

《心理学报》审稿意见与作者回应

题目：音乐训练促进诗句韵律整合加工的神经过程

作者：张政华，韩梅，张放，李卫君

第一轮

审稿人 1 意见：

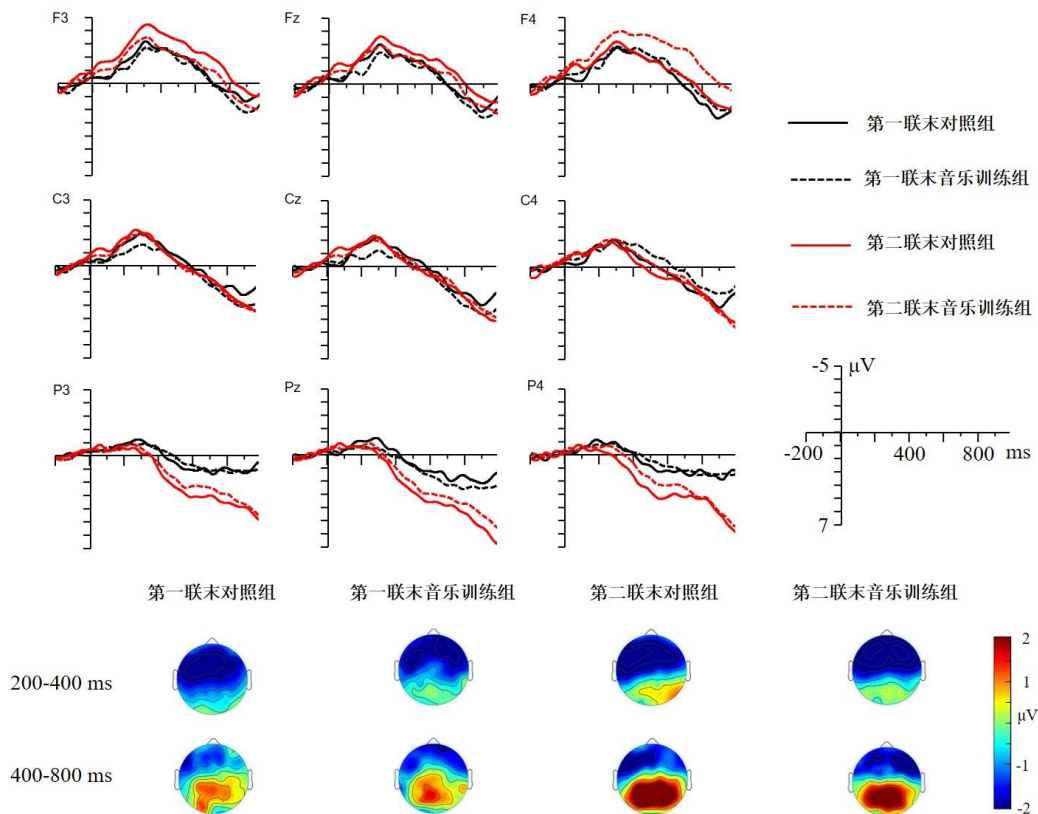
本研究通过对七律古诗中的韵律信息操控，探索了不同音乐背景经历对韵律整合的异同。实验设计新颖。所探讨的问题比较有科学意义。我的建议主要集中在数据分析和解释上。希望我的建议可以帮助作者精进此研究。

意见 1：“七言绝句有严格的韵律规则，其中包括押韵，即通常要求诗句第一句、第二句、第四句最后一个字的韵母相同，且声调为平声。”因此第一句最后一个音可以作为奠定韵律的基础，而第一联最后一个音(即本实验操控的)是首次押此韵的音，而第二联的最后一个音是第二次押此韵。因此第二联的最后一个音的整合至少可以有两种方式：其和第一句的最后一个音整合，或是和第一联最后一个音整合。而现有的脑电分析锁定的起始点为每首诗中第二联最后一个字的开始位置，因此，所观察到的 300~700ms 反应在两组被试中的不同有可能是整合方式的不同 -- 一组偏向和最近的整合—recency effect；另一组偏向对一开始的整合—primacy effect。

回应：感谢审稿人提出两种整合方式的可能性，使我们进一步思考和分析当前数据。根据诗句押韵特点，绝句末尾确实可能存在您所说的两种整合方式(recency effect, primacy effect)。不过，我们更倾向于认为这两种整合方式是共存的，并且两组被试在本实验中，可能需要同时完成以上提及的两种整合方式。对于违反押韵规则的诗句，一旦第一联末音违反了诗句押韵规则，被试即可判断该绝句为押韵违反，并不再以此位置的音为标准对诗句末的音进行判断。两类被试在绝句末整合应该同时采用两种整合方式。对于押韵合适的诗句，如果被试间存在基于不同位置的整合方式的差异，那么通过比较两类被试在第一联末和第二联末的 ERP 结果，即可获得第二联末 ERP 结果中的被试间差异是否由于整合方式不同造成。因此，我们在 200~400 ms 和 400~800 ms 两个可能存在差异的时间窗口，对两类被试在第一联末和第二联末两个位置全合适条件下诱发的 ERP 进行了以被试(音乐训练组，对照组)，位置(第一联末；第二联末)，脑区(前，中，后)和半球(左，中，右)为自变量，各时间窗口平均波幅

为因变量的重复测量方差分析。结果发现，在 200~400 ms，脑区和位置交互作用显著， $F(2, 96) = 17.69, p < 0.001, \eta_p^2 = 0.27$ 。简单效应分析表明，第二联末相比于第一联末在前部诱发了波幅更大的负效应， $F(1,48) = 10.30, p = 0.002, \eta_p^2 = 0.18$ ，在后部诱发了波幅更大的正效应， $F(1,48) = 6.84, p = 0.012, \eta_p^2 = 0.13$ 。被试主效应以及被试与位置、脑区的交互作用均不显著($ps > 0.1$)。在 400~800 ms，发现脑区和位置交互作用显著， $F(2, 96) = 18.22, p < 0.001, \eta_p^2 = 0.28$ 。简单效应分析表明，第二联末相比于第一联末在后部诱发了波幅更大的正效应， $F(1,48) = 18.80, p < 0.001, \eta_p^2 = 0.28$ 。被试主效应以及被试与位置、脑区的交互作用均不显著($ps > 0.1$)(如补充图 1 所示)，总之，我们发现第二联末和第一联末在两时间窗均诱发了差异显著的脑电效应，但是两个位置均不存在被试间差异，表明两组被试在绝句内部和末尾使用了相似的整合方式。

综上，本研究在诗歌第一联和第二联末均设置了可能出现声调或/和韵母违反的情况；尽管前者的比率高于后者，但是在实验过程中，被试必须以绝句第一句最后一个音作为奠定韵律的基础，同时考虑第一联末和第二联末两个位置声调和韵母的合适性，才能完成最终的押韵合适性判断任务，两组被试均需在绝句末完成与第一句和第一联末末音进行整合的过程。



补充图 1 音乐训练组(25 个)和对照组(25 个)被试加工七言绝句在第一联末和第二联末两个位置诱发脑电

效应的总平均波形(上)和地形图(下)。

意见 2: “对句内发生韵律违反，句末完成韵律信息整合”。(首先要明确使用诗、联、句。不要混淆产生混乱。这里应该是诗内)。

回应: 非常感谢审稿人的建议。我们将“对句内发生韵律违反，句末完成韵律信息整合”。改为“对诗句内发生韵律违反，诗句末完成韵律信息整合”。此外，根据审稿人意见，修改稿中其他可能产生混淆的表达也已作出相应修改。

意见 3: 违反和整合均需要前提，即上面所说的要有一个音奠定韵律基础，之后的音才可以和其比较。因此不存在‘内’和‘末’之分。作者所说的‘内’也是第一联的末，也会和之前第一句比较和整合。因此，挽救此项目的一个方法是分析 ERP 对第一联最后一个音，即作者操控的地方的 ERP 效应。应，并将其与对第二联最后一个音的反应做对比。Erp 的分析和解释中的正负极性及其相对大小，是受到其具体位置所影响的。比如，所观察到的反应幅度区别也许是因为脑中的源的位置朝向变化产生了脑地形图(电势分布)的变化，而反应幅度其实没有变化。如果要做选电极的分析，需要先确定不同条件下脑地形图相似，因此之后发现的差别就只是反应幅度的区别。或者变成无极性不受电极位置的测量量(比如 GFP，所有点击的几何平均)来更好的探测操控变量如何影响 erp 反应幅度；

回应: 非常感谢审稿人的建议。我们同意审稿人关于诗歌“内”和“末”的界定。本研究中我们说的“内”确实也是绝句第一联的末，并且只有和第一句末音比较才会确定其押韵的合适性。如上对问题 1 的回复，我们根据审稿人建议分析了第一联最后一个音的 ERP 效应，并与第二联最后一个音的反应进行比较。结果发现在 200~400 ms，第二联末相比于第一联末在前部诱发了波幅更大的负效应，在后部诱发了波幅更大的正效应。在 400~800 ms，第二联末相比于第一联末在后部诱发了波幅更大的正效应，与以往研究结果一致(李卫君，杨玉芳，2010)。

此外，如审稿人所言，ERP 的分析和解释中正负极性及其相对大小，确实受到其具体位置影响。就当前实验结果，从补充图 1 的地形图可以看到音乐训练组和对照组在完全合适的诗句中，在第一联末和第二联末诱发脑电效应地形图相似，仅在幅度上存在差别。其他的效应也可以看到各条件差异主要体现在反应幅度，而非地形图上(如图 3 所示)。不过，我们仍然根据您的建议，对 100~300ms 时间窗的 ERP 数据进行了 GFP 分析。具体结果请见问题 10 的回复。

意见 4: 图 3 中需把 V+T+这个基线条件的地形图也画出来。加画柱状图表示统计结果。

回应: 已经根据审稿人意见画出 V+T+这个基线条件的地形图(详见图 3), 并加画柱形图(详见图 4)表示统计结果。

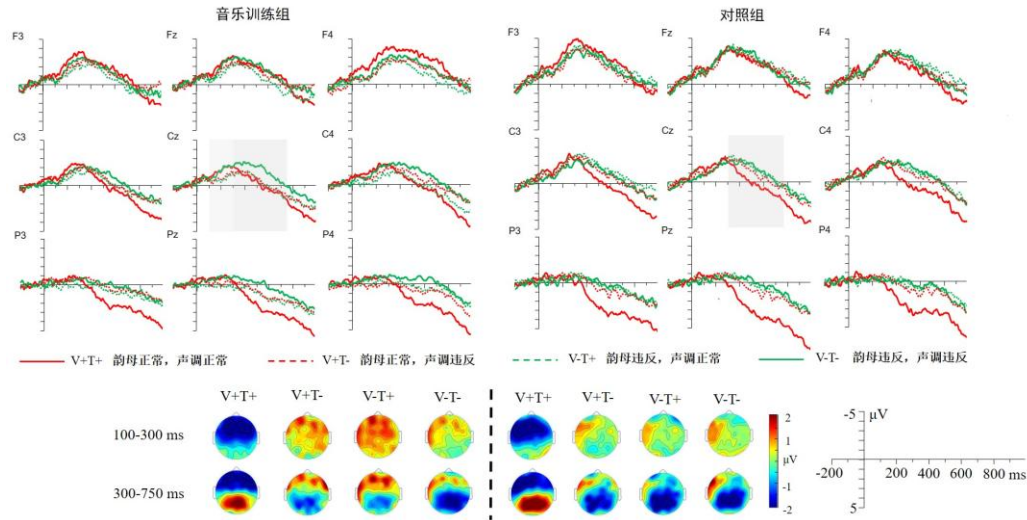


图 3 25 个音乐训练者(左)和 25 个对照组(右)加工七言绝句押韵信息的总平均波形(上)和地形图(下)。ERP 分析的起始时间为第二联末字开始位置。地形图为全合适条件的地形图以及各违反条件与全合适条件的差异波地形图。

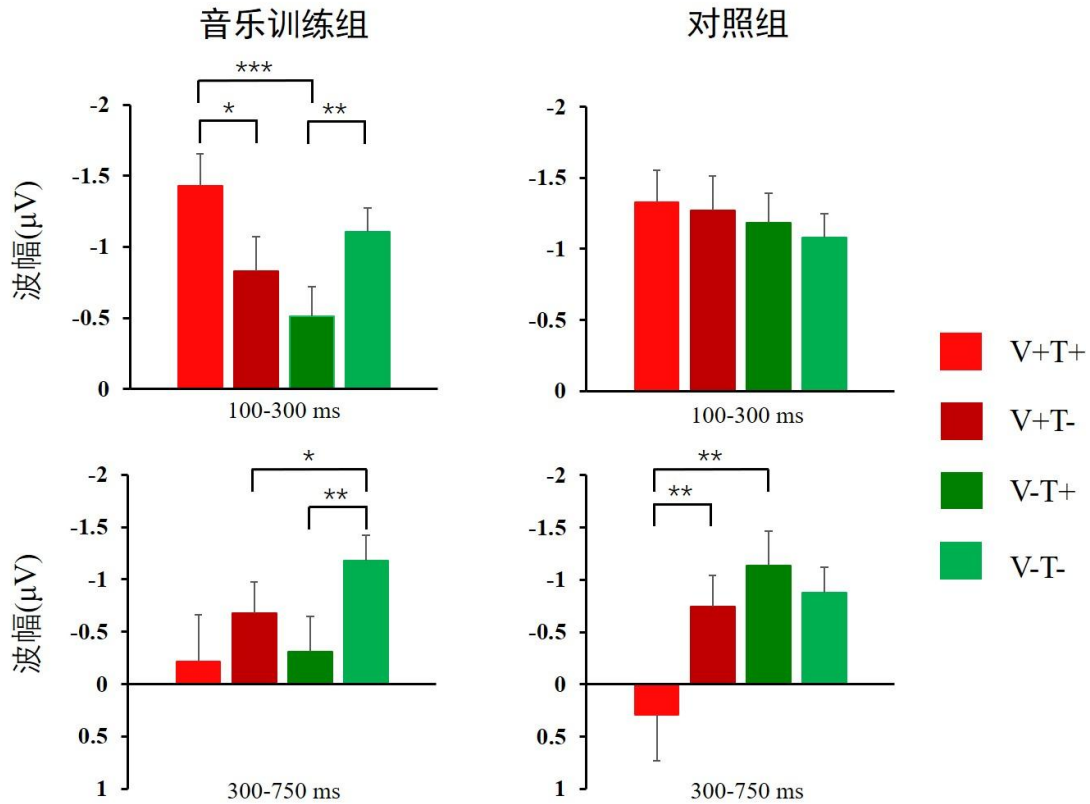


图 4 25 个音乐训练者(左)和 25 个对照组被试(右)加工七言绝句押韵信息时大脑两侧 100-300 ms (上)和 300-750 ms (下)诱发脑电效应的平均波幅。V+T+: 韵母正常, 声调正常; V+T-: 韵母正常, 声调违反; V-T+: 韵母违反, 声调正常; V-T-: 韵母违反, 声调违反。

意见 5: 低熟悉度的七言绝句—如何量化和确定为低熟悉度。

回应: 低熟悉度七言绝句量化标准具体为: 首先通过网络和书籍选择实验者不熟悉的 300 首七言绝句, 然后进行预实验。预实验中, 16 名大学生阅读依次呈现的绝句, 并要求他们判断这些绝句是否熟悉, 并在 7 点量表上表明该绝句意义上的可理解程度。最终我们选择被试认为不熟悉并且意义性较低(5 分及其以下, 主要为了平衡最小语义干扰和开展 ERP 实验所需试次的) 240 首七言绝句作为正式实验材料。以上内容已经增加在修改稿中方法部分(第 20 页)。

意见 6: “通常要求诗句第一句、第二句、第四句最后一个字的韵母相同, 且声调为平声”, 是否确定了所有材料均符合, 尤其是第一句、第二句、第四句最后一个字的韵母相同。

回应: 谢谢审稿人提出的疑问。由于韵母完全相同且熟悉度低的诗句数量有限, 因此我们使用的绝句除了第一句、第二句、第四句最后一个字的主要韵母相同的材料外, 还包括了三种其他押韵方式的材料: eng, ing: 如“生, 轻, 行”; en, in: 如“侵, 阴, 深”; ou, iu: 如“流, 幽, 鸥”(占总材料 20.6%)(见修改稿第 20 页)。

在练习中, 我们将上述三种其他押韵类型材料增加在练习实验中, 并告知被试受历史影响这些诗句仍属于押韵合适(见修改稿第 21 页)。实验材料通过听觉方式呈现, 至少相比韵母违反条件, 被试更容易知觉到这些押韵的音之间的相似性, 因此对于比较两类被试绝句末韵律整合加工结果应该不会产生影响。本研究所有实验材料文字版请参见文末“附表: 实验材料”。

意见 7: 材料为录制, 正常朗读, 如何控制所有语音因素, 比如语速, 节奏, 重音等? 100~300ms 反应以及不像典型的 N1/P2 听觉反应了。这很有可能是录音中连续音节产生的影响。那这种印象是否干扰了相关测量, 进而影响了结果的可行性?

回应: 根据之前研究对于七言绝句声音特点的分析表明, 节奏和重音方面七言绝句有较为固定的节奏模式, 至少同一个发音人朗读的不同绝句在语速, 节奏和重音上的差异不大(Li & Yang, 2010)。当前实验我们邀请了一名男性专业发音人录制所有实验材料。对于声调和韵母全合适的绝句, 发音人会完整朗读每首绝句; 声调和韵母任何一个维度违反的绝句, 发音人仅朗读诗句第一联, 然后通过语音拼接技术将第一联末字拼接完全合适的绝句中。因此, 对于同一绝句的不同条件, 仅第一联末字不同, 其他部分完全相同。通过统计发现, 该发音人朗读每首合适绝句总体发音时间在 14s 到 17s 之间(见第 20 页)。

如审稿人所言，受到听觉呈现材料连续语流的影响，本实验 100-300ms 诱发脑电效应确实不像典型的 N1/P2 听觉反应。不过如上所述，在绝句末我们对不同条件之间比较的位置，实验材料是完全相同的，因此脑电上产生的任何差异都不可能是由单纯的物理刺激引起的。而且当前研究在 100-300 ms 诱发的脑电效应与以往研究中使用连续语流考察语言信息加工在早期诱发的效应类似(Marie, Magne, & Besson, 2011)。如果连续语流产生的效应会干扰相关测量，那么这种连续语流对各个条件的影响是均等的，因此我们倾向于认为不会影响结果的可信度。

参考文献：

Li, W., & Yang, Y. (2010). Perception of Chinese poem and its electrophysiological effects. *Neuroscience*, 168(3), 757-768.

意见 8：“声音刺激播放完毕后会 1/3 的试次后出现探测界面，要求被试既快又准地判断诗句是否押韵。”因主操控在第一联末，而填充材料的违反在第二联末。那被试要完成任务是否需要均注意这两个地方？但出现违反的次数在第一联末是明显多于第二联末的，那被试是否会更关注第一联末，从而影响脑电的结果？

回应：本实验将第二联末出现违反的材料纳入实验(作为填充材料)，主要是为了平衡被试对绝句不同位置的关注。如果只有绝句内部出现违反，那么被试在发现绝句内部押韵合适或者违反时，就已经能够完成实验任务。绝句末违反的填充材料能让被试持续关注整个绝句(即使一首绝句第一联末押韵合适，仍然有可能在绝句末出现违反)，从而起到增加被试对诗句末位置关注的作用。受 ERP 实验时间不能过长的限制，我们选择违反出现在第二联末的材料确实少于第一联末，被试有可能更关注第一联末。不过，本实验中绝句末违反的材料仅作为本实验的填充材料，我们分析的为绝句内部存在违反和完全合适的材料。因此，被试即使更多关注第一联末，应该不会对当前我们分析的绝句末的脑电效应产生影响。

意见 9：作者可以把正确率和反应时按第一联末和第二联末分开，检查一下。而且为什么合适的条件比违反的条件正确率会低很多，而且反应时会慢很多？

回应：感谢审稿人对数据分析提出的建议。本文中的行为结果均是对第一联末的统计，没有包含第二联末(即填充材料)的反应结果。我们根据审稿人意见将正确率和反应时按第一联末和第二联末分开检查。对第一联末的总正确率的统计在文中已经给出，即‘对正确率进行独立样本 T 检验发现，音乐训练组($M = 83.24\%$, $SE = 6.64\%$)和对照组($M = 82.64\%$, $SE = 9.15\%$)

完成押韵合适性判断的正确率不存在显著差异, $t(48) = 0.27, p = 0.791$ 。对第一联末的总反应时进行独立样本 T 检验, 发现音乐训练组($M = 899.15, SE = 319.82$)和对照组($M = 873.07, SE = 300.29$)完成押韵合适性判断的正确率不存在显著差异, $t(48) = 0.30, p = 0.768$ 。因为第二联末为填充材料, 并没有对各个条件进行更为细致的标记, 所以仅能统计总的正确率和反应时。对第二联末的正确率和反应时进行独立样本 T 检验, 发现音乐训练组[正确率: $M = 79.81\%, SE = 6.69\%$; 反应时: $M = 1029.77, SE = 557.17$]和对照组[正确率: $M = 79.16\%, SE = 9.29\%$; 反应时: $M = 879.06, SE = 240.27$]完成押韵合适性判断的正确率[$t(48) = 0.29, p = 0.776$]和反应时[$t(48) = 1.24, p = 0.220$]均不存在显著差异。总之, 将第一联末和第二联末分开统计反应时和正确率, 发现两类被试在正确率和反应时上均无显著差异。

当前研究中合适条件比违反条件的正确率低、反应时慢的结果与以往研究一致(Schirmer, Tang, Penney, Gunter, & Chen, 2005, 使用高完型概率的粤语句子; Hu, Gao, Ma, & Yao, 2012, 使用成语; 及 Li, Wang, & Yang, 2014, 使用诗句)。当前实验对于合适条件和违反条件在正确率和反应时上产生的差异, 可能是因为, 被试只要探测到第一联末出现不合适情况, 就能够判定该绝句属于押韵违反, 相对容易; 但是对于合适情况则需要听完整个绝句, 确定绝句每一个地方都合适, 才能做出合适判断, 相对更难。以上内容我们已经增加在修改稿讨论部分第 26 页。

意见 10: 为什么发现的 100~300ms 的违反脑电反应会是更小的正波? 这和所有的违反范式的结果, 比如 MMN, prediction error 等都不一致。也许是因为选取电极不同做分析所致。因此用 GFP 的结果也许会更加清晰。

回应: 感谢审稿人提出的疑问。在 100~300 ms, 对照组没有诱发任何显著的效应, 音乐训练组在声调/韵母合适条件下, 对韵母/声调的一致性进行深入分析, 并诱发了更大的正波。在韵母违反条件下, 发现声调违反相比声调合适诱发了更小的正波(见第 23 页)。该结果应该不是由选取电极不同造成, 而是因为: 本研究中绝句末字不存在条件间差异, 虽然诱发的正波在时间上出现的较早, 但是实验控制的条件发生在绝句内, 所以它更可能反映的是对诗句内发生押韵违反的迅速探测和对韵律信息的初步整合。对于音乐训练组被试, 相比于仅一种违反(如韵母违反, 声调合适), 全违反条件(即韵母违反, 声调违反)下的整合难度较小, 诱发的正波波幅相应更小; 同时, 相较于全合适条件(韵母合适, 声调合适), 被试整合仅一种违反时(即韵母合适, 声调违反或韵母违反, 声调合适)消耗的认知资源更多, 诱发的正波波幅也更大。相反, 对照组在此时间窗口没有诱发任何显著的效应, 可能表明普通人无法在

绝句末迅速完成对绝句内出现的任何押韵违反的探测和整合(见第 29 页)。

我们使用 GFP 对全脑电极点进行分析,发现在 100-300 ms 声调违反($M = 1.64, SE = 0.07$)相比声调合适($M = 1.76, SE = 0.08$)诱发更小的正波, $F(1, 48) = 4.09, p = 0.049, \eta_p^2 = 0.08$ 。不存在其他显著的主效应和交互作用($ps > 0.116$)。如审稿人所述, GFP 的分析方法不考虑极性,是所有电极的几何平均,并且不受电极位置的影响。与传统计算方式相比, GFP 分析方法虽排除了电极点位置对数据的影响,但同时也掩盖了不同时间点诱发脑电效应极性上的变化,从而造成两种分析方法结果上的显著差别。本研究可能不适合采用此方法进行分析。

.....

审稿人 2 意见:

研究采用 ERP 技术,通过让被试判断七言绝句受否押韵的判断任务,考察了音乐训练组和普通组在绝句韵律信息的整合加工过程,具体意见如下:

意见 1: 采用 ERP 技术和声调、韵母违反范式研究言语韵律的整合加工过程,这方面的研究比较多,引言中并没有提及;

回应: 根据审稿人的意见,我们在修改稿引言部分增加了采用 ERP 技术和声调、韵母违反范式研究言语韵律的整合加工过程的研究:目前采用违反范式探究言语韵律的研究主要考察了声调、韵母信息违反的即时加工过程。大多研究采用行为和脑电技术,使用不同的材料(如,单字对、成语、高限制语义句子、诗句等)和不同的实验任务(如,词汇判断、语义判断、词汇决策等),探究声调和韵母在词汇识别和限制语义整合加工过程中的时间进程(Hu, Gao, Ma, & Yao, 2012; Huang, Liu, Yang, Zhao, & Zhou, 2018; Li, Wang, & Yang, 2014; Schirmer, Tang, Penney, Gunter, & Chen, 2005)。虽然尚未得出一致结论,但是大多数研究发现相较于声调信息,韵母信息的加工相对更快;即使母语是声调语言者,韵母在词汇识别和语义限制方面仍具有更强的影响力和约束力(Tong, Francis, & Gandour, 2008; Hu et al., 2012; Huang et al., 2018; Li et al., 2014)(详见第 18 页)。

意见 2: 为什么要选用音乐韵律,音乐韵律和言语韵律的区别在哪?文中没有阐述;

回应: 谢谢审稿人提出的疑问。通过仔细思考,我们认为在文章中使用音乐韵律的说法欠妥。已经删除该类说法。在修改稿中我们重新阐述了此部分内容(修改稿第 16 页):“音乐同语言一样,作为人类获取听觉信息的主要来源,可以通过音调的高低、节奏的快慢等特征实现表情、表意、表态的功能(Dilley, Mattys, & Vinke, 2010; Juslin & Laukka, 2003; Patel & Daniele,

2003; 蒋存梅, 2016)。有研究发现作曲家对音乐旋律的创作受言语韵律的影响(Patel, Iversen, & Rosenberg, 2006), 具有音乐障碍的先天性失乐症患者在感知、识别言语韵律时更加困难(Liu, Patel, Fourcin, & Stewart, 2010; Nan, Sun, & Peretz, 2010), 表明音乐和语言在韵律方面有着诸多联系(Hausen et al., 2013; Patel, Peretz, Tramo, & Labreque, 1998; Patel & Daniele, 2003)。”。

意见 3: 在绝句选取方面, 有的绝句在押韵方面特别明显, 例如第一句、第二句和第四句的最后一个字分别是桃、高、袍, 但是有些绝句可能受到历史演变的影响, 如今看来并不是特别明显, 例如第一句、第二句和第四句的最后一个字分别是迎、生、行, 在对绝句熟悉度低的情况下, 被试对这类绝句在改变韵母情景下的整合加工与前一类语料相比是否会存在区别? 在正式实验前, 语料是否找人进行过押韵与否的合适度判断;

回应: 本研究的实验材料均从现存的七言绝句中选取。韵母完全相同且熟悉度低的诗句数量有限。因此我们使用的绝句除了第一句、第二句、第四句最后一个字的主要韵母相同的材料外, 还包括了三种其他押韵方式的材料: eng, ing: 如“生, 轻, 行”; en, in: 如“侵, 阴, 深”; ou, iu: 如“流, 幽, 鸥”(占总材料 20.6%)(见修改稿第 20 页)。这三种类型的押韵在加工上可能相比“桃, 高, 袍”这种韵母部分完全相同的材料在整合上难度略有增加, 不过实验材料通过听觉方式呈现, 至少相比韵母违反条件, 被试更容易知觉到这些押韵的音之间的相似性, 同样也较容易理解并接受它们属于押韵合适, 因此我们倾向于认为被试在加工不同押韵类型材料时不会存在区别。

在正式实验前, 语料并未找人进行押韵与否的合适度判断。不过, 本研究中诗句押韵与否有较为明确和客观的标准, 实验材料也均符合标准。在练习中, 我们将上述三种其他押韵类型材料增加在练习实验中, 并告知被试受历史影响这些诗句仍属于押韵合适(见第 21 页)。所以我们倾向于认为被试掌握了押韵合适度判断的标准, 受到不同押韵类型的影响较小。

意见 4: 本文的被试是音乐训练组和普通组, 音乐训练组被试的筛选依据是音乐训练的年龄和训练时长, 普通组的被试是否进行过筛选? 有没有使用蒙特利尔失歌症测试表对普通组的音高感知表现进行评定, 排除普通组中的一些失歌症患者和对音高感知特别敏感的被试;

回应: 感谢审稿人提出的疑问。参照以往探究音乐训练对韵律加工影响的研究(Zhang, Jiang, Zhou, & Yang, 2016), 本研究中对照组的筛选依据就是没有接受过系统音乐训练, 并且在年

龄和性别与音乐训练组匹配的普通人。虽然并没有进行失歌症者和音高特别敏感者的筛选，但是我们认为对研究结果影响不大。有以下几点原因：1、失歌症者比较罕见，有研究指出在 117 名中国大学生中失歌症发生率仅有 3.4% (Nan, Sun, & Peretz, 2010)。绝对音高者更为罕见，在西方国家，绝对音高者的发生率是 0.1‰ (Takeuchi & Hulse, 1993)。虽然在亚洲的发生率较高，但更多的出现在音乐专业学生中 (Gregersen, Kowalsky, Kohn, & Marvin, 2001; Miyazaki, 2004)。因此 25 名普通被试存在失歌症和绝对音高者的可能性极小。2、有研究发现并非所有的失歌症者在感知词汇声调方面存在不足，汉语普通话失歌症者中约仅有三分之一的人存在声调感知缺陷 (Jiang et al., 2012; Nan, Huang, Wang, Liu, & Dong, 2016)，表明有声调感知缺陷的失歌症者在普通人群中占比极少，其次音乐中的音高和语言中的声调加工可能存在不同的神经基础。本研究主要探究被试间加工声调而非音乐音高的神经过程，因此被试是否有失歌症或绝对音高能力对声调加工的结果影响不大。3、观察对照组声调违反条件下的行为结果发现 25 名被试的变异较小 ($M = 77.97\%$, $SD = 5.57\%$)，没有明显表现更好和更差的被试。

参考文献：

- Peretz, I. (2001). Brain specialization for music: New evidence from congenital amusia. *Annals of the New York Academy of Sciences*, 930(1), 153-165.
- Jiang, C., Hamm, J. P., Lim, V. K., Kirk, I. J., & Yang, Y. (2011). Fine-grained pitch discrimination in congenital amusics with Mandarin Chinese. *Music Perception: An Interdisciplinary Journal*, 28(5), 519-526.
- Jiang, C., Hamm, J. P., Lim, V. K., Kirk, I. J., & Yang, Y. (2010). Processing melodic contour and speech intonation in congenital amusics with Mandarin Chinese. *Neuropsychologia*, 48(9), 2630-2639.
- Jiang, C., Hamm, J. P., Lim, V. K., Kirk, I. J., & Yang, Y. (2012). Impaired categorical perception of lexical tones in mandarin-speaking congenital amusics. *Memory & Cognition*, 40(7), 1109-1121.
- Nan, Y., Sun, Y., & Peretz, I. (2010). Congenital amusia in speakers of a tone language: association with lexical tone agnosia. *Brain*, 133(9), 2635-2642.
- Nan, Y., Huang, W. T., Wang, W. J., Liu, C., & Dong, Q. (2016). Subgroup differences in the lexical tone mismatch negativity (mmn) among mandarin speakers with congenital amusia. *Biological Psychology*, 113, 59-67.
- Tillmann, B., Burnham, D., Nguyen, S., Grimault, N., Gosselin, N., & Peretz, I. (2011). Congenital amusia (or tone-deafness) interferes with tone language perception. *Frontiers in Psychology*, 2,
- Liu, F., Jiang, C., Thompson, W. F., Xu, Y., Yang, Y., & Stewart, L. (2012). The mechanism of speech processing in

congenital amusia: Evidence from Mandarin speakers. *PLoS ONE*, 7, e30374.

Burnham, D. K., & Brooker, R. (2002). Absolute pitch and lexical tones: Tone perception by non-musician, musician, and absolute pitch non-tonal language speakers. In *Seventh International Conference on Spoken Language Processing*.

Takeuchi, A. H., & Hulse, S. H. (1993). Absolute pitch. *Psychological bulletin*, 113(2), 345.

Miyazaki, K. I. (2004). Recognition of transposed melodies by absolute pitch possessors. *Japanese Psychological Research*, 46(4), 270-282.

Zhang, J., Jiang, C., Zhou, L., & Yang, Y. (2016). Perception of hierarchical boundaries in music and its modulation by expertise. *Neuropsychologia*, 91, 490-498.

意见 5: 条件操控方面，韵母违反具体是怎么操控的？韵母可以细分为介音、主要元音和韵尾，修改其中的任何一样都可以是违反，实验研究中修改介音、主要元音和韵尾的都有，有没有具体的修改标准？

回应: 谢谢审稿人提出的疑问。在修改稿中，我们在方法部分进一步阐述了韵母违反的修改标准(见第 20 页)：在我们操纵韵母违反时并没有细致地从介音、主要元音和韵尾的角度考虑，仅是以违反押韵规则为标准将韵母看作一个整体进行变化，使其与符合押韵规则的韵母在感知上存在显著差别。

意见 6: 普通组加工诗句韵律的整合过程和已有研究中加工言语韵律的过程是否有差别，论文中 没有将实验结果与已有的研究进行比较；

回应: 感谢审稿人的建议。目前尚没有研究考察普通组加工诗句韵律的整合过程。以往采用声调、韵母违反范式探究言语韵律的研究主要考察了声调、韵母信息违反的即时加工过程。大多数研究发现相较于声调信息，韵母信息的加工相对更快；即使母语是声调语言者，韵母在词汇识别和语义限制方面仍具有更强的影响力和约束力(Tong, Francis, & Gandour, 2008; Hu et al., 2012; Huang et al., 2018; Li et al., 2014)。与本研究最为相关的研究(Li, Wang & Yang, 2014, 使用普通被试)也发现了韵母信息的加工优势。研究者通过变化七言绝句第一联末字声调和韵母信息产生韵律违反,发现在 300-500ms 韵母违反相比韵母合适诱发了波幅更大的负波(N450); 在 600~1000ms, 声调违反相比声调合适在中后部诱发波幅更大的正波(LPC), 并且在声调合适条件下, 韵母违反相比韵母合适在后部诱发波幅更大的正波(LPC)。本研究主要考察了音乐训练组和对照组被试诗句末的韵律整合过程。对照组的结果显示, 在

300~750ms, 声调合适条件下, 韵母违反相比韵母合适诱发波幅更大的负波; 韵母合适条件下, 声调违反相比声调合适诱发波幅更大的负波。两研究结果比较, 可以发现在对诗句中出现的押韵违反信息的即时加工过程中, 人们探测韵母违反相较于声调违反更加快速, 声调和韵母的交互作用发生在晚期时间窗口(600~1000 ms)。诗句末韵律整合加工是对绝句内出现的声调、韵母信息违反进行的深入整合, 在绝句末关键字出现后 300 ms 即同时开始了声调和韵母的整合加工, 并诱发了中后部分布的持久负波。总之, 对照组诗句末韵律的整合过程不同于诗句内韵律信息的即时加工, 体现了诗句末韵律加工的独特过程。以上内容已经增加到修改稿的讨论部分第 27 页。

意见 7: 三种条件操控下, 即声调违反、韵母违反, 声调和韵母同时违反, 音乐训练组和普通组 对诗句韵律整合加工的神经差异并没有在讨论中解释;

回应: 已经根据审稿人的意见, 在修改稿的讨论中增加两组被试对诗句在不同条件下诗句韵律整合加工的神经差异(详见修改稿第 29 页): 本研究中绝句末字不存在条件间差异, 虽然诱发的正波在时间上出现的较早, 但是实验控制的条件发生在绝句内, 所以它更可能反映的是对诗句内发生押韵违反的迅速探测和对韵律信息的初步整合。对于音乐训练组被试, 相比于仅一种违反(如韵母违反, 声调合适), 全违反条件(即韵母违反, 声调违反)下的整合难度较小, 诱发的正波波幅相应更小; 同时, 相较于全合适条件(韵母合适, 声调合适), 被试整合仅一种违反时(即韵母合适, 声调违反或韵母违反, 声调合适)消耗的认知资源更多, 诱发的正波波幅也更大。相反, 对照组在此时间窗口没有诱发任何显著的效应, 可能表明普通人无法在绝句末迅速完成对绝句内出现的任何押韵违反的探测和整合。

不过, 对照组在绝句末也会对绝句内发生的押韵违反进行深入分析, 体现为声调和韵母合适时, 韵母和声调不一致会诱发一个 300ms 开始出现的类似于 N400 的负效应。不管该效应反映了语义(Magne et al., 2007; Marie et al., 2011; Perrin & García-Larrea, 2003)还是韵律(Chen et al., 2016; Li et al., 2014)的整合加工, 当前结果与以往考察声调和韵母即时加工的研究诱发了相似的效应(Schirmer, Tang, Penney, Gunter, & Chen, 2005; Hu, Gao, Ma, & Yao, 2012; Li, Wang, & Yang, 2014)。在此时间窗口, 音乐训练组仅在韵母和声调违反时才会继续对声调、韵母一致性进行深入分析。这表明音乐训练组在上一阶段已经基本完成了韵律信息的整合; 在此时间窗对声调和韵母违反的深入分析则体现了他们对韵律信息的感知加工更为精细。研究发现长期的音乐训练可以促进工作记忆广度和刷新能力的提升(Nutley, Darki, & Klingberg, 2014; Slevc et al., 2016), 包括言语工作记忆(Clayton et al., 2016; Hansen, Wallentin,

& Vuust, 2013; Roden, Kreutz, & Bongard, 2012)。为了完成当前押韵判断任务，听者需要动态存储诗句押韵信息并且在诗句末提取相关信息做出行为反应。音乐训练组可能相比对照组能更有效调动储存在工作记忆的信息，从而完成对声调和韵母更为精细和深入的加工过程。

以往采用被动 *oddball* 范式，探究音乐训练影响前注意阶段声调加工的研究发现，音乐训练组在声调变化时相比普通人会诱发更大的 MMN、P3a 成分(Nan et al., 2018; Tang et al., 2016)。当前研究中，我们仅在音乐训练组发现声调违反条件下的反应时显著短于声调合适条件；即使韵母违反，音乐训练组仍会深入加工声调的合适性，体现在早期的正效应和晚期的负效应上。这可能表明在有意识注意状态下音乐训练组对声调信息仍然更为敏感。以往研究发现，不论音乐训练者的母语为声调还是非声调语言，其声调识别能力均优于非音乐训练者(Delogu, Lampis, & Belardinelli, 2010; Gottfried, Staby, & Ziemer, 2004; Cooper & Wang, 2012; Tang, Xiong, Zhang, Dong, & Nan, 2016)。本研究结果则进一步表明，无论是在即时的声调加工过程中，还是在延迟的整合加工时，音乐训练者均相比普通人对声调投入更多注意资源，也更为敏感。

意见 8： 论文的理论依据以及在理论意义上的阐述还较欠缺。

回应： 根据审稿人建议，在修改稿中进一步细化了文章的理论依据(见修改稿第 17 页)：“长期的音乐训练不仅提高了对低水平声学信息加工的敏感性，而且对一些高级认知加工起促进作用，如注意(Wang et al., 2015)、言语记忆(Franklin et al., 2008)、执行功能(陈杰, 刘雷, 王蓉, 沈海洲, 2017)、语音分段(François, Chobert, Besson, & Schön, 2012)、语言表达技能(Milovanov, Huotilainen, Välimäki, Esquef, & Tervaniemi, 2008)等。基于以往结果，*OPERA* 假说解释了为何音乐训练可能通过共享的神经资源对语言加工起促进作用：当音乐对这两个领域共享的感官和认知加工机制提出的要求比语音更高时，音乐能力可以增强语音加工；这些更高要求的结合以及与音乐能力相关的情感奖励，重复和注意激活了神经可塑性，改变了影响语音加工的神经结构和功能，从而促进语言加工(Patel, 2014)。因此，长时间系统的音乐训练可能提高人们对言语韵律的敏感性。”。

根据审稿人意见，在讨论部分增加本研究的理论意义(第 30 页)：“*OPERA* 理论指出音乐训练使得音乐训练者在基础声学线索的感知、一般认知能力等方面有很大的提高，并且伴随脑结构和功能的变化，为音乐能力在语言领域的迁移提供基础。本研究发现长期的音乐训练影响了人们对诗句韵律整合加工，使得音乐训练组在脑皮层水平上表现出了更加迅速、精细的神经整合过程。音乐训练在语言领域的这一迁移效应支持了 *OPERA* 理论，并且将现有

探究音乐训练促进言语韵律信息即时加工的问题拓展到了对整体韵律信息的整合加工中。此外，本研究在有意识的主动注意状态下发现母语为声调语言的音乐训练组仍具加工优势，表明音乐训练的影响广泛，不仅体现在简单的声音感知方面，还体现在更加复杂、高级的认知加工过程。”

第二轮

审稿人 1 意见：

作者按照评审意见进行了相应修改，论文有了较大提升，建议发表。

审稿人 2 意见：

作者已经回答了我的所有问题。谢谢。

编委复审

同意该稿件发表

主编终审

经审阅，Xb19-491 达到了发表的水平，同意发表