รายงานวิจัยฉบับสมบูรณ์

โครงการเชื่อมโยงการวิจัยภาควิชาวิศวกรรมคอมพิวเตอร์ สู่ภาคอุตสาหกรรม ปี 2546

โครงการย่อยที่ 3 การพัฒนาระบบการรู้จำเสียงพูดภาษาไทย

ภาควิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์
จุฬาลงกรณ์มหาวิทยาลัย

บุญเสริม กิจศิริกุล
ณัฐกร ทับทอง

การพัฒนาระบบการรู้จำเสียงพูดภาษาไทย

Development of Thai Speech Recognition Systems

คำนำ

เอกสารนี้เป็นรายงานวิจัยฉบับสมบูรณ์โครงการเชื่อมโยงการวิจัยภาควิชาวิศวกรรมคอมพิวเตอร์สู่ภาคอุตสาหกรรม ปี 2546 โครงการย่อยที่ 3 “การพัฒนาระบบการรู้จำเสียงพูดภาษาไทย” ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย งานวิจัยนี้ทำการศึกษาวิจัยระบบรู้จำเสียงพูดภาษาไทยและได้พัฒนาต้นแบบระบบรู้จำเสียงพูดภาษาไทยต้นแบบขึ้นสองระบบคือ (1) ระบบการโอนสายโทรศัพท์อัติโนมัติจากเสียงพูดชื่อไทย และ (2) ระบบสอบถามรายนามผู้ใช้โทรศัพท์โดยผ่านระบบโทรศัพท์

ขอขอบคุณผู้ช่วยวิจัยและผู้มีส่วนร่วมในโครงการนี้ทุกท่าน ซึ่งได้แก่ ประเสริฐศักดิ์ ผุงประเสริฐยิ่ง, รุ่งศรัทธา ด่านศิริวิโรจน์, รณชัย มีฤทธิ์ และอภิวัฒน์ ตรีเลิศมาลา

บุญเสริม กิจศิริกุล
ณัฐกร ทับทอง
กันยายน 2548

สารบัญ

1. บทนำ

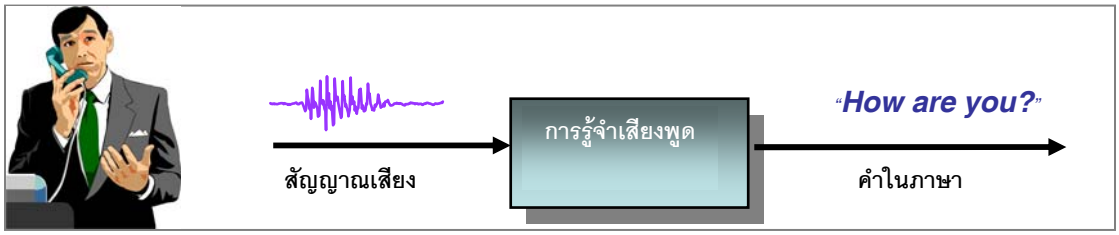
บทนี้นำเสนอความรู้ทั่วไปในการรู้จำเสียงพูด เริ่มจากปัญหาการรู้จำเสียงพูด ความสำคัญของการรู้จำเสียงพูด ปัจจัยต่างๆ ที่มีผลต่อการพัฒนาระบบรู้จำเสียงพูด ประวัติความเป็นมาของระบบรู้จำเสียงพูดและวัตถุประสงค์ของโครงการ

1.1 ปัญหาการรู้จำเสียงพูด

1.1.1 นิยามปัญหา [1]

การรู้จำเสียงพูด (Speech Recognition) เป็นกระบวนการสกัดลำดับของคำ (Sequence of Words ที่อยู่ในสัญญาณเสียงออกมา ดังรูปที่ 1.1 การรู้จำเสียงพูดเป็นเทคโนโลยีด้านเสียงพูด (Speech Technology) ซึ่งเทคโนโลยีด้านเสียงพูดมีหลายด้าน เช่น

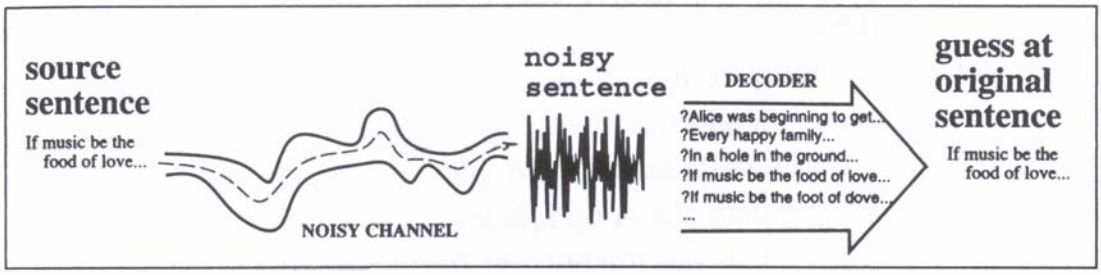
- การทำความเข้าใจเสียงพูด (Speech Understanding) เป็นการทำความเข้าใจความหมายที่มีอยู่ในลำดับของคำที่สกัดมาจากสัญญาณเสียงพูด
- การสังเคราะห์เสียงพูด (Speech Synthesis) เป็นกระบวนการตรงข้ามกับการรู้จำเสียงพูด นั่นคือเป็นการสังเคราะห์สัญญาณเสียงจากลำดับของคำในภาษา
- การระบุตัวผู้พูด (Speaker Identification) เป็นกระบวนการที่รับสัญญาณเสียงแล้วระบุว่าผู้ที่พูดสัญญาณเสียงนี้เป็นผู้ใด



รูปที่ 1.1 การรู้จำเสียงพูด

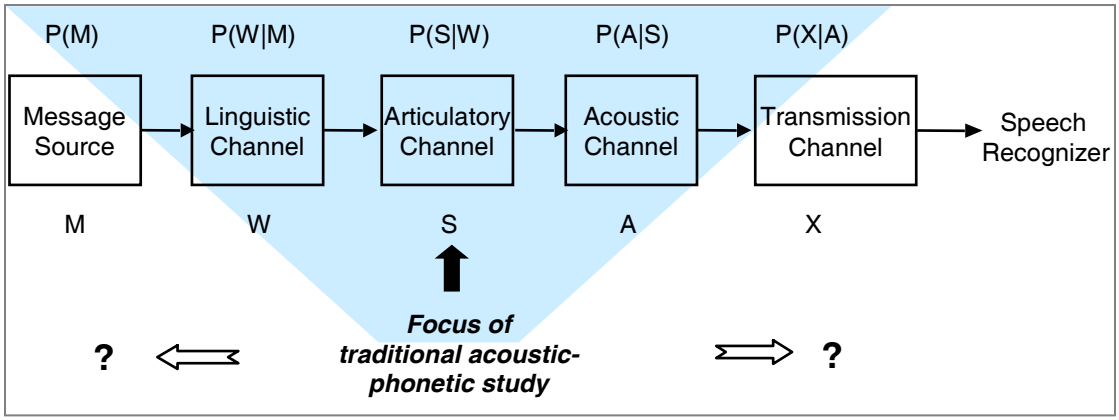
1.1.2 การรู้จำเสียงพูดเมื่อมองในเชิงทฤษฎีสารสนเทศ

ในเชิงทฤษฎีสารสนเทศ (Information Theory) เราพิจารณาเสียงพูดที่ได้ยินว่าเป็นสัญญาณที่ถูกรบกวนผ่านทางช่องสัญญาณรบกวน (Noisy Channel) และการรู้จำเสียงพูดคือการถอดรหัส (Decode) สัญญาณนั้น [2] ดังรูปที่ 1.2



รูปที่ 1.2 การรู้จำเสียงพูดเมื่อมองในเชิงทฤษฎีสารสนเทศ

ซึ่งช่องรบกวนในที่นี้สามารถจำแนกได้ตามช่องสื่อสาร [3] ดังรูปที่ 1.3



รูปที่ 1.3 ช่องรบกวนซึ่งจำแนกตามช่องสื่อสาร

จากรูปที่ 1.3 จะเห็นว่าช่องสื่อสารมีลำดับดังนี้

- **แหล่งกำเนิดข้อความ** (Message Source) เป็นความคิด (M) ที่ผู้ส่งสารต้องการจะสื่อไปยังผู้รับสาร
- **ช่องภาษา** (Linguistic Channel) ที่ช่องนี้ ความคิด (M) จะถูกเปลี่ยนเป็นภาษา หรือลำดับของคำ (W) เพื่อใช้ในการสื่อสาร โดยลำดับของคำหลายชุดอาจสามารถแทนความคิดเดียวกันได้
- **ช่องการออกเสียง** (Articulatory Channel) ที่ช่องนี้ ลำดับของคำ (W) จะถูกเปลี่ยนเป็นเสียงพูด (S) โดยอวัยวะที่ใช้ในการออกเสียง ซึ่งลำดับคำหนึ่งๆ อาจสามารถออกเสียงได้หลายเสียง
- **ช่องตัวกลาง** (Acoustic Channel) เสียงพูด (S) จะเดินทางผ่านช่องนี้ไปสู่ผู้รับสาร ช่องนี้ถือเป็นตัวกลางนำเสียงพูด ซึ่งตัวกลางนำเสียงพูดของมนุษย์โดยทั่วไปคืออากาศ เมื่อผ่านช่องนี้ เสียงพูดที่ไปสู่ผู้รับสาร (A) อาจถูกบิดเบือนจากเสียงพูดที่ผู้ส่งสารพูดออกมา เนื่องจากสัญญาณรบกวนที่มีอยู่

- **ช่องถ่ายทอดเสียง** (Transmission Channel) ผู้รับสารจะเปลี่ยนเสียงพูดที่มาถึง (A) เป็นสัญญาณอย่างใดอย่างหนึ่ง (X) เพื่อใช้ในการรู้จำและทำความเข้าใจต่อไป โดยช่องนี้สำหรับคนคือหู ซึ่งเปลี่ยนเสียงที่ได้รับเป็นสัญญาณไฟฟ้าในระบบประสาท สำหรับในเครื่องคอมพิวเตอร์ ช่องนี้คือไมโครโฟน และวงจรที่แปลงจากสัญญาณอนาล็อกเป็นสัญญาณดิจิตอล

ปัจจุบันการวิจัยเรื่องการรู้จำเสียงพูดได้เน้นที่ช่องการออกเสียง ซึ่งเป็นการศึกษาเสียงต่างๆ ที่ได้จากการออกเสียง (Acoustic-Phonetic Study)

1.1.3 การหาลำดับของคำที่ดีที่สุดของข้อมูลทางเสียงที่สังเกตได้ [2] [4]

ปัญหาการรู้จำเสียงพูดหรือการถอดรหัสของสัญญาณเสียงพูด สามารถมองได้ว่าเป็นการหาลำดับของคำ (Sequence of Words) ที่ดีที่สุดสำหรับลำดับของข้อมูลทางเสียงที่สังเกตได้ (Observation Sequence) โดยกำหนดให้ สัญลักษณ์ของลำดับข้อมูลที่สังเกตได้เป็น $O = o_1 o_2 o_3 ... o_t$ สัญลักษณ์ของลำดับของคำเป็น $W = w_1 w_2 w_3 ... w_n$ และ L แทนภาษาที่พิจารณา

ให้ $\hat{W}$ เป็นลำดับของคำที่ดีที่สุด เพราะฉะนั้นเราจะได้ว่า

$$\hat{W} = \arg\max_{W \in L} P(W \mid O)$$

ซึ่งในทางปฏิบัติแล้ว การหาค่า $P(W \mid O)$ โดยตรงทำได้ยาก จึงเขียนสูตรข้างต้นใหม่โดยใช้กฎของเบย์ได้ดังนี้

$$\hat{W} = \arg\max_{W \in L} \frac{P(O \mid W) P(W)}{P(O)}$$

$$= \arg\max_{W \in L} P(O \mid W) P(W)$$

ในที่นี้ $P(W)$ คือความน่าจะเป็นที่ลำดับของคำ W จะเกิดขึ้นในภาษา จึงเรียก $P(W)$ ว่าความรู้ก่อนหน้า (Prior) หรือแบบจำลองทางภาษา (Language Model) ส่วน $P(O \mid W)$ เป็นความน่าจะเป็นที่ลำดับของข้อมูลที่สังเกตได้เป็น O เมื่อลำดับของคำที่เป็นที่มาของ O คือ W และเรียก $P(O \mid W)$ ว่าความเป็นไปได้ (Likelihood) หรือแบบจำลองทางเสียง (Acoustic Model)

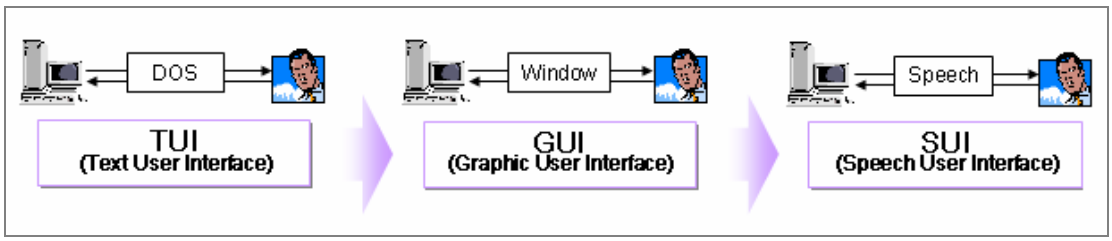
1.2 ความสำคัญ

การรู้จำเสียงพูดเป็นเทคโนโลยีที่สามารถนำไปใช้ประโยชน์ได้หลากหลาย บทบาทของเทคโนโลยีการรู้จำเสียงพูดที่สำคัญในปัจจุบันมีดังต่อไปนี้

1.2.1 เป็นตัวเชื่อมประสานกับผู้ใช้ (User Interface)

เมื่อนำการรู้จำเสียงพูดมาเป็นตัวเชื่อมประสานกับผู้ใช้ การติดต่อระหว่างมนุษย์กับคอมพิวเตอร์ (Human Computerl Interaction) จะเป็นไปในทางที่เป็นธรรมชาติ และอำนวยความสะดวกต่อมนุษย์มากขึ้น เนื่องจากเสียงพูดเป็นช่องทางหลัก และเป็นช่องทางที่สะดวกที่สุดในการติดต่อสื่อสารของมนุษย์ [5] ดังจะเห็นจากแผนภาพการวิวัฒน์ในรูปที่ 1.4 ที่จากเดิมการติดต่อ

ระหว่างมนุษย์กับคอมพิวเตอร์อยู่ในรูปแบบตัวอักษร ก่อนที่จะพัฒนามาเป็นการใช้รูปภาพ และเทคโนโลยีเกี่ยวกับเสียงพูดตามลำดับ



รูปที่ 1.4 วิวัฒนาการของการติดต่อระหว่างมนุษย์กับคอมพิวเตอร์

นอกจากจะช่วยอำนวยความสะดวกในการติดต่อระหว่างมนุษย์กับคอมพิวเตอร์แล้ว เทคโนโลยีการรู้จำเสียงพูดยังเป็นเทคโนโลยีทางเลือกที่ทำให้มนุษย์ติดต่อกับคอมพิวเตอร์ได้ในสภาวการณ์ที่คับขันหรือในกรณีที่การติดต่อระหว่างมนุษย์กับคอมพิวเตอร์วิธีอื่นใช้การไม่ได้ เช่น

- ขณะที่มือไม่ว่าง
- ต้องการความคล่องตัว
- สายตาไม่ว่าง
- ไม่ต้องการใช้คีย์บอร์ด
- ทัศนวิสัยไม่ดี
- มีข้อจำกัดทางด้านร่างกาย
- ฯลฯ

1.2.2 เป็นตัวเชื่อมประสานกับผู้ใช้ในโปรแกรมประยุกต์ทางโทรศัพท์

การรู้จำเสียงพูดสามารถนำมาใช้เพื่อการสอบถามและสืบค้นข้อมูลทางโทรศัพท์ ทำให้เกิดความสะดวกรวดเร็ว และลดค่าใช้จ่ายในการจ้างบุคลากร ปัจจุบันโปรแกรมประยุกต์ทางโทรศัพท์ที่ใช้เทคโนโลยีเกี่ยวกับเสียงพูดมีมูลค่าทางการตลาดมหาศาล และถูกนำไปใช้ในงานต่างๆ หลายประเภท เช่น

- การสอบถามข้อมูลต่างๆ เช่น ข้อมูลเที่ยวบิน [6] เส้นทางและสภาพการจราจร [7] ร้านอาหาร [8] การซื้อขายรถยนต์ [9] สภาพดินฟ้าอากาศ [10] [11] เป็นต้น
- การจองประเภทต่างๆ เช่น ตั๋วเครื่องบิน [12] เป็นต้น

1.2.3 เป็นตัวเชื่อมประสานกับผู้ใช้ในโปรแกรมประยุกต์อื่น ๆ

ผู้ใช้สามารถใช้งานโปรแกรมประยุกต์หลายชนิดได้สะดวกและรวดเร็วขึ้นเมื่อติดต่อด้วยเสียงพูด เช่น

- โปรแกรมประมวลผลคำ (Word Processing) ซึ่งจะเร็วกว่าถ้าผู้ใช้สามารถใช้การบอกจด (Dictation) แทนการพิมพ์ [13] [14]
- โปรแกรมแปลภาษาแบบทันกาล (Real-time Translation) ซึ่งรับเสียงพูดจากภาษาหนึ่งและแปลเป็นอีกภาษาหนึ่งโดยทันที [15]
- ฯลฯ

1.3 ปัจจัยในการพัฒนาระบบ [1] [16]

ความยากง่ายในการพัฒนา และประสิทธิภาพของระบบรู้จำเสียงพูดขึ้นอยู่กับปัจจัยต่างๆ ดังต่อไปนี้

1.3.1 จำนวนคำศัพท์ (Vocabulary Size)

ระบบที่รองรับจำนวนคำศัพท์น้อย เช่น รู้จำเสียงพูดของตัวเลข ศูนย์ ถึง เก้า สามารถพัฒนาได้ง่ายกว่าและให้ความผิดพลาดน้อยกว่าระบบที่ต้องรองรับจำนวนคำศัพท์มาก เช่น รู้จำเสียงพูดของทุกคำในภาษา ซึ่งอาจมีมากถึง 20,000 คำ

1.3.2 ความขึ้นต่อผู้พูด (Speaker Dependency)

ระบบรู้จำเสียงพูดสามารถแบ่งตามความขึ้นต่อผู้พูดที่ระบบสามารถรองรับได้ ได้แก่

- **ระบบที่ขึ้นกับผู้พูด** (Speaker Dependent) ระบบนี้รู้จำได้เฉพาะเสียงพูดของผู้ใช้ที่มีจำนวนจำกัด อาจเป็นผู้ใช้เพียงคนเดียว หรือผู้ใช้เป็นกลุ่ม ซึ่งระบบนี้จำเป็นต้องมีการฝึกโดยใช้เสียงพูดของผู้ใช้ก่อน
- **ระบบที่ไม่ขึ้นกับผู้พูด** (Speaker Independent) จะรู้จำเสียงพูดโดยไม่ขึ้นอยู่กับว่าใครเป็นผู้พูด ซึ่งพัฒนาได้ยากกว่าและให้ความผิดพลาดมากกว่า เพราะผู้ใช้แต่ละคนย่อมมีลักษณะของเสียงที่ต่างกันออกไป

1.3.3 รูปแบบการพูด (Speech Style)

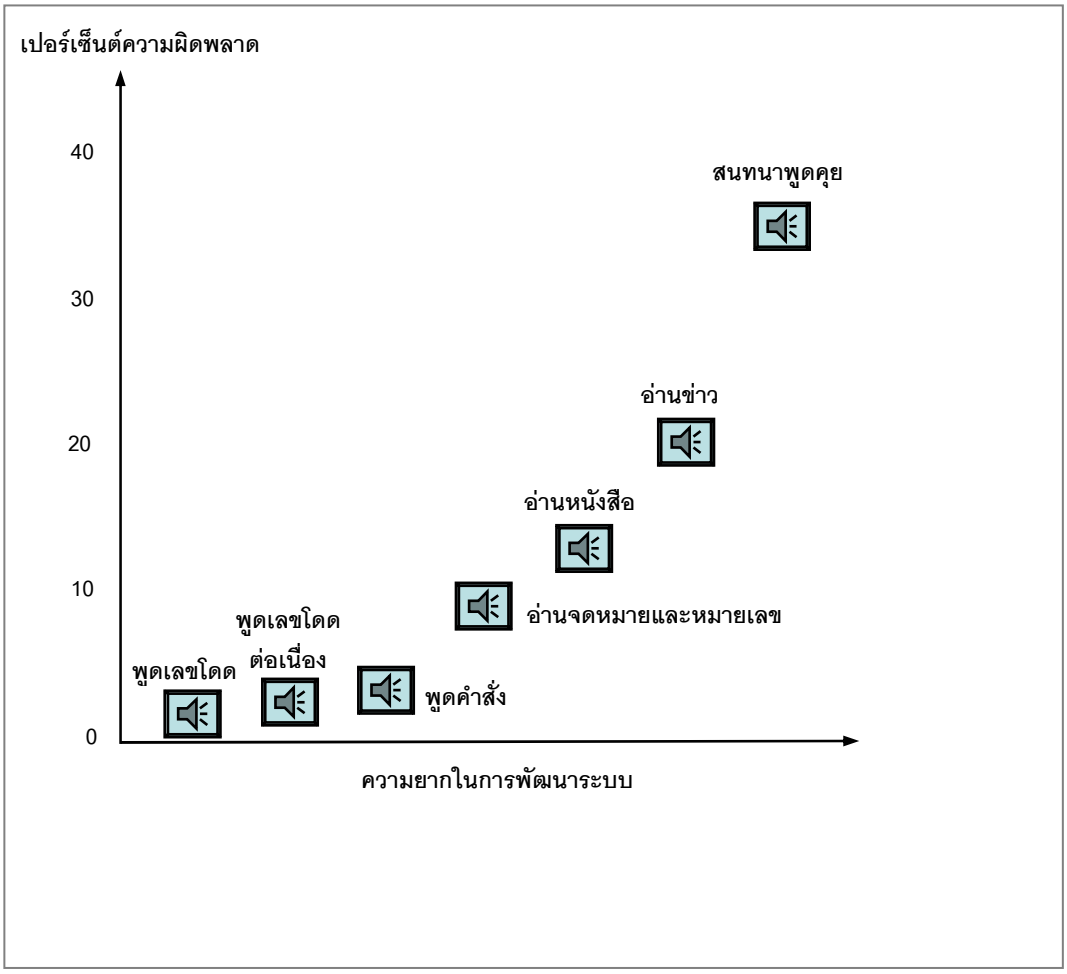
ระบบรู้จำเสียงพูดสามารถแบ่งตามรูปแบบการพูดที่ระบบสามารถรองรับได้ ตั้งแต่

- **เสียงพูดคำเดี่ยว** (Isolated Speech) เป็นการพูดทีละคำอย่างชัดเจน จุดเริ่มต้นและจุดสิ้นสุดของคำเป็นเสียงเงียบ จึงสามารถระบุขอบเขตของคำได้อย่างแน่ชัด รวมทั้งเสียงของคำไม่เพี้ยนมากนัก ทำให้รู้จำได้ง่ายที่สุด ตัวอย่างการพูดรูปแบบนี้ได้แก่ การพูดตัวเลขเดี่ยว การพูดชื่อคน เป็นต้น
- **เสียงพูดคำต่อเนื่อง** (Continuous Speech) เป็นการพูดหลายคำสั้นๆ ติดกัน โดยขอบเขตของแต่ละคำจะแยกจากกันไม่ชัดเจน นอกจากนี้แต่ละคำที่พูดมีจะมีความหลากหลายในการออกเสียง เนื่องจากมีการพูดที่เป็นธรรมชาติ และได้รับผลกระทบจากเสียงของคำอื่น ทำให้การรู้จำทำได้ยากกว่า ตัวอย่างการพูดรูปแบบนี้ได้แก่ การพูดหมายเลขโทรศัพท์ การออกคำสั่ง เป็นต้น
- **เสียงพูดคำอ่าน** (Read Speech) เป็นการพูดต่อเนื่องอย่างยาวนาน เสียงที่พูดมีความเป็นธรรมชาติกว่าการพูดคำต่อเนื่อง และมักมีจำนวนคำศัพท์มาก ทำให้การรู้จำทำได้ยากขึ้น ตัวอย่างการพูดรูปแบบนี้ได้แก่ การอ่านข่าวกระจายเสียง การอ่านนิทาน เป็นต้น
- **เสียงพูดแบบสนทนา** (Conversational Speech) เป็นรูปแบบการพูดที่ทำการรู้จำได้ยากที่สุด เนื่องจากเป็นการพูดที่ไม่เป็นทางการ และเป็นธรรมชาติที่สุด คำศัพท์ที่ใช้พูดอาจเป็นคำศัพท์ที่ระบบไม่รู้จัก นอกจากนี้ยังมีเสียงอื่นๆ คอยแทรก เช่น เสียงหัวเราะ และเสียงอุทาน ตัวอย่างการพูดรูปแบบนี้ได้แก่ การสนทนาทางโทรศัพท์ การพูดคุยระหว่างเพื่อนฝูง เป็นต้น

1.3.4 สภาพแวดล้อมของสัญญาณรบกวน (Noise Environment)

ระบบที่ทำการรู้จำในสภาพแวดล้อมที่เงียบสนิทจะพัฒนาได้ง่ายกว่าและให้ความผิดพลาดน้อยกว่าระบบที่ทำการรู้จำในสภาพแวดล้อมที่มีเสียงรบกวน ซึ่งในกรณีหลัง ระบบจะต้องทำการแยกแยะเสียงรบกวนออกจากเสียงพูดของผู้ใช้งาน

ปัจจัยของลักษณะการพูดแบบต่าง ๆ ก่อให้เกิดความยากในการพัฒนาระบบและความผิดพลาดของระบบ [1] ดังแสดงในรูปที่ 1.5 ซึ่งสังเกตได้ว่า ยิ่งการพูดมีลักษณะเป็นธรรมชาติมากขึ้น ความยากในการพัฒนาระบบและความผิดพลาดของระบบจะยิ่งมีมากขึ้นเป็นเงาตามตัว



รูปที่ 1.5 ความยากในการพัฒนาระบบและเปอร์เซ็นต์ความผิดพลาดในการรู้จำของการพูดลักษณะต่าง ๆ

1.4 ความเป็นมาของระบบรู้จำเสียงพูดในระดับสากล [17] [18] [19]

พัฒนาการของระบบรู้จำเสียงพูดในระดับสากลสามารถแบ่งเป็นช่วงและเรียงตามลำดับเวลาได้ดังนี้

1.4.1 ช่วงก่อนปี ค.ศ. 1950

ช่วงนี้เป็นช่วงที่ยังไม่มีการพัฒนาระบบที่เป็นระบบรู้จำเสียงพูดอย่างแท้จริง แต่มีหลายเทคโนโลยีที่เป็นพื้นฐานสำคัญของการพัฒนาระบบรู้จำเสียงพูด ซึ่งเรียงตามลำดับเวลาได้ดังนี้

- ค.ศ. 1876 เบลล์ประดิษฐ์โทรศัพท์ [20] [21]
- ค.ศ. 1875 เอดิสันประดิษฐ์เครื่องบันทึกเสียง (Phonograph) [22]
- ค.ศ. 1913 คิดค้นแบบจำลองมาคอฟ (Markov Models) [23]
- ค.ศ. 1922 ของเล่นเรดิโอเร็กซ์ (Radio Rex) ดังแสดงในรูปที่ 1.6 ถือเป็นระบบแรกในโลกที่สามารถรู้จำเสียงพูด ของเล่นชิ้นนี้ประกอบด้วยสุนัขเซลลูลอยด์ติดสปริงนั่งอยู่ในบ้านและถูกยึดติดกับฐานบ้านไว้ด้วยแรงแม่เหล็กไฟฟ้า แรงนี้จะส่งผ่านโลหะสองชิ้นเพื่อยึดสุนัขและฐานบ้านเข้าด้วยกัน แต่เมื่อโลหะสองชิ้นนี้ได้รับเสียงที่มีค่าพลังงานสูงในช่วงความถี่ 500 เฮิรตซ์ เช่นเสียง “เร็กซ์” โลหะจะสั่นและปล่อยสุนัขเด้งออกมาจากตัวบ้าน [24]

รูปที่ 1.6 เรดิโอเร็กซ์

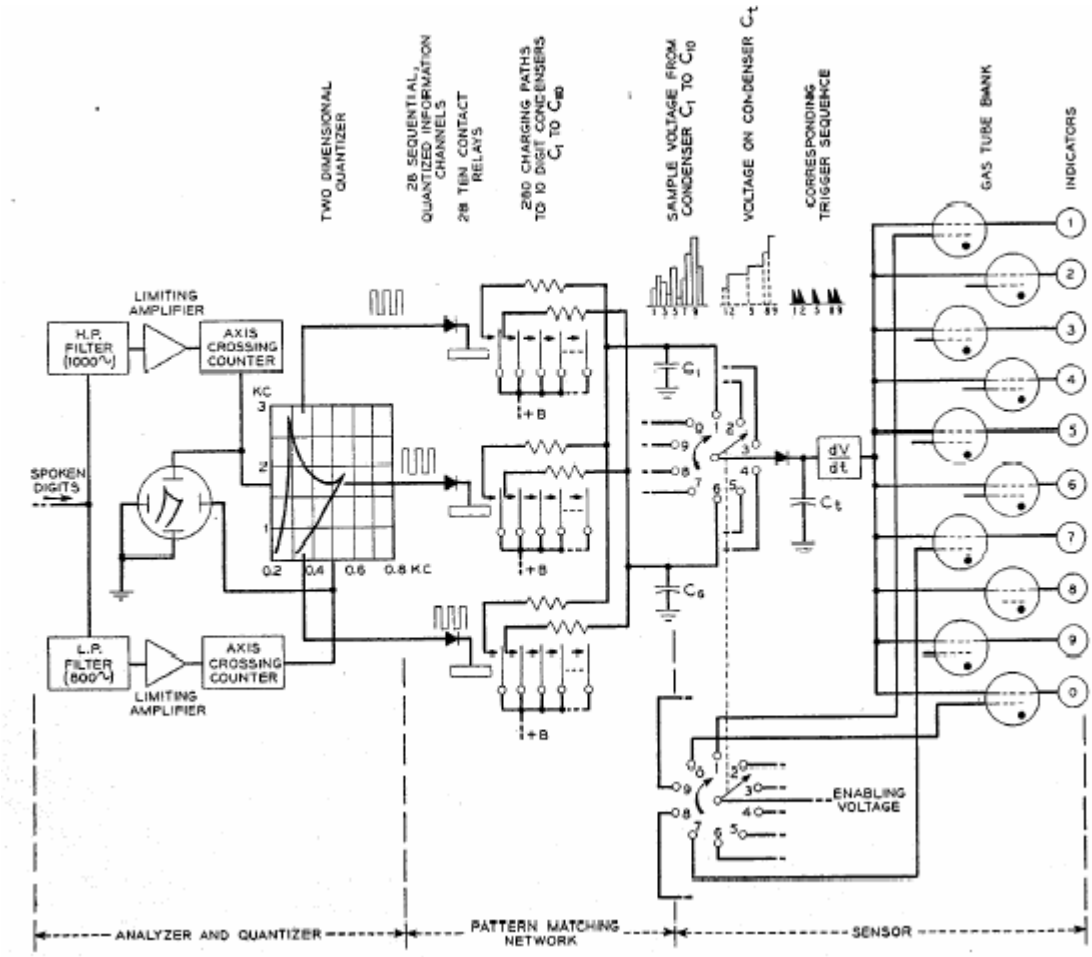
- ค.ศ. 1946 ENIAC ซึ่งเป็นคอมพิวเตอร์ดิจิตอลเครื่องแรกได้ถือกำเนิดขึ้น
- ค.ศ. 1946 ประดิษฐ์ซาวด์สเปคโตรกราฟ (Sound Spectrograph) [25]
- ค.ศ. 1948 กำเนิดทฤษฎีสารสนเทศโดยแชนนอน (Shannon) [26] ซึ่งแบบจำลองช่องรบกวนในทฤษฎีนี้ ภายหลังได้กลายมาเป็นบรรทัดฐานอันสำคัญของเทคโนโลยีทางภาษาและเทคโนโลยีการรู้จำเสียงพูด

1.4.2 ช่วงทศวรรษ1950

เป็นช่วงเริ่มต้นในการพัฒนาระบบรู้จำเสียงพูด ซึ่งมีเหตุการณ์สำคัญเกิดขึ้นดังนี้

- ค.ศ. 1952 ที่ห้องปฏิบัติการวิจัยเบลล์มีการสร้างระบบรู้จำตัวเลขเดี่ยวสำหรับผู้พูดคนเดียว โดยใช้การวิเคราะห์สเปกตรัมของทั้งคำ และประมาณค่าความถี่ฟอร์แมนต์ (Formant frequences) เพื่อนำไปใช้แยกแยะเสียงพูด ถึงแม้เทคนิคที่ใช้เป็นเทคนิคการประมาณที่ค่อนข้างหยาบ และสูญเสียข้อมูลเชิงเวลาไป แต่นี่ถือเป็นจุดเริ่มต้นที่ดี และเป็น

การเปิดแนวทางใหม่ให้วงการการรู้จำเสียงพูด โดยระบบนี้ถูกพัฒนาขึ้นโดยใช้วงจรไฟฟ้าอนาล็อก [27] ดังรูปที่ 1.7



รูปที่ 1.7 ระบบรู้จำเสียงพูดตัวเลขเดี่ยวของเบลล์แล็บส์ในปี ค.ศ. 1952

- **ค.ศ. 1956** ที่ห้องปฏิบัติการวิจัยอาร์ซีเอ โอลสันและเบลาร์ (Olson and Belar) สร้างระบบรู้จำเสียงพูดพยางค์เดี่ยว 10 พยางค์ ของคำพยางค์เดี่ยว 10 คำ สำหรับผู้พูดคนเดียว ระบบนี้ใช้การวัดสเปคตรัมซึ่งผ่านมาจากคลังตัวกรอง (Filter Bank) แบบอนาล็อก [28] [29]
- **ค.ศ. 1958** ดัดลีย์ (Dudley) ใช้วิธีวิเคราะห์สเปคตรัมแบบต่อเนื่องทีละช่วง ซึ่งเป็นวิธีใหม่แทนที่วิธีเดิมที่วิเคราะห์เพียงฟอร์แมนต์หรือคุณลักษณะอื่นๆ ของทั้งเสียง [30]
- **ค.ศ. 1959** ดีนส์ (Denes) ใช้แบบจำลองทางภาษาแบบไบแกรม (Bigram Language Model) เข้ามาช่วยในการรู้จำหน่วยเสียง (Phoneme) นี่เป็นครั้งแรกที่มีการนำข้อมูลอื่นที่ไม่ใช่ข้อมูลทางเสียงเข้ามาใช้เพื่อช่วยในการรู้จำ ซึ่งหน่วยเสียงที่รู้จำได้จะไม่ขึ้นอยู่กับลักษณะของเสียงในหน่วยเสียงนั้นอย่างเดียว หากแต่ขึ้นอยู่กับหน่วยเสียงก่อนหน้าด้วย [31] [32] [33]

1.4.3 ช่วงทศวรรษ1960

ช่วงนี้เป็นช่วงที่เกิดพัฒนาขึ้นอย่างมากมายในการรู้จำเสียงพูด ทั้งทางด้านการสกัดลักษณะสำคัญ (Feature Extraction) และเทคนิคที่ใช้ในการรู้จำ อันได้แก่

- **ค.ศ. 1963** โบเกิร์ต (Bogert) เสนอกระบวนการเซปสตรัม (Cepstral Processing) เพื่อวิเคราะห์แผ่นดินไหว [34] ซึ่งภายหลัง ออพเพนไฮม์ เชฟเฟอร์ และสตอคแฮม (Oppenheim, Schafer and Stockham) นำมาประยุกต์ใช้ในการรู้จำเสียงพูด [35]
- **ค.ศ. 1964** มาร์ติน (Martin) พัฒนาวิธีการจัดเรียงและปรับบรรทัดฐานทางเวลา (Time Alignment and Normalization) [36]
- **ค.ศ. 1965** การแปลงฟูเรียร์แบบเร็ว (Fast Fourier Transform, FFT) โดยคูลีย์และทูคีย์ (Cooley and Tukey) [37]
- **ค.ศ. 1966-72** ความก้าวหน้าทางทฤษฎีฟังก์ชันความน่าจะเป็นของโซ่มาร์คอฟ (Probabilistic Function of Markov Chains) โดยบอม (Baum) และคณะจากไอดีเอ (IDA, Institute for Defense Analysis) ซึ่งภายหลังใช้ชื่อว่าแบบจำลองฮิดเดนมาร์คอฟ (Hidden Markov Models) [38] [39] [40] [41]
- **ค.ศ. 1966** ราช เรดดี (Raj Reddy) ใช้การตามหน่วยเสียงแบบพลวัต (Dynamics Tracking of Phonemes) เพื่อช่วยในการรู้จำ [42] นอกจากนั้น เรดดียังเป็นผู้วางรากฐานให้มหาวิทยาลัยคาร์เนกีเมลลอน (Carnegie Mellon University) เป็นผู้นำในด้านเทคโนโลยีการรู้จำเสียงพูดตลอดมา
- **ค.ศ. 1968** การผ่านเวลาแบบพลวัต (Dynamic Time Warp) ถูกประยุกต์ใช้ครั้งแรกโดย วินต์สยัค (Vintsyuk) [43][44] [44] และซาโกเอะ (Sakoe) [45] การผ่านเวลาแบบพลวัตเป็นการใช้การโปรแกรมแบบพลวัต (Dynamic Programming) ทำการปรับบรรทัดฐานทางเวลาเพื่อนำมาจับคู่รูปแบบ (Pattern Matching)
- **ค.ศ. 1969** รหัสทำนายเชิงเส้น (Linear Predictive Coding, LPC) ถูกนำมาประยุกต์กับการรู้จำเสียงพูดโดย อิตาคุระ (Itakura) [46] อะทัลและโชรเดอร์ (Atal&Shroeder) [47] และมาร์เคล (Markel) [48] รหัสทำนายเชิงเส้นเป็นการจำลองสัญญาณเสียงให้ใกล้เคียงกับสัญญาณเสียงที่ผ่านช่องเสียง (Vocal Tract) มากที่สุด โดยใช้แบบจำลองอัตตาถดถอย (Autoregressive Model) [49]
- **ค.ศ. 1969** จอห์น เพียร์ซ (John Pierce) มีจดหมายวิจารณ์วงการอย่างเจ็บๆ โดยกล่าวว่าการวิจัยสนใจเรื่องกระบวนการวิเคราะห์สัญญาณ (Signal Processing) มากเกินไป ควรหันมามองเรื่องการทำความเข้าใจเสียงพูดมากกว่า นอกจากนี้ความพยายามปรับพารามิเตอร์นิดๆ หน่อยๆ เพื่อให้ประสิทธิภาพในการรู้จำดีขึ้นเป็นความพยายามของนักวิทยาศาสตร์สติเฟื่อง (Mad Scientist) มากกว่านักวิทยาศาสตร์ตัวจริง (Serious Scientist) [50]

1.4.4 ช่วงทศวรรษ1970

ในช่วงนี้ ระบบรู้จำเสียงพูดได้ถูกพัฒนาอย่างต่อเนื่อง โดยมีงบประมาณสนับสนุน ซึ่งมีรายละเอียดดังนี้

- **ค.ศ. 1971-76** โครงงานอาร์พา (ARPA Project) ครั้งแรกได้เริ่มต้นขึ้น โดยอาร์พา (Advanced Research Project Agency, ARPA) ให้เงินจำนวน 15 ล้านดอลลาร์เพื่อพัฒนาระบบรู้จำและทำความเข้าใจเสียงพูด ซึ่งมีเป้าหมายคือให้รู้จำคำพูดต่อเนื่องที่มีผู้พูดจำนวนน้อย มีไวยากรณ์ชัดเจน และมีคำศัพท์ 1000 คำ หน่วยงานวิจัยหลักของโครงงานนี้ได้แก่ บริษัท System Development Corporation มหาวิทยาลัยคาร์เนกีเมลลอน และโบลต์ เบราเนค และนิวแมน (Bolt, Beranek and Newman, BBN) นอกจากนี้ยังมีหน่วยงานอื่น เช่น ห้องปฏิบัติการวิจัยลินคอล์น (Lincoln Labs) เอสอาร์ไอ (SRI) และมหาวิทยาลัยแคลิฟอร์เนียร์เบิร์กเลย์ (University of California Berkeley) ภายหลังแคล็ท (Klatt) [51] รายงานว่าระบบที่บรรลุเป้าหมายของโครงงานคือระบบฮาร์ปี (Harpy) จากมหาวิทยาลัยคาร์เนกีเมลลอน
- **ค.ศ. 1975** มีการประยุกต์แบบจำลองฮิดเดนมาร์คอฟเพื่อใช้ในการรู้จำเสียงพูดกันอย่างแพร่หลาย เช่น จอห์น เฟอร์กูสัน (John Ferguson) ที่ไอดีเอ [52] จิม และเจเน็ต เบเกอร์ (Jim and Janet Baker) ที่มหาวิทยาลัยคาร์เนกีเมลลอน (ซึ่งได้ผลลัพธ์เป็นระบบดรากอน [53]) และเฟร็ด เจลิเนค (Fred Jelinek) ที่ไอบีเอ็ม (IBM) [54] [55] [56] [57] [58] แบบจำลองฮิดเดนมาร์คอฟถือเป็นการเปลี่ยนกระบวนทัศน์ (Paradigm Shift) ของเทคโนโลยีการรู้จำเสียงพูด แม้ในปัจจุบันระบบรู้จำเสียงพูดส่วนใหญ่ยังคงถูกพัฒนาด้วยแบบจำลองฮิดเดนมาร์คอฟ [59]
- **ค.ศ. 1977** เดมป์สเตอร์ แลร์ด และรูบิน (Dempster, Laird and Rubin) ค้นพบรูปแบบทั่วไปของการปรับปรุงบอมเวล์ช (Baum-Welch Update) และให้ชื่อว่าอัลกอริทึมคาดหวัง-ทำให้สูงสุด (Expectation-Maximization Algorithm) [60]

1.4.5 ช่วงทศวรรษ1980

ในระยะนี้ระบบรู้จำเสียงพูดได้ถูกพัฒนาเพื่อให้มีประสิทธิภาพมากขึ้น และสามารถรองรับงานรู้จำเสียงพูดที่ยากและซับซ้อนขึ้น มีการให้ทุนสนับสนุน และการรวบรวมคลังข้อมูล (Corpus) ดังนี้

- **ค.ศ. 1984** อาร์พาให้ทุนโครงงานที่สอง โดยให้ทำการรู้จำงานจัดการทรัพยากร โดยให้รู้จำประโยคคำถามและคำสั่งที่ใช้ในการจัดการฐานข้อมูลกองทัพเรือโดยไม่ขึ้นกับผู้พูด ทำให้เกิดคลังข้อมูลการจัดการทรัพยากร (Resource Management, RM) ขึ้น โดยมีคำศัพท์ 1000 คำ พูดแบบอ่าน และมีผู้พูด 160 คน [61]
- **ค.ศ. 1986** เริ่มรวบรวมคลังข้อมูลทิมิท (TIMIT Corpus) โดยมีเสียงพยัญชนะ 61 เสียง และประโยคที่พูดจะมีเสียงแต่ละพยัญชนะอย่างสมดุล ใช้ผู้พูด 630 คน พูดคนละ 10 ประโยค คลังข้อมูลนี้เป็นคลังข้อมูลแรกที่ใช้กันอย่างแพร่หลาย [62]
- **ค.ศ. 1986-7** โครงงานอาร์พาให้รู้จำเสียงอ่านวารสารวอลสตรีท (Wall Street Journal) และเสียงอ่านค้นข้อมูลในระบบข้อมูลการท่องเที่ยวทางอากาศ (Air Travel Information System, ATIS) ก่อให้เกิดคลังข้อมูลวอลสตรีทที่มีคำศัพท์ 5000 คำ และภายหลังพัฒนา

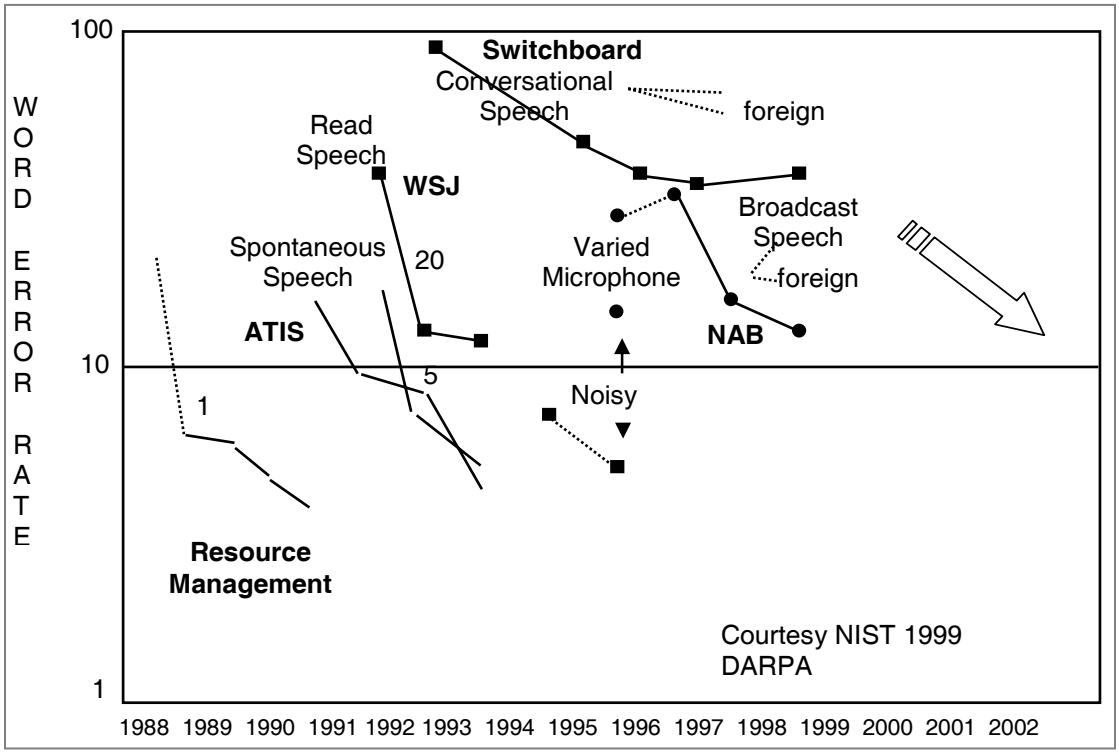
เป็น 20000 คำ ส่วนระบบข้อมูลการท่องเที่ยวทางอากาศนั้น ค่อนข้างจะเป็นการทำความเข้าใจเสียงพูดมากกว่าการรู้จำเสียงพูด เนื่องจากผู้ใช้สามารถพูดประโยคใดก็ได้เพื่อสอบถามหรือจองตั๋วเที่ยวบิน

- **1988-90** ลักษณะสำคัญของเสียงใหม่ๆ ได้ถูกนำมาใช้ เช่น เมลเซปสตรัม (Mel Cepstrum) โดยบริดเดิล (Bridle) [63] พีแอลพี (PLP) โดยเฮอร์มานสกี (Hermansky) [64] และสัมประสิทธิ์เดลตา เดลตา-เดลตา (Delta / Delta-Delta Coefficients) โดยฟุรุอิ (Furui) [65] ซึ่งช่วยเพิ่มประสิทธิภาพในระบบรู้จำที่มีขนาดใหญ่ และมีเสียงรบกวนมาก
- การกลับมาอีกครั้งของโครงข่ายประสาทเทียม (Artificial Neural Networks) ทำให้มีการนำเทคนิคนี้มาใช้ในการรู้จำเสียงพูดอย่างแพร่หลาย [66] [67] [68] ซึ่งในบางระบบได้นำโครงข่ายประสาทเทียมมาประยุกต์ใช้ร่วมกับแบบจำลองฮิดเดนมาร์คอฟ โดยโครงข่ายประสาทเทียมทำหน้าที่รู้จำหน่วยเสียง ขณะที่แบบจำลองฮิดเดนมาร์คอฟทำหน้าที่ค้นหาลำดับคำที่น่าจะเป็นที่สุด [69] [70] [71]
- ระบบฐานความรู้ (Knowledge-Based) ซึ่งเป็นที่นิยมในวงการปัญญาประดิษฐ์ (Artificial Intelligence) ได้ถูกนำมาใช้เพื่อช่วยในการรู้จำเสียงพูด เช่นการนำความรู้ทางด้านเสียง-สัทศาสตร์ (Acoustic-phonetic Knowledge) มาช่วยในการแยกเสียงพยัญชนะ [72][73] [73] ข้อได้เปรียบของวิธีนี้คือลักษณะสำคัญที่ใช้ในการรู้จำเสียงพูดไม่ได้จำกัดอยู่เพียงลักษณะทางเสียงของเฟรมที่สังเกตอยู่เท่านั้น [74]

1.4.6 ยุคปัจจุบัน

- มีความพยายามในการเพิ่มความทนทาน (Robustness) ของระบบ ต่อช่องสัญญาณที่ต่างกัน และต่อเสียงรบกวนต่างๆ [75] 1[76] [77] [78]
- มีความต้องการที่จะพัฒนาระบบให้รู้จำเสียงพูดที่เป็นธรรมชาติมากขึ้น เช่น เสียงการสนทนาทางโทรศัพท์ โดยมีคลังข้อมูลสวิตช์บอร์ด (Switchboard Corpus) ที่เก็บรวบรวมเสียงพูดชนิดนี้ อย่างไรก็ดี นี่เป็นงานที่ยาก และประสิทธิภาพของระบบในการรู้จำเสียงพูดชนิดนี้ยังไม่เป็นที่น่าพอใจนัก

โดยประสิทธิภาพของระบบรู้จำเสียงพูดในแต่ละปีเมื่อทดสอบกับคลังข้อมูลชุดต่างๆ [79] สรุปได้ดังรูปที่ 1.8



รูปที่ 1.8 ประสิทธิภาพของระบบรู้จำเสียงพูดในแต่ละปีเมื่อทดสอบกับคลังข้อมูลชุดต่าง ๆ

ความท้าทายและทิศทางในการวิจัยด้านเทคโนโลยีเสียงพูดในปัจจุบัน [80] สามารถสรุปได้ดังนี้

- **ด้านความทนทาน** (Robustness) เพื่อให้ความถูกต้องของการรู้จำไม่ลดลงมากนักเมื่อข้อมูลที่ส่งมาเกิดความผิดเพี้ยนหรือหายไปเนื่องจากเสียงรบกวนในช่องรบกวน
- **ด้านการเรียนรู้และปรับปรุงตัวเองโดยอัตโนมัติ** (Automatic Training and Adaptation) เพื่อให้ระบบสามารถเรียนรู้และปรับปรุงตัวเองให้เข้ากับโดเมนใหม่ๆ ได้อย่างรวดเร็ว ประหยัด และง่ายดาย
- **ด้านการรู้จำเสียงพูดที่เป็นธรรมชาติ** (Spontaneous Speech) เพื่อให้ระบบสามารถรับรู้สำเนียงการพูด (Prosody) จังหวะการพูด อารมณ์ และพฤติกรรมการพูดชนิดต่างๆ
- **ด้านการสนทนา** (Dialogue Models) เพื่อให้ระบบสามารถเข้าใจบทสนทนาของผู้ใช้
- **ด้านการสร้างภาษาโต้ตอบ** (Natural Language Response Generation) เพื่อให้ระบบสามารถสร้างภาษาโต้ตอบกับผู้ใช้ โดยภาษาที่สร้างขึ้นต้องสอดคล้องและเหมาะสมกับเรื่องที่กำลังสนทนา
- **ด้านการสังเคราะห์และสร้างเสียงพูด** (Speech Synthesis and Generation) เพื่อให้ระบบสามารถสังเคราะห์เสียงพูด และสนทนาโต้ตอบกับผู้ใช้
- **ระบบหลายภาษา** (Multilingual Systems) เพื่อการเข้าถึงข้อมูลข้ามภาษา และการแปลภาษาแบบทันกาลจากเสียงพูด

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

1.5 วัตถุประสงค์ของโครงการ

โครงการนี้มีวัตถุประสงค์เพื่อศึกษาวิจัยระบบการรู้จำเสียงพูดภาษาไทยและได้พัฒนาระบบการรู้จำเสียงพูดอัตโนมัติต้นแบบขึ้นมา 2 ระบบ ดังต่อไปนี้

- ระบบการโอนสายโทรศัพท์อัตโนมัติจากเสียงพูดชื่อไทย (รายละเอียดอยู่ในบทที่ 3)
- ระบบสอบถามรายนามผู้ใช้โทรศัพท์โดยผ่านระบบโทรศัพท์ (รายละเอียดอยู่ในบทที่ 4)

2. ทฤษฎีและแนวคิดที่เกี่ยวข้อง

บทนี้นำเสนอทฤษฎีและแนวคิดที่เกี่ยวข้องกับการพัฒนาระบบรู้จำเสียงพูด โดยเริ่มจากทฤษฎีทางสัทศาสตร์ (Phonetics) ซึ่งเป็นทฤษฎีที่ศึกษาปรากฏการณ์ของเสียงพูดในด้านต่างๆ ดังนี้

