



การสังเคราะห์เสียงพูดภาษาไทยด้วยแบบจำลองอวัยวะที่ใช้ในการเปล่งเสียง
Thai Speech Synthesis by Articulatory Modeling

กীরติ อินทวิเศษ
Keerati Inthavisas

วิทยานิพนธ์วิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมคอมพิวเตอร์
มหาวิทยาลัยสงขลานครินทร์
Master of Engineering Thesis in Computer Engineering
Prince of Songkla University

เลขหมู่ TK 789.5.565 664 8549 น.1
Bib Key 24/99
25 อ.ย. 2547

ชื่อวิทยานิพนธ์ การสังเคราะห์เสียงพหูภาษาไทยด้วยแบบจำลองอวัยวะที่ใช้ในการเปล่งเสียง
ผู้เขียน นายกীরติ อินทวิเศษ
สาขาวิชา วิศวกรรมคอมพิวเตอร์
ปีการศึกษา 2546

บทคัดย่อ

วิทยานิพนธ์ฉบับนี้เสนอการสังเคราะห์เสียงพหูภาษาไทย โดยแทนช่องทางเดินเสียงด้วยท่อทรงกระบอกความยาว 1.7 เซนติเมตรนำมาต่อเรียงกัน โดยแต่ละท่อมีพื้นที่หน้าตัดต่าง ๆ กัน ตามความยาวและลักษณะของช่องทางเดินเสียงของมนุษย์ขณะเปล่งเสียงนั้น ๆ โดยได้ทำการสังเคราะห์เสียงสระตามเสียงวรรณยุกต์ ประกอบด้วยสระเดี่ยวเสียงสั้นยาวจำนวน 18 หน่วยเสียง และสระประสมเสียงสั้นยาวจำนวน 6 หน่วยเสียง ในการสังเคราะห์ตามเสียงวรรณยุกต์ ทุกเสียงประกอบไปด้วยเสียงวรรณยุกต์ 5 หน่วย คือ สามัญ เอก โท ตรี และจัตวา รวมเสียงสังเคราะห์ทั้งสิ้น 120 หน่วยเสียง ในส่วนของพารามิเตอร์ประกอบไปด้วยพื้นที่หน้าตัดจำนวน 9 ชุด สำหรับสระเดี่ยว ร่วมกับพารามิเตอร์ของความถี่หลักมูลสำหรับเปลี่ยนเสียงสั้นยาวตามวรรณยุกต์และพารามิเตอร์พลังงานของเสียงจำนวน 10 ชุด หลังจากนั้นนำพารามิเตอร์เหล่านี้ไปสังเคราะห์เป็นสระประสมด้วยการเพิ่มลดขนาดช่องทางเดินเสียงจากสระที่ 1 ไปสระที่ 2 ในระหว่างการสังเคราะห์เสียงไปพร้อมกัน

ประเมินสระเดี่ยวที่สังเคราะห์ได้ โดยเทียบกับค่ามาตรฐานได้เปอร์เซ็นต์ความแตกต่างของฟอร์แมนท์ 1 และ 2 เป็น 9.89% และ 7.44% ตามลำดับ ส่วนการสังเคราะห์สระประสมใช้วิธีเปรียบเทียบระหว่างการเพิ่มลดขนาดช่องทางเดินเสียงด้วยฟังก์ชัน 3 รูปแบบคือ ฟังก์ชันเชิงเส้น ฟังก์ชันลอกการิทึม และฟังก์ชันเลขชี้กำลัง ปรากฏว่าฟังก์ชันเลขชี้กำลังเหมาะสมกับการสังเคราะห์เสียง เอียะ, เอีย อีก 4 สระคือ เอือะ, เอือ และ อัวะ, อิว เหมาะสมกับการเพิ่มลดขนาดช่องทางเดินเสียงแบบลอกการิทึม

Thesis Title Thai Speech Synthesis by Articulatory Modeling
Author Mr.Keerati Inthavisas
Major Program Computer Engineering
Academic Year 2003

Abstract

This thesis presents the Thai speech synthesis by representing the vocal-tract as a series of concatenated tubes of 1.7 centimeters which are different in cross sectional areas. The number of tubes and the cross sectional areas are corresponding to the vocal-tract of human while uttering. We had synthesized all of vowels that are classified as either monophthong and diphthong. Monophthongs and diphthongs comprise nine short-long vowel pairs and three short-long vowel pairs, respectively. By adjusting fundamental frequency or pitch contour, the 5 tones of these vowels can be synthesized totally 120 phonemes. For synthesis of monophthongs, the parameters consist of 9 set of cross sectional areas together with 10 set of fundamental frequency and energy or amplitude envelope of sound. After that, diphthongs were synthesized which these parameters by raising or lowering the coinciding cross sectional areas from one vowel to another.

The synthesized monophthongs have been evaluated by comparing first and second formants to standard in term of different percentage. The results are 9.89% and 7.44% respectively. For synthesis of diphthongs, we have varied the vocal-tract using linear, logarithm and exponential functions to compare the result. Exponential functions are satisfactory for synthesizing diphthongs such as [ia], [iia]. Logarithm function are better for synthesis of [ua], [uuua] and [ua], [uua].