บทที่ 1

บทนำ

ความสำคัญและความเป็นมา

สัญญาณเสียงหรือข้อมูลเสียงในแบบดิจิตอลมีข้อดีหลายประการ ได้แก่ ความทนทานต่อสัญญาณรบกวนด้วยเหตุที่สัญญาณดิจิตอลมีค่าเพียงศูนย์หรือหนึ่งเท่านั้นและระดับสัญญาณไฟฟ้าที่แสดงค่าทั้งสองนี้ก็สามารถอยู่ในช่วงที่ห่างกันได้มาก สัญญาณรบกวนจะต้องมีความแรงค่อนข้างมากจึงจะทำให้การตีความหมายของสัญญาณผิดพลาด นอกจากนี้ข้อมูลเสียงในแบบดิจิตอลยังสามารถบันทึกและทำซ้ำได้โดยไม่บั่นทอนคุณภาพของสัญญาณ รวมทั้งสามารถใช้คอมพิวเตอร์มาปรับแต่งสัญญาณเสียงได้ ด้วยเหตุผลเหล่านี้จึงเกิดการประยุกต์ใช้งานในหลาย ๆ ด้าน ตัวอย่างได้แก่ การรับส่งสัญญาณเสียงในระบบโทรศัพท์ซึ่งมีการแปลงสัญญาณเสียงให้อยู่ในรูปดิจิตอลเพื่อทำให้สามารถส่งสัญญาณผ่านอุปกรณ์ทวนสัญญาณได้หลาย ๆ ครั้ง นอกจากนี้ยังทำให้สามารถมัลติเพล็กซ์ (multiplex) สัญญาณเสียงหลาย ๆ สัญญาณเพื่อส่งผ่านไปทางบนช่องทางสื่อสารเดียวกันได้โดยอาศัยวิธีการมัลติเพล็กซ์เชิงเวลาที่ง่ายและราคาถูก ซึ่งถ้าเป็นสัญญาณในแบบอนาล็อกการมัลติเพล็กซ์สัญญาณจะต้องอาศัยการมัลติเพล็กซ์ในเชิงความถี่และการส่งสัญญาณผ่านตัวทวนสัญญาณก็จะทำให้เกิดการบั่นทอนคุณภาพของสัญญาณลง อีกตัวอย่างหนึ่งได้แก่การประยุกต์ในระบบที่ให้ข้อมูลแก่ผู้ใช้โดยมีส่วนติดต่อกับผู้ใช้เป็นเสียงพูด ในระบบนี้ต้องมีการบันทึกเสียงพูดเอาไว้ก่อน เช่นระบบที่ให้ข้อมูลผ่านทางโทรศัพท์ พจนานุกรมพูดได้ และอื่น ๆ อีกมาก

ข้อเสียที่พบได้เสมอของเสียงในแบบดิจิตอลก็คือเรื่องของขนาดข้อมูล พิจารณาการแปลงข้อมูลเสียงพูดสำหรับระบบโทรศัพท์ เสียงของคนในช่วงความถี่ 200-3400 Hz ใช้อัตราการแปลงข้อมูล 8000 ตัวอย่างต่อวินาที หลังจากผ่านการทำคอมแพนดิง (companding) ซึ่งเป็นวิธีหนึ่งที่ช่วยลดจำนวนบิตของรหัสที่ให้แทนตัวอย่างข้อมูลเสียง รหัสที่ได้สำหรับแต่ละตัวอย่างจะมีขนาด 8 บิต ทำให้ได้ข้อมูลในอัตรา 64 Kbps ถ้าต้องการเก็บเสียงไว้ในช่วงเวลาหนึ่งนาทีก็จะต้องใช้หน่วยความจำถึง 480 Kbytes และยิ่งถ้าเป็นกรณีที่บันทึกเสียงดิจิตอลที่ต้องการคุณภาพของเสียงที่ดีกว่านี้ เช่นเสียงดนตรีหรือเสียงที่บันทึกในสื่อของภาพยนต์ ก็ต้อง

ใช้จำนวนบิตต่อตัวอย่างเสียงและอัตราการแปลงข้อมูลสูงขึ้นไปอีกทำให้ปริมาณข้อมูลผลลัพธ์ที่ได้สูงมาก การบีบขนาดข้อมูลเสียงจึงเป็นเรื่องสำคัญทั้งนี้เพื่อทำให้สามารถจัดเก็บสัญญาณเสียงลงในอุปกรณ์บันทึกต่าง ๆ หรือส่งสัญญาณเสียงไปในช่องสื่อสารได้อย่างมีประสิทธิภาพเป็นการใช้สายนำสัญญาณที่มีแบนด์วิดธ์ (bandwidth) จำกัดได้อย่างคุ้มค่ามากขึ้น แต่ทั้งนี้การบีบข้อมูลเสียงดังกล่าวยังต้องรักษาคุณภาพของเสียงไว้ที่ระดับที่ยอมรับได้ด้วย

