# รายการอ้างอิง


ภาษาไทย

ชัยศรี เอี่ยมอำไพ, การตรวจหาจุดเริ่มต้นและจุดสิ้นสุดของคำโดดๆ. สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง, ม.ป.ท.

พิณทิพย์ ทวยเจริญ, สัทศาสตร์และสัทศาสตร์ปฏิบัติ. 2000 เล่ม, พิมพ์ครั้งที่ 1, คณะศิลปศาสตร์ มหาวิทยาลัยธรรมศาสตร์ กรุงเทพมหานคร : โรงพิมพ์มหาวิทยาลัยธรรมศาสตร์, 2533

ไพศาล ธรรมโพธิทอง, ระบบการรับรู้เสียงพูดแบบต่างบุคคล. วิทยานิพนธ์ปริญญามหาบัณฑิต จุฬาลงกรณ์มหาวิทยาลัย, 2533.

สุนิสา จันทวีกูล, การวิเคราะห์สเปสกตรัมกำลังของสัญญาณโดยใช้ดีสครีตฮาร์ดเลย์ทรานส์ฟอร์ม. วิทยานิพนธ์ปริญญามหาบัณฑิต จุฬาลงกรณ์มหาวิทยาลัย, 2536.

ภาษาอังกฤษ

Bocchieri, E. L., and Doddington, G. R., "Frame-specific statistical features for speaker independents speech recognition", IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-34, No. 4, pp 755-764, Aug. 1986.

Dante, H. M., and Sarma, V. V. S., "Autcmatic speaker identification for a large population" , IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-27, No. 3, pp 255-263, June 1979.

Das, S. K., "Some experiments in Discrete Utterance Recognition", IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-30, No. 5, pp 766-770, Oct. 1982.

Furui, S., Advances in speech signal processing. Marcel Dekker, New York, 1989.

_____ ., "A Training procedure for isolated word recognition systems" , IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-28, No. 2, pp 129-136 , Apr. 1980.

_____ ., and Sondhi M. M., Digital speech processing synthesis and recognition , Marcel Dekker, New York, 1991.

Lamel, L. F., Rabiner, L. R., Rosenberg, A. E., and Wilpon, J. G., "An improved endpoint detector for isolated word recognition", IEEE Trans. Acoust., Speech, Signal Processing. vol. ASSP-29, No. 4, pp 777-785, Aug. 1981

Markel, J. D., Oshika, B. T., and Gray, A. H., "Long-term feature averaging for speaker recognition", IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-25 , No. 4, pp 330-337, Aug. 1977.

Myers, C., Rabiner, L. R., and Rosenberg, A. E., "Performance tradeoffs in dynamic time warping algorithms for isolated word recognition", IEEE Trans.Acoust.,Speech, Signal Processing, vol. ASSP-28, No. 6, pp 623-635, Dec. 1980.

Rabiner, L. R., and Levinson, S. E., "Isolated and connected word recognition-theory and selected applications", IEEE Trans. Comm.,vol. COM-29, No. 5, pp 621-659, May 1981.

_____ ., RosenBerg, A. E., and Levinson, S. E., "Considerations in dynamic time warping algorithms for discrete word recognition", IEEE Trans. Acoust.,Speech, Signal Processing, vol. ASSP-26, No. 6 , pp 575-582, Dec. 1978.

_____ ., and Sambur, M. R., "Some Preliminary Experiments in the Recognition of Connected Digits", IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-24, No. 2, pp 170-182, Apr. 1976.

_____ ., and Schmidt, C. E., "Application of dynamic time warping to connected digit recognition", IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-28, No. 4, pp 377-388, Aug. 1980.

Silverman,H. F., and Dixon N. R., "A comparison of several speech-spectra classification methods", IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-24, No. 4, pp 289-295, Aug. 1976.

Sound Blaster User Reference Manual. Creative Labs, Inc. 1989.

The Developer Kit for Sound Blaster Series. Creative Labs, Inc. 1991

Turbo C User's guide and Reference guide. version 2.0, Borland Internation, Inc

Ukita, T., Saito E., Nitta, T., and Watanabe, S., "A speaker-independent connected digit recognition system concatenating statistically discriminated words", IEEE Trans. Signal Proessing, vol.40, No. 40, pp 2414-2424 , Oct. 1992.

