

《心理学报》审稿意见与作者回应

题目：音乐训练与抑制控制的关系：来自 ERPs 的证据

作者：陈洁佳、周翊、陈杰

第一轮

编委意见：

该研究使用两个经典的认知控制范式进行了比较简单的 ERP 分析，考察音乐训练经验对其的影响，就技术方法而言并没有大的创新性。

意见 1: 作为一个横断性研究，音乐训练部分的分析显得比较粗糙，研究结果并不能做出“音乐训练提升了反应抑制中冲突监控和运动抑制能力”这样的因果推断，文章的陈述需要进行大的修改。

回应: 感谢您的建议。根据您的意见，我们修改了全文关于因果关系的陈述，并在讨论部分对横断研究的局限性进行了说明以及对未来研究进行了展望。

修改如下（见正文第 11 页，第 2 段）：

“值得注意的是，本研究采用的是音乐训练组和控制组进行对比分析的横断设计，不能得出音乐训练与抑制控制之间的因果关系，同时也不能完全控制两组被试固有的个体差异，例如人格特质和认知能力等(Corrigall, Schellenberg, & Misura, 2013; Schellenberg, 2015)。因此，在本研究中我们也尽可能控制了对抑制控制有一定影响的额外变量，例如年龄、智力和家庭经济地位等。未来研究可以通过纵向设计来考察音乐训练和抑制控制的因果联系。”

相关参考文献：

Corrigall, K. A., Schellenberg, E. G., & Misura, N. M. (2013). Music training, cognition, and personality. *Front Psychol*, 4, 10.

Schellenberg, E. G. (2015). Music training and speech perception: a gene-environment interaction. In E. Bigand, B. Tillmann, I. Peretz, R. J. Zatorre, L. Lopez & M. Majno (Eds.), *Neurosciences and Music V: Cognitive Stimulation and Rehabilitation* (Vol. 1337, pp. 170-177). Oxford: Blackwell Science Publ.

意见 2: 既然是研究音乐训练效应，作者为何选择视觉而非听觉 Go-go，听觉 stroop 任务？

对比视觉和听觉任务是否能更好地回答领域一般性 vs 领域特异性问题？

回应：感谢您宝贵的意见。以往探讨音乐训练和抑制控制关系的研究主要采用视觉任务，并且多见于行为研究，因此，我们研究的主要目的是采用 ERP 技术和视觉任务，以期通过电生理指标来揭示音乐训练和抑制控制的关系，这样也能更好的与以往视觉任务研究进行对比。您提的这个建议非常好，引入了一个更为有趣、有价值的研究问题。根据您的意见，我们进一步查阅了相关文献，据我们所知，目前只有两项研究同时考察了音乐训练与视觉、听觉抑制控制能力的关系（Bialystok & Depape, 2009; Slevc, Davey, Buschkuehl, & Jaeggi, 2016），这两项研究均为行为研究，且发现了不一致的结果。Bialystok 和 Depape 的研究采用了音乐训练组与控制组的对比分析，结果发现音乐训练组只在听觉任务上比控制组表现出更强的抑制控制能力；Slevc 等人的研究采用音乐能力量表测量了所有被试的音乐能力，然后通过相关和回归分析发现音乐能力与视觉和听觉通道的抑制控制能力并没有显著的相关关系。并且这两项研究的听觉任务都采用 Stroop 任务，而视觉任务都采用 Simon 任务。尽管这两项研究采用了相同的实验任务，但是它们考察音乐训练效应的方法不一样，Bialystok 和 Depape 的研究是通过组间对比分析，Slevc 等是通过音乐能力与抑制控制能力的相关和回归分析，这可能是导致两个研究结果不一致的原因。此外，Stroop 任务和 Simon 任务具有一定的差异性。因此，这两项研究也不能较明确回答领域一般性 vs 领域特异性问题。未来研究可以采用视觉 Stroop 和听觉 Stroop 任务，或者视觉 Go/No-go 和听觉 Go/No-go 任务的对比分析来考察这个问题。

根据您的意见，我们在讨论部分对此问题进行了展望。

修改如下（见正文第 11 页，第 2 段）：

“最后，本研究仅从视觉通道上考察音乐训练和抑制控制的关系。目前，在一个实验里同时考察音乐训练与视觉、听觉抑制控制能力关系的研究还很少（Bialystok & Depape, 2009; Slevc, Davey, Buschkuehl, & Jaeggi, 2016）。因此，未来研究可以同时采用视觉和听觉 Stroop 任务，以及视觉和听觉 Go/No-go 任务来考察，从而更好地回答音乐训练和抑制控制的关系是领域一般性还是领域特异性。”

相关参考文献：

Bialystok, E., & Depape, A. M. (2009). Musical expertise, bilingualism, and executive functioning. *Journal of Experimental Psychology Human Perception & Performance*, 35(2), 565-574.

Slevc, L. R., Davey, N. S., Buschkuehl, M., & Jaeggi, S. M. (2016). Tuning the mind: Exploring the connections between musical ability and executive functions. *Cognition*, 152, 199-211.

意见 3: ERP 成分与行为指标之间有何对应关系? 两个任务间有无相关性?

回应: 根据您的意见, 我们对 Go/No-go 和 Stroop 任务中行为指标和 ERP 成分之间的相关关系, 以及两个任务行为指标之间的相关关系进行了分析。结果发现:

(1) 在 Go/No-go 中, 被试的 d' 分数与 N2 差异波 (No-go 减 Go 条件) 和 P3 差异波 (No-go 减 Go 条件) 均无显著相关关系 ($ps > 0.1$); 在 Stroop 任务中, 在正确率和反应时指标上的干扰效应 (不一致减一致条件) 和 N450 差异波 (不一致减一致条件)、SP 差异波 (不一致减一致条件) 均无显著的相关 ($ps > 0.1$)。

(2) 被试的 Go/No-go 任务 d' 分数与 Stroop 任务中正确率上的干扰效应和反应时上的干扰效应均无显著相关 ($ps > 0.1$)。这个结果在一定程度上说明了反应抑制和冲突控制是抑制控制中两个不同的认知过程。

由于这些指标之间的相关都不显著, 并且这些指标之间的关系不是本研究所关注的核心问题, 本文的篇幅已较长, 所有我们就没有在正文中呈现这些结果了。希望您能采纳我们的意见。

意见 4: 被试的音乐训练经历和其他信息需要更详细的介绍 (最好列表), 包括每个人的训练起始年纪、训练时长 (年)、训练类型、训练频率 (每周练习时间), 智力 (研究发现音乐家与非音乐家在智力上存在差异, 且可解释其他一些认知测试的差异)、社会经济地位。

回应: 根据您的意见, 我们把每位音乐组被试的音乐训练经历主要包括训练起始年龄、训练时长、所学器乐类型、每周练习时间和每位被试的个人信息包括年龄、社会经济地位、智力, 通过表格汇总呈现在文中 (见表 1)。另外, 我们最初没有测量被试的智力水平, 根据您的意见, 我们又单独联系了每位被试, 采用瑞文高级智力测验简化版, 通过电子问卷的方式测量了被试的智力水平 (Arthur & Day, 1994; 张政华, 韩梅, 张放, 李卫君, 2020), 结果显示两组被试在智力水平上没有显著差异 $t(44) = -1.37, p = 0.177$ 。

新增加的表 1 如下 (见正文第 19 页):

表 1 被试人口学资料

编号	年龄 (岁)		社会经济地位		智力		训练开始年龄 (岁)		训练时长 (年)		乐器 (首个为主修乐器)	每周练习时间 (小时)		
	M	C	M	C	M	C	M	C	M	C				
1	19	23	3.5	1.5	10	9	5	无	8	无	小提琴, 吉他	无	18	无
2	18	18	3.0	2.5	10	9	9	无	8	无	古筝, 钢琴, 陶笛	无	5	无
3	18	18	3.5	2.5	11	7	5	无	8	无	钢琴	无	6	无
4	22	20	3.5	3.5	9	12	6	无	15	无	小提琴, 钢琴	无	9	无
5	23	19	3.0	2.5	11	7	6	无	12	无	钢琴, 小提琴	无	14	无
6	22	23	2.0	4.0	7	10	12	无	8	无	钢琴	无	10	无
7	20	20	3.0	2.0	11	9	10	无	8	无	古筝, 钢琴, 手风琴	无	5	无
8	20	17	2.5	3.5	7	6	5	无	14	无	钢琴	无	13	无
9	21	19	3.0	2.0	3	12	5	无	15	无	钢琴	无	16	无
10	19	19	2.0	3.0	8	7	7	无	11	无	古筝, 钢琴	无	12	无
11	20	18	3.5	3.0	10	6	11	无	8	无	钢琴	无	10	无
12	20	20	2.0	2.5	6	10	10	无	9	无	大提琴	无	9	无
13	20	19	3.5	3.5	11	11	5	无	14	无	中提琴, 小提琴, 吉他	无	8	无
14	20	22	4.0	2.5	8	9	5	无	13	无	钢琴	无	11	无
15	19	24	3.0	2.0	7	7	9	无	10	无	小提琴	无	10	无
16	19	19	3.0	1.5	9	9	6	无	12	无	钢琴, 扬琴	无	15	无
17	20	20	3.0	4.0	7	6	5	无	14	无	钢琴	无	12	无
18	20	19	3.5	3.0	11	7	9	无	8	无	古筝, 钢琴, 扬琴, 大提琴	无	13	无
19	20	18	3.5	1.5	11	8	12	无	8	无	笙, 钢琴, 葫芦丝, 唢呐, 箫	无	10	无
20	19	20	3.0	2.5	10	8	6	无	12	无	钢琴	无	8	无
21	19	18	3.0	2.5	10	8	7	无	10	无	小提琴, 中提琴, 钢琴	无	6	无
22	20	19	4.0	4.5	12	5	6	无	13	无	钢琴	无	10	无
23	20	25	2.5	3.0	11	9	6	无	12	无	小提琴, 钢琴	无	5	无

注: M = 音乐训练组(Music group), C = 控制组(Control group)

相关参考文献:

Arthur, W., & Day, D. V. (1994). Development of a short form for the raven advanced progressive matrices test. *Educational & Psychological Measurement, 54*(2), 394–403.

张政华, 韩梅, 张放&李卫君.(2020).音乐训练促进诗句韵律整合加工的神经过程. *心理学报, 52*(07):847-860

意见 5: 作为一个横断性研究, 抑制控制上的差异可能来源于个体间先天的差异或者其他认知因素 (如智力), 因此要建立音乐经验与抑制控制间的关联需要进行行为和 ERP 指标与音乐训练量 (粗糙一点可以用训练年限, 精确一点可以用年限*每周频率) 的相关分析, 就算发现了显著的相关, 也不能做出音乐训练提高了抑制控制能力这样的因果性结论, 只能使用关联性的描述。

回应: 根据您的意见, 首先, 我们修改了全文关于因果关系的陈述, 并在讨论部分对横断研究的局限性进行了说明以及对未来研究进行了展望。其次, 我们对音乐训练量与抑制控制的行为指标和 ERP 指标进行了相关分析, 这些指标之间均没有显著的相关, 具体结果如下:

音乐训练量与抑制控制行为和 ERP 指标的相关结果

	Go/No-go				Stroop		
	<i>d'</i> 分数	N2 差异波	P3 差异波	反应时的干扰效应	反应时的干扰效应	N450 差异波	SP 差异波
开始学习器乐的年龄	.063	-.235	-.214	-.075	.053	-.146	-.173
学习器乐年限	.199	.130	-.091	-.313	-.241	.005	-.054
每周练习时长	-.265	-.203	.109	-.112	-.059	.096	-.100
音乐训练量 (年限*每周训练时长)	-.075	-.138	.043	-.226	-.200	.060	-.069

注: * $p < .05$, ** $p < .01$, *** $p < .001$ 。

相关结果可能反映了音乐训练经验和抑制控制能力并不存在线性的相关关系，但是这个结果也有可能与被试的选取有关，在我们的研究中，为了突出音乐训练组与对照组之间的区分，我们选取的音乐训练组被试都是具有丰富音乐训练经验的被试（平均训练年限约为 11 年），被试的同质性较高，因此就没有发现音乐训练经验与行为和脑电指标的线性相关。例如，Slevc 等（2016）的研究为了较好考察音乐经验与执行功能（包含抑制控制）的线性关系，确保被试的音乐经验值具有广泛的分布，特意挑选了两种类型的音乐训练组，一种是正式音乐训练年限低于 2 年的被试，一种是正式音乐训练年限长于 5 年的被试。由于本文有两个 ERP 实验，篇幅已经比较长了，所以我们没有把这些相关的结果以及对这些结果的解释呈现在正文中了，我们着重关心音乐训练组与对照组的差异，以及这些差异体现在抑制控制加工的哪个阶段。同时，根据您的建议，我们也避免根据组间差异做出因果推断。非常感谢您宝贵的意见。

相关参考文献：

Slevc, L. R., Davey, N. S., Buschkuehl, M., & Jaeggi, S. M. (2016). Tuning the mind: Exploring the connections between musical ability and executive functions. *Cognition*, 152, 199-211.

意见 6: 在样本量计算时采用了一个大的效应量，这有何依据？结果中组间效应似乎都是小到中的效应量。

回应: 非常感谢您的意见，我们基于 η_p^2 效应量，并把效应量设为中等，重新估算了样本量。

结果显示假定效应量 $f = 0.25$ (中等)、 $\text{power} = 0.95$ 和任务测量指标间估计相关为 0.70 的情况下, 在重复测量方差分析中, 探测到显著的组内效应或者交互作用最少需要 34 名被试, 即两组各 17 名。本实验计划样本量每组 25 名被试, 因 4 名被试 ERP 伪迹过高未纳入统计分析, 实际样本量为每组各 23 名。综上, 本研究被试量满足需求。

我们的 ERP 方差分析结果发现交互作用的 η_p^2 的范围为 0.09-0.11, t 检验中 Cohen's d 的范围为 0.62-0.70, 约对应中到大的效应量(Cohen, 1988)。同时我们也查阅了音乐训练和执行功能(包含抑制控制)文献中样本量的估计, 与我们的类似, 均为中到大的的效应量(D'Souza, Moradzadeh & Wiseheart, 2018; Joret, Germeys, & Gidron, 2016)。

相关参考文献:

Cohen, J. (1988). *Statistical power analysis for the behavior sciences*, (2nd ed.,).Mahwah: Lawrence Erlbaum Associates

Joret, M.-E., Germeys, F., & Gidron, Y. (2016). Cognitive inhibitory control in children following early childhood music education. *Musicae Scientiae*, 21(3), 303-315.

D'Souza, A. A., Moradzadeh, L., & Wiseheart, M. (2018). Musical training, bilingualism, and executive function: working memory and inhibitory control. *Cognitive Research-Principles and Implications*, 3, 18.

.....

审稿人 1 意见:

该文通过采用 Go-go 和 Stroop 任务, 使用 ERP 技术, 考察了音乐训练影响抑制控制能力的认知神经过程。研究具有一定的理论意义和现实价值。全文写作通顺流畅, 可读性强。

意见 1: 作者在前言部分介绍音乐训练影响抑制控制的相关研究结果并不一致, 除了作者在第四段提到的采用不同任务和实验手段的原因之外, 是否还有其他可能? 在讨论中除了讨论支持本实验结果的研究, 也应该进一步阐明为何一些研究会出现与当前研究不一致的研究结果。

回应: 感谢审稿专家非常好的建议。我们合并了您提的这两个问题, 在讨论部分一起进行了讨论。

修改如下(见正文见第 11 页, 最后一段):

“此外, 也有一些研究并没发现音乐训练与抑制控制存在关联。例如, Sachs 等(2017)

的研究使用 Stroop 任务和 fMRI 技术发现 8-10 岁音乐训练组和对照组儿童在抑制控制的行为学指标上并无显著差异，但是在神经学指标上，音乐训练组儿童在辅助运动区、额下回、前扣带回有更强的激活；Okada（2016）使用量表来测量大学生的音乐训练经验，结果发现音乐训练量与 Stroop 任务成绩并没有显著相关；但是，Seinfeld 等（2013）的研究对 60-84 岁老年人进行了为期 4 个月的音乐训练，训练后老年人在 Stroop 任务上的成绩有所提升。因此，被试的年龄、音乐训练经验以及抑制控制的测量指标等因素都可能对实验结果产生影响，未来的研究需要考虑这些因素。”

相关参考文献：

Okada, B. M. (2016). *Musical training and executive functions* (Unpublished master's thesis). University of Maryland, College Park.

Sachs, M., Kaplan, J., Sarkissian, A. D., & Habibi, A. (2017). Increased engagement of the cognitive control network associated with music training in children during an fMRI Stroop task. *PLOS ONE*, 12(10), 29.

Seinfeld, S., Figueroa, H., Ortiz-Gil, J., & Sanchez-Vives, M. V. (2013). Effects of music learning and piano practice on cognitive function, mood and quality of life in older adults. *Front Psychol*, 4, 810.

意见 2：前言部分最后一段的预期像是根据已经得到结果写的。正如作者在前面介绍的，也有研究没有发现音乐训练提升抑制控制，这里应该考虑到其他可能性。

回应：感谢审稿专家的意见，根据您的意见，我们对研究的预期进行了修改。

修改如下（见正文见第 3 页，最后一段）：

“本研究预期：1)若音乐训练组被试在反应抑制中的具有更强的冲突监控或运动抑制能力，那么在 Go/No-go 任务中他们的 N2 差异波或 P3 差异波波幅要大于对照组被试；若音乐训练组被试没有反应抑制能力优势音，则两组被试的 N2 和 P3 差异波不会出现显著差异；2) 若音乐训练组被试在冲突控制中具有更强的冲突监控或者冲突解决能力，那么在 Stroop 任务中他们的 N450 差异波或 SP 差异波波幅要大于对照组被试；若音乐训练组被试没有冲突控制能力优势，则两组被试在 N450 差异波和 SP 差异波上不会出现显著差异。”

意见 3：在“被试”部分，作者需要报告两组被试智力水平是否存在差异。

回应：两组被试的智力水平不存在显著差异 $t(44) = -1.37, p = 0.177$ 。此外，根据您和另外审稿专家的意见，我们通过表格详细列出了每位被试的人口学资料，如下表（见正文第 19 页）：

表 1 被试人口学资料

编号	年龄(岁)		社会经济地位		智力		训练开始年龄(岁)		训练时长(年)		乐器(首个为主修乐器)	每周练习时间(小时)			
	M	C	M	C	M	C	M	C	M	C		M	C	M	C
1	19	23	3.5	1.5	10	9	5	无	8	无	小提琴, 吉他	无	18	无	
2	18	18	3.0	2.5	10	9	9	无	8	无	古筝, 钢琴, 陶笛	无	5	无	
3	18	18	3.5	2.5	11	7	5	无	8	无	钢琴	无	6	无	
4	22	20	3.5	3.5	9	12	6	无	15	无	小提琴, 钢琴	无	9	无	
5	23	19	3.0	2.5	11	7	6	无	12	无	钢琴, 小提琴	无	14	无	
6	22	23	2.0	4.0	7	10	12	无	8	无	钢琴	无	10	无	
7	20	20	3.0	2.0	11	9	10	无	8	无	古筝, 钢琴, 手风琴	无	5	无	
8	20	17	2.5	3.5	7	6	5	无	14	无	钢琴	无	13	无	
9	21	19	3.0	2.0	3	12	5	无	15	无	钢琴	无	16	无	
10	19	19	2.0	3.0	8	7	7	无	11	无	古筝, 钢琴	无	12	无	
11	20	18	3.5	3.0	10	6	11	无	8	无	钢琴	无	10	无	
12	20	20	2.0	2.5	6	10	10	无	9	无	大提琴	无	9	无	
13	20	19	3.5	3.5	11	11	5	无	14	无	中提琴, 小提琴, 吉他	无	8	无	
14	20	22	4.0	2.5	8	9	5	无	13	无	钢琴	无	11	无	
15	19	24	3.0	2.0	7	7	9	无	10	无	小提琴	无	10	无	
16	19	19	3.0	1.5	9	9	6	无	12	无	钢琴, 扬琴	无	15	无	
17	20	20	3.0	4.0	7	6	5	无	14	无	钢琴	无	12	无	
18	20	19	3.5	3.0	11	7	9	无	8	无	古筝, 钢琴, 扬琴, 大提琴	无	13	无	
19	20	18	3.5	1.5	11	8	12	无	8	无	笙, 钢琴, 葫芦丝, 唢呐, 箫	无	10	无	
20	19	20	3.0	2.5	10	8	6	无	12	无	钢琴	无	8	无	
21	19	18	3.0	2.5	10	8	7	无	10	无	小提琴, 中提琴, 钢琴	无	6	无	
22	20	19	4.0	4.5	12	5	6	无	13	无	钢琴	无	10	无	
23	20	25	2.5	3.0	11	9	6	无	12	无	小提琴, 钢琴	无	5	无	

注: M = 音乐训练组(Music group), C = 控制组(Control group)

相关参考文献:

Arthur, W., & Day, D. V. (1994). Development of a short form for the raven advanced progressive matrices test.

Educational & Psychological Measurement, 54(2), 394–403.

张政华, 韩梅, 张放&李卫君.(2020).音乐训练促进诗句韵律整合加工的神经过程.*心理学报, 52(07):847-860*

意见 4: 作者使用“音乐训练组”和“控制组”应该好于“音乐家”和“非音乐家”。应该在全文统一稳定使用。

回应：谢谢专家细致地审查，“音乐训练组”和“控制组”的表述更为准确，根据您的意见，已在文中统一使用“音乐训练组”和“控制组”。

意见 5：该研究中 stroop 任务采用刺激为“红、绿、蓝、黄四种不同颜色的汉字“红”、“绿”、“蓝”、“黄”。共计 16 种刺激材料，其中词色一致的材料有四种，词色不一致的材料 12 种。”可以看到词色一致的材料要明显少于不一致的材料。但是后面作者又提到“本实验共计 240 个试次，字色和字义一致为一致条件，字色和字义不一致则为不一致条件，概率相等。”作者究竟是怎么做的需要交代清楚。

回应：谢谢专家对文章细致的审查。非常抱歉，我们在文中对刺激材料的类型和呈现次数没做详细介绍。根据您的意见，已经在文中进行补充。

修改如下（见正文见第 5 页，2.2.2 Stroop 任务）：

“本实验共计 240 个试次，一致条件（每种材料 30 个试次*4 种材料=120 个试次）和不一致条件（每种材料 10 个试次*12 种材料=120 个试次）各为 120 次。”

意见 6：音乐组的 Stroop 干扰效应在正确率上（一致减不一致正确率）显著小于控制组，这个结果该如何解释？“在 Stroop 任务中的行为指标上，音乐组的 Stroop 干扰效应更小”这个结论是怎么得到的？

回应：感谢审稿人的意见，是我们语言表述不够清楚从而造成了误解。Stroop 任务的行为核心指标为 Stroop 干扰效应（Stroop interference effect），用来评估被试的冲突控制能力，即一致条件和不一致条件之间正确率和反应时之差(MacLeod, 1991; Lansbergen, van Hell, & Kenemans, 2007; Tillman, & Wiens, 2011)。差值越小，说明相对于一致条件，被试在完成不一致条件任务时成绩下降较小，从而反映了更强的冲突控制能力 (MacLeod, 1991; Lansbergen et al., 2007; Tillman et al., 2011)。

本研究的行为结果发现，音乐训练组被试在正确率上的 Stroop 干扰效应显著小于控制组被试，说明音乐训练组的冲突控制能力更强。

相关参考文献：

MacLeod, C. M. (1991). Half a Century of Research on the Stroop Effect: An Integrative Review. *Psychological Bulletin*, 109(2), 163-203.

