

《心理学报》审稿意见与作者回应

题目：面孔宽高比的自动加工

作者：汪海玲 陈恩光 连玉净 李晶晶 王丽薇

第一轮

审稿人 1 意见：

该研究采用失匹配负波为指标，采用了中性面孔、愤怒面孔和恐惧面孔，考察了人们是否能够自动加工面孔宽高比，以及这些情绪的影响。本研究有一定的研究意义，实验设计控制比较严格，然而本论文还存在以下问题。

意见 1：本研究的逻辑比较清晰，但结果非常繁杂。如果在报告结果前对各个 ERP 成分的意义有进一步说明(比如，为何要测 P1，预期结果如何；为何测 N170，预期结果如何；为何能用失匹配负波来检测自动化加工？)，会使得逻辑更清晰，帮助不熟悉这些成分的读者更容易读懂结果。

回应：非常感谢审稿人对本研究的肯定和宝贵意见！全文所有修改部分已用红色字体标注。为了使本研究的逻辑更清晰，结果易解读，我们对前言和结果部分都进行了调整。

(1)在前言部分，我们把原来首段关于 vMMN 的介绍移至面孔宽高比内容介绍完，根据面孔宽高比已有研究，我们提出本研究要解决的问题“虽然上述研究从不同视角探讨了面孔宽高比与个体攻击性行为水平间的关系，但是，关于面孔宽高比自身的加工机制知之甚少，尤其是其自动加工机制，这对于理解个体快速识别威胁性信号有重要的社会价值。”在介绍完 vMMN 内容后我们增加了一段说明关于面孔研究中 P1 和 N170 成分的内容，也在此增加了本研究关于 P1 和 N170 对面孔宽高比敏感性的预期结果。

(2)为了增加结果的可读性，根据审稿人后续的意见，我们在结果图 2 和 3 中补充了其他条件出现 vMMN 的时间段，同时在两个实验结果的文字描述部分也增加了我们对 vMMN 的分析思路，补充了一些便于对实验结果理解的说明。

意见 2：另外，如果将失匹配负波的值(偏差刺激与标准刺激的差值)作为因变量，有关的结果可能更好理解，更能体现是否支持假设。比如，在引言中预期实验一得到高 fWHR 诱发的 vMMN 大于低 fWHR 的。从图 2 看，似乎确实得到了这样的结果，但正文却并没有报告这样的结果，而是集中在报告是否诱发了 vMMN。同理，如果实验 2 得到情绪和 fWHR 交互影响 vMMN，则能更直接得到“愤怒情绪促进高 fWHR 的自动加工，而恐惧情绪促进低 fWHR 的自动加工”的符合实验 2 假设的结论。

回应：非常感谢审稿人的意见，抱歉由于结果部分描述的不清晰而使得重要结果不容易被捕捉到。审稿人提到的实验 1 和实验 2 中与预期对应的实验结果在原有的文稿中都有提到，为了使重要结果更加清晰，我们对结果部分的呈现内容进行调整，希望能使读者准确、清晰地找到我们的重要结果。首先，我们在两个实验 vMMN 分时间段结果报告前增加了下面内容，用来说明我们的分析思路。实验 1 “如果能得到刺激类型与其他自变量间显著的交互作用，尤其是 fWHR × 刺激类型间显著的交互作用，并且简单效应分析能得到高、低 fWHR 都能诱发 vMMN (偏差刺激诱发的波幅较标准刺激更负)，那么我们将以 vMMN (偏差刺激诱发的波幅减去标准刺激诱发的波幅)为指标做进一步分析，考察高、低 fWHR 自动加工程度的差异。”

实验 2 “能反映本实验预期的结果是情绪 \times fWHR \times 刺激类型出现显著的交互作用, 并且简单效应分析能得到愤怒和恐惧情绪高、低 fWHR 都能诱发 vMMN(偏差刺激诱发的波幅较标准刺激更负), 那么我们将以 vMMN (偏差刺激诱发的波幅减去标准刺激诱发的波幅)为指标做进一步分析, 考察愤怒和恐惧情绪高、低 fWHR 自动加工程度的差异。”其次, 根据前面的分析思路, 我们在具体结果部分增加了“以 vMMN 为指标进行分析发现”的表述, 增加具体结果和研究假设结合的清晰程度。

意见 3: 实验 2 对于实验材料的选取控制得比较严格, 保证了愤怒情绪和恐惧情的高 fWHR 的值差不多, 都在 2 左右, 而低 fWHR 也差不多, 都在 1.5 左右。然而, 如果能给出具体的统计结果, 则更有说服力。两种情绪的情绪强度, 也同理。

回应: 非常感谢审稿人的意见! 我们分别在实验 1 和实验 2 的材料部分补充了中性、愤怒和恐惧情绪高低 fWHR 的统计结果, 具体表现为“中性 $t(4) = 8.669, p = 0.001, \text{Cohen's } d = 0.962$, 愤怒 $t(10) = 6.262, p < 0.001, \text{Cohen's } d = 0.874$, 恐惧 $t(10) = 6.568, p < 0.001, \text{Cohen's } d = 0.885$ ”。此外, 我们呈现了中性、愤怒和恐惧情绪高、低 fWHR 的情绪认同度和强度值, 并对两个实验分别做了相应的统计分析, 以及在两个实验对比时也做了相应分析, 但是所有结果未发现任何显著差异, 说明面孔情绪信息得到一定控制。

意见 4: 与第 3 点有关的一个问题是, 实验 1 中的高 fWHR 的值在 1.7 左右(与实验 2 中的两种情绪的 2 左右的值不同), 而低 fWHR 的值在 1.5 左右(与实验 2 类似)。那么, 进行实验间的比较是否合理? 是否是造成中性情绪和其他两种情绪差异的原因?

