



**THAI TONE RECOGNITION
BY THE POLYNOMIAL REGRESSION TECHNIQUE**

PLYTEP SUKAPUN

≪

อภิรักษ์นันทนาการ

จาก

บัณฑิตวิทยาลัย มหาวิทยาลัยมหิดล

**A THESIS SUBMITTED IN PARTIAL
FULFILLMENT OF THE REQUIREMENTS FOR
THE DEGREE OF MASTER OF SCIENCE
(COMPUTER SCIENCE)
FACULTY OF GRADUATE STUDIES
MAHIDOL UNIVERSITY**

2001

TH
P 738 t
2001
C.2

ISBN 974-04-1985-2

COPYRIGHT OF MAHIDOL UNIVERSITY

3937004 SCCS/M : MAJOR : COMPUTER SCIENCE; M.Sc.
(COMPUTER SCIENCE)

KEY WORDS : TONE RECOGNITION / POLYNOMIAL REGRESSION
MODEL / LINEAR REGRESSION MODEL / FUNDAMENTAL
FREQUENCY

PLYTEP SUKAPUN : THAI TONE RECOGNITION BY THE
POLYNOMIAL REGRESSION TECHNIQUE. THESIS ADVISORS: SUPACHAI
TANGWONGSAN, Ph.D., CHOMTIP PORNPANOMCHAI, Ph.D., SUKANYA
PHONGSUPHAP, Ph.D., 70 P. ISBN 974-04-1985-2

This research studies the solution to Thai tone recognition. Tone recognition is the task of determining acoustic speech signals into one of tone mask sound. It explores the strategies and techniques to model the variability of tones and improve the recognition accuracy of tone recognition. Speech application has been a goal of worldwide research for many years. Its objective is to achieve flexible, convenient and efficient man-machine communication through natural speech. In tonal language, a complete speech recognition system requires a tone recognition system to properly recognize the meaning of the syllable. Tone information is extracted from the voiced portion of a syllable. The slope, shape, and level of pitch contour are the important properties to identify a type of tone.

In this research, the polynomial regression model was proposed as the alternative solution rather than the linear regression model for the tone recognition system. The template models of each tone were obtained by polynomial regression training. The distance measures and slope of first degree of polynomial were used to determine a type of tone. The duration normalization is performed in order to obtain the equal length of pitch contours. The variability of tones is significantly reduced by the duration normalization. The polynomial regression is also proposed as the most accurate technique to smooth the pitch contour. The autocorrelation with center clipping method is used to extract the fundamental frequency from the speech signal.

This research focused on speaker dependence and isolated syllable. There were 95 utterances of each tone used for training and 770 utterances of each tone used for testing. The best degree of polynomial model was chosen for each tone in order to obtain the highest recognition accuracy. The recognition accuracy of mid, low, falling, high, and rising tone is 92.20, 87.01, 97.79, 99.22, and 95.71 % respectively. It outperforms the linear regression approach, which has the recognition rate of 64.67, 66.62, 93.24, 97.53, and 91.16 % respectively.

The experiment results prove that the capability of the proposed model is satisfactory. More improvement is needed to extend the scope and idea of the present research in order to be used in real world applications.

3936997 SCCS/M : สาขาวิชา : วิทยาการคอมพิวเตอร์; วท.ม. (วิทยาการคอมพิวเตอร์)

พลาเยเทพ ศุขะพันธ์ : การรู้จำเสียงวรรณยุกต์ในภาษาไทยโดยใช้วิธีโพลีโนเมียลรีเกรสชันโมเดล (THAI TONE RECOGNITION BY THE POLYNOMIAL REGRESSION TECHNIQUE).
คณะกรรมการควบคุมวิทยานิพนธ์ : ศุภชัย ตั้งวงศ์สานต์, Ph.D., ชมทิพ พรพนมชัย, Ph.D., สุภัฏญา พงศ์สุภาพ, Ph.D. 70 หน้า. ISBN 974-04-1985-2

ในงานวิจัยนี้ได้นำเสนอแนวทางการแก้ปัญหาของการรู้จำเสียงวรรณยุกต์ในภาษาไทย จุดประสงค์ของการทำการวิจัยนี้คือการหาเทคนิคและวิธีการเพื่อใช้ในการวิเคราะห์ความไม่คงที่ของโทนเสียงวรรณยุกต์และปรับปรุงความถูกต้องของการรู้จำโทนเสียงวรรณยุกต์ให้สูงขึ้น

ได้มีการทำการวิจัยทางด้านเสียงโดยมีวัตถุประสงค์เพื่อให้เกิดความสะดวกและมีประสิทธิภาพในการติดต่อระหว่างมนุษย์และคอมพิวเตอร์โดยใช้เสียงพูด ในภาษาที่ใช้โทนเสียงระบบของการรู้จำเสียงพูด (Speech recognition system) จะประกอบด้วยระบบของการรู้จำโทนเสียงด้วย เพื่อที่จะสามารถรู้ความหมายของพยางค์เสียงได้อย่างถูกต้อง ข้อมูลของโทนเสียงมาจากส่วนที่เป็นสระของพยางค์ ความชัน รูปร่าง และระดับของเส้นความถี่เสียง (Pitch contour) เป็นคุณสมบัติที่สำคัญในการรู้จำโทนเสียง

ในการทำวิจัยนี้ได้เสนอวิธีโพลีโนเมียลรีเกรสชัน โมเดล เพื่อใช้ในการแก้ปัญหาของการรู้จำโทนเสียงนอกเหนือจากวิธีการของลิเนียร์รีเกรสชัน โดยตัวของแต่ละโทนเสียงจะได้มาจากการ training ของโพลีโนเมียลรีเกรสชัน การแบ่งแยกชนิดของโทนจะใช้วิธีการของ Distance measure และ ความชันของ โพลีโนเมียล ดีกรี 1 และมีการทำให้ความยาวของเส้นความถี่เสียงเท่ากันเพื่อที่จะลดความไม่คงที่ของรูปร่างของโทนเสียงลง นอกจากนี้วิธีโพลีโนเมียลรีเกรสชันยังได้ถูกเสนอเพื่อแก้ไขค่าที่ผิดพลาดของเส้นความถี่ด้วย และได้ใช้วิธีการของออคโตคอร์ริเลชัน และ เซ็นเตอร์คลิปปิง ในการหาความถี่มูลฐานของเสียง

ขอบเขตของการวิจัยนี้ได้กำหนดให้เป็นแบบขึ้นกับผู้พูดและเป็นเสียงของพยางค์เดี่ยว โดยในแต่ละโทนเสียงจะใช้ 95 เสียงในการ Training และใช้ 770 เสียงในการ Testing ความถูกต้องของการรู้จำเสียงวรรณยุกต์ สามัญ เอก โท ตรี และจัตวา คือ 92.20, 87.01, 97.79, 99.22, และ 95.71 % โดยลำดับ ซึ่งวิธีการที่ได้เสนอนี้มีประสิทธิภาพดีกว่าวิธีการของลิเนียร์รีเกรสชัน (Linear regression) ซึ่งได้ความถูกต้อง 64.67, 66.62, 93.24, 97.53, และ 91.16 % ตามลำดับ

จากการทดลองพิสูจน์ว่าความสามารถของโมเดลที่ได้เสนอนี้ได้ผลเป็นที่น่าพอใจ แต่จำเป็นต้องมีการปรับปรุงโดยการขยายขอบเขตและแนวความคิด เพื่อที่จะสามารถนำไปใช้ในงานจริงได้