199760

ความถูกต้องของการระบุตำแหน่งขอบเขตของเสียงมีบทบาทสำคัญที่สามารถเพิ่มประสิทธิภาพของการตัดแบ่งเป็นหน่วยเสียงในการรู้จำเสียง และเพิ่มคุณภาพของเสียงในการเลือกหน่วยเสียงสำหรับการสังเคราะห์เสียง การที่สามารถระบุขอบเขตเสียงได้โดยอัตโนมัติ จึงช่วยลดปัญหาของการสิ้นเปลืองแรงงานคนและลดเวลาในการพัฒนาการสร้างฐานข้อมูลเสียงที่ระบุตำแหน่งหน่วยเสียงด้วยคน

วิทยานิพนธ์นี้นำเสนอวิธีการในการหาตำแหน่งขอบเขตเสียง โดยแบ่งเป็นสองขั้นตอนคือ 1) การระบุตำแหน่งขอบเขตเสียงแบบบังคับจากแบบจำลองฮิดเด็นมาคอฟเพื่อหาตำแหน่งที่มีโอกาสเป็นขอบเขตเสียง 2) นำเสนอการปรับหาตำแหน่งขอบเขตเสียงเพื่อปรับและหาตำแหน่งขอบเขตเสียงที่ได้จากขั้นตอนแรกโดยละเอียด งานวิจัยนี้ได้ใช้ตัวจำแนกกลุ่มด้วยการวิเคราะห์ดิสคริมิแนนต์เชิงเส้นที่ขึ้นกับบริบทเพื่อตรวจหาตำแหน่งขอบเขตเสียง และได้ทำการจำแนกกลุ่มด้วย 21 ตัวจำแนกตามชนิดขอบเขตเสียง และในที่สุดได้เลือกตำแหน่งที่มีค่าความน่าจะเป็นมากที่สุดมาเป็นผลลัพธ์ ซึ่งค่าความน่าจะเป็นคำนวณจากระยะทางของพื้นที่ที่แยกโดยฟังก์ชันดิสคริมิแนนต์

งานวิจัยนี้ใช้ฐานข้อมูลเสียงโลตัสในการประเมินประสิทธิภาพของวิธีที่เสนอ ซึ่งประกอบด้วยข้อมูลสัทลักษณ์ และข้อมูลแสดงเวลากำกับตำแหน่งขอบเขตเสียงสำหรับทุกหน่วยเสียงที่ระบุด้วยคน วิธีการหาตำแหน่งขอบเขตเสียงที่เสนอได้รับความแม่นยำของการตรวจหาตำแหน่งขอบเขตเสียงเท่ากับ 80.22% เมื่อใช้ระดับที่ยอมรับได้ 10 มิลลิวินาทีเพื่อนับเป็นตำแหน่งที่ถูกต้อง ความผิดพลาดการหาตำแหน่งขอบเขตเสียงลดลง 43.42% เมื่อเทียบกับแบบอ้างอิงจากแบบจำลองฮิดเด็นมาคอฟ ค่าเฉลี่ยความคลาดเคลื่อนของตำแหน่งขอบเขตเสียงซึ่งเป็นจำนวนตำแหน่งเฟรมของตำแหน่งขอบเขตเสียงที่ตรวจหาได้คลาดเคลื่อนจากตำแหน่งขอบเขตเสียงที่ระบุด้วยคน มีค่าเฉลี่ยลดลงจาก 1.42 เฟรมเป็น 1.00 เฟรม เมื่อใช้เฟรมพิจารณาขนาด 10 มิลลิวินาที

199760

Precise phone boundary labeling plays an important role in improving segmentation performance in speech recognition, and increasing sound quality of unit selection in speech synthesis. Automatic phone alignment techniques are proposed to reduce the human efforts and time in the development of manually labeled speech corpus.

This thesis proposes an automatic method for locating acoustic boundaries. They can be divided into two steps: 1) HMM forced alignment is used to find the candidates phone boundaries, and 2) refinement of phone boundaries is proposed to adjust and fine-tune the boundaries obtained from the first step. The context-dependent Linear Discriminant Analysis (LDA) classifiers are used for phone boundary detection. The 21 specialized phone boundary classifiers are applied. The frame with maximum probability, calculated from distances in the space spanned by associated discriminant functions, is chosen as the output.

The LOTUS corpus (Large vOcabulary Thai contInUous Speech recognition Corpus) is used to evaluate the proposed performance. It contains manual transcriptions with phone boundary information for every speech utterance. The proposed method yields the detection accuracy of 80.22% using 10 milliseconds tolerance level, considered as correct. The proposed refinement results in a 43.42% error reduction in locating phone boundaries compared to the baseline. The average deviation, the number of frame of the detected boundaries deviated from their corresponding manually labeled boundaries, is reduced from 1.42 to 1.0 frame when the frame size used is 10 milliseconds.