

Title	音声知覚における脳内の音声エンコーディング・デコーディングのプロセスの研究
Author(s)	周, 迪
Citation	
Issue Date	2023-09
Type	Thesis or Dissertation
Text version	ETD
URL	<a href="http://hdl.handle.net/10119/18781">http://hdl.handle.net/10119/18781</a>
Rights	
Description	Supervisor: 鶴木祐史, 先端科学技術研究科, 博士

氏名	Zhou Di		
学位の種類	博士 (情報科学)		
学位記番号	博情第 516 号		
学位授与年月日	令和 5 年 9 月 22 日		
論文題目	Speech encoding and decoding processes in the brain during speech perception		
論文審査委員	鶴木 祐史	北陸先端科学技術大学院大学	教授
	SAKTI Sakriani Watiasri	同	准教授
	吉高 淳夫	同	准教授
	岡田 将吾	同	准教授
	党 建武	天津大学	教授
	ZHANG Gaoyan	同	准教授

## 論文の内容の要旨

### Research Background

In recent years, there has been a growing trend towards using naturalistic setups in speech research. These setups involve presenting participants with continuous speech, such as stories or conversations, instead of isolated words or sentences. This provides a more realistic understanding of speech processing and comprehension in real-world conditions. However, fMRI studies have found that naturalistic speech processing elicits more widespread brain activation compared to traditional setups. This poses a challenge for traditional neurocognitive models of language, which are based on isolated words or simple sentences and have a strong left hemisphere bias. These models fail to explain how naturalistic speech is processed in the brain.

### Research Methodology and Purpose

This study is guided by the theory of systems, which allows us to explore and assess the functioning of the brain system by estimating its encoding function. This study focuses on the process of semantic processing during speech perception and investigates the extraction of the temporal amplitude envelope (TAE) from the speech signal. This TAE carries crucial semantic information and is encoded by specific brain regions involved in semantic processing. This study aims to describe the encoding of TAE in the brain using a linear time-invariant (LTI) model. Given the linearity of the system, it can reverse this transformation and recover the original semantic information of TAE. The research philosophy is based on the belief that by accurately estimating the encoding function and capturing the information at the semantic processing stage, it may be possible to infer what is being heard solely from the decoding process, without directly hearing the sound.

To validate this philosophy, this study aims to (1) identify the brain regions responsible for semantic processing, (2) determine the semantic representations in the brain regions, and (3) explore the possibility of recovering the original semantic information from these representations. In the experiment, this study exposed participants to both normal speech, which contained semantic information, and time-reversed speech, which disrupted the semantic content. The aim was to compare the brain activity patterns between the two conditions and identify the specific brain regions responsible for processing semantic information. Then, it could proceed to reconstruct the original TAE information from these identified brain regions. And this study employed a hyper-alignment method

to map EEG signals from the scalp to the cortex level to overcome the spatial limitation of EEG.

### **Results**

- (1) Semantic processing engages frontal, temporal, and cingulate brain areas, it involves a widespread distribution of brain regions beyond the traditional temporal and frontal areas.
- (2) Semantic representations can be found in these brain areas. Moreover, this study has observed a stronger correlation between neural oscillations in the delta and gamma frequency bands and semantic processing.
- (3) Through the reverse decoding process, this study successfully reconstructed the TAEs of speech from brain activity and recovered the semantic information using the NVS technique.

### **Research significance**

Our study is unique in that it combines both temporal and spatial dimensions to offer new insights into the natural language processing that occurs in the brain. Using an innovative methodology, we have been able to estimate encoding functions across various brain regions - a feat that has previously been challenging in both fMRI and EEG research. By challenging traditional models, we've emphasized the extensive involvement of multiple brain regions and the dynamic nature of their encoding capabilities. Our findings suggest that the desynchronization between different subnetworks, especially within the frontal and temporal areas, plays a crucial role in the brain's semantic information processing mechanism. In our study, we were able to successfully reconstruct the TAEs of speech from brain activity and recovered semantic information using noise-vocoded speech (NVS). While restoring speech intelligibility on unknown data remains a challenge due to noise and inter-subject variability, we were able to achieve a perfect fit of TAEs and recovered semantic information to a certain level of speech intelligibility in the known training dataset. This contributes to a deeper understanding of language processing and has potential implications for technologies like speech-brain interfaces in the future.

**Keywords:** electroencephalography (EEG), speech encoding/decoding, temporal response function, source localization, neural entrainment

## 論文審査の結果の要旨

音声生成知覚に関する神経科学の研究の重要な狙い一つは、知覚された音声を脳がどのように理解しているかを明らかにすることである。これまでに、孤立単語や句を使用して人間の脳における音声理解のメカニズムを調査した研究があり、**dual-stream** という言語モデルが提案された。しかし、このようなタスクは日常生活における音声理解とかけ離れており、**dual-stream** モデルでは自然な音声理解のメカニズムを解釈し難い。そのため、連続音声の処理過程に着目する必要がある。

本研究では、被験者に星新一の短編小説（中国語バージョン）を聴かせながら彼らの脳波信号（**EEG**）を収集し、音声の振幅包絡線（**TAE**）に含まれる語彙情報が脳内でどのように処理されるか、その過程に着目し、**TAE** から **EEG** 信号へのエンコーディング過程を線形時不変システムでモデル化した。このエンコーディングは音声理解のメカニズムの解明に大きな役割を果たしている。次に、本研究では、**EEG** の空間分解能が低い問題を改善しつつ、空間情報に係る脳皮質レベルのエンコーディングの機能を推定した。その結果、自然言語の語彙処理に関与する重要な脳領域を特定することができた。また、これらの脳領域から得られた **EEG** から、意味表現に基づいた音声に含まれる語彙情報を再構築できることも明らかにした。これは、実際に音声を聞かなくても、その人の脳活動から、その人が聞いた音声の内容を知ることができることを意味する。最後に、本研究では、自然な音声理解のための動的な脳ネットワークを構築した。この構築した脳ネットワークでは、自然言語の語彙処理が、**dual-stream** モデルが指摘したような、特定の脳領域に限定されるのではなく、左右大脳半球の広い範囲にわたって行われていることを示唆した。また、推定したエンコーディングの機能関数から、脳内に二つの脱同期化サブネットワークが存在することが示唆された。この脱同期化の脳活動は自然言語の語彙処理と関わる重要なメカニズムであることを示している。

本研究の結果は、聴覚障害者の音声知覚能力の改善に役立てられ、効果的なリハビリ方法や補助技術の開発に役を立つものと考えられる。また、言語理解のメカニズムの深い検討を進めることで、人工知能分野への波及効果も期待することができるため、スマートで自然な音声認識や自然言語処理システムの開発に貢献するものと考えられる。

以上、本論文は、音声知覚に係る脳活動から音声のエンコーディング・デコーディング過程の理解を示した。本技術は、音声生成知覚メカニズムの解明の一助になり、応用範囲が広く、学術的に貢献するところも大きい。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。