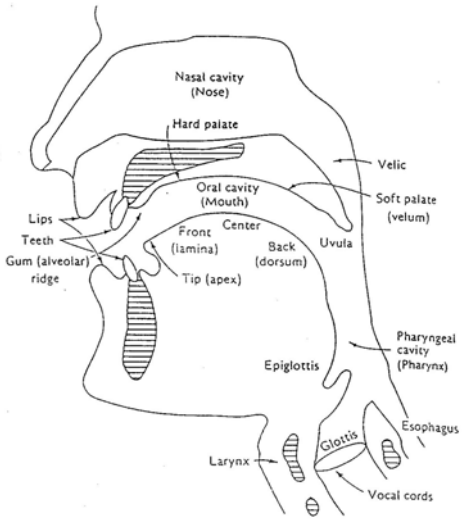
- **สรีรสัทศาสตร์** (Articulatory Phonetics) เป็นการศึกษาเสียงพูดจากอวัยวะ และการเคลื่อนไหวของอวัยวะที่ทำให้เกิดเสียงพูด
- **สวนสัทศาสตร์** (Acoustic Phonetics) เป็นการศึกษาเสียงพูดจากลักษณะของคลื่นเสียงที่ผู้พูดเปล่งออกมา
- **โสตสัทศาสตร์** (Auditory Phonetics) เป็นการศึกษากลไกการรับรู้เสียงพูด

จากนั้นจะนำเสนอทฤษฎีการสกัดลักษณะสำคัญ และเทคนิคการรู้จำที่ใช้ในการพัฒนาระบบ ซึ่งในที่นี้คือการทำนายเชิงเส้นแบบรับรู้ โครงข่ายประสาทเทียม การสังเคราะห์เสียงพูด และทฤษฎีวิภัชนัย ตามลำดับ รวมทั้งวิธีการหาขอบเขตของเสียงพูดที่ใช้

2.1 สรีรสัทศาสตร์

2.1.1 อวัยวะการออกเสียง (Speech Organs) [82]

อวัยวะการออกเสียงทั้งหมดของมนุษย์สามารถแสดงได้ดังรูปที่ 2.1 [81]



รูปที่ 2.1 อวัยวะการออกเสียง

อวัยวะการออกเสียงทั้งหมดมีดังนี้

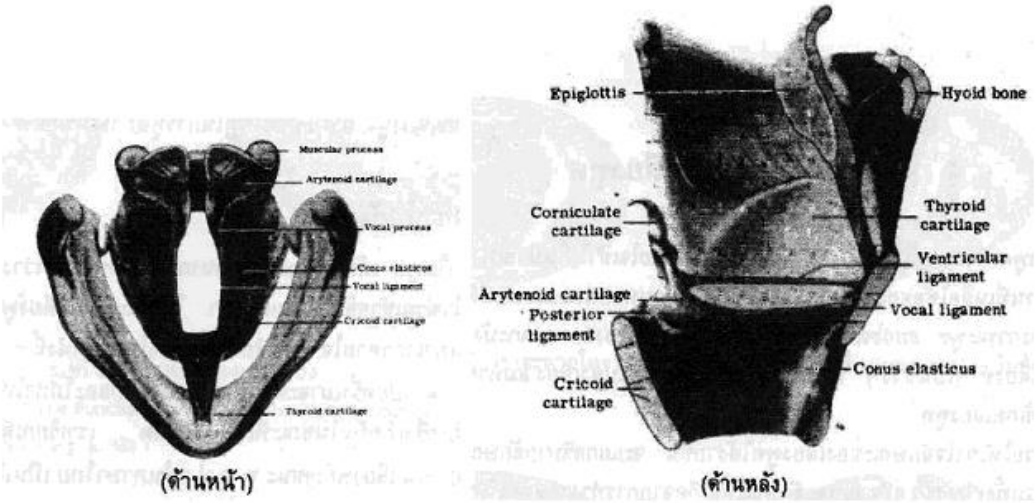
- **ริมฝีปาก** (Lips) เป็นอวัยวะส่วนที่สามารถเคลื่อนไหวได้มากและทำให้เสียงแตกต่างกันได้มาก เราอาจจะบังคับริมฝีปากให้ปิดสนิท ให้เปิดเล็กน้อย ให้เปิดกว้างขึ้น ให้ยื่นออกมา ให้ห่อกลมหรือทำเป็นรูปรีก็ได้ ลักษณะต่างๆ ของริมฝีปากล้วนมีผลต่อการออกเสียงทั้งสิ้น เสียงพยัญชนะที่เกิดจากการกักที่ริมฝีปากเรียกว่าเสียงโอฐชะ (Bilabial Sound)
- **ฟัน** (Teeth) เป็นอวัยวะที่เป็นฐานหรือตำแหน่งที่เกิดของเสียงหลายชนิด เช่น เมื่อฟันบนกดลงบนริมฝีปากล่าง ลมที่ผ่านออกมาโดยแรงจะลอดช่องที่พอจะผ่านได้ออกมา ทำให้เกิด

เป็นเสียงชนิดที่เรียกว่า เสียงเสียดแทรกที่เกิดระหว่างฟันกับริมฝีปาก ถ้าฟันบนกดกับฟันล่าง ลมที่ผ่านออกมาโดยแรงจะทำให้ได้เสียงเสียดแทรกที่เกิดที่ฟัน เป็นต้น นอกจากนี้เนื่องจากปลายลิ้นอยู่ใกล้กับฟัน ปลายลิ้นจึงมักจะทำอาการต่างๆ บริเวณฟันและหลังฟันบ่อยๆ ทำให้เกิดเสียงทันตชะ (Dental Sound)

- **ปุ่มเหงือก** (Alveolus, Gum Ridge, Tooth Ridge) เป็นส่วนที่นูนออกมาตรงบริเวณหลังฟันด้านบน ถ้าเอาลิ้นแตะดูจะรู้สึกว่ามีลักษณะเป็นคลื่น ลิ้นอาจแตะหรือวางอยู่ใกล้บริเวณปุ่มเหงือก ซึ่งทำให้เกิดเสียงมุทธชะ (Alveolar Sound)
- **เพดานแข็ง หรือเพดานปาก** (Palate, Hard Palate) หมายถึงส่วนโค้งของเพดานปากส่วนที่เป็นกระดูกแข็ง ซึ่งอยู่ถัดจากปุ่มเหงือกเข้ามา ถ้าลิ้นแตะหรือวางใกล้เพดานแข็งจะทำให้เกิดเสียงตาลุชะ (Palatal Sound)
- **เพดานอ่อน** (Velum, Soft Palate) คือ ส่วนของเพดานที่อยู่ต่อเพดานแข็งเข้าไปข้างในเป็นกระดูกอ่อนที่ขยับขึ้นลงได้เล็กน้อย เวลาหายใจเพดานอ่อนและลิ้นไก่ซึ่งอยู่ปลายเพดานอ่อนจะลดระดับลงมาเปิดช่องให้ลมออกทางจมูก ฉะนั้นเวลาที่ไม่พูด เพดานอ่อนและลิ้นไก่จะลดระดับลงมา เวลาพูดส่วนใหญ่เพดานอ่อนและลิ้นไก่จะถูกยกขึ้นไปจดกับผนังคอ จะมีแต่เวลาออกเสียงนาสิกเท่านั้นที่เพดานอ่อนจะลดระดับลงมาเพื่อให้ลมออกไปทางจมูกได้ ถ้าลิ้นแตะหรือวางใกล้เพดานอ่อนจะทำให้เกิดเสียงมุทธชะ (Velar Sound)
- **ลิ้นไก่** (Uvula) เป็นก้อนเนื้อเล็กๆ อยู่ต่อจากปลายเพดานอ่อนเข้าไปข้างใน และห้อยอยู่ตรงกลางปาก สามารถสั่นรัวได้ เวลาอ้าปากมักจะเห็น ลิ้นไก่ใช้ออกเสียงในบางภาษาเช่น ภาษาเยอรมัน ฝรั่งเศส นอร์เวย์ อาหรับ และอิสราเอล เป็นต้น
- **ช่องจมูก** (Nasal Cavity) หมายถึง โพรงในช่องจมูก ซึ่งอยู่เหนือลิ้นไก่ขึ้นไป เป็นช่องที่ลมซึ่งผ่านเส้นเสียงขึ้นมาจะผ่านออกไปทางจมูกได้เมื่อเวลาหายใจและเวลาออกเสียงนาสิก ในเวลาเปล่งเสียงอื่นๆ ลิ้นไก่จะถูกยกขึ้นไปปิดช่องจมูกเพื่อให้ลมออกมาทางช่องปาก
- **ลิ้น** (Tongue) เป็นส่วนที่เคลื่อนไหวได้มากที่สุดในการออกเสียงพูด ส่วนที่เคลื่อนไหวของลิ้นแต่ละส่วนมีผลต่อการออกเสียง เราจึงแบ่งลิ้นออกเป็น 3 ส่วนด้วยกันตามหน้าที่ที่มีในการออกเสียงคือ
  - ปลายลิ้น (Tip of the Tongue) หรือ ลิ้นส่วนปลายสุด หมายถึงส่วนปลายของลิ้น ซึ่งสามารถจะยกขึ้นไปแตะอวัยวะส่วนต่างๆ ในปากตอนบนได้โดยง่าย
  - หน้าลิ้น (Blade of the Tongue) หรือ ลิ้นส่วนหน้า หมายถึงลิ้นส่วนที่อยู่ตรงข้ามกับเพดานแข็ง ในขณะที่วางลิ้นราบกับปากตอนไม่ได้พูด
  - หลังลิ้น (Back of the Tongue) หรือ ลิ้นส่วนหลัง หมายถึงส่วนของลิ้นที่อยู่ตรงข้ามกับเพดานอ่อน ในขณะที่วางลิ้นราบกับปากตอนไม่ได้พูด
- **แผ่นเนื้อปากหลอดลม** (Epiglottis) หรือ ลิ้นปิดกล่องเสียง เป็นก้อนเนื้อเล็กๆ คล้ายลิ้นไก่อยู่ต่อโคนลิ้นลงไปในคอ มีหน้าที่ปิดเปิดช่องหลอดลม เพื่อป้องกันมิให้อาหารตกลงไปในหลอดลม ในเวลาที่กลืนอาหาร แผ่นเนื้อปากหลอดลมปิดลงให้อาหารผ่านไปลงหลอดอาหาร แต่ในเวลาที่พูด แผ่นเนื้อนี้จะเปิดออกเพื่อให้ลมจากหลอดลมออกมา
- **โพรงคอ** (Pharynx) เป็นโพรงซึ่งอยู่ถัดปากลงไปจากช่องปากจนถึงเส้นเสียงหรือสายเสียง
- **เส้นเสียง หรือสายเสียง** (Vocal Cords) เป็นอวัยวะสำคัญที่ทำให้เกิดเสียง เส้นเสียงประกอบด้วยเส้นเอ็นและกล้ามเนื้อเป็นแผ่น 2 แผ่น มีความยาวประมาณ 1.2-1.7

เซนติเมตร กว้างประมาณ 0.2-0.3 เซนติเมตร ปิดขวางอยู่ตรงปากของช่องหลอดลม โดยจะวางตัวจากด้านหลังมายังด้านหน้าอยู่ตรงกลางของกล่องเสียง เส้นเสียงทั้งสองสามารถที่จะดึงออกให้ห่างจากกันหรือดึงเข้ามาให้ชิดกันก็ได้ เส้นเสียงเป็นส่วนสำคัญที่ทำให้เกิดเสียงพูด โดยจะเปิดให้ลมผ่านในเวลาหายใจตามปกติ แต่จะอยู่ชิดกันเมื่อมีการเปล่งเสียง

- **กล่องเสียง** (Larynx) ตั้งอยู่ตอนบนของหลอดลมตรงตำแหน่งที่เรียกว่าลูกกระเดือก (Adam's Apple) กล่องเสียงประกอบด้วยกระดูกอ่อนหลายส่วนด้วยกัน ส่วนที่อยู่ด้านหน้า คือ กระดูกอ่อนไทรอยด์ (Thyroid Cartilage) ปลายด้านหนึ่งของเส้นเสียงทั้งสองจะเชื่อมอยู่กับกระดูกอ่อนไทรอยด์นี้และอยู่ชิดกัน ส่วนปลายอีกด้านหนึ่งของเส้นเสียงทั้งสอง จะเชื่อมอยู่กับกระดูกอ่อนอาริตินอยด์ (Arytenoids Cartilages) ซึ่งเป็นกระดูกอ่อนอีกสองชิ้น กระดูกอ่อนอาริตินอยด์และกล้ามเนื้อในกล่องเสียงจะทำให้เส้นเสียงทั้งสองอยู่ชิดติดกันหรือห่างจากกันได้ เมื่อเส้นเสียงอยู่ห่างจากกันจะเกิดเป็นช่องสามเหลี่ยม ซึ่งเป็นทางให้ลมผ่านเข้าไปถึงปอด หรือผ่านออกมาจากปอดได้ ดังรูปที่ 2.2 [82]



รูปที่ 2.2 กล่องเสียง

- **ช่องระหว่างเส้นเสียง** (Glottis) จะเปิดอยู่ระหว่างที่หายใจเข้าออกตามปกติ แต่จะปิดลงเมื่อมีการเปล่งเสียง ก่อให้เกิดการสั่น และเป็นเสียงดังขึ้น
- **ช่องปาก** (Oral Cavity) ทำหน้าที่เป็นช่องกำทอน (Resonant Chamber) ซึ่งสามารถเปลี่ยนให้มีรูปร่างต่างๆ กัน ตามท่าทางของอวัยวะภายในช่องปาก โดยอวัยวะภายในช่องปากอาจสามารถแบ่งได้ดังนี้
  - อวัยวะส่วนกระทำอาการ (Articulator) หมายถึงอวัยวะส่วนที่เคลื่อนไหวเพื่อผลักหรือกักลมในที่ต่างๆ อวัยวะส่วนกระทำอาการที่สำคัญคือลิ้น ซึ่งเคลื่อนไหวได้มากที่สุด อวัยวะส่วนกระทำอาการอาจเรียกว่า "กรณ์"
  - อวัยวะส่วนเกิดอาการ (Point of Articulation) หมายถึง ตำแหน่งที่อวัยวะส่วนกระทำอาการเคลื่อนไหวไป เพื่อผลักหรือกักลมไว้ อาจเรียกอวัยวะส่วนนี้ว่า "ฐาน" ที่เกิดของหน่วยเสียงต่างๆ ฐานภายในช่องปากที่สำคัญได้แก่ ริมฝีปาก ฟัน ปุ่มเหงือก เพดานแข็ง และเพดานอ่อน
- **หลอดลม** (Trachea) เป็นทางเดินอากาศจากปอดถึงกล่องเสียง

2.1.2 เสียงพยัญชนะในภาษาไทย (Thai Consonants) [82]

<u>เสียงพยัญชนะ</u>

เสียงพยัญชนะ (Consonant) หมายถึงเสียงของลมที่ผ่านปอดขึ้นมายังกล่องเสียงแล้วปะทะกับอวัยวะต่างๆ ในช่องปาก ทำให้ลมเพียงส่วนหนึ่งหรือทั้งหมดพบกับอุปสรรคที่อยู่เหนือช่องของเส้นเสียง โดยอุปสรรคเหล่านี้เกิดจากการทำงานประสานกันของอวัยวะในช่องปาก เสียงพยัญชนะที่เกิดขึ้นมาจึงมีหลายแบบแตกต่างกัน ซึ่งเสียงที่แตกต่างกันมักจะทำให้ความหมายของเสียงในภาษาแตกต่างกันไปด้วย คุณสมบัติที่ทำให้เสียงพยัญชนะแตกต่างกันมีดังนี้

- **คุณสมบัติความก้องของเสียง** ความก้องของเสียงเป็นคุณสมบัติที่ใช้ในการแบ่งแยกเสียงพยัญชนะออกได้เป็นสองชนิด คือ
  - เสียงพยัญชนะโฆษะ (Voiced) หรือเสียงก้อง เป็นเสียงพยัญชนะที่เส้นเสียงสั่นสะเทือนขณะที่เปล่งเสียง
  - เสียงพยัญชนะอโฆษะ (Voiceless) หรือเสียงไม่ก้อง เป็นเสียงพยัญชนะที่เส้นเสียงไม่สั่นสะเทือนขณะที่เปล่งเสียง
- **ลักษณะของลมที่ผ่านเส้นเสียง** เสียงพยัญชนะสามารถแบ่งตามลักษณะลมที่ผ่านเส้นเสียงออกมาได้ดังนี้
  - เสียงพยัญชนะหยุด (Stop) อาจแบ่งออกเป็น 2 ลักษณะย่อยๆ ได้แก่ เสียงพยัญชนะระเบิด (Plosive Stop) และเสียงพยัญชนะกัก (Unreleased Stop) เสียงพยัญชนะระเบิดเกิดจากการที่ลมซึ่งเปล่งออกมาถูกกักเอาไว้ ณ ที่ใดที่หนึ่งในช่องปาก แล้วช่องที่กักนั้นเปิดให้ลมพุ่งออกมา เสียงพยัญชนะระเบิดแบ่งออกได้อีกเป็นเสียงพยัญชนะระเบิดมีลม (Aspirated Plosive) หรือธนิต ซึ่งจะมีลมหายใจพ่นออกมาหลังเปล่งเสียง และเสียงพยัญชนะระเบิดไม่มีลม (Unaspirated Plosive) หรือสิถิล ซึ่งไม่มีลมหายใจพ่นออกมา ส่วนเสียงพยัญชนะกักเกิดจากลมซึ่งเปล่งออกมาถูกกักไว้ ณ ที่ใดที่หนึ่งในช่องปาก โดยเสียงพยัญชนะกักนี้มักจะเป็นเสียงตัวสะกดท้ายพยางค์
  - เสียงพยัญชนะเสียดแทรก (Fricative) เป็นเสียงพยัญชนะที่เมื่อออกเสียงแล้วลมที่ผ่านขึ้นมาถูกบังคับให้ต้องบีบตัวผ่านช่องแคบๆ ที่ใดที่หนึ่งในช่องปาก ซึ่งเสียงเสียดแทรกนี้เราจะทำค้างไว้นานเท่าใดก็ได้ ตราบเท่าที่ลมหายใจจะอำนวย
  - เสียงพยัญชนะนาสิก (Nasal) เป็นเสียงพยัญชนะที่มีลมผ่านออกมาทางจมูก ซึ่งเกิดจากการที่ลมมากักอยู่ในช่องปาก แล้วเพดานอ่อนและลิ้นไก่ลดระดับลง
  - เสียงพยัญชนะข้างลิ้น (Lateral) เป็นเสียงที่เกิดจากการนำลิ้นปิดบริเวณปุ่มเหงือกและเพดานแข็งส่วนกลางไว้ แล้วปล่อยให้ลมผ่านออกมาทางข้างลิ้น
  - เสียงพยัญชนะรัว (Trill) เกิดจากการที่อวัยวะส่วนใดส่วนหนึ่งในช่องปากกระทบกับอวัยวะอีกส่วนหนึ่งในขณะที่ลมถูกพ่นผ่านอวัยวะนั้นออกมาอย่างรุนแรง
  - เสียงพยัญชนะกึ่งสระ (Semi-vowel, Approximant) หรืออรรธสระ เป็นเสียงเลื่อน (Glide) ที่เกิดขึ้นระหว่างเสียงสระสองเสียง ในการเปล่งเสียงพยัญชนะกึ่งสระ อวัยวะที่ใช้ในการออกเสียงจะอยู่ในตำแหน่งของการออกเสียงสระใดสระหนึ่งก่อน แล้วจึงเปล่งเสียงออกมาขณะที่เปลี่ยนตำแหน่งอวัยวะไปสู่การออกเสียงของอีกสระหนึ่ง

- **ฐานที่เกิดของเสียง** ไม่ว่าลมที่ใช้ในการออกเสียงพยัญชนะนั้นจะมาถูกกัก ถูกดัดแปลง จนเกิดการกัก หรือการเสียดแทรก จำเป็นต้องมีตำแหน่งที่เกิดอยู่ด้วยเสมอในช่องปาก

เสียงพยัญชนะภาษาไทย

พยัญชนะในภาษาไทยมีทั้งหมด 44 รูป 21 หน่วยเสียง แบ่งเป็น 2 กลุ่มใหญ่ๆ คือ พยัญชนะกัก (Stop Consonants) 11 หน่วยเสียง และพยัญชนะไม่กัก (Non-stop Consonants) 10 หน่วยเสียง ดังแสดงในตารางที่ 2.1 ทั้งนี้หน่วยเสียงพยัญชนะทั้ง 21 หน่วยเสียง สามารถที่จะอยู่ในต้นพยางค์ได้ทุกหน่วยเสียง แต่จะมีหน่วยเสียงพยัญชนะที่ปรากฏท้ายพยางค์ได้เพียง 9 หน่วยเสียงเท่านั้น คือ เสียงพยัญชนะกัก 4 หน่วยเสียง (/p/, /t/, /k/, /ʔ/) เสียงพยัญชนะนาสิก 3 หน่วยเสียง (/m/, /n/, /ŋ/) และเสียงพยัญชนะกึ่งสระ 2 หน่วยเสียง (/w/, /j/) ส่วนพยัญชนะต้นควบกล้ำในภาษาไทยแท้เป็นได้ 11 หน่วยเสียง คือ /pr/, /pʰr/, /pl/, /pʰl/, /tr/, /kr/, /kʰr/, /kl/, /kʰl/, /kw/, /kʰw/ ส่วนพยัญชนะต้นควบกล้ำในภาษาไทยทับศัพท์อังกฤษมีได้ 6 หน่วยเสียง คือ /br/, /bl/, /dr/, /fr/, /fl/, /tr/ ส่วนคำไทยที่ยืมมาจากภาษาสันสกฤตก็ควบ /tr/ ได้เช่นกัน

ตารางที่ 2.1 เสียงพยัญชนะภาษาไทย

[1] (*) ปรากฏท้ายพยางค์ได้

[2] (/…/) ปรากฏเฉพาะในคำไทยทับศัพท์อังกฤษ

[2] [/…/] ปรากฏในคำไทยทับศัพท์อังกฤษ หรือคำไทยที่ยืมมาจากภาษาสันสกฤต

หน่วยเสียง[1]	หน่วยเสียงควบกล้ำ[2]	ลักษณะของลม	การพ่นลม	ความก้อง	ฐานที่เกิด	รูปพยัญชนะ
/p/ (*)	/pr/, /pl/	กัก	ไม่พ่นลม	ไม่ก้อง	ริมฝีปาก	ป
/pʰ/	/pʰr/, /pʰl/	กัก	พ่นลม	ไม่ก้อง	ริมฝีปาก	ผ พ ภ
/b/	(/br/), (/bl/)	กัก	ไม่พ่นลม	ก้อง	ริมฝีปาก	บ
/t/ (*)	/tr/	กัก	ไม่พ่นลม	ไม่ก้อง	ฟัน หรือ ปุ่มเหงือก	ฏ ต
/tʰ/	[/tʰr/]	กัก	พ่นลม	ไม่ก้อง	ฟัน หรือ ปุ่มเหงือก	ฐ ฑ ฒ ถ ท ธ
/d/	(/dr/)	กัก	ไม่พ่นลม	ก้อง	ฟัน หรือ ปุ่มเหงือก	ฎ ฑ ด
/c/		กัก	ไม่พ่นลม	ไม่ก้อง	เพดานแข็ง	จ
/cʰ/		กัก	พ่นลม	ไม่ก้อง	เพดานแข็ง	ฉ ช ฌ
/k/ (*)	/kr/, /kl/, /kw/	กัก	ไม่พ่นลม	ไม่ก้อง	เพดานอ่อน	ก
/kʰ/	/kʰr/, /kʰl/, /kʰw/	กัก	พ่นลม	ไม่ก้อง	เพดานอ่อน	ข ฃ ค ฅ ฆ
/ʔ/ (*)		กัก	ไม่พ่นลม	ไม่ก้อง	เส้นเสียง	อ
/m/ (*)		นาสิก		ก้อง	ริมฝีปาก	ม
/n/ (*)		นาสิก		ก้อง	ฟัน หรือ ปุ่มเหงือก	ณ น
/ŋ/ (*)		นาสิก		ก้อง	เพดานอ่อน	ง
/f/	(/fr/), (/fl/)	เสียดแทรก		ไม่ก้อง	ริมฝีปาก	ฝ ฟ
/s/		เสียดแทรก		ไม่ก้อง	ฟัน หรือ ปุ่มเหงือก	ซ ศ ษ ส
/h/		เสียดแทรก		ไม่ก้อง	เส้นเสียง	ห ฮ
/r/		รัว		ก้อง	ฟัน หรือ ปุ่มเหงือก	ร
/l/		ข้างลิ้น		ก้อง	ฟัน หรือ ปุ่มเหงือก	ล ฬ
/w/ (*)		กึ่งสระ		ก้อง	ริมฝีปาก – เพดานอ่อน	ว
/j/ (*)		กึ่งสระ		ก้อง	เพดานแข็ง	ญ ย

2.1.3 เสียงสระในภาษาไทย (Thai Vowels) [82]

<u>เสียงสระ</u>

เสียงสระ (Vowel) เป็นเสียงซึ่งถูกเปล่งออกมาทางช่องปากหรือช่องจมูกโดยไม่มีอวัยวะส่วนใดในปากมาเป็นอุปสรรคปิดกั้นทางลมไว้เลย เสียงสระเกิดจากการที่ลมผ่านเส้นเสียงในตำแหน่งที่เส้นเสียงทั้งสองอยู่ชิดกันมากจนเกือบปิดสนิท ทำให้ลมต้องดันตัวออกมาอย่างรุนแรงจนเส้นเสียงเกิดการสั่นสะเทือน และส่งผลทำให้เกิดเสียงดังที่เป็นเสียงก้อง โดยคุณสมบัติที่ทำให้เสียงสระมีความแตกต่างกันมีดังนี้

- **ส่วนของลิ้นที่ใช้ในการเปล่งเสียง** (Place of Articulation) จากการศึกษาภาพถ่ายเอ็กซเรย์ช่องปากมนุษย์ในขณะที่ออกเสียงสระต่างๆ พบว่ามีลิ้นหลายส่วนที่ใช้ในการออกเสียงสระ ไม่ว่าจะเป็นลิ้นส่วนหน้า ลิ้นส่วนกลาง หรือลิ้นส่วนหลัง โดยลิ้นส่วนนั้นๆ จะยกขึ้นใกล้เพดานปากในขณะที่ออกเสียงสระหนึ่งๆ ก่อให้เกิดเสียงสระที่แตกต่างกัน โดยถ้าลิ้นส่วนหน้ายกขึ้นให้จุดสูงสุดอยู่ใกล้เพดานแข็ง เราก็จะเรียกเสียงสระนั้นว่าเสียงสระส่วนเพดานแข็ง หรือสระหน้า (Front Vowel) เช่น สระอี สระเอ สระแอ เป็นต้น แต่ถ้าการออกสียงสระใดใช้ลิ้นส่วนหลัง โดยทำการยกลิ้นส่วนหลังขึ้นให้จุดสูงสุดอยู่ใกล้เพดานอ่อน เราก็จะเรียกเสียงสระนั้นว่าเป็นเสียงสระส่วนเพดานอ่อน หรือสระหลัง (back vowel) เช่น สระอู สระโอ สระออ เป็นต้น ส่วนถ้าในการออกเสียงสระใดลิ้นส่วนกลางถูกยกขึ้นไปยังส่วนกลางของเพดานปาก เราก็จะเรียกเสียงสระนั้นว่า สระกลาง (Central Vowel) เช่น สระอือ สระเออ สระอา เป็นต้น
- **ระยะห่างระหว่างลิ้นและเพดานปาก หรือความสูงของลิ้น** (Degree of Stricture) ระยะห่างระหว่างลิ้นและเพดานปากเป็นลักษณะที่สำคัญอย่างหนึ่งในการแบ่งชนิดของเสียงสระ โดยระยะห่างนี้จะเป็นตัวกำหนดว่าเสียงสระที่เปล่งออกมาเป็นสระเปิดหรือสระปิด ถ้าหากลิ้นอยู่ห่างจากเพดานปากมาก หรือลิ้นอยู่ในระดับต่ำ ทำให้ช่องโพรงปากกว้าง ลมก็จะผ่านออกมาได้มาก เสียงสระที่ได้จะเป็นสระเปิด (Open Vowel) เช่น สระอา ในทางตรงกันข้าม ถ้าตำแหน่งของลิ้นอยู่ใกล้กับเพดานปากมาก หรือลิ้นอยู่ในระดับสูง ช่องโพรงในปากก็จะแคบ ทำให้ลมผ่านออกมาได้น้อย เสียงสระที่ได้จะเป็นสระปิด (Close Vowel) เช่น สระอี สระอู เป็นต้น แต่ถ้าระยะห่างระหว่างลิ้นกับเพดานปากอยู่ในระหว่างสระเปิดและสระปิด เช่นเสียงสระที่เปิดกว้างกว่าสระปิดเล็กน้อย เราก็จะเรียกว่าเป็นสระกลางปิด หรือสระกึ่งปิด (Close-mid, Half-close Vowel) เช่น สระเอ สระโอ เป็นต้น แต่ถ้าเปิดกว้างขึ้นอีก จะเรียกว่าเป็นสระกลางเปิด หรือสระกึ่งเปิด (Open-mid, Half-open Vowel) เช่น สระ แอ สระออ เป็นต้น
- **การห่อริมฝีปาก** (Labialization) หมายถึงการที่ริมฝีปากทั้งสองเคลื่อนไหวโดยยื่นตัวไปข้างหน้า แล้วห่อกลมมากน้อยเพียงใด ถ้าริมฝีปากยื่นออกไปข้างหน้าแล้วห่อกลมมาก เสียงสระที่ได้จะเรียกว่าสระกลม (Rounded Vowel) เช่น สระอู สระโอ สระออ เป็นต้น แต่ถ้าริมปากทั้งสองฉีกออกหรือไม่ห่อกลมขณะเปล่งเสียง สระที่ได้ก็จะเป็นสระไม่กลม (Unrounded Vowel) เช่น สระอี สระเอ สระแอ สระอา เป็นต้น

- **ลักษณะนาสิก** (Nasalization) เป็นลักษณะในการออกเสียงสระที่ทำให้เกิดเสียงสระขึ้นจมูกหรือสระนาสิก (Nasal Vowel) ขึ้น ซึ่งจะทำให้เสียงแตกต่างจากสระโอฐชะ (Oral Vowel) กล่าวคือ ในการเปล่งเสียงสระโอฐชะนั้น เพดานอ่อนจะยกขึ้นปิดโพรงจมูก อากาศจึงไม่สามารถผ่านออกไปทางช่องจมูกได้ แต่ออกมาทางปากทั้งหมด สำหรับสระนาสิกนั้น เพดานอ่อนจะลดต่ำลง และปล่อยให้อากาศผ่านออกทางช่องจมูกด้วยในเวลาเดียวกัน โดยในการเขียนสัทอักษร เราจะใช้เครื่องหมาย ~ (Tilde) กำกับอยู่เหนือสระที่ออกเสียงแบบสระนาสิก เช่นในภาษาฝรั่งเศสจะมีหน่วยเสียงนาสิกอยู่ 4 หน่วยเสียงด้วยกัน แต่สำหรับภาษาไทย ตามปกติแล้วไม่มีการออกเสียงสระนาสิก แต่ในบางครั้งก็อาจได้รับอิทธิพลจากการเปล่งเสียงพยัญชนะนาสิกที่อยู่ใกล้เคียง เช่น คำว่า นั้น เป็นต้น
- **ความยาวในการออกเสียง** (Duration) ความสั้นยาวของการออกเสียงนั้นมีความสำคัญมากในภาษาไทย เพราะหน่วยเสียงที่ใช้ความยาวในการออกเสียงต่างกันจะทำให้ความหมายของพยางค์แตกต่างกันได้ เช่นคำว่า ขุด และคำว่า ขูด โดยสระจะถูกแบ่งออกเป็นสองประเภทตามความสั้นยาวของสระ คือ สระเสียงสั้น (รัสสระ) และสระเสียงยาว (ฑีฆสระ) โดยในการเขียนสัทอักษร เราจะใช้สัญลักษณ์ : (Length Mark) เพื่อแสดงว่าเสียงสระนั้นถูกยืดออกไป

<u>เสียงสระภาษาไทย</u>

สระในภาษาไทยตามไวยากรณ์ดั้งเดิม [83] มีทั้งหมด 21 รูป 32 หน่วยเสียง แบ่งเป็น 3 กลุ่มใหญ่ๆ คือ

- **สระเดี่ยว** (Monophthongs) เป็นสระเสียงแท้ ซึ่งการออกเสียงสระตั้งแต่เริ่มต้นจนสิ้นสุดไม่มีการเปลี่ยนรูปร่างของลิ้นและช่องปาก สระเดี่ยวในภาษาไทยมีทั้งสิ้น 18 หน่วยเสียง เป็นสระเสียงสั้น 9 หน่วยเสียง และสระเสียงยาว 9 หน่วยเสียง
- **สระประสม** (Diphthongs) เป็นสระที่เกิดจากการออกเสียงผสมกันของสระแท้ โดยลิ้นและช่องปากจะเปลี่ยนจากรูปร่างการออกเสียงของสระหนึ่งไปยังอีกสระหนึ่งอย่างค่อนข้างกลมกลืนและรวดเร็ว สระประสมในภาษาไทยมีทั้งสิ้น 6 หน่วยเสียง เป็นสระเสียงสั้น 3 หน่วยเสียง และสระเสียงยาว 3 หน่วยเสียง
- **สระเกิน** (Vowel Letter) ในภาษาไทยมีรูปสระที่เกิดจากการรวมของเสียงสระกับตัวสะกดหรือคำควบเข้าไว้ด้วยกัน ซึ่งมีทั้งหมด 8 หน่วยเสียง

  เสียงสระในภาษาไทยสามารถสรุปได้ดังตารางที่ 2.2

ตารางที่ 2.2 เสียงสระภาษาไทย

หน่วยเสียง	ส่วนของลิ้นที่ใช้เปล่งเสียง	ความสูงของลิ้น	การห่อริมฝีปาก	ความยาวเสียง	รูปสระ
สระเดี่ยว					
/i/	หน้า	ปิด	ไม่ห่อ	สั้น	อิ
/i:/	หน้า	ปิด	ไม่ห่อ	ยาว	อี
/e/	หน้า	กึ่งปิด	ไม่ห่อ	สั้น	เอะ
/e:/	หน้า	กึ่งปิด	ไม่ห่อ	ยาว	เอ
/ɛ/	หน้า	กึ่งเปิด	ไม่ห่อ	สั้น	แอะ
/ɛ:/	หน้า	กึ่งเปิด	ไม่ห่อ	ยาว	แอ
/ɯ/	หลัง ค่อนมาทางกลาง	ปิด	ไม่ห่อ	สั้น	อึ
/ɯ:/	หลัง ค่อนมาทางกลาง	ปิด	ไม่ห่อ	ยาว	อือ
/ɤ/	หลัง ค่อนมาทางกลาง	กึ่งปิด	ไม่ห่อ	สั้น	เออะ
/ɤ:/	หลัง ค่อนมาทางกลาง	กึ่งปิด	ไม่ห่อ	ยาว	เออ
/a/	กลาง	Open	ไม่ห่อ	สั้น	อะ
/a:/	กลาง	Open	ไม่ห่อ	ยาว	อา
/u/	หลัง	ปิด	ห่อ	สั้น	อุ
/u:/	หลัง	ปิด	ห่อ	ยาว	อู
/o/	หลัง	กึ่งปิด	ห่อ	สั้น	โอะ
/o:/	หลัง	กึ่งปิด	ห่อ	ยาว	โอ
/ɔ/	หลัง	กึ่งเปิด	ห่อ	สั้น	เอาะ
/ɔ:/	หลัง	กึ่งเปิด	ห่อ	ยาว	ออ
สระประสม	ส่วนประกอบ				
/ia/	/i/ + /a/			สั้น	เอียะ
/i:a/	/i:/ + /a/			ยาว	เอีย
/ɯa/	/ɯ/ + /a/			สั้น	เอือะ
/ɯ:a/	/ɯ:/ + /a/			ยาว	เอือ
/ua/	/u/ + /a/			สั้น	อัวะ
/u:a/	/u:/ + /a/			ยาว	อัว
สระเกิน	ส่วนประกอบ				
/am/	/a/ + /m/			สั้น	อำ
/aj/	/a/ + /j/			สั้น	ไอ ใอ
/aw/	/a/ + /w/			สั้น	เอา
/ri/, /rɯ/	/r/ + /i/ , /r/ + /ɯ/			สั้น	ฤ
/ri:/, /rɯ:/	/r/ + /i:/ , /r/ + /ɯ:/			ยาว	ฤา
/li/, /lɯ/	/l/ + /i/ , /l/ + /ɯ/			สั้น	ฦ
/li:/, /lɯ:/	/l/ + /i:/ , /l/ + /ɯ:/			ยาว	ฦา

2.1.4 เสียงวรรณยุกต์ในภาษาไทย (Thai Tones) [82]

เสียงวรรณยุกต์นั้นคือเสียงสูงต่ำในภาษา ซึ่งเกิดจากการสั่นสะเทือนของเส้นเสียงในอัตราความถี่ที่ต่างกันไป ดังนั้นเสียงวรรณยุกต์จะปรากฏอยู่ในส่วนของเสียงสระ เพราะเสียงสระเป็นเสียงที่เกิดจากการสั่นของเส้นสียง นอกจากนี้ยังอาจมีเสียงวรรณยุกต์ปรากฏอยู่บ้างในบางส่วนของเสียงพยัญชนะ แต่จะต้องเป็นส่วนของเสียงพยัญชนะที่เป็นเสียงก้องหรือพยัญชนะนาสิกเท่านั้น เพราะเสียงพยัญชนะไม่ก้องนั้นไม่ได้เกิดจากการสั่นของเส้นเสียง จึงไม่สามารถมีเสียงวรรณยุกต์อยู่ด้วยได้

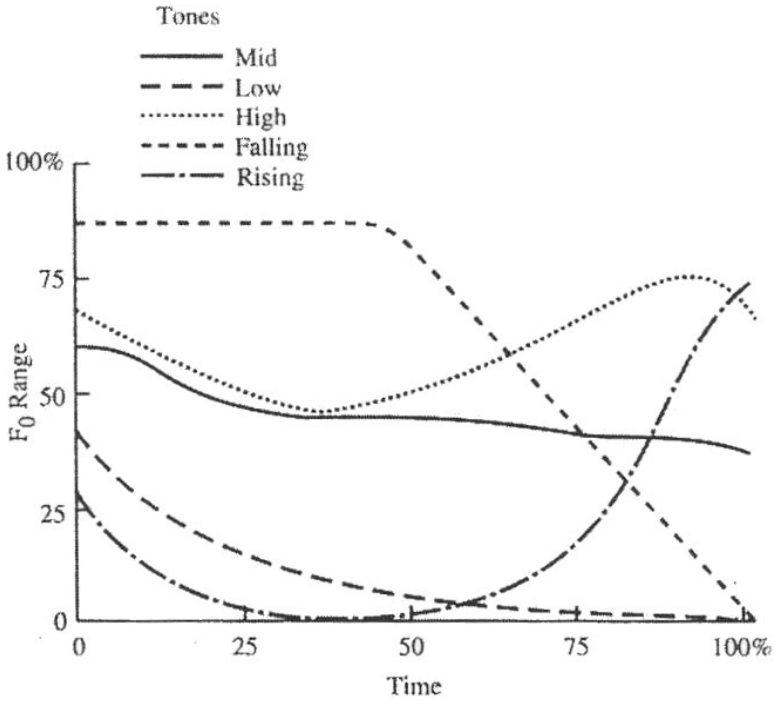
สำหรับในภาษาไทย วรรณยุกต์นั้นถือได้ว่าเป็นหน่วยเสียงที่สำคัญ เพราะสามารถใช้แยกแยะความแตกต่างทางความหมายของคำในภาษาไทยได้ ตรงกันข้ามกับบางภาษา เช่น ภาษาอังกฤษ ซึ่งไม่จัดว่าเสียงวรรณยุกต์เป็นหน่วยเสียงในภาษา เพราะไม่ว่าเราจะพูดภาษาอังกฤษด้วยเสียงสูงต่ำอย่างไร ผู้ฟังก็สามารถเข้าใจได้เหมือนกัน แต่ก็อาจจะต้องมีการใช้เสียงวรรณยุกต์ประกอบบ้าง ทั้งนี้เพื่อสื่ออารมณ์ของผู้พูดเท่านั้น ภาษาไทยจึงจัดได้ว่าเป็นภาษามีวรรณยุกต์ (Tonal Language) เสียงวรรณยุกต์ภาษาไทยสามารถแบ่งออกเป็น 2 ชนิดใหญ่ๆ คือ

- **เสียงวรรณยุกต์ระดับ** (Level Tone) เป็นเสียงวรรณยุกต์ที่มีระดับความถี่ค่อนข้างคงที่ตลอดพยางค์ ถึงแม้ว่าในการออกเสียงพูดโดยปกตินั้น เสียงต้นพยางค์มักจะไม่ได้มีความถี่และความดังเท่ากันกับเสียงท้ายพยางค์ โดยเสียงต้นพยางค์มักมีมีระดับความถี่สูงกว่าและดังกว่าเสียงท้ายพยางค์ แต่ในทางสัทศาสตร์แล้ว ความถี่ที่ต่างกันหรือการเปลี่ยนแปลงของระดับเสียงนี้ถือว่าเล็กน้อยมาก เมื่อเทียบกับการเปลี่ยนระดับความถี่ของเสียงในพยางค์อีกจำพวกหนึ่งซึ่งจะได้กล่าวต่อไป สำหรับเสียงวรรณยุกต์ระดับในภาษาไทยนั้น มีอยู่ด้วยกัน 3 เสียงดังนี้คือ
  - เสียงวรรณยุกต์สามัญ (Mid Tone) เสียงวรรณยุกต์นี้มีระดับความถี่ปานกลางประมาณ 120 เฮิรตซ์ และคงที่อยู่ที่ระดับนั้นจนกระทั่งปลายพยางค์ จึงจะลดต่ำลงมาจนเกือบถึงประมาณ 110 เฮิรตซ์ เสียงวรรณยุกต์สามัญนี้จะไม่ปรากฏในพยางค์ที่มีพยัญชนะกักเป็นพยัญชนะท้าย หรือที่เรียกกันว่าคำตาย
  - เสียงวรรณยุกต์เอก (Low Tone) เสียงวรรณยุกต์นี้มีระดับความถี่ต้นเสียงปานกลางประมาณ 120 เฮิรตซ์ แล้วลดต่ำลงมาเหลือประมาณ 100 เฮิรตซ์อย่างรวดเร็ว และคงที่อยู่ในระดับนี้ สำหรับเสียงวรรณยุกต์เอกจะปรากฏกับพยางค์ได้ทุกรูปแบบ ทั้งคำเป็นและคำตาย
  - เสียงวรรณยุกต์ตรี (High Tone) เสียงวรรณยุกต์นี้มีระดับความถี่ค่อนข้างสูง โดยจะค่อยๆ สูงขึ้นทีละน้อยจากต้นพยางค์ซึ่งมีความถี่ประมาณ 125 เฮิรตซ์ ไปจนถึงประมาณ 135 – 140 เฮิรตซ์เมื่อสิ้นพยางค์ หรืออาจจะลดต่ำลงตอนปลายพยางค์มาอยู่ที่ประมาณ 130 เฮิรตซ์ก็ได้ ขึ้นอยู่กับว่าพยางค์นั้นๆ จบลงด้วยเสียงประเภทใด ถ้าพยางค์นั้นคำเป็น ระดับของเสียงตอนปลายของพยางค์จะไม่ลดต่ำลงมา แต่ถ้าพยางค์นั้นเป็นคำตาย ระดับเสียงตอนปลายจะลดต่ำลงอย่างรวดเร็ว
- **เสียงวรรณยุกต์เปลี่ยนระดับ** (Contour Tone) เป็นเสียงวรรณยุกต์ที่มีระดับความถี่ของการออกเสียงเปลี่ยนแปลงมากในช่วงพยางค์หนึ่งๆ เช่น ต้นพยางค์ออกเสียงให้มีระดับสูงแล้วลดระดับเสียงลงอย่างรวดเร็วไปสู่ระดับต่ำที่ท้ายพยางค์ หรือต้นพยางค์ออกเสียงให้มีระดับต่ำ แล้วเพิ่มระดับเสียงอย่างรวดเร็วไปเป็นระดับสูงที่ท้ายพยางค์ นอกจากนี้ ยังอาจะ

เกิดจากการเปลี่ยนระดับเสียงจากสูงแล้วไปต่ำแล้วไปสูงอีก หรือเปลี่ยนจากต่ำแล้วไปสูงแล้วไปต่ำอีกก็ได้ สำหรับในภาษาไทยนั้นมีเสียวรรณยุกต์เปลี่ยนระดับอยู่ 2 เสียงดังนี้

- เสียงวรรณยุกต์โท (Falling Tone) ระดับเสียงจะเริ่มต้นที่ระดับความถี่ประมาณ 140 เฮิรตซ์ แต่เมื่อถึงประมาณ 1 ใน 4 ของความยาวช่วงพยางค์ ระดับความถี่จะเริ่มลดลงเรื่อยๆ จนต่ำกว่า 100 เฮิรตซ์ที่ปลายพยางค์ หรืออาจจะมีการเปลี่ยนระดับความถี่สูงขึ้นจากต้นพยางค์เล็กน้อยก่อนที่จะลดระดับเสียงลงอย่างรวดเร็วก็ได้ เสียงวรรณยุกต์โทนี้จะไม่ปรากฏในคำตาย ยกเว้นในคำเลียนเสียงธรรมชาติ หรือคำลงท้ายประโยคบางคำ เช่น "พลัก" หรือ "ละ" เป็นต้น
- เสียงวรรณยุกต์จัตวา (Rising Tone) ระดับเสียงจะเริ่มที่ระดับความถี่ประมาณ 110 เฮิรตซ์ แล้วมักจะลดลงเล็กน้อยก่อนจะเพิ่มความถี่ขึ้นอย่างรวดเร็วจนสูงถึงประมาณ 140 เฮิรตซ์ที่ท้ายพยางค์ เสียงวรรณยุกต์จัตวานี้จะไม่ปรากฏที่คำตาย

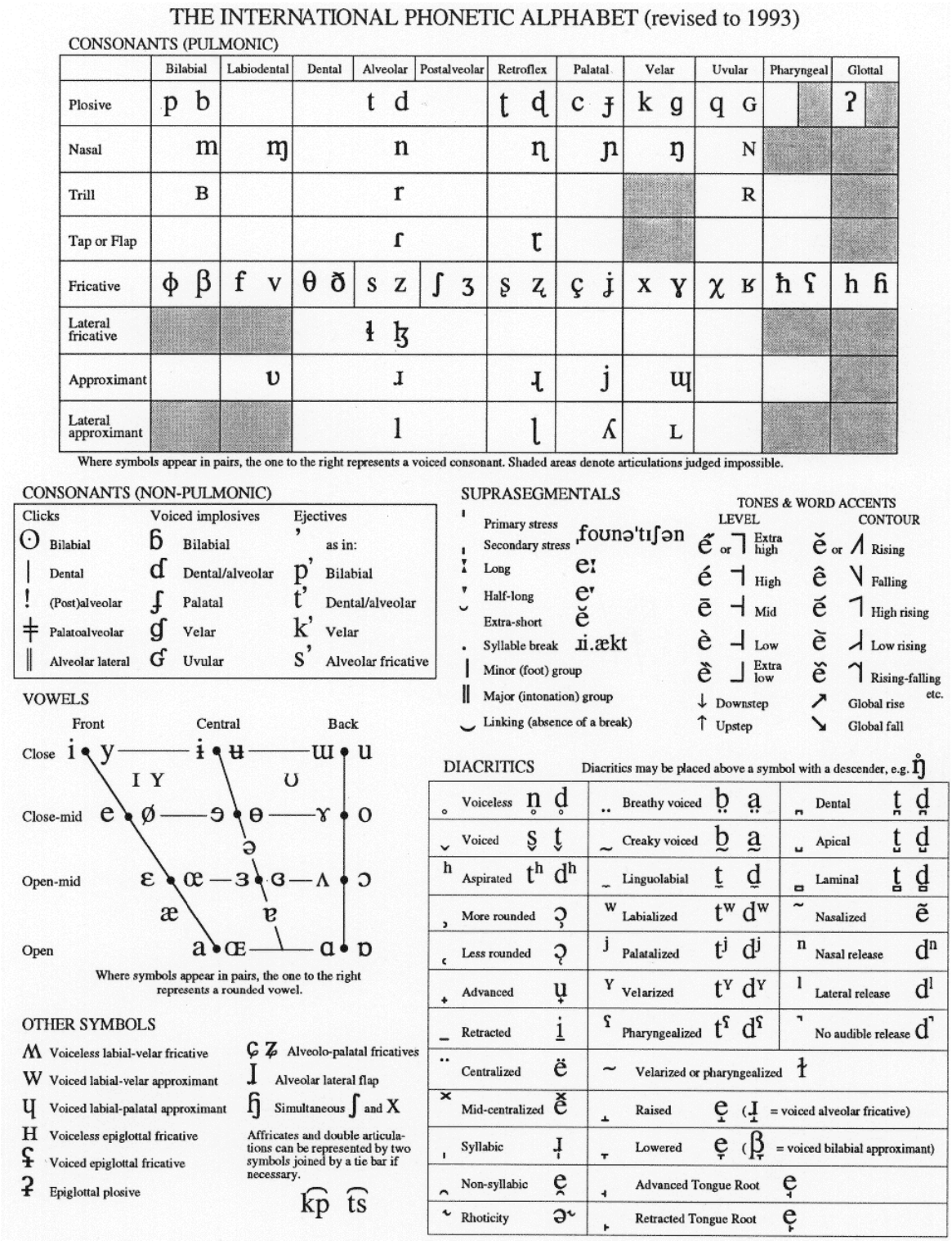
การเปลี่ยนแปลงความถี่ของเสียงในวรรณยุกต์ภาษาไทยสามารถแสดงได้ดังรูปที่ 2.3



รูปที่ 2.3 การเปลี่ยนแปลงความถี่ของเสียงวรรณยุกต์ในภาษาไทย

2.1.5 สัทอักษรสากล

เสียงในภาษาทั้งหมดที่มีในโลก สามารถเขียนแทนด้วยสัทอักษรสากล ดังรูปที่ 2.4 [84]

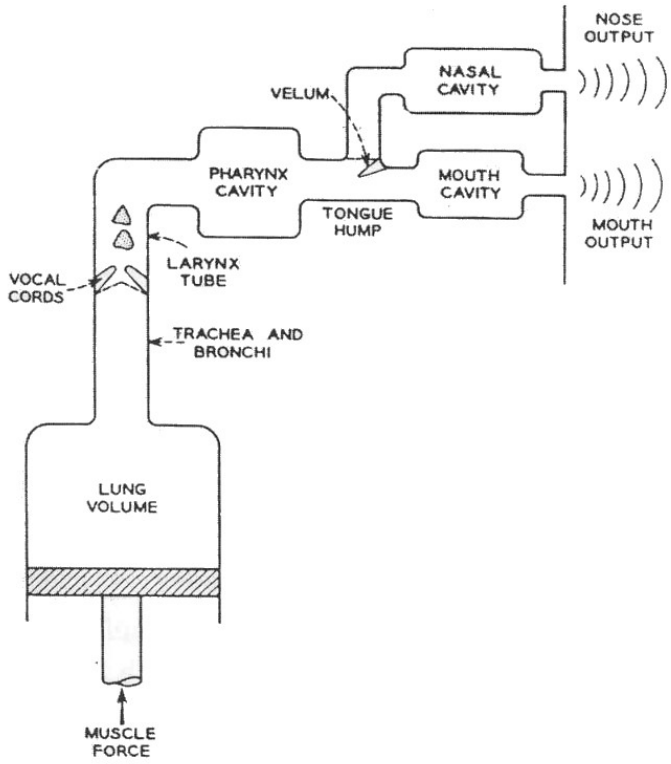


รูปที่ 2.4 สัทอักษรสากล

2.2 สวนสัทศาสตร์

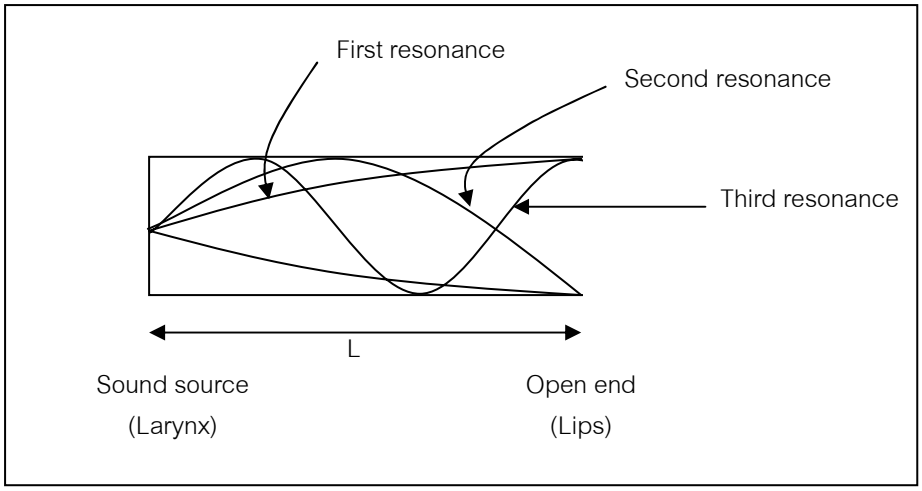
2.2.1 กระบวนการสร้างเสียงพูด (Speech Production)

กระบวนการสร้างเสียงพูดสามารถพิจารณาในเชิงสวนสัทศาสตร์ได้อย่างง่ายๆ ว่าประกอบไปด้วยลำดับของท่อและช่อง ซึ่งเปรียบได้กับทางเดินของเสียงจากปอดไปยังปากและจมูก ท่อและช่องนี้มีความยาวโดยรวมประมาณ 7 นิ้ว โดยเส้นเสียงจะอยู่ในตำแหน่งปลายสุด ทำหน้าที่ควบคุมการไหลของลมจากปอดให้เข้าสู่ช่องทางเดินเสียง ส่วนประกอบของช่องทางเดินเสียงที่มีลักษณะเป็นท่อจะสามารถเปลี่ยนแปลงรูปร่างได้ในอัตราสูงถึง 10 ครั้งต่อวินาที ส่วนเส้นเสียงนั้นจะสามารถเปิดและปิดได้ด้วยอัตราเร็วประมาณ 100 – 300 ครั้งต่อวินาที ซึ่งการเปลี่ยนแปลงรูปร่างของช่องทางเดินเสียงและการเปลี่ยนแปลงรูปร่างและตำแหน่งของสื่อกลางที่ทำให้เกิดเสียงดังกล่าวนี้ รวมเรียกว่ากระบวนการสร้างเสียงพูด (Speech Production) [85] แบบจำลองของกระบวนการสร้างเสียงพูดสามารถแสดงได้ดังรูปที่ 2.5 [18]



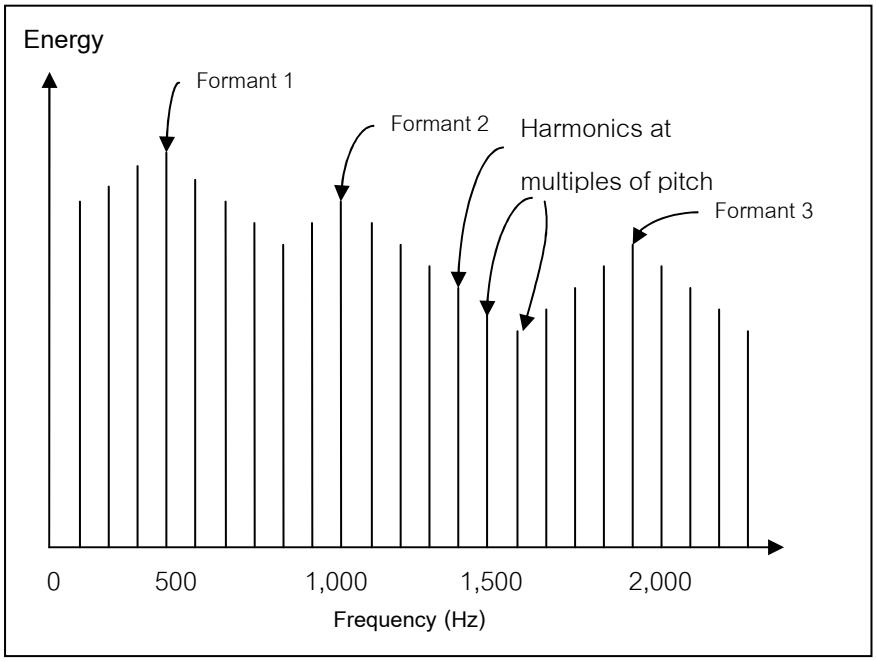
รูปที่ 2.5 แบบจำลองกระบวนการสร้างเสียงพูด

แบบจำลองอย่างง่ายของช่องทางเดินเสียงอาจมองได้เป็นลักษณะของท่อทรงกระบอกที่มีต้นกำเนิดเสียงอยู่ที่ปลายปิดข้างหนึ่ง (กล่องเสียง) ส่วนปลายอีกข้างหนึ่งจะเปิด (ปากและจมูก) ดังรูปที่ 2.6 ดังนั้นจึงเกิดเรโซแนนท์ (Resonant) ภายในท่อได้ที่ความยาวเท่ากับ 4L, 4L/3, 4L/5, ... เมตร โดยที่ L คือ ความยาวของท่อ ถ้าคิดเป็นความถี่ที่เกิดเรโซแนนท์จะได้ความถี่ที่ c/4L, 3c/4L, 5c/4L, ... เฮิรตซ์ โดยที่ c คือ ค่าความเร็วของเสียงในอากาศ และถ้าจะคำนวณหาความถี่ในการเรโซแนนท์ของช่องทางเดินเสียงของคน ซึ่งปกติช่องทางเดินเสียงของคนเราจะมีความยาวประมาณ 7 นิ้ว หรือ 17 เซนติเมตร และ c มีค่าเท่ากับ 340 เมตรต่อวินาที ดังนั้น จึงมีเรโซแนนท์ที่ความถี่ประมาณ 500, 1,500, 2,500 เฮิรตซ์ เป็นต้น [86]



รูปที่ 2.6 การเกิดเรโซแนนท์ภายในแบบจำลองของช่องทางเดินเสียง

เมื่อกล่องเสียงกระตุ้นให้เกิดคลื่นที่ประกอบไปด้วยฮาร์มอนิกต่าง ๆ มากมาย เรโซแนนท์ของช่องทางเดินเสียงนี้จะสร้างรูปคลื่นที่มียอดสูงเด่นขึ้นมาเมื่อดูจากสเปกตรัมพลังงานของรูปคลื่น ซึ่งเรียกว่า ฟอร์แมนท์ (Formant) ของเสียง ดังรูปที่ 2.7



รูปที่ 2.7 สเปกตรัมพลังงานของเสียง

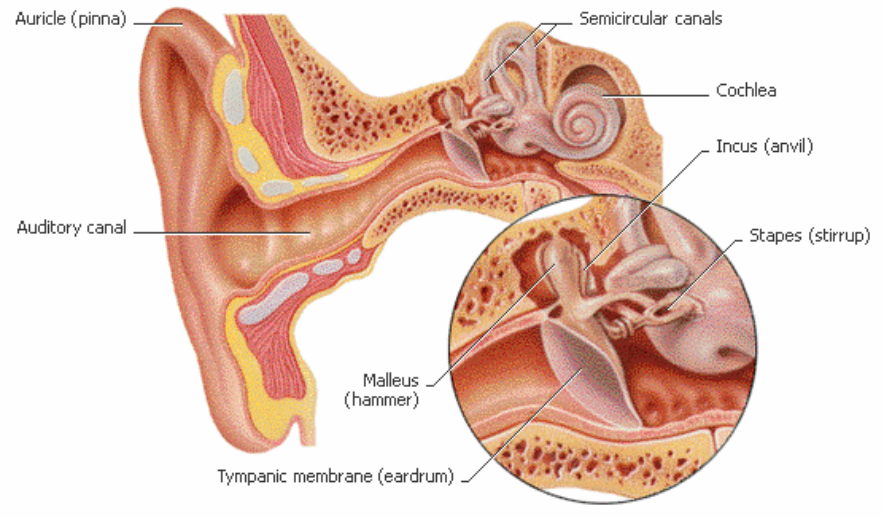
ฟอร์แมนท์ที่มีความถี่ต่ำที่สุดจะเรียกว่า ฟอร์แมนท์ที่หนึ่ง ซึ่งจะมีค่าประมาณ 200 – 1,000 เฮิรตซ์ ทั้งนี้ขึ้นอยู่กับขนาดของช่องทางเดินเสียงด้วย ส่วนฟอร์แมนท์ที่สองที่อยู่ถัดไปก็จะมีค่าประมาณ 500 – 2,500 เฮิรตซ์ และฟอร์แมนท์ที่สามมีค่าประมาณ 1,500 – 3,500 เฮิรตซ์ เป็นต้น โดยฟอร์แมนท์ที่หนึ่งและสองเป็นคุณสมบัติที่สำคัญมากคุณสมบัติหนึ่งที่สามารถบ่งชี้เสียงสระ

2.3 โสตสัทศาสตร์

อวัยวะสำหรับการรับฟังของมนุษย์คือหู ซึ่งมีโครงสร้างและกระบวนการทำงานดังนี้

2.3.1 กายวิภาคศาสตร์ของหู (Ear Anatomy) [87]

โครงสร้างของหูมนุษย์แบ่งได้เป็น 3 ส่วน คือ หูชั้นนอก หูชั้นกลาง และหูชั้นใน ดังรูปที่ 2.8 [87]



รูปที่ 2.8 โครงสร้างของหูมนุษย์

หูชั้นนอก

หูชั้นนอกประกอบด้วยส่วนต่างๆ ดังนี้

- ใบหู (Auricle, Pinna) เป็นส่วนโค้งภายนอกของหูที่ยึดติดกับส่วนข้างของหัวด้วยเส้นเอ็นและกล้ามเนื้อเล็กๆ ใบหูประกอบไปด้วยกระดูกอ่อนที่ยืดหยุ่นได้ซึ่งมีรูปร่างที่ช่วยในการรับฟังเสียง สำหรับด้านล่างของหูหรือติ่งหูนั้นส่วนใหญ่จะประกอบไปด้วยเนื้อเยื่อไขมัน
- ช่องหู (Auditory Canal) ช่องหูส่วนนอกยาวประมาณ 3 เซนติเมตร มีลักษณะเป็นท่อเชื่อมจากใบหูไปสู่แก้วหู ประกอบด้วยขนเล็กๆ และต่อมที่หลั่งสารซีรูเมน (Cerumen) ซีรูเมนเป็นของเหลวคล้ายขี้ผึ้ง ทำหน้าที่ดักจับฝุ่นละอองไม่ให้ผ่านไปถึงส่วนของหูที่อยู่ชั้นใน
- แก้วหู (Eardrum, Tympanic Membrane) เป็นเนื้อเยื่อตึงๆ บางๆ รูปร่างกลม กว้างประมาณ 10 มิลลิเมตร เป็นส่วนที่แบ่งขอบเขตระหว่างหูชั้นนอกและหูชั้นกลาง การสั่นของแก้วหูทำให้คลื่นเสียงถูกส่งไปยังหูส่วนต่อไปได้

หูชั้นกลาง

หูชั้นกลางเป็นท่ออากาศเล็กๆ ยาวประมาณ 15 มิลลิเมตร ประกอบด้วยส่วนต่างๆ ดังนี้

- ท่อออสเตเชียน (Eustachian Tube) เป็นท่อที่เชื่อมจากหูชั้นกลางไปยังคอและส่วนหลังของจมูก ทำหน้าที่ปรับความดันของหูชั้นนอกและหูชั้นกลางให้เท่ากัน เพื่อป้องกันการบาดเจ็บของแก้วหู
- กระดูกสามชิ้น (Three Ossicles) ในท่อของหูชั้นกลางประกอบด้วยกระดูกขนาดเล็กมากสามชิ้นเชื่อมต่อกัน และถูกตั้งชื่อตามรูปร่างของมัน ได้แก่ กระดูกค้อน (Malleus,

Hammer) กระดูกทั่ง (Incus, Anvil) และกระดูกโกลน (Stapes, Stirrup) ซึ่งเป็นกระดูกชิ้นที่เล็กที่สุดในร่างกาย โดยกระดูกค้อนจะเชื่อมต่อกับแก้วหู และกระดูกโกลนจะเชื่อมต่อกับหน้าต่างรูปไข่ (Oval Window) ซึ่งเป็นเนื้อเยื่อที่อยู่ด้านหน้าของหูชั้นใน การสั่นไหวของแก้วหูจะทำให้กระดูกค้อนเคลื่อนไหว การเคลื่อนไหวของกระดูกค้อนก่อให้เกิดการเคลื่อนไหวของกระดูกทั่ง และการเคลื่อนไหวของกระดูกทั่งทำให้เกิดการเคลื่อนไหวของกระดูกโกลน เนื่องจากหูชั้นในซึ่งเป็นของเหลวรับการสั่นสะเทือนได้ยาก การเคลื่อนไหวของกระดูกเหล่านี้ถือเป็นการรวบรวมความสั่นสะเทือนให้เข้มข้นขึ้นที่หน้าต่างรูปไข่

หูชั้นใน

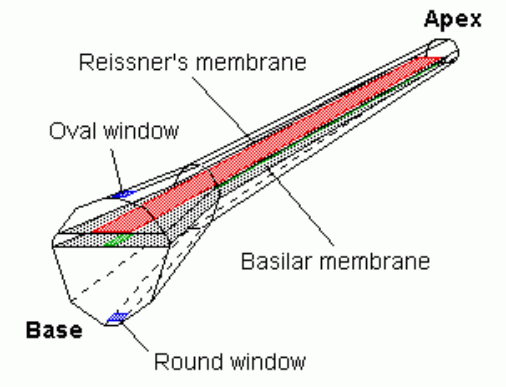
หูชั้นในทำหน้าที่ทั้งในด้านการได้ยินและการรักษาสมดุลของร่างกาย ประกอบด้วยส่วนต่างๆ ดังนี้

- **คอเคลีย** (Cochlea) มีรูปร่างเป็นท่อม้วนขดคล้ายเปลือกหอยทาก ประกอบด้วยช่องของเหลวสามช่อง ได้แก่ ช่องเวสทิบูลาร์ (Vestibular Canal) ช่องคอเคลีย (Cochlear Canal) และช่องทิมพานิค (Tympanic Canal) โดยมีเนื้อเยื่อบาสิลา (Basilar Membrane) แบ่งระหว่างช่องคอเคลียและช่องทิมพานิค และมีออร์แกนของคอร์ติ (Organ of Corti) ติดอยู่กับเนื้อเยื่อบาสิลา ที่ออร์แกนของคอร์ติจะมีเซลล์เส้นผม (Hair Cell) คอยรับการสั่นสะเทือนและแปลงเป็นสัญญาณเพื่อส่งไปยังสมองต่อไป
- **เวสทิบูล** (Vestibule) และ **ช่องเซมิเซอร์คูลา** (Semicircular Canals) เป็นส่วนที่ช่วยรักษาความสมดุลของร่างกาย โดยจะส่งสถานะการทรงตัวไปยังสมอง

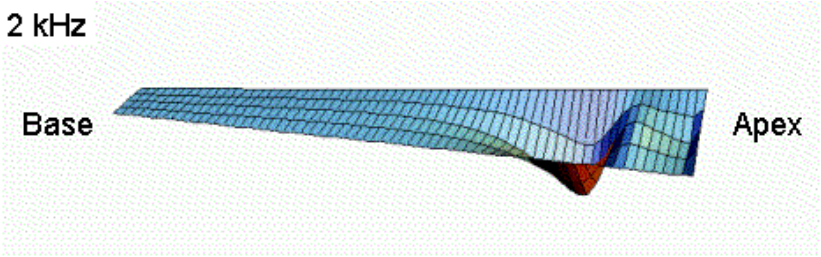
2.3.2 กลไกในการรับฟังเสียงพูดของมนุษย์ [17]

เมื่อหน้าต่างรูปไข่สั่นสะเทือน การสั่นสะเทือนนั้นจะถูกส่งเข้าไปยังของเหลวที่อยู่ในคอเคลีย และทำให้เนื้อเยื่อบาสิลาสั่นสะเทือนด้วย เซลล์เส้นผมที่อยู่ด้านบนของเนื้อเยื่อบาสิลาจะแปลงการสั่นสะเทือนให้เป็นไฟกระชาก (Spike) ส่งไปทางเส้นประสาทการได้ยิน (Auditory Nerve)

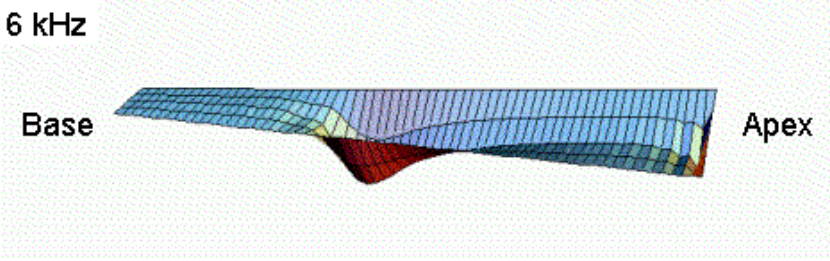
เนื้อเยื่อบาสิลาในคอเคลียด้านที่ติดกับหน้าต่างรูปไข่ หรือด้านเบส (Base) จะค่อนข้างแคบและแข็ง ขณะที่เนื้อเยื่อบาสิลาอีกด้านหนึ่ง หรือด้านเอเพ็กซ์ (Apex) จะกว้างและอ่อนกว่า ดังรูปที่ 2.9 การสั่นสะเทือนที่ความถี่ต่างกันจะทำให้เนื้อเยื่อบาสิลามีการสั่นสูงสุดที่ตำแหน่งต่างกัน โดยเสียงความถี่สูงจะทำให้เนื้อเยื่อบาสิลามีการสั่นสูงสุดที่ตำแหน่งใกล้ด้านเบส ดังรูปที่ 2.10 ขณะที่เสียงความถี่ต่ำจะทำให้เนื้อเยื่อบาสิลามีการสั่นสูงสุดที่ตำแหน่งใกล้ด้านเอเพ็กซ์ ดังรูปที่ 2.11 [88]



รูปที่ 2.9 แบบจำลองอย่างง่ายของคอเคลีย

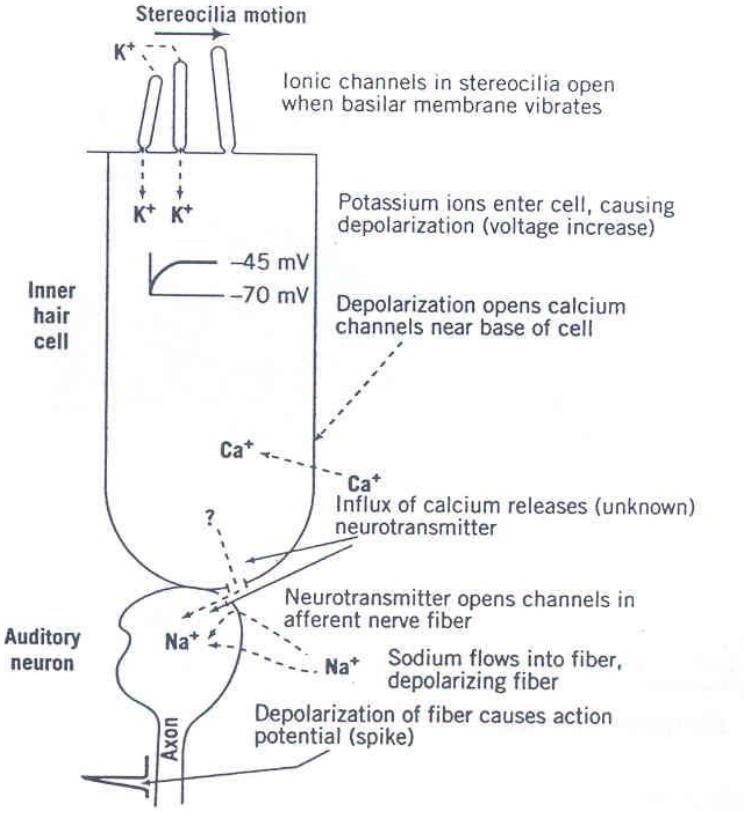


รูปที่ 2.10 แบบจำลองการสั่นของเนื้อเยื่อบาสิลาเมื่อได้รับเสียงความถี่ 2000 เฮิรตซ์



รูปที่ 2.11 แบบจำลองการสั่นของเนื้อเยื่อบาสิลาเมื่อได้รับเสียงความถี่ 6000 เฮิรตซ์

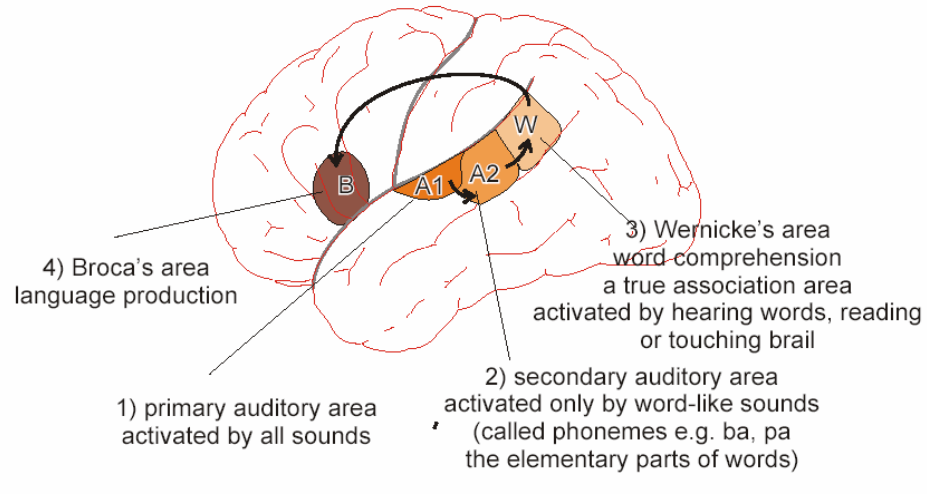
รูปที่ 2.12 แสดงการเคลื่อนไหวของสเตอเรโอซิเลีย (Stereocilia) ทำให้เกิดไฟกระชากในเส้นประสาทการได้ยินที่อยู่ติดกับเซลล์เส้นผม ซึ่งสมองจะรับสัญญาณนี้และนำไปตีความต่อไป



รูปที่ 2.12 การเกิดไฟกระชากจากเซลล์เส้นผม

สมองส่วนที่เกี่ยวข้องกับการรู้จำเสียงพูดแสดงได้ดังรูปที่ 2.13 โดยสมองส่วนแรกในการได้ยิน (Primal Auditory Area, A1) จะถูกกระตุ้นจากทุกเสียง แต่สมองส่วนที่สองในการได้ยิน

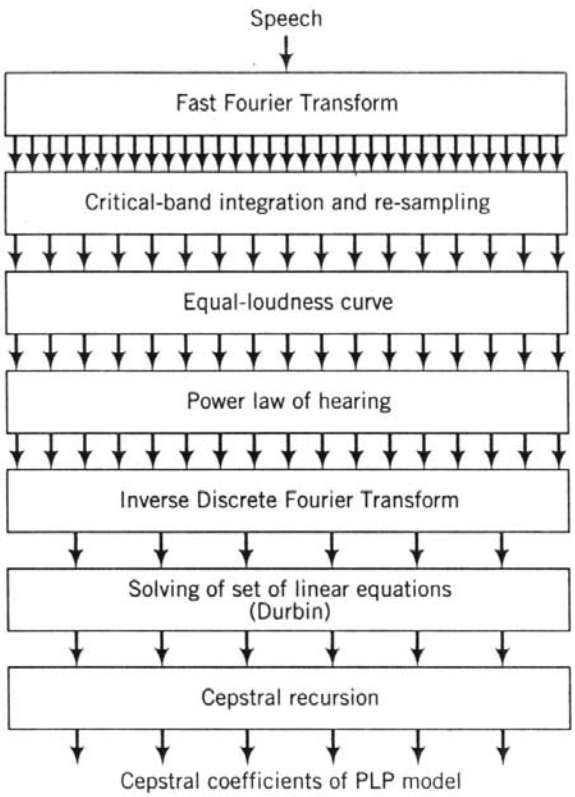
(Secondary Auditory Area, A2) จะถูกกระตุ้นเฉพาะเสียงที่เป็นหน่วยคำ และสมองส่วนเวอร์นิค (Wernicke's area, W) จะทำความเข้าใจความหมายของคำที่เข้ามา ทั้งจากการอ่าน การฟัง หรือการสัมผัส โดยมีสมองส่วนโบรคา (Broca's Area) ทำหน้าที่ผลิตภาษา [89]



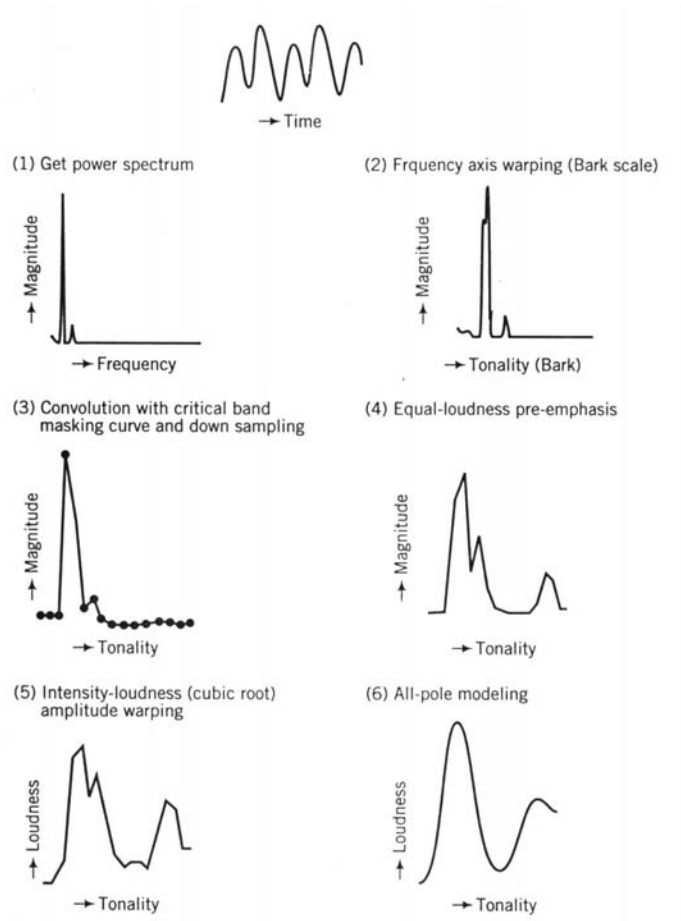
รูปที่ 2.13 สมองส่วนที่ทำหน้าที่เกี่ยวกับการรู้จำเสียงพูด

2.4 การทำนายเชิงเส้นแบบรับรู้ [64]

การทำนายเชิงเส้นแบบรับรู้ – พีแอลพี (Perceptual Linear Prediction, PLP) เป็นการสกัดลักษณะสำคัญของเสียงพูดโดยมีพื้นฐานมาจากการได้ยินของมนุษย์ ซึ่งมีขั้นตอนดังรูปที่ 2.14 และในแต่ละขั้นตอนจะทำให้สเปกตรัมของเสียงเปลี่ยนไปดังรูปที่ 2.15 [17]



รูปที่ 2.14 ขั้นตอนการคำนวณพีแอลพี



รูปที่ 2.15 ผลกระทบจากแต่ละขั้นตอนของพีแอลพีที่มีต่อสเปกตรัม

การทำนายเชิงเส้นแบบรับรู้แต่ละขั้นตอน ซึ่งมีรายละเอียดดังต่อไปนี้

2.4.1 การแปลงฟูเรียร์แบบเร็ว

เริ่มต้นด้วยคำนวณค่าประมาณของสเปกตรัมกำลัง (Power Spectrum) สำหรับหน้าต่างวิเคราะห์ (Analysis Window) ซึ่งทำโดยการเอาหน้าต่างไปใส่ในขอบเขตที่จะวิเคราะห์ โดยการคูณแต่ละค่าของสัญญาณในกรอบข้อมูลเสียงด้วยค่าฟังก์ชันกรอบ เช่น หน้าต่างแฮมมิง (Hamming Window) ดังสมการ

$$W(n) = 0.54 + 0.46\cos\left(\frac{2\pi m}{N-1}\right)$$

เมื่อ N คือ จำนวนข้อมูลต่อ 1 กรอบ จากนั้นนำสัญญาณที่ได้ผ่านกระบวนการแปลงฟูเรียร์แบบเร็ว และคำนวณขนาดกำลังสองของมันพร้อมทั้งสเปกตรัมกำลัง ดังสมการ

$$P(\omega) = \mathrm{Re}[S(\omega)]^2 + \mathrm{Im}[S(\omega)]^2$$

โดยที่ $S(\omega)$ คือ สัญญาณในโดเมนความถี่ที่ผ่านการแปลงฟูเรียร์

ขั้นตอนนี้จะเทียบเท่ากับขั้นตอนแรกในรูปที่ 2.14 และผลที่ได้ก็คือสเปกตรัมกำลังซึ่งแสดงได้ดังรูปย่อยที่ (1) ของรูปที่ 2.15

2.4.2 การหาปริพันธ์ของแถบวิกฤตและการชักตัวอย่างใหม่

ในการหาปริพันธ์ของแถบวิกฤตและการชักตัวอย่างใหม่ (Critical-Band Integration and Re-sampling) ตัวกรองซึ่งถูกทำให้เป็นรูปสี่เหลี่ยมคางหมู จะถูกนำมาใช้ที่ช่วงห่างประมาณ 1 บาร์ก (Bark) ซึ่งแกนของบาร์กนั้นจะได้มาจากแกนความถี่ โดยใช้ฟังก์ชันวาร์ป (Warping Function) ของโชรเดอร์ และ $P(\omega)$ ในแกนความถี่เฮิรตซ์จะถูกแปลงในอยู่ในแกนความถี่บาร์ก โดยใช้สมการ

$$\Omega(\omega) = 6\ln\left\{\frac{\omega}{1200\pi} + \left[\left(\frac{\omega}{1200\pi}\right)^2 + 1\right]^{0.5}\right\}$$

โดยที่ ω คือ ความถี่เชิงมุม ในหน่วยของเรเดียนต่อวินาที (rad/s)

สำหรับหน้าต่างรูปสี่เหลี่ยมคางหมูนั้นก็คือการประมาณสเปกตรัมกำลังของเส้นโค้งแถบวิกฤต (Critical Band Curve) ซึ่งจะเป็นดังสมการ

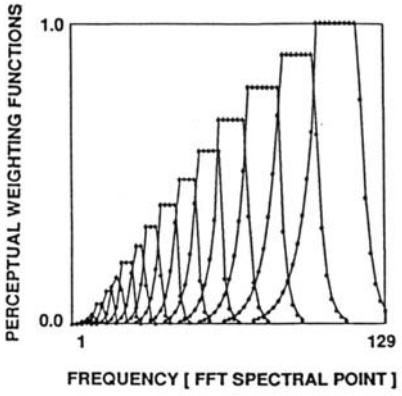
$$\Psi(\Omega) = \begin{cases} 0 & for \quad \Omega < -1.3 \\ 10^{2.5(\Omega+0.5)} & for -1.3 \le \Omega \le -0.5 \\ 1 & for \ -0.5 < \Omega < 0.5 \\ 10^{-1.0(\Omega-0.5)} & for \ \ 0.5 \le \Omega \le 2.5 \\ 0 & for \quad \Omega > 2.5 \end{cases}$$

$\Omega(\omega)$ จะถูกกระทำด้วยเส้นโค้งแถบวิกฤตโดยใช้สมการข้างต้น จากนั้นคำนวณ $\Theta(\Omega)$ ดังสมการ

$$\Theta(\Omega_i) = \sum_{\Omega=-1.3}^{2.5} P(\Omega - \Omega_i)\Psi(\Omega)$$

$\Theta(\Omega)$ ที่ได้เรียกว่าสเปกตรัมกำลังแถบวิกฤต (Critical-Band Power Spectrum) ซึ่งจะมีทั้งหมด 18 ค่า ครอบคลุม 0-16.9 บาร์ก (0-5 กิโลเฮิรตซ์) และแต่ละค่าจะปรากฏที่ตำแหน่งต่างกัน 0.994 บาร์ก

สำหรับผลกระทบสุทธินั้นจะใช้เพื่อลดความไวทางความถี่ของการประมาณค่าสเปกตรัมดั้งเดิม โดยเฉพาะอย่างยิ่งที่ความถี่สูง ซึ่งความถี่สูงๆ นั้นจะถูกเน้นย้ำโดยใช้ความกว้างแถบกรอง (Filter Bandwidth) ที่กว้างขึ้น ซึ่งผลที่ได้ดังแสดงในรูป 2.16



รูปที่ 2.16 ตัวกรองรูปสี่เหลี่ยมคางหมูของพีแอลพี

2.4.3 โค้งความดังเทียบเท่า

ทำการเน้นสเปกตรัมอีกครั้งหนึ่งเพื่อประมาณค่าความไวที่ไม่สมดุลของการได้ยินของมนุษย์ ณ ความถี่ต่างๆกัน ขั้นตอนนี้จะถูกทำให้เกิดผลด้วยการถ่วงน้ำหนักในของส่วนของสเปกตรัมแถบวิกฤต ซึ่งขั้นตอนนี้ก็เทียบได้กับขั้นตอนของโค้งความดังเทียบเท่า (Equal-Loudness Curve) ในรูปที่ 2.14

$\Theta(\Omega(\omega))$ จะถูกเน้นสัญญาณโดยเส้นโค้งความดังเท่าจำลอง (Simulated Equal-Loudness Curve) โดยใช้สมการ

$$\Xi[\Omega(\omega)] = E(\omega)\Theta[\Omega(\omega)]$$

โดยที่ $E(\omega)$ คือ ค่าประมาณของความไวในการรับเสียงของหูมนุษย์ที่ความถี่ต่างๆ ซึ่งถูกจำลองที่ความดัง 40 เดซิเบล และมีรูปแบบดังสมการ

$$E(\omega) = \frac{[(\omega^2 + 56.8\times10^6)\omega^4]}{[(\omega^2 + 6.3\times10^6)^2 \times (\omega^2 + 0.38\times10^9)(\omega^6 + 9.58\times10^{26})]}$$

2.4.4 กฎกำลังของการได้ยิน

กฎกำลังของการได้ยิน (Power Law of Hearing) เป็นการบีบแอมปลิจูดรากที่สาม (Cubic-Root Amplitude Compression) ทำการบีบอัดขนาดของสเปกตรัม ซึ่งโดยทั่วไปแล้วจะมีการหาค่าลอการิทึมหลังจากการหาปริพันธ์ โดยการทำนายเชิงเส้นแบบรับรู้จะใช้การหารากที่สามแทนการหาค่าลอการิทึม แสดงได้ดังสมการ

$$\Phi(\Omega) = \Xi(\Omega)^{0.33}$$

การคำนวณนี้เป็นการประมาณด้วยกฎกำลัง (Power Law) เพื่อที่จะจำลองความสัมพันธ์แบบไม่เชิงเส้นระหว่างความเข้มของเสียง (Intensity) และความรู้สึกถึงความดังของเสียง (Loudness) ซึ่งกระบวนการนี้จะช่วยลดความแปรปรวนของขนาด (Amplitude) ของสเปกตรัมแถบวิกฤต (Critical-Band Spectrum) หรือการสั่นพ้องของสเปกตรัม (Spectral Resonances)

2.4.5 การแปลงฟูเรียร์แบบไม่ต่อเนื่องผกผัน

จะทำการแปลงฟูเรียร์แบบไม่ต่อเนื่องผกผัน (Inverse Discrete Fourier Transform, IDFT) สำหรับการทำนายเชิงเส้นแบบรับรู้นี้ เนื่องจากค่าลอการิทึมไม่ได้ถูกคำนวณ ดังนั้นผลที่ได้จึงมักจะคล้ายกับค่าสัมประสิทธิ์อัตสหสัมพันธ์ (Autocorrelation Coefficients) มากกว่าถึงแม้ว่ามันจะมาจากสเปกตรัมที่ถูกบีบอัดก็ตาม และเนื่องจากค่าสเปกตรัมกำลังนั้นเป็นจำนวนจริงและเป็นเลขคู่ ดังนั้นจึงไม่จำเป็นที่จะต้องทำการคำนวณส่วนประกอบโคไซน์ของการแปลงฟูเรียร์แบบไม่ต่อเนื่องผกผันก็ได้

$\Phi(\Omega)$ จะถูกประมาณโดยสเปกตรัมของแบบจำลองทุกขั้ว (All-Pole) ด้วยวิธีอัตสหสัมพันธ์

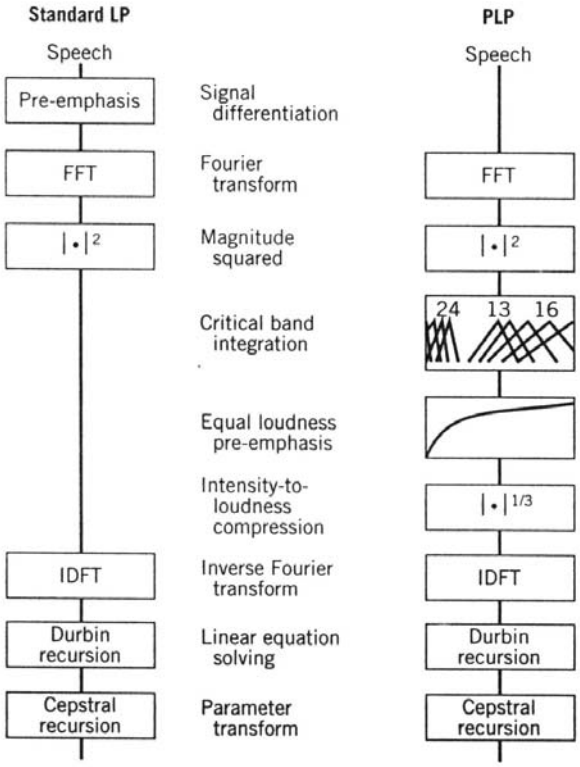
2.4.6 การแก้ชุดสมการเชิงเส้น

การหาปริพันธ์นั้นจะมีประโยชน์ต่อการลดผลกระทบของต้นกำเนิดซึ่งไม่เกี่ยวข้องกับทางภาษาศาสตร์ของความแปรปรวนในสัญญาณเสียงพูดได้ แบบจำลองอัตตาถดถอย จะได้มาจากผลเฉลยของสมการเชิงเส้น ซึ่งถูกสร้างขึ้นจากอัตสหสัมพันธ์ของขั้นตอนก่อนหน้านี้นั่นเอง โดยแบบจำลองอัตตาถดถอยนี้จะถูกใช้ในการขัดเกลาสเปกตรัมแถบวิกฤตซึ่งถูกบีบอัดแล้ว เหมือนกับในการทำรหัสทำนายเชิงเส้นแบบดั้งเดิม โดยสเปกตรัมผลลัพธ์ซึ่งถูกขัดเกลาแล้วจะเหมาะสมกับยอดสเปกตรัม (Spectral Peak) มากกว่าเหวสเปกตรัม (Spectral Valley) นักวิจัยหลายท่านได้พบแล้วว่าวิธีนี้จะนำไปสู่ความทนทานต่อเสียงรบกวน และการไม่ขึ้นกับผู้พูดได้ดีกว่าการตัดปลายเซปสตรัม สำหรับรูปย่อยที่ (6) ในรูปที่ 2.15 นั้นจะแสดงถึงตัวแทนซึ่งถูกขัดเกลาแล้วของยอดซึ่งเป็นชนิดหนึ่งของผลที่ได้มาจากกระบวนการทำนายเชิงเส้นแบบรับรู้นั่นเอง

2.4.7 การเวียนเกิดเซปสตรัม

การเวียนเกิดเซปสตรัม (Cepstral Recursion) จะทำการใช้ตัวแทนเชิงตั้งฉาก (Orthogonal Representation) สำหรับการทำนายเชิงเส้นแบบรับรู้ สัมประสิทธิ์อัตตาถดถอยจะถูกแปลงให้เป็นตัวแปรเซปสตรัมแทน

รูปที่ 2.17 เป็นการเปรียบเทียบขั้นตอนต่างๆ ของการทำการทำนายเชิงเส้นแบบธรรมดากับการทำนายเชิงเส้นแบบรับรู้ [17]



รูปที่ 2.17 การเปรียบเทียบระหว่างแอลพีซีกับพีแอลพี

2.4.8 อนุพันธ์ของการทำนายเชิงเส้นแบบรับรู้

ลักษณะสำคัญของเสียงที่หาได้จากการทำนายเชิงเส้นแบบรับรู้สามารถนำมาหาค่าอนุพันธ์อันดับที่หนึ่งของการทำนายเชิงเส้นแบบรับรู้ (Delta PLP) และอนุพันธ์อันดับที่สองของการทำนายเชิงเส้นแบบรับรู้ (Delta-Delta PLP) ได้

อนุพันธ์อันดับที่หนึ่งใช้สมการ

$$\Delta\hat{c}_l(m) = \left[\sum_{k=-K}^{K} k\hat{c}_{l-k}(m)\right] \;;1 \le m \le Q$$

เมื่อ Q คือจำนวนอันดับการทำนายเชิงเส้นแบบรับรู้

$2K+1$ เป็นจำนวนกรอบที่ใช้ในการคำนวณอนุพันธ์

l เป็นกรอบปัจจุบัน ที่นำมาคำนวณ

$\Delta\hat{c}$ เป็นสัมประสิทธิ์การทำนายเชิงเส้นแบบรับรู้

$\Delta\Delta\hat{c}$ เป็นอนุพันธ์อันดับที่หนึ่งของสัมประสิทธิ์การทำนายเชิงเส้นแบบรับรู้

อนุพันธ์อันดับที่สองใช้สมการ

$$\Delta\Delta\hat{c}_l(m) = \left[\sum_{k=-K}^{K} k\Delta\hat{c}_{l-k}(m)\right] \;;1 \le m \le Q$$

เมื่อ Q คือ จำนวนอันดับการทำนายเชิงเส้นแบบรับรู้

$2K+1$ เป็นจำนวนกรอบที่ใช้ในการคำนวณอนุพันธ์

l เป็นกรอบปัจจุบัน ที่นำมาคำนวณ

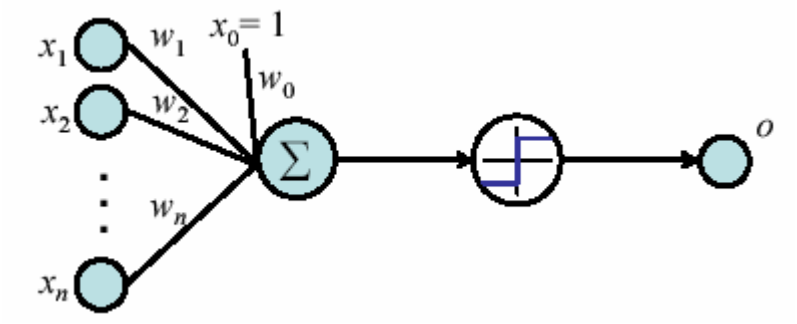
$\Delta\hat{c}$ เป็นอนุพันธ์อันดับที่หนึ่งของสัมประสิทธิ์การทำนายเชิงเส้นแบบรับรู้

$\Delta\Delta\hat{c}$ เป็นอนุพันธ์อันดับที่สองของสัมประสิทธิ์การทำนายเชิงเส้นแบบรับรู้

2.5 โครงข่ายประสาทเทียม [90] [91]

2.5.1 เพอร์เซปตรอน

หน่วยพื้นฐานของโครงข่ายประสาทเทียม คือ เพอร์เซปตรอน (Perceptron) ซึ่งจำลองมาจากการทำงานของเซลล์สมอง (Neuron) ของมนุษย์ มีลักษณะดังรูปที่ 2.18



รูปที่ 2.18 เพอร์เซปตรอน

เพอร์เซปตรอนจะรับอินพุตเป็นเวคเตอร์ของจำนวนจริง ($x_1, x_2, ..., x_n : \vec{x}$) เข้ามา แล้วคำนวณหาผลรวมเชิงเส้น (Linear Combination) แบบถ่วงน้ำหนักของอินพุต ($w_1x_1 + w_2x_2 + ... + w_nx_n : \vec{w} \cdot \vec{x}$) โดยถ้าค่าผลรวมเชิงเส้นแบบถ่วงน้ำหนักนี้มีค่ามากกว่าค่าขีดแบ่งค่าหนึ่ง (θ) เพอร์เซปตรอนจะให้ผลลัพธ์เป็น 1 มิฉะนั้น ผลลัพธ์ที่ออกมาจะมีค่าเป็น -1 ซึ่งค่าของผลลัพธ์นี้อาจเปลี่ยนแปลงได้ตามฟังก์ชันกระตุ้น (Activation Function) ที่ใช้ ซึ่งในที่นี้ใช้ฟังก์ชันสองขั้น (Bipolar Function, 

) หรืออาจเขียนสมการได้เป็น

$$o(x_1, x_2, ..., x_n) = \begin{cases} 1, & w_1x_1 + w_2x_2 + ... + w_nx_n > \theta \\ -1, & w_1x_1 + w_2x_2 + ... + w_nx_n < \theta \end{cases}$$

ถ้าย้าย θ มา พร้อมจัดรูปใหม่โดยใช้ w_0 จะได้สมการ

$$o(x_1, x_2, ..., x_n) = \begin{cases} 1, & w_0 + w_1x_1 + w_2x_2 + ... + w_nx_n > 0 \\ -1, & w_0 + w_1x_1 + w_2x_2 + ... + w_nx_n < 0 \end{cases}$$

การเรียนรู้ของเพอร์เซปตรอนทำโดยปรับเวคเตอร์ของน้ำหนักจนกระทั่งแยกแยะข้อมูลอินพุตได้ถูกต้องทุกตัว ถ้าให้ข้อมูลการเรียนรู้อยู่ในรูป ($\vec{x}$, t) เมื่อ $\vec{x}$ เป็นเวคเตอร์อินพุต และ t เป็นค่าเอาต์พุตของข้อมูล การเรียนรู้ของเพอร์เซปตรอนจะมีกระบวนการดังนี้

- สุ่มน้ำหนัก w_i เริ่มต้น
- สำหรับทุกข้อมูลอินพุต
  - คำนวณค่า $o(x_1, x_2, ..., x_n)$
  - ถ้า $o(x_1, x_2, ..., x_n) = t$ ไม่ต้องปรับค่าน้ำหนัก
  - ถ้า $o(x_1, x_2, ..., x_n) \neq t$ ปรับค่าน้ำหนัก โดย

$$w_i \leftarrow w_i + \Delta w_i$$

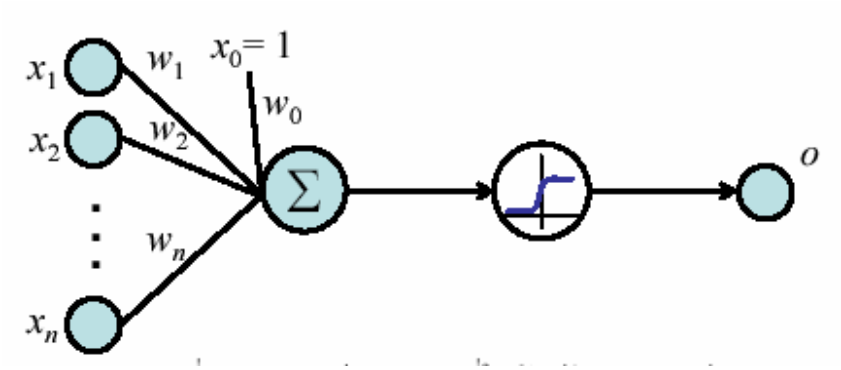
$$\Delta w_i = \alpha(t - o)x_i$$

โดยที่ α เป็นค่าอัตราการเรียนรู้

อย่างไรก็ตาม เพอร์เซปตรอนจะเรียนรู้ได้เฉพาะฟังก์ชันแยกเชิงเส้น (Linear Separable Function) เท่านั้น สำหรับฟังก์ชันแยกเชิงเส้นไม่ได้ (Linearly Non-separable Function) อาจทำการเรียนรู้โดยเปลี่ยนฟังก์ชันกระตุ้นให้เป็นฟังก์ชันที่หาอนุพันธ์ได้ แล้วปรับค่าน้ำหนักตามกฎเดลตา (Delta Rule) เพื่อให้ความผิดพลาดในการเรียนรู้มีน้อยที่สุด

2.5.2 โครงข่ายประสาทเทียม

โครงข่ายประสาทเทียม (Artificial Neural Network) เป็นการนำเพอร์เซปตรอนหลายตัวมาต่อเชื่อมกัน ส่งผลให้พื้นผิวการตัดสินใจ (Decision Surface) มีความซับซ้อนมากกว่าเพอร์เซปตรอน ทำให้สามารถเรียนรู้และแยกแยะข้อมูลได้ดีกว่า เพอร์เซปตรอนในโครงข่ายประสาทเทียมจะใช้ฟังก์ชันซิกมอยด์ (σ) เป็นฟังก์ชันกระตุ้น ดังรูป 2.19



รูปที่ 2.19 เพอร์เซปตรอนที่ใช้ฟังก์ชันซิกมอยด์

เมื่อใช้ฟังก์ชันซิกมอยด์ จะได้ว่า

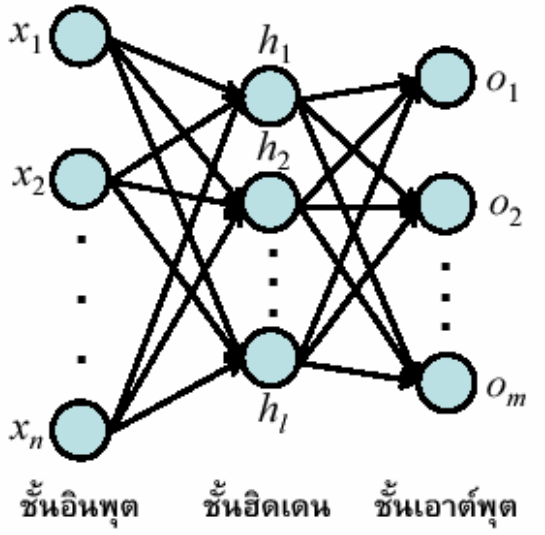
$$o = \sigma(net) = \frac{1}{1+e^{-net}}$$

เมื่อ $net = \sum_{i=0}^{n} w_i x_i$

ซึ่งฟังก์ชันซิกมอยด์ถือเป็นฟังก์ชันที่มีลักษณะพิเศษคือ

$$\frac{d\sigma(y)}{dy} = \sigma(y)(1-\sigma(y))$$

เทคนิคการเรียนรู้ที่ใช้กันอย่างแพร่หลายในโครงข่ายประสาทเทียมเป็นการเรียนรู้แบบแบ็กพรอพาเกชัน (Backpropagation) ซึ่งแบ่งเครือข่ายออกเป็นสามชั้น ได้แก่ ชั้นอินพุต ชั้นฮิดเดน และชั้นเอาต์พุต ดังรูปที่ 2.20



รูปที่ 2.20 โครงข่ายประสาทเทียม

ถ้าให้ข้อมูลในการเรียนรู้อยู่ในรูปคู่เวคเตอร์ $(\vec{x},\vec{t})$ เมื่อ $\vec{x}$ เป็นเวคเตอร์อินพุต และ $\vec{t}$ เป็นเวคเตอร์เอาต์พุต โดยที่อินพุตและค่าน้ำหนักของหน่วย j ซึ่งมาจากหน่วย i เขียนแทนด้วย x_{ji} และ w_{ji} ตามลำดับ การเรียนรู้แบบแบ็กพรอพาเกชันมีขั้นตอนดังนี้

- สร้างโครงข่ายประสาทเทียมตามโครงสร้างที่ต้องการ
- สุ่มน้ำหนัก w_{ji} เริ่มต้น
- สำหรับทุกข้อมูลอินพุต
  - ทำการคำนวณหาค่า $\vec{o}$
  - คำนวณหาค่า δ_k ของทุกหน่วย k ในชั้นเอาต์พุต โดย

$$\delta_k = o_k(1-o_k)(t_k - o_k)$$

  - คำนวณค่าความคลาดเคลื่อน δ_h ของทุกหน่วย h ในชั้นฮิดเดน โดย

$$\delta_h = o_h(1-o_h)\sum_{k\in outputs} w_{kh}\delta_k$$

  - ปรับค่าน้ำหนักของเส้นเชื่อม w_{ji} โดย

$$w_{ji} = w_{ji} + \Delta w_{ji}$$
$$\text{ซึ่ง } \Delta w_{ji} = \eta\delta_j x_{ji} + \alpha\Delta w_{ji}$$

เมื่อ η คือค่าอัตราการเรียนรู้ และ α คือค่าโมเมนตัม ซึ่งมีหน้าที่รักษาน้ำหนักให้ไปในทิศทางเดิม

2.6 การหาขอบเขตของเสียงพูด [92]

