先端科学技術研究科　修士論文要旨

所属研究室 (主指導教員)	自然言語処理学 (渡辺　太郎　(教授))		
学籍番号	2411125	提出日	令和 7年 7月 24日
学生氏名	坂上　温紀		
論文題目	Exploring Language Models' Understanding of Vowel Articulation: From a Visual Information Perspective		

要旨

How human vocalizations are articulated can be described by analyzing the tongue position.
Researchers have discovered this through lived experience and detailed observation, including by MRI.
Using this knowledge and personal experience, teachers can understand and explain the relationship between tongue positions and vowels, thus helping language students to learn pronunciation.
Our preliminary studies suggest that language models (LMs), trained on extensive data from the linguistic and medical fields, can explain the mechanisms of vowel pronunciation.
However, it is unclear whether multimodal LMs, such as Large-scale Vision Language Models (LVLMs), sufficiently align textual information with visual information.
From this, the research question arises: Do LVLMs associate real tongue positions with vowel articulation?
To investigate whether visual information can help LVLMs understand vowel articulation based on tongue positions, this study created video and image datasets from real-time MRI samples.
We discuss how LVLMs predict vowels by analyzing several experimental results.
Our findings suggest that LVLMs can potentially identify the interrelationship between vowels and tongue positions when reference examples are provided, but have difficulty without them. LVLMs also perform better when inferring directly from visual information than when offered text descriptions before inferring.