การทำขั้นตอนวิธีบีบข้อมูลเสียงพูดโดยการประมวลผลสัญญาณดิจิตอล

นายมีลาภ เรืองรัตนวิชา

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
ภาควิชาวิศวกรรมคอมพิวเตอร์
บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2539

ISBN 974-635-914-2

ลิขสิทธิ์ของบัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย

I 17138450

AN IMPLEMENTATION OF A SPEECH COMPRESSION ALGORITHM BY DIGITAL SIGNAL PROCESSING

Mr.Meelarp Ruangratanawicha

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science
Graduate School
Chulalongkorn University
Academic Year 1996

ISBN 974-635-914-2

หัวข้อวิทยานิพนธ์	การทำขั้นตอนวิธีบีบข้อมูลเสียงพูดโดยการประมวลผลสัญญาณดิจิตอล
โดย	นายมีลาภ เรืองรัตนวิชา
ภาควิชา	วิศวกรรมคอมพิวเตอร์
อาจารย์ที่ปรึกษา	อาจารย์ ดร. ประภาส จงสถิตย์วัฒนา

บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

..... คณบดีบัณฑิตวิทยาลัย
(ศาสตราจารย์ นายแพทย์ ศุภวัฒน์ ชุติวงศ์)

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ
(ผู้ช่วยศาสตราจารย์ บุญชัย โสวรรณวณิชกุล)

..... อาจารย์ที่ปรึกษา
(อาจารย์ ดร. ประภาส จงสถิตย์วัฒนา)

..... กรรมการ
(อาจารย์ นงลักษณ์ โควาวิสารัช)

..... กรรมการ
(อาจารย์ ดร. สืบสกุล พิภพมงคล)

..... กรรมการ
(อาจารย์ ดร. ฐิต ศิริบูรณ์)

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

พิมพ์ต้นฉบับบทคัดย่อวิทยานิพนธ์ภายในกรอบสีเขียวนี้เพียงแผ่นเดียว

หอสมุดกลาง จุฬาลงกรณ์มหาวิทยาลัย

มีลาภ เรืองรัตนวิชา : การทำขั้นตอนวิธีบีบข้อมูลเสียงพูดโดยการประมวลผลสัญญาณดิจิตอล (AN IMPLEMENTATION OF A SPEECH COMPRESSION ALGORITHM BY DIGITAL SIGNAL PROCESSING) อ. ที่ปรึกษา : อ.ดร. ประภาส จงสถิตย์วัฒนา, 92 หน้า. ISBN 974-635-914-2.

การวิจัยครั้งนี้มีวัตถุประสงค์หลักในการพัฒนาโปรแกรมบีบข้อมูลเสียงพูดซึ่งให้คุณภาพของเสียงพูดในระดับที่สูงเพียงพอสำหรับการประยุกต์ด้านการสื่อสาร หลักการทำงานของโปรแกรมที่พัฒนาขึ้นมีพื้นฐานมาจากวิธีการเข้ารหัสแบบทำนายเชิงเส้น (Linear Predictive Coding - LPC) วิธีการที่ใช้ได้แก่ แอลพีซี10 (LPC10) ซีอีแอลพี (CELP - Code Excited Linear Prediction) และ อาร์พีอี-แอลทีพี (RPE-LTP - Residual Pulse Excitation - Linear Prediction) ในขั้นแรกของการวิจัยได้พัฒนาโปรแกรมให้ทำงานอยู่บนคอมพิวเตอร์พีซีและรับข้อมูลเสียงพูดที่อยู่ในแฟ้มข้อมูลเวฟ (wave file) เป็นข้อมูลเข้า

