

《心理学报》审稿意见与作者回应

题目：外语口语焦虑对言语互动质量的影响：fNIRS 超扫描研究

作者：徐楚言;朱麟;王芸萍;王瑞冰;刘聪慧

第一轮

审稿人 1 意见：

该文章使用基于近红外成像技术的超扫描技术，对口语焦虑如何影响言语互动质量以及言语互动过程中的脑间同步性展开了研究。该文章的研究思路和文字撰写都很清晰，两个实验的结果表现出了较好的可重复性。但是，该文章在立论上和分析上可能存在一定的提升空间。具体而言：

意见 1：文章在引言部分需要进一步澄清分析脑间同步性的合理性和必要性。

回应：感谢评审专家的建议。本研究探索外语焦虑对第二语言沟通质量的影响，而脑间同步性是预测与反映个体互动质量的可靠脑神经指标。为了更为清晰的呈现脑间同步性在本研究中的合理性和必要性，我们在正文中增加了该部分的论述。具体如下：

“分析 INS 可以加深我们对于言语互动的理解。首先，言语互动过程中的脑间同步性能够反映被试在当下的互动质量，如 Jiang 等(2015)采用无领导小组讨论的方法，发现在言语交流中，领导者-跟随者之间左侧颞顶联合区的 INS 显著高于跟随者-跟随者，这一区域与心理理论密切相关；相应地，相比起与跟随者-跟随者，跟随者-领导者的交流和关系也更加密切，因此更能形成有效的言语互动。此外，INS 也可以对个体的互动质量进行预测，如在师生互动的研究中，发现师生之间的脑间同步性可以预测学生的课堂参与度(Dikker et al., 2017)和学习程度(Zhu et al., 2021)。综上所述，我们认为在言语交流过程中，个体的 INS 可能与其言语互动的质量密切相关，对于研究第二语言的互动质量及其影响因素，是一个有效的脑神经指标。在已有研究中，言语交流相关的 INS 多发生在左侧前额叶(Jiang et al., 2012)和左侧颞顶联合区域(Hirsch et al., 2018)，这两个脑区与言语的加工和表达(张清芳, 杨玉芳, 2003; Kelsen et al., 2020)、心理理论密切相关(Sassa et al, 2007)，且也涉及到了第二语言的加工(Jeong et al., 2016)。综上，在言语沟通过程中，与人际神经同步相关的脑区多集中在额叶和颞顶区域，且这些脑区的脑活动耦合程度越高，倾听者对讲述者表达的内容理解程度也越高 (Stephens et al., 2010)。” (引言部分第四段)

意见 2：文章需要解释如何选定目前近红外测量所覆盖的脑区范围。而且，在现有的覆盖范围相对局限在左侧半脑的情况下，一些经典研究表明，外语加工可能会比母语加工更多地涉及右侧半脑。因此，现有的覆盖范围可能遗漏一些外语加工的机制。这一点需要进行进一步的解释和讨论。

回应：感谢评审专家的建议和新的视角。在选择近红外测量的脑区过程中，我们主要进行了如下考虑：言语互动过程包括两个部分，第一部分涉及言语的理解和产生，和额叶和颞叶皮层有关；第二部分涉及人际之间的心理理论，和颞顶联合区、内侧前额叶等区域有关。相关论述详见引言部分第三段。

另外，确如专家所言，右侧脑区对于外语的加工和学习起着重要的作用，但是我们使用的设备条件有限(日本岛津公司生产的功能性近红外光谱脑成像仪(LABNIRS 系统, Shimadzu Corp., Kyoto, Japan)), 对于每位被试, 仅有发射和接收光极各 7 个, 共 18 个通道, 难以覆盖全脑, 确实缺失存在遗漏重要脑区的可能。我们也已经将这一缺憾写入研究的局限部分, 具体如下(粗体字为新增论述):

“本研究存在以下不足之处。第一, 因硬件条件的限制, 本研究并没有记录被试间右脑的同步情况, 而右侧颞顶联合区(Zheng et al., 2018)及右侧额叶皮层(Baker et al., 2016)的脑间活动同步水平同样可能和言语互动有关; **还有研究发现相比起母语, 对外语的加工更多涉及到大脑右半球, 尤其是对于不够熟练的双语使用者(Weber-Fox and Neville, 1997), 如右侧额下回与外语的语音感知和学习有关(Qi et al., 2019)。**因此, **本研究可能存在某些重要脑区的遗漏, 未来研究可尝试覆盖更多的脑区。”** (4.3 第二段)

意见 3: 文章的研究重点是口语焦虑对于言语互动质量的影响。依照文章逻辑, 实验 2 由于评价压力的存在而可能会导致更高的口语焦虑。因此, 文章可能需要对实验 1 和实验 2 的数据进行直接的比较才能更好地回答文章所关心的问题。

