

社交焦虑对跨模态情绪信息加工的影响

王亚丽^{1*}, 石茹²

¹浙江财经大学马克思主义学院, 浙江 杭州

²包头医学院党委学生工作部(处), 内蒙古 包头

收稿日期: 2026年2月1日; 录用日期: 2026年2月27日; 发布日期: 2026年3月12日

摘要

跨模态情绪信息的准确识别是社会交往的重要基础。既有研究表明, 社交焦虑个体在情绪识别与判断中可能存在偏差, 但其跨模态情绪加工特征仍有待进一步探讨。本研究筛选高、低社交焦虑大学生, 考察其在视觉、听觉及视听情绪信息加工中的表现。结果显示, 无论在正确率还是反应时指标上, 视听一致条件均显著优于单视觉、单听觉及视听不一致条件, 呈现典型的双通道加工优势; 视听不一致条件正确率最低、反应时最长, 表现出显著的跨模态冲突效应。进一步分析发现, 在愤怒情绪条件下, 高、低社交焦虑组均表现出双通道优势, 其中高社交焦虑组的整合效应更为突出; 在中性情绪条件下, 两组均出现跨模态不一致效应, 但低社交焦虑组的不一致干扰更为明显。研究结果揭示了社交焦虑水平对跨模态情绪加工的调节作用, 为理解社交焦虑相关的情绪感知偏差提供了行为学证据。

关键词

社交焦虑, 跨通道情绪加工, 整合优势, 冲突效应

The Impact of Social Anxiety on Cross-Modal Emotional Processing

Yali Wang^{1*}, Ru Shi²

¹School of Marxism Studies, Zhejiang University of Finance and Economics, Hangzhou Zhejiang

²Student Affairs Department (Office) of Party Committee, Baotou Medical College, Baotou Inner Mongolia

Received: February 1, 2026; accepted: February 27, 2026; published: March 12, 2026

Abstract

Accurate recognition of cross-modal emotional information is a fundamental prerequisite for successful social interaction. Previous studies have suggested that individuals with social anxiety may exhibit

*通讯作者。

文章引用: 王亚丽, 石茹(2026). 社交焦虑对跨模态情绪信息加工的影响. *心理学进展*, 16(3), 338-350.

DOI: 10.12677/ap.2026.163148

biases in emotion recognition and judgment; however, their characteristics of cross-modal emotional processing remain insufficiently understood. The present study recruited university students with high and low levels of social anxiety to examine their performance in processing visual, auditory, and audiovisual emotional information. The results showed that, in terms of both accuracy and reaction time in emotion recognition, the audiovisual congruent condition significantly outperformed the unimodal visual, unimodal auditory, and audiovisual incongruent conditions, demonstrating a typical dual-channel processing advantage. In contrast, the audiovisual incongruent condition yielded the lowest accuracy and the longest reaction times, indicating a robust cross-modal conflict effect. Further analyses revealed that under angry emotion conditions, both high and low social anxiety groups exhibited a dual-channel advantage, with a more pronounced integration effect observed in the high social anxiety group. Under neutral emotion conditions, both groups showed significant cross-modal incongruency effects; however, the interference caused by incongruent information was more pronounced in the low social anxiety group. These findings indicate that social anxiety level modulates cross-modal emotional processing and provide behavioral evidence for understanding emotion perception biases associated with social anxiety.

Keywords

Social Anxiety, Cross-Modal Emotional Processing, Integration Advantage, Conflict Effect

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

对情绪性信息的准确识别和判断是成功社会交往的重要能力。在真实的社交情境中,个体往往需要在较短时间内同时接收来自面部表情、语音语调及姿态动作等多种情绪线索,并对其进行跨模态的整合与加工(Klasen et al., 2012; Robins et al., 2009)。因此,跨模态情绪整合被认为是实现有效社会沟通和准确情绪理解的关键认知过程(Filippi, 2016; Schirmer & Adolphs, 2017)。然而,这一跨模态情绪整合过程并非在所有个体中都同样高效,尤其是在涉及社会评价或潜在威胁的情境下,部分个体可能表现出情绪加工偏向或整合困难(Gan & Li, 2023; 宋素涛等, 2023)。

在典型的双通道情绪加工情境中,当来自不同感觉通道的情绪线索保持一致时,个体通常能够有效整合多通道信息,从而表现出显著的跨模态整合优势。具体而言,行为研究发现,在双通道情绪信息加工中,当面部表情与情绪性声音保持一致时,个体通常表现出显著的整合优势,即反应时更短、识别准确率更高(Collignon et al., 2008; Gan & Li, 2023; Klasen et al., 2011; Li et al., 2024; Peschard & Philippot, 2017; Schelenz et al., 2013)。来自神经影像与脑电研究的证据进一步表明,不同感觉通道的情绪信息并非独立加工,而是在感知、认知及反应选择等多个加工层次发生相互作用并实现整合(Proverbio & De Benedetto, 2018; Yeh et al., 2016)。这种整合过程不仅有助于提高情绪识别效率,还可降低判断的不确定性,使个体更依赖整合后的情绪线索进行决策。

然而,当来自不同通道的情绪线索不一致时,跨模态情绪整合往往受到干扰,从而产生典型的跨模态冲突效应(cross-modal conflict) (Watson et al., 2013)。已有研究发现,当面部表情与声音情绪不匹配时,个体在情绪判断任务中通常表现为反应时延长、正确率下降(Dolan et al., 2001; Müller et al., 2011),该行为层面的损耗被认为源于更高水平的冲突检测(conflict monitoring)的需求(Proverbio et al., 2020)。与之相

对应的神经影像研究显示, 这类冲突通常伴随着前额-顶叶网络及双侧尾状核等脑区激活的增强, 提示在工作记忆维持与情绪编码阶段, 情绪信息的处理需要额外的认知资源(Klasen et al., 2011)。

跨模态情绪信息的整合与冲突加工并非在所有个体中都以相同方式展开, 而会受个体差异因素的影响(Liu et al., 2021), 其中焦虑水平被认为是重要的调节变量之一(Heffer et al., 2022; Koizumi et al., 2011)。相较于低焦虑个体, 高焦虑者在信息加工过程中更倾向于优先注意并放大威胁相关线索, 这种偏向不仅体现在单一感觉通道的情绪加工中, 也进一步影响跨模态情绪整合与冲突加工(Heffer et al., 2022; Koizumi et al., 2011)。

与特质焦虑类似, 社交焦虑个体在社会情境中同样表现出对负性或威胁性社会线索的注意与解释偏向(Schofield et al., 2012; 宋素涛等, 2023)。然而, 现有关于社交焦虑的研究多数仍集中于单一感觉通道的情绪加工, 例如对情绪面孔或情绪性声音的识别(Peschard & Philippot, 2017), 而对多通道情绪线索同时呈现下的整合与冲突加工, 系统性研究仍不足。

近年来, 少量研究开始关注社交焦虑与跨模态情绪加工的关系, 但相关证据仍存在不一致。一方面, 有研究发现, 在一致的威胁性视听情绪条件下, 社交焦虑个体可能表现出更强的整合优势, 如反应时缩短或正确率提高(Gan & Li, 2023; Yuan et al., 2026), 并伴随与情绪整合相关的脑电成分(如 P3、LPP)波幅的增强(Gan & Li, 2023)以及双侧上颞沟等脑区激活的增强(Kreifelts et al., 2020)。另一方面, 也有研究未发现社交焦虑对跨模态情绪整合效应的显著调节作用(Peschard & Philippot, 2017)。这些不一致结果提示, 仅考察整合优势的存在可能不足以揭示社交焦虑对跨模态情绪加工的真实影响, 需进一步系统探讨社交焦虑在一致与不一致情绪条件下的加工特点。

实证研究提示, 社交焦虑对跨模态情绪加工的影响可能并非简单表现为整体整合能力的增强或削弱, 而更可能反映在个体对不同情绪线索组合的加工偏差上。已有研究表明, 在涉及社会威胁的负性情绪条件下, 社交焦虑个体更易对威胁性信息保持高度敏感(Yang et al., 2025), 这可能进一步影响他们在情绪线索一致与不一致情境中的加工表现。然而, 目前尚缺乏直接而系统的实证研究, 探讨社交焦虑个体在不同情绪效价条件下如何调节跨模态情绪整合优势与不一致干扰效应, 尤其是行为表现的差异。为此, 本研究采用视听跨模态情绪判断任务, 系统操纵情绪一致性(一致 vs. 不一致)和情绪效价(中性 vs. 生气), 比较高、低社交焦虑个体在跨模态情绪整合与冲突加工过程中的行为差异, 以期揭示社交焦虑影响多通道情绪信息整合与冲突加工的认知机制。

2. 方法

2.1. 被试

采用整群抽样的方式在内蒙古某高校招募大学生被试 270 名。所选被试均身体健康, 视力或矫正视力正常, 无精神疾病史或药物依赖史。根据社交焦虑量表得分, 选取前、后各 15% 的被试作为高、低社交焦虑候选组, 每组各 40 名。随后进行二次筛选, 为提高分组稳定性, 仅保留连续两次测量中社交焦虑得分均低于 38 分的被试作为低社交焦虑组, 得分均高于 52 分的被试作为高社交焦虑组。同时剔除有效实验试次比例低于 70% 的无效被试, 最终纳入被试共 70 名, 高、低社交焦虑组各 35 名(女性 22 名, 平均年龄为 22.65 ± 1.83 岁; 男性 13 名, 平均年龄为 22.63 ± 1.83 岁)。本研究经所在单位伦理委员会审查批准, 且被试在实验前均签署知情同意书。

2.2. 量表

社交焦虑水平采用 Leary (1983) 编制、彭纯子等(2004)修订的《交往焦虑量表》进行测量。该量表共包含 15 个条目, 采用 1 (完全不符合)~5 (非常符合) 5 点计分方式, 其中第 3、6、10、15 题为反向计分, 得分越高代表个体的社交焦虑水平越高。本研究中, 该量表的 Cronbach's α 系数为 0.80, 表明量表具有

良好的内部一致性。

2.3. 实验设计和材料

实验采用 2(情绪效价: 中性、生气) × 4(通道性质: 视觉、听觉、视听一致、视听不一致) × 2(分组: 高社交焦虑组、低社交焦虑组) 的混合实验设计, 其中情绪效价与通道性质为组内变量, 分组为组间变量。

根据通道构成, 实验刺激可分为三类: 视觉刺激、听觉刺激和视听刺激, 其中视听刺激进一步区分为一致与不一致条件。视觉刺激选自 NimStim 的面孔数据库(Tottenham et al., 2009), 共选取 4 名个体(2 男、2 女)的中性和生气面孔。所有面孔图片均通过图像处理软件(Adobe Photoshop 2019, Adobe Inc., San Jose, CA, USA)进行预处理, 去除非面部信息(如头发、服饰等), 并对亮度、对比度和大小进行标准化处理。听觉刺激选自蒙特利尔情绪声音刺激库(Montreal Affective Voices; Belin et al., 2008)。同样选取 4 名个体(2 男、2 女)的中性和生气的情绪声音。所有声音刺激均通过 Adobe Audition CC 软件进行处理, 时长统一为 1000 ms, 音量控制在约 65 dB。视听刺激由上述视觉和听觉材料组合而成。其中, 视听一致条件同时呈现情绪一致的面孔和声音(中性 - 中性或生气 - 生气), 视听不一致条件则呈现情绪不一致的面孔和声音(中性面孔 - 生气声音或生气面孔 - 中性声音)。

2.4. 实验设计和材料

实验程序采用 E-Prime 3.0 (Psychology Software Tools Inc., Sharpsburg, PA, USA)进行编制, 所有的视觉刺激均呈现在一台 14 英寸的笔记本电脑上, 屏幕分辨率为 1920 × 1080。实验过程中, 被试与屏幕之间的距离约为 60 cm, 对应的刺激视角为 5.8° (水平视角) × 7.2° (垂直视角)。听觉刺激通过头戴式耳麦以双声道方式呈现。实验包含 8 种组内实验条件, 每种条件设置 48 个试次, 共 384 个试次。所有试次被分为 4 个 block 依次呈现, 每个 block 结束后安排短暂休息, 以减轻被试疲劳。单个试次的流程如下: 首先呈现 300 ms 的注视点, 随后为 700 ms 的实验刺激, 刺激结束后呈现 1000 ms 的空屏。刺激呈现后, 被试需判断刺激所表达的情绪类型(中性或生气), 并尽快做出按键反应。其中, 一半被试按“F”键表示中性、按“J”键表示生气, 另一半被试的按键则相反, 以平衡按键偏好带来的影响。对于视听双通道刺激, 被试被明确要求以面孔表情为判断依据。

鉴于面孔情绪相较于声音情绪更易被识别(Joassin et al., 2004), 并参考既有研究的实验设置(Peschard & Philippot, 2017), 在视听双通道条件下, 声音刺激先行呈现 100 ms, 随后面孔刺激与声音刺激同时呈现。正式实验开始前, 被试需完成 30 个练习试次, 以熟悉实验流程和任务要求。

2.5. 数据分析

数据分析在 SPSS 19.0 (IBM, Somers, USA)中进行。对正确率和反应时分别进行 2(情绪效价: 中性、生气) × 4 通道性质(视觉、听觉、视听一致、视听不一致) × 2(分组: 高社交焦虑组、低社交焦虑组)的重复测量方差分析, 显著性水平设为 0.05。对于不满足球形假设的数据, 采用 Greenhouse-Geisser 校正, 多重比较采用 Bonferroni 校正法。

3. 结果

3.1. 正确率

重复测量方差分析结果显示, 通道性质的主效应显著, $F(3, 204) = 37.15, p < 0.001, \eta_p^2 = 0.35$ 。进一步的事后比较表明, 视听一致条件下($M \pm SD: 0.84 \pm 0.11$)的正确率显著高于单视觉通道($0.81 \pm 0.10, t = 4.25, p < 0.001, \text{Cohen's } d = 0.33$)和单听觉通道($0.75 \pm 0.12, t = 9.89, p < 0.001, \text{Cohen's } d = 0.80$), 表现出明显的双通道加工优势。相反, 视听不一致条件下(0.69 ± 0.15)的正确率显著低于视听一致($0.84 \pm 0.11, t =$

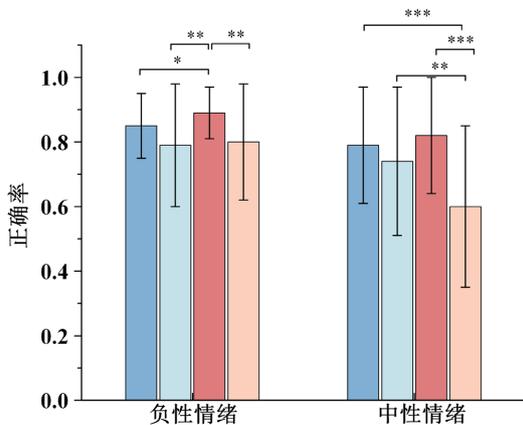
-8.05, $p < 0.001$, Cohen's $d = -1.20$)、单视觉通道(0.81 ± 0.10 , $t = -7.00$, $p < 0.001$, Cohen's $d = -0.97$)以及单听觉通道(0.75 ± 0.12 , $t = -2.78$, $p = 0.043$, Cohen's $d = -0.49$), 表明视听信息不一致会显著降低情绪识别的准确性。此外, 情绪效价的主效应显著, $F(1, 68) = 6.41$, $p = 0.014$, $\eta_p^2 = 0.09$, 生气情绪条件下(0.81 ± 0.12)的正确率显著高于中性情绪条件(0.74 ± 0.17 , $t = 2.52$, $p = 0.014$, Cohen's $d = 0.48$)。

通道性质与情绪效价的交互作用显著, $F(3, 204) = 17.98$, $p < 0.001$, $\eta_p^2 = 0.21$ 。简单效应分析显示, 在生气情绪条件下, 视听一致条件(0.88 ± 0.12)的正确率显著高于单视觉通道(0.83 ± 0.12 , $t = 4.40$, $p < 0.001$, Cohen's $d = 0.37$)、单听觉通道(0.74 ± 0.21 , $t = 6.67$, $p < 0.001$, Cohen's $d = 0.48$)以及视听不一致条件(0.78 ± 0.16 , $t = 5.44$, $p < 0.001$, Cohen's $d = 0.83$), 表现出显著的双通道优势效应; 而在中性情绪条件下, 视听不一致条件(0.60 ± 0.24)的正确率显著低于单视觉通道(0.78 ± 0.18 , $t = -7.08$, $p < 0.001$, Cohen's $d = -0.89$)、单听觉通道(0.77 ± 0.21 , $t = -5.63$, $p < 0.001$, Cohen's $d = -0.77$)以及视听一致条件(0.81 ± 0.18 , $t = -7.74$, $p < 0.001$, Cohen's $d = -1.00$), 呈现出显著的跨模态不一致干扰效应。

