arXiv:2611.32825v1 [cs.CV] 24 Nov 2026

EEG-World：观看中的大脑的上下文条件动态

Junxiang Zhang[1,5,*] JiayuZuo[1,*] Wenhao Ye[2] Junfeng Xia[1]Xiang Chen[3] MoWang[4,†] Quanying Liu[1,5,†]

[1]南方科技大学，深圳，中国 [2]深圳大学，深圳，中国[3]Omni Intelligence，深圳，中国 [4]华威大学，考文垂，英国[5]深圳环路区域研究院，深圳，中国

摘要

脑编码模型可根据视听输入预测神经反应，但通常忽略持续脑活动如何塑造个体的反应。我们提出***EEG-World***，这是一种循环世界模型，可利用前*6*秒的脑电图以及随后呈现的视听刺激，预测未来*10*秒的脑电频带功率动态。*EEG-World*在连续脑电图上进行预训练，学习共享的神经动力学，并通过轻量级的受试者特异性读出层适应未见过的个体。通过使用来自*78*名参与者的*2,584*小时脑电图数据，我们表明，大规模非线性循环预测优于强大的历史感知线性基线。*EEG-World*的预测相关性达到*0.261*，而最强的历史与视频线性模型为*0.233*，其预测时域优势为*3.3*秒。学习到的动态特性可迁移至未见过的参与者；经过校准后，在*24*名个体中的*23*名上均优于线性基线。尽管持续脑电图历史数据解释了大部分预测性能，但受控刺激分析揭示了额外的视频依赖成分。该成分还支持对视觉表征进行评估：与*CLIP*和*DINOv2*特征结合相比，*V-JEPA 2*使片段锁定预测提升了*35%*。我们的结果证明了可迁移神经动力学对于个体脑电预测的价值，并强调了区分状态驱动与刺激驱动可预测性的重要性。

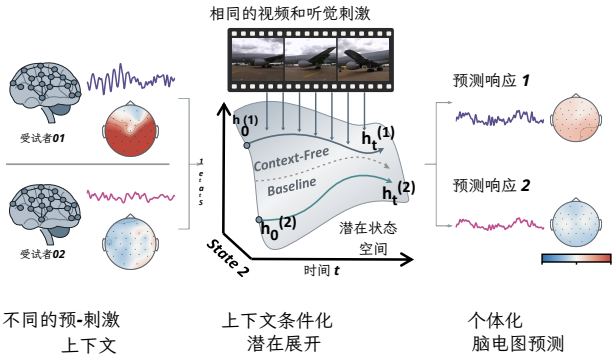


图1。**同一视频片段，不同观看者**。不同受试者在不同的大脑状态下开始观看同一视频片段（左）。EEG-World 根据视频片段向前推演每位观看者的状态，并预测该观看者的自身反应（右）；无上下文编码模型则为每个人预测相同的轨迹（虚线）。轨迹根据预测带的频带功率绘制。

1. 引言

自然情境视频观看期间的大脑记录已成为视觉模型的重要测试平台：能够预测神经活动的表征被视为与人类视觉加工相一致的证据 [22, 41, 49]。脑电图使这类测量更易开展，并能以时间分辨的方式观察大脑活动。然而，预测某种刺激引发的神经反应，并不等同于预测特定观看者在观看过程中脑活动将如何演变。即使观看同一段视频，不同观看者也可能沿着不同的神经轨迹发展，因为他们起始时处于不同的脑状态（图1）。因此，要预测这些轨迹，不仅需要对屏幕上的内容建模，还需要对大脑持续进行的动态变化建模。

传统的刺激到大脑编码模型根据视觉或视听特征预测神经反应 [23,26, 34, 35]。尽管这类模型能够有效刻画与刺激相关的反应，但通常不会以个体此前的神经活动为条件。对于脑电图而言，这一遗漏尤为重要，因为持续节律和时间上的持续性使得近期活动能够高度预测接下来几秒内的活动。因此，模型可能能够准确预测脑电图，却几乎没有学到有关刺激本身的信息。这

*贡献相同。
†共同通讯作者：12250099@mail.sustech.edu.cn, liuqy@sustech.edu.cn

引出了两个问题：非线性模型能否捕捉到超出强大的、考虑历史数据的线性模型所能解释范围之外的神经动态特性？以及，如何将这种由状态驱动的可预测性与视频所带来的贡献区分开来？

我们从世界模型的视角来处理这一问题：世界模型预测系统在外部输入作用下如何从当前状态演化 [15, 16, 18]。在我们的设定中，大脑是动力系统，脑电图提供其状态的部分观测，而视频则充当外生输入。我们提出 EEG-World，从先前的脑电图中推断潜在脑状态，并在视听条件化下将其向前推进，以预测个体神经活动。循环动力学核心在连续脑电图上进行预训练，轻量级的受试者特异性读出层则将共享动力学适配到新个体。由于仅凭神经活动历史便能解释未来活动的很大一部分，我们将其与接收相同 EEG历史和视频的强大线性预测器进行比较，而不是与仅基于刺激的编码模型比较。

学习和评估此类动态特性需要现有脑电图资源很少能同时具备的数据：用于学习时间规律的大规模记录、向许多观看者呈现的相同视频以区分刺激驱动活动和状态驱动活动，以及足够多的参与者，以便留出部分参与者测试迁移能力。为满足这些要求，我们收集了 VAEEG 数据集，其中包含 78 名参与者观看自然视听片段时的 2,584 小时脑电图。在该数据集上，当合并数百小时的数据后，非线性循环预测的表现超过强大的线性基线方法；随着线性模型趋于饱和，其性能仍在持续提升。其动力学可迁移至未见参与者：经过 33分钟的校准，即可达到使用多 10 倍数据拟合的线性模型的水平。尽管持续脑电图动力学主导了预测性能，但受控刺激分析揭示了额外的视频依赖成分；该成分经过计算机模拟调谐后与记录的脑电相匹配，从而能够更有信息量地比较视觉表征与人类神经反应。

贡献。
• 我们提出 EEG-World，这是一种用于个体脑电图预测的循环世界模型，结合了预训练神经动态特性、视听条件化以及轻量级的受试者特异性适应。

• 我们使用来自 78 名参与者的 2,584 小时自然场景脑电图数据，证明非线性循环预测在大规模条件下超越了强大的历史感知线性基线，且学习到的动态特性能够迁移到未见参与者。

•我们建立了历史感知基线和持久性控制评估，将神经状态的可预测性与刺激驱动反应区分开来，从而能够更有针对性地比较视觉表征与大脑。

2. 相关工作

脑编码模型。 编码模型根据刺激特征预测功能性磁共振成像、脑磁图或脑电图反应，通常使用正则化线性映射 [2, 23, 26, 34, 35], , 而视觉模型的特征对大脑的预测效果则用于将其与人类视觉进行比较 [41, 49]。近期的挑战和模型利用多模态特征预测观看电影时的全脑功能性磁共振成像反应 [8, 13]；对于脑电图和视频，大多数研究都是从反应中解码刺激 [30] ，或将反应与视觉模型的表征进行比较 [40]。所有这些方法都将反应视为仅由刺激决定的函数。它们无法说明特定大脑接下来会做什么，因为塑造随后几秒钟的持续状态（唤醒、注意、内源性节律）会因观看者和时刻而异，却被当作噪声处理。

脑世界模型。 世界模型学习环境如何在智能体动作作用下演化 [4, 15–18, 33]；在其通常的设置中，观测信息丰富，而动作强烈决定接下来发生的事情。对于大脑而言，潜在动力学模型从神经元放电活动中推断神经状态 [38, 50], , NeuroWorld [9] 则以刺激作为动作，从单个观测到的体积数据开始，对功能性磁共振成像进行展开预测。这两者都无法直接迁移到脑电图。首先，头皮信号是主要由持续性活动驱动的噪声混合信号，因此必须根据数秒的历史数据而非单个快照推断状态，而视频相对于强大的内部动力学而言只是较弱的输入。其次，持续振荡的相位在数秒后是不可预测的，因此像像素世界模型重建帧那样重建原始信号并不适定；频带功率则是可预测的。第三，不同受试者的通道和频带尺度存在差异，因此单一读出层无法适用于所有人，而新受试者只能提供几分钟的数据。第四，缓慢且具有持续性的状态会使复制起始点获得较高收益，因此必须相对于具有相同历史数据的基线来衡量增益；NeuroWorld 唯一的线性基线仅使用刺激，其识别指标也未明确候选池，这些选择会如我们所示夸大增益（第 5 节和第 6.5节）。脑电图基础模型 [10, 24, 47]和功能性磁共振成像基础模型 [48]主要通过分类而非预测进行评估，并且脑电表征与视频与语言模型的对齐程度仅略高于时间平移控制条件 [46]。EEG-World 正是基于这些特性构建的。它从数秒的脑电图中推断潜在状态，预测频带功率，将视频作为零初始化的动作加入在连续脑电图上预训练的动态模型中，并通过共享线性头和受试者特异性线性头进行读出。

3. VAEEG数据集

语料库。 我们收集了 VAEEG 数据：受试者观看屏幕上的短视频片段时录制的 2,584 小时脑电图数据，来自 78 名受试者，共有 618,106 次可用呈现（表 1）。这是



图 2. **与在视听刺激期间录制的公开脑电图数据集进行对比。**屏幕上显示刺激时的脑电图小时数，按受试者汇总；气泡面积表示每位受试者的小时数。

在视听刺激期间录制的最大公开脑电数据集的 4.8 倍 [43], , 且按受试者计算，每人 33 小时的录制时长约为其他数据集的五倍 [5, 12, 14, 27, 30, 52]（图 2）。这些时长中，2,122 小时是在四个公开视频数据集上录制的；其余时长主要来自较早的录制阶段，录制内容为其他短视频。四个视频数据集、将 AVE 片段划分为 3,312 个训练片段、391 个验证片段和 394 个测试片段的划分（辅助文件按公开片段标识符列出了这些片段），以及所有预处理、模型和评估设置，均已公开或完整说明。

采集范式。脑电图通过一个32通道湿电极脑电帽（BrainCo）进行录制，并通过局域网传输至呈现计算机。该脑电帽在标准10–20系统和10–10系统位置设有31个头皮电极（图7；最早的记录中为30个），以及一个眼电图通道（IO），我们将其排除。参考电极位于FCz，接地（偏置）电极位于AFz。记录按每轮约30分钟组织，中间安排由受试者自行决定的休息时间；受试者按照采集中心工作人员的指示进行观看。公开片段来自四个视频数据集，即AVE（包含28个类别的10秒视听事件片段）[44], MUSIC-AVQA [29], 感知测试 [39] 以及EEG Moments 数据集（EMD；带声音的3秒自然情境视频）[14]。片段大多按照固定顺序一个接一个地呈现，因此许多受试者会看到每个片段（每个AVE测试片段的观看者中位数为33人）。AVE片段每12秒开始一次（10秒片段，2秒空白）。脑电图以250 Hz采样，经过带通滤波（0.1–96 Hz）和50/60 Hz陷波滤波器处理后，重采样至256 Hz；我们保留FCz参考。

时序。 片段起始时刻记录在演示计算机的时钟上，而脑电图记录则使用设备自身的时钟，该时钟的漂移率为

表 1. **VAEEG 及所用子集。** †仅训练会话。‡包含未见受试者的训练数据；第 6.3 节中的模型从未接触过这些数据。测试数据不会用于训练。

子集	受试者	小时	Use
完整语料库	78	2,584	–
AVE 呈现	70	1,016	视频配对试验
连续池	70‡	2,012	预训练、扩展
核心受试者	15	908†	扩展、视频、ID
未见受试者	24	516	单次迁移
视频微调	65‡	630	226,979 次片段观看