回应: 非常感谢审稿人的意见! 这个问题我们在对比实验一和实验二时也曾考虑过, 中性情绪与愤怒和恐惧情绪间宽高比的差异可能会影响两个实验间对比的可靠性, 因而进行两个实验具体结果对比前我们对中性情绪与愤怒和恐惧情绪间宽高比值的差异进行了 2 (fWHR: 低、高) \times 3 (面孔情绪: 中性、愤怒、恐惧) 的统计分析, 结果尚未发现二者的交互作用, 也就是高、低 fWHR 的值在不同情绪间不存在差异。具体内容请参考“4 实验一 vs 实验二”的第一段内容。

意见 5: ERP 数据的 trials 是被试正确反应的 trials 吗? 还是所有 trials?

回应: 根据已有 vMMN 的研究(Kecskés-Kovács et al., 2013; Stefanics et al., 2012; Wang et al., 2022), 在实验中设置三类刺激, 标准刺激、偏差刺激和目标刺激, 目标刺激是被试需要反应的刺激, 标准刺激和偏差刺激不需要被试反应, 而能反映 vMMN 成分的是标准刺激和偏差刺激, ERP 分析通过考察标准刺激和偏差刺激诱发的脑电波差异来得到 vMMN。由于标准刺激和偏差刺激不涉及被试反应正确与否。因而, 我们在“脑电记录与处理”部分只报告了标准和偏差刺激经过去伪迹等过程后的叠加 trials 数, 没有说明被试反应情况。

意见 6: 图 2A 中, P3/PO5/PO7 的结果也有显著的 vMMN, 也应该画上灰色阴影?

回应: 非常感谢审稿人的意见! 我们已在结果图 2 和图 3 中用灰色阴影标注了所有出现 vMMN 的实验条件, 增强主要结果的可视化, 具体请参考当前文稿中的图 2 和图 3。

意见 7: 建议作者在做出结论时更严谨。由于该研究没有让被试评价攻击性, 也并没有评价感知到的愤怒和恐惧情绪, 因此做出“本研究发现与个体感知攻击性水平密切相关的面孔宽高比信息能进行自动加工, 具体表现为, 与被评价为低攻击性的低 fWHR 相比, 被评价为高攻击性的高 fWHR 的自动加工时程较长, 程度较大。这可能与分别被感知为愤怒和恐惧情绪的作用有关”的结论时应明确区分哪些是已有文献的贡献, 哪些是本研究的贡献。

回应：非常感谢审稿人的意见！我们对结论部分进行了更正，“本研究不仅发现面孔宽高比信息能进行自动加工，而且发现高fWHR比低fWHR的自动加工时程更长，程度更大。”从而凸显本研究在面孔宽高比信息自动加工机制上的贡献。

意见 8：摘要中“面孔宽高比(facial width-to-height ratio, fWHR)在表征男性个体攻击性上存在差异”不通顺。

回应：面孔非常感谢审稿人的意见。我们已将上述不通顺的内容更改为“面孔宽高比(facial width-to-height ratio, fWHR)在表征或预测男性个体的攻击性行为水平上存在差异，高fWHR较低fWHR个体不仅有更强的攻击性，也被他人评价为有更高的攻击倾向。”，同时也对文章其他部分进行再次确认，更改不通顺的语句或错别字。

意见 9：“Merlhiot 等(2021)还发现面孔情绪也会影响面孔宽高比; 具体表现为, 与中性情绪面孔相比, 愤怒情绪使得面孔宽高比增加, 而恐惧情绪使得面孔宽高比降低。”面孔宽高比是一个客观的物理量, 如何受情绪的影响? 作者想要表达的是否是“被试知觉到的宽高比”? 还是“愤怒情绪面孔的宽高比大于恐惧情绪面孔的宽高比”?

回应：抱歉由于我们表述的不准确而给审稿人带来的困惑。Merlhiot 等(2021)实验一考察的是情绪对实际测量面孔宽高比的影响, 不是被试感知到的面孔宽高比大小。具体内容我们在前言部分进行了更正, 请参考前言第三段红色字体部分

.....

审稿人 2 意见：

本研究采用脑电技术，探究了非注意条件下高低 fWHR 引起视觉失匹配负波的差异，对于理解面孔宽高比加工的神经机制具有一定的意义。作者可以从以下几个方面进一步完善论文。

意见 1：在前言中，可以进一步凝练研究意义。作者提到对隶属于面孔构形信息的面孔宽高比自动加工的机制知之甚少, 从而以 vMMN 为指标研究面孔宽高比在前注意加工阶段的神经机制。这不足以成为本研究的亮点。包括在自检报告中，亮点应突出创新性贡献而不是简单罗列研究结果。

