ต้นแบบเครื่องสังเคราะห์เสียงพูดด้วยวิธีเข้ารหัสแบบลิเนียร์พรีดิกทีฟ

หอสมุดกลาง สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

นายกฤษดา เรเยส

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรมหาบัณฑิต

ภาควิชาวิศวกรรมไฟฟ้า

บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย

พ.ศ. 2530

ISBN 974-568-132-6

ลิขสิทธิ์ของบัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย

012651

A SPEECH SYNTHESIZER PROTOTYPE
USING LINEAR PREDICTIVE CODING

หอสมุดกลาง สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Mr. Krisada Reyes

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Engineering

Department of Electrical Engineering
Graduate School
Chulalongkorn University

1987

หัวข้อวิทยานิพนธ์ ต้นแบบเครื่องสังเคราะห์เสียงพูดด้วยวิธีเข้ารหัสแบบลิเนียร์พรีดิกทีฟ

โดย นายกฤษดา เรเยส

ภาควิชา วิศวกรรมไฟฟ้า

อาจารย์ที่ปรึกษา รองศาสตราจารย์ ดร.ประสิทธิ์ ประพิณมงคลการ

อาจารย์ที่ปรึกษาร่วม รองศาสตราจารย์ กฤษดา วิศวธีรานนท์

หอสมุดกลาง สถาบันวิทยบริการ จุฬาลงกรณ์มหาวิทยาลัย

บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

.... คณบดีบัณฑิตวิทยาลัย

(ศาสตราจารย์ ดร.ถาวร วัชราภัย)

คณะกรรมการสอบวิทยานิพนธ์

.... ประธานกรรมการ

(ศาสตราจารย์ อาภรณ์ เก่งพล)

.... กรรมการ

(ศาสตราจารย์ ดร.มงคล เดชนครินทร์)

.... กรรมการ

(รองศาสตราจารย์ ดร.ชาตรี ศรีไพพรรณ)

.... กรรมการ

(รองศาสตราจารย์ ดร.สมชาย จิตะพันธ์กุล)

.... กรรมการ

(รองศาสตราจารย์ ดร.ประสิทธิ์ ประพิณมงคลการ)

.... กรรมการ

(รองศาสตราจารย์ กฤษดา วิศวธีรานนท์)

หัวข้อวิทยานิพนธ์	ต้นแบบเครื่องสังเคราะห์เสียงพูดด้วยวิธีเข้ารหัสแบบลิเนียร์พรีดิกทีฟ
ชื่อนิสิต	นายกฤษดา เรเยส
อาจารย์ที่ปรึกษา	รองศาสตราจารย์ ดร.ประสิทธิ์ ประพิณมงคลการ
อาจารย์ที่ปรึกษาร่วม	รองศาสตราจารย์ กฤชดา วิศวธีรานนท์
ภาควิชา	วิศกรรมไฟฟ้า
ปีการศึกษา	2529

บทคัดย่อ

วัตถุประสงค์ของงานวิจัยนี้คือ การสร้างระบบวิเคราะห์และสังเคราะห์เสียงพูดโดยอาศัยวิธีเข้ารหัสแบบลิเนียร์พรีดิกทีฟซึ่งเป็นวิธีการที่แพร่หลายวิธีหนึ่ง ระบบวิเคราะห์และสังเคราะห์เสียงพูดที่สร้างขึ้นประกอบด้วย ระบบไมโครคอมพิวเตอร์ ภาคแปลงสัญญาณระหว่างสัญญาณอนาลอกกับสัญญาณเชิงเลข และภาคประมวลผลสัญญาณ การทดลองในงานวิจัยนี้จะสุ่มตัวอย่างเสียงพูดด้วยความถี่ 10 kHz นำมาบันทึกลงหน่วยความจำภายในเครื่องไมโครคอมพิวเตอร์ แล้วคำนวณหาค่าพารามิเตอร์ซึ่งเป็นตัวแทนของสัญญาณเสียง พารามิเตอร์เหล่านี้สามารถนำไปผลิตเป็นสัญญาณเสียงพูดได้ด้วยกรรมวิธีสังเคราะห์เสียง จากการประยุกต์ใช้วงจรรวมประมวลผลสัญญาณ ส่วนของการสังเคราะห์เสียงพูดสามารถทำงานได้ในลักษณะ Real-time ในการทำให้อัตราข้อมูลเสียงน้อยที่สุดโดยไม่ใช้กระบวนการ Optimum Parameter Quantization จะอาศัยการทดลองโดยเปลี่ยนแปลงจำนวนออร์เดอร์ของฟิลเตอร์ระหว่าง 2 ถึง 15 และเปลี่ยนแปลงขนาดของเฟรมระหว่าง 100 ถึง 300 แซมเปิล จากนั้นเสียงที่ได้จากการสังเคราะห์จะถูกประเมินคุณภาพด้วยค่า Normalized Error และจากการฟัง เพื่อหาองค์ประกอบในการสังเคราะห์เสียงที่ให้อัตราข้อมูลเสียงน้อยที่สุด ท้ายสุดเป็นการแสดงการทำงานของระบบโดยการสังเคราะห์เสียงเป็นคำ ๆ จากตัวอย่างของคำพูดที่ได้ทำการทดลอง