中位数为 −980百万分率；若不进行校正，在一次录制结束时，起始时刻的偏差中位数将达到 1.5 秒，最高可达 10 秒。对于每条记录，我们估计能够最大化片段起始时刻和结束时刻诱发反应的时钟偏移和漂移率（在奇数片段上拟合，在偶数片段上测试），将每个起始时刻转换到脑电图时钟，并排除拟合未通过质量标准的记录（验证集见第 B节）。数据将剩余误差限制在我们 1 秒的分辨率以内。后验 α 在片段之间持续 2 秒的空白期间上升，并在下一个起始时刻下降；在 1,820 条记录中的 1,765 条（97%）中，这一反应都出现在预期的秒数内。另外 55 条记录（试次的 3%）偏移了 1–4 秒；所有模型看到的都是相同的数据。

特征与数据划分。 每秒表示为一个包含 155 维的向量，其内容为根据 1 秒多锥形窗频谱计算的对数频带功率，涵盖 31 个通道 × 5 频带（δ 1–4、 θ 4–8、α 8–13、 β 13–30、γ 30–45 Hz），并按每段录制进行 Z 分数标准化。若任一通道的峰峰值振幅超过 500 μV，则将该秒标记为不良；包含不良秒的时间窗将被跳过。核心组受试者的连续录制按整段录制划分（75% 用于训练、10% 用于选择、15% 用于测试）。试次采用语料库的刺激划分，每个试次包含片段起始前 6 秒以及片段的 10 秒内容。所有验证集或测试集片段周围的脑电图数据均从所有训练数据中屏蔽。每个片段的每一秒由 24 个主成分描述，涵盖 OpenCLIP [6]和 DINOv2-L [37]的帧嵌入、响度、谱通量以及片段开启标志。

4. EEG-World

任务。 令 $x_t \in \mathbb{R}^D$ 表示时间步t 的脑电图特征（每个通道和频带的频带功率；见第3节），令 a_t 表示屏幕上片段的特征（起始前的$a_t = 0$）。给定上下文 $x_{1:T_c}$ 和动作 $a_{1:T}$，其中$T = T_c + T_f$，模型在一次自由滚动预测中预测 $\hat{x}_{T_c+1:\,T}$。对于不含视频的连续脑电图，在长度为 T的每个窗口上执行相同的任务。

动力学核心（图 3）。编码器 e 将每个观测步骤映射到 $\mathbb{R}^H$，而门控循环单元 [7] 更新潜在脑状态

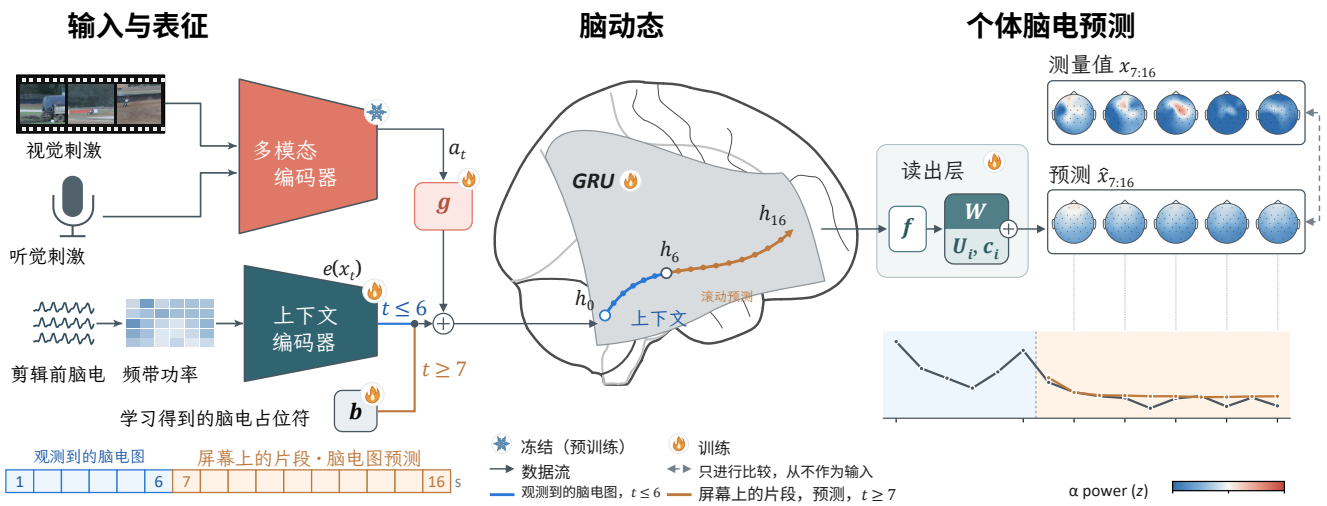


图3。 **EEG-World**。 每秒执行一次门控循环单元步骤。在上下文中，编码后的频带功率 $e(x_t)$ 会更新潜在状态；在滚动预测中，学习得到的空白向量 b 会取而代之（公式（1））。片段通过视频通路 g 作用于冻结特征，共享读出层 W 加上被试专属输出头 $(U_i,\ c_i)$ 给出预测。右侧为选取用于说明的一次真实测试观看，其技能位于第99.8百分位。

状态 $h_t \in \mathbb{R}^H$。在滚动预测过程中不会观测到脑电图，因此编码器输出会被一个学习得到的空白向量 b替代：

$$h_t = \mathrm{GRU}\big(z_t + g(a_t),\, h_{t-1}\big), \qquad z_t = \begin{cases} e(x_t), & t \le T_c, \\ b, & t > T_c. \end{cases} \tag{1}$$

主干网络 f 和共享读出层 W 给出预测，再加上逐受试者的线性头 i，

$$\hat{x}_t = W f(h_t) + U_i f(h_t) + c_i, \tag{2}$$

其中 $U_i,\ c_i$ 初始化为零。预测头吸收受试者特定的通道和频带尺度，因此动力学核心由所有受试者共享。

视频通路。 视频片段的每一秒都由冻结特征描述，即 OpenCLIP 和 DINOv2 帧嵌入、响度及谱通量（第 3 节）；其他编码器将在第 6.4 节中进行比较。视频通路 g 是一个多层感知机，其最后一层初始化为零，因此预训练模型在微调开始时保持不变。它的输出会在每一步加到门控循环单元的输入上，因此视频片段充当控制输入，使潜在状态发生变化。

训练。 阶段1使用自由滚动预测的均方误差，在连续脑电图上预训练核心组、读出层和预测头（AdamW [32], 根据验证集选择误差进行早停）。每个训练轮次的步数与数据量成正比，因此模型在不同规模下看到数据的次数相同。阶段2加入视频通路，并在视频试次上进行微调；每个批次中一半样本取自连续时间窗，以保持预训练动力学；预训练参数使用比新参数更小的学习率。超参数见第A节。

未见受试者。 对于训练中未见过的受试者，训练好的模型保持冻结。由于 $f(h_t)$ 在给定上下文和视频片段后即被固定，方程（2）中的新头部 $(U_i,\ c_i)$ 是基于该受试者自身训练试次进行的岭回归，并通过其自身验证试次选择惩罚项，使其向共享读出层收缩。

线性基线。每项结果都与每位受试者可用的最强线性模型进行比较。历史岭回归将 $T_c \times D$ 上下文直接映射到 $T_f \times D$ 未来，可按受试者分别拟合、在受试者间合并拟合，或采用分层方式拟合，并将受试者级权重向合并权重收缩；这相当于共享核心加预测头的线性对应形式。加入一组片段特征后，便得到历史数据 +视频编码模型。对每位受试者而言，在其自身验证集数据上表现最佳的三种模型中，最优者即为最佳线性模型。我们还使用 EEG-World 的架构和训练方法训练了线性世界模型，但移除了所有非线性。

无状态非线性基线。 为了探究这种增益是否需要循环状态，还是仅需非线性网络，我们采用两个直接预测器：多层感知机和 iTransformer [31]，后者将每个特征的历史数据嵌入为一个词元；二者都使用 EEG-World 的输入、被试专属输出头和步数预算，将上下文以及试次中的片段特征一次性映射为全部 T_f 步的预测。

5. 评估方案

我们的基准测试包含五项任务，即多步预测、通过片段交换和片段锁定r利用视频、片段识别、向未见受试者迁移以及数据扩展，并固定了三个会决定结论的选择。

使用 EEG历史的基线方法。 每项任务都会报告上下文均值、仅历史岭回归和历史数据 + 视频岭回归，分别采用按受试者、合并和分层方式；在线性模型上，大脑数据 [20, 42] 和时间序列 [51]方面很难超越线性模型。仅刺激线性模型，即通常的编码基线 [9, 13], 在我们的试次中，每个预测步的相关性仅达到 0.04，而加入 EEG历史后则达到 0.23；与其比较会将任何预测模型的增益夸大数倍。在试次中，分层岭回归比按受试者岭回归高出 0.008 的技能得分，因此仅与按受试者岭回归比较会使世界模型的表观增益翻倍。

评估指标。 主要评估指标是预测相关性 r，与脑编码 [9, 13]中的定义相同。它是每个预测步和每个特征在测试窗口上的 Pearson r 相关性，再对特征和预测步取平均。由于 r 忽略了幅度，我们还报告技能，即相对于上下文均值预测 $\bar{x}^{\text{ctx}}$的平方误差，

$$\text{skill} = 1 - \frac{\sum_{t=T_c+1}^{T}\sum_{d=1}^{D}(\hat{x}_{t,d} - x_{t,d})^2}{\sum_{t=T_c+1}^{T}\sum_{d=1}^{D}(\bar{x}_d^{\text{ctx}} - x_{t,d})^2}, \qquad (3)$$

其奖励经过校准的幅度。预注册的决策采用技能；在任一指标下达到显著的每项比较，在两项指标下的方向都一致，但零样本迁移除外（第6.3节）。为使差异更直观，我们还将其表示为时间。The 预测时域增益 表示基线模型的逐步 r 下降到模型在最后一步达到的水平时所提前的秒数（步之间采用线性插值，按受试者计算）。试次结果先在每个测试片段内取平均，再跨片段取平均，因此频繁出现的片段不会占据主导。

持久性控制识别。 为考察预测是否知道正在观看的内容，我们固定真实上下文，并针对每个测试片段各进行一次预测；真实片段应产生与记录的脑电最接近的预测。忽略视频的模型得分恰好等于机会水平。我们还根据可比研究的评估方案报告结果 [9] （第6.5节）。

统计与预注册。 推断单位是受试者。我们在 15 名核心受试者上进行精确的符号翻转检验，在另外 24 名未见过的受试者上进行 100,000 次随机符号翻转；对于试次，还在 394 个测试片段上进行符号翻转检验。

每个假设及其决策规则（平均差异 > 0，$p < 0.05$，至少 11/15 或 16/24 名受试者呈阳性）以及每次测试集的使用，都在运行前记录下来 [36]；第二次尝试使用 $p < 0.025$。所有这些，包括失败的尝试，都列在第K节中。

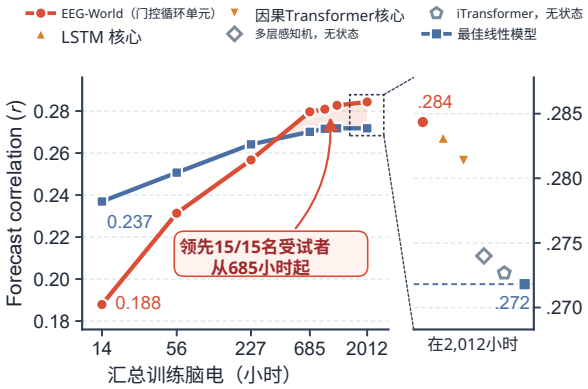


图4. **连续脑电图上的扩展。** 在 15 名核心受试者的留出记录上，预测相关性与汇总训练脑电进行比较；阴影部分表示 EEG-World 的领先优势。右侧为 2,012 小时数据下的所有核心模型及无状态预测器。

