

Title	音声生成と理解における神経振動メカニズムに関する研究
Author(s)	趙, 彬
Citation	
Issue Date	2022-03
Type	Thesis or Dissertation
Text version	ETD
URL	http://hdl.handle.net/10119/17792
Rights	
Description	Supervisor: 鷗木 祐史, 先端科学技術研究科, 博士

氏名	ZHAO, Bin		
学位の種類	博士 (情報科学)		
学位記番号	博情第 469 号		
学位授与年月日	令和 4 年 3 月 24 日		
論文題目	Investigation on the neural oscillatory mechanisms of speech production and comprehension		
論文審査委員	主査	鵜木 祐史	北陸先端科学技術大学院大学 教授
		赤木 正人	同 教授
		党 建武	同 教授
		吉高 淳夫	同 准教授
		日高 昇平	同 准教授
		王 龍標	天津大学 教授

論文の内容の要旨

Speech communication is a trait of human species. It is the most natural and convenient means of interpersonal interaction and information acquisition. Normally, we can quickly and accurately understand the meaning of spoken language based on our acquired linguistic knowledge. Also, we can precisely control articulatory movements via appropriate speech planning, motor programming and auditory feedback. These seemingly quite flexible and easy operations, however, require extremely complex brain network coordination. With the mutual development and continuous interactions, there is a tight link formed between the speech comprehension and production systems, called the 'speech chain'. Understanding the principles governing the neural dynamics related to the speech chain remains to be an unmet goal within neuroscience. Recent brain imaging advances have contributed abundant of anatomical and functional cortical localizations and neural pathways. However, it is not sufficient for us to form a comprehensive understanding of the organizational principles underlying the local computation and long-range communication across multi-scale brain networks. Our study aims to unravel the speech production and comprehension mystery by probing into the neural oscillatory mechanisms and reconstructing the temporal-spatial-spectral brain network dynamics in a series of listening and speaking tasks. Technically, we utilized a multi-modal data acquisition system, which includes a high-density (128-channel) electroencephalography (EEG) recorder, an eye movement tracking system, and an electret condenser microphone. Experimentally, we designed a word listening task and a sentence oral reading task to explore speech processing at different linguistic scales and along successive functional stages. Materially, we take advantage of distinctive Chinese properties (e.g., uniformity and rhythmicity) to address some linguistic controversies that have not been settled in other languages. Computationally, we performed EEG artifact reduction, source reconstruction, effective connectivity estimation, and cross-frequency coupling (CFC) evaluation, etc. These developed methods enable us to investigate some interesting topics, such as the motor theory of speech perception, the syntactic and semantic effects in sentence phrase building, and the interactive nature of the speech perception-production loops. By incorporating prevailing theoretical and computational frameworks with our results in the brain network dynamics and CFC patterns, we proposed a neurofunctional model of speech production and comprehension (SPAC) to explain the dynamic, hierarchical and interactive organization of speech functions. The SPAC model (1) considers the active

nature of the speech functions with top-down regulations from higher cognition; (2) extends the framework up to the sentence level with the consideration of syntactic and semantic effects on low-level sensorimotor processing. (3) complements earlier anatomical and functional models with spatiotemporal brain network dynamics. (4) indicates bidirectional information flows in the dorsal stream for speech production and ventral stream for speech comprehension. (5) explains with CFC mechanisms for the bridging of linguistic form diversities and representational hierarchies, as well as the bidirectional interactions between the bottom-up sensory input and top-down cognitive regulations. The SPAC model is supposed to forward our understanding of the speech chain from a more comprehensive perspective.

Key words: Speech production and comprehension, cross-frequency coupling, brain network dynamics, EEG source reconstruction, SPAC, model

論文審査の結果の要旨

本論文は、音声生成と音声理解の脳活動の計測とその計測結果から神経振動メカニズムを検討したものであり、音声聴取・音声発話の一連のタスクにおける音声生成と音声理解のメカニズムとその役割を明らかにすることを狙いとしている。本研究では、発話音声を音素、語、文といった段階で捉え、高密度脳波計測器（128 ch ElectroEncephaloGraphy, EEG）、眼球運動測定器、コンデンサマイクで構成されるマルチモーダルデータ観測システムを利用して、中国語（北京語）を母語とする 20 名の大学生を対象に、連続文の朗読における神経生理学的データと行動学的データを計測した。また、計測結果から、朗読時の音声処理における脳活動のダイナミクス特性を説明するために、脳のどの部位がどのようなタイミングで活動し情報を受け渡して、音声の生成と理解を行っているかについて、朗読時の眼球運動をタイミングの指標とし、脳波計測による活動部位の特定、脳波の振幅および位相の相関分析による部位間のネットワークの特定、および、脳波周波数の分類による活動の意味推定を行った。特に、神経振動のクロス周波数カップリングを用いて、時間・空間・周波数の次元から神経振動メカニズムを考察し、位相・振幅カップリングによる脳の各部位間で異なる周波数帯域の関係や、位相・位相カップリングによる脳の各部位の連結を明らかにした。

これらの結果に基づき、音声生成と音声理解の神経機能的モデル（SPAC モデル）を提案した。提案モデルは、従来の言語二重通路モデルと階層状態フィードバック制御モデルを拡張して、音素、単語と文レベルの音声知覚と生成の相互作用およびトップダウンとボトムアップのメカニズムを説明することができる。また、提案モデルは、言語形式の多様性と表現階層を橋渡するクロス周波数カップリングだけでなく、感覚入力からのボトムアップと認知処理からのトップダウンの相互作用も説明できる。そのため、提案法モデルは、従来モデルと比較して、脳活動計測から得られた知見を説明できるだけでなく、現在知られているモデルに新たな知見を加えることもできる。

以上、本論文は、音声生成と音声理解に関する神経機能的な研究について新たな知見を与えるものである。また、音声知覚も含めた言葉の鎖における音声生成・知覚におけるメカニズムの解明に向けて、基礎的かつ重要な知見を与えるものであり、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。