

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

氏名	吴西愉		
学位の種類	博士(情報科学)		
学位記番号	博情第300号		
学位授与年月日	平成26年3月24日		
論文題目	A Study on Control Strategy of a Physiological Articulatory Model for Speech Production (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)		
論文審査委員	主査	党建武	北陸先端科学技術大学院大学 教授
		赤木 正人	同 教授
		田中 宏和	同 准教授
		鵜木 祐史	同 准教授
		本多 清志	中国天津大学 教授

論文の内容の要旨

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 antagonist-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.

論文審査の結果の要旨

音声コミュニケーションには、人がどのように発声・発話器官を制御して音声を生成

しているかが、音声科学において重要な課題である。発話過程では、人間が発話目標に基づいて発話器官を迅速で正確に制御することにより発話運動を実現している。そのなか、迅速な動きはフィードフォワード制御により、正確な動きをフィードバックにより実現しているとされる。本論文では、フィードフォワード制御のため、発話目標から調音筋の活動パターンへのマッピングファンクションを研究しておく。

これまでに、人間の運動制御に関する計算モデル（たとえば川人モデル）は、主に骨格的な運動を対象としたが、発話時舌のような軟弾性体の変形による運動の制御を取り上げるのは少ない。骨格的な運動の制御と異なり、軟弾性体の変形による運動においては、調音筋の間の拮抗関係と協調関係を動き箇所によって変わり、空間依存性を持っている。どうやって空間依存の調音筋から自動的に正確な筋活動パターンを推定するかが挑戦的な難題である。本論文では、動的な主成分（PCA）空間を提案し、任意の動き箇所に調音筋間の相対関係を正確に把握できて、適切な筋活動パターンを推定することが可能となった。その上、フィードバックロープを用いて、発話目標から調音筋の活動パターンへのフィードフォワードマッピングを学習した。筋電図により計測された筋の活動パターンと比較して、フィードフォワードマッピングの正確さを確認した。

筋電図は一般的に大きな外舌筋の計測に有効で、小さな内舌筋の計測に難しい。それ故、本論文では、発話モデルを用いて、発話時内舌筋の声道形状の形成への貢献を定量的に評価した。さらに、体性感覚に基いたフィードバックを導入した後、外力による摂動への補正機能を考察した。その結果より、学習したフィードフォワードマッピングと体性感覚によるフィードバックを取り入れた発話機構モデルは、より安定で正確な発話を実現することを確認した。

以上、本論文は、3次元生理学的発話機構モデルの制御を目指して、フィードバックロープを用いて、発話目標から調音筋の活動パターンへのフィードフォワードマッピングの学習を研究したものであり、学術的に貢献するところが大きい。よって博士（情報科学）の学位論文として十分価値あるものと認めた。