回应: 感谢审稿专家的建议, 我们已在文中补充了相应的数据分析结果。结果显示, 实验 2(有评价)被试的自评外语焦虑水平显著高于实验 1(无评价); 但实验 2(有评价)被试的外语言语互动质量与实验 1(无评价)则无显著差异, 对于自评和他评的结果皆是如此。我们新增加了此部分内容, 具体内容如下:

“综合两个实验的结果, 可以发现, 不管外界评价是否存在, 被试的外语口语焦虑对其言语互动质量的影响模式是相同的。此处我们将两个实验的结果放在一起分析, 进一步探索外界评价压力对于实验结果的影响。

独立样本 t 检验结果显示, 实验 1(无评价)中被试的口语焦虑自评得分(3.74 ± 0.67)显著低于实验 2(有评价)的分数(4.24 ± 0.68), $t(61) = -2.91$, $p = 0.005$, Cohen's $d = 0.75$ 。这证明外界评价的引入确实有效唤起了被试更高的外语口语焦虑。但评价的存在并未显著影响被试的言语互动质量, 两个实验中被试的自评言语互动质量分数差异不显著, $t(61) = -0.77$, $p = 0.444$; 对于他评分数也是如此, $t(61) = 0.16$, $p = 0.875$ 。” (3.8.5 部分)

意见 4: 在表 1 中, 文章一共定义出 25 组脑区对的排布情况。但由于本文中的被试并不区分说者和听者的角色, 因此对角线上的脑区对(例如脑区对 2 和 6; 脑区对 3 和 11)是否可以考虑合并或者平均以后再进行分析。

回应: 非常感谢专家的建议。本研究没有对这种类型的脑区对取平均计算, 主要是参考了先前关于言语互动的研究的经验, 尤其是轮流言语互动的研究(Liu et al., 2019; Lu et al., 2021; Jiang et al., 2012)。

参考文献:

- Liu, W., Branigan, H. P., Zheng, L., Long, Y., Bai, X., Li, K., Zhao, H., Zhou, S., Pickering, M. J. & Lu, C. (2019). Shared neural representations of syntax during online dyadic communication. *NeuroImage*, 198, 63–72.
- Lu, K., Qiao, X., Yun, Q., & Hao, N. (2021). Educational diversity and group creativity: Evidence from fNIRS hyperscanning. *NeuroImage*, 243, 118564.
- Jiang, J., Dai, B. H., Peng, D. L., Zhu, C. Z., Liu, L., & Lu, C. M. (2012). Neural synchronization during

face-to-face communication. *Journal of Neuroscience*, 32(45), 16064–16069.

意见 5: 图 3 中, 本文是根据什么标准选定的 0.02-0.04Hz 和 0.14-0.2Hz 两个频带? 仅仅通过观察? 可能更严谨的做法是基于一些参数统计 (例如基于某个阈值) 或者非参数统计 (cluster permutation) 的方法来进行选定。

回应: 非常感谢评审专家的建议。本研究的频段选取主要依据前人的 fNIRS 超扫研究。首先, 参考先前关于高级社会认知过程的多项 fNIRS 研究, 选取可规避全局生理信号的较大频段(0.01~0.2Hz)(Jiang et al., 2012, 2015), 接着对频段内每一帧进行逐一计算, 寻找与本研究任务有关的频段范围(Liu et al., 2019; Pan et al., 2018; Xue et al., 2018)。其次, 本研究最关注被试在使用英语进行交流时的 INS, 所以在选取频段时, 将静息阶段与英语任务阶段的 INS 进行配对样本 *t* 检验, 并发现在英语任务阶段, 0.02~0.04Hz 及 0.14~0.20Hz 两个频段的 INS 显著高于静息阶段。因此, 这两个阶段最有可能和本研究关注的认知过程有关。最后, 这两个频段中, 0.02~0.04Hz 对应的时长约为 25~50 秒, 与本研究单张图片的叙述时长相近, 较可能是任务相关的频段; 0.14~0.20Hz 这一频段在先前研究中被发现可能与言语互动有关(Nozawa et al., 2016), 同样属于本研究关注的问题。所以, 本研究选择了这两个频段进行具体分析。以上频段选择的过程在许多认知神经科学的超扫研究中均被广泛使用, 如咨询对话(Zhang et al., 2018)、师生互动(Liu et al., 2019; Pan et al., 2018; Zheng et al., 2018)和多人头脑风暴(Lu et al., 2018)等。

参考文献:

