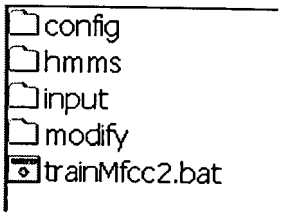
ภาคผนวก

ภาคผนวก ก: การสร้างแบบจำลองเสียง

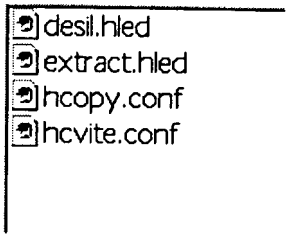
เป็นการใช้งานชุดโปรแกรมรู้จำเสียงพูด HTK บนเครื่องคอมพิวเตอร์ที่ใช้ระบบปฏิบัติการวินโดวส์ โดยใช้วิธีการเขียน Batch File เพื่อควบคุมการทำงานของชุดโปรแกรมทั้งหมด วิธีการทดลองนี้จะกล่าวถึงวิธีการมาตรฐานที่ใช้ในทุกการทดลองคือ หน่วยเสียงในภาษาไทย การสร้างแบบจำลองเสียง การทดสอบแบบจำลองเสียง วิธีการกำหนดค่าพารามิเตอร์สำหรับการทดลองในแต่ละการทดลอง

ได้กำหนดโครงสร้างการจัดเก็บข้อมูลดังภาพประกอบ ก1 ภายในแต่ละไดเรคทอรีจะมีการเก็บข้อมูลที่แตกต่างกันซึ่งจะอธิบายเป็นหัวข้อดังต่อไปนี้



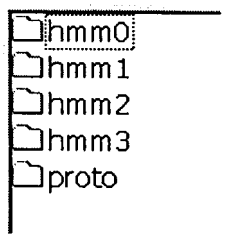
ภาพประกอบ ก1 โครงสร้างข้อมูลในการสร้างแบบจำลองเสียง

- Config เก็บค่าพารามิเตอร์ของระบบรู้จำเสียงพูดทั่วไปเช่น การกำหนดวิธีการคำนวณค่าคุณลักษณะเด่น ว่าเป็น MFCC,LPC, LPCC หรืออื่นๆ และ การกำหนดคุณลักษณะของข้อมูลเสียงที่ใช้ว่าเป็นมาตรฐานใด ค่าความถี่ในการสุ่มสัญญาณ ความละเอียดของสัญญาณ ดังภาพประกอบ ก2



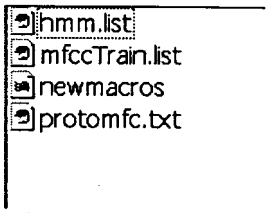
ภาพประกอบ ก2 ไฟล์ข้อมูลคอนฟิกของการสร้างแบบจำลองเสียง

- Hmms เก็บแบบจำลองเสียงในแต่ละขั้นตอนของการสร้างแบบจำลองเสียง ดังภาพประกอบ ก3



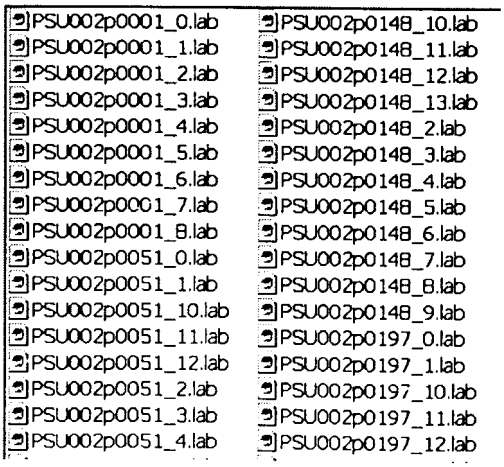
ภาพประกอบ ก3 ที่เก็บแบบจำลองเสียงในแต่ละขั้นตอน

- Input เก็บค่าตัวแปรต้นสำหรับการสร้างแบบจำลองเสียง ตั้งแต่รายชื่อข้อมูลเสียงฝึกฝน ต้นแบบสำหรับสร้างแบบจำลองเสียง ชุดหน่วยเสียงที่ใช้ เป็นส่วนสำคัญที่ผู้ใช้ต้องปรับแก้ให้ถูกต้องก่อนการสร้างแบบจำลองเสียงในแต่ละครั้ง ดังภาพประกอบ ก4



ภาพประกอบ ก4 ไฟล์ที่ผู้ใช้ต้องกำหนดเพื่อสร้างแบบจำลองเสียง

- Modify เก็บข้อมูลกำกับเสียง (Transcription File) ของเสียงในแต่ละไฟล์เพื่อใช้ในการแบ่งช่วงของเสียงในไฟล์เพื่อจดจำลักษณะของช่วงเสียงกับหน่วยเสียงภายในข้อมูลกำกับเสียงได้อย่างถูกต้อง ดังภาพประกอบ ก5



ภาพประกอบ ก5 ไฟล์ข้อมูลกำกับเสียงในการสร้างแบบจำลองเสียง

- TrainMfcc2.bat เป็น Batch File ที่ใช้สำหรับการสร้างแบบจำลองเสียงภายในประกอบด้วยคำสั่งต่อไปนี้

ขั้นตอนการสร้างแม่แบบของแบบจำลองเสียง

```
perl MakeProto.pl input\protomfc.txt
xcopy /q hmms\proto\* hmms\hmm0
xcopy /q hmms\proto\* hmms\hmm1
```

ขั้นตอนการปรับปรุงแบบจำลองเสียงตามข้อมูลเสียงฝึกฝน โดยที่ข้อมูลเสียงถูกแบ่งโดยช่วงเวลาที่ระบุในข้อมูลกำกับเสียง โดยใช้วิธีการคำนวณแบบ Hinit

```
\htklab\bin.win32\hhinit input\hmm.list hmms\proto -L modify -S
input\mfccTrain.list -M hmms\hmm0
```

ขั้นตอนการปรับปรุงแบบจำลองเสียงตามข้อมูลเสียงฝึกฝน โดยที่ข้อมูลเสียงถูกแบ่งโดยช่วงเวลาที่ระบุในข้อมูลกำกับเสียง โดยใช้วิธีการคำนวณแบบ Hrest

```
\htklab\bin.win32\hhrest input\hmm.list hmms\hmm0 -L modify -S
input\mfccTrain.list -M hmms\hmm1
```

