

早稲田大学 人間科学学術院 人間科学会 諸費用補助成果報告書 (Web 公開用)

申請者 (ふりがな)	大浦 杏奈 (おおうら あんな)
所属・資格 (※学生は課程・学年を記載。卒業生・修了生は卒業・修了年月も記載)	博士後期課程 1 年
発表年月 または事業開催年月	2024 年 9 月
発表学会・大会 または事業名・開催場所	INTERSPEECH 2024
発表者 (※学会発表の場合のみ記載、共同発表者の氏名も記載すること)	Anna Oura, Hideaki Kikuchi, Tetsunori Kobayashi
発表題目 (※学会発表の場合のみ記載)	Preprocessing for acoustic-to-articulatory inversion using real-time MRI movies of Japanese speech
発表の概要と成果 (抄録を公開している URL がある場合、「概要・成果」を記載した上で、URL を末尾に記してください。また、抄録 PDF は別途ご提出ください。なお、抄録 PDF は Web 上には公開されません。)	
— Summary —	
<p>Acoustic-to-articulatory inversion (AAI) estimates the articulation movements by using acoustic speech signals. AAI has been investigated using various approaches for decades. The traditional AAI relies on indirect estimation using articulatory models. However, recent advancements have proposed the use of machine learning models to directly output real-time MRI (rtMRI) movies.</p> <p>Previous studies on rtMRI movie estimation have identified the lack of preprocessing of input data as a key issue. Additionally, while prior research explored models using LSTM, it is anticipated that implementing a Bidirectional LSTM (BLSTM) could further enhance the smoothness of the estimated rtMRI movie outputs.</p> <p>In this study, we constructed the model used in previous research with rtMRI movies of Japanese speech and proposed normalization and filtering processes as preprocessing methods. We also investigated the estimation accuracy of a BLSTM model by modifying certain aspects of the existing LSTM model.</p>	
— Results —	
<p>Three key findings were obtained:</p> <ul style="list-style-type: none"> - Applying normalization stabilizes the facial position, enabling more accurate estimations. - Filtering allows the model to focus on articulators, resulting in high-precision estimations. - The BLSTM model achieves smooth rtMRI movie estimations even without preprocessing. <p>Based on these findings, we concluded that the preprocessing methods proposed in this study are effective for AAI. Moving forward, we aim to develop an AAI model capable of speaker-independent estimation.</p>	

※無断転載禁止