- Jiang, J., Chen, C., Dai, B., Shi, G., Ding, G., Liu, L., & Lu, C. (2015). Leader emergence through interpersonal neural synchronization. *Proceedings of the National Academy of Sciences*, 112(14), 4274–4279.
- Jiang, J., Dai, B., Peng, D., Zhu, C., Liu, L., & Lu, C. (2012). Neural synchronization during face-to-face communication. *Journal of Neuroscience*, 32(45), 16064–16069.
- Liu, J., Zhang, R., Geng, B., Zhang, T., Yuan, D., Otani, S., & Li, X. (2019). Interplay between prior knowledge and communication mode on teaching effectiveness: Interpersonal neural synchronization as a neural marker. *Neuroimage*, 193, 93–102.
- Lu, K., Xue, H., Nozawa, T., & Hao, N. (2019). Cooperation makes a group be more creative. *Cerebral Cortex*, 29(8), 3457–3470.
- Pan, Y., Novembre, G., Song, B., Li, X., & Hu, Y. (2018). Interpersonal synchronization of inferior frontal cortices tracks social interactive learning of a song. *Neuroimage*, 183, 280–290.
- Xue, H., Lu, K., & Hao, N. (2018). Cooperation makes two less-creative individuals turn into a highly-creative pair. *Neuroimage*, 172, 527–537.
- Zhang, Y., Meng, T., Hou, Y., Pan, Y., & Hu, Y. (2018). Interpersonal brain synchronization associated with working alliance during psychological counseling. *Psychiatry Research: Neuroimaging*, 282, 103–109.
- Zheng, L., Chen, C., Liu, W., Long, Y., Zhao, H., Bai, X., Zhang, Z., Han, Z., Liu, L., Guo, T. & Chen, B. (2018). Enhancement of teaching outcome through neural prediction of the students' knowledge state. *Human Brain Mapping*, 39(7), 3046–3057.

意见 6: 本文在一开始选定的频率范围是 0.01-0.2Hz。现在的脑间分析的研究可能会关注更大范围的频带，上限可能会涉及到 0.5Hz 以上 (Long et al., Cereb Cortex, 2022; Li et al., Cereb Cortex, 2022)。而且从图 3 来看，许多通道对的频率到 0.2Hz 时仍然非常显著。因此，文章可以考虑将频率上限进一步延展。

回应: 感谢评审专家的建议。本研究没有考虑更高频段，主要是因为 0.2Hz 以上频段可能会受到呼吸等生理活动的噪声影响(Liu et al., 2019; Long et al., 2022)。此外，有研究者认为，在较大的频段范围中，仅有部分频段与**对话任务**有关，且低频段血氧信号对于反映脑间同步性比高频段更加可靠，也更可能与认知活动有关(Achard et al., 2006; Lu et al., 2019; Nozawa et al., 2016; Zuo et al., 2010)。因此，本研究没有选择更高的频段，也是为了能够更好地反映研究关注的问题。

参考文献:

- Achard, S., Salvador, R., Whitcher, B., Suckling, J., & Bullmore, E. D. (2006). A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs. *Journal of Neuroscience*, 26(1), 63–72.
- Liu, J., Zhang, R., Geng, B., Zhang, T., Yuan, D., Otani, S., & Li, X. (2019). Interplay between prior knowledge and communication mode on teaching effectiveness: Interpersonal neural synchronization as a neural marker. *Neuroimage*, 193, 93–102.
- Long, Y., Chen, C., Wu, K., Zhou, S., Zhou, F., Zheng, L., Zhao, H., Zhai, Y. & Lu, C. (2022). Interpersonal conflict increases interpersonal neural synchronization in romantic couples. *Cerebral Cortex*, 32(15), 3254–3268.
- Lu, K., Qiao, X., & Hao, N. (2019). Praising or keeping silent on partner's ideas: Leading brainstorming in particular ways. *Neuropsychologia*, 124, 19–30.
- Nozawa, T., Sasaki, Y., Sakaki, K., Yokoyama, R., & Kawashima, R. (2016). Interpersonal frontopolar neural synchronization in group communication: An exploration toward fNIRS hyperscanning of natural interactions. *Neuroimage*, 133, 484–497.
- Xue, H., Lu, K., & Hao, N. (2018). Cooperation makes two less-creative individuals turn into a highly-creative pair. *Neuroimage*, 172, 527–537.
- Zuo, X. N., Di Martino, A., Kelly, C., Shehzad, Z. E., Gee, D. G., Klein, D. F., Castellanos, F. X., Biswal, B. B., & Milham, M. P. (2010). The oscillating brain: Complex and reliable. *Neuroimage*, 49(2), 1432–1445.

意见 7: 0.14-0.20Hz 频带的信号在很大程度上会受到呼吸节律的影响，文章需要讨论这一点。

回应: 感谢专家的建议。在先前研究中，这一频段的信号也被认为有一定的意义，尤其是对于言语互动(Nozawa et al., 2016)。但确实无法完全排除呼吸节律的影响，我们已经在文章中补充讨论，具体如下：

“此外，也有部分研究表明，0.14~0.20Hz 频段也可能会受到呼吸的影响(Liu et al., 2019; Long et al., 2022)，且此频段结果在两个实验间不够稳定，因此对于本研究的这一结果应谨慎看待。未来需要对该频段信号进行针对性的深入研究，以考察其稳定性和有效性。” (4.2 第二段末尾)

意见 8：文章在一开始确定频带的时候使用的是英语对话-静息态的对比。那在这一频带上，中文对话-静息态的对比是否也显著？

回应：非常感谢评审专家的提醒。我们在文中增加了该问题的阐述。内容如下：

“因此，选取 0.02~0.04Hz、0.14~0.20Hz 为本研究的感兴趣频段。此外，在这两个频段上，中文对话和静息态的 INS 对比也是显著的。”(2.6.2 第二段末尾，粗体字为新增内容)

“此外，在该频段下，左侧缘上回、左侧颞中回与左侧额极在双语任务中的相干值均显著高于静息阶段(经 FDR 校正，all $p_s < 0.005$)，即上述脑区在双语任务中(相较于静息阶段)均出现了显著的脑间活动同步水平。” (2.7.3 第二段)

“此外，在上述两个频段下，左侧缘上回、左侧颞中回与左侧额极在双语任务中的相干值均显著高于静息阶段(经 FDR 校正，all $p_s < 0.005$)，即上述脑区在双语任务中(相较于静息阶段)均出现了显著的脑间活动同步水平。” (3.7.3 第三段)

意见 9：图 7 的 C/D 图的展示不是特别合理，因为本文在计算脑间同步性时已经将多个通道进行平均了。此时绘图通道层面的结果不是特别准确。

回应：感谢专家的建议，这里确实不太合适，我们已将通道层面的同步性脑图删去。

意见 10：表 2 和表 4 中的结果是否经过了多重校正？

回应：非常感谢评审专家的建议。表 2 和表 4(现表 5)中的结果未经过多重校正，我们使用了多种方法(包括：holm、hochberg、hommel、Bonferroni、BH、BY、FDR)对相关数据进行了校正，结果发现，表 2 仅有英语互动质量自评分和英语口语焦虑自评分相关系数显著，表 4(现表 5)仅有英语互动质量自评分和英语口语焦虑自评分、英语口语焦虑自评分和英语课堂焦虑总分相关系数显著，我们把经 Bonferroni 校正后的结果补充到了结果部分(详见 2.7.2 表 3 和 3.7.2 表 6)。

表 3 实验 1 中各变量的描述统计及相关分析结果(校正后)

变量	$M \pm SD$	1	2	3	4	5	6	7
1.英语课堂焦虑总分	3.03±0.43							
2.英语口语焦虑	4.01±0.62	0.03						
3.口语焦虑自评(英语)	3.74±0.67	0.33	0.36					
4.口语焦虑自评(汉语)	1.01±0.01	0.10	0.11	0.18				
5.言语互动质量自评(英语)	4.85±0.75	-0.37	-0.35	-0.77***	-0.20			
6.言语互动质量自评(汉语)	6.55±0.12	-0.11	-0.20	-0.03	-0.08	0.25		
7.言语互动质量他评(英语)	5.18±0.57	-0.29	-0.28	-0.30	-0.14	0.46	0.20	
8.言语互动质量他评(汉语)	6.53±0.40	0.18	0.19	0.38	0.10	-0.31	0.19	-0.24

注：*** $p < 0.001$ 。

表 6 实验 2 中变量的描述统计及相关分析结果(校正后)

变量	$M \pm SD$	1	2	3	4	5	6	7
1.英语课堂焦虑总分	3.07 \pm 0.39							
2.英语口语焦虑	4.29 \pm 0.76	0.57*						
3.口语焦虑自评(英语)	4.24 \pm 0.68	0.003	0.21					
4.口语焦虑自评(汉语)	1.07 \pm 0.10	-0.17	0.04	-0.04				
5.言语互动质量自评(英语)	5.07 \pm 1.36	-0.13	-0.15	-0.62**	0.18			
6.言语互动质量自评(汉语)	6.56 \pm 0.12	0.24	0.02	-0.01	-0.10	-0.12		
7.言语互动质量他评(英语)	5.16 \pm 0.61	-0.10	-0.37	-0.37	0.01	0.22	-0.20	
8.言语互动质量他评(汉语)	6.40 \pm 0.37	0.19	0.10	0.05	-0.31	-0.16	0.54	-0.23

注: * $p < 0.05$, ** $p < 0.01$ 。

审稿人 2 意见:

该研究采用功能性近红外光谱成像超扫描技术考察了英语交流期间的焦虑水平、互动质量以及脑间神经同步水平。结果发现,与汉语交流相比,英语交流出现了更高的口语焦虑和更低的互动质量,以及更低的脑间神经同步。其中,左侧缘上回的脑间神经同步水平在口语焦虑水平和言语互动质量之间存在部分中介效应。这些结果对于深入理解外语交流的认知过程和神经基础有重要价值。

总之,以往大部分研究主要探讨了母语交流。该研究将交流拓展到了二语,并揭示了二语特有的交流过程和神经基础,因此有重要意义。研究方法合理,结果和讨论也较为恰当。

以下建议供作者参考。

意见 1: 个别文字表述不够准确。例如,“言语互动往往需要两人或多人共同参与,在此过程中,个体表达自己的看法并跟随他人观点”第一个逗号应为句号。“但是很少有研究考察外语焦虑与言语互动质量之间的神经机制。”应为“之间关系的神经机制”,或“外语交流影响言语互动质量的神经机制”。另外,建议将“神经机制”改为“神经基础”或“神经过程”,因为此类研究尚难以触及机制层面。

回应: 非常感谢专家的指正,我们已在全文中做出修改。修改的细节如下:

“言语互动往往需要两人或多人共同参与。”(引言第一段,逗号改为句号)

“但是很少有研究考察外语焦虑影响言语互动质量的**神经基础**。”(引言第一段)

“综上,虽然大量的研究发现外语口语焦虑和口语成绩之间存在负性相关(Radić-Bojanić & Topalov, 2021),并对该机制进行了解释和理论构建(Eysenck et al., 2007; Eysenck & Calvo, 1992; Sarason, 1988),但大多局限在行为层面,对于其内部的**神经过程**,尤其是互动双方大脑同步的**神经基础**则很少关注。”(引言第二段)

“从多脑的角度拓展了言语交流的**神经基础**”(引言第四段)

“综上,本研究拟使用基于 fNIRS 的超扫技术,通过两个实验探究英语口语焦虑对英语言语互动质量的影响,以及相关的认知**神经基础**。”(引言第六段)

“以往很多研究表明英语口语焦虑是影响英语互动和学习的重要变量,但很少有研究考

察其**神经基础**。”(总讨论第一段)

“因此，在外语言交流中，谈话者左侧缘上回的 INS 可能是外语口语焦虑影响言语交流质量的重要**神经基础**。”(4.3 第一段)

“本研究使用近红外超扫技术，首次在真实的对话情境下，揭示了外语口语焦虑对言语互动质量的影响及相关脑区同步的**神经基础**”(4.4 第一段)

意见 2: 近红外光谱成像 (NIRS) 应为功能性近红外光谱成像 (fNIRS)，因为 NIRS 同时包括功能性和结构性两种。

回应: 非常感谢专家的指正，我们已在全文中做出修改。修改的细节如下：

“本研究使用**功能性**近红外成像超扫(Hyperscanning)技术”(摘要)

“基于**功能性**近红外光谱成像(**functional** Near Infrared Spectroscopy, fNIRS)的超扫描技术(hyperscanning)已经被广泛应用于社会认知神经科学相关研究中(Babiloni & Astolfi, 2014)”(引言第四段)

“本研究拟使用基于 **fNIRS** 的超扫技术”(引言第六段)

意见 3: 引用和参考文献不一致，引用不够严谨。例如，Jeong et al., 2018 在文献列表中为 2016。而且该研究并不是 hyperscanning，也没有分析 INS。同样，Stephens et al., 2010 也不是 hyperscanning，没有分析 INS。

回应: 非常感谢专家的指正，我们对这两篇文献的处理如下：

1. Jeong et al., 2016 的研究虽然不是超扫研究，但探究的是第二语言和外语焦虑的单脑神经基础，对于我们研究中脑区的选取有参考价值，故在文中引用；我们已经将与该文献相关的论述移到更合适的位置，即言语交流的单脑机制。具体如下：

“言语互动是一种复杂的认知过程，往往涉及两人或多人，需要个体能够理解他人所表达的内容并进行跟随才能顺利完成(Celce-Murcia, 2001)。这一过程不仅涉及个体自身的言语产生和接收，还需要理解他人表达的信息才能顺利推进(Kelsen et al., 2020; Schoot et al., 2016)，这两种认知活动涉及众多脑区。言语的产生和接收与大脑的语义系统、认知控制系统紧密相关(程士静，何文广, 2020)，主要涉及部分额叶(布洛卡区)和颞叶皮层(威尔尼克区)(刘丽红 等, 2004; 张清芳, 杨玉芳, 2003; Kelsen et al., 2020)；跟随与理解他人意图则涉及到对他人心理状态的预测和解释，即心理理论的运用(Corbetta et al., 2008)，这一过程主要涉及颞顶联合区、内侧前额叶、额极等脑区(Sassa et al, 2007)，这些脑区也都被证明和言语及社会互动有密切关系(Liang et al., 2022)。最近有研究者使用 **fMRI (functional magnetic resonance imaging, Jeong et al., 2016)** 考察了外语焦虑影响言语沟通的单脑神经基础，发现外语沟通任务会激活左侧缘上回，还发现眶额皮层(包括左侧脑岛)的激活会随着语言焦虑水平的增加而降低。”(引言部分第三段，加粗部分为修改内容)

2. Stephens et al., 2010 的研究确实不是严格意义上的超扫研究，但也涉及到言语的表达和理解，以及在此过程中多脑活动的关联，对于本研究的脑区选取和理论建构有一定的贡献。我们已经修改对这篇文献论述的措辞，具体如下：

“综上，在言语沟通过程中，与人际神经同步相关的脑区多集中在额叶和颞顶区域，且

这些脑区的脑活动耦合程度越高，倾听者对讲述者表达的内容理解程度也越高(Stephens et al., 2010)。”(引言部分第四段)