Warren, J. H., "A pattern classification technique for speech recognition", IEEE Trans., Audio. Electroacoustics, vol. AU-19, No. 4. pp 281-285, Dec. 1971.

White, G. M., and Neely, R. B., "Speech recognition experiments with linear predication, bandpass filtering, and dynamic programming", IEEE Trans. Acoust., Speech. Signal Processing, vol. ASSP-24, No. 2, pp 183-188, Apr. 1976.

Zelinski, R., and Class, F., "A segmentation algorithm for connected word recognition based on Estimation Principles", IEEE Trans. Acoust., Speech. Signal Processing, vol. ASSP-31, No. 4, pp 818-827 , Aug. 1983.

ภาคผนวก ก

## การวิเคราะห์หาพารามิเตอร์ที่เหมาะสม

1. ค่าของพารามิเตอร์ที่เหมาะสมในการตัดคำ

ตารางที่ ก.1 แสดงผลของพารามิเตอร์ a, b, และ c ของการตัดคำที่มีผลต่อการรู้จำ โดยแบบอ้างอิงจำนวน 20 คน แบบทดสอบจำนวน 20 คน

| พารามิเตอร์ทดสอบ (c=0.5) | | ความถูกต้องของการรู้จำของแบบทดสอบ B ต่อแบบอ้างอิง A1 (%) |
|---|---|---|
| a | b | |
| 0.1 | 0.1 | 69.75 |
| | 0.2 | 67.00 |
| | 0.3 | 64.75 |
| | 0.4 | 64.50 |
| 0.2 | 0.1 | 70.00 |
| | 0.2 | 68.75 |
| | 0.3 | 67.00 |
| | 0.4 | 65.25 |
| 0.3 | 0.1 | 76.75* |
| | 0.2 | 70.25 |
| | 0.3 | 67.00 |
| | 0.4 | 65.00 |
| 0.4 | 0.1 | 77.75* |
| | 0.2 | 69.50 |
| | 0.3 | 69.00 |
| | 0.4 | 65.75 |

ในการวิเคราะห์การดัดคำ จะมีพารามิเตอร์ 3 ตัว ดังในสมการที่ 2.2.13 คือ $k_1$ $k_2$ และ $k_3$ ซึ่งจะแทนระดับของพลังงานที่ใช้ในการอ้างอิงของการดัดคำ โดยจะกำหนดผ่าน ค่า a,b, และ c ตามลำดับ ในการทดสอบนี้จะกำหนดให้ค่าของ c มีค่าเท่ากับ 0.5 ซึ่งจะกำหนดค่าของ $k_3$ มีค่าเป็น 50 % ของระดับพลังงานสูงสุดของสัญญาณ ส่วนค่าของ a และ b นั้นจะกำหนดให้อยู่ในช่วง 0.1 ถึง 0.5 ซึ่งทำให้ค่าของ k1 และ k2 จะอยู่ในช่วง 10 % ถึง 40 % ของค่าระดับพลังงานสูงสุดของสัญญาณ โดยผลการทดสอบจะแสดงดังในตารางที่ ก.1 โดยใช้รูปแบบการกำหนดพารามิเตอร์แบบที่ 2 ดังที่กำหนดในตารางที่ 3.3.1

จากผลการทดสอบในตารางที่ ก.1 จะเห็นได้ว่าค่าของระดับพลังงานที่เหมาะสมในการกำหนดระดับอ้างอิงของการตัดคำ โดยในการดัดคำของส่วนหน้าของสัญญาณเสียง k1 ที่เหมาะสมจะอยู่ที่ประมาณ 30-40 % ของระดับพลังงานสูงสุดของสัญญาณ และระดับพลังงานอ้างอิงที่ใช้ในการตัดคำส่วนหลัง k2 จะอยู่ที่ 10 % ของระดับพลังงานสูงสุดของสัญญาณ จะทำให้ผลการของรู้จำมีค่าดีที่สุด

2. การเลือกรูปแบบของการกำหนดพารามิเตอร์ในการรู้จำ

จะเป็นการทดสอบหารูปแบบของการแทนพารามิเตอร์เพื่อใช้ในการรู้จำที่เหมาะสมที่สุด ซึ่งในการแทนรูปแบบของพารามิเตอร์นี้ จะกำหนดตามในตารางที่ 3.3.1 โดยใช้ผู้ทดสอบจำนวน 20 คน ให้กลุ่มอ้างอิงเป็น A1 และกลุ่มทดสอบเป็น A2,A3 พารามิเตอร์ของการตัดคำ a,b มีค่า 0.3 และ 0.1 ตามลำดับ ช่วงของลำดับพารามิเตอร์แสดงความถี่ 10-127 ผลการทดสอบจะแสดงดังในตารางที่ ก.2

ตารางที่ ก.2 แสดงผลการทดสอบการเลือกใช้รูปแบบพารามิเตอร์

| กลุ่มข้อมูล จำนวน 20 คน | ความถูกต้อง (%) | |
|---|---|---|
| | รูปแบบพารามิเตอร์ที่ใช้ในการรู้จำ | |
| | แบบที่ 1 | แบบที่ 2 |
| แบบอ้างอิง (A1) | 90.50 | 93.50 |
| แบบทดสอบ (A2,A3) | 82.00 | 85.50 |

จากตารางที่ ก.2 เปรียบเทียบการกำหนดค่าของพารามิเตอร์เพื่อใช้ในการรู้จำ จะพบว่าการเลือกใช้พารามิเตอร์รูปแบบที่ 2 จะช่วยเพิ่มความถูกต้องในการรู้จำขึ้น เนื่องจากในแบบแรกนั้นกำหนดระดับของแอมปลิจูดของสัญญาณมีค่ามาก จะให้ผลการรู้จำต่ำกว่าในแบบที่ 2 ซึ่งในแบบนี้จะกำหนดระดับของแอมปลิจูดของสัญญาณมีค่าต่ำกว่า ดังนั้นการกำหนดระดับของสัญญาณเพื่อสร้างพารามิเตอร์ที่เหมาะสมจะทำให้ระบบการรู้จำมีอัตราความถูกต้องเพิ่มขึ้น

3. การวิเคราะห์ช่วงความถี่ที่เหมาะสม

ในการวิเคราะห์นี้ พารามิเตอร์ที่ใช้ในการทดสอบจะมีจำนวนพารามิเตอร์ทั้งหมด 256 ตัว โดยจะมีลำดับของพารามิเตอร์ 0 ถึง 255 ซึ่งตัวเลขนี้จะแสดงถึงช่วงความถี่ในช่วง 0 ถึง 4 kHz ดังนั้นในการทดสอบนี้จะทำการหาช่วงของความถี่ที่เหมาะสม เพื่อให้ผลของการรู้จำที่ดี อีกทั้งจะช่วยลดเวลาของการคำนวณลง ดังในตารางที่ ก.2 จะแสดงผลของการรู้จำของกลุ่มทดสอบ B ต่อ กลุ่มอ้างอิง A1 ที่ช่วงความถี่ต่าง ๆ กัน

ตารางที่ ก.3 แสดงผลการรู้จำของแบบทดสอบ B ต่อแบบอ้างอิง A1 ในช่วงความถี่ต่าง ๆ

| ลำดับของพารามิเตอร์ | ความถูกต้องของการรู้จำของกลุ่มทดสอบ B ต่อกลุ่มอ้างอิง A1 (%) |
|---|---|
| 0 - 255 | 67.50 |
| 0 - 191 | 67.50 |
| 0 - 127 | 67.25 |
| 5 - 127 | 69.25 |
| 10 - 127 | 76.75 |
| 20 - 127 | 78.25 |
| 30 - 127 | 74.25 |
| 40 - 127 | 74.25 |
| 50 - 127 | 72.25 |
| 20 - 191 | 79.25* |
| 30 - 191 | 77.25 |
| 20 - 223 | 79.25 |

จากผลในตารางที่ ก.3 จะแสดงถึงผลของการเลือกกลุ่มของความถี่ที่นำมาใช้ในการวัดเพื่อหาค่าของ distance ซึ่งจะเห็นได้ว่า ในช่วงของลำดับพารามิเตอร์ที่ 20-191 ซึ่งแทนข้อมูลในช่วงของความถี่ 300 ถึง 3000 Hz ก็เพียงพอและให้ผลของการรู้จำดีที่สุด

4 การปรับขนาดของความยาวเสียงพูด

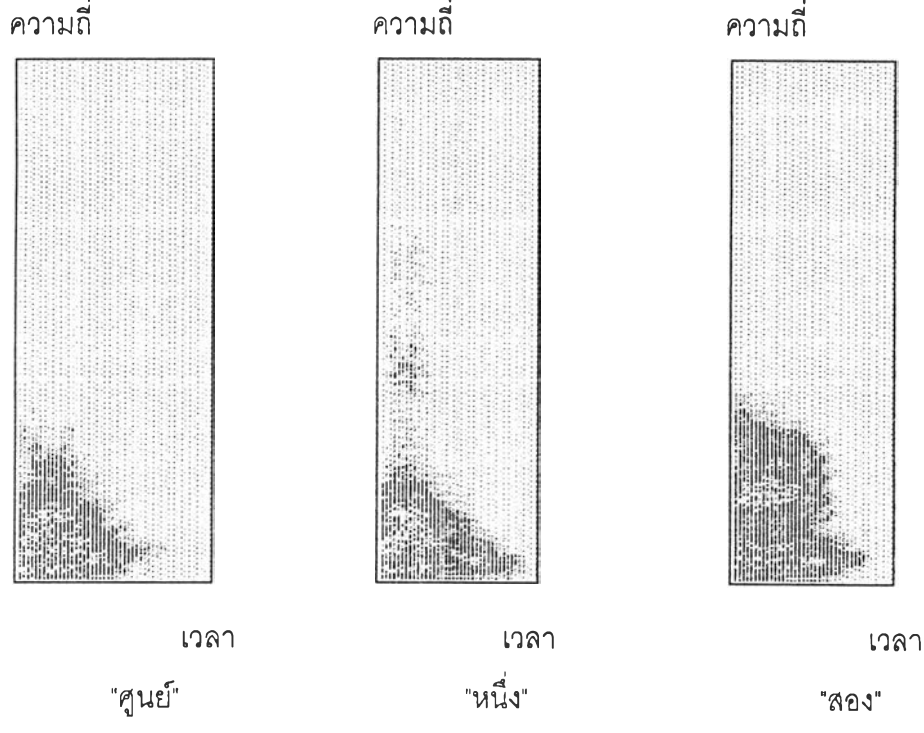
ในส่วนการทดสอบนี้ หลังจากที่ได้กำหนดพารามิเตอร์ของการรู้จำแล้ว จะทำการปรับขนาดของข้อมูลเสียงตามสมการที่ 3.2.16 เพื่อให้ขนาดของเสียงพูดมีความยาวเท่ากัน โดยการทดสอบจะให้มีความยาวเท่ากับ 9,10,20,และ 30 เฟรม ช่วงพารามิเตอร์แสดงความถี่มีค่าของ k ในช่วง 10-127 ผลการทดสอบจะเป็นดังในตารางที่ ก.4 ซึ่งจะเห็นได้ว่าแบบที่ไม่มีการปรับขนาดของจำนวนเฟรมจะให้ผลที่ดีที่สุด

ตารางที่ ก.4 แสดงการปรับขนาดรูปขนาดความยาวของเสียงพูดของแบบอ้างอิง

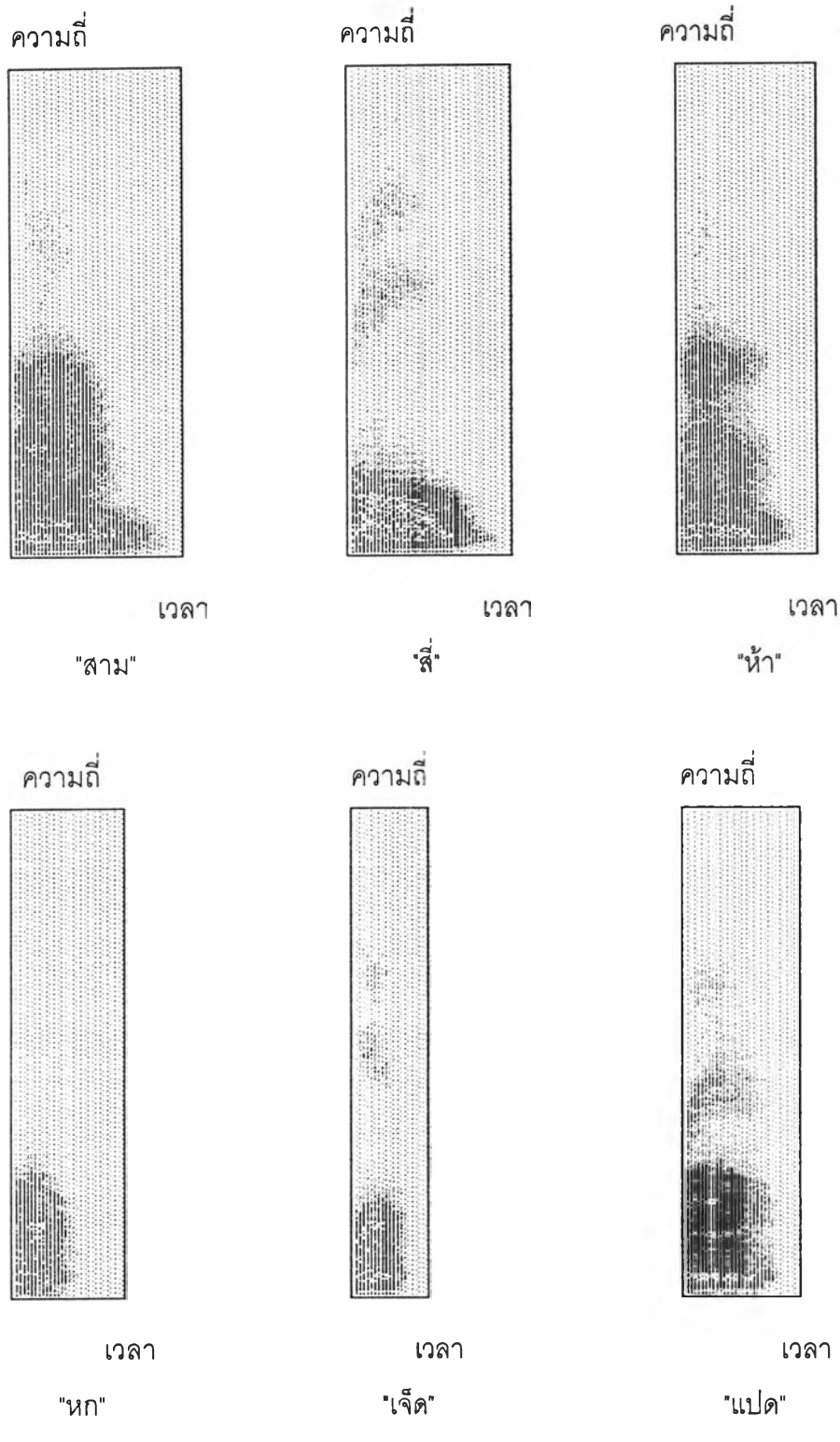
| จำนวนเฟรม | ความถูกต้อง (%) |
|---|---|
| ไม่ปรับขนาด | 93.50 |
| 9 | 89.00 |
| 10 | 88.00 |
| 20 | 85.50 |
| 30 | 71.00 |

ภาคผนวก ข.

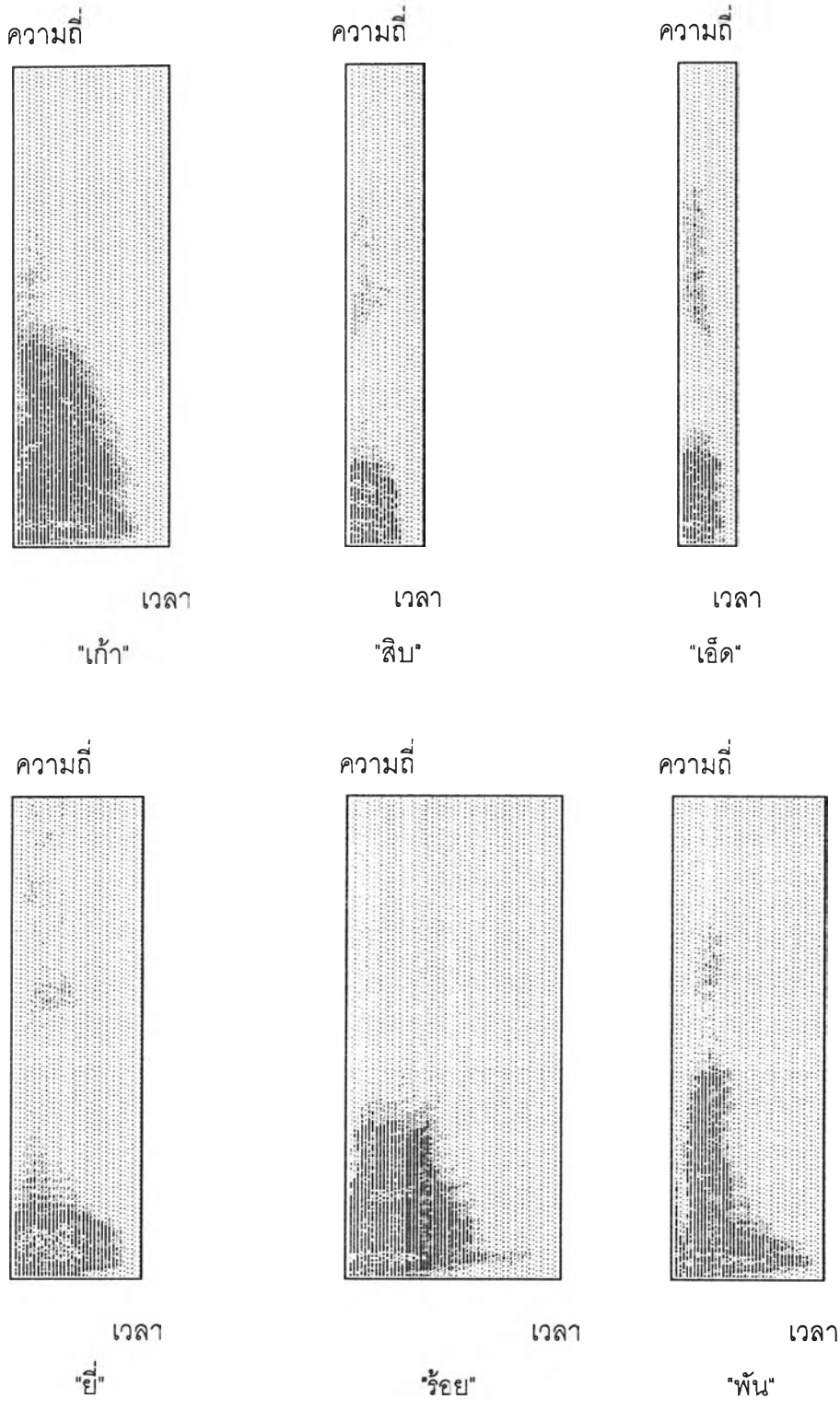
แสดงลักษณะของแบบอ้างอิงของเสียงตัวเลขคำต่าง ๆ ที่ใช้ในการทดสอบครั้งนี้ โดยที่แกนในแนวตั้งจะแสดงช่วงของพารามิเตอร์ k ว่า 0-255 ที่แสดงช่วงความถี่ 0 ถึง 4000 Hz แกนในแนวนอนจะเป็นแกนของเวลาที่เปลี่ยนไป แบบอ้างอิงนี้สร้างจากผู้พูดจำนวน 20 คน



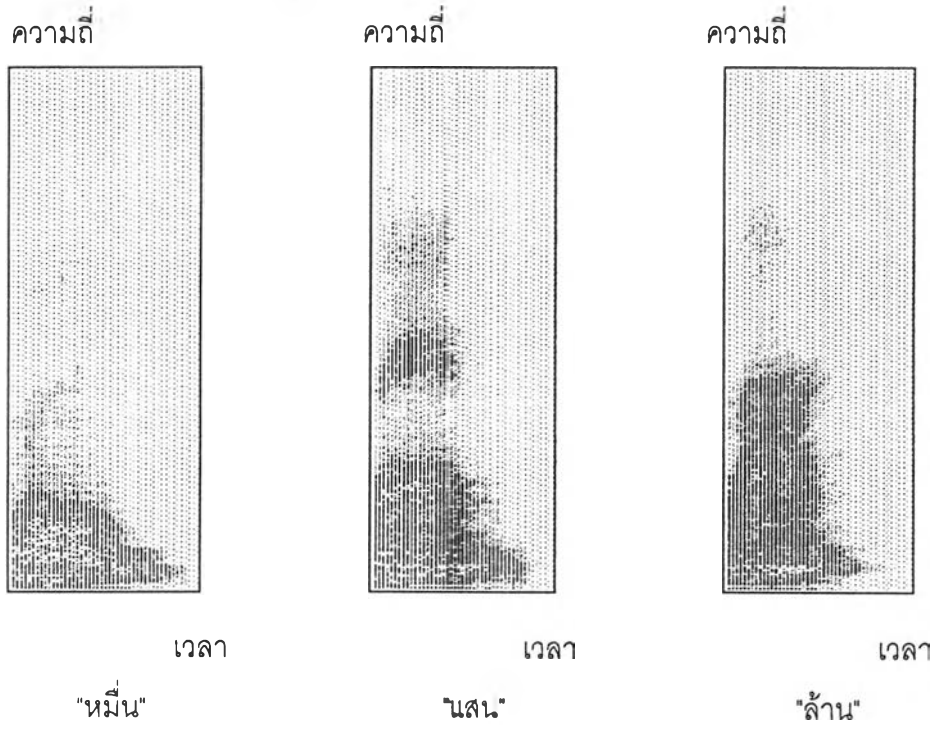

รูปที่ ข.1 แสดงแบบอ้างอิงที่สร้างจากผู้พูดจำนวน 20 คน




รูปที่ ข.1 แสดงแบบอ้างอิงที่สร้างจากผู้พูดจำนวน 20 คน (ต่อ)




รูปที่ ข.1 แสดงแบบอ้างอิงที่สร้างจากผู้พูดจำนวน 20 คน (ต่อ)




รูปที่ ข.1 แสดงแบบอ้างอิงที่สร้างจากผู้พูดจำนวน 20 คน (ต่อ)

ภาคผนวก ค

# รูปแบบของไฟล์เสียง

รูปแบบของข้อมูลเสียง (creative voice file format) นี้จะสามารถแบ่งออกได้ 2 ส่วนคือ Header Block และ Data Block ซึ่งมีรายละเอียดดังนี้คือ (Sound Blaster User Reference Manual, 1989)

1. Header Block

ในส่วน Header Block นี้จะเป็นส่วนต้นของ creative voice file ซึ่งมีจำนวนข้อมูลทั้งหมด 20 ไบต์ โดยมีรายละเอียดดังนี้คือ

ก) ไบต์ที่ 0H-13H จะเป็นส่วนของ file description มีขนาด 20 ไบต์ จะเป็นส่วนที่ใช้เก็บข้อความแสดงรายละเอียดทั่วไป โดยจะเก็บข้อความดังนี้คือ "Creative voice file",1AH

ข) offset to voice data location จะใช้ในงานประยุกต์ ในกรณีที่ Header Block มีขนาดเปลี่ยนแปลงไป ในที่นี้จะมีค่าเป็น 1AH ในสวนนี้จะมีขนาด 2 ไบต์ คือไบต์ที่ 14H-15H

ค) voice file format version number ไบต์ที่ 16H-17H ขนาด 2 ไบต์ จะมีค่าต่างกันไปตามแต่ละ version เพื่อกำหนดรูปแบบของ format ในอนาคตข้างหน้า

ง) voice file identification code ไบต์ที่ 18H-19H เป็น code ที่แสดงว่าไฟล์ ที่ load ขึ้นมานั้นเป็นไฟล์เสียง (voice file) จริง ซึ่ง code นี้หาได้จากค่า complement ของ file format version number บวกกับ 1234H เช่น

version ที่ใช้ : comp(010AH) + 1234H = 01129H

2. Data Block

ในไบต์แรกของ data block นี้จะใช้ในการกำหนดชนิดของ data block (BLKTYPE) ส่วนอีกสามไบต์ถัดไปจะแสดงถึงขนาดของความยาวของ block (BLKLEN) นอกนี้แล้วยังมีส่วนของรายละเอียดของข้อมูลเสียงอื่นอีก เช่น อัตราการสุ่ม (sampling rate) , วิธีการบีบข้อมูล (packing) , ข้อมูลเสียง (voice data) หรืออาจจะเป็นข้อมูลอื่น เช่น markers / remark ซึ่งข้อมูลต่าง ๆ เหล่านี้จะขึ้นอยู่กับชนิดของ data block (BLKTPYE) block ชนิดต่าง ๆ แบ่งออกได้ดังนี้

ก) Block Type 0 ที่เรียกว่า "END BLOCK" จะเป็น block ที่แสดงการสิ้นสุดของข้อมูลเสียง สัญญาณเสียงที่ออกมาจะหยุด block นี้จะเป็น block สุดท้าย ซึ่งแสดงดังในรูปข้างล่างนี้

| 00 |
|---|

ข) Block Type 1 หรือที่เรียกว่า "New voice block" block ชนิดนี้จะแสดงถึงกลุ่มของข้อมูลเสียงเริ่มต้น ซึ่งอาจมีค่าของ sampling rate หรือ packing method ที่ต่างกัน โดยที่ไบต์แรกที่ต่อจาก BLKLEN จะแสดงถึง sampling rate code (SR) ซึ่งจะมีความสัมพันธ์ดังนี้

SR = 256 - (1000000/ความถี่ที่ใช้ในการสุ่ม)

ในส่วนของไบต์ถัดไปจะแสดงถึงรูปแบบการบีบข้อมูล (packing format,PACK) ซึ่งมีค่าต่าง ๆ ดังนี้

| ค่าของ PACK | ความหมาย |
|---|---|
| 0 | 8 bits unpack |
| 1 | 4 bits packed |
| 2 | 2.5 bits packed |
| 3 | 2 bits packed |

และจำนวนของข้อมูลเสียงจะหาได้จาก

ความยาวของข้อมูลเสียง (จำนวนไบต์) = BLKLEN - 2

| 01 | BLKLEN | SR | PACK | VOICE DATA |
|---|---|---|---|---|

ค) Block Type 2 "sub sequent voice block" เป็น block ที่แสดงถึงความต่อเนื่องของข้อมูลเสียงที่ต่อจาก BLKTYPE ที่ 1 จะใช้ในกรณีที่ข้อมูลเสียงมีมากจนไม่สามารถเก็บไว้ภายใน block เดียวได้ ซึ่งจะมี format ดังนี้

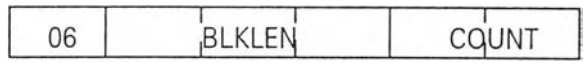
| 02 | BLKLEN | VOICE DATA |
|---|---|---|

ง) Block Type 3 "Silence Block" จะแสดงถึงสภาวะเงียบ (silence period) ของข้อมูลเสียง (voice data) ความยาวของ silence ɔeriod นี้จะแสดงอยู่ต่อจากส่วนของ BLKLEN ดังในรูป ซึ่งจะมีค่าเท่ากับ 3 เสมอ

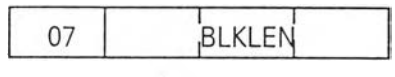
จ) Block Type 4 "maker Block" .เป็น block พิเศษที่สามารถใช้ส่งผ่าน maker นี้ไปยังโปรแกรมประยุกต์อื่นได้ในขณะที่มี output ของเสียง เพื่อใช้ในการ synchronization control

ฉ) Block Type 5 "message block" จะเก็บข้อมูลที่เป็น ASCII โดยข้อมูล ASCII นี้จะแสดงการสิ้นสุดของข้อมูล ASCII โดยใช้ NULL STRING (ASCII 0) block ชนิดนี้ส่วนใหญ่ไม่ใช้กัน แต่จะใช้เมื่อโปรแกรมประยุกต์ต้องการที่จะระบุ ชื่อ ,ชนิด หรือข้อมูลของเสียงลงไปเป็นข้อมูล ASCII

ช) Block Type 6 "Repeat B ock" จะทำการเริ่มต้นของข้อมูลเสียงใน Block ที่อยู่ระหว่าง Block Type 6 กับ Block Type 7 ตามจำนวนของค่า COUNT ภายใน Block Type 6 นี้ ซึ่งจะทำการเล่นเสียงซ้ำจำนวน COUNT + 1 ครั้ง BLKLEN จะมีค่าเท่ากับ 2




ซ) Block Type 7 "Repeat end block" จะเป็น block ที่กำหนดการสิ้นสุดของการเล่นซ้ำ เมื่อ block นี้มาถึง การ process จะเริ่มซ้ำตามที่ได้กล่าวมาในข้อ ช) โดยที่ค่าของ BLKLEN จะมีค่าเท่ากับ 0



# ประวัติผู้เขียน

นาย ระพีพัฒน์ เพ็ญศิริ เกิดเมื่อวันที่ 25 มีนาคม พ.ศ. 2512 ที่นครสวรรค์ จบการศึกษาขั้นปริญญาวิศวกรรมศาสตร์บัณฑิต ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเชียงใหม่ ปีพ.ศ. 2533 เข้าศึกษาต่อในหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต สาขาวิชากรรมวิธีสัญญาณเชิงดิจิตัล ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย เมื่อปีพ.ศ. 2535