回应：非常感谢审稿人的宝贵意见和建议！全文所有修改部分已用红色字体标注。

(1)为了更好地突出本研究在已有研究基础上的创新点，引出本研究的问题，我们调整了前言部分的内容结构，把原来首段关于 vMMN 的介绍移至面孔宽高比内容介绍完，根据面孔宽高比已有研究，我们提出本研究要解决的问题“虽然上述研究从不同视角探讨了面孔宽高比与个体攻击性行为水平间的关系，但是，关于面孔宽高比自身的加工机制知之甚少，尤其是其自动加工机制，这对于理解个体快速识别威胁性信号有重要的社会价值。”最后总结本研究的目的是“总之，本研究主要以 vMMN 为指标研究面孔宽高比在前注意加工阶段的神经机制，同时结合 ERP 技术高时间分辨率的特点以期揭示高低面孔宽高比的自动加工机制在时序上的变化。”。

(2)根据审稿人的意见我们调整了自检报告中第一条研究亮点内容的表述，删除了本研究的结果，主要从研究问题和方法上突出了本研究的创新性。

意见 2：作者最好区分下感知到的以及实际的 fWHR，以及其与攻击性行为的关系。前言中提到“相比于拥有低 fWHR 的男性而言，拥有高 fWHR 的男性被认为更具有行为上的攻击性”。这是感知到的攻击性行为。紧接着作者又提到“这一结果可能与雄性激素睾酮有关，因

为辜酮与攻击行为有关。也就是说,辜酮可能既增大了 fWHR 也增强了个体的攻击性行为,进而使研究者得到高 fWHR 个体攻击性水平较高的结果”。这应该属于实际的攻击性行为。那么,辜酮到底解释了感知到攻击性行为还是实际的攻击性行为?

回应: 非常感谢审稿人的意见!已有研究发现面孔宽高比既影响个体实际的攻击性也影响个体被感知到的攻击性(Carré & McCormick, 2008; Carré et al., 2009), 这部分内容我们在前言和摘要部分都进行了修改。辜酮主要解释的是对个体实际攻击性的影响,但由于研究者发现个体实际攻击性水平与被感知的攻击性水平存在较高相关性(Carré et al., 2009), 因而辜酮也可能使得高 fWHR 个体被感知的攻击性水平也较高。具体内容请参考前言第 2 段红色字体部分。

意见 3: 此外,在实验 1vs.实验 2 部分,作者提到与中性情绪面孔相比,愤怒情绪会增加面孔宽高比,恐惧情绪会降低面孔宽高比(Merlhiot et al., 2021),但是,对本研究实验一和实验二使用的不同情绪面孔宽高比的值进行 2 (fWHR: 低、高) × 2 (面孔情绪: 中性、愤怒、恐惧) 的统计分析并未发现情绪对 fWHR 值的影响, $F(2, 4) = 1.119, p = 0.411$ 。按照 Merlhiot 论文题目,他们应该探究的是感知到的 fWHR,而本文测到的是实际的 fWHR,两者没有可比性。

回应: 抱歉我们对 Merlhiot et al. (2021)的研究没有表述清楚, Merlhiot et al. (2021)的实验一探究情绪对面孔宽高比的影响,用的是实际测量的 fWHR,而不是被试感知到的 fWHR。具体内容我们在前言部分进行了更正“Merlhiot 等(2021)通过测量同一个个体的不同情绪下的面孔宽高比发现,个体表现出来的面孔情绪也会影响其面孔宽高比大小;具体表现为,与中性情绪相比,愤怒和恐惧情绪分别使得眉心下移和上移,进而分别减少和增加了宽高比中的高度,而不影响宽高比中的宽度,因而,愤怒情绪使得面孔宽高比增加,而恐惧情绪使得面孔宽高比降低”。因而,我们在对比实验一和实验二时也对实际测量的不同情绪的 fWHR 进行统计。但是,我们发现与 Merlhiot et al. (2021)相反的实验结果,中性、愤怒和恐惧情绪的宽高比值不存在显著差异,这可能与我们的两个实验使用的是不同身份面孔有关,同时,这在一定程度上增加了我们两个实验间比较的有效性,中性和愤怒情绪高宽高比自动加工程度的差异较少或不受受不同情绪宽高比差异的影响。

意见 4: 研究方法可以提供更多的细节。比如,目标刺激呈现的是高还是低 fWHR 面孔?在实验 1 中,从 113 张中性面孔中最终得到 6 张实验刺激,是指一个标准差外的刺激只有这 6 张吗,还是进行了进一步删选?标准和偏差刺激中的各三张面孔呈现几次?刺激呈现顺序如何,随机的吗?高低 fWHR 情绪面孔在情绪认同度和情绪强度上有差异吗(这对于理解实验 2 的结果非常重要,即是不是情绪强度的差异而不是 fWHR 的差异影响结果)?为什么实验 2 不同刺激的比例与实验 1 不同,它会不会导致实验 1 和实验 2 结果的差异?如何计算波幅(峰基线值)和峰潜伏期?

回应: 非常感谢审稿人的详细建议!根据审稿人的意见,我们在两个实验的研究方法部分都补充了一些细节信息。

实验材料: (1)实验一经过面孔宽高比的测量最终得到标准差外高、低宽高比面孔各张,根据图片库提供的面孔情绪认同度和强度的信息,高、低宽高比在这两个维度上不存在显著差异,因而没有再进一步删选图片。实验二的材料也是同样的操作。(2)我们已经在两个实验材料部分补充了中性、愤怒和恐惧情绪高、低宽高比在情绪认同度和强度上的具体值和统计分析结果。我们对实验二的愤怒和恐惧情绪高低 fWHR 的认同度和情绪强度值都进行 2 (fWHR: 高、低) × 2 (面孔情绪: 愤怒、恐惧) 统计分析,结果未发现任何显著结果($ps \geq 0.359$);这说明,额外变量高、低宽高比面孔情绪强度和认同度的差异在一定程度上得到控制,较少或不影响我们对实验主要结果宽高比机制的分析。

实验程序：(1)目标刺激的宽高比与每个 block 中的标准刺激保持一致，我们在两个实验的所有 block 中都补充了相关信息。(2)实验一中标准和偏差刺激图片在每个 block 中各呈现 90 次和 10 次，实验二中标准和偏差刺激图片在每个 block 中各呈现 80 次和 20 次。(3)所有刺激 block 内随机呈现，block 顺序也在被试间进行平衡。(4)关于两个实验中刺激比例的问题，我们基于两个实验各自的图片数量确定了总的试次数和每张图片的重复次数。但根据 GPower 计算，两个实验的被试数量有较大的差别，实验二的被试数量较实验一少，为了能在实验二中得到偏差刺激诱发的较稳定的波形，因而，我们调整了实验二各刺激类型的比例，增加了偏差刺激的比例。我们实验中使用的两种刺激比例的分配情况在已面孔 vMMN 研究中经常用到(Kecskés-Kovács et al., 2013; Stefanics et al., 2012; Wang et al., 2022)。此外，尚未有研究说明不同刺激的比例会影响 vMMN(Stefanics et al., 2014, a review for vMMN)，因而，我们对两个实验的刺激比例进行了调整。

脑电记录与分析：P1 和 N170 的波幅采用峰基线的测量方法，指的是每个被试每个实验条件的 P1 和 N170 分别在 90-120 ms 和 130-200 ms 内波幅最高点和最低点分别到基线 X 轴的垂直距离；峰潜伏期指的是每个被试每个实验条件的 P1 和 N170 分别在 90-120 ms 和 130-200 ms 内波幅最高点和最低点出现的时间点。

意见 5：统计方法和结果可以报告得更详实完整。比如，可以把不显著的结果也呈现出来。喝多交互作用也不明确，如 fWHR × 刺激类型 × 电极 × 半球的交互作用显著， $F(2, 80) = 6.189, p = 0.005, \eta^2 p = 0.134$ ，后续分析中半球的交互作用体现在哪？同样，fWHR × 刺激类型 × 电极交互作用显著， $F(2, 80) = 7.493, p = 0.004, \eta^2 p = 0.158$ ，交互体现在哪？低 fWHR 偏差刺激的结果呢？希望作者能呈现完整的结果，包括不显著的结果。图 2 地形图中的颜色代表的值可以明确下。

回应：非常感谢审稿人的建议！

(1)实验一 vMMN 结果部分与实验目的对应的主要结果是 fWHR × 刺激类型的交互作用，这一结果可以让我们进一步分析高、低宽高比在诱发 vMMN 上的差异。实验二 vMMN 结果部分与实验目的对应的主要结果是情绪 × fWHR × 刺激类型的交互作用，这一结果可以让我们进一步分析愤怒和恐惧情绪高、低宽高比在诱发 vMMN 上的差异。针对交互作用显著后的简单效应检验分析，我们补充了上述与研究目的对应的主要结果的不显著结果表现。比如，审稿人提到的实验一 vMMN 分析结果发现 fWHR × 刺激类型 × 电极交互作用显著， $F(2, 80) = 7.493, p = 0.004, \eta_p^2 = 0.158$ ，在之前的文稿中我们只报告了显著结果，高 fWHR 偏差刺激与标准刺激诱发波幅在不同电极上的结果，对于不显著的结果没有报告；在此基础上，我们增加了重要结果低 fWHR 偏差刺激与标准刺激诱发波幅的差异($ps \geq 0.456$)，从而说明交互作用的表现。其他部分不显著结果的补充请参考修改后文稿的结果部分。

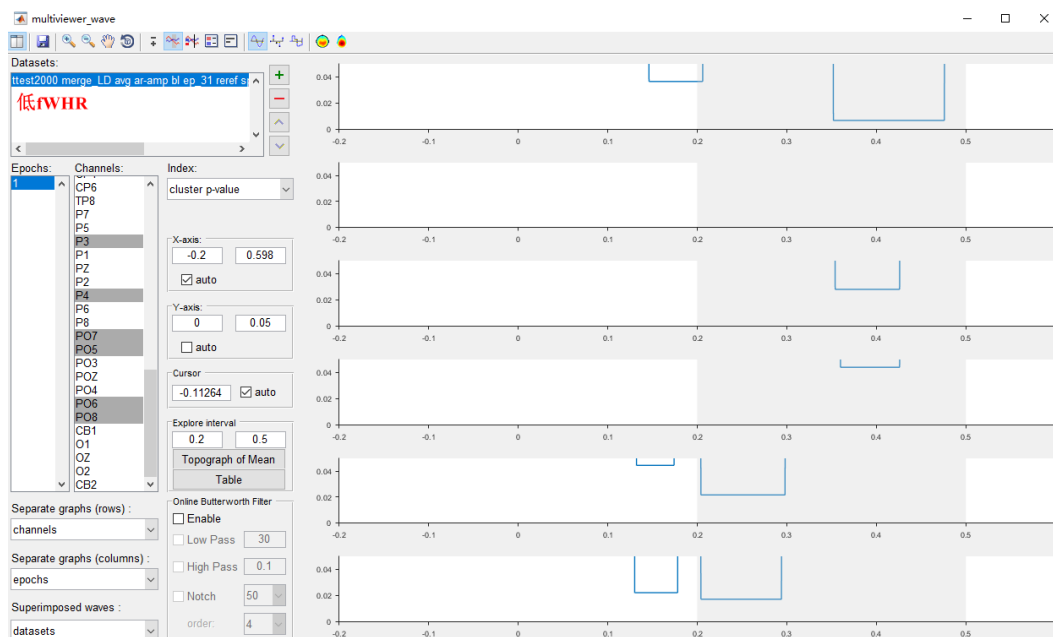
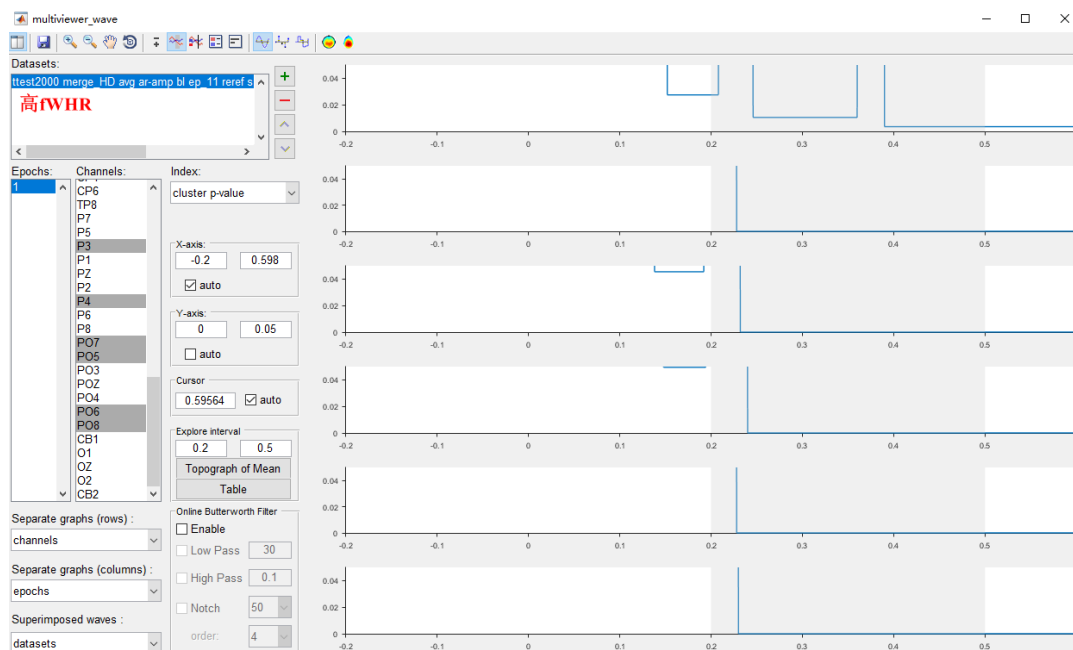
(2)图 2 地形图的颜色代表相应时间段的波幅值，我们已在颜色图例部分增加了波幅单位 μV 。非常感谢审稿人对研究细节的把握！

意见 6：揭示时序上的差异非常值得提倡，也是本研究的亮点。但作者采用的方法需要进一步澄清。比如，为什么选择 50 毫秒作为窗口，有何依据？多窗口分析有没有做多重比较校准？为了提升数据的时间精度，作者可以考虑滑动窗口的方式分析数据(比如 200-250, 201-251 等等，即数据间有重合)。但我更推荐作者针对每个数据点做统计，而不是汇总某段数据。然后使用数据驱动的方法揭示在哪个时间段会出现差异，而不是提前人为确定好时间窗。作者可以考虑以下方法：基于丛聚的置换检验(cluster-based permutation test)进行多重比较校准，使用生存分析，或是 generalized additive mixed modeling。

回应：(1)非常感谢审稿人对本研究创新点的肯定！我们在论文自检报告第一条研究亮点中

增加了此条作为本研究的贡献,同时,我们也在前言表述本研究目的时增加了该条内容,“*同时结合 ERP 技术高时间分辨率的特点以期揭示高低面孔宽高比的自动加工机制在时序上的变化。*”。

(2)在方法上,我们的做法与审稿人的意见相似。在确定 vMMN 分段时间窗口之前,我们使用了 Ietwave7 中提供的数据驱动方法对两个实验中 200-500ms 不同条件下标准刺激和偏差刺激诱发的波幅进行点对点差异检验,并用了 cluster-based permutation test 进行多重比较校正(cluster threshold 0.05; number of permutation 2000),结果发现不同条件下标准和偏差刺激诱发波幅差异显著的时间段较多且针对不同的实验条件有一定的分散性,为了能较全面体现这些时间点上的差异,因而,我们根据点对点检验的结果并结合已有研究的分段(Wang et al., 2022),基于假设的驱动对本研究 vMMN 采用 50ms 作为窗口分段分析。下面两张图是我们实验一进行数据驱动分析后显示的 P3/P4/PO7/PO5/PO4/PO8 电极上高低 fWHR 标准和偏差刺激诱发波幅差异显著的时间段,纵坐标表示的是 cluster p-value,我们用阴影标注了 200-500ms 的时间段。



意见 7: 其他意见: 样本量的确定需要更详细的细节。比如它是基于方差分析吗, 具体包含哪些自变量及水平数。效应量是基于主效应还是交互效应等。正文中需要写明伦理。实验一和实验二是不是使用不同的人脸(当然最好是同一个人表现出不同表情)? 是不是由于面孔不一样导致实验 1 和实验 2 结果不一样。希望能解释下为什么学生毕业论文数据就不能预先备案了?

回应: (1) 实验一样本量的确定是根据刺激类型主效应, 也就是偏差刺激与标准刺激诱发 ERP 波幅的差异, 效应量是基于刺激类型主效应。实验二样本量的确定是根据刺激类型与 fWHR 的交互作用, 即高低 fWHR 偏差刺激与标准刺激诱发 ERP 波幅的差异, 效应量是基于刺激类型与 fWHR 的交互作用。

(2) 本研究已获得所属单位伦理研究委员会的批准和同意, 该内容已在被试介绍部分进行补充。

(3) 实验一和二使用的是不同身份的面孔材料。审稿人可能会质疑两个实验使用不同身份的面孔在情绪强度和认同度上存在差异可能影响实验间比较的有效性, 但是, 我们对两个实验面孔情绪强度和认同度分别进行 2 (fWHR: 低、高) \times 3 (面孔情绪: 中性、愤怒、恐惧) 的统计分析并未发现任何显著差异结果($p \geq 0.232$), 这说明两个实验材料的情绪强度和认同度不存在差异。另外, 使用不同身份的面孔在一定程度上可以控制刺激材料低水平物理差异对额外神经活动的影响, 这一方法在面孔情绪研究中已经普遍被研究者们所接受, 因而我们结合图片库的实际情况也使用了不同身份面孔。但是, 审稿人所说两个实验使用同一人不同情绪在一定程度上可以控制面孔身份信息带来的可能的影响, 这将是进一步研究需要考虑的问题。

(4) 本研究作为学生毕业论文设计的部分内容, 我们在 2020 年下半年开始进行实验设计和准备实验材料, 2021 年集中实施实验, 在我们完成实验设计前本刊对实验室研究的预备案尚未做出明确要求, 因而, 我们没有提前在本刊备案。将来我们会多关注本刊的投稿要求。

第二轮

审稿人 1 意见:

作者做了令人比较满意的修改和回复。

只是对于第 8 点, 我仍然觉得摘要中“面孔宽高比(facial width-to-height ratio, fWHR)在表征或预测男性个体的攻击性行为水平上存在差异”不通顺。谁和谁有差异? 一般来说, 可以说“高低面孔宽高比在表征或预测男性个体的攻击性行为水平上存在差异”或者“面孔宽高比可以预测男性个体的攻击性行为”。供参考。

回应: 再次感谢审稿人的意见和建议! 我们已按照审稿人的建议修改了上述不通顺的语句, 修改为“高低面孔宽高比(facial width-to-height ratio, fWHR)在表征某男性个体实际攻击性或预测他人评价该个体的攻击性行为水平上存在差异”。

备注: 为了与第一轮修改内容(红色)作区分, 本论所有修改内容用蓝色字体标注。

审稿人 2 意见: 感谢作者的回复, 论文也有完善。但有些问题依然需要进一步补充说明。

意见 1: 很多时候, 作者仅仅只是回复审稿人, 并没有把相关内容补充进正文中。比如:

1) 本研究中的面孔宽高比结果与 Merlhiot et al., 2021 不一致需要在讨论中讨论。

2) 实验 1 和实验 2 使用不同刺激比例的理由或者相关讨论也需要在正文中说明。

3)波幅(峰基线值)和峰潜伏期计算方法也应该在正文中明确。

相关内容的补充不仅能让更多读者对方法进行评判以及促进结果可重复性。同理，希望作者能及时上传数据，能让其他读者进行验证。

回应：再次感谢审稿人为完善本论文提出的详细意见和建议！我们已经把对审稿人在第一轮提出相关问题的回复补充在正文中的适当位置。

1)关于“本研究中的面孔宽高比结果与 Merlhiot et al., 2021 不一致”内容的阐述和解释我们放在了“4 实验一 vs. 实验二”第一段不同情绪面孔宽高比的值的统计结果之后，放在这里的主要原因是，这个结果可以为接下来对本研究两个实验间 MMN 比较的有效性提供依据，即中性与愤怒情绪高宽高比以及中性与恐惧情绪低宽高比自动加工程度的差异较少或不受不同情绪宽高比差异的影响。

2)关于“实验 1 和实验 2 使用不同刺激比例的理由或者相关讨论”的内容我们放在了正文的讨论部分的最后一段，也就是描述本研究局限性的内容处。虽然本研究使用的这两种刺激比例的分配情况在已面孔 vMMN 研究中经常用到(Kecskés-Kovács et al., 2013; Stefanics et al., 2012; Wang et al., 2022), 并且尚未有研究说明不同刺激的比例会影响 vMMN(Stefanics et al., 2014, a review for vMMN), 但是, 这一问题在将来的研究中还是需要注意, 使用同一刺激比例可能更能增加实验间比较的说服力。

3)波幅(峰基线值)和峰潜伏期的测量方法在实验一“脑电记录与处理”的方法部分进行了补充说明。

4)原始数据和实验程序已经完成了上传，具体信息如下。

我的数据		相关数据					
标题 ▾	相关期刊 / 组或项目 ▾	更新	状态 ▾	类型 ▾	分享状态 ▾	查看	操作
面孔宽高比自动加工脑电数据	心理学报	2023-03-29 16:38:06 GMT+8	审核通过	数据集/文件集	限制性获取	预览 私有访问链接 查看更多	

意见 2：在时序分析上，多窗口分析有没有做多重比较校准依然没有回复。作者提到基于假设驱动使用 50ms 窗口，到底是何假设？很多研究者会选择使用多种窗口对数据分析，验证结果的稳定性。考虑到本研究主要关注 fVHR、刺激类型、和面孔情绪之间的交互，为什么不使用 cluster-based permutation test 做方差分析，而仅仅只是对比标准刺激和偏差刺激的差异？

回应：非常感谢审稿人详细的建议！

(1)我们在所有数据进行分析时，当适用非球形性时我们报告的是 Greenhouse-Geisser 校正 p 值，利用 Bonferroni 校正来解释多重比较和事后分析。在两个实验数据的分析部分我们做了说明。

(2)抱歉我们上一轮的回复没有说清楚 50ms 窗口设定的标准，这个窗口的设定不是假设驱动，是我们根据本研究的结果图，同时参考以往 vMMN 研究设定的窗口。已有研究在考察某些刺激诱发 vMMN 时序上的变化时，多是采用类似的方法对 vMMN 进行分时间窗分析，因而，本研究也采用相似的方法设定分段时间窗。这部分内容我们在“2.2 脑电记录与处理”进行了补充说明。

(3)参考已有采用从聚的置换检验的研究，常把这种分析作为 ERP 传统分析的补充或探索性分析，因而，我们也按照已有研究的方法补充了从聚的置换检验的分析。在“2.2 脑电记录与处理”最后一段我们补充了使用置换检验分析的原因和具体方法。虽然本研究两实验是多因素实验设计，但是我们还是对感兴趣的两个实验条件进行了配对样本 t 检验。原因有二，

一是研究者认为基于从聚的置换检验不能保证方差分析的有效性(Groppe et al., 2011), 二是已有采用基于从聚的置换检验的研究也是进行的配对样本 t 检验分析(Vormbrock et al., 2023; Zochowska et al., 2021), 这些研究也都是多因素实验设计。在实验一和二的结果部分补充了置换检验的具体结果, 请参考“2.3.3 基于从聚的置换检验”和“3.3.3 基于从聚的置换检验”的内容。在讨论部分我们增加了一段内容讨论该检验与 ERP 传统分析结果的异同及原因, 请参考讨论部分倒数第二段内容。

意见 3: 样本量的确定是不是存在问题? 比如作者提到实验一样本量的确定是根据刺激类型主效应, 不应该是刺激类型与 fWHR 的交互作用吗?

回应: 非常感谢审稿人的意见! 我们重新检查了实验数据采集前样本容量的确定方法, 实验一是根据 2 (fWHR: 高 vs. 低) \times 2 (刺激类型: 标准 vs. 偏差) 重复测量方差分析的交互作用确定的被试量, 具体参数为 $\eta_p^2 = 0.06$, 检验力: 80%, $\alpha = 0.01$, 通过计算需要至少 33 名被试。实验二采用同样的参数计算至少需要 22 名被试才能得到 2 (情绪: 愤怒 vs. 恐惧) \times 2 (fWHR: 高 vs. 低) \times 2 (刺激类型: 标准 vs. 偏差) 重复测量方差分析的交互作用。针对上述问题我们已在文中做了修改, 最终两个实验招募的被试量都符合计算出的被试数要求。

意见 4: 正文中还存在一些实际行为和被他人感知行为之间的混淆, 作者可以进一步明确。比如, “大量研究表明, 男性面孔宽高比是预测个体不良社会行为水平的可靠指标”, 个体不良社会行为水平应该是个体实际的行为, 但这句话之后的描述都是基于他人的评价或者感知。别人知觉到某个个体不可信不代表那个个体实际上真的不可信。

回应: 非常感谢审稿人的意见! 我们已对上述问题进行了全文调整和修改, 把“预测”改成了“预测他人评价该个体”, 具体内容请参考正文部分蓝色字体区域。

意见 5: 其他意见:

(1) 作者为什么会预期“与低 fWHR 诱发的 vMMN 相比, 高 fWHR 诱发的 vMMN 波幅较大, 时程较长”? 是基于已有威胁性刺激的脑电结果做出的推论吗? 相关假设的推理可以在前言中说明

回应: 非常感谢审稿人的建议! 我们在前言最后一段补充了实验一结果预期的依据, 包括两个方面, 一是根据已有研究, 相比于低 fWHR, 个体对高 fWHR 的攻击性感知较高(Carré & McCormick, 2009); 二是结合研究发现的负面情绪高低宽高比面孔在预测他人评价该个体攻击上的差异可能源于其识别为愤怒和恐惧情绪(Deska et al., 2018), 以及愤怒较恐惧情绪更可能表达直接威胁性信息(Adams et al., 2003)。我们进一步预期, 与低 fWHR 诱发的 vMMN 相比, 具有更高威胁性信息的高 fWHR 诱发的 vMMN 波幅较大, 时程较长。同时, 我们也对实验二的预期结果进行了补充和完善。

(2) 行为结果(目标刺激的击中率)需要报告下高低 fWHR 的差异以保证脑电差异不会因为行为结果不同引起

回应: 非常感谢审稿人的意见! 我们对实验一和二的行为结果分别进行了配对样本 t 检验和 2 (fWHR: 低、高) \times 2 (面孔情绪: 愤怒、恐惧) 重复测量方差分析, 对实验一和实验二行为结果的对比进行了 2 (fWHR: 低、高) \times 3 (面孔情绪: 中性、愤怒、恐惧) 重复测量方差分析。在正确率上未发现任何显著结果, 在反应时上有个别显著结果, 具体内容请参考各个结果部分, 同时, 我们也在讨论部分对此反应时显著结果进行了解释。

(3) 作者在讨论中提到高低 fWHR 分别表征的愤怒和恐惧情绪可能导致了实验一的结果, 是

指被试会从高 fwhr 面孔解读出愤怒情绪吗？但实验 2 又发现愤怒情绪较中性情绪减弱了高 fWHR 的自动加工程度。如何解释这一矛盾。

回应：非常感谢审稿人的意见！这一矛盾的结果我们也注意到了，因而在讨论部分做了解释。之前的稿件中我们主要从以下方面进行了解释，根据已有情绪自动加工的研究发现，具体情绪面孔较中性情绪面孔更容易进行自动加工(曾宪卿 等, 2021)，这在一定程度上可能减弱了作为知觉信息面孔宽高比的自动加工程度。在这轮的修改中我们增加了另外一种可能的解释，即本研究对不同情绪面孔宽高比的控制。已有研究发现，与中性情绪面孔相比，个体表现出的愤怒情绪会增加其面孔宽高比(Merlhiot et al., 2021)，进而可能会提高其被感知的攻击性水平。但是，本研究通过对比实验一二面孔宽高比的值发现，中性与愤怒情绪高宽高比的值不存在显著差异，这在一定程度上可能减弱了个体对愤怒情绪面孔攻击性的感知，进而也可能减弱其自动加工程度。具体内容请参考讨论部分第 5 段内容。

(4)“众所周知，我们的大脑可以快速自动地完成对外界信息的加工，尤其是对个体生存具有重要意义

的社会信息”，需要参考文献
回应：非常感谢审稿人的建议！我们已在该内容后增加对应参考文献“(Haselton & Funder, 2006; Kovarski et al., 2017)”。

(5)实验 2 中，作者对 200-500ms 整体进行分析时，并没有发现 fWHR，刺激类型和面孔情绪之间的交互，这个结果与实验 1 不同，如何解释

回应：非常感谢审稿人的意见！正如审稿人所说能与实验二研究目的对应的结果确实是情绪 \times fWHR \times 刺激类型出现显著的交互作用，我们在实验二分时间段进行结果分析前也说明了这点。遗憾的是，我们在对整个时间窗口分析时我们并未发现显著的交互作用，而仅在分时间窗口分析时发现了与上述交互作用有关的结果。但是，我们在 200-500ms 分析时发现了刺激类型的主效应，进一步分析发现此时间段确实存在 vMMN 成分。这说明，非中性情绪面孔宽高比信息的自动加工可能与时间段有关。其原因可能与情绪的加工有关，情绪自动加工的元分析发现，非中性情绪面孔在 200-400ms 能诱发 vMMN，进行自动加工，但是，不同情绪自动加工的时间窗又存在差别(曾宪卿 等, 2021)，这可能在一定程度上影响了作为知觉信息的 fWHR 的自动加工，因而，我们在实验二的整体分析时未得到上述交互作用。对于这一结果的解释，我们已放在了实验二结果的最后。

第三轮

审稿人 2 意见：

作者对论文进一步的完善值得肯定。我还有一个问题：作者提到通过观察结果图确定分析时间窗会导致循环分析，具体参见 Ten common statistical mistakes to watch out for when writing or reviewing a manuscript。以我个人经验，cluster-based permutation test 是解决这一问题的最好方法。作者不一定需要对全脑的通道进行分析，可以限制分析在所关心的通道上，主要用数据驱动的方法去发现到底哪些时间段显著。

回应：再次感谢审稿人的意见和建议！根据审稿人的建议，把 cluster-based permutation test 作为一种数据驱动的方法用来找到高低宽高比在哪些时间段诱发 vMMN，我们调整了上一轮修改中补充的“基于从聚的置换检验”结果在两个实验中的位置。上一轮修改我们把这部分单独列为结果的一部分，在这轮修改中我们把这部分结果放到 vMMN 结果标题下，同时调整了结果图的顺序。结合本研究置换检验的结果和已有文献(200-440ms, Wang et al., 2022)，我们确定 vMMN 分析时间窗为 200-500ms。此外，大部分研究在进行基于从聚的置

换检验时进行的是全脑通道检验(Zochowska et al., 2023), 本研究中所关注的电极点位置分散, 如果只对关注区域进行检验则结果可能不具备参考价值。因此, 我们采用全脑通道检验, 同时, 我们增加了所关注的电极点在置换检验中显著的具体时间段, 以此来进一步推断解释 vMMN 分时间窗口分析的依据。本研究在最初定位 vMMN 时间窗口时使用了数据驱动方法但并未于正文中呈现细节, 所以, 虽然基于从聚的置换检验在正文中作用及位置有所调整, 但是 ERP 重要结果以及推论并未发生改变。

备注: 本论所有修改内容用绿色字体标注。

第四轮

审稿人 2 意见:

建议发表

编委意见:

Okay. Please accept this paper for publication.

主编意见:

同意外审和编委意见, 建议录用。