“Stephen 等人(2010)在关于言语互动的研究中发现，当说者用听者不熟悉的语言进行讲述时，仅在初级听觉皮层发现二者脑活动的耦合，和心理理论相关的脑区则并没有明显的耦合，表明听者如果讲述的内容存在理解困难，就无法完成信息的传递和理解。”(4.2 第一段)

意见 4: 直接将英语和汉语交流下的互动质量和脑同步进行比较是否合适？二者的熟练度肯定会有所不同，如何排除英语互动质量降低不是由于熟练度造成的？

回应: 非常感谢评审专家的质疑。在中英两种任务中，直接比较互动质量和脑同步，确实很难完全排除语言熟练度造成的影响，因为熟练度可能和语言焦虑共同影响言语互动质量。同时，熟练度和语言焦虑之间也存在关联，

许多研究都发现，个体的外语水平越低，语言焦虑也越强烈(Liu, 2016; Oteir & Al-Otaibi, 2019; Tang & Tian, 2019; 张日昇, 袁莉敏, 2004)。外语能力较高的个体更有信心，更少担心在使用外语时发生失败，焦虑水平也就更低。

虽然我们无法通过直接比较英语和汉语交流下的互动质量和脑同步，得出互动质量差异和脑同步的根源，但是可以通过中英两种任务的比较获得基础数据(可能和外语有关的脑区)。另外，本研究在分析外语口语焦虑和外语言语互动质量之间关系时，通过协变量控制了个体的第二语言的熟练度，具体见问题 3 的回应。

参考文献:

张日昇, 袁莉敏. (2004). 大学生外语焦虑、自我效能感与外语成绩关系的研究. *心理发展与教育*, 03, 56-61.

Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *The Modern Language Journal*, 70, 125-132

Liu, M. (2006). Anxiety in Chinese EFL students at different proficiency levels. *System*, 34(3), 301-316.

Oteir, I. N., & Al-Otaibi, A. N. (2019). Foreign language anxiety: A systematic review. *Arab World English Journal*, 10(3), 309-317.

Tang, M.L., & Tian, J. (2019, January). The Influences of EFL Graduate Students' Gender, Major and English Proficiency on FLA. In *2nd International Conference on Social Science, Public Health and Education (SSPHE 2018)* (pp. 40-43). Atlantis Press.

意见 5: 一个有趣的问题是，如何验证是外语焦虑影响了言语活动，而不是言语活动诱发了外语焦虑？如果在互动前就知道采用英语交流，且焦虑水平在互动前就高于汉语，则可以说焦虑影响了言语活动。否则似乎很难将二者的关系厘清。如果无法厘清，则需要问题提出、结果报告以及讨论等表述上避免使用包含因果关系的字句。

回应: 非常感谢评审人提出的新视角。我们在实验开始前告知被试将使用外语进行言语互动，且在实验开始前也对被试的口语焦虑进行测量，并发现此时被试的外语口语焦虑已经显著高于汉语口语焦虑，这一部分的结果，实验 1 见 2.7.1 第一段末尾，实验 2 见 3.7.1 第一段末尾；所以可以推测在言语互动开始前，被试的外语焦虑要高于说汉语引起的焦虑。

意见 6: 同样，脑同步的差异到底是语言熟练度造成的还是互动行为引起的，亦或是焦虑带来的？中介模型无法回答这一问题。

回应: 非常感谢评审专家的提问。同问题 1 一样，我们无法完全排除熟练度等因素的影响，但我们计算了两个实验中被试自评的英语口语能力分数与各个脑区 INS 的相关值，结果显示在两个实验中，均没有发现显著的相关性(见表 1)。

表 1 被试英语口语能力自评分与各脑区 INS 的相关分析结果			
实验 1			
	左侧缘上回 INS	左侧颞中回 INS	左侧额极 INS
英语口语能力自评分	-0.009	-0.002	0.008
实验 2(0.02-0.04Hz)			
英语口语能力自评分	-0.095	-0.19	-0.095
实验 3(0.14-0.20Hz)			
英语口语能力自评分	-0.092	-0.186	-0.077

此外，我们重新计算了中介模型，将被试自评的英语口语能力分数作为控制变量纳入模型中，结果显示中介模型仍然成立。

实验 1: 中介模型分析结果见表 2。模型直接效应(95% CI = [-1.20, -0.61]，效应值占比为 59.7%)和间接效应(95% CI = [-0.88, -0.21]，效应值占比为 40.3%)皆显著。具体中介路径见图 1。

