์ชั่นการทำงานพิมพ์ข้อความด้วยเสียงพูดโดย ที่ Google เป็นผู้พัฒนาขึ้น โดยทำงานผ่านการเชื่อมต่ออินเทอร์เน็ตแบบ Real Time บนเว็บบราวเซอร์ Chrome และระบบปฏิบัติการ Android สามารถรองรับการใช้งานภาษาไทยและภาษาอื่นรวมมากกว่า 80 ภาษา ซึ่งไม่มีการคิดค่าใช้จ่าย

โดยในส่วนของการทำงานบนเว็บบราวเซอร์ Chrome จะสามารถใช้งานได้ในหลายลักษณะ คือ

1) การป้อนข้อมูลด้วยเสียงผ่านโปรแกรมที่จัดเตรียมไว้แล้ว ซึ่งจะปรากฏรูปสัญลักษณ์ไมโครโฟนในบริเวณที่ต้องการป้อนข้อมูล เช่น ในการค้นหาของ Google เป็นต้น

2) การพิมพ์ข้อความด้วยเสียงผ่าน Google เอกสาร (Docs) ซึ่งเป็นบริการหนึ่งบน Cloud Storage ที่เรียกว่า Google Drive ในการจัดทำไฟล์เอกสาร

3) การสร้าง Application ขึ้นเองตามความต้องการ โดยเชื่อมต่อผ่าน Speech API ใน Google Cloud Platform ตัวอย่างเช่น https://dictation.io

สำหรับการทำงานบนระบบปฏิบัติการ Android (รองรับ Android 4.1 Jelly bean ขึ้นไป) จะทำงานผ่าน Google Keyboard ซึ่งจะมีรูปสัญลักษณ์ไมโครโฟน ให้เรียกใช้งานสำหรับพิมพ์ข้อความลงในโปรแกรมต่างๆ แทนแป้นพิมพ์

2.4 Siri

Siri เป็นโปรแกรมที่พัฒนาโดย Apple เพื่อทำหน้าที่เป็นผู้ช่วยส่วนตัวอัจฉริยะ (intelligent personal assistant) และเป็นผู้ตอบคำถามความรู้ต่างๆ (knowledge navigator) ที่มาพร้อมกับระบบระบบปฏิบัติการ iOS ไม่ว่าจะเป็นเครื่องคอมพิวเตอร์ (Macintosh) แท็บเล็ต (iPad) สมาร์ทโฟน (iPhone) หรือบนอุปกรณ์อื่นๆ ได้แก่ Apple Watch Apple TV และ iPod

ทั้งนี้ Siri มีความสามารถในการพิมพ์ข้อความด้วยเสียงพูดด้วย ผ่านการเชื่อมต่อทางอินเทอร์เน็ตแบบ Real Time รองรับภาษาไทย และภาษาอื่นรวมมากกว่า 60 ภาษา ซึ่งสามารถเรียกใช้งานให้พิมพ์ข้อความลงในโปรแกรมที่ทำงานบนเครื่องคอมพิวเตอร์หรืออุปกรณ์นั้น

นอกจากนี้ ใน iOS 10 ทาง Apple เปิดให้นักพัฒนาสามารถพัฒนาโปรแกรมเชื่อมต่อผ่าน API ของ Siri ในชื่อ SirKit API โดยไม่มีการคิดค่าใช้จ่าย แต่ทั้งนี้จะใช้งานได้กับโปรแกรมบางประเภทเท่านั้น

2.5 Windows Speech Recognition

Windows Speech Recognition เป็นโปรแกรมที่พัฒนาโดย Microsoft สำหรับการรู้จำเสียงเพื่อสั่งการให้คอมพิวเตอร์ทำงานต่างๆ เช่น เปิดปิดโปรแกรม ส่งอีเมล ค้นหาและพิมพ์ข้อความลงในโปรแกรมที่ให้มาพร้อมกับระบบปฏิบัติการ Windows (Windows 7 ขึ้นไป) บนเครื่องคอมพิวเตอร์ ที่สำคัญคือสามารถทำงานร่วมกับชุดโปรแกรม Microsoft Office และทำงานแบบ Offline ได้ แต่ทั้งนี้ยังไม่รองรับการใช้งานภาษาไทย

ทั้งนี้ Microsoft ได้มีการพัฒนา Cortana เพื่อทำหน้าที่เป็นผู้ช่วยส่วนตัว (personal assistant) ที่มาพร้อมกับระบบปฏิบัติการ Windows Phone 8.1 ขึ้นไป โดยเฉพาะใน Windows 10 ซึ่งสามารถใช้งานบนทุกอุปกรณ์ไม่ว่าจะเป็นคอมพิวเตอร์ แท็บเล็ต หรือสมาร์ทโฟน โดยการใช้งานบางอย่างจะต้องเชื่อมต่อกับอินเทอร์เน็ต และคุณลักษณะบางอย่างสามารถใช้งานได้เพียงบางประเทศเท่านั้น และปัจจุบันยังไม่รองรับภาษาไทย

นอกจากนี้ ทาง Microsoft ได้เปิดให้นักพัฒนาสามารถพัฒนาโปรแกรมเชื่อมต่อผ่าน API ที่เรียกว่า Bing Speech API ใน Microsoft Cognitive Services โดยไม่มีการคิดค่าใช้จ่าย แต่ยังไม่รองรับภาษาไทย

2.6 Dragon

Dragon เป็นโปรแกรมที่พัฒนาโดย Nuance Communications ในการรู้จำเสียงพูดและใช้งานต่างๆ รองรับภาษาไทยและภาษาอื่นรวมมากกว่า 80 ภาษา แต่มีการคิดค่าใช้จ่ายในลักษณะเป็นการใช้งานส่วนบุคคล และทางธุรกิจ

โดยแบบการใช้งานส่วนบุคคลจะเป็นชุดโปรแกรมสำเร็จรูป สามารถติดตั้งบนเครื่องคอมพิวเตอร์ได้ทั้งระบบปฏิบัติการ Windows และ iOS ซึ่งมีค่าใช้จ่ายตามลักษณะการใช้งาน (ราคาประมาณ 3,000 ถึง 20,000 บาท) ทั้งนี้ ที่น่าสนใจคือมีซอฟต์แวร์ด้านกฎหมายโดยเฉพาะ ซึ่งฝึกสอนจากคำศัพท์ทางกฎหมาย 400 ล้านคำ จากเอกสารกฎหมาย แต่ยังไม่รองรับภาษาไทย

ส่วนแบบการใช้งานทางธุรกิจ จะเป็นลักษณะแยกตามประเภทธุรกิจ และมีค่าใช้จ่าย (ต้องติดต่อกับตัวแทนจัดจำหน่าย) ทั้งนี้ ที่น่าสนใจคือมีประเภทธุรกิจที่เป็นงานราชการ งานกฎหมาย และงานรักษาความปลอดภัยสาธารณะ (การตรวจจับ เฝ้าระวัง และตรวจพิสูจน์)

นอกจากนี้ ยังมีโปรแกรมที่ทำงานบนอุปกรณ์เคลื่อนที่บนระบบปฏิบัติการ iOS และ Android ให้ใช้งานโดยไม่เสียค่าใช้จ่าย โดยบางโปรแกรมรองรับแค่บางระบบปฏิบัติการ

2.7 Speech to Text

IBM ได้เปิดบริการการเชื่อมต่อผ่าน API กับ Speech to Text ภายใต้ IBM Watson Developer Cloud ซึ่งทำงานผ่านอินเทอร์เน็ตบนแพลตฟอร์มโมบายล์คลาวด์ที่เรียกว่า IBM Bluemix โดยเป็น Platform as a Service ที่สามารถรองรับธุรกิจขนาดเล็กจนถึงองค์กรขนาดใหญ่ ด้วยแนวคิด Open standards และ Cloud ลดความยุ่งยากในการพัฒนา และมีบริการเครื่องมืออย่างครบวงจร

บริการดังกล่าวามารถปรับแต่งการทำงานได้ตามที่ต้องการ และมีการคิดค่าใช้จ่ายในการบริการ ทั้งนี้ปัจจุบันยังไม่รองรับภาษาไทย

2.8 พาที (PARTY)

พาที (PARTY) เป็นระบบรู้จำเสียงพูดภาษาไทย ที่พัฒนาโดยศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ หรือ NECTEC ซึ่งเป็นหน่วยงานในสังกัดสำนักงานพัฒนาวิทยาศาสตร์และเทคโนโลยีแห่งชาติ (สวทช.) หรือ NSTDA กระทรวงวิทยาศาสตร์และเทคโนโลยี

ระบบรู้จำเสียงพูดภาษาไทย “พาที (PARTY)” เวอร์ชัน 1.0 ทำงานบนอุปกรณ์แท็บเล็ตหรือสมาร์ทโฟนที่มีระบบปฏิบัติการ iOS มีคุณสมบัติสำคัญดังนี้

- ไม่จำกัดเนื้อหา (Open domain) – พาที ได้ใช้วิทยาการใหม่ที่สร้างขึ้น โดยมีพจนานุกรมในระบบขนาดเพียง 40,000 คำ บรรจุหน่วยคำผสมกับหน่วยพยางค์ที่ใช้บ่อยในภาษาไทย หน่วยพยางค์สามารถผสมเป็นคำใหม่ๆ ที่ระบบไม่รู้จัก ส่งผลให้ครอบคลุมคำศัพท์ได้ใกล้เคียงกับระบบที่มีพจนานุกรมขนาดเกิน 140,000 คำ ทั้งยังช่วยลดปริมาณทรัพยากรที่ต้องใช้ในการคำนวณอย่างมากด้วยเช่นกัน

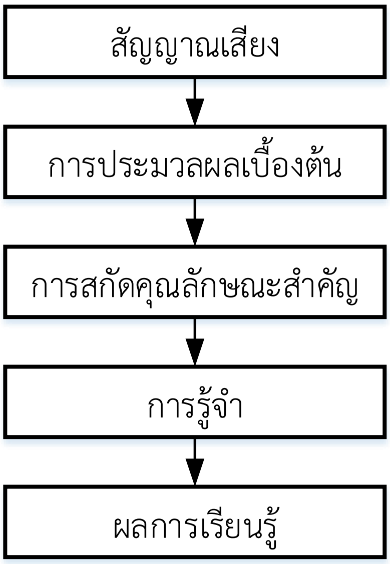
- ความแม่นยำ 80% (Accuracy) – ภายใต้การทดสอบกับเสียงพูดผ่านช่องทางข้อมูล (data channel) ด้วยสมาร์ทโฟน โดยไม่กำหนดเนื้อหา ผู้พูด หรือรูปแบบการพูด ระบบให้ความถูกต้องของการรู้จำเกือบ 80% ซึ่งใกล้เคียงกับบริการจากต่างประเทศ (ทดสอบเมื่อพฤษภาคม 2557)
- ตอบสนองภายใน 1.5xRT (Response time) – การทดสอบความเร็วในการตอบสนองภายใต้เครือข่าย WiFi และเครือข่าย 3G จำลอง พบว่าพาทีสามารถตอบสนองได้ภายในเวลาไม่เกิน 1.5 เท่าของความยาวของเสียงอินพุต ซึ่งใกล้เคียงกับการตอบสนองของบริการรู้จำเสียงพูดภาษาไทยจากต่างประเทศ
- เว็บบริการพร้อมติดตั้งและปรับแต่งได้ (Customizable) – จุดเด่นสำคัญของ พาทีคือความเชี่ยวชาญของทีมวิจัยและพัฒนาในการปรับแต่งระบบตามความต้องการ ปัจจุบันพาทีอยู่ในรูปแบบเว็บบริการพร้อมติดตั้งในหน่วยงานที่ต้องการใช้งาน สามารถขยายบริการตามปริมาณการใช้งาน และปรับแต่งระบบให้ถอดความได้ถูกต้องมากขึ้นในเนื้อหาหรือผู้พูดที่กำหนดได้

ทั้งนี้ NECTEC มีแผนที่จะนำระบบดังกล่าวมาประยุกต์ใช้ในด้านต่างๆ ได้แก่

- การสื่อสารโทรคมนาคม (Telecommunication) โดยการนำระบบรู้จำเสียงพูดไปใช้แปลงเสียงพูดลูกค้าที่ติดต่อเข้าที่ศูนย์บริการลูกค้าหรือ Contact center ผ่านทางโทรศัพท์เพื่อวิเคราะห์ความต้องการของลูกค้า
- การอำนวยความสะกวดในการป้อนข้อมูล (Voice data input) โดยการประยุกต์ใช้ระบบรู้จำเสียงพูดช่วยในศูนย์บริการถ่ายทอดการสื่อสารสำหรับผู้บกพร่องทางการได้ยิน (Thailand Telecommunication Relay Service หรือ TTRS) นอกจากนี้ยังสามารถประยุกต์ใช้ในนวัตกรรมที่สร้างโอกาสทางธุรกิจได้อีกมาก เช่น ความต้องการป้อนข้อมูลอย่างรวดเร็วผ่านอุปกรณ์พกพาหรือสมาร์ทโฟนภายใต้สถานการณ์ฉุกเฉิน การป้อนข้อมูลเพื่อบริหารคลังสินค้าขนาดใหญ่ ตลอดจนการป้อนข้อมูลด้วยเสียงเพื่อสั่งการอุปกรณ์ต่างๆ ภายในอาคารบ้านเรือน โดยเล็งเห็นถึงช่องทางในอนาคตที่นวัตกรรมนี้จะมาทดแทนการใช้รีโมทคอนโทรลที่ใช้อยู่ในปัจจุบัน
- การถอดความข้อมูลเสียง (Audio transcription) จากความต้องการถอดความเพื่อนำข้อมูลข่าวสารไปใช้ประโยชน์ต่อก็ต้องมีการลงทุนเพิ่มขึ้นอย่างมากตามไปด้วย ตัวอย่างเช่น ความต้องการในการจัดทำรายงานการประชุมรัฐสภาของสำนักงานเลขาธิการรัฐสภาและวุฒิสภา ที่ต้องสำเร็จออกเป็นร่างรายงานภายในเวลาอันสั้น ความต้องการจัดทำรายงานการไต่สวนคดีในศาลยุติธรรม เพื่อให้ผู้ที่เกี่ยวข้องลงนามรับรองหลังจบการไต่สวน ความต้องการถอดความเสียงบันทึกเทปต่างๆ เพื่อการสืบค้นย้อนหลังได้ง่ายขึ้น เป็นต้น ความต้องการเหล่านี้ล้วนมีความเป็นไปได้ในการประยุกต์ใช้เทคโนโลยีการรู้จำเสียงพูด โดยในปัจจุบันเนคเทคได้รับการติดต่อจากหน่วยงานที่ให้ความสนใจดังกล่าว และอยู่ระหว่างศึกษาความเป็นไปได้ในการประยุกต์ใช้ให้มีประสิทธิภาพ

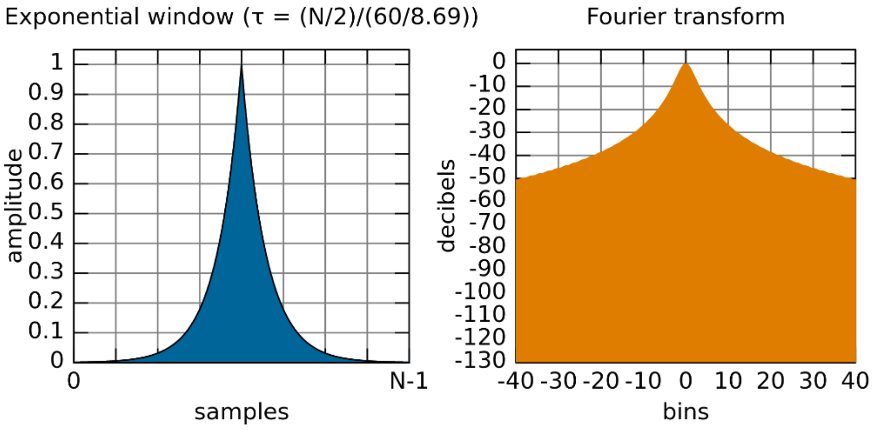
2.9 การพัฒนาระบบรู้จำเสียงพูด

หลักการรู้จำเสียงพูดภาษาไทยในปัจจุบันได้มีงานวิจัยจำนวนมาก ที่มีการวิจัยเกี่ยวกับการรู้จำเสียงพูดภาษาไทย ไม่ว่าจะเป็นเสียงพูดภาษาไทยกลาง หรือแม้กระทั่งเสียงพูดภาษาถิ่น โดยเน้นหลักการสร้างระบบรู้จำเสียงพูดซึ่งได้แสดงไว้ในรูปที่ 2.2 ซึ่งการทำงานจะประกอบด้วยการประมวลผลเบื้องต้น (Preprocessing) การสกัดค่าลักษณะสำคัญ (Feature extraction) และการรู้จำ (Recognition)



ภาพที่ 2.2 หลักการโดยทั่วไปของระบบระบุคำพูด

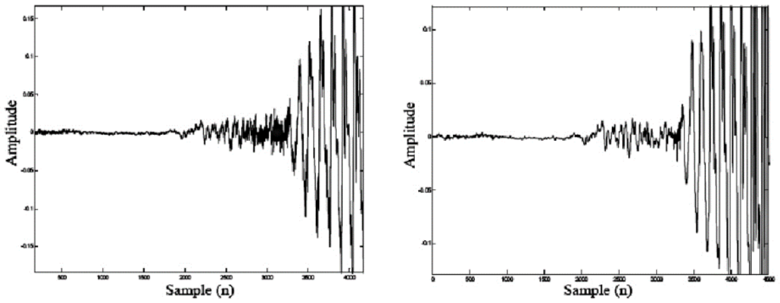
การรู้จำเสียงพูดจะเริ่มต้นด้วยขั้นตอนของการประมวลสัญญาณเบื้องต้นซึ่งเป็นขั้นตอนในการจัดเตรียมเสียงพูด โดยเริ่มตั้งแต่การแปลงสัญญาณเสียงพูดที่ได้มาจากการบันทึกเสียงมาเป็นสัญญาณเชิงเลขซึ่งได้เป็นข้อมูลเสียงพูด เพื่อที่จะนำไปใช้ในการประมวลผลในขั้นตอนต่อไปได้ เนื่องจากสัญญาณเสียงพูดเป็นค่าทางสถิติที่เปลี่ยนแปลงตามเวลา ทำให้ไม่สามารถจำลองสัญญาณเสียงพูดเป็นค่าทางสถิติได้ ด้วยเหตุนี้ในการประยุกต์ใช้งานเสียงพูดกับกรรมวิธีสัญญาณดิจิตอลจึงจำเป็นต้องแบ่งสัญญาณเสียงพูดออกเป็นส่วนย่อยสั้นๆ เรียกว่า กรอบเสียงพูดโดยในแต่ละส่วนย่อยนั้นจะมีความยาวประมาณ 10-40 มิลลิวินาที ทำให้กรอบเสียงพูดแต่ละกรอบมีค่าทางสถิติเปลี่ยนแปลงตามเวลาน้อยมาก จนถือได้ว่าในแต่ละกรอบเสียงพูดมีค่าทางสถิติไม่เปลี่ยนแปลงตามเวลา จึงสามารถประมวลผลโดยใช้ค่าทางสถิติกับสัญญาณเสียงพูดในแต่ละกรอบได้ ซึ่งขั้นตอนของการวางกรอบสัญญาณนี้ คือการแบ่งสัญญาณเสียงพูดออกเป็นส่วนย่อย ๆ เพื่อใช้ในการหาค่าสหสัมพันธ์ ซึ่งมีหลายชนิดด้วยกัน เช่น หน้าต่างแฮมมิ่ง หน้าต่างแฮนนิ่ง เป็นต้น ซึ่งโดยในงานวิจัยทั่วไป ส่วนใหญ่จะเลือกใช้ฟังก์ชันกรอบแบบหน้าต่างแฮมมิ่ง ดังแสดงในรูปที่ 2.3



ภาพที่ 2.3 ฟังก์ชันกรอบชนิด Hamming Windows

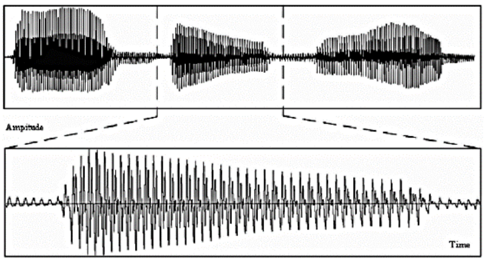
การวางกรอบขนาดสัญญาณ สามารถทำได้ด้วยการลดทอนแอมปลิจูดอย่างช้าๆ ที่บริเวณปลายแต่ละข้างของกรอบข้อมูลเสียงพูดเพื่อป้องกันการเปลี่ยนแปลงที่ไม่ต่อเนื่องอย่างกระทันหันที่ส่วนปลายกรอบสัญญาณและเพื่อเป็นการสร้างค่าการประสานสำหรับผลจากการแปลงฟูริเยร์ของฟังก์ชันกรอบและแถบสเปตรัมของเสียงพูด โดยสัญญาณเสียงที่ผ่านการแปลงสัญญาณเป็นดิจิตอลแล้วผ่านการประมวลผลเบื้องต้น สามารถสรุปเป็นขั้นตอนหลัก ๆ เพื่อให้เกิดความเข้าใจได้ง่ายขึ้นดังนี้

1) การกรองทางความถี่ (Filtering) เป็นขั้นตอนในการกรองสัญญาณในช่วงความถี่ที่ไม่ต้องการออกโดยอาศัยตัวกรองแบบดิจิตอล ดังแสดงในรูปที่ 2.4



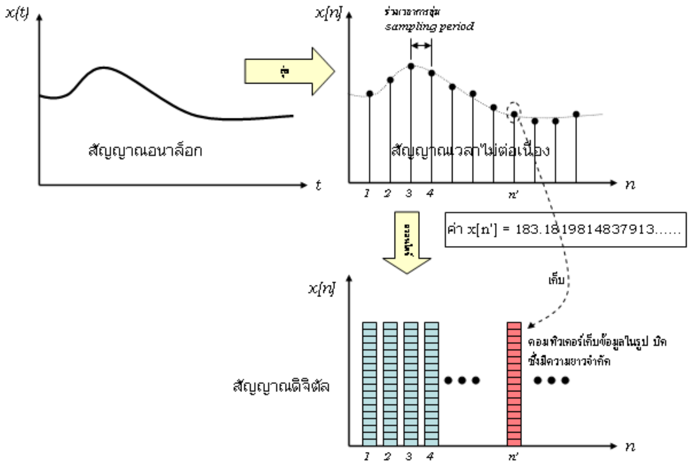
ภาพที่ 2.4 การกรองความถี่สัญญาณเสียงพูดแบบ low-pass filter

การตัดหัว-ท้ายเสียง (Endpoint detection) เป็นขั้นตอนในการกำหนดจุดเริ่มต้นและจุดสิ้นสุดของเสียง โดยการแยกส่วนที่เป็นคำพูดออกจากส่วนที่ไม่ใช่คำพูด ดังแสดงในรูปที่ 2.5 วิธีในการตัดหัว-ท้ายเสียงมีหลายวิธี เช่น ใช้ค่าระดับพลังงาน (Energy level) ใช้อัตราการตัดศูนย์ (Zero-crossing rate) เป็นต้น

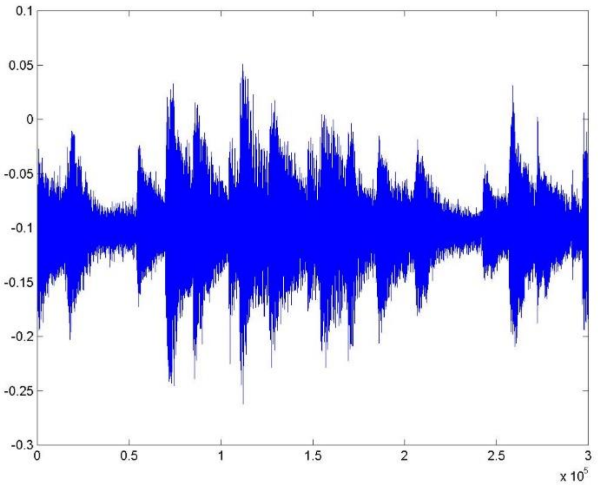


ภาพที่ 2.5 การตัดหัว-ท้ายเสียง ของสัญญาณเสียงพูดที่ต่อเนื่อง

2) การนอร์มอลไลซ์ทางเวลา (Time normalization) เป็นขั้นตอนการเพิ่มหรือลดขนาดความยาวของสัญญาณในเชิงเวลา เพื่อปรับแต่งขนาดความยาวของสัญญาณให้เหมาะสมตามต้องการ ทั้งนี้จะขึ้นอยู่กับกระบวนการในการรู้จำเสียงว่าจำเป็นต้องนอร์มอลไลซ์สัญญาณให้เท่ากันหรือไม่ วิธีการนอร์มอลไลซ์ทางเวลามีหลายวิธี เช่น การเปลี่ยนอัตราการชักตัวอย่าง (Sampling rate changing) การประมาณค่าในช่วงเชิงเส้น (Linear interpolation) และการเหลื่อมและรวมส่วนย่อยแบบซิงโครไนซ์ (Synchronized overlap-and-add) เป็นต้น



ภาพที่ 2.6 การสุ่มสัญญาณ (signal sampling) และการควอนไทซ์ (quantization)



ภาพที่ 2.7 ตัวอย่างเสียงพูด

จากรูปที่ 2.6 และรูปที่ 2.7 เป็นตัวอย่างของการการสุ่มสัญญาณ (signal sampling) และการควอนไทซ์ (quantization) จากตัวอย่างเสียงพูด โดยที่โดเมนของ การแปลง (transformed domain) แบ่งออกเป็น

1) ความถี่ (frequency) หรือฟูริเยร์ (Fourier)
    a. การแปลงฟูริเยร์ (Fourier transform –FT)
    b. การแปลงฟูริเยร์ไม่ต่อเนื่อง (Discrete Fourier transform –DFT)
    c. อัลกอริทึมในการแปลงฟูริเยร์ไม่ต่อเนื่องอย่างเร็ว (Fast Fourier transform –FFT)
    d. การแปลงโคซายน์ไม่ต่อเนื่อง (Discrete Cosine Transform-DCT)
2) เวลาและความถี่ (time-frequency)
    a. การแปลงฟูริเยร์ในเวลาช่วงสั้น (Short time Fourier transform – STFT)
3) เวลาและสเกล (time-scale) หรือเวฟเลต (wavelets)
    a. การแปลงเวฟเลต (Wavelet transform)

ขั้นตอนที่สำคัญอีกขั้นตอนหนึ่ง คือการสกัดค่าลักษณะสำคัญเป็นการวิเคราะห์สัญญาณเสียงพูดเพื่อหาค่าที่เหมาะสม สำหรับใช้เป็นตัวแทนข้อมูลเสียงพูดซึ่งเก็บรวบรวมลักษณะสำคัญของเสียงพูดแต่ละเสียง โดยจะนำค่าเหล่านี้ไปฝึกฝนระบบให้รับรู้ถึงความแตกต่างของเสียงพูดแต่ละเสียงและใช้ในการเปรียบเทียบเพื่อแบ่งแยกความแตกต่างของเสียงพูดแต่ละเสียงออกจากกัน ซึ่งการวิเคราะห์หาค่าที่ใช้แทนสัญญาณเสียง ที่นำไปใช้ในขั้นตอนการรู้จำ แบ่งได้เป็น 3 กลุ่มหลัก กลุ่มแรกเป็นค่าลักษณะสำคัญระดับสูง (High level feature) ได้แก่ สำเนียงการพูด รูปแบบในการพูด และความเร็วในการพูด เป็นต้น กลุ่มที่สอง จะใช้ค่าลักษณะสำคัญทางฉันทลักษณ์ (Prosodic feature) เช่น ค่าความถี่มูลฐาน (Fundamental frequency) ความถี่ฟอร์แมนท์ (Formant frequency) และระดับพลังงาน (Energy profile) เป็นต้น ถึงแม้ว่าค่าลักษณะสำคัญแบบนี้จะมีประสิทธิภาพสูงในการรู้จำ แต่ยากในการสกัดจากสัญญาณ กลุ่มสุดท้ายเรียกว่าค่าลักษณะสำคัญแบบเอนเวโลปของสเปกตรัม(Spectral envelop feature) เป็นกลุ่มที่นิยมใช้กันมาก เนื่องจากค่าลักษณะสำคัญส่วนใหญ่สำหรับการรู้จำเสียงจะรวมอยู่ในข้อมูลเชิงสเปกตรัมนี้ อีกทั้งยังง่ายและสะดวกในการคำนวณหาค่าด้วย ตัวอย่างค่าลักษณะสำคัญแบบนี้ได้แก่

สัมประสิทธิ์การประมาณพันธะเชิงเส้น (Linear prediction coefficients: LPC) ซึ่ง เป็นเทคนิคที่นิยมนำมาใช้ในการวิเคราะห์คุณสมบัติทางกายภาพของสัญญาณ เช่น Spectral magnitude ของสัญญาณ เทคนิคนี้พิจารณาได้รวดเร็ว แม่นยำและมีประสิทธิภาพซึ่งก็ใช้ได้ดีสำหรับสัญญาณเสียงพูดและนำมาประยุกต์ใช้กับระบบการรู้จำเสียงพูดได้เป็นอย่างดี แบบจำลองการประมาณพันธะเชิงเส้นเป็นแบบจำลองแบบใช้ค่าพารามิเตอร์ (Parametric Model) ซึ่งจำลองให้ค่าพารามิเตอร์ที่ได้นี้เกินจากช่องทางเดินเสียงเพื่อจำลองเป็นสัญญาณเสียงพูด

นอกจากนั้นยังมีวิธีอื่นๆอีก เช่น การหาสัมประสิทธิ์เซปสตรัม (Cepstral coefficient) การหาสัมประสิทธิ์เซปสตรัมบนสเกลเมล (Mel frequency cepstral coefficients: MFCC) เซปสตรัมแบบหักลบค่าเฉลี่ย (Cepstral mean substraction:CMS) และเซปสตรัมแบบผ่านตัวกรองภายหลัง (Post filtered cepstrum: PFL) เป็นต้น ทั้งยังมีการคำนวณค่าการเปลี่ยนแปลง (Derivative หรือ Delta) ของสัมประสิทธิ์เหล่านี้มาใช้เป็นค่าลักษณะสำคัญเพิ่มเติมได้ด้วย

ขั้นตอนสุดท้ายคือการรู้จำ (Recognition) สำหรับขั้นตอนของการรู้จำนี้จะประกอบด้วย 2 หน้าที่หลัก คือการนำเวกเตอร์ของค่าลักษณะสำคัญของสัญญาณเสียง ที่อยู่ในชุดอ้างอิงหรือชุดฝึกฝน มาทำการเรียนรู้ เมื่อเรียนรู้แล้วเวกเตอร์ของสัญญาณเสียงที่ต้องการทดสอบการรู้จำจะถูกนำเข้ามาเทียบเคียงเพื่อรู้จำ ขั้นตอนในการเรียนรู้นั้นขึ้นอยู่กับวิธีในการรู้จำของระบบนั้นๆ บางวิธีก็เพียงแค่เก็บข้อมูลชุดเรียนรู้ไว้เปรียบเทียบกับข้อมูลชุดทดสอบเท่านั้น เช่น วิธีการรู้จำแบบหาค่าระยะห่างยูคลิเดียน (Euclidean distance) วิธีไดนามิกไทม์วาร์ปปิง (Dynamic time warping: DTW) เป็นต้น ในขณะที่บางวิธี จะนำข้อมูลชุดเรียนรู้ไปแปลงเป็นค่าอ้างอิงที่ต้องการ เช่น โครงข่ายประสาทเทียม (Artificial neural networks: ANN) จะนำข้อมูลชุดเรียนรู้ไปผ่านโครงข่ายที่สร้างขึ้น เพื่อจดจำรูปแบบ และเก็บเป็นค่าน้ำหนัก (Weight) แทน วิธีควอนไตซ์แบบเวกเตอร์ (Vector quantization: VQ) ซึ่งจะแทนเวกเตอร์ทั้งหมด ของแต่ละสัญญาณเสียงอ้างอิงด้วยเวกเตอร์จำนวนไม่มาก หรือการใช้แบบจำลองฮิดเดนมาร์คอฟ (Hidden markov model: HMM) โดยนำข้อมูลชุดฝึกฝนไปผ่านแบบจำลองที่สร้างขึ้นเพื่อจดจำรูปแบบ และเก็บค่าทางสถิติและค่าความน่าจะเป็นของแต่ละสถานะไว้ เป็นต้น แต่ทั้งหมดจะมีพื้นฐานอยู่ที่การคำนวณระยะห่างของรูปแบบที่จะรู้จำ และนำค่าระยะห่างที่ได้ไปใช้รู้จำตามแต่ละวิธีนั้นๆ การเลือกใช้วิธีการรู้จำขึ้นอยู่กับข้อกำหนดของงาน เช่น วิธี DTW และ ANN เหมาะสมกับระบบแบบกำหนดคำพูดตายตัว ในขณะที่วิธี VQ และ HMM จะเหมาะสมกับระบบงานที่เป็นแบบไม่กำหนดคำพูดมากกว่า ซึ่งในหลักของความเป็นจริงแล้วในการรู้จำ เราไม่ได้นำเอาสัญญาณเสียงพูดจากผู้พูดไปเทียบโดยตรง แต่จะดึงเฉพาะค่าสำคัญของเสียงออกมา เราเรียกค่าสำคัญนี้ว่า Speech feature เพื่อนำไปเป็นตัวแทนของเสียงในการรู้จำได้เป็นอย่างดี ขั้นตอนการแปลงนี้ ซึ่งค่าที่ได้จะถูกเก็บเวกเตอร์ โดยที่เวกเตอร์หนึ่งเวกเตอร์จะแทนสัญญาณเสียงยาวประมาณ 20 มิลลิวินาที แต่ละเวกเตอร์ก็แทนสัญญาณเสียง ที่ค่อยๆ เลื่อนไปแบบคาบเกี่ยวกัน เช่นเลื่อนไปทีละ 10 มิลลิวินาที ดังรูปที่ 2.11 ดังนั้นหากมีเสียงที่ยาว 1 วินาทีเข้ามา ก็จะแทนด้วยเวกเตอร์จำนวน 100 อัน

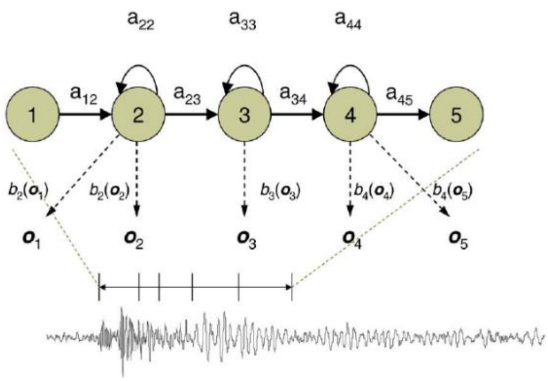
การเรียงลำดับ (Sequence) ของเวกเตอร์ที่แทนสัญญาณเสียงนี้เรียกว่าลำดับการสังเกตการณ์ (Observation sequence) นอกจากนั้น ในการรู้จำยังต้องเข้าใจคุณลักษณะของแต่ละคำที่เราต้องใช้ในการรู้จำคือ Phone ซึ่งหมายถึง โมเดลการออกเสียง (Pronunciation model) โมเดลเสียง (Acoustic model) และโมเดลภาษา (Language model)

1) Phone คือหน่วยย่อยสุดทางเสียง ตัวอย่างเช่น คำว่า “การ” อ่านออกเสียงด้วยเสียง “ก” ตามด้วยสระ “า” และลงท้ายด้วย เสียงตัวสะกด “น” คือ Phone ในทางภาษาศาสตร์ จะมีสัญลักษณ์มาตรฐานแทนเสียง Phone แต่ละเสียง ตัวอย่างเช่น “k” แทน เสียง “ก” “aa” แทนสระ “า”
2) โมเดลการออกเสียง (Pronunciation model) จะบอก Sequence ของ Phone เช่น “การ” ออกเสียงว่า “k aa n^” “ขนม” ออกเสียงว่า “kh a n o m^”
3) โมเดลเสียง (Acoustic model) ซึ่งโดยปกติเราจะมีโมเดลเสียง 1 โมเดล ต่อ 1 Phone เมื่อเราป้อน Observation sequence เข้าไปยังโมเดลเสียงใดๆ มันจะคำนวณค่าความน่าจะเป็นที่ Observation sequence นั้นจะเป็นเสียงของ Phone นั้นๆ ความน่าจะเป็นที่ว่านี่เขียนสั้นๆว่า P(O|p) โดยที่ p คือโมเดลเสียงของ Phone ใดๆ

โมเดลภาษา (Language model) คือตัวบอกให้ทราบว่า คำ (Word) นี้ ตามด้วยคำนี้ได้หรือไม่ หรือในบางโมเดลจะบอกค่าความน่าจะเป็นที่คำใดๆ จะพูดต่อกัน เช่น โมเดลภาษาอาจจะบอกว่า “จะ ไป” ได้ แต่ “ไป จะ” ไม่ได้ หรืออาจจะบอกเป็นค่าความน่าจะเป็นว่า “จะ ไป” มีโอกาสเกิดได้ 0.8 แต่ “ไป จะ” มีโอกาสเกิดได้แค่ 0.01 เป็นต้น โมเดลภาษาแท้จริงไม่เพียงบอกโอกาสที่คำสองคำจะเกิดคู่กันเท่านั้น ยังสามารถบอกด้วยว่า ทั้งประโยคมีโอกาสเกิด ได้เท่าไหร่ สมมุติว่าเรามีประโยคซึ่งประกอบด้วยคำต่อๆ กันหลายๆ คำ เช่น W = (w1…wM) โดยที่ w แทนคำแต่ละคำ โมเดลภาษาจะบอกว่า W สามารถเกิดได้หรือไม่ หรือบอกเป็น ค่าความน่าจะเป็นว่ามีโอกาสเกิดมากน้อยแค่ไหน ขอแทนค่าความน่าจะเป็นด้วย P(W) ขั้นตอนของการรู้จำมีขั้นตอนหลักๆ คือ ระบบรับ Observation sequence ที่ ต้องการรู้จำเข้ามา โดยมันจะเริ่มด้วยการเดาว่าเป็นคำใดต่อๆ กัน จะเป็นประโยคว่า “ฉัน รัก เธอ” ประโยคว่า “ฉัน หิว ข้าว” หรือ “อาหาร อร่อย ดี” ฯลฯ หลังจากเดาประโยคขึ้นมาแล้ว จะส่งประโยคนั้นเข้าไปยัง Language model ได้ค่าความน่าจะเป็น P(W) ที่จะเกิดประโยคดังกล่าว แล้วจะทำการแปลงประโยคเป็นเสียงอ่านโดยอาศัย Pronunciation model เมื่อได้ Sequence ของ Phone แล้ว ก็จะเอา Acoustic model ของแต่ละ Phone มาต่อกันแล้วทำการป้อน Observation sequence เข้าไปยัง Acoustic model ของทั้งประโยค จะหาค่าความน่าจะเป็น P(O|W) ซึ่งเกิดจาก P(O|p) ของแต่ละ Phone คูณกัน และสุดท้ายก็จะนำเอา P(W) มาคูณกับ P(O|W) ได้เป็น P(O,W) ซึ่งหมายถึง โอกาสที่สัญญาณเสียงดังกล่าวจะเป็นเสียงประโยค W แล้วก็ทำอย่างนี้กับทุกๆ ประโยคที่เดาขึ้นมา และเทียบค่า P(O,W) ว่าประโยค ไหนมีโอกาสสูงที่สุด ก็ตอบเป็นประโยคนั้น ในกรณีที่ว่าประโยคที่เป็นไปได้มีหลาก หลายล้านล้านแบบ และไม่กำหนดว่าประโยคยาวเท่าไหร่ วิธีการแก้ไขก็คือ การสร้าง Word network โดยเอาคำมาต่อๆ กันในลักษณะของ Network ระหว่างคำก็กำกับด้วยโอกาสที่แต่ละคำจะต่อกัน หรือ P(wi|wi-1) และในแต่ละคำก็ประกอบด้วย Acoustic model ของ Phone ที่ต่อกันเป็นเสียงอ่านของคำนั้นๆ แล้วเวลาทำงานก็จะผ่านสัญญาณเสียงเข้าไป ในขณะที่ผ่าน Node ของ Network แต่ละ Node ก็จะ มีการคูณค่าความน่าจะเป็น P(O,W) ต่อๆ ไปเรื่อยๆ หากในเส้นทางใดที่ค่าความน่าจะเป็นรวมขณะนั้น ตกต่ำกว่าค่า Threshold ที่กำหนด ก็ให้เลิก วิ่งไปเส้นทางนั้น เท่านี้ก็จะช่วยลดจำนวนประโยคที่จะต้องคำนวณลงได้มาก วิธีนี้จะเรียกว่า Beam search หรือการ Search ภายใน Beam ที่กำหนดเท่านั้น ซึ่งนอกจากนั้นยังมีอีก วิธีในการกำหนด Beam ของการ Search โดยกำหนดให้ ณ ขณะใดๆ

จะมีเส้นทางที่วิ่ง ไปได้ ไม่เกิน N เส้นทาง วิธีนี้ก็ช่วยลดจำนวนประโยคที่ต้องคำนวณลงมหาศาลเช่นกัน เราเรียกวิธีที่สองนี้ว่า N-best search

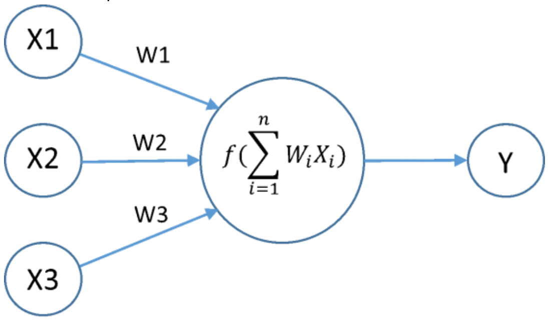
วิธีที่ได้รับการนิยมอีกวิธีคือ แบบจำลองฮิดเดนมาร์คอฟ (HMM) แบบต่อเนื่อง ซึ่งงานวิจัยส่วนมากนิยมใช้เป็นแบบจำลองในการจดจำเสียงพูด โดยทำการสร้างแบบจำลองฮิดเดนมาร์คอฟเป็นแบบหน่วยพื้นฐานของเสียง (phoneme) และแต่ละแบบจำลองจะใช้แบบจำลองฮิดเดนมาร์คอฟแบบ 5 สถานะ (state) มีการเปลี่ยนสถานะแบบซ้ายไปขวาและแต่ละสถานะเป็นแบบ 1 เกาส์เซียน ดังแสดงในรูปที่ 2.8



ภาพที่ 2.8 แผนภาพแสดง HMM แบบ 5 สถานะ ที่มีการเปลี่ยนแปลงสถานะแบบซ้ายไปขวา

ส่วน K-nearest neighbor (K-NN) เป็นเทคนิคที่เหมาะกับปัญหาแบบการแบ่งกลุ่มข้อมูล (classification) เทคนิคนี้แตกต่างจากเทคนิคอื่นตรงที่มันไม่ได้ใช้ข้อมูลฝึกหัด (training data) ในการสร้างแบบจำลอง แต่จะใช้ข้อมูลนั้นมาเป็นตัวแบบจำลองเลย ในการใช้งานขั้นตอนวิธีการค้นหาเพื่อนบ้านใกล้สุด (K-NN algorithm) นั้นเราต้องระบุค่าตัวเลข จำนวนเต็มบวกให้กับ k ด้วย ซึ่งค่านี้จะเป็นตัวบอกจำนวนของกรณี (case) ที่จะต้องค้นหาในการทำนายกรณีใหม่ algorithm แบบ K-NN ได้แก่ 1-NN, 2-NN, 3-NN, ... K-NN โดยที่ k แทนเลขจำนวนเต็มบวก เช่น 4-NN หมายถึง algorithm นี้จะค้นหา 4 กรณีที่มีลักษณะ ใกล้เคียงกับกรณีใหม่ (4 nearest cases) ในการทำนายกรณีใหม่การเรียนรู้แบบเบย์อย่างง่าย (Naïve-Bayes) เป็นเทคนิคที่ถูกตั้งชื่อตามโทมัส เบส์ (Thomas Bayes) เทคนิคแบบการเรียนรู้แบบเบย์อย่างง่าย (Naïve-Bayes) ใช้ทฤษฎี Bayes Theorem ในการคำนวณความน่าจะเป็นซึ่งถูกใช้ในการทำนายผล เมื่อทำการวิเคราะห์กรณีใหม่ การทำนายผลทำได้โดยการรวมผลของตัวแปรอิสระ (independent variable) ที่มีต่อตัวแปรตาม (dependent variable) การเรียนรู้แบบเบย์อย่างง่าย (Naïve-Bayes) เป็นเทคนิคในการแก้ปัญหาแบบ classification ที่ทั้งสามารถคาดการณ์ผลลัพธ์ได้และสามารถอธิบายได้ด้วย มันจะทำการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรอิสระแต่ละตัวกับตัวแปรตามเพื่อใช้ในการสร้างเงื่อนไขความน่าจะเป็นสำหรับแต่ละความสัมพันธ์ ในทางทฤษฎีแล้วการทำนายผลของ Naïve-Bayes จะถูกต้องถ้าตัวแปรอิสระทั้งหมดเป็นอิสระต่อกัน ไม่ขึ้นกับตัวแปรอิสระตัวใดตัวหนึ่ง ซึ่งในความเป็นจริงแล้วมีกรณีไม่มากนักที่ตัวแปรอิสระทั้งหมดเป็นอิสระต่อกัน เทคนิคการเรียนรู้แบบเบย์อย่างง่าย (Naïve-Bayes) ยังไม่รองรับข้อมูลที่เป็นข้อมูลต่อเนื่อง (continuous data) ด้วย ดังนั้น ตัวแปรอิสระหรือตัวแปรตามที่มีค่าเป็นค่าต่อเนื่องจะต้องถูกแบ่งเป็นช่วง ซึ่งการแบ่งช่วงนั้น ถ้าแบ่งไม่เหมาะสม ก็จะมีผลต่อคุณภาพของแบบจำลองที่สร้างขึ้น แต่ถ้าไม่คำนึงถึงข้อจำกัดนี้แล้ว เทคนิคการ

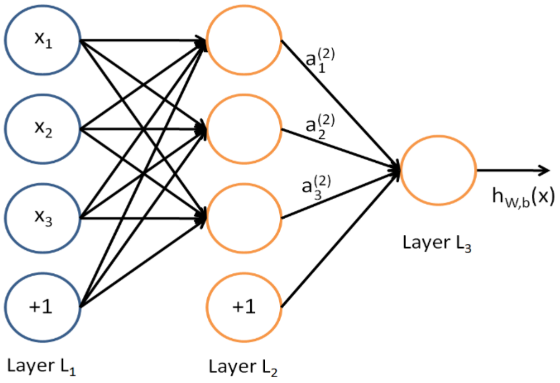
เรียนรู้แบบเบย์อย่างง่าย (Naïve-Bayes) สามารถให้ผลลัพธ์ที่ดีและรวดเร็วได้ ความง่ายและความเร็วทำให้เทคนิคนี้เป็นเครื่องมือที่ดีในการสร้างแบบจำลองและหารูปแบบความสัมพันธ์ที่ไม่ซับซ้อน โครงข่ายประสาทเทียม (Neural networks) มีพื้นฐานมาจากแบบจำลองการทำงานของสมองมนุษย์ และก็สามารถใช้ได้ดีกับปัญหาการแบ่งกลุ่มข้อมูล และการวิเคราะห์การถดถอย เป็นระบบที่จำลองการทำงานของสมองมนุษย์มาใช้ในการเรียนรู้ แยกแยะ ตัดสินใจในสิ่งต่าง โดยสถาปัตยกรรมพื้นฐานของโครงข่ายประสาทเทียม (Neural Network Architecture) จากการทำงานของเซลล์ประสาทได้ถูกนำประยุกต์เป็นโครงสร้างทางคณิตศาสตร์ ดังแสดงดังรูปที่ 2.9 ซึ่งสามารถแยกโครงสร้างของโครงข่ายประสาทเทียมออกเป็นหลักๆได้ 2 โครงสร้างดังนี้



ภาพที่ 2.9 แบบจำลองทางคณิตศาสตร์ของสมองมนุษย์

1) โครงข่ายประสาทเทียมแบบชั้นเดียว (Single layer artificial neural network)
2) โครงข่ายประสาทเทียมแบบหลายชั้น (Multilayer artificial neural network)

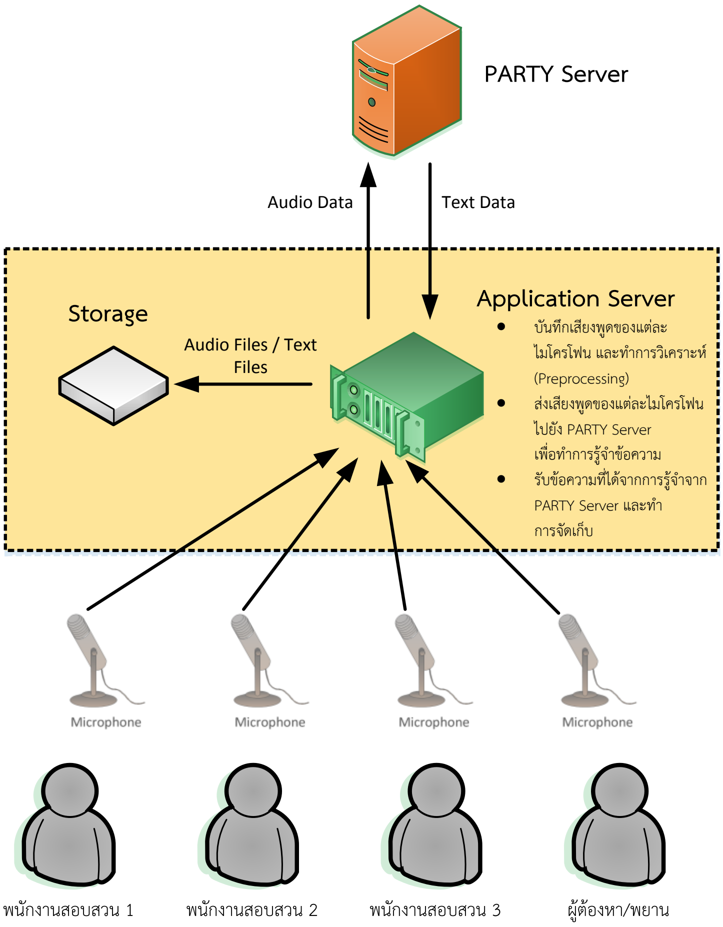
โดยในโครงสร้างต่างๆจะประกอบด้วยเลเยอร์ดังนี้ Input Layer, Hidden Layer, Output Layer ซึ่งเราจะนำตัวอย่างของการอาศัยอัลกอริทึมในการสอนให้กับโครงข่ายประสาทเทียมใช้เทคนิคแบ็กพรอบพาเกชั่น โดยเทคนิคนี้เป็นการใช้โครงสร้างแบบหลายชั้นในการเรียนรู้แบบมีผู้สอน (Supervised Learning) มีการกำหนดค่าเป้าหมาย (Target) ที่ต้องการไว้ และใช้โครงข่ายประสาทเทียม ในการปรับค่าน้ำหนัก(Weight) ให้เหมาะ ดังแสดงรูปที่ 2.10



ภาพที่ 2.10 โครงข่ายประสาทเทียมแบบหลายชั้น ที่ใช้การสอนแบบอัลกอริทึมการแพร่ย้อนกลับ (Backpropagation Neural Network)

ดังนั้นโครงการนี้เป็นอีกแนวทางหนึ่งในการอำนวยความสะดวกสำหรับเจ้าหน้าที่สอบสวนของกรมสอบสวนคดีพิเศษ เพื่อลดเวลาในการบันทึกข้อมูลการสอบสวนจากเดิมที่ใช้การพิมพ์สัมผัสตามเสียงพูดด้วยพนักงานสอบสวน โดยมีการปรับเปลี่ยนมาเป็นระบบพิมพ์ตัวหนังสืออัตโนมัติตามเสียงพูด ระบบสามารถรองรับผู้ใช้ได้มากกว่า 1 คนในเวลาเดียวกันและเป็นแบบ Real Time และยังสามารถจัดเก็บลงในระบบฐานข้อมูลโดยอัตโนมัติเพื่อใช้ประโยชน์ในการสืบค้นภายหลังได้ โดยขั้นตอนการทำงานของระบบตามรูปที่ 2.11

เริ่มจากในขั้นตอนของการสอบสวนพนักงานสอบสวน ผู้ต้องหาหรือพยาน ทำการสอบสวนหรือสนทนาผ่านไมโครโฟน โดยจะมีไมโครโฟนประจำตัวของแต่ละคน จากนั้นขอมูลเสียงของการสอบสวนหรือสนทนาทำการประมวลผล(Preprocessing) เพื่อตรวจสอบคุณลักษณะของเสียงที่ไม่พึงประสงค์ออกไป เช่นเสียงรบกวน เป็นต้น จากนั้นระบบทำการส่งข้อมูลเสียงแยกตามไมโครโฟนของแต่ละคนไปยังโปรแกรมพาที(PARTY) ที่พัฒนาโดยศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ (NECTEC) เพิ่มทำการรู้จำให้ออกมาเป็นข้อความตัวหนังสืออัตโนมัติตามเสียงพูด จากนั้นระบบทำการจัดเก็บข้อมูลเสียงและข้อความจากการสอบสวนหรือสนทนา เพื่อใช้ประโยชน์ในการสืบค้นได้ในอนาคต



ภาพที่ 2.11 ภาพรวมขั้นตอนการใช้งาน

บทที่ 3
การวิเคราะห์และออกแบบระบบ

3.1 ศึกษา วิเคราะห์สภาพปัญหา และความต้องการของผู้ใช้งาน

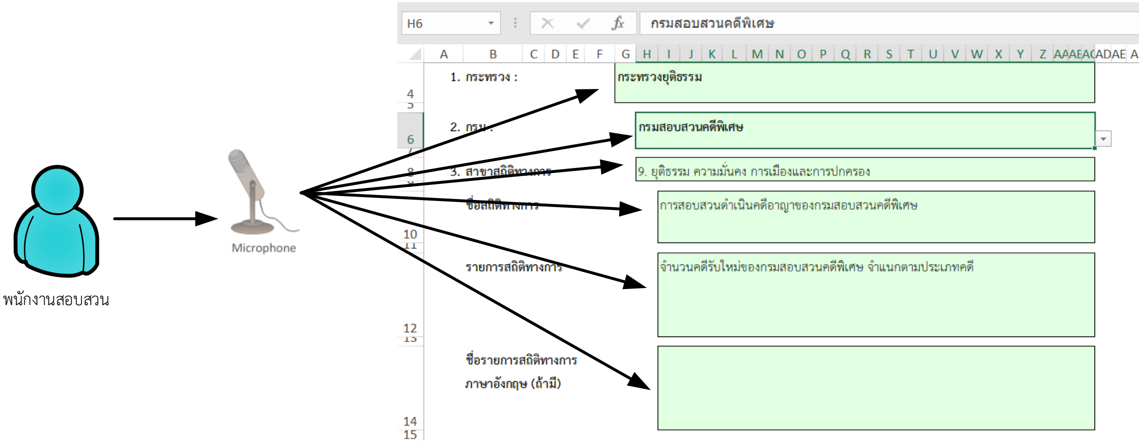
จากเก็บข้อมูลผู้ใช้งาน (พนักงานสอบสวนจากกองคดีต่างๆ) ด้วยแบบสอบถามปลายเปิด (Open-ended questionnaire) ซึ่งเป็นการรวบรวมข้อมูลการทำงานในปัจจุบัน ปัญหา อุปสรรค และความต้องการต่อการพัฒนาระบบรู้จำเสียงพูดของผู้ใช้งาน โดยมีรูปแบบคำถามดังนี้

1) ลักษณะการทำงานในปัจจุบัน
2) ปัญหาและอุปสรรคในการสอบสวนผู้ต้องหาและผู้เสียหาย
3) ข้อจำกัดและข้อควรระวังในการสอบสวนผู้ต้องหาและผู้เสียหาย
4) ข้อแตกต่างของปัญหาอุปสรรค์และข้อควรระวัง ในการสอบสวนผู้ต้องหาและผู้เสียหายของแต่ละหน่วยงาน
5) ความต้องการในการแก้ปัญหางานด้านการสอบสวน ในส่วนของการพัฒนาระบบรู้จำเสียงพูด
6) มีข้อเสนอแนะพิเศษในส่วนของการพัฒนาระบบรู้จำเสียงพูด

ซึ่งจากการรวบรวมข้อมูลดังกล่าวทำให้ผู้วิจัยสามารถสรุปปัญหา และอุปสรรคของกลุ่มผู้ใช้งานต่อการพัฒนาระบบรู้จำสียงพูด ได้ดังนี้

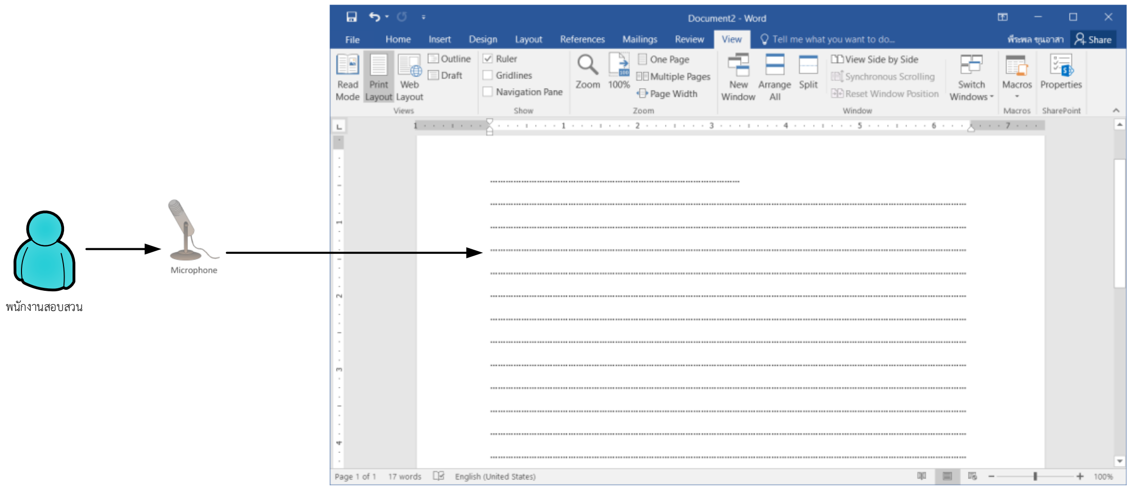
1) การออกเสียงของผู้ใช้มีความชัดเจนมากแค่ไหน ถ้าผู้ใช้งานออกเสียงไม่ชัดเจนสามารถทำให้ระบบรู้จำเสียงพูดอัตโนมัติ (Automatic Speech Recognition) แปลเป็นข้อความตัวอักษรที่ผิดพลาดได้
2) เสียงรบกวนจากสิ่งแวดล้อมรอบข้าง ถือเป็นส่วนที่สำคัญเนื่องมาจากเสียงรบกวนนั้นจะถูกบันทึกรวมไปด้วยกับเสียงพูดสั่งการหรือเสียงสัมภาษณ์ ซึ่งเมื่อนำไปประมวลผลแล้วสามารถทำให้ระบบรู้จำเสียงพูดอัตโนมัติ (Automatic Speech Recognition) แปลเป็นข้อความตัวอักษรที่ผิดพลาดได้
3) คำศัพย์ที่โปรแกรมรู้จำเสียงพูดอัตโนมัติ (Automatic Speech Recognition) ได้เรียนรู้ไปครอบคลุมกับการรูปประโยค ในการใช้งานหรือไม่
4) ระยะเวลาในการประมวลผลของโปรแกรมรู้จำเสียงพูดอัตโนมัติ (Automatic Speech Recognition) ซึ่งในบางครั้งมีผู้ใช้จำนวนมากเข้าใช้งานพร้อมกัน หรือเสียงพูดสั่งการเสียงสัมภาษณ์ มีความยาวในการบันทึกที่มาก อาจทำให้ระบบมีการตอบสนองได้ช้ากว่าที่ผู้ใช้ต้องการได้
5) สำหรับเนื้องานจริงในขั้นตอนการใช้งานนั้น ระบบควรมีกระบวนการของการทำให้แน่ใจในความเป็นส่วนตัว สามารถรักษาความลับได้ ความปลอดภัยในการรับส่งข้อมูลเนื่องจากข้อมูลมีความสำคัญทางกฎหมาย อาจมีการใช้ใช้การเข้ารหัสในการรับส่งข้อมูล
6) การใช้งานระบบควรระวัง และควรตรวจสอบขั้นตอนของข้อกฎหมายก่อนแสมอ

นอกจากนั้น ยังสามารถสรุปความต้องการข้องผู้ใช้งานได้ ดังนี้



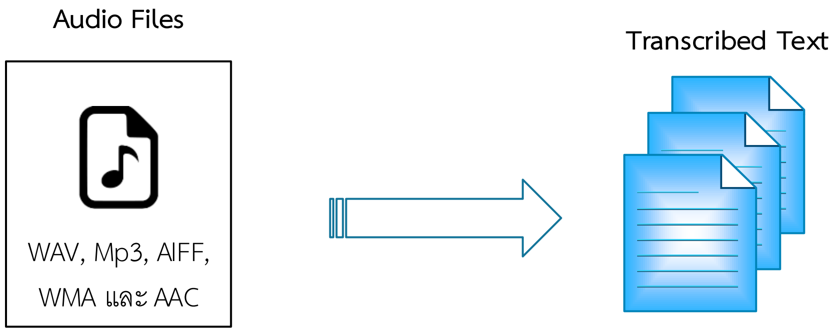
ภาพที่ 3.1 การใช้งานกับเอกสารต้นแบบ (template)

1) สามารถใช้งานกับเอกสารต้นแบบ (template) ที่มีตัวอย่างดังรูปที่ 3.1 (ตามเอกสารที่แสดงใน ภาคผนวก) โดยมีคุณสมบัติดังต่อไปนี้
   1.1) สามารถใช้งานในรูปแบบหนังสือแบบฟอร์มราชการของแต่ละหน่วยงานเช่น บันทึกถ้อยคำ เป็นต้น โดยจะเป็นการแปลงข้อความให้ตามที่ผู้ใช้ต้องการ จากนั้นข้อความผลลัพธ์ จะถูกเติมลงในแบบฟอร์มที่กำหนดไว้ ดังรูปที่ 3.1
   1.2) ระบบสามารถขึ้นคำถามหรือข้อความแบบอัตโนมัติ ที่ใช้งานบ่อยครั้งในแต่ละแบบฟอร์มต่างได้เพื่อลดเวลาและเพิ่มความถูกต้อง เช่น การแจ้งสิทธิ์ทางกฎหมาย หรือ คำถามเฉพาะตามระเบียบกฎหมาย เป็นต้น
   1.3) ข้อความอัตโนมัติที่ขึ้นนั้นไม่สามารถแก้ไขได้ มีการป้องกันการแก้ไข เพื่อเพิ่มความถูกต้อง ลดโอกาสผิดพลาดในการเขียนคำถามผิดทำให้เกิดข้อผิดพลาดในทางกฎหมายได้
   1.4) ระบบควรมีความสามารถในการแทรกรูปภาพได้ขณะใช้งาน
   1.5) เมื่อแปลงเสียงเป็นข้อความใส่ในรูปแบบหนังสือแบบฟอร์มราชการของแต่ละหน่วยงาน แล้ว สามารถแก้ไขได้อย่างสะดวด โดยผลลัพธ์อาจอยู่ในรูปแบบของไฟล์ที่สามารถแก้ไขได้สะดวดเช่น Microsoft Word
2) ใช้งานกับเอกสารที่ไม่มีต้นแบบ (template) โดยจะเป็นการแปลงข้อความให้ตามที่ผู้ใช้ต้องการดังรูปที่ 2 เช่น ใบลา บันทึกการประชุม บันทึกข้อความทางราชการ และเอกสารอื่นตามที่ผู้ใช้งาน เป็นต้น



ภาพที่ 3.2 ใช้งานกับเอกสารที่ไม่มีต้นแบบ (template)

3) สามารถนำไฟล์เสียง เข้าสู่ระบบและแปลงเป็นข้อความได้ และสนับสนุนไฟล์เสียงได้หลากหลายรูปแบบ เช่น WAV, Mp3, AIFF, WMA และ AAC เป็นต้น



ภาพที่ 3.3 สามารถนำไฟล์เสียงเข้าสู่ระบบและแปลงเป็นข้อความ

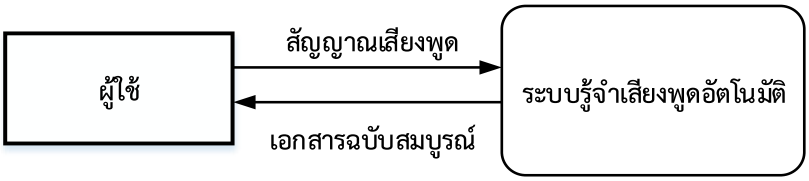
4) ระบบสามารถทำการบันทึกเสียงและสามารถจัดเก็บได้หลากหลายรูปแบบ เช่น WAV, Mp3, AIFF, WMA และ AAC เป็นต้น
5) ระบบควรรองรับไมโครโฟนที่มีหลายขนาดได้เพื่อความสะดวดในการใช้งาน

ภาพที่ 3.4 ไมโครโฟนลักษณะต่าง

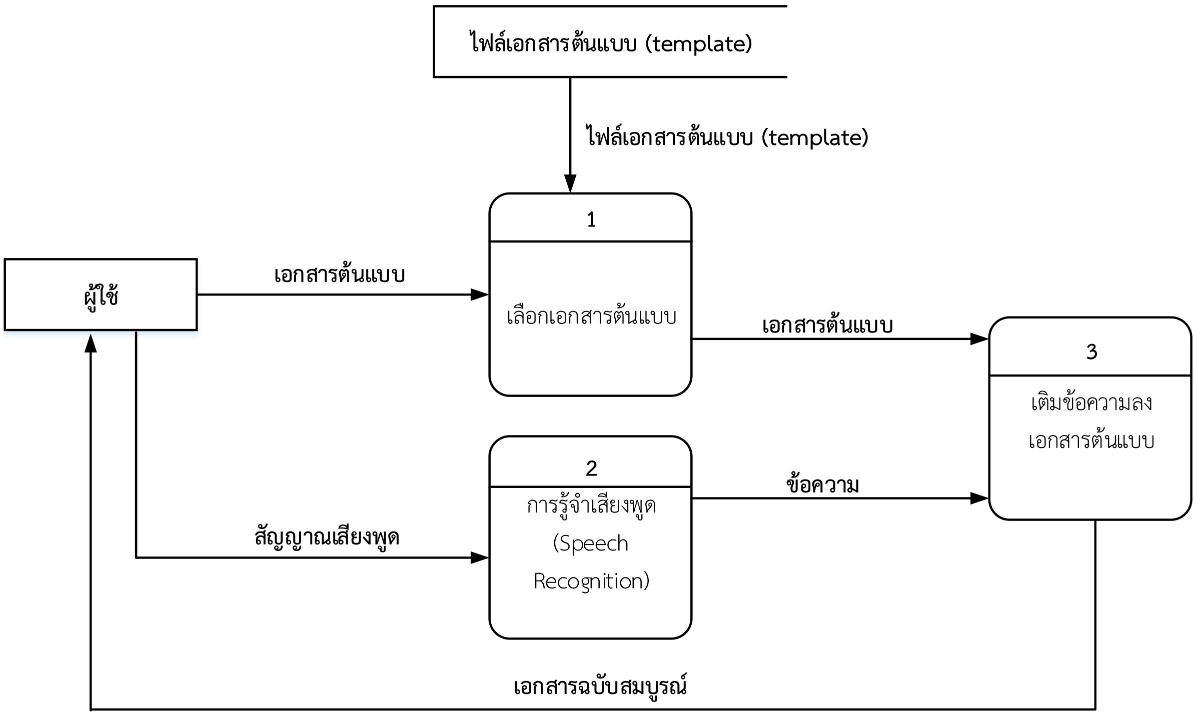
6) ระบบมีขนาดเล็กติดตั้งง่าย สามารถติดตั้งในแล็ปท็อปหรือโน๊ตบุ๊คได้
7) ระบบสามารถสามารถตอบสนองความต้องการของผู้ใช้ได้อย่างรวดเร็วและมีประสิทธิภาพตอบสนองได้อย่างรวดเร็ว
8) ใช้งานง่าย ส่วนการนำเข้าและส่วนแสดงผลลัพธ์ไม่มีความสลับซับซ้อน มีความชัดเจน
9) ไฟล์ที่ได้หลังจากการประมวลผลสามารถสั่งพิมพ์ได้ทันที
10) ไฟล์ที่ได้หลังจากการทำงานควรอยู่ในรูปแบบที่แก้ไขได้ง่าย เช่น Microsoft Word เป็นต้น
11) ระบบควรมีฟังก์ชันการใช้งานแบบ off-line
12) ระบบควรสนับสนุนการใช้งานได้หลายภาษา ภาษาถิ่น ภาษาพื้นบ้าน และภาษาอังกฤษ เป็นต้น

3.2 การวิเคราะห์และออกแบบระบบ

สำหรับขั้นตอนการวิเคราะห์ระบบนั้น ผู้พัฒนาได้ศึกษาเครื่องมือที่ใช้ในการออกแบบและพัฒนาเพื่อให้ตอบสนองกับความต้องการของผู้ใช้และสามารถใช้งานได้จริง โดยเลือกใช้เครื่องมือการวิเคราะห์ และการออกแบบเชิงโครงสร้าง (Structured) ซึ่งเป็นการอธิบายด้วยแผนภาพ (Diagram) ได้แก่ แผนภาพบริบท (Context Diagram) แผนภาพกระแสข้อมูล (Data Flow Diagram level 1) และ คำอธิบายการประมวลผลข้อมูล (Process Description) การจัดการเอกสารในระบบงานใหม่สามารถอธิบายขั้นตอนการปฏิบัติงานในลักษณะของ แผนภาพแสดงการไหลของข้อมูล (Data Flow Diagram) ดังนี้



ภาพที่ 3.5 แผนภาพบริบท (Context Diagram): ระบบรู้จำเสียงพูดอัตโนมัติเพื่องานสอบสวน



ภาพที่ 3.6 Data Flow Diagram Level 1 : ระบบรู้จำเสียงพูดอัตโนมัติเพื่องานสอบสวน

เริ่มจากในขั้นตอนของการสอบสวน พนักงานสอบสวนใช้คอมพิวเตอร์หรือ Notebook ที่มีไมโครโฟนพร้อมใช้งานทำการเลือกชนิดของเอกสารที่ต้องการใช้งาน ไม่ว่าจะเป็นเอกสารต้นแบบ (template) ตามภาคผนวก ข หรือ เอกสารที่ไม่มีต้นแบบ ที่สามารถเรียกได้อีกอย่างว่าเอกสารเปล่า จากนั้นพนักงานสอบสวนทำการสอบสวนหรือสั่งการผ่านไมโครโฟน โดยจะมีไมโครโฟนประจำตัวของแต่ละคน จากนั้นข้อมูลเสียงของการสอบสวนหรือสนทนาทำการประมวลผล (Preprocessing) เพื่อตรวจสอบคุณลักษณะของเสียงที่ไม่พึงประสงค์ออกไป เช่นเสียงรบกวน เป็นต้น จากนั้นระบบทำการส่งข้อมูลเสียงแยกตามไมโครโฟนของแต่ละคนไปยังโปรแกรมพาที (PARTY) ที่พัฒนาโดยศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ (NECTEC) เพิ่มทำการรู้จำให้ออกมาเป็นข้อความตัวหนังสืออัตโนมัติตามเสียงพูด

เมื่อได้ตัวหนังสือตามที่พนักงานสอบสวนได้พูดหรือสั่งการแล้ว ระบบจะทำการจับคู่ข้อความกับตำแหน่งของข้อความในไฟล์เอกสารต้นแบบ โดยจะเป็นการเติมข้อความแบบอัตโนมัติ ซึ่งหลักการจะคล้ายคลึงกับระบบสั่งการด้วยเสียงอัตโนมัติ (Voice Typing) จากนั้นระบบทำการจัดเก็บข้อมูลเสียงและข้อความจากการสอบสวนหรือสนทนา เพื่อใช้ประโยชน์ในการสืบค้นได้ในอนาคต

ตารางที่ 1 คำอธิบายการประมวลผลของโปรเซสที่ 1.0 : เลือกเอกสารต้นแบบ

Process Description	
System :	ระบบรู้จำเสียงพูดอัตโนมัติเพื่องานสอบสวน
DFD Number :	1
Process Name :	เลือกเอกสารต้นแบบ
Input Data Flow :	เอกสารต้นแบบ (template) ตามตารางที่ 1 2 และ 3 หรือ เอกสารที่ไม่มีต้นแบบ
Output Data Flow :	ไฟล์เอกสารต้นแบบ (template) ตามตารางที่ 1 2 และ 3 หรือ เอกสารที่ไม่มีต้นแบบ
Data Stored Used :	ไฟล์เอกสารต้นแบบ (template)
Description :	พนักงานสอบสวนทำการเลือกชนิดของเอกสารที่ต้องการใช้งาน ไม่ว่าจะเป็นเอกสารต้นแบบ (template) ตามตารางที่ 1 2 และ 3 หรือเอกสารที่ไม่มีต้นแบบ ที่สามารถเรียกได้อีกอย่างว่าเอกสารเปล่า

ตารางที่ 2 คำอธิบายการประมวลผลของโปรเซสที่ 2.0 : การรู้จำเสียงพูด (Speech Recognition)

Process Description	
System :	ระบบรู้จำเสียงพูดอัตโนมัติเพื่องานสอบสวน
DFD Number :	2
Process Name :	การรู้จำเสียงพูด (Speech Recognition)
Input Data Flow :	ข้อมูลเสียง
Output Data Flow :	ข้อความตัวหนังสือ
Data Stored Used :	-
Description :	พนักงานสอบสวนทำการสอบสวนหรือสั่งการผ่านไมโครโฟน โดยจะมีไมโครโฟนประจำตัวของแต่ละคน จากนั้นข้อมูลเสียงของการสอบสวนหรือสนทนาทำการประมวลผล(Preprocessing) เพื่อตรวจสอบคุณลักษณะของเสียงที่ไม่พึงประสงค์ออกไป เช่นเสียงรบกวน เป็นต้น จากนั้นระบบทำการส่งข้อมูลเสียงแยกตามไมโครโฟนของแต่ละคนไปยังโปรแกรมพาที(PARTY) ที่พัฒนาโดยศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ (NECTEC) เพิ่มทำการรู้จำให้ออกมาเป็นข้อความตัวหนังสืออัตโนมัติตามเสียงพูด

ตารางที่ 3 คำอธิบายการประมวลผลของโปรเซสที่ 3.0 : การรู้จำเสียงพูด (Speech Recognition)

Process Description	
System :	ระบบรู้จำเสียงพูดอัตโนมัติเพื่องานสอบสวน
DFD Number :	3
Process Name :	เติมข้อความลงเอกสารต้นแบบ
Input Data Flow :	ข้อความตัวหนังสือ, ไฟล์เอกสารต้นแบบ (template) ตามตารางที่ 1 2 และ 3 หรือ เอกสารที่ไม่มีต้นแบบ
Output Data Flow :	เอกสารฉบับสมบูรณ์
Data Stored Used :	-
Description :	เมื่อได้ตัวหนังสือตามที่พนักงานสอบสวนได้พูดหรือสั่งการแล้ว ระบบจะทำการจับคู่ข้อความกับตำแหน่งของข้อความในไฟล์เอกสารต้นแบบ โดยจะเป็นการเติมข้อความแบบอัตโนมัติ

บทที่ 4
บทสรุป

4.1 สรุปผลการดำเนินงาน

จากการศึกษาผลงานวิจัยเรื่อง “การศึกษาระบบรู้จำเสียงพูดอัตโนมัติ” เป็นการศึกษาความต้องการของของผู้ใช้งาน ซึ่งเป็นการเก็บข้อมูลโดยแบบสอบถามปลายเปิด (Open-end questionnaire) ซึ่งจากการเก็บข้อมูลดังกล่าว พบว่าผู้ใช้งานมีความต้องการให้ระบบดังกล่าวสามารถใช้กับเอกสารที่มีต้นแบบและไม่มีต้นแบบ นำไฟล์เสียงเข้าสู่ระบบแปลงข้อความได้ ระบบทำการบันทึกไฟล์เสียงได้หลายรูปแบบ รองรับไมโครโฟนได้หลายรูปแบบ ระบบมีขนาดเล็กติดตั้งง่าย ตอบสนองต่อความต้องการของผู้ใช้อย่างรวดเร็ว ใช้งานได้ง่าย ไฟล์ที่ได้หลังจากการประมวลผลสามารถสั่งพิมพ์ได้ทันที ไฟล์ที่ได้หลังเสร็จสิ้นกระบวนการควรอยู่ในรูปแบบที่แก้ไขได้ง่าย สามารถใช้งานได้ในที่ที่ไม่มีอินเตอร์เน็ต และสามารถรองรับภาษาถิ่นได้หลากหลาย

ซึ่งจากข้อมูลดังกล่าวทำให้ผู้วิจัยสามารถวิเคราะห์และออกแบบระบบให้สอดคล้องกับความต้องการของผู้ใช้งาน โดยเริ่มจากในขั้นตอนของการสอบสวนพนักงานสอบสวนใช้คอมพิวเตอร์หรือ Notebook ที่มีไมโครโฟนพร้อมใช้งานทำการเลือกชนิดของเอกสารที่ต้องการใช้งาน ไม่ว่าจะเป็นเอกสารต้นแบบ (template) ตามภาคผนวก ก หรือ เอกสารที่ไม่มีต้นแบบ ที่สามารถเรียกได้อีกอย่างว่าเอกสารเปล่า จากนั้นพนักงานสอบสวนทำการสอบสวนหรือสั่งการผ่านไมโครโฟน โดยจะมีไมโครโฟนประจำตัวของแต่ละคน จากนั้นข้อมูลเสียงของการสอบสวนหรือสนทนาทำการประมวลผล (Preprocessing) เพื่อตรวจสอบคุณลักษณะของเสียงที่ไม่พึงประสงค์ออกไป เช่นเสียงรบกวนเป็นต้น จากนั้นระบบทำการส่งข้อมูลเสียงแยกตามไมโครโฟนของแต่ละคนไปยังโปรแกรมพาที (PARTY) ที่พัฒนาโดยศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ (NECTEC) เพิ่มทำการรู้จำให้ออกมาเป็นข้อความตัวหนังสืออัตโนมัติตามเสียงพูด

เมื่อได้ตัวหนังสือตามที่พนักงานสอบสวนได้พูดหรือสั่งการแล้ว ระบบจะทำการจับคู่ข้อความกับตำแหน่งของข้อความในไฟล์เอกสารต้นแบบ โดยจะเป็นการเติมข้อความแบบอัตโนมัติ ซึ่งหลักการจะคล้ายคลึงกับระบบสั่งการด้วยเสียงอัตโนมัติ (Voice Typing) จากนั้นระบบทำการจัดเก็บข้อมูลเสียงและข้อความจากการสอบสวนหรือสนทนา เพื่อใช้ประโยชน์ในการสืบค้นได้ในอนาคต

4.2 ปัญหาและอุปสรรค

1) จำนวนการเก็บข้อมูลของผู้ใช้งานไม่เพียงพอ ทำให้ไม่สามารถทราบปัญหาและแนวทางการออกแบบระบบรู้จำเสียงพูดได้ทั้งหมด
2) ข้อจำกัดด้านระยะเวลาในการดำเนินงาน

4.3 แนวทางการดำเนินการต่อไป

จากการศึกษางานวิจัยดังกล่าว เป็นการศึกษาความต้องการของผู้ใช้บริการเพื่อนำมาออกแบบระบบรู้จำเสียงพูดอัตโนมัติเพื่อนำมาใช้ในงานสืบสวนและสอบสวนคดีพิเศษ ซึ่งเมื่อผู้วิจัยได้ออกแบบ

ระบบดังกล่าวแล้ว ได้มีกิจกรรมสำรวจความเห็นผู้ใช้งานที่มีต่อการพัฒนาระบบรู้จำเสียงพูดอัตโนมัติ เมื่อวันที่ 10 กรกฎาคม 2561 ต่อมาได้มีการจัดกิจกรรมเสนอร่างผลงานวิจัยและรับฟังความเห็นที่มีต่อร่างผลงานวิจัย เรื่อง "การศึกษาระบบรู้จำเสียงพูดอัตโนมัติ" ในวันที่ 11 กันยายน 2561 ซึ่งได้มีการเสนอแนะความคิดเห็นต่อระบบและแนวทางในการพัฒนา ดังนี้

1) การแก้ปัญหาเรื่องการแบ่งแยกเสียงหลักกับเสียงรบกวนอื่นๆ ให้ชัดเจน

2) การพัฒนาระบบที่สามารถแก้ไขคำผิดได้อย่างอัตโนมัติโดยไม่ต้องแก้ไขภายหลัง

3) เนื่องจากในการสอบสวนบางครั้งจะมีการสนทนาที่รวดเร็วและพร้อมกันหลายๆคน จึงควรจะพัฒนาระบบให้มีการประมวลผลอย่างรวดเร็วเพื่อตอบสนองต่อการสอบสวน

จากข้อเสนอแนะและความคิดเห็นจากการจัดกิจกรรมดังกล่าว จะเป็นแนวทางที่จะนำไปสู่การพัฒนาให้เกิดเป็น ระบบรู้จำเสียงพูดอัตโนมัติ ที่สอดคล้องกับการใช้งานด้านการสืบสวนสอบสวนคดีพิเศษของกรมสอบสวนคดีพิเศษต่อไป

บรรณานุกรม

บุญเสริม กิจศิริกุล และ ณัฐกร ทับทอง. (2548). **การพัฒนาระบบการรู้จำเสียงพูดภาษาไทย.** กรุงเทพฯ: ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ. (2016). **PARTY: พาที ระบบรู้จำเสียงพูดภาษาไทย.** [ออนไลน์]. เข้าถึงได้จาก http://www.nectec.or.th/innovation/innovation-software/party.html สืบค้น 2 สิงหาคม 2559

Agarwal, A. Dictation (2016). **Online Speech Recognition.** [ออนไลน์]. เข้าถึงได้จาก https://dictation.io สืบค้น 3 สิงหาคม 2559.

Google. (2016). **Android.** [ออนไลน์]. เข้าถึงได้จาก https://www.android.com/ สืบค้น 3 สิงหาคม 2559.

Google. (2016). **Google Cloud Platform.** [ออนไลน์]. เข้าถึงได้จาก https://cloud.google.com สืบค้น 3 สิงหาคม 2559.

Google. (2016). **Google Voice Typing.** [ออนไลน์]. เข้าถึงได้จาก https://www.google.co.th/intl/en/about/products สืบค้น 2 สิงหาคม 2559.

IBM. (2016). **Speech to Text.** [ออนไลน์]. เข้าถึงได้จาก https://www.ibm.com/watson/developercloud/speech-to-text.html สืบค้น 5 สิงหาคม 2559

IBM. (2016). **Speech to Text.** [ออนไลน์]. เข้าถึงได้จาก https://console.ng.bluemix.net/catalog/services/speech-to-text สืบค้น 5 สิงหาคม 2559

Microsoft. (2016). **Windows Speech Recognition.** [ออนไลน์]. เข้าถึงได้จาก https://www.microsoft.com สืบค้น 5 สิงหาคม 2559.

Nuance Communications. (2016). **Dragon.** [ออนไลน์]. เข้าถึงได้จาก http://www.nuance.com/dragon สืบค้น 2 สิงหาคม 2559

Karpagavalli S. and Chandra E. (2016). **A Review on Automatic Speech Recognition Architecture and Approaches.** International Journal of Signal Processing, Image Processing and Pattern Recognition Vol.9, No.4, pp.393-404

Chai Wutiwiwatchai, Sadooki Furui, Thai speech processing technology: A review, Speech communication, Volume 49, Issue 1, January 2007, Pages 8-27, ISSN 0167-6293.

S. Jitapunkul, S. Luksaneeyanawin, V. Ahkuputra, E. Maneenoi, S. Kasuriya and P. Amornkyl (1998), "Recent advances of Thai speech recognition in Thailand," Circuits and Systems. IEEE APCCAS 1998. The 1998 IEEE Asia-Pacific Conference on, Chiangmai 1998, pp. 173-176.

V. Ahkuputra, S. Jitapunkul, E. Maneenoi, S. Kasuriya and P. Amornkul, (1998) "Comparison of different techniques on Thai speech recognition," Circuits and Systems, 1998. IEEE APCCAS 1998. THE 1998 IEEE Asia-Pacific Conference on, Chiangmai, pp. 177-180.

W. Rochkittchareon, A. Suchato and P. Punyabukkana, (2012), "Broad phonetic class segmentation study for Thai automatic speech recognition," Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), 2012 9th International Conference on, Phetchaburi pp. 1-4.

เจษฎา กานต์ประชา) .2545). การรู้จำเสียงพูดภาษาไทยอย่างคงทนโดยใช้สัมประสิทธิ์เมลฟรีเคว็นซีเซปสตอรลของค่าอัตสหสัมพันธ์ของเสียงพูดที่มีสัญญาณรบกวน. จุฬาลงกรณ์มหาวิทยาลัย กรุงเทพมหานคร.

ฐนียา สัตยพานิช) .2541). ระบบรู้จำเสียงพูดภาษาไทยต่อเนื่องแบบเฉพาะบุคคลสำหรับการใช้งานอีเมลล์.มหาวิทยาลัยเกษตรศาสตร์:กรุงเทพมหานคร.

ศุภชัย ตั้งวงศ์ศานต์ และ วิจิตรธเนศานุรักษ์. (2544). การรู้จำเสียงพยางค์ภาษาไทย แบบขึ้นกับผู้พูดโดยวิธี Segmental Probability Model, Department of Computer Science, Faculty of science, Mahidol University, Bangkok.

S. Tangwongsan and R. Phoophuangpairoj, (2008), "Boosting Thai Syllable Speech Recognition Using Acoustic Models Combination," Computer and Electrical Engineering, 2008. ICCEE 2008. International Conference on, Phuket, pp.568-572.

A. Suchato, P. Punyabukkana, P. Ariyakornwijit and T. Namchaisawatwong, (2011), "Automatic sppech recognition of Thai person names from dynamic name lists," Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), 2011 8th International Conference on, Khon Kaen, pp. 962-966.

C. Wutiwiwatchai, K. Thangthai and P. Sertsi, (2012), "Thai ASR Development for network-based speech translation," Speech Database and Assessments (Oriental COCOSDA), 2012 International Conference on, Macau, 2012, pp. 92-96

S. Klaithin, P. Chootrakool and K. Kosawat, (2010), “LEXiTRON-Pro Editor: An integrated tool for developing Thai pronunciation dictionary,” ComputerScience and Information Technology (IMCSIT), Proceedings of the 2010 International Multiconference on, Wisla, pp. 429-433.

“PARTY: พาทีระบบรู้จำเสียงพูดภาษาไทย”, (2559), (ออนไลน์). แหล่งที่มา: http://www.nectec.or.th/innovation/innovation-solfware/party.html

“List of speech recognition software”, (2016), (ออนไลน์). แหล่งที่มา: http://www.capterra.com/speech-recognition-software/

Imai et al. (2004.) Speech Recognition for Subtitling Japanese Live Broadcasts. ICA 2004, Vol.l , pp.165-168

H.Shinichi, K. Akio and O. Takahiro, (2018), “New Real-time Closed-Captioning System for Japanese Broadcast News Programs.” 11th International Conference, ICCHP 2008, Linz, Austria, July 9-11, 2008. Proceedings pp. 651-654.

ภาคผนวก

ผนวก ก. เอกสารที่ใช้ในการทำสำนวนคดีพิเศษ (เป็นเอกสารที่ควรมีอยู่ในเอกสารต้นแบบของระบบ)

รหัสของแบบฟอร์ม	รายการ
คพ.๐๐๑	**ปกสำนวนการสอบสวน**
	หนังสือส่งสำนวนการสอบสวน
คพ.๐๐๒	๑ หนังสือส่งสำนวนคดีพิเศษที่เห็นควรงดการสอบสวน/ให้งดการสอบสวน
คพ.๐๐๓	๒ หนังสือส่งสำนวนคดีพิเศษที่เห็นควรสั่งฟ้อง (ผู้ต้องหาหลบหนี)
คพ.๐๐๔	๓ หนังสือส่งสำนวนคดีพิเศษที่เห็นควรสั่งฟ้อง/สั่งไม่ฟ้อง
	บัญชีสำนวนการสอบสวน
คพ.๐๐๕	๑ บัญชีสำนวนการสอบสวน
คพ.๐๐๕	๒ บัญชีสำนวนการสอบสวน (เพิ่มเติม)
คพ.๐๐๖	**รายงานการสอบสวน**
	เอกสารเกี่ยวกับคำให้การผู้กล่าวหา
คพ.๐๐๗	๑ บันทึกคำให้การผู้กล่าวหา
คพ.๐๐๘	๒ บันทึกสอบสวนผู้กล่าวหา (เพิ่มเติม)
	๓ เอกสารยื่นแทนคำให้การผู้กล่าวหา (ถ้ามี)
	๔ สำเนาเอกสารประกอบคำให้การผู้กล่าวหาเฉพาะที่เป็นสาระสำคัญทางคดี ที่สามารถบ่งชี้ถึงข้อเท็จจริงหรือพยานหลักฐานในคดี (ถ้ามี)
	เอกสารเกี่ยวกับคำให้การผู้ต้องหา
คพ.๐๐๙	๑ บันทึกคำให้การผู้ต้องหา
คพ.๐๑๐	๒ บันทึกสอบสวนผู้ต้องหา (เพิ่มเติม)
	๓ เอกสารยื่นแทนคำให้การผู้ต้องหา (ถ้ามี)
	๔ สำเนาเอกสารประกอบคำให้การผู้ต้องหาเฉพาะที่เป็นสาระสำคัญทางคดีที่สามารถบ่งชี้ถึงข้อเท็จจริงหรือพยานหลักฐานในคดี (ถ้ามี)
	เอกสารเกี่ยวกับคำให้การพยาน (รายบุคคล)
คพ.๐๑๑	๑ บันทึกคำให้การพยาน (เรียงตามลำดับความสำคัญของพยานและ/หรือเหตุการณ์)
คพ.๐๐๘	๒ บันทึกสอบสวนพยาน (เพิ่มเติม)
	๓ เอกสารยื่นแทนคำให้การพยาน (ถ้ามี)
	๔ สำเนาเอกสารประกอบคำให้การพยานเฉพาะที่เป็นสาระสำคัญทางคดี ที่สามารถบ่งชี้ถึงข้อเท็จจริงหรือพยานหลักฐานในคดี (ถ้ามี)
คพ.๐๑๒	รายการพยานเอกสารและพยานวัตถุ
	รายละเอียดเกี่ยวกับของกลาง
คพ.๐๑๓	๑ บัญชีของกลางคดีอาญา
คพ.๐๑๔	๒ ภาพถ่ายของกลาง
	รายละเอียดเกี่ยวกับทรัพย์ถูกประทุษร้าย
คพ.๐๑๕	๑ บัญชีทรัพย์ถูกประทุษร้าย
คพ.๐๑๖	๒ ภาพถ่ายทรัพย์ที่ถูกประทุษร้าย
คพ.๐๑๗	๓ บัญชีมูลค่าความเสียหาย
คพ.๐๑๘	๔ บัญชีทรัพย์ถูกประทุษร้ายได้คืน

รหัสของแบบฟอร์ม	รายการ
คพ.๐๑๙	๕ ภาพถ่ายทรัพย์ที่ถูกประทุษร้ายได้คืน
คพ.๐๒๐	**บันทึกพนักงานสอบสวน**
คพ.๐๒๑	**รายงานการประชุมคณะพนักงานสอบสวนคดีพิเศษ**
	เอกสารเกี่ยวกับสถานที่เกิดเหตุและสถานที่เกี่ยวข้อง
คพ.๐๒๒	๑ บันทึกการตรวจสถานที่เกิดเหตุ
คพ.๐๒๓	๒ แผนที่สังเขปแสดงสถานที่เกิดเหตุ
คพ.๐๒๔	๓ ภาพถ่ายประกอบการตรวจสถานที่เกิดเหตุ
	๔ รายงานการตรวจสถานที่เกิดเหตุของหน่วยงานหรือผู้เชี่ยวชาญด้านนิติวิทยาศาสตร์ (ถ้ามี)
คพ.๐๒๕	๕ บันทึกการนำชี้สถานที่ประกอบคำให้การของผู้กล่าวหา ผู้ต้องหา หรือพยาน
คพ.๐๒๖	๖ ภาพถ่ายการนำชี้สถานที่ประกอบคำให้การของผู้กล่าวหา ผู้ต้องหา หรือพยาน
	บันทึกส่งผู้บาดเจ็บหรือศพฯ หนังสือนำส่งและรายงานผลการตรวจพิสูจน์พยานหลักฐานทางนิติวิทยาศาสตร์
คพ.๐๒๗	๑ บันทึกส่งผู้บาดเจ็บหรือศพให้แพทย์ตรวจชันสูตรและผลการตรวจชันสูตรบาดแผลหรือศพ ของแพทย์
คพ.๐๒๘	๒ หนังสือนำส่งการตรวจพิสูจน์ยาเสพติด
คพ.๐๒๙	๓ หนังสือนำส่งเอกสารและวัตถุของกลางตรวจพิสูจน์
	เอกสารเกี่ยวกับพยาน
คพ.๐๓๐	๑ หมายเรียกพยาน
คพ.๐๓๑	๒ หนังสือขอเชิญมาให้ถ้อยคำ
คพ.๐๓๒	๓ หนังสือขอตรวจสอบข้อมูลการทำธุรกรรมทางการเงิน
คพ.๐๓๓	๔ หนังสือขอตรวจสอบข้อมูลเกี่ยวกับการจดทะเบียนนิติบุคคล
	๕ เอกสารโต้ตอบ (เรียงตามลำดับเวลาแต่ละหน่วยงาน) (ถ้ามี)
	เอกสารที่เกี่ยวข้องกับการใช้มาตรการพิเศษตาม พระราชบัญญัติการสอบสวนคดีพิเศษและกฎหมายอื่น
คพ.๐๓๔	๑ หนังสือขออนุมัติเข้าถึงข้อมูลข่าวสาร
คพ.๐๓๕	๒ คำสั่งอนุมัติให้เข้าถึงข้อมูลข่าวสาร
คพ.๐๓๖	๓ หนังสือขอเข้าถึงข้อมูลข่าวสาร
คพ.๐๓๗	๔ คำขอให้ได้มาซึ่งข้อมูลข่าวสาร
คพ.๐๓๘	๕ หนังสือรายงานผลการดำเนินการเข้าถึงข้อมูลข่าวสาร
คพ.๐๓๙	๖ หนังสือขอใช้ประโยชน์จากเอกสารหรือข้อมูลข่าวสาร
คพ.๐๔๐	๗ หนังสือรับรองการรักษาความลับ
	เอกสารเกี่ยวกับการค้น
	๑ คำร้องขอหมายค้น
	๒ คำให้การพยานประกอบคำร้อง
	๓ รายงานกระบวนพิจารณา

รหัสของแบบฟอร์ม	รายการ
คพ.๐๔๑ คพ.๐๔๒ คพ.๐๔๓ คพ.๐๔๔ คพ.๐๔๕ คพ.๐๔๖	๔ แบบหมายค้น ๕ บันทึกการตรวจค้น ๖ บัญชีรายละเอียดสิ่งของแนบบันทึกการตรวจค้น ๗ หนังสืออนุมัติให้ดำเนินการตามมาตรา ๒๔ ๘ บันทึกเหตุสงสัยตามสมควรและเหตุอันควรเชื่อที่ทำให้สามารถเข้าค้นได้ (ค้นไม่มีหมาย) ๙ หนังสือรายงานผลการปฏิบัติตามหมายค้น ๑๐ บันทึกการยึดหรืออายัดสิ่งของ
คพ.๐๔๗	**เอกสารที่มาของอำนาจการสอบสวน เช่น เอกสารที่แสดงถึงอำนาจในการสืบสวนสอบสวน และหนังสือแจ้งมติ กคพ. (ตั้งแต่รายงานการสืบสวน, การขอรับเป็นคดีพิเศษ จนถึงการแต่งตั้งคณะพนักงานสอบสวนและพนักงานอัยการ รวมทั้งที่ปรึกษา) คำร้องขอสืบพยานก่อนฟ้อง** ๑ สำนวนการสืบสวนหรือตรวจสอบข้อเท็จจริง ๒ แบบเสนอเรื่องเพื่อมีมติให้การกระทำความผิดทางอาญาอื่นเป็นคดีพิเศษตามมาตรา ๒๑ วรรคหนึ่ง (๒) แห่งพระราชบัญญัติการสอบสวนคดีพิเศษ พ.ศ. ๒๕๔๗ (แบบ กคพ.๑) ๓ แบบเสนอเรื่องที่คณะอนุกรรมการคดีพิเศษเห็นควรไม่มีมติให้การกระทำความผิดทางอาญาอื่นเป็นคดีพิเศษตามมาตรา ๒๑ วรรคหนึ่ง (๒) แห่งพระราชบัญญัติการสอบสวนคดีพิเศษ พ.ศ. ๒๕๔๗ (แบบ กคพ.๒) ๔ แบบเสนอเรื่องต่อคณะกรรมการคดีพิเศษเพื่อพิจารณามีมติให้คดีความผิดอาญาที่ค้างดำเนินการและคดียังไม่ถึงที่สุดเป็นอำนาจหน้าที่ของพนักงานสอบสวนคดีพิเศษ ตามมาตรา ๔๔ แห่งพระราชบัญญัติการสอบสวนคดีพิเศษ พ.ศ. ๒๕๔๗ (แบบ กคพ.๓) ๕ แบบเสนอเรื่องต่อคณะกรรมการคดีพิเศษเพื่อพิจารณาชี้ขาดตามมาตรา ๒๑ วรรคท้าย (แบบ กคพ.๔) ๖ สำเนารายงานการประชุมคณะกรรมการคดีพิเศษ (กคพ.) ที่มีมติรับเป็นคดีพิเศษ ๗ ประกาศกรมสอบสวนคดีพิเศษ เรื่องมติคณะกรรมการคดีพิเศษให้คดีความผิดทางอาญาอื่นเป็นคดีพิเศษ ๘ คำสั่งอธิบดีกรมสอบสวนคดีพิเศษ อนุมัติให้ทำการสอบสวนเป็นคดีพิเศษ ตามมาตรา ๒๑ วรรคหนึ่ง (๑) ๙ หนังสือจากสำนักงานอัยการสูงสุดมอบหมายให้กรมสอบสวนคดีพิเศษทำการสอบสวนความผิดทางอาญาที่กระทำนอกราชอาณาจักรและเอกสารที่เกี่ยวข้อง ๑๐ คำสั่งแต่งตั้งคณะพนักงานสอบสวนคดีพิเศษ (แบบ ศบพ.๒) ๑๑ คำสั่งแต่งตั้งคณะพนักงานสอบสวนคดีพิเศษ สำหรับคดีความผิดซึ่งมีโทษตามกฎหมายไทยได้กระทำลงนอกราชอาณาจักรไทย (แบบ ศบพ.๓) ๑๒ คำสั่งแต่งตั้งคณะพนักงานสอบสวนคดีพิเศษ เพิ่มเติม ๑๓ แบบคำร้องขอออกเลขคดีพิเศษ (แบบ ศบพ.๙) ๑๔ คำสั่งแต่งตั้งผู้ช่วยเหลือพนักงานสอบสวนคดีพิเศษหรือเจ้าหน้าที่คดีพิเศษ ๑๕ หนังสือขอสืบพยานก่อนฟ้อง
	เอกสารการปฏิบัติเกี่ยวกับตัวผู้ต้องหา

รหัสของแบบฟอร์ม	รายการ
คพ.๐๔๘	๑ หมายเรียกผู้ต้องหาพร้อมหลักฐานนำส่ง
คพ.๐๔๙	๒ บันทึกแจ้งข้อหา
คพ.๐๕๐	๓ บันทึกการชี้รูปผู้ต้องหา
	๔ คำร้องขอหมายจับและรายงานกระบวนพิจารณา
	๕ หมายจับ
คพ.๐๕๑	๖ ตำหนิรูปพรรณผู้กระทำความผิด
คพ.๐๕๒	๗ หนังสืองดการสืบจับบุคคล
คพ.๐๕๓	๘ บันทึกการจับกุม
คพ.๐๕๔	๙ บันทึกการมอบตัว/จับกุม/แจ้งข้อหา
	๑๐ คำร้องตรวจสอบการจับกุมตัวผู้ต้องหาศาลเยาวชนและครอบครัว
คพ.๐๕๕	๑๑ หนังสือแจ้งการจับกุมผู้ต้องหาที่เป็นทหาร
คพ.๐๕๖	๑๒ บันทึกผลการชี้ตัวผู้ต้องหา
คพ.๐๕๗	๑๓ บันทึกการแจ้งสิทธิในการรับตัวผู้ต้องหา
คพ.๐๕๘	๑๔ หนังสือขออายัดตัวผู้ต้องขัง
คพ.๐๕๙	๑๕ หนังสือขอพบผู้ต้องขังเพื่อแจ้งข้อกล่าวหา สอบสวนปากคำ และพิมพ์ลายนิ้วมือ
คพ.๐๖๐	๑๖ บันทึกการควบคุมผู้ต้องหา
	๑๗ คำร้องขอผัดฟ้อง/ฝากขังผู้ต้องหา(ศาลแขวง) ครั้งที่ ๑
	๑๘ คำร้องขอผัดฟ้อง/ฝากขังผู้ต้องหา(ศาลแขวง) ครั้งที่ ๒ – ๕ และคำให้การพยานชั้นไต่สวนฝากขัง
	๑๙ คำให้การพยานชั้นไต่สวนผัดฟ้อง/ฝากขัง ครั้งที่.....
	๒๐ คำร้องฝากขังผู้ต้องหา(ศาลจังหวัด) ครั้งที่ ๑
	๒๑ คำร้องฝากขังผู้ต้องหา(ศาลจังหวัด) ครั้งที่.....
คพ.๐๖๑	๒๒ บันทึกเสนอสัญญาประกัน
คพ.๐๖๒	๒๓ สัญญาประกันตัวผู้ต้องหา
คพ.๐๖๓	๒๔ คำร้องขอประกันและตารางนัดหมาย
คพ.๐๖๔	๒๕ ใบนัดส่งตัวผู้ต้องหา
	๒๖ ภาพถ่ายผู้ต้องหา/ประวัติและแบบพิมพ์ลายนิ้วมือผู้ต้องหา/ผลคดี
คพ.๐๖๕	๒๗ คำยินยอมกรณีผู้ให้สัญญาค้ำประกันมีคู่สมรส
คพ.๐๖๖	๒๘ คำร้องขอคืนสิ่งของ
คพ.๐๖๗	๒๙ สัญญาประกันและรับมอบสิ่งของ
คพ.๐๖๘	บัญชีพยานบุคคล
	เอกสารอื่นๆ ที่เกี่ยวข้องกับการดำเนินคดี เช่น การส่งตัว ,ผู้ต้องหาไปควบคุม ,การฝากของกลาง ,การเบิกสำนวน ,การส่งเก็บสำนวน ให้นำไปใส่ไว้ใน “ร่างสำนวนการสอบสวน”