เนื่องจากวิธีการในการเข้ารหัสเพื่อบีบขนาดข้อมูลเสียงที่สามารถบีบข้อมูลได้มาก ๆ มักต้องอาศัยกระบวนการคำนวณที่ซับซ้อน และต้องการความสามารถในการประมวลผลในระดับสูง คอมพิวเตอร์ที่ใช้ในงานทั่วไปจึงมักไม่สามารถบีบและคลายสัญญาณเสียงได้อย่างรวดเร็วเพียงพอหรือมิฉะนั้นก็มีราคาสูงเกินไปไม่เหมาะในงานประยุกต์ที่ต้องผลิตเป็นจำนวนมาก อุปกรณ์ซึ่งได้รับการออกแบบมาเป็นพิเศษเพื่อใช้ในงานประมวลผลสัญญาณที่ต้องมีการคำนวณในปริมาณมาก ๆ ในเวลาที่มีจำกัดได้แก่ตัวประมวลผลสัญญาณดิจิตอลหรือที่เรียกย่อกันว่าดีเอสพี (Digital Signal Processor - DSP) งานวิทยานิพนธ์นี้มุ่งศึกษาการบีบข้อมูลเสียงพูดให้สามารถทำงานได้อย่างรวดเร็วในระดับที่เรียกว่า 'ทันที' (real-time) โดยทั่วไปสำหรับวิธีที่ศึกษาจะมีการแบ่งสัญญาณหรือข้อมูลเสียงพูดที่จะนำมาบีบออกเป็นช่วง ๆ ละประมาณ 20 mS แต่ละช่วงเรียกว่าเฟรม (frame) การบีบข้อมูลเสียงแบบทันทีหมายถึงตัวประมวลผลจะต้องทำการบีบข้อมูลเสียงในเฟรมปัจจุบันให้เสร็จสิ้นก่อนที่เฟรมข้อมูลเสียงเฟรมใหม่จะเข้ามาเพื่อหลีกเลี่ยงการสะสมของข้อมูลเสียงเมื่อเวลาผ่านไป

วัตถุประสงค์ในการวิจัย

ในการวิจัยครั้งนี้มีจุดประสงค์ดังต่อไปนี้

1. เรียนรู้วิธีการบีบและคลายข้อมูลเสียงพูด

ศึกษาวิธีการต่าง ๆ ในการบีบขนาดข้อมูลเสียง โดยเน้นที่วิธีที่เหมาะกับเสียงพูดและการประยุกต์ทางด้านการสื่อสารกล่าวคือเป็นวิธีที่ให้คุณภาพของเสียงภายหลังการบีบและคลายแล้วในระดับที่สามารถสื่อสารเข้าใจได้ ในงานวิทยานิพนธ์นี้คำว่า 'เสียง' หากมิได้เจาะจงเฉพาะว่าเป็นเสียงใดจะหมายถึงเสียงพูดเป็นหลัก วิธีบีบข้อมูลเสียงที่ต้องการคุณภาพเสียงในระดับสูงเช่นที่ใช้กับดนตรีหรือภาพยนต์อยู่นอกเหนือขอบเขตการวิจัยนี้ การศึกษาเน้นวิธีที่สามารถบีบข้อมูลเสียงพูดได้มากและสามารถรักษาคุณภาพของเสียงพูดไว้ที่ระดับที่ยังฟังแล้วเข้าใจได้โดยง่าย วิธีการเหล่านี้ได้แก่วิธีที่มีพื้นฐานมาจากการวิเคราะห์และเข้ารหัสที่เรียกว่าการเข้ารหัสแบบทำนายเชิงเส้นหรือเรียกย่อว่าแอลพีซี (Linear Predictive Coding - LPC) ได้แก่

วิธีแอลพีซี10 (LPC10) วิธีซีอีแอลพี (CELP - Code-Excited Linear Prediction) และวิธีอาร์พีอี-แอลทีพี (RPE-LTP - Residual Pulse Excitation - Long-Term Prediction) เป็นต้น

