Peer-Reviewed Structured Abstract

Visual-Auditory Multimodal Decoding from Multi-Channel Human Intracranial EEG

Baowen Cheng[1], Ke Chen[1], Haoyu Hua[2], Zehan Wu[3], Zhuofan Zhao[1], Liang Chen[3], Ying Mao[3], Meng Li[1,2]

[1] Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Sciences, Shanghai, China. Email: chengbaowen23@mails.ucas.ac.cn, ck97@mail.sim.ac.cn, zhaozhuofan23@mails.ucas.ac.cn, limeng.braindecoder@gmail.com

[2] INSIDE Institute for Biological and Artificial Intelligence, Shanghai, China. Email: huahy@stonehill-tech.com, limeng.braindecoder@gmail.com

[3] Department of Neurosurgery, Huashan Hospital, Fudan University, Shanghai, China.
Email: zehanred11@163.com, hschenliang@fudan.edu.cn, maoying@fudan.edu.cn

Summary

The human brain's response to video stimuli involves the engagement of numerous sensory modalities, including visual and auditory systems. Nevertheless, the neural underpinnings governing the processing of video stimuli remain incompletely understood. One critical factor contributing to this knowledge gap is the scarcity of corresponding human electrophysiology datasets, particularly those collected from deep brain regions. This work presented an intracranial EEG (iEEG) dataset acquired from 66 electrodes in an epilepsy patient over the course of 300 minutes while they viewed customized video stimuli. The dataset encompassed a wide range of cortical and subcortical brain regions, with the goal of enhancing our comprehension of the perceptual processing mechanisms involved during video observation. Based on this dataset, we employed a classification experiment to assess the significance of brain regions involved in the processing of diverse videos, auditory stimuli, and visual scenes within the human brain.

Background

Substantial progress in brain-computer interface technology has enabled remarkable achievements in linguistic and visual decoding tasks [1,2]. Accumulating evidence indicates that neurological signals originating from deep brain regions convey significant information pertinent to various brain functions [3,4,5]. Real-world video stimuli demand the integration of multiple sensory processes, and research examining brain activity during multisensory stimulation remains limited.

iEEG recordings are commonly obtained during the treatment course of epilepsy patients to identify the epileptogenic foci [6]. In this study, we developed a virtual reality (VR)-based paradigm that integrated multisensory stimuli, which was conducted with an epilepsy patient who had implanted iEEG electrodes. iEEG signals were recorded through 66 channels over a total duration of approximately 300 minutes.

We further conducted a detailed investigation into the respective contributions of disparate brain regions to various decoding tasks, and the findings revealed that numerous brain areas were engaged, including temporal lobe, hippocampus, temporo-occipital junction, inferior frontal gyrus, insula, fusiform gyrus. Thus, video stimuli offer a wealth of contextual details that facilitate a more comprehensive understanding of how the brain processes and responds to complex situations.

Methodology

Participants and Tasks

Participants. A female individual diagnosed with epilepsy participated in our study. The participant underwent implantation of iEEG electrodes as part of her clinical treatment for epilepsy. The participant voluntarily enrolled in the study and provided written informed consent. The experimental protocol and data collection procedures were reviewed and approved by the institutional review board of Huashan Hospital, Fudan University.

Video Stimuli. Two well-known audiovisual stimuli, a Michael Jackson music video and a Harry Potter film, were selected. From each, a 15-second clip was extracted, resulting in a total of two video clips. For each video, five variations were designed: the complete video, removal of specific frames, removal of sounds, the absence of audio content with intermittent gaps in the middle frames, as well as the absence of visual frames with partial sound removal in the middle. These five variations were presented as a coherent sequence, played in order. The videos were repeated three times per trial, yielding a total duration of 10 minutes per trial (Fig. 1). During the experiment, the participant watched videos using VR goggles, ensuring their field of vision was limited to the intended content (Fig. 1a).

iEEG Recoding

Electrode Localization and Visualization. The number and positioning of the electrodes were selected based on clinical requirements. Anatomical designations were obtained through Freesurfer's cortical parcellation [7].

Neural signal recording. Neural activity and timestamp of each video chips was recorded using a multi-channel electrophysiological recording system (EEG-1200C, Nihon Kohden, Tokyo, Japan) with a sampling rate of 2000 Hz. Pico 4 VR headset utilized for viewing the video.

Data preprocessing. The neural signals were extracted based on the data labels recorded during the experiment, and then downsampled from 2 kHz to 400 Hz and notch filtered at 50 Hz, 100 Hz, and 150 Hz. All neural signals underwent bipolar referencing and normalization.

Visual-Auditory Multimodal Decoding

Model structure. Our model includes a TransformerEncoder architecture and a fully connected layer to facilitate the classification task.

Decoding setup. Neural data was partitioned into training and validation subsets, maintaining an approximate 4:1 ratio. Adam algorithm with an initial learning rate of 0.00001, and a cosine decay strategy was employed to govern the learning rate schedule. The training process spanned ~500 epochs on an NVIDIA 4090 GPU.

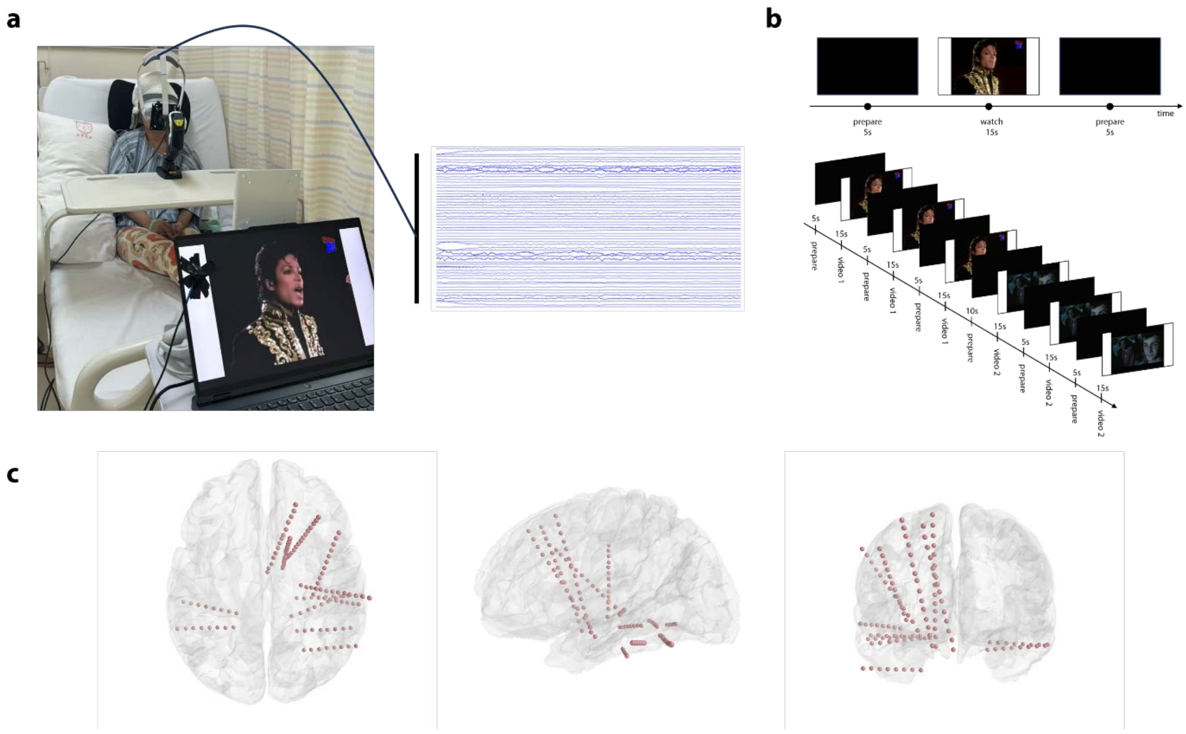

Fig. 1. iEEG Recording and Experiment Setup. (a) The participant used a VR headset to watch the video stimuli while iEEG signals were simultaneously recorded. (b) Video stimulus experimental paradigm. (c) The distribution of iEEG electrodes.

