

《心理学报》审稿意见与作者回应

题目：虚拟现实学习环境对外语词汇产出的影响：来自行为和脑电的证据

作者：刘聪；刘秋霞；祝梦蕊；焦鲁；王瑞明

第一轮

审稿人 1 意见：

本文使用脑电技术考查了虚拟环境下和词图匹配环境下二语学习效果的差异，从行为层面和神经生理层面进行了比较，结果表明，VR 环境下二语词汇习得效果要优于词图匹配环境。该文研究问题明确、研究思路清晰，实验任务设计合理、实验结果稳定可靠、结论的得出在某种程度上也是根植于研究结果。但该文也存在如下几个问题：

意见 1：该文的题目是 VR 环境下外语词汇习得的神经机制：基于词汇产出的证据。从题目上看，该文的核心问题是探究 VR 和词图匹配环境下外语词汇习得神经生理机制的差异，只是把词汇产出表现作为两种学习环境存在差异的考查依据。所以本文呈现的结果都是词汇产出层面，而探讨的神经生理活动，也仅仅是两种不同学习环境下词汇产出时行为和神经生理基础的差异，并没有真正探究外语词汇习得时两种环境下神经生理活动模式，因为学习过程和学习结果产出过程是不同的心理阶段，拥有不同的心理机制。所以该文有些文不对题。建议应当从“外语词汇不同习得环境对词汇产出影响”的角度撰写该文。该问题是本文存在的最大问题，也是致命问题；

回应：非常感谢审稿人的建议。根据您的建议，我们将标题修改为“虚拟现实学习环境对外语词汇产出的影响：来自行为和脑电的证据”。同时，我们修改了正文中的相关表述，明确本研究的核心问题是虚拟现实环境对词汇产出的影响，以便读者理解。其中最主要的修改内容举例如下：“本研究关注的问题是 VR 环境与图词环境下所学习外语词汇的产出效果是否存在差异？这种差异背后的神经机制是什么？”（第 14 页）。“为了检验 VR 学习环境对外语词汇产出的影响，本研究以传统图-词配对学习环境为参照，采用图片命名任务，一方面通过行为指标探讨 VR 环境和图-词环境下所学外语词汇的产出效果差异，另一方面结合相关脑电时域特征（P200 和 LPC 成分）和时频域特征（ μ 频段和 θ 频段）揭示这一影响背后的神经机制。”（第 15 页）

意见 2: 文中指出“目前相关的脑电证据均采用事件相关电位 (event-related potential, ERP) 分析脑电时域特征, 虽然区分了 VR 环境影响外语词汇习得的时间进程, 但忽略了词汇加工过程中的神经振荡活动 (neural oscillations) 的影响”, 但作者并没有指出已有研究忽略该影响带来的弊端或在揭示问题上局限, 需要作者进一步说明该局限的影响;

回应: 感谢审稿人的建议。我们对这一局限性进行了更详细的论述, 并在正文中进行了相应的补充和修改。修改内容如下: “其二, 在神经机制层面, 目前相关的脑电证据均采用事件相关电位 (event-related potential, ERP) 分析脑电时域特征, 忽略了词汇加工过程中的神经振荡活动 (neural oscillations)。传统的 ERP 分析主要反映锁时且锁相活动 (time-locked and phased-locked activity), 揭示认知加工的时间进程 (Bastiaansen et al., 2012); 时频分析则能够较好地反映出锁时非锁相活动 (time-locked and non-phased-locked activity), 揭示认知加工过程中神经振荡模式的变化 (Schneider et al., 2021), 但这些活动在经过叠加平均后往往被抵消, 无法通过 ERP 分析显现。因此, 有必要结合 ERP 和时频分析, 从而更好地揭示 VR 环境影响外语词汇产出的神经机制。” (第 14 页)

Bastiaansen, M., Mazaheri, A., & Jensen, O. (2012). Beyond ERPs: oscillatory neuronal dynamics. *The Oxford Handbook of Event-Related Potential Components*, 31-50.

Schneider, J. M., Abel, A. D., Momsen, J., Melamed, T. C., & Maguire, M. J. (2021). Neural oscillations reveal differences in the process of word learning among school-aged children from lower socioeconomic status backgrounds. *Neurobiology of Language*, 2(3), 372-388.