2. พัฒนาโปรแกรมในการบีบและคลายข้อมูลเสียงพูดสำหรับคอมพิวเตอร์พีซี

พัฒนาโปรแกรมที่ทำงานบนคอมพิวเตอร์พีซี สำหรับบีบและคลายข้อมูลเสียงพูดโดยอาศัยหลักการตามวิธีที่ได้ศึกษาไว้ในข้อหนึ่ง โปรแกรมที่พัฒนารับข้อมูลเสียงพูดที่อยู่ในแฟ้มข้อมูลชนิดเวฟ (.wav) ซึ่งสร้างได้จากแผงวงจรเสียง (sound card) ทั่วไป โปรแกรมจะสร้างแฟ้มข้อมูลผลลัพธ์ที่เก็บรหัสที่ใช้แทนข้อมูลเสียงพูด ส่วนโปรแกรมที่ใช้คลายข้อมูลก็จะทำงานกลับกันโดยอ่านแฟ้มข้อมูลที่เก็บรหัสข้อมูลเสียงพูดแล้วสร้างแฟ้มข้อมูลเสียงพูดกลับคืนมา โปรแกรมต่าง ๆ ที่พัฒนาใช้วิธี แอลพีซี แอลพีซี10 ซีอีแอลพี และ อาร์พีอี-แอลทีพี

3. พัฒนาโปรแกรมบีบข้อมูลเสียงพูดสำหรับตัวประมวลผลสัญญาณดิจิตอลที่ทำงานแบบทันที

พัฒนาโปรแกรมบีบข้อมูลเสียงพูดที่ทำงานแบบทันทีโดยนำวิธีการที่ให้คุณภาพของเสียงพูดในระดับที่ดีและมีความซับซ้อนในการทำงานในระดับที่เหมาะสมและยังสามารถทำงานแบบทันทีได้โดยตัวประมวลผลสัญญาณดิจิตอล ซึ่งจากการศึกษาพบว่า วิธีอาร์พีอี-แอลทีพี เป็นวิธีหนึ่งที่มีความเหมาะสม วิธีดังกล่าวแม้จะไม่ใช่วิธีที่ให้อัตราการบีบข้อมูลสูงสุดแต่คุณภาพของเสียงพูดที่ได้ก็จัดอยู่ในเกณฑ์ดี อีกทั้งขั้นตอนวิธีก็ไม่ซับซ้อนจนเกินไป สำหรับตัวประมวลผลสัญญาณดิจิตอลที่เลือกใช้คือ ADSP2101 เนื่องจากเป็นตัวประมวลผลที่มีความสามารถในการประมวลผลสูงและมีภาษาแอสเซมบลีที่ง่ายและทำงานได้อย่างมีประสิทธิภาพ โปรแกรมที่พัฒนาทำงานอยู่บนฮาร์ดแวร์สำหรับการบีบและคลายข้อมูลเสียงพูดซึ่งได้พัฒนาขึ้นมาสำหรับการวิจัยครั้งนี้

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ขั้นตอนการวิจัย

1. การศึกษาวิธีการบีบและคลายข้อมูลเสียงพูด

ในขั้นแรกได้ทำการศึกษารายละเอียดของการทำงานของการบีบและคลายข้อมูลเสียงพูดในวิธีต่าง ๆ ได้แก่ แอลพีซี ซีอีแอลพี และ อาร์พีอี-แอลทีพี เป็นต้น วิธีการศึกษาเพื่อให้สามารถเข้าใจขั้นตอนวิธีได้อย่างลึกซึ้งจึงทำการเขียนโปรแกรมตามหลักการของการเข้ารหัสแบบทำนายเชิงเส้นข้างต้นเพื่อเข้ารหัสข้อมูลเสียงพูดและสังเคราะห์สัญญาณเสียงพูดคืนจาก

รหัสข้อมูลที่บีบไว้ พารามิเตอร์ต่าง ๆ ของการเข้ารหัสและถอดรหัสในแต่ละวิธีเป็นสิ่งที่สนใจศึกษาและถูกนำมาเปรียบเทียบกันเพื่อใช้เป็นข้อมูลในการวิจัยขั้นต่อไปพารามิเตอร์ที่สนใจได้แก่ อัตราส่วนการบีบอัดข้อมูล เนื่องจากผลลัพธ์ของโปรแกรมในขั้นตอนการวิจัยนี้เป็นแฟ้มรหัสข้อมูล พารามิเตอร์ตัวนี้จึงถูกเปรียบเทียบในลักษณะอัตราส่วนของขนาดของแฟ้มข้อมูลต้นฉบับเทียบกับขนาดแฟ้มข้อมูลรหัสที่ได้ พารามิเตอร์ที่สนใจศึกษาอีกตัวหนึ่งก็คือคุณภาพของข้อมูลเสียงพูดเมื่อผ่านการบีบและคลายแล้ว