6. 实验

我们使用 $T_c = 6$，$T_f = 10$ 步长为 1 秒，$D = 155$ 以及$H = 512$；其余超参数见第A节。对于连续脑电图，所有结果均基于 15 名核心受试者的留出记录；对于试次，则基于他们对 394 个测试片段的 10,970 次观看。下文中的每项比较均按受试者进行测试；区间、受试者数量和 p 值，以及哪些测试已预注册，列于第J 节和第K节。

6.1. 连续脑电图预测

与规模化研究一样， [21, 25], 我们在嵌套的连续脑电图数据池上进行不使用视频的预训练，并在每个规模下，将其与使用相同数据拟合的最佳线性历史模型进行比较（图4）。随着规模扩大，排名发生逆转。在几十个小时的数据规模下，线性模型明显更优；当合并数据达到 227 至 685 小时之间时，世界模型超过线性模型，此后在每名受试者上均胜出。因此，只有在跨受试者合并数百小时的数据后，脑电图动态特性的非线性部分才变得可学习；这也解释了为什么在线性预测器通常难以在每名受试者仅有几小时数据的典型脑电数据集上取得超越。

线性模型趋于饱和；世界模型则不会。超过交叉点后，两个模型开始分化。最佳线性模型在几百小时后趋于平稳，因此更多数据并不能带来任何线性模型可从近期过去中利用的新信息，而世界模型一直改进到 2,012 小时。后期改进约有一半来自更大的训练预算，另一半来自额外的数据（见第F节）。

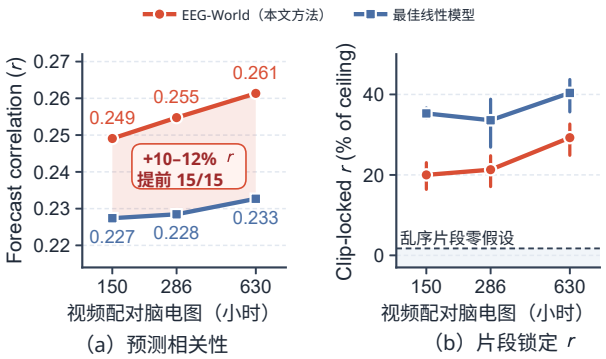


图5。**视频试次**。15 名核心受试者的留出片段与视频配对训练脑电图数量的关系。（a）预测相关性。（b）片段锁定相关（SLv）：每个预测结果中由片段驱动的部分与其他观看者的平均脑电图之间的相关性，以噪声上限的百分比表示；阴影区域表示低于 +2SD 的置换片段零分布。

表2。**视频试次**。片段起始模型在相同试次上进行微调，但仅获知每个片段何时开始；无状态模型是第 4 节中的直接预测器；交换成本是将片段替换为另一个测试片段时损失的相关性。

模型（15名受试者，测试片段）	r	技能	交换成本（r）
上下文均值	.180	0	–
线性，仅脑电图	.225	.1909	–
线性，脑电图 + 视频	.233	.1928	.012
多层感知机，无状态	.242	.1948	–
iTransformer ，无状态	.248	.1982	–
EEG-World，脑电图 + 片段起始	.255	**.2034**	–
EEG-World ，脑电图 +视频	**.261**	.2030	**.026**

关键在于一种贯穿时间传递的状态。移除所有非线性的相同架构表现低于闭式岭回归，因此这种增益并非训练造成的假象。一个 MLP 和一个 iTransformer [31] 在使用相同的预测头和预算，将相同的上下文直接映射到未来时，相比线性模型几乎没有提升，因此仅有非线性还不够。LSTM 和因果 Transformer 核心与 GRU 一样，都会逐步展开一个潜在序列，其增益几乎与 GRU 相当（见图4，右）。因此，线性模型所遗漏的是大脑状态如何演化，而不是从过去映射到未来的更丰富方式。GRU 既是最佳核心，也是最小的核心，因此在这一规模下，紧凑的循环状态比容量更重要。

6.2. 将视频作为动作进行预测

在留出片段上，EEG-World 的表现优于最强的历史 + 视频编码模型 12%，达到 r （表2）。这一优势在第一秒较小，并随预测时域增长。若要在最后一步达到与 EEG-World 相同的准确度，线性模型必须提前 3.3 秒停止。上下文之后的第一秒—主要由紧接着的过去决定，任何预测器都可以复制这一点；世界模型的价值在于将状态进一步带到未来。无状态非线性预测器在试次上的帮助大于在连续脑电图上的帮助，但其表现仍远低于 EEG-World，因此循环状态对于视频的重要性与无视频时一样。

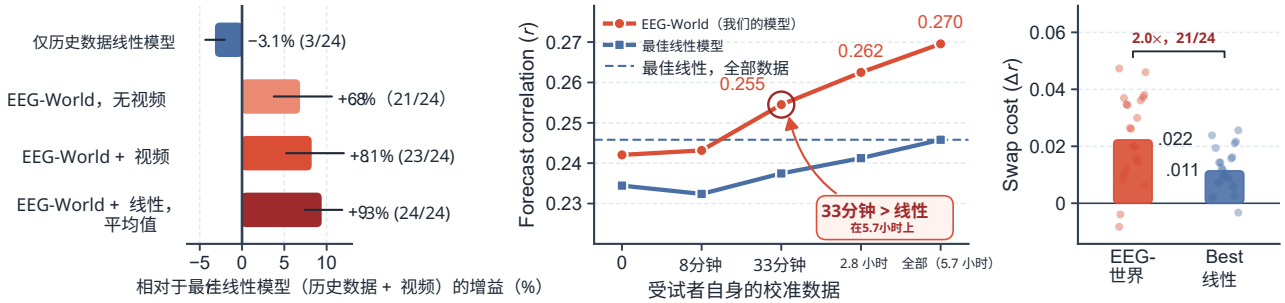
个体轨迹与平均响应。更多视频配对数据对世界模型的帮助超过对线性模型的帮助，其片段锁定响应也随之增强（图5）。然而，线性编码模型更贴近观看者的平均响应（图 5b），而世界模型对每位观看者的预测更准确（图 5a）。这两个目标并不相同，因为对观看者求平均会消除世界模型所预测的信号中依赖状态的部分。因此，仅根据平均响应对模型评分的基准测试会偏向编码模型，也无法体现预测个体大脑活动的模型所带来的增益。

6.3. 迁移至未见受试者

我们对24名在开发过程中从未使用过的受试者各测试了一次，并预先固定了分析方案；这些受试者和片段对模型而言都是全新的。在冻结视频模型、仅为每名受试者拟合一个线性头的情况下，EEG-World在24名受试者中的23名上击败了最佳历史 + 视频线性模型（图6a）；在其余41名受试者上重新训练的模型复现了这一结果（图 6b）。

能够迁移的是动态特性。一个从未看过视频的世界模型，其迁移效果几乎与视频模型一样好，尽管视频模型确实使用了片段（图 6c）。对于新的大脑，主要能够迁移的是从连续脑电图中学习到的动态特性；在训练队列上学习到的刺激映射对新受试者的作用很小。对世界模型和线性模型的预测取平均，其效果优于两者，因此二者产生的错误部分独立。在另外两种脑电设备上，冻结模型仅配备新头部时，其表现只能与线性模型持平（第I节），因此动态特性可以在同一记录设置内跨个体迁移，但目前尚不能跨设备迁移。

校准。对于预测头而言，仅使用新受试者的**33分钟**数据，重新训练的模型就达到了 $r = 0.255$，高于0。246 的线性模型；该模型使用该受试者的全部数据拟合，耗时中位数为**5.7小时**，约为前者的十倍（图6b）。即使不进行校准，它的相关性也已经优于合并线性模型，但平方误差并未更低，因为共享读出层能正确捕捉轨迹的形状，却未能正确捕捉其尺度。人与人之间的差异主要体现在通道和频带尺度上，线性头能够吸收这类差异；而动态特性的差异小得多，共享核心能够捕捉这一点。



(a) 在24名未见受试者上的增益　(b) 未见受试者的校准　(c) 片段交换成本

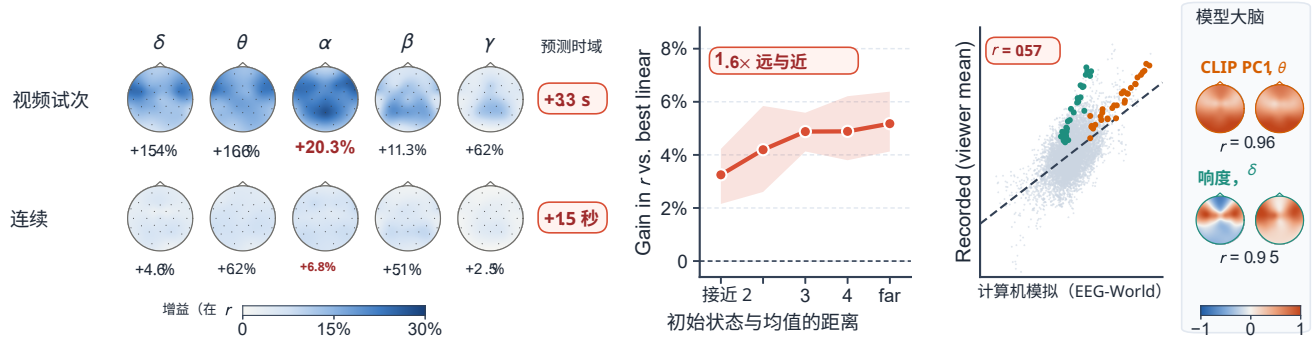
图6. **未见受试者**。 24名从未参与开发的受试者；冻结模型搭配逐受试者线性头。(a) 相较于最佳历史数据 + 视频线性模型的增益，以其 r 的百分比表示，受试者按增益排序。(b) 不使用这些受试者重新训练的模型，与每名受试者自身的校准数据进行比较；虚线表示基于全部校准数据拟合的最佳线性模型。(c) 更换片段后损失的相关性；圆点表示受试者。

6.4. 世界模型学到了什么

增益遍及整个头皮，在 α 处最强。在连续脑电图和视频试次中，所有电极和频带上的增益均为正（图7a）。如果模型学到的是局部反应，例如视觉输入引发的枕叶 α，就会呈现局部图谱；增益遍布整个头皮，则表明这是一种全局脑状态。在视频试次中，增益在 α (+20%)处最大；这种节律与视觉注意和唤醒的关联最为密切，其增益约为连续脑电图中的三倍。

增益在远离大脑通常状态的位置最大。线性预测器只能让每个特征以固定速率向其均值弛豫。如果大脑的回归取决于初始位置，那么当初始状态不寻常时，非线性模型应能获得最大的增益。事实上，增益随着初始状态与受试者均值之间的距离增加而稳步增长，最远五分位的初始状态获得的增益 1.6× 与最近五分位一样多 r (+5.2% 相比于 +3.2%；图 7b），并且 3.8× 也同样具备相当的技能。其优势集中在最不适合线性弛豫的区域，即远离平衡态的状态。预测结果平滑，但并未向均值收缩（第 G节）。

计算机模拟调谐与记录的脑电图相匹配。 我们将 EEG-World 用作虚拟受试者。在保留真实上下文的情况下，我们提高整段视频中的某一项视频或音频特征，并记录预测的变化。在留出验证视频片段上，我们根据每个片段所有观看者的平均响应，测量记录的脑电图中的相同调谐。两者在不同特征和通道之间均一致（图 7c）；对于最强的特征——第一项 CLIP 分量，它能区分拨弦乐器与车辆——其 θ 频带图几乎完全重合。该模型仅接受过预测单次观看的训练，从未见过平均响应，因此这种一致性并非模型预先具备的。对单个观看者进行预测，迫使视频通路学习刺激—响应关系



(a) 各电极和频带相较于最佳线性模型的增益　(b) 增益与初始状态的关系　(c) 计算机模拟与实测调谐

图7. **何处、何时、何种**。 (a) 相较于各电极和频带上的最佳线性模型的增益，以其 r 的百分比表示，以及预测时域增益。(b) 增益（百分比）与起始状态偏离受试者均值的距离之间的关系。(c) 计算机模拟调谐：提高一个输入特征时预测值的变化，与各观看者记录的脑电图的调谐情况进行比较；比较涵盖每个特征和电极—频带单元；右侧为两个特征的映射。面板(c)仅使用验证数据。



图8。**根据预测识别片段**。每次观看均保留其真实上下文，并以394个测试片段中的每一个作为动作进行预测；根据与记录的脑电图的拟合度为候选片段排名。（a）一个测试片段示例，以一帧画面及其事件标签呈现。（b）真实片段排名居中的那次观看。（c）对该常见重复片段的全部183次观看取平均。（d）汇总随机抽取的观看子集，涵盖全部测试片段。事后分析。

表3。**片段特征排名**。视频通路每次输入一组特征，每组含24个主成分；使用三个随机种子。SLv定义同图。5b，未进行噪声上限归一化。

片段特征	模态	SLv	r
V-JEPA 2 [3]	视频	**.188**	**.261**
CLIP +DINOv2（主要视觉输入）	图像	.139	.260
CLIP [6]	图像	.133	**.261**
DFN [11]	图像	.126	.255
VideoMAE [45]	视频	.115	.255
MAE [19]	图像	.110	.255
DINOv2 [37]	图像	.109	.253
对数梅尔频谱图	音频	.062	.249
响度 + 谱通量	音频	.061	.259
仅片段起始	–	0	.253

而这种关系正是群体平均所揭示的。

视觉编码器排名。 由于片段仅通过其特征输入 EEG-World，因此更换编码器就能将模型变成对视觉模型的测试（表 3）。不同编码器的预测相关性几乎没有变化，因为它主要由大脑自身动力学决定；而片段锁定响应分离出了预测中由片段驱动的部分，因而清晰地呈现出它们之间的差异。V-JEPA 2 [3], 经过训练，能够在其潜在空间中预测视频的遮蔽部分，使这一响应表现最佳，比我们的 CLIP + DINOv2 输入高出 35%。仅凭时间输入无法解释这一点，因为 VideoMAE 虽然也是视频模型，却经过训练以重建像素，其排名与图像模型相当。因此，在潜在空间中预测视频这一目标似乎比感知运动更重要。主要结果仍采用预注册的 CLIP +DINOv2 输入。

6.5. 刺激识别

在 NeuroWorld 的轨迹匹配协议下， [9],EEG-World 在受试者的测试观看中找出正确试次的表现远高于机会，并且比线性模型略为频繁。但一种忽略视频、重复上下文均值的“模型”已经达到 12.5×机会水平，而将候选池扩大到所有观看次数后，相同的预测从 16× 变为 145×机会水平。该协议主要衡量起始状态的持续性以及候选池的抽取方式。在我们的持久性控制协议中，忽略视频时的得分恰好为机会水平，EEG- World 和线性模型都能以高于机会的水平识别片段，但二者没有拉开差距，因为单次观看在频带功率中携带的片段特异性信息很少。这些信息会在观看者之间累积（图8）。将每个候选片段在测试片段的所有观看次数中的得分取平均，并让每个预测都基于其自身的上下文，使 394 个片段中的 Top-1 识别率从 0.8% 提升至 3.8%（机会水平为 0.25%；线性模型为 3.3%）；在 13% 的片段中，得分最高的候选片段与真实片段属于同一类别（机会水平为 4%）。这一合并分析属于事后分析。

7. 讨论与局限性

大脑就是世界。 在每项分析中，对接下来几秒脑电图的最强预测因素都是大脑自身的状态；迁移到未见受试者的是从连续脑电图中学习到的动态特性。忽略这些动态特性的视觉大脑基准测试，会把属于大脑的部分归因于刺激；历史感知基线和持久性控制指标成本低廉，应成为标准。由于优势会随着训练脑电图的增加而扩大，预测是脑电图基础模型一种自然的预训练目标。

局限性。 我们预测的是一种表征（每秒的对数频带功率）；更精细的时间尺度和源空间目标仍有待研究。视频有助于训练队列，但在未见受试者上没有带来净增益。核心受试者的测试集被用于数项预注册分析，详见第K节；确认性迁移结果使用了24名受试者，每人仅接受一次测试。片段大体按固定顺序呈现，因此前一片段造成的延续效应仅通过交换分析加以控制。

伦理。 本研究已获得南方科技大学科学与技术伦理委员会的批准，

（批准编号：20260399）。所有参与者均提供了书面知情同意，并获得了报酬。

大语言模型使用声明

大语言模型助手在作者的指导下使用。Claude（Anthropic）用于编写和调试分析代码、起草图表，以及润色本文文本。ChatGPT 和 Codex（OpenAI）以及 Gemini（Google）用于审阅文本。所有假设、决策规则和结果均由作者审阅，作者对本文内容承担全部责任。

参考文献

[1] Lindsay M. Alexander、Jasmine Escalera、Lei Ai、Charissa Andreotti、Karina Febre、Alexander Mangone 等。用于儿科心理健康和学习障碍跨诊断研究的开放资源。Scientific Data，4 170181，2017 15：。[2] E mily J. Allen、Ghislain St-Yves、Yihan Wu、Jesse L. Breedlove、Jacob S. Prince、Logan T. Dowdle、Matthias Nau、Brad Caron、Franco Pestilli、Ian Charest、J. Benjamin Hutchinson、Thomas Naselaris 和 Kendrick Kay。用于连接认知神经科学与人工智能的大规模 7T 功能性磁共振成像数据集。Nature Neuroscience，25 126，2022 1 116 2（）：–。[3] Mido Assran、Adrien Bardes、David Fan、Quentin Garrido、Russell Howes、Mojtaba Komeili、Matthew Muckley、Ammar Rizvi、Claire Roberts、Koustuv Sinha、Artem Zholus 等。V-JEPA 2：自监督视频模型实现理解、预测与规划。arXiv 预印本 arXiv:2506.09985，2025。8、15[4] Jake Bruce、Michael Dennis、Ashley Edwards、Jack Parker-Holder、Yuge Shi、Edward Hughes 等。Genie：生成式交互环境。载于国际机器学习会议（ICML），2024。2[5] Jingjing Chen、Xiaobin Wang、Chen Huang、Xin Hu、Xinke Shen 和 Dan Zhang。大规模细粒度情感计算脑电图数据集。Scientific Data，10:740，2023。3[6] Mehdi Cherti、Romain Beaumont、Ross Wightman、Mitchell Wortsman、Gabriel Ilharco、Cade Gordon、Christoph Schuhmann、Ludwig Schmidt 和 Jenia Jitsev。对比式语言-图像学习的可复现缩放定律。载于 IEEE/CVF计算机视觉与模式识别会议（CVPR），2023。3、8、12[7] Kyunghyun Cho、Bart van Merriënboer、Caglar Gulcehre、Dzmitry Bahdanau、Fethi Bougares、Holger Schwenk 和 Yoshua Bengio。使用循环神经网络编码器-解码器学习用于统计机器翻译的短语表示。载于自然语言处理经验方法会议（EMNLP），2014。3[8] Stéphane d’Ascoli、Jérémy Rapin、Yohann Benchetrit、Hubert Banville 和 Jean-Rémi King。TRIBE：用于全脑功能性磁共振成像响应预测的三模态脑编码器。载于国际学习表征会议（ICLR），2026。2

[9]Zijian Dong、Jianxiong Zhou、Kwun Kei Ng、Jan Paolo Macapinlac Balagtas、Zhizhou Li、Zijiao Chen 和 Juan Helen Zhou。NeuroWorld：用于刺激条件化人脑动态的潜在脑世界模型。arXiv预印本 arXiv:2608.01773，2026年。2、5、8、15[10] Yassine El Ouahidi、Jonathan Lys、Philipp Thölke、Nicolas Farrugia、Bastien Pasdeloup、Vincent Gripon、Karim Jerbi 和 Giulia Lioi。REVE：一种脑电图基础模型，通过在25,000名受试者身上进行大规模预训练，适配任意设置。载于《神经信息处理系统进展》（NeurIPS），2025年。2[11] Alex Fang、Albin Madappally Jose、Amit Jain、Ludwig Schmidt、Alexander Toshev 和 Vaishaal Shankar。数据过滤网络。载于国际学习表征会议（ICLR），2024年。8[12] Jianxiong Gao、Yichang Liu、Baofeng Yang、Jianfeng Feng 和 Yanwei Fu。CineBrain：用于脑条件视频生成的大规模多模态视听脑数据集。载于IEEE/CVF计算机视觉与模式识别会议（CVPR），第36224–36234页，2026年。3[13] Alessandro T. Gifford、Domenic Bersch、Marie St-Laurent、Basile Pinsard、Julie Boyle、Lune Bellec、Aude Oliva、Gemma Roig 和 Radoslaw M. Cichy。Algonauts项目2025年挑战赛：人脑如何理解多模态电影。arXiv预印本 arXiv:2501 00504，2025年 2、5 。 .[14] Alessandro T. Gifford、Pablo Oyarzo、Anne W. Zonneveld、Christina Sartzetaki、Iris I. A. Groen 和 Radoslaw M. Cichy。一个包含人类对短自然主义视频脑电图反应的大型数据集，用于研究动态视觉事件处理。arXiv预印本 arXiv:2608.28768，2026年。3、12、16[15] David Ha 和 Jürgen Schmidhuber。循环世界模型有助于策略演化。载于《神经信息处理系统进展》（NeurIPS），2018年。2[16] Danijar Hafner、Timothy Lillicrap、Ian Fischer、Ruben Villegas、David Ha、Honglak Lee 和 James Davidson。学习潜在动态，以实现基于像素的规划。载于国际机器学习会议（ICML），2019年。2[17] Danijar Hafner、Timothy Lillicrap、Jimmy Ba 和 Mohammad Norouzi。梦境控制：通过潜在想象学习行为。载于国际学习表征会议（ICLR），2020年。[18] Danijar Hafner、Jurgis Pasukonis、Jimmy Ba 和 Timothy Lillicrap。通过世界模型掌握多样化的控制任务。《Nature》，640(8059)：647–653，2025年。2[19] Kaiming He、Xinlei Chen、Saining Xie、Yanghao Li、Piotr Dollár 和 Ross Girshick。掩码自编码器是可扩展的视觉学习器。载于IEEE/CVF计算机视觉与模式识别会议（CVPR），2022年。8[20] Tong He、Ru Kong、Avram J. Holmes、Minh Nguyen、Mert R. Sabuncu、Simon B. Eickhoff、Danilo Bzdok、Jiashi Feng 和 B. T. Thomas Yeo。深度神经网络与核回归在行为和人口统计学特征的功能连接预测方面准确率相当。《神经影像》，206：116276，2020年。5[21] Jordan Hoffmann、Sebastian Borgeaud、Arthur Mensch、Elena Buchatskaya、Trevor Cai 等。计算最优大语言模型训练的实证分析

。载于《神经信息处理系统进展》（NeurIPS），第30016–30030页，2022年。5[22] Minyoung Huh、Brian Cheung、Tongzhou Wang 和 Phillip Isola。柏拉图式表征假说。载于国际机器学习会议（ICML），2024年。1[23] Alexander G. Huth、Wendy A. de Heer、Thomas L. Griffiths、Frédéric E. Theunissen 和 Jack L. Gallant。自然语音揭示了在人类大脑皮层中铺展的语义图谱。《Nature》，532(7600)：453–458，2016年。1、2[24] Wei-Bang Jiang、Li-Ming Zhao 和 Bao-Liang Lu。利用海量脑电图数据在脑机接口中学习通用表征的大型脑模型。载于国际学习表征会议（ICLR），2024年。2[25] Jared Kaplan、Sam McCandlish、Tom Henighan、Tom B. Brown、Benjamin Chess、Rewon Child、Scott Gray、Alec Radford、Jeffrey Wu 和 Dario Amodei。神经语言模型的缩放定律。arXiv预印本 arXiv:2001 08361，2020年 5 。.[26] Kendrick N. Kay、Thomas Naselaris、Ryan J. Prenger 和 Jack L. Gallant。从人脑活动中识别自然图像。《Nature》，452(7185)：352–355，2008年。1、2[27] Sander Koelstra、Christian Mühl、Mohammad Soleymani、Jong-Seok Lee、Ashkan Yazdani、Touradj Ebrahimi、Thierry Pun、Anton Nijholt 和 Ioannis Patras。DEAP：一个使用生理信号进行情感分析的数据库。《IEEE 情感计算汇刊》，3(1)：18–31，2012年。3[28] Yann LeCun。迈向自主机器智能之路。OpenReview预印本，2022年。15[29] Guangyao Li、Yake Wei、Yapeng Tian、Chenliang Xu、Ji-Rong Wen 和 Di Hu。学习在动态视听场景中回答问题。载于IEEE/CVF计算机视觉与模式识别会议（CVPR），2022年。3、12[30] Xuan-Hao Liu、Yan-Kai Liu、Yansen Wang、Kan Ren、Hanwen Shi、Zilong Wang、Dongsheng Li、Bao-Liang Lu 和 Wei-Long Zheng。EEG2Video：从脑电图信号解码动态视觉感知。载于《神经信息处理系统进展》（NeurIPS），2024年。2、3[31] Yong Liu、Tengge Hu、Haoran Zhang、Haixu Wu、Shiyu Wang、Lintao Ma 和 Mingsheng Long。iTransformer：反转变换器可有效用于时间序列预测。载于国际学习表征会议（ICLR），2024年。4、6、12[32] Ilya Loshchilov 和 Frank Hutter。解耦权重衰减正则化。载于国际学习表征会议（ICLR），2019年。4、12[33] Vincent Micheli、Eloi Alonso 和 François Fleuret。变换器是样本高效的世界模型。载于国际学习表征会议（ICLR），2023年。2[34] Thomas Naselaris、Kendrick N. Kay、Shinji Nishimoto 和 Jack L. Gallant。功能性磁共振成像中的编码与解码。《神经影像》，56(2)：400–410，2011年。1、2[35] Shinji Nishimoto、An T. Vu、Thomas Naselaris、Yuval Benjamini、Bin Yu 和 Jack L. Gallant。重建由自然影片诱发的脑活动所对应的视觉体验。《当代生物学》，21(19)：1641–1646，2011年。1、2

[36] Brian A. Nosek、Charles R. Ebersole、Alexander C. De- Haven 和 David T. Mellor。预注册革命。《美国国家科学院院刊》，115(11)：2600–2606，2018年。5[37] Maxime Oquab、Timothée Darcet、Théo Moutakanni、Huy Vo 等。DINOv2：无需监督地学习鲁棒视觉特征。《机器学习研究汇刊》，2024年。3、8、12[38] Chethan Pandarinath、Daniel J. O’Shea、Jasmine Collins、Rafal Jozefowicz、Sergey D. Stavisky、Jonathan C. Kao、Eric M. Trautmann、Matthew T. Kaufman、Stephen I. Ryu、Leigh R. Hochberg、Jaimie M. Henderson、Krishna V. Shenoy、L. F. Abbott 和 David Sussillo。使用序列自编码器推断单试次神经群体动态特性。《自然·方法》，15(10)：805–815，2018年。2[39] Viorica Pˇatrˇaucean、Lucas Smaira、Ankush Gupta、Adrià Re- casens Continente、Larisa Markeeva、Dylan Banarse、Skanda Koppula、Joseph Heyward、Mateusz Malinowski、Yi Yang 等。感知测试：多模态视频模型的诊断基准。在《神经信息处理系统进展》（NeurIPS）数据集与基准赛道，2023年。3、12[40] Christina Sartzetaki、Anne W. Zonneveld、Pablo Oyarzo、Alessandro T. Gifford、Radoslaw M. Cichy、Pascal Mettes 和 Iris I. A. Groen。作为视频理解中动态专家混合模型的人脑。在国际学习表征会议（ICLR），2026年。2[41] Martin Schrimpf、J onas Kubilius、Ha Hong、Najib J. Ma- jaj、Rishi Rajalingham、Elias B. Issa、Kohitij Kar、Pouya Bashivan、Jonathan Prescott-Roy、Franziska Geiger、Kailyn Schmidt、Daniel L. K. Yamins 和 James J. DiCarlo。Brain-Score：用于物体识别的哪种人工神经网络最像大脑? bioRxiv 407007，2018年。1、2[42] Marc-Andre Schulz、B. T. Thomas Yeo、Joshua T. Vogel- stein、Janaina Mourão-Miranda、Jakob N. Kather、Konrad Kording、Blake Richards 和 Danilo Bzdok。UKBiobank脑影像与机器学习数据集中线性模型和深度学习的不同缩放方式。《自然通讯》，11：4238，2020年。5[43] Seyed Yahya Shirazi、Alexandre Franco、Maurício Scopel Hoffmann、Nathalia B. Esper、Dung Truong、Arnaud De- lorme、Michael P. Milham 和 Scott Makeig。HBN-EEG：健康大脑网络（HBN）脑电图数据集的 FAIR 实现。bioRxiv预印本 2024.10.03.615261，2024年。3[44] Yapeng Tian、Jing Shi、Bochen Li、Zhiyao Duan 和 Chenliang Xu。无约束视频中的视听事件定位。在欧洲计算机视觉会议（ECCV），2018年。3、12[45] Zhan Tong、Yibing Song、Jue Wang 和 Limin Wang。VideoMAE：掩码自编码器是用于自监督视频预训练的数据高效学习器。在《神经信息处理系统进展》（NeurIPS），2022年。8[46] Nicolas Waehner、Enzo Tagliazucchi、Jacobo Sitt、Demian Wassermann、Giovanni Franco Gabriel Marraffini 和 Trinidad Borrell。柏拉图式大脑：脑电图基础模型中的表征趋同。arXiv预印本 arXiv:2611.05097，2026年。2

[47] Jiquan Wang、Sha Zhao、Zhiling Luo、Yangxuan Zhou、Hait- eng Jiang、Shijian Li、Tao Li 和 Gang Pan。CBraMod：用于脑电图解码的交叉式脑基础模型。在国际学习表征会议（*ICLR*），2025年。2[48] Mo Wang、Wenhao Ye、Junfeng Xia、Junxiang Zhang、Xu- anye Pan、Minghao Xu、Haotian Deng、Hongkai Wen 和 Quanying Liu。Omni-fMRI：通用无图谱功能磁共振成像基础模型。在国际机器学习会议（*ICML*），2026年。2[49] Daniel L. K. Yamins、Ha Hong、Charles F. Cadieu、Ethan A. Solomon、Darren Seibert 和 James J. DiCarlo。性能优化的分层模型能够预测高级视觉皮层中的神经反应。《美国国家科学院院刊》，111(23)：8619–8624，2014年。1、2[50] Joel Ye 和 Chethan Pandarinath。使用神经数据变换器学习神经群体活动的表征。神经元、行为、数据分析与理论，5(3)，2021年。2[51] Ailing Zeng、Muxi Chen、Lei Zhang 和 Qiang Xu。变换器对于时间序列预测是否有效？在*AAAI*人工智能会议，第11121–11128页，2023年5–。[52] Wei-Long Zheng 和 Bao-Liang Lu。利用深度神经网络研究基于脑电图的情绪识别中的关键频带和通道。*IEEE*自主心理发展汇刊，7(3)：162–175，2015年。3

A. 实现细节

规模。 上下文 $T_c = 6$ 和滚动预测 $T_f = 10$ 步，每步 1 秒（16 秒时间窗）； $D = 155$ 特征（31 个通道 $\times$ 5 频带的对数功率）； $H = 512$。编码器 e 为线性层、LayerNorm、GELU 和 Dropout 0.1；主干网络 f 由线性层和 GELU 构成；视频通路 g 是一个两层多层感知机，最后一层初始化为零。线性世界模型以线性递归取代门控循环单元，并去掉所有 GELU。门控循环单元核心组有 160 万个参数，共享模型有 200 万个参数（加上视频通路后为 230 万个）；每个受试者头增加 8 万个参数，因此包含 70 个头部的 2,012 受试者模型共有 760 万个参数。

优化。 两个阶段都仅针对 T_f 滚动预测步计算均方误差；梯度会通过全部 16 个门控循环单元步反向传播，包括上下文阶段；滚动预测期间不会输入观测到的脑电图（不使用教师强制）。阶段1：AdamW [32], 学习率10^{-3}，权重衰减为 0.01，批次大小为 64，经过预热后，在 80 个训练轮次内采用余弦衰减，并根据选择误差早停；在 908小时数据上每个训练轮次为 250 步，在 2,012小时数据上为 555 步。阶段2：使用 65 名受试者的视频试次（226,979 次观看，630 小时），每个批次的一半取自连续时间窗（根据验证集选择）；预训练参数的学习率 3×10^{-4}（在验证片段上从四个候选学习率中选出），视频通路和新头部的学习率为 10^{-3} 。

基线方法。 选择数据为每位受试者连续脑电图记录的 10%，以及视频试次中对 391 个验证片段的观看数据。所有模型，无论是否为线性模型，均使用这些数据选择超参数；训练过的模型还使用这些数据确定停止轮次；不会在这些数据上重新拟合。岭回归惩罚项按受试者设定，取设计矩阵平均特征值的 10^{-4} 至 10^{5} 倍；对于分层岭回归向合并权重收缩的惩罚项，取值范围扩展至 10^{7}；对于历史数据， + 视频岭回归的视频模块尺度则从 1、0.3、0.1、0.03 和 0.01 中选取。无状态基线方法——一个三层多层感知机和一个 iTransformer——以相同方式选择宽度、深度和学习率 [31], 。

未见受试者模型。 2,012 小时的连续池和 65 名受试者的视频集包含 24 名未见受试者的训练数据，因此第 6.3 节中的模型在不使用这些受试者数据的情况下进行训练。图 6a 使用的是在纳入这些受试者之前、基于 54,133 次试次微调的冻结视频模型；图 6b 和零样本测试使用的是从同一预训练核心出发、在视频集其余 41 名受试者的数据上微调的模型。合并式和分层线性基线方法均在与各模型相同的受试者数据上重新拟合，因此除各自的校准数据外，任何比较都未涉及这 24 名受试者的数据。

B. 数据

公开数据资源。 在 2,584 小时的脑电图数据中，有 2,122 小时记录于四个公开视频数据集：AVE [44]（28 个事件类别中的 4,097 个 10 秒片段；1,016 小时）、MUSIC- AVQA [29]（536 小时）、感知测试 [39]（510 小时）和 EEG Moments 数据集 [14]（60 小时）。其余 462 小时主要来自较早的录制阶段，记录于其他短视频中。只有 AVE片段会进入视频实验；其他三个公开数据集提供连续脑电图。辅助文件 ave_clip_split.csv 列出了每个 AVE片段的公开标识符，以及其属于训练（3,312 个）、验证（391 个）还是测试（394 个）片段。所有刺激特征均来自冻结的公开编码器。

内部资源。 脑电图记录属于人体受试者数据，是在经伦理审查批准的方案下收集的，并已取得书面知情同意。所有分析仅使用假名代码标识参与者。

脑电图预处理。 脑电帽以 250 Hz 采样，参考电极位于 FCz；语料预处理采用 0.1–96 Hz 带通滤波，应用 50 和 60 Hz 陷波滤波器，并将采样率重采样至 256 Hz。片段起始时刻按下文所述映射至脑电图时钟。每段录制均沿校正后的脑电图时间轴切分为 1 秒的时间箱，从其首个片段起始时刻之前 6 秒处开始。对于每个时间箱和通道，功率谱采用多窗估计法计算，使用时间—半带宽积为 2 的离散长球面序列窗，保留谱集中度高于 0.9 的窗；去除均值后，根据特征值对窗加权。频带功率为 δ 1–4、 θ4–8、 α 8–13、 β 13–30 和 γ 30–45 Hz 范围内各频率箱谱功率的均值；特征取其以 10 为底的对数，每秒得到 $31\times5 = 155$ 个特征。若任一通道的峰峰值振幅超过 $500\,\mu$V，则将该时间箱标记为不良；在每段录制的不良时间箱之外，对特征进行 Z 分数标准化，并跳过包含不良时间箱的时间窗。

刺激特征。对于片段的每一秒，Open-CLIP [6]和 DINOv2-L [37] 图像嵌入均取自四帧的平均值；每种嵌入都降至 24 个主成分，这些主成分在训练片段上拟合，并在训练片段上缩放至单位方差。音轨转换为 16 kHz 单声道后，每秒的响度以其均方根电平（单位为 dB）表示，低于 16位静音的值则设为 16位静音；谱通量则是对数短时功率正向变化的均值，步长为 10 ms；两者均根据训练片段进行 Z 分数标准化。另设一个标记，用于标明片段出现在屏幕上的秒数；片段起始前的所有特征均为零。表 3 中比较的编码器采用相同方式，其中视频模型使用每秒 16 帧。

时钟校正。设备时钟与呈现计算机时钟的差值中位数为 −980ppm（四分位距−1，430 至 −760）。若不进行校正，一段录制（时长中位数为 31 分钟）末尾的片段起始时刻将偏差中位数为 1.5 秒，90 百分位数为 4.5 秒，最大可达 10 秒。对每段记录的时钟偏移和漂移率，使用匹配滤波器根据片段起始和结束时刻前后的 1–30 Hz 脑电图信号进行拟合，并在奇数片段上拟合、偶数片段上测试。如果采集软件也记录了设备采样计数，两种估计结果一致（差值中位数为 1 ppm，$r = 0.994$，501 段记录）。第 3 节中的数据驱动时序检查（在 2 秒空白处的后验 α，1,820 段记录中有 1,765 段的时间落在预期秒数内）使用至少有 40 次观看的记录中的训练片段。未经校正时，线性模型的视频增益几乎减半。恒定偏移（中位数为 51 ms）包含平均显示延迟和传输延迟；绝对延迟未经光电二极管测量。

C. 观看者状态在视频片段中保持持续

图 1基于这一观察：观看者以不同状态开始观看同一视频片段，而这种差异会持续超过片段的起始阶段。图S1 在每一次测试集中观看中检验了这一点。我们根据某项特征在起始前6秒内的平均值，将每个视频片段的观看次数分为两组，并跟踪片段中前10%和后10%的观看次数。在155项特征中，有141项的两组在片段第2—10秒期间、所有15名受试者中仍存在差异；而对全部155项特征而言，至少有13名受试者存在这种差异（图S1c）。仅凭视频进行的预测——其他受试者对同一视频片段的平均响应——几乎不包含这种差异（中位数仅为测量值的0.3%），因此仅基于刺激的编码模型无法表达这种差异。

最持久的方向在片段前和片段中的平均值之间保持着 λ =0。81 的相关性。其权重集中在 β 和 γ 颞区电极（T7、T8）处的功率上（图S1a），这是一种典型的头皮肌肉活动模式，因此其中一部分持续存在的信号很可能反映的是肌张力，而非神经活动。第一主成分对所有特征赋予大致相等的权重，并保持0.38的相关性；后验 α 仅保持0.25。在这三个方向上，EEG-World以及使用EEG历史和视频的线性模型，都使两组之间的距离保持在测量值的水平（持久成分为2.3和2.4个标准差，而测量值为2.4个标准差）；而仅基于视频的预测则无法做到这一点（0.01个标准差；图 S1d）。

由此可以得出两个结论。每位观看者的反应中有很大一部分延续自片段播放之前，这正是我们的基线方法看到相同脑电图历史，而不只是刺激的原因。将状态向前传递本身并不能区分世界模型与线性模型，因为任何历史感知预测器都会这样做；世界模型的优势在于状态如何演化，尤其是当状态初始时远离受试者的通常状态时（图7b）。

D. 增益的构成

表 S1. 在核心受试者的测试观看中，每次只进行一项改动的**消融实验**，每项使用三个随机种子。

变体（视频试次）	r	Δr [95% CI]	完整前瞻
EEG-World（完整模型）	.260	–	–
无个体专属头部	.240	−.020 [−.023, −.017]	15/15
3 秒上下文（而非 6 秒）	.249	−.011 [−.013, −.009]	15/15
无预训练	.251	−.009 [−.011, −.007]	14/15
仅片段起始	.253	−.007 [−.012, −.002]	10/15
无连续时间窗	.258	−.0016 [−.0030, −.0004]	11/15
空白 $b = 0$	.259	−.0006 [−.0012, +.0002]	11/15

逐一移除一个成分（表S1）表明，个体专属头部最为重要（−0.020 在 r，15/15；技能−0.010）：单个读出层无法适配每个人的通道和频带尺度。只读取6个上下文秒数中的最后3秒会损失0。011 （15/15），因此状态需要数秒的历史数据；仅在视频试次上训练，而不预先在连续脑电图上进行训练，则会损失0。009 （14/15）。这三项损失中的每一项都大于片段内容带来的损失（0。007）。每个微调批次中的连续时间窗会带来少许帮助（0。0016，11/15），而在两个阶段都将空白固定为 $b = 0$ 只会使r 改变 −0.0006，其区间包含零，因此学习得到的空白向量并非必需。

片段通过两种方式进入模型：作为其开始的时刻，以及作为其内容。仅告知模型片段何时开始，就已经带来了几乎全部的技能增益（表 2）：片段起始时发生的许多变化，都是所有片段共有的脑状态一般性转变。内容带来的提升较小，在 r 左右，与其对线性模型带来的提升大致相当。之所以仍使用内容，是因为替换片段对 EEG-World 的代价是对线性模型的两倍；然而，即使没有内容，EEG-World 仍优于线性 EEG +视频模型（+0.021 在 r，15/15；技能 +0.009，14/15；三个随机种子的均值，描述性分析）。世界模型的优势在于刺激所扰动的动态特性，而不在于更好的刺激编码。

E. 片段特征作用于何处

为了了解视频通路使用哪些特征，我们分别仅用视觉特征、仅用音频特征或两者都不用来重新训练该通路，并测量每类特征带来的增益（图S2）。该通路依赖视觉特征。视觉特征带来的最大增益出现在 δ 和 θ，位于额叶外侧电极（F7、F8、FT9、FT10；最高达 +0.03，在 r 中），这些部位可反映水平眼动；此外还出现在 θ 和 α，位于顶枕电极（+0.008在 r 处，11/15，预注册）。在 γ 中，它们略微降低了 r （−0.005）。因此，部分视觉效应可能与眼动有关，因为片段内容既能预测观看者如何移动眼睛，也能预测其视觉皮层如何响应；而顶枕区的增益并未显著高于颞叶的增益（$p = 0.07$）。音频

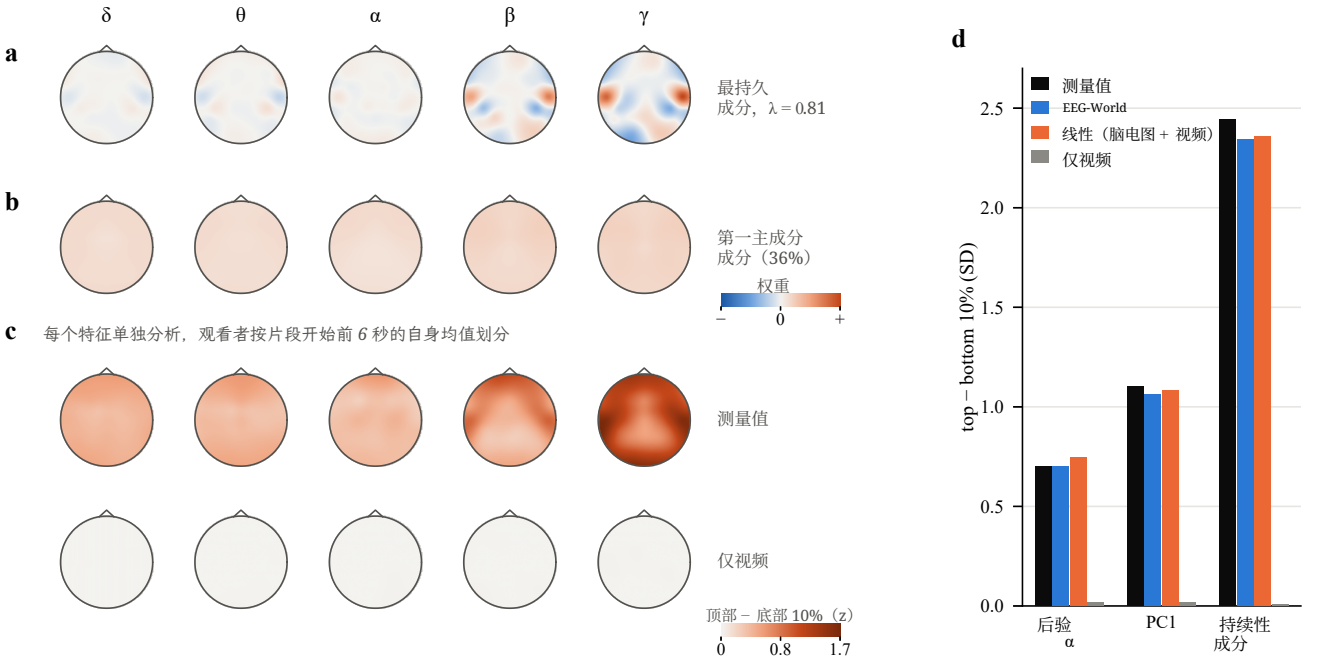


图 S1. **以不同状态开始观看片段的观看者，会持续保持不同状态。** 15 名核心受试者的全部 10,970 次测试观看。(a) 仅根据训练观看次数学习得到的 155 个特征中最持久成分的权重：该方向在片段开始前 6 秒内的均值，与片段第 2–10 秒内的均值相关性最高；计算时已扣除其他观看者对每个片段的平均响应（λ 即该相关性）。(b) 片段前均值的第一主成分（占其方差的 36%）。(c) 每个特征单独考察：该特征片段前均值位于最高和最低 10% 的观看者，在片段第 2–10 秒内的差异；分别显示测量值，以及仅根据视频预测的差异（其他 14 名受试者对同一视频片段的平均响应）。(d) 三个状态变量对应的相同差异。

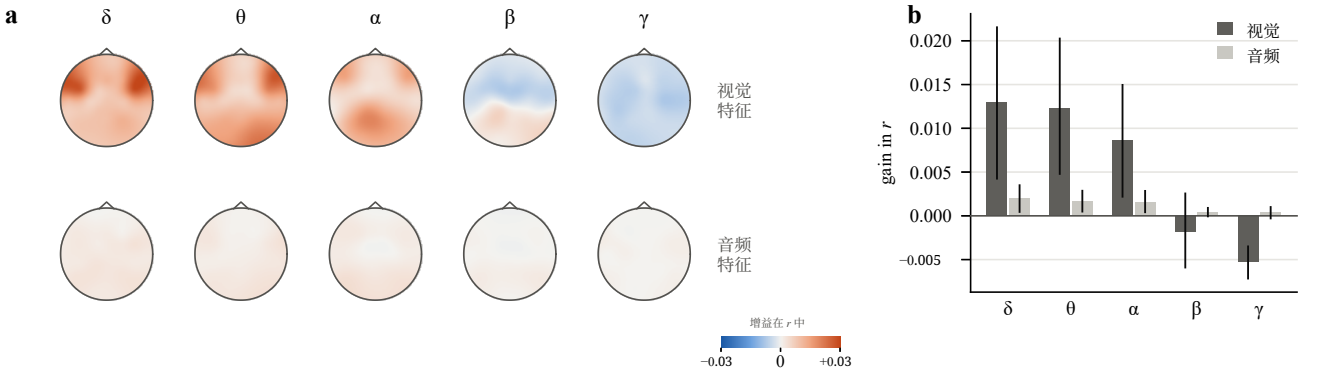


图 S2. **视觉和音频特征。** 各类特征在测试片段上的预测增益 r：在一个 2×2 设计中，重新训练的模型分别使用视觉特征、音频特征、两者兼用或均不使用（每种情况使用三个随机种子），并对 10 个预测步取平均。(a) 各电极和频带上的受试者均值，使用统一量尺。(b) 所有电极上的频带均值，以及基于 15 名受试者计算的自助法 95% 区间。

特征带来的增益很小（+0.001 在颞叶部位， 12/15，预注册，且不高于其他部位），而视觉特征存在时，音频特征不再带来任何增益。仅替换视觉特征会使每个区域的 r 降低0.022–0.029；仅替换音频特征， <0.001。

F. 规模、计算与架构

从 908 小时提升到 2,012 小时所带来的增益中，约有一半来自计算量，因为使用 908 小时数据训练、但采用 2,012 小时步数预算的模型达到了 $r = 0.282$，而更大的数据池 0。284 （表 S2）。LSTM 和因果 Transformer 核心组（2 层或 4 层，学习率分别调优；最多达到 4× 核心参数）优于

表 S2. **连续脑电图，无视频。** 15 名核心受试者的留出会话，随机种子均值；仅核心组的参数。 *EEG-World 的架构与无非线性的训练。

		908小时		2,012小时	
模型	参数	r	技能	r	技能
最佳线性模型（历史岭回归）	–	.2716	.1745	.2718	.1746
线性世界模型*	–	.2572	.1655	–	–
EEG-World（门控循环单元）	1.6M	**.2809**	**.1787**	.2844	.1804
+JEPA潜在损失	1.6M	–	–	**.2859**	**.1813**
908小时数据，2,012小时预算	1.6M	.2824	.1795	–	–
LSTM 核心	2.1M	.2784	.1773	.2831	.1797
因果 Transformer 核心	6.3M	.2788	.1775	.2813	.1787
非线性、无状态（直接进行 *10* 秒预测）					
MLP	5.7M	–	–	.2740	.1745
iTransformer	3.6M	–	–	.2727	.1740

在 15/15 受试者中的最佳线性模型（908 小时和 2,012 小时）表现，因此增益并不依赖于门控循环单元；门控循环单元是性能最佳且规模最小的模型，而 Transformer 在 2,012 小时时低 0。003于 r（0/15）。在指数移动平均目标编码器的潜在状态上加入辅助损失 [3, 28] 可带来小幅且一致的增益（技能 15/15）。

G. 校准：平滑预测并非向均值塌缩

预测比脑电图更平滑，这是对不确定未来进行预测时必然如此的。为检查预测是否只是向均值收缩，我们按受试者、步和特征，对测量到的脑电图进行预测回归。在连续脑电图数据上，每一步的斜率均为 0.93–0.99，且预注册标准（在验证集和测试数据中，至少 13 名受试者的全部 10 步斜率都在 $[0.8, 1.25]$ 范围内）得以满足（14/15 和 15/15）。在视频试次中，斜率从第一秒的 0.93 降至第十秒的 0.81，因此仅在 6/15中满足该标准：较晚的预测略微过冲，这与塌缩恰恰相反，而最佳线性模型也有相同表现（从 0.95 降至 0.79）。在这两种情况下，预测方差都接近测量方差的 r^2倍，符合校准良好的预测。

H. 识别

在轨迹匹配协议下（图 S3），EEG-World 的前十名命中率达到 28.5%（16× 机会），最佳线性模型为 26.5%（+1.9 个百分点， 11/15， $p = 0.04$），但上下文均值已达到 12.5×机会。将 15 名受试者的全部 10,970 次观看作为候选项时，相同预测达到 145× 机会（上下文均值为 90×）。在持久性控制协议下，EEG-World 以 5.5% 的准确率识别片段（2.2× 机会， 15/15高于机会水平），线性模型则为 4.7 %（受试者层面的 $p = 0.15$，预先注册的检验未通过；片段层面 $p = 0.024$）：单次观看的频带功率中几乎不含片段特异性信号。上下文

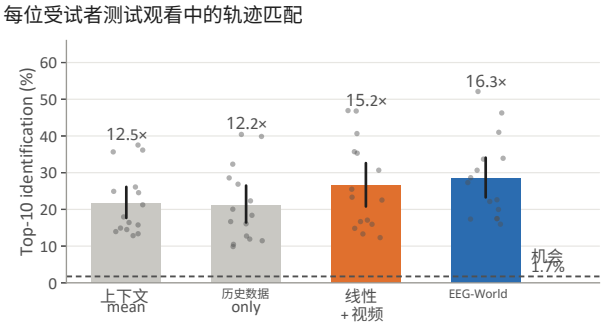


图 S3. **轨迹匹配** [9] 在每位受试者的测试观看中进行：按受试者统计前十准确率，并以机会的倍数表示均值和 95% 置信区间。上下文均值不考虑视频。

均值在轨迹匹配下得分如此之高，是因为第 C 节所示的持久性；无论片段是什么，一次观看的过去都与其未来相似。

I. 其他脑电设备

仅通过添加一个新的线性预测头即可实现，与另外两种设备上的最佳线性历史模型相比，冻结模型表现相当：+0.004 在 r $[-0.002, 0.009]$ 59 通道实验室电极帽上（12 名成年人观看一部故事片； 9/12；预注册技能检验未通过），以及 +0.0005 $[-0.003, 0.004]$ 128 通道电极网上（94 名儿童观看来自健康大脑网络的电影 [1]； 48/94，描述性分析）。

J. 正文结果的统计

表 S3 和S4 列出了第6节每项比较背后的统计数据。差异为受试者均值，方括号内为参与者自助法 95% 区间； k/n 表示差异为正的受试者人数； p 表示精确的受试者层面符号翻转检验（15 名核心受试者）或 100,000 次随机符号翻转（24 名未见受试者）。预注册决策采用技能。最后一列标明该行对应的预注册检验（第K节）。

K. 所有预注册检验

表格 S5 和 S6 列出了在本文数据上运行的每一项预注册测试，即 H1 至 H60，包括正文未讨论的测试。在每一轮测试之前，我们都预先记录了其假设、模型、验证集、决策规则以及将用于评分的测试数据，并且只在此后对测试数据进行评分。除非另有说明，测试均按照第 5节的规则，根据技能进行判定；对于 H12–H14，受试者标准是 24 人中超过一半（采用 16/24 时结局相同），而在实验室脑电帽上则为 9/12。非劣效检验要求 95%区间的下限高于规定的界值，测试失败后的第二次尝试使用 $p < 0.025$。标识符 H7、H39 和 H40 从未注册。核心组的

受试者的测试数据经过多轮评分，因此之后在这些受试者上得到的结果并不是相互独立的重复验证；24 名未见受试者首先在验证性测试（H12–H14）中接受评分，随后又在 H24–H26、H53 和 H53b 中接受评分。在该语料库存在之前，先导模型仅使用实验室脑电帽数据和 EEG Moments 数据集进行训练 [14]，每位受试者几小时的数据，但并未胜过线性预测器；H41–H45 使用预训练模型重复了实验室脑电帽比较。

在60项测试中，36项满足标准，23项失败；H35为双侧检验。若干失败结果塑造了本文。第一个从头训练的视频模型不敌线性模型（H4）；在连续脑电图上进行预训练有所帮助（H9），但需要第二次尝试才能击败最佳线性模型（H8、H11）。线性模型的视频项比世界模型的视频项更能匹配观看者之间的平均响应（H18、H20），尽管世界模型的片段锁定响应会随数据量增加而增强（H19）。在未见受试者的连续脑电图上，冻结模型仅名义上胜过最佳线性模型（H12）；零样本迁移失败（H24、H25）；多设备预训练向新受试者的迁移效果略差（H53、H53b），因此主要模型均在同一设备上训练。在2,012小时核心组上使用908小时方案对视频通路进行微调，效果更差（H32、H33），因为早停使8次运行中的7次在第8个训练轮次结束，而此时视频通路尚未收敛。在实验室电极帽上，预训练使试点实验中的损失达到持平，但并未转化为胜出（H41–H45）。识别未能区分EEG-World与线性模型（H23、H34；第H节），视觉增益和听觉增益都不特异于匹配的感觉电极（H49、H50；第E节），且校准标准在视频试次上失败（H60；第G节）。

表 S3。 **正文比较的统计数据，第 6.1 和 6.2 节。** 以“vs.”开头的行将 EEG-World 与指定模型进行比较。效应：差值 r 及其 95%区间，除非给出了单位或该行写有“in skill”；rel.：相对于另一模型 r的百分比； k/n：领先的受试者数。最佳线性模型：基于 EEG历史的最佳逐受试者、合并或分层岭模型，在试次上使用视频特征；对于连续脑电图，该模型在 908、1,142 和 2,012 小时达到 $r = 0.272$ 。SLv：片段锁定相关。测试集：决定该行结果的预注册检验（表 S5 和 S6）；其他各行属于次要分析或描述性分析。

比较	效应 [95% CI]	rel.	k/n	p	Test
连续脑电图（第*6.1*节，图*4*）					
vs. 最佳线性模型，14 h		−20.7%	0/15		
对比最佳线性模型，56 h		−7.7%	0/15		
对比最佳线性模型，227 小时		−2.8%	2/15		
对比最佳线性模型，685小时		+3.5%	15/15	6×10^{-5}	H10
在技能方面	+0.0042				
与最佳线性模型相比，908小时		+3.4%	15/15		
对比最佳线性模型，1,142小时		+4.0%	15/15		
与最佳线性模型相比，2,012 小时		+4.6%	15/15		
领先优势增长，从 908 小时增至 2,012 小时	+0.0032 [0.0021, 0.0046]		15/15	10^{-4}	H27
技能中	+0.0016				
线性世界模型与 EEG-World 对比	−0.024		0/15		H6
线性世界模型与闭式岭回归对比	较低（表 S2）				
长短期记忆网络和因果 Transformer 与最佳线性模型对比	见表。S2		每项 15/15		H36（Tr.）
多层感知机 vs. 最佳线性模型，2,012 小时	+0.002		12/15		
iTransformer 对比最佳线性模型，2,012 小时	+0.001		7/15	n.s.	
对比多层感知机	+0.010 [0.009, 0.012]		15/15	6×10^{-5}	H54
对比 iTransformer	+0.012 [0.010, 0.014]		15/15	6×10^{-5}	H55
视频试次（第 *6.2*节，图 *5*，表 *2*）					
与最佳历史数据 +视频相比	+0.029 [0.026, 0.031]	+12.3%	15/15	6×10^{-5}	H21
在技能检验中	+0.0103				
提前测试片段			285/394		
在第一秒领先		+4%			
在第十秒领先		+18%			
预测时域增益	3.3 秒 [2.2，4.5]		14/15		
相比多层感知机	+0.018 [0.016，0.020]		15/15		H56
与 iTransformer 对比	+0.012 [0.009, 0.015]		15/15		H57
150 至 630 小时的试验脑电图，EEG-World	+0.012		13/15		H22
150 至 630 小时的试验脑电图，线性	+0.005		11/15		
SLv，与最佳历史数据相比 + 视频	−0.058		2/15		H20（失败）
SLv，150 至 630 小时		+46%	15/15		H19
交换成本，EEG-World	0.026		15/15		
交换成本，线性	0.012		15/15		
仅片段起始 vs. 预训练模型	+0.005		13/15		
技能方面	几乎全部增益				
截取起始点之后的内容，EEG-World	+0.007		10/15		
技能方面	n.s.				
截取起始点之后的内容，线性	+0.008		14/15		

表 S4。 **正文比较的统计，见第 6.3 至 6.5 节**，如表 S3 所示。片段：在 394 个测试片段上进行测试（H13 使用符号翻转法，H58 使用自助法）。 *与基于受试者全部数据拟合的线性模型相比，中位数为 2,046 个试次（5.7 小时）；描述性分析。

比较	效应 [95% CI]	rel.	k/n	p	Test
未见受试者（第 6.3 节，图 6）					
冻结的视频模型（图 6a）与最佳线性模型	$+0.020$ [0.013, 0.027]	$+8.1\%$	23/24	CLIP $< 10^{-5}$	H13
在技能中	$+0.0083$				
重新训练的模型（图 6b），全部为自有数据		$+9.7\%$	24/24		H26
无视频世界模型与最佳线性模型		$+6.8\%$	21/24		
视频与无视频世界模型	$+0.003$			0.20	
EEG-World 与最佳线性模型的平均值		$+9.3\%$	24/24		H14
交换成本，EEG-World	0.022		22/24		
交换成本，最佳线性模型	0.011		23/24		
交换成本，EEG-World 优于最佳线性模型	$2.0\times$		21/24		
来自 200 个试次的预测头（≈33分钟）*		$+3.5\%$	15/24		
来自 1 个试次的预测头，000 次试次*			23/24		
零样本 vs. 合并线性模型		$+3.2\%$	20/24		
在技能中	没有更好		8/24		H24（失败）
模型学到了什么（第6.4节，图7，表3）					
具有正增益的特征	155/155		15/15		
FDR显著特征，连续 / 视频	154 / 155				
视频试次的最大频带增益（α）	$+0.034$	$+20\%$			
连续脑电图上的相同频带		$+6.8\%$			
连续脑电图的预测时域增益	1.5 秒 [1.1，1.9]				
增益，初始状态的最远五分位（$1.6\times$）		$+5.2\%$			
Gain, 初始状态的最近五分位		$+3.2\%$			
技能中的最远五分位减最近五分位（$3.8\times$）	$+0.0072$ [0.0057，0.0088]		15/15	10^{-4}	H37
计算机模拟与实测调谐，50×155 细胞	$r = 0.57$			置换检验 0.001	H38
具有正一致性的特征	47/50				
最强特征（CLIP 第一主成分）	$r = 0.94$				
其 θ 图；响度 δ 图	$r = 0.96$；0.95				
SLv、V-JEPA 2 对比 CLIP + DINOv2	$+35\%$ [25, 47]		15/15	clip $< 10^{-4}$	H58
两者的 SLv	0.188 对比 0.139				
SLv，视频模型与图像模型	$+0.031$		15/15		H59
识别（第 6.5节，图 8）					
轨迹匹配、前十名与线性	28.5% 对比 26.5%		11/15	0.04	
EEG-World / 上下文均值	$16\times$/ $12.5\times$机会				
将全部 10,970 次观看作为候选项，与线性方法对比	13.2% 对比 11.9%		10/15	0.065	H34（失败）
EEG-World / 上下文均值	$145\times$ / $90\times$ 机会				
持久性控制，相较于线性	5.5% 对比 4.7%		11/15	0.15	H23（失败）
EEG-World 对比机会	$2.2\times$		15/15		

表 S5。 **关于连续脑电图和实验室电极帽的预注册检验。** 除非另有说明，技能差异均有利于前述模型； k/n 统计差异为正的受试者数量；线性模型同表 S3。 ‡在相同数据的基础上，另外使用来自 128 通道电极网的 142小时数据和来自 EEG Moments 数据集的 71 小时数据进行预训练。

Test	假设	结果	k/n	结局
连续脑电图，*15* 名核心受试者的留出会话				
H1	逐受试者世界模型与按受试者岭回归的比较，每者 ≈46 小时	$-0.0021 (p = 0.009)$	2/15	失败
H2	H1 的差距从每位受试者 1 条记录缩小到 48 条记录	-0.052 至 -0.002	15/15	met
H3	基于 15 名受试者（908小时）合并训练的世界模型，与逐受试者岭回归的比较	$+0.0050$	15/15	met
H5	汇 pooled 世界模型与最佳线性模型的比较	$+0.0043$	14/15	met
H6	汇 pooled 世界模型与线性世界模型的比较	$+0.025$	15/15	met
H6b	采用最终 80轮训练计划的两个模型的 H6	$+0.013$	15/15	met
H10	EEG-World与最佳线性模型在685小时时的对比	$+0.0042$	15/15	met
H29	EEG-World与最佳线性模型在2 ,012小时	$+0.0058$	15/15	met
H27	相较于最佳线性模型的领先优势从 908 小时扩大到 2,012 小时	$+0.0016$	15/15	met
H28	EEG-World 在 2 ,012 小时，相比之下为 908 小时	$+0.0017$	14/15	met
H30	在相同步数预算下，2,012 小时的数据与 908 小时的数据	$+0.0009$	11/15	met
H15	908小时时，JEPA 风格潜在损失与无潜在损失相比	$+0.0010$	15/15	met
H16	在227小时也是如此	$+0.0011$	13/15	met
H31	在2也是如此 ,012 小时	$+0.0009$	15/15	met
H17	使用潜在损失时，EEG-World 在 227 小时时就已胜过最佳线性模型	-0.0058		失败
H35	因果Transformer 与门控循环单元核心组对比，2,012 小时（双侧）	-0.0016	0/15	门控循环单元更优
H36	因果 Transformer 核心组 vs. 最佳线性模型, 2,012 小时	$+0.0042$	15/15	met
H54	EEG-World vs. MLP直接预测器 , in r	$+0.010$	15/15	met
H55	EEG-World vs. iTransformer直接预测器 , in r	$+0.012$	15/15	met
H37	相较于最佳线性模型的技能提升，初始状态最远五分位与最近五分位的比较	$+0.0072$	15/15	met
H46	多设备预训练‡ 不劣于单设备（界值 0.001）	$+0.0001$ $[-0.00002, 0.0003]$		met
H51	r 中相同（界值 0.003）	$+0.0003$ [≈0, 0。0008]		met
H60	$[0.8, 1.25]$ 中所有步骤在 $\geq 13/15$ 受试者中的校准斜率（验证集、测试集）	14/15、15/15；视频 6/15		失败（视频）
59 通道实验室电极帽，*12* 名成年人观看一部故事片				
H41	微调后的 EEG-World 与最佳线性模型对比	$+0.0022 (p = 0.13)$	8/12	失败
H42	相同模型与仅使用这些数据训练的模型	$+0.0136$	12/12	met
H43	冻结的 EEG-World +新头部与最佳线性模型	$+0.0022 (p = 0.11)$	8/12	失败
H44	多设备‡ 与单设备预训练相比，冻结 + 新头部	$+0.0000$	5/12	失败
H45	多设备‡ 预训练与最佳线性模型相比	$+0.0023 (p = 0.11)$	9/12	失败

表 S6. **视频试次的预注册检验**，针对 15 名核心受试者和 24 名未见受试者，与表 S5相同。SLv：片段锁定相关性。 †在片段层面（对 394 个测试片段进行符号翻转）并结合受试者数量进行判定。 ‡ 与表 S5 相同。 § 三个随机种子之一在第 0 个训练轮次停止。

Test	假设	结果	k/n	结局
15 名核心受试者的视频试次测试观看				
H4†	从头训练的视频模型（150 小时）与逐受试者历史记录 + 视频岭回归	-0.004	3/15	失败
H8†	从连续核心数据微调的视频模型 vs. 最佳线性模型	$+0.0007$	8/15	失败
H11†	在 H8 上的第二次尝试（更宽的学习率网格；片段级 $p < 0.025$）	$+0.0047$（片段 $p = 0.0019$）	12/15	met
H9†	微调模型 vs. 从头训练的模型	$+0.0128$	15/15	met
H21†	EEG-World vs. 最佳线性模型，630 小时	$+0.0103$	15/15	met
H22†	630 小时 vs. 150 小时的视频配对脑电图	$+0.0054$	14/15	met
H32†	第 2 个视频模型 ，012小时核心组与 908小时核心组对比	-0.0042	1/15	失败
H18	SLv、视频模型（150小时）与最佳线性模型对比	-0.079	0/15	失败
H20	SLv、EEG-World（630 小时）与最佳线性模型	-0.058	2/15	失败
H19	SLv，630 小时与 150 小时的视频配对脑电图	0.152 与 0.104	15/15	met
H33	SLv, 第 2 个视频上的视频模型，012小时与908小时核心组的对比	0.128 与 0.152		失败
H58	SLv，验证集上最佳视觉编码器（V-JEPA 2）与 CLIP + DINOv2	$+0.049$ [0.035，0.065]	15/15	met
H59	SLv，视频与图像编码器	$+0.031$ [0.024，0.037]	15/15	met
H23	持久性控制识别，394个片段中的前十名，与最佳线性模型相比	5.5% 对比 4.7%（$p = 0.15$）	11/15	失败
H34	以全部 10,970 次观看作为候选进行轨迹匹配，与最佳线性模型比较	$+1.3$点（$p = 0.065$）	10/15	失败
H38	计算机模拟调谐与实测调谐相关（验证片段）	$r = 0.57$(置换。 $p = 0.001$)		met
H47	视觉特征改善顶枕区 r	$+0.0076$	11/15	met
H48	音频特征改善颞叶 r	$+0.0014$	12/15	met
H49	视觉增益在顶枕区大于颞叶时间部位	$+0.0023$($p = 0.074$)	13/15	失败
H50	颞叶部位的听觉增益大于顶枕部位	-0.0005	7/15	失败
H52	多设备预训练‡ 在 r 中具有非劣性（界值 0.003）	-0.0029 [-0.0053, -0.0003]		失败§
H52b	第二次尝试，早停从训练轮次 20 开始计算	$+0.0013$ [-0.0001, 0。0026]	11/15	met
H56	EEG-World 与多层感知机直接预测器的比较, in r	$+0.018$	15/15	met
H57	EEG-World 与 iTransformer直接预测器 , in r	$+0.012$	15/15	met
未见受试者，*24* 名受试者的测试观看（*H12*：连续脑电图）				
H12	冻结的连续模型 +新头部 与 最佳线性模型	$+0.0015$($p = 0.067$)	15/24	失败
H13†	冻结的视频模型 + 新头部与最佳线性模型对比（图6a）	$+0.0083$	23/24	met
H14†	两项预测的平均值优于两者	$+0.0106$/ $+0.0023$	24/24, 18/24	met
H26	在其余 41 名受试者上重新训练的模型 + 新头部与最佳线性模型	$+0.0097$	24/24	met
H24	相同模型，零样本，与合并线性模型比较	-0.0043	8/24	失败
H25	相同模型，持久性控制识别	-0.58分（$p = 0.11$）	6/24	失败
H53	多设备预训练‡ 在 r 中不劣（界值 0.003）	-0.0018 [-0.0058, 0.0022]	10/24	失败
H53b	第二次尝试，三个随机种子	-0.0020 [-0.0053, 0.0012]	11月 24日	失败