

การทำนายค่าความถี่มูลฐานสำหรับการสังเคราะห์เสียงพูดภาษาไทย

F0 Prediction for Thai Speech Synthesis

เปรมวดี อุดมเรืองเกียรติ , รัชฎา คงคะจันทร์

บทคัดย่อ

การสร้างการสังเคราะห์เสียงพูดให้มีความเป็นธรรมชาติสูงนั้น จะต้องมีส่วนของสัทสัมพันธ์ (Prosody) เป็นองค์ประกอบหลัก ค่าความถี่มูลฐานก็เป็นหนึ่งในส่วนประกอบที่สำคัญของสัทสัมพันธ์เสียงพูด ความแตกต่างของค่าความถี่มูลฐานในคำพูดต่อเนื่องประโยคภาษาไทย มีผลต่อการควบคุมสัทสัมพันธ์ในระบบการสังเคราะห์เสียงภาษาไทย การทดลองนี้เลือกใช้ เทคนิคต้นไม้ตัดสินใจ (Decision Tree) ชื่อว่า Classification and Regression Trees (CART) มาช่วยในการสร้างแบบจำลองค่าความถี่มูลฐาน และได้มีการประยุกต์เทคนิคนี้ เพื่อทำการเปรียบเทียบผลกับเทคนิคต้นไม้ตัดสินใจอื่น เช่น C4.5 และ เทคนิคข่ายงานประสาทเทียมแบบแพร่กระจายย้อนกลับหลายชั้น (Multilayer Backpropagation Neural Network) โดยการทดลองได้ทำกับคลังข้อมูลเสียงภาษาไทย ชื่อว่า 'TSynC1' ซึ่งถูกพัฒนาโดย ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ (NECTEC) ประโยคคำพูด 5,200 ประโยคถูกใช้ในการเรียนรู้และทดสอบระบบ ในอัตรา 90 ต่อ 10 ของจำนวนประโยคคำพูดในคลังข้อมูลเสียง ผลการทดลองแสดงให้เห็นว่าเทคนิคต้นไม้ตัดสินใจสามารถทำนายค่าความถี่มูลฐานได้ดีกว่าเทคนิคข่ายงานประสาทเทียม

Abstract

For the generation of highly natural synthetic speech, the control of prosody is of primary importance. The fundamental frequency (F0) is one of the most important components of speech prosody. The variation of F0 in continuous Thai speech, with the goal of establishing an effective mechanism of prosody control in Thai text-to-speech (TTS) application. In this thesis, Classification And Regression Tree (CART) is employed for modeling the fundamental frequency (F0). This applied technique is compared with the well-known decision tree as C4.5 and the popular neural network as Multilayer Back-Propagation Neural Network (BPNN). The evaluation is done with Thai speech corpus, named TSYNC1. This corpus is developed by National Electronics and Computer Technology Center (NECTEC). Five thousands and two hundreds sentences of speech data are used for training and testing with the proportion as 90:10. The result shows that modeling F0 by decision tree technique is has higher performance than neural network technique.