รายการอ้างอิง

ภาษาไทย

ทวี ปทุมทาน. การตรวจรู้จำเสียงพูดภาษาไทยโดยใช้หน่วยพยางค์ สาขาวิศวกรรมคอมพิวเตอร์ จุฬาลงกรณ์มหาวิทยาลัย. วิทยานิพนธ์ปริญญามหาบัณฑิต จุฬาลงกรณ์มหาวิทยาลัย, 2529.

ไพศาล ธรรมโพธิ์ทอง. ระบบการรู้จำเสียงพูดแบบต่างบุคคล สาขาวิทยาศาสตร์คอมพิวเตอร์ จุฬาลงกรณ์มหาวิทยาลัย. วิทยานิพนธ์ปริญญามหาบัณฑิต จุฬาลงกรณ์มหาวิทยาลัย, 2532.

ระพีพัฒน์ เพ็ญศิริ. การรู้จำเสียงตัวเลขภาษาไทยโดยไม่ขึ้นต่อผู้พูดโดยการใช้ไดนามิคไทม์วาร์ปปิง สาขาวิศวกรรมไฟฟ้าสื่อสาร จุฬาลงกรณ์มหาวิทยาลัย. หัวข้อวิทยานิพนธ์ปริญญามหาบัณฑิต จุฬาลงกรณ์มหาวิทยาลัย, 2537.

สุเธียร เกียรติสุนทร. การประมาณพันธะเชิงเส้นเสียงพูด (LINEAR PREDICTION OF SPEECH) สาขาวิศวกรรมไฟฟ้า สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหาร ลาดกระบัง. วิทยานิพนธ์ปริญญามหาบัณฑิต สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหาร ลาดกระบัง, 2525.

ภาษาอังกฤษ

Forney, G. David. The Viterbi Algorithm. Proceeding of the IEEE 61 (March 1973) : 268-278.

Furui, Sadaoki. Digital speech processing, systhysis and recognition. New York and Basel : Marcel Dekker, 1989.

Kim, H. R., and Lee, H. S. Postprocessor using Fuzzy Vector Quantizer in HMM-Based speech recognition. Electronics Letter 24th 27 (October 1991).

Koo, J. M., and UN, C. K, Fuzzy smoothing of HMM parameters in speech recognition. Electronics Letter 24th 26 (May 1990): 734-744.

Lamel, L. F., Rabiner, L. R., Rosenberg, A. E., and Wilpon, J. G. An improved endpoint detector for isolated word recognition. IEEE Transaction on Speech and Signal Processing ASSP-29 (Aug. 1981) : 777-785.

Lee, Kai-Fu., and Hon, Hsiao-Wuen. Speaker-independent phone recognition using Hidden Markov Models. IEEE Transaction on Acoustics, Speech and Signal Processing 37 (Nov. 1989): 1641-1648.

Levinson, S. E., Rabiner, L. R., and Sondhi, M. M. An introduction to the application of the theory of probability functions of a Markov process to automatic speech recognition. The Bell System Technical Journal 62 (April 1983).

Picone, Joseph. Cotinuous speech recognition using Hidden Markov Model. IEEE ASSP Magazine (July 1990): 26-40.

Rabiner, L. R., and Levinson, S.E. Isolated and connected word recognition - theory and selected application," IEEE Transaction on Communication COM-29 (May 1981) : 621-659.

_____ , L. R., Levinson, S. E., and Sondhi, M. M. On the application of Vector Quantization and Hidden Markov Model to speaker independent isolated word recognition. The Bell System Technical Journal 62 (April 1983) : 1075-1106.

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก ก.