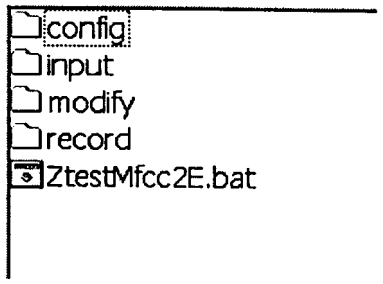
ขั้นตอนการปรับปรุงแบบจำลองเสียงตามข้อมูลเสียงฝึกฝน โดยที่ข้อมูลเสียงถูกแบ่งโดยวิธีการคำนวณหาการเปลี่ยนแปลงสถานะของเสียง โดยใช้วิธีการคำนวณแบบ Hrest จำนวน 2 ครั้ง

```
\htklab\bin.win32\herest -t 250 150 1000 -S input\mfccTrain.list -L modify -M
hmms\hmm2 -d hmms\hmm1 input\hmm.list
\htklab\bin.win32\herest -t 250 150 1000 -S input\mfccTrain.list -L modify -M
hmms\hmm3 -H hmms\hmm2\newmacros input\hmm.list
copy /y hmms\hmm3\* input
```

สุดท้ายจะได้แบบจำลองเสียงอยู่ในไฟล์ newmacros

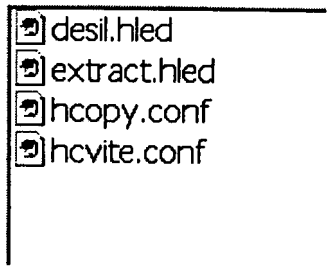
ภาคผนวก ข: การทดสอบแบบจำลองเสียง

ได้กำหนดโครงสร้างการจัดเก็บข้อมูลดังภาพประกอบ ข1 ภายในแต่ละไดเรคทอรีจะมีการเก็บข้อมูลที่แตกต่างกันซึ่งจะอธิบายเป็นหัวข้อดังต่อไปนี้



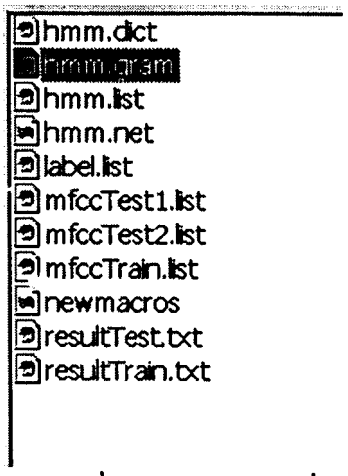
ภาพประกอบ ข1 โครงสร้างหลักของระบบการทดสอบการรู้จำ

- Config เก็บค่าพารามิเตอร์ของระบบรู้จำเสียงพูดทั่วไปเช่น การกำหนดวิธีการคำนวณค่าคุณลักษณะเด่น MFCC,LPC, LPCC หรืออื่นๆ และ การกำหนดคุณลักษณะของข้อมูลเสียงที่ใช้ว่าเป็นมาตรฐานใด ค่าความถี่ในการสุ่มสัญญาณ ความละเอียดของสัญญาณ ดังภาพประกอบ ข2



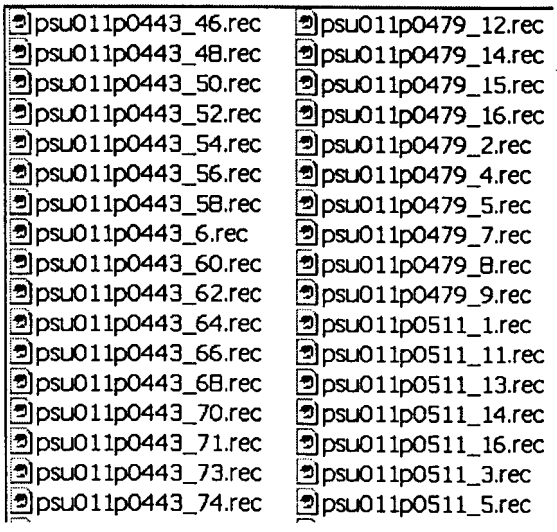
ภาพประกอบ ข2 ไฟล์ข้อมูลคอนฟิกของการทดสอบการรู้จำ

- Input เก็บค่าตัวแปรต้นสำหรับการรู้จำเสียงพูด ตั้งแต่รายชื่อข้อมูลเสียงทดสอบ แบบจำลองเสียง ชุดหน่วยเสียงที่ใช้ พจนานุกรม ไฟลไวยากรณ์ ถือว่าเป็นส่วนสำคัญที่ผู้ใช้ต้องปรับแก้ให้ถูกต้องก่อนการรู้จำเสียงพูดในแต่ละครั้ง ดังภาพประกอบ ข3



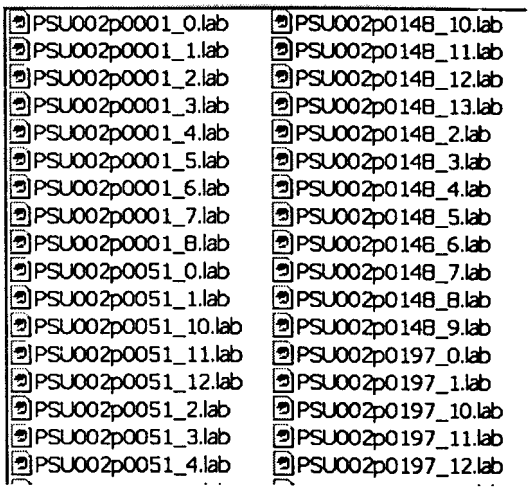
ภาพประกอบ ข3 ไฟล์ที่ผู้ใช้ต้องกำหนดเพื่อทดสอบการรู้จำ

- Record เก็บผลการรู้จำมีลักษณะเช่นเดียวกับข้อมูลกำกับเสียง ในการรู้จำเสียงพูดระบบรู้จำจะบันทึกผลการรู้จำเสียงพูดแต่ละไฟล์ลงในไดเรคทอรีนี้ในชื่อเดียวกับไฟล์เสียงที่ทำการรู้จำ มีนามสกุลเป็น .rec ดังภาพประกอบ ข4



