一种可准确预测语音可懂度的多解码器神经追踪方法

Rien Sonck[1], Bernd Accou[1], Tom Francart[1],Jonas Vanthornhout[1]

[1] ExpORL，神经科学系，鲁汶大学，鲁汶，比利时
jonas.vanthornhout@kuleuven.be

arXiv:2602.03624v1 [eess.SP] 3 Feb 2026

摘要

目的：基于脑电图（EEG）的方法在预测通过行为测试测得的语音可懂度方面展现出潜力，但其准确性和稳健性通常不及标准行为重测程序；后者得出的重测差异往往小于 1 分贝。我们提出多解码器方法，用于根据语音聆听时的 EEG记录预测语音接收阈值（SRT）。该方法专为无法进行行为测试的人群设计，可在各种聆听环境中客观评估语音可懂度，并为临床决策、助听器验配或意识障碍（DoC）患者评估提供参考。

方法：多解码器方法汇总了数百个解码器的信息。每个解码器都采用不同的语音特征组合和脑电预处理配置进行训练，以计算神经追踪（NT），量化各解码器从脑电图中重建语音信号的能力。为训练和测试解码器，我们使用了39名参与者（7名男性、32名女性）的数据，参与者年龄为18至24岁，每位参与者的脑电记录总时长为29分钟。每位参与者首先完成一项行为任务，以估算其语音接收阈值（SRT）。随后，在参与者聆听以六种不同信噪比（SNR）呈现的语音以及在安静环境中播放的旁白故事时，采集脑电记录。接着，将所有解码器得出的NT 值合并，为每位受试者生成一个高维特征向量。随后训练支持向量回归（SVR）模型，以学习这些NT向量与行为 SRT之间的关系，旨在根据NT向量预测个体的SRT。

主要结果：采用多解码器方法进行的预测与行为 SRT 显著相关（r=0.647, $p < 0.001$），归一化均方根误差为 =0.19；所有差异均小于 1 分贝。SHAP（SHapley 加性解释）分析表明，尽管所有特征均有贡献，但 θ 波和 δ 波频带、早期时滞以及被试特异解码器对模型预测的影响略大。此外，我们证明，利用预训练的受试者独立（SI）解码器，可将新参与者的脑电数据采集时间缩短至仅 15 分钟，其中包括 3 分钟的故事聆听，以及在六种信噪比条件下各聆听 40 个矩阵句、总计 12 分钟的任务，且不会牺牲预测准确性。

重要性：本研究提出的多解码器方法优于以往的基于脑电图的方法。未来研究应在更多样化的临床人群中验证该方法，包括听力损失者。

1. 引言

评估语音可懂度，即听者理解口语词语和句子的程度，是评估人类听觉功能的一个基本方面。广泛使用的一项指标是语音接收阈值（SRT），其定义为听者能够正确复述所呈现语音材料中50%的信噪比（SNR）。语音接收阈值（SRT）可用于评估在具有挑战性的聆听环境中使用或未使用助听器的听力受损者（Festen & Plomp, 1986; Plomp, 1986; Verschuure & van Benthem, 1992），也可用于评估特定助听器算法的有效性（van Dijkhuizen et al., 1989; Zaar et al., 2024）。

针对多种语言，研究人员开发了行为语音接收阈值任务，以评估噪声中的语音可懂度。一个早期例子是瑞典矩阵测试（Hagerman, 1982），此后该测试启发了多种语言版本的开发，包括荷兰语和弗拉芒语版本（Houben et al., 2014; Luts et al., 2014）。然而，这些行为评估要求参与者主动参与，因此在测试儿童或某些临床群体时可能颇具挑战，甚至难以实施。因此，基于神经影像的方法在测量语音可懂度方面日益受到关注。

神经影像学研究采用了神经追踪的反向（即解码器）和前向（即编码器）模型（综述见 Gillis et al., 2022）。这些模型旨在建立语音刺激特征（如包络、声学起始和音位）与相应脑反应之间的线性映射。反向模型从多通道脑电图（EEG）信号中重建语音特征，而前向模型则根据刺激特征预测各通道的 EEG 信号。对于反向模型，重建特征与实际特征之间的皮尔逊相关称为神经追踪（NT）值。对于前向模型，该值指重建的 EEG 信号与实际 EEG 信号之间的皮尔逊相关。

语音包络感知对语音可懂度至关重要（Peelle & Davis, 2012; Shannon et al., 1995; Smith et al., 2002）；研究还发现，行为测得的语音可懂度与语音包络的神经追踪之间存在很强的相关性，进一步支持了这一观点（Vanthornhout et al., 2018）。然而，尽管宽带包络感知是实现可懂度的必要条件，却并非总是充分条件；时频精细结构也发挥着重要作用，尤其是在嘈杂或具有挑战性的聆听条件下（Ding et al., 2014; Lorenzi et al., 2006; Moore, 2008）。语音包络的解码还会受到频谱退化（Karunathilake et al., 2023; K¨osem et al., 2023）、语速（Verschueren et al., 2022）、注意（Vanthornhout et al., 2019）和双耳去掩蔽（Dieudonn´e et al., 2025）等因素的影响。值得注意的是，即使对于外语或不可懂的语音，也会出现包络的神经追踪现象

（Gillis et al., 2023; Or-tiz Barajas et al., 2021; Reetzke et al., 2021; Song & Iverson, 2016; Zou et al., 2019）；尽管如此，包络仍然是言语理解的一项重要声学特征。

尽管语音包络是语音可懂度的一项重要特征，声学起始在支持言语理解方面也发挥作用。声学起始由语音包络的半波整流一阶导数导出。声学起始对于语音的切分与组织十分重要，并携带关键的语音学信息（Brodbeck et al., 2018年; Kluender et al., 2003）。增强人工耳蜗中的声学起始已被证明可提高语音可懂度（Koning & Wouters, 2012, 2016）。然而，与包络一样，无论语音可懂度如何，听觉系统都会处理声学起始（Karunathilake et al., 2023）。

目前，仅有少数研究提出了用于估计语音可懂度的客观基于脑电图的方法。Iotzov和Parra（2019）实施了一种结合正向建模与反向建模的混合方法来评估语音可懂度；然而，他们并未尝试估计语音接收阈。他们发现，语音检测随可懂度提高而改善，表现为脑电图—语音相关性增加。专门用于从脑电图数据预测语音接收阈值的方法始于Vanthornhout 等（2018）的工作，他们采用了反向建模方法。在该方法中，神经追踪值在一系列信噪比条件下计算，并对这些值拟合Sigmoid函数，类似于在行为实验中拟合心理测量曲线。拟合曲线的中点称为相关阈值，被用作行为语音接收阈值的估计，二者之间的相关性为0.69，中位数绝对差为1.7分贝，标准差为3.17分贝。然而，该方法并非对所有参与者都成功：在21%的情况下（24名受试者中的5名），相关阈值无法可靠确定。Lesenfants 等（2019）研究了一种采用正向建模的替代方法，评估了多种语音特征，包括包络、语谱图、音位、语音学特征以及一个称为FS的组合特征集（语音学特征 + 语谱图的音位）。FS特征表现最佳，且无需Sigmoid拟合。相反，他们提出了FS零交叉点，定义为神经追踪值由负转正时的信噪比。该指标可有效预测行为 SRT，并达到0.97分贝的均值绝对差。另一项采用正向建模的研究由Borges 等（2025）开展，他们使用包络特征并同样应用Sigmoid拟合。其方法的中位数差为0.38分贝，标准差为1.45分贝，但未能可靠拟合9%的参与者（22名中的2名）。Muncke 等（2022）通过将前向模型与均方根方法相结合来预测行为 SRT，从而无需Sigmoid拟合，并达到1.2分贝的均值差。

尽管大多数关于深度学习与语音可懂度的研究并不直接估计语音接收阈，但 Accou 等（2021）的研究是一个值得注意的例外。在这项研究中，研究人员使用匹配—失配范式训练了一种空洞卷积神经网络，并通过对不同语音接收阈下的分类准确率进行Sigmoid拟合来估计语音接收阈。该方法与行为 SRT 的相关系数为 0.59，绝对差中位数为 3.64 分贝，标准差为 1.68 分贝。

尽管近期取得了进展，仍有两项主要挑战。首先，许多现有方法仍难以在个体层面提供可靠的 SRT 预测。相比之下，行为矩阵测试始终能够准确测量 SRT，研究报告的重测变异性非常低——中位差约为 0.4 分贝（Boon，2014年），标准差为 0.7 分贝（Decruy 等人，2018年）。虽然基于脑电图的方法不断进步，并逐渐缩小这一准确度差距，但其稳健性仍未达到行为基准的水平。其次，由于语音材料各不相同，而且参与者的行为 SRT 分布存在差异，因此难以比较不同研究中的 SRT 预测表现。