เป็นการหาค่าพารามิเตอร์ที่เหมาะสมเพื่อใช้ในการทดลอง

1. จำนวน order ของ LPC

มีการวิจัยการประมาณเชิงเส้นเสียงพูด (สุเธียร เกียรติสุนทร 2525) ซึ่งเป็นงานวิจัยที่มีการนำค่าสัมประสิทธิ์การสะท้อนที่หาโดยการใช้วิธีอัตสัมพันธ์ มาทำการสังเคราะห์เสียงขึ้นมาใหม่ จากการทดสอบดังกล่าว ให้ผลเป็นที่น่าพอใจ งานวิจัยชิ้นนั้นมีการใช้ LPC ด้วยจำนวน order 10 ซึ่งเพียงพอสำหรับการหาสัมประสิทธิ์การสะท้อนเพื่อการสังเคราะห์เสียงที่ต้องใช้ความละเอียดมาก ในวิทยานิพนธ์นี้ได้นำ LPC มาใช้ในการทดลองเช่นเดียวกัน จำนวน order ที่ใช้คือ 10 วิทยานิพนธ์นี้เป็นการรู้จำเสียงเท่านั้น ย่อมต้องการความละเอียดในการวิเคราะห์น้อยกว่าการสังเคราะห์เสียง ดังนั้น order 10 จึงเพียงพอสำหรับการรู้จำเสียง

ในรูป ก.1 เป็นการเปรียบเทียบค่า residual ของสัญญาณ เมื่อใช้ order ในการหาสัมประสิทธิ์การสะท้อนต่าง ๆ กัน โดยใช้ order 10, 12, 14 ถ้าใช้ order ยิ่งมากความละเอียดในการวิเคราะห์จะมากไปด้วย สัญญาณที่ใช้วิเคราะห์นี้เป็นสัญญาณเสียงจริงที่ใช้ในการรู้จำ ซึ่งนำมาวิเคราะห์เพียง 1 เฟรมเพื่อให้เห็นภาพชัดเจนยิ่งขึ้น

<u>ขั้นตอนในการวิเคราะห์</u>

1. นำเสียง 1 เสียงมาผ่านขั้นตอนการตัดหัวท้ายคำ หลังจากนั้นนำมาหาค่าสัมประสิทธิ์การสะท้อนตามทฤษฎีในบทที่ 2 มีขั้นตอนตามบทที่ 3

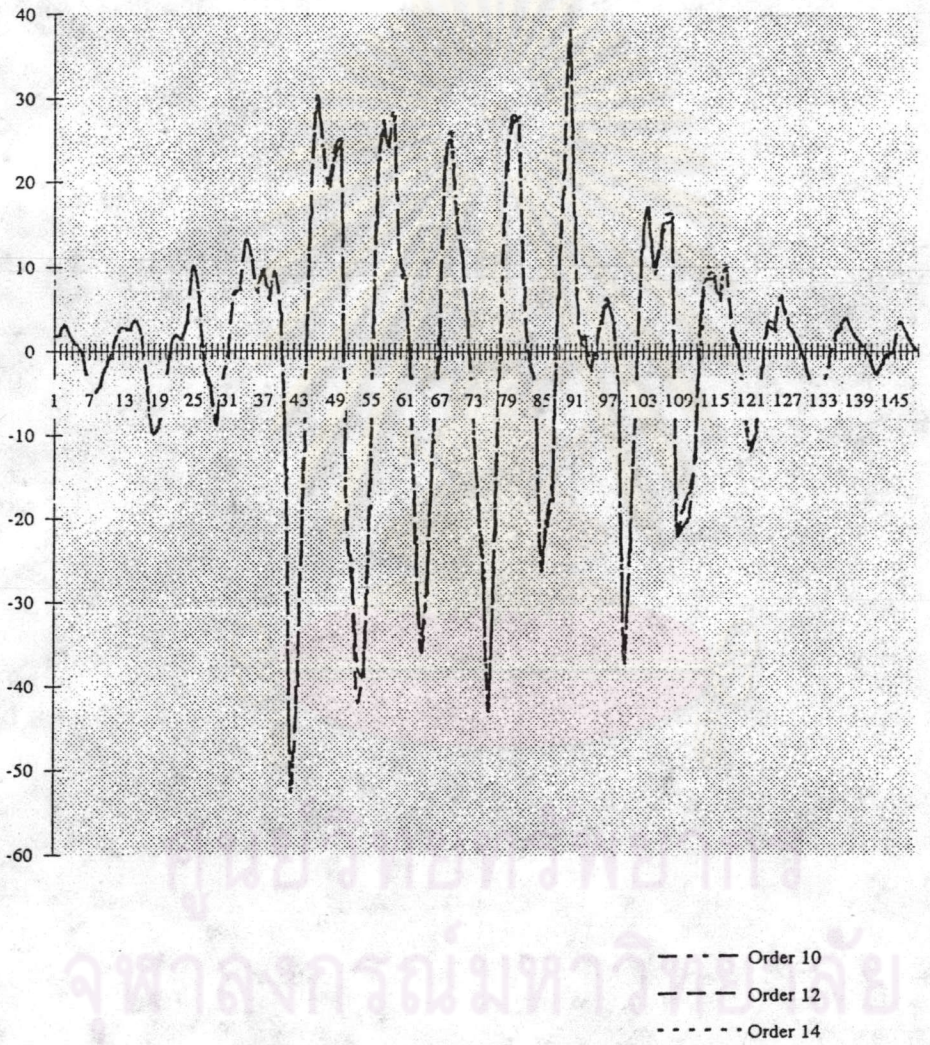
2. เมื่อผ่านขั้นตอนการหาสัมประสิทธิ์การสะท้อน จะได้ค่าสัมประสิทธิ์การสะท้อน 1 ชุด/ 1 เฟรม