意见 3: 文中指出“以往记忆研究发现, 额区诱发的 P200 成分往往被认为与无关信息抑制能力相关, 例如随着工作记忆刷新功能的训练降低了 P200 波幅”, 审稿人任务该例子与语言加工任务关联性不强, 作为支撑二语加工脑成分指标的力度不够;

回应: 我们非常同意审稿人的建议, 我们已在正文中删除不恰当的例子, 并结合外语加工研究, 修改了 P200 指标的相关论述。修改内容如下: “P200 成分是刺激呈现后 200ms 左右出现的正走向波, 反映了刺激加工的早期阶段, 与选择性注意、工作记忆密切联系 (Hackley et al., 1990; Lijffijt et al., 2009)。外语加工的相关研究表明, 相较母语词汇阅读任务, 外语词汇诱发的 P200 波幅会增加, 这可能是因为外语词汇的加工难度更大、注意卷入程度更高 (Midgley et al., 2009)。” (第 14 页)

Hackley, S. A., Woldorff, M., & Hillyard, S. A. (1990). Cross - modal selective attention effects on retinal, myogenic, brainstem, and cerebral evoked potentials. *Psychophysiology*, 27(2), 195-208.

Lijffijt, M., Lane, S. D., Meier, S. L., Boutros, N. N., Burroughs, S., Steinberg, J. L., ... & Swann, A. C. (2009). P50, N100, and P200 sensory gating: relationships with behavioral inhibition, attention, and working memory. *Psychophysiology*, 46(5), 1059-1068.

Midgley, K. J., Holcomb, P. J., & Grainger, J. (2009). Masked repetition and translation priming in second

language learners: A window on the time - course of form and meaning activation using ERPs. *Psychophysiology*, 46(3), 551-565.

意见 4: “我们预期，如果 VR 环境能够促进新学习外语词汇的产出过程，那么相较于传统学习环境，VR 环境与所学外语词汇诱发的 P200 和 LPC 波幅出现变化，并且体现在频段和频段的神经振荡强度。”该实验预期不合理，应当建构定向研究假设，预期....环境下比....什么环境下震荡强或若，而不是有变化，而应当指出如何变化，是增强还是减弱等；

回应: 根据审稿人的建议，我们已经对相关内容进行详细论述，指出研究假设的方向，并在正文中进行了相应修改。修改内容如下：“结合具身认知等理论以及以往研究证据，本研究假设如下：(1) VR 学习环境会促进外语词汇产出，其命名正确率和命名速度优于图词环境；(2) VR 环境下外语词汇产出过程的提取难度更小，诱发的 P200 和 LPC 波幅小于图词环境；(3) VR 学习环境所呈现的感知-运动信息可能影响外语词汇产出，表现出 μ 波能量抑制效应和 θ 波能量增强效应。”(第 15 页)。

意见 5: 为了保证学习阶段与测试阶段的图片材料一致性，对于图词环境下所学词汇的测试图片与学习阶段相同，而 VR 环境下所学词汇的测试图片为 3D 客体的平面截图。两种测试环境下测试图片不一致，存在干扰变量，可能导致测试效果存在混淆；

回应: 感谢审稿人的建议。本研究之所以控制学习阶段和测试阶段的材料一致性，主要有两方面的原因。一是因为记忆领域的编码特异性原则强调，如果语义提取时的情境与最初编码时的情境相一致，那么其回忆效果会更好；二是参考了既往研究的实验设计，例如 Legault 等人 2009 年的研究中，四选一测试任务所呈现的图片材料与学习阶段一致。

当然，我们非常同意审稿人的观点，不同环境下的测试图片差异有可能会产生干扰效应。针对这一不足，我们已在正文中进行补充和阐述，并尝试在未来研究中深入探索。修改内容如下：“同时，鉴于记忆研究领域的编码特异性原则 (Tulving & Thomson, 1973)，当前研究尽可能控制了测试阶段与学习阶段的材料一致性，但这使得 VR 条件和图词条件的测试图片存在差异。未来研究应设置不同类型的测试图片，深入探究学习环境影响外语词汇产出的特异性与稳定性。”(第 25 页)

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, 80(5), 352-373.

意见 6: “首先删除反应错误数据、极端数据 (< 200ms 或 > 5000ms)，以及 $\pm 2.5SD$ 之外的反应时数据。”，此处需要给出删除数据的比例；

回应：根据审稿人的建议，我们已经将相关的数据删除比例补充在正文中（第 18 页）。

意见 7：“VR 学习环境的沉浸式和交互性特征为外语词汇习得提供了真实的感觉、运动等非语言信息，增强了多模态感官信息与语义网络的整合，这有助于被试在命名任务中实现更为准确的词汇提取”，该结论的得出么有结果支撑，因为作者考查的是产出阶段的神经活动，而不是学习阶段的脑神经活动模式，有些推论过度；

回应：非常感谢审稿人的建议，我们修改了正文中的相关表述。修改内容如下：“在本研究中，即时测试观察到的外语词汇产出差异可能与 VR 环境的沉浸式和交互性特征相关。相较传统图词学习环境，VR 环境所提供的视觉、听觉、运动等非语言信息可能增强了外语词汇知识的记忆和表征，使得被试在命名任务中实现更为准确的词汇提取。这一推测与既往实证研究 (Legault et al., 2019a) 和具身认知理论 (Barsalou, 2008) 相呼应。”（第 22 页）

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, 59(1), 617-645.