การหาของเขตของเสียงพูด (Speech Boundary Detection) เป็นการตรวจจับว่าเสียงพูดที่เข้ามานั้นเริ่มต้นที่ใด และสิ้นสุดที่ใด โดยจะทำการวิเคราะห์ค่าพลังงานของเสียงทีละเฟรมในช่วงเวลาสั้นๆ เริ่มจากการคำนวณค่าพลังงานของแต่ละจุดในเฟรมด้วยสมการ

$$E_j = \sqrt{\left(\sum_{i=j}^{j+n} x_i^2\right)/n}$$

โดยที่

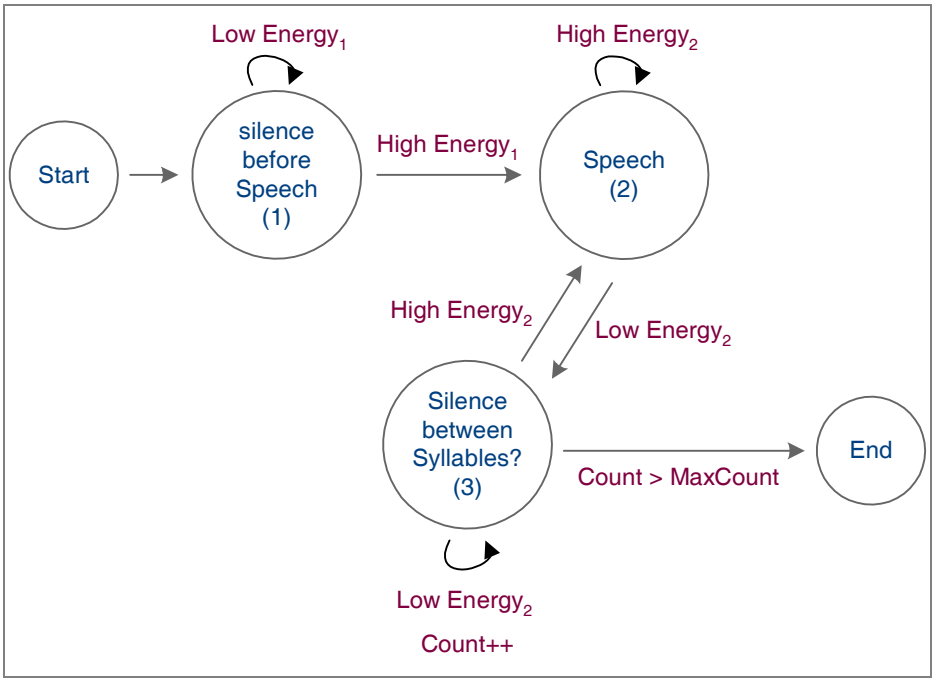
- E_j คือค่าพลังงานของจุดที่ j
- x_i คือค่าแอมปลิจูดของสัญญาณเสียงตำแหน่งที่ i
- n เป็นจำนวนจุดที่นำมาใช้ในการคำนวณ

สำหรับค่าเฉลี่ยพลังงานของเฟรมใดๆ จะคำนวณโดย

$$E_{avg}(F) = \left(\sum_{j\in F} E_j\right)/|F|$$

เมื่อ F คือเฟรม หรือเซตของจุดที่ต้องการคำนวณ

ในการหาขอบเขตของเสียงพูด จำเป็นต้องกำหนดสถานะของระบบ ดังรูปที่ 2.21



รูปที่ 2.21 สถานะในการหาขอบเขตของเสียงพูด

เริ่มแรก ระบบจะอยู่ในสถานะ (1) ซึ่งเป็นสถานะก่อนที่จะมีเสียงพูดเข้ามา จากนั้นจะทำการคำนวณเฟรมที่รับเข้ามาว่าเป็นจุดเริ่มต้นของเสียงพูดหรือไม่ โดยการนับจำนวนจุดที่มีค่าพลังงานสูงกว่าค่าพลังงานขีดจำกัด (Threshold Energy) ว่ามีมากพอหรือไม่ ถ้ามีมากพอ เราจะสรุปว่าเฟรมนี้มีเสียงพูด สถานะของระบบจะเปลี่ยนไปสู่ (2) นั่นคือเริ่มมีเสียงพูดเข้ามาแล้ว และเก็บเฟรมนี้ไว้เป็นเสียงพูด แต่ถ้าไม่ใช่ ระบบจะยังคงอยู่ที่สถานะเดิม คือยังไม่มีเสียงพูดเข้ามา แต่ระบบจะเก็บเฟรมนี้ไว้เพราะว่าอาจจะเป็นส่วนของพยัญชนะต้นในเสียงพูดที่มีค่าพลังงานไม่สูง พร้อมคำนวณค่าพลังงานเฉลี่ยของเฟรมก่อนหน้าให้เป็นค่าพลังงานของสภาวะแวดล้อม (Surround Energy) เพื่อนำไปใช้พิจารณาหาจุดสิ้นสุดของเสียงพูดต่อไป

เมื่อระบบเข้าสู่สถานะ (2) แล้ว ทุกเฟรมที่เข้ามาจะถูกนำมาคำนวณเพื่อหาว่าเป็นเสียงเงียบหรือไม่ โดยการนับจำนวนจุดที่มีค่าพลังงานต่ำกว่าค่าพลังงานของสภาวะแวดล้อมว่ามีมากพอหรือไม่ ถ้ามีมากพอ ระบบจะเปลี่ยนสถานะเป็น (3) ซึ่งเป็นสถานะของเสียงเงียบ แต่ถ้าไม่ใช่ ระบบจะอยู่ที่สถานะนี้ต่อไป พร้อมรวมเฟรมที่รับเข้ามานี้ไว้เป็นเสียงพูด

ที่สถานะ (3) ระบบจะทำการตรวจว่าเฟรมนี้เป็นเฟรมที่เป็นจุดสิ้นสุดของเสียงพูดจริงหรือไม่ หรือเป็นเพียงเสียงเงียบระหว่างพยางค์เท่านั้น โดยจะทำการคำนวณทุกเฟรมที่เข้ามาว่าเป็นเสียงเงียบหรือไม่ ถ้าทุกเฟรมที่เข้ามาเป็นเสียงเงียบติดต่อกันเกินค่าค่าหนึ่ง จะสรุปว่าเฟรมที่เป็นเสียงเงียบที่พบครั้งแรกเป็นจุดสิ้นสุดของเสียงพูด แต่ถ้าหากพบเฟรมที่ไม่เป็นเสียงเงียบก่อน ก็จะสรุปว่าเสียงเงียบที่พบเป็นเสียงเงียบระหว่างพยางค์ พร้อมรวมเสียงทั้งหมดเป็นเสียงพูด และกลับไปอยู่ที่สถานะ (2)

2.7 การสังเคราะห์เสียงพูด

โครงสร้างของระบบสังเคราะห์เสียงโดยทั่วไป สามารถแบ่งการทำงานได้เป็น 3 ส่วนได้แก่

2.7.1 ส่วนการวิเคราะห์ข้อความ (Text Analysis)

ส่วนนี้จะมีหน้าที่วิเคราะห์ข้อความอินพุทเพื่อแปลงข้อมูล เสียงอ่าน (Phoneme) ของคำนั้น และส่งต่อให้ส่วนของการสังเคราะห์เสียง (Speech Synthesis) ต่อไป นอกจากนี้ส่วนนี้ยังทำหน้าที่อย่างอื่น เช่น การแบ่งประโยคจากข้อความที่ยาว (Sentence Breaking) การหาขอบเขตของวลีของการอ่านในประโยค

2.7.2 ส่วนการวิเคราะห์สัทสัมพันธ์ (Prosody Analysis)

ส่วนนี้ทำหน้าที่ในการวิเคราะห์และสังเคราะห์ข้อมูล สัทสัมพันธ์ (Prosody) ของประโยคใดๆ จากข้อมูลเสียงอ่าน และข้อความ ข้อมูลสัทสัมพันธ์ที่วิเคราะห์ออกมาได้ในระบบทั่วไป เช่น

ช่วงเวลาตัดแบ่ง (Segment Duration) หมายถึง ความยาวของเสียงย่อยที่ต้องการสังเคราะห์ ค่านี้จะมีผลต่อจังหวะของเสียงที่ทำการสังเคราะห์ เช่น ถ้ากำหนดให้ค่าความยาวของเสียงย่อยที่ต้องการสังเคราะห์มีขนาดสั้น เสียงที่ทำการสังเคราะห์ก็จะเหมือนกับการพูดเร็ว

เส้นขอบระดับเสียง (Pitch Contour) หมายถึง ค่าความสัมพันธ์ของความถี่มูลฐานกับเวลา ค่านี้จะมีผลต่อเสียงสูงต่ำ

2.7.3 ส่วนการสังเคราะห์เสียง

ส่วนนี้ทำหน้าที่ในการสร้างสัญญาณคลื่นเสียง จากข้อมูลเสียงอ่าน (phonetic transcription) และข้อมูลสัทสัมพันธ์ (Prosody Transcription) จากส่วนของการวิเคราะห์ข้อความ และส่วนของการวิเคราะห์สัทสัมพันธ์ และส่งออกสู่ลำโพง เพื่อให้เราได้ยินเสียงพูดประโยคนั้นๆโดยทั่วไป ส่วนนี้สามารถแบ่งตามเทคนิควิธีการสังเคราะห์เสียง ได้ 3 ประเภท คือ

- Formant Synthesis

เทคนิคการสังเคราะห์วิธีแบบนี้ ข้อมูลเสียงอ่านใดๆจะถูกกำหนด ไว้อยู่ในรูปของความถี่ฟอร์แมนท์ต่างๆ (เช่น F1, F2, F3) ของเสียงนั้นๆ เมื่อต้องการสังเคราะห์เสียงใดๆก็นำข้อมูลเหล่านี้มาทำการสังเคราะห์ให้เป็นสัญญาณเสียง ซึ่งวิธีการนี้จะมีข้อดีที่สามารถควบคุมค่าความเปลี่ยนแปลงของความถี่ฟอร์แมนท์ (Formant transition) ที่บริเวณรอยต่อระหว่างเสียงได้ง่าย แต่มีข้อเสียคือ การจะแทนเสียงใดๆด้วยค่าฟอร์แมนท์ทำได้ยากจะต้องมีกฎในการสังเคราะห์เสียงจำนวนมาก และเสียงที่สังเคราะห์ออกมาได้จะไม่ค่อยเป็นธรรมชาติ

- Articulation Synthesis

สำหรับวิธีนี้ข้อมูลเสียงที่ต้องการสังเคราะห์ จะอยู่ในรูปของค่าพารามิเตอร์ของโครงสร้างทางกายภาพของการเคลื่อนไหวของอวัยวะในช่องปากที่ทำให้เกิดเสียงต่างๆ วิธีนี้ค่อนข้างยากในแง่การโมเดลเสียงต่างๆ ซึ่งจะต้องศึกษาจากอวัยวะในการออกเสียงจริงๆ

- Concatenation Synthesis

เสียงที่ทำการสังเคราะห์ขึ้นเกิดจากการนำหน่วยเสียงย่อย ที่ทำการเก็บไว้ก่อนแล้วมาต่อกันเป็นเสียงพูดที่ต้องการ โดยทั่วไปหน่วยเสียงย่อยที่ทำเก็บไว้จะอยู่ระดับต่ำกว่าคำ เช่น หน่วยเสียงพยางค์ หน่วยเสียงครึ่งพยางค์ (demesyllable) หน่วยเสียงของเสียงคู่เสียง (diphone) เป็นต้น

2.8 ทฤษฎีวิภัชนัย

ในงานทั่วๆไปที่ต้องมีการตัดสินใจ โดยที่ปัจจัยที่นำมาพิจารณาในการตัดสินใจมีความซับซ้อนกำกวม ยากต่อการตัดสินใจ จะพบว่ามนุษย์สามารถทำการตัดสินใจได้ดีกว่าเครื่องจักรกล ตัวอย่างเช่น ต้องการจะคัดส้มเพื่อแบ่งแยกราคาขาย ปัจจัยที่ใช้ในการพิจารณาไม่ใช่เพียงขนาดของส้มเท่านั้น แต่ยังต้องพิจารณาสีผิวภายนอก น้ำหนักหรือความแน่นของเนื้อส้มและน้ำส้มในผลส้ม นอกจากนี้ยังอาจจะมีปัจจัยนอกเหนือจากที่คาดคิดออกไปได้อีก เช่น สภาวะเศรษฐกิจ สถานะทางบ้าน เป็นต้น ในกรณีดังกล่าวนี้เราไม่สามารถสร้างสมการทางคณิตศาสตร์ที่ครอบคลุมปัจจัยเหล่านี้เพื่อใช้ในการตัดสินใจได้ง่าย ทฤษฎีทางด้านวิภัชนัยถูกสร้างขึ้นมาเพื่อกรณีเหล่านี้

ข้อมูลที่พบอยู่ในโลกปัจจุบันสามารถแบ่งออกเป็นข้อมูลที่แน่นอน (Certain Information) และข้อมูลที่มีความไม่แน่นอน (Uncertain Information) ซึ่งทฤษฎีเซตวิภัชนัยเป็นทฤษฎีทางคณิตศาสตร์ที่ใช้กับข้อมูลที่มีความไม่แน่นอน เช่นเดียวกับทฤษฎีทางสถิติ (Statistic Theory)

2.8.1 เซตวิภัชนัย

เซตวิภัชนัย คือ เซตที่มีการสนใจระดับมากน้อยของการเป็นสมาชิกแต่ละตัวต่างกับเซตคริสป์ (Crisp Set) ที่สนใจเพียงว่าสมาชิกที่สนใจอยู่ในเซตหรือไม่ โดยจะอธิบายดังต่อไปนี้

เซตแบบแน่นอน (Crisp Set)

กำหนดให้ A และ B เป็นเซตที่ประกอบด้วยสมาชิก (Member) ดังนี้

$$A = \{a,b,c\}$$

$$B = \{b,c,d\}$$

กำหนดให้ $\mu_X(y)$ เป็นค่าสมาชิกภาพ (Membership Value) ของ y ในเซต X ซึ่งมีความหมายคือ $\mu_X(y)$ เป็นตัวเลขที่แสดงความเป็นสมาชิกของ y ในเซต X ถ้า $\mu_X(y)$ มีค่ามากแสดงว่า y มีความเป็นสมาชิกในเซต X มาก เมื่อพิจารณาเซตแบบแน่นอน A และ B จะพบว่า a, b และ c เป็นสมาชิกของเซต A แน่นอน และ b, c และ d เป็นสมาชิกของเซต B แน่นอน ดังนั้นค่าสมาชิกภาพจะเป็นดังนี้

$$\mu_A(a) = 1 \qquad \mu_B(a) = 0$$
$$\mu_A(b) = 1 \qquad \mu_B(b) = 1$$
$$\mu_A(c) = 1 \qquad \mu_B(c) = 1$$
$$\mu_A(d) = 0 \qquad \mu_B(d) = 1$$

หรือสามารถเขียนรวมอยู่ในเซต ได้ดังนี้

$$A = \left\{\frac{a}{1},\frac{b}{1},\frac{c}{1},\frac{d}{0}\right\}$$

$$B = \left\{\frac{a}{0},\frac{b}{1},\frac{c}{1},\frac{d}{1}\right\}$$

ดังนั้นสำหรับเซตแบบแน่นอน ค่าสมาชิกภาพ $\mu_X(y)$ ของสมาชิกในเซตจะมีเพียง 1 หรือ 0 เท่านั้น สามารถเขียนสมการแสดงค่าสมาชิกอยู่ในรูปทั่วไปดังนี้

$$\mu_X(y) = \begin{cases} 1, & y \in X \\ 0, & y \notin X \end{cases}$$

เซตวิภัชนัย (Fuzzy Set) ประกอบด้วยสมาชิกที่มีค่าสมาชิกภาพ $\mu_X(y) \in \{0,1\}$ ยกตัวอย่างเช่น

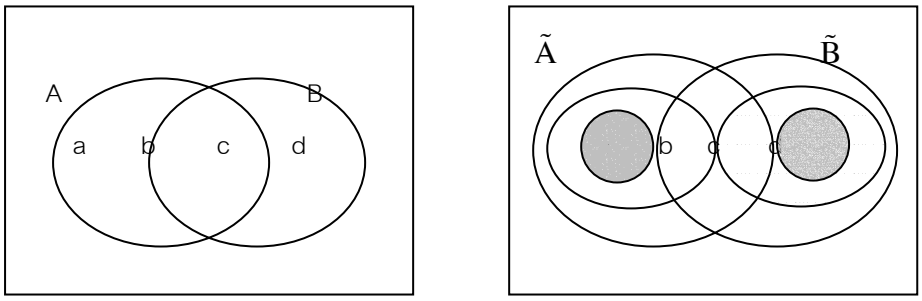
$$\tilde{A} = \left\{ \frac{a}{1}, \frac{b}{1}, \frac{c}{0.5}, \frac{d}{0} \right\}$$

$$\tilde{B} = \left\{ \frac{a}{0}, \frac{b}{0.4}, \frac{c}{0.9}, \frac{d}{1} \right\}$$

ซึ่งสามารถเขียนเซตวิภัชนัยอยู่ในรูปทั่วไปได้ดังนี้

$$\tilde{X} = \left\{ \frac{y}{\mu_{\tilde{X}}(y)} \right\}, \qquad \mu_{\tilde{X}}(y) \in \{0,1\}$$

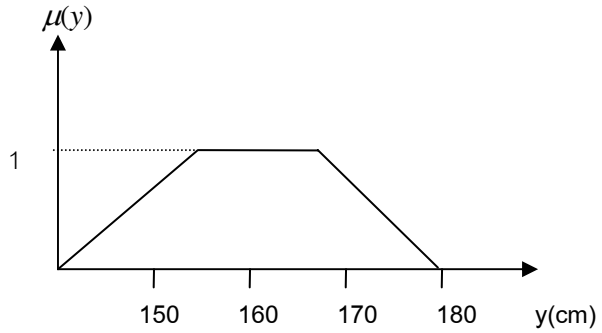
และสามารถเขียนเป็นแผนภาพของเซต A, B ซึ่งเป็นเซตแบบแน่นอน และเซต $\tilde{A}, \tilde{B}$ ซึ่งเป็นเซตวิภัชนัยได้ดังรูปที่ 2.22



รูปที่ 2.22 แผนภาพแสดงสมาชิกในเซต A,B และ $\tilde{A}, \tilde{B}$

2.8.2 ฟังก์ชันสมาชิกภาพแบบวิภัชนัย (Fuzzy Membership Function)

ฟังก์ชันสมาชิกภาพแบบวิภัชนัย คือ ฟังก์ชันที่ใช้หาค่าสมาชิกภาพ $\mu_X(y)$ ของสมาชิก y ในเซตแบบวิภัชนัย $\tilde{X}$ ในกรณีที่สมาชิกในเซตมีจำนวนจำกัด และไม่มากเกินไปสามารถเขียนแบบแจกแจงพร้อมทั้งแสดงค่าสมาชิกภาพได้ แต่ถ้าสมาชิกในเซตเป็นค่าต่อเนื่อง เช่น เซต $\tilde{A}$ เป็นเซตของความสูงเป็นเซนติเมตรของคนที่จัดว่าเป็นคนสูงปานกลาง สมาชิกของเซตนี้อาจจะมีค่าสมาชิกภาพเป็นศูนย์ในช่วงความสูงน้อยกว่า 150 เซนติเมตร และความสูงมากกว่า 180 เซนติเมตร ส่วนในช่วงความสูงระหว่าง 150 ถึง 180 เซนติเมตร จัดอยู่ในเซตนี้ด้วยค่าสมาชิกภาพต่าง ๆกัน ซึ่งนิยมแสดงฟังก์ชันสมาชิกภาพแบบวิภัชนัยเป็นรูปกราฟความสัมพันธ์ของค่าสมาชิกแบบวิภัชนัยเทียบกับค่าของสมาชิกทั้งหมดในเซต ดังแสดงกราฟในรูปที่ 2.23

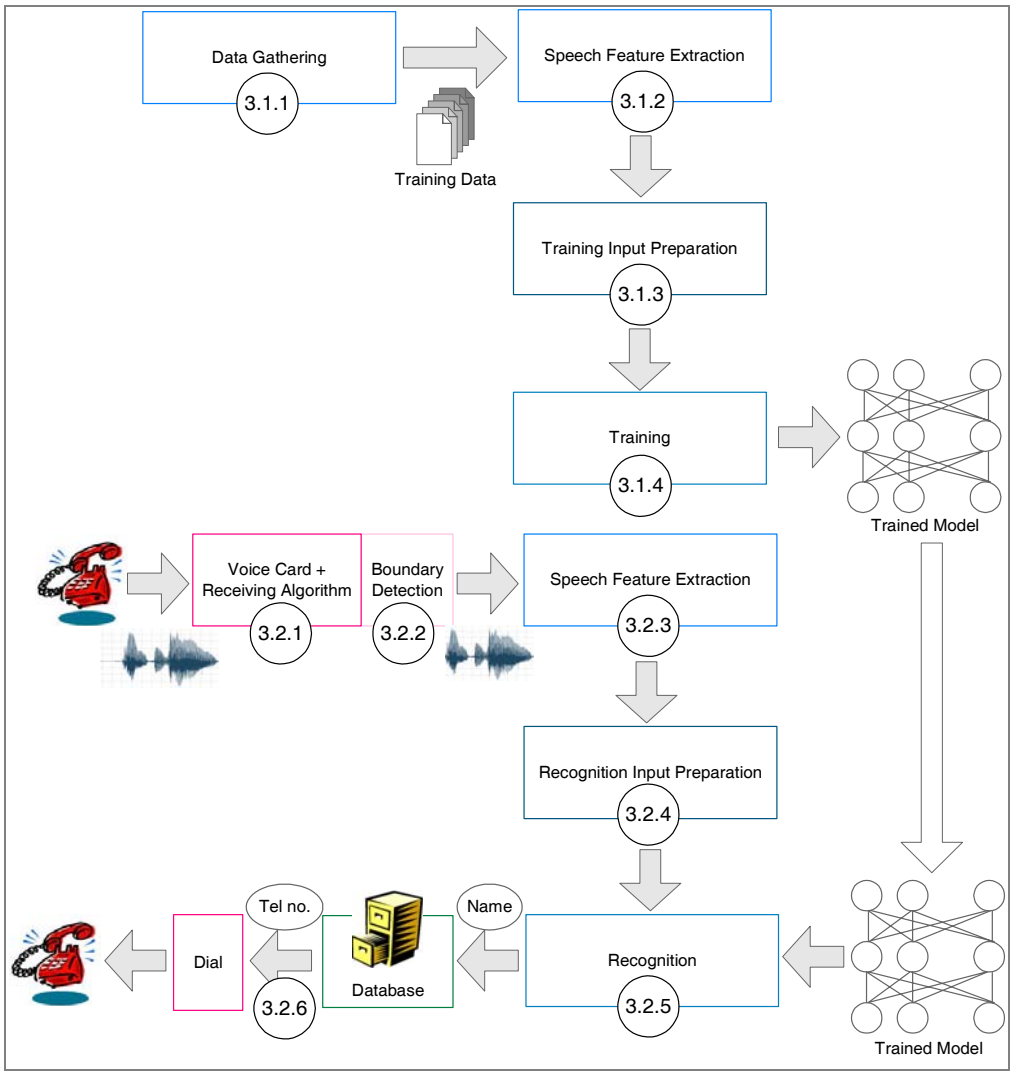


รูปที่ 2.23 กราฟความสัมพันธ์ของค่าสมาชิกแบบวิภัชนัยกับค่าของสมาชิกในเซต $\tilde{A}$ ซึ่งเป็นเซตของความสูงของคนที่จัดว่าสูงปานกลาง

3. ระบบโอนสายโทรศัพท์จากเสียงพูดชื่อไทย

จุดประสงค์หนึ่งของโครงการนี้ คือ การพัฒนาโปรแกรมประยุกต์ที่สามารถรู้จำเสียงพูดชื่อไทยทางโทรศัพท์ โดยให้รู้จำเสียงพูดชื่ออาจารย์และบุคลากรของภาควิชาวิศวกรรมคอมพิวเตอร์ จุฬาลงกรณ์มหาวิทยาลัย จากผู้ที่โทรเข้ามา แล้วทำการโอนสายไปยังห้องอาจารย์หรือบุคลากรคนนั้นโดยอัตโนมัติ ทำให้ผู้ที่โทรเข้ามาติดต่อภาควิชาไม่จำเป็นต้องจำเบอร์ภายในทั้งหมดของภาควิชา ขณะเดียวกันก็เป็นการลดภาระของโอเปอเรเตอร์ในภาควิชาไปในตัว หรือแม้แต่การโทรภายในภาควิชาระหว่างอาจารย์และบุคลากรก็ก่อให้เกิดความสะดวกรวดเร็ว เนื่องจากไม่ต้องจำเบอร์ของอาจารย์และบุคลากรด้วยกัน

ระบบรู้จำเสียงพูดที่ได้พัฒนาขึ้นทำการรู้จำเสียงพูดระดับคำโดยไม่ขึ้นกับผู้พูด โดยมีสถาปัตยกรรมของระบบเป็นดังรูปที่ 3.1



รูปที่ 3.1 สถาปัตยกรรมของระบบ

ระบบโอนสายอัตโนมัติจากเสียงพูดชื่อไทย แบ่งออกเป็นสองส่วน คือส่วนที่เป็นกระบวนการเรียนรู้ และส่วนที่เป็นกระบวนการรู้จำ ซึ่งแต่ละส่วนมีกระบวนการต่างๆ ดังนี้

3.1 กระบวนการเรียนรู้

กระบวนการเรียนรู้มีขั้นตอนต่างๆ ดังนี้

3.1.1 การเก็บตัวอย่างเสียงพูดเพื่อการเรียนรู้

กระบวนการเรียนรู้เริ่มจากการเก็บตัวอย่างเสียงพูดมาเป็นข้อมูลการเรียนรู้ (Training Data) โดยจะเก็บข้อมูลเหล่านี้ผ่านทางโทรศัพท์ เพื่อให้ข้อมูลที่ใช้ฝึกสอนมีรูปแบบเหมือนกันกับข้อมูลในการรู้จำเมื่อนำไปใช้งานจริง ซึ่งการ์ดเสียง (Voice Card) จะทำการแปลงเสียงพูดทางโทรศัพท์ที่เข้ามาเป็นไฟล์เสียง (Wav File) ที่มีอัตราการชักตัวอย่าง (Sampling Rate) 11,025 ตัวอย่างต่อวินาที แต่ละตัวอย่างแทนด้วยหน่วยความจำ 8 บิต

การ์ดเสียงที่ใช้เป็นการ์ดเสียงของ Dialogic® รุ่น D/4PCI ซึ่งจะควบคุมการทำงานผ่านโปรแกรมประสานประยุกต์ (Application Program Interface Reference, API Reference) [93] ในการเก็บตัวอย่างเสียงพูดเพื่อทำการเรียนรู้ การ์ดเสียงจะทำงานตามลำดับดังนี้

- **การเปิดช่องอุปกรณ์** ในการทำงานร่วมกับการ์ดเสียง ผู้เขียนโปรแกรมจะต้องทำการเปิดช่องอุปกรณ์ (Device Channel) ก่อน โดยใช้คำสั่ง

  ```
  chdev = dx_open("dxxxB1C1", NULL);
  ```

  โดยคำสั่ง dx_open นี้ จะทำการเปิดช่องสัญญาณ dxxxB1C1 ซึ่งเป็นช่องสำหรับต่อเอาสายโทรศัพท์เข้าเครื่องคอมพิวเตอร์โดยผ่านการ์ดเสียง การ์ดเสียงรุ่นที่ใช้นี้จะมีช่องสำหรับต่ออยู่ทั้งหมด 4 ช่อง ได้แก่ dxxxB1C1, dxxxB1C2, dxxxB1C3 และ dxxxB1C4 ซึ่งสามารถนำคู่สายโทรศัพท์เข้ามาต่อได้ทั้งหมด 4 หมายเลข ด้วยคำสั่งนี้ จะคืนค่าหมายเลขอ้างอิงช่องอุปกรณ์มาคือ chdev เพื่อนำไปใช้อ้างอิงในส่วนอื่นๆ ของโปรแกรมเมื่อมีการเรียกใช้อุปกรณ์ แต่หากเปิดไม่สำเร็จ จะคืนค่า -1 ออกมาแทน สาเหตุที่ไม่สามารถเปิดช่องอุปกรณ์ได้ อาจจะเกิดจากพารามิเตอร์ที่ใช้ไม่ถูกต้อง ด้วยคำสั่ง dx_open() นี้ทำให้เราได้หมายเลขสำหรับอ้างอิงมา เพื่อจะได้กระทำคำสั่งอื่นๆ ต่อไป เช่น การเล่นเสียง การบันทึกเสียง การโอนสาย เป็นต้น

- **การตอบรับสัญญาณเรียกเข้า** ในขณะที่รอการโทรเข้ามาของผู้ใช้ระบบ โปรแกรมจะต้องรอรับสัญญาณสายเรียกเข้าและเมื่อมีสายเรียกเข้าเข้ามา โปรแกรมจะต้องกระทำการบางอย่างเพื่อตอบรับการเรียกเข้านั้น คำสั่งสำหรับรอสัญญาณเรียกเข้าก็คือ

  ```
  dx_wtring(chdev, 2, DX_OFFHOOK, -1);
  ```

  คำสั่งนี้สามารถระบุจำนวนครั้งของสัญญาณเรียกเข้าได้ว่าดังกี่ครั้งจึงจะกระทำการต่อไป เช่น ปล่อยให้สัญญาณดัง 2 ครั้งจึงยกหูโทรศัพท์ (OFFHOOK) ส่วน -1 หมายความว่าเมื่อไม่มีสัญญาณเรียกเข้าเข้ามา ก็ให้รอไปเรื่อยๆ โดยไม่มีกำหนดเวลา

- **การบันทึกเสียง** การบันทึกเสียงด้วยชุดคำสั่งของ Dialogic® สามารถบันทึกเป็นไฟล์หรือบันทึกลงบัฟเฟอร์บนหน่วยความจำหลักก็ได้ สำหรับตัวอย่างเสียงพูดจะบันทึกลงไฟล์โดยใช้คำสั่ง

  ```
  dx_recwav(chdev, pathName, tptRec, (DX_XPB *)NULL, PM_TONE | EV_SYNC);
  ```

  ซึ่งผู้เขียนโปรแกรมสามารถระบุพารามิเตอร์ต่างๆ ตามต้องการดังนี้

  - tptRec ระบุสาเหตุที่ทำให้หยุดการบันทึกเสียงลง เช่น การกดปุ่ม การบันทึกเสียงเกินเวลาที่กำหนด หรือการเกิดเสียงเงียบเกินเวลาที่กำหนด
  - pathName เป็นชื่อไฟล์ที่ต้องการบันทึก

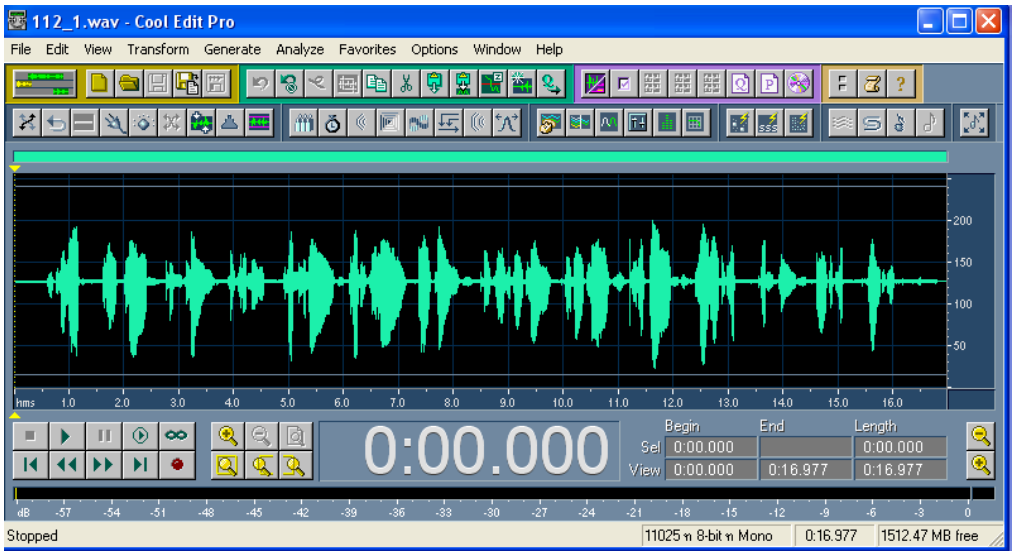
- **การปิดช่องอุปกรณ์** เมื่อไม่มีความต้องการที่จะใช้ช่องอุปกรณ์อีก ผู้เขียนโปรแกรมจะต้องปิดช่องอุปกรณ์โดยใช้คำสั่งดังนี้

  ```
  dx_close(chdev);
  ```

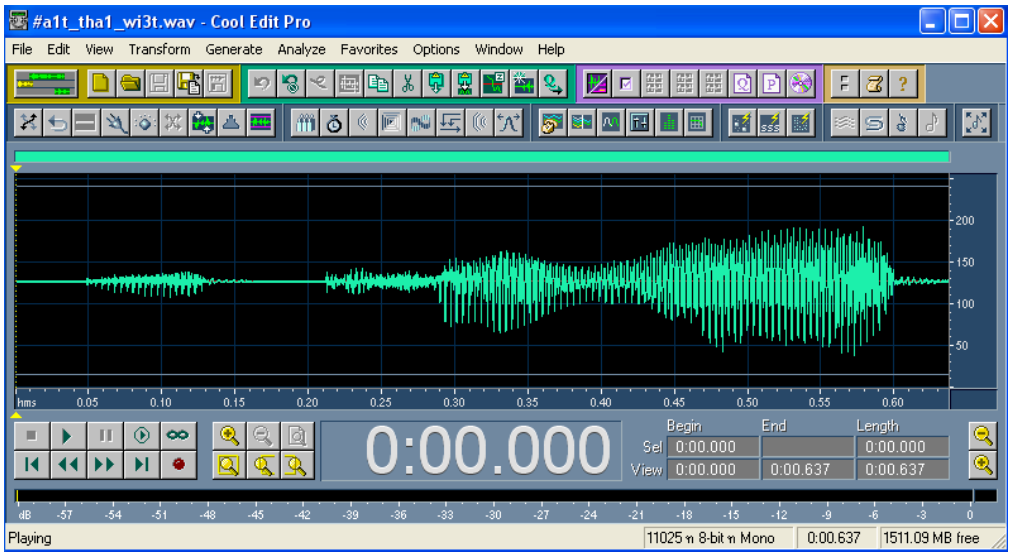
  โดย chdev คือหมายเลขอ้างอิงช่องอุปกรณ์ที่เปิดเอาไว้และต้องการจะปิด

จากนั้นจะทำการตัดไฟล์เสียงให้เหลือแต่เฉพาะส่วนที่เป็นเสียงพูดชื่อไทย และทำการบอกว่าตัวอย่างนั้นคือเสียงพูดชื่อใคร (Labeling) โดยใช้โปรแกรม Cool Edit Pro 1.0

การเก็บตัวอย่างเสียงพูดเพื่อการเรียนรู้เริ่มจากการที่ผู้บันทึกเสียงโทรศัพท์เข้ามาแล้วพูดทุกชื่อ การ์ดเสียงจะบันทึกเสียงพูดลงในไฟล์ ซึ่งมีลักษณะดังรูปที่ 3.2 จากนั้นจะใช้คนมาตัดเสียงให้เหลือเพียงไฟล์ละชื่อ แล้วทำการตั้งชื่อไฟล์ให้เป็นชื่อเดียวกับชื่อที่ถูกพูดในไฟล์นั้น ดังรูปที่ 3.3



รูปที่ 3.2 ไฟล์เสียงพูดชื่อไทยที่บันทึกผ่านทางโทรศัพท์



รูปที่ 3.3 ไฟล์เสียงพูดชื่อไทยที่ผ่านการตัดเสียงแล้ว

3.1.2 การหาลักษณะสำคัญของเสียงเพื่อการเรียนรู้

การหาลักษณะสำคัญของเสียงในที่นี้ใช้วิธีการทำนายเชิงเส้นแบบรับรู้ดังที่กล่าวไว้ในหัวข้อ 2.4 โดยใช้โปรแกรมเอ็กซ์ตราราสต้าพีแอลพี (ExtraRastaPLP) ซึ่งดัดแปลงจากโปรแกรมต้นแบบของ International Computer Science Institute (ICSI) [94]

โปรแกรมเอ็กซ์ตราราสต้าพีแอลพีจะทำกระบวนการ 3.1.2 และ 0 ไปพร้อมกัน การหาลักษณะสำคัญของเสียงโดยใช้การทำนายเชิงเส้นแบบรับรู้ต้องปรับค่าพารามิเตอร์ต่างๆ ซึ่งจะระบุไว้ในไฟล์ RastaPLPParam.cfg ดัง

ตารางที่ 3.1

ตารางที่ 3.1 ค่าพารามิเตอร์ในการหาลักษณะสำคัญของเสียง

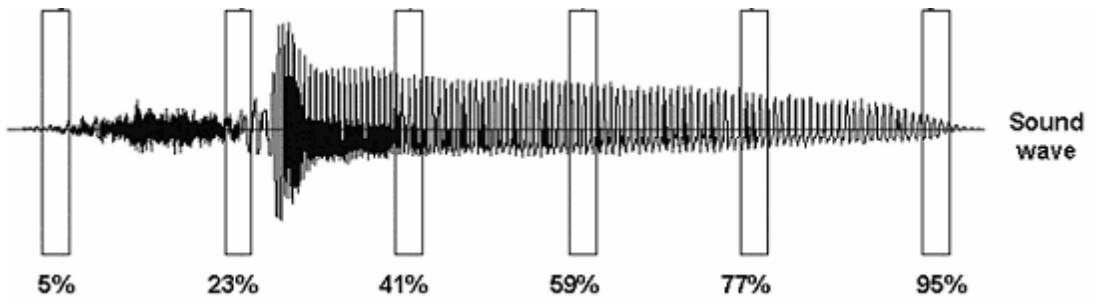
```
***** Configuration file for ExtraRastaPLP program
*  float winsize; /* analysis window size in msec */
#w 25
*  float stepsize; /* analysis step size in msec */
#s 12.5
*  int padInput; /* if true, pad input so the (n * steppts)-th sample is centered in the n-th frame */
#y 0
*  int sampfreq; /* sampling frequency in Hertz */
#S 11025
*  int nfilts; /* number of critical band filters used */
#c 0
*  int trapezoidal; /* set true if the auditory filters are trapezoidal */
#v 1
*  float winco; /* window coefficient */
```

```
#W 0.54
*  float polepos; /* rasta integrator pole position */
#P 0.94
*  int order; /* LPC model order */
#m 9
*  int nout; /* length of final feature vector*/
#n 0
*  int gainflag; /* flag that says to use gain */
#g 1
*  float lift; /* cepstral lifter exponent */
#l 0.6
*  int lrasta; /* set true if log rasta used */
#L 0
*  int jrasta; /* set true if jah rasta used */
#J 0
*  int cJah; /* set true if constant Jah used */
#C 0
*  char *mapcoef_fname; /* Jah Rasta mapping coefficients input text file name */
#f 0
*  int crbout; /* set true if critical band values after bandpass filtering instead of
*                 cepstral coefficients are desired as outputs */
#R 0
*  int comcrbout; /* set true if critical band values after cube root compression
*                       and equalization are desired as outputs. */
#P 0
*  float rfrac; /* fraction of rasta mixed with plp */
#r 1
*  float jah; /* Jah constant */
#j 1.0e-6
*  char *infname; /* Input file name, where "-" means stdin */
#i *.wav
*  char *outfname; /* Output file name, where "-" means stdout */
#o *.plp
*  int ascin; /* if true, read ascii in */
#a 0
*  int ascout; /* if true, write ascii out */
#A 0
*  int spherein /* read SPHERE input */
#z 0
*  int abbotIO; /* read and write CUED's Abbot wav format */
```

```
#k 0
*  int inswapbytes; /* swap bytes of input */
#T 0
*  int outswapbytes; /* swap bytes of output */
#U 0
*  int debug; /* enable debug info */
#d 1
*  int smallmask; /* add small constant to power spectrum */
#M 0
*  int online; /* online, not batch file processing */
#O 0
*  int HPfilter; /* highpass filter on waveform is used */
#F 0
*  int history;   /* use stored noise level estimation and RASTA filter history for initialization */
#h 0
*  char *hist_fname; /* history filename */
*#H history.out
*    /* incorporated delta additions */
*    int deltawindow; /* number of frames to use in delta calc */
#Q 9
*    int deltaorder; /* 0=no deltas, 1=just deltas, 2=d+dd */
#q 0
```

3.1.3 การเตรียมข้อมูลเพื่อนำไปใช้ในการเรียนรู้

โปรแกรมเอ็กซ์ตราราสต้าพีแอลพีนอกจากจะทำการหาลักษณะสำคัญของเสียงพูดแล้ว ยังทำการเตรียมข้อมูลเพื่อนำไปใช้ในการเรียนรู้ด้วยโครงข่ายประสาทเทียม โดยเมื่อทำการหาลักษณะสำคัญของเสียงพูดในทุกเฟรมแล้ว เอ็กซ์ตราราสต้าพีแอลพีจะทำการเลือกเฟรมสำหรับนำไปใช้ในการเรียนรู้ โดยเลือกให้กระจายครอบคลุมช่วงทั้งหมดของเสียงพูด ดังรูปที่ 3.4 ซึ่งแสดงถึงการเลือก 6 เฟรมในตำแหน่งต่างๆ กันของเสียงพูดตั้งแต่ต้นจนจบโดยเทียบเป็นเปอร์เซ็นต์



รูปที่ 3.4 การเลือกเฟรมเพื่อนำไปใช้ในการเรียนรู้

การใช้โปรแกรมเอ็กซ์ตราราสต้าพีแอลพีจะต้องกำหนดค่าพารามิเตอร์ดังตารางที่ 3.2

ตารางที่ 3.2 ค่าพารามิเตอร์ของโปรแกรมเอ็กซ์ตราราสต้าพีแอลพี

```
ExtraRastaPLP RastaPLPParam.cfg *.wav DirTrainList.txt FrameList.txt NameList.txt Nico\NicoTr.in
Nico\NicoTr.tar

ExtraRastaPLP RastaPLPParam.cfg *.wav DirTestList.txt FrameList.txt NameList.txt Nico\NicoTe.in
Nico\NicoTe.tar
```

โดยที่

- RastaPLPParam.cfg เป็นไฟล์ที่กำหนดค่าพารามิเตอร์ในการหาลักษณะสำคัญของเสียง ดังแสดงใน
- ตารางที่ 3.1
- *.wav บอกว่าให้โปรแกรมทำการประมวลผลกับทุก wav ไฟล์
- DirTrainList.txt เป็นไฟล์ที่บอกชื่อไดเรคทอรีที่มีไฟล์เสียงสำหรับการเรียนรู้
- DirTestList.txt เป็นไฟล์ที่บอกชื่อไดเรคทอรีที่มีไฟล์เสียงสำหรับการทดสอบ
- FrameList.txt เป็นไฟล์ที่บอกตำแหน่งของเฟรม โดยบอกเป็นเปอร์เซ็นต์เทียบกับความยาวของเสียงพูด
- NameList.txt เป็นไฟล์ที่บอกชื่อผลลัพธ์ หรือชื่อไทยทั้งหมดที่ต้องการรู้จำ
- Nico\NicoTr.in บอกชื่อไฟล์ผลลัพธ์ซึ่งเป็นอินพุตของโครงข่ายประสาทเทียมสำหรับข้อมูลที่ใช้ในการเรียนรู้
- Nico\NicoTe.in บอกชื่อไฟล์ผลลัพธ์ซึ่งเป็นอินพุตของโครงข่ายประสาทเทียมสำหรับข้อมูลที่ใช้ในการทดสอบ
- Nico\NicoTr.tar บอกชื่อไฟล์ผลลัพธ์ซึ่งเป็นเอาต์พุตของโครงข่ายประสาทเทียมสำหรับข้อมูลที่ใช้ในการเรียนรู้
- Nico\NicoTe.tar บอกชื่อไฟล์ผลลัพธ์ซึ่งเป็นเอาต์พุตของโครงข่ายประสาทเทียมสำหรับข้อมูลที่ใช้ในการทดสอบ

รูปที่ 3.5 แสดงการรันโปรแกรมเอ็กซ์ตราราสต้าพีแอลพี

รูปที่ 3.5 การรันโปรแกรมเอ็กซ์ตราราสต้าพีแอลพี

3.1.4 การเรียนรู้

การเรียนรู้ในที่นี้ใช้โครงข่ายประสาทเทียมดังที่กล่าวไว้ในหัวข้อ 2.5 โดยใช้โปรแกรม Nico Toolkit [95] [96] เป็นเครื่องมือสำหรับสร้างโครงข่ายประสาทเทียมเพื่อใช้ในการเรียนรู้ และเป็นเครื่องมือสำหรับการเรียนรู้ด้วยโครงข่ายประสาทเทียมนั้น โดยโครงข่ายประสาทเทียมจะมีจำนวนโนดในชั้นเอาต์พุตเท่ากับจำนวนชื่อไทยทั้งหมดที่ต้องการรู้จำ จำนวนโนดในชั้นฮิดเดนตามความเหมาะสม และจำนวนโนดในชั้นอินพุตเท่ากับจำนวนเฟรมที่ใช้คูณด้วยจำนวนลักษณะสำคัญของเสียงที่หาได้ในแต่ละเฟรม โดยจำนวนลักษณะสำคัญของเสียงที่หาได้ในแต่ละเฟรมจะเท่ากับอันดับในการทำนายเชิงเส้นแบบรับรู้ จากนั้นทำการเรียนรู้ด้วยวิธีแบ็กพรอพาเกชัน จนได้ผลการเรียนรู้เพื่อที่จะนำไปใช้ในกระบวนการรู้จำต่อไป

ชุดคำสั่งสำหรับสร้างโครงข่ายประสาทเทียม และเรียนรู้ด้วยโครงข่ายประสาทเทียมนั้นแสดงได้ดังตารางที่ 3.3

ตารางที่ 3.3 ชุดคำสั่งในการสร้างโครงข่ายประสาทเทียมและเรียนรู้ด้วยโครงข่ายประสาทเทียม

```
CreateNet NicoTr NicoTr.rtd

AddStream -x in -d . -F binary 423 r INPUT NicoTr.rtd
AddGroup input NicoTr.rtd
AddUnit -i -u 423 input NicoTr.rtd
LinkGroup INPUT input NicoTr.rtd

AddStream -x tar -d . -F binary 45 t OUTPUT NicoTr.rtd
AddGroup output NicoTr.rtd
AddUnit -o -u 45 -S ..\AjarnList.txt output NicoTr.rtd
LinkGroup OUTPUT output NicoTr.rtd

AddGroup hidden NicoTr.rtd
AddUnit -u 100 hidden NicoTr.rtd

NormStream -s INPUT -d 1.64 NicoTr.rtd NicoTr.in

Connect input hidden NicoTr.rtd
Connect hidden output NicoTr.rtd
Display NicoTr.rtd

BackProp -U 80.0 -d –E -m 0.7 -g 1e-4 -i 1000 -P NicoTr.log 1 1 NicoTr.rtd NicoTr

CResult -c NicoTr.rtd NicoTr.in>tr1.out
CResult -c NicoTr.rtd NicoTe.in>te1.out
CResult -c NicoTr.upd NicoTr.in>trb1.out
CResult -c NicoTr.upd NicoTe.in>teb1.out
```

โดยที่

- CreateNet เป็นคำสั่งสำหรับสร้างโครงข่ายประสาทเทียมขึ้นมาตัวหนึ่ง ซึ่งมีรูปแบบการใช้คือ

  CreateNet $NetName $NetFileName

  - $NetName เป็นชื่อของโครงข่ายประสาทเทียม
  - $NetFileName เป็นชื่อไฟล์ที่เก็บข้อมูลของโครงข่ายประสาทเทียมที่สร้างขึ้นมา

- AddStream สำหรับรับข้อมูลจากไฟล์มาเก็บไว้เป็นเวคเตอร์ของจำนวนจริง หรือสตรีม เพื่อเตรียมพร้อมในการเรียนรู้ต่อไป ซึ่งมีรูปแบบการใช้คือ

  AddStream [options] nSize $Mode $StreamName $NetFileName

  - nSize เป็นข้อมูลที่จะถูกอ่านในแต่ละรอบของการเรียนรู้ ซึ่งจะมีจำนวนเท่ากับจำนวนโนดในชั้นอินพุตในกรณีที่เป็นสตรีมของอินพุต และจะมีจำนวนเท่ากับจำนวนโนดในชั้นเอาต์พุตในกรณีที่เป็นสตรีมของเอาต์พุต
  - $Mode มีค่าเป็น r สำหรับสตรีมของอินพุต และมีค่าเป็น t สำหรับสตรีมของเอาต์พุต
  - $StreamName เป็นชื่อของสตรีม
  - $NetFileName เป็นชื่อไฟล์ของโครงข่ายประสาทเทียมที่ต้องใช้สตรีมนี้

  สำหรับ [options] มีดังนี้
  - -x $FileExtension บอกให้อ่านไฟล์นามสกุล $FileExtension เข้ามาในสตรีม
  - -d $Dir บอกไดเรคทอรีของไฟล์ที่จะอ่าน
  - -F $Format บอกรูปแบบของไฟล์ที่จะอ่าน

- AddGroup สำหรับสร้างกลุ่มต่างๆ ในโครงข่ายประสาทเทียม ซึ่งมีรูปแบบการใช้คือ

  AddGroup [options] $GroupName $NetFileName

  - $GroupName เป็นชื่อกลุ่มที่ต้องการสร้าง
  - $NetFileName เป็นชื่อไฟล์ของโครงข่ายประสาทเทียมที่จะบรรจุกลุ่มที่สร้าง

- AddUnit สำหรับสร้างโนดในกลุ่มต่างๆ ของโครงข่ายประสาทเทียม ซึ่งมีรูปแบบการใช้คือ

  AddUnit [options] $GroupName $NetFileName

  - $GroupName เป็นชื่อกลุ่มที่ต้องการสร้างโนด
  - $NetFileName เป็นชื่อไฟล์ของโครงข่ายประสาทเทียมที่จะบรรจุโนดในกลุ่มที่สร้าง

  สำหรับ [options] มีดังนี้
  - -i บอกว่าเป็นโนดอินพุต ซึ่งจะต้องเชื่อมต่อกับอินพุตสตรีม
  - -o บอกว่าเป็นโนดเอาต์พุต ซึ่งจะต้องเชื่อมต่อกับเอาต์พุตสตรีม
  - -u nSize บอกให้สร้างโนดในกลุ่มขึ้นมาจำนวน nSize โนด
  - -S $File เป็นการอ่าน $File ซึ่งจะบอกชื่อของแต่ละโนด

- LinkGroup สำหรับเชื่อมกลุ่มและสตรีมเข้าด้วยกัน ซึ่งมีรูปแบบการใช้คือ

  LinkGroup $StreamName $GroupName $NetFileName

  - $StreamName เป็นชื่อสตรีมที่ต้องการเชื่อมกับกลุ่ม
  - $GroupName เป็นชื่อกลุ่มที่ต้องการเชื่อมกับสตรีม
  - $NetFileName เป็นชื่อไฟล์ของโครงข่ายประสาทเทียมที่บรรจุสตรีมและกลุ่มที่จะนำมาเชื่อมกัน

- NormStream เป็นการปรับบรรทัดฐานของสตรีมให้เกือบทุกค่าอยู่ในช่วง [-1,1] ซึ่งจะเหมาะสมกับการเรียนรู้ด้วยโครงข่ายประสาทเทียมมากกว่าค่าของสตรีมที่กระจายในช่วงกว้าง มีรูปแบบการใช้คือ

  NormStream [options] $NetFileName [$InputFileName]

  - $NetFileName เป็นชื่อไฟล์ของโครงข่ายประสาทเทียมที่บรรจุสตรีมที่จะนำมาปรับบรรทัดฐาน
  - $InputFileName บอกชื่อไฟล์ที่จะนำมาปรับบรรทัดฐาน

  สำหรับ [options] มีดังนี้
  - -s $StreamName บอกชื่อสตรีมที่จะทำการปรับบรรทัดฐาน
  - -d mult บอกว่าค่าอินพุตที่อยู่ในช่วง $\bar{x} \pm \sigma \cdot mult$ จะถูกปรับบรรทัดฐานให้อยู่ในช่วง [-1,1] เช่นถ้า $mult = 1.64$ จะมีอินพุตอยู่ในช่วง [-1,1] ประมาณ 90% เป็นต้น

- Connect สำหรับเชื่อมกลุ่มต่างๆ เข้าด้วยกัน ซึ่งมีรูปแบบการใช้คือ

  Connect [options] $FromGroupName $ToGroupName $NetFileName

  - $FromGroupName เป็นชื่อกลุ่มที่ต้องการเชื่อมไป
  - $ToGroupName เป็นชื่อกลุ่มที่ต้องการเชื่อมด้วยกับกลุ่ม $FromGroupName
  - $NetFileName เป็นชื่อไฟล์ของโครงข่ายประสาทเทียมที่จะบรรจุโนดในกลุ่มที่จะเชื่อมต่อกันนั้น

- Display สำหรับแสดงรายละเอียดในโครงสร้างของโครงข่ายประสาทเทียมที่สร้างขึ้น

- BackProp เป็นคำสั่งหลักเพื่อใช้ในการเรียนรู้ด้วยโครงข่ายประสาทเทียมแบบแบ็กพรอพาเกชัน ซึ่งมีรูปแบบการใช้คือ

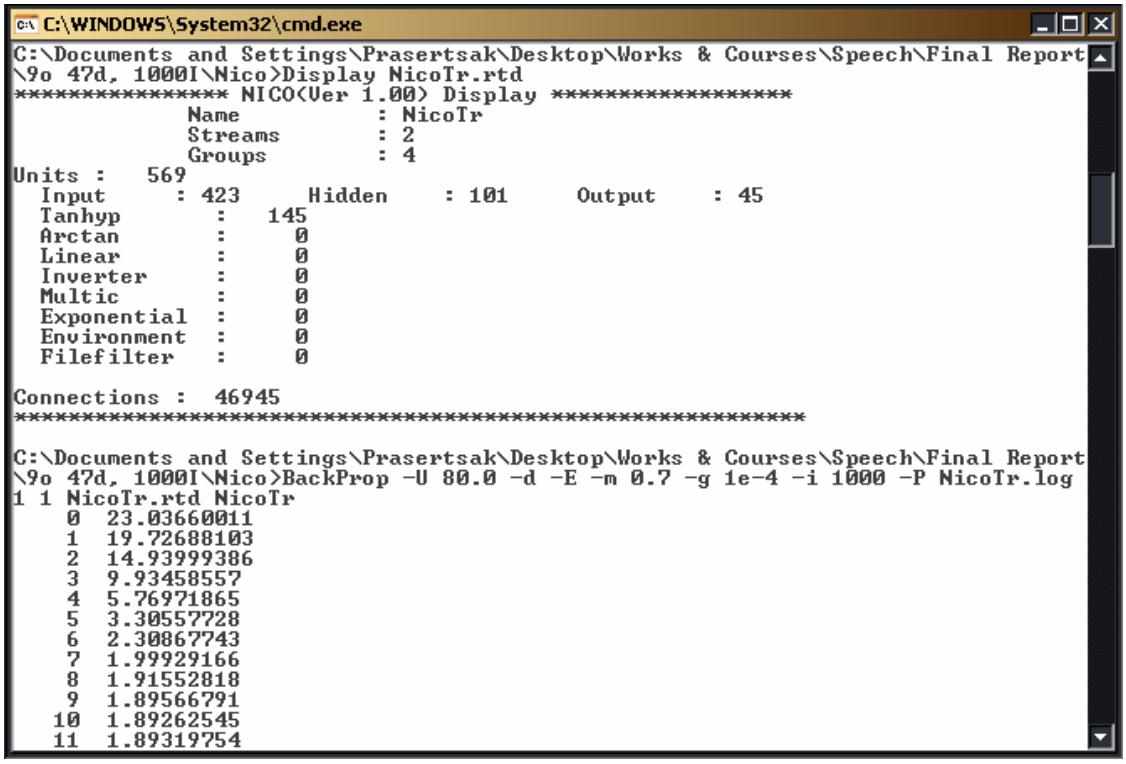
  BackProp [options] $NetFileName $Input

  - $NetFileName เป็นชื่อไฟล์ของโครงข่ายประสาทเทียมที่จะนำมาทำการเรียนรู้
  - $Input เป็นชื่อไฟล์ที่ใช้ในกาเรียนรู้

  สำหรับ [options] มีดังนี้
  - -U percent บอกให้เก็บโครงข่ายประสาทเทียมที่ให้ค่าความถูกต้องเกิน percent ไว้ที่ไฟล์ .upd
  - -E กำหนดให้ทำการปรับปรุงน้ำหนักของโครงข่ายประสาทเทียมเมื่อข้อมูลทุกตัวถูกนำเข้ามาประมวลผลแล้ว
  - -d หมายถึงให้เก็บข้อมูลที่ใช้ในการเรียนรู้ไว้ที่หน่วยความจำหลักเพื่อความสะดวกรวดเร็วในการเรียนรู้
  - -m momentum กำหนดค่าโมเมนตัมที่ใช้ในการเรียนรู้
  - -g gain กำหนดค่าอัตราการเรียนรู้ของโครงข่ายประสาทเทียม
  - -i iter กำหนดว่าจะวนเรียนรู้กี่รอบ

- -P $LogFileName logupdate netupdate เป็นการกำหนดให้รายงานความก้าวหน้าในการเรียนรู้ไว้ที่ไฟล์ชื่อ $LogFileName โดยจะปรับปรุงไฟล์นี้ทุกๆ logupdate รอบของการเรียนรู้ ในขณะเดียวกันไฟล์ที่เก็บข้อมูลของโครงข่ายประสาทเทียมก็จะได้รับการปรับปรุงทุกๆ netupdate รอบของการเรียนรู้

รูปที่ 3.6 แสดงการรันโปรแกรมเพื่อการเรียนรู้โดยใช้โครงข่ายประสาทเทียม



รูปที่ 3.6 การรันโปรแกรมเพื่อการเรียนรู้โดยใช้โครงข่ายประสาทเทียม

3.2 กระบวนการรู้จำ

กระบวนการรู้จำเป็นการนำผลลัพธ์จากกระบวนการเรียนรู้ ซึ่งในที่นี้คือโครงข่ายประสาทเทียมที่ได้ทำการเรียนรู้แล้ว ไปใช้งานจริงเพื่อรู้จำเสียงพูดชื่อไทยทางโทรศัพท์และโอนสายไปยังเบอร์ภายในของชื่อไทยนั้นโดยอัตโนมัติ กระบวนการรู้จำในที่นี้ทำผ่านโปรแกรมสปีชคอล (SpeechCall) ซึ่งมีวิธีใช้ดังที่จะอธิบายใน ภาคผนวก – โปรแกรมโอนสายอัตโนมัติจากเสียงพูดชื่อไทยทางโทรศัพท์

กระบวนการรู้จำที่ใช้ภายในโปรแกรมสปีชคอลมีขั้นตอนต่างๆ ดังนี้

3.2.1 การรับเสียงพูดทางโทรศัพท์

เมื่อมีผู้ใช้โทรศัพท์เข้ามายังเลขหมายที่ให้บริการและพูดชื่อไทยที่ต้องการติดต่อ การ์ดเสียงจะแปลงเสียงพูดทางโทรศัพท์เป็นข้อมูลเพื่อใช้ประมวลผล การทำงานของการ์ดเสียงในการรับเสียงพูดทางโทรศัพท์จะคล้ายกับการทำงานของการ์ดเสียงในการเก็บตัวอย่างเสียงพูดเพื่อการเรียนรู้ นั่นคือมีการเปิดช่องอุปกรณ์ การตอบรับสัญญาณเรียกเข้า และการปิดช่องอุปกรณ์ ส่วนการบันทึกเสียงจะ

ทำแตกต่างออกไป คือแทนที่จะบันทึกเสียงพูดลงไฟล์ ก็จะบันทึกลงบัฟเฟอร์แทน เพื่อนำข้อมูลเสียงพูดนั้นมาใช้ประมวลผลต่อไป การบันทึกเสียงพูดลงบัฟเฟอร์ทำได้โดยใช้คำสั่ง

```
dx_reciotttdata(chdev, iott, tptRec, xpb, mode);
```

โดยที่

- iott ระบุข้อมูลของบัฟเฟอร์ที่จะบันทึก
- tptRec ระบุสาเหตุที่ทำให้หยุดการบันทึกเสียงลง เช่น การกดปุ่ม การบันทึกเสียงเกินเวลาที่กำหนด หรือการเกิดเสียงเงียบเกินเวลาที่กำหนด
- xpb ระบุพารามิเตอร์ของเสียงที่บันทึก เช่น อัตราการชักตัวอย่าง และจำนวนบิตของแต่ละตัวอย่าง เป็นต้น ซึ่งค่าเหล่านี้จะต้องตรงกันกับค่าของข้อมูลที่ใช้ในการเรียนรู้ ซึ่งในที่นี้ใช้อัตราการชักตัวอย่าง 11025 ตัวอย่างต่อวินาที และแต่ละตัวอย่างแทนด้วยหน่วยความจำ 8 บิต

นอกจากนี้ ในการรับเสียงพูดทางโทรศัพท์ จำเป็นต้องใช้การ์ดเสียงทำงานอื่นๆ ดังนี้

- **การตรวจจับการกดปุ่ม** การตรวจจับการกดปุ่มทำให้โปรแกรมทราบว่าผู้ใช้กดปุ่มใดบนแผงแป้นตัวเลข ซึ่งเป็นประโยชน์ในการโต้ตอบกับผู้ใช้ การ์ดเสียงสามารถตรวจจับการกดปุ่มได้โดยดูจากสัญญาณ DTMF ที่ตู้สาขาส่งเข้ามา Dialogic® ได้เตรียมชุดคำสั่งไว้สำหรับรับสัญญาณการกดปุ่มคือ

  ```
  dx_getdig(chdev, tptDig, &digp, EV_SYNC);
  ```

  โดยที่

  - tptDig ระบุสาเหตุที่คำสั่งเสร็จสิ้น เช่น กดปุ่มครบตามจำนวนครั้งที่กำหนด หรือไม่กดปุ่มเป็นระยะเวลาเกินกว่าที่กำหนด เป็นต้น
  - digp เป็นโครงสร้างข้อมูลที่ใช้สำหรับเก็บข้อมูลที่ได้จากการกดปุ่ม
  - EV_SYNC ใช้เพื่อให้กระทำคำสั่งนี้แบบสมวาร

  ในที่นี้ การ์ดเสียงจะเก็บข้อมูลการกดปุ่มไว้ในบัฟเฟอร์ตลอดเวลา การล้างข้อมูลในบัฟเฟอร์ทำได้โดยใช้คำสั่ง

  ```
  dx_clrdigbuf(chdev);
  ```

  การเสร็จสิ้นของคำสั่ง dx_getdig() อาจเกิดจากหลายสาเหตุดังที่ระบุไว้ใน tptDig ซึ่งสามารถตรวจสอบสาเหตุของการเสร็จสิ้นได้ด้วยคำสั่ง

  ```
  ATDX_TERMMSK(chdev);
  ```

  โดยค่าที่คืนออกมาจะบ่งบอกถึงสาเหตุที่คำสั่ง dx_getdig() เสร็จสิ้น เช่น เกินเวลาที่กำหนด (TM_MAXTIME) หรือ กดปุ่มครบจำนวนที่ต้องการแล้ว (TM_MAXDIGIT) เป็นต้น
- **การเล่นไฟล์เสียง** การ์ดเสียงที่นำมาใช้นี้สามารถใช้เล่นไฟล์เสียงประเภท *.wav และ *.vox โดยชุดคำสั่งที่ใช้สำหรับเล่นไฟล์เสียงที่ Dialogic® จัดไว้ให้มีหลากหลาย โดยสามารถเล่นได้ทั้งจากบัฟเฟอร์ หรือจากไฟล์บนฮาร์ดดิสก์ สำหรับการเล่นไฟล์เสียงในสปีชคอลทำโดยเล่นจากไฟล์บนฮาร์ดดิสก์ โดยใช้คำสั่ง

```
dx_playwav(chdev, pathName, &tptPlay, EV_SYNC);
```

โดยที่

- pathName เป็นชื่อไฟล์เสียงที่ต้องการเล่น
- tptPlay จะต้องประกาศเป็น DV_TPT เพื่อระบุเหตุการณ์ที่จะทำให้หยุดการเล่นไฟล์เสียง เช่น เมื่อโทรศัพท์ถูกกดปุ่มหนึ่งครั้ง เป็นต้น

3.2.2 การหาขอบเขตของเสียงพูด

จากเสียงพูดที่รับเข้ามา จะทำการหาขอบเขตของเสียงพูดนั้นโดยใช้วิธีดังที่กล่าวไปในหัวข้อ 2.6 โดยค่าพารามิเตอร์ที่ต้องกำหนดเพื่อการหาขอบเขตของเสียงพูดมีดังนี้

- nFrameWidth บอกว่าในเฟรมหนึ่งๆ กว้างเท่าใด
- nNumCalPoint เป็นจำนวนจุดที่นำมาใช้ในการคำนวณหาค่าพลังงาน
- fHighEnergyThresh เป็นค่าพลังงานขีดจำกัด สำหรับแบ่งจุดที่มีพลังงานสูงและพลังงานต่ำออกจากกัน โดยจุดที่ถือว่ามีพลังงานสูงจะมีค่าพลังงานมากกว่าค่านี้ ส่วนจุดที่ถือว่ามีพลังงานต่ำจะมีค่าพลังงานน้อยกว่า
- fPercentNumHighPointThresh เป็นค่าที่ตัดสินว่าเฟรมนี้มีเสียงพูดหรือไม่ โดยเฟรมที่มีเสียงพูดจะมีจำนวนจุดที่มีพลังงานสูงเทียบกับจำนวนจุดทั้งหมดในเฟรมเป็นเปอร์เซนต์แล้วมากกว่าค่าค่านี้
- nSilenceFrameThresh เป็นค่าที่ใช้บอกจุดสิ้นสุดของเสียงพูด โดยถ้าจำนวนเฟรมที่มีเสียงเงียบเข้ามาติดต่อกันเกินกว่าค่าค่านี้ แสดงว่าเฟรมแรกที่เกิดเสียงเงียบเป็นจุดสิ้นสุดของเสียงพูด

3.2.3 การหาลักษณะสำคัญของเสียงเพื่อการรู้จำ

เมื่อผ่านการหาขอบเขตของเสียงพูดแล้ว จะนำเสียงพูดที่ผ่านการหาขอบเขตนั้นมาทำการหาลักษณะสำคัญของเสียง ซึ่งใช้วิธีการทำนายเชิงเส้นแบบรับรู้ดังที่กล่าวไว้ในหัวข้อ 2.4 โดยค่าพารามิเตอร์ที่ใช้ในการหาลักษณะสำคัญของเสียงในที่นี้จำเป็นต้องเหมือนกับค่าที่ใช้ในการหาลักษณะสำคัญของเสียงเพื่อการเรียนรู้ การตั้งค่าพารามิเตอร์ในการหาลักษณะของเสียงของโปรแกรมสปีชคอลสามารถดูได้ที่ พารามิเตอร์สำหรับการรู้จำเสียงพูด ใน ภาคผนวก – โปรแกรมโอนสายอัตโนมัติจากเสียงพูดชื่อไทยทางโทรศัพท์

3.2.4 การเตรียมข้อมูลเพื่อนำไปใช้ในการรู้จำ

เสียงพูดเมื่อนำมาหาลักษณะสำคัญแล้วจะทำการเตรียมเพื่อนำไปใช้ในการรู้จำต่อไป การเตรียมข้อมูลเพื่อนำไปใช้ในการรู้จำจะทำเช่นเดียวกับ

การเตรียมข้อมูลเพื่อนำไปใช้ในการเรียนรู้การหาลักษณะสำคัญของเสียงเพื่อการเรียนรู้ โดยจะทำการเลือกเฟรมสำหรับนำไปใช้ในการรู้จำให้กระจายครอบคลุมช่วงทั้งหมดของเสียงพูด ซึ่งค่าพารามิเตอร์ที่ใช้ในการเตรียมข้อมูลเพื่อนำไปใช้ในการรู้จำในที่นี้จำเป็นต้องเหมือนกับค่าที่ใช้ในการเตรียมข้อมูลเพื่อนำไปใช้ในการเรียนรู้การหาลักษณะสำคัญของเสียงเพื่อการเรียนรู้ การตั้งค่าพารามิเตอร์สำหรับการเตรียมข้อมูลเพื่อนำไปใช้ในการรู้จำของโปรแกรมสปีชคอลสามารถดูได้ที่

พารามิเตอร์สำหรับการรู้จำเสียงพูด ใน ภาคผนวก – โปรแกรมโอนสายอัตโนมัติจากเสียงพูดชื่อไทยทางโทรศัพท์

3.2.5 การรู้จำ

การรู้จำจะทำโดยการป้อนข้อมูลที่ได้จากกระบวนการ 3.2.4 เข้าเป็นอินพุตของโครงข่ายประสาทเทียมที่ได้จากการเรียนรู้ โครงข่ายประสาทเทียมจะทำการคำนวณหาค่าของทุกโนดเอาต์พุต ซึ่งชื่อไทยที่รู้จำได้จะมาจากโนดเอาต์พุตที่มีค่าสูงสุด การนำเข้าโครงข่ายประสาทเทียมของโปรแกรมสปีชคอลสามารถดูได้ที่ พารามิเตอร์สำหรับการรู้จำเสียงพูด ใน ภาคผนวก – โปรแกรมโอนสายอัตโนมัติจากเสียงพูดชื่อไทยทางโทรศัพท์

3.2.6 การค้นหาชื่อ และการโอนสาย

เมื่อได้ชื่อไทยมาแล้ว โปรแกรมจะทำการค้นหาเบอร์โทรศัพท์ของชื่อนั้น โดยรายชื่อและเบอร์โทรศัพท์สามารถนำเข้ามาได้ทางโปรแกรมสปีชคอล การนำเข้ารายชื่อและเบอร์โทรศัพท์ดูได้ที่ พารามิเตอร์สำหรับการโอนสายโทรศัพท์ ใน ภาคผนวก – โปรแกรมโอนสายอัตโนมัติจากเสียงพูดชื่อไทยทางโทรศัพท์ ทั้งนี้ ลำดับรายชื่อที่ใช้ในการค้นหาจะต้องเรียงตัวเหมือนกับลำดับรายชื่อที่ใช้ในการเรียนรู้ ซึ่งลำดับรายชื่อที่ใช้ในการเรียนรู้ในที่นี้คือ NameList.txt ในกระบวนการ 0

สำหรับการโอนสายหรือโทรออก จะใช้การ์ดเสียงส่งสัญญาณ DTMF ไปยังตู้สาขาเพื่อดำเนินการโอนสายหรือโทรออกไปยังหมายเลขอื่นๆ ด้วยคำสั่ง

```
dx_dial(chdev, number, (DX_CAP *)NULL, DX_CALLP | EV_SYNC);
```

โดยที่ number คือหมายเลขที่ต้องการโอนสายหรือโทรออก โดยในกรณีการโอนสายจะต้องส่งสัญญาณ “&” หรือ แฟลช ออกไปก่อนเพื่อบอกตู้สาขาว่าต้องการโอนสาย แล้วค่อยตามด้วยหมายเลขที่ต้องการจะโอนไป

หลังจากโอนสายแล้ว ผลการโอนสายสามารถติดตามได้โดยดูจากค่าที่คำสั่ง dx_dial() คืนออกมา ซึ่งผลการโอนสายอาจเป็นได้หลายแบบ เช่น “ติดต่อได้สำเร็จ” (CR_CNCT) “สายไม่ว่าง” (CR_BUSY) หรือ “เกิดความผิดพลาด” (CR_ERROR) เป็นต้น

3.3 การทดลอง

ในการทดลองได้ให้ระบบสามารถรู้จำชื่ออาจารย์และบุคลากรภาควิชาวิศวกรรมคอมพิวเตอร์ จุฬาลงกรณ์มหาวิทยาลัย จำนวน 45 ชื่อ ดังตารางที่ 3.4
ตารางที่ 3.4

การทดลองนี้จะแบ่งไฟล์เสียงที่บันทึกไว้เป็นสองส่วน คือส่วนที่ใช้สำหรับการเรียนรู้ และส่วนที่ใช้สำหรับการทดสอบ ในการทดลองได้บันทึกไฟล์เสียงจากผู้บันทึกทั้งหมด 20 คน ซึ่งจะแบ่งไฟล์เสียงจากผู้บันทึก 14 คนเป็นข้อมูลที่ใช้สำหรับการเรียนรู้ และไฟล์เสียงจากผู้บันทึก 6 คนเป็นข้อมูลที่ใช้สำหรับการทดสอบ โดยมีรายละเอียดต่างๆ ดังตารางที่ 3.5 และตารางที่ 3.6

โดยในการทดลองที่ 3.3.1 - 3.3.4 จะใช้ข้อมูลที่ผู้บันทึกเสียงแต่ละคนพูดชื่อไทย 45 ชื่อ ชื่อละครั้ง ทำให้ได้จำนวนข้อมูลทั้งหมดสำหรับการเรียนรู้เท่ากับ 630 และจำนวนข้อมูลทั้งหมดสำหรับการรู้จำเท่ากับ 270

ส่วนในการทดลองที่ 3.3.5 จะใช้ข้อมูลที่ผู้บันทึกเสียงแต่ละคนพูดชื่อไทย 45 ชื่อ ชื่อละ 5 ครั้ง ทำให้ได้จำนวนข้อมูลทั้งหมดสำหรับการเรียนรู้เท่ากับ 3150 และจำนวนข้อมูลทั้งหมดสำหรับการรู้จำเท่ากับ 1350 ซึ่งในการทดลองที่ 3.3.5 จะนำข้อมูลจำนวนนี้ไปทำการเรียนรู้และเปรียบเทียบกับผลการเรียนรู้ของชุดข้อมูลในตารางที่ 3.5

ตารางที่ 3.4 ชื่อไทยที่ใช้ในการรู้จำ

อรรถวิทย์	#a1t_tha1_wi3t	วีระ	wii0_ra3
โปรดปราน	proo1d_praa0n	ประภาส	pra1_phaa2t
ธงชัย	tho0ng_cha0j	เศรษฐา	see1t_thaa4
วิษณุ	wi3t_sa1_nu3	วิวัฒน์	wi3_wa3t
มัณฑนา	ma0n_tha3_naa0	พรศิริ	ph@@0n_si1_ri1
ทักษิณา	tha3k_si1_naa0	เชษฐ	chee2t
ฐานิศรา	thaa4_ni3t_sa1_raa0	อรรถสิทธิ์	#a1t_tha1_si1t
ญาใจ	jaa0_ca0j	กอบกุล	k@@1p_ku0n
ผู้ช่วยหัวหน้าภาค	phuu2_chuua2j_huua4_naa2_phaa2k	นครทิพย์	na3_kh@@0n_thi3p
หัวหน้าภาค	huua4_naa2_phaa2k	วิชาญ	wi3_chaa0n
ธุรการ	thu3_ra3_kaa0n	สุเมธ	su1_mee2t
ธาราทิพย์	thaa0_raa0_thi3p	ธนาวรรณ	tha3_naa0_wa0n
บุญเสริม	bu0n_sqq4m	วันพร	wa0n_ph@@0n
อาทิตย์	#aa0_thi3t	ชัยศิริ	cha0j_si1_ri1
สาธิต	saa4_thi3t	ทวิตีย์	tha3_wi3t_tii0
เฉลิมเอก	cha1_lqq4m_#ee1k	ชัย	cha0j
สืบสกุล	svv1p_sa1_ku0n	ณัฐวุฒิ	na3t_tha1_wu3t
บุญชัย	bu0n_cha0j	จารุมาตร	caa0_ru3_maa2t
นงลักษณ์	no0ng_la3k	ยรรยง	ja0n_jo0ng
ฐิต	thi1t	วันชัย	wa0n_cha0j
ชูชีพ	chuu0_chii2p	พิษณุ	pi3t_sa1_nu3
เมธี	mee0_thii0	เกริก	krqq1k
สมชาย	so4m_chaa0j		

ตารางที่ 3.5 รายละเอียดของข้อมูลที่ใช้ในการทดลองที่ 3.3.1 - 3.3.4 และ 3.3.5

	ข้อมูลสำหรับการเรียนรู้	ข้อมูลสำหรับการทดสอบ
จำนวนผู้บันทึก	14	6
จำนวนข้อมูลที่ได้จากแต่ละผู้บันทึก	45	45
จำนวนข้อมูลทั้งหมด	630	270

ตารางที่ 3.6 รายละเอียดของข้อมูลที่ใช้ในการทดลองที่ 3.3.5

	ข้อมูลสำหรับการเรียนรู้	ข้อมูลสำหรับการทดสอบ
จำนวนผู้บันทึก	14	6
จำนวนข้อมูลที่ได้จากแต่ละผู้บันทึก	45*5	45*5
จำนวนข้อมูลทั้งหมด	3150	1350

การทดลองทำโดยปรับค่าพารามิเตอร์ต่างๆ ที่มีผลต่อการรู้จำ ดังนี้

3.3.1 การทดลองเพื่อเปรียบเทียบผลของอันดับการทำนายเชิงเส้นแบบรับรู้

ทำการทดลองโดยเปลี่ยนอันดับการทำนายเชิงเส้นแบบรับรู้ โดยคงค่าพารามิเตอร์ตัวอื่นๆ ไว้ดังนี้

- ไม่ใช้อนุพันธ์การทำนายเชิงเส้นแบบรับรู้
- จำนวนเฟรมการวิเคราะห์ = 47
- จำนวนรอบในการวนปรับน้ำหนัก = 1000
- ข้อมูลที่ใช้ในการทดลอง ดังในตารางที่ 3.5

ผลลัพธ์จากการเปลี่ยนอันดับการทำนายเชิงเส้นแบบรับรู้แสดงได้ดังตารางที่ 3.7

ตารางที่ 3.7 ผลลัพธ์จากการเปลี่ยนอันดับการทำนายเชิงเส้นแบบรับรู้

อันดับการทำนายเชิงเส้นแบบรับรู้	ความถูกต้องของข้อมูลที่ใช้ในการเรียนรู้	ความถูกต้องของข้อมูลที่ใช้ในการทดสอบ
9	630/630 (100%)	257/270 (95.19%)
12	630/630 (100%)	255/270 (94.44%)
15	630/630 (100%)	256/270 (94.81%)
18	630/630 (100%)	258/270 (95.56%)
21	630/630 (100%)	248/270 (91.85%)

3.3.2 การทดลองเพื่อเปรียบเทียบผลของอนุพันธ์การทำนายเชิงเส้นแบบรับรู้

ทำการทดลองโดยเปลี่ยนอนุพันธ์การทำนายเชิงเส้นแบบรับรู้ โดยคงค่าพารามิเตอร์ตัวอื่นๆ ไว้ดังนี้

- อันดับการทำนายเชิงเส้นแบบรับรู้ = 9
- จำนวนเฟรมการวิเคราะห์ = 47
- จำนวนรอบในการวนปรับน้ำหนัก = 1000
- ข้อมูลที่ใช้ในการทดลอง ดังในตารางที่ 3.5

ผลลัพธ์จากการเปลี่ยนอนุพันธ์การทำนายเชิงเส้นแบบรับรู้แสดงได้ดังตารางที่ 3.8

ตารางที่ 3.8 ผลลัพธ์จากการเปลี่ยนอนุพันธ์การทำนายเชิงเส้นแบบรับรู้

อนุพันธ์การทำนายเชิงเส้นแบบรับรู้	ความถูกต้องของข้อมูลที่ใช้ในการเรียนรู้	ความถูกต้องของข้อมูลที่ใช้ในการทดสอบ
ไม่ใช้	630/630 (100%)	257/270 (95.19%)
อันดับที่หนึ่ง	630/630 (100%)	257/270 (95.19%)
อันดับที่สอง	630/630 (100%)	248/270 (91.85%)

3.3.3 การทดลองเพื่อเปรียบเทียบผลของจำนวนเฟรมการวิเคราะห์

ทำการทดลองโดยเปลี่ยนจำนวนเฟรมการวิเคราะห์ โดยคงค่าพารามิเตอร์ตัวอื่นๆ ไว้ดังนี้

- อันดับการทำนายเชิงเส้นแบบรับรู้ = 9
- ไม่ใช้อนุพันธ์การทำนายเชิงเส้นแบบรับรู้
- จำนวนรอบในการวนปรับน้ำหนัก = 1000
- ข้อมูลที่ใช้ในการทดลอง ดังในตารางที่ 3.5

ผลลัพธ์จากการเปลี่ยนจำนวนเฟรมการวิเคราะห์แสดงได้ดังตารางที่ 3.9

ตารางที่ 3.9 ผลลัพธ์จากการเปลี่ยนจำนวนเฟรมการวิเคราะห์

เฟรมการวิเคราะห์	ความถูกต้องของข้อมูลที่ใช้ในการเรียนรู้	ความถูกต้องของข้อมูลที่ใช้ในการทดสอบ
47	630/630 (100%)	257/270 (95.19%)
29	629/630 (99.84%)	253/270 (93.70%)
15	629/630 (99.84%)	256/270 (94.81%)

3.3.4 การทดลองเพื่อเปรียบเทียบผลของจำนวนรอบในการวนปรับน้ำหนักของโครงข่ายประสาทเทียม

ทำการทดลองโดยเปรียบเทียบผลของจำนวนรอบในการวนปรับน้ำหนักของโครงข่ายประสาทเทียม โดยคงค่าพารามิเตอร์ตัวอื่นๆ ไว้ดังนี้

- อันดับการทำนายเชิงเส้นแบบรับรู้ = 9
- ไม่ใช้อนุพันธ์การทำนายเชิงเส้นแบบรับรู้
- จำนวนเฟรมการวิเคราะห์ = 47
- ข้อมูลที่ใช้ในการทดลอง ดังในตารางที่ 3.5

ผลลัพธ์จากการเปลี่ยนจำนวนรอบในการวนปรับน้ำหนักแสดงได้ดังตารางที่ 3.10

ตารางที่ 3.10 ผลลัพธ์จากการเปลี่ยนจำนวนรอบในการวนปรับน้ำหนัก

จำนวนรอบ	ความถูกต้องของข้อมูลที่ใช้ในการเรียนรู้	ความถูกต้องของข้อมูลที่ใช้ในการทดสอบ
1000	630/630 (100%)	257/270 (95.19%)
10000	630/630 (100%)	255/270 (94.44%)

3.3.5 การทดลองเพื่อเปรียบเทียบผลของจำนวนชุดข้อมูลที่ใช้ในการเรียนรู้

ทำการทดลองโดยเปรียบเทียบผลของจำนวนชุดข้อมูลที่ใช้ในการเรียนรู้ ระหว่างชุดข้อมูลเดิมในตารางที่ 3.5 ที่ผู้พูดพูดคนละครั้งในทุกชื่อ กับชุดข้อมูลในตารางที่ 3.6 ที่ผู้พูดพูดคนละ 5 ครั้งในทุกชื่อ โดยคงค่าพารามิเตอร์ตัวอื่นๆ ไว้ดังนี้

- อันดับการทำนายเชิงเส้นแบบรับรู้ = 9
- ไม่ใช้อนุพันธ์การทำนายเชิงเส้นแบบรับรู้
- จำนวนเฟรมการวิเคราะห์ = 47
- จำนวนรอบในการวนปรับน้ำหนัก = 1000

ผลลัพธ์จากการเปลี่ยนจำนวนชุดข้อมูลที่ใช้ในการเรียนรู้แสดงได้ดังตารางที่ 3.11

ตารางที่ 3.11 ผลลัพธ์จากการเปลี่ยนจำนวนชุดข้อมูลที่ใช้ในการเรียนรู้

ชุดข้อมูล	ความถูกต้องของข้อมูลที่ใช้ในการเรียนรู้	ความถูกต้องของข้อมูลที่ใช้ในการทดสอบ
ดังตารางที่ 3.5	630/630 (100%)	257/270 (95.19%)
ดังตารางที่ 3.6	3143/3150 (99.78%)	1300/1350 (96.30%)

3.3.6 การทดลองนำผลการเรียนรู้มาใช้งานจริงกับโปรแกรมประยุกต์ทางโทรศัพท์

ทำการเรียนรู้โครงข่ายประสาทเทียม โดยใช้ค่าพารามิเตอร์ต่างๆ ดังนี้

- อันดับการทำนายเชิงเส้นแบบรับรู้ = 9
- ไม่ใช้อนุพันธ์การทำนายเชิงเส้นแบบรับรู้
- จำนวนเฟรมการวิเคราะห์ = 47
- จำนวนรอบในการวนปรับน้ำหนัก = 1000
- ข้อมูลที่ใช้ในการเรียนรู้ ดังในตารางที่ 3.6

เมื่อนำโครงข่ายประสาทเทียมที่ได้จากการเรียนรู้ไปใช้จริงกับโปรแกรมประยุกต์ทางโทรศัพท์ พบว่ามีความถูกต้อง 79.33 เปอร์เซ็นต์

3.4 สรุปผลการทดลอง

จากการทดลองสรุปได้ว่าอันดับการทำนายเชิงเส้นแบบรับรู้ที่ให้ผลดีที่สุดคืออันดับที่ 18 โดยมีอันดับที่ 9 ที่ให้ผลดีเกือบจะพอๆ กัน ส่วนจำนวนเฟรมการวิเคราะห์เท่ากับ 47 จะให้ผลดีกว่า

จำนวนอื่น อาจเป็นเพราะจำนวนเฟรมที่มากขึ้นจะทำให้คุณลักษณะสำคัญของเสียงพูดถูกดึงมาใช้ในการเรียนรู้และรู้จำได้มากขึ้น ขณะที่การวนรอบปรับน้ำหนักของเครือข่ายประสาทเทียม 1000 รอบ จะให้ผลดีกว่าการวนปรับน้ำหนักที่มากกว่านั้นคือ 10000 รอบ ซึ่งการวนปรับน้ำหนักมากรอบอาจทำให้เครือข่ายประสาทเทียมที่ได้มีความเจาะจงกับข้อมูลที่ใช้เรียนรู้จนเกินไป [90] ส่วนการเรียนรู้ที่ใช้ข้อมูลในการเรียนรู้มากกว่า จะให้ผลดีกว่าการเรียนรู้ที่ใช้ข้อมูลในการเรียนรู้น้อยกว่า เนื่องจากข้อมูลที่ใช้เรียนรู้มีความหลากหลายมากกว่า [90] และสุดท้าย การไม่ใช้อนุพันธ์เชิงเส้นแบบรับรู้ให้ผลดีกว่าการใช้อนุพันธ์เชิงเส้นแบบรับรู้อันดับที่หนึ่งและอันดับที่สอง

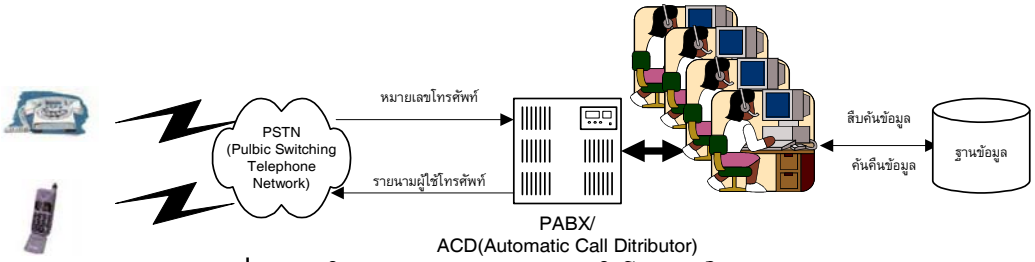
เมื่อนำโปรแกรมไปใช้จริงพบว่าให้ความผิดพลาดมากกว่าการทดลองโดยใช้ข้อมูลการทดสอบ ทั้งนี้อาจเป็นเพราะว่าผู้ที่โทรเข้ามามีเสียงและสำเนียงการพูดชื่อคนที่หลากหลายกว่า รวมทั้งอาจเกิดความผิดพลาดจากการหาขอบเขตของเสียงพูด ซึ่งในข้อมูลการทดสอบมีการหาขอบเขตของเสียงพูดไว้เรียบร้อยแล้ว

4. ระบบการสอบถามรายนามผู้ใช้โทรศัพท์โดยผ่านระบบโทรศัพท์

ในปัจจุบัน การให้บริการทางด้านข้อมูลข่าวสารกับลูกค้าโดยผ่านทางระบบโทรศัพท์เริ่มเข้ามามีบทบาทสำคัญต่อการให้บริการกับลูกค้า เพื่อเพิ่มความพึงพอใจและความสะดวกรวดเร็วในการให้บริการ การให้บริการสอบถามรายนามและเลขหมายผู้ใช้โทรศัพท์ก็เป็นหนึ่งในบริการที่ผู้ให้บริการทางด้านโทรศัพท์ต้องให้บริการแก่ลูกค้า โดยข้อมูลของผู้ใช้โทรศัพท์ที่จะสอบถามได้จะต้องได้รับอนุญาตจากผู้ใช้โทรศัพท์ที่ต้องแจ้งให้ผู้ให้บริการทราบว่าสามารถเปิดเผยข้อมูลได้ การให้บริการสอบถามรายนามผู้ใช้โทรศัพท์ก็เป็นการให้บริการ เมื่อผู้ใช้บริการต้องการทราบชื่อเจ้าของหมายเลขโทรศัพท์โดยแจ้งหมายเลขโทรศัพท์แก่พนักงาน พนักงานจะตอบกลับมาเป็นชื่อและนามสกุลของผู้ใช้โทรศัพท์ ทั้งนี้อาจเพื่อวัตถุประสงค์ต่าง ๆของผู้ใช้บริการ เช่น ใช้ในการตรวจสอบความถูกต้องของหมายเลขโทรศัพท์ว่าเจ้าของหมายเลขโทรศัพท์เป็นผู้ใด ส่วนการให้บริการสอบถามเลขหมายโทรศัพท์เป็นการให้บริการเมื่อผู้ใช้บริการต้องการทราบหมายเลขโทรศัพท์ของเจ้าของโทรศัพท์นั้น โดยแจ้งชื่อและนามสกุลแก่พนักงาน พนักงานจะตอบกลับมาเป็นเลขหมายโทรศัพท์

นอกจากนี้ยังมีบริการต่าง ๆ ที่คาดว่าทางผู้ให้บริการจะเปิดให้บริการผ่านทางระบบโทรศัพท์ต่อไปในอนาคต เช่น การให้บริการข่าวสารประจำวัน การให้บริการพยากรณ์อากาศ การให้บริการข่าวสารทางด้านตลาดหลักทรัพย์ เป็นต้น การบริการทั้งหมดที่กล่าวมานี้สามารถนำหลักการของระบบรู้จำเสียงพูดหรือระบบสังเคราะห์เสียงพูดมาประยุกต์ใช้งานได้เป็นอย่างดี เพื่อให้เกิดความรวดเร็วทันต่อเหตุการณ์และลดค่าใช้จ่ายในการเพิ่มประสิทธิภาพของระบบงาน ซึ่งระบบการรู้จำเสียงพูดแบบต่อเนื่องและการสังเคราะห์เสียงพูดภาษาไทยในปัจจุบัน ได้มีการพัฒนาไปได้ระดับหนึ่งแล้ว สามารถนำมาประยุกต์ ใช้งานกับชีวิตประจำวันของมนุษย์ได้จริงในปัจจุบัน

โครงการนี้ได้พัฒนาต้นแบบเพื่อเพิ่มประสิทธิภาพของระบบสอบถามรายนามผู้ใช้โทรศัพท์ผ่านทางระบบโทรศัพท์ให้สูงยิ่งขึ้น ซึ่งในปัจจุบันระบบสอบถามรายนามผู้ใช้โทรศัพท์ยังจำเป็นต้องมีพนักงานตอบรับโทรศัพท์คอยประจำอยู่ที่ศูนย์บริการ เพื่อทำหน้าที่รับโทรศัพท์จากผู้ใช้บริการที่ต้องการทราบรายนามผู้ใช้โทรศัพท์ตลอดทั้ง 24 ชั่วโมง ดังนั้นเมื่อผู้ใช้บริการโทรศัพท์ไปยังศูนย์บริการแล้วแจ้งหมายเลขโทรศัพท์ที่ต้องการทราบกับพนักงานรับโทรศัพท์ พนักงานรับโทรศัพท์จะค้นหารายนามผู้ใช้โทรศัพท์โดยพิมพ์ข้อมูลลงไปในระบบ เพื่อจะนำข้อมูลนั้นไปทำการค้นคืนในฐานข้อมูล เมื่อพบข้อมูลที่ต้องการแล้วจึงแจ้งให้ผู้ใช้บริการทราบดังแสดงขั้นตอนในรูปที่ 4.1



รูปที่ 4.1 การให้บริการสอบถามรายนามผู้ใช้โทรศัพท์ในปัจจุบัน

ขั้นตอนดังกล่าวทำให้เกิดความล่าช้าของเวลาที่ใช้ไปต่อหนึ่งการเรียกใช้บริการ เมื่อมีผู้ใช้บริการเข้ามาเรียกใช้งานในเวลาพร้อมกันทำให้เกิดแถวรอคอยเพื่อรับบริการเป็นจำนวนมาก เป็นเหตุให้ผู้ใช้บริการต้องรอคอยและอาจสร้างความไม่พึงพอใจขึ้นได้ ซึ่งระบบสอบถามรายนามผู้ใช้โทรศัพท์โดยใช้การรู้จำเสียงพูดและการสังเคราะห์เสียงพูดภาษาไทยแบบอัตโนมัตินี้ จะสามารถลดระยะเวลาในการใช้บริการ และลดค่าใช้จ่ายในการเพิ่มประสิทธิภาพของระบบงานเนื่องจากไม่จำเป็นต้องเพิ่มจำนวนพนักงานตอบรับโทรศัพท์เพื่อที่จะรองรับให้เท่ากับปริมาณการเรียกใช้บริการที่ผู้ใช้เรียกเข้ามาสู่ระบบสูงสุด

การพัฒนาโปรแกรมต้นแบบของระบบสอบถามรายนามผู้ใช้โทรศัพท์แบบอัตโนมัติบนเครื่องไมโครคอมพิวเตอร์นั้นต้องประกอบด้วยส่วนประกอบ 2 ส่วนใหญ่ คือ

ก. การรู้จำเสียงพูดตัวเลขภาษาไทยแบบต่อเนื่องและไม่ขึ้นกับผู้พูด ซึ่งใช้ชุดหมายเลขโทรศัพท์ศูนย์ถึงเก้า

ข. สังเคราะห์เสียงพูดสำหรับชื่อเฉพาะภาษาไทย โดยพูดชื่อและนามสกุลจากรายนามผู้ใช้โทรศัพท์

สำหรับโครงการวิจัยนี้มุ่งเน้นในการพัฒนาเพื่อสร้างโปรแกรมต้นแบบของระบบสอบถามรายนามผู้ใช้โทรศัพท์แบบอัตโนมัติที่สามารถรู้จำเสียงพูดตัวเลขภาษาไทยศูนย์ถึงเก้าแบบต่อเนื่องที่เป็นหมายเลขโทรศัพท์ด้วยการรู้จำเสียงพูดต่อเนื่องภาษาไทยระดับพยางค์ โดยลักษณะสำคัญทางสวนศาสตร์ที่นำมาใช้ได้แก่ ราสต้า-พีแอลพี ส่วนเทคนิคของการเรียนรู้ที่นำมาใช้คือ ข่ายงานระบบประสาทเทียม แล้วนำเลขหมายที่ได้ไปทำการค้นคืนรายนามผู้ใช้โทรศัพท์ในฐานข้อมูลได้เป็นชื่อและนามสกุล แล้วทำการสังเคราะห์เสียงพูดออกมาเป็นภาษาไทย โดยใช้โมดูลการต่อหน่วยเสียงย่อย (Concatenation) โดยจำลองระบบบนเครื่องไมโครคอมพิวเตอร์ ใช้ไมโครโฟนเป็นอุปกรณ์รับข้อมูลขาเข้าและลำโพงเป็นอุปกรณ์ส่งข้อมูลขาออก ทำการบันทึกเสียงเก็บเป็นแฟ้มข้อมูลเสียง (WAV file) แบบโมโน 16 บิต ที่อัตราการชักข้อมูล (sampling rate) 11,025 เฮิรตซ์

4.1 ขั้นตอนการพัฒนาระบบสอบถามรายนามผู้ใช้โทรศัพท์

การเพิ่มความสามารถและประสิทธิภาพให้กับระบบสอบถามรายนามผู้ใช้โทรศัพท์ โดยให้สามารถทำงานแบบอัตโนมัติได้นั้นต้องอาศัยเทคโนโลยีหลักสองอย่างเข้ามาใช้ร่วมกัน กล่าวคือ การรู้จำเสียงพูด (Speech Recognition) และ การสังเคราะห์เสียงพูด (Speech Synthesis) ซึ่งในส่วนนี้ได้จำลองระบบบนเครื่องไมโครคอมพิวเตอร์ โดยมีข้อมูลขาเข้าเป็นเสียงพูดตัวเลขผ่านทางไมโครโฟน และข้อมูลขาออกเป็นเสียงพูดชื่อและนามสกุลผ่านทางลำโพง เพื่อใช้เป็นต้นแบบ (Prototype) ซึ่งมีรายละเอียดการใช้งานต่างๆ ดังแสดงในภาคผนวก 9 ระบบดังกล่าวมีส่วนประกอบหลัก ดังต่อไปนี้

1. เริ่มต้นเมื่อรับเสียงพูดตัวเลขต่อเนื่องภาษาไทยที่ต้องการทราบถึงรายนามผู้ใช้โทรศัพท์ผ่านทางไมโครโฟนเมื่อผู้พูดพูดจบประโยค เสียงที่ได้ทั้งหมดจะผ่านส่วนที่ทำการตัดหัวท้ายหน่วยหลังจากนั้นก็นำเสียงที่ได้มาผ่านส่วนการตัดแบ่งพยางค์ออกมาเป็นตัวเลขเดี่ยว ในกรณีที่เสียงพูดนั้นไม่ชัดเจน อาจจะเกิดจากระบบไม่สามารถตัดแบ่งพยางค์ได้ตามจำนวนหลักของตัวเลขที่ระบบกำหนด ระบบจะให้ผู้ใช้พูดเสียงตัวเลขซ้ำใหม่อีกครั้งหนึ่ง
2. เมื่อรับเสียงพูดที่ชัดเจน (มีจำนวนพยางค์ครบตามจำนวนหลักของตัวเลขที่ระบบกำหนด) มาแล้ว จากนั้นก็นำเสียงพูดที่ได้นั้นมาผ่านส่วนของการรู้จำเสียงพูดภาษาไทย ซึ่งจะให้ผลคือ

ได้หมายเลขโทรศัพท์ออกมา และแสดงออกมาทางจอภาพเพื่อช่วยให้ผู้ใช้สามารถตรวจสอบความถูกต้องของหมายเลขโทรศัพท์ที่พูดไปได้อีกทางหนึ่ง

3. ส่วนของการค้นคืนรายนามผู้ใช้โทรศัพท์จากฐานข้อมูล ในกรณีที่พบรายนามผู้ใช้โทรศัพท์ก็จะพูดชื่อและนามสกุลที่เป็นคำอ่านของหมายเลขโทรศัพท์นั้นออกมาจากฐานข้อมูล ในกรณีที่ไม่พบรายนามผู้ใช้โทรศัพท์ระบบก็จะแจ้งให้ทราบว่าไม่พบหมายเลขโทรศัพท์นั้นในฐานข้อมูลที่มีอยู่
4. ส่วนการสังเคราะห์เสียงพูดภาษาไทย ซึ่งจะนำเอาคำอ่านจากข้อ 3 มาแล้วทำการค้นคืนในพจนานุกรมเสียงที่มีอยู่แล้วออกเสียงทางลำโพง

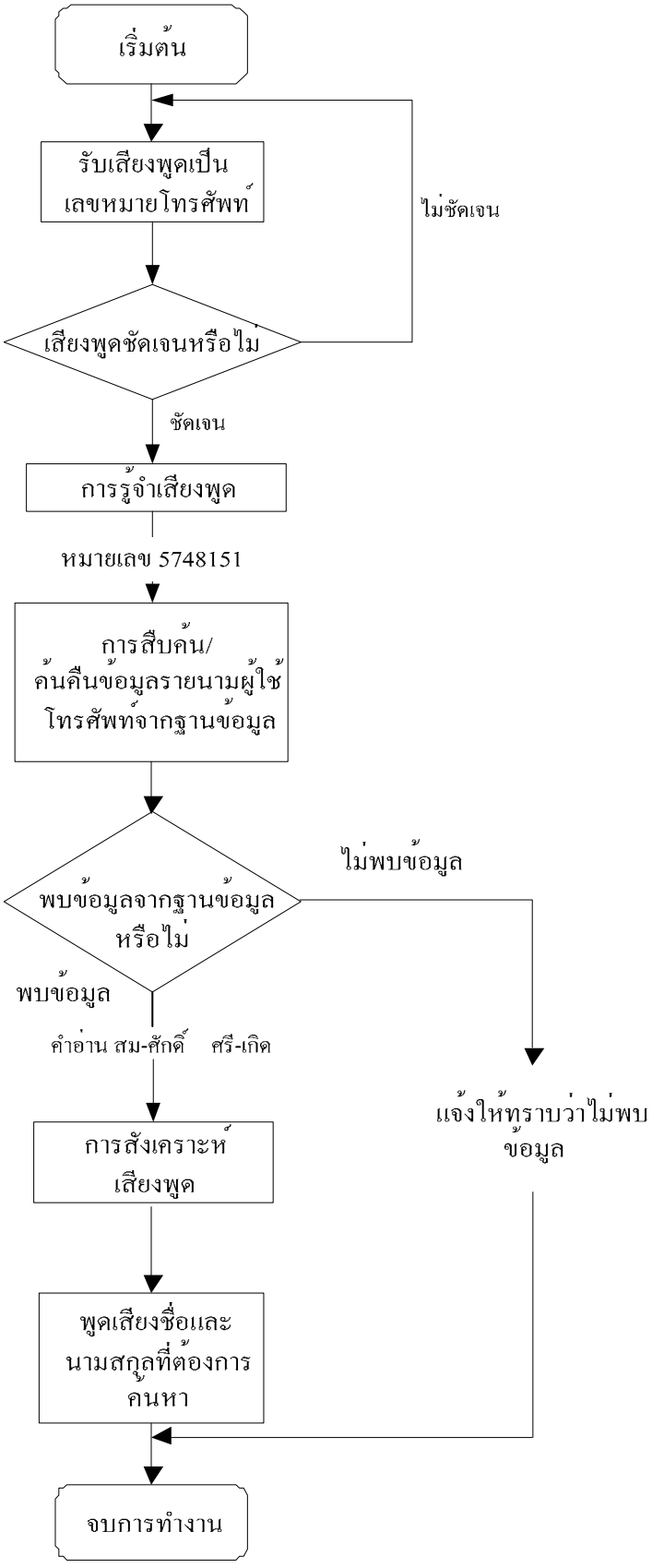
การทำงานของระบบสอบถามรายนามผู้ใช้โทรศัพท์แบบอัตโนมัติสามารถแสดงดังรูปที่ 4.2 ขั้นตอนในการพัฒนาระบบสอบถามรายนามผู้ใช้โทรศัพท์อัตโนมัติดังกล่าว จึงแบ่งออกเป็น 4 ขั้นตอนหลักๆ คือ

1) ทำการตัดแบ่งพยางค์และตัดหัวท้ายของหน่วย โดยพิจารณาจากค่าพลังงานและค่าอัตราตัดผ่านระดับกำหนด เพื่อให้ได้พยางค์ของตัวเลขเดี่ยวออกมาเพื่อจะนำไปทำการรู้จำในระดับพยางค์

2) จากนั้นนำเอาเสียงที่ได้เป็นพยางค์ของตัวเลขเดี่ยวมาผ่านขั้นตอนการรู้จำเสียงพูด โดยผ่านวิธีการทางสวนศาสตร์คือ ราสต้า-พีแอลพี เพื่อหาคุณลักษณะสำคัญของเสียง จากนั้นนำมาผ่านขั้นตอนข่ายงานระบบประสาทเทียม เพื่อทำการเรียนรู้และฝึกฝนพยางค์ของเสียงนั้นๆ

3) พัฒนาการสืบค้นรายนามผู้ใช้โทรศัพท์ โดยออกแบบโครงสร้างข้อมูลของฐานข้อมูลเป็นแบบแฮชชิงเนื่องจากต้องการจำกัดเวลาในการค้นหาข้อมูล เพื่อไม่ให้เป็นตัวแปรที่สำคัญในเรื่องเวลาของระบบทั้งหมดที่ใช้คือตั้งแต่พูดหมายเลขโทรศัพท์จนถึงพูดชื่อและนามสกุลของผู้ใช้ออกมา

4) การสังเคราะห์เสียงพูด คือจะนำชื่อที่ได้ผ่านกระบวนการสังเคราะห์เสียงโดยใช้วิธีการตัดคำและแบ่งพยางค์ รวมถึงฐานข้อมูลของพจนานุกรมเสียงของอัจจิมา [97] มาพัฒนาโดยทำการเพิ่มเติมคำศัพท์และพจนานุกรมหน่วยเสียง และได้ปรับปรุงวิธีการประมาณค่าความใกล้เคียงของเสียงซึ่งทำให้สามารถค้นหาคำพ้องเสียงในพจนานุกรมได้แม้คำนั้นจะไม่ได้บรรจุไว้ในพจนานุกรม ซึ่งจะกล่าวขั้นตอนโดยละเอียดดังต่อไปนี้



รูปที่ 4.2 การทำงานของระบบสอบถามรายนามผู้ใช้โทรศัพท์

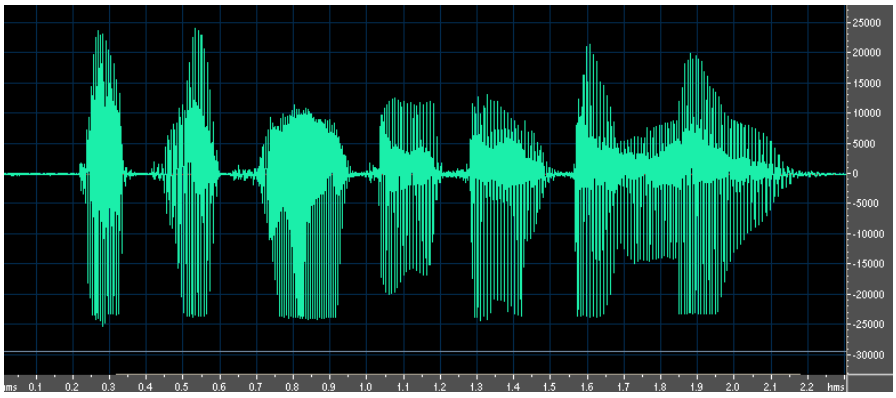
4.1.1 การตัดหัวท้ายหน่วยและการตัดแบ่งพยางค์

ส่วนของการตัดหัวท้ายหน่วยและการตัดแบ่งพยางค์ เป็นส่วนที่สำคัญและจำเป็นในการรู้จำเสียงพูดต่อเนื่องภาษาไทย เนื่องจากในงานวิจัยนี้ได้ใช้หน่วยเสียงพยางค์เป็นตัวทาบเทียบ (Syllable Based) โดยทั่วไปเสียงพูดต่อเนื่องจะประกอบด้วยพยางค์หลายๆ พยางค์ติดกันทำให้เกิดความหลากหลายของเสียงพูดและทำให้ยากต่อการฝึกฝนและการรู้จำของคอมพิวเตอร์ ดังนั้นจึงจำเป็นต้องมีการตัดแบ่งพยางค์เพื่อให้สามารถแบ่งคำพูดที่ต่อเนื่องออกเป็นพยางค์เดี่ยว แล้วก็นำเอาพยางค์นั้นเข้าไปฝึกและทำการรู้จำได้ง่ายขึ้น

