หอสมุดกลาง สถาบันวิทยบริการ จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 5

สรุปผลการวิจัยและข้อเสนอแนะ

5.1 สรุปผลการวิจัย

การสร้างวงจรเข้าร หัสดิจิตอลจากสัญญาณเสียงนี้ อาศัยการวิเคราะห์ข้อมูลเสียงเชิงความถี่เป็นหลักในการแยกประเภทของกลุ่มคำ ทำให้สา มารถลดปริมาณหน่วยความจำที่ต้องใช้เก็บแบบแผนอ้างอิง รวมทั้งข้อมูลที่ต้องนำมาใช้เปรียบเทียบลงได้ ตาราง ที่ 5.1 แสดงผลการทดลองเมื่อให้ระบบจดจำคำพูดจากผู้วิจัย

Experiment Report
Automatic Speech Recognition System
Case : A = 7 vocabularies: B = 10 vocabularies

Vocabulary	No. of Training	No. of Try		No. of Success		No. of Failure	
		Case A	Case B	Case A	Case B	Case A	Case B
0	10	20	20	18	12	2	8
1	10	--	20	--	14	--	6
2	10	--	20	--	8	--	12
3	10	20	20	17	14	3	6
4	10	--	20	--	10	--	10
5	10	20	20	18	18	2	2
6	10	20	20	19	20	1	0
7	10	20	20	19	18	1	2
8	10	20	20	18	16	2	4
9	10	20	20	18	18	2	2

ตารางที่ 5.1 แสดงผลการทดลองเมื่อให้ระบบจำคำพูดจำนวน 10 คำ

ความถูกต้องในการจดจำของระบบจะขึ้นกับลักษณะการพูดของผู้พูด ว่าจะมีความใกล้เคียงกับตัวอย่างที่ให้ไว้เพียงใด รวมทั้งทิศทางและระยะทางการพูดจากไมโครโฟน เนื่องจากสิ่งเหล่านี้มีผลต่อการปรากฏของพลังงานเสียงในแต่ละแถบความถี่

หลักการที่ใช้ในการตัดสินใจของระบบอาศัยวิธีการ 3 ประการ คือ

5.1.1 ระยะเวลาในการออกเสียง ใช้แยกกลุ่มเป็นคำสั้น คำยาว

5.1.2 ค่าช่วงกว้างของค่าเฉลี่ยสัญญาณเสียงที่ปรากฏในแต่ละแถบความถี่ ใช้หากลุ่มของคำอ้างอิงที่มีความเป็นไปได้ว่าจะเป็นกลุ่มเดียวกันกับสัญญาณคำพูดที่ไม่ทราบความหมาย โดยเปรียบเทียบค่าเฉลี่ยของสัญญาณที่ไม่ทราบความหมายว่าตกอยู่ในช่วงค่าเฉลี่ยของสัญญาณของคำพูดใดบ้าง

5.1.3 ค่าเฉลี่ยของค่าช่วงกว้างของค่าเฉลี่ยสัญญาณเสียงที่ปรากฏในแต่ละแถบความถี่ ใช้เป็นค่าแทนของแต่ละคำอ้างอิง เพื่อใช้หาตัวอย่างอ้างอิงที่มีตำแหน่งใกล้เคียงที่สุดกับสัญญาณคำพูดที่ไม่ทราบความหมาย โดยนำค่าผลรวมของความแตกต่างยกกำลังสองระหว่างค่าเฉลี่ยของสัญญาณเสียงที่ไม่ทราบความหมายกับค่าเฉลี่ยของค่าช่วงกว้างสัญญาณของแต่ละคำพูดอ้างอิง

