

การสังเคราะห์เสียงพูดจากข้อความภาษาไทย



นาย อากาศ นันทิกุล

ศูนย์วิจัยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

วิทยานี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

ภาควิชาวิศวกรรมคอมพิวเตอร์

บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย

พ.ศ. 2533

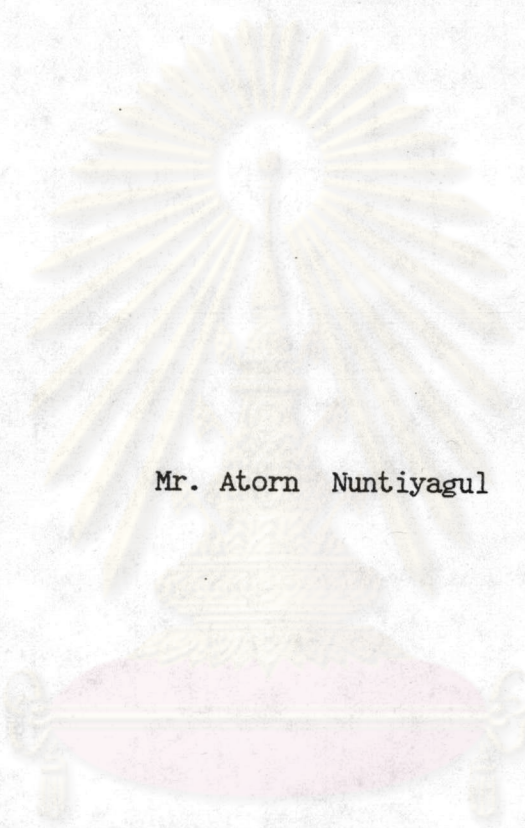
ISBN 974-577-629-7

ลิขสิทธิ์ของบัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย

016630

i 10309317

THAI TEXT TO SPEECH SYNTHESIS



Mr. Atorn Nuntiyagul

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

A Thesis Submitted in Partial Fulfillment of the Requirements

for the Degree of Master of Science

Department of Computer Engineer

Graduate School

Chulalongkorn University


1990

ISBN 974-577-629-7

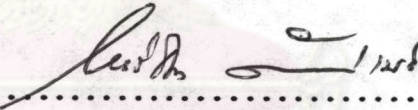


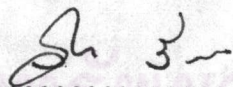
หัวข้อวิทยานิพนธ์ การสังเคราะห์เสียงพูดจากข้อความภาษาไทย
โดย นาย อาทร นันทิกุล
ภาควิชา วิศวกรรมคอมพิวเตอร์
อาจารย์ที่ปรึกษา ผู้ช่วยศาสตราจารย์ ดร. วีระ ธีรวิทักษ์

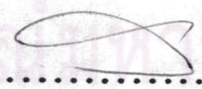
บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้บัณฑิตวิทยาลัยฉบับนี้ เป็นส่วนหนึ่ง
ของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

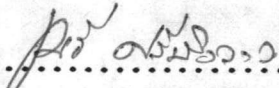

..... คณะบดีบัณฑิตวิทยาลัย
(ศาสตราจารย์ ดร. อวาร ธีรวิทักษ์)

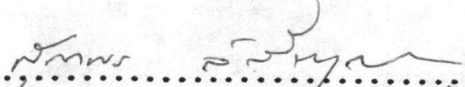
คณะกรรมการสอบวิทยานิพนธ์


..... ประธานกรรมการ
(รองศาสตราจารย์ ไกรวิชิต ตันติเมธ)


..... อาจารย์ที่ปรึกษา
(ผู้ช่วยศาสตราจารย์ ดร. วีระ ธีรวิทักษ์)


..... กรรมการ
(รองศาสตราจารย์ สมชาย ทยานยง)


..... กรรมการ
(ผู้ช่วยศาสตราจารย์ เมธี ศรีสังวาล)


..... กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร. สุตานร ลักขณีนาวิน)



อาทร นันทิชกุล : การสังเคราะห์เสียงพูดจากข้อความภาษาไทย (THAI TEXT TO SPEECH SYNTHESIS) อ.ที่ปรึกษา : ผศ.ดร.วีระ รวีวิทักษ์ , 120 หน้า.
ISBN 974-577-629-7

การสังเคราะห์เสียงพูดจากข้อความภาษาไทยเป็นการสร้างเสียงพูดขึ้นมาจากข้อความภาษาไทยที่ถูกป้อนเป็นอินพุตเข้าสู่ระบบ ซึ่งได้ประยุกต์ลงบนไมโครคอมพิวเตอร์และมีภาคการแปลงสัญญาณระหว่างสัญญาณอนาล็อกกับสัญญาณดิจิทัลรวมอยู่ด้วย โดยใช้หลักของการวิเคราะห์หน่วยย่อยของเสียงพูดคือ พยางค์ ของเสียงต้นแบบมาทำการตัดตัวอย่างด้วยความถี่ 10 kHz และนำมาทำการวิเคราะห์ด้วยเทคนิคการทำนายแบบเชิงเส้น (LPC) แบบออร์เดอร์ 10 ทีละเฟรม (เฟรมละ 200 จุดหรือ 20 มิลลิวินาที) ได้เอาที่พุดออกมาเป็นชุดพารามิเตอร์ประกอบด้วย 1) ค่าความผิดพลาดเฉลี่ย 2) ค่าคาบของพิทช์ และ 3) ค่าสัมประสิทธิ์ของการทำนาย (10 ค่า) ต่อหนึ่งเฟรม เก็บเอาไว้ในหน่วยเก็บความจำสำรองในรูปแบบของพจนานุกรมข้อมูลซึ่งสามารถจะทำการแก้ไขพารามิเตอร์เพื่อปรับปรุงให้ได้เสียงสังเคราะห์ที่ดีขึ้น จากนั้นก็จะทำการสร้างเสียงพูดสังเคราะห์ขึ้นมา โดยนำเอาชุดพารามิเตอร์ของหน่วยย่อยของเสียงที่ได้มาจากการค้นหาในพจนานุกรมข้อมูลตามข้อความที่ป้อนเข้ามา มาทำการสังเคราะห์ผ่านตัวกรองแลตทิส (Lattice filter) และภาคการแปลงสัญญาณดิจิทัลเป็นสัญญาณอนาล็อกออกมาเป็นเสียงพูด โดยคุณภาพของเสียงที่สังเคราะห์ออกมาอยู่ในระดับปานกลางซึ่งวัดผลได้จากการรับฟังของกลุ่มตัวอย่าง และเสียงสังเคราะห์นั้นจะมีคุณสมบัติของเสียงวรรณยุกต์ซึ่งเป็นลักษณะสำคัญของภาษาไทยด้วย

การป้อนข้อความอินพุตจะป้อนในรูปแบบของตัวสะกดตามคำอ่านเท่านั้น ระบบการสังเคราะห์เสียงพูดจากข้อความภาษาไทยนี้ยังต้องมีการพัฒนาในส่วนของการแปลงข้อความที่เป็นภาษาเขียนให้กลายเป็นคำอ่านอีก แต่อย่างไรก็ตามระบบการสังเคราะห์เสียงพูดนี้ก็ ได้ใช้เทคนิคการสังเคราะห์เสียงพูดขึ้นมาจากชุดพารามิเตอร์โดยใช้เป็นซอฟต์แวร์ทั้งหมด ซึ่งทำให้สามารถนำไปประยุกต์ใช้ได้ง่ายและกว้างขวางสำหรับระบบการสังเคราะห์เสียงพูดแบบฐานความรู้ทั่วไป

จุฬาลงกรณ์มหาวิทยาลัย

ภาควิชา วิศวกรรมคอมพิวเตอร์
สาขาวิชา วิทยาการศาสตร์คอมพิวเตอร์
ปีการศึกษา 2532

ลายมือชื่อนิสิต Dr. วีระ
ลายมือชื่ออาจารย์ที่ปรึกษา Sh. 3

ลายมือชื่ออาจารย์ที่ปรึกษาร่วม



ATORN NUNTIYAGUL : THAI TEXT TO SPEECH SYNTHESIS. THESIS
ADVISOR : ASST. PROF.VEERA REWPITAK. 120 pp.

This Thai text to speech synthesis system reproduces speech from Thai text input. The system is developed on microcomputers. The microcomputer used needs only an analog-digital signal converter unit. The knowledge representation for the system is syllable based. This representation is sampled at 10 KHZ and then analyzed by the Linear Prediction Coding (LPC) technique with order 10 frame by frame (1 frame is 200 samples or 20 mS). The outputs from the LPC analysis consists of a set of parameters 1) the root mean-square error value 2) pitch period and 3) the value of predictor coefficient (10/frame). These arithmetic values are then stored in a storage device in terms of a data dictionary. This stored data can be edited for the improvement of the synthesized speech. The synthesis process starts from searching and retrieving of the set of parameters from the dictionary as text input then passing the set of values to the lattice filter and to the D/A converter. The synthesized speech is well recognized by test subjects, the quality is quite good, especially the tonal feature of the Thai language is very clear.