在本研究中，我们提出一种应对第一个挑战的新方法：多解码器方法。其关键创新在于整合大量解码器的信息；每个解码器均针对不同的刺激特征配置（包络和声学起始）、积分窗口（范围为 0 至 500 毫秒（ms））、频带（δ 波、θ 波和宽频带），以及被试特异和被试无关的解码器配置进行训练。这样便无需预先为所有受试者选定这些参数。此外，我们的方法将传统的 Sigmoid拟合替换为误差函数与支持向量回归（SVR）相结合的方式，从而能够为所有参与者稳健、准确地预测 SRT。我们还分析了维持模型最佳性能所需的脑电图数据量。具体而言，我们希望了解，预训练的被试无关解码器能否减少新参与者所需的脑电图数据量，同时保持模型预测其 SRT 的性能不变——这是迈向临床环境实际应用的关键一步。

针对第二个挑战：为比较不同语音材料及SRT分布下的SRT预测性能，我们主张各研究采用更加标准化的报告方式。为此，我们建议使用归一化均方根误差（归一化均方根误差）作为指标，因为它考虑了数据分布的变异性，从而使SRT预测性能的比较更有意义。

2. 材料与方法

2.1. 参与者

共纳入39名参与者（7名男性，32名女性），年龄在18至24岁之间（$\bar{x} = 21$，标准差 = 1.8），数据来自Accou等人于2024年描述的数据集。所有参与者的听阈均处于正常至轻度听力损失范围内，这一点已由纯音气导测听确认。在125–8000 Hz范围内，大多数听阈为 $\leq$ 20 分贝听力级（听力级），少数介于21至30分贝听力级之间。参与前，所有受试者均对鲁汶大学医院/研究医学伦理委员会（鲁汶大学，比利时；编号S57102）批准的方案提供了书面知情同意。

2.2. 实验程序

每位参与者均完成了 Accou 等人（2024年）详细介绍的一系列任务。在本研究中，我们分析了一个行为任务（下文称为矩阵行为任务）和两个脑电图任务的数据：故事任务和矩阵EEG任务。所有听觉刺激均以双耳方式呈现，声压级为 62 dBA。脑电图数据通过64个头皮电极以 8192 Hz的采样率记录。有关脑电图采集的更多细节，请参阅 Accou 等人（2024年）。实验程序概览见图*1* 。

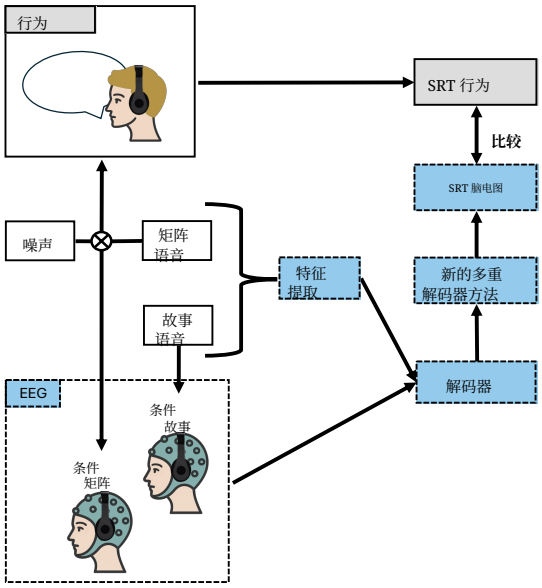

图1： 实验程序概览。 每位参与者完成了三项主要任务：首先是行为任务，他们聆听被语音加权噪声掩蔽的矩阵句，并复述这些句子以确定其行为语音接收阈值；随后完成两项脑电图任务，均为被动聆听——一项聆听旁白故事，另一项聆听在不同语音加权噪声水平下播放的矩阵句。特征提取聚焦于干净刺激的包络和声学起始。随后，利用旁白故事和矩阵句的刺激特征及 *EEG*记录，训练并测试了数百种不同的解码器配置。接着，使用我们提出的多解码器方法对这些解码器进行组合，整合所有解码器的信息来预测语音接收阈值，并与行为语音接收阈值进行比较。

2.2.1. 矩阵行为任务

佛兰芒语矩阵行为任务（Luts et al., 2014）采用自适应程序来确定每位参与者的语音接收阈值。每个测试列表由20个标准化句子组成（一个矩阵列表），并在固定为65分贝声压级的语音加权噪声背景下呈现。有关任务的完整细节，请参阅 Luts et al.（2014）。

2.2.2. 故事任务

在故事任务中，参与者聆听了一段时长为15分钟的弗拉芒语故事《Milan》。该故事由男性讲述者 Stijn Vranken 撰写并朗读，同时记录参与者的脑电图。

2.2.3. 矩阵脑电图任务

在矩阵脑电图任务期间，参与者专注地聆听以固定信噪比呈现于语音加权噪声中的矩阵句： -12.5、 -9.5、 -6.5、 -3.5、 -0.5 和 2。5 分贝，以及静音条件。这些条件统称为信噪比条件。研究人员告知参与者被动聆听。每个信噪比条件包含两个矩阵列表，共计 40 个句子，句子之间有短暂停顿。所有参与者均完成了每个信噪比条件，各条件的顺序针对每位参与者随机安排。每个条件持续约两分钟，因此任务总时长约为 14 分钟。

2.3. 语音特征提取

语音信号的采样率为 $48\ kHz$。从故事和矩阵刺激材料的无噪声版本中提取语音包络和声学起始。

2.3.1. 包络

为提取时间包络（Env），使用包含 28 个子带的伽马通滤波器组对音频进行处理，中心频率范围为 $50\ Hz$ 至 $5000\ Hz$，随后进行幂律压缩：每个子带包络通过取每个采样点的绝对值并将其提升至 0.6 次幂来计算。最终的宽带包络通过对所有子带取平均获得，所使用的滤波器组配置与 Biesmans 等人（2017 年）和 Vanthornhout 等（2018 年）所述的相同。

2.3.2. 声学起始

声学起始（AEs）计算为语音包络的一阶导数（即连续包络样本之间的差值），并施加半波整流，仅保留正向变化。

2.4. 信号预处理

脑电图数据首先使用截止频率为0.5 Hz的一阶零相位巴特沃斯滤波器进行高通滤波。随后将数据降采样至 1024 Hz。使用多通道维纳滤波器（Somers等人，2018年）去除眨眼伪迹，并对该次脑电图记录中识别出的任何坏通道进行线性插值。信号被重新参考至公共平均参考。

脑电图、刺激的包络以及刺激的 AE 均被下采样至 64 Hz。分别应用三个最小二乘带通滤波器（阶数为 2000）提取 δ 波（0.5–4 Hz）、θ 波（4–8 Hz）和宽频带（0.5–30 Hz）频带；每个滤波器的阻带衰减均设为 80 分贝，通带波纹为 1 分贝。最后，所有脑电图通道

并对整个录音中的语音特征进行 z 分数归一化。

2.5. 解码器

为了评估大脑对语音特征（尤其是旁白故事和矩阵句的包络与声学起始）的追踪或对齐程度，我们使用了解码器（即反向模型）。这种方法通过神经数据重建语音特征，并将其与原始语音特征进行比较，以此衡量神经追踪程度；该程度以实际特征与重建特征之间的皮尔逊相关表示（全面综述见 Gillis et al., 2022）。解码器训练完成后，可应用于未见过的神经数据，以重建相应的刺激特征。所得相关系数即为神经追踪值：相关性越高，表示从神经数据重建刺激特征的准确度越高；对于参与者而言，这可作为语音可懂度的指标（Vanthornhout et al., 2018）。

2.5.1. 解码器配置

解码器配置由一组用于设置解码器训练和测试的参数定义，包括脑电图任务（故事或矩阵）、语音特征（包络或声学起始）、积分窗口（27个窗口）、频带（θ 波、δ 波和宽频带）以及解码器类型（被试无关或被试特异）。总计得到648种不同的解码器配置。每种解码器配置均用于训练和测试7个解码器，每个信噪比条件对应一个：-12.5, -9.5, -6.5, -3.5, -0.5、2.5 分贝和静音，共计4536个解码器。对于每个解码器，我们使用岭回归来估计最优权重集（Machens 等，2004），将脑电图数据映射到所关注的语音特征，即语音包络或声学起始。虽然解码器可使用矩阵或故事脑电图任务的数据进行训练，但所有解码器均仅在矩阵脑电图任务上进行测试，因为该任务的表现提供了预测语音接收阈值所需的信息。

