arxiv:2506.00854v3 中译本

脑电波文本对齐研究：通过大型语言模型和对比学习探索开放词汇中文文本与脑电波的对齐关系
一项利用大型语言模型和对比学习在 ChineseEEG 上进行开放词汇中文文本-脑电波对齐的探索性研究

Jacky Tai-Yu Lu* [1], **Jung Chiang*** [1,2], **Chi-Sheng Chen*** [3],
Anna Nai-Yun Tung[1,4], **Hsiang Wei Hu**[1], **Yuan Chiao Cheng**[1,6]
[1]Department of Artificial Intelligence in Healthcare,
International Academia of Biomedical Innovation Technology, Reno, USA
[2]National Taiwan University
[3]Neuro Industry Research, Neuro Industry, Inc.
[4]Department of Biomedical Engineering,
University of Southern California
[5]Taiwan Artificial Intelligence Association
b0090100@gmail.com, Lloyd4127@icloud.com,m50816m50816@gmail.com,
naiyun@usc.edu, Hw.Hsiang.wei@gmail.com, yuanchiao.cheng@gmail.com

Abstract

我们提出了 EEG2TEXT-CN，据我们所知，这代表了最早的针对中文的开放词汇量 EEG 到文本生成框架之一。该架构基于一个生物学基础的 EEG 编码器（NICE-EEG）和一个紧凑的预训练语言模型（MiniLM），通过掩码预训练和对比学习将多通道脑电信号与自然语言表示对齐。使用 ChineseEEG 数据集的一个子集，其中每个句子包含大约十个中文字符，并且这些字符与以 256 Hz 记录的 128 通道 EEG 信号对齐，我们将 EEG 分割为每个字符的嵌入，并在零样本设置下预测完整的句子。解码器通过教师强迫和填充掩码训练来适应可变长度序列。通过对超过 1,500 个训练-验证句子和 300 个保留测试样本进行评估显示了有希望的词汇对齐，最佳 BLEU-1 得分为 6.38%。虽然句法流畅性仍然是一个挑战，但我们的研究结果证明了从 EEG 中解码非语音、跨模态语言的可行性。这项工作开启了多语言脑到文本研究的新方向，并为未来中文认知语言接口奠定了基础。

1 介绍

将大脑活动转化为自然语言已经成为脑机接口 (BCIs)、认知神经科学和神经语言建模等领域的重要目标。在各种神经成像方法中，由于其非侵入性、精细的时间分辨率以及现实世界部

*Equal contribution

Preprint. Under review.

署的可行性，脑电图 (EEG) 仍然是一个有吸引力的选择。然而，大多数基于 EEG 的解码系统受限于狭窄的词汇量、刺激锁定范式和浅层建模流程。这些限制阻碍了它们捕捉大脑中语言处理连续性和丰富上下文的能力。

深度学习和自监督表示学习的最新进展引入了开放词汇 EEG 到文本生成的可能性，其中神经活动可以转化为自由形式、连贯的文本输出。这一演变标志着范式转变——从分类离散的 EEG 事件到学习语境化的脑语言映射。

1.1 文献回顾

开创性的脑电图到文本解码工作包括 EEG2TEXT 框架，该框架引入了掩码脑电图建模策略和多视图变压器架构，从句子级别的脑电图信号 [1] 中生成结构化的文本。这种方法摒弃了刺激对齐的解码方式，直接学习跨大脑区域原始脑电段落的语义映射。此外，DeWave 应用量化变分编码来推导脑电图的离散代码表示，提高了模型应对个体差异和会话噪声的鲁棒性 [2]。这些进展突显了无监督学习和对比学习在从高维脑电图数据中提取有意义潜在结构的作用。

在一个互补的方向上，[3] 探索了通过将脑电图特征与预训练的语言模型对齐来进行零样本的脑电图情感分类，展示了语义对齐在基于脑电图的自然语言处理任务中的潜力。同时，NICE-EEG 框架通过引入空间和图形注意力机制增强了从脑电图中进行视觉解码的能力，提供了一种以生物学为依据的时空神经动力学编码方式 [4]。NICE-EEG 使用对比学习与语义嵌入的方法，为跨受试者和会话的解码提供了一个可泛化的机制。