3. ใช้สัมประสิทธิ์การสะท้อนที่ได้จากข้อ 2 มาสร้างสัญญาณเสียงใหม่ขึ้นมา

4. เปรียบเทียบสัญญาณเสียงที่ได้จากการคำนวณกับสัญญาณเสียงจริง ที่ใช้หาสัมประสิทธิ์การสะท้อน

5. เปลี่ยน order ในการพิจารณาโดยใช้ order 10, 12, 14

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย



รูป ก. 1 เปรียบเทียบค่า residual ของสัญญาณ

2. ขนาดของ codebook

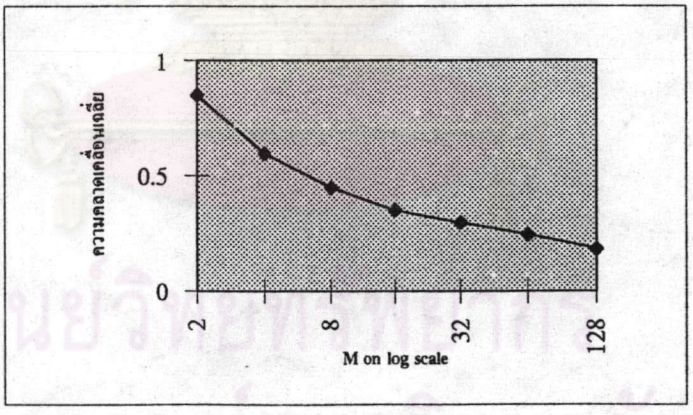
ในการทดสอบหาค่าขนาดของ codebook ที่เหมาะสม (Rabiner, Levinson, and Sondhi, 1982) เป็นการทดสอบโดยใช้เวกเตอร์ของ LPC จำนวน 39708 เวกเตอร์ จากคน 100 คน(ชาย 50 หญิง 50) พูดตัวเลขภาษาอังกฤษ 0-9 จากนั้นนำเวกเตอร์มาหาค่าควอนไตซ์ที่ codebook ขนาดต่าง ๆ M = 2, 4, 8, 16, 32, 64, 128 ใช้การพิจารณาดังนี้

ค่าความคลาดเคลื่อนเฉลี่ย $\|D_m\|$

$$\|D_m\| = \min_{\hat{a}_m} \left\{ \frac{1}{I} \sum_{i=1}^{I} \min_{1 \le m \le M} \left[d(\hat{a}_m, a_i) \right] \right\} \qquad \text{(ก.1)}$$

โดย a_i เป็นเวกเตอร์ LPC i = 1, 2, ... , I $\hat{a}_m$ เป็น codebook LPC เวกเตอร์ m = 1, 2, ... , M

จากผลการทดสอบจะได้ดังรูป ก.2 โดยรูป ก.2 เป็นการ plot ค่า $\|D_m\|$ กับค่า M(ขนาด codebook โดยใช้ log scale) โดย M มีค่า 2-128 เราจะเห็นว่าที่ M ≥ 32 ค่าความคลาดเคลื่อนเฉลี่ยจะต่ำกว่า 0.3 M = 64 จะมีค่า $\|D_m\|$ ประมาณ 0.2 ซึ่งค่า $\|D_m\|$ < 0.3 เป็นค่าความคลาดเคลื่อนที่น้อยมากสำหรับ vector quantization (Levinson, Rabiner, et al., 1979) ดังนั้นสำหรับวิธีของ HMM ควรใช้ M = 64 เพราะจะมี ค่าความคลาดเคลื่อนเฉลี่ยน้อย เมื่อ M = 64 - 128 (Rabiner, Levison, and Sondhi, 1982)



รูป ก. 2 เปรียบเทียบค่าความคลาดเคลื่อนเฉลี่ยกับขนาด codebook

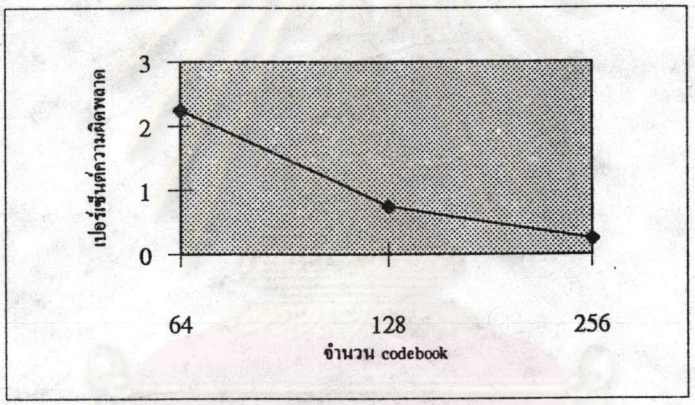
จากผลการทดสอบโดยใช้โปรแกรมที่พัฒนาขึ้น นำข้อมูลจากกลุ่มต้นแบบกลุ่มที่ 1 และกลุ่มที่ 2 จำนวน 20 คน มาหาโมเดลต้นแบบและทำการทดสอบความถูกต้องในการรู้จำ

1. ใช้กลุ่มที่ 1 และกลุ่มที่ 2 ทดสอบกับโมเดลต้นแบบที่คำนวณได้ เปรียบเทียบกับขนาด codebook ดังตารางที่ ก.1 และกราฟรูปที่ ก.3

ตาราง ก. 1 เปรียบเทียบความผิดพลาดเป็นร้อยละกับขนาด codebook ของกลุ่มต้นแบบ

จำนวน codebook	ความผิดพลาด (%)
64	2.25
128	0.75
256	0.25

รูปที่ ก.3 เปรียบเทียบความผิดพลาดเป็นร้อยละกับขนาด codebook ของกลุ่มต้นแบบ

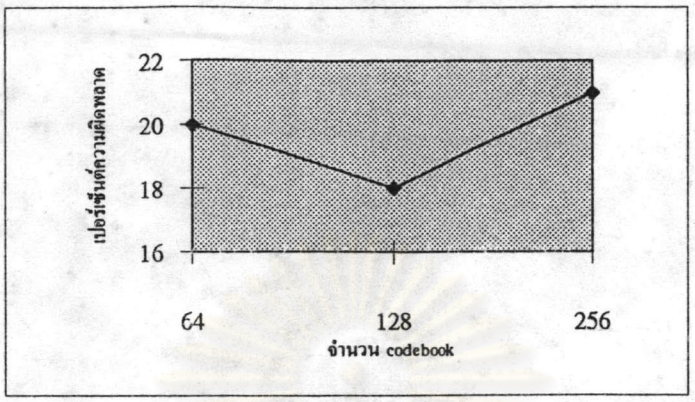


2. ใช้กลุ่มทดสอบมาทดสอบกับโมเดลต้นแบบที่คำนวณได้ ซึ่งเป็นบุคคลที่ไม่มีการฝึกฝน ดังแสดงผลความผิดพลาดในการรู้จำที่ขนาด codebook ต่างๆ ดังตารางที่ ก.2 และรูปที่ ก.4

ตารางที่ ก.2 เปรียบเทียบความผิดพลาดเป็นร้อยละกับขนาด codebook ของกลุ่มทดสอบ

จำนวน codebook	ความผิดพลาด (%)
64	20
128	18
256	21

รูปที่ ก.4 เปรียบเทียบความผิดพลาดเป็นร้อยละกับขนาด codebook ของกลุ่มทดสอบ

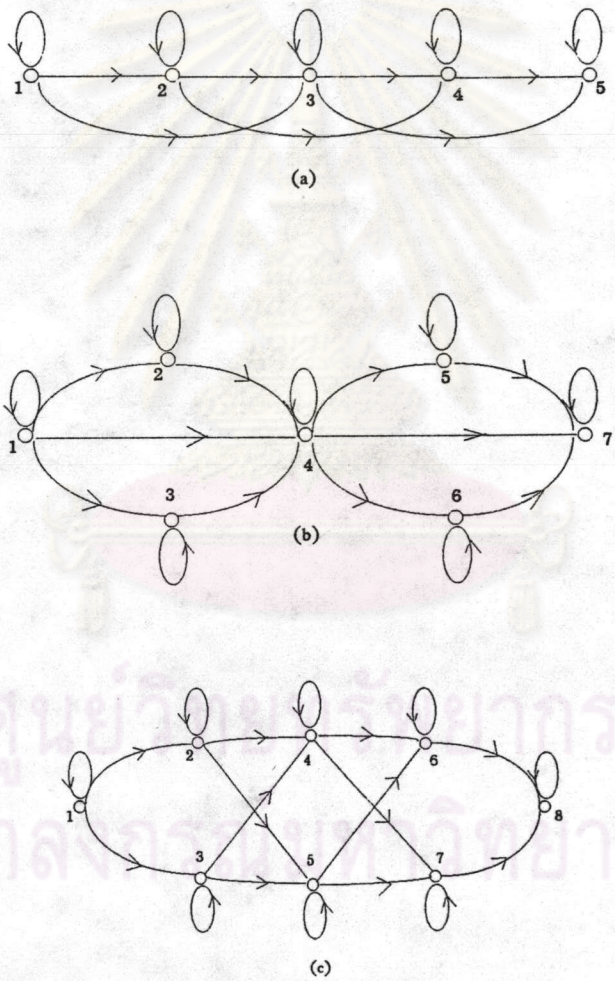


ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

3. โครงสร้างของ HMM (Hidden Markov Model)

เป็นที่ทราบกันแล้วว่ามีพารามิเตอร์ต่างๆ ที่มีผลต่อค่าความถูกต้องในการวิเคราะห์ รูปแบบโครงสร้างของ HMM ก็เป็นอีกส่วนหนึ่งที่มีผลกระทบสำคัญ

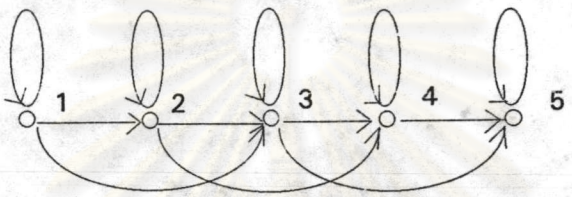
จากรูป ก.5 ได้แสดงโครงสร้างของ HMM แบบต่างๆ รูป ก.5 (a) เป็น contrained serial HMM ที่มี 5 state รูป ก.5 (b) เป็น constrained parallel ที่มี 7 state รูป ก.5 (c) เป็น constrained parallel ที่มี 8 state ใช้โครงสร้างทั้ง 3 โครงสร้างในการหาโมเดลของแต่ละเสียง โดยใช้เซ็ตทดสอบเซ็ตเดียวกันทดสอบทั้ง 3 โครงสร้าง ผลที่ได้คือมีค่าความคลาดเคลื่อนใกล้เคียงกันทั้ง 3 โครงสร้าง (3.5 % $\pm$0.1%) จึงสรุปได้ว่าโครงสร้างแบบ serial ก็เพียงพอในการพิจารณา (Rabiner, Levison, and Sondhi, 1982)



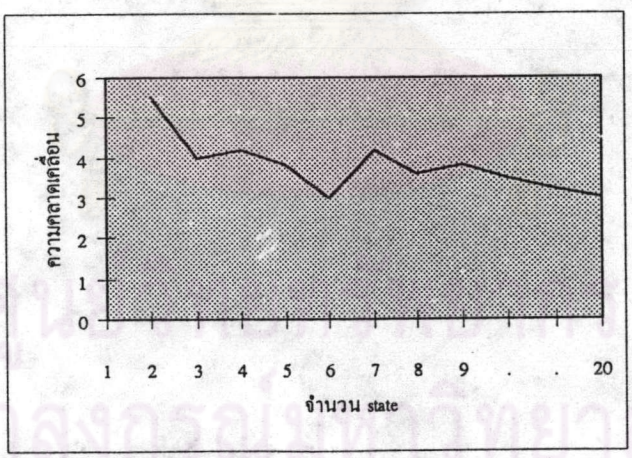
รูป ก.5 โครงสร้างที่ทดสอบ

4. จำนวน state

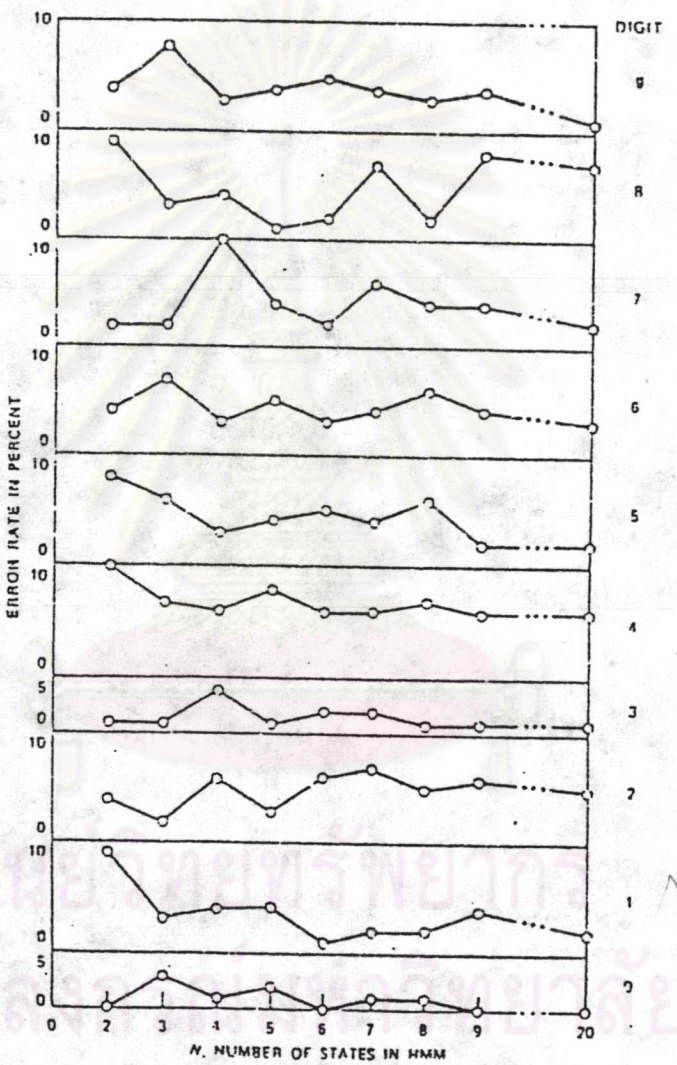
จากรูป ก.6 เป็นโครงสร้างแบบ serial ที่มี 5 state ที่จะใช้พิจารณา นำโครงสร้างนี้มาหาโมเดล โดยใช้จำนวน state ตั้งแต่ 2 - 9 และ 20 โดยจะทดสอบด้วยเซ็ตทดสอบเดียวกัน ผลการทดสอบเป็นดังรูป ก.7 และ รูป ก.8 เป็นการแสดงค่าความคลาดเคลื่อนของแต่ละตัวเลขภาษาอังกฤษและค่าจำนวน state จากรูป ก.7 และ รูป ก.8 สามารถกล่าวได้ว่าความสัมพันธ์ของความถูกต้องในการวิเคราะห์ จำนวนเสียงที่รู้จำ และจำนวน state ไม่ใช่ความสัมพันธ์ที่ชัดเจน (Rabiner, Levison, and Sondhi, 1982)