Thesis Title	A SPEECH SYNTHESIZER PROTOTYPE USING LINEAR PREDITIVE CODING
Name	Mr. Krisada Reyes
Thesis Advisor	Associate Professor Prasit Prapinmongkolkarn, Ph.D.
Co-advisor	Associate Professor Krisada Visavatheeranond, M.Eng.
Department	of Electrical Engineering
Academic Year	1986

หอสมุดกลาง สถาบันวิทยบริการ จุฬาลงกรณ์มหาวิทยาลัย

ABSTRACT

The objective of this study is to construct a speech analysis and synthesis system with application of Linear Predictive Coding, one of the most widely used techniques. The system consists of a microcomputer system, an analog-to-digital, digital-to-analog converter and a signal processor section. In the experiment, the speech waveform is sampled at 10 kHz sampling rate then stored and processed in the microcomputer system to calculate LPC speech parameters representing the speech waveform. From these parameters, speech waveform can be reproduced by synthesis algorithm. By using a signal processor integrated circuit, the system can provide real-time calculation of synthesis speech samples. To minimize data rate without using Optimum Parameter Quantization, in the experiment, the filter order is varied between 2 to 15 and the frame length is varied between 100 to 300 samples. Then the synthesized sound quality is evaluated by means of Normalized Error Analysis and listening test to obtain minimized data rate synthesis configuration. The research will include a group of synthesized utterance as an example of the system performance.

กิตติกรรมประกาศ

ข้าพเจ้าขอขอบคุณท่านอาจารย์ที่ปรึกษาคือ รองศาสตราจารย์ ดร. ประสิทธิ์ ประพิณมงคลการ และรองศาสตราจารย์ กฤษดา วิศวธีรานนท์ ในฐานะอาจารย์ที่ปรึกษาร่วม ที่ได้กรุณาสละเวลาให้คำปรึกษาชี้แนะ ตลอดจนช่วยผลักดันให้วิทยานิพนธ์นี้สำเร็จลุล่วงไปได้ นอกจากท่านทั้งสองนี้แล้ว ข้าพเจ้าขอขอบคุณท่านที่มีส่วนสนับสนุนงานวิทยานิพนธ์ฉบับนี้ให้สำเร็จลุล่วงไปด้วยดี ดังมีรายนามต่อไปนี้

1. น.ส.ภาสนันท์ จิรฐานกุล
2. นายอนิรุทธิ์ รัชตะวราห์
3. เจ้าหน้าที่ประจำแผนกเครื่องฝึกบินจำลอง บริษัทการบินไทย จำกัด

