T 130518

| | |
|---|---|
| หัวข้อวิทยานิพนธ์ | การรู้จำเสียงพูดที่เกิดจากการเน้นด้วยลักษณะสำคัญที่ต่างกัน |
| หน่วยกิตของวิทยานิพนธ์ | 12 หน่วย |
| โดย | นายภคพงศ์ อมรกุล |
| อาจารย์ที่ปรึกษา | รศ.ดร. โกสินทร์ จำนงไทย |
| ระดับการศึกษา | วิศวกรรมศาสตรมหาบัณฑิต |
| ภาควิชา | วิศวกรรมไฟฟ้า |
| ปีการศึกษา | 2544 |

## บทคัดย่อ

การเน้นที่เกิดจากอารมณ์หรือสภาพแวดล้อมต่างๆ เช่นโกรธ และผลของลอมบาร์ด นั้นทำให้ผู้พูดออกเสียงที่เปลี่ยนแปลงไปจากปกติ การเปลี่ยนแปลงที่เกิดขึ้นทำให้เกิดความผิดพลาดในระบบรู้จำเสียงพูดและลดสมรรถนะของระบบลง ในวิทยานิพนธ์นี้เสนอวิธีการรู้จำเสียงพูดสำหรับการเน้นหลายๆแบบ โดยการหาวิธีการแยกลักษณะสำคัญที่เหมาะสมสำหรับแต่ละการเน้นเพื่อใช้ในการหาลักษณะสำคัญของเสียงที่ให้อัตราการรู้จำที่สูงในแต่ละการเน้น จากนั้นนำลักษณะสำคัญที่ได้มาใช้ในการฝึกฝนแบบจำลองฮิดเดนมาร์คอฟที่กำหนดเฉพาะกับเสียงเน้นแต่ละแบบ ผลที่ได้จากแบบจำลองแต่ละแบบประกอบด้วยคำที่มีความน่าจะเป็นสูงสุดซึ่งกำหนดไว้ในรูปของคะแนน สุดท้ายคำและคะแนนจะถูกนำมาหาความเหมือนซึ่งมีค่าอยู่ระหว่าง 0-1 คำตอบที่เหมือนกันมากกว่าครึ่งหนึ่งจะเป็นคำตอบของระบบ ถ้าคำตอบไม่เหมือนกันหรือเหมือนกันน้อยกว่าครึ่งหนึ่งจะพิจารณาจากคะแนนที่สูงที่สุดเป็นคำตอบ ในการหาลักษณะสำคัญที่เหมาะสมสำหรับแต่ละการเน้นนั้น ทำโดยการฝึกฝนระบบรู้จำด้วยลักษณะสำคัญที่แตกต่างกัน เช่น LPC, MFCC และ LPCEP วิธีการแยกลักษณะสำคัญที่ให้ผลการรู้จำที่ดีที่สุดสำหรับการเน้นนั้นจะถูกเลือกมาใช้

วิธีการที่เสนอได้นำไปใช้สร้างและทดสอบในระบบที่ประกอบด้วยการเน้น 4 แบบ (3 การเน้น+ 1 เสียงปกติ) คือ โกรธ ดัง ผลของลอมบาร์ด และปกติ ระบบรู้จำสำหรับแต่ละการเน้นได้รับการฝึกฝนด้วยเสียงเน้นจำนวน 41 เสียงจากผู้พูดจำนวน 7 คน และทดสอบด้วยเสียงเน้น 41 เสียงจากผู้พูดคนอื่นจำนวน 3 คน ผลการรู้จำแสดงอัตราการรู้จำที่ร้อยละ 95.12 เพิ่มขึ้นจากระบบดั้งเดิมร้อยละ 0.68

TE 130518

| | |
|---|---|
| Thesis Title | Stress Speech Recognition with Different Feature Extraction |
| Thesis Credits | 12 |
| Candidate | Mr. Phakapong Amornkul |
| Supervisor | Assoc. Prof. Dr. Kosin Chamnongthai |
| Degree of Study | Master of Engineering |
| Department | Electrical Engineering |
| Academic Year | 2001 |

## Abstract

Stresses caused by emotions or environment such as anger and Lombard effect vary the speech patterns of speakers from their normal ones. This change causes errors in a speech recognition system and reduces its recognition performance. This thesis purposes a speech recognition method for multiple stress speeches. A feature extraction method appropriate for each stress is found and applied to extract speech features that will yield a high recognition rate for the stress . Then, the extracted features are fed to a Hidden Markov Model (HMM) trained with a set of the specific stress speeches. The result from each HMM is a recognized word with the highest probability given as a score. Finally, these words are use to find the similarity whose value is between 0-1. The selected word is the word that has the similarity more than 0.5. If they have the lower or the same similarity, the scores are used to compare in such way that the recognized word with maximum score is the winner. To find a suitable feature for each stress, recognition systems are trained with different feature extraction methods such as LPC, MFCC and LPCEP. The method that gives the highest recognition rate for the stress is chosen to be its feature extraction method.

The proposed method is implemented and tested using 4 kinds of speeches (3 stresses + neutral) namely angry, loud, Lombard effect and neutral speeches. Each recognition line is trained with 41 stress speeches from 7 people. The testing sets of speeches comprise of 41 speeches from 3 different people. The results show that the recognition rate is 95.12 percent, which improved by 0.68 percent from conventional method.