Title	音声生成を目指した生理学的発話モデルの制御法に関する研究
Author(s)	Wu, Xiyu
Citation	
Issue Date	2014-03
Type	Thesis or Dissertation
Text version	ETD
URL	http://hdl.handle.net/10119/12101
Rights	
Description	Supervisor:党　建武，情報科学研究科，博士

Abstract

In speech production, articulatory apparatus are the final organs that execute efferent motor commands from the central nervous system. Although the articulators play important roles in speech production, the mechanisms of how the motor commands control the articulators to generate speech sounds are not still very clear. Physiological articulatory models together with their control strategy provide a means to investigate the mechanisms of speech production.

In this thesis, a full 3D physiological articulatory model including the tongue, jaw, hyoid bone and vocal tract wall was constructed based on continuum finite element modeling. This model comprises articulatory muscles with realistic properties and geometrical arrangements. In order to control the physiological articulatory model more accurately, not only the extrinsic genioglossus muscle but also some intrinsic muscles are divided into smaller units according to their functions.

A control framework consists of a feedforward mapping, and a feedback learning loop was realized. In speech production, the feedforward mapping is used to find muscle activation pattern directly according to given articulatory targets and feedback learning loop is used to establish and maintain the feedforward mapping. In this control framework, the articulatory targets were defined by the entire posture of the tongue and jaw in the midsagittal plane, which was reduced to a six-dimensional vector with the principal component analysis (PCA).

Different from the musculoskeletal system, in the muscular-hydrostat system angonist-antagonist muscle pairs varied during articulation, which make it difficult to adjust muscle activations to minimize the distances between target positions and realized ones. In this study, the adjustment of muscle activations was guided by a dynamic PCA workspace that was used to predict individual muscle functions in given positions. This dynamic PCA workspace was estimated based on an interpolation of eight reference PCA workspaces.

In order to construct the feedforward mapping, the articulations of five Japanese vowels from magnetic resonance images were used as the targets for the learning process. The articulatory targets of five Japanese vowels were achieved, which proved that the proposed feedback learning loop was effective for the model control. According to the learning process by using the feedback loop, the feedforward mapping was established. This learned mapping function was assessed through an open set test, and reasonable vocal tract shapes were obtained compared with the target as a result. For the minorities that the articulatory targets cannot achieve perfectly, the implementation of the somatosensory feedback loop can further improve the control accuracy. Besides the improvement of control accuracy, the mapping established by a learning process makes the control strategy the ability to adapt to the external forces added as a perturbation. In order to evaluate the adaptation ability, a vertically downward external force was exerted to the jaw when producing Japanese vowels /i/ and /o/, by implementing the feedback loop, the articulatory targets can be re-achieved, which shows the adaptation ability.

The midsagittal contour including the tongue and jaw was used as the articulatory target, instead of using three crucial points. We expect that by using the articulatory posture as a target, the accuracy of model control for speech production will be improved, because the detailed characteristics of speech sounds depend on the whole vocal tract shape rather than the constriction positions alone.

The physiological articulatory model together with the framework of the control strategy can be implemented in the following aspects: 1) Investigating human speech production mechanism including estimating motor commands from observed articulation, exploring the "economy of effort", "saturation effect", "motor equivalence", etc. 2) Medical treatment. 3) Generating speech sounds by simulating the speech production process of human.

Keywords: Speech production, Physiological articulatory model, Muscle activation, Motor control, Motor learning, Somatosensory feedback