จากตารางที่ 4.4 จะเห็นได้ว่ามีคำบางคำซึ่งช่วงกว้างของค่าเฉลี่ยสัญญาณเสียงจากแต่ละแถบความถี่มีความเหลื่อมกัน เช่น "0" "2" และ "3" หรือ "1" และ "4" ซึ่งจะพบว่ามีความเป็นไปได้ที่คำพูด "2" อาจให้ค่าเฉลี่ยของสัญญาณตกอยู่ในกลุ่มของคำว่า "0" หรือ "3" มากกว่า "2" เป็นต้น ซึ่งทำให้การเลือกความหมายของระบบผิดพลาดได้ ในกรณีที่ค่าเฉลี่ยของสัญญาณเสียงที่ปรากฏในแต่ละแถบความถี่ของคำแต่ละคำมีความแตกต่างกันอย่างมาก ระบบจะสามารถจดจำได้อย่างถูกต้อง

จากผลการทดลองระบบจะสามารถจดจำคำพูด 0, 3, 5, 6, 7, 8, 9 ได้ถูกต้องมากกว่าร้อยละ 70 ของคำทดลอง และความถูกต้องจะลดลงมาประมาณ 50% เมื่อให้ระบบจดจำคำพูด 1, 2 และ 4

5.2 ปัญหาในการวิจัย

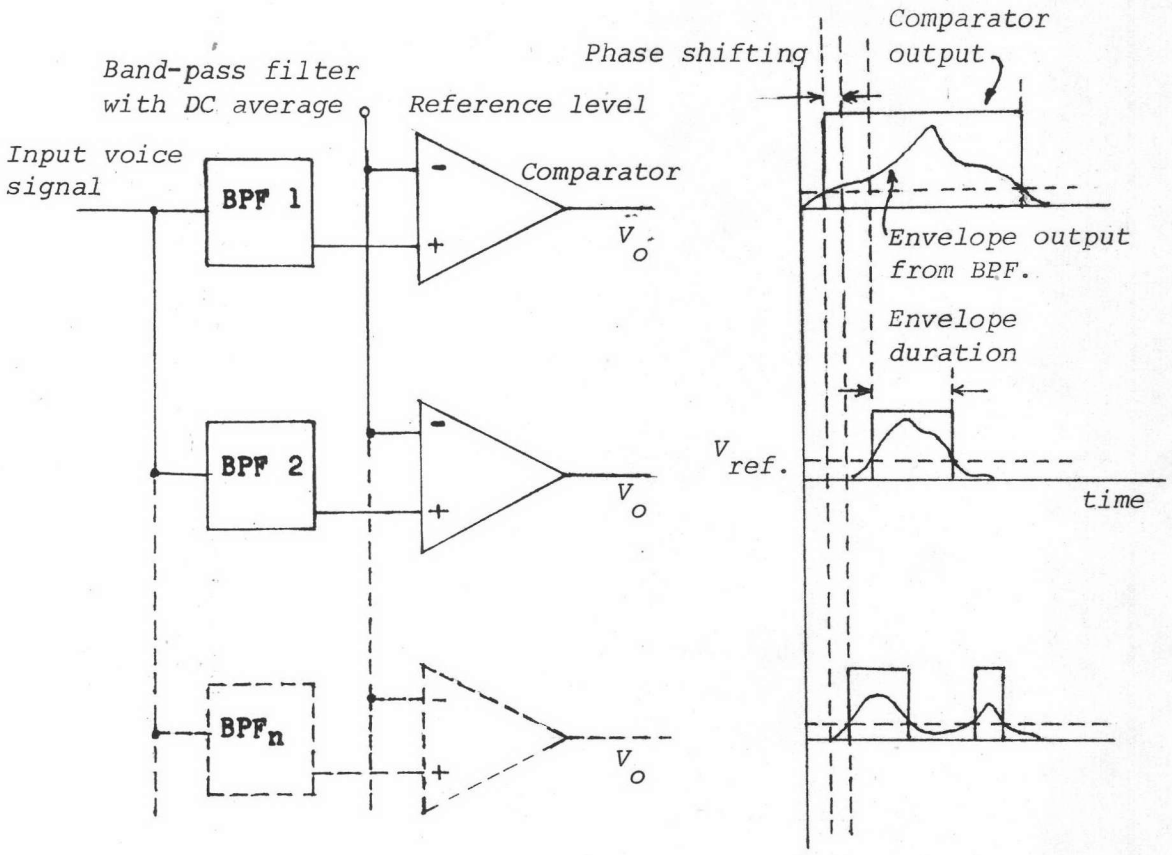
ในการวิจัยครั้งนี้ ผู้วิจัยยัง มีความรู้ด้าน เสียง และการวิเคราะห์สัญญาณน้อย หนังสือที่เกี่ยวกับระบบจดจำคำพูดหาได้ยาก การค้นคว้าวิจัยและหาแนวทางในการพัฒนาจึงอาศัยผลจากการทดลองเป็นหลัก ซึ่งจะทำให้ระบบที่พัฒนาขึ้นมายังไม่สมบูรณ์เพียงพอ ปัญหาที่สำคัญของระบบก็คือ ข้อมูลที่เข้าสู่ระบบ (เสียงพูดของคน) มีความแปรปรวนมากอัน