Results and Observations

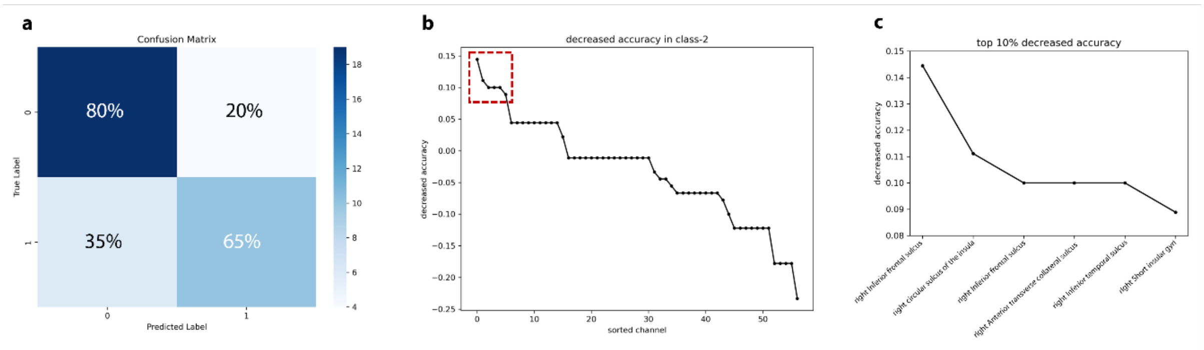

Fig. 2. Decoding Analysis. (a) Classification performance of two video clips based on iEEG signals. (b) The iEEG channel's contribution to classification task by ablation experiments. (c)Top six contributed brain regions.

Datasets

iEEG data was collected from one patient over five days. The inclusion of markers enabled flexible utilization of the data according to researchers' specific requirements.

Decoding Analysis Revealed the Regional Contributions in Representing Videos

We first verified the correspondence between EEG data and video stimuli by using EEG to predict video clips. The classification results demonstrated that iEEG could effectively predict the content of the video (Fig. 2a). We employed ablation experiments to assess the contribution of brain regions to the classification performances. We systematically set the input layer weights of individual electrodes to zero and observed the impact of this manipulation on the final result. As shown in Fig. 2b and c, our results indicated that multiple brain regions located in the right cerebral hemisphere demonstrated enhanced significance in the performance of video classification tasks.

Conclusion

In this study, we developed a video stimulus paradigm using VR goggles. This resulted in a dataset comprising simultaneously recorded iEEG signals and video stimuli. This dataset can be leveraged for various investigations of the human brain's responses to multisensory inputs. Furthermore, we decoded video information from iEEG and showed that the right cerebral hemisphere played a more crucial role in video decoding.

Acknowledgement

The following funding sources provided to M.L. are acknowledged: the National Key Research and Development Program of China (2021YFF1200700), the National Natural Science Foundation of China (32371083), the National Key Research and Development Program of China (2021YFC2501500). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

References

1. Ikegawa Y, Fukuma R, Sugano H, Oshino S, Tani N, Tamura K. (2024). Text and image generation from intracranial electroencephalography using an embedding space for text and images. *J Neural Eng*, 21(3). Available from: National Library of Medicine [accessed 30 July 2024].
2. Nicholas S. Card, Maitreyee Wairagkar, Carrina Iacobacci. (2024). An Accurate and Rapidly Calibrating Speech Neuroprosthesis. *N Engl J Med*, 391, pp.609-618. Available from: The new England Journal of Medicine [accessed 1 September 2024].
3. Hu R, Eskandar E, Williams Z. (2009). Role of deep brain stimulation in modulating memory formation and recall. *Neurosurg Focus*, 27(1), pp.E3. Available from: National Library of Medicine [accessed 12 April 2023].
4. Fortin NJ, Agster KL, Eichenbaum HB. (2002). Critical role of the hippocampus in memory for sequences of events. *Nat Neurosci*, 5(5), pp.458-62. Available from: Nature Neuroscience [accessed 12 April 2023].
5. Wolf-Julian Neumann, Andreas Horn, Andrea A. Kuhn. (2023). Insights and opportunities for deep brain stimulation as a brain circuit intervention. *Trends in Neurosciences*, 46(6), pp.472-487. Available from: National Library of Medicine [accessed 12 September 2023].
6. Mercier MR, Dubarry AS, Tadel F, Avanzini P, Axmacher N. (2022). Advances in human intracranial electroencephalography research, guidelines and good practices. *Neuroimage*, 260, pp.119438. Available from: National Library of Medicine [accessed 12 September 2024].
7. Bruce Fischl. (2012). FreeSurfer. *NeuroImage*, 62(2), pp.774-781. Available from: ScienceDirect [accessed 1 July 2022].