

การทำขันตอนวิธีเป็บข้อมูลเสียงพูดโดยการประมวลผลสัญญาณดิจิตอล

นายมีลาก เรืองรัตนวิชา



# ศูนย์วิทยทรัพยากร

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิชาศาสตรมหาบัณฑิต

ภาควิชาศึกกรรมคอมพิวเตอร์

บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2539

ISBN 974-635-914-2

ลิขสิทธิ์ของบัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย

AN IMPLEMENTATION OF A SPEECH COMPRESSION ALGORITHM  
BY DIGITAL SIGNAL PROCESSING

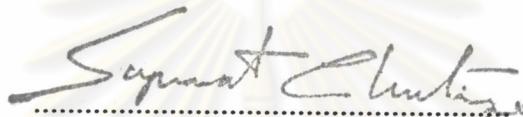
Mr.Meelarp Ruangratanawicha

ศูนย์วิทยบรพยากร  
จุฬาลงกรณ์มหาวิทยาลัย  
A Thesis Submitted in Partial Fulfillment of the Requirements  
for the Degree of Master of Science

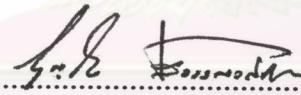
Graduate School  
Chulalongkorn University  
Academic Year 1996  
ISBN 974-635-914-2

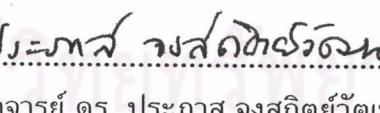
หัวข้อวิทยานิพนธ์ การทำขึ้นตอนวิธีเป็นข้อมูลเสียงพูดโดยการประมวลผลสัญญาณดิจิตอล  
โดย นายมีลาก เรืองรัตนวิชา  
ภาควิชา วิศวกรรมคอมพิวเตอร์  
อาจารย์ที่ปรึกษา อาจารย์ ดร. ประภาส จงสถิตย์วัฒนา

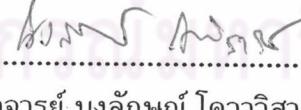
บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วน  
หนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

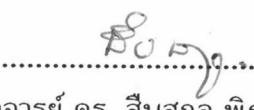
  
..... คณบดีบัณฑิตวิทยาลัย  
( ศาสตราจารย์ นายแพทย์ ศุภวัฒน์ ชุติวงศ์ )

คณะกรรมการสอบวิทยานิพนธ์

  
..... ประธานกรรมการ  
( ผู้ช่วยศาสตราจารย์ บุญชัย ஸوار眷ณิชกุล )

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

  
..... กรรมการ  
( อาจารย์ นงลักษณ์ โควิสารัช )

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

  
..... กรรมการ  
( อาจารย์ ดร. สุริต ศิริบูรณ์ )

## พิมพ์ต้นฉบับที่คัดย่อวิทยานิพนธ์ภาษาไทยในกรอบสีเขียวนี้เพียงแผ่นเดียว



มีลักษณะเป็นรัตนวิชา : การทำขันตอนวิธีบีบข้อมูลเสียงพูดโดยการประมวลผลสัญญาณดิจิตอล  
(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 หลักการทำงานของโปรแกรมที่พัฒนาขึ้นมีพื้นฐานมาจากวิธีการเข้ารหัสแบบทำงานเชิงเส้น (Linear Predictive Coding - LPC) วิธีการที่ใช้ได้แก่ แอลพีชี10 (LPC10) ซีอีเอลพี (CELP - Code Excited Linear Prediction) และ อาร์พีอี-แอลทีพี (RPE-LTP - Residual Pulse Excitation - Linear Prediction) ในขั้นแรกของการวิจัยได้พัฒนาโปรแกรมให้ทำงานอยู่บนคอมพิวเตอร์พีซีและรับข้อมูลเสียงพูดที่อยู่ในไฟล์ข้อมูลเวฟ (wave file) เป็นข้อมูลเข้า

สิ่งที่สนใจคือการได้แก่ตัวข้อมูลของรหัสที่ได้หลังการบีบข้อมูล ความซับซ้อนของขันตอนวิธี และคุณภาพของสัญญาณเสียงพูดที่ได้ หัวข้อเหล่านี้เป็นหลักเกณฑ์สำคัญในการเลือกวิธีที่เหมาะสมสำหรับการพัฒนาเป็นโปรแกรมบีบข้อมูลเสียงที่ทำงานแบบทันที (real time) ซึ่งจะทำงานบนตัวประมวลผลสัญญาณดิจิตอล ADSP2101 อัตราข้อมูลที่ได้สำหรับวิธีแอลพีชี10 คือ 2.4 Kbps สำหรับวิธีซีอีเอลพีคือ 4.8 Kbps และสำหรับวิธีอาร์พีอี-แอลทีพีคือ 13 Kbps พบว่าวิธีซีอีเอลพีเป็นวิธีที่มีความซับซ้อนในการทำงานมากที่สุด รองลงมาคือวิธีอาร์พีอี-แอลทีพี ส่วนวิธีแอลพีชี10 มีความซับซ้อนน้อยที่สุด การเปรียบเทียบคุณภาพของเสียงพูดที่ได้ใช้คะแนนความเห็นจากผู้ทดสอบจำนวน 12 ท่าน ตัวอย่างเสียงพูดที่ใช้ในการทดลองเป็นตัวอย่าง และเสียงหญิงหนึ่งตัวอย่าง วิธีแอลพีชี10 ได้คะแนนเฉลี่ย 5.3 วิธีซีอีเอลพีได้ 6.7 คะแนน ส่วนวิธีอาร์พีอี-แอลทีพีได้คะแนนสูงสุดคือ 8.1 คะแนน ตั้งนั้นจึงได้เลือกวิธีอาร์พีอี-แอลทีพีในการพัฒนาโปรแกรมบีบข้อมูลเสียงพูดแบบทันทีเนื่องจากมีเป็นวิธีที่ให้คุณภาพของเสียงพูดดีกว่าและการทำงานไม่ซับซ้อนมากนัก โปรแกรมดังกล่าวใช้เวลาประมาณ 16.3 ms ในการบีบและคลายข้อมูลเสียงต่อหนึ่งเฟรม (frame) ซึ่งมีความยาว 20 ms

จุฬาลงกรณ์มหาวิทยาลัย

พิมพ์ดันฉบับทัศน์อวิทยานิพนธ์ภาษาไทยในกรอบสีเขียวนี้เพียงแผ่นเดียว

# # 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

สาขาวิชา วิทยาศาสตร์คอมพิวเตอร์

ปีการศึกษา 2539

ลายมือชื่อนิสิต ฐานะ ปริญญา

ลายมือชื่ออาจารย์ที่ปรึกษา ประภาส วงศ์กุลธรรมนนท์

ลายมือชื่ออาจารย์ที่ปรึกษาร่วม

## กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลงได้ด้วยความร่วมมือและการสนับสนุนจากหลาย ๆ ท่าน ผู้วิจัยขอขอบพระคุณ อาจารย์ ดร. ประภาส จงสถิตย์วัฒนา อาจารย์ที่ปรึกษาวิทยานิพนธ์ ที่ได้ให้ข้อมูลและชี้แนะแนวทางในการวิจัยอย่างดีเยี่ยมอมา ขอขอบคุณอาจารย์คณะกรรมการที่ได้ให้คำแนะนำอันมีประโยชน์อย่างต่อเนื่อง ตลอดจนผู้วิจัยจะขาดการกล่าวถึงเสียไม่ได้ดีอ บริษัท เช็นจูรี่ เทเลคอม จำกัด ที่ได้ให้การสนับสนุนทางด้านการศึกษาแก่พนักงานโดยตลอด ผู้วิจัยขอขอบคุณผู้ร่วมงานและพนักงานบริษัททุกท่านที่ช่วยให้คะแนนความเห็นในช่วงการทำกราฟด่อง

ผู้วิจัยขอขอบคุณ คุณนพวรรณ คงเกษม ที่ค่อยอำนวยความสะดวกและติดต่อประสานงานทั้งกับทางภาควิชา และทางบัณฑิตวิทยาลัย ทำให้การวิจัยเป็นไปอย่างราบรื่น และท้ายสุดนี้ผู้วิจัยขอขอบคุณญาติพี่น้อง และเพื่อนสนิทที่ค่อยให้กำลังใจมาโดยตลอด

ศูนย์วิทยทรัพยากร  
จุฬาลงกรณ์มหาวิทยาลัย

## สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ง
บทคัดย่อภาษาอังกฤษ .....	จ
กิตติกรรมประกาศ .....	ฉ
สารบัญตาราง .....	ณ
สารบัญภาพ .....	ญ
คำอธิบายคำย่อ.....	ภ
บทที่	
1. บทนำ.....	1
ความสำคัญและความเป็นมา.....	1
วัตถุประสงค์ในการวิจัย.....	2
ขั้นตอนการวิจัย .....	3
ขอบเขตการวิจัย.....	4
ประโยชน์ที่ได้รับ .....	4
2. แนวความคิดและทฤษฎี .....	6
วิธีการทั่วไปในการบีบขนาดข้อมูล.....	6
วิธีการในการบีบข้อมูลเสียง.....	7
การบีบข้อมูลเสียงพุดโดยวิธีการทำนายเชิงเส้น .....	9
การทำนายเชิงเส้นโดยการกระตุนด้วยรหัส .....	21
การทำนายเชิงเสียงพุดโดยวิธีการทำนายช่วงยาวโดยการกระตุนด้วยพัลส์เศษค้าง .....	22
3. การบีบข้อมูลเสียงพุดโดยวิธีการทำนายช่วงยาวโดยการกระตุนด้วยพัลส์เศษค้าง .....	24
การบีบข้อมูลเสียงด้วยวิธีอาร์พีอี-แอลทีพี .....	25
การคลายข้อมูลเสียงที่บีบด้วยวิธีอาร์พีอี-แอลทีพี.....	40
4. ตัวประมวลผลสัญญาณดิจิตอล.....	43
โครงสร้างภายในของ ADSP2101 .....	44
คณิตศาสตร์ที่นิยมติดตัวสำหรับ ADSP2101 .....	48
การควบคุมลำดับคำสั่ง.....	53

5. การวิจัยและการพัฒนา.....	54
การพัฒนาโปรแกรมบีบข้อมูลเสียงพูดด้วยวิธีการเข้ารหัสแบบท่านายเชิงเส้น.....	54
โปรแกรมอื่น ๆ ที่ใช้ในการบีบข้อมูลเสียงพูดบนพีซี.....	58
ฮาร์ดแวร์สำหรับบีบข้อมูลเสียงพูดด้วยตัวประมวลผลสัญญาณดิจิตอล .....	60
โปรแกรมบีบข้อมูลเสียงพูดด้วยวิธีอาร์ฟีอี-แอลทีพีบันตัวประมวลผลสัญญาณ ดิจิตอล ADSP2101 .....	61
โปรแกรมโลลด์เดอร์.....	64
6. การวัดคุณภาพของสัญญาณเสียงพูดโดยการใช้คะแนนความเห็นเฉลี่ย .....	65
ขั้นตอนการเก็บข้อมูลและผลที่ได้ .....	65
การวิเคราะห์ผลของคะแนนความเห็นที่ได้.....	69
7. บทสรุปและข้อเสนอแนะ .....	78
สรุป .....	78
ข้อเสนอแนะ.....	79
รายการอ้างอิง.....	80
ภาคผนวก.....	81
ประวัติผู้เขียน.....	92

# ศูนย์วิทยบริการ จุฬาลงกรณ์มหาวิทยาลัย

## สารบัญตาราง

หน้า

ตารางที่ 3.1 แสดงค่า A B ค่าสูงสุด ต่ำสุด และจำนวนบิตที่ใช้ในการควบคุมตัวชี้และเข้ารหัสอัตราส่วนล็อกแอเรีย.....	30
ตารางที่ 3.2 แสดงค่า B MIC และ INVA ซึ่งใช้ในการแปลง LARc เป็น LAR .....	31
ตารางที่ 3.3 แสดงล้มประลิทธีของตัวกรองเวท .....	38
ตารางที่ 5.1 แสดงคุณสมบัติการบีบข้อมูลเสียงพูดด้วยวิธีแอลพีชี 10 ตามมาตรฐานเฟดเดอร์ล 1015 .....	58
ตารางที่ 5.2 แสดงคุณสมบัติการบีบข้อมูลเสียงพูดด้วยวิธี CELP ตามมาตรฐาน เฟดเดอร์ล 1015 .....	60
ตารางที่ 6.1 แสดงคะแนนคุณภาพของสัญญาณเสียงพูดที่ได้จากโปรแกรมบีบและคลายข้อมูลเสียงวิธีต่าง ๆ .....	67
ตารางที่ 6.2 แสดงคะแนนคุณภาพสัญญาณเสียงพูดเปรียบเทียบระหว่างชายสองคน ที่มีความชัดเจนของคำพูดต่างกัน .....	68
ตารางที่ 6.3 คะแนนของคุณภาพเสียงพูดที่ได้สำหรับการบีบและคลายสัญญาณแบบทันที ...	68
ตารางที่ 6.4 คะแนนของคุณภาพของเสียงพูดที่ได้จากการทดลองของ Finke.....	75
ตารางที่ 7.1 สรุปลักษณะเด่นของวิธีการบีบข้อมูลเสียงที่ศึกษา .....	78

ศูนย์วิทยทรัพยากร  
จุฬาลงกรณ์มหาวิทยาลัย

## สารบัญภาพ

หน้า

รูปที่ 2.1 แสดงแบบจำลองในการเปลี่ยนเสียงของมนุษย์.....	12
รูปที่ 2.2 แสดงแบบจำลองของการเปลี่ยนเสียงในรูปของฟังก์ชันโอนยาัย .....	14
รูปที่ 2.3 แสดงจุดที่มีการนำพารามิเตอร์ต่าง ๆ ไปใช้ในการสังเคราะห์เสียงด้วยวิธีแอลพีซี....	16
รูปที่ 3.1 แสดงหลักการทำงานของการบีบข้อมูลเสียงโดยวิธี อาร์พีอี-แอลทีพี .....	25
รูปที่ 3.2 แสดงขั้นตอนอย่างของการวิเคราะห์แอลพีซี.....	26
รูปที่ 3.3 แสดงการแปลงสัมประสิทธิ์สะท้อนไปเป็นสัดส่วนล็อกแອเรียโดยการประมาณค่า ..	29
รูปที่ 3.4 แสดงการคำนวนหาสัมประสิทธิ์สะท้อนทั้ง 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 แสดงโครงสร้างพื้นฐานของดีอีสพี ADSP2101.....	45
รูปที่ 5.1 แสดงขั้นตอนการทำงานของของโปรแกรมบีบข้อมูลเสียงพุด sc.exe .....	55
รูปที่ 5.2 แสดงส่วนหนึ่งของสัญญาณเสียงพุดที่ใช้ในการทดสอบโปรแกรม LPC.....	56
รูปที่ 5.3 แสดงสัญญาณอโตเครีเลชันที่ได้จากการบีบข้อมูลเสียงพุด .....	56
รูปที่ 5.4 แสดงขั้นตอนการทำงานของโปรแกรมถอดรหัสข้อมูลเสียงพุด su.exe .....	57
รูปที่ 5.5 แสดงสัญญาณเสียงที่ได้หลังจากการบีบและคลายด้วยวิธีแอลพีซีแล้ว .....	57
รูปที่ 5.6 เปรียบเทียบรูปคลื่นของเสียงพุดที่ได้จากวิธีการบีบข้อมูลเสียงแต่ละวิธี .....	59
รูปที่ 5.7 แสดงโครงสร้างฮาร์ดแวร์สำหรับการบีบข้อมูลเสียงพุดด้วยวิธีอาร์พีอี-แอลทีพี ด้วย เอเดิลสพี2101 .....	61
รูปที่ 6.1 กราฟคะแนนคุณภาพของเสียงพุดที่ได้จากวิธีบีบและคลายต่าง ๆ .....	74

## คำอธิบายคำย่อ

AC-3	Audio-Compression-3 วิธีการบีบข้อมูลเสียงพัฒนาโดยห้องปฏิบัติการดอลบี (Dolby) เพื่อการส่งสัญญาณเสียงในระบบโทรทัศน์
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

VSELP

วิธีเข้ารหัสข้อมูลเสียง

Vector Sum Excitation Long Term Prediction

วิธีเข้ารหัสข้อมูลเสียง



# ศูนย์วิทยทรัพยากร จุฬาลงกรณ์มหาวิทยาลัย