สิ่งที่สนใจศึกษาได้แก่อัตราข้อมูลของรหัสที่ได้หลังการบีบข้อมูล ความซับซ้อนของขั้นตอนวิธี และคุณภาพของสัญญาณเสียงพูดที่ได้ หัวข้อเหล่านี้เป็นหลักเกณฑ์สำคัญในการเลือกวิธีที่เหมาะสมสำหรับการพัฒนาเป็นโปรแกรมบีบข้อมูลเสียงที่ทำงานแบบทันที (real time) ซึ่งจะทำงานบนตัวประมวลผลสัญญาณดิจิตอล ADSP2101 อัตราข้อมูลที่ได้สำหรับวิธีแอลพีซี10 คือ 2.4 Kbps สำหรับวิธีซีอีแอลพีคือ 4.8 Kbps และสำหรับวิธีอาร์พีอี-แอลทีพีคือ 13 Kbps พบว่าวิธีซีอีแอลพีเป็นวิธีที่มีความซับซ้อนในการทำงานมากที่สุด รองลงมาคือวิธีอาร์พีอี-แอลทีพี ส่วนวิธีแอลพีซี10 มีความซับซ้อนน้อยที่สุด การเปรียบเทียบคุณภาพของเสียงพูดที่ได้ใช้คะแนนความเห็นจากผู้ทดสอบจำนวน 12 ท่าน ตัวอย่างเสียงพูดที่ใช้ในการทดลองเป็นเสียงชายหนึ่งตัวอย่างและเสียงหญิงหนึ่งตัวอย่าง วิธีแอลพีซี10 ได้คะแนนเฉลี่ย 5.3 วิธีซีอีแอลพีได้ 6.7 คะแนน ส่วนวิธีอาร์พีอีแอลทีพีได้คะแนนสูงสุดคือ 8.1 คะแนน ดังนั้นจึงได้เลือกวิธีอาร์พีอี-แอลทีพีในการพัฒนาโปรแกรมบีบข้อมูลเสียงพูดแบบทันทีเนื่องจากมีเป็นวิธีที่ให้คุณภาพของเสียงพูดจัดว่าดีและการทำงานไม่ซับซ้อนมากนัก โปรแกรมดังกล่าวใช้เวลาประมาณ 16.3 ms ในการบีบและคลายข้อมูลเสียงต่อหนึ่งเฟรม (frame) ซึ่งมีความยาว 20 ms

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ภาควิชา วิศวกรรมคอมพิวเตอร์ ลายมือชื่อนิสิต มีลาภ เรืองรัตนวิชา

สาขาวิชา วิทยาศาสตร์คอมพิวเตอร์ ลายมือชื่ออาจารย์ที่ปรึกษา ประภาส จงสถิตย์วัฒนา

ปีการศึกษา 2539 ลายมือชื่ออาจารย์ที่ปรึกษาร่วม .....

พิมพ์ต้นฉบับบทคัดย่อวิทยานิพนธ์ภายในกรอบสีเขียวนี้เพียงแผ่นเดียว

C618273 : MAJOR COMPUTER SCIENCE
KEY WORD: SPEECH COMPRESSION / LPC / CELP / RPE-LTP

MEELARP RUANGRATANAWICHA : SPEECH COMPRESSION BY DIGITAL SIGNAL PROCESSING APPROACH. THESIS ADVISOR : PRABHAS JONGSATHITWATANA, Ph.D. 92 pp. ISBN 974-635-914-2

The main objective of this research is to develop speech compression programs of which the speech quality levels are high enough for communication applications. The programs employ methods based on linear predictive coding (LPC) i.e. LPC10, CELP (Code Excited Linear Prediction) and RPE-LTP (Residual Pulse Excitation - Long Term Prediction). In the first phase, the programs were developed for working on a PC and accepted speech in wave file format (.wav) as the input.

The characteristics of each compression method e.g. the compression ratio or data rate after compression, algorithm complexity and the quality of the speech were studied and used as the criteria to choose one of them to implement as the real-time version which intended to run on the digital signal processor ADSP2101. The data rate after compression is 2.4 Kbps for LPC10, 4.8 Kbps for CELP and 13 Kbps for RPE-LTP. CELP is found as the most complicated method, RPE-LTP is the second and LPC10 is the least of the three methods. The quality of the speech of each method was compared by using the opinion score from 12 listeners. The experiment was performed with two sample files of man and woman speech. LPC10 got the average score of 5.3, CELP got 6.7 and RPE-LTP got the highest score of 8.1. RPE-LTP was therefore selected for real-time implementation due to its good quality of speech and moderate complexity. The estimated compression and decompression time of the real-time compression program is 16.3 ms for 20 ms speech frame.