2. ศึกษาตัวประมวลผลสัญญาณดิจิตอล

ศึกษาตัวประมวลผลสัญญาณดิจิตอล ในแง่สถาปัตยกรรม การทำงาน ความสามารถในการประมวลผลสัญญาณ การเชื่อมโยงกับอุปกรณ์ที่แปลงสัญญาณเสียงพูดในแบบอนาล็อกเป็นดิจิตอลและกลับกัน การเขียนโปรแกรม และความเหมาะสมที่จะนำมาประยุกต์ใช้ทางด้านการบีบข้อมูลเสียงพูด ตัวประมวลผลสัญญาณดิจิตอลที่นำมาศึกษาและใช้ในการวิจัยครั้งนี้คือ ตัวประมวลผลสัญญาณดิจิตอลของบริษัทอนาล็อกดีไวส์ ADSP2101 ด้วยเหตุผลดังที่กล่าวมาข้างต้น

3. พัฒนาโปรแกรมและฮาร์ดแวร์ที่จำเป็น

เขียนโปรแกรมซึ่งทำงานบนตัวประมวลผลสัญญาณดิจิตอล ADSP2101 เพื่อบีบและคลายสัญญาณเสียงพูดแบบทันที รวมทั้งพัฒนาฮาร์ดแวร์ตามความจำเป็นเพื่อให้โปรแกรมสามารถทำงานได้

ขอบเขตการวิจัย

เสียงที่สนใจศึกษาในการวิจัยจำกัดเฉพาะเสียงพูดเท่านั้น ในส่วนของการเปรียบเทียบคุณภาพของเสียงพูดที่ได้จากการบีบและคลายวิธีต่าง ๆ ใช้เฉพาะภาษาไทยในการทดสอบ ส่วนตัวประมวลผลสัญญาณดิจิตอลที่สนใจศึกษาสำหรับขั้นตอนการพัฒนาโปรแกรมบีบข้อมูลเสียงพูดแบบทันทีก็จำกัดเฉพาะ ADSP2101 เท่านั้น

ประโยชน์ที่ได้รับ

1. ได้รับความรู้ทางด้านวิธีการบีบข้อมูลเสียงพูด

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ได้รับความรู้เกี่ยวกับเทคนิคต่าง ๆ ที่ใช้ในการบีบข้อมูลเสียงพูดว่าแต่ละแบบมีการทำงานอย่างไร มีข้อดีข้อเสียทั้งในด้านคุณภาพของสัญญาณเสียงพูดและอัตราการบีบอัดข้อมูลแตกต่างกันอย่างไร สามารถนำไปใช้ในการตัดสินใจเพื่อเลือกเทคนิคให้เหมาะสมกับงานเพื่อให้การบีบข้อมูลเสียงพูดมีประสิทธิภาพมากที่สุด

2. ได้โปรแกรมสำหรับบีบข้อมูลเสียงพูดที่ทำงานบนพีซี

ได้โปรแกรมสำหรับบีบข้อมูลเสียงพูดโดยเทคนิคต่าง ๆ ได้แก่ แอลพีซี ซีอีแอลพี และ อาร์พีอี-แอลทีพี ซึ่งสามารถใช้เป็นเครื่องมือสำหรับบีบแฟ้มข้อมูลเสียงพูดชนิดเวฟให้มีขนาดเล็กลง สามารถจัดเก็บข้อมูลเสียงพูดได้มากขึ้นในหน่วยจัดเก็บ (storage device) ที่มีเนื้อที่จำกัด

3. ได้โปรแกรมสำหรับใช้ในการบีบและคลายข้อมูลเสียงพูดโดยทำงานแบบทันทีสำหรับตัวประมวลผลสัญญาณดิจิตอล ADSP2101

โปรแกรมมีความสามารถในการบีบสัญญาณเสียงพูดเป็นรหัสข้อมูลผลลัพธ์ในทันทีที่สัญญาณเสียงพูดเข้ามา รหัสข้อมูลที่ได้สามารถส่งไปในช่องทางสื่อสารที่มีแบนด์วิดธ์จำกัดได้ หรืออาจบันทึกลงในหน่วยจัดเก็บข้อมูลได้โดยใช้เนื้อที่น้อยลง สำหรับในโหมดการทำงานที่ทำการคลายข้อมูลโปรแกรมสามารถรับรหัสข้อมูลเสียงพูดที่บีบไว้แล้วและแปลงรหัสข้อมูลเป็นสัญญาณเสียงพูดได้ทันที

นอกจากนี้โปรแกรมต่าง ๆ ที่พัฒนาก็ถือได้ว่าเป็นเครื่องมือพื้นฐานสำหรับใช้ในการวิจัยที่เกี่ยวข้องกับทางด้านเสียงที่อาจมีขึ้นในอนาคต

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย