การสังเคราะห์เสียงพูดจากข้อความภาษาไทย

นาย อาทร นันทิยกุล

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
ภาควิชาวิศวกรรมคอมพิวเตอร์
บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย
พ.ศ. 2533

ISBN 974-577-629-7

THAI TEXT TO SPEECH SYNTHESIS

Mr. Atorn Nuntiyagul

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 นันทิยกุล
ภาควิชา วิศวกรรมคอมพิวเตอร์
อาจารย์ที่ปรึกษา ผู้ช่วยศาสตราจารย์ ดร. วีระ ริ้วพิทักษ์

บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

..... คณบดีบัณฑิตวิทยาลัย
(ศาสตราจารย์ ดร. ถาวร วัชราภัย)

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ
(รองศาสตราจารย์ ไกรวิชิต ตันติเมธ)

..... อาจารย์ที่ปรึกษา
(ผู้ช่วยศาสตราจารย์ ดร. วีระ ริ้วพิทักษ์)

..... กรรมการ
(รองศาสตราจารย์ สมชาย ทยานยง)

..... กรรมการ
(ผู้ช่วยศาสตราจารย์ เมธี ศรีสังวาล)

..... กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร. สุดาพร ลักษณียนาวิน)

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

การสังเคราะห์เสียงพูดจากข้อความภาษาไทยเป็นการสร้างเสียงพูดขึ้นมาจากข้อความภาษาไทยที่ถูกป้อนเป็นอินพุทเข้าสู่ระบบ ซึ่งได้ประยุกต์ลงบนไมโครคอมพิวเตอร์และมีภาคการแปลงสัญญาณระหว่างสัญญาณอนาล็อกกับสัญญาณดิจิตอลรวมอยู่ด้วย โดยใช้หลักของการวิเคราะห์หน่วยย่อยของเสียงพูดคือ พยางค์ ของเสียงต้นแบบมาทำการคัดตัวอย่างด้วยความถี่ 10 kHz และนำมาทำการวิเคราะห์ด้วยเทคนิคการทำนายแบบเชิงเส้น (LPC) แบบออร์เดอร์ 10 ทีละเฟรม (เฟรมละ 200 จุดหรือ 20 มิลลิวินาที) ได้เอาท์พุทออกมาเป็นชุดพารามิเตอร์ประกอบด้วย 1) ค่าความผิดพลาดเฉลี่ย 2) ค่าคาบของพิทช์ และ 3) ค่าสัมประสิทธิ์ของการทำนาย (10 ค่า) ต่อหนึ่งเฟรม เก็บเอาไว้ในหน่วยเก็บความจำสำรองในรูปแบบของพจนานุกรมข้อมูลซึ่งสามารถจะทำการแก้ไขพารามิเตอร์เพื่อปรับปรุงให้ได้เสียงสังเคราะห์ที่ดีขึ้น จากนั้นก็จะทำการสร้างเสียงพูดสังเคราะห์ขึ้นมา โดยนำเอาชุดพารามิเตอร์ของหน่วยย่อยของเสียงที่ได้มาจากการค้นหาในพจนานุกรมข้อมูลตามข้อความที่ป้อนเข้ามา มาทำการสังเคราะห์ผ่านตัวกรองแลททิส (Lattice filter) และภาคการแปลงสัญญาณดิจิตอลเป็นสัญญาณอนาล็อกออกลำโพงกลับคืนมาเป็นเสียงพูด โดยคุณภาพของเสียงที่สังเคราะห์ออกมาอยู่ในระดับปานกลางซึ่งวัดผลได้จากการรับฟังของกลุ่มตัวอย่าง และเสียงสังเคราะห์นี้จะมีคุณสมบัติของเสียงวรรณยุกต์ซึ่งเป็นลักษณะสำคัญของภาษาไทยด้วย

การป้อนข้อความอินพุทจะป้อนในรูปของตัวสะกดตามคำอ่านเท่านั้น ระบบการสังเคราะห์เสียงพูดจากข้อความภาษาไทยนี้ยังต้องมีการพัฒนาในส่วนของการแปลงข้อความที่เป็นภาษาเขียนให้กลายเป็นคำอ่านอีก แต่อย่างไรก็ตามระบบการสังเคราะห์เสียงพูดนี้ก็ได้ใช้เทคนิคการสังเคราะห์เสียงพูดขึ้นมาจากชุดพารามิเตอร์โดยใช้เป็นซอฟแวร์ทั้งหมด ซึ่งทำให้สามารถนำไปประยุกต์ใช้ได้ง่ายและกว้างขวางสำหรับระบบการสังเคราะห์เสียงพูดแบบฐานความรู้ทั่วๆไป

ภาควิชา วิศวกรรมคอมพิวเตอร์ ลายมือชื่อนิสิต .....

สาขาวิชา วิทยาศาสตร์คอมพิวเตอร์

ปีการศึกษา 2532 ลายมือชื่ออาจารย์ที่ปรึกษา .....

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.

ภาควิชา วิศวกรรมคอมพิวเตอร์ ลายมือชื่อนิสิต .....

สาขาวิชา วิทยาศาสตร์คอมพิวเตอร์

ปีการศึกษา 2532 ลายมือชื่ออาจารย์ที่ปรึกษา .....

ลายมือชื่ออาจารย์ที่ปรึกษาร่วม .....

กิตติกรรมประกาศ

ผู้วิจัยขอขอบพระคุณท่านอาจารย์ที่ปรึกษา ผศ.ดร. วีระ ริ้วพิทักษ์ ที่ได้ให้คำปรึกษา แนะนำแนวทางที่เป็นประโยชน์ต่อการวิจัย และคอยผลักดันให้กำลังใจในการทำวิทยานิพนธ์ฉบับนี้จนสำเร็จลุล่วงด้วยดี ขอขอบคุณ ผศ.ดร. สุดาพร ลักษณียนาวิน ที่ช่วยให้คำปรึกษาและตรวจสอบ แก้ไขข้อบกพร่องวิทยานิพนธ์ฉบับนี้ทำให้เนื้อหามีความถูกต้องมากขึ้น อ่านได้ง่าย ไม่ติดขัด

นอกจากนี้ขอขอบคุณ คุณไพศาล ธรรมโพธิทอง คุณวัลนพ ตันฤดี เพื่อนร่วมโครงการวิจัย และ คุณกฤษดา เรเยส ที่ได้คอยช่วยเหลือค้นคว้าเอกสาร แลกเปลี่ยนความรู้ ให้คำแนะนำต่างๆและช่วยแก้ปัญหาด้วยดีตลอดมา ขอขอบคุณ คุณเหมือนฝัน สุธีธร ที่คอยให้กำลังใจและเป็นแรงใจที่สำคัญต่อผู้วิจัยมาตั้งแต่ต้นจนประสบกับความสำเร็จในที่สุด ขอขอบคุณคณาจารย์และบรรดาพี่ๆน้องๆชาววิทยาศาสตร์คอมพิวเตอร์จุฬาลงกรณ์มหาวิทยาลัยทุกรุ่น ที่คอยให้ความร่วมมือด้วยดีและให้กำลังใจอย่างอบอุ่นต่อผู้วิจัยเสมอมา

สุดท้ายนี้ผู้วิจัยขอกราบขอบพระคุณ บิดา มารดา ซึ่งเป็นผู้มีพระคุณสูงสุดอันหาที่เปรียบมิได้ที่ได้ให้การอุปการะผู้วิจัยมาตลอด

สารบัญ

หน้า

สารบัญตาราง

ตารางที่ หน้า

สารบัญภาพ

ภาพที่ หน้า