脑电图任务在第 2.2 节中定义，语音特征在第 2.3 节中定义，频带在第 2.4 节中定义。下面将详细介绍后整合窗口和解码器类型。

2.5.2. 刺激后积分窗口

由于听觉处理时延会导致神经对声音的反应出现不同延迟，因此将脑电图数据扩展为多个自身的时滞版本。该刺激后积分窗口在 0 到 500 ms 之间系统地变化。通过从 0 开始逐步增大最大时滞，构建一系列积分窗口：先从 0 到 5 个采样点，再从 0 到 6 个采样点，依此类推。在 64 Hz 的采样率下，这些窗口对应的时滞范围分别为 0–78 ms、0–94 ms、0–109 ms 等；每增加一个窗口，最大时滞就延长约 15 至 16 ms，直至 500 ms。最终共得到 27 个窗口。

2.5.3. 解码器训练

我们使用了被试无关（SI）和被试特异（SS）解码器。SI 解码器采用留一被试交叉验证（CV）方法，在所有参与者上进行训练和评估（见图*2A*）。在该流程中，每位参与者的数据均有一次作为测试数据，而来自其余参与者的数据则用于训练。解码器使用来自故事任务或矩阵任务单一信噪比条件的脑电图数据进行训练。如前所述，每个解码器仅在矩阵任务脑电图数据的一个对应信噪比条件上进行测试。

SS解码器使用单个参与者的数据进行训练和测试（见图*2B*）。采用了两种训练配置。第一种配置中，解码器在该参与者故事任务的脑电图数据上训练，并在该参与者矩阵任务的一个信噪比条件上测试。第二种配置中，解码器在同一矩阵任务的数据上既进行训练也进行测试。由于每个信噪比条件包含40个矩阵句子，因此采用留一法按句子交叉验证：用39个句子训练解码器，剩余句子用于测试和特征重建。该过程重复进行，直至每个句子均被重建一次。随后，将全部40个测试句子的重建刺激特征拼接起来（排除静音间隔），并与相应的实际刺激特征进行相关，从而得到单个神经追踪值。

2.6. 预测语音接收阈值

2.6.1. 调整后的神经追踪值

每种解码器配置都用于训练和测试七个解码器，每个信噪比条件对应一个，因此得到七个神经追踪（NT）值：-12.5, -9.5, -6.5, -3.5, -0.5、2.5 分贝和静音，如（图*3A*）所示。由于噪声中的语音所引发的神经反应与静音中的语音所引发的神经反应存在根本差异（Di Liberto & Lalor, 2017），因此后续分析不纳入静音条件。选择 -12.5 分贝信噪比条件作为噪声基线，因为该条件的信噪比远低于所有参与者的语音接收阈值。将此基线下的 NT 值从其他信噪比条件下的 NT 值中减去，由此得到信噪比分别为 -9.5, -6.5, -3.5, -0.5 和 2.5 分贝时的五个校正后神经追踪值，见图*3B*。

2.6.2. 特征：经误差函数校正的 *NT* 向量

由于每种解码器配置在不同信噪比下会产生五个校正后的神经追踪值，因此总计有 3240 个校正后的神经追踪值。对于每位受试者，这些校正后的神经追踪值会拼接成一个 NT 向量（见 图*3C*）。随后，使用误差函数（ERF）（Abramowitz & Stegun，1972）对 NT 向量进行变换（见 图*3D*），从而得到经 ERF 校正的 NT 向量。ERF 取代了 Sigmoid 拟合，将曲线转换为平滑的 S 形曲线，类似于心理测量函数，但无需进行被试特定曲线拟合。

2.6.3. 模型：支持向量回归

每位参与者均对应一个经 ERF 校正的 NT 向量。为根据这些向量预测相应的 SRT 值，研究采用了线性支持向量回归模型（Smola & Sch¨olkopf，2004），并结合岭正则化以及一种基于 L-BFGS 的高效优化算法（Ho & Lin，2012；Hoerl & Kennard，1970）。该方法对经 ERF 校正的 NT 向量与语音接收阈值结果之间的线性关系进行建模，同时通过正则化控制过拟合。

采用嵌套留一交叉验证（CV）程序评估模型性能，以确保超参数调优和测试过程不受偏差影响。在外层循环中，留出一名受试者作为测试样本，其余受试者构成训练集。在该训练集中进行内层留一交叉验证，每次使用一名受试者作为验证集，

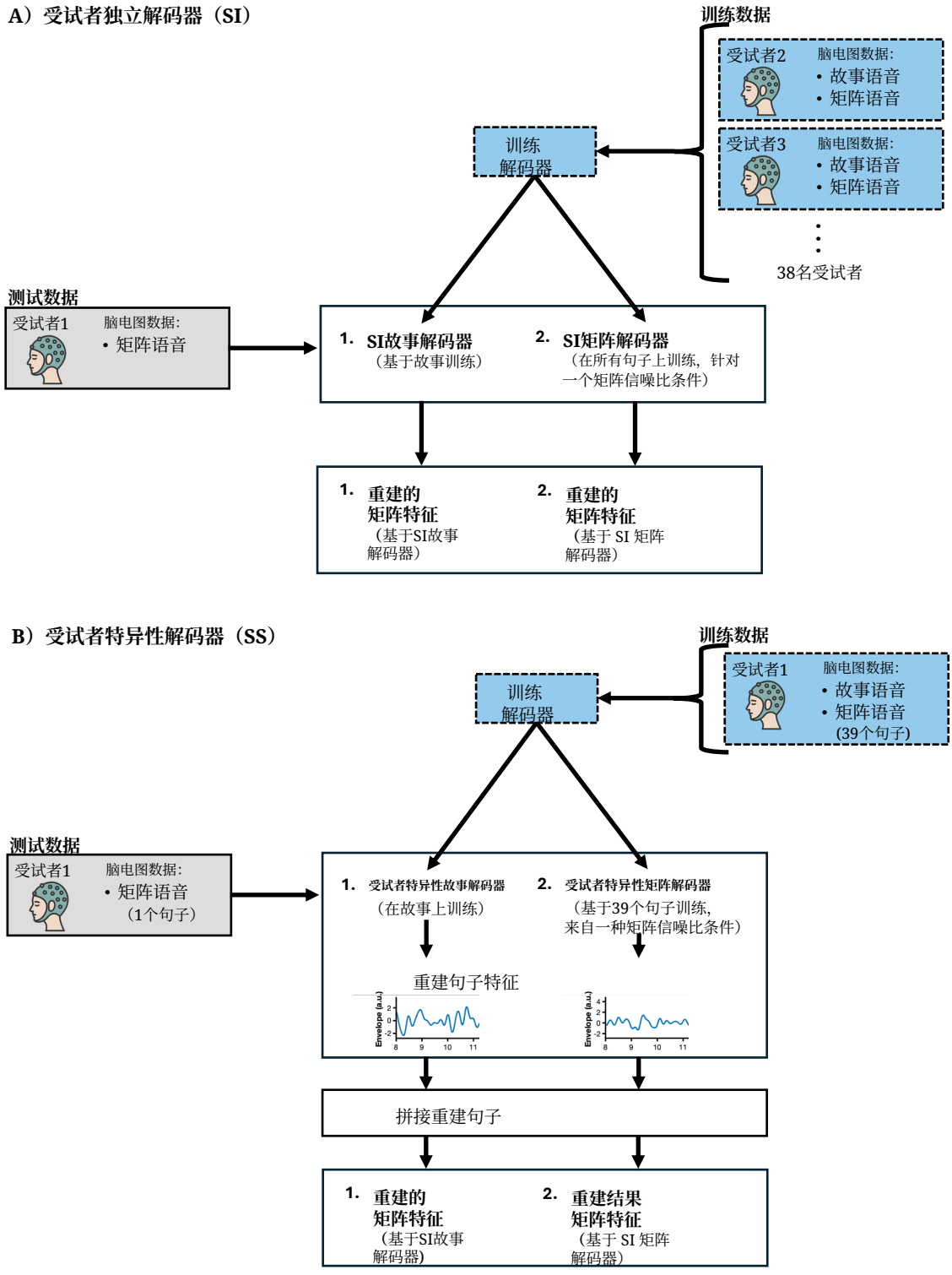

图2： 被试无关和被试特异解码器。 *(A)* 对于 *SI* 解码器，采用留一被试交叉验证；此面板展示了单个受试者的处理过程，代表交叉验证中的一折。图中展示了两个 *SI* 解码器：一个使用故事聆听任务中的脑电图数据训练（*SI*故事解码器），另一个使用矩阵聆听任务中特定信噪比条件下的脑电图数据训练（*SI*矩阵解码器）。每个信噪比条件包含两个矩阵句子列表，共计*40*个句子。*(B)* 对于 *SS* 解码器，在单个信噪比条件下的矩阵句上进行了留一交叉验证。与 *SI* 解码器类似，训练了两个 *SS* 解码器：一个使用故事脑电图数据训练（*SS*故事解码器），另一个使用*40*个矩阵句中的*39*个句子的脑电图数据训练（*SS*矩阵解码器）。在交叉验证的各折中重建全部*40*个句子的特征后，将重建的句子特征拼接成一个特征向量。

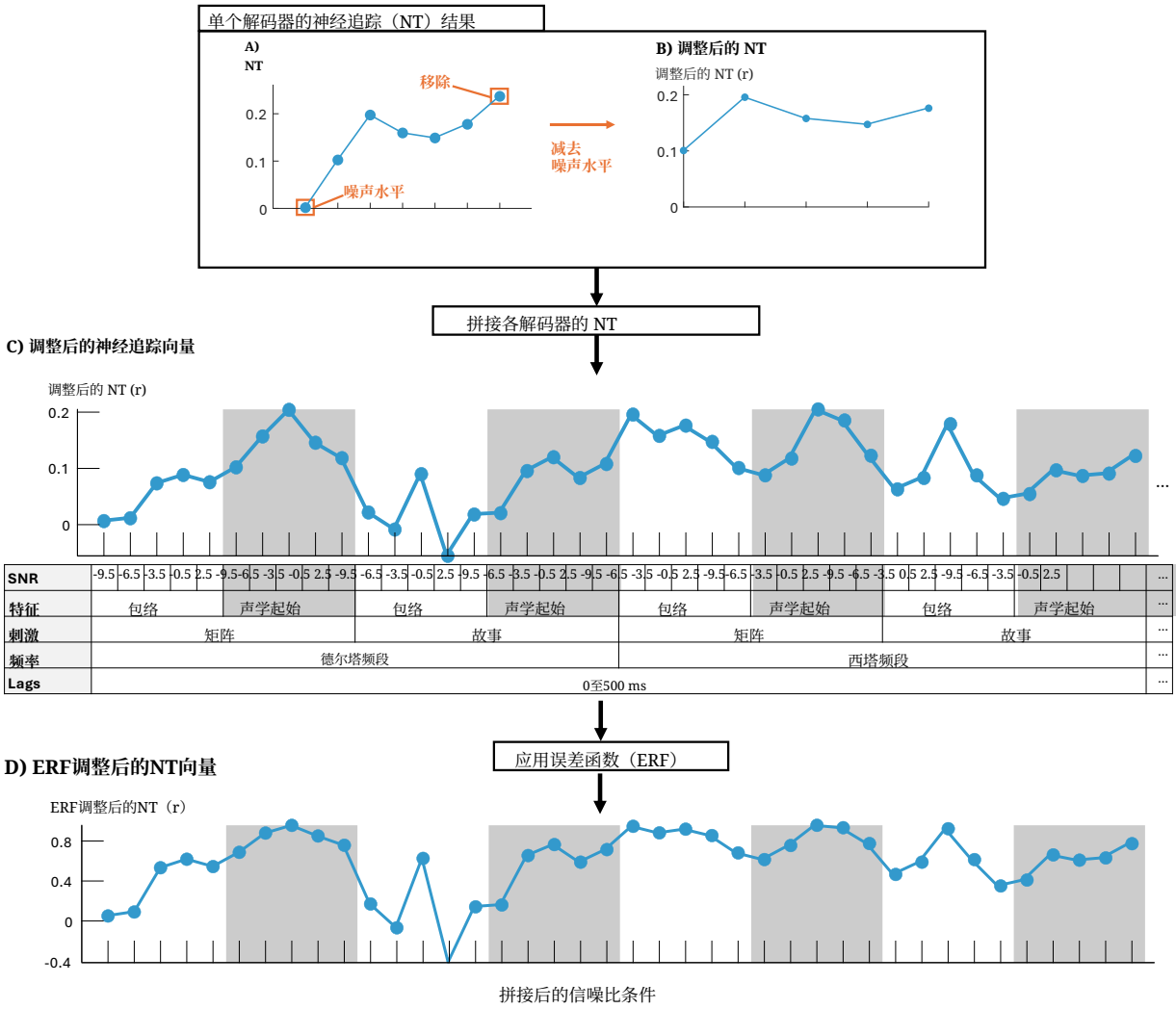

图3：***ERF***调整后的神经追踪向量。*(A)* 显示了在特定解码器配置下（例如，*θ*频带、故事聆听、刺激包络、滞后不超过*250*毫秒）七种信噪比条件的神经追踪值。针对每种信噪比条件分别训练和测试一个解码器，且所有解码器使用相同的解码器配置，因此每种条件对应一个神经追踪值。信噪比为*–12.5*分贝时的神经追踪值作为噪声基线，静音条件不纳入后续分析。*(B)* 通过从其余信噪比条件下的神经追踪值中减去噪声基线（信噪比为*–12.5*分贝），计算校正后的神经追踪值。*(C)* 每种解码器配置都会在各种信噪比条件下产生一组校正后的神经追踪值。将所有解码器配置所对应的这些神经追踪值拼接成一个完整的校正后神经追踪向量。交替的灰色和白色背景标示了不同信噪比条件下对应于相同解码器配置的校正后神经追踪值。为清晰起见，图中未显示解码器类型参数。*(D)* 使用误差函数（*ERF*）将每位受试者的校正后神经追踪向量转换为*ERF*调整后的向量。

优化超参数 σ，该参数控制S形误差函数曲线的陡峭程度。该曲线将预期神经追踪性能 $\Psi(l)$建模为刺激信噪比水平 l的平滑Sigmoid函数，并以 μ（即阈值或中心信噪比值）为中心：

$$\Psi(l) = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{l-\mu}{\sigma\sqrt{2}}\right)\right], \quad (1)$$

按照公式(1)，将该ERF变换逐元素应用于NT向量的每个解码器配置（(图*3C*中的灰白背景方块）。重要的是，陡峭度参数 σ 是在该向量的所有解码器配置上全局拟合的。这意味着，虽然每个解码器配置都单独经过ERF函数变换，但所有解码器配置始终使用相同的 σ 拟合值。

在内层留一交叉验证循环的每一折中，测试一个不同的候选值（范围为1到10）。对于每个候选值，将调整后的神经追踪向量变换为ERF调整后的NT向量。随后，在内层训练集上训练支持向量回归（SVR）模型，并在验证受试者上进行评估。通过最小化内层循环中所有受试者的真实语音接收阈值（y_i）与预测语音接收阈值（$\hat{y}_i$）之间的平均绝对误差（MAE），确定最优 σ：

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \quad (2)$$

其中， N 是内层循环中的受试者数量。随后选取产生最低 MAE 的 σ，并将采用此最优 σ 的 ERF 应用于训练集内所有受试者调整后的 NT 向量。之后，在整个外层训练集上重新训练 SVR，最后在留出的测试受试者上进行测试，从而得到该受试者的预测语音接收阈值。

2.6.4. 模型性能

通过计算预测的 SRT 值与相应行为 SRT 测量值之间的相关性来评估模型性能。为确定观测到的相关性在统计学上是否显著，我们构建了模型性能的零分布（图 *A1*）。我们通过在参与者之间随机打乱行为 SRT，从而打破行为 SRT 与 ERF调整后的NT向量之间的关系，来生成该零分布。使用这些打乱后的数据，我们通过留一交叉验证重新训练 SVR 模型，得到一组预测 SRT，再将其与行为 SRT 进行相关性分析。重复这一过程 1000 次，得到相关值的零分布。由于相关系数的取值范围为 -1 到 +1，且呈非正态分布，因此我们对零分布应用 Fisher Z变换。所得的 Z 转换值近似服从正态分布，因此可以对零分布数据拟合高斯曲线并进行统计检验。具体而言，我们通过将观测相关性的 Fisher Z 转换值与零分布进行比较，计算双尾p值。

评估模型性能的另一种方法，是计算预测语音接收阈值与行为语音接收阈值之间的绝对差值，如 Lesenfants 等，2019 所示。理想情况下，该差值应小于 1 分贝，与行为矩阵任务本身观察到的重测变异性相当。

2.6.5. 评估数据缩减对模型性能的影响

在本分析中，我们旨在确定维持与使用每位受试者完整 27 分钟数据（不包括未使用的静音条件）相当的模型性能所需的最少脑电图数据量。为此，我们系统地减少了故事和矩阵EEG任务中的数据。

对于矩阵EEG任务，我们逐步将每个信噪比条件下的矩阵句数量从40句减少至5句，每次减少5句，同时保留完整的15分钟故事脑电数据。此次缩减会影响涉及矩阵句的训练数据集和测试数据集。

对于故事EEG任务，我们以3分钟为间隔，将时长从15分钟缩短至3分钟，同时保留完整的40个矩阵句。此次缩减仅影响解码器的训练数据，因为所有测试均仅使用矩阵EEG任务数据。

2.6.6. 与其他论文的性能比较

为了能够将模型性能与其他论文中获得的性能进行比较，我们提出使用归一化均方根误差（NRMSE）作为指标，因为它可以对不同论文中行为语音接收阈值（SRT）分布的变异性进行标准化：

$$\mathrm{NRMSE} = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}}{y_{\max} - y_{\min}} \quad (3)$$