1. การตัดหัวท้ายหน่วย

ขั้นตอนในการตัดหัวท้ายหน่วยสามารถแบ่งเป็นขั้นตอนได้ดังนี้

1) รับสัญญาณเสียงพูดที่เป็นตัวเลขต่อเนื่องเข้ามาโดยจะรับเสียงเข้ามาจากไมโครโฟนเพื่อประมวลผล โดยพิจารณาพลังงานเสียงที่เข้ามาในแต่ละ 200 มิลลิวินาที ว่ามีสัญญาณเสียงหรือไม่ ถ้าไม่ใช่จะคำนวณไว้เป็นค่าสภาวะแวดล้อม
2) ถ้าเป็นสัญญาณเสียงเข้ามาซึ่งก็จะพิจารณาได้จากค่าพลังงานของเสียงนั้นมีค่าเกินกว่าค่าที่กำหนด ก็จะทำการรวมสัญญาณเสียงที่เข้ามาใหม่นี้กับรอบก่อนหน้านี้เพื่อรอสัญญาณเสียงต่อไป
3) ตรวจสอบว่าเสียงนั้นจบประโยคแล้วหรือไม่ โดยจะพิจารณาจากค่าพลังงานของเสียงที่เข้ามาว่าน้อยกว่าค่าที่กำหนดและนานเกินกว่าระยะเวลาที่กำหนด (ในระบบกำหนดระยะเวลาไว้เท่ากับ 600 มิลลิวินาที)
4) ทำวนรอบซ้ำไปจนกระทั่งพบว่าไม่มีสัญญาณเสียงพูด แล้วเก็บเสียงพูดที่ได้ลงในหน่วยความจำสำรองและหน่วยความจำหลักเพื่อที่จะไปทำการตัดแบ่งพยางค์ต่อไป ซึ่งแฟ้มข้อมูลที่ได้จากการเก็บลงหน่วยความจำหลักเมื่อนำมาเปิดดูเพื่อตรวจสอบความถูกต้องของการตัดสามารถแสดงดังรูปที่ 4.3



รูปที่ 4.3 สัญญาณเสียงพูดต่อเนื่องที่ทำการตัดหัวท้ายหน่วยแล้ว

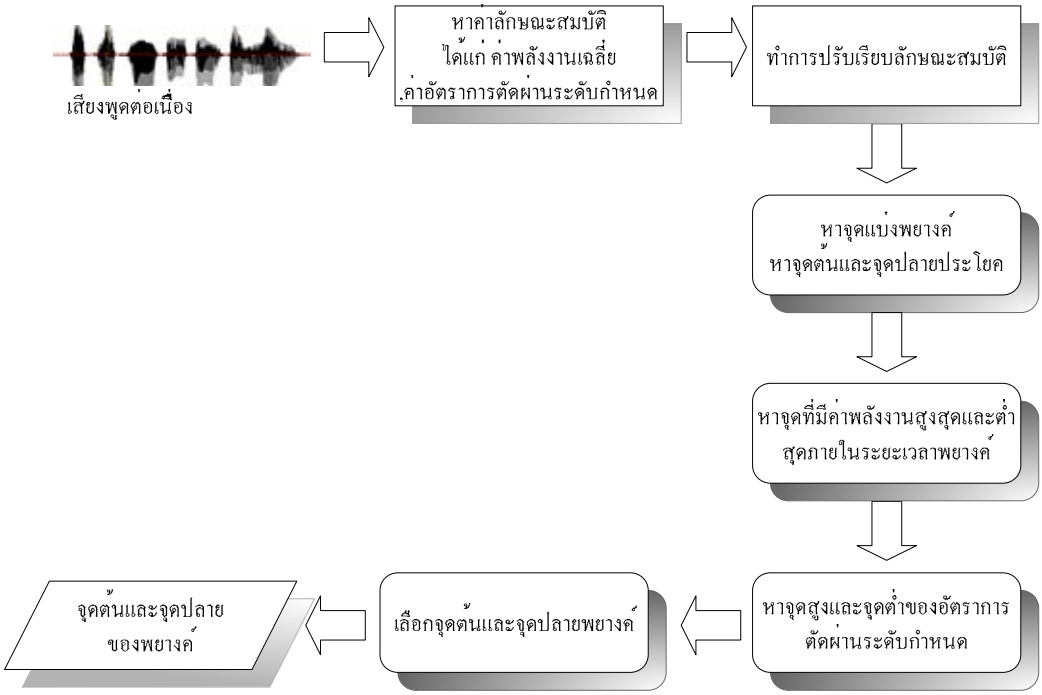
2. การตัดแบ่งพยางค์

ขั้นตอนการตัดแบ่งพยางค์มีขั้นตอนดังต่อไปนี้

1) รับสัญญาณเสียงพูดที่เป็นประโยคคำพูดต่อเนื่องเข้ามาแล้ววิเคราะห์หาค่าคุณลักษณะของสัญญาณเสียงพูด โดยกำหนดคุณลักษณะที่ใช้ในระบบนี้คือ พลังงานเฉลี่ย และอัตราการตัดผ่านระดับกำหนด ซึ่งกำหนดค่าระดับกำหนด (L) ไว้มีค่าเท่ากับ 0
2) ทำการปรับเรียบคุณลักษณะของสัญญาณเสียงพูด โดยใช้วิธีปรับเรียบด้วยค่าเฉลี่ยเคลื่อนไหว
3) พิจารณาหาจุดแบ่งพยางค์ โดยใช้ค่าพลังงานเฉลี่ยและอัตราการตัดผ่านระดับกำหนดเป็นตัวพิจารณา การหาจุดแบ่งพยางค์สามารถแบ่งเป็นขั้นตอนย่อยๆ ได้ดังนี้
   (3.1) ทำการหาจุดต้น (Beginning Point) และจุดปลาย (Ending Point) ของประโยค โดยพิจารณาจากค่าพลังงานของเสียงทั้งหมด ค่าของพลังงานเสียงที่มากกว่าค่าที่กำหนดระดับให้เป็นจุดต้นของประโยคจะเป็นจุดเริ่มต้นของประโยคนั้น ส่วนค่าพลังงานเสียงที่น้อยกว่าค่าที่กำหนดระดับให้เป็นจุดปลายของประโยค ก็จะเป็นจุดปลายของประโยคนั้น โดยการกำหนดค่าระดับของจุดต้นกับจุดปลายประโยคนั้นจะต้องทำการหา และปรับค่าบรรทัดฐานจากระบบจริงเนื่องจากค่าที่ได้จะไม่เท่ากันถ้าสิ่งแวดล้อมของระบบเปลี่ยนไป
   (3.2) หาจุดพลังงานสูงสุดภายในระยะเวลาพยางค์ คือ การหาค่าพลังงานที่มากที่สุดในระยะเวลาพยางค์ โดยจุดที่ทำการพิจารณาจะเป็นจุดสูงสุดในระยะเวลาพยางค์ ถ้าจุดนั้นมีค่าพลังงานมากที่สุดในกรอบหน้าต่างเท่ากับระยะเวลาพยางค์ โดยมีจุดดังกล่าวเป็นจุดกึ่งกลางกรอบหน้าต่าง
   (3.3) หาจุดพลังงานต่ำสุดภายในระยะเวลาพยางค์ คือ การหาจุดที่มีค่าพลังงานต่ำสุดระหว่างจุดสูงสุดภายในระยะเวลาพยางค์ 2 จุดที่อยู่ต่อเนื่องกัน โดยเมื่อทำการหาจุดต่ำสุดภายในระยะเวลาพยางค์ได้แล้ว ให้เลื่อนจุดสูงสุดภายในระยะเวลาพยางค์ไปทางขวาทีละจุด แล้วพิจารณาหาจุดต่ำสุดภายในระยะเวลาพยางค์ตัวต่อไป
   (3.4) หาจุดสูงและจุดต่ำของอัตราการตัดผ่านระดับกำหนด โดยจุดต่ำของอัตราการตัดผ่านระดับกำหนดคือ จุดที่มีอัตราการตัดผ่านระดับกำหนดน้อยที่สุดที่อยู่ระหว่างหาจุดพลังงานสูงสุดภายในระยะเวลาพยางค์ 2 จุดที่อยู่ต่อเนื่องกัน ส่วนจุดสูงของอัตราการตัดผ่านระดับกำหนด คือ จุดที่มีค่าอัตราการตัดผ่านระดับกำหนดมากที่สุดที่อยู่ระหว่างจุดพลังงานสูงสุดภายในระยะเวลาพยางค์ กับจุดต่ำของอัตราการตัดผ่านระดับกำหนด
   (3.5) พิจารณาเลือกจุดต้นและจุดปลายของพยางค์ โดยจะพิจารณาจากอัตราการตัดผ่านระดับกำหนดของจุดที่พิจารณาเป็นอันดับแรก ถ้ามีค่าน้อยกว่าระดับที่กำหนดไว้ให้ไปพิจารณาค่าความต่างของอัตราการตัดผ่านระดับกำหนด ถ้ามากเกินกว่าระดับที่กำหนดไว้ แสดงว่าจุดนั้นคือจุดปลายพยางค์ ถ้าไม่มากเท่าระดับที่กำหนดไว้ให้ไปพิจารณาค่าความต่างของพลังงานระหว่างจุดพลังงานสูงสุดและจุดพลังงานต่ำสุดที่จุดที่พิจารณาอยู่ระหว่างจุดทั้งสอง หากมีค่ามากกว่าระดับที่กำหนดไว้ จุดนั้นก็เป็นจุดปลายพยางค์ ซึ่งเมื่อหาจุดปลายพยางค์ได้แล้วจุดนั้นก็จะเป็นจุดต้นของพยางค์ต่อไป

4) ทำการปรับปรุงขอบเขตของพยางค์ โดยการตรวจสอบว่ามีพยางค์ไหนบ้างที่ไม่สมบูรณ์ เช่น มีระยะเวลาพยางค์สั้นเกินกว่าค่าที่กำหนด ก็ให้ตัดพยางค์นั้นออกไป แล้วนำไปรวมกับพยางค์ก่อนหน้านี้หนึ่งพยางค์

ขั้นตอนและวิธีในการตัดแบ่งพยางค์แสดงได้ดังรูปที่ 4.4 ต่อไปนี้



รูปที่ 4.4 ขั้นตอนการตัดแบ่งพยางค์

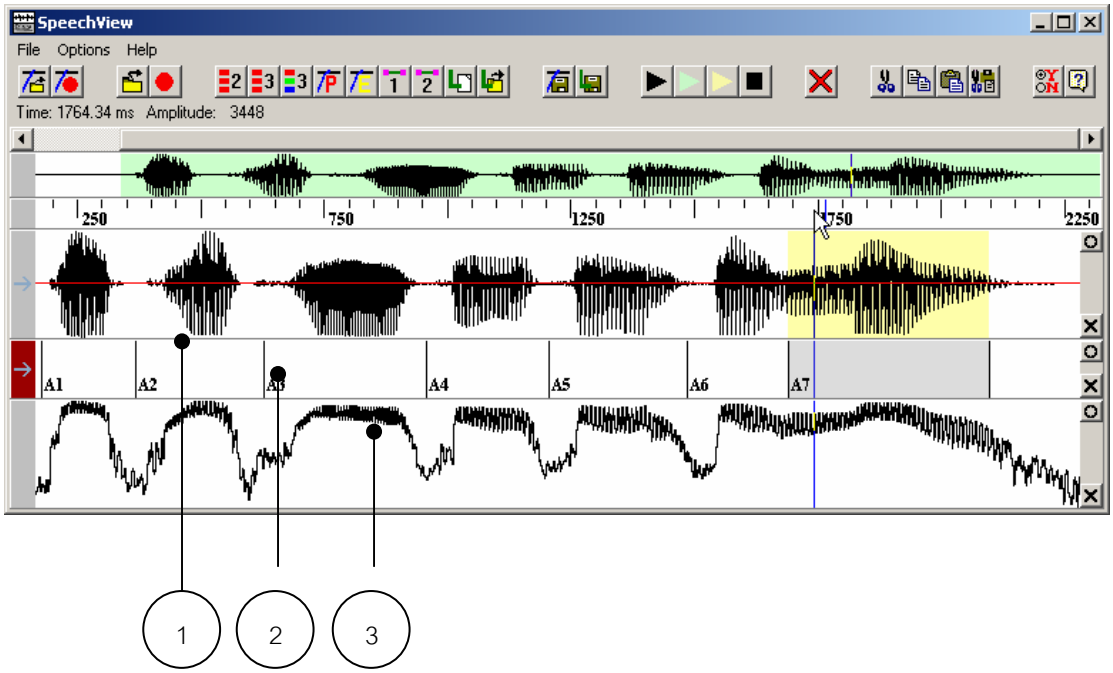
ในการตัดแบ่งพยางค์แต่ละพยางค์จะได้ค่าจุดต้นและจุดปลายของพยางค์ซึ่งมีหน่วยเป็นเวลา เพื่อให้เป็นการง่ายและสะดวกในการตรวจสอบความถูกต้องในการตัดแบ่งพยางค์ จึงได้ใช้โปรแกรม SpeechView ซึ่งอยู่ในชุดโปรแกรม CSLU toolkit รุ่น 2.0.0 พัฒนาโดย Center for Spoken Language Understanding, Oregon Graduate Institute of Science and Technology เป็นเครื่องมือที่ช่วยในการตรวจสอบความถูกต้องในการตัดแบ่งพยางค์ โดยโปรแกรมดังกล่าวจะเก็บค่าจุดต้นและจุดปลายของพยางค์ไว้ในแฟ้มที่มีนามสกุล .phn และรูปแบบภายในแฟ้มมีลักษณะดังนี้

```
MillisecondsPerFrame: 1.0
END OF HEADER
100     420     A1
420     740     A2
740     1100    A3
1100    1390    A4
1390    1620    A5
1620    1980    A6
1980    2520    A7
```

ภายในแฟ้มประกอบด้วยส่วนของหัวของแฟ้ม (Header) ซึ่งแสดงถึงหน่วยเวลาที่ใช้ต่อหนึ่งเฟรมมีหน่วยเป็นมิลลิวินาที และส่วนที่สองแสดงถึงเวลาของจุดต้นและจุดปลายของพยางค์ใน

แต่ละพยางค์นั้นๆ ซึ่งสดมภ์แรกเป็นเวลาของจุดต้นของพยางค์แต่ละพยางค์ สดมภ์ที่สองเป็นเวลาของจุดปลายของพยางค์นั้นๆ และสดมภ์ที่สามเป็นชื่อของพยางค์แต่ละพยางค์

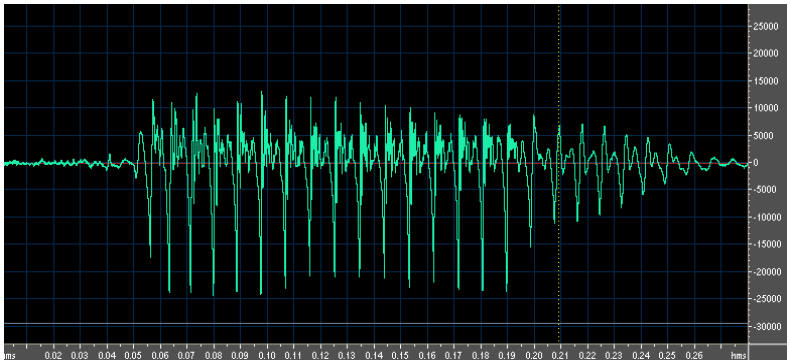
การใช้งานโปรแกรม SpeechView [14] เพื่อเป็นเครื่องมือที่ช่วยในการตรวจสอบความถูกต้องในการตัดแบ่งพยางค์ แสดงดังรูปที่ 4.5



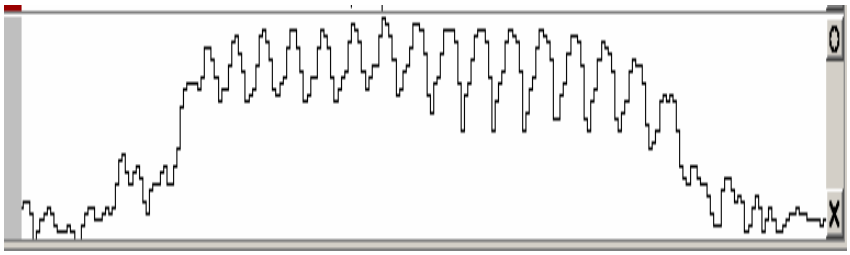
รูปที่ 4.5 การใช้โปรแกรม SpeechViewเป็นเครื่องมือตรวจสอบความถูกต้องในการแบ่งพยางค์

จากรูปที่ 4.5 ในส่วนที่ 1 แสดงค่าของแอมปลิจูดของเสียงที่ทำการตัดหัวท้ายหน่วยแล้ว ส่วนที่ 2 แสดงชื่อของพยางค์และขอบเขตของพยางค์แต่ละพยางค์ ซึ่งนำค่าของจุดต้นและจุดปลายที่อ่านค่าจากแฟ้มนามสกุล .phn มาใช้งาน และส่วนที่ 3 แสดงค่าพลังงานของเสียงทั้งหมด

แฟ้มข้อมูลของเสียงที่ได้จากการตัดแบ่งพยางค์ จะถูกเก็บไว้ในทั้งหน่วยความจำสำรองและหน่วยความจำหลักเพื่อการตรวจสอบความถูกต้องจะทำได้ง่ายขึ้น ดังนั้นเราจึงสามารถนำเอาแฟ้มข้อมูลของเสียงที่ตัดแบ่งพยางค์แล้วมาเปิดดูเพื่อตรวจสอบความถูกต้องว่าตัดแบ่งพยางค์ได้ถูกต้องมากเพียงใดหรืออาจตรวจสอบจากค่าพลังงานของเสียงที่ตัดแบ่งพยางค์แล้วนั้นๆ ดังรูปที่ 4.6 และรูปที่ 4.7



รูปที่ 4.6 สัญญาณเสียงที่ได้จากการตัดแบ่งพยางค์



รูปที่ 4.7 พลังงานของเสียงที่ได้จากการตัดพยางค์

4.1.2 การรู้จำเสียงพูด

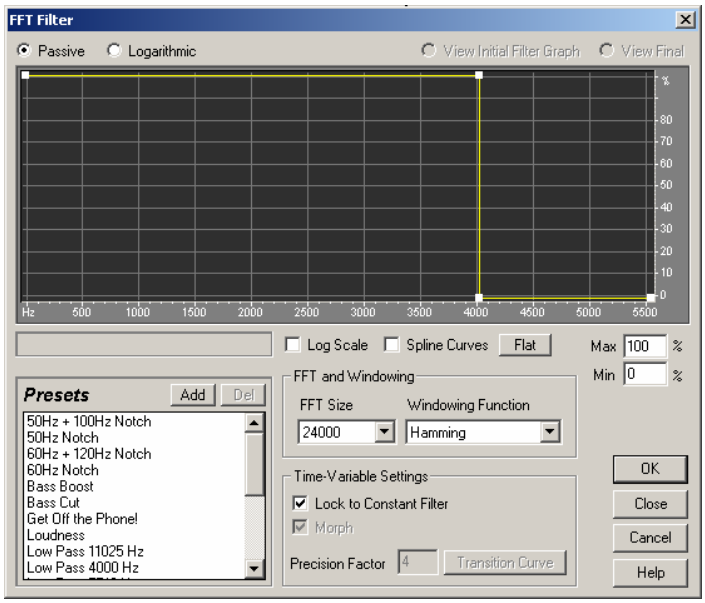
การรู้จำเสียงพูดนั้นจัดเป็นการจับคู่รูปแบบ (Pattern Matching) ชนิดหนึ่งแต่ค่อนข้างจะมีความซับซ้อนมาก และต้องอาศัยการเปรียบเทียบหลายขั้นตอน โดยจะนำสัญญาณทางสวนศาสตร์ (acoustic signal) มาตรวจสอบและจัดโครงสร้างให้เป็นลำดับขั้น (Hierarchy) ของหน่วยเสียง คำ พยางค์ โดยในแต่ละขั้นก็จะค่อยๆ เพิ่มข้อจำกัด (Constraint) เข้าไปเรื่อยๆ เช่น การออกเสียงของคำที่รู้จัก หรือการเรียงคำที่ถูกหลักไวยากรณ์ ซึ่งจะช่วยให้สามารถทดแทนกับความผิดพลาดหรือความไม่แน่นอนในระดับที่ต่ำกว่าได้ โดยลำดับขั้นของข้อจำกัดนี้จะเป็นประโยชน์ต่อการตัดสินใจโดยอาศัยหลักของความน่าจะเป็น ณ ระดับที่ต่ำกว่าทั้งหมด ทำให้ได้ผลลัพธ์ที่ถูกต้องที่สุดในระดับที่สูงที่สุด

งานวิจัยนี้ได้นำเสนอการรู้จำเสียงพูดภาษาไทยโดยใช้โครงข่ายประสาทเทียมเป็นเทคนิคของการเรียนรู้ที่นำมาใช้

4.1.2.1 ข้อมูลทางเสียง

ข้อมูลเสียงได้ถูกเก็บมาดังต่อไปนี้

1) เก็บข้อมูลเสียงพูดจากผู้พูดที่ใช้ภาษาไทย ไม่จำกัดสำเนียง ทั้งชายและหญิงอายุระหว่าง 18 – 45 ปี จำนวน 50 คน เพศละ 25 คน แต่ละคนให้อ่านหมายเลขจากชุดข้อมูลตัวเลข 7หลัก ที่สร้างเองชุดละ 5 หมายเลข โดยใช้รูปแบบการพูดปกติ 2 รอบ ดังนั้นข้อมูลจะประกอบด้วยหมายเลขโทรศัพท์ทั้งสิ้น 500 ชุดหมายเลข (3,500 ตัวเลข)
2) บันทึกเสียงลงเครื่องคอมพิวเตอร์ Packard Bell รุ่น Legend 1008T AP โดยใช้โปรแกรม Cool Edit Pro รุ่น 2.1 (ปัจจุบันเปลี่ยนชื่อเป็นโปรแกรม Adobe Audition) ของบริษัท Adobe Systems Incorporated จัดเก็บเป็นแฟ้มข้อมูลเสียง (Sound File) ในรูปแบบของ Wav 1 แฟ้ม แบบโมโน 16 บิต ด้วยอัตราการชักข้อมูล (Sampling Rate) 11,025 เฮิรตซ์ โดยแต่ละแฟ้มจะประกอบด้วยเสียงตัวเลขต่อเนื่อง 1 ชุด ถ้ามีประโยคใดที่ผู้พูดออกเสียงไม่ดีหรือมีความผิดพลาดก็จะอัดทับเฉพาะคำนั้นๆใหม่
3) หลังจากทำการเก็บบันทึกเสียงแล้ว จะทำการกรองสัญญาณที่ไม่ได้ต้องการทิ้ง ด้วย FFT Filter โดยกำหนดค่าพารามิเตอร์ต่างๆ คือ Hamming Window ขนาด 24000 และช่วงความถี่ที่ผ่านได้ (Band Pass) คือ 60-4000 เฮิรตซ์ ซึ่งจะช่วยทำให้การนำไปเรียนรู้นั้นได้ผลดียิ่งขึ้น การทำ FFT Filter นั้นจะใช้โปรแกรม Cool Edit Pro เช่นเดียวกับการบันทึกเสียง



รูปที่ 4.8 วิธีการกรองสัญญาณเสียงด้วยโปรแกรม Cool Edit

4.1.2.2 การแบ่งชุดข้อมูล

ข้อมูลเสียงพูดที่เก็บไว้ประกอบด้วยข้อมูลจากผู้พูดทั้งหมด 50 คน เป็นชาย 25 คนและหญิง 25 คน จากนั้นทำการแบ่งข้อมูลเป็น 2 ชุด คือ ชุดฝึก และชุดทดสอบ

ชุดฝึก (Training Set) ประกอบด้วยข้อมูลจากผู้พูดทั้งหมด 40 คน แบ่งเป็นชาย 20 คน หญิง 20 คน รวมทั้งสิ้น 2,800 ตัวเลข

ชุดทดสอบ (Test Set) ประกอบด้วยข้อมูลจากผู้พูดทั้งหมด 10 คน แบ่งเป็นชาย 5 คน หญิง 5 คน รวมทั้งสิ้น 700 ตัวเลข

4.1.2.3 วิธีการทางสวนศาสตร์

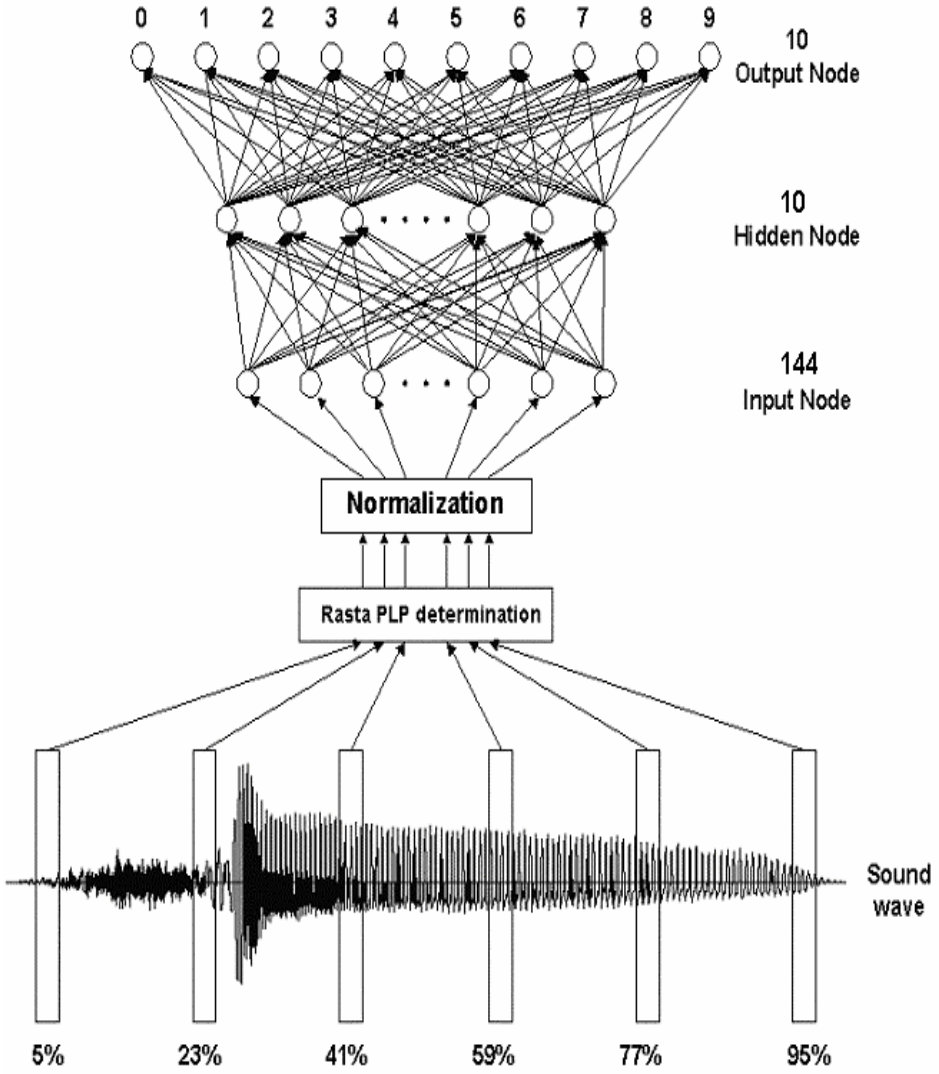
การหาคุณลักษณะที่สำคัญของสัญญาณเสียงที่ใช้ในงานวิจัยนี้คือ ราสต้า-พีแอลพี และอนุพันธ์อันดับหนึ่ง ซึ่งมีค่าพารามิเตอร์ดังนี้

1. อันดับของราสต้า-พีแอลพี ซึ่งใช้ราสต้า-พีแอลพีอันดับ 12
2. จำนวนกรอบ เพื่อให้ทราบว่าจำนวนกรอบเท่าไรจึงให้ค่าความถูกต้องได้ดีที่สุด จำนวนกรอบที่เลือกใช้ได้แก่ 6, 9, 12, 15 กรอบ โดยแต่ละกรอบจะมีความยาว 25 มิลลิวินาที ตำแหน่งในการเลือกกรอบมีดังนี้
   - จำนวนกรอบ 6 กรอบ เลือกกรอบจากตำแหน่งที่ 5, 23, 41, 59, 77, 95 เปอร์เซ็นต์ ตามแกนเวลาของข้อมูลเสียง
   - จำนวนกรอบ 9 กรอบ เลือกกรอบจากตำแหน่งที่ 5, 16, 28, 39, 50, 61, 73, 84, 95 เปอร์เซ็นต์ ตามแกนเวลาของข้อมูลเสียง
   - จำนวนกรอบ 12 กรอบ เลือกกรอบจากตำแหน่งที่ 5, 13, 21, 30, 38, 46, 54, 62, 70, 79, 87, 95 เปอร์เซ็นต์ ตามแกนเวลาของข้อมูลเสียง
   - จำนวนกรอบ 15 กรอบ เลือกกรอบจากตำแหน่งที่ 5, 11, 18, 24, 31, 37, 44, 50, 56, 63, 69, 76, 82, 89, 95 เปอร์เซ็นต์ ตามแกนเวลาของข้อมูลเสียง

4.1.3 โครงข่ายประสาทเทียม

โครงข่ายประสาทเทียมที่ใช้ในส่วนนี้ใช้วิธีความผิดพลาดแบบแพร่กระจายย้อนกลับ (Error Back-Propagation) และกำหนดค่าโมเมนตัมเท่ากับ 0.9 และค่าอัตราการเรียนรู้ที่ใช้เท่ากับ 0.0001 และ 0.00001 ตามลำดับ

ในการฝึกฝนโครงข่ายประสาทเทียมมีหลักการทำงานของโปรแกรมก็คือ สร้างโครงข่ายที่ประกอบด้วยระดับชั้นข้อมูลเข้า (Input Layer) ระดับชั้นฮิดเดน (Hidden Layer) และระดับชั้นข้อมูลออก (Output Layer) ซึ่งแต่ละชั้นติดต่อกันหมด จากนั้นทำการฝึกฝนให้โครงข่ายข่าย โดยการอ่านข้อมูลเข้าของชุดฝึก กำหนดค่าน้ำหนักเริ่มต้นโดยการสุ่ม แล้วนำค่าน้ำหนักที่ได้มาหาค่าของข้อมูลออก แล้วมาเปรียบเทียบกับค่าเป้าหมายเพื่อหาเปอร์เซ็นต์ความถูกต้องและค่าความผิดพลาด เพื่อนำมาใช้ในการปรับค่าน้ำหนักสำหรับรอบถัดไป ค่าน้ำหนักที่ได้ในแต่ละรอบจะถูกนำไปใช้ในการหาข้อมูลขาออกของชุดทดสอบ แล้วนำมาเทียบกับค่าเป้าหมาย หาเปอร์เซ็นต์ความถูกต้อง นั่นคือ หนึ่งรอบของการฝึก จากนั้นจะทำการฝึกเรื่อยๆ ซึ่งเป็นการปรับค่าน้ำหนักเพื่อให้ได้เปอร์เซ็นต์ความถูกต้องเมื่อเทียบกับชุดทดสอบให้มากที่สุด

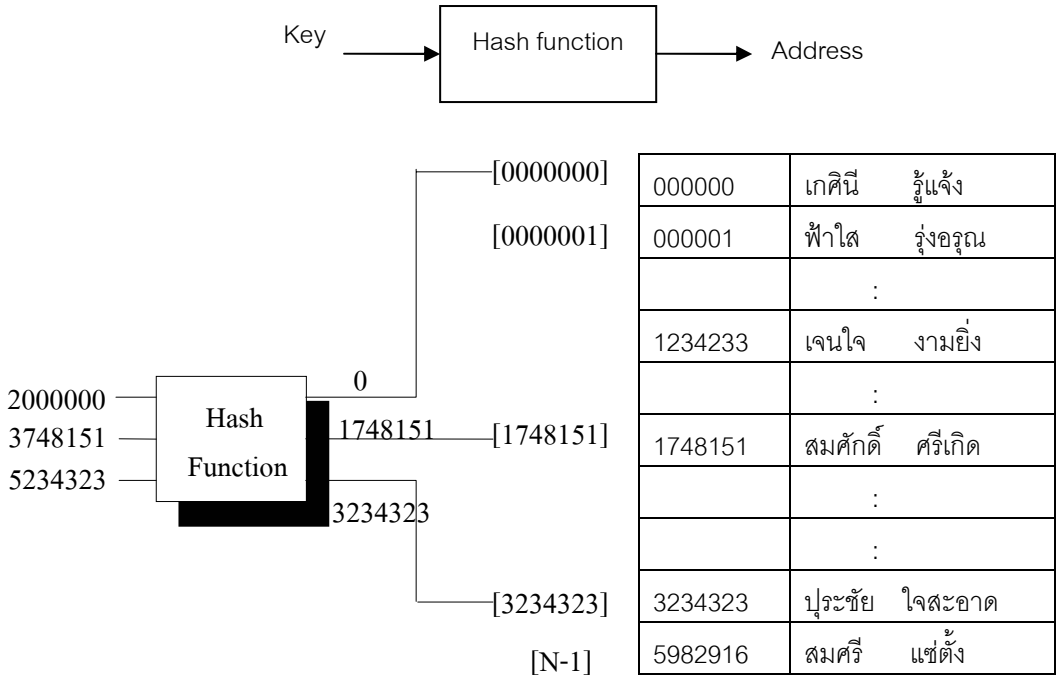


รูปที่ 4.9 ขั้นตอนการรู้จำเสียงเมื่อใช้ข้อมูล 6 กรอบ

รูปที่ 4.9 แสดงขั้นตอนในการรู้จำเสียงของตัวเลขที่ทำการแบ่งพยางค์แล้ว โดยใช้จำนวนกรอบ 6 กรอบ เลือกกรอบจากตำแหน่งที่ 5, 23, 41, 59, 77, 95 เปอร์เซ็นต์ ตามแกนเวลาของข้อมูลเสียง กรอบละ 25 มิลลิวินาที แต่ละกรอบจะผ่านขั้นตอนการหาลักษณะสำคัญของเสียงต่างๆ และกรรมวิธีปรับบรรทัดฐาน จากนั้นจะนำข้อมูลที่ได้เข้าสู่โครงข่ายประสาทเทียม ซึ่งใช้จำนวนบัพของระดับชั้นข้อมูลออกเท่ากับ 10 บัพ เนื่องจากผลลัพธ์ที่ต้องการคือเลข 0-9 จำนวนสิบตัวเลข ส่วนจำนวนบัพของระดับชั้นข้อมูลเข้าคือ 144 บัพ ได้จากจำนวนกรอบคูณกับจำนวนอันดับของค่าคุณลักษณะสำคัญแล้วนำมาคูณด้วยสอง เนื่องจากการเพิ่มค่าอนุพันธ์อันดับหนึ่งเข้าไป และระดับชั้นฮิดเดนเท่ากับ 10 บัพ

4.1.4 การสืบค้นข้อมูลของรายนามผู้ใช้โทรศัพท์

การสืบค้นข้อมูลที่มีข้อมูลเป็นจำนวนมาก สิ่งที่ต้องคำนึงถึงที่เป็นส่วนสำคัญคือ ระยะเวลาที่ใช้ในการค้นคืนของข้อมูล ซึ่งในปัจจุบันมีผู้คิดค้นเทคนิคที่ใช้ในการค้นหาข้อมูลอยู่มากมายขึ้นอยู่กับความเหมาะสมและการนำไปประยุกต์ใช้งาน เทคนิคที่ใช้ในการค้นคืนและการเก็บข้อมูลของรายนามผู้ใช้โทรศัพท์ได้แก่ เทคนิคแบบแฮช (Hashing) เนื่องจากมีความเหมาะสมกับการค้นหาข้อมูลที่มีปริมาณมากและใช้เวลาในการทำงานเป็น O(1) โดย นำค่าคีย์ (key) ซึ่งได้แก่เลขหมายโทรศัพท์มาทำการหาค่าผลต่างกับค่า 2000000 ซึ่งเป็นเลขหมายโทรศัพท์เริ่มต้นในส่วนของโทรศัพท์พื้นฐานในเขตนครหลวงและปริมณฑลเนื่องจากงานวิจัยนี้ใช้ข้อมูลในส่วนของพื้นที่ในเขตนครหลวงและปริมณฑลเท่านั้น ผลลัพธ์ที่ได้จะเป็นค่า address ที่ใช้ในการเก็บของข้อมูลนั้น ซึ่งวิธีนี้จะสามารถทำให้ไม่เกิดการชนกันของข้อมูล ดังแสดงตัวอย่างในรูปที่ 4.10



รูปที่ 4.10 การใช้เทคนิคแบบแฮชชิ่งในการค้นหาข้อมูลรายนามผู้ใช้โทรศัพท์

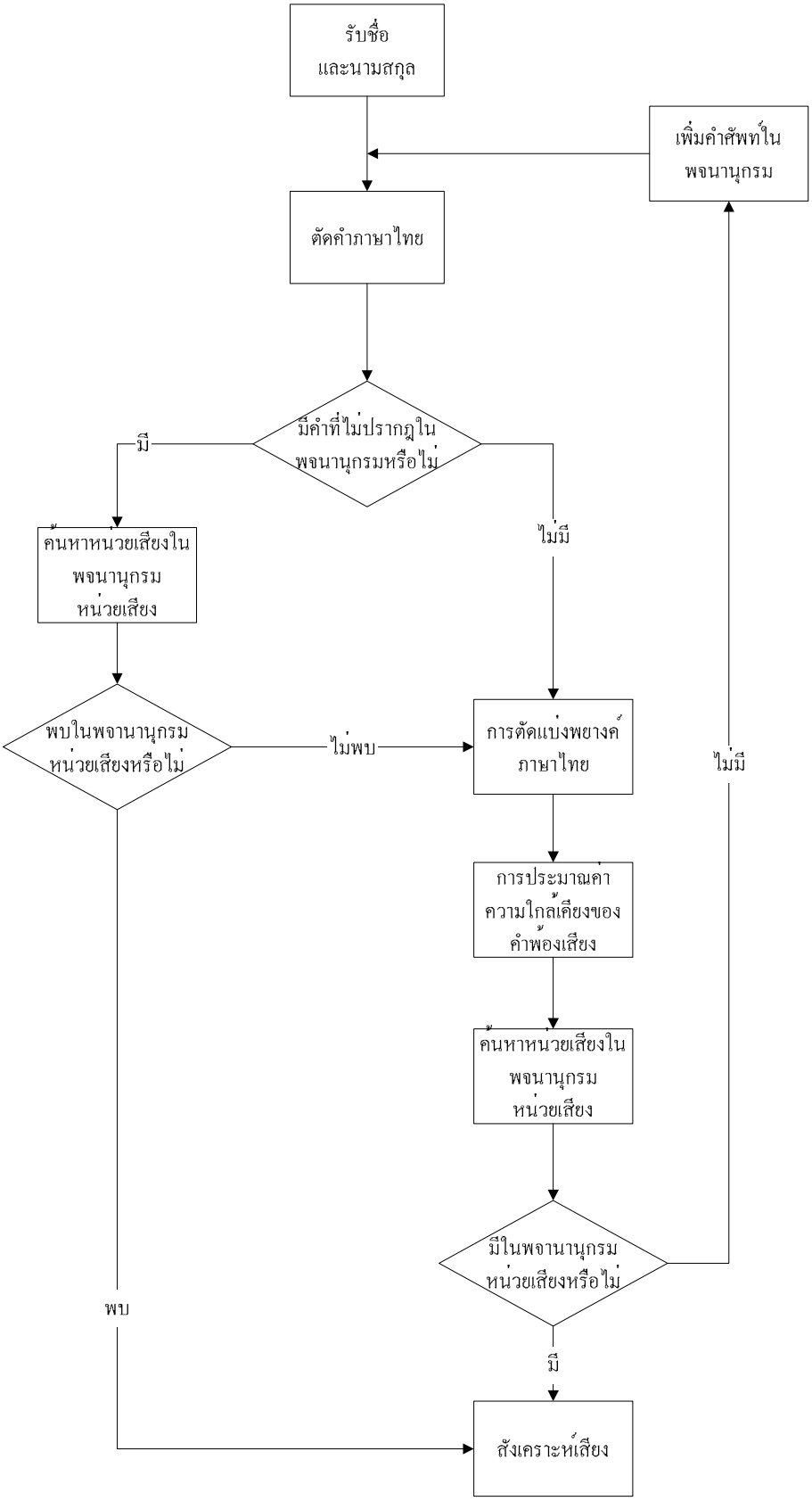
4.1.5 การสังเคราะห์เสียงพูดรายนามผู้ใช้โทรศัพท์

ในระบบสอบถามรายนามผู้ใช้โทรศัพท์แบบอัตโนมัติได้นำเอาหลักการสังเคราะห์เสียงพูดมาประยุกต์ใช้เพื่อใช้ในการแสดงผลออกมาทางเสียงที่ประกอบด้วยชื่อและนามสกุลของผู้ใช้โทรศัพท์ สำหรับในงานวิจัยนี้ได้นำโปรแกรม การสังเคราะห์เสียงพูดภาษาไทยสำหรับคำทับศัพท์ภาษาอังกฤษและคำนามเฉพาะ ของอัจจิมา ตันสกุล [97] มาประยุกต์ใช้งานร่วมกับระบบ โดยโปรแกรมจะมีหลักการทำงาน คือ นำเสียงที่เก็บไว้ก่อนแล้วนำมาต่อกันเป็นเสียงพูดที่ต้องการและมีหลักการตัดคำโดยใช้พจนานุกรม และวิธีตัดคำให้ยาวที่สุด (Longest Matching) ซึ่งการตัดคำโดยใช้พจนานุกรมที่ทำการหาขอบเขตของหน่วยคำที่เป็นวิสามานยนาม (ชื่อเฉพาะ) นั้น ถ้าหากเก็บทุกชื่อหรือนามสกุลที่มีอยู่ในฐานข้อมูลลงในพจนานุกรมทั้งหมด จากนั้นก็ค้นหาและเปรียบเทียบหาคำศัพท์นั้นๆ ว่ามีอยู่ในพจนานุกรมหรือไม่ เพียงเท่านั้นก็จะสามารถหาขอบเขตของคำแต่ละคำได้ทั้งหมด แต่ในความเป็นจริงแล้ว ถ้ามีข้อมูลจำนวนมากหรือเพิ่มข้อมูลให้มากขึ้นจะทำให้เกิดความยุ่งยากในการจัดเก็บลงในพจนานุกรมเป็นอย่างมาก ดังนั้นการบรรจุคำไว้ในพจนานุกรมจะมีทั้งหน่วยคำที่ย่อยที่สุดที่มีหรือไม่มีความหมาย อาจเป็นคำประสม หรือเป็นคำวิสามานยนามนั้นทั้งคำถ้าหากมีชื่อนั้นปรากฏในฐานข้อมูลเป็นจำนวนมาก

ในงานวิจัยนี้ ได้เพิ่มขั้นตอนของการสังเคราะห์ชื่อและนามสกุลเสียงพูดภาษาไทยเพื่อให้มีความถูกต้องมากยิ่งขึ้นจึงทำการเพิ่มเขตของข้อมูล (field) ของคำอ่านออกเสียงภาษาไทยของชื่อและนามสกุลจากฐานข้อมูลเดิมที่มีเพียงแค่ชื่อและนามสกุลเท่านั้น เพื่อช่วยทำให้การอ่านของชื่อและนามสกุลมีความถูกต้องใกล้เคียงความเป็นจริงมากยิ่งขึ้น เนื่องจากผู้ดูแลระบบจะสามารถแก้ไขคำอ่านให้ถูกต้องได้ โดยก่อนที่จะนำเอาฐานข้อมูลของชื่อและนามสกุลเข้านั้น ผู้ดูแลระบบจะต้องนำเอาชื่อและนามสกุลเหล่านั้น มาให้โปรแกรมสังเคราะห์เสียงพูดทำการอ่านออกเสียงแล้วผู้ดูแลระบบจะเป็นผู้ตัดสินใจว่าโปรแกรมอ่านออกเสียงได้ถูกต้องหรือไม่ ถ้าไม่ถูกต้องผู้ดูแลระบบจะทำการแก้ให้เป็นเสียงของคำอ่านที่ถูกต้องแล้วจึงนำเข้าในฐานข้อมูล

นอกจากนี้ยังได้เพิ่มส่วนของการอ่านคำพ้องเสียงที่ไม่พบในพจนานุกรม เนื่องจากในส่วนของการอ่านชื่อและนามสกุลนั้นบางครั้งพบคำที่ไม่ได้บรรจุไว้ในพจนานุกรมแต่เป็นคำพ้องเสียงของคำที่บรรจุไว้ในพจนานุกรม จึงทำให้ผู้ดูแลระบบต้องทำการเพิ่มคำเข้าไปในพจนานุกรมหรืออาจจะต้องไปแก้ในส่วนของคำอ่านให้เป็นคำที่บรรจุไว้ในพจนานุกรม ดังนั้นในงานวิจัยนี้จึงนำทฤษฎีเซตวิภัชนัยมาใช้เพื่อคำนวณหาค่าสัดส่วนของความเป็นสมาชิกของเซตวิภัชนัย เพื่อเปรียบเทียบหาพยางค์หรือคำที่พ้องเสียงที่ให้เสียงได้เหมือนกันที่สุด

ขั้นตอนของการสังเคราะห์เสียงพูดของรายนามผู้ใช้โทรศัพท์ แสดงดังรูปที่ 4.11



รูปที่ 4.11 ขั้นตอนของการสังเคราะห์เสียงพูด

4.1.5.1 การตัดคำภาษาไทย

ลักษณะทางโครงสร้างของภาษาอังกฤษจะมีประโยคซึ่งประกอบด้วยคำหลายๆ คำเรียงต่อกันไป โดยมีช่องว่าง เป็นตัวคั่นคำ การตัดคำจึงทำค่อนข้างสะดวก แต่สำหรับโครงสร้างทางภาษาไทยและภาษาอื่นๆ ในภูมิภาคเอเชียซึ่งมีความใกล้เคียงทางโครงสร้างภาษา เช่น ภาษาลาว ภาษาจีน ภาษาญี่ปุ่นจะมีความซับซ้อนมากกว่า โดยการเขียนประโยคภาษาไทยเป็นการเขียนติดต่อกันเป็นส่วนใหญ่ จึงต้องมีวิธีการให้คอมพิวเตอร์มีความสามารถที่จะเรียนรู้ขอบเขตของคำได้

หลักการตัดคำในภาษาไทยสามารถแบ่งออกได้เป็น 2 หลักการใหญ่คือ

ก. หลักการตัดคำโดยใช้กฎเกณฑ์ ซึ่งขั้นตอนการตัดคำในยุคแรกๆ จะใช้วิธีการตรวจสอบกฎเกณฑ์ของคำภาษาไทย เช่น กฎเกณฑ์ของตัวอักษรที่อยู่ติดกัน หรือกฎเกณฑ์ที่กำหนดโดยราชบัณฑิตยสถาน วิธีการนี้มีข้อจำกัดมากนั่นคือผลของการตัดคำอาจได้กลุ่มของคำ ซึ่งในความเป็นจริงยังสามารถตัดคำแยกย่อยออกไปได้อีก นั่นคือความถูกต้องของคำหลังการตัดต่ำ

ข. หลักการตัดคำโดยใช้พจนานุกรม ในยุคต่อมาขั้นตอนวิธีการตัดคำภาษาไทยโดยส่วนใหญ่ จะใช้พจนานุกรมเข้ามาช่วย วิธีการนี้ถึงแม้จะใช้เนื้อที่ของหน่วยความจำหลักมาก แต่ก็เป็นวิธีที่ให้ความถูกต้องในการตัดคำสูงวิธีหนึ่ง ซึ่งในงานวิจัยนี้ได้เลือกขั้นตอนการตัดคำโดยใช้พจนานุกรมนี้

4.1.5.2 พจนานุกรมหน่วยเสียง

คำศัพท์ที่เก็บอยู่ในพจนานุกรมหน่วยเสียงนั้นแตกต่างจากคำศัพท์ในพจนานุกรมทั่วไป โดยจะเก็บคำศัพท์ทั้งระดับคำและระดับพยางค์ เพราะเก็บเป็นพยางค์ตามการอ่านออกเสียง เช่น คำว่า สิญชัย นำสิน พจนานุกรมการอ่านออกเสียงจะแยกเก็บเป็นคำว่า สิน นำ และ ชัย ทั้งนี้พยางค์ที่พ้องเสียงกันคือ สิน กับ สิญ พจนานุกรมจะทำการเก็บพยางค์ สิน เพียงหนึ่งพยางค์ เนื่องจากออกเสียงเหมือนกันสามารถใช้ข้อมูลเสียงร่วมกันได้ ซึ่งข้อมูลเสียงที่ใช้ในงานวิจัยนี้ได้นำบางส่วนมาจากงานวิจัยเรื่อง การสังเคราะห์ข้อความเสียงพูดภาษาไทยสำหรับคำทับศัพท์ภาษาอังกฤษและคำนามเฉพาะ ของอัจจิมา ตันสกุล [11] และได้ทำการแก้ไขเพิ่มเติมในส่วนที่เสียงไม่ชัดเจนคุณภาพไม่ดีและเพิ่มความเป็นธรรมชาติมากยิ่งขึ้น นอกจากนี้ยังได้ทำการเพิ่มคำศัพท์บางคำที่ใช้ในส่วนของชื่อและนามสกุลของรายนามผู้ใช้โทรศัพท์เป็นจำนวนหลายครั้งลงในพจนานุกรมหน่วยเสียง รวมทั้งสิ้น 4,682 เสียง โดยบันทึกเสียงบนเครื่องคอมพิวเตอร์ รุ่น Pentium4 การ์ดเสียงของบริษัท Creative Technology รุ่น Creative SB Audio PCI ไมโครโฟน AKG รุ่น D50S ใช้โปรแกรม Cool Edit2000 Version1.1 ในการเก็บแฟ้มข้อมูลเสียงแบบโมโน 16 บิต ที่อัตราการชักตัวอย่าง 11,025 เฮิรตซ์

4.1.6 การประมาณค่าความใกล้เคียงของคำพ้องเสียง

การเปรียบเทียบในการประมาณค่าความใกล้เคียงของคำพ้องเสียงเพื่อคำนวณหาค่าสัดส่วนของความเป็นสมาชิกของเซตวิภัชนัย ทำได้โดยการกำหนดให้มีการจัดกลุ่มอักขระตามการพ้องเสียง ทั้งนี้จะมีสัดส่วนของเซตวิภัชนัยที่ไม่เท่ากัน โดยกำหนดกลุ่มของอักขระดังต่อไปนี้

กลุ่มของพยัญชนะต้น

การจัดกลุ่มของพยัญชนะไทยที่ใช้ในงานวิจัยนี้ใช้วิธีการจัดกลุ่มโดยนำพยัญชนะ 16 ตัวที่ไม่ได้เกี่ยวข้องกับการผันมา นำมาจากพยัญชนะเดิม 13 ตัวซึ่งไม่มีส่วนในการผันวรรณยุกต์ได้แก่ ฆ ฌ ญ ฏ ฐ ฑ ฒ ณ ธ ภ ฬ ศ ษ (มีเพียง 15 คำที่ปรากฏในคำไทยที่มีการผัน) และ ฃ ฅ ฎ แม้จะเป็นกลุ่มพยัญชนะเติม ฎ ก็ใช้เขียนคำไทยแต่โบราณคำเดียวคือ กฎ นอกนั้นใช้เขียนคำแผลงจาก ฎ ในภาษาบาลี-สันสกฤต และมักใช้เป็นพยัญชนะต้น ส่วน ฃ ฅ แม้จะผันวรรณยุกต์ได้ แต่เมื่อเป็นพยัญชนะต้นที่เลิกใช้แล้ว ก็เท่ากับไม่มีส่วนในการผันวรรณยุกต์ โดยกลุ่มคำเหล่านี้มาพิจารณาถ้าคำใดมีพยัญชนะเสียงซ้ำกับตัวใดใน 28 ตัว ก็จัดเข้าหมู่เดียวกับตัวนั้นๆ ไป เป็นการกำหนดพื้นเสียงของคำ [98] (เสียงวรรณยุกต์ของคำที่ยังไม่ได้ผันจะเป็นเสียงวรรณยุกต์ใดขึ้นอยู่กับชนิดตัวอักษร ลักษณะพยางค์ (คำเป็น-คำตาย) และเสียงสระสั้น-ยาว) เพื่อใช้ในการอ่านและการพูดเท่านั้น ไม่ใช่เพื่อการผัน ซึ่งแสดงการจำแนกกุ่มของพยัญชนะไทยได้ในตารางที่ 4.1