เนื่องมาจากสภาวะร่างกาย อารมณ์ และสิ่งแวดล้อมรอบตัวผู้พูด การวิจัยเชิงความถี่ของเสียงจำเป็นต้องใช้เครื่องวิเคราะห์อะคูสติกเสปคตรัม (acoustic spectrum analyzer) จึงสามารถวิเคราะห์ถึงองค์ประกอบเชิงความถี่ของเสียงได้อย่างละเอียด ซึ่งผู้วิจัยไม่สามารถหาใช้ได้ จึงทำให้เป็นการยากพอสมควรที่จะหาวิธีการที่เหมาะสมในการแยกประเภทกลุ่มของข้อมูล โดยครอบคลุมความแปรปรวนของข้อมูลได้ อย่างไรก็ดี จากการศึกษาการวิจัยอื่น ๆ ที่ได้กระทำมาแล้วนั้น ยังมีวิธีการอื่น ๆ อีกหลายวิธีที่ใช้ในระบบจดจำคำพูด ซึ่งผู้ทำวิจัยจำเป็นต้องมีความรู้เกี่ยวกับเสียงและการวิเคราะห์สัญญาณดีพอสมควร การทำวิจัยลักษณะดังกล่าว จะให้ผลดีอย่างแท้จริง จะต้องมีการลงทุนและใช้เวลานานมาก เพื่อหาวิธีการที่เหมาะสม และโครงการประเภทนี้ควรจะมีการดำเนินการโดยผู้เชี่ยวชาญทั้งทางวิเคราะห์เสียงและทางคอมพิวเตอร์ทำงานร่วมกันเป็นเวลานาน

5.3 ข้อเสนอแนะ

จากผลการวิจัย ระบบสามารถจดจำคำพูดได้ดีพอสมควร สามารถนำไปใช้เป็นแนวทางในการวิจัยต่อไปได้ การพัฒนาในขั้นต่อไปอาจทำโดยการเพิ่มวงจรกรองผ่านแถบความถี่ เพื่อศึกษาลักษณะองค์ประกอบของเสียงพูดในแถบความถี่อื่น ๆ เพิ่มเติม ซึ่งจากรูปที่ 4.4 จะเห็นได้ว่า เสียงพูดแต่ละคำมีลักษณะองค์ประกอบเด่นชัดในหลาย ๆ แถบความถี่ด้วยกัน นอกจากนี้อาจเพิ่มกฎเกณฑ์ในการวัดองค์ประกอบของเสียงและเพิ่มตัวแปรที่ใช้ในการแบ่งประเภทของกลุ่มคำ ซึ่งจะช่วยเพิ่มความถูกต้องในการตัดสินใจ ได้แก่

5.3.1 การวัดตำแหน่งและช่วงเวลาที่มีการปรากฏของพลังงานเสียงในแต่ละแถบความถี่และการหาความสัมพันธ์ของการปรากฏเทียบระหว่างแถบความถี่และการเกิดคำ ทำได้โดยการเพิ่มวงจรตรวจวัดระดับสัญญาณ แล้วนำสัญญาณเอ็นเวล็อบ(envelope) จากแต่ละวงจรกรองแถบความถี่มาเทียบกับระดับสัญญาณอ้างอิง ทำให้ได้ความสัมพันธ์ของการปรากฏของพลังงานเสียงในแต่ละแถบความถี่รวมทั้งช่วงเวลาของการปรากฏ



รูปที่ 5.1 แสดงการวัดตำแหน่งและการปรากฏของพลังงานเสียงในแถบความถี่ต่าง ๆ

5.3.2 การตัดตอนแยกช่วงเวลาในการหาค่าเฉลี่ยของสัญญาณเสียง ในการวิจัยนี้การหาค่าเฉลี่ยของสัญญาณเสียงคิดจากผลรวมของขนาดสัญญาณที่ปรากฏของแต่ละแถบความถี่หารด้วยระยะเวลาของคำพูด ซึ่งให้เป็นค่าเฉลี่ยของสัญญาณทั้งหมด จากผลการทดลองที่ 4.1.2 จะเห็นได้ว่า ลักษณะของเอ็นเวล็อบที่ปรากฏมีความโน้มเอียงที่แตกต่างกัน เช่น มีสัญญาณปรากฏมากในช่วงเวลาแรกของการออกเสียง หรืออาจไปปรากฏมากในช่วงท้ายของคำ ซึ่งทำให้ได้คุณลักษณะของสัญญาณเพิ่มเติม

5.3.3 การวิเคราะห์อัตราการผ่านศูนย์ของคำ อาจช่วยแยกประเภทของเสียงออกได้ ในวงจรที่สร้างขึ้นสำหรับการวิจัยนี้ ผู้วิจัยได้สร้างวงจรตรวจจับการผ่านศูนย์ไว้ซึ่งจากข้อมูลที่ปรากฏ (ดังแสดงในภาคผนวก) จะเห็นได้ว่าการออกเสียงบางเสียง เช่น ส, ฝ ซึ่งเป็นเสียงที่ไม่มีการออกเสียง (unvoiced) จะให้อัตราการผ่านศูนย์

ค่อนข้างสูง ในขณะที่การออกเสียงอื่น ๆ ให้อัตราการผ่านศูนย์ที่ต่ำกว่าอย่างเด่นชัด

วิธีการอื่น ๆ ที่ใช้ในการหาพารามิเตอร์ของเสียงยังมีอีกหลายวิธี เช่น การหาลักษณะของพลังงานเสียง ซึ่งบ่งลักษณะเสียงสระ เสียงพยัญชนะ การนำเอากลวิธีของการเข้ารหัสจากการพยากรณ์เชิงเส้น (Linear predictive coding) เข้ามาช่วยในการวิเคราะห์สัญญาณเสียง จะช่วยให้การจดจำถูกต้องมากขึ้น ซึ่งวิธีการเหล่านี้สามารถค้นคว้าได้จากหนังสือของ IEEE เรื่อง Automatic Speech & Speaker Recognition

นอกจากนี้ ในระบบปัจจุบันในการใช้งานทุกครั้ง จำเป็นต้องสอนให้ระบบเรียนรู้ เมื่อปิดเครื่องข้อมูลอ้างอิงจะหายไป เนื่องจากข้อมูลอ้างอิงที่ระบบได้สร้างขึ้นถูกเก็บไว้ในหน่วยความจำแบบ RAM หากนำมาเชื่อมโยงกับระบบไมโครคอมพิวเตอร์จะสามารถจัดเก็บข้อมูลไว้ในหน่วยความจำสำรอง เช่น แผ่นจานบันทึก ซึ่งเวลาใช้งานจะสามารถนำแบบแผนอ้างอิงของบุคคลที่ต้องการคืนเข้าสู่ระบบ ทำให้ใช้งานได้ทันทีโดยไม่ต้องสอนให้ระบบเรียนรู้ใหม่ สำหรับระบบที่สร้างขึ้นนี้ เมื่อไม่ต่อเครื่องพิมพ์สามารถส่งข้อมูลออกแบบขนานได้ทางพอร์ท 09 และรับข้อมูลเข้าทางพอร์ท A ของ 8255