其中， N 为观测数量， y_i 为行为语音接收阈值，$\hat{y}_i$ 为预测语音接收阈值， $y_{\max}$ 和 $y_{\min}$分别为行为语音接收阈值的最大值和最小值。

为获取其他研究中的行为语音接收阈值和预测语音接收阈值，我们使用 WebPlotDigitizer（Automeris）从已发表论文的图表中提取数据点（Rohatgi，2015年）。我们成功从以下论文中获取了这些数值：Accou 等人，2021年；Borges 等人，2025年；Vanthornhout 等，2018。

*2.6.7. SHAP*分析

采用 Shapley 加性解释（SHAP）分析（Lundberg 和 Lee，2017年）来解释每个误差函数调整后的 NT 值对语音接收阈值结果预测的贡献。为进一步探究模型的可解释性，我们根据参数组合对这些 NT 值进行分组，从而评估每个时滞、解码器类型、刺激特征、脑电图任务和频带的相对重要性。

3. 结果

3.1. SRT 预测

3.1. 行为

参与者在行为矩阵测试中的平均 SRT 为 -9.07 分贝，标准差为 0.56 分贝。

3.2. 模型结果

支持向量回归模型经过训练，以学习经误差函数调整后的NT向量与行为 SRT 之间的线性映射，在预测语音接收阈值与行为语音接收阈值之间达到了0.647的皮尔逊相关（$p < 0.001$）（见图*4A*）。预测语音接收阈值与行为语音接收阈值之间的绝对差值中位数为

0.29 分贝，最大差值为0.91 分贝，如图*4B*所示。

3.3. SHAP分析结果

从 SHAP分析（图*5A-B*）可以看出，尽管所有参数都有贡献，但 theta/delta频带、早期时滞、故事刺激和 SS解码器对模型预测的影响最大。

3.4. 数据缩减结果

数据缩减分析显示，如图*6A*所示，减少矩阵EEG任务中的矩阵句数量会立即导致模型性能下降。相比之下，缩短故事EEG任务的时长不会影响性能；相对平缓的曲线表明，当故事脑电数据从15分钟缩减至3分钟时，模型准确率仍保持稳定，见图*6B*。图*6* 还表明，结合 SI 和 SS 解码器是实现最佳模型性能的必要条件，其表现优于单独使用任一解码器。

3.5. 与其他论文的模型结果比较

表 *1*总结了各项研究的指标。尽管当前研究并未呈现最高相关性，但考虑到数据集的变异性，其归一化均方根误差最低，为0.19。

4. 讨论

我们提出了一种新方法——多解码器方法，用于根据受试者在聆听矩阵句和连续故事时获得的脑电图记录预测语音接收阈值（SRT）。该方法的核心是引入 ERF 调整后的 NT 向量，这些向量能够捕捉多种不同解码器配置下的 NT 信息。

具体而言，我们通过改变解码器的关键参数来构建 NT 向量，这些参数包括脑电图任务及相应的语音刺激（故事或矩阵句）、语音特征（包络或声学起始）、后整合窗口（最长至500 ms）、频带（θ 波、δ 波和宽频带），以及解码器类型（被试无关或被试特异）。由此得到高维 NT 向量。在扣除噪声水平并通过误差函数进行变换后，将所得的 ERF 校正的 NT 向量用于支持向量回归（SVR）模型，以预测个体的行为 SRT。

使用该模型，我们成功预测了全部 39 名参与者的 SRT；预测的基于脑电图的 SRT 与行为语音接收阈值之间的皮尔逊相关系数为 0.647，且具有统计学显著性。参与者预测 SRT 与行为 SRT 之间绝对差值的中位数为 0.29 分贝，标准差为 0.26 分贝，并且每位参与者的差值均低于 1 分贝。

我们方法的主要创新在于整合来自多种解码器和特征集的信息。相比之下，以往的方法通常只依赖单一类型的解码器和特征集（Borges 等人，2025年；Lesenfants 等，2019；Muncke 等人，2022；Vanthornhout 等，2018）。我们的方法无需进行 Sigmoid 拟合，而是采用支持向量回归（SVR）。这一改变克服了部分早期方法的局限，因为这些方法往往难以拟合所有受试者。Vanthornhout 等（2018）无法拟合 21% 的受试者（24 人中有 5 人）。Borges 等（2025）无法拟合 9% 的受试者（2 人在 22 人中），Accou 等（2021）无法拟合 20% 的受试者（20 人中有 4 人），而 Lesenfants 等（2019）平均无法拟合 17% 的参与者。不过，Lesenfants 等（2019）和 Muncke 等（2022）都提出了无需进行 Sigmoid 拟合的替代方法，从而能够为所有参与者预测语音接收阈值（SRT）。Lesenfants 等（2019）将时间对齐的语音特征与语谱图相结合（Di Liberto 等，2015），而 Muncke 等（2022）则采用了时间窗均方根方法。

在比较用于预测行为语音接收阈的基于脑电图的模型时，必须考虑行为语音接收阈值本身的准确性。对于质量良好的语音材料，行为重测差异通常约为 0.4 分贝（中位数），标准差为 0.7 分贝（Boon，2014年；Decruy 等人，2018年）。如表 *1*所示，不同数据集的行为 SRT 数据存在差异，这给跨研究比较模型性能带来了挑战。包括我们自己的研究在内，SRT 变异性较低的研究，在降低行为语音接收阈值与预测语音接收阈值之间的分贝差值中位数方面可能具有先天优势。例如，与变异性较高的数据集相比，在标准差较小的数据集上，始终预测平均 SRT 的模型会得到更低的分贝差值中位数。为实现更公平的比较，我们建议采用能对数据集变异性进行归一化的指标，例如 NRMSE。尽管此前的研究未直接报告 NRMSE 指标，我们还是对其已发表结果进行了追溯计算，以便进行标准化比较。重新计算所得的数值分别为 0.86（Vanthornhout 等，2018）、0.52（Borges 等人，2025年）和 1.26（Accou 等人，2021年）。相比之下，我们的多解码器方法取得了显著更低的 NRMSE，为 0.19。

支持向量回归模型的 SHAP分析表明，所有参数都会影响模型的整体性能；然而，某些组别——具体而言，θ 波和 δ 波频带、早期时滞、故事刺激以及 SS解码器——对模型预测的影响略大。早期时滞的重要性并不意外，因为本研究采用的两种语音特征——包络和声学起始——主要通过自下而上的方式加工，较少受到启动效应等自上而下效应的影响（Karunathilake et al., 2023）。δ 波和 θ 波频带在言语理解中都发挥着关键作用，反映了大脑对句子、短语、词、音节和音位等语言单位的层级加工（Ding et al., 2016）。这与自然言语的调制频谱相符，其中 δ 波（1–4 Hz）和 θ 波（4–8 Hz）范围内的能量尤为显著（Varnet et al., 2017）。

这进一步得到了先前研究结果的支持：在 Vanthornhout 等，2018 中，德尔塔频段的预测性最强，而在 Lesenfants 等，2019 中，西塔频段取得了最佳结果。在比较解码器类型时，出现了一个有趣的规律：SI 解码器在各信噪比下始终获得更高的 NT 值。这表明，其优势可能源于可获得更大量的训练数据，并且如果提供等量数据，SS 解码器也有可能达到相近的 NT 值。相比之下，SS 解码器在模型中产生更高的 SHAP 值，凸显了它们对预测个体行为结果的更大贡献。这一点进一步得到以下结果的支持：仅使用 SS 解码器的多解码器方法，其表现远优于仅使用 SI 解码器的多解码器方法。这表明，SS 解码器更善于捕捉独特的、个体特异性的神经特征，而这些特征对于预测每位受试者的语音接收阈值尤其具有信息价值。