สารบัญ

หอสมุดกลาง สถาบันวิทยบริการ จุฬาลงกรณ์มหาวิทยาลัย

หน้า

บทคัดย่อภาษาไทย .... ง

บทคัดย่อภาษาอังกฤษ .... จ

กิตติกรรมประกาศ .... ฉ

สารบัญตาราง .... ฌ

สารบัญภาพ .... ญ

บทที่

1. บทนำ 1

2. การสังเคราะห์เสียงพูด 5

2.1 เสียงพูด 5

2.2 หลักการสังเคราะห์เสียงพูด 9

2.2.1 การนำสัญญาณเข้ารหัสเชิงเลข 9

2.2.2 การวิเคราะห์ในโดเมนเวลา 11

2.2.3 การวิเคราะห์และสังเคราะห์ทางสเปกตรัม 12

2.2.4 การกรองแบบโฮโมมอร์ฟิค 16

2.2.5 การเข้ารหัสแบบลิเนียร์พรีดิกทีฟ 18

2.2.6 สรุปหลักการสังเคราะห์เสียงพูด 21

2.3 การหาฟิลเตอร์พารามิเตอร์ของวิธีแอลพีซี 23

2.3.1 แบบจำลองการผลิตเสียงพูดตามวิธีแอลพีซี 23

2.3.2 การหาคำตอบด้วยวิธีโควาเรียนซ์ 26

2.3.4 วิธีพาร์คอร์ 30

2.3.5 หลักการอินเนอร์โพรดักท์และคุณสมบัติออโธโกนัลของตัวทำนาย 32

2.3.6 การหาสัมประสิทธิ์พาร์คอร์ 34

2.3.7 อินเวอสฟิลเตอร์จากสัมประสิทธิ์ 37

2.3.8 กรรมวิธีคำนวณค่าสัมประสิทธิ์พาร์คอร์ 39

2.4 การหาคาบของสัญญาณเสียงพูด 46

2.4.1 สัญญาณค่าผิดพลาดของสัญญาณเสียง 46

2.4.2 หลักการหาคาบของสัญญาณ 47

2.4.3 กรรมวิธี SIFT 49

บทที่ หน้า

2.5 การคำนวณเพื่อสังเคราะห์เสียง 57
2.5.1 ขนาดของสัญญาณ Excitation 57
2.5.2 ฟิลเตอร์สังเคราะห์ตามวิธี PARCOR 59
3. ระบบวิเคราะห์และเครื่องต้นแบบสังเคราะห์เสียงพูด 64
3.1 มาสเตอร์อินเตอร์เฟซ 67
3.2 ภาคอนาลอกโปรเซสเซอร์ 74
3.3 ภาคอนาลอกอินเตอร์เฟซ 79
3.4 ภาคประมวลผลสัญญาณ 85
3.4.1 รายละเอียดส่วนประกอบของบอร์ด SP-B 86
3.4.2 รายละเอียดส่วนประกอบของบอร์ด SP-A 88
3.4.3 วงจรภาคประมวลผลสัญญาณบอร์ด SP-B 90
3.4.4 วงจรภาคประมวลผลสัญญาณบอร์ด SP-A 91
4. ซอฟท์แวร์ 100
4.1 โปรแกรมวิเคราะห์-สังเคราะห์เสียงพูดในมาสเตอร์คอมพิวเตอร์ 102
4.1.1 โปรแกรม SIGNAL EDITOR 103
4.1.2 โปรแกรม LPCX 108
4.1.3 โปรแกรม SIFTX 110
4.1.4 โปรแกรม SYNTX 112
4.2 โปรแกรมที่เกี่ยวกับการสังเคราะห์เสียงด้วยภาคประมวลผลสัญญาณ ... 114
4.2.1 โปรแกรม DCONX 114
4.2.2 โปรแกรม SPCONT 118
4.2.3 โปรแกรม SPX 120
5. การทดลองและประเมินผล 128
5.1 การทดลองเปรียบเทียบคุณภาพเสียงที่ได้จากการสังเคราะห์ด้วยจำนวนออร์เดอร์ต่างๆ 131
5.2 การทดลองเปรียบเทียบคุณภาพเสียงที่ได้จากการสังเคราะห์ด้วยจำนวนแซมเปิลในหนึ่งเฟรมต่างๆ 142
5.3 การทดลองวิเคราะห์และสังเคราะห์เสียงพูดของตัวเลขหนึ่งถึงสิบ 151
6. บทสรุป 152
เอกสารอ้างอิง 155
ภาคผนวก 158
ประวัติ 164

สารบัญตาราง

ตาราง หน้า

3.1 รายละเอียดของพอร์ทที่ใช้ในการติดต่อและควบคุมระบบ 68

3.2 ผลการทดลองการทำงานของวงจรกรองความถี่ 77

3.3 หน้าที่ของแต่ละบิทในรีจิสเตอร์ควบคุมและสัญญาณที่ผลิต 87

4.1 ตำแหน่งในหน่วยความจำที่ใช้เก็บข้อมูลของโปรแกรม SIGNAL EDITOR 105

5.1 ค่าลดทอนข้อมูลเมื่อกำหนดค่า M และ N ต่าง ๆ 143

สารบัญภาพ

รูป หน้า

รูป		หน้า
2.1	ส่วนประกอบของอวัยวะผลิตเสียง	6
2.2	เป็นตัวอย่างในการศึกษาคุณสมบัติของเสียงพูด	7
2.3	สเปกตรัมของเสียงสระ "e"	8
2.4	แผนภาพหลักการทำงานของดิฟเฟอร์เรนเชียลควอนไทเซชั่น	10
2.5	ส่วนประกอบของ Channel Vocoder	13
2.6	ส่วนประกอบของ Formant Vocoder ยุคแรก	14
2.7	Formant Synthesizer ยุคปัจจุบัน	15
2.8	ส่วนประกอบของภาควิเคราะห์เสียงพูดด้วยวิธีการกรองแบบโฮโมมอร์ฟิค ..	16
2.9	ส่วนประกอบของภาคสังเคราะห์เสียงพูดด้วยวิธีการกรองแบบโฮโมมอร์ฟิค ..	17
2.10	แบบจำลองการผลิตเสียงพูดด้วยวิธีแอลพีซี	19
2.11	แบบจำลองของการผลิตเสียงพูด	21
2.12	การเปรียบเทียบประสิทธิภาพของวิธีผลิตเสียงพูด	22
2.13	แบบจำลองการวิเคราะห์และสังเคราะห์เสียงพูดด้วยวิธีแอลพีซี	25
2.14	ชุดของสัญญาณสุ่มและการทำนายทั้งสองทาง	30
2.15	อินเนอร์โพรดักท์ของฟิลเตอร์ F(z) กับ G(z)	32
2.16	โครงสร้างของอินเวิร์สฟิลเตอร์ที่ใช้วิธีพาร์คอร์	38
2.17.1	ชื่อตัวแปรในโปรแกรม AUTO เทียบกับสัญญลักษณ์ในสมการ	43
2.17.2	โปรแกรมย่อย AUTO	44
2.18	สัญญาณเสียงพูดและสัญญาณค่าผิดพลาดของคำว่า "Shade"	46
2.19	สเปกตรัมของสัญญาณเสียง และสัญญาณค่าผิดพลาดของเสียงสระอี	47
2.20	สัญญาณเสียงและค่าผิดพลาดของเสียง /n/ ในคำว่า "Linear"	48
2.21	สัญญาณค่าผิดพลาดของเสียง /n/ ที่ผ่านการคูณด้วยฟังค์ชั่น $1 - Z^{-1}$	48
2.22	ขั้นตอนของกรรมวิธี SIFT	49
2.23	โครงสร้างของฟิลเตอร์แบบ Direct Form	51
2.24	โปรแกรมย่อย Step1	52
2.25	โปรแกรมย่อย Direct	54
2.26	โปรแกรมย่อย Step2	55
2.27	โครงสร้างของฟิลเตอร์สังเคราะห์	59

รูป หน้า

2.28 โปรแกรมย่อย TWOMUL 61

2.29 โปรแกรมย่อย SYNT 62

3.1 ส่วนประกอบของระบบวิเคราะห์และเครื่องต้นแบบสังเคราะห์เสียงพูด 65

3.2 ส่วนประกอบของมาสเตอร์อินเตอร์เฟซ 67

3.3.1 วงจรมาสเตอร์อินเตอร์เฟซ 71-72

3.3.2 ตำแหน่งของสัญญาณต่างๆ ที่ Connector สำหรับส่งไปสู่ภาคประมวลผลสัญญาณ 73

3.4 ส่วนประกอบของภาคสัญญาณเสียงเข้า 74

3.5 ส่วนประกอบของภาคสัญญาณเสียงออก 75

3.6.1 วงจรสร้างสัญญาณนาฬิกาสำหรับวงจรกรองความถี่ 76

3.6.2 วงจรกรองความถี่ผ่านต่ำ 78

3.7 ส่วนประกอบของภาคแปลงสัญญาณอนาลอกเป็นสัญญาณเชิงเลข 79

3.8 ส่วนประกอบของภาคแปลงสัญญาณเชิงเลขเป็นสัญญาณอนาลอก 80

3.9.1 วงจรภาคแปลงสัญญาณอนาลอกเป็นสัญญาณเชิงเลข 82

3.9.2 วงจรภาคแปลงสัญญาณเชิงเลขเป็นสัญญาณอนาลอก 83

3.9.3 อุปกรณ์และตำแหน่งของสัญญาณที่ภาคอนาลอกอินเตอร์เฟซ 84

3.10 ส่วนประกอบของภาคประมวลผลสัญญาณ 85

3.11 ตำแหน่งสัญญาณใน MAIN BUS CONNECTOR 93

3.12.1 วงจรภาคประมวลผลสัญญาณบอร์ด SP-B 94-95

3.13.1 วงจรภาคประมวลผลสัญญาณบอร์ด SP-A 96-99

4.1 โปรแกรมที่ใช้ในการวิเคราะห์ และสังเคราะห์เสียง 101

4.2 คำสั่งในโปรแกรม SIGNAL EDITOR 103

4.3 แผนผังการทำงานของโปรแกรมย่อย AIN 106

4.4 แผนผังการทำงานของโปรแกรมย่อย AOUT 107

4.5 แผนผังการทำงานของโปรแกรม LPCX 109

4.6 แผนผังการทำงานของโปรแกรม SIFTX 111

4.7 แผนผังการทำงานของโปรแกรม SYNTX 113

4.8 คำสั่งในโปรแกรม DCONX 115

4.9 การจัดเรียงข้อมูลที่ทำการแปลงแล้ว 117

4.10 คำสั่งในโปรแกรม SPCONT 118

รูป		หน้า

4.11 การจัดข้อมูลเสียงพูดสำหรับโปรแกรม SPX 121
4.12 แผนผังการทำงานของโปรแกรมหลัก (MAIN PROGRAM) 123
4.13 แผนผังการทำงานส่วนของโปรแกรมสำหรับออกเสียงพูด (SPEAK) 125
4.14 แผนผังการทำงานส่วนของโปรแกรมสำหรับคำนวณสัญญาณเสียง (NLOOP) .. 127
5.1 สัญญาณและสเปกตรัมของเสียงสระอาต้นแบบ 133
5.2 สัญญาณและสเปกตรัมของเสียงสระอา ที่ได้จากการสังเคราะห์ เมื่อ M=2... 134
5.3 สัญญาณและสเปกตรัมของเสียงสระอา ที่ได้จากการสังเคราะห์ เมื่อ M=4... 135
5.4 สัญญาณและสเปกตรัมของเสียงสระอา ที่ได้จากการสังเคราะห์ เมื่อ M=6... 136
5.5 สัญญาณและสเปกตรัมของเสียงสระอา ที่ได้จากการสังเคราะห์ เมื่อ M=8... 137
5.6 สัญญาณและสเปกตรัมของเสียงสระอา ที่ได้จากการสังเคราะห์ เมื่อ M=10.. 138
5.7 สัญญาณและสเปกตรัมของเสียงสระอา ที่ได้จากการสังเคราะห์ เมื่อ M=12.. 139
5.8 สัญญาณและสเปกตรัมของเสียงสระอา ที่ได้จากการสังเคราะห์ เมื่อ M=15.. 140
5.9 ผลการทดลองคำนวณค่าผิดพลาดนอร์มัลไลซ์ของเสียงสระอา 141
5.10 สัญญาณและความถี่หลักมูลของเสียงพูดคำว่า"ก้า" 144
5.11 ค่าผิดพลาดนอร์มัลไลซ์ที่ส่วนต่างๆ ของคำว่า"ก้า" เมื่อ N=100 145
5.12 ค่าผิดพลาดนอร์มัลไลซ์ที่ส่วนต่างๆ ของคำว่า"ก้า" เมื่อ N=150 146
5.13 ค่าผิดพลาดนอร์มัลไลซ์ที่ส่วนต่างๆ ของคำว่า"ก้า" เมื่อ N=200 147
5.14 ค่าผิดพลาดนอร์มัลไลซ์ที่ส่วนต่างๆ ของคำว่า"ก้า" เมื่อ N=250 148
5.15 ค่าผิดพลาดนอร์มัลไลซ์ที่ส่วนต่างๆ ของคำว่า"ก้า" เมื่อ N=300 149
5.16 เปรียบเทียบค่าเฉลี่ยค่าผิดพลาดนอร์มัลไลซ์จากทุกส่วนของคำพูด"ก้า" รวมทั้งค่าลดทอนข้อมูล 150
5.17 ผลการทดลองวิเคราะห์-สังเคราะห์คำพูด "หนึ่ง" ถึง "สิบ" 151