

ABSTRACT

Project Code : RSA/03/2541

Project Title : Tone Classification and Its Impact on the Design of Speaker-independent Automatic Speech Recognition System for Continuous Thai Speech

Investigator : Colonel Dr. Siripong Potisuk
Department of Electrical Engineering, Academic Division,
Chulachomklao Royal Military Academy

E-mail Address : srppts@hotmail.com

Project Period : 1 Dec 1998 – 31 May 2002 (3 years 6 months)

This research is aimed at studying the process of classifying Thai tones in connected speech and how the process affects the design of a Thai automatic speech recognition (ASR) system. Tone classification is considered one of the crucial components of the system because tone is a distinctive feature of Thai, which helps signal differences in lexical meaning. Every Thai syllable carries a lexically-contrastive tone. This research is divided into the following three major parts:

First, two acoustic experiments were designed to empirically study the acoustic characteristics of stressed and unstressed syllables in terms of their F_0 realization and the effects the interactions between stress and tonal coarticulation have on the height and shape of F_0 contours. Findings were used to guide the implementation of a tone classifier.

In the second part, an analysis-by-synthesis approach to tone classification have been modified to account for all of the linguistic factors affecting F_0 contours of Thai tones in continuous speech. They are: continuity effect due to syllable structure, stress, tonal coarticulation, and declination. Our tone classifier achieved 81.7% accuracy. We also addressed the prosody generation aspect of a Thai text-to-speech system using our mathematical model for F_0 contours, which is based on an extension to Fujisaki's model.

The final portion of this research deals with the question of how to best integrate the tone classification process into a design of a Thai ASR system. The integration process collectively results in a component called a lexical decoder. We propose a novel three-stage lexical decoder utilizing successive hidden Markov models. Due to limited time and resources at our disposal, we were unable to implement a full-scale design, and thus, were unable to assess the performance of the system.

Keywords: Tone classification, Thai speech recognition, Thai speech synthesis, Prosody.

บทคัดย่อ

รหัสโครงการ : RSA/03/2541

ชื่อโครงการ : การจำแนกเสียงวรรณยุกต์และผลกระทบที่มีต่อการออกแบบระบบรู้จำเสียงพูด
อัตโนมัติชนิดไม่จำกัดผู้พูดสำหรับภาษาไทยแบบต่อเนื่อง

ชื่อนักวิจัย : พ.อ.ดร.ศิริพงษ์ โพธิสุข

กองวิชาวิศวกรรมไฟฟ้า ส่วนการศึกษา โรงเรียนนายร้อยพระจุลจอมเกล้า

E-mail Address : srppts@hotmail.com

ระยะเวลาโครงการ : 1 ธันวาคม 2541 — 31 พฤษภาคม 2545 (3 ปี 6 เดือน)

งานวิจัยฉบับนี้มุ่งเน้นที่จะศึกษากระบวนการจำแนกเสียงวรรณยุกต์ในภาษาไทยแบบต่อเนื่อง และผลกระทบที่มีต่อการออกแบบระบบรู้จำเสียงพูดอัตโนมัติ การจำแนกเสียงวรรณยุกต์นั้นถือว่าเป็นขั้นตอนหนึ่งที่มีความสำคัญต่อการออกแบบระบบดังกล่าว ทั้งนี้เนื่องจากภาษาไทยเป็นภาษาวรรณยุกต์ ความแตกต่างของความหมายของพยางค์ในภาษาไทยขึ้นอยู่กับเสียงวรรณยุกต์เป็นส่วนใหญ่ งานวิจัยฉบับนี้แบ่งออกเป็นสามส่วนใหญ่ ๆ ดังนี้

งานวิจัยในส่วนที่หนึ่งเป็นการดำเนินการทดลองทางสัทศาสตร์สองการทดลอง การทดลองที่หนึ่งมีความมุ่งหมายที่จะศึกษาการเปลี่ยนแปลงรูปแบบเส้นโค้งความถี่พื้นฐานของพยางค์ภายใต้อิทธิพลของการเน้นเสียงหนักเบาในประโยค ส่วนการทดลองที่สองเป็นการศึกษาผลกระทบของปฏิสัมพันธ์ระหว่างการเน้นเสียงหนักเบาและการผสมผสานกันของเส้นโค้งความถี่พื้นฐานของวรรณยุกต์ที่อยู่ติดกันในประโยค ผลการทดลองที่ได้จากการทดลองทั้งสองจะเป็นประโยชน์ต่อการออกแบบตัวจำแนกเสียงวรรณยุกต์อัตโนมัติซึ่งเป็นงานในส่วนที่สองต่อไป

ในส่วนที่สอง ผู้วิจัยได้ทำการพัฒนาและปรับปรุงตัวจำแนกเสียงวรรณยุกต์ซึ่งอาศัยหลักการวิเคราะห์จากการสังเคราะห์ โดยผู้วิจัยสามารถสร้างกฎเกณฑ์ทางภาษาศาสตร์ที่เกี่ยวข้องกับการแปรเปลี่ยนเส้นโค้งความถี่พื้นฐานของแต่ละเสียงวรรณยุกต์ อาทิเช่น ความต่อเนื่องของเส้นโค้งความถี่อันเนื่องมาจากเสียงโฆษะ การเน้นเสียงหนักเบา การผสมผสานของเส้นโค้งความถี่ และการถดถอย ตัวจำแนกเสียงวรรณยุกต์นี้มีความถูกต้องแม่นยำ 81.7% นอกจากนี้ผู้วิจัยยังได้พัฒนาในส่วนของการสังเคราะห์ท่วงทำนองเสียงของระบบสังเคราะห์เสียงภาษาไทยอัตโนมัติอีกด้วย โดยอาศัยแบบจำลองทางคณิตศาสตร์ของเส้นโค้งความถี่พื้นฐานที่ปรับปรุงจากแบบจำลองของฟูจิซากิ

ส่วนสุดท้ายของงานวิจัยฉบับนี้มุ่งเน้นที่จะตอบคำถามเกี่ยวกับวิธีที่ดีที่สุดในการออกแบบระบบรู้จำเสียงพูดภาษาไทยอัตโนมัติ โดยเฉพาะอย่างยิ่งบทบาทและหน้าที่ของตัวจำแนกเสียงวรรณยุกต์ที่มีต่อระบบโดยรวม ผู้วิจัยนำเสนอคำตอบในรูปของตัวถอดรหัสคำ ซึ่งมีส่วนประกอบหลัก ๆ สามส่วน เนื่องจากเวลาและทรัพยากรการวิจัยที่จำกัด ทำให้ผู้วิจัยไม่สามารถทำการจำลองระบบแบบเต็มรูปแบบได้ ส่งผลให้ไม่สามารถประเมินประสิทธิภาพของระบบโดยรวมได้

คำหลัก : การจำแนกเสียงวรรณยุกต์ การรู้จำเสียงพูดภาษาไทยอัตโนมัติ การสังเคราะห์เสียงพูด