ภาพประกอบ ข4 ผลการรู้จำจากการทดสอบการรู้จำ

- Modify เก็บข้อมูลกำกับเสียง (Transcription File) ของเสียงในแต่ละไฟล์เพื่อใช้ในการแบ่งช่วงของเสียงในไฟล์เพื่อจดจำลักษณะของช่วงเสียงกับหน่วยเสียงภายในข้อมูลกำกับเสียงได้อย่างถูกต้อง ดังภาพประกอบ ข5



ภาพประกอบ ข5 ไฟล์ข้อมูลกำกับเสียงในการทดสอบการรู้จำ

- TrainMfcc2.bat เป็น Batch File ที่ใช้สำหรับการสร้างแบบจำลองเสียงภายในประกอบด้วยคำสั่งต่อไปนี้

เริ่มจากการรู้จำเสียงพูด แล้วบันทึกผลลงใน Record ไดเรคทอรี

```
\htklab\bin.win32\hparse input\hmm.gram input\hmm.net
\htklab\bin.win32\hvite -l record -w input\hmm.net -H input\newmacros
input\hmm.dict input\hmm.list -S input\mfccTest1.list
```

ปรับปรุงผลการรู้จำก่อนการเปรียบเทียบโดยทำการลบหน่วยเสียงที่ไม่มีเสียง (เสียงเงียบ) ออกจากผลการรู้จำและข้อมูลกำกับเสียง

```
move record\*.rec record\source
xcopy /q /y record\source\*.rec record\modify
xcopy /q /y phone\*.lab record\modify
echo on
dir /s /b record\modify\*.rec > record\record.list
dir /s /b record\modify\*.lab > record\label.list
\htklab\bin.win32\hled -X rec config\desil.hled -S record\record.list
\htklab\bin.win32\hled config\desil.hled -S record\label.list
```

เปรียบเทียบผลการรู้จำกับข้อมูลกำกับเสียงแล้วบันทึกค่าความแม่นยำลงไฟล์

```
\htklab\bin.win32\hresults -t -f record\label.list -S record\record.list >
input\resultTest.txt
```

ภาคผนวก ค: ตัวอย่างข้อมูลจากชุดหน่วยเสียง $C_i+V_t+C_f$

การทดสอบแบบจำลอง $C_i+V_t+C_f$ โดยใช้เทคนิคการจำแนกหน่วยเสียงได้มีการกำหนดค่าภายในต่างๆ ดังภาพประกอบ ค1 – ค8 การแสดงผลอาจไม่ครอบคลุมทั้งหมดเนื่องจากข้อมูลบางอย่างมีจำนวนมากไม่สามารถนำมาแสดงให้หมดได้จึงขอแสดงเพียงบางส่วน

```
8  hsKind: P
9  covKind: D
0  nStates: 3
1  nStreams: 1
2  sWidths: 36
3  mixes: 8
4  parmKind: MFCC_D_A
5  vecSize: 36
6  outDir: hmms\proto
7  hmmList: input\hmm.list
```

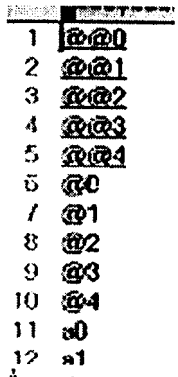
ภาพประกอบ ค1 ข้อมูลภายในไฟล์โครงสร้าง HMM (protomfc.txt)

```
1  J:\HTKlab\resource\mfcc\mfccda\psu002p0001.mfcc
2  J:\HTKlab\resource\mfcc\mfccda\psu002p0051.mfcc
3  J:\HTKlab\resource\mfcc\mfccda\psu002p0099.mfcc
4  J:\HTKlab\resource\mfcc\mfccda\psu002p0148.mfcc
5  J:\HTKlab\resource\mfcc\mfccda\psu002p0197.mfcc
6  J:\HTKlab\resource\mfcc\mfccda\psu002p0246.mfcc
7  J:\HTKlab\resource\mfcc\mfccda\psu002p0295.mfcc
8  J:\HTKlab\resource\mfcc\mfccda\psu002p0341.mfcc
9  J:\HTKlab\resource\mfcc\mfccda\psu002p0390.mfcc
10 J:\HTKlab\resource\mfcc\mfccda\psu002p0439.mfcc
11 J:\HTKlab\resource\mfcc\mfccda\psu002p0476.mfcc
```

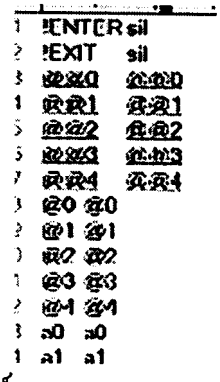
ภาพประกอบ ค2 ข้อมูลภายในไฟล์ที่เก็บข้อมูลฝึกฝน (mfccTrain.list)

```
1  ~o
2  <STREAMINFO> 1 36
3  <VECSIZE> 36<NULLD><MFCC_D_A><DIAGC>
4  ~h "xx"
5  <BEGINHMM>
6  <NUMSTATES> 5
7  <STATE> 2
8  <NUMMIXES> 8
9  <MIXTURE> 1 9.038652e-002
10 <MEAN> 36
11 -3.075530e+000 -1.026187e-001 1.174172e-000 -9.964042e-000 -1.086123e
12 <VARIANCE> 36
13 5.286293e+000 1.584463e+001 1.276852e+001 1.383253e+001 2.843183e+0
```

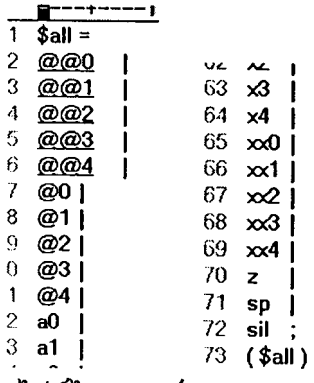
ภาพประกอบ ค3 ข้อมูลภายในไฟล์แบบจำลองเสียง (newmacros)



ภาพประกอบ ค4 ข้อมูลภายในไฟล์รายชื่อหน่วยเสียงของชุดหน่วยเสียง Ci+Vt+Cf (hmm.list)



ภาพประกอบ ค5 ข้อมูลภายในไฟล์พจนานุกรมของชุดหน่วยเสียง Ci+Vt+Cf (hmm dict)



