การศึกษาการรู้จำเสียงพูดตัวเลขภาษาไทยแบบแยกคำชนิดไม่ขึ้นกับผู้พูด โดยใช้โครงข่ายประสาทที่มีการเรียนรู้แบบแพร่กลับ

Study of Speaker Independent Isolated Word Recognition of Thai Digits using Backpropagation Neural Networks

ปฏิมากร กิ้มสวัสดิ์

Patimakorn Kimsawad

วิทยานิพนธ์วิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมไฟฟ้า

มหาวิทยาลัยสงขลานครินทร์

Master of Engineering Thesis in Electrical Engineering

Prince of Songkla University

2544

ชื่อวิทยานิพนธ์	การศึกษาการรู้จำเสียงพูดตัวเลขภาษาไทยแบบแยกคำชนิดไม่ขึ้นกับผู้พูด โดยใช้โครงข่ายประสาทที่มีการเรียนรู้แบบแพร่กลับ
ผู้เขียน	นางสาว ปฏิมากร กิ้มสวัสดิ์
สาขาวิชา	วิศวกรรมไฟฟ้า
ปีการศึกษา	2543

บทคัดย่อ

วิทยานิพนธ์ฉบับนี้นำเสนอการศึกษาการรู้จำสัญญาณเสียงพูดตัวเลข 0-9 แบบแยกคำชนิดไม่ขึ้นกับผู้พูดโดยใช้โครงข่ายประสาทที่มีการเรียนรู้แบบแพร่กลับ โดยเริ่มทำการศึกษาการรู้จำสัญญาณเสียงพูดตัวเลข 0-9 แบบแยกคำชนิดขึ้นกับผู้พูดก่อน เพื่อหาพารามิเตอร์ที่เป็นตัวแทนสัญญาณเสียงพูดที่เหมาะสมสำหรับการศึกษาในกรณีไม่ขึ้นกับผู้พูดต่อไป ทั้งนี้พารามิเตอร์ที่ใช้ในการศึกษาประกอบด้วย พารามิเตอร์ซึ่งเคยมีผู้นำมาใช้ในการรู้จำก่อนแล้ว คือ ค่าสัมประสิทธิ์พหุนามอันดับสองของคาบเวลาพิตซ์ ค่าความหนาแน่นกลุ่มความถี่ฟอร์แมนต์ ค่าสัมประสิทธิ์การประมาณพันธะเชิงเส้น ร่วมกับพารามิเตอร์ที่ได้เสนอขึ้นมาใหม่ซึ่งยังไม่เคยมีผู้นำมาใช้ในการรู้จำเสียงพูดตัวเลขภาษาไทยมาก่อน คือ ค่าระยะเวลาในการออกเสียง และ ค่าผลรวมของความหนาแน่นของสเปกตรัมเชิงกำลังในแต่ละกลุ่มความถี่ จากการทดลองศึกษาการรู้จำแบบขึ้นกับผู้พูดกับข้อมูลกลุ่มตัวอย่างและกลุ่มทดสอบกลุ่มละ 50 ตัวอย่าง พบว่า พารามิเตอร์ที่เหมาะสมในการรู้จำ คือ ค่าสัมประสิทธิ์การประมาณพันธะเชิงเส้น ร่วมกับ ค่าผลรวมของความหนาแน่นของสเปกตรัมเชิงกำลังในแต่ละกลุ่มความถี่ โดยให้เปอร์เซนต์การรู้จำถูกต้องสูงถึง 98.0 และ 94.0 เปอร์เซนต์สำหรับข้อมูลกลุ่มตัวอย่างและกลุ่มทดสอบ ตามลำดับ ส่วนในการศึกษาการรู้จำแบบไม่ขึ้นกับผู้พูดของผู้ชายและผู้หญิงกับข้อมูลกลุ่มตัวอย่างและกลุ่มทดสอบเพศละ 300 และ 300 ตัวอย่าง ตามลำดับ พบว่า พารามิเตอร์ที่เหมาะสมในการรู้จำ คือ ค่าระยะเวลาในการออกเสียง ร่วมกับ ค่าสัมประสิทธิ์พหุนามอันดับสองของคาบเวลาพิตซ์ และ ค่าผลรวมของความหนาแน่นของสเปกตรัมเชิงกำลังในแต่ละกลุ่มความถี่ ซึ่งเมื่อใช้โครงข่ายประสาททำการรู้จำด้วยพารามิเตอร์เหล่านี้ร่วมกับการพิจารณาแยกสัญญาณเสียงพูดตัวเลข 5 และ 9 แล้ว พบว่า สำหรับข้อมูลกลุ่มตัวอย่างและกลุ่มทดสอบของผู้ชายมีเปอร์เซนต์ความถูกต้อง 99.0 และ 85.3 เปอร์เซนต์ ตามลำดับ และมีเปอร์เซนต์ความ

ถูกต้องเท่ากับ 99.7 และ 82.0 เปอร์เซ็นต์ สำหรับข้อมูลกลุ่มตัวอย่างและกลุ่มทดสอบของผู้หญิง ตามลำดับ

Thesis Title	Study of Speaker Independent Isolated Word Recognition of Thai Digits using Backpropagation Neural Networks
Author	Ms. Patimakorn Kimsawad
Major Program	Electrical Engineering
Academic Year	2000

Abstract

This thesis presents the study of speaker independent isolated word recognition of 0-9 Thai digits using Backpropagation Neural Networks. The investigation began with studying of isolated word 0-9 Thai digits of speaker dependent recognition in order to find suitable parameters for studying in case of speaker independent recognition. The parameters which using in this thesis consist of the parameters, previously used for recognition, are the 2nd order Polynomial Coefficient of Pitch Period, the Density of Formant Frequency Groups and the Linear Predictive Coding Coeffcient. Together with these parameters, the new parameters were introduced, the Duration of Speech and the Summation of Power Spectrum Density in each Frequency Groups, which have never been used in Thai digits speech recognition. From the result of using speaker dependent recognition with 50 examples of training and testing data each, it showed that the suitable parameters was the Linear Predictive Coding Coefficient with the Summation of Power Spectrum Density in each Frequency Groups. The correction percentage of using these two parameters for speaker dependent recognition was 98.0 and 94.0 for training and testing data, respectively. Then in the case of speaker independent recognition, 300 examples of training and testing data each for male and 300 examples training and testing data each for female were tested. It showed that the suitable parameters were the Duration of Speech with the 2nd order Polynomial Coefficient of Pitch Period and the Summation of Power Spectrum Density in each Frequency Groups. After distinguish the speech of the number 5 and 9 from the results of recognition by Backpropagation Neural Networks, the correction percentage for male

was 99.0 and 85.3 for training and testing data, respectively. In the case of female, it was 99.7 and 82.0 percent for training and testing data, respectively.