วิทยานิพนธ์นี้เสนอระบบการสังเคราะห์เสียงซึ่งดัดแปลงการสังเคราะห์เสียงที่อิงแบบจำลองฮิดเดนมาร์คอฟให้สามารถรองรับการกำหนดลักษณะของสัญญาณแหล่งกำเนิดจากเส้นเสียง และเสียงรบกวนลมหายใจได้โดยตรง ทำให้สามารถสร้างเสียงสังเคราะห์เพื่อเลียนแบบลักษณะเสียงมนุษย์ชนิดต่าง ๆ ได้โดยมิต้องทำการประมวลผลสัญญาณที่ได้จากการสังเคราะห์ในภายหลัง เพื่อสร้างแบบจำลองเสียงจากแหล่งกำเนิดเส้นเสียง วิทยานิพนธ์นี้ได้วิเคราะห์ค่าพารามิเตอร์แอลเอฟแบบแปลง ซึ่งใช้เป็นแบบจำลองของแหล่งกำเนิดจากเส้นเสียง และศึกษาการหาค่าระดับเสียงรบกวนลมหายใจโดยใช้วิธีการลดสัญญาณรบกวนในระบบโดยเวฟเลท ซึ่งเป็นวิธีการสร้างสัญญาณใหม่จากสัญญาณที่ถูกรบกวน เพื่อสกัดสัญญาณรบกวนในสัญญาณเส้นเสียงได้ นอกจากนี้วิทยานิพนธ์นี้ได้เสนอวิธีหาฟังก์ชันการหาจุดเปลี่ยนแบบใหม่ในขั้นตอนการลดสัญญาณรบกวนในระบบโดยเวฟเลท

จากการวัดความเป็นธรรมชาติของเสียงสังเคราะห์ คะแนนความเป็นธรรมชาติของระบบที่นำเสนอไม่แตกต่างกับระบบอ้างอิง ชี้ให้เห็นว่าเสียงสังเคราะห์จากการใช้แหล่งกำเนิดจากเส้นเสียงเป็นอินพุทของแบบจำลองสามารถเทียบเคียงได้กับการใช้ขบวนอิมพัลส์เป็นอินพุทดังปรากฏในการสังเคราะห์เสียงด้วยแบบจำลองฮิดเดนมาร์คอฟทั่ว ๆ ไป นอกจากนี้เสียงที่ได้จากระบบสังเคราะห์เสียงที่นำเสนอของวิทยานิพนธ์นี้สามารถเลียนแบบเสียงลมหายใจ (Breathy) และเสียงบีบ (Creaky) ได้ดีกว่าระบบอ้างอิง และสามารถสังเคราะห์ลักษณะเสียงที่มีระดับความแตกต่างหลากหลายกว่าระบบอ้างอิง

234013

This thesis proposes a modified HMM-based speech synthesis system in which characteristic of the glottal source signal and aspiration noise can be manipulated explicitly. It can synthesize speech signals with different voice qualities without post processing of the synthetic speech signals. In order to model the glottal source, the transformed LF-model was used to represent the glottal waveform, while the aspiration noise level was estimated by a wavelet denoising algorithm. This thesis also proposes a new threshold function for evaluating threshold values used during the denoising process.

Results show that the synthetic speech signals produced by applying the glottal source as the input to the system is comparable to ones from a traditional HMM-based speech synthesis system that uses a pulse train as its input in terms of their naturalness. The proposed method can also mimic the breathiness and the creakiness of the synthetic speech with more flexibility than the baseline HMM-based system.