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ภาควิชา วิศวกรรมคอมพิวเตอร์ ลายมือชื่อนิสิต มีลา เรืองรัตน์

สาขาวิชา วิทยาศาสตร์คอมพิวเตอร์ ลายมือชื่ออาจารย์ที่ปรึกษา ประภาส จงสถิตย์วัฒนา

ปีการศึกษา 2539 ลายมือชื่ออาจารย์ที่ปรึกษาร่วม

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลงได้ด้วยความร่วมมือและการสนับสนุนจากหลาย ๆ ท่าน ผู้วิจัยขอขอบพระคุณ อาจารย์ ดร. ประภาส จงสถิตย์วัฒนา อาจารย์ที่ปรึกษาวิทยานิพนธ์ที่ได้ให้ข้อมูลและชี้แนะแนวทางในการวิจัยอย่างดียิ่งเสมอมา ขอขอบคุณอาจารย์คณะกรรมการที่ได้ให้คำแนะนำอันมีประโยชน์ต่อการวิจัย และที่ผู้วิจัยจะขาดการกล่าวถึงเสียมิได้คือ บริษัท เซ็นจูรี่ เทเลคอม จำกัด ที่ได้ให้การสนับสนุนทางด้านการศึกษาแก่พนักงานมาโดยตลอด ผู้วิจัยขอขอบคุณผู้ร่วมงานและพนักงานบริษัททุกท่านที่ช่วยให้คะแนนความเห็นในช่วงการทำการทดลอง

ผู้วิจัยขอขอบคุณ คุณนพวรรณ คงเกษม ที่คอยอำนวยความสะดวกและติดต่อประสานงานทั้งกับทางภาควิชา และทางบัณฑิตวิทยาลัย ทำให้การวิจัยเป็นไปอย่างราบรื่น

และท้ายสุดนี้ผู้วิจัยขอขอบคุณญาติพี่น้อง และเพื่อนสนิทที่คอยให้กำลังใจมาโดยตลอด

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญ

หน้า

บทคัดย่อภาษาไทย ..... ง
บทคัดย่อภาษาอังกฤษ ..... จ
กิตติกรรมประกาศ ..... ฉ
สารบัญตาราง ..... ฌ
สารบัญภาพ ..... ญ
คำอธิบายคำย่อ ..... ฎ

บทที่

1. บทนำ 1
   - ความสำคัญและความเป็นมา 1
   - วัตถุประสงค์ในการวิจัย 2
   - ขั้นตอนการวิจัย 3
   - ขอบเขตการวิจัย 4
   - ประโยชน์ที่ได้รับ 4
2. แนวความคิดและทฤษฎี 6
   - วิธีการทั่วไปในการบีบขนาดข้อมูล 6
   - วิธีการในการบีบข้อมูลเสียง 7
   - การบีบข้อมูลเสียงพูดโดยวิธีการเข้ารหัสแบบทำนายเชิงเส้น 9
   - การบีบข้อมูลเสียงพูดโดยวิธีการทำนายเชิงเส้นโดยการกระตุ้นด้วยรหัส 21
   - การบีบข้อมูลเสียงพูดโดยวิธีการทำนายช่วงยาวโดยการกระตุ้นด้วยพัลส์เศษค้าง ... 22
3. การบีบข้อมูลเสียงพูดโดยวิธีการทำนายช่วงยาวโดยการกระตุ้นด้วยพัลส์เศษค้าง 24
   - การบีบข้อมูลเสียงด้วยวิธีอาร์พีอี-แอลทีพี 25
   - การคลายข้อมูลเสียงที่บีบด้วยวิธีอาร์พีอี-แอลทีพี 40
4. ตัวประมวลผลสัญญาณดิจิตอล 43
   - โครงสร้างภายในของ ADSP2101 44
   - คณิตศาสตร์ทศนิยมตายตัวสำหรับ ADSP2101 48
   - การควบคุมลำดับคำสั่ง 53