ภาพประกอบ ค6 ข้อมูลภายในไฟล์ไวยากรณ์ของชุดหน่วยเสียง Ci+Vt+Cf (hmm.gram)

```
9507  REC: ng_
9508  Aligned transcription: modify/PSU055p0787_39.lab vs J:/HTKlab/process/Testl
9509  LAB: a3
9510  REC: qq0
9511  ====================== HTK Results Analysis ==================
9512   Date: Tue Jun 15 20:11:16 2004
9513   Ref : modify
9514   Rec : J:\HTKlab\process\TestRecog\record\PSU085p0043_2.rec
9515      : J:\HTKlab\process\TestRecog\record\PSU085p0043_5.rec
9516      : J:\HTKlab\process\TestRecog\record\PSU085p0043_6.rec
9517      : J:\HTKlab\process\TestRecog\record\PSU085p0043_7.rec
9518      : J:\HTKlab\process\TestRecog\record\PSU085p0043_9.rec
9519  ------------------------ Overall Results --------------------------
9520  SENT: %Correct=58.19 [H=4412, S=3170, N=7582]
9521  WORD: %Corr=58.19, Acc=58.19 [H=4412, D=0, S=3170, I=0, N=7582]
9522  ============================================================
```

ภาพประกอบ ค7 ข้อมูลภายในไฟล์ผลการรู้จำแบบที่ 1 (resultTest1.list)

```
6901  ====================== HTK Results Analysis ==================
6902   Date: Tue Jun 15 20:32:11 2004
6903   Ref : modify
6904   Rec : J:\HTKlab\process\TestRecog\record\psu010p0005_2.rec
6905      : J:\HTKlab\process\TestRecog\record\psu010p0005_3.rec
6906      : J:\HTKlab\process\TestRecog\record\psu010p0005_5.rec
6907      : J:\HTKlab\process\TestRecog\record\psu010p0005_6.rec
6908      : J:\HTKlab\process\TestRecog\record\psu010p0005_7.rec
6909  ------------------------ Overall Results --------------------------
6910  SENT: %Correct=73.86 [H=6500, S=2300, N=8800]
6911  WORD: %Corr=73.86, Acc=73.86 [H=6500, D=0, S=2300, I=0, N=8800]
6912  ============================================================
```

ภาพประกอบ ค8 ข้อมูลภายในไฟล์ผลการรู้จำแบบที่ 2 (resultTest2.list)

ภาคผนวก ง: เอกสารที่ตีพิมพ์

เอกสารที่ตีพิมพ์ระหว่างการทำวิทยานิพนธ์มีเอกสาร 1 ฉบับคือ

การจำแนกหน่วยเสียงสำหรับการรู้จำเสียงพูดภาษาไทย ตีพิมพ์ในงานประชุมวิชาการ NCSEC 2004 (The 8th National Computer Science and Engineering Conference) ที่อำเภอหาดใหญ่ จังหวัดสงขลา จัดโดยภาควิชาวิศวกรรมคอมพิวเตอร์ มหาวิทยาลัยสงขลานครินทร์ ระหว่างวันที่ 20-22 ตุลาคม 2547

การจำแนกหน่วยเสียงสำหรับการรู้จำเสียงพูดภาษาไทย
Phonetic Classification for Thai Speech Recognition

พงศ์ธร ตันธนกิจ, มนตรี กาญจนะเดชะ
ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยสงขลานครินทร์
อำเภอหาดใหญ่, จังหวัดสงขลา 90110, ประเทศไทย
e-mail: pongtorn@hospital-os.com

บทคัดย่อ

บทความนี้กล่าวการจำแนกหน่วยเสียงภาษาไทยโดยใช้แบบจำลองฮิดเดนมาร์คอฟเพื่อหาค่าความแม่นยำของแบบจำลองเสียงที่สร้างขึ้น ซึ่งจะนำไปวิเคราะห์ และนำไปเป็นค่าอ้างอิงสำหรับการพัฒนาระบบรู้จำเสียงพูดภาษาไทยแบบคำต่อเนื่อง ฐานข้อมูลที่ใช้ประกอบด้วยประโยค 580 ประโยคจากผู้พูด 29 คน ผลการจำแนกหน่วยเสียงทั้งหมด 170 หน่วยเสียงมีค่าความแม่นยำ 56.1% และเมื่อพิจารณาค่าความแม่นยำตามกลุ่มพยัญชนะ กลุ่มสระ กลุ่มตัวสะกด พบว่ามีค่า 69.1% 34.1% และ 72.9% ตามลำดับ จากผลการทดลองแสดงให้เห็นว่าแบบจำลองเสียงที่สร้างขึ้นไม่สามารถจำแนกหน่วยเสียงกลุ่มสระได้ดีเนื่องจากความผิดพลาดในการจำแนกวรรณยุกต์ ซึ่งจำเป็นจะต้องหาวิธีการที่เหมาะสมต่อไป

Abstract

This paper describes the development of a Thai phonetic classification system using Hidden Makov Model. The objective is to determine best Thai speech modeling technique, which could be used for speech recognition. The speech database used in our experiments consisted of 580 sentences, spoken by 29 speakers. The classifier was designed to classify 170 Thai phones and a classification accuracy of 56.1% was obtained. When the phones were divided into 3 groups according to theirs function, the accuracy were 69% for first consonants, 34% for vowels and 72% for final consonants. There was high confusion among tones. We suggest Thai modeling technique for vowels needs to be further studied.

1. บทนำ

ระบบรู้จำเสียงพูดทั่วไปจะต้องประกอบด้วยองค์ประกอบที่สำคัญได้แก่ ชุดหน่วยเสียงที่ครอบคลุมหน่วยเสียงทั้งหมด ข้อมูลเสียงที่มีการออกเสียงชัดเจน การสร้างข้อมูลกำกับเสียงที่ระบุตำแหน่งของการออกเสียงถูกต้อง สำหรับระบบรู้จำเสียงพูดภาษาไทยสิ่งหล่านี้ยังไม่มีการวางรากฐานที่ชัดเจน หมายถึงการกำหนดมาตรฐานกลางขององค์ประกอบดังกล่าวยังไม่แน่นอน ดังนั้นการพัฒนาจึงเป็นไปในทิศทางเดียวกัน เพราะผู้พัฒนาแต่ละกลุ่มก็มีการกำหนดองค์ประกอบตามมาตรฐานของตนและพัฒนาไปในทิศทางที่แตกต่างกัน มาตรฐานกลางที่กำหนดขึ้นจะทำให้ได้ค่าความแม่นยำพื้นฐานสำหรับเป็นค่าอ้างอิงเพื่อใช้ในการพัฒนาประสิทธิภาพของการรู้จำให้ดีขึ้น

งานวิจัยที่เกี่ยวกับการรู้จำเสียงพูดส่วนใหญ่จะเน้นการศึกษาทางด้านเทคนิคและวิธีการใหม่ๆ ในการรู้จำเสียงพูดให้แม่นยำยิ่งขึ้น [1][2][3][4] ในทางกลับกันยังมีงานวิจัยอีกกลุ่มหนึ่งที่ปรับปรุงระบบรู้จำเดิมให้มีประสิทธิภาพสูงสุดโดยการ

ปรับปรุงองค์ประกอบต่างๆ ของระบบรู้จำ [5][6][7] ได้แก่การปรับปรุงชุดหน่วยเสียงให้เหมาะสมกับระบบรู้จำภาษาไทย การปรับปรุงข้อมูลเสียงสำหรับการฝึกฝนระบบรู้จำ

ชุดหน่วยเสียงถือเป็นปัจจัยเริ่มต้นสำหรับการพัฒนาระบบรู้จำเสียงพูดภาษาไทย การกำหนดมาตรฐานกลางของชุดหน่วยเสียงจึงเป็นสิ่งที่จำเป็น งานวิจัยชิ้นนี้ได้เลือกชุดหน่วยเสียงภาษาไทยจากงานวิจัยที่เกี่ยวข้อง[5] เพื่อนำมาทดสอบประสิทธิภาพในการจำแนกหน่วยเสียงภาษาไทยตามกลุ่มของหน่วยเสียง ได้แก่ กลุ่มหน่วยเสียงพยัญชนะ สระ ตัวสะกด และค่าความแม่นยำรวมทั้งหมด นอกจากนี้ยังเปรียบเทียบการรู้จำเสียงฝึกฝนและเทียบกับเสียงทดสอบ ซึ่งผลการทดลองที่ได้สามารถนำมาใช้ในการกำหนดชุดหน่วยเสียงที่เหมาะสมกับระบบรู้จำเสียงพูดภาษาไทยได้ นอกจากนี้ยังชี้ให้เห็นถึงปัญหาและสาเหตุที่เกิดขึ้น สำหรับนำไปแก้ไขต่อไปในอนาคต

ภายในเอกสารชิ้นนี้จะกล่าวถึง 5 หัวข้อหลักได้แก่ ชุดหน่วยเสียงภาษาไทย ฐานข้อมูลเสียง วิธีการจำแนกหน่วยเสียง ผลการทดลอง การวิเคราะห์และสรุปปัญหา สรุปผล

2. หน่วยเสียงภาษาไทย (Thai Phonemes)

ได้นำชุดหน่วยเสียงจากงานวิจัยหน่วยเสียงภาษาไทย [8] แบ่งหน่วยเสียงออกเป็นกลุ่มๆ ดังนี้ กลุ่มพยัญชนะ กลุ่มสระเดี่ยว กลุ่มตัวสะกด ภายในงานวิจัยนี้ได้เพิ่มหน่วยเสียงสระในระดับวรรณยุกต์ต่างๆ อีก 4 ระดับรวมหน่วยเสียงทั้งสิ้น 170 หน่วยเสียง ดังแสดงในตารางที่ 1

ตารางที่ 1 ชุดหน่วยเสียงภาษาไทย 170 หน่วยเสียง

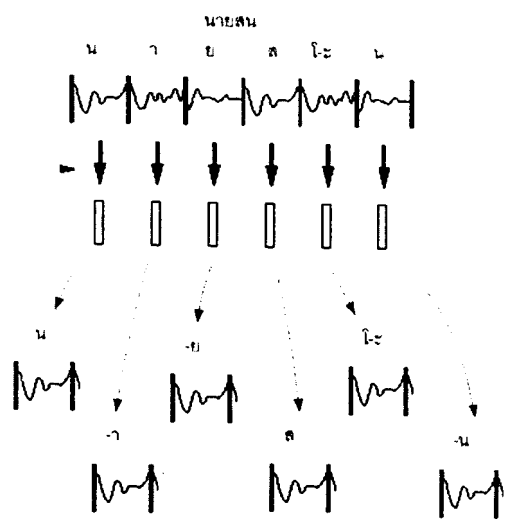
พยัญชนะ	ก ค ง จ ช ด ต ท น บ ป พ ฟ ม ย ร ล ว ส อ ฮ กร กล คร คล ดว ตร ทร ปร ปล พร พลบร บล ฟร ฟล คร
สระและวรรณยุกต์	-ะ -า อิ อี อึ อือ อุ อู เ-ะ เ- แ-ะ แ- โ-ะ โ- เ-าะ –อ เ-อะ เ-อ เอียะ เอีย เอือะ เอือ อัวะ อัว.....เอื่อ อั่ว
ตัวสะกด	-บ -ด -ก -น -ม -ง -ย -ว -ฟ -ล -ส –ช

3. ฐานข้อมูลเสียง (Speech Database)

ข้อมูลเสียงทั้งหมด 580 ประโยคโดยผู้พูด 29 คน พูดคนละ 20 ประโยค เป็นเสียงของผู้พูดชาย 18 คน หญิง 11 คน แบ่งเป็นเสียงสำหรับสอน 480 ประโยค โดยผู้พูดชาย 15 คน หญิง 9 คน เสียงสำหรับทดสอบ 100 ประโยค โดยผู้พูดชาย 3 คน หญิง 2 คน เป็นชุดประโยคจากงานวิจัย [8] บันทึกเสียงในห้องที่ไม่มีสัญญาณรบกวน ความถี่ในการสุ่มสัญญาณ (Sampling Rate) 16 กิโลเฮิร์ต ความละเอียดของข้อมูล (Resolution/Bit Rate) 16 บิต เป็นข้อมูลที่มีการกำกับหน่วยเสียงแล้วโดยผู้กำกับเสียง และแบ่งข้อมูลเสียงออกเป็น 2 ส่วน คือ เสียงฝึกฝนสำหรับการสอนระบบ และเสียงทดสอบสำหรับการจำแนกหน่วยเสียง

4. การจำแนกหน่วยเสียง

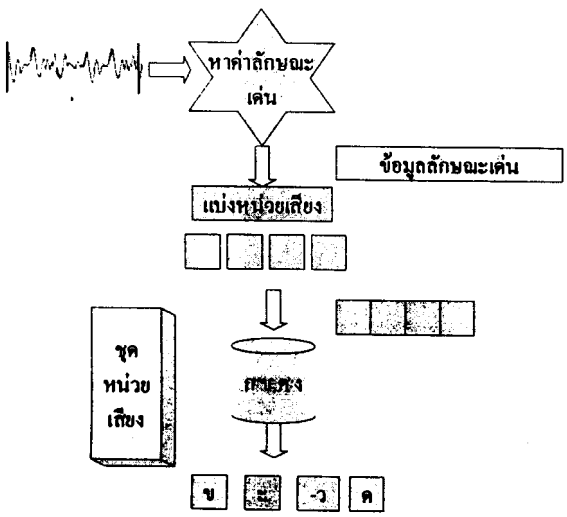
การรู้จำเสียงพูดแบบคำต่อเนื่องประกอบด้วยขั้นตอนหลัก 3 ขั้นตอนคือ การหาค่าลักษณะเด่น (Feature Extraction) การแบ่งแยกหน่วยเสียง (Segmentation) และการจำแนกหน่วยเสียง (Classification) ขั้นตอนการจำแนกหน่วยเสียงถือว่าเป็นขั้นตอนที่สำคัญสำหรับการรู้จำ



รูปที่ 1 การแบ่งข้อมูลเสียงออกเป็นหน่วยเสียงย่อย

ในขั้นตอนของการจำแนกหน่วยเสียง จะประกอบด้วยขั้นตอนของการแบ่งข้อมูลเสียงทั้งประโยคเป็นช่วงหน่วยเสียงย่อยๆ ตามข้อมูลกำกับเสียงดังรูปที่ 1 แล้วนำหน่วยเสียงเหล่านั้นมาเปรียบเทียบกับข้อมูลในแบบจำลองเสียงจนได้ผลการจำแนกหน่วยเสียงออกมา จากนั้นจึงนำผลการจำแนกไปเปรียบเทียบกับข้อมูลกำกับเสียง (Transcription Label) ได้เป็นค่าความแม่นยำในการจำแนกหน่วยเสียงออกมา ค่านี้จะบ่งบอกถึงประสิทธิภาพในการจำแนกหน่วยเสียงของแบบจำลองเสียงนั้นๆ

งานวิจัยนี้ได้ใช้ชุดโปรแกรมรู้จำเสียงพูด HTK ในการพัฒนาควบคู่กับวิธีการหาค่าลักษณะเด่นแบบค่าสัมประสิทธิ์เคปสตรัมความถี่เมล[9] (Mel Scale Frequency Cepstral Coeficient) ที่มีค่าอนุพันธ์อันดับที่หนึ่ง (Delta Term) และค่าอนุพันธ์อันดับที่สอง (Acceleration Term) หรือ MFCC+DA โดยกำหนดระยะระหว่างเฟรมเป็น 10 มิลลิวินาที ขนาดของวินโดวส์เป็น 25 มิลลิวินาที แบบจำลองเสียงแบบฮิดเดนมาร์คอฟ 3 สถานะ 8 ส่วนประกอบ (3 State 8 Mixture) ลักษณะความปรวนแปรแบบเมตริกทะแยง (Diagonal Covariance)

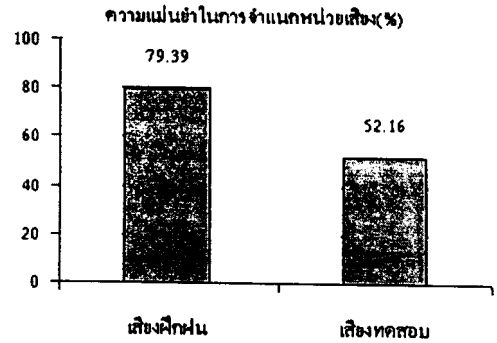


รูปที่ 2 ขั้นตอนการจำแนกหน่วยเสียง

สำหรับการสร้างแบบจำลองเสียง เริ่มจากการนำข้อมูลเสียงฝึกฝน ข้อมูลกำกับหน่วยเสียงตามชุดหน่วยเสียงที่กำหนด เข้าสู่กระบวนการเรียนรู้เพื่อสร้างแบบจำลองเสียงโดยมีขั้นตอนเช่นเดียวกับการสร้างแบบจำลองพื้นฐานตัวอย่างของชุดโปรแกรมฮิดเดนมาร์คอฟ HTK [10]

5. การทดลอง (Experiment)

ได้ทำการทดลองเพื่อหาประสิทธิภาพในการจำแนกหน่วยเสียงทั้งหมดโดยแยกตามกลุ่มเสียงฝึกฝนและกลุ่มเสียงทดสอบ แล้วนำกลุ่มเสียงทั้งสองเข้าสู่กระบวนการจำแนกหน่วยเสียงทีละกลุ่ม

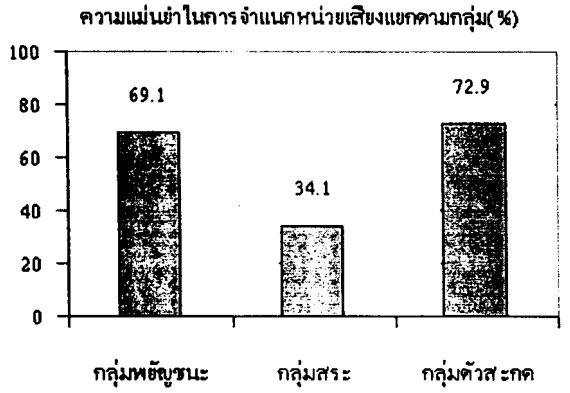


รูปที่ 3 ความแม่นยำในการจำแนกหน่วยเสียง

ผลการจำแนกหน่วยเสียงจากรูปที่ 3 ค่าความแม่นยำของเสียงทดสอบ 52.16 % เสียงฝึกฝน 79.39% พบว่ามีความแตกต่างระหว่างความแม่นยำในการจำแนกเสียงทดสอบกับเสียงฝึกฝนยังมีค่ามาก ซึ่งแสดงให้เห็นว่าความเที่ยงตรง (Precision) ของแบบจำลองเสียงยังไม่ดีพอ จึงได้ทำการทดลองต่อไปเพื่อพิจารณาความแม่นยำในแต่ละกลุ่มหน่วยเสียง โดยการกำหนดชุดหน่วยเสียงตามกลุ่มที่ต้องการพิจารณา 3 กลุ่มคือ กลุ่มพยัญชนะ กลุ่มสระ และกลุ่มตัวสะกด

จากรูปที่ 4 ค่าความแม่นยำแบบแยกลุ่มชี้ให้เห็นได้อย่างชัดเจนว่ากลุ่มสระมีการจำแนกได้ไม่ดีเท่าที่ควร อาจเนื่องมาจากจำนวนหน่วยเสียงในกลุ่มสระที่มีมากถึง 120 หน่วยเสียง หรือความใกล้เคียงกันในการออกเสียงของหน่วยเสียงแต่ละหน่วยตัวอย่าง เมื่อพิจารณาผลการรู้จำอย่างละเอียดพบว่าการ

รู้จำที่ผิดพลาดส่วนใหญ่มาจาก การรู้จำเสียง ร เป็นเสียง ล การรู้จำคลาดเคลื่อนของการออกเสียงสระสั้น-ยาวและระดับวรรณยุกต์ และตัวสะกด -ง -น –ม มีการรู้จำสลับกัน



รูปที่ 4 ความแม่นยำในการจำแนกหน่วยเสียงแบบแยกกลุ่ม

6. วิเคราะห์ผลการทดลอง

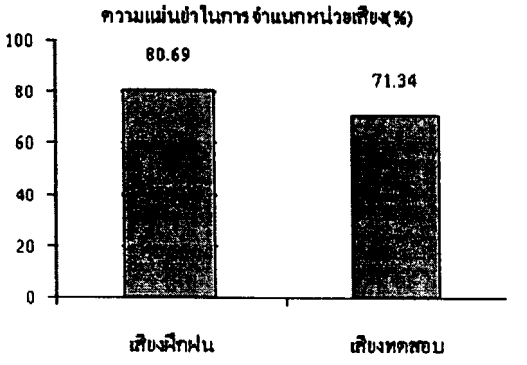
จากผลการทดลองพบว่าความแม่นยำในกลุ่มสระยังไม่ดีเท่าที่ควร มีหลายสาเหตุที่ทำให้การจำแนกหน่วยเสียงผิดพลาด ได้แก่ ปัญหาจากข้อมูลเสียง การสนทนาทั่วไป คำศัพท์บางคำอาจมีการออกเสียงไม่ตรงตามหลักภาษา บางอย่างเป็นเพราะผู้พูดออกเสียงผิด หรือความเคยชินของผู้พูด เช่น คำว่า มหาวิทยาลัย อ่านว่า มะ หา วิด ทะ ยา ลัย สังเกตคำว่า มะ ที่ต้องออกเสียงตรี แต่คนส่วนใหญ่ออกเสียงสามัญ ซึ่งความคลาดเคลื่อนในจุดนี้จะทำให้ชุดหน่วยเสียงในงานวิจัยนี้ไม่สามารถแก้ปัญหาได้

ด้วยเหตุนี้จึงได้ออกแบบการทดลองเพื่อพิสูจน์ว่าความผิดพลาดของการจำแนกหน่วยเสียงในกลุ่มสระนั้นมาจากสาเหตุความซ้ำซ้อนของหน่วยเสียงหรือไม่ โดยการรวมหน่วยเสียงที่มีความคล้ายคลึงกันเข้าด้วยกันและแยกหน่วยเสียงที่มีการออกเสียงต่างกันออกจากกันดังนี้ ลำดับแรกแยกเสียงพยัญชนะควบออกจากกัน ลำดับต่อมารวมสระชุดเดียวกันเข้าด้วยกันได้แก่ สระเสียงสั้นและสระเสียงยาวทุกวรรณยุกต์ให้เหลือเพียงหน่วยเสียงเดียว สุดท้ายตัดหน่วยเสียงตัวสะกดที่เป็นคำทับศัพท์ออกให้เหลือเพียงตัวสะกด 8 ตัวสะกดตามมาตรฐานของภาษาไทย รวมแล้วมีหน่วยเสียงทั้งสิ้น 41 หน่วยเสียง ดังตารางที่ 2

ตารางที่ 2 ชุดหน่วยเสียงภาษาไทยแบบ 41 หน่วยเสียง

พยัญชนะ	ก ค ง จ ช ด ต ท น บ ป พ ฟ ม ย ล ว ส อ ฮ
สระเดี่ยว	อะ อิ อึ อุ เอะ แอะ โอะ เอาะ เออะ เอียะ เอือะ อัวะ
ตัวสะกด	-บ -ด -ก -น -ม -ง –ย –ว

หน่วยเสียงชุดนี้ได้ตัดหน่วยเสียงออกไปหลายหน่วยเพื่อลดความซ้ำซ้อนของการออกเสียงในหน่วยเสียงที่ซ้ำซ้อนกัน ทำให้มีจำนวนตัวอย่างเสียงมากขึ้น และลดปัญหาความผิดพลาดในการจำแนกหน่วยเสียงในส่วนของวรรณยุกต์ และความสั้นยาวของสระ จากนั้นจึงทำการทดลองเช่นเดียวกับการทดลองแรก

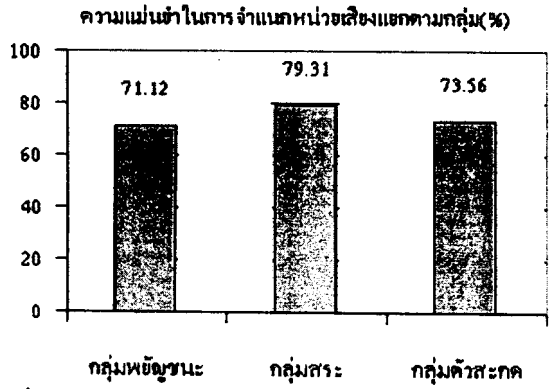


รูปที่ 5 ความแม่นยำรวมในการจำแนกหน่วยเสียงชุดที่ 2

จากผลการทดลองในรูปที่ 5 ความแม่นยำของเสียงทดสอบ 71.34 % แสดงให้เห็นว่า การลดความซ้ำซ้อนของหน่วยเสียงทำให้ระบบรู้จำมีประสิทธิภาพเพิ่มขึ้นเกือบ 20% ความแม่นยำในการรู้จำเสียงทดสอบและเสียงฝึกฝนใกล้เคียงกันมากขึ้นแสดงถึงความเที่ยงตรงของแบบจำลองเสียงที่ดีขึ้น

เมื่อพิจารณาในแต่ละกลุ่มหน่วยเสียงแล้วพบว่าประสิทธิภาพในการจำแนกโดยรวมมีค่าสูงขึ้น จากรูปที่ 6 กลุ่มพยัญชนะเพิ่มขึ้นจาก 69.05% เป็น 71.12% กลุ่มสระเพิ่มขึ้นจาก 34.1% เป็น 79.31% ซึ่งสูงขึ้นอย่างชัดเจน ส่วนกลุ่มตัวสะกดเพิ่มขึ้นจาก 72.9 % เป็น 73.56% แสดงให้เห็นว่าการลดความซ้ำซ้อนของหน่วยเสียงทำให้ระบบรู้จำถูกต้องมากขึ้น

แต่ยังพบความผิดพลาดในบางกลุ่มหน่วยเสียงได้แก่ การรู้จำเสียง จ เป็นเสียง ช การรู้จำเสียง พ เป็นเสียง ท การรู้จำเสียง ท เป็นเสียง พ กลุ่มสระ มีการรู้จำสลับกันระหว่างสระ ออ และ สระ อะ รู้จำเสียงสระอะ เป็นเสียงสระ แอะ สำหรับตัวสะกด -ง -น –ม มีการรู้จำที่ซ้ำซ้อนกันพอสมควร ซึ่งต้องมีการปรับปรุงในอนาคต



รูปที่ 6 ความแม่นยำตามกลุ่มในการจำแนกหน่วยเสียงชุดที่ 2

8. สรุปผลการทดลอง

เมื่อลดความซ้ำซ้อนของชุดหน่วยเสียงทำให้ประสิทธิภาพการรู้จำสูงขึ้น สรุปได้ว่าปัญหาสำคัญของชุดหน่วยเสียงภาษาไทยคือความแตกต่างระหว่างหน่วยเสียงที่มีน้อยมาก เช่น คำเดียวกันต่างวรรณยุกต์กันมีการออกเสียงคล้ายกันมาก ข้อมูลเสียงสำหรับการสอนระบบรู้จำจึงจำเป็นต้องมีความชัดเจน สำหรับปัญหาในการจำแนกระดับวรรณยุกต์อาจเนื่องมาจากวิธีการคำนวณค่าลักษณะเด่นที่ไม่สามารถรู้จำระดับวรรณยุกต์ได้ดีพอ แต่สามารถแก้ปัญหาได้โดยการเพิ่มค่าพิตช์ (Pitch) ในขั้นตอนการคำนวนค่าลักษณะเด่น ดังเช่นงานวิจัยเรื่องการรู้จำเสียงวรรณยุกต์ในภาษาไทย[11] ที่ใช้ค่าพิตช์เป็นค่าลักษณะเด่นในการรู้จำ ผลความแม่นยำที่ได้คือ 79.1%

9. เอกสารอ้างอิง

[1] วิศรุต อาชุบุตร, รศ.ดร. สมชาย จิตะพันธ์สกุล, ผศ.ดร. สุดาพร ลักษณียนาวิน,"ระบบการรู้จำคำไทยหลายพยางค์แบบไม่ขึ้นกับผู้พูด โดยใช้แบบจำลองฮิดเดน มาร์คอฟ",การประชุมวิชาการไฟฟ้า ครั้งที่ 19, Vol. 1, 2539, หน้า DS-56 - DS-59.

[2] Virach Sornlertlamvanich, Tanapong Potipiti, Chai Wutiwiwatchai and Pradit Mittrapiyanuruk, "The State of the Art in Thai Language Processing", National Electronics and Computer Technology Center (NECTEC), 2000.

[3] วุฒิพงษ์ พรสุขจันทรา, รศ.ดร. สมชาย จิตะพันธ์สกุล," การรู้จำตัวเลขภาษาไทยแบบไม่ขึ้นกับผู้พูดโดยใช้ แอล พีซี และ นิวรอลเน็ตเวิร์คแบบแบ็กพรอพาเกชัน",การประชุมวิชาการไฟฟ้า ครั้งที่ 19, Vol. 1, 2539, หน้า DS-67 - DS-71.

[4] ชินภัทร นันทจิวากรชัย, พลศาสตร์ เลิศประเสริฐ, ประภากร สุวรรณา, มนัส สังวรศิลป์,"การรู้จำเสียงพูดโดยใช้ลักษณะแบ่งความต่าง",การประชุมวิชาการไฟฟ้า ครั้งที่ 16, Vol. 1, 2536, หน้า 244 - 247.

[5] Supphanat Kanokphara, "Syllable Structure Base Phonetic Units for Context-Dependent Continuous Thai Speech Recognition," EUROSPEECH 2003-Geneva

[6] N. Thatphithakkul, S. Kanokphara, "Hmm Parameter Optimization using Tabu Search," Communications and Information Technologies 2004 (ISCIT 2004)

[7] C.Wutiwiwatchai, P. Cotsomrong, S. Suebvisai, S. Kanokphara, "Phonetically Distributed Continuous Speech Corpus for Thai Language", LREC2002, vol. 3, 869-872.

[8] Thongprasert, R., Sornlertlamvanich, V., Cotsompong, P., Suebrisai, S., and Kanokphara, S., "Process Reportion Corpus Development and Speech Technology in Thailand", Proc. Of the Joint International Conference of SNLP Oriental COCOSDA 2002,pp. 300,306, Prachuapkirikhan, Thailand, 9-11 May, 2002

[9] Gowdy, J. N; Tufekci, Z," Mel-scaled discrete wavelet coefficients for speech recognition", Acoustics, Speech, and Signal Processing, 2000. ICASSP 00. Proceedings.2000 IEEE International Conference, vol3

[10] Department of Electronic Engineering The Chinese University of Hong Kong," Mini-project: Speech Recognition Using HTK", ELE 7080 Digital Processing of Speech Signals (2000 – 2001)

[11] Li Tan, Montri Karnjanadecha, "Pitch Detection Algorithm: Auto Correlation Method and AMDF," ISCIT2003, Songkhla, Thailand