在本研究中，我们最初使用了 29 分钟的脑电图

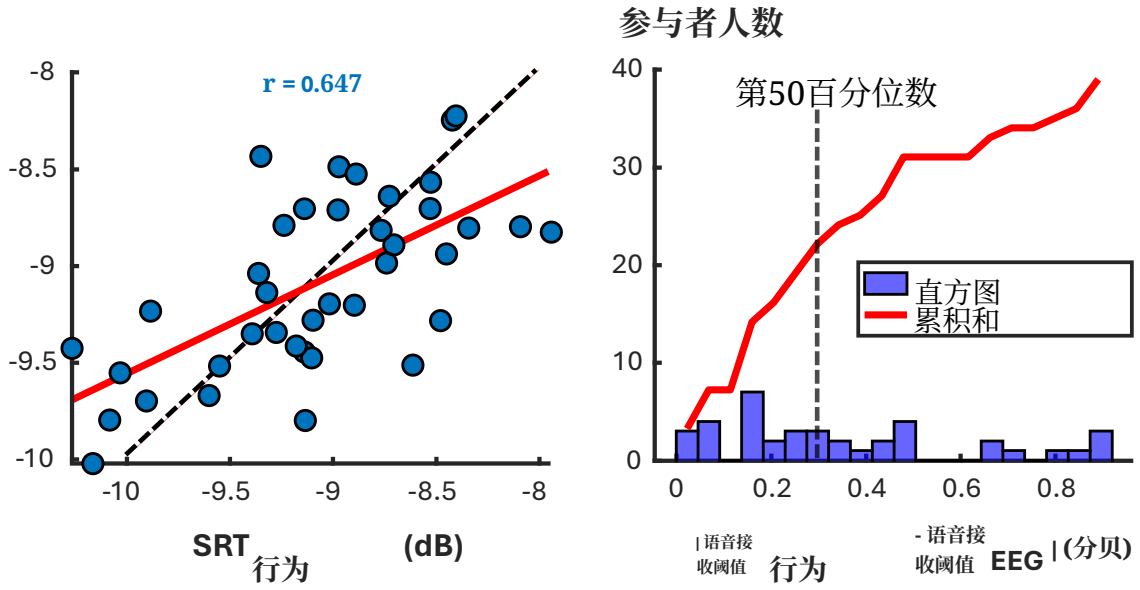

图 4： 模型性能结果。 *(A)* 行为语音接收阈值（$SRT_{behavioral}$）与预测语音接收阈值（SRT_{eeg}）之间的相关性（$p < 0.001$）。蓝色点表示受试者，红线为回归拟合线，虚线为恒等线。*(B)* $SRT_{behavioral}$ 与 SRT_{eeg}之间绝对差值的直方图。

表 1： 各研究中行为指标和模型性能指标的汇总，包括语音接收阈值（*SRT*）的均值和标准差（*SD*）、相关系数、绝对分贝差值的中位数和标准差，以及归一化均方根误差（*NRMSE*）。在此表中，*Sonck* 等，*2026* 指当前研究。

N	研究	平均语音接收阈值	SD 语音接收阈值	相关性	分贝差异中位数	分贝差异标准差	归一化均方根误差
24	Vanthornhout 等人 , 2018	-7.6	1.48	0.69	1.7	3.17	0.86
19	Lesenfants 等, 2019		0.95				
18	Muncke 等人 , 2022		1.2				
20	Accou 等人 , 2023	-8.73	0.83	0.59	3.64	1.68	1.26
22	Borges 等人，2025年	-5.34	0.6		0.38	1.45	0.52
39	Sonck 等人, 2026	-9.07	0.56	0.65	0.29	0.91	0.19

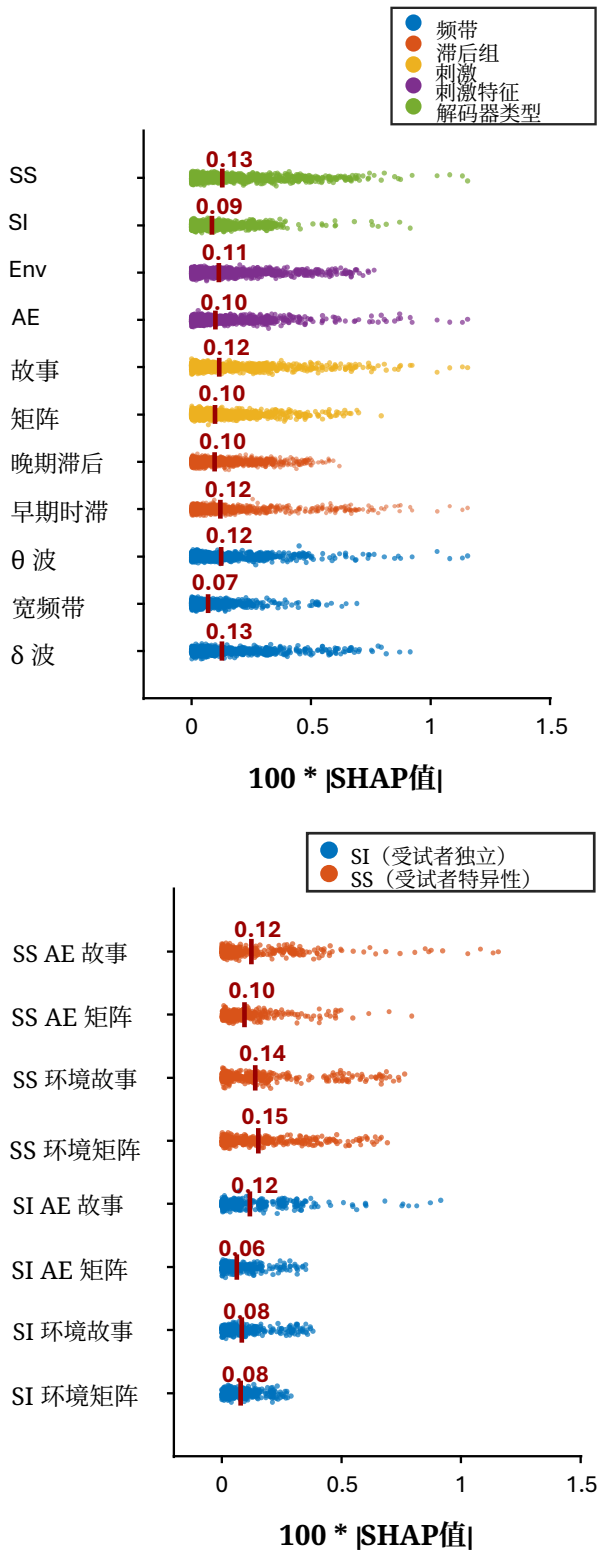

图5： ***SHAP***分析结果。 每个点表示解码器配置中某个特定参数对应的一个 *ERF*调整后的*NT*值的 *SHAP*值。*SHAP* 值分组如下： （*A*）按参数组对 *SHAP* 值进行整理。如果某个 *SHAP* 值对应多个参数（例如 *SS* 和 *Env*），则会出现在每个相关参数组中。（*B*）根据 *SS* 和 *SI* 解码器类型对 *SHAP* 值进行分组。

每位受试者的数据。不过，由于矩阵EEG任务中排除了静音条件，每位参与者的有效记录时间缩短至27分钟。我们的时间缩减分析表明，故事EEG任务的时长可从15分钟缩短至3分钟，同时不影响模型性能，从而将所需的脑电图总记录时间缩短至约15分钟。相反，不建议减少矩阵EEG任务中的矩阵句数量，因为这会导致模型准确率下降。因此，如果可以使用本数据集预训练的 SI 解码器，仅需记录15分钟脑电图，即可预测新参与者的语音接收阈值。虽然这一时长超过了标准行为语音接收阈值测试所需的时间，但对于无法进行行为评估的人群而言，这提高了测试的可行性。

因此，我们提出的方法显示出临床应用潜力，尤其适用于主动参与受限或无法主动参与的人群，并有助于评估各种助听器策略。为进一步提高生态效度，未来研究可省略矩阵句，完全依靠以不同信噪比水平呈现的有声书材料来预测语音接收阈，正如Borges等（2025）所建议的那样。此外，在听力障碍个体和老年人中验证多解码器方法，将确立其在更广范围语音接收阈上的可推广性，因为目前的研究结果仅基于正常听力参与者。

5. 结论

在本研究中，我们提出了多解码器方法，这是一种基于脑电图的方法，利用来自多种解码器及不同解码器配置的ERF校正的NT向量来预测语音接收阈。该方法能够为所有参与者实现准确的SRT 预测，归一化均方根误差为0.19。SHAP分析表明，尽管所有参数均对模型性能有所贡献，但某些组别——具体而言是θ波和δ波频带、早期时滞以及SS解码器——的影响略大。这些结果凸显了该方法的临床潜力，尤其适用于行为测试困难或不切实际的人群。此外，我们证明，利用预训练的SI解码器可将新参与者的脑电数据采集缩短至仅15分钟——包括3分钟的故事聆听，以及在六种信噪比条件下各使用40个矩阵句、共计12分钟——且不牺牲预测准确性。未来研究应在更多样化的临床人群中评估并验证该方法，包括听力损失人群。

6. 参考文献

Abramowitz, M., & Stegun, I. A. (1972, 12月). 《数学函数手册：公式、图表和数学表格》。Accou, B., Bollens, L., Gillis, M., Verheijen, W., Van hamme, H., & Francart, T. (2024年). SparrKULee：来自 KU Leuven 的语音诱发听觉反应资料库，包含 85 位参与者的脑电图。*Data*, *9*(8), 94. https://doi.org/10.3390/data9080094