รูป ก.6 โครงสร้างแบบ serial



รูป ก.7 เปรียบเทียบความคลาดเคลื่อนกับจำนวน state



รูป ก.8 เปรียบเทียบความคลาดเคลื่อนของแต่ละเสียงตัวเลข(ภาษาอังกฤษ)

จากรูป ก.6 จะนำค่า state ที่เท่ากับ 3 มาใช้เนื่องจากเป็นจุดที่เริ่มจะมีค่าความคลาดเคลื่อนต่ำ ดังนั้นโครงสร้างของ HMM ที่ใช้ในวิทยานิพนธ์นี้จึงเป็นดังรูป 3.6 ในบทที่ 3

ภาคผนวก ข.

เป็นการนำผลการรู้จำเสียงตัวเลขที่มีการทดสอบโดยวิธีต่าง ๆ ที่มีการวิจัยการรู้จำเสียงตัวเลขภาษาอังกฤษ

จากตาราง ข.1 เป็นการเปรียบเทียบเปอร์เซ็นต์ความถูกต้องของการทดสอบเสียง 0-9 ของบุคคล 100 คน พูดคนละครั้งโดยผู้พูดที่นำมาทดสอบเป็นบุคคลเดียวกับที่นำมาหาโมเดล แต่พูดห่างกันหลายวัน โดยวิธีที่นำมาเปรียบเทียบมี 3 วิธี คือ HMM/VQ, LPC/DTW และ LPC/DTW/VQ (Rabiner, Levison, and Sondhi, 1982)

	RECOGNIZER		
DIGIT	LPC/HMM/VQ	LPC/DTW	LPC/DTW/VQ
0	98	99	99
1	98	98	99
2	96	100	96
3	99	99	97
4	93	97	96
5	97	96	93
6	96	100	94
7	99	100	94
8	92	98	96
9	95	98	96
AVARAGE	96.3	98.5	96.5

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

จากตาราง ข.2 เป็นการเปรียบเทียบเปอร์เซ็นต์ความถูกต้องของการทดสอบเสียง 0-9 ของ บุคคล 10 คน พูดคนละ 20 ครั้งต่อ 1 เสียงโดยผู้พูดที่นำมาทดสอบไม่ใช่บุคคลเดียวกับที่นำมาหาโมเดล โดยวิธี ที่นำมาเปรียบเทียบมี 3 วิธี คือ HMM/VQ, LPC/DTW และ LPC/DTW/VQ (Rabiner, Levison, and Sondhi, 1982)

	RECOGNIZER		
TALKER	LPC/HMM/VQ	LPC/DTW	LPC/DTW/VQ
1	74.5	96	87
2	99.5	100	99.5
3	94	99	97
4	89	99	91
5	95.5	100	100
6	95.5	99	99.5
7	100	100	99.5
8	91.5	96	93
9	91.5	99.5	93.5
10	96.5	98.5	98
AVARAGE	92.8	98.7	95.5

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ประวัติผู้เขียน

นางสาว เสาวลักษณ์ อารีย์พงศา เกิดเมื่อวันที่ 20 มีนาคม พ.ศ. 2514 ที่กรุงเทพมหานคร จบการศึกษาระดับปริญญาวิศวกรรมศาสตร์บัณฑิต สาขาวิชาวิศวกรรมอิเล็กทรอนิกส์ ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์ สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหาร ลาดกระบัง ปีพ.ศ. 2536 เข้าศึกษาต่อในหลักสูตรปริญญาวิศวกรรมศาสตร์มหาบัณฑิต สาขาวิชาวิศวกรรมไฟฟ้า ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย เมื่อปีพ.ศ. 2536

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย