

《心理学报》审稿意见与作者回应

题目：人类镜像系统参与音乐情绪的自动加工：来自 EEG 的证据

作者：赵怀阳 江俊 周临舒 蒋存梅

第一轮

审稿人 1 意见：

意见 1：本文“人类镜像神经元系统参与音乐情绪的加工：来自 EEG 的证据”试图利用 EEG 技术探究人类镜像系统在音乐情绪加工中的作用。研究思路新颖，但是文章的理论假设较为牵强，无论是实验设计还是实验结果都无法回答作者提出的科学问题。

回应：非常感谢评审专家提出的宝贵意见！基于已有研究尚未探讨人类镜像系统与音乐情绪自动加工的关系，本研究试图通过 EEG 手段探究这一科学问题。下面我们将从实验设计和实验结果两个方面，解释它们是如何回答本研究的科学问题。

在实验设计方面，我们使用跨通道的情绪启动范式考察本研究的科学问题，原因如下：

(1) 情绪启动是用于考察情绪加工的一种实验范式。该范式已被广泛运用到已有音乐情绪研究中(Goerlich et al., 2012; Logeswaran & Bhattacharya, 2009; Peynircioğlu, March, Surprenant, & Neath, 2013; Sollberge, Rebe, & Eckstein, 2003)。其理论依据源自激活扩散理论(Collins & Loftus, 1975)。也就是说，概念是以节点的方式储存在联结网络之中，对启动刺激概念表征的加工可以在概念水平上预先激活与概念相关的其他节点的表征，从而促进对相关表征的加工。类似地，在音乐情绪启动范式中，对音乐情绪的识别可以自动地激活与之相关的目标刺激的情绪表征，从而产生情绪启动效应 (Steinbeis & Koelsch, 2008, 2011)。基于此，以音乐作为启动刺激诱发的情绪启动范式可以反映出听者对音乐情绪的加工。更重要的是，在情绪启动范式下，由于启动刺激和目标刺激在一致和不一致两个条件都出现，两个条件的唯一区别仅在于启动刺激和目标刺激配对的方式（一致或不一致）不同，在这种情况下，对比一致和不一致条件（也就是情绪启动效应）可以排除刺激本身感知觉加工的影响。因此，情绪启动范式已成为音乐情绪知觉研究的理想实验范式，被用于音乐情绪加工的研究中 (Goerlich et al., 2012; Logeswaran & Bhattacharya, 2009; Peynircioğlu, March, Surprenant, & Neath, 2013; Sollberge, Rebe, & Eckstein, 2003)。

(2) 由于情绪的模仿具有自动化的特点(Cracco et al., 2018)，本研究的实验设计采用 200 ms 的 SOA，旨在探究自动的情绪启动效应。的确已有研究发现 SOA 为 200 ms 的时

候，体现出对音乐情绪的自动加工 (Steinbeis & Koelsch, 2008, 2011)。这是因为，当 SOA 时间过长，被试就会形成意识反应策略，这种策略将影响随后的反应(Musch & Klauer, 2003)。另外，与先前研究一致(Steinbeis & Koelsch, 2008, 2011)，被试的任务是判断目标刺激的情绪类型，并不要求被试有意识地对音乐刺激进行加工，由此揭示音乐情绪的自动化加工机制。

(3) 在以上两点的基础上，本研究采用听觉-视觉跨通道的情绪启动范式，以音乐和弦作为启动刺激，以情绪面孔图片作为目标刺激。这一实验设计主要由于情绪启动范式的 SOA 时间较短，因此不能采用音乐-音乐的同通道启动，否则目标刺激与启动刺激容易产生重合。

在实验结果方面，本研究聚焦于 μ 抑制波，并发现 μ 抑制波的 α 频段和 β 频段能量在两个实验条件下的差异，该结果表明，镜像系统参与到音乐情绪的自动加工。其原因如下：

首先，已有研究认为，大脑中线出现的 μ 抑制波是镜像系统活动的电生理指标(Fox et al., 2016; Hobson & Bishop, 2016; Muthukumaraswamy, Johnson, & Mcnair, 2004)。这一论断主要基于以下实证研究发现：(1) μ 抑制波的出现与镜像系统核心区域的激活具有相关。的确，一些 EEG-fMRI 的同步研究表明， μ 抑制波抑制活动与镜像系统相关脑区的血氧水平依赖 (BOLD)信号变化高度相关 (Arnstein, Cui, Keysers, Maurits, & Gazzola, 2011; Mizuhara, 2012; Yin, Liu, & Ding, 2016)。(2) μ 抑制波体现出镜像系统的活动规律。的确，在运动观察、运动执行和运动想象活动中，已有研究不仅发现了镜像系统相关脑区的激活(Caspers et al., 2010; Corbo & Orban, 2017; Sakreida et al., 2018; Simos et al., 2017)，而且也观察到 μ 抑制波的活动(Déy & Lepage, 2013; Fox et al., 2016; Hobson & Bishop, 2016; McFarland, Miner, Vaughan, & Wolpaw, 2000; Pfurtscheller, Brunner, Schlögl, & Lopes da Silva, 2006)。基于以上原因，大量研究通过分析 EEG 信号中 μ 抑制波的变化，考察镜像系统是否参与认知或情绪加工的问题(Arnett, Roach, Elzy, & Jelsone-Swain, 2018; Fox et al., 2016; McFarland et al., 2000; Muthukumaraswamy et al., 2004)。基于以上论述，本研究预期，音乐情绪自动加工将激活大脑中线的 μ 抑制波。具体来说，如果 β 频段与运动计划相关(X. Chen et al., 2016; Tzagarakis, Ince, Leuthold, & Pellizzer, 2010)，那么，我们预期情绪一致与不一致条件下 β 频段能量将会存在差异；如果 α 频段与动作执行相关 (Fox et al., 2016; Frenkel-Toledo et al., 2013; Muthukumaraswamy & Johnson, 2004)，那么，我们预期，在音乐聆听阶段， α 频段的能量会下降。

的确，我们的研究结果也验证了我们的假设。具体来说，就 β 频段而言，与情绪不一致条件相比，情绪一致条件诱发了更低的 β 能量。事实上，在一致与不一致两种条件下，目标刺激（情绪面孔）都是相同的，因此结果不可能是由于图片情绪的差异引起的，它体现出听

者对音乐情绪的自动加工。的确,已有研究发现, β 频段的抑制与运动准备有关(Cheyne, 2013; Kilavik et al., 2012; Zhang, Chen, Bressler, & Ding, 2008)。比如,当在做食指伸展时,若在食指上施加负重,此时需要准备肌肉收缩程度增强,在这种情况下,运动前 β 频段抑制则会增加(Stančák, Rimpl, & Pfurtscheller, 1997);类似地,在速度—准确率权衡任务中,与准确率反应相比,由于速度反应需要快速做出运动准备,由此体现为运动前 β 频段的抑制增强(Pastötter, Berchtold, & Bäuml, 2012)。可见,以上两个研究中 β 频段只在运动前有差异,该频段的抑制始于动作/运动执行前的准备(Cheyne, 2013)。在本研究中,当音乐情绪与面孔情绪一致时,由于被试无需再次进行相应情绪表征动作的准备工作,因而减少了动作表征的准备阶段;相反,当音乐情绪与面孔情绪不一致时,由于被试需要再次进行相应情绪表征动作的准备工作,因而增加了动作表征的准备阶段,体现在 β 频段的能量下降。

另一方面,就 α 频段而言,在音乐出现300~450 ms时,情绪一致与不一致条件下源于人类镜像系统的 α 频段能量均降低。尽管在该时间段内,音乐和面孔图片同时出现,但是该结果归因于对音乐情绪而不是面孔情绪的加工,这是因为面孔情绪加工所诱发的 α 频段能量变化一般是在情绪面孔呈现后的500 ms左右出现(A. Moore et al., 2012)。因此,这个时间窗口内 α 频段能量的变化不可能源于对面孔情绪的加工。同时,该结果也不是由音乐情绪和情绪面孔图片的整合引起的,因为跨通道信息整合主要体现在 γ 频段能量的变化上(Lense, Gordon, Key, & Dykens, 2014; Schneider, Debener, Oostenveld, & Engel, 2008)。因此,该频段的能量变化只能归因于音乐情绪的自动加工,的确,已有研究发现, α 频段的抑制与动作执行或模仿相关(D'Ásy & Lepage, 2013; Fox et al., 2016; Frenkel-Toledo et al., 2013; Hobson & Bishop, 2016)。在本研究中,无论在一致,还是不一致条件下,都出现 α 频段的能量下降,这表明,音乐刚开始就建立了情绪动作表征,这种运动表征体现在 α 频段的活动。

在本次修改稿中,我们已经对引言、方法,以及结果讨论部分进行了详细的修改(请见第15-17页,第24-25页)。

Cracco, Emiel, Bardi, Lara, Desmet, Charlotte, Genschow, Oliver, Rigoni, Davide, De Coster, Lize, . . . Brass,

Marcel. (2018). Automatic imitation: A meta-analysis. *Psychological bulletin*, 144(5), 453-500.

意见 2: EEG 技术是否是回答这个问题的最优手段也需要进一步讨论。

回应: 尽管已有一些研究采用 fMRI 探究音乐加工的神经机制(比如, Blood & Zatorre, 2001; Fujisawa & Cook, 2011; Koelsch et al., 2013),但是,从本质上说,音乐是一门时间的艺

术，音乐事件随着时间而展开，因此，从音乐心理探究来说，时间分辨率很高的 EEG 技术要比时间分辨率较低的 fMRI 技术更具有优势。另一方面，在听觉研究中，由于 fMRI 扫描会产生很大的噪音，这对于考察音乐情绪自动加工的研究来说是不利的，因此本研究采用 EEG 技术进行探究。

Blood, A. J., & Zatorre, R. J. (2001). Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. *Proceedings of the National Academy of Sciences*, 98(20), 11818-11823.

Fujisawa, T. X., & Cook, N. D. (2011). The perception of harmonic triads: an fMRI study. *Brain Imaging & Behavior*, 5(2), 109-125.

Koelsch, S., Skouras, S., Fritz, T., Herrera, P., Bonhage, C., Kussner, M. B., & Jacobs, A. M. (2013). The roles of superficial amygdala and auditory cortex in music-evoked fear and joy. *Neuroimage*, 81, 49-60.

意见 3: 作者利用启动范式（音乐作为启动刺激），对目标刺激（视觉信息）后的脑电响应进行分析，并根据时频分析的结果得到结论，认为音乐情绪加工与人类镜像系统的激活相关。但是这其中的逻辑不够明确，作者需要提供更多的证据或者理论来解释为什么认为时频结果反应的是音乐情绪而不是图片情绪的差异，或者是两者的交互？

回应: 非常感谢评审专家提出的宝贵建议！该问题与第一个问题相似。详细内容请参见我们的回应（第 1 至第 4 页）。下面我们从实验设计与实验结果两个方面简略解释为什么时频分析结果反映的是听者对音乐情绪的自动加工：

从实验设计来说，本研究考察音乐情绪的自动化加工。为了探究这一问题，我们采用 SOA 为 200 ms 的跨通道的情绪启动范式：以音乐和弦作为启动刺激，以情绪面孔图片作为目标刺激。这种方式的理论依据来源于激活扩散理论(Collins & Loftus, 1975)。也就是说，在音乐情绪启动范式中，对音乐情绪的加工可以自动地激活与之相关的目标刺激的情绪表征，从而产生情绪启动效应(Steinbeis & Koelsch, 2011)。在本研究中，音乐启动刺激和目标图片刺激在一致和不一致两个条件都出现，两个条件的唯一区别仅在于，启动刺激和目标刺激配对是否一致。这种关系意味着，一致和不一致条件相减得出的时频结果应该归因于音乐刺激的作用：它是促进，还是抑制了目标图片的加工？也就是说，音乐情绪的自动化加工是否启动目标图片的加工？

时频结果也可以进一步说明这一点。就 β 频段而言，时频结果显示，与情绪不一致条件相比，情绪一致条件诱发了更低的 β 能量。事实上，在一致与不一致两种条件下，目标刺激

(情绪面孔)都是相同的,因此结果不可能是由于图片情绪的差异引起的。就 α 频段而言,在音乐出现300~450 ms时,情绪一致与不一致条件下源于人类镜像系统的 α 频段能量均降低。尽管在该时间段内,音乐和面孔图片同时出现,但是该结果归因于对音乐情绪而不是面孔情绪的加工,这是因为面孔情绪加工所诱发的 α 频段能量变化一般是在情绪面孔呈现后的500 ms左右出现(A. Moore et al., 2012)。因此,这个时间窗口内 α 频段能量的变化不可能源于对面孔情绪的加工。同时,该结果也不是由音乐情绪和情绪面孔图片的整合引起的,因为跨通道信息整合主要体现在 γ 频段能量的变化上(Lense, Gordon, Key, & Dykens, 2014; Schneider, Debener, Oostenveld, & Engel, 2008)。综上所述,本研究时频结果反映的是音乐情绪而不是图片情绪的差异,或者是两者的交互。

在本次修改稿中,我们已经对这一问题进行了详细的解释(请见第24-25页)。

意见 4: 作者使用图片刺激前650~250 ms作为基线,但是该时间段内声音刺激仍在呈现。因此,目标刺激的脑电响应(尤其是 α 频段信息)会因为基线校正而被抵消,从而无法进一步确定观察到的目标刺激后的差异是由于基线校正还是由于刺激呈现导致的

回应: 谢谢评审专家提出的意见。在本研究中,由于SOA是200 ms,因此,在图片刺激前200ms,音乐开始呈现。也就是说,图片刺激前650~250 ms为静息状态。因此实验观察到的目标刺激后的差异归因于刺激呈现而不是基线校正。

意见 5: 建议作者在前言部分给出详细推理过程以方便读者理解 μ 抑制波是如何选作为镜像系统活动的电生理指标的,以及 μ 波在情绪加工中的变化是如何与镜像系统活动相关的。因为根据文章中提供的信息,不能确定 μ 波的变化是镜像系统活动的充要条件。

回应: 谢谢评审专家提出的宝贵意见。已有研究认为,大脑中线出现的 μ 抑制波是镜像系统活动的电生理指标(Fox et al., 2016; Hobson & Bishop, 2016; Muthukumaraswamy, Johnson, & McNair, 2004)。这一论断主要基于以下实证研究发现:(1) μ 抑制波的出现与镜像系统核心区域的激活具有相关。的确,许多EEG-fMRI的同步研究表明, μ 抑制波抑制活动与镜像系统相关脑区的血氧水平依赖(BOLD)信号变化高度相关(Arnstein, Cui, Keysers, Maurits, & Gazzola, 2011; Mizuhara, 2012; Yin, Liu, & Ding, 2016)。(2) μ 抑制波体现出镜像系统的活动规律。在运动观察、运动执行和运动想象等认知活动中,已有研究不仅发现了镜像系统相关脑区的激活(Caspers et al., 2010; Corbo & Orban, 2017; Sakreida et al., 2018; Simos et al., 2017),而且也观察到 μ 抑制波的活动(Désy & Lepage, 2013; Fox et al., 2016; Hobson & Bishop,

2016; McFarland, Miner, Vaughan, & Wolpaw, 2000; Pfurtscheller, Brunner, Schlögl, & Lopes da Silva, 2006)。基于以上原因,大量研究通过分析 EEG 信号中 μ 抑制波的变化,考察镜像系统是否参与认知或情绪加工的问题(Arnett, Roach, Elzy, & Jelsone-Swain, 2018; Fox et al., 2016; McFarland et al., 2000; Muthukumaraswamy et al., 2004)。

在此次修改稿中,我们在前言部分对 μ 抑制波是如何体现镜像系统活动进行了阐述(请见修改稿第 15 页)。也可参见我们的回应(请见第 2-3 页)。

意见 6: 讨论部分过于简单,结论的得出也多依赖于猜测。

回应: 谢谢评审专家提出的宝贵意见。在本次稿件中,我们对讨论部分做了详细的修改。我们结合已有的文献论证了本研究结果的可靠性,并增加了对研究结果的解释和论证。具体内容请参照修改稿的讨论部分(请见第 24-25 页)。

意见 7: 文中并没有具体指出样本量是如何确定的。

回应: 为了确保足够的统计检验力,我们事先使用 G*Power 3.1.9.2 软件(Faul, Erdfelder, Lang, & Buchner, 2007)确定样本量。根据已有同类研究(Steinbeis & Koelsch, 2011),我们预测本实验将有一个大的效应量($\eta^2 = 0.35$)。根据软件计算,如果实验要在 0.05 的显著性水平上达到 80%的统计检验力,样本量至少为 17。因此,我们招募了 18 名被试。但是,由于 3 名被试身体动作较大,导致采集到的信号较差,有效试次较少,因此数据分析不包含他们的数据。

意见 8: 频谱分析的时间窗和电极点是如何确定的?依据是什么?

回应: 在本研究中,频率分析的时间窗和电极点是根据已有研究结果,并结合目测确定。对于 α 频段来说,已有研究发现, α 频段大约在刺激呈现 300 ms 时出现,主要集中在大脑中线电极(Fu & Franz, 2014)。对于 β 频段来说,已有研究发现, β 频段在刺激呈现后 200~600 ms 出现于中线电极(X. Chen et al., 2016)。通过对本研究结果的观察,我们将 α 和 β 频段的时间窗口分别确定为启动刺激和目标刺激呈现后的 300~450 ms 之间,电极点确定在中线 Fz、Cz、Pz 三个电极点。我们已经在修改稿的方法部分解释了时间窗和电极点的选择依据(请见第 20 页)。

意见 9: 是否对音频材料评分的唤醒度进行评分? 是否能排除音乐材料唤醒度不同对实验结果差异性的影响? 是否在实验结束后让参加正式实验的被试对音乐刺激和视觉刺激再次进行评估, 以确定实验被试对刺激情绪的判断符合常模?

回应: 为了确定协和与不协和和弦在唤醒度上没有差异, 8 名没有受过音乐训练的被试(没有参加正式试验), 对每条和弦的唤醒度在 7 点量表上进行评价(1 非常安静, 7 非常激动)。配对样本 T 检验显示, 对协和和弦的愉悦度评价 ($M = 4.99, SD = 0.77$) 与不协和和弦 ($M = 4.88, SD = 0.85$) 不存在差异, $t(7) = 0.54, p = .610$ 。这表明两类和弦的唤醒度没有存在显著的差异。

遗憾的是, 在实验结束后, 被试并没有对音乐刺激和视觉刺激再次进行评估。但是, 由于音乐前测被试与正式实验被试均是随机选取, 且两组被试来自同一样本群, 我们推测, 正式实验被试的情绪判断与前测被试的判断相似。

意见 10: 全文存在大量书写不规范, 格式错误的问题。如字号, 字体不一致; 又如 2.5.1 频率分析: “..., 并以目标刺激前 -650-250 ms 为基线对分段进行校正”, 既然已经提到目标前, 650 之前无需加负号, 而且 250 ms 是指刺激前是刺激后 250 ms?

回应: 非常抱歉! 在修改稿中, 我们已经修正了以上提及的书写不规范之处, 并对文本中其他书写不规范之处进行了认真的检查和修改。另外, 250 ms 是指刺激前。

意见 11: 提供的图片质量不高。如图 2 中 (a) 和 (b) 图大小不一致, 不方便比较, 而且需要提供时间 0 点作为参考。

回应: 我们已经根据建议, 上传了更新后的图片(请见第 21-22 页)。

意见 12: 建议作者提供原始行为学数据。

回应: 按照《心理学报》投稿要求, 在目前阶段, 我们不能上传原始数据, 只有在论文被接受时, 才能上传。

.....

审稿人 2 意见:

意见 1: 如果面孔和音乐刺激的呈现顺序与本实验顺序相反, 可能解释会更有说服力。换言之, 面孔情绪的认识涉及镜像神经元是可以理解的, 这可能反映了模仿的过程; 但音乐情绪

的识别为什么会激活镜像神经元？音乐情绪的识别需要借助面孔模仿来实现？作者需要给出更为合理和明确的解释。

回应：感谢评审专家提出的宝贵意见。首先，本研究探究的问题是镜像系统是否参与音乐情绪的自动加工。为了探究该问题，我们采用 SOA 为 200 ms 的跨通道的情绪启动范式：以音乐和弦作为启动刺激，以情绪面孔图片作为目标刺激。这种范式的理论依据来源于激活扩散理论(Collins & Loftus, 1975)。也就是说，概念是以节点的方式储存在联结网络之中，对启动刺激概念表征的加工可以在概念水平上预先激活与概念相关的其他节点的表征，从而促进对相关表征的加工。类似地，在音乐情绪启动范式中，对音乐情绪的加工可以自动地激活与之相关的目标刺激的情绪表征，从而产生情绪启动效应(Steinbeis & Koelsch, 2011)。基于此，以音乐作为启动刺激诱发的情绪启动效应可以反映听者对音乐情绪的加工。更重要的是，在情绪启动范式下，由于启动刺激和目标刺激在一致和不一致两个条件都出现，两个条件的唯一区别仅在于启动刺激和目标刺激配对的方式（一致或不一致）不同，在这种情况下，对比一致和不一致条件（也就是情绪启动效应）可以排除刺激本身感知觉加工的影响。因此，情绪启动范式成为音乐情绪知觉研究中一种理想的实验范式，已广泛运用于音乐情绪研究中(Goerlich et al., 2012; Logeswaran & Bhattacharya, 2009; Peynircioğlu, March, Surprenant, & Neath, 2013; Sollberge, Rebe, & Eckstein, 2003)。

其次，如果改变刺激呈现顺序(先面孔后音乐刺激)，那么就会将本研究的研究问题改成“镜像神经系统是否参与到面孔情绪的自动加工”。也就是说，该范式的核心问题在于，启动刺激是否自动地激活与之相关的目标刺激的情绪表征，其中，“是否激活”的判断标准就是启动刺激是否促进或抑制目标刺激的加工，因此，尽管在情绪启动范式中，被试任务大多是识别目标刺激的情绪，但是，这种实验范式真正考察的是被试对启动刺激的加工。

最后，音乐情绪自动加工之所以与镜像系统相关，原因如下：

1) 音乐是一种情绪的语言或一门情绪的艺术(Pratt, 1948)。研究者认为，音乐表现情绪是通过模仿人类表达性行为(Kivy, 1990; Meyer, 1956)或内部心理状态(Koelsch, 2011, 2013)来实现的。同时，音乐诱发的情绪体验也是听者通过模仿情绪产生的(Juslin, 2013; Juslin & Västfjäll, 2008)。这些观点与模仿理论相一致(Leman, 2007)。该理论认为，音乐本质意义上就是一种模仿：通过模仿某种对象，传达音乐信息。如果听者就是通过模仿加工音乐情绪，那么，有理由推测，为了将音乐这种物理声音转化成具体的情绪表征，人类镜像系统可能参与音乐情绪的自动加工(Molnar-Szakacs & Overy, 2006)。此外，尽管有些研究发现，人类镜像系统与观看钢琴演奏(Haslinger et al., 2005; Hou et al., 2017)和敲击节奏 (J. L. Chen,

Penhune, & Zatorre, 2008)等活动相关,但是尚未有研究探讨人类镜像系统与音乐情绪的自动加工的关系。

2) 我们的研究结果也验证了我们假设。就 β 频段而言,与情绪不一致条件相比,情绪一致条件诱发了更低的 β 能量。事实上,在一致与不一致两种条件下,目标刺激(情绪面孔)都是相同的,因此该结果不可能是由于图片情绪的差异引起的,它体现出听者对音乐情绪的自动加工。的确,已有研究发现, β 频段的抑制与运动准备有关(Cheyne, 2013; Kilavik et al., 2012; Pastötter et al., 2012; Zhang et al., 2008)。比如,当在做食指伸展时,若在食指上施加负重,此时需要准备肌肉收缩程度增强,在这种情况下,运动前 β 频段抑制则会增加(Stančák, Riml, & Pfurtscheller, 1997);类似地,在速度—准确率权衡任务中,与准确率反应相比,由于速度反应需要快速做出运动准备,由此体现为运动前 β 频段的抑制增强 (Pastötter, Berchtold, & Bäuml, 2012)。可见,以上两个研究中 β 频段只在运动前有差异,该频段的抑制始于动作/运动执行前的准备(Cheyne, 2013)。在本研究中,当音乐情绪与面孔情绪一致时,由于被试无需再次进行相应情绪表征动作的准备工作,因而减少了动作表征的准备阶段;相反,当音乐情绪与面孔情绪不一致时,由于被试需要再次进行相应情绪表征动作的准备工作,因而增加了动作表征的准备阶段,体现在 β 频段的能量下降。

另一方面,就 α 频段而言,在音乐出现 300~450 ms 时,情绪一致与不一致条件下源于人类镜像系统的 α 频段能量均降低。尽管在该时间段内,音乐和面孔图片同时出现,但是该结果归因于对音乐情绪而不是面孔情绪的加工,这是因为面孔情绪加工所诱发的 α 频段能量变化是在情绪面孔呈现后的 500 ms 左右出现 (A. Moore et al., 2012)。因此,这个时间窗口内 α 频段能量的变化不可能源于对面孔情绪的加工。同时,该结果也不是由音乐情绪和情绪面孔图片的整合引起的,因为跨通道信息整合主要体现在 γ 频段能量的变化上(Lense, Gordon, Key, & Dykens, 2014; Schneider, Debener, Oostenveld, & Engel, 2008)。因此,该频段的能量变化归因于音乐情绪的自动加工。的确,已有研究发现, α 频段的抑制与动作执行或模仿相关(Désy & Lepage, 2013; Fox et al., 2016; Frenkel-Toledo et al., 2013; Hobson & Bishop, 2016)。在本研究中,无论在一致,还是不一致条件下,都出现 α 频段的能量下降,这表明,音乐刚开始就建立了情绪动作表征,这种运动表征体现在 α 频段的活动。(具体解释也可见对审稿专家 1 的第 3 个问题的回答)

在本次修改稿中,我们已经对前言和讨论部分进行修订(请见第 15-17 页,第 24-25 页),详细内容也可参见本回应(请见第 1-4 页)。

意见 2: 实验被试是未受过专业音乐训练的被试，听到时间很短（800 ms）的音乐片段会产生正面或者负面情绪？特别是实验还是启动范式的情况下。从一般读者的角度来看，似乎与常识不符。为什么不招募音乐专业被试完成实验？

回应: 非常感谢审稿专家的意见。事实上，情绪知觉与情绪体验是不一样的。的确，听到很短的音乐片段很难让人产生情绪体验，音乐情绪的体验通常需要 30 s (Eerola & Vuoskoski, 2013)，然而，情绪知觉或理解可以在很短时间内实现(比如，Steinbeis & Koelsch, 2008, 2011)。另一方面，已有研究表明，无论是音乐家还是非音乐家被试，他们都能准确地对 1 s 的音乐片段进行情绪分类(Bigand, Filipic, & Lalitte, 2005)。即便是对于时长只有 800 ms 的和弦刺激，音乐家和非音乐家都能区分协和与不协和和弦的愉悦度，而且两组被试对和弦愉悦度的评定没有显著差异(Steinbeis & Koelsch, 2011)。本研究的前测结果也验证，非音乐家能够区分 800 ms 的协和与不协和和弦的愉悦度。基于以上这些研究结果，本研究并不旨在考察音乐训练效应，而是以非音乐家为被试，考察镜像神经系统是否参与到音乐情绪的自动加工。

Bigand, E, Filipic, S, & Lalitte, P. (2005). The time course of emotional responses to music. *Annals of the New York Academy of Sciences*, 1060(1), 429-437.

Eerola, Tuomas, & Vuoskoski, Jonna K. (2013). A review of music and emotion studies: Approaches, emotion models, and stimuli. *Music Perception An Interdisciplinary Journal*, 30(3), 307-340.

第二轮

审稿人 1 意见:

意见 1: 为什么认为音乐情绪的加工涉及人类镜像系统，以及为什么以 μ 抑制波作为量化指标？

回应: 非常感谢评审专家提出的宝贵意见！很抱歉在上一稿中我们没有将这个问题阐述清楚。下面我们分别对这两个问题进行解释。

就音乐情绪加工与镜像系统的关系而言，我们的实验假设主要基于以下两个方面。第一，音乐情绪本质意义上就是模仿人类心理活动的变化(Koelsch, 2013)。这种论述主要基于音乐的模仿属性 (Leman, 2007)。由于对心理活动的模仿，导致音乐情绪的表达与人类的心理活动状态具有“异质同构”的关系(Langer, 1953)，比如，快乐音乐与大调式、音区宽广、协和的

和声、快速、音量大、音高上行等特点相关，悲伤音乐与小调式、音区较窄、慢速、音量小、轻度不协和、音高下行等特点相关(Juslin& Sloboda, 2013)。这些特征与人们在快乐或悲伤的行为特征相吻合。研究发现，当处于快乐的状态时，人们走得快、说话声音大、语调轮廓向上，且变化幅度较大；当处于悲伤的状态时，人们走得慢、说话声音小、语调轮廓变化向下，且变化幅度较小(Johnson-Laird & Oatley, 2016)。可见，音乐情绪与心理活动的“异质同构”主要基于声学要素变化与动作或运动的相关(Sievers, Polansky, Casey, & Wheatley, 2013)。最近一个脑成像研究进一步发现，当呈现上行或下行旋律时，听者对其意义的知觉会激活镜像系统(Jiang, Liu, Zhou, & Jiang, 2019)。该研究结果揭示出音乐声学要素知觉与镜像系统激活的相关。

第二，音乐情绪知觉会引起相应的生理反应，由此激活镜像神经系统。按照具身认知理论，知觉加工会“激活”生理变化(Cox, 2011; Maes, Leman, Palmer, & Wanderley, 2014)。的确，当人们判断动态信息（包括言语）情绪时，其面部也呈现出相应的表情(Hess & Blairy, 2001)。研究进一步发现，当听者对人声情绪进行判断时，与中性声音条件相比，情绪声音条件下镜像系统激活更强(Warren et al., 2006)。由于音乐情绪与情绪韵律(Juslin&Laukka, 2003)和情绪性的运动(Sievers et al., 2013)使用相似的声学线索表现情绪，因此我们预测音乐情绪加工也会激活镜像系统。的确，音乐也被认为是对人类面部表情、语调表情和姿态表情的模仿(Davies, 1994; Kivy, 1990; Koelsch, 2013; Leman, 2007)。研究也发现，音乐情绪理解会自动地诱发相应的面部表情动作(Chan, Russo, & Livingstone, 2013)。后续 EEG 研究进一步表明，当判断音乐情绪时，听者的镜像系统会出现 μ 抑制波现象(McGarry et al., 2015)。

以上阐述本研究为什么假设音乐情绪加工会激活镜像神经系统。的确，本研究发现的 μ 抑制波验证了我们的理论假设。具体论述请见修改稿第 21-23 页与第 30-32 页。

就 μ 抑制波而言，与已有研究一致（见 Fox, Bakermans-Kranenburg et al. 2016, Bowman, Bakermans-Kranenburg et al. 2017 综述），我们将其解释为镜像系统的电生理指标，主要基于以下原因：

首先， μ 抑制波与镜像系统的脑区具有相同的功能。镜像系统的最初研究认为，对同一动作的执行和观察都激活了相同的脑区，那么这些脑区就称为镜像系统的脑区(Iacoboni et al., 2005; Rizzolatti and Craighero, 2004; Molenberghs et al., 2012)。基于此，有研究者进一步探究了 EEG 活动是否也能反映出镜像系统脑区的功能。大量 EEG 研究证据表明，被试对相同动作进行观察和执行时， μ 抑制波的活动均增强，且脑电波的活动模式相似(Avanzini et al., 2012; Cannon et al., 2014; S. Cochin, ., Barthelemy, Roux, & Martineau, 1999; S. Cochin,

Barthelemy, Roux, & Martineau, 2010; Müller-Putz et al., 2007)。近年 2 篇元分析文章进一步验证, μ 抑制波的活动在动作观察和动作执行具有较高的稳定性(Fox, Bakermans-Kranenburg et al. 2016, Bowman, Bakermans-Kranenburg et al. 2017)。

后续研究进一步发现, μ 抑制波具有镜像系统的其他功能, 比如, 镜像系统参与了言语加工(Corina & Knapp, 2006; Lotto, Hickok, & Holt, 2009), 类似地, μ 抑制波也参与了语言加工(Guan, et al., 2013; Shebani & Puvemuller, 2013); 同时, 与动物动作相关的语言条件相比, μ 抑制波在人类动作有关的语言条件下(如 “The athlete jumped over the fence”)的活动更强(Guan, et al., 2013)。镜像系统也参与了社会交流(Dapretto et al., 2006; Le Bel, Pineda, & Sharma, 2009), 类似地, 研究也发现, 无论在移情(Carr, Iacoboni, Dubeau, Mazziotta, & Lenzi, 2003 ; Leslie, Johnson- Frey, & Grafton, 2004), 还是在心理理论(Williams et al., 2006 ; Williams, Whiten, Suddendorf, & Perrett, 2001)中, 均发现 μ 抑制波的活动。

第二, μ 抑制波与镜像系统脑区具有相同的活动脑区。EEG-fMRI 同步研究发现, 无论在动作观察(Arnstein, Cui, Keysers, Maurits, & Gazzola, 2011; Braadbaart, Williams, & Waiter, 2013), 还是在动作执行(Arnstein et al., 2011; Mizuhara, 2012)中, 镜像系统脑区 BLOD 信号的增加都与 α 和 β 频段的抑制具有显著相关, 这表明 μ 抑制波与镜像系统脑区具有相似的活动规律。此外, 动作执行和观察中诱发 μ 抑制波的源估计结果显示, 其发生源均集中于感觉运动区中央沟周围 (Hari, Salmelin, Makela, Salenius, & Helle, 1997; Salmelin & Hari, 1994a, 1994b)以及大脑皮层的中心和顶叶区域(Salmelin, Hamalainen, Kajola, & Hari, 1995; Manshanden, De Munck, Simon, & Lopes da Silva, 2002; Thorpe, Cannon, & Fox, 2015), 的确, 在同类任务(动作执行和观察)下, fMRI 研究也发现这些镜像系统脑区的激活(Molenberghs et al., 2012)。此外, 脑磁图(MEG)研究结果也验证了 μ 抑制波源定位在镜像系统的脑区(Hari, 2006; Muthukumaraswamy & Singh, 2008)。以上证据表明 μ 抑制波与镜像系统脑区具有相同的活动脑区。

第三, μ 抑制波活动与镜像系统脑区之间存在因果联系。研究进一步使用经颅磁刺激(TMS)的研究方法验证 μ 抑制波活动与镜像系统脑区之间的因果联系(Berntsen, Cooper, & Romei, 2017; Keuken et al., 2011)。这些研究分别通过对额下回(Keuken et al., 2011)或顶下小叶(Berntsen et al., 2017)施加抑制的磁刺激, 观察被试诱发的 μ 抑制波活动情况。结果均显示, 在抑制条件下, μ 抑制波活动显著减弱或消失, 这些结果均表明, μ 抑制波的活动与镜像系统的激活具有因果关系(Berntsen et al., 2017; Keuken et al., 2011)。

基于以上三个方面的研究证据以及已有研究的观点, 本文将 μ 抑制波作为人类镜像神

经系统的电生理指标。在本次修改稿中，我们已经在引言部分对这两个问题进行进一步的补充（第 19-22 页）。

Corina, D. P., & Knapp, H. (2006). Sign language processing and the mirror neuron system. *Cortex*, 42(4), 529-539.

Dapretto, M., Davies, M. S., Pfeifer, J. H., Scott, A. A., Sigman, M., Bookheimer, S. Y., & Iacoboni, M. (2006). Understanding emotions in others: mirror neuron dysfunction in children with autism spectrum disorders. *Nature neuroscience*, 9(1), 28.

Le Bel, R. M., Pineda, J. A., & Sharma, A. (2009). Motor–auditory–visual integration: the role of the human mirror neuron system in communication and communication disorders. *Journal of communication disorders*, 42(4), 299-304.

Lotto, A. J., Hickok, G. S., & Holt, L. L. (2009). Reflections on mirror neurons and speech perception. *Trends in cognitive sciences*, 13(3), 110-114.

意见 2: 考虑到具体实验中仅用到了协和和非协和的和弦，不建议使用“音乐”作为目的关键词，而且实验材料本身可能无法反映作者想要强调的“音乐情绪”。而且，非协和音和音乐情绪中的悲伤和愤怒等负性情绪也不能等同，作者需要针对这一点进行讨论和解释。

回应: 感谢评审专家提出的宝贵意见。根据专家建议，我们已经将文中的关键词“音乐”换成了“和弦”，同时也对文中相关措辞进行了修改。

的确，就负性情绪而言，悲伤和愤怒情绪在唤醒度上并不相同，但是，本研究旨在从效价维度上将协和和弦与不协和和弦区分开来，主要原因在于，在情绪二维模型中，效价维度被认为是人类的一个基本的，且普遍的情绪体验(Lindquist et al., 2016)，与之相反，唤醒维度概念很模糊，它在不同测量中具有不同的操作定义，如注意力的提高，行为接触，情感体验的强度，生理激活，或激活的主观感受等(Lindquist et al., 2016)。由于本研究采用单和弦（协和/不协和）作为音乐刺激，较难激活唤醒维度上的显著变化，因此适用于探究情绪效价上的差异(Dellacherie, Roy, Hugueville, Peretz, & Samson, 2011)。在这次修改稿中，我们已经在刺激部分对此进行说明（第 24 页）。

意见 3: 作者选用了跨通道情绪启动范式并提供了选择的原因。但是，这种范式涉及的心理过程不仅包括情绪加工，还包括了情绪冲突加工。如何确定所得结果仅仅是反映了音乐情绪

的加工？

回应：感谢评审专家提出的宝贵意见。虽然跨通道情绪启动范式的确可能包含情绪冲突加工，但是本实验结果与情绪冲突加工无关，主要原因如下：一方面，本实验的 SOA 为 200 ms，这种较短的 SOA 与情绪的自动化加工相关。的确，已有研究发现，较短的 SOA 只能易化情绪加工，不太可能产生情绪加工的冲突，这可能因为，冲突更多体现在意识层面的加工(de Groot, 1984; Neely, 1977)。另一方面，如果本实验结果是由于情绪冲突加工引起的，那么我们应该观察到位于额中线的 θ 频段活动。因为大量脑电研究表明，冲突加工与额中线的 θ 频段活动有关(Cavanagh & Frank, 2014; Cohen & Cavanagh, 2011; Cohen & Ridderinkhof, 2013; Ertl, Hildebrandt, Ourina, Leicht, & Mulert, 2013; Nigbur, Cohen, Ridderinkhof, & Stürmer, 2012; Tang, Hu, & Chen, 2013; Zavala et al., 2013)，但是，在本研究中，我们并未发现该频段活动的变化。

在这次修改稿中，我们已经在讨论部分对此进行了说明（第 30 页）。

意见 4：全文关于时间窗的描述让人难以理解。比如 2.5.1 最后一段：“我们将 α 和 β 频段的时间窗口分别确定为启动刺激和目标刺激呈现后的 300~450 ms 之间”，所以 α 频段信息指的是听觉刺激呈现完毕后（offset）300~450 ms 内 8-13 Hz 的信息， β 频段的信息是视觉刺激呈现后（onset）300~450 ms 内的 15-25 Hz 信息吗？如果是这样作者应将“启动刺激呈现后”这种表述改为“启动刺激呈现结束后”，否则会让读者误解。还是说，这里指的都是视觉刺激呈现后的 300-450 ms？

回应：非常感谢审稿专家的意见。很抱歉我们在上一稿中没有表述清楚。事实上， α 频段信息指的是听觉刺激呈现开始后（onset）300~450 ms 内 8-13 Hz 的信息， β 频段的信息指的是视觉刺激呈现后（onset）300~450 ms 内的 15-25 Hz 信息。在这次修改稿中，我们已经在结果部分对此进行了说明（第 26 页）。

.....

审稿人 2 意见：

意见 1：同意作者针对审稿意见所做的回复和修改。

回应：谢谢审稿专家！