Accou, B., Jalilpour Monesi, M., Van hamme, H., & Francart, T. (2021年). 在非线性分类范式下基于脑电图预测语音可懂度*。《神经工程学杂志》, *18*(6), 066008. https://doi.org/ 10.1088/1741-2552/ac33e9

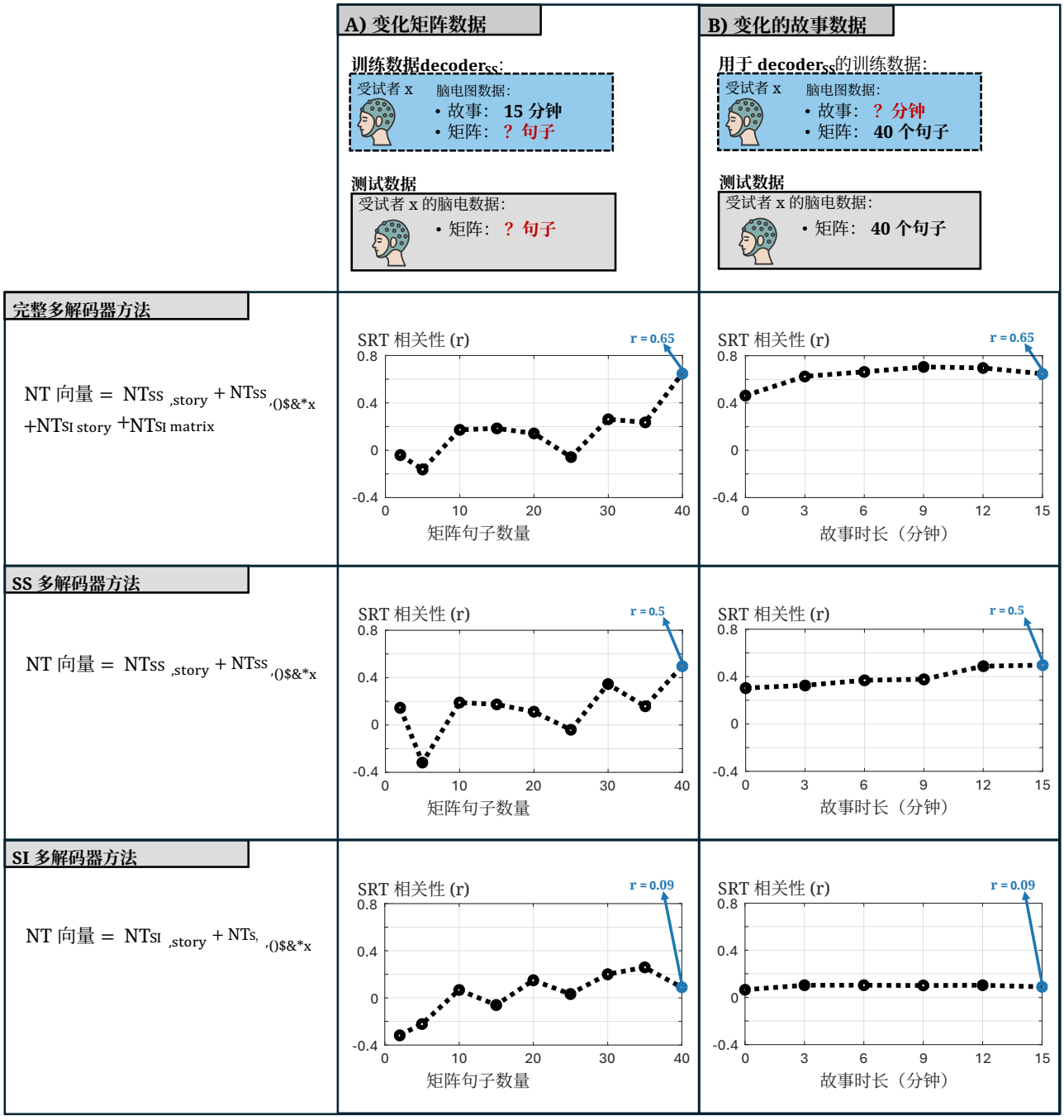

图6：脑电数据缩减结果。*(A)* 展示了在将故事脑电数据保持为 *15* 分钟不变的情况下，减少矩阵脑电数据量（以模型可用的矩阵句数量衡量）所产生的影响。句子数量不能少于 *2*，因为 *SS*解码器至少需要一个句子用于训练，另一个句子用于测试。在完整多解码器方法中，*NT*特征向量由 *SI* 解码器和 *SS* 解码器的输出构建。*SS* 多解码器方法仅使用 *SS* 解码器，而 *SI* 多解码器方法仅使用 *SI* 解码器。*(B)* 展示了在矩阵句数量保持为 *40* 不变的情况下，减少故事脑电数据量对模型性能的影响。

Accou, B ̇, Vanthornhout, J ̇, Hamme, H ̇ V., & Francart, T ̇ (2023). 使用 VLAAI 深度神经网络从脑电图中解码语音包络。《科学报告》, *13*(1), 812. https://doi.org/10.1038/s41598- 022-27332-2 Biesmans, W., Das, N., Francart, T., & Bertrand, A. (2017). 受听觉启发的语音包络提取方法，可改进鸡尾酒会情景下基于脑电图的听觉注意检测。《*IEEE* 神经系统与康复工程汇刊》, *25*(5), 402–412. https://doi.org/10.1109/TNSRE. 2016.2571900 Boon, S. (2014). 新荷兰语矩阵测试的多中心评估与验证 [硕士论文, 鲁汶大学]。未发表；无在线版本。Borges, H. B., Zaar, J., Alickovic, E., Christensen, C. B., & Kidmose, P. (2025). 基于脑电图连续语音包络重建的言语接收阈估计。《欧洲神经科学杂志》, *61*(6), e70083. https://doi.org/10.1111/ejn. 70083 Brodbeck, C., Hong, L. E., & Simon, J. Z. (2018). 连续语音从听觉表征到语言表征的快速转换。当代生物学：*CB*, *28*(24), 3976–3983.e5. https://doi.org/10.1016/j.cub.2018. 10.042 Decruy, L., Das, N., Verschueren, E., & Francart, T. (2018). 自我评估 B´ekesy 程序：一种测量连贯语篇可懂度的方法的验证。听力趋势, *22*, 2331216518802702. https://doi.org/10.1177/2331216518802702 Di Liberto, G. M., & Lalor, E. C. (2017). 在音位层面量化自然言语的皮层夹带：应用研究的方法学考量。听力研究, *348*, 70–77. https://doi.org/10. 1016/j.heares.2017.02.015 Di Liberto, G. M., O' Sullivan, J. A., & Lalor, E. C. (2015). 对语音的低频皮层夹带反映了音位级加工。当代生物学, *25*(19), 2457–2465. https://doi.org/10.1016/j.cub. 2015.08.030 Dieudonn´e, B., Decruy, L., & Vanthornhout, J. (2025). 语音包络的神经追踪可预测双耳去掩蔽。《欧洲神经科学杂志》, *61*(1), e16638. https://doi.org/10.1111/ejn.16638 Ding, N., Chatterjee, M., & Simon, J. Z. (2014). 对语音包络的稳健皮层夹带依赖于时频精细结构。神经影像, *88*, 41– 46. https://doi.org/10.1016/j.neuroimage.2013.10. 054 Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2016). 连贯言语中对层级语言结构的皮层追踪。自然神经科学, *19*(1), 158–164. https://doi.org/10.1038/nn.4186 Festen, J. M., & Plomp, R. (1986). 使用一台或两台助听器时的噪声中言语接收阈。《美国声学学会志》, *79*(2), 465–471. https://doi.org/10.1121/1.393534 Gillis, M., Van Canneyt, J., Francart, T., & Vanthornhout, J. (2022). 神经追踪作为评估听觉通路的诊断工具。听力研究, *426*, 108607 10 1016 2022 . https : / / doi . org / . / j . heares . . 108607

Gillis, M., Vanthornhout, J., & Francart, T. (2023). 听到了还是理解了？在可理解故事、不可理解故事和词表中对语言特征进行神经追踪。*eNeuro*, *10*(7). https://doi.org/10. 1523/ENEURO.0075-23.2023 Hagerman, B. (1982). 用于测试噪声中语音可懂度的句子。《斯堪的纳维亚听力学》, *11*(2), 79–87. https://doi.org/10.3109/01050398209076203 Ho, C.-H., & Lin, C.-J. (2012). 大规模线性支持向量回归。《机器学习研究杂志》, *13*(107), 3323–3348. Hoerl, A. E., & Kennard, R. W. (1970). 岭回归：非正交问题的有偏估计。《技术计量学》, *12*(1), 55–67. https : / / doi . org / 10 . 2307 / 1267351 Houben, R., Koopman, J., Luts, H., Wagener, K. C., van Wierin- gen, A., Verschuure, H., & Dreschler, W. A. (2014). 开发用于评估噪声中语音可懂度的荷兰语矩阵句测试。《国际听力学杂志》, *53*(10), 760–763. https://doi.org/10. 3109/14992027.2014.920111 Iotzov, I., & Parra, L. C. (2019). 脑电图可以预测语音可懂度。《神经工程学杂志》, *16*(3), 036008. https://doi.org/10.1088/1741-2552/ab07fe Karunathilake, I. M. D., Kulasingham, J. P., & Simon, J. Z. (2023). 语音可懂度的神经追踪指标：在保持声学特征不变的情况下操控可懂度。《美国国家科学院院刊》, *120*(49), e2309166120. https: // doi.org/10.1073/pnas.2309166120 Kluender, K. R., Coady, J. A., & Kiefte, M. (2003). 对语音感知变化的敏感性。《语音传播》, *41*(1), 59–69. https://doi.org/10.1016/S0167- 6393(02)00093-6 Koning, R., & Wouters, J. (2012). 起始增强在提高听觉假体语音可懂度方面的潜力。《美国声学学会杂志》, *132*(4), 2569–2581. https://doi.org/ 10.1121/1.4748965 Koning, R., & Wouters, J. (2016). 语音起始增强可改善人工耳蜗使用者在不利聆听条件下的可懂度。听力研究, *342*, 13–22. https://doi.org/10.1016/j.heares.2016.09. 002 K¨osem, A., Dai, B., McQueen, J. M., & Hagoort, P. (2023). 语音包络的神经追踪并不一定反映可懂度。神经影像, *272*, 120040 10 1016 2023 . https://doi.org/ . / j.neuroimage. . 120040 Lesenfants, D., Vanthornhout, J., Verschueren, E., Decruy, L., & Francart, T. (2019). 根据声学层面和语音层面语音表征的皮层追踪，预测个体语音可懂度。听力研究,*380*, 1–9. https://doi.org/10.1016/j.heares. 2019.05.006 Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., & Moore, B. C. J. (2006). 听力受损者的语音感知问题反映了其无法利用时间精细结构。《美国国家科学院院刊》, *103*(49), 18866–18869. https : / / doi . org / 10 . 1073 / pnas . 0607364103

Lundberg, S. M., & Lee, S.-I. (2017). 解释模型预测的统一方法。神经信息处理系统进展, *30*. Luts, H., Jansen, S., Dreschler, W., & Wouters, J. (2014). 弗拉芒语/荷兰语矩阵测试的开发与规范数据。在线获取：https://lirias.kuleuven.be/retrieve/293640. Machens, C. K., Wehr, M. S., & Zador, A. M. (2004). 使用自然声音测量皮层感受野的线性。《神经科学杂志》, *24*(5), 1089–1100. https://doi.org/10.1523/JNEUROSCI.4445-03.2004 Moore, B. C. J. (2008). 时间精细结构处理在正常听力者和听力受损者音高感知、掩蔽和语音感知中的作用。*JARO*：耳鼻喉科学研究协会期刊, *9*(4), 399–406. https: // doi.org/10.1007/s10162-008-0143-x Muncke, J., Kuruvila, I., & Hoppe, U. (2022). 利用脑电图对噪声中句子的反应预测语音可懂度。《神经科学前沿》, *16*. https://doi.org/10.3389/fnins.2022.876421 Ortiz Barajas, M. C., Guevara, R., & Gervain, J. (2021). 生命最初几个月语音包络跟踪的起源与发展。《发育认知神经科学》, *48*, 100915. https : / / doi . org / 10 . 1016/j.dcn.2021.100915 Peelle, J. E., & Davis, M. H. (2012). 神经振荡将言语节律传递至理解过程。《心理学前沿》, *3*. https://doi.org/10.3389/fpsyg.2012. 00320 Plomp, R. (1986). 用于描述听力受损者语音接收阈值的信噪比模型。《言语与听力研究杂志》, *29*(2), 146–154. https://doi.org/10.1044/jshr.2902.146 Reetzke, R., Gnanateja, G. N., & Chandrasekaran, B. (2021). 语音包络的神经追踪受到注意和语言经验的差异性调节。《脑与语言》, *213*, 104891. https://doi. org/10.1016/j.bandl.2020.104891 Rohatgi, A. (2015). Webplotdigitizer [版本3.9, 访问于11月 2025]。Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., & Eke- lid, M. (1995). 利用主要的时间线索进行语音识别。《科学》, *270*(5234), 303–304. https: // doi.org/10.1126/science.270.5234.303 Smith, Z. M., Delgutte, B., & Oxenham, A. J. (2002). 嵌合声音揭示了听觉感知中的二分现象。《自然》, *416*(6876), 87–90. https://doi.org/10. 1038/416087a Smola, A. J., & Sch¨ olkopf, B. (2004). 支持向量回归教程。《统计学与计算》, *14*(3), 199– 222. https://doi.org/10.1023/B:STCO.0000035301. 49549.88 Somers, B., Francart, T., & Bertrand, A. (2018). 一种基于多通道维纳滤波器的通用脑电图伪迹去除算法。《神经工程学杂志》, *15*, 036007. https : / / doi . org / 10 . 1088 / 1741 - 2552 / aaac92 Song, J., & Iverson, P. (2016). 连续言语的跨语言感知：对语音振幅包络的神经同步。《美国声学学会杂志》, *140*(4 Supplement), 3335. https:// doi.org/10.1121/1.4970635 van Dijkhuizen, J. N., Festen, J. M., & Plomp, R. (1989). 改变振幅-频率响应对听力受损者句子掩蔽语音接收阈值的影响。《美国声学学会志》, *86*(2), 621–628. https://doi.org/10.1121/1.398240 Vanthornhout, J., Decruy, L., & Francart, T. (2019). 任务和注意对语音神经追踪的影响。《神经科学前沿》, *13*. Vanthornhout, J., Decruy, L., Wouters, J., Simon, J. Z., & Fran- cart, T. (2018). 根据语音包络的神经同步预测语音可懂度。耳鼻喉科学研究协会期刊, *19*(2), 181–191. https : / / doi . org / 10 . 1007 / s10162 - 018-0654-z Varnet, L., Ortiz-Barajas, M. C., Erra, R. G., Gervain, J., & Lorenzi, C. (2017). 语音调制谱的跨语言研究。《美国声学学会志》, *142*(4), 1976. https://doi.org/10. 1121/1.5006179 Verschueren, E., Gillis, M., Decruy, L., Vanthornhout, J., & Francart, T. (2022). 在语速操控范式中，语音理解对声学和语言神经追踪产生相反影响。《神经科学杂志》, *42*(39), 7442–7453. https://doi.org/ 10.1523/JNEUROSCI.0259-22.2022 Verschuure, J., & van Benthem, P. P. (1992). 助听器对噪声环境中语音感知的影响。听力学：国际听力学会官方刊物, *31*(4), 205–221. https://doi.org/10.3109/ 00206099209081656 Zaar, J., Simonsen, L. B., & Laugesen, S. (2024). 一种用于预测佩戴助听器的听力受损者言语接收能力的时频调制测试。听力研究, *443*, 108949. https://doi.org/10. 1016/j.heares.2024.108949 Zou, J., Feng, J., Xu, T., Jin, P., Luo, C., Zhang, J., Pan, X., Chen, F., Zheng, J., & Ding, N. (2019). 听觉和语言因素对噪声环境中语音特征神经编码的贡献。神经影像, *192*, 66–75. https://doi.org/10.1016/j.neuroimage. 2019.02.047

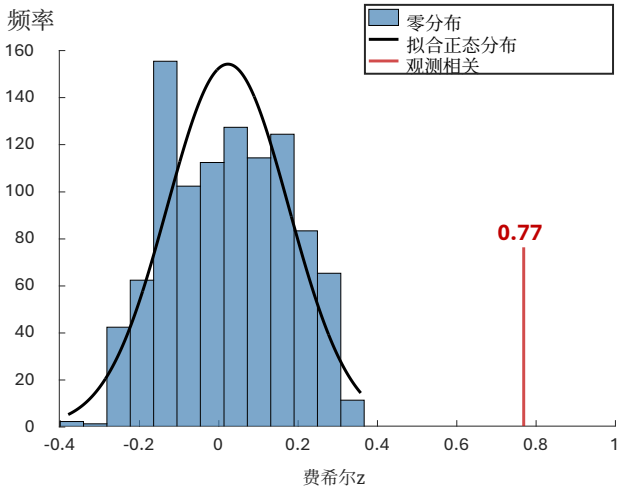

图 A1：零分布。经费希尔z变换后的零相关分布，均值为 *0.02*，标准差为 *0.15*。观测相关同样经过费希尔z变换，并根据未扰乱底层数据结构的模型计算得出（红线所示），其值显著落在该零分布之外。这表明存在超越偶然因素的、具有统计意义的关系。