表 2 实验 1 中左侧缘上回 INS 的中介模型分析结果(纳入控制变量)						
预测变量	结果变量					
	左侧缘上回 INS			言语互动质量自评分(英语)		
	β	SE	t	β	SE	t
口语焦虑自评分(英语)	-0.07(-0.72)	0.02	-4.81***	-0.54(-0.48)	0.16	-3.31**
左侧缘上回 INS				4.90 (0.45)	1.45	3.38**
R^2	0.44			0.60		
F	11.57**			22.41***		

注: ** $p < 0.01$, *** $p < 0.001$, 括号内为标准化系数

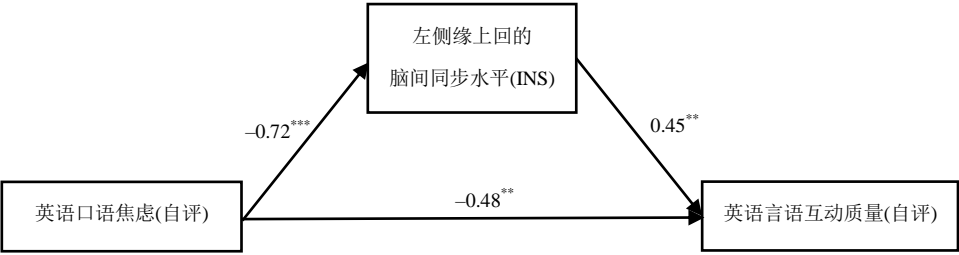


图 1 实验 1 中左侧缘上回的脑间同步水平的中介作用(标准化系数, 纳入控制变量)。注: ** $p < 0.01$, *** $p < 0.001$ 。

实验 2: 中介模型分析结果见表 3。0.02~0.04Hz 频段下, 模型直接效应(95% CI = [-1.60, -0.26], 效应值占比为 73.2%)和间接效应(95% CI = [-0.80, -0.06], 效应值占比为 26.8%)皆显著。0.14~0.20Hz 频段下, 模型直接效应(95% CI = [-1.60, -0.30], 效应值占比为 74.6%)和间接效应(95% CI = [-0.73, -0.07], 效应值占比为 25.4%)皆显著。具体中介路径见图 2。

表 3 不同频段下中介模型的分析结果(纳入控制变量)

不同频段	结果变量					
	左侧缘上回 INS			言语互动质量自评得分(英语)		
	<i>B</i> (β)	<i>SE</i>	<i>t</i>	<i>B</i> (β)	<i>SE</i>	<i>t</i>
0.02~0.04Hz						
口语焦虑自评得分(英语)	-0.03(-0.48) **	0.04	-2.80	-0.93(-0.46) **	0.33	-2.85
左侧缘上回 INS				11.98(0.35) *	5.42	2.21
<i>R</i> ²	0.23			0.49		
<i>F</i>	8.46**			8.20***		
0.14~0.20Hz						
口语焦虑自评得分(英语)	-0.02(-0.44) *	0.01	-2.51	-0.95(-0.47) **	0.32	-3.01
左侧缘上回 INS				13.57(0.37) *	5.75	2.36
<i>R</i> ²	0.20			0.50		
<i>F</i>	3.30*			8.57***		

注: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, 括号内为标准化系数。

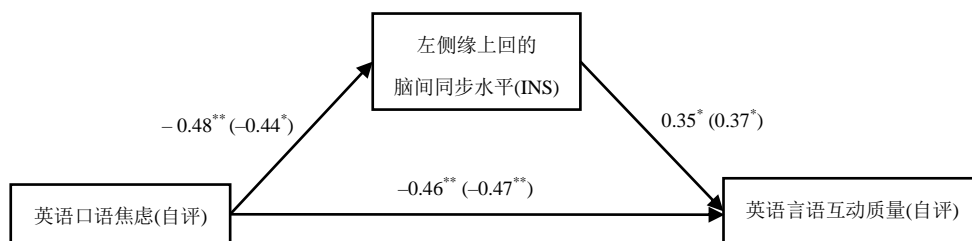


图2 实验2中0.02~0.04Hz(0.14~0.20Hz)左侧缘上回的脑间同步水平的中介作用(标准化系数,(纳入控制变量))。注: * $p < 0.05$, ** $p < 0.01$

值得注意的是,在本研究中,我们没有对被试的语言熟练度进行客观的测量,只让被试对自己的英语能力进行了简单的自评。相比起系统的语言能力测评,这种简单的自评分可能并不能非常完整地反映被试的英语能力。该不足已经在讨论不仅进行了阐述:

“第四,本研究虽然在中介模型中对被试的外语熟练度进行了控制,但使用的是被试自评的英语能力分数。相比起系统的语言能力测评,这种简单的自评分可能并不能完整地反映被试的英语能力。未来研究可以使用更专业的测评手段,如教育部和国家语言文字工作委员会 2018 年发布的《中国英语能力等级量表》,对被试的英语能力进行更客观的评价。”(4.3 最后一段)

另外,关于脑同步的差异是否为互动行为造成,我们更倾向于认为脑同步可能为互动行为的原因;但是,也不排除互动行为是脑同步的原因,亦或者脑同步和互动行为会受到第三方变量的影响。如想理清这些变量的因果关系,可能需要其他的研究手段,例如经颅磁刺激(TMS)和经颅直流电刺激(tDCS),通过刺激特定的脑区,以探索脑与行为之间的因果模式。

意见 7: 被试的外语水平均为自评。是否有更为客观的评价? 另外自评与他评是否一致? 每一对两名被试的外语水平是否有差异? 被试的外语水平与外语焦虑、言语互动质量是否有关?

回应: 非常感谢评审人的意见和新视角。

(1) 被试的外语水平均为自评。是否有更为客观的评价?

回复: 除了让被试对自己英语能力的四个维度进行自评外,虽然没有对被试的英语能力进行系统的测评,但我们搜集了被试的英语四六级分数,作为被试英语能力的客观指标。

(2) 另外自评与他评是否一致?

回复: 实验 1 中的英语言语互动自评与他评分数显著相关: $r = 0.47$, $p = 0.006$; 但实验 2 中的英语言语互动自评与他评分数相关性没有达到显著性水平: $r = 0.21$, $p = 0.258$ 。由此可见,两个实验中的自评和他评分数一致性不是很高,可能也正是因为,本研究的中介模型仅当因变量为他评分数时成立。英语口语能力的自评和他评分数可能受到不同因素的影响,他评分数往往仅依据于被评价者的外在表现,而自评分数还会受到个体内在心理状态的影响(Dong, et al., 2022),造成二者出现不一致的现象。

(3) 每一对两名被试的外语水平是否有差异?

回复: 对两个实验中每一对两名被试的英语水平自评分(四个维度)进行独立样本 t 检验,

四个维度的分数差异均不显著。对于本研究中最关注的口语能力，实验 1 中： $t(64) = -0.80$ ， $p = 0.425$ ；实验 2 中： $t(58) = -0.37$ ， $p = 0.712$ 。由此可见，每一对的两位被试外语水平差异不大。此外，每一对两名被试的口语焦虑自评分数差异也没有达到显著水平(实验 1： $t(64) = -0.38$ ， $p = 0.704$ ；实验 2： $t(58) = -0.04$ ， $p = 0.967$)。

(4) 被试的外语水平与外语焦虑、言语互动质量是否有关？

回复：我们使用被试在实验开始前自评的英语口语能力分数进行计算。

英语口语能力和口语焦虑自评：实验 1 中，二者有显著的相关性： $r = -0.34$ ， $p = 0.005$ ；实验 2 中二者的相关性呈边缘显著， $r = -0.23$ ， $p = 0.077$ 。

英语口语能力和言语互动质量：实验 1 中，英语口语能力和自评分相关性不显著， $r = 0.24$ ， $p = 0.171$ ；但他评分相关性显著， $r = 0.49$ ， $p = 0.004$ ；实验 2 中，英语口语能力与自评、他评分数均不显著，对于自评分数， $r = -0.03$ ， $p = 0.896$ ；对于他评分数， $r = -0.03$ ， $p = 0.871$ 。

由此可见，在实验开始前收集的被试的英语水平数据与其外语焦虑、言语互动质量均存在相关性，但相关关系不够稳定，可能和被试数量较少有关。

意见 8： 实验 2 有评价下的焦虑高于实验 1 无评价，但是并没有做统计比较，因此得出评价的影响效应这一结论并不可靠。

回应：感谢审稿专家的建议，我们已在文中补充了相应的数据分析结果。结果显示，实验 2(有评价)被试的自评外语焦虑水平显著高于实验 1(无评价)；但实验 2(有评价)被试的外语言语互动质量与实验 1(无评价)则无显著差异，对于自评和他评的结果皆是如此。我们新增加了此部分内容，具体内容如下：

“综合两个实验的结果，可以发现，不管外界评价是否存在，被试的外语口语焦虑对其言语互动质量的影响模式是相同的。此处我们将两个实验的结果放在一起分析，进一步探索外界评价压力对于实验结果的影响。

独立样本 t 检验结果显示，实验 1(无评价)中被试的口语焦虑自评得分(3.74 ± 0.67)显著低于实验 2(有评价)的分数(4.24 ± 0.68)， $t(61) = -2.91$ ， $p = 0.005$ ，Cohen's $d = 0.75$ 。这证明外界评价的引入确实有效唤起了被试更高的外语口语焦虑。但评价的存在并未显著影响被试的言语互动质量，两个实验中被试的自评言语互动质量分数差异不显著， $t(61) = -0.77$ ， $p = 0.444$ ；对于他评分数也是如此， $t(61) = 0.16$ ， $p = 0.875$ 。” (3.8.5 部分)

第二轮

审稿人 1 意见：作者很好地回答了我的问题，同意发表。

审稿人 2 意见：作者已经回答了我的问题，没有进一步意见了。

编委复审：已阅，同意发表。

主编终审：同意外审和编委意见，建议录用。