5. การวิจัยและการพัฒนา.....54
การพัฒนาโปรแกรมบีบข้อมูลเสียงพูดด้วยวิธีการเข้ารหัสแบบทำนายเชิงเส้น.....54
โปรแกรมอื่น ๆที่ใช้ในการบีบข้อมูลเสียงพูดบนพีซี.....58
ฮาร์ดแวร์สำหรับบีบข้อมูลเสียงพูดด้วยตัวประมวลผลสัญญาณดิจิตอล 60
โปรแกรมบีบข้อมูลเสียงพูดด้วยวิธีอาร์พีอี-แอลทีพีบนตัวประมวลผลสัญญาณดิจิตอล ADSP2101 61
โปรแกรมโหลดเดอร์.....64
6. การวัดคุณภาพของสัญญาณเสียงพูดโดยการใช้คะแนนความเห็นเฉลี่ย 65
ขั้นตอนการเก็บข้อมูลและผลที่ได้ 65
การวิเคราะห์ผลของคะแนนความเห็นที่ได้.....69
7. บทสรุปและข้อเสนอแนะ 78
สรุป 78
ข้อเสนอแนะ.....79
รายการอ้างอิง.....80
ภาคผนวก.....81
ประวัติผู้เขียน.....92

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญตาราง

หน้า

ตารางที่ 3.1 แสดงค่า A B ค่าสูงสุด ต่ำสุด และจำนวนบิตที่ใช้ในการควอนไตซ์และเข้ารหัสอัตราส่วนล็อกแอเรีย 30
ตารางที่ 3.2 แสดงค่า B MIC และ INVA ซึ่งใช้ในการแปลง LARc เป็น LAR 31
ตารางที่ 3.3 แสดงสัมประสิทธิ์ของตัวกรองเวท 38
ตารางที่ 5.1 แสดงคุณสมบัติการบีบข้อมูลเสียงพูดด้วยวิธีแอลพีซี10 ตามมาตรฐานเฟดเดอรัล 1015 58
ตารางที่ 5.2 แสดงคุณสมบัติการบีบข้อมูลเสียงพูดด้วยวิธี CELP ตามมาตรฐานเฟดเดอรัล 1015 60
ตารางที่ 6.1 แสดงคะแนนคุณภาพของสัญญาณเสียงพูดที่ได้จากโปรแกรมบีบและคลายข้อมูลเสียงวิธีต่าง ๆ 67
ตารางที่ 6.2 แสดงคะแนนคุณภาพสัญญาณเสียงพูดเปรียบเทียบระหว่างชายสองคนที่มีความชัดเจนของคำพูดต่างกัน 68
ตารางที่ 6.3 คะแนนของคุณภาพเสียงพูดที่ได้สำหรับการบีบและคลายสัญญาณแบบทันที ... 68
ตารางที่ 6.4 คะแนนของคุณภาพของเสียงพูดที่ได้จากการทดลองของ Finke 75
ตารางที่ 7.1 สรุปลักษณะเด่นของวิธีการบีบข้อมูลเสียงที่ศึกษา 78