Legault, J., Fang, S. Y., Lan, Y. J., & Li, P. (2019a). Structural brain changes as a function of second language vocabulary training: Effects of learning context. *Brain and Cognition*, 134, 90-102.

意见 8：“本研究利用 VR 环境实现仿真化、互动式的词汇学习过程，促进了被试在编码阶段将非语言信息整合到语义网络，加深了词汇信息与非语言感觉特征（图片、动作等）的联结程度。”该结论也无法从实验结果中获得，更多是推测，缺乏足够的证据支持。

回应：非常感谢审稿人的建议，我们在正文中进行了修改。修改内容如下：“在本研究中，P200 波幅的变化可能与学习阶段的信息编码相关，相较于图词环境下的单模态信息编码，VR 学习环境可能有助于多模态信息与语义网络的整合，加深词汇信息与非语言感觉信息（图片、动作等）之间的联结。因此，当测试阶段呈现图片时，图词环境下所学词汇只能激活抽象的母语语言符号或单一的图片信息，而 VR 环境下所学词汇可能会激活更多的非语言信息，降低了词汇提取和产出的难度，表现为 P200 波幅减小。需要注意的是，由于本研究未能采集学习编码阶段的神经活动，这一推测值得进一步考察。”（第 22-23 页）

.....

审稿人 2 意见：

目前，虚拟现实对外语学习的影响是心理学、语言学和外语教育学等领域的热点议题。在这些研究中，该文探讨“语言产出”的视角，赋予其一定的创新性。

整体看，该文文献回顾详实，实验设计合理，统计方法运用得当，讨论较为深入。然而，该文仍有一定的提升空间。

意见 1: 摘要中, 行为学测量有即时和延时。而脑电测试, 没有提及。需要强调“即时测试”的表述。

回应: 感谢审稿人的建议。我们已经将摘要部分的相关表述修改为“即时测试阶段的脑电结果发现”(第 12 页)。

意见 2: 性别差异: 该文最终选取 29 名被试, 男性 4 人, 女性 25 人。其中, 男女性别不平衡, 至少在后述的结论中, 应该提及这一设计不足。

回应: 我们非常同意审稿人的建议, 已在正文的讨论部分补充这一不足之处。修改内容如下:
“由于被试取样范围的限制, 本研究仅让汉语母语者学习了德语, 且在一定程度上存在性别不平衡性。未来研究拟扩大取样范围, 结合人口学变量 (如性别、年龄) 和外语词汇类型 (如日语词汇、假词), 进一步探索 VR 环境影响外语词汇学习和产出的神经机制。”(第 25 页)

意见 3: 该文的实验材料为 40 个德语词汇。作者只提及“学习材料在不同环境之间进行平衡”。作者并没有告知读者, 如何平衡。此外, 40 个德语词汇, 与英语的同源性如何? 我们知道, 被试的第二外语均是英语, 而英语与德语拼写相似的词汇很多。

回应: 感谢审稿人的建议。关于学习材料平衡这一问题, 我们已经在正文中补充了相关内容, 以便读者理解。修改内容如下: “在学习阶段, 被试需要在两种不同环境下学习 40 个德语词汇。所有德语词汇分为两组, 其中一组 (20 个词汇) 在图词环境中进行学习, 另一组 (20 个词汇) 在 VR 环境中进行学习。学习材料在不同环境之间进行平衡, 即一半被试在 VR 环境中学习第一组德语词汇, 在图词环境中学习第二组德语词汇, 而另一半被试在图词环境中学习第一组德语词汇, 在 VR 环境中学习第二组德语词汇。”(第 16 页)

关于德语词汇与英语词汇的相似性这一问题, 我们在材料评定阶段, 招募了语言水平相当的汉语-英语双语者评价德语词汇是否与其所知道的任何汉语或英语词汇相似。结果表明, 本研究所使用的德语词汇与英语词汇并不相似。我们在正文中补充了这一细节, 内容如下:
“此外, 本实验招募了一组与正式实验被试英语水平相当的汉-英双语者, 要求他们评定所筛选的德语词汇是否与他们已知的汉语或英语词汇相似。基于五点量表 (1 = 非常不相似, 5 = 非常相似) 的评定结果表明, 本实验的德语词汇与汉语和英语词汇均不相似 ($p < 0.001$), 排除了已有的英语词汇知识对德语词汇学习的影响。”(第 16 页)

意见 4: 该文在讨论中, 在论述理论贡献时, 只运用了“具身认知理论”和“二语社会学习理论”。这样论述, 忽略了前面文献回顾专门提及的“认知负荷理论”。其实, 具身认知理论和“二语社会学习理论”都是具身认知研究的理论。如果想都用, 虽然有些勉强, 但是

否考虑补充一段关于“认知负荷理论”的分析？

回应：非常感谢审稿人的建议。首先，“具身认知理论”和“二语社会学习理论”都与具身认知观点密切联系，为了便于读者理解，我们已经在正文的讨论部分进行整合与修改。其次，根据审稿人的建议，我们围绕“认知负荷理论”补充了相关论述（第 23-24 页）。

修改内容如下：“最后，从理论贡献层面，一方面，本研究揭示了 VR 环境中所学外语词汇的产出优势，为具身认知理论 (Barsalou, 2008) 以及二语社会学习理论 (Li & Jeong, 2020) 提供了新的实证支持。VR 环境中的感知-运动交互使得学习者能够将外语词汇信息与多模态的感知觉-运动信息进行深度整合，增强词汇表征与感官特征之间的联结，从而促进外语词汇的习得与产出 (杨静, 王晓婵, 2023)。这一推论与以往相关脑成像研究相一致，例如，Legault 等人 (2019a) 发现 VR 环境中的外语学习效果与右侧顶上小叶皮层厚度密切相关，而该区域正是负责感知-动作整合与具身表征的关键脑区。这些研究证据共同表明，感知-运动交互可能是 VR 环境促进外语词汇产出的重要机制。另一方面，本研究有助于我们深入认识 VR 学习中认知负荷的作用模式。虽然认知负荷理论 (Swell, 2011) 指出学习环境中的多模态冗余信息可能增加外在认知负荷并干扰学习，但 VR 环境的沉浸感和交互性可以提升学习者的兴趣、动机与自我效能感，进而增加内在认知负荷并对学习产生积极影响 (Makransky & Petersen, 2021)。当内在认知负荷大于外在认知负荷时，学习者能够将认知资源更多地投入到学习内容本身的相关要素，从而形成有意义的生成式学习 (Mayer, 2005)。因此，我们推测 VR 环境的影响并非简单地增加或减少认知负荷，而是不同类型认知负荷共同作用的结果，未来研究有必要直接探讨不同类型认知负荷对 VR 环境下外语词汇学习的影响。”

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, 59(1), 617-645.

Legault, J., Fang, S. Y., Lan, Y. J., & Li, P. (2019a). Structural brain changes as a function of second language vocabulary training: Effects of learning context. *Brain and Cognition*, 134, 90-102.

Li, P., & Jeong, H. (2020). The social brain of language: grounding second language learning in social interaction. *npj Science of Learning*, 5(1), 8.

Makransky, G., & Petersen, G. B. (2021). The cognitive affective model of immersive learning (Camil): A theoretical research-based model of learning in immersive virtual reality. *Educational Psychology Review*, 33(3), 937-958.

Mayer, R. E. (2005). Cognitive theory of multimedia learning. *The Cambridge handbook of multimedia learning*, 41(1), 31-48.

Sweller, J. (2011). Cognitive load theory. In *Psychology of learning and motivation* (Vol. 55, pp. 37-76). Academic Press.

杨静, 王晓婵. (2023). 双语心理词汇通达模型: 70 年回顾与展望. *解放军外国语学院学报*, 46(02), 1-10.

意见 5: 结论中需要论述该文研究的不足。

回应: 结合您的第二条建议, 一方面我们在讨论部分详细论述了本研究的不足和局限 (第 25 页), 另一方面也在结论部分进行了修改和补充, 修改内容如下: “本研究在被试取样、材料匹配方面存在一定的不足, 未来需要结合不同的人口学和认知相关因素, 进一步探索 VR 学习环境对外语词汇产出的影响”。

.....

审稿人 3 意见:

论文聚焦于 VR 学习环境对外语词汇产出的影响, 以行为结合相关脑电时域特征和时频特征为指标, 比较了传统的图词配对学习与 VR 环境下的外语词汇产出。研究发现, 在行为层面, 只在正确率方面, 在即时测验时, 图词学习和 VR 学习存在差异, VR 环境的正确率更高。研究更丰富的结果主要是体现在脑电分析方面。研究发现, 相较于图词条件, VR 条件下所学外语词汇诱发了更小的 P200 成分和 LPC 成分。在脑电时频层面也有一些发现。研究的选题切合时代, 有较好的创新性。写作清晰, 文字流畅。但还需要有以下几个小的问题, 需要作者再进一步回应一下。

意见 1: 在引言部分第 5 页, 作者提到前人研究忽略了词汇加工过程中的神经震荡活动。前人忽略了, 并不代表这个就值得研究。建议在这里还要把神经震荡活动的意义再充分的阐述一下。

回应: 非常感谢审稿人的建议, 我们在正文中详细论述了神经振荡活动的意义。修改内容如下: “其二, 在神经机制层面, 目前相关的脑电证据均采用事件相关电位 (event-related potential, ERP) 分析脑电时域特征, 忽略了词汇加工过程中的神经振荡活动 (neural oscillations)。传统的 ERP 分析主要反映锁时且锁相活动 (time-locked and phased-locked activity), 揭示认知加工的时间进程 (Bastiaansen et al., 2012); 时频分析则能够较好地反映出锁时非锁相活动 (time-locked and non-phased-locked activity), 揭示认知加工过程中神经振荡模式的变化 (Schneider et al., 2021), 但这些活动在经过叠加平均后往往被抵消, 无法通过 ERP 分析显现。因此, 有必要结合 ERP 和时频分析, 从而更好地揭示 VR 环境影响外语词汇产出的神经机制。” (第 14 页)

Bastiaansen, M., Mazaheri, A., & Jensen, O. (2012). Beyond ERPs: oscillatory neuronal dynamics. *The Oxford*

Handbook of Event-Related Potential Components, 31-50.

Schneider, J. M., Abel, A. D., Momsen, J., Melamed, T. C., & Maguire, M. J. (2021). Neural oscillations reveal differences in the process of word learning among school-aged children from lower socioeconomic status backgrounds. *Neurobiology of Language*, 2(3), 372-388.

意见 2: 图词学习环境，相应的学习材料为黑白线条画。从跟这个 VR 环境匹配的条件来看的话，更真实的图片是不是会更加的跟这个 VR 条件匹配啊？这样 VR 与图词的差异，就不是材料差异带来的，更多的是操作以及环境沉浸的差异？

回应: 非常感谢审稿人的建议，我们在未来研究中将尝试以真实图片为参照，深入探索 VR 环境影响外语词汇产出的机制。同时，我们已经在正文中补充这一不足之处，修改内容如下：“本研究的 VR 学习环境虽然能够有效模拟真实环境，但 VR 环境的三维客体与图词环境的线条图片在色彩丰富度、视觉复杂性方面存在一定的差异。后续研究可以考虑采用真实图片作为对照条件，进而控制材料差异对 VR 环境下外语词汇学习的影响。” (第 25 页)

意见 3: 建议修改一下结论。结论说，本研究发现，相较传统的图词学习环境，沉浸式 VR 学习环境有利于外语词汇产出。这个呢，仅限于即时词汇产出这个层面。虽然有神经指标的证据，神经指标本身并不具有指向性，难以说明某个神经指标的高低对词汇产出而言是利还是有弊。所以神经指标的意义（指向性）需要跟行为做关联分析才能真正说明。

回应: 我们同意审稿人的观点，已将相关表述修改为“行为结果表明，相较传统的图词学习环境，VR 环境下所学外语词汇在即时测试阶段表现出一定程度的词汇产出优势。” (第 25 页)。

意见 4: 研究仅使用了德语，是否可以泛化到外语，需要进一步考虑。

回应: 我们同意审稿人的观点，已经在正文中补充这一不足之处，内容如下：“由于被试取样范围的限制，本研究仅让汉语母语者学习了德语，且在一定程度上存在性别不平衡性。未来研究拟扩大取样范围，结合人口学变量（如性别、年龄）和外语词汇类型（如日语词汇、假词），进一步探索 VR 环境影响外语词汇学习和产出的神经机制。” (第 25 页)

意见 5: 格式上，请作者注意一下中英文括号的使用问题。

回应: 非常感谢审稿人的仔细阅读，我们已经检查全文并进行修改。

第二轮

审稿人 2 意见: 已按要求修改，建议发表。

审稿人 3 意见：感谢作者的修改，我没有更进一步的意见啦。

编委意见：建议录用。

主编意见：同意外审和编委意见，建议录用。