The text input is not the ordinary Thai writing system but rather the pronunciation spelling. Research on letter to sound conversion rules need to be done for the development of the system. However, the system provides the technique on parametric software synthesis which is very flexible for any knowledge based synthesis system.

จุฬาลงกรณ์มหาวิทยาลัย

ภาควิชา วิศวกรรมคอมพิวเตอร์
สาขาวิชา วิทยาศาสตร์ คอมพิวเตอร์
ปีการศึกษา 2532

ลายมือชื่อนิสิต
ลายมือชื่ออาจารย์ที่ปรึกษา
ลายมือชื่ออาจารย์ที่ปรึกษาร่วม



กิตติกรรมประกาศ

ผู้วิจัยขอขอบพระคุณท่านอาจารย์ที่ปรึกษา ผศ.ดร. วีระ ธีรวิทักษ์ ที่ได้ให้คำปรึกษา แนะนำแนวทางที่เป็นประโยชน์ต่อการวิจัย และคอยผลักดันให้กำลังใจในการทำวิทยานิพนธ์ฉบับนี้ จนสำเร็จลุล่วงด้วยดี ขอขอบคุณ ผศ.ดร. สุธาพร ลักษณะินาวิน ที่ช่วยให้คำปรึกษาและตรวจสอบ แก้ไขข้อผิดพลาดวิทยานิพนธ์ฉบับนี้ทำให้เนื้อหามีความถูกต้องมากขึ้น อ่านได้ง่ายไม่ติดขัด

นอกจากนี้ขอขอบคุณ คุณไพศาล ธรรมโพธิทอง คุณวัลลภ ดันฤดี เพื่อนร่วมโครงการวิจัย และ คุณกาญจนา เรขยศ ที่ได้คอยช่วยเหลือค้นหาเอกสาร แลกเปลี่ยนความรู้ ให้คำแนะนำต่างๆและช่วยแก้ปัญหาด้วยดีตลอดมา ขอขอบคุณ คุณเหมือนฝัน สุวิธร์ ที่คอยให้กำลังใจ และเป็นแรงใจที่สำคัญต่อผู้วิจัยมาตั้งแต่ต้นจนประสบความสำเร็จในที่สุด ขอขอบคุณคุณอาจารย์และบรรดาพี่น้องชาวศึกษาศาสตร์คอมพิวเตอร์จุฬาลงกรณ์มหาวิทยาลัยทุกรุ่น ที่คอยให้ความร่วมมือด้วยดีและให้กำลังใจอย่างอบอุ่นต่อผู้วิจัยเสมอมา

สุดท้ายนี้ผู้วิจัยขอกราบขอบพระคุณ บิดา มารดา ซึ่งเป็นผู้มีพระคุณสูงสุดอันหาที่เปรียบมิได้ที่ได้ให้การอุปการะผู้วิจัยมาตลอด

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย



สารบัญ

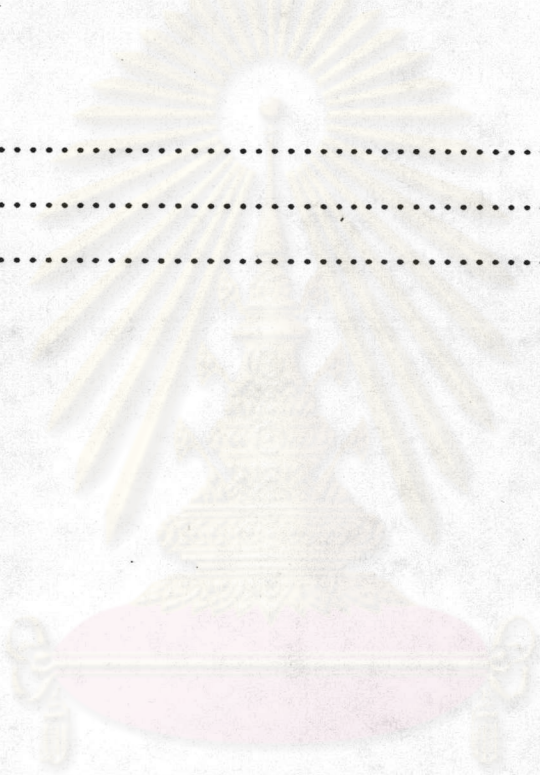
	หน้า
บทคัดย่อภาษาไทย	ง
บทคัดย่อภาษาอังกฤษ	จ
กิตติกรรมประกาศ	ฉ
สารบัญตาราง	ญ
สารบัญภาพ	ฎ

บทที่

1. บทนำ	1
1.1 แนวเหตุผลและสมมติฐาน	1
1.2 หลักการเบื้องต้น	2
1.3 วัตถุประสงค์	5
1.4 ขั้นตอนการวิจัย	6
1.5 ขอบเขตการวิจัย	6
2. หลักการวิเคราะห์และสังเคราะห์เสียงพูด	7
2.1 ทฤษฎีการสร้างเสียงพูด	7
2.1.1 กระบวนการทำให้เกิดเสียง	7
2.1.2 สรุปหลักสำคัญของ การเกิดเสียงพูด	12
2.2 วิธีการเข้ารหัสสัญญาณเสียงแบบดิจิทัล	13
2.2.1 พีซีเอ็ม	14
2.3 หลักการวิเคราะห์เสียงพูด	16
2.3.1 การวิเคราะห์ในโดเมนเวลา	16
2.3.2 การวิเคราะห์ทางสเปกตรัม	23
2.3.3 การหาคาบของพิทช์	29

2.4	การวิเคราะห์โดยใช้เทคนิคการทำนายแบบเชิงเส้นหรือ แอลพีซี	35
2.4.1	รูปแบบจำลองของการสร้างสัญญาณเสียงโดยเทคนิค การทำนายแบบเชิงเส้น	35
2.4.2	รูปแบบของการทำนายแบบเชิงเส้น	37
2.4.3	วิธีการหาค่าสัมประสิทธิ์ของการทำนาย	42
2.5	การสังเคราะห์เสียงพูดโดยใช้เทคนิคการทำนายแบบ เชิงเส้น	61
2.5.1	ต้นกำเนิดเสียงและอัตราการขยายในการ สังเคราะห์	63
2.5.2	โครงสร้างของการสังเคราะห์เสียงพูด	65
2.5.3	โครงสร้างของการสังเคราะห์เสียงพูดแบบ แลททิส	66
3.	การจำลองลงบนระบบไมโครคอมพิวเตอร์	72
3.1	ภาคของการวิเคราะห์เสียงพูด	75
3.1.1	ขั้นตอนสำหรับการคิดตัวอย่างสัญญาณเสียง ต้นแบบ	75
3.1.2	โปรแกรมสำหรับคำนวณหาชุดพารามิเตอร์ ของสัญญาณเสียง	76
3.1.3	โปรแกรมสำหรับคำนวณหาคาบของพิทช์	79
3.2	ภาคของการสังเคราะห์เสียงพูด	81
3.2.1	โปรแกรมสังเคราะห์เสียงพูด	81
3.2.2	ขั้นตอนการสร้างเสียงพูด	83
3.3	ภาครับอักขระอินพุตภาษาไทย	84
3.4	โปรแกรมอรรถประโยชน์	84
3.4.1	โปรแกรมสำหรับแก้ไขสัญญาณข้อมูล	85
3.4.2	โปรแกรมบำรุงรักษาพจนานุกรมข้อมูล	85
4.	วิเคราะห์รายละเอียดและผลการทดสอบ	86
4.1	ส่วนของการวิเคราะห์เสียงต้นแบบ	86

4.2 ส่วนของการแก้ไขพารามิเตอร์	89
4.3 ส่วนของการสังเคราะห์เสียงพูด	94
5. บทสรุปและข้อเสนอแนะ	102
5.1 สรุป	102
5.2 ข้อเสนอแนะและแนวทางในการวิจัยต่อ	104
เอกสารอ้างอิง	107
ภาคผนวก	110
ประวัติผู้เขียน	120



ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย



สารบัญตาราง

ตารางที่		หน้า
4.1	รายละเอียดพยางค์ตัวอย่างที่ใช้ในพจนานุกรมข้อมูล	88
4.2	เวลาที่ใช้ในการคำนวณสังเคราะห์เสียงกับเครื่องไมโครคอมพิวเตอร์ ชนิดต่างๆ	94
4.3	ผลการรับรู้แบบคำพยางค์เดี่ยว	99
4.4	ผลการรับรู้แบบคำพูดหลายพยางค์	100

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย



สารบัญภาพ

ภาพที่	หน้า
1.1.1 การสร้างเสียง โดยการเก็บบันทึกเสียงพูด	4
1.1.2 การสร้างเสียง โดยการสังเคราะห์เสียงพูด	4
2.1.1 รูปแบบอย่างง่าย ๆ ของช่องทางเดินเสียงของมนุษย์	8
2.1.2 อวัยวะภายในของระบบการพูดของมนุษย์	8
2.1.3 แสดงการเกิดการกำทอนภายในแบบจำลองของช่องเสียง	11
2.1.4 สเปคตรัมของพลังงานของเสียง	11
2.2.1 การเข้ารหัสสัญญาณเสียงด้วยเทคนิคพีซีเอ็ม	15
2.3.1 รูปพลังงานของคำว่า /six/	18
2.3.2 ช่องวิเคราะห์แบบสี่เหลี่ยม	20
2.3.3 ลักษณะของช่องวิเคราะห์บางชนิด	21
2.3.4 ตัวอย่างแสดงฟังก์ชันออโตคอร์เรเลชันของเสียงก้องและเสียงไม่ก้อง	23
2.3.5 การวิเคราะห์ฟิลเตอร์แบงด์	24
2.3.6 แชลแนลไวโคดเดอร์	25
2.3.7 รูปแสดงการแปลงฟูเรียร์ ฟูเรียร์ซีรี่ และการแปลงดีสครีทฟูเรียร์	27
2.3.8 ภาพของสเปคโตรแกรมไวร์แบนด์และนาโรแบนด์	28
2.3.9 ลักษณะสมบัติของสัญญาณเมื่อผ่านกระบวนการเช่นเตอร์คลิปปิง	31
2.3.10 ลักษณะสมบัติของเทคนิคเช่นเตอร์คลิปปิง	32
2.3.11 ตัวอย่างการประมวลผลของสัญญาณใน 1 เฟรม	34
2.4.1 รูปแบบจำลองการกำเนิดเสียง โดยใช้เทคนิคการทำนายแบบเชิงเส้น	35
2.4.2 การตัดตัวอย่างของสัญญาณเสียงทุกช่วงเวลา T	38
2.4.3 บล็อกไดอะแกรมของแบบจำลองของการวิเคราะห์และการสังเคราะห์ของ การทำนายแบบเชิงเส้น	40
2.4.4 ขอบเขตของตัวอย่างที่ใช้ในวิธีการโควาเรียนซ์	42
2.4.5 หลักการของวิฟาร์คอร์	46
2.4.6 แสดงถึงอินเนอร์โปรดักต์ของตัวกรอง $F(z)$ กับ $G(z)$	48
2.4.7 โครงสร้างของส่วนกลับของตัวกรอง $\{A(z)\}$ ในรูปแบบของโครงสร้าง	

แลททิส	55
2.5.1 แผนภาพของการสังเคราะห์เสียงโดยใช้เทคนิคการทำนายเชิงเส้น	62
2.5.2 ภาพแสดงการไหลของสัญญาณของตัวกรอง	69
2.5.3 แสดงกราฟการไหลของสัญญาณของโครงสร้างแลททิสแบบตัวคูณสองตัว ...	71
3.1 ขั้นตอนการทำงานรวมของระบบ	74
3.2 ขั้นตอนสำหรับทำการตัดตัวอย่างสัญญาณเสียง	75
3.3 พังงานของโปรแกรมคำนวณหาชุดพารามิเตอร์	77
3.4 ภาพแสดงโครงสร้างของระบบพจนานุกรมข้อมูล	79
3.5 พังงานของโปรแกรมคำนวณหาคาบของพิทซ์	80
3.6 พังงานของโปรแกรมสังเคราะห์เสียงพูด	82
3.7 ขั้นตอนการสร้างเสียงพูด	83
4.1 รูปสัญญาณเสียงต้นแบบของพยางค์ "สอง"	89
4.2 รูปสัญญาณเสียงพยางค์ "สอง" ต้นแบบกับที่สังเคราะห์ได้จากพารามิเตอร์ ก่อนแก้ไข	90
4.3 แสดงค่าพารามิเตอร์ของพยางค์ "สอง" ที่ได้มาจากการวิเคราะห์เสียง ...	91
4.4 รูปสัญญาณเสียงพยางค์ "สอง" ต้นแบบกับที่สังเคราะห์ได้จากพารามิเตอร์ หลังแก้ไข	92
4.5 รูปแสดงสเปคตรัมของเฟรมที่ 20 ของสัญญาณเสียงต้นแบบและสัญญาณเสียง สังเคราะห์	92
4.6 แนวทางการเปลี่ยนแปลงความถี่ของเสียงวรรณยุกต์ไทย	93
4.7 รูปสัญญาณของวลี "สิบเจ็ด" และ "เจ็ดสิบสตางค์"	96
4.8 รูปสัญญาณของวลี "ยี่สิบเอ็ด" และ "ห้าล้านบาท"	97
4.9 รูปสัญญาณของวลี "ราคาขึ้น" และ "ราคาเปิด"	97