ตารางที่ 4.1 การจัดกลุ่มเสียงพยัญชนะไทยเพื่อใช้ในการอ่านและพูด

ฎ ด	ฌ ช	
ฏ ต	ญ ย	
ฐ ถ	ฑ ฒ ธ ท	
ศ ษ ส	ณ น	
ฆ ค	ภ พ	
ฬ ล	ฅ ค	ฃ ข

กลุ่มของตัวสะกดไทย

มีหลักเกณฑ์การจัดกลุ่มตามมาตรา คือ แม่บทแจกลูกอักษรตามหมวดคำที่มีตัวสะกดหรือออกเสียงอย่างเดียวกัน แบ่งเป็น 8 มาตรา โดยนำมาประยุกต์ใช้เพื่อเปรียบเทียบตัวสะกด ถ้าอยู่ในกลุ่มนี้จะได้ค่าคะแนนความใกล้เคียงของคำพ้องเสียง ดังแสดงในตารางที่ 4.2

ตารางที่ 4.2 การจัดกลุ่มของเสียงตัวสะกดไทยที่สอดคล้องกัน 8 กลุ่ม

ก ข ค ฆ	น ญ ณ ร ล ฬ	ง
ด จ ช ซ ฌ ฎ ฏ ฐ ฑ ฒ ต ถ ท ธ ศ ษ ส	ย	ม
บ ป พ ฟ ภ	ว	

4.1.6.1 การกำหนดค่าตัววัดความใกล้เคียงของคำพ้องเสียง

จากหลักไวยากรณ์ภาษาไทยที่ใช้ในการประสมพยางค์หรือคำ ซึ่งพยางค์หนึ่งมีส่วนประสมต่างๆ เรียงตามลำดับดังนี้

1) สระนำ

2) พยัญชนะต้น

3) พยัญชนะควบกล้ำ

4) สระตาม

5) ตัวสะกด

6) วรรณยุกต์

การกำหนดค่าความใกล้เคียงของคำพ้องเสียงนั้น เราได้กำหนดตามความเหมาะให้กับคำพ้องเสียงของพยัญชนะต้นและสระไว้ดังนี้

ค่าตัววัดความใกล้เคียงของคำพ้องเสียงของพยัญชนะต้น = 6

ค่าตัววัดความใกล้เคียงของคำพ้องเสียงของตัวสะกด = 4

ส่วนประสมของพยางค์หรือคำที่เหลือได้แก่ สระนำ พยัญชนะควบกล้ำ สระตามวรรณยุกต์ ไม่ได้นำมาคำนวณหรือนำมาเป็นตัวเปรียบเทียบ ดังนั้นรูปของ สระนำ พยัญชนะควบกล้ำ สระตาม และวรรณยุกต์ จะยังคงรูปเดิมไม่เปลี่ยนแปลง

4.1.6.2 ความสัมพันธ์แบบวิภัชนัยของการประมาณค่าความใกล้เคียงของคำพ้องเสียง

การคำนวณหาค่าสัดส่วนของความเป็นสมาชิกของเซตวิภัชนัย เพื่อเปรียบเทียบหาพยางค์หรือคำที่มีค่าความใกล้เคียงของคำพ้องเสียงสูงที่สุด โดยได้กำหนดสมการได้ดังนี้

$$F = \frac{\left(\left(\sum C[i] \times M_i\right) \times 100\right)}{\left(\sum M_i \times n\right)}$$

โดยที่ F = ค่าสัดส่วนของความเป็นสมาชิกของเซตวิภัชนัย

$C[i]$ = สัดส่วนของเซตวิภัชนัยตามกลุ่มของตัวอักขระ

M_i = ค่าตัววัดความใกล้เคียงของคำพ้องเสียง

n = จำนวนครั้งของการใช้ค่าตัววัดความใกล้เคียงของคำพ้องเสียง

สัดส่วนของเซตวิภัชนัยที่ทำให้เกิดเป็นคำพ้องเสียง ตามกลุ่มของตัวอักขระต่างๆ จะอยู่ในช่วง [0,1] ดังตัวอย่างต่อไปนี้

คำที่ให้อ่านคือ สุญ คำที่ค้นพบในพจนานุกรม คือ สุน ซึ่งแสดงความสัมพันธ์ของเซตวิภัชนัยได้ตามรูปที่ 4.12

M_i	M_1	M_2
M_1	1	0
M_2	0	0.9

รูปที่ 4.12 ความสัมพันธ์แบบวิภัชนัยบน {M_1, M_2}

จากรูป M_1 เป็นค่าตัววัดความใกล้เคียงของพยัญชนะที่พ้องเสียงจากการจำแนกกลุ่มพยัญชนะภาษาไทย และ M_2 เป็นค่าตัววัดความใกล้เคียงของพยัญชนะที่พ้องเสียงจากการจำแนกกลุ่มตัวสะกดภาษาไทย ดังนั้นค่าสัดส่วนของเซตวิภัชนัยตามกลุ่มของตัวอักขระ = [1, 0.9] โดยที่ส่วนประกอบที่เหลือของคำเช่นสระ วรรณยุกต์ จะต้องไม่เปลี่ยนแปลงหรือผันเสียงในกรณีที่เป็นวรรณยุกต์ แล้วนำมาคำนวณหาค่าสัดส่วนของความเป็นสมาชิกของเซตวิภัชนัย

$$F = \frac{((1\times 6)+(0.9\times 4))\times 100}{((6\times 1)+(4\times 1))}$$

$$= \quad 96.6$$

และนำผลที่ได้จากการหาค่าสัดส่วนของความเป็นสมาชิกของเซตวิภัชนัยนั้นมาพิจารณา โดยค่าของคำที่เข้าใกล้ค่า 100 มากที่สุดจะเป็นคำที่มีค่าความใกล้เคียงของคำพ้องเสียงสูงที่สุด แต่ในกรณีที่มีค่าเท่ากันสองคำ จะเลือกคำแรกที่ค้นคืนได้ก่อนเป็นอันดับแรก

4.1.7 การหาค่าอัตราความถูกต้องของระบบ

เนื่องจากในระบบสอบถามรายนามผู้ใช้โทรศัพท์แบบอัตโนมัติแบ่งส่วนออกเป็น 2 ส่วน ใหญ่ๆ คือ การรู้จำเสียงพูด และสังเคราะห์เสียงพูด ดังนั้นในการคำนวณหาอัตราความถูกต้องของทั้งระบบก็สามารถหาได้โดยพูดหมายเลขที่ต้องการเข้าไปในระบบ แล้วให้ระบบพูดชื่อของเจ้าของหมายเลขนั้นออกมา จากนั้นหาค่าเปอร์เซ็นต์ความถูกต้องที่ระบบสามารถไปหาชื่อของรายนามผู้ใช้โทรศัพท์ออกมาแล้วออกเสียงชื่อและนามสกุลได้ถูกต้อง ซึ่งโดยความเป็นจริงแล้วในงานวิจัยนี้ส่วนของการสังเคราะห์เสียง วิจัยมีเจตนาที่จะให้ผลของการอ่านออกเสียงชื่อและนามสกุลมีค่าความถูกต้อง 100 เปอร์เซ็นต์ จึงได้เพิ่มเติมในส่วนของคำอ่านของชื่อและนามสกุล ถ้าชื่อหรือนามสกุลใดที่ระบบสังเคราะห์เสียงพูด พูดผิดหรือบางคำไม่มีในพจนานุกรมก็ให้ผู้ดูแลระบบเป็นผู้แก้ไขหรือทำการเพิ่มคำศัพท์เข้าไปในพจนานุกรม ซึ่งผลของค่าอัตราความถูกต้องจะขึ้นกับค่าอัตราความถูกต้องของการรู้จำเสียงพูดนั่นเอง

ดังนั้นอัตราความถูกต้องของระบบจะเท่ากับอัตราความถูกต้องของการรู้จำเสียงพูดในระดับคำต่อเนื่องซึ่งสามารถคำนวณได้จากวิธีการดังต่อไปนี้

กำหนดให้ A_i เป็นคำตอบที่ได้จากโครงข่ายประสาทเทียมตามลำดับของชุดหมายเลข

หลังจากที่หาค่า A_i ได้แล้ว ก็คำนวณหาความถูกต้อง โดยให้ค่าคะแนนความถูกต้อง (Score)กับชุดหมายเลขที่ถูกต้อง คือถ้าคำตอบที่ได้ตรงกับคำตอบที่ถูกต้องทุกตัวก็ให้คะแนนเท่ากับ 1 แต่ถ้ามีตัวหนึ่งตัวใดไม่ตรงกับคำตอบที่ถูกต้องจะให้คะแนนเท่ากับ 0 ดังสมการต่อไปนี้

$$score_j = \begin{cases} 1; & A_i = C_j \\ 0; & A_i \neq C_j \end{cases}$$

เมื่อ j คือ ลำดับของชุดหมายเลข

i คือ ลำดับของตัวเลขในชุดหมายเลข

C_j คือ คำตอบที่ถูกต้องทั้งลำดับและตัวเลขของชุดหมายเลข

ดังนั้นเมื่อทำการให้ค่าคะแนนความถูกต้องไปจนครบทุกชุดหมายเลขแล้ว ก็นำค่าคะแนนรวมของชุดหมายเลขทั้งหมด แล้วนำมาคำนวณหาค่าอัตราความถูกต้องของระบบได้ดังสมการต่อไปนี้

$$\text{อัตราความถูกต้องของระบบ} = \frac{\sum score_j \times 100}{\text{จำนวนชุดหมายเลขทั้งหมด}}$$

4.2 การทดลองและผลการทดลอง

ในหัวข้อนี้จะกล่าวถึงวิธีการทดลองและผลการทดลองของระบบสอบถามรายนามผู้ใช้โทรศัพท์อัตโนมัติซึ่งเป็นการทดลองหาผลของอัตราความถูกต้องของระบบทั้งหมด โดยในส่วนของการรับเสียงที่ทำหน้าที่รับเสียงพูดเข้ามาในระบบนั้น ในการทดลองเสียงที่ได้มานั้นมาจากสองแหล่ง คือ ไมโครโฟน และจากแฟ้มข้อมูลเสียงซึ่งเก็บเสียงพูดไว้เพื่อความสะดวกในการตรวจสอบความถูกต้อง

4.2.1 วิธีการทดลอง

1. นำแฟ้มข้อมูลเสียงของชุดทดสอบซึ่งประกอบด้วยข้อมูลจากผู้พูดทั้งหมด 10 คน แบ่งเป็นชาย 5 คน หญิง 5 คน จำนวนชุดหมายเลข 100 ชุด รวมทั้งสิ้น 700 ตัวเลข มาทำการทดสอบโดยนำเข้าโปรแกรมแล้วให้โปรแกรมทำการรู้จำเสียงแล้วแสดงผลการรู้จำที่ได้ออกมาทางจอภาพแล้วตรวจสอบอัตราความถูกต้องที่ได้
2. ทำการปรับค่าพารามิเตอร์ต่างๆ เพื่อหาค่าที่เหมาะสมเพื่อให้ได้ค่าอัตราความถูกต้องที่สูงเพียงพอ ค่าพารามิเตอร์ที่ใช้ในการรู้จำเสียงพูดที่ใช้ในงานวิจัยมีดังต่อไปนี้
   - เสียงที่ใช้บันทึกที่อัตราการชักข้อมูล 11,025 เฮิรตซ์ 16 บิต บันทึกแบบช่องสัญญาณเดี่ยว (Mono channels)
   - ลักษณะสำคัญของเสียงใช้ราสต้า-พีแอลพีอันดับที่ 12 จำนวนกรอบ 6, 9, 12, และ15กรอบตามลำดับ โดยแต่ละกรอบจะมีความยาว 25 มิลลิวินาที
   - โครงข่ายประสาทเทียมที่ใช้ในการเรียนรู้เป็นแบบเพอร์เซปตรอนหลายชั้น และวิธีการเรียนรู้แบบแพร่กระจายย้อนกลับ
   - ค่าโมเมนตัม 0.9 ค่าอัตราการเรียนรู้ 0.0001 และ 0.00001
   - จำนวนรอบที่นำมาทำการทดลอง 500, 2,000, และ 5,000 รอบ
   - จำนวนบัพของชั้นข้อมูลเข้าที่ใช้ 144, 216, 288, 360 บัพ
   - จำนวนบัพชั้นข้อมูลฮิดเดนที่ใช้ 10, 20, 50, 100 บัพ
   - จำนวนบัพของชั้นข้อมูลออก 10 บัพ
3. ทำการหาค่าความผิดพลาดซึ่งเกิดจากการตัดแบ่งพยางค์ โดยนำแฟ้มข้อมูลเสียงทั้งหมดมาผ่านขั้นตอนการตัดแบ่งพยางค์แล้วนำมาตรวจสอบโดยใช้โปรแกรม SpeechView เป็นเครื่องมือช่วยแล้ววิเคราะห์พยางค์ใดที่ตัดออกมาแล้วมีส่วนประกอบของพยางค์ไม่ครบซึ่งได้แก่ ส่วนของพยัญชนะ ส่วนสระ และส่วนตัวสะกด ถ้าพยางค์ใดขาดส่วนหนึ่งส่วนใดไปถือว่าเกิดความผิดพลาดของการตัดแบ่งพยางค์

ในงานวิจัยนี้เราได้แบ่งการทดลองออกเป็น 9 ชุดการทดลองดังตารางที่ 4.3 โดยแต่ละชุดการทดลองจะมีการปรับเปลี่ยนค่าพารามิเตอร์ตามความสนใจและความเหมาะสม ถ้าหากค่าใดที่ทำให้ค่าอัตราความถูกต้องสูงที่สุด จะนำค่านั้นมาใช้ในงานทดลองต่อไป

ตารางที่ 4.3 ค่าพารามิเตอร์ที่ปรับเปลี่ยนของแต่ละชุดการทดลอง

การทดลองชุดที่	1	2	3	4	5	6	7	8	9
จำนวนกรอบ	6	9	12	15	12	12	12	12	12
จำนวนชั้นฮิดเดน	1	1	1	1	1	1	2	2	3
ค่าอัตราการเรียนรู้	10^{-4}	10^{-4}	10^{-4}	10^{-4}	10^{-4}	10^{-4}	10^{-5}	10^{-5}	10^{-5}
จำนวนบัพของชั้นข้อมูลขาเข้า	144	216	288	360	288	288	288	288	288
จำนวนบัพของชั้นฮิดเดนชั้นที่หนึ่ง	10	10	10	10	50	100	10	50	10
จำนวนบัพของชั้นฮิดเดนชั้นที่สอง	-	-	-	-	-	-	20	100	50
จำนวนบัพของชั้นฮิดเดนชั้นที่สาม	-		-	-	-	-	-	-	100
จำนวนบัพของชั้นข้อมูลขาออก	10	10	10	10	10	10	10	10	10
จำนวนรอบที่ใช้ในการฝึก	500	2000	2000	2000	2000	2000	5000	5000	5000

4.2.2 ผลการทดลอง

จากการทดลองดังกล่าวข้างต้นได้ผลการทดลองดังตารางที่ 4.4

ตารางที่ 4.4 อัตราความถูกต้องของแต่ละชุดการทดลอง

การทดลองชุดที่	อัตราความถูกต้องของระบบ(%)	อัตราความถูกต้องระดับพยางค์(%)
1	69	92.9
2	71	92.9
3	73	93
4	73	92.9
5	73	93.6
6	73	93.6
7	75	94
8	75	94
9	73	93.6

การคำนวณหาค่าอัตราความถูกต้องของระบบ คำนวณได้จากนำค่าคะแนนรวมของชุดหมายเลขทั้งหมด ส่วนค่าอัตราความถูกต้องระดับพยางค์ คำนวณจากผลรวมของหมายเลขที่ถูกต้องคูณด้วยหนึ่งร้อยแล้วนำมาหารด้วยจำนวนหมายเลขเดี่ยวทั้งหมดที่ทำการทดลอง

ค่าความถูกต้องซึ่งเกิดจากการตัดแบ่งพยางค์มีค่าเท่ากับ 86 เปอร์เซนต์ ซึ่งได้จากการนำแฟ้มข้อมูลเสียงจำนวน 500 ชุดหมายเลข ที่บันทึกเมื่อนำเข้าขั้นตอนการตัดแบ่งพยางค์แล้วปรากฏว่าตัดแบ่งพยางค์ผิดพลาดคือมีส่วนประกอบของพยางค์ไม่ครบและต้องนำมาทำการแก้ไขก่อนนำไปสู่ในขั้นตอนฝึกของโครงข่ายประสาทเทียม จำนวน 70 ชุดหมายเลข

4.2.3 วิเคราะห์ผลการทดลอง

จากการทดลองชุดที่ 1, 2, 3 และ 4 เป็นการเปลี่ยนจำนวนกรอบจาก 6 เป็น 9 ,12 และ15 กรอบตามลำดับ ซึ่งผลปรากฏว่าจำนวนกรอบที่เท่ากับ 12 กรอบให้ค่าอัตราความถูกต้องของระบบและอัตราความถูกต้องระดับพยางค์สูงที่สุดคือ 73 และ 93 เปอร์เซ็นต์ ตามลำดับ ดังนั้นจึงใช้จำนวนกรอบ 12 กรอบในชุดการทดลองต่อไป ในการทดลองชุดที่ 5 และ 6 เป็นการทดลองเพิ่มจำนวนบัพของชั้นฮิดเดนของโครงข่ายประสาทเทียมเป็น 50 บัพ และ 100 บัพ ตามลำดับ ผลการทดลองปรากฏว่าให้ค่าอัตราความถูกต้องของระบบและอัตราความถูกต้องระดับพยางค์เพิ่มขึ้น แต่ทำเวลาในการประมวลผลเมื่อนำไปใช้งานจริงเพิ่มขึ้นเป็น 2 เท่า คือจากเดิมใช้เวลาในส่วนของการรู้จำเสียงประมาณ 1 วินาที เพิ่มขึ้นเป็นประมาณเกือบ 2 วินาที ซึ่งก็เป็นเวลาที่ยอมรับได้เมื่อนำไปใช้ในระบบจริง ส่วนในการทดลองที่ 7 และ 8 จึงเป็นการเพิ่มจำนวนชั้นของชั้นฮิดเดนและทดลองปรับจำนวนบัพในแต่ละชั้นที่ต่างกัน นอกจากนั้นยังทำการปรับค่าอัตราการเรียนรู้กับจำนวนรอบที่ใช้ในการฝึกเพื่อให้ได้ค่าอัตราการรู้จำที่ยอมรับได้โดยใช้เวลาไม่นาน ซึ่งผลปรากฏว่ามีค่าอัตราความถูกต้องของระบบและค่าอัตราความถูกต้องระดับพยางค์เพิ่มขึ้นมาเป็น 75 และ 94 เปอร์เซ็นต์

ตามลำดับ หลังจากนั้นได้ทำการทดลองชุดที่ 9 โดยการเพิ่มจำนวนชั้นของชั้นฮิดเดนเป็น 3 ชั้น ผลปรากฏว่าให้ค่าอัตราความถูกต้องลดลงเล็กน้อย ซึ่งเมื่อทำการทดลองทั้ง 9 ชุดแล้วได้ค่าพารามิเตอร์ที่เหมาะสมที่สุด ได้แก่ ใช้ลักษณะทางสวนศาสตร์คือ ราสต้า-พีแอลพี และอนุพันธ์อันดับหนึ่ง โดยใช้จำนวนกรอบ 12 กรอบ ส่วนการเรียนรู้ของโครงข่ายประสาทเทียมใช้จำนวนชั้นของข้อมูลขาเข้า 288 บัพ จำนวนชั้นฮิดเดน 2 ชั้น ชั้นที่หนึ่ง 50 บัพ และชั้นที่สอง 100 บัพ และจำนวนชั้นข้อมูลขาออก 10 บัพ

จากผลการทดลองพบว่า การปรับค่าทั้งส่วนของลักษณะทางสวนศาสตร์และส่วนของการเรียนรู้ของโครงข่ายประสาทเทียมเพิ่มค่าอัตราความถูกต้องของระบบไม่มากนัก ดังนั้นถ้าต้องการที่จะเพิ่มค่าอัตราความถูกต้องให้มากขึ้น จึงควรปรับปรุงในส่วนของการตัดพยางค์ให้มีค่าความผิดพลาดให้น้อยที่สุด โดยอาจใช้ค่าต่างๆร่วมประกอบในการพิจารณาหาจุดตัดของพยางค์ เช่น ค่าสัมประสิทธิ์อัตสหสัมพันธ์ (Autocorrelation Coefficients) และปรับปรุงส่วนของการตัดพยางค์ให้มีความละเอียดขึ้นเนื่องจากในงานวิจัยนี้จุดปลายของพยางค์หนึ่งเป็นจุดต้นของพยางค์หนึ่ง ซึ่งในกรณีที่ผู้พูดเว้นระยะเวลาในการพูดแต่ละพยางค์นานก็จะทำให้การตัดพยางค์รับเอาเสียงเงียบที่อยู่ระหว่างพยางค์เข้าไปด้วย

4.3 สรุปผลของระบบการสอบถามรายนามผู้ใช้โทรศัพท์

งานส่วนนี้เป็นการนำเอาเทคโนโลยีทางด้านการรู้จำเสียงพูดและการสังเคราะห์เสียงพูดภาษาไทยมาพัฒนาแล้วนำมาประยุกต์ใช้กับระบบสอบถามรายนามผู้ใช้โทรศัพท์แบบอัตโนมัติโดยจำลองระบบลงบนเครื่องไมโครคอมพิวเตอร์รับเสียงข้อมูลขาเข้าโดยผ่านทางไมโครโฟนและส่งเสียงข้อมูลขาออกโดยผ่านทางลำโพง

ระบบสอบถามรายนามผู้ใช้โทรศัพท์แบบอัตโนมัตินั้นจะประกอบด้วย 2 ขั้นตอนใหญ่ๆ ได้แก่ การรู้จำเสียงพูดภาษาไทย และการสังเคราะห์เสียงพูดภาษาไทย

ในขั้นตอนของการรู้จำเสียงพูดซึ่งเป็นเสียงพูดของตัวเลขต่อเนื่องนั้น สามารถแบ่งเป็นขั้นตอนย่อยได้แก่ ขั้นตอนการหาจุดตัดหัวท้ายหน่วยและการตัดแบ่งพยางค์ ซึ่งผู้วิจัยได้เสนอวิธีในการตัดแบ่งโดยใช้ค่าพลังงานและค่าอัตราตัดผ่านแกนศูนย์ร่วมกันในการพิจารณาหาจุดแบ่ง ซึ่งให้ผลของอัตราความถูกต้องในการตัดแบ่งพยางค์เท่ากับ 86 เปอร์เซ็นต์ ส่วนขั้นตอนการรู้จำเสียงพูดได้ใช้คุณลักษณะสำคัญของเสียงคือราสต้า-พีแอลพี อันดับที่ 12 และอนุพันธ์อันดับที่ 1 และใช้โครงข่ายประสาทเทียมช่วยในการฝึกฝนและการรู้จำ ซึ่งได้ค่าพารามิเตอร์ที่เหมาะสมสำหรับการเรียนรู้จำเสียงพูดตัวเลขต่อเนื่องคือ จำนวนกรอบ 12 กรอบ จำนวนชั้นข้อมูลฮิดเดน 2 ชั้น จำนวนบัพของชั้นข้อมูลขาเข้า 288 บัพ จำนวนบัพของชั้นฮิดเดนที่หนึ่ง 50 บัพ จำนวนบัพของชั้นฮิดเดนที่สอง 100 บัพ จำนวนบัพของชั้นข้อมูลขาออก 10 บัพ ซึ่งได้ค่าอัตราความถูกต้องของระบบเท่ากับ 75 เปอร์เซ็นต์และได้ค่าอัตราความถูกต้องระดับพยางค์ 94 เปอร์เซ็นต์

ส่วนในขั้นตอนการสังเคราะห์เสียงพูดใช้วิธีการสังเคราะห์เสียงโดยใช้วิธีการตัดคำโดยใช้เปรียบเทียบจากพจนานุกรมแล้วนำหน่วยเสียงที่ได้จากพจนานุกรมหน่วยเสียงมาต่อกันเป็นเสียงพูด ในงานวิจัยนี้ได้เพิ่มเติมส่วนที่เป็นคำอ่านของชื่อและนามสกุลเพื่อให้มีค่าความถูกต้องมากที่สุดคือ 100 เปอร์เซ็นต์ และได้เสนอวิธีการประมาณค่าความใกล้เคียงของคำพ้องเสียงในกลุ่มของพยัญชนะต้นและตัวสะกด โดยใช้ทฤษฎีเซตวิภัชนัยเข้ามาช่วยเพื่อลดจำนวนการเพิ่มคำศัพท์ในพจนานุกรม ซึ่งในชื่อและนามสกุลจะพบคำพ้องเสียงอยู่ 20 เปอร์เซ็นต์โดยประมาณ

5. สรุปโครงการ

โครงการนี้มีวัตถุประสงค์เพื่อศึกษาวิจัยระบบการรู้จำเสียงพูดภาษาไทยและได้พัฒนาระบบการรู้จำเสียงพูดอัตโนมัติต้นแบบขึ้นมา 2 ระบบ คือ

- ระบบการโอนสายโทรศัพท์อัติโนมัติจากเสียงพูดชื่อไทย
- ระบบสอบถามรายนามผู้ใช้โทรศัพท์โดยผ่านระบบโทรศัพท์

5.1 ระบบการโอนสายโทรศัพท์อัติโนมัติจากเสียงพูดชื่อไทย

ระบบการโอนสายสายโทรศัพท์อัติโนมัติจากเสียงพูดชื่อไทย เป็นตัวอย่างของการประยุกต์ระบบการรู้จำเสียงพูดอัตโนมัติเพื่ออำนวยความสะดวกแต่ผู้ใช้บริการ

งานวิจัยนี้ได้พัฒนาต้นแบบเพื่อใช้สำหรับการโอนสายโทรศัพท์ด้วยการพูดชื่ออาจารย์และบุคลากรของภาควิชาวิศวกรรมคอมพิวเตอร์ จุฬาลงกรณ์มหาวิทยาลัย จำนวน 45 ชื่อ เพียงผู้ใช้บริการโทรเข้ามาและพูดชื่ออาจารย์หรือบุคคลาการของภาคฯ ระบบจะทำการโอนสายไปยังห้องอาจารย์หรือบุคลากรคนนั้นโดยอัตโนมัติ ผู้ใช้บริการที่โทรเข้ามาติดต่อภาควิชาไม่จำเป็นต้องจำเบอร์ภายในทั้งหมดของภาควิชา ขณะเดียวกันก็เป็นการลดภาระของโอเปอเรเตอร์ในภาควิชาไปในตัว หรือแม้แต่การโทรภายในภาควิชาระหว่างอาจารย์และบุคลากรก็ก่อให้เกิดความสะดวกรวดเร็วเนื่องจากไม่ต้องจำเบอร์ของอาจารย์และบุคลากรด้วยกัน

ระบบฯ แบ่งออกเป็นสองส่วน คือส่วนที่เป็นกระบวนการเรียนรู้ และส่วนที่เป็นกระบวนการรู้จำ โดยลักษณะสำคัญทางสวนศาสตร์ และเทคนิกการเรียนที่ใช้ คือ การทำงานเชิงเส้นแบบรับรู้และโครงข่ายประสาทเทียมตามลำดับ

ผลการทดลองด้วยการปรับเปลี่ยนพารามิเตอร์ต่างๆ ได้ผลว่าอันดับการทำนายเชิงเส้นแบบรับรู้ที่ให้ผลดีที่สุดคืออันดับที่ 18 โดยไม่ใช้อนุพันธ์เชิงเส้นแบบรับรู้ให้ผลดีกว่าการใช้อนุพันธ์เชิงเส้นแบบรับรู้อันดับที่หนึ่งและอันดับที่สอง ส่วนจำนวนเฟรมการวิเคราะห์เท่ากับ 47 จะให้ผลดีกว่าจำนวนอื่น ขณะที่การวนรอบปรับน้ำหนักของโครงข่ายประสาทเทียม 1000 รอบ จะให้ผลดีกว่าการวนปรับน้ำหนักที่มากกว่านั้นคือ 10000 รอบ ผลการทดลองที่ดีที่สุดคือ 95.56 เปอร์เซ็นต์

เมื่อนำระบบไปใช้จริงเลือกใช้การทำนายเชิงเส้นแบบรับรู้อันดับที่ 9 เนื่องจากผลการรู้จำไม่ต่างจากอันดับ 18 น้อยมาก แต่ประมวลผลได้รวดเร็วกว่า พบว่าว่ามีความถูกต้องเพียง 79.33 เปอร์เซ็นต์ ทั้งนี้อาจเป็นเพราะว่าผู้ที่โทรเข้ามามีเสียงและสำเนียงการพูดชื่อคนที่หลากหลายกว่า รวมทั้งอาจเกิดความผิดพลาดจากการหาขอบเขตของเสียงพูด ซึ่งในข้อมูลการทดสอบมีการหาขอบเขตของเสียงพูดไว้เรียบร้อยแล้ว

5.2 ระบบสอบถามรายนามผู้ใช้โทรศัพท์โดยผ่านระบบโทรศัพท์

ในปัจจุบัน การให้บริการทางด้านข้อมูลข่าวสารกับลูกค้าโดยผ่านทางระบบโทรศัพท์เริ่มเข้ามามีบทบาทสำคัญต่อการให้บริการกับลูกค้า การให้บริการสอบถามรายนามผู้ใช้โทรศัพท์ก็เป็นการให้บริการหนึ่งที่สำคัญ ซึ่งปัจจุบันพนักงานตอบรับโทรศัพท์เป็นผู้ใหบริการ อย่างไรก็ตาม การนำหลักการของระบบรู้จำเสียงพูดหรือระบบสังเคราะห์เสียงพูดมาประยุกต์ใช้งาน จะก่อให้เกิดความรวดเร็วทันต่อเหตุการณ์และลดค่าใช้จ่ายในการเพิ่มประสิทธิภาพของระบบงาน ซึ่งระบบการรู้จำ

เสียงพูดแบบต่อเนื่องและการสังเคราะห์เสียงพูดภาษาไทยในปัจจุบัน ได้มีการพัฒนาไปได้ระดับหนึ่งแล้ว สามารถนำมาประยุกต์ ใช้งานกับชีวิตประจำวันของมนุษย์ได้จริงในปัจจุบัน

โครงการวิจัยนี้ได้นำเทคนิคของการรู้จำเสียงพูดและสังเคราะห์เสียงพูดภาษาไทยมาประยุกต์ใช้ในการพัฒนาระบบดังกล่าว เพื่อให้ผู้ใช้งานได้เกิดความพึงพอใจและความสะดวกรวดเร็วในการให้บริการเพิ่มมากขึ้น ส่วนประกอบสำคัญของระบบประกอบด้วย ส่วนการรู้จำเสียงพูดตัวเลขต่อเนื่องภาษาไทยระดับพยางค์ และส่วนของการสังเคราะห์เสียงพูดชื่อและนามสกุลของผู้ใช้โทรศัพท์ภาษาไทย

ในขั้นตอนการตัดแบ่งพยางค์อัตโนมัติของเสียงพูดต่อเนื่อง ได้นำหลักเกณฑ์ของค่าพลังงานของเสียงและค่าอัตราการตัดผ่านระดับกำหนดมาใช้เป็นเกณฑ์ในการแบ่งพยางค์ ส่วนลักษณะสำคัญทางสวนศาสตร์ที่นำมาใช้ได้แก่ ราสต้า-พีแอลพี และอนุพันธ์อันดับที่หนึ่ง และเทคนิคการเรียนรู้ที่ใช้ในการรู้จำคือ โครงข่ายประสาทเทียม ซึ่งใช้การฝึกแบบแพร่กระจายความผิดพลาดย้อนกลับ ส่วนของการสังเคราะห์เสียงพูดใช้วิธีการนำหน่วยเสียงย่อยที่ทำการเก็บไว้ในพจนานุกรมหน่วยเสียง แล้วนำมาต่อกันเป็นเสียงพูดชื่อและนามสกุล และได้นำเอาทฤษฎีเซตวิภัชนัยมาช่วยคำนวณหาค่าความใกล้เคียงของคำพ้องเสียง เพื่อลดจำนวนคำศัพท์ที่เพิ่มขึ้นในพจนานุกรมหน่วยเสียง ผลการทดลองของทั้งระบบปรากฏว่าให้ค่าความถูกต้องของระบบ 75 เปอร์เซ็นต์ และให้ค่าความถูกต้องระดับพยางค์ 94 เปอร์เซ็นต์

6. บรรณานุกรม

[1] P. Price and J. Picone, “Automatic Speech Recognition: Better than Text?”, Invited Talk, American Association for the Advancement of Science, 2000.

[2] D. Jurafsky and J. H. Martin. Speech and Language Processing. Prentice Hall. New Jersey. 2000.

[3] B. H. Juang, “Progress & Challenges in Automatic Recognition and Understanding of Spoken Language”, Invited Talk, In Proc. of International Symposium toward the Realization of Spontaneous Speech Engineering, 2000.

[4] F. Jelinek, Statistical Methods for Speech Recognition, MIT Press, 1997.

[5] B. H. Juang and S. Furui, “Automatic Recognition and Understanding of Spoken Language – A First Step toward Natural Human-Machine Communication”, In Proc. of IEEE, Vol. 88, No. 8, pp. 1142-1165, 2000.

[6] V. Zue, et al., “PEGASUS: A Spoken Language Interface for On-line Air Travel Planning”, Speech Communication, Vol. 15, pp. 331-340, 1994.

[7] J. R. Glass, et al., “Multilingual Spoken Language Understanding in the MIT VOYAGER System”, Speech Communication, Vol. 17, pp. 1-18, 1995.

[8] S. Seneff and J. Polifroni, “A New Restaurant Guide Conversational System: Issues in Rapid Prototyping for Specialized Domains”, In Proc. of ICSLP’96, 1996.

[9] H. Meng, et al., “WHEELS: A Conversational System in the Automobile Classifieds Domain”, In Proc. of ICSLP’96, 1996.

[10] J. R. Glass and T. J. Hazen, “Telephone-based Conversational Speech Recognition in the Jupiter Domain”, In Proc. of ICSLP’98, 1998.

[11] J. R. Glass, T. J. Hazen and I. L. Hetherington, “Real-time Telephone-based Speech Recognition in the Jupiter Domain”, In Proc. of ICASSP’99, Vol. 1, pp. 61-64, 1999.

[12] S. Seneff and J. Polifroni, “Dialogue Management in the Mercury Flight Reservation System”, In Proc. of ANLP-NAACL, 2000.

[13] IBM, ViaVoice, http://www-306.ibm.com/software/voice/viavoice/.

[14] ScanSoft, Dragon NaturallySpeaking, http://www.scansoft.com/naturallyspeaking/.

[15] A. Smailagic, D. Siewiorek, R. Martin and D. Reilly, “CMU Wearable Computers for Real-time Speech Translation”, In Proc. of ISWC’99, 1999.

[16] K. Sasipoka, S. Suebvisai, K. Kamutpat, ”A Comparison of Hidden Markov Models and Neural Networks for Phoneme-based Continuous Thai Speech Recognition”, Senior Project Report, Department of Computer Engineering, Chulalongkorn University, 2000.

[17] B. Gold and N. Morgan, Speech and Audio Signal Processing, Wiley Press, 1999.

[18] L. R. Rabiner and B.-H. Juang, Fundamental of Speech Recognition, Prentice Hall, 1993.

[19] S. Roweis, “Speech Processing Background”, November 1998.

[20] A. G. Bell, Patent no. 174,465, U.S., Patent Office, February 14 1876.

[21] A. G. Bell, Prehistoric Telephone Days, “National Geographic Magazine”, Vol. 41, pp. 223-242, 1922.

[22] T. A. Edison, Patent no. 200,521, U. S., Patent Office, February 19 1878.

[23] A. A. Markov. “An Example of Statistical Investigation in the Text of ‘Eugene Onyegen’ Illustrating Coupling of ‘Tests’ in Chains”, In Proc. of the Academy of Science, St. Petersburg, Vol. 7, pp. 153-162. 1913.

[24] E. David and O. Selfridge, “Eyes and Ears for Computers”, In Proc. of the IRE, pp. 1093-1101. May 1962.

[25] R. Koenig, H. K. Dunn, and L. Y. Lacy. “The Sound Spectrograph”, Journal of Acoustic Society of America, Vol. 18, pp. 19-49, 1946.

[26] C. E. Shannon, “A Mathematical Theory of Communication”, Bell System Technical Journal, Vol. 27, Issue. 3, pp. 379-423, 1948.

[27] K. Davis, R. Biddulph and S. Balashek, “Automatic Recognition of Spoken Digits”, Journal of Acoustic Society of America, Vol. 24, pp. 637-642, 1952.

[28] H. F. Olson and H. Bellar, “Phonetic Typewriter”, Journal of Acoustic Society of America, Vol. 28, Issue. 6, pp. 1072-1081, 1956.

[29] H. F. Olson, Music, Physics and Engineering, Dover, 1967.

[30] H. Dudley and S. Balashek, “Automatic Recognition of Phonetic Patterns in Speech”, Journal of Acoustic Society of America, Vol. 30, pp. 721-732, 1958.

[31] D. B. Fry and P. Denes, “The Solution of Some Fundamental Problems in Mechanical Speech Recognition”, Language and Speech, Vol. 1, pp. 35-58, 1958.

[32] P. Denes and M. V. Mathews, “Spoken Digit Recognition Using Time-frequency Pattern Matching”, Journal of Acoustic Society of America, Vol. 32, pp. 1450-1455, 1960.

[33] P. Denes, “On the Statistics of Spoken English”, Journal of Acoustic Society of America, Vol. 35, Issue. 6, pp. 892-905, 1963.

[34] B. Bogert, M. Healy and J. Tukey, “The Quefrency Analysis of Time Series for Echoes”, In M. Rosenblatt, ed., Proc. Symp. on Time Series Analysis, Chap. 15, Wiley, New York, pp. 209-243, 1963.

[35] A. V. Oppenheim, R. W. Schafer and T. G. Jr. Stockham, “Nonlinear Filtering of Multiplied and Convolved Signals”, In Proc. of IEEE, Vol. 56, pp. 1264-1291, 1968.

[36] T. B. Martin, A. L. Nelson and H. J. Zadel, “Speech Recognition by Feature Abstraction Techniques”, Tech. Report AL-TDR-64-176, Air Force Avionics Lab, 1964.

[37] J. W. Cooley and J. W. Tukey, “An Algorithm for the Machine Computation of Complex Fourier Series”, Mathematical Computations, Vol. 19, pp. 297-301, 1965.

[38] L. E. Baum and T. Petrie, “Statistical Inference for Probabilistic Functions of Finite State Markov Chains”, Annals of Mathematical Statistics, Vol. 37, pp. 1554-1563, 1966.

[39] L. E. Baum and J. A. Eagon. “An Inequality with Applications to Statistical Estimation for Probabilistic Functions of Markov Processes and to a Model for Ecology”, Bulletins of American Mathematical Society, Vol. 73, pp. 360-363, 1967.

[40] L. E. Baum, T. Petrie, G Soules and N. Weiss, “A maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains”, Annals of Mathematical Statistics, Vol. 41. Issue. 1, pp. 164-171, 1970.

[41] L. E. Baum, “An Inequality and Associated Maximization Technique in Statistical Estimation of Probabilistic Functions of a Markov Process”, Inequalities, Vol. 3, pp. 1-8, 1972.

[42] D. R. Reddy, “An Approach to Computer Speech Recognition by Direct Analysis of the Speech Wave”, Tech. Report No. C549, Computer Science Dept., Stanford Univ., September 1966.

[43] T. K. Vintsyuk, “Speech Discrimination by Dynamic Programming”, Kibernetka (Cybernetics), Vol 4, pp. 81-88, January-February 1968.

[44] T. K. Vintsyuk. “Element-wise Recognition of Continuous Speech Consisting of Words from a Specified Vocabulary”, Kibernetka (Cybernetics), Vol. 7, pp. 133-143, 1971.

[45] H. Sakoe and S. Chiba, “Dynamic Programming Algorithm Optimization for Spoken Word Recognition”, IEEE Transactions on Acoustics Speech and Signal Processing, Vol. 26, pp. 43-49, February 1978.

[46] F. Itakura and S. Saito, “Analysis-synthesis Telephony Based on the Maximum-likelihood Method”, In Proc. of the 6th International Congress Acoustics, 1968.

[47] B. Atal and S. Hanauer, “Speech Analysis and Synthesis by Prediction of the Speech Wave”, Journal of Acoustic Society of America, Vol. 50, pp. 637-655, 1972.

[48] J. D. Markel and Jr. A. H. Gray, “Linear Prediction of Speech”, Springer-Verlag, Berlin, 1976.

[49] J. Markhoul, “Linear Prediction: A Tutorial Overview”, In Proc. of IEEE, Vol. 63, No. 4, April 1975.

[50] J. Pierce, “Whither Speech Recognition?”, Journal of Acoustic Society of America, Vol. 46, pp. 1049-1051, 1969.

[51] D. Klatt, “Review of the ARPA Speech Understanding Project”, Journal of Acoustic Society of America, Vol. 62, pp. 1345-1366, 1977.

[52] J. Ferguson, Hidden Markov Models for Speech, Princeton, NJ, 1980.

[53] J. Baker, “The DRAGON System – An Overview”, IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. 23, pp. 24-29, 1975.

[54] C. Tappert, N. Dixon, A. Rabinowitz and W. Chapman, “Automatic Recognition of Continuous Speech Utilizing Dynamic Segmentation, Dual Classification, Sequential Decoding, and Error Recovery”, IBM Tech. Report, RADC-TR-71-146, Yorktown Heights, New York, 1971.

[55] R. Bakis, “Continuous-speech Word Spotting via Centisecond acoustic States”, IBM Res. Report, RC 4788, Yorktown Heights, New York, 1974.

[56] L. Bahl and F. Jelinek, “Decoding for Channels with Insertions, Deletions, and Substitutions with Applications to Speech Recognition”, IEEE Transactions on Information Theory, IT-21, pp. 404-411, 1975.

[57] F. Jelinek, L. Bahl and R. Mercer, “The Design of a Linguistic Statistical Decoder for the Recognition of Continuous Speech”, IEEE Transactions on Information Theory, IT-21, pp. 250-256, 1975.

[58] F. Jelinek, “Continuous Recognition by Statistical Methods”, In Proc. of IEEE, Vol. 64, pp. 250-256, 1975.

[59] L. R. Rabiner, “A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition”, In Proc. of IEEE, Vol. 77, No. 2, pp. 257-286, February 1989.

[60] A. P. Dempster, N. M. Laird and D. B. Rubin, “Maximum Likelihood from Incomplete Data via the EM Algorithm (With Discussion)”, Journal of the Royal Statistical Society series B, Vol. 39, pp. 1-38, 1977.

[61] P. Price, W. Fisher, J. Bernstein and D. Pallett, “The DARPA 1000-word Resource Management Database for Continuous Speech Recognition”, In Proc. of ICASSP’88, New York, pp. 651-654, 1988.

[62] National Institute of Standards and Technology, “TIMIT Acoustic-Phonetic Continuous Speech Corpus”, Speech Disc 1-1.1, NIST Order No. PB91-505065, 1990.

[63] S. Davis and P. Mermelstein, “Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences”, IEEE Transactions on Acoustics Speech and Signal Processing, Vol. 28, pp. 357-366, February 1980.

[64] H. Hermansky, “Perceptual Linear Predictive (PLP) Analysis of Speech”, Journal of Acoustic Society of America, Vol. 87, pp. 1738-1752, 1990.

[65] S. Furui, “Speaker Independent Isolated Word Recognizer using Dynamic Features of Speech Spectrum”, IEEE Transactions on Acoustics Speech and Signal Processing, Vol. 34, pp. 52-59, February 1986.

[66] S. Makino, T. Kawabata and K. Kido, “Recognition of Consonants Based on the Perceptron Model”, In Proc. of ICASSP’83, Boston, pp. 738-741, 1983.

[67] A. Waibel, T. Hanazawa, G. Hinton, K. Shikano and K. Lang, “Phoneme Recognition: Neural Networks vs. Hidden Markov Models”, In Proc. of ICASSP’88, New York, pp. 107-110, 1988.

[68] R. Lippmann and B. Gold, “Neural Classifiers Useful for Speech Recognition”, In IEEE Proc. of the First International Conference on Neural Networks, pp. 417-422, 1987.

[69] H. Bourlard and N. Morgan, Connectionist Speech Recognition – A Hybrid Approach, Kluwer Academic Publishers, 1994.

[70] N. Morgan and H. Bourlard, “Continuous Speech Recognition: An Introduction to the Hybrid HMM/Connectionist Approach”, Signal Processing Magazine, Vol. 12, pp. 25-42, 1995.

[71] T. Robinson, M. Hochberg and S. Renals, “The Use of Recurrent Neural Networks in Continuous Speech Recognition”, In C. H. Lee, F. K. Soong, and K. K. Paliwal, eds., Automatic Speech and Speaker Recognition, Kluwer, Boston, 1996.

[72] C. Weinstein, S. McCandless, L. Mondshein and V. Zue, “A System for Acoustic-phonetic Analysis of Continuous Speech”, IEEE Transactions on Acoustics Speech and Signal Processing, Vol. 23, pp. 54-67, 1975.

[73] V. Zue, “The Use of Speech Knowledge in Automatic Speech Recognition”, In Proc. of IEEE, Vol. 73, pp. 1602-1615, 1985.

[74] R. Cole, R. Stern and M. Lasry, “Performing Fine Phonetic Distinctions: Templates Versus Features”, In J. S. Perkell and D. M. Klatt, eds., Variability and Invariance in Speech Processes, Erlbaum, Hillsdale, New Jersey, 1986.

[75] H. Hermansky and N. Morgan, “RASTA Processing of Speech”, IEEE Transactions on Acoustics Speech and Signal Processing, pp. 578-589, 1994. (special issue on robust speech recognition)

[76] R. Stern, A. Acero, F.-H. Liu, Y. Ohshima, “Signal Processing for Robust Speech Recognition”, In C. H. Lee, F. K. Soong, and K. K. Paliwal, eds., Automatic Speech and Speaker Recognition, Kluwer, Boston, 1996.

[77] M. Gales and S. Young, “Robust Speech Recognition in Additive and Convolutional Noise using Parallel Model Combination”, Computer Speech and Language, Vol. 9, pp. 289-307, 1995.

[78] C.-H. Lee, “On Stochastic Feature and Model Compensation Approaches to Robust Speech Recognition”, Speech Communication, Vol. 25, pp. 29-48, 1998.

[79] D. Pallett, et al., ”DARPA HUB-4 Report”, National Institute of Science and Technology, February 1999.

[80] R. Cole, L. Hirschman, L. Atlas, M. Beckman, A. Bierman, M. Bush, J. Cohen, O. Garcia, B. Hanson, H. Hermansky, S. Levinson, K. McKeown, N. Morgan, D. Novick, M. Ostendorf, S. Oviatt, P. Price, H. Silverman, J. Spitz, A. Waibel, C. Weinstein, S. Zahorian, V. Zue, "The challenge of spoken language systems: Research direction for the nineties", IEEE Transactions on Speech and Audio Processing, Vol. 3, No.1, pp. 1-21, 1995.

[81] P. Ladefoged, A Course in Phonetics, Harcourt Brace Jovanovich Inc., 1975.

[82] กาญจนา นาคสกุล, ระบบเสียงภาษาไทย, พิมพ์ครั้งที่ 4, โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย, 2541.

[83] อุปกิตศิลปสาร, พระยา, หลักภาษาไทย, ไทยวัฒนาพานิช, 2533.

[84] The International Phonetic Association, http://www.arts.gla.ac.uk/ipa/ipa.html.

[85] S. Veltri, How to Make Your Computer Talk, McGraw-Hill, 1985.

[86] I. H. Witten, Principles Computer Speech, Academic Press Inc., 1982.

[87] F. L. Lederer, “Ear”, Microsoft Encarta Reference Library, 2003.

[88] H. Wada, “Dynamic Animation of Basilar Membrane in Cochlea when Otoacoustic Emissions are Generated”, In Proc. of the 4th China-Japan-USA-Singapore Conference on Biomechanics, pp. 203-206, 1995.

[89] T. Vilis, “Auditory Physiology”, Lecture Notes on Neurophysiology, University of Western Ontorio, 2004.

[90] T. Mitchell, Machine Learning, McGraw Hill, 1997.

[91] B. Kijsirikul, Artificial Intelligence, Lecture Notes on 2110654, Artificial Intelligence, Chulalongkorn University, 2003.

[92] S. Lertvilai, S. Wongthongserm, ”A Real Time Thai Isolated Digit Speech Recognition”, Senior Project Report, Department of Computer Engineering, Chulalongkorn University, 2002.

[93] Dialogic, Intel Corporation, System Release 5.1.1 for Windows Online Bookshelf, http://resource.intel.com/telecom/support/releases/winnt/SR511/docs/htmlfiles.

[94] The International Computer Science Institute, http://www.icsi.berkeley.edu.

[95] N. Storm, “Phoneme Probability Estimation with Dynamic Sparsely Connected Artificial Networks”, The Free Speech Journal, Issue No. 5, 1997.

[96] N. Storm, “The NICO Toolkit for Artificial Neural Networks”, http://www.speech.kth.se/NICO, 1996.

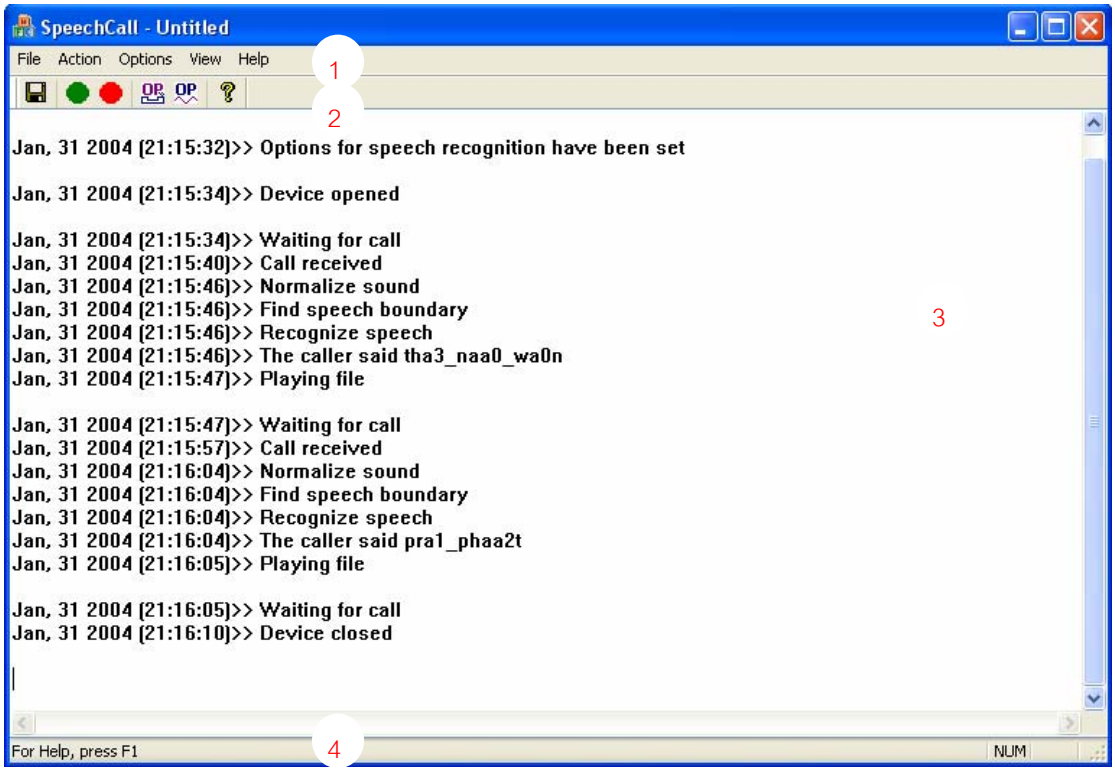
[97] อัจจิมา ตันสกุล, “การสังเคราะห์ข้อความเสียงพูดภาษาไทยสำหรับคำทับศัพท์ภาษาอังกฤษและคำนามเฉพาะ” วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย, 2544.

[98] พิศศรี กมลเวชช, ครบครันเรื่องวรรณยุกต์ พิมพ์ครั้งที่ 2. กรุงเทพมหานคร : หจก.หอรัตนชัยการพิมพ์, 2544.

7. ภาคผนวก – โปรแกรมโอนสายอัตโนมัติจากเสียงพูดชื่อไทยทางโทรศัพท์

7.1 ตัวเชื่อมประสานกับผู้ใช้

ตัวเชื่อมประสานกับผู้ใช้ของโปรแกรมเป็นดังรูปที่ 7.1



รูปที่ 7.1 ตัวเชื่อมประสานกับผู้ใช้

ตัวเชื่อมประสานกับผู้ใช้ประกอบด้วยส่วนต่างๆ ดังนี้

7.1.1 รายการ

ประกอบด้วยรายการหลักดังนี้

- **File** เป็นรายการที่ใช้สำหรับจัดการไฟล์และควบคุมการทำงานของโปรแกรมหลัก ประกอบด้วยรายการย่อยคือ การบันทึกข้อมูล และการปิดโปรแกรม
- **Action** เป็นรายการสำหรับเริ่มต้นและหยุดการทำงานของระบบโอนสายโทรศัพท์อัตโนมัติ ประกอบด้วยรายการย่อยคือ การเริ่มต้นระบบโอนสายโทรศัพท์อัตโนมัติ และการหยุดระบบโอนสายโทรศัพท์อัตโนมัติ

- **Options** เป็นรายการสำหรับการกำหนดค่าพารามิเตอร์ที่จำเป็นต้องใช้ในระบบ ประกอบด้วยรายการย่อยคือ การกำหนดค่าพารามิเตอร์ที่ต้องใช้สำหรับการโอนสายโทรศัพท์ และการกำหนดค่าพารามิเตอร์ที่ต้องใช้สำหรับการรู้จำเสียงพูด
- **View** เป็นรายการสำหรับกำหนดรูปแบบของตัวเชื่อมประสานกับผู้ใช้ ประกอบด้วยรายการย่อยคือ การกำหนดรูปแบบของแถบเครื่องมือ และการกำหนดรูปแบบของแถบสถานะ
- **Help** เป็นรายการสำหรับอธิบายส่วนต่างๆ ของโปรแกรมให้ผู้ใช้ทราบ ประกอบด้วยรายการย่อยคือ การอธิบายวิธีใช้โปรแกรม และการบอกข้อมูลเกี่ยวกับโปรแกรม

7.1.2 แถบเครื่องมือ

แถบเครื่องมือทำหน้าที่เช่นเดียวกับรายการ โดยแถบเครื่องมือที่ปรากฏเรียงจากซ้ายไปขวาได้แก่ แถบข้อมูลสำหรับการบันทึกข้อมูล แถบข้อมูลสำหรับการเริ่มต้นระบบโอนสายโทรศัพท์อัตโนมัติ แถบข้อมูลสำหรับการหยุดระบบโอนสายโทรศัพท์อัตโนมัติ แถบข้อมูลสำหรับการกำหนดค่าพารามิเตอร์ที่ต้องใช้สำหรับการโอนสายโทรศัพท์ แถบข้อมูลสำหรับการกำหนดค่าพารามิเตอร์ที่ต้องใช้สำหรับการรู้จำเสียงพูด และแถบข้อมูลสำหรับอธิบายวิธีใช้โปรแกรม

7.1.3 หน้าต่างข้อความ

หน้าต่างข้อความจะแสดงข้อความบอกถึงขั้นตอนการทำงานต่างๆ ของโปรแกรม

7.1.4 แถบสถานะ

แถบสถานะจะแสดงสถานะของโปรแกรมในช่วงเวลาหนึ่งๆ

7.2 วิธีใช้โปรแกรม

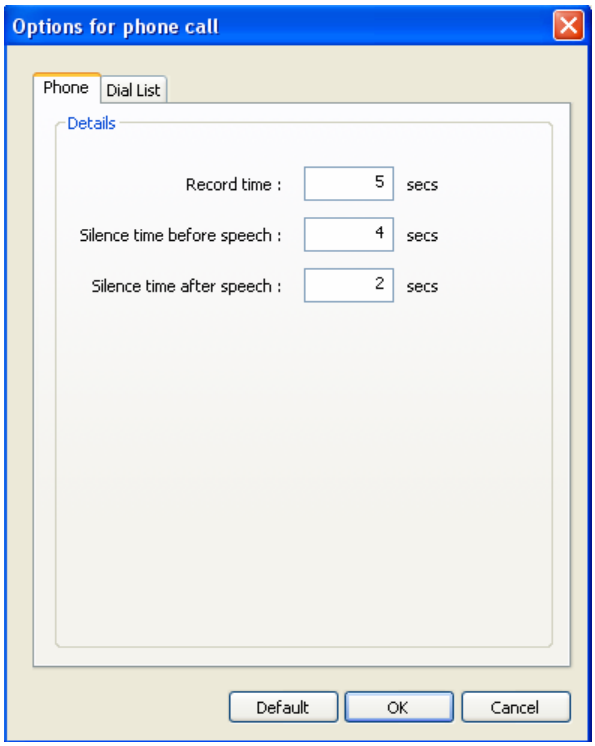
7.2.1 การปรับค่าพารามิเตอร์

ก่อนที่จะทำการเริ่มต้นระบบโอนสายโทรศัพท์อัตโนมัติ จำเป็นต้องกำหนดค่าพารามิเตอร์ให้ระบบเสียก่อน ซึ่งพารามิเตอร์ที่ใช้สามารถแบ่งได้เป็นสองจำพวก ได้แก่ พารามิเตอร์สำหรับการโอนสายโทรศัพท์ พารามิเตอร์สำหรับการรู้จำเสียงพูด

พารามิเตอร์สำหรับการโอนสายโทรศัพท์

พารามิเตอร์สำหรับการโอนสายโทรศัพท์สามารถกำหนดได้โดยการเลือกรายการ **Options->Phone Call...** หรือกดปุ่ม  บนแถบเครื่องมือ

เมื่อกระทำการข้างต้นแล้ว จะปรากฏกล่องโต้ตอบดังรูปที่ 7.2



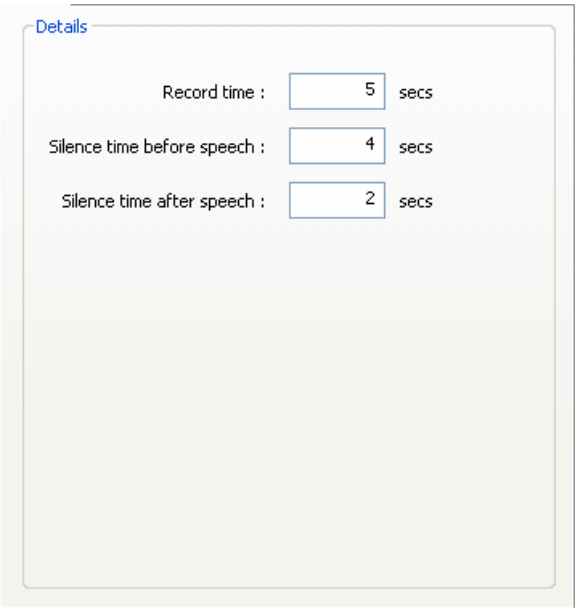
รูปที่ 7.2 กล่องโต้ตอบการปรับค่าพารามิเตอร์สำหรับการโอนสายโทรศัพท์

กล่องโต้ตอบสำหรับการปรับค่าพารามิเตอร์ที่ใช้ในการโอนสายโทรศัพท์ประกอบด้วยหน้าต่างๆ ดังนี้

- **Phone** เป็นหน้าสำหรับปรับค่าพารามิเตอร์สำคัญที่ต้องใช้ในการโอนสายโทรศัพท์อัตโนมัติ ซึ่งประกอบด้วย
  - **Record time** เป็นเวลาทั้งหมดที่ใช้ในการบันทึกเสียงพูดจากการโทรเข้า มีหน่วยเป็นวินาที
  - **Silence time before speech** เป็นเวลามากสุดสำหรับเสียงเงียบที่มาก่อนเสียงพูดจากการโทรเข้า มีหน่วยเป็นวินาที ถ้าไม่มีการพูดภายในระยะเวลาที่กำหนดนี้ การโทรเข้านั้นจะถูกตัดทันที

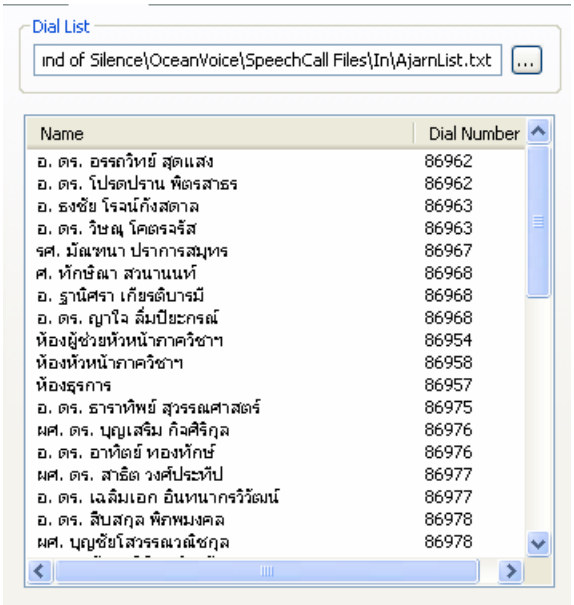
– **Silence time after speech** เป็นเวลามากสุดสำหรับเสียงเงียบที่มาหลังจากเสียงพูดจากการโทรเข้า มีหน่วยเป็นวินาที ในการโทรเข้า ถ้าหลังจากพูดไปแล้วเกิดการเงียบจนถึงเวลาค่านี้ การโทรเข้านั้นจะถูกตัดทันที

โดยหน้า Phone แสดงได้ดังรูปที่ 7.3



รูปที่ 7.3 หน้าการปรับค่าพารามิเตอร์ Phone

- **Dial List** เป็นหน้าสำหรับนำเข้าเบอร์โทรศัพท์เพื่อใช้ในการติดต่อและโอนสายภายใน โดยหน้า Dial List แสดงได้ดังรูปที่ 7.4

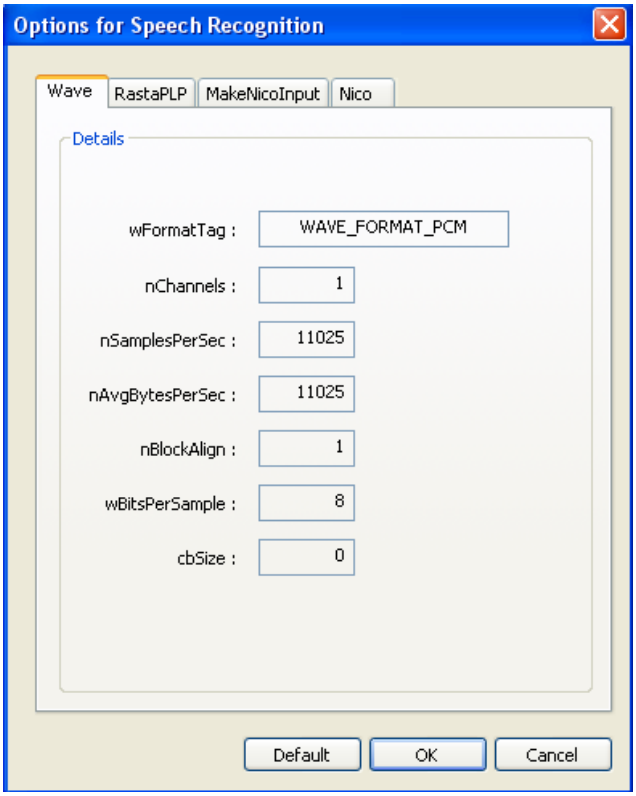


รูปที่ 7.4 หน้าการปรับค่าพารามิเตอร์ Dial List

พารามิเตอร์สำหรับการรู้จำเสียงพูด

พารามิเตอร์สำหรับการรู้จำเสียงพูดสามารถกำหนดได้โดยการเลือกรายการ **Options->Speech Recognition...** หรือกดปุ่ม  บนแถบเครื่องมือ

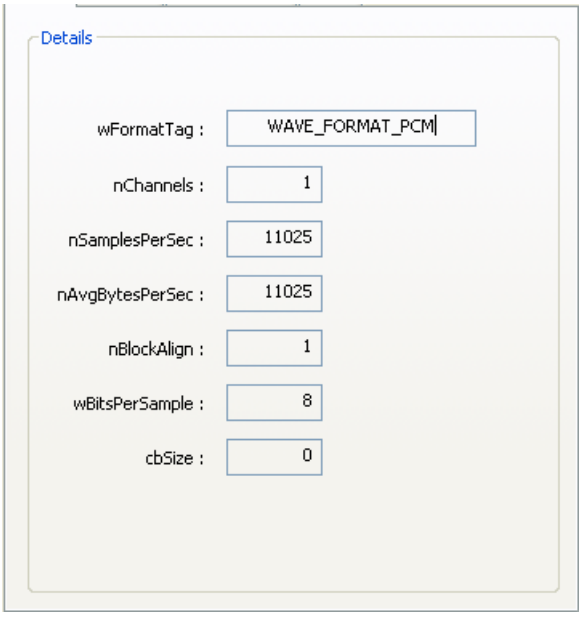
เมื่อกระทำการข้างต้นแล้ว จะปรากฏกล่องโต้ตอบดังรูปที่ 7.5



รูปที่ 7.5 กล่องโต้ตอบการปรับค่าพารามิเตอร์สำหรับการรู้จำเสียงพูด

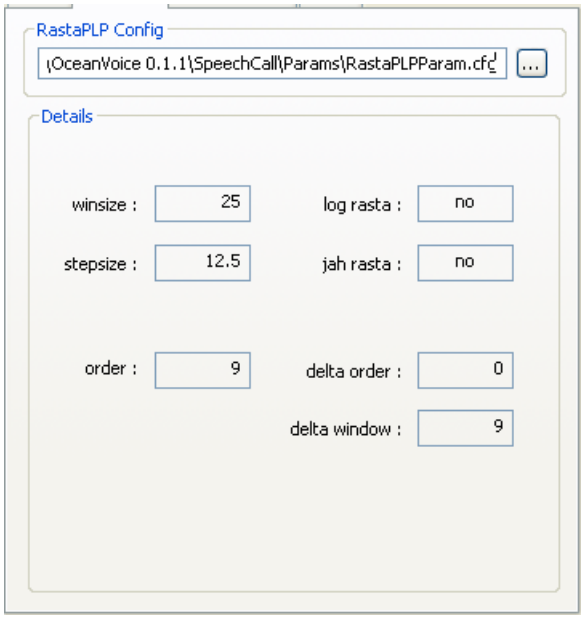
กล่องโต้ตอบสำหรับการปรับค่าพารามิเตอร์ที่ใช้ในการรู้จำเสียงพูดประกอบด้วยหน้าต่างๆ ดังนี้

- **Wave** เป็นพารามิเตอร์ที่บอกลักษณะของไฟล์เสียงที่ใช้ในการรู้จำ โดยจำเป็นต้องสอดคล้องกับลักษณะของไฟล์เสียงที่ใช้ในการเรียนรู้ ซึ่งพารามิเตอร์นี้จะถูกกำหนดตายตัว เนื่องจากข้อจำกัดในการรับเสียงพูดผ่านทางโทรศัพท์ผ่านทางการ์ดเสียง หน้า Wave แสดงได้ดังรูปที่ 7.6



รูปที่ 7.6 หน้าการปรับค่าพารามิเตอร์ Wave

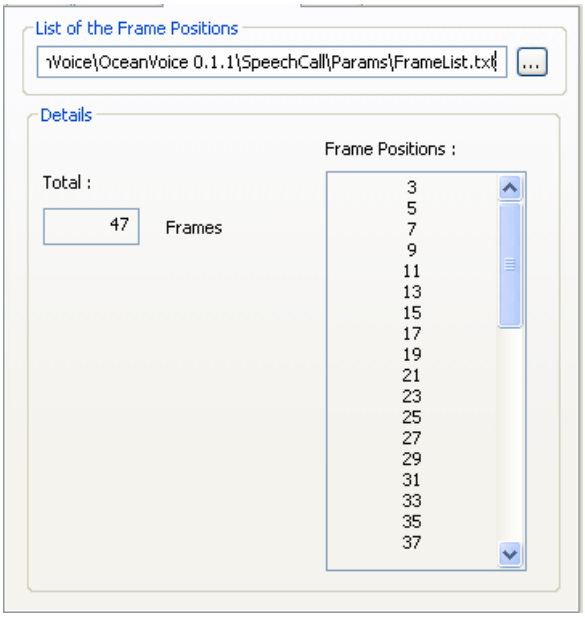
- **RastaPLP** เป็นพารามิเตอร์ที่ใช้ในการหาลักษณะสำคัญของเสียงเพื่อการรู้จำ ซึ่งจำเป็นต้องสอดคล้องกับที่ใช้ในส่วนของการเรียนรู้ หน้า RastaPLP แสดงได้ดังรูปที่ 7.7 โดยผู้ใช้สามารถเลือกไฟล์ที่บอกพารามิเตอร์ที่ใช้ในการหาลักษณะสำคัญของเสียงที่ส่วนบนของหน้า ส่วนล่างของหน้าแสดงถึงรายละเอียดต่างๆ ภายในไฟล์นั้น



รูปที่ 7.7 หน้าการปรับค่าพารามิเตอร์ RastaPLP

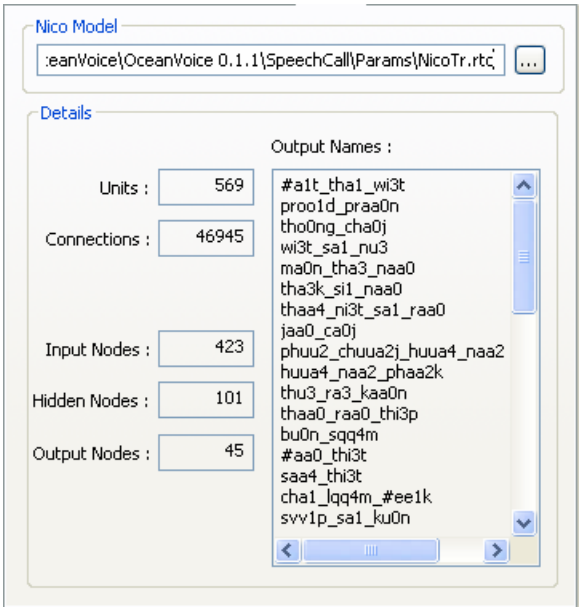
- **MakeNicoInput** เป็นพารามิเตอร์ที่ใช้เพื่อเตรียมลักษณะสำคัญของเสียงให้พร้อมก่อนจะเข้าสู่กระบวนการรู้จำโดยโครงข่ายประสาทเทียม ซึ่งจำเป็นต้องสอดคล้องกับที่ใช้ในส่วนของการเรียนรู้ หน้า MakeNicoInput แสดงได้ดังรูปที่ 7.8 โดยผู้ใช้สามารถเลือกไฟล์ที่

บอกพารามิเตอร์ที่ใช้เพื่อเตรียมลักษณะสำคัญของเสียงที่ส่วนบนของหน้า ส่วนล่างของหน้าแสดงถึงรายละเอียดต่างๆ ภายในไฟล์นั้น



รูปที่ 7.8 หน้าการปรับค่าพารามิเตอร์ MakeNicoInput

- Nico เป็นโครงข่ายประสาทเทียมซึ่งเป็นผลลัพธ์ที่ได้จากการเรียนรู้ หน้า Nico แสดงได้ดังรูปที่ 7.9 โดยผู้ใช้สามารถเลือกไฟล์โครงข่ายประสาทเทียมที่ส่วนบนของหน้า ส่วนล่างของหน้าแสดงถึงรายละเอียดต่างๆ ภายในไฟล์นั้น



รูปที่ 7.9 หน้าการปรับค่าพารามิเตอร์ Nico

7.2.2 การโอนสายโทรศัพท์อัตโนมัติ

ก่อนเริ่มต้นระบบโอนสายโทรศัพท์อัตโนมัติจำเป็นต้องเรียกใช้บริการของการ์ดเสียง และกำหนดค่าพารามิเตอร์สำหรับการโอนสายโทรศัพท์และการรู้จำเสียงพูดเสียก่อน จากนั้นสามารถเริ่มต้นระบบได้โดยเลือกรายการ **Action->Start** หรือกดปุ่ม  บนแถบเครื่องมือ เมื่อต้องการหยุดระบบสามารถทำได้โดยเลือกรายการ **Action->Stop** หรือกดปุ่ม  บนแถบเครื่องมือ

เมื่อเริ่มต้นระบบ ระบบจะรอรับโทรศัพท์ และทำการรู้จำเสียงพูดจากการโทรเข้าโดยอัตโนมัติ โดยกระบวนการทำงานของระบบจะถูกแสดงออกมาที่หน้าจอแสดงผล

8. ภาคผนวก – ความผิดพลาดในการแยกแยะข้อมูลที่ใช้ทดสอบของโครงข่ายประสาทเทียมของระบบการโอนสายอัตโนมัติจากเสียงพูดชื่อไทยทางโทรศัพท์

ความผิดพลาดในการแยกแยะข้อมูลที่ใช้ทดสอบของโครงข่ายประสาทเทียมในการทดลองต่าง ๆ สามารถแสดงผลได้ดังนี้

8.1 การทดลองที่ 3.3.1

เมื่อใช้อันดับการทำนายเชิงเส้นแบบรับรู้เท่ากับ 18 ความผิดพลาดในการแยกแยะข้อมูลที่ใช้ทดสอบของโครงข่ายประสาทเทียมสามารถแสดงได้ดังตารางที่ **8.1**

ตารางที่ 8.1 ความผิดพลาดในการแยกแยะข้อมูลที่ใช้ทดสอบของโครงข่ายประสาทเทียมในการทดลองที่ 3.3.1 เมื่อใช้อันดับการทำนายเชิงเส้นแบบรับรู้เท่ากับ 18

	#a1t_tha1_wi3t	proo1d_praa0n	tho0ng_cha0j	wi3t_sa1_nu3	ma0n_tha3_naa0	tha3k_si1_naa0	thaa4_ni3t_sa1_raa0	jaa0_ca0j	phuu2_chuua2j_huua4_naa2_phaa2k	huua4_naa2_phaa2k	thu3_ra3_kaa0n	thaa0_raa0_thi3p	bu0n_sqq4m	#aa0_thi3t	saa4_thi3t	cha1_lqq4m_#ee1k	svv1p_sa1_ku0n	bu0n_cha0j	no0ng_la3k	thi1t	chuu0_chii2p	mee0_thii0	so4m_chaa0j	wii0_ra3	pra1_phaa2t	see1t_thaa4	wi3_wa3t	ph@@0n_si1_ri1	chee2t	#a1t_tha1_si1t	k@@1p_ku0n	na3_kh@@0n_thi3p	wi3_chaa0n	su1_mee2t	tha3_naa0_wa0n	wa0n_ph@@0n	cha0j_si1_ri1	tha3_wi3t_tii0	cha0j	na3t_tha1_wu3t	caa0_ru3_maa2t	ja0n_jo0ng	wa0n_cha0j	pi3t_sa1_nu3	krqq1k
#a1t_tha1_wi3t	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
proo1d_praa0n	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
tho0ng_cha0j	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
wi3t_sa1_nu3	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
ma0n_tha3_naa0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
tha3k_si1_naa0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
thaa4_ni3t_sa1_raa0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
jaa0_ca0j	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
phuu2_chuua2j_huua4_naa2_phaa2k	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
huua4_naa2_phaa2k	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
thu3_ra3_kaa0n	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
thaa0_raa0_thi3p	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
bu0n_sqq4m	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
#aa0_thi3t	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
saa4_thi3t	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
cha1_lqq4m_#ee1k	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
svv1p_sa1_ku0n	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
bu0n_cha0j	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
no0ng_la3k	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
thi1t	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
chuu0_chii2p	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
mee0_thii0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
so4m_chaa0j	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
wii0_ra3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
pra1_phaa2t	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
see1t_thaa4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
wi3_wa3t	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ph@@0n_si1_ri1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
chee2t	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
#a1t_tha1_si1t	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
k@@1p_ku0n	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0
na3_kh@@0n_thi3p	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0
wi3_chaa0n	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0
su1_mee2t	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0
tha3_naa0_wa0n	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0
wa0n_ph@@0n	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0
cha0j_si1_ri1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0
tha3_wi3t_tii0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0
cha0j	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0
na3t_tha1_wu3t	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0
caa0_ru3_maa2t	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0
ja0n_jo0ng	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0
wa0n_cha0j	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0
pi3t_sa1_nu3	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0
krqq1k	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6

8.2 การทดลองที่ 3.3.2

เมื่อใช้อนุพันธ์การทำนายเชิงเส้นแบบรับรู้อันดับที่หนึ่ง ความผิดพลาดในการแยกแยะข้อมูลที่ใช้ทดสอบของโครงข่ายประสาทเทียมสามารถแสดงได้ดัง

ตารางที่ **8.2**

ตารางที่ 8.2 ความผิดพลาดในการแยกแยะข้อมูลที่ใช้ทดสอบของโครงข่ายประสาทเทียมในการทดลองที่ 3.3.2 เมื่อใช้อนุพันธ์การทำนายเชิงเส้นแบบรับรู้อันดับที่หนึ่ง

	#a1t_tha1_wi3t	proo1d_praa0n	tho0ng_cha0j	wi3t_sa1_nu3	ma0n_tha3_naa0	tha3k_si1_naa0	thaa4_ni3t_sa1_raa0	jaa0_ca0j	phuu2_chuua2j_huua4_naa2_phaa2k	huua4_naa2_phaa2k	thu3_ra3_kaa0n	thaa0_raa0_thi3p	bu0n_sqq4m	#aa0_thi3t	saa4_thi3t	cha1_lqq4m_#ee1k	svv1p_sa1_ku0n	bu0n_cha0j	no0ng_la3k	thi1t	chuu0_chii2p	mee0_thii0	so4m_chaa0j	wii0_ra3	pra1_phaa2t	see1t_thaa4	wi3_wa3t	ph@@0n_si1_ri1	chee2t	#a1t_tha1_si1t	k@@1p_ku0n	na3_kh@@0n_thi3p	wi3_chaa0n	su1_mee2t	tha3_naa0_wa0n	wa0n_ph@@0n	cha0j_si1_ri1	tha3_wi3t_tii0	cha0j	na3t_tha1_wu3t	caa0_ru3_maa2t	ja0n_jo0ng	wa0n_cha0j	pi3t_sa1_nu3	krqq1k
#a1t_tha1_wi3t	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
proo1d_praa0n	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
tho0ng_cha0j	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
wi3t_sa1_nu3	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ma0n_tha3_naa0	0	0	0	0	5	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
tha3k_si1_naa0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
thaa4_ni3t_sa1_raa0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
jaa0_ca0j	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
phuu2_chuua2j_huua4_naa2_phaa2k	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
huua4_naa2_phaa2k	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
thu3_ra3_kaa0n	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
thaa0_raa0_thi3p	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
bu0n_sqq4m	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
#aa0_thi3t	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
saa4_thi3t	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
cha1_lqq4m_#ee1k	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
svv1p_sa1_ku0n	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
bu0n_cha0j	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
no0ng_la3k	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
thi1t	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
chuu0_chii2p	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
mee0_thii0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
so4m_chaa0j	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
wii0_ra3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
pra1_phaa2t	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
see1t_thaa4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
wi3_wa3t	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ph@@0n_si1_ri1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
chee2t	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
#a1t_tha1_si1t	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
k@@1p_ku0n	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
na3_kh@@0n_thi3p	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0
wi3_chaa0n	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0
su1_mee2t	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0
tha3_naa0_wa0n	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0
wa0n_ph@@0n	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0
cha0j_si1_ri1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0
tha3_wi3t_tii0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0
cha0j	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0
na3t_tha1_wu3t	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0
caa0_ru3_maa2t	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0
ja0n_jo0ng	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0
wa0n_cha0j	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0
pi3t_sa1_nu3	0	0	0	2	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0
krqq1k	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6

8.3 การทดลองที่ 3.3.3

เมื่อใช้จำนวนเฟรมการวิเคราะห์เท่ากับ 47 ความผิดพลาดในการแยกแยะข้อมูลที่ใช้ทดสอบของโครงข่ายประสาทเทียมสามารถแสดงได้ดัง

ตารางที่ **8.3**

ตารางที่ 8.3 ความผิดพลาดในการแยกแยะข้อมูลที่ใช้ทดสอบของโครงข่ายประสาทเทียมในการทดลองที่ 3.3.3 เมื่อใช้จำนวนเฟรมการวิเคราะห์เท่ากับ 47

	#a1t_tha1_wi3t	proo1d_praa0n	tho0ng_cha0j	wi3t_sa1_nu3	ma0n_tha3_naa0	tha3k_si1_naa0	thaa4_ni3t_sa1_raa0	jaa0_ca0j	phuu2_chuua2j_huua4_naa2_phaa2k	huua4_naa2_phaa2k	thu3_ra3_kaa0n	thaa0_raa0_thi3p	bu0n_sqq4m	#aa0_thi3t	saa4_thi3t	cha1_lqq4m_#ee1k	svv1p_sa1_ku0n	bu0n_cha0j	no0ng_la3k	thi1t	chuu0_chii2p	mee0_thii0	so4m_chaa0j	wii0_ra3	pra1_phaa2t	see1t_thaa4	wi3_wa3t	ph@@0n_si1_ri1	chee2t	#a1t_tha1_si1t	k@@1p_ku0n	na3_kh@@0n_thi3p	wi3_chaa0n	su1_mee2t	tha3_naa0_wa0n	wa0n_ph@@0n	cha0j_si1_ri1	tha3_wi3t_tii0	cha0j	na3t_tha1_wu3t	caa0_ru3_maa2t	ja0n_jo0ng	wa0n_cha0j	pi3t_sa1_nu3	krqq1k
#a1t_tha1_wi3t	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
proo1d_praa0n	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
tho0ng_cha0j	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
wi3t_sa1_nu3	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ma0n_tha3_naa0	0	0	0	0	5	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
tha3k_si1_naa0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
thaa4_ni3t_sa1_raa0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
jaa0_ca0j	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
phuu2_chuua2j_huua4_naa2_phaa2k	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
huua4_naa2_phaa2k	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
thu3_ra3_kaa0n	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
thaa0_raa0_thi3p	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
bu0n_sqq4m	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
#aa0_thi3t	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
saa4_thi3t	0	0	0	0	0	0	0	0	0	0	0	0	1	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
cha1_lqq4m_#ee1k	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
svv1p_sa1_ku0n	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
bu0n_cha0j	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
no0ng_la3k	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
thi1t	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
chuu0_chii2p	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
mee0_thii0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
so4m_chaa0j	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
wii0_ra3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
pra1_phaa2t	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
see1t_thaa4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
wi3_wa3t	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ph@@0n_si1_ri1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
chee2t	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0
#a1t_tha1_si1t	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0
k@@1p_ku0n	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0
na3_kh@@0n_thi3p	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0
wi3_chaa0n	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0
su1_mee2t	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0
tha3_naa0_wa0n	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0
wa0n_ph@@0n	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0
cha0j_si1_ri1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0
tha3_wi3t_tii0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0
cha0j	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0
na3t_tha1_wu3t	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0
caa0_ru3_maa2t	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0
ja0n_jo0ng	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0
wa0n_cha0j	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6
pi3t_sa1_nu3	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0
krqq1k	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6

8.4 การทดลองที่ 3.3.4

เมื่อใช้จำนวนรอบในการวนปรับน้ำหนักเท่ากับ 10000 ความผิดพลาดในการแยกแยะข้อมูลที่ใช้ทดสอบของโครงข่ายประสาทเทียมสามารถแสดงได้ดังตารางที่ 8.4

ตารางที่ 8.4 ความผิดพลาดในการแยกแยะข้อมูลที่ใช้ทดสอบของโครงข่ายประสาทเทียมในการทดลองที่ 3.3.4 เมื่อใช้จำนวนรอบในการวนปรับน้ำหนักเท่ากับ 10000

	#a1t_tha1_wi3t	proo1d_praa0n	tho0ng_cha0j	wi3t_sa1_nu3	ma0n_tha3_naa0	tha3k_si1_naa0	thaa4_ni3t_sa1_raa0	jaa0_ca0j	phuu2_chuua2j_huua4_naa2_phaa2k	huua4_naa2_phaa2k	thu3_ra3_kaa0n	thaa0_raa0_thi3p	bu0n_sqq4m	#aa0_thi3t	saa4_thi3t	cha1_lqq4m_#ee1k	svv1p_sa1_ku0n	bu0n_cha0j	no0ng_la3k	thi1t	chuu0_chii2p	mee0_thii0	so4m_chaa0j	wii0_ra3	pra1_phaa2t	see1t_thaa4	wi3_wa3t	ph@@0n_si1_ri1	chee2t	#a1t_tha1_si1t	k@@1p_ku0n	na3_kh@@0n_thi3p	wi3_chaa0n	su1_mee2t	tha3_naa0_wa0n	wa0n_ph@@0n	cha0j_si1_ri1	tha3_wi3t_tii0	cha0j	na3t_tha1_wu3t	caa0_ru3_maa2t	ja0n_jo0ng	wa0n_cha0j	pi3t_sa1_nu3	krqq1k
#a1t_tha1_wi3t	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
proo1d_praa0n	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
tho0ng_cha0j	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
wi3t_sa1_nu3	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0
ma0n_tha3_naa0	0	0	0	0	5	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
tha3k_si1_naa0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
thaa4_ni3t_sa1_raa0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
jaa0_ca0j	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
phuu2_chuua2j_huua4_naa2_phaa2k	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
huua4_naa2_phaa2k	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
thu3_ra3_kaa0n	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
thaa0_raa0_thi3p	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
bu0n_sqq4m	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
#aa0_thi3t	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
saa4_thi3t	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
cha1_lqq4m_#ee1k	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
svv1p_sa1_ku0n	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
bu0n_cha0j	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
no0ng_la3k	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
thi1t	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
chuu0_chii2p	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
mee0_thii0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
so4m_chaa0j	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
wii0_ra3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
pra1_phaa2t	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
see1t_thaa4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
wi3_wa3t	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ph@@0n_si1_ri1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
chee2t	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
#a1t_tha1_si1t	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0
k@@1p_ku0n	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0
na3_kh@@0n_thi3p	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0
wi3_chaa0n	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0
su1_mee2t	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0
tha3_naa0_wa0n	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0
wa0n_ph@@0n	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0
cha0j_si1_ri1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0
tha3_wi3t_tii0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0
cha0j	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0
na3t_tha1_wu3t	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0
caa0_ru3_maa2t	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0
ja0n_jo0ng	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0
wa0n_cha0j	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0
pi3t_sa1_nu3	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0
krqq1k	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6

8.5 การทดลองที่ 3.3.5

เมื่อใช้ชุดข้อมูลในตารางที่ 3.6 ความผิดพลาดในการแยกแยะข้อมูลที่ใช้ทดสอบของโครงข่ายประสาทเทียมสามารถแสดงได้ดังตารางที่ 8.5

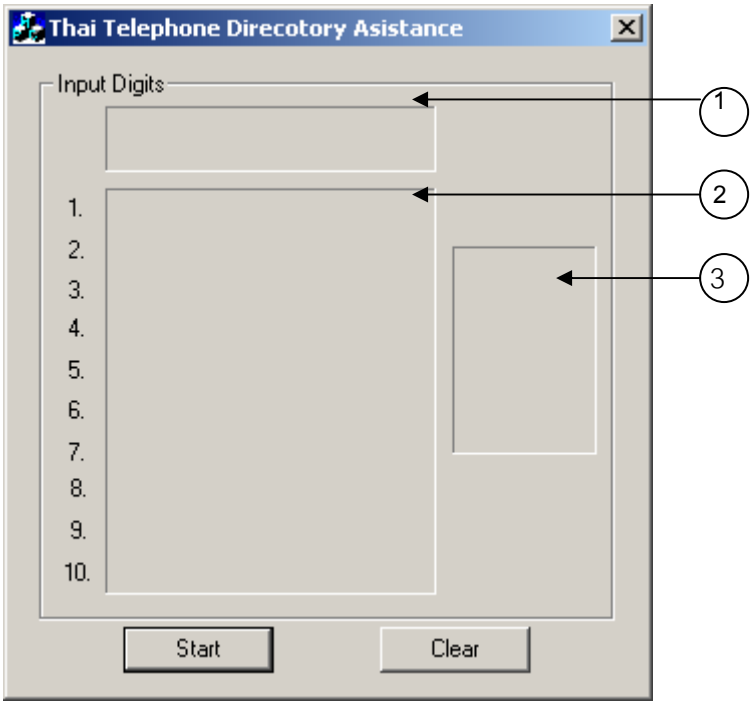
ตารางที่ 8.5 ความผิดพลาดในการแยกแยะข้อมูลที่ใช้ทดสอบของโครงข่ายประสาทเทียมในการทดลองที่ 3.3.5 เมื่อใช้ชุดข้อมูลในตารางที่ 3.6

	#a1t_tha1_wi3t	proo1d_praa0n	tho0ng_cha0j	wi3t_sa1_nu3	mae0_tha3k_si1	thaa_nia_naa	thaa4_ni3t_sa1_raa0	jaa0_ca0j	phuu2_chuua2j_huua4_naa2_phaa2k	huua4_naa2_phaa2k	thu3_ra3_kaa0n	thaa0_raa0_thi3p	bu0n_sqq4m	#aa0_thi3t	saa4_thi3t	cha1_lqq4m_#ee1k	svv1p_sa1_ku0n	bu0n_cha0j	no0ng_la3k	thi1t	chuu0_chii2p	mee0_thii0	so4m_chaa0j	wii0_ra3	pra1_phaa2t	see1t_thaa4	wi3_wa3t	ph@@0n_si1_ri1	chee2t	#a1t_tha1_si1t	k@@1p_ku0n	na3_kh@@0n_thi3p	wi3_chaa0n	su1_mee2t	tha3_naa0_wa0n	wa0n_ph@@0n	cha0j_si1_ri1	tha3_wi3t_tii0	cha0j	na3t_tha1_wu3t	caa0_ru3_maa2t	ja0n_jo0ng	wa0n_cha0j	pi3t_sa1_nu3	krqq1k
#a1t_tha1_wi3t	30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
proo1d_praa0n	0	29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
tho0ng_cha0j	0	0	30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
wi3t_sa1_nu3	0	0	0	25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	4	0	0
ma0n_tha3_naa0	0	0	0	0	27	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
tha3k_si1_naa0	0	0	0	0	0	30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
thaa4_ni3t_sa1_raa0	0	0	0	0	0	0	29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
jaa0_ca0j	0	0	0	0	0	0	0	27	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
phuu2_chuua2j_huua4_naa2_phaa2k	0	0	0	0	0	0	0	0	29	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
huua4_naa2_phaa2k	0	0	0	0	0	0	0	0	1	29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
thu3_ra3_kaa0n	0	0	0	0	0	0	0	0	0	0	29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
thaa0_raa0_thi3p	0	0	0	0	0	0	0	0	0	0	0	29	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
bu0n_sqq4m	0	0	0	0	0	0	0	0	0	0	0	0	30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
#aa0_thi3t	0	0	0	0	0	0	0	0	0	0	0	1	0	26	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
saa4_thi3t	0	0	0	0	0	0	0	0	0	0	0	4	0	5	21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
cha1_lqq4m_#ee1k	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
svv1p_sa1_ku0n	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
bu0n_cha0j	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
no0ng_la3k	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
thi1t	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
chuu0_chii2p	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
mee0_thii0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
so4m_chaa0j	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
wii0_ra3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
pra1_phaa2t	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
see1t_thaa4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
wi3_wa3t	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ph@@0n_si1_ri1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
chee2t	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
#a1t_tha1_si1t	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
k@@1p_ku0n	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
na3_kh@@0n_thi3p	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	30	0	0	0	0	0	0	0	0	0	0	0	0	0	0
wi3_chaa0n	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	30	0	0	0	0	0	0	0	0	0	0	0	0	0
su1_mee2t	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	27	0	0	0	0	0	0	0	0	0	0	0	0
tha3_naa0_wa0n	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	30	0	0	0	0	0	0	0	0	0	0	0
wa0n_ph@@0n	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	30	0	0	0	0	0	0	0	0	0	0
cha0j_si1_ri1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	28	0	1	0	0	0	0	0	0	0
tha3_wi3t_tii0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	30	0	0	0	0	0	0	0	0
cha0j	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	30	0	0	0	0	0	0	0
na3t_tha1_wu3t	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	29	0	0	0	0	0	0
caa0_ru3_maa2t	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	30	0	0	0	0	0
ja0n_jo0ng	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	30	0	0	0	0
wa0n_cha0j	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	30	0	0	0
pi3t_sa1_nu3	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	27	0	0
krqq1k	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	30

9. ภาคผนวก – โปรแกรมระบบสอบถามรายนามผู้ใช้โทรศัพท์โดยผ่านระบบโทรศัพท์

ระบบสอบถามรายนามผู้ใช้โทรศัพท์แบบอัตโนมัติที่นำเสนอในงานวิจัยนี้เป็นต้นแบบ (Prototype) ของระบบ โดยทำการจำลองระบบลงบนเครื่องไมโครคอมพิวเตอร์ รับเสียงข้อมูลขาเข้าโดยผ่านทางไมโครโฟนและส่งเสียงข้อมูลขาออกโดยผ่านทางลำโพง โดยลักษณะการทำงานจริงจะเป็นรูปแบบที่ผ่านทางสายโทรศัพท์ผู้ใช้จะติดต่อผ่านระบบโดยใช้เสียงเท่านั้น แต่เนื่องจากต้องการให้สามารถตรวจสอบความถูกต้องได้ง่ายขึ้นจึงออกแบบให้สามารถแสดงผลออกทางหน้าจอโดยแสดงเป็นหมายเลขโทรศัพท์

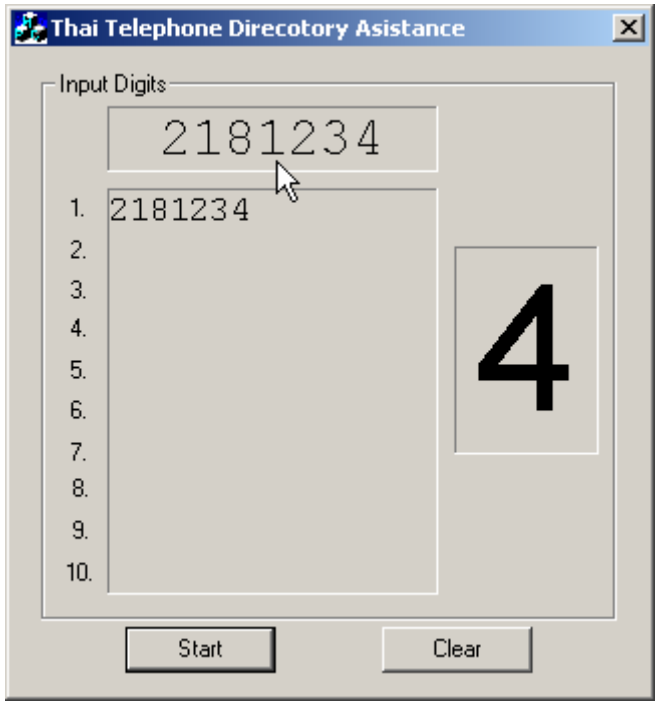
เริ่มต้นโปรแกรมโดยสั่งงานจากโปรแกรมชื่อ TTDA.exe จะปรากฏหน้าต่างโปรแกรมดังรูปที่ 9.1



รูปที่ 9.1 โปรแกรมระบบสอบถามรายนามผู้ใช้โทรศัพท์แบบอัตโนมัติ

จากรูป หมายเลข 1 แสดงผลของชุดหมายเลขที่พูดแต่ละครั้ง หมายเลข 2 แสดงผลของชุดหมายเลขทั้งหมดจำนวน 10 ชุด หมายเลข 3 แสดงส่วนของหมายเลขเดี่ยวของชุดหมายเลขตัวสุดท้าย ปุ่ม Start เป็นปุ่มเริ่มต้นการทำงานของระบบ ปุ่ม Clear เป็นปุ่มล้างชุดหมายเลขทั้งหมดในกรณีที่ชุดหมายเลขมีจำนวนเกิน 10 ชุด

เมื่อเริ่มต้นโปรแกรมโดยกดปุ่ม Start ระบบจะกล่าวต้อนรับและบอกให้พูดหมายเลขโทรศัพท์ที่ต้องการทราบชื่อผู้ใช้โทรศัพท์หลังได้ยินเสียงสัญญาณ เมื่อผู้พูดบอกเลขหมายโทรศัพท์ที่ต้องการทราบ ระบบจะทำการประมวลผลและแจ้งกลับมาในส่วนของตัวเลขแสดงทางจอภาพ และพร้อมกันนั้นระบบจะไปทำการค้นหารายชื่อผู้ใช้โทรศัพท์จากฐานข้อมูล นำกลับมาโดยพูดชื่อและนามสกุลนั้นออกมา ดังแสดงในรูปที่ 9.2



รูปที่ 9.2 โปรแกรมแสดงผลของการรู้จำหมายเลขโทรศัพท์

ในส่วนของฐานข้อมูลผู้ใช้สามารถเข้าไปเรียกดูและสามารถเพิ่มหรือลบรายนามของผู้ใช้โทรศัพท์ได้โดยเรียกจากโปรแกรมชื่อ TTDADB.exe จะแสดงรายการเมนูหลักของคำสั่งดังแสดงในรูปที่ 9-3 ถึง 9-5 ซึ่งประกอบด้วย

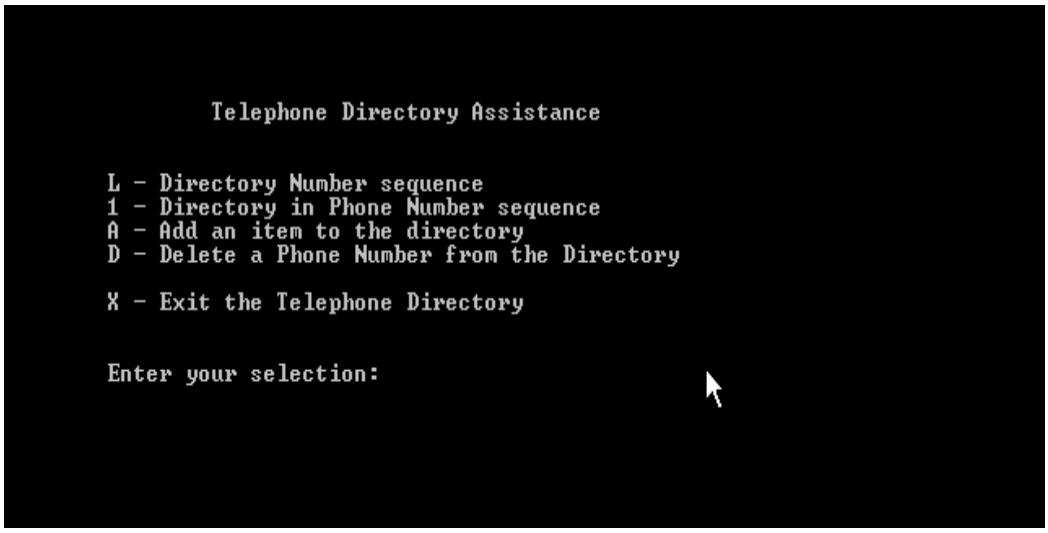
L แสดงรายการของหมายเลขโทรศัพท์และรายนามผู้ใช้โทรศัพท์ตามลำดับการนำข้อมูลเข้า

1 แสดงรายการของหมายเลขโทรศัพท์และรายนามผู้ใช้โทรศัพท์ตามลำดับเลขหมาย

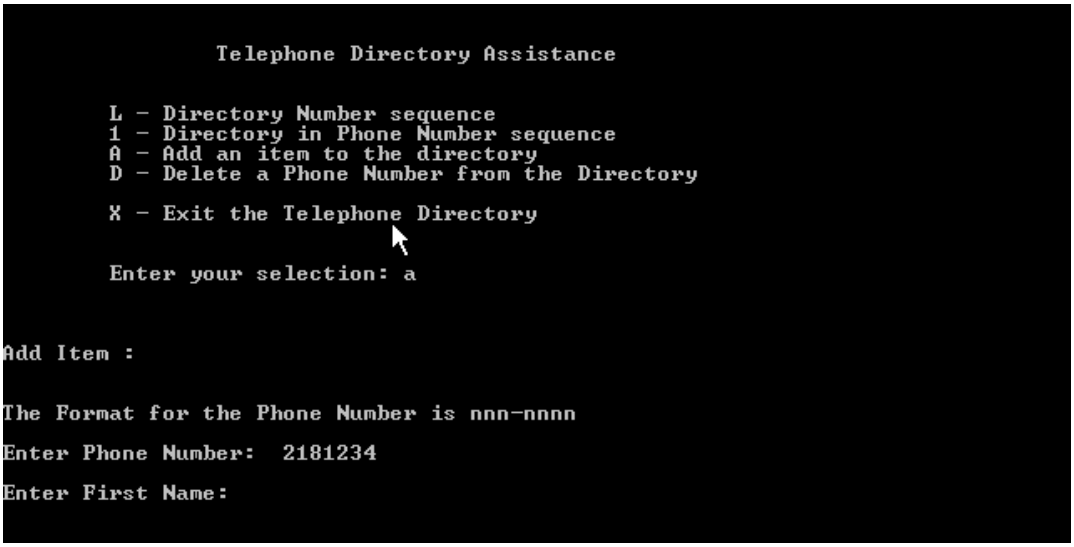
A คำสั่งในการเพิ่มหมายเลขโทรศัพท์ลงในฐานข้อมูล

D คำสั่งในการลบหมายเลขโทรศัพท์ออกจากฐานข้อมูล

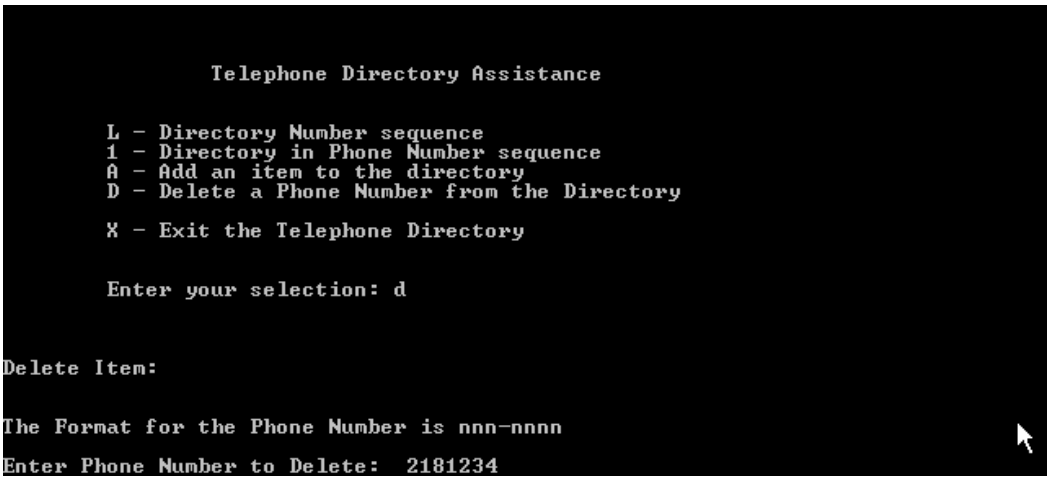
X ออกจากระบบ



รูปที่ 9.3 เมนูแสดงคำสั่งที่ใช้กับฐานข้อมูล

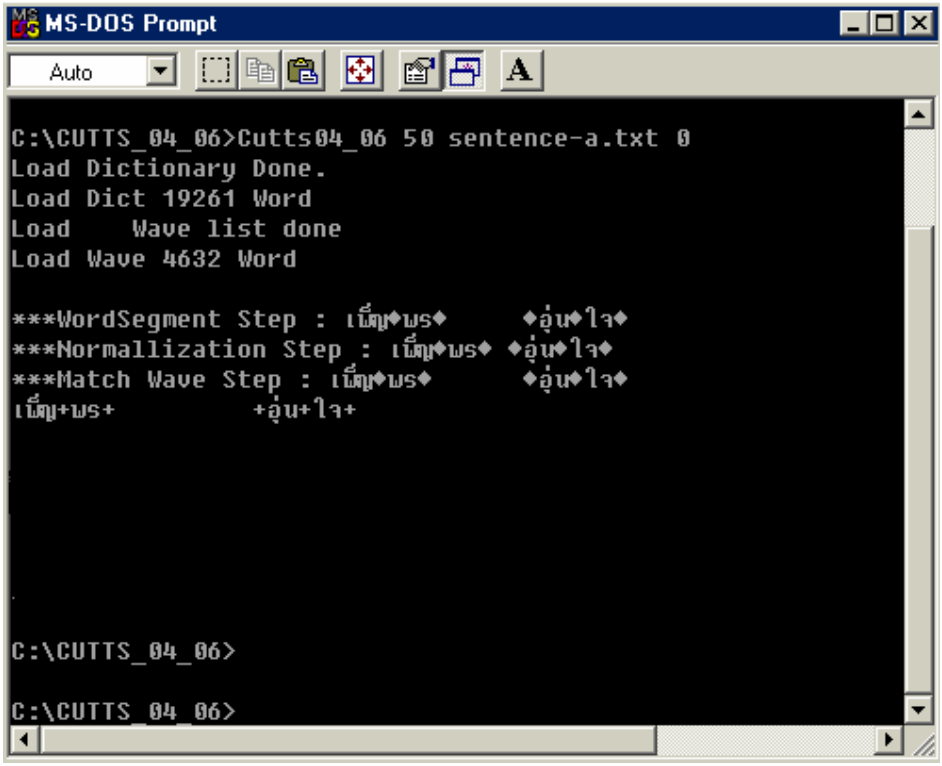


รูปที่ 9.4 การเพิ่มหมายเลขโทรศัพท์ลงในฐานข้อมูล



รูปที่ 9.5 การลบหมายเลขโทรศัพท์ออกจากฐานข้อมูล

ในส่วนของการสังเคราะห์เสียงผู้ดูแลระบบจะต้องนำชื่อและนามสกุลมาผ่านขั้นตอนการสังเคราะห์เสียงก่อนเพื่อตรวจสอบความถูกต้องในการอ่านออกเสียง โดยการฟังจากลำโพง ดังแสดงในรูปที่ ง-6 แต่ถ้าการอ่านออกเสียงนั้นผิดพลาดก็ให้ผู้ดูแลระบบแก้ไขโดยใส่คำอ่านที่ถูกต้อง



รูปที่ 9.6 การตรวจสอบการอ่านออกเสียงชื่อและนามสกุล