然而，大多数这些研究都集中在英语数据集和字母语言结构上。最近发布的 ChineseEEG 数据集 [5] 为在表意字符语言环境中调查脑电图解码提供了独特的机会，这涉及不同的认知和神经编码过程。尽管其内容丰富，但此前没有模型充分利用先进的解码架构以及该数据集支持中文句子级别神经解码的潜力。

1.2 EEG2TEXT-CN：多视角语义对齐的中文脑电图

在这项研究中，我们介绍了 EEG2TEXT-CN，一个将 NICE-EEG 的生物基础编码与中文预训练语言模型的语义表示能力相结合的统一框架。我们的方法首次探索了从脑电图（EEG）进行开放词汇中文句子对齐的研究，使用的是来自中国 EEG 语料库中母语者阅读两本小说时记录的完整句子数据。

所提出的架构包含两个关键组件：

- 一个空间感知的 EEG 编码器，基于 NICE-EEG 改编，集成了自注意力和图注意力层以解决 EEG 地形问题；
- 一个受 EEG2TEXT 启发的掩码预训练模块，使模型能够在无视觉设置中从受损的脑电图-语言配对中学习语义关联。

为了增强对齐，我们将预训练的中文语言模型（我们在本工作中使用 all-MiniLM-L12-v2[6]）作为 EEG 对齐嵌入的语义锚点。这种多模态融合使我们的模型能够在零样本设置下将稀疏的 EEG 信号映射到有意义的语言表示。

2 相关工作

尽管最近的研究进展导致了脑电图到图像对齐研究的激增——例如 NICE [4], MUSE [7], QMCL [8], NERV [9] 和 ATM [10]，但相对较少的研究解决了将脑电图与自然语言对齐这

一更具挑战性的任务。相比引发强烈视觉皮层反应的图像刺激，语言刺激涉及更抽象、时间上更为延长的神经过程，使得解码问题变得更加复杂。因此，脑电图到文本的对齐仍然很少被探索，尤其是在开放词汇句子生成的背景下。

脑电图到文本是一个位于神经科学和人工智能交叉领域的新兴领域，旨在直接从大脑信号中解码自然语言。得益于深度神经网络的进步和改进的脑电图记录技术，该领域已经从闭合词汇识别发展到了更加开放生成的语言模型。然而，迄今为止大多数研究仅专注于英语，这在多语种脑电图解码方面留下了一个显著空白。在此部分，我们将回顾脑电图到文本系统的关键进展，重点关注模型设计、学习范式和可用数据集，并确定那些推动我们研究的未满足挑战。

2.1 闭词汇与开词汇的 EEG 解码

初步的脑电图到文本的研究集中在语义空间有限的封闭词汇任务上。最近的努力转向了开放词汇生成，其中模型试图从脑电图输入中生成自由格式的文本。例如，EEG2TEXT [1] 引入了一种在脑电图数据上预训练的多视角变压器，与之前的基准相比，在 BLEU 和 ROUGE 分数上提高了 5%。这些结果突出了结合脑电图特征提取与大型语言模型（LLM）的潜力。然而，这种开放词汇系统在多种语言中的有效性仍然研究不足。

2.2 对比学习和自动编码器-based 方法

为了弥合脑电图信号与文本之间的模态差距，若干研究利用了对比学习和自动编码策略。CET-MAE 和 E2T-PTR [11] 通过多流架构联合嵌入 EEG 和文本。E2T-PTR 使用基于 BART 的解码器，并在 ZuCo 数据集上显示出显著的改进（ROUGE-1 F1 提升了 8.34%，BLEU-4 提升了 32.21%）。尽管这些方法提高了对齐和生成质量，但它们通常假设输入是良好分段的且以英语为中心，并很少在不同语言或书写系统中进行验证。

2.3 想象语音解码和无声通信

另一个重要的研究方向涉及从 EEG 解码想象或无声的言语。Xiong 等人 [12] 提出了 ETS，这是一个端到端的 EEG 到语音框架，使用 X 形 DNN 从想象输入生成可理解的语音。该模型在 13 名参与者中实现了平均 91.23%的准确率。尽管对于无声交流很有前景，但这类模型通常依赖于音素级对齐或基于音频的监督，这些方法在转换到汉字这样的表意语言时不太适用。

2.4 自监督和离散编码方法

为了提高泛化能力和数据效率，提出了自监督方法。BELT [13, 14] 引入了一种 D-Conformer，将 EEG 映射为离散表示，指导与文本序列的对齐。DeWave [2] 结合了离散变分编码和预训练语言模型来解决标记不匹配和主体变异问题。这些工作提高了鲁棒性，但主要还是针对英语单词级别的解码，并且缺乏在表意文字语言中的评估。

2.5 数据集和语言限制

该领域目前严重依赖 ZuCo 数据集 [15]，其中包含了英语读者的 EEG 和眼动追踪数据。虽然丰富且注释良好，但 ZuCo 仅支持英语刺激物。迄今为止，尽管汉语具有独特的语言和视觉特性，还没有大型公开的 EEG 数据集专注于中文。缺乏此类资源给开发多语言脑机接口带来了挑战。

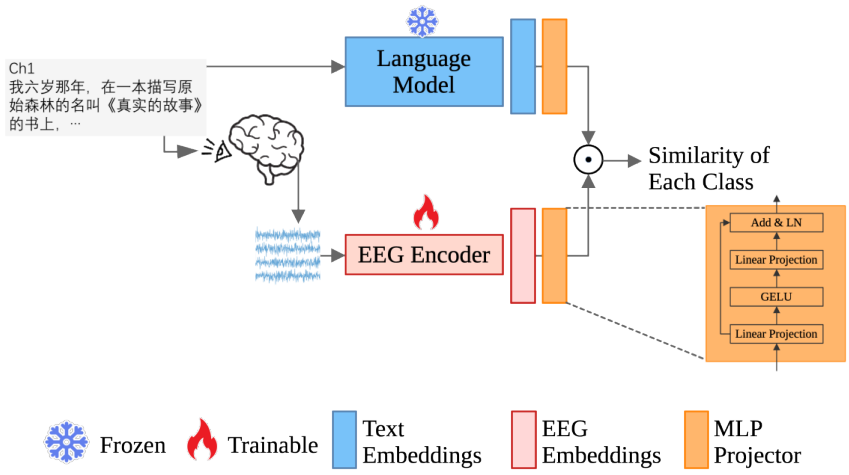

图 1: 我们 EEG2TEXT-CN 对齐模型概述。该模型接受中英文本及其相应的 EEG 信号同步输入对。语言模型（冻结，蓝色）将字符级文本编码为上下文化嵌入，而 EEG 编码器（可训练，红色）将 PCA 压缩的 EEG 段映射到相同的嵌入空间。两个嵌入通过一个共享的 MLP 头进行投影，并使用余弦相似度进行比较以计算类级别对齐。该模型采用对比目标进行训练。

2.6 语音合成与多模态系统

其他努力旨在直接从 EEG 信号中重建语音。BTS [16] 使用 EEG 和音素感知的语音模型生成用户的自己的声音。类似地，使用低成本设备探索了音素级别的 BCI [17]，实现了高达 97% 的准确性。这些研究更侧重于低层次的语音单元而不是全文生成，并且它们对音频语料库的依赖限制了其在非音位书写系统中的应用。

2.7 面向多语言 EEG 解码

总体而言，这些研究强调了从 EEG 解码自然语言方面取得的显著进展。然而，几乎所有先前的研究都局限于字母语言，尤其是英语。处理如中文等表意文字语言的挑战——其中字符级编码、笔画顺序和语法差异很大——尚未得到解决。据我们所知，目前尚无研究探索从 EEG 生成开放词汇的中文句子。

在这项工作中，我们提出了首个中文的脑电图到文本模型。利用新发布的 ChineseEEG 语料库，我们引入了一个基于生物学的脑电图编码器，并将其与语言模型结合进行句子级别的生成。我们的工作扩展了脑电图解码在多语言领域的前沿，为未来关于中文语义解码和脑电波到文本通信的研究奠定了基础。

3 方法

3.1 数据集

我们利用了 ChineseEEG 数据集 [5]，这是一个用于自然中文阅读任务中的语义对齐和神经解码的高密度 EEG 语料库。该数据集包含了来自 10 位参与者的超过 13 小时的 EEG 和眼动追踪记录，每位参与者都在无声地阅读两部中国小说：小王子和加内特 梦想。每个字符以

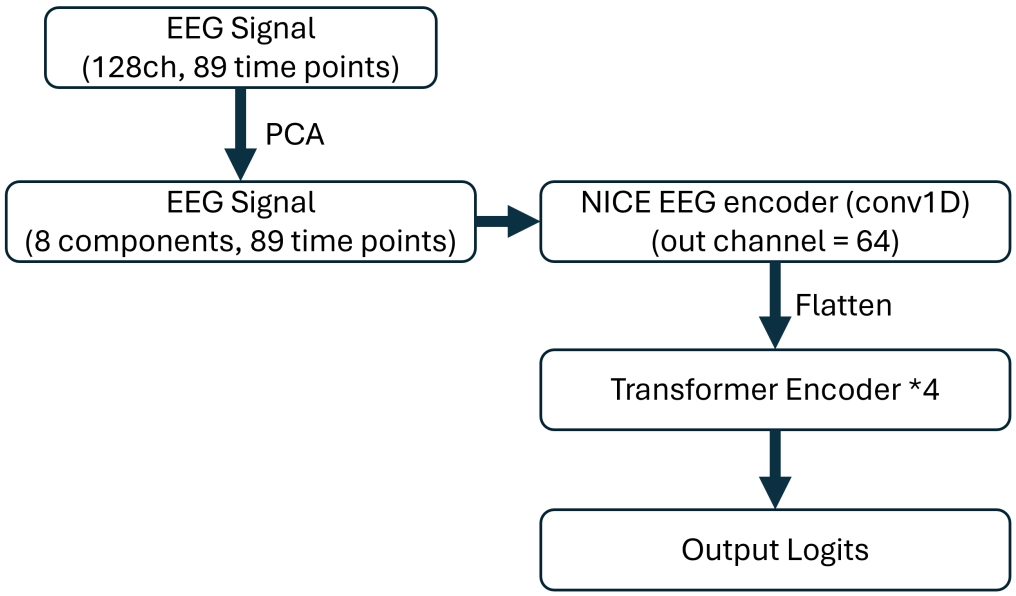

图 2: EEG 编码器-解码器管道。该管道始于形状为 (128 通道, 89 时间点) 的原始 EEG 信号，通过 PCA 减少到 8 个主成分。压缩后的 EEG 数据然后通过使用具有 64 个输出通道的一维卷积层的 NICE EEG 编码器进行传递。得到的特征被展平并通过一组 4 个 Transformer 编码器层进行处理。最后，输出被投影到目标类别对数几率上。

350 毫秒的时间间隔被依次高亮显示且不包含标点符号，从而能够精确对齐文本与 EEG 信号。EEG 数据使用 128 通道的 EGI 几何传感网系统以 256 赫兹的采样率进行记录。

在我们的探索性研究中，我们选择了 sub-04 到 sub-09 在阅读加内特 梦想的第 1–18 章时的数据，共得到了 77,815 个汉字。每个脑电波-文本配对对应最多包含 10 个字符的句子。基于固定的字符持续时间（0.35 秒），我们将脑电波每字符分割为 90 个时间点（$256 \times 0.35 \approx 89.6$，四舍五入到 90），从而得到每个句子的张量形状为（字符数量，128，90）。

为了降低计算成本，我们在通道维度上应用了主成分分析（PCA）[18]，将每个样本压缩到形状（字符数量，90，8）。在训练过程中，不同长度的句子被零填充到最多 10 个字符。在前向传播中使用了填充掩码来防止对填充标记进行梯度更新。然而，这种填充设计使得预测变得更加困难，特别是在推理时句子长度未知的情况下。

3.2 模型架构

我们采用了 NICE-EEG[4] 作为我们的脑电图编码器，因为其在脑电图-图像对齐任务中表现出的性能。这种架构有效地捕捉了高密度脑电图数据中的空间和时间动态。细节如图 2 所示。

对于语言编码器，我们使用了 Hugging Face 提供的轻量级变压器模型 all-MiniLM-L12-v2，该模型提供了高效的上下文嵌入。由于计算限制，我们采用了字符级别的输入策略。尽管分词器可能将频繁出现的字符序列（例如，“全世界”）合并为一个标记，但我们通过强制执行按字符输入和输出来规避这种歧义，而不是依赖于学习到的子词边界。

解码以自回归方式进行，假设已知句子开始和固定长度的输入。

3.3 模型构建

我们将 EEG2TEXT-CN 公式化为一个序列到序列的编码器-解码器模型，该模型从脑电图信号生成中文文本。每个句子由最多 N 个字符组成，每个字符与一个脑电段 $\boldsymbol{X}_i \in \mathbb{R}^{T\times C}$ 相关联，其中包含 $T = 89$ 个时间点（对应于 256 赫兹下 350 毫秒）和通过主成分分析获得的 $C = 8$ 个空间分量。一句话的完整输入表示为 $\boldsymbol{X} = [\boldsymbol{X}_1, \ldots, \boldsymbol{X}_N] \in \mathbb{R}^{N\times T\times C}$。详情如图 1 所示。

为了编码 EEG 片段，我们使用基于 NICE-EEG 架构的卷积编码器 f_{enc}。每个片段 $\boldsymbol{X}_i$ 被转换为固定长度的向量 $\boldsymbol{h}_i = f_{\text{enc}}(\boldsymbol{X}_i) \in \mathbb{R}^d$，从而得到编码序列 $\boldsymbol{H} = [\boldsymbol{h}_1, \ldots, \boldsymbol{h}_N] \in \mathbb{R}^{N\times d}$。在将 $\boldsymbol{H}$ 传递给解码器之前添加位置编码。

解码器是一个基于编码的 EEG 序列和先前生成的标记条件下的 Transformer。在训练过程中，我们应用教师强制：解码器接收移位后的 ground truth 标记序列 $\boldsymbol{Y}_{<t}$ 并输出下一个标记的预测对数几率：

$$\hat{y}_t = \text{Softmax}(W_o \cdot \hat{\boldsymbol{y}}_t + b_o), \tag{1}$$

其中 $W_o \in \mathbb{R}^{|V|\times d}$ 和 $|V| = 250002$ 表示包括 [BOS] 和 [EOS] 标记在内的词汇表大小。

我们使用目标序列上的交叉熵损失来优化模型：

$$\mathcal{L} = -\sum_{t=1}^{N} \log p(y_t \mid y_{<t}, \boldsymbol{X}). \tag{2}$$

应用填充掩码以确保仅在有效字符位置上计算损失。该框架通过利用脑电波数据的空间和时间动态，实现了从 EEG 信号自回归生成中文文本。

3.4 训练和评估

我们从加内特 梦想的 3 个章节的最初 100 个句子片段中构建了我们的训练和验证集，涉及五名参与者。该模型训练了 50 个 epoch。这产生了 1,500 个总样本，随机划分为 1,200 个用于训练和 300 个用于验证。我们在训练期间使用了 teacher forcing[19]，逐步将真实字符输入到解码器中以稳定学习并提高收敛性。

为了评估，我们选择了第六位参与者（sub-06），抽取了相同的 3 章，每章 100 个片段，得到了一个包含 300 个样本的测试集。这种基于个体的划分确保了对未见过的个体的一般化，并避免了数据泄露。

我们在 ChineseEEG 数据集上通过分析每个预测句子中单个字符的输出概率分布来评估 EEG2TEXT-CN。

为了实现公平且可解释的比较，我们选择了表现最佳的模型检查点，并计算了 top-1 预测的 BLEU 分数（BLEU-1 到 BLEU-4）。

BLEU（双语评估辅助）[20] 测量预测序列与真实序列之间的 n 元语法精度，并附带一个简短惩罚以处罚过短的输出。BLEU 分数计算为：

$$\text{BLEU} = \text{BP} \cdot \exp\left(\sum_{n=1}^{N} w_n \log p_n\right) \tag{3}$$

这里，p_n 是修改后的 n-gram 精度直到阶数 N，w_n 是权重（通常是均匀的），而 BP 是简洁性惩罚：

$$\text{BP} = \begin{cases} 1 & \text{if } c > r \\ \exp(1 - r/c) & \text{if } c \leq r \end{cases} \tag{4}$$

其中 c 是候选（预测）句子的长度，r 是参考句子的长度。

具体而言：

- **BLEU-1:** 考虑单字（单字符）重叠；
- **BLEU-2:** 包括二元词组（两个字符序列）精度；
- **BLEU-3:** 考虑三元组；
- **BLEU-4:** 使用四字序列评估流畅度。

这一细致的评估反映了模型捕捉词汇准确性和短距离流畅性的能力。尽管存在开放词汇和低资源设置，我们的模型在 EEG 与文本之间展示了有前景的字符级对齐。这些结果表明，我们的框架为跨模态脑电到语言的中文建模提供了一个可行的基础，开启了未来多语种 EEG 到文本研究的新方向。

表 1: 测试集上使用最佳性能模型（第 49 个周期）的 BLEU-n 分数。

度量	n 元级别	得分
BLEU-1	Unigram	0.0638
BLEU-2	Bigram	0.0212
BLEU-3	Trigram	0.0152
BLEU-4	4-gram	0.0132

表 2: 地面真实值与预测结果的对比（使用 BLEU 分数）。第一个结果来自第 6 个周期，第二个结果来自第 49 个周期。

地面真相（GT）	预测结果（PR）	蓝得分-1	BLEU-2	BLEU-3	BLEU-4
全世界的狼都有一个共	在草原东北端一块马蹄甲	0.0909	0.0302	0.0216	0.0189
全世界的狼都有一个共	石和一块块岩。这样奔跑	0.0909	0.0302	0.0216	0.0189
同的习性，在严寒的冬天	石。终于草棚上的猎人和	0.0909	0.0302	0.0216	0.0189
同的习性，在严寒的冬天	在草原东北端一块马蹄甲	0.0909	0.0302	0.0216	0.0189
集合成群，平时单身独处。	石。终于草棚上的猎人和	0.0830	0.0275	0.0197	0.0172
集合成群，平时单身独处。	在草原东北端一块马蹄甲	0.0000	0.0000	0.0000	0.0000
眼下正是桃红柳绿的春	石。终于草棚上的猎人和	0.0909	0.0302	0.0216	0.0189
眼下正是桃红柳绿的春	在草原东北端一块马蹄甲	0.0000	0.0000	0.0000	0.0000

4 结果分析

我们在中文 EEG 测试集上使用 BLEU-1 到 BLEU-4 评分评估了 EEG2TEXT-CN 模型，结果见表 1 和表 2。表现最佳的检查点（第 49 轮）产生的结果如下：BLEU-1 = 0.0638, BLEU-2 = 0.0212, BLEU-3 = 0.0152，和 BLEU-4 = 0.0132。这些分数表明模型在字符级别上表现出更好的性能，但在较长的 n 元语法一致性和句法流畅性方面存在问题。

为了更好地理解这些结果，我们对预测句子与地面真实（GT）句子进行了定性分析。在少数情况下，模型成功捕捉到了与地面真实（GT）中的重叠字符或语义片段（例如，“石。于草棚上的人和”部分匹配地面真实句子），这导致了非零的 BLEU 分数。然而，在大多数情况下，预测结果仅与地面真实松散相关，并且通常默认为训练语料库中的高频词汇，表明存在强烈的词汇偏见。

这些观察表明，即使在没有视觉线索的零样本设置中，EEG2TEXT-CN 也能学习基本的词汇模式和有限的语义关联。然而，该模型在生成更长、连贯的句子方面仍然面临困难。预测中常见单词的频繁重复也表明，该模型倾向于高概率的标记而不是生成多样化的输出。

尽管 BLEU 分数较低，它们与定性对齐质量呈正相关，证实了 BLEU 可以作为评估 EEG 到文本生成的合理基准指标。在未来的工作中，诸如 BERTScore 或基于嵌入的余弦相似度等指标可能会提供对语义对齐更细致的理解，尤其是在中文方面。

5 消融研究：解码器的影响

为了研究解码器在我们的 EEG2TEXT-CN 架构中的作用，我们通过移除基于变压器的解码器并训练仅包含编码器的变体来进行了一项消融研究。在此设置中，EEG 编码器从 EEG 输入生成一系列嵌入，并且每个嵌入都是通过线性分类头直接映射到一个预测字符上，没有进行任何自回归生成。

形式上，给定输入的 EEG 片段 $\boldsymbol{X} = [\boldsymbol{X}_1, \ldots, \boldsymbol{X}_N] \in \mathbb{R}^{N \times T \times C}$，编码器产生特征向量 $\boldsymbol{H} = [\boldsymbol{h}_1, \ldots, \boldsymbol{h}_N] \in \mathbb{R}^{N \times d}$，这些特征向量通过一个共享的线性分类器：

$$\hat{y}_i = \text{Softmax}(W_c \cdot \boldsymbol{h}_i + b_c),$$

其中 $W_c \in \mathbb{R}^{|V| \times d}$ 是权重矩阵，$|V|$ 是词汇表大小。与完整模型不同，这个仅编码器的基线缺乏词间依赖关系建模和语言生成能力。

如表 3 所示，详情见表 4，编码器-only 模型的表现明显劣于完整的编码器-解码器模型，特别是在较高阶的 BLEU 分数上。这表明了解码器在捕捉 EEG 到文本生成中的序列和语言依赖关系的关键作用。

表 3: 编码器仅模型的 BLEU 分数（消融研究）。

模型	**BLEU-1**	**BLEU-2**	**BLEU-3**	**BLEU-4**
EncoderOnly	0.0674	0.0199	0.0139	0.0120

表 4: 真实值与预测结果的对比（使用 BLEU 分数的仅编码器模型）

地面真相（GT）	**预测结果（PR）**	**BLEU-1**	**BLEU-2**	**BLEU-3**	**BLEU-4**
1	的	0.0000	0.0000	0.0000	0.0000
全世界的狼都有一个共	的的的的的的的的的的	0.1000	0.0333	0.0240	0.0211
同的习性，在严寒的冬天	的的的的的的的的的的	0.1810	0.0427	0.0274	0.0227
集合成群，平时单身独处。	的的的的的的的的的的	0.0000	0.0000	0.0000	0.0000
眼下正是桃红柳绿的春	的的的的的的的的的的	0.1000	0.0333	0.0240	0.0211
天，日曲卡雪山的狼群按	的的的的的的的的的的	0.0905	0.0302	0.0218	0.0191
自然属性解体了，	的的的的的的的	0.0000	0.0000	0.0000	0.0000

6 结论

在此工作中，我们介绍了 EEG2TEXT-CN，首个开放词汇的脑电波到中文生成模型。基于多视角脑电编码器和通过对比学习对齐预训练语言模型构建，我们的系统在零样本条件下运行，并能从原始脑电数据成功预测完整的中文句子，无需访问视觉刺激或特定任务微调。

该模型在中国 EEG 数据集上进行了训练和评估，其中每个 EEG 序列对应一个字幕级别的中文句子。尽管绝对的 BLEU 分数仍然很低，但模型展示了将部分词汇内容与神经信号对齐的能力，并标志着非语音 EEG 到语言研究的重要进展。

未来的研究方向包括：(1) 改进长范围上下文建模和记忆；(2) 结合语义感知的评估指标和人类判断；以及 (3) 通过增加更多参与者和更长文本来扩展数据集。EEG2TEXT-CN 建立了多语言神经解码的有前景的基础，并为在较少探索的语言领域中实现脑电到语言生成提供了新路径。

参考文献

[1] Hanwen Liu, Daniel Hajialigol, Benny Antony, Aiguo Han, and Xuan Wang. Eeg2text: Open vocabulary eeg-to-text decoding with eeg pre-training and multi-view transformer. *arXiv preprint arXiv:2405.02165*, 2024.

[2] Yiqun Duan, Jinzhao Zhou, Zhen Wang, Yu-Kai Wang, and Chin-teng Lin. Dewave: Discrete encoding of eeg waves for eeg to text translation. *Advances in Neural Information Processing Systems*, 36:9907–9918, 2023.

[3] Zhenhailong Wang and Heng Ji. Open vocabulary electroencephalography-to-text decoding and zero-shot sentiment classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 5350–5358, 2022.

[4] Yonghao Song, Bingchuan Liu, Xiang Li, Nanlin Shi, Yijun Wang, and Xiaorong Gao. Decoding natural images from eeg for object recognition. *arXiv preprint arXiv:2308.13234*, 2023.

[5] Xinyu Mou, Cuilin He, Liwei Tan, Junjie Yu, Huadong Liang, Jianyu Zhang, Yan Tian, Yu-Fang Yang, Ting Xu, Qing Wang, et al. Chineseeeg: A chinese linguistic corpora eeg dataset for semantic alignment and neural decoding. *Scientific Data*, 11(1):550, 2024.

[6] Nils Reimers and Iryna Gurevych. Sentence-bert: Sentence embeddings using siamese bert-networks. `https://huggingface.co/sentence-transformers/all-MiniLM-L12-v2`, 2021. Accessed: 2024-03-01.

[7] Chi-Sheng Chen and Chun-Shu Wei. Mind's eye: Image recognition by eeg via multimodal similarity-keeping contrastive learning. *arXiv preprint arXiv:2406.16910*, 2024.

[8] Chi-Sheng Chen, Aidan Hung-Wen Tsai, and Sheng-Chieh Huang. Quantum multimodal contrastive learning framework. *arXiv preprint arXiv:2408.13919*, 2024.

[9] Chi-Sheng Chen. Necomimi: Neural-cognitive multimodal eeg-informed image generation with diffusion models. *arXiv preprint arXiv:2410.00712*, 2024.

[10] Dongyang Li, Chen Wei, Shiying Li, Jiachen Zou, Haoyang Qin, and Quanying Liu. Visual decoding and reconstruction via eeg embeddings with guided diffusion. *arXiv preprint arXiv:2403.07721*, 2024.

[11] Pengfei Wang, Huanran Zheng, Silong Dai, Yiqiao Wang, Xiaotian Gu, Yuanbin Wu, and Xiaoling Wang. A survey of spatio-temporal eeg data analysis: from models to applications. *arXiv preprint arXiv:2410.08224*, 2024.

[12] Wenjing Xiong, Lin Ma, and Haifeng Li. Synthesizing intelligible utterances from eeg of imagined speech. *Frontiers in Neuroscience*, 19:1565848, 2025.

[13] Jinzhao Zhou, Yiqun Duan, Yu-Cheng Chang, Yu-Kai Wang, and Chin-Teng Lin. Belt: Bootstrapped eeg-to-language training by natural language supervision. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2024.

[14] Jinzhao Zhou, Yiqun Duan, Fred Chang, Thomas Do, Yu-Kai Wang, and Chin-Teng Lin. Belt-2: Bootstrapping eeg-to-language representation alignment for multi-task brain decoding. *arXiv preprint arXiv:2409.00121*, 2024.

[15] Nora Hollenstein, Jonathan Rotsztejn, Marius Troendle, Andreas Pedroni, Ce Zhang, and Nicolas Langer. Zuco, a simultaneous eeg and eye-tracking resource for natural sentence reading. *Scientific data*, 5(1):1–13, 2018.

[16] Young-Eun Lee, Seo-Hyun Lee, Sang-Ho Kim, and Seong-Whan Lee. Towards voice reconstruction from eeg during imagined speech. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 6030–6038, 2023.

[17] John LaRocco, Qudsia Tahmina, Sam Lecian, Jason Moore, Cole Helbig, and Surya Gupta. Evaluation of an english language phoneme-based imagined speech brain computer interface with low-cost electroencephalography. *Frontiers in neuroinformatics*, 17:1306277, 2023.

[18] Eran Elhaik. Principal component analyses (pca)-based findings in population genetic studies are highly biased and must be reevaluated. *Scientific reports*, 12(1):14683, 2022.

[19] Ronald J Williams and David Zipser. A learning algorithm for continually running fully recurrent neural networks. *Neural computation*, 1(2):270–280, 1989.

[20] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, pages 311–318, 2002.