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญภาพ

หน้า

รูปที่ 2.1 แสดงแบบจำลองในการเปล่งเสียงของมนุษย์ 12
รูปที่ 2.2 แสดงแบบจำลองของการเปล่งเสียงในรูปของฟังก์ชันโอนย้าย 14
รูปที่ 2.3 แสดงจุดที่มีการนำพารามิเตอร์ต่าง ๆไปใช้ในการสังเคราะห์เสียงด้วยวิธีแอลพีซี 16
รูปที่ 3.1 แสดงหลักการทำงานของการบีบข้อมูลเสียงโดยวิธี อาร์พีอี-แอลทีพี 25
รูปที่ 3.2 แสดงขั้นตอนย่อยของการวิเคราะห์แอลพีซี 26
รูปที่ 3.3 แสดงการแปลงสัมประสิทธิ์สะท้อนไปเป็นสัดส่วนล็อกแอเรียโดยการประมาณค่า .. 29
รูปที่ 3.4 แสดงการคำนวณหาสัมประสิทธิ์สะท้อนทั้ง 4 ชุดจาก LAR ในเฟรมปัจจุบัน
และ LAR ในเฟรมที่แล้ว 32
รูปที่ 3.5 แผนผังการไหลของสัญญาณในตัวกรองวิเคราะห์แบบผลึก 33
รูปที่ 3.6 แสดงการคลายข้อมูลเสียงที่บีบด้วยวิธี อาร์พีอี-แอลทีพี 37
รูปที่ 3.7 แผนผังการไหลของสัญญาณในตัวกรองสังเคราะห์ที่สร้างแบบผลึก 41
รูปที่ 4.1 แสดงโครงสร้างพื้นฐานของดีเอสพี ADSP2101 45
รูปที่ 5.1 แสดงขั้นตอนการทำงานของของโปรแกรมบีบข้อมูลเสียงพูด sc.exe 55
รูปที่ 5.2 แสดงส่วนหนึ่งของสัญญาณเสียงพูดที่ใช้ในการทดสอบโปรแกรม LPC 56
รูปที่ 5.3 แสดงสัญญาณออโตโครีเลชันที่ได้จากตัวอย่างสัญญาณเสียงพูด 56
รูปที่ 5.4 แสดงขั้นตอนการทำงานของโปรแกรมถอดรหัสข้อมูลเสียงพูด su.exe 57
รูปที่ 5.5 แสดงสัญญาณเสียงที่ได้หลังจากการบีบและคลายด้วยวิธีแอลพีซีแล้ว 57
รูปที่ 5.6 เปรียบเทียบรูปคลื่นของเสียงพูดที่ได้จากวิธีการบีบข้อมูลเสียงแต่ละวิธี 59
รูปที่ 5.7 แสดงโครงสร้างฮาร์ดแวร์สำหรับการบีบข้อมูลเสียงพูดด้วยวิธีอาร์พีอี-แอลทีพี
ด้วย เอดีเอสพี2101 61
รูปที่ 6.1 กราฟคะแนนคุณภาพของเสียงพูดที่ได้จากวิธีบีบและคลายต่าง ๆ 74

คำอธิบายคำย่อ

AC-3	Audio-Compression-3 วิธีการบีบข้อมูลเสียงพัฒนาโดยห้องปฏิบัติการดอลบี (Dolby) เหมาะสำหรับการสร้างเสียงในระบบรอบทิศทาง
ADPCM	Adaptive Differential Pulse Code Modulation วิธีเข้ารหัสข้อมูลเสียง
APCM	Adaptive Pulse Code Modulation วิธีเข้ารหัสข้อมูลเสียง
bps	bit per second บิตต่อวินาที
CCITT	International Telegraph & Telecommunication Consultative Committee คณะกรรมการที่ปรึกษาทางด้านโทรเลขและโทรคมนาคมนานาชาติ เป็นส่วนหนึ่งของสหภาพโทรคมนาคมนานาชาติ (International Telecommunication - ITU) มีหน้าที่ออกมาตรฐานทางด้านโทรคมนาคม
CELP	Code Excited Linear Prediction วิธีเข้ารหัสข้อมูลเสียง
CVSD	Continuously Variable Slope Delta Modulation วิธีเข้ารหัสข้อมูลเสียง
DSP	Digital Signal Processor ตัวประมวลผลสัญญาณดิจิตอล
Hz	Hertz หน่วยวัดความถี่
HF	High Frequency ย่านความถี่สูง
LD-CELP	Low-Delay - Code Excited Linear Prediction วิธีเข้ารหัสข้อมูลเสียง
LPC	Linear Predictive Coding การเข้ารหัสแบบทำนายเชิงเส้น
MOS	Mean Opinion Score คะแนนความเห็น(ต่อคุณภาพเสียง)เฉลี่ย
PCM	Pulse Code Modulation วิธีเข้ารหัสข้อมูลเสียง
QAM	Quadrature Amplitude Modulation วิธีการผสมสัญญาณ
RPE-LTP	Residual Pluse Excitation - Long-Term Prediction

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

	วิธีเข้ารหัสข้อมูลเสียง
VSELP	Vector Sum Excitation Long Term Prediction
	วิธีเข้ารหัสข้อมูลเสียง

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย