รายการอ้างอิง

[1] Glass, J.R. A probabilistic frameworkfor segment-based speech recognition. Computer Speech and Language 17 (2003): 137–152.

[2] Kvale, K. On the Connection Between Manual Segmentation Conventions and Errors Made by Automatic Segmentation. Proceedings of ICSLP'94, pp. 1667–1670, 1994.

[3] Zue, V., Glass, J.R., Phillips, M., and Sene, S. The MIT SUMMIT speech recognition system: a progress report. Proceedings of the Speech and Natural Language Workshop, pp. 179–189, 1989.

[4] Lee, S.C. Probabilistic Segmentation for Segment-Based Speech Recognition. Master's Thesis, Department of Electrical Engineering and Computer Science, MIT, Cambridge, 1998.

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

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

[7] Stevens, K.N. Acoustic Phonetics. Cambridge, MA: MIT Press, 1999.

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

[9] Chomsky, N., Halle, N. The Sound Pattern of English. Cambridge, MA: MIT Press, 1968.

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

[11] Mammone, R.J., Zhang, X., and Ramachandran, R.P. Robust Speaker Recognition, A Feature-based Approach, IEEE Signal Processing Magazine, pp. 58-71, 1996.

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

[13] Roe, D.B., Wilpon, J.G. Whither Speech Recognition: the next 25 years, IEEE Comm. Magazine, Vol. 31, 1993.

[14] Ostendorf, M., Digalakis, V., and Kimball, O. From HMMs to Segment Models: A Unified View of Stochastic Modeling for Speech Recognition. IEEE Transactions on Speech and Audio Proceesing, pp. 360-378, 1996.

[15] บุญเสริม กิจศิริกุล, การเรียนรู้ของเครื่อง. โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย, 2549

[16] Vapnik, V. Statistical Learning Theory. New York: Wiley.

[17] Kipp, A., Wesenick, M. B., and Schiel, F. Automatic Detection and Segmentation of Pronunciation Variants in German Speech Corpora. Proc. of ICSLP'96, pp. 106-109, 1996.

[18] Brugnara, F., Falavigna, D., and Omologo, M. Automatic Segmentation and Labeling of Speech Based on Hidden Markov Models. Speech Communication (1993): 357-370.

[19] Hosom, J.P. Automatic Time Alignment of Phonemes Using Acousitc-Phonetic Information. Ph.D. Thesis, Oregon Graduate Institute of Science and Technology, 2000.

[20] Kim, Y.J., and Conkie, A. Automatic Segmentation Combining an HMM-based Approach and Spectral Boundary Correction. In Proc. ICSLP'2002, pp. 145-148, 2002.

[21] Leelaphattarakij, P., Punyabukkana, P., and Suchato, A. Locating Phone Boundaries from Acoustic Discontinuities using a Two-staged Approach. The Ninth International Conference on Spoken Language Processing: Interspeech2006, Pittsburgh, PA, 2006.

[22] Chang, J., Glass, J.R. Segmentation and modeling in segment-based recognition, Proc. Eurospeech'97, pp. 1199-1202, 1997.

[23] Cole, R., Fanty, M. Spoken letter recognition. In Proc. Third DARPA Speech and Natural Language Workshop, pp. 385-390, 1990.

[24] Wang, D., Lu, L., and Zhang H.-J. Speech Segmentation without Speech Recognition. Proc. of IEEE ICASSP'03, pp. 468-471, 2003.

[25] Glass, J.R., Zue, V.W. Multi-Level Acoustic Segmentation of Continuous Speech. Proc. of IEEE ICASSP'88, pp. 429-432, 1988.

[26] Glass, J.R. Finding Acoustic Regularities in Speech: Application to Phonetic Recognition. Ph.D Thesis, MIT press, 1988.

[27] Sorin, D., Rabiner, L.R. On the Relation between Maximum Spectral Transition Positions and Phone Boundaries. The Ninth International Conference on Spoken Language Processing: Interspeech2006, Pittsburgh, PA, 2006.

[28] Kasuriyam, S., Sornlertlamvanich, V., Cotsomrong, P., Kanokphara, S., and Thatphithakkul, N. Thai Speech Corpus for Thai Speech Recognition. Proc. COCOSDA'03, pp. 54-61, 2003.

[29] Young, S., Jansen, J., Odell, J. Ollason, D., and Woodland, P. The HTK Book (for HTK Version 3.3). Cambridge, 2005.

[30] Makashay, M. J., Wightman, C. W., Syrdal, A. K., and Conkie, A. Perceptual evaluation of automatic segmentation in text-to-speech synthesis. In Proc. ICLSP2000, pp. 431-434, 2000.

[31] Kominek, J. K., Bennett, C., and Black, A. W. Evaluating and Correcting Phoneme Segmentation for Unit Selection Synthesis. In Proc. EUROSPEECH'2003, pp. 313-316, 2003.

[32] Kawai, H., and Toda, T. An evaluation of automatic phone segmentation for concatenative speech synthesis. In Proc. ICASSP'2004, pp.677–680, Quebec, Canada, May 2004.

[33] Huckvale, M. Speech Filing System Vs3.0 – Computer Tools for Speech Research, London, U.K.: University College, 1996.

[34] Scholkopf, B. Smola, A. J., and Muller, K. R. Kernel principal component analysis. Advances in Kernel Methods - Support Vector Learning. MIT Press, 1999.

ภาคผนวก

ภาคผนวก

ในหัวข้อนี้จำนำเสนอเกี่ยวกับขั้นตอนการเรียนรู้และการรู้จำของเครื่องรู้จำเสียงพูดที่นำมาใช้ในการตรวจหาขอบเขตของหน่วยเสียง การสกัดลักษณะสำคัญที่ได้จากการใช้สารสนเทศสวนสัทศาสตร์ รวมถึงนำเสนอเกี่ยวกับการเรียนรู้และการจำแนกข้อมูลด้วยซัพพอร์ตเวกเตอร์แมชชีน

ขั้นตอนการเรียนรู้และการรู้จำของเครื่องรู้จำเสียงพูด

ในวิทยานิพนธ์นี้ทำการสร้างเครื่องรู้จำเสียงพูดโดยใช้โปรแกรม Hidden Markov Toolkit - HTK [29] ซึ่งเป็นชุดเครื่องมือสำหรับสร้างเครื่องรู้จำเสียงพูดแบบอาศัยแบบจำลองฮิดเดนมาร์คอฟ ในหัวข้อนี้จะนำเสนอเกี่ยวกับขั้นตอนการการเรียนรู้เสียงพูด การรู้จำเสียงพูด และการตรวจหาขอบเขตของหน่วยเสียงจากผลการรู้จำเสียงพูด

1. กระบวนการเรียนรู้

กระบวนการเรียนรู้มีขั้นตอนต่างๆดังนี้

1.1 การหาลักษณะสำคัญของเสียงเพื่อการเรียนรู้

ลักษณะสำคัญของเสียงเพื่อการเรียนรู้ในที่นี้คือ สัมประสิทธิ์เซปสตรัมบนสเกลเมลดังที่กล่าวไว้ในหัวข้อการสกัดลักษณะสำคัญ โดยใช้โปรแกรม HCopy ซึ่งเป็นเครื่องมือที่อยู่ในชุดเครื่องมือ HTK มาใช้ในการสกัดลักษณะสำคัญจากสัญญาณเสียง โดยเลือกใช้พารามิเตอร์ต่างๆ ซึ่งจะระบุไว้ในไฟล์ code.config ดังนี้

```
# HTK Configuration Parameters for Generating MFCC_E _D_A

SOURCEFORMAT=WAV        # source forwat is WAV
SOURCEKIND = WAVEFORM # simple waveform
SOURCERATE = 625        # source sampling frequency is [16kHz]

# mel-frequency cepstral coefs, energy, their deltas, and accelerations

TARGETKIND = MFCC_E _D_A
```

```
TARGETRATE=100000.0     # frame interval is 10msec
WINDOWSIZE=250000.0     # window length is 25msec
USEHAMMING=T            # use HAMMING window
PREEMCOEF=0.97          # apply highpass filtering
NUMCHANS=24             # number of filterbank for MFCC is 24
NUMCEPS=12              # number of parameters for MFCC presentation

# Rather local Parameters

ENORMALISE=T            # normalize energy
ESCALE=1.0              # energy scale is 1.0
```

เมื่อใช้พารามิเตอร์ต่างๆที่กำหนด โปรแกรม HCopy จะคำนวณหาสัมประสิทธิ์เซปสตรัมบนสเกลเมลโดยมีค่าพลังงานรวมอยู่ด้วย อัตราการเปลี่ยนแปลง (delta) และความเร่ง (accelerations) จากสัญญาณเสียงทุกๆกรอบเวลายาว 25 มิลลิวินาที โดยแต่ละกรอบเวลาจะมีระยะเวลาห่างกัน 10 มิลลิวินาที ได้ออกมาเป็นเวกเตอร์ลักษณะสำคัญที่มีขนาด 39 มิติ โดยต่อจากนี้จะขออ้างอิงชุดลักษณะสำคัญที่ด้วยสัญลักษณ์ MFCC_E_D_A

การใช้งานโปรแกรม HCopy จะต้องเรียกใช้ด้วยการพิมพ์คำสั่งดังนี้

```
HCopy -C code.config -S CCHCopy.scp
```

โดยที่

- code.config เป็นไฟล์ที่กำหนดค่าพารามิเตอร์ในการหาลักษณะสำคัญของเสียง
- –C code.config เป็นการกำหนดให้ใช้ไฟล์ code.config เป็นไฟล์กำหนดค่าพารามิเตอร์ในการหาลักษณะสำคัญของเสียง
- CCHCopy.scp เป็นไฟล์ข้อความซึ่งแต่ละบรรทัดเป็นชื่อของไฟล์สัญญาณเสียง .wav ที่เป็นอินพุต และไฟล์ลักษณะสำคัญที่เป็นเอาท์พุตโดยมีเครื่องหมายเว้นวรรคคั่นกลาง ตัวอย่างเช่น CCF001_Pa001_001.wav CCF001_Pa001_001.mfc เป็นการกำหนดให้ใช้ไฟล์ CCF001_Pa001_001.wav เป็นไฟล์สัญญาณเสียงอินพุต และให้เขียนไฟล์เอาท์พุตชื่อ CCF001_Pa001_001.mfc ออกมา

- -S CCHCopy.scp เป็นการกำหนดให้ใช้ไฟล์ CCHCopy.scp เป็นไฟล์กำหนดไฟล์สัญญาณเสียง .wav ที่เป็นอินพุต และไฟล์ลักษณะสำคัญที่เป็นเอาท์พุต ในการหาลักษณะสำคัญของเสียง

1.2 การเตรียมข้อมูลเพื่อนำไปใช้ในการเรียนรู้

ข้อมูลที่ต้องเตรียมเพื่อนำไปใช้ในการเรียนรู้การรู้จำเสียงพูดแบ่งออกเป็นสองส่วนคือ ส่วนที่เป็นไฟล์กำหนดพารามิเตอร์และลักษณะการเรียนรู้ซึ่งได้แก่

- ไฟล์พจนานุกรม (Dictionary File) เป็นไฟล์ข้อความแสดงรายการของคำศัพท์ของเครื่องรู้จำเสียงพูด โดยในงานวิจัยนี้ต้องการสร้างเครื่องรู้จำที่รู้จำเสียงพูดในระดับหน่วยเสียง ดังนั้นพจนานุกรมที่ใช้จึงอยู่ในรูปหน่วยเสียงทั้งหมดแทนที่จะอยู่ในรูปของคำ โดยในที่นี้จะใช้ชื่อไฟล์เป็น CCphn.dict
- ไฟล์รายการหน่วยเสียง (Phone List File) เป็นไฟล์ข้อความแสดงรายการหน่วยเสียงทั้งหมด โดยแต่ละบรรทัดจะแทนแต่ละหน่วยเสียง โดยในที่นี้จะใช้ชื่อไฟล์เป็น CCphn.list
- ไฟล์ต้นแบบทอพอโลยีของแบบจำลองเสียง (HMM Prototype File) เป็นไฟล์ที่เก็บพารามิเตอร์ตั้งต้นของแบบจำลองฮิดเดนมาร์คอฟเอาไว้ โดยจะนำมาใช้เป็นไฟล์ต้นแบบสำหรับสร้างแบบจำลองเสียง เราสามารถกำหนดทอพอโลยี ปรับเปลี่ยนพารามิเตอร์ตั้งต้น และจำนวนสถานะของแบบจำลองฮิดเดนมาร์คอฟได้จากไฟล์ต้นแบบนี้ ในที่นี้เลือกใช้แบบจำลองฮิดเดนมาร์คอฟที่มีจำนวนสถานะ 5 สถานะโดยมีทอพอโลยีเป็นแบบจากซ้ายไปขวา โดยในที่นี้จะใช้ชื่อไฟล์เป็น proto5s

ข้อมูลอีกส่วนหนึ่งจะเป็นไฟล์ข้อมูลอินพุตเพื่อการเรียนรู้ซึ่งได้แก่

- ไฟล์ลักษณะสำคัญ ซึ่งเป็นไฟล์ที่ได้จากขั้นตอนในหัวข้อการหาค่าลักษณะเพื่อการเรียนรู้ โดยใช้ไฟล์เสียงจากชุดข้อมูลเสียงเพื่อการเรียนรู้ดังที่ได้กล่าวไว้ในหัวข้อฐานข้อมูลเสียงเพื่อการแบ่งเสียงพูดเป็นเซกเมนต์เป็นอินพุต รวมแล้วจะได้ไฟล์ลักษณะสำคัญเพื่อการเรียนรู้ 840 ไฟล์
- ไฟล์กำกับหน่วยเสียง (Label File) ซึ่งเป็นไฟล์ข้อความที่แสดงลำดับของหน่วยเสียงและระยะเวลาของการเกิดหน่วยเสียงต่างๆของแต่ละไฟล์เสียง โดยไฟล์เหล่านี้จะมีมาให้อยู่แล้วในชุดข้อมูลเสียงเพื่อการเรียนรู้ รวมแล้วจะได้ไฟล์กำกับหน่วยเสียง 840

ไฟล์เท่ากันกับจำนวนของไฟล์ลักษณะสำคัญ และ ไฟล์เสียงของชุดข้อมูลเสียงเพื่อการเรียนรู้

- ไฟล์กำกับหน่วยเสียงสำคัญ (Master Label File) ซึ่งเป็นไฟล์ข้อความที่รวบรวมไฟล์กำกับหน่วยเสียงย่อยทั้งหมดมาไว้ในไฟล์เดียว เพื่อสะดวกต่อการนำไปใช้ในการเรียนรู้ โดยในที่นี้จะใช้ชื่อไฟล์เป็น CCphn.mlf

1.3 การเรียนรู้

การเรียนรู้ในที่นี้ใช้แบบจำลองฮิดเดนมาร์คอฟดังที่กล่าวไว้ในบทที่ 2 โดยใช้ชุดเครื่องมือ HTK เป็นเครื่องมือสำหรับสร้างเครื่องรู้จำเสียงพูด โดยนำมาใช้ในการเรียนรู้แบบจำลองภาษา และแบบจำลองเสียงพูด โดยแบบจำลองเสียงพูดที่เรียนรู้จะมีจำนวนเท่ากับจำนวนหน่วยเสียงทั้งหมดที่ต้องการรู้จำ ขั้นตอนการเรียนรู้ในที่นี้แบ่งออกเป็นสองขั้นตอนคือ การเรียนรู้แบบจำลองภาษา และการเรียนรู้แบบจำลองเสียง โดยมีรายละเอียดดังนี้

1.3.1 การเรียนรู้แบบจำลองภาษา

การเรียนรู้แบบจำลองภาษาในที่นี้ ใช้แบบจำลองภาษาแบบอาศัยค่าความน่าจะเป็นของการที่หน่วยเสียงหนึ่งจะปรากฏอยู่ติดกับอีกหน่วยเสียงหนึ่ง (Bigram Language Model) โดยใช้โปรแกรม HLStats มาวิเคราะห์เก็บรวบรวมสถิติค่าความน่าเป็นออกมาจากไฟล์กำกับหน่วยเสียงสำคัญ การใช้งานโปรแกรม HLStats จะต้องเรียกใช้ด้วยการพิมพ์คำสั่งดังนี้

```
HLStats –C code.config -b CCphn.big CCphn.lst CCphn.mlf
```

โดยที่

- -C code.config เป็นการกำหนดให้ใช้ไฟล์ code.config เป็นไฟล์กำหนดค่าพารามิเตอร์ในการหาลักษณะสำคัญของเสียง
- -b CCphn.big เป็นการกำหนดให้เก็บสถิติค่าความเป็นของการที่หน่วยเสียงหนึ่งจะปรากฏอยู่ติดกับอีกหน่วยเสียงหนึ่งเก็บอยู่ในไฟล์ชื่อ CCphn.big
- CCphn.list คือไฟล์รายการหน่วยเสียงที่ใช้เป็นอินพุต
- CCphn.mlf คือไฟล์กำกับหน่วยเสียงสำคัญที่ใช้เป็นอินพุต

แบบจำลองภาษาที่ใช้จะอยู่ในรูปของโครงข่ายไวยากรณ์ของการเกิดหน่วยเสียงต่างๆ โดยใช้โปรแกรม HBuild มารับไฟล์สถิตค่าความเป็นของการที่หน่วยเสียงหนึ่งจะปรากฏอยู่ติดกับอีกหน่วยเสียงเข้ามาเป็นอินพุตแล้วแปลงให้อยู่ในรูปของโครงข่ายไวยากรณ์

การใช้งานโปรแกรม HBuild จะต้องเรียกใช้ด้วยการพิมพ์คำสั่งดังนี้

```
HBuild -C code.config -n CCphn.big CCphn.lst CCphn.net
```

โดยที่

- –C code.config เป็นการกำหนดให้ใช้ไฟล์ code.config เป็นไฟล์กำหนดค่าพารามิเตอร์ในการหาลักษณะสำคัญของเสียง
- -n CCphn.big เป็นการกำหนดให้ใช้ไฟล์ CCphn.big เป็นไฟล์เก็บสถิตค่าความเป็นของการที่หน่วยเสียงหนึ่งจะปรากฏอยู่ติดกับอีกหน่วยเสียง
- CCphn.list คือไฟล์รายการหน่วยเสียงที่ใช้เป็นอินพุต
- CCphn.net คือไฟล์แบบจำลองภาษาที่เป็นเอาท์พุต

1.3.2 การเรียนรู้แบบจำลองเสียงพูด

การเรียนรู้แบบจำลองเสียงพูดในที่นี้ จะสร้างแบบจำลองเสียงพูดสำหรับหน่วยเสียงในภาษาไทยทั้งหมด ให้เป็นแบบจำลองเสียงพูดแบบไม่ขึ้นกับบริบทรอบข้าง (Context-independent Phone Model) โดยจะใช้การกระจายของค่าความน่าจะเป็นแบบเกาส์เซียนที่มีเมทริกซ์ความแปรปรวนร่วมเกี่ยวตามแนวทแยง (Diagonal Covariance Gaussian Distribution) มาประมาณความน่าจะเป็นของการเกิดลักษณะสำคัญในแบบจำลองฮิดเดนมาร์คอฟ การเรียนรู้แบบจำลองเสียงมีขั้นตอนต่างๆดังนี้

1. การกำหนดลักษณะทอพอโลยี

ในลำดับแรกเราจะต้องกำหนดลักษณะทอพอโลยีตั้งต้นของแบบจำลองเสียงพูด โดยใช้โปรแกรม HCompV และไฟล์ต้นแบบทอพอโลยีของแบบจำลองเสียง ดังที่กล่าวไว้ในหัวข้อการเตรียมข้อมูลเพื่อใช้ในการเรียนรู้ การใช้งานโปรแกรม HCompV จะต้องพิมพ์คำสั่งดังนี้

```
HCompV -C train.config -f 0.01 -m –S CCTrain.scp -M hmm_0 proto5s
```

โดยที่

- -C train.config เป็นการกำหนดให้ใช้ไฟล์ train.config เป็นไฟล์กำหนดค่าพารามิเตอร์ในการเรียนรู้ โดยจะระบุพารามิเตอร์สามอย่างได้แก่ SOURCEKIND TARGETKIND และ RAWENERGY โดยใช้ค่าเดียวกันกับพารามิเตอร์ในไฟล์ code.config ทุกประการ
- -f 0.01 เป็นการกำหนดให้ค่าความแปรปรวนชักลง (Variance floor) เป็น 0.01
- -m เป็นการกำหนดให้มีการคำนวณพารามิเตอร์ ค่าเฉลี่ย เช่นเดียวกันกับคำนวณค่าความแปรปรวน
- -S CCTrain.scp เป็นการกำหนดให้ใช้ไฟล์ลักษณะสำคัญที่แสดงเป็นรายการอยู่ในไฟล์ CCTrain.scp มาใช้ในการเรียนรู้
- -M hmm_0 เป็นการกำหนดให้เก็บเอาท์พุตแบบจำลองเสียงพูดที่ได้ไว้ในไดเร็กทอรีชื่อ hmm_0
- proto5s เป็นการกำหนดให้ใช้ลักษณะทอพอโลยีที่กำหนดไว้แล้วในไฟล์ proto5s

ผลลัพท์ที่ได้คือไฟล์เป็นภาพรวมของแบบจำลองเสียงชื่อ macros และไฟล์ค่าความแปรปรวนของแต่ละลักษณะสำคัญชื่อ vFloor ซึ่งอยู่ในไดเร็กทอรี hmm_0

2. **การกำหนดค่าพารามิเตอร์ตั้งต้น**

ขั้นตอนถัดไปจะเป็นการกำหนดค่าตั้งต้นให้กับพารามิเตอร์ต่างๆให้กับแบบจำลองเสียงพูดของแต่ละหน่วยเสียง โดยใช้โปรแกรม HInit มากำหนดพารามิเตอร์ตั้งต้นจากลักษณะสำคัญที่ได้จากแต่ละหน่วยเสียง

การใช้งานโปรแกรม HInit จะต้องเรียกใช้ด้วยการพิมพ์คำสั่งดังนี้

```
HInit -I CCphn.mlf -S CCTrain.scp -H hmm_0/macros -C train.config -M hmm_1 -l $phn
proto5s
```

โดยที่

- -I CCphn.mlf เป็นการกำหนดให้โปรแกรม HInit ไปอ่านข้อมูลจากไฟล์กำกับหน่วยเสียงสำคัญชื่อ CCphn.mlf ขึ้นมาเพื่อดูว่าหน่วยเสียงแต่ละแบบอยู่ในช่วงเวลาใดบ้าง เพื่อที่จะได้เลือกลักษณะสำคัญในช่วงเวลาดังกล่าวมากำหนดค่าพารามิเตอร์ของแบบจำลองเสียงพูดของหน่วยเสียงนั้นๆ
- -S CCTrain.scp เป็นการกำหนดให้ใช้ไฟล์ลักษณะสำคัญที่แสดงเป็นรายการอยู่ในไฟล์ CCTrain.scp มาใช้ในการเรียนรู้

- -H hmm_0/macros เป็นการกำหนดให้อ่านไฟล์ภาพรวมของแบบจำลองเสียงที่อยู่ในไดเร็กทอรี hmm_0 เข้ามา
- -C train.config เป็นการกำหนดให้ใช้ไฟล์ train.config เป็นไฟล์กำหนดค่าพารามิเตอร์เพื่อการเรียนรู้
- -M hmm_1 เป็นการกำหนดให้เก็บเอาท์พุตแบบจำลองเสียงพูดที่ได้ไว้ในไดเร็กทอรีชื่อ hmm_1
- -l *$phn* เป็นการกำหนดพารามิเตอร์ให้แบบจำลองเสียงพูดของหน่วยเสียง *$phn*
- proto5s เป็นการกำหนดให้ใช้ลักษณะทอพอโลยีที่กำหนดไว้แล้วในไฟล์ proto5s

โดยจะต้องเรียกใช้โปรแกรม HInit เพื่อประมาณพารามิเตอร์ตั้งต้นของแบบจำลองเสียงพูดของหน่วยเสียงทุกๆหน่วยเสียง เพื่อให้ได้แบบจำลองเสียงพูดครบตามที่ต้องการ

ผลที่ได้จะเป็นไฟล์แบบจำลองเสียงพูดของแต่ละหน่วยเสียงโดยมีชื่อไฟล์ตรงกับสัญลักษณ์หน่วยเสียงนั้นๆเก็บไว้ในไดเร็กทอรี hmm_1

3. **การประมาณค่าพารามิเตอร์**

ขั้นตอนถัดไปจะเป็นการประมาณค่าพารามิเตอร์ต่างๆของแบบจำลองเสียงพูด โดยใช้โปรแกรม HRest มาประมาณพารามิเตอร์จากลักษณะสำคัญที่ได้จากแต่ละหน่วยเสียง

การใช้งานโปรแกรม HRest จะต้องเรียกใช้ด้วยการพิมพ์คำสั่งดังนี้

```
HRest –I CCphn.mlf -S CCTrain.scp -H hmm_1\macros -M hmm_2 -l $phn hmm_1\$phn
```

โดยที่

- -I CCphn.mlf เป็นการกำหนดให้โปรแกรม HInit ไปอ่านข้อมูลจากไฟล์กำกับหน่วยเสียงสำคัญชื่อ CCphn.mlf ขึ้นมาเพื่อดูว่าหน่วยเสียงแต่ละแบบอยู่ในช่วงเวลาใดบ้าง เพื่อที่จะได้เลือกลักษณะสำคัญในช่วงเวลาดังกล่าวมากำหนดค่าพารามิเตอร์ของแบบจำลองเสียงพูดของหน่วยเสียงนั้นๆ
- -S CCTrain.scp เป็นการกำหนดให้ใช้ไฟล์ลักษณะสำคัญที่แสดงเป็นรายการอยู่ในไฟล์ CCTrain.scp มาใช้ในการเรียนรู้
- -H hmm_1\macros เป็นการกำหนดให้อ่านไฟล์ภาพรวมของแบบจำลองเสียงที่อยู่ในไดเร็กทอรี hmm_1 ซึ่งได้จากขั้นตอนที่แล้วเข้ามา

- -M hmm_2 เป็นการกำหนดให้เก็บเอาท์พุตแบบจำลองเสียงพูดที่ได้ไว้ในไดเร็กทอรีชื่อ hmm_2
- -l *$phn* เป็นการกำหนดพารามิเตอร์ให้แบบจำลองเสียงพูดของหน่วยเสียง *$phn*
- hmm_1*$phn* เป็นการกำหนดให้นำไฟล์แบบจำลองเสียงพูดของหน่วยเสียง *$phn* ที่ได้จากขั้นตอนที่แล้วมาใช้เป็นอินพุต

ผลที่ได้คือไฟล์แบบจำลองเสียงพูดของแต่ละหน่วยเสียงเก็บไว้ในไดเร็กทอรี hmm_2

ต่อจากนั้นจะเป็นการประมาณพารามิเตอร์ซ้ำๆกันอีกหลายครั้งเพื่อให้ได้ค่าพารามิเตอร์ที่ใกล้เคียงกันกับลักษณะสำคัญของเสียงพูดมากยิ่งขึ้น โดยในครั้งนี้จะใช้โปรแกรม HERest มาประมาณพารามิเตอร์แทนการใช้โปรแกรม HRest เนื่องจากโปรแกรม HERest สามารถประมาณค่าพารามิเตอร์ของทุกๆแบบจำลองเสียงไปพร้อมกันได้แบบคู่ขนาน โดยในที่นี้เราจะประมาณค่าพารามิเตอร์ซ้ำเป็นจำนวน 10 รอบ ได้ผลลัพท์ออกมาเป็นแบบจำลองเสียงพูดของแต่ละหน่วยเสียงเก็บอยู่ในไดเร็กทอรี hmm_12

การใช้งานโปรแกรม HERest จะต้องเรียกใช้ด้วยการพิมพ์คำสั่งดังนี้

```
HERest -C train.config -I CCphn.mlf -S CCTrain.scp -H hmm_$i\macros -M hmm_$j
CCphn.list
```

โดยที่

- -C train.config เป็นการกำหนดให้ใช้ไฟล์ train.config เป็นไฟล์กำหนดค่าพารามิเตอร์ในการเรียนรู้ โดยจะระบุพารามิเตอร์สามอย่างได้แก่ SOURCEKIND TARGETKIND และ RAWENERGY โดยใช้ค่าเดียวกันกับพารามิเตอร์ในไฟล์ code.config ทุกประการ
- -I CCphn.mlf เป็นการกำหนดให้โปรแกรม HInit ไปอ่านข้อมูลจากไฟล์กำกับหน่วยเสียงสำคัญชื่อ CCphn.mlf ขึ้นมาเพื่อดูว่าหน่วยเสียงแต่ละแบบอยู่ในช่วงเวลาใดบ้าง เพื่อที่จะได้เลือกลักษณะสำคัญในช่วงเวลาดังกล่าวมากำหนดค่าพารามิเตอร์ของแบบจำลองเสียงพูดของหน่วยเสียงนั้นๆ
- -S CCTrain.scp เป็นการกำหนดให้ใช้ไฟล์ลักษณะสำคัญที่แสดงเป็นรายการอยู่ในไฟล์ CCTrain.scp มาใช้ในการเรียนรู้
- -H hmm_*$i*\macros เป็นการกำหนดให้อ่านไฟล์ภาพรวมของแบบจำลองเสียงที่อยู่ในไดเร็กทอรี hmm_*$i* เข้ามา

- -M hmm_$j เป็นการกำหนดให้เก็บเอาท์พุตแบบจำลองเสียงพูดที่ได้ไว้ในไดเร็กทอรี ชื่อ hmm_$j เมื่อให้ตัวแปร $j = $i – 1
- CCphn.list คือไฟล์รายการหน่วยเสียงที่ใช้เป็นอินพุต

2. กระบวนการรู้จำ

กระบวนการรู้จำเป็นการนำผลที่ได้จากขั้นตอนการเรียนรู้ ซึ่งในที่นี้คือแบบจำลองเสียงพูดและแบบจำลองภาษาที่ผ่านการเรียนรู้แล้ว รวมกันเป็นเครื่องรู้จำเสียงพูดที่รู้จำเสียงพูดได้ในระดับหน่วยเสียง

กระบวนการรู้จำมีขั้นตอนต่างๆดังต่อไปนี้

2.1 การหาค่าลักษณะสำคัญของเสียงเพื่อการรู้จำ

ลักษณะสำคัญของเสียงเพื่อการรู้จำจะใช้สัมประสิทธิ์เซปสตรัมบนสเกลเมลเหมือนกันกับค่าลักษณะสำคัญของเสียงเพื่อการเรียนรู้ โดยค่าพารามิเตอร์ที่ใช้ในการหาลักษณะสำคัญของเสียงในที่นี้จำเป็นต้องเหมือนกับค่าที่ใช้ในการหาลักษณะสำคัญของเสียงเพื่อการเรียนรู้ทุกประการ

2.2 การเตรียมข้อมูลเพื่อนำไปใช้ในการรู้จำ

ข้อมูลที่ต้องเตรียมเพื่อนำไปใช้ในการรู้จำเสียงพูดในที่นี้ได้แก่

- ไฟล์ลักษณะสำคัญ ซึ่งเป็นไฟล์ที่ได้จากขั้นตอนการหาค่าลักษณะสำคัญของเสียงเพื่อการรู้จำโดยใช้ไฟล์เสียงจากชุดข้อมูลเสียงเพื่อการทดสอบดังที่ได้กล่าวไว้ในหัวข้อฐานข้อมูลเสียงเพื่อการแบ่งเสียงพูดเป็นเซกเมนต์ รวมแล้วจะได้ไฟล์ลักษณะสำคัญเพื่อการรู้จำ 840 ไฟล์
- ไฟล์กำกับหน่วยเสียง (Label File) ซึ่งเป็นไฟล์ข้อความที่แสดงลำดับของหน่วยเสียงและระยะเวลาของการเกิดหน่วยเสียงต่างๆของแต่ละไฟล์เสียง โดยไฟล์เหล่านี้จะมีมาให้อยู่แล้วในชุดข้อมูลเสียงเพื่อการทดสอบ รวมแล้วจะได้ไฟล์กำกับหน่วยเสียง 840 ไฟล์

2.3 การรู้จำ

การรู้จำจะทำโดยการป้อนข้อมูลที่ได้จากกระบวนการเตรียมข้อมูลเพื่อนำไปใช้ในการรู้จำ เข้าเป็นอินพุตของเครื่องรู้จำเสียงพูดที่ได้จากการเรียนรู้ เครื่องรู้จำเสียงพูดจะคำนวณหาลำดับของหน่วยเสียงที่ให้ค่าความน่าจะเป็นสูงสุด N อันดับ ซึ่งคิดมาจากความน่าจะเป็นของแบบจำลองเสียง และแบบจำลองภาษาร่วมกัน โดยการรู้จำจะใช้โปรแกรม HVite มารู้จำเสียงพูดด้วยการพิมพ์คำสั่งดังนี้

```
HVite -S CCTest.scp -H hmm_12/macros -w CCphn.net -n $i N -l '*' -i recout.mlf
CCphn.dict CCphn.list
```

โดยที่

- -S CCTest.scp เป็นการกำหนดให้ใช้ไฟล์ลักษณะสำคัญที่แสดงเป็นรายการอยู่ในไฟล์ CCTest.scp มาใช้ในการเรียนรู้
- -H hmm_12\macros เป็นการกำหนดให้อ่านไฟล์ภาพรวมของแบบจำลองเสียงที่อยู่ในไดเร็กทอรี hmm_12 เข้ามา
- -w CCphn.net เป็นการกำหนดใช้ใช้ไฟล์แบบจำลองภาษาชื่อ CCphn.net
- -n i N เป็นการกำหนดให้เครื่องรู้จำทำงานโดยให้ผลลัพธ์ที่มีค่าความน่าจะเป็นสูงสุด N อันดับแรกออกมา
- -l '*' -i recout.mlf เป็นการกำหนดเก็บผลลัพธ์การรู้จำไว้ที่ไฟล์ recout.mlf โดยไฟล์นี้จะมีรูปแบบเนื้อความเหมือนกับไฟล์กำกับหน่วยเสียงสำคัญ
- CCphn.dict คือไฟล์พจนานุกรมหน่วยเสียง
- CCphn.list คือไฟล์รายการหน่วยเสียง

3. การพิจารณาขอบเขตของหน่วยเสียงจากผลการรู้จำเสียงพูด

การตรวจหาขอบเขตของหน่วยเสียงแบบอาศัยเครื่องรู้จำเสียงพูด จะอาศัยผลที่ได้จากการรู้จำเสียงพูด โดยเลือกเฉพาะข้อมูลส่วนที่เป็นขอบเขตทางเวลาของหน่วยเสียงออกมาเป็นขอบเขตของหน่วยเสียง ซึ่งขอบเขตของหน่วยเสียงนี้จะมีความละเอียดอยู่ในระดับเดียวกันกับระยะห่างของแต่ละกรอบเวลาที่กำหนดเป็นพารามิเตอร์ไว้ในขั้นตอนการสกัดลักษณะสำคัญเพื่อการเรียนรู้และการรู้จำ โดยในที่นี้ความละเอียดจะอยู่ที่ 10 มิลลิวินาที

ตัวอย่างการตรวจหาขอบเขตของหน่วยเสียง เช่น เมื่อเนื้อความในไฟล์ผลการรู้จำมีลักษณะเป็นดังนี้

```
"*/CCF003_Pa003_012.rec"
0 2100000 sil -1024.688843
2100000 3100000 thr -780.772339
3100000 3800000 l -575.756409
3800000 5300000 x -904.084290
5300000 6100000 b -618.631958
6100000 7000000 a -599.543152
7000000 8100000 ng^ -784.267822
8100000 8800000 sp -459.062744
8800000 9300000 p -439.476685
9300000 13600000 ii -2227.795166
13600000 15700000 sp -1454.947998
15700000 17600000 f -1381.702393
17600000 18300000 o -565.726929
18300000 19900000 n^ -1211.035156
19900000 20700000 k -639.093384
20700000 21300000 a -452.169922
21300000 22200000 m -695.446106
22200000 23900000 xx -1044.781738
23900000 24700000 l -573.536621
24700000 28800000 aa -1923.821899
28800000 29600000 f^ -599.704895
29600000 32200000 sil -1300.840332
///
0 2100000 sil -1024.688843
2100000 3100000 thr -780.772339
3100000 3800000 l -575.756409
3800000 5300000 x -904.084290
5300000 6100000 b -618.631958
6100000 7000000 a -599.543152
7000000 8100000 ng^ -784.267822
8100000 8800000 sp -459.062744
8800000 9300000 p -439.476685
9300000 13600000 ii -2227.795166
13600000 15700000 sp -1454.947998
15700000 17600000 f -1381.702393
17600000 18300000 o -565.726929
18300000 19900000 n^ -1211.035156
19900000 20700000 k -639.093384
20700000 21300000 a -452.169922
21300000 22100000 m -617.664978
22100000 23900000 aa -1122.411133
23900000 24700000 l -575.246582
24700000 28800000 aa -1923.821899
28800000 29600000 f^ -599.704895
29600000 32200000 sil -1300.840332
.
```

จะได้ลำดับของขอบเขตของหน่วยเสียง $L = \{0, 21, 31, 38, 53, 61, 70, 81, 88, 93, 136, 157, 176, 183, 199, 207, 213, 221, 222, 239, 247, 288, 296, 322\}$ โดยที่ขอบเขตของหน่วยเสียงแต่ละอันจะเป็นตัวเลขบอกตำแหน่งทางเวลาซึ่งอยู่ในหน่วย 10 มิลลิวินาที ตัวอย่างเช่นขอบเขตของหน่วยเสียงที่สามของ L แทนด้วยเลข 31 จะหมายความว่าขอบเขตของหน่วยเสียงนั้นอยู่ที่เวลา 310 มิลลิวินาทีของสัญญาณเสียง

การสกัดลักษณะสำคัญที่ได้จากการใช้สารสนเทศสวนสัทศาสตร์

การคำนวณหาลักษณะสำคัญที่ได้จากการใช้สารสนเทศสวนสัทศาสตร์ในตารางที่ 3.1 จะใช้โปรแกรม Speech Filling System – SFS [33] ซึ่งเป็นเครื่องมือสำหรับวิเคราะห์และสกัดลักษณะสำคัญจากสัญญาณเสียง การสกัดค่าพลังงานในช่วงความถี่ต่างๆจะใช้คำสั่ง mktrack โดยเรียกใช้ด้วยการพิมพ์คำสั่งดังนี้

```
mktrack -l lowfreq -h highfreq -n 4 -r 100 file
```

โดยที่

- -l lowfreq –h highfreq เป็นการกำหนดให้หาค่าพลังงานในช่วงความถี่ตั้งแต่ lowfreq ไปจนถึง highfreq (หน่วยเฮิร์ต)
- -n 4 เป็นการกำหนดให้ใช้จำนวนอันดับตัวกรองสัญญาณทั้งสิ้น 4 อันดับ
- -r 100 เป็นการกำหนดอัตราสุ่มของค่าพลังงานเป็น 100 เฮิร์ต
- file เป็นการกำหนดไฟล์อินพุตที่ต้องการหาค่าพลังงาน

การสกัดค่าระดับการสั่นสะเทือนเส้นเสียงหรือความก้องของเสียงจะอาศัยค่าการกระจายของพลังงาน (Energy distribution) อัตราการตัดศูนย์ (Zero-crossing rate) และอัตสหสัมพันธ์ (Autocorrelation) มาประมาณระดับความก้องของสัญญาณเสียงโดยเรียกใช้คำสั่ง vdegree ด้วยการพิมพ์คำสั่งดังนี้

```
vdegree file
```

โดยที่

- file เป็นการกำหนดไฟล์อินพุตที่ต้องการหาค่าพลังงาน

การสกัดค่าระดับของภาวะความไม่เป็นคาบของจะใช้คำสั่ง noisanal มาประมาณระดับของภาวะความไม่เป็นคาบในสัญญาณเสียง โดยเรียกใช้ด้วยการพิมพ์คำสั่งดังนี้

```
noisanal -w 0.025file
```

โดยที่

- -w windowsize เป็นการกำหนดขนาดของกรอบเวลาที่ทำการวิเคราะห์ให้เป็น 25 มิลลิวินาที

การเรียนรู้และการจำแนกข้อมูลด้วยซัพพอร์ตเวกเตอร์แมชชีน

ในงานวิทยานิพนธ์นี้เลือกใช้เอสวีเอ็มไลท์ SVM^{light} [34] มาเป็นเครื่องมือสำหรับสร้างตัวแบ่งแยกเอสวีเอ็ม โดยรายละเอียดเกี่ยวกับใช้งานเครื่องมือเพื่อการเรียนรู้และการจำแนกข้อมูลมีดังนี้

1. การเรียนรู้ของซัพพอร์เวกเตอร์แมชชีน

ข้อมูลที่ต้องเตรียมเพื่อนำไปใช้ในการเรียนรู้ในที่นี้ จะอาศัยไฟล์ลักษณะสำคัญที่ได้จากขั้นตอนการสกัดลักษณะสำคัญของเสียงในบทที่ 4 มาสร้างไฟล์เวกเตอร์ลักษณะสำคัญเพื่อการเรียนรู้ของเอสวีเอ็ม โดยไฟล์เวกเตอร์ลักษณะสำคัญจะเป็นไฟล์ข้อความที่แต่ละบรรทัดแทนลักษณะสำคัญของแต่ละกรอบเวลา ซึ่งแต่ละบรรทัดแสดงด้วยรูปแบบดังนี้

```
<target> <feature>:<value> <feature>:<value> ... <feature>:<value>
```

โดยที่

- <target> คือฉลากกำกับเวกเตอร์ลักษณะสำคัญโดยในที่นี้คือ +1 หรือ -1
- <feature> คือตัวเลขจำนวนเต็มสำหรับบอกลำดับของลักษณะสำคัญ
- <value> คือค่าตัวเลขจำนวนจริงของลักษณะสำคัญ

ตัวอย่างเช่น

```
-1 1:0.432 :0.12  3:0.2
```

แทนเวกเตอร์ที่มีฉลากกำกับเป็น -1 โดยเป็นเวกเตอร์ 3 มิติที่มีลักษณะสำคัญลำดับที่หนึ่งเป็น 0.43 ลักษณะสำคัญลำดับที่สองเป็น 0.12 และลำดับสุดท้ายเป็น 0.2

โดยในที่นี้เลือกเรียนรู้เอสวีเอ็มสองแบบคือแบบที่ใช้ฟังก์ชันเคอร์เนลเชิงเส้น และแบบที่ใช้ฟังก์ชันเคอร์เนลพหุนาม โดยใช้โปรแกรม svm_learn ที่มีมากับชุดเครื่องมือ SVM^{light}

การใช้งานโปรแกรม svm_learn จะต้องเรียกใช้ด้วยการพิมพ์คำสั่งดังนี้

```
svm_learn -t 0 train_sonorant.tok linear_sonorant.svm
```

โดยที่

- -t 0 เป็นการกำหนดฟังก์ชันเคอร์เนลที่ต้องการใช้งานเป็นแบบเชิงเส้น ถ้าต้องการใช้ฟังก์ชันเคอร์เนลแบบพหุนามให้กำหนดเป็น 1
- train_sonorant.tok เป็นไฟล์เวกเตอร์ลักษณะสำคัญที่นำมาใช้เป็นตัวอย่างข้อมูลเพื่อการเรียนรู้เอสวีเอ็ม
- linear_sonorant.svm คือไฟล์แบบจำลองที่เป็นเอาท์พุต

2. การจำแนกข้อมูลด้วยซัพพอร์เวกเตอร์แมชชีน

การจำแนกข้อมูลด้วยซัพพอร์เวกเตอร์แมชชีน จะใช้โปรแกรม svm_classify ซึ่งมีอยู่ในชุดเครื่องมือ SVM^{light} โดยการใช้งานโปรแกรม svm_classify จะต้องเรียกใช้ด้วยการพิมพ์คำสั่งดังนี้

```
svm_classify example_file model_file output_file
```

โดยที่

- example_file คือไฟล์เวกเตอร์ลักษณะสำคัญที่ต้องการทดสอบ
- model_file คือไฟล์แบบจำลองเอสวีเอ็มที่ต้องการนำมาใช้แบ่งแยกข้อมูล
- output_file คือไฟล์ผลลัพธ์ค่าของฟังก์ชันตัดสินใจที่เป็นเอาท์พุต

ประวัติผู้เขียนวิทยานิพนธ์

นายไพโรจน์ ลีลาภัทรกิจ เกิดเมื่อวันที่ 27 มกราคม พ.ศ. 2527 ที่จังหวัดกรุงเทพฯ สำเร็จการศึกษาระดับมัธยมศึกษาตอนต้นและตอนปลายจากโรงเรียนเทพศิรินทร์ สำเร็จการศึกษาระดับปริญญาบัณฑิต ในสาขาวิชาวิศวกรรมคอมพิวเตอร์ จากคณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัยในปีการศึกษา 2548