Tillman, C. M., & Wiens, S. (2011). Behavioral and ERP indices of response conflict in Stroop and flanker tasks.

Psychophysiology, 48(10), 1405-1411.

Lansbergen, M. M., van Hell, E., & Kenemans, J. L. (2007). Impulsivity and conflict in the Stroop task - An ERP study. *Journal of Psychophysiology*, 21(1), 33-50.

意见 7: 我认为作者的统计存在一定问题。以 go 和 no-go 任务中脑电指标为例，在两种任务中均以条件和被试作为自变量，以平均波幅作为因变量进行统计。那以这个统计方法，是无法获得两类被试在 N2 和 P3 差异波上是否存在差别的，作者应该是做了相减以后进行了 t 检验，但是没写。

回应: 感谢审稿专家的意见。在数据分析中，我们首先是进行两因素的方差分析，当存在交互作用，进行简单效应分析时。然后进一步对 No-go 和 Go 条件的 N2 和 P3 波幅进行了相减，再进行独立样本 t 检验。在文中我们没有表述清楚，根据你的意见已经进行修改。

修改如下（见正文见第 7 页，3.2 ERP 结果）:

“独立样本 t 检验结果发现，音乐训练组的 N2 差异波（No-go 减 Go）显著大于控制组 $t(44) = -2.33, p = 0.025, 95\% \text{ CI} = [-2.89, -0.21], \text{Cohen's } d = 0.69$ （见图 2c 和 2d）。”

“独立样本 t 检验结果发现，音乐训练组的 P3 差异波（No-go 减 Go）显著大于控制组 $t(44) = 2.37, p = 0.022, 95\% \text{ CI} = [0.28, 3.47], \text{Cohen's } d = 0.70$ （见图 3c 和 3d）。”

“独立样本 t 检验结果发现，音乐训练组的 N450 差异波（不一致减一致）显著大于控制组 $t(44) = -2.11, p = 0.041, 95\% \text{ CI} = [-2.06, 0.04], \text{Cohen's } d = 0.62$ （见图 4c 和 4d）。”

“独立样本 t 检验结果发现：音乐训练组的 SP 差异波（不一致减一致）与控制组无显著差异 $t(44) = 0.34, p = 0.738$ （见图 5c 和 5d）。”

意见 8: 我认为讨论中 4.1 音乐训练对反应抑制的影响“冲突监控模型。。。”这一段内容应该紧接“与以往研究一致。。。”这一段。

回应: 感谢您的细致审查，我们已在文中修改。

意见 9: uV 应该改为 $\mu \text{ V}$ 。

回应: 感谢您的细致审查，已在文中和图中修改。

意见 10: 英文摘要需要再好好改一下，中文一些表达上存在的问题标在文中，可能不止这些地方。

回应：我们再次仔细检查了全文的语言，已认真修改英文摘要和中文表达。

.....

审稿人 2 意见：

音乐训练对抑制控制的行为学研究已有很多，也有一些相关的神经科学研究。作者将抑制控制进行进一步划分（反应抑制和冲突控制），并通过两个行为测验任务（Go/No-go、Stroop）以及脑电技术进行观察（各阶段加工特征），这是该研究的新颖点。全篇内容安排合理、研究依据和观点翔实明确、方法合理、写作通俗易懂。审稿人近些年主要通过脑成像进行研究，但在 10 年前做过一点 ERP 研究，因此以下意见供作者参考。

意见 1: Stroop 任务中，作者主要使用 N450 和 SP 来观察冲突控制，但有些研究也通过 N1、P1、P2（例如 Yu et al., 2015）、N2（如 Naylor et al., 2012; Yu et al., 2015）、P3 和 N4（如 Sahinoglu & Dogan 2016）、global field power (GFP)（Badzakova-Trajko et al., 2009）等指标观察 Stroop 测验。虽然不同的文献选用了不同的研究任务（如语言的、听觉的 Stroop 等），请问在该研究中作者只选择 N450 和 SP 的原因？这是基于大多数的参考文献还是仅限于音乐训练的相关研究？同样，在 Go/No-go 任务中，也有研究观察了 N4（如 Lee et al., 2019）和 N1（如 Jia et al., 2017），但很多文献基本关注 N2 和 P3（与作者研究一致）。

回应：谢谢审稿人的宝贵意见，我们认真拜读了您提供的文献和经典视觉 Stroop、Go/No-go 任务的文献，发现视觉色-词 Stroop 任务往往会诱发与冲突控制有关的成分 N450 和 SP (Gajewski & Falkenstein, 2015; Liotti et al., 2000; West, 2004)。N450 成分反映冲突监测的认知过程，SP 成分反映冲突解决的认知过程(Larson, Clayson, & Clawson, 2014; West & Alain, 2000)。正如您说的，也有一些 Stroop 研究考察 N1、P1、P2、N2、P3 和 N4 等成分，这些研究与本研究考察的 ERP 成分不同，主要有以下几个原因。1)不同感觉通道的 Stroop 任务。例如 Yu 学者（2015）的研究采用了听觉 Stroop 任务，考察了 N1、P1、P2 等听觉早期成分，主要反映了被试的选择注意和知觉冲突检测等早期认知控制阶段。鉴于感觉通道的差异性，视觉 Stroop 任务研究却发现早期加工阶段与视觉冲突检测无关，所以视觉 Stroop 任务较少考察早期成分（Yu et al., 2015）。2）不同的 Stroop 范式。例如 Naylor（2012）的研究采用的是双语 Stroop 任务，即完成不同语言的色-词 Stroop 任务，主要用来研究双语者每种语言相互的干扰程度。N2 成分主要反映双语 Stroop 任务中对反应相关语言的识别过程(Naylor et al., 2012)，并不是冲突控制的经典成分。3)Stroop 任务被试不同的响应方式。Sahinoglu 和 Dogan

(2016) 发现, 在 Stroop 任务中当被试要求说出单词颜色或在心中默念单词颜色时与按键反应的加工过程不同, N4 成分可能反映了语言加工对冲突解决过程的影响。4) 数据的不同分析方法。Badzakova-Trajko (2009) 主要使用 EEG 微状态分析, 与我们 ERP 时域分析方法不同。global field power (GFP)是指某个时间点地形图所有电极电压值的标准差, 用于描述某个地形图电场的强度。而地形图的强度 (GFP) 增大、降低或极性变化反应了 EEG 的微状态 (Badzakova-Trajko et al., 2009)。综上, 因为我们使用了视觉的经典色-词 Stroop 任务考察音乐组和非音乐组反应抑制的差异, 且被试做按键反应, 综合前人文献, 选择了冲突控制的相关指标 ERP 时域指标 N450 和 SP。

对于 Go/No-go 任务, 正如您提到的, 大量文献关注 N2 和 P3 成分, N2 成分反映了冲突监测能力, P3 成分反映了运动抑制能力(Donkers & van Boxtel, 2004; Enriquez-Geppert, Konrad, Pantev, & Huster, 2010; Folstein & Van Petten, 2008)。在 Lee (2019) 研究中使用基于 Go/No-go 选择的词汇判断任务 (go/no-go choice in lexical decision task), N4 是与语言加工有关的 ERP 成分, 主要反映词频等词汇和语义特征, 不是反应抑制的核心成分 (Lee et al., 2019)。Jia (2017) 考察了 Go/No-go 任务中与视觉刺激处理相关的视觉 P1 和 N1 成分, 而这些早期成分并不是反应抑制的核心成分。综上, 鉴于本研究使用的经典视觉 Go/No-go 任务来考察音乐组和非音乐组反应抑制的差异, 选择了反应抑制相关的 ERP 成分 N2 和 P3。

相关参考文献:

- Badzakova-Trajkov, G., Barnett, K. J., Waldie, K. E., & Kirk, I. J. (2009). An ERP investigation of the Stroop task: the role of the cingulate in attentional allocation and conflict resolution. *Brain Res*, 1253, 139-148.
- Donkers, F. C. L., & van Boxtel, G. J. M. (2004). The N2 in go/no-go tasks reflects conflict monitoring not response inhibition. *Brain and Cognition*, 56(2), 165-176.
- Enriquez-Geppert, S., Konrad, C., Pantev, C., & Huster, R. J. (2010). Conflict and inhibition differentially affect the N200/P300 complex in a combined go/nogo and stop-signal task. *NeuroImage*, 51(2), 877-887.
- Folstein, J. R., & Van Petten, C. (2008). Influence of cognitive control and mismatch on the N2 component of the ERP: A review. *Psychophysiology*, 45(1), 152-170.
- Gajewski, P. D., & Falkenstein, M. (2015). Long-term habitual physical activity is associated with lower distractibility in a Stroop interference task in aging: Behavioral and ERP evidence. *Brain and Cognition*, 98, 87-101.
- Jia, H. B., Li, H. Y., & Yu, D. C. (2017). The relationship between ERP components and EEG spatial complexity in a visual Go/Nogo task. *Journal of Neurophysiology*, 117(1), 275-283.
- Larson, M. J., Clayson, P. E., & Clawson, A. (2014). Making sense of all the conflict: A theoretical review and

- critique of conflict-related ERPs. *International Journal of Psychophysiology*, 93(3), 283-297.
- Lee, H., Lee, Y., Tae, J., & Kwon, Y. (2019). Advantage of the go/no-go task over the yes/no lexical decision task: ERP indexes of parameters in the diffusion model. *PLOS ONE*, 14(7).
- Liotti, M., Woldorff, M. G., Perez, R., & Mayberg, H. S. (2000). An ERP study of the temporal course of the Stroop color-word interference effect. *Neuropsychologia*, 38(5), 701-711.
- Naylor, L. J., Stanley, E. M., & Wicha, N. Y. Y. (2012). Cognitive and electrophysiological correlates of the bilingual Stroop effect. *Front Psychol*, 3.
- Sahinoglu, B., & Dogan, G. (2016). Event-Related Potentials and the Stroop Effect. *Eurasian Journal of Medicine*, 48(1), 53-57.
- West, R. (2004). The effects of aging on controlled attention and conflict processing in the Stroop task. *Journal of Cognitive Neuroscience*, 16(1), 103-113.
- West, R., & Alain, C. (2000). Effects of task context and fluctuations of attention on neural activity supporting performance of the stroop task. *Brain Res*, 873(1), 102-111.
- Yu, B., Wang, X., Ma, L., Li, L., & Li, H. (2015). The Complex Pre-Execution Stage of Auditory Cognitive Control: ERPs Evidence from Stroop Tasks. *PLOS ONE*, 10(9).

意见 2: 在 2.1 被试中:

- (1) 25 名音乐专业被试主要学习/演奏何种乐器 (钢琴、小提琴...)?
- (2) “至今仍在练习”: 练习的时间和频率是? (如每周平均几天练习、每次练习时间)
- (3) 两组被试的“社会经济地位”显示了父母受教育程度无差异。受教育程度信息是如何获得 (询问或正式问卷调查)? 是否还有家庭经济收入等询问或调查?
- (4) 两组被试的年龄、社会经济地位、经济收入 (若有这方面数据) 等信息, 最好通过表格呈现, 列入平均值、标准差等... 要添加表格 (在附近呈现吧)

回应: 根据您和另外审稿专家的意见, 我们把每位音乐组被试的音乐训练经历主要包括训练起始年龄、训练时长、所学器乐类型、每周练习时间和每位被试的个人信息包括年龄、社会经济地位、智力, 通过表格汇总呈现在文中。具体内容见表 1 被试人口学资料。我们使用父母的受教育程度来衡量社会经济地位, 1-5 点计分: “小学及以下” =1; “初中” =2; “高中或中专” =3; “大学” =4; “研究生及以上” =5)。被试的个人信息是通过问卷调查形式收集。此外, 我们没有调查被试的家庭经济收入。

新增加的表 1 如下 (见正文第 19 页):

表 1 被试人口学资料

编号	年龄 (岁)		社会经济地位		智力		训练开始年龄 (岁)		训练时长 (年)		乐器 (首个为主修乐器)	每周练习时间 (小时)		
	M	C	M	C	M	C	M	C	M	C		M	C	
1	19	23	3.5	1.5	10	9	5	无	8	无	小提琴, 吉他	无	18	无
2	18	18	3.0	2.5	10	9	9	无	8	无	古筝, 钢琴, 陶笛	无	5	无
3	18	18	3.5	2.5	11	7	5	无	8	无	钢琴	无	6	无
4	22	20	3.5	3.5	9	12	6	无	15	无	小提琴, 钢琴	无	9	无
5	23	19	3.0	2.5	11	7	6	无	12	无	钢琴, 小提琴	无	14	无
6	22	23	2.0	4.0	7	10	12	无	8	无	钢琴	无	10	无
7	20	20	3.0	2.0	11	9	10	无	8	无	古筝, 钢琴, 手风琴	无	5	无
8	20	17	2.5	3.5	7	6	5	无	14	无	钢琴	无	13	无
9	21	19	3.0	2.0	3	12	5	无	15	无	钢琴	无	16	无
10	19	19	2.0	3.0	8	7	7	无	11	无	古筝, 钢琴	无	12	无
11	20	18	3.5	3.0	10	6	11	无	8	无	钢琴	无	10	无
12	20	20	2.0	2.5	6	10	10	无	9	无	大提琴	无	9	无
13	20	19	3.5	3.5	11	11	5	无	14	无	中提琴, 小提琴, 吉他	无	8	无
14	20	22	4.0	2.5	8	9	5	无	13	无	钢琴	无	11	无
15	19	24	3.0	2.0	7	7	9	无	10	无	小提琴	无	10	无
16	19	19	3.0	1.5	9	9	6	无	12	无	钢琴, 扬琴	无	15	无
17	20	20	3.0	4.0	7	6	5	无	14	无	钢琴	无	12	无
18	20	19	3.5	3.0	11	7	9	无	8	无	古筝, 钢琴, 扬琴, 大提琴	无	13	无
19	20	18	3.5	1.5	11	8	12	无	8	无	笙, 钢琴, 葫芦丝, 唢呐, 箫	无	10	无
20	19	20	3.0	2.5	10	8	6	无	12	无	钢琴	无	8	无
21	19	18	3.0	2.5	10	8	7	无	10	无	小提琴, 中提琴, 钢琴	无	6	无
22	20	19	4.0	4.5	12	5	6	无	13	无	钢琴	无	10	无
23	20	25	2.5	3.0	11	9	6	无	12	无	小提琴, 钢琴	无	5	无

注: M = 音乐训练组(Music group), C = 控制组(Control group)

(5) 由于两组被试在性别数量上不匹配, 那么在行为和脑电分析中, 性别因素是否考虑作为协变量?

回应: 感谢您的意见。我们把性别作为协变量进行方差分析, 行为结果和脑电结果均与原来的结果一致。因此在正文中我们没有把性别因素作为协变量进行了分析。以往研究发现在执行功能 (包括抑制控制能力) 上并不存在明显的性别差异 (见 Grissom & Reyes, 2019 的综述性论文), 也没有研究发现性别因素会调节音乐训练对抑制控制的影响, 所以当初在做实验时没有严格匹配性别数量, 在未来的研究中, 我们会注意这个问题。非常感谢您的意见。

相关参考文献:

Grissom, N. M., & Reyes, T. M. (2019). Let's call the whole thing off: evaluating gender and sex differences in executive function. *Neuropsychopharmacology*, 44(7), 1344-1344.

意见 3: 在 3.1 行为和脑电结果中: 都给出了方差和 p 值。作者是否可列出一个表, 输入相应的平均值和标准差, 并在后面也给出 F 值 (F 值的 df 请下标)。

回应: 回应: 根据您的意见, 我们绘制了表格, 如下 (见正文见第 19 页, 表 2):

表 2 两组被试在 Go/No-go 和 Stroop 任务中的正确率(%), 反应时(ms)和 ERP 波幅 (μV) 的差异比较

执行功能	控制组 <i>M</i> (<i>SD</i>)	音乐训练组 <i>M</i> (<i>SD</i>)	<i>F</i> _(1, 44)	<i>p</i>
Go/No-go				
行为				
Go 正确率	99.35 (1.72)	99.83 (0.65)	1.55	0.219
No-go 正确率	96.22 (4.40)	96.00 (3.93)	0.03	0.861
ERP				
Go N2	2.72 (3.33)	7.85 (4.50)	19.24	0.001***
No-go N2	1.69 (3.62)	5.27 (5.55)	6.69	0.013*
Go P3	6.49 (3.62)	11.83 (4.76)	18.37	0.001***
No-go P3	8.05 (4.12)	15.27 (5.98)	22.70	0.001***
Stroop				
行为				
一致 正确率	89.30 (7.89)	93.26 (5.11)	4.07	0.050*
不一致 正确率	75.70 (12.91)	84.48 (8.64)	7.36	0.010**
一致 反应时	636.48 (41.50)	584.39 (61.18)	11.42	0.002***
不一致 反应时	700.90 (42.88)	657.93 (70.24)	6.27	0.016*
ERP				
一致 N450	1.20(3.00)	4.13 (5.13)	5.62	0.022*
不一致 N450	0.41 (2.56)	2.29 (4.19)	3.39	0.072
一致 SP	-5.08 (3.26)	-4.84 (4.40)	0.04	0.835
不一致 SP	-3.54 (3.14)	-3.49 (3.88)	0.003	0.959

注: * $p < .05$, ** $p < .01$, *** $p < .001$ 。

意见 4: 在 3.2.1 Go/No-go 结果中: N2 和 P3 都未有被试类型主效应结果。请问是否不显著?

回应: 根据您的意见, 我们在结果部分补充了 Go/No-go 任务中 N2 和 P3 成分被试类型主效应的结果和 Stroop 任务中 N450 成分被试类型主效应的结果。结果发现 Go/No-go 任务中, N2 和 P3 成分的被试类型主效应显著, 音乐组的 N2 波幅小于控制组被试, 但 P3 波幅大于控制组被试。Stroop 任务中, N450 成分的被试类型主效应显著, 音乐组的 N450 波幅大于

控制组被试。

修改如下（见正文见第 7 页，3.2 ERP 结果）：

“被试类型主效应显著 $F(1, 44) = 12.39, p = 0.001, \eta_p^2 = 0.22$ ，音乐训练组的 N2 波幅小于控制组（见表 2）。”

“被试类型主效应显著 $F(1, 44) = 22.31, p < 0.001, \eta_p^2 = 0.34$ ，音乐训练组的 P3 波幅大于控制组（见表 2）。”

“被试类型主效应显著 $F(1, 44) = 4.73, p = 0.035, \eta_p^2 = 0.10$ ，音乐训练组的 N450 波幅大于控制组（见表 2）。”

意见 5：（1）2.2.1 将“要求被试快速且准确地做出反应”挪到“而当 No-go 试次出现时无需按键”之前；

（2）2.2.2（请将标题改为 2.2.2）Stroop 任务的最后一句：请将”一致试次和不一致试次之间正确率和反应时之差被”，去掉“被”字；

（3）2.3 数据采集与分析的第三行：“使用 ERPlab 插件对对数据进行离线分析”去掉一个“对”字。

回应：感谢您认真的审读和检查，我们再次仔细检查了全文的语言，错别字等语言问题均已经在文中修改。

第二轮

审稿人 1 意见：

本文作者根据审稿老师意见对文章做了较好的修改和回复。目前文章存在问题如下：

意见 1：前言中对本文涉及到的脑电成分如 N2, P3, N450 的介绍可以完善一下，比如 N450 的头皮分布等，作者可以作进一步的补充。

回应：感谢您的建议。根据您的意见，我们在前言部分对 N2、P3 和 N450 成分进行了介绍，

修改如下（见正文第 3 页）：

“在脑电成分上，Go/No-go 任务会诱发比较明显的额-中央区分布的 N2 和 P3 成分，并且 No-go 条件比 Go 条件会诱发更大的 N2 和 P3 波幅(Donkers & van Boxtel, 2004)。No-go N2 效应被认为与反应抑制过程中的冲突监控有关，一般用 No-go 条件诱发的 N2 波幅或 No-go 减去 Go 条件的 N2 差异波幅（N2d）来衡量，并且 No-go N2 或 N2d 的波幅越大则可能反映

了越强的冲突监控能力 (Cheng et al., 2017; Donkers et al., 2004; Nieuwenhuis et al., 2003; Pandey et al., 2012; Smith et al., 2010)。No-go P3 效应被认为与反应抑制过程中的运动抑制有关, 一般用 No-go 条件诱发的 P3 振幅或 No-go 条件减 Go 条件的 P3 差异波幅 (P3d) 来衡量, No-go P3 或者 P3d 的波幅越大则可能反映了越强的运动抑制能力 (Enriquez-Geppert et al., 2010; Guan et al., 2015 ; Jonkman, 2006; Smith et al., 2008; Smith et al., 2013)。”

“Stroop 任务往往会诱发比较明显的 N450 和冲突相关的慢电位(conflict-sensitive slow potential, SP) (Gajewski & Falkenstein, 2015; West & Alain, 2000b; West, 2004)。N450 一般出现在刺激呈现后 300-600 ms 的时间窗口, 在 450ms 左右达到峰值, 并在头皮额-中央区比较明显, 不一致条件比一致条件会诱发更大的 N450 波幅(Gajewski et al., 2015; Larson et al., 2014)。N450 成分与冲突控制过程中的冲突监控有关, 并且 N450 差异波(不一致条件减一致条件)越大则可能反映了越强的冲突监控能力(Holmes & Pizzagalli, 2008; West, 2004; West & Alain, 2000a)。”

意见 2: 目前文中仍然有语句表达问题, 如: 抑制控制能力是人类非常重要的认知功能之一, 它对是个体适应环境的具有重要的意义。其他部分也存在, 请认真修改。

回应: 感谢您的建议, 我们对全文的语句表达进行了认真检查和修改。

审稿人 2 意见: 作者认真回复了三位审稿人的意见, 对每一项问题都给予了细致和认真的回答; 写作和修改态度诚恳细致。全文写作流畅、语句通顺。虽然音乐训练和执行功能的国内外文章已有很多, 但该论文将执行功能进行了进一步细分, 虽然理论意义适中, 但应用价值比较高, 因为当前音乐学习和训练的意义受到广泛关注, 该论文的研究结果有助更加细致和深入了解音乐学习和训练的意义。

回应: 感谢审稿专家的评价和宝贵建议。

第三轮

编委复审意见: 同意该文发表。

主编终审意见: 本人研读了该篇论文, 基本同意几位审稿专家的意见, 同意该文发表。