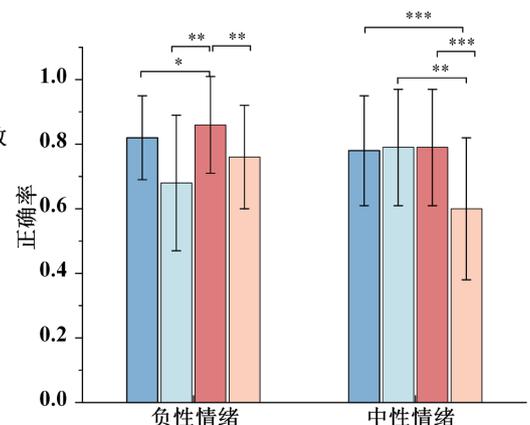
此外, 通道性质、情绪效价和分组的三重交互作用也显著(见图 1), $F(3, 204) = 3.14$, $p = .048$, $\eta_p^2 = 0.04$ 。低社交焦虑组中, 在生气情绪条件下, 通道效应显著: 视听一致条件(0.86 ± 0.15)的正确率显著高于单视觉通道(0.82 ± 0.13 , $t = 2.67$, $p = 0.044$, Cohen's $d = 0.29$)、单听觉通道(0.68 ± 0.21 , $t = 6.10$, $p < 0.001$, Cohen's $d = 0.41$)以及视听不一致条件(0.76 ± 0.16 , $t = 3.96$, $p = 0.001$, Cohen's $d = 0.64$), 在中性情绪条件下, 则表现出显著的不一致干扰效应, 视听不一致条件(0.60 ± 0.22)的正确率显著低于其他三种通道条件(单视觉通道: 0.78 ± 0.17 , $t = -4.86$, $p < 0.001$, Cohen's $d = -0.91$; 单听觉通道: 0.79 ± 0.18 , $t = -4.71$, $p < 0.001$, Cohen's $d = -0.98$; 视听一致: 0.79 ± 0.18 , $t = -5.21$, $p < 0.001$, Cohen's $d = -0.99$)。

高社交焦虑中, 在生气情绪条件下同样观察到显著的通道效应, 视听一致(0.89 ± 0.08)的正确率显著高于单视觉通道(0.85 ± 0.10 , $t = 3.13$, $p = 0.012$, Cohen's $d = 0.55$)、单听觉通道(0.79 ± 0.19 , $t = 3.52$, $p = 0.005$, Cohen's $d = 0.72$)以及视听不一致条件(0.80 ± 0.17 , $t = 3.92$, $p = 0.002$, Cohen's $d = 0.76$)。在中性情绪条件下, 高社交焦虑组同样表现出显著的跨模态不一致效应, 即视听不一致(0.60 ± 0.25)的正确率显著低于单视觉通道(0.79 ± 0.18 , $t = -5.11$, $p < 0.001$, Cohen's $d = -0.87$)、单听觉通道(0.74 ± 0.23 , $t = -3.36$, $p = 0.007$, Cohen's $d = -0.60$)以及视听一致条件(0.82 ± 0.18 , $t = -5.76$, $p < 0.001$, Cohen's $d = -1.01$)。正确率的描述性统计结果见表 1。

A. 高社交焦虑组



B. 低社交焦虑组



注: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$ 。

Figure 1. Accuracy in high and low social anxiety groups
图 1. 高低社交焦虑组的正确率

Table 1. Descriptive statistical table for accuracy**表 1.** 正确率的描述性统计表

实验条件	高社交焦虑组		低社交焦虑组	
	负性情绪($M \pm SD$)	中性情绪($M \pm SD$)	负性情绪($M \pm SD$)	中性情绪($M \pm SD$)
视觉条件	0.85 (0.10)	0.79 (0.18)	0.82 (0.13)	0.78 (0.17)
听觉条件	0.79 (0.19)	0.74 (0.23)	0.68 (0.21)	0.79 (0.18)
视听一致	0.89 (0.08)	0.82 (0.18)	0.86 (0.15)	0.79 (0.18)
视听不一致	0.80 (0.18)	0.60 (0.25)	0.76 (0.16)	0.60 (0.22)

3.2. 反应时

3.2.1. 正确反应时

在分析正确反应时前,剔除了错误反应、无反应以及平均值 ± 3 个标准差的异常反应。重复测量方差分析结果显示,通道性质的主效应显著, $F(3, 204) = 65.37, p < 0.001, \eta_p^2 = 0.49$ 。事后比较表明,视听一致条件(581.19 ± 93.87 ms)的反应时显著短于单听觉通道($667.18 \pm 129.39, t = -10.56, p < 0.001, \text{Cohen's } d = -0.77$)及视听不一致条件($604.15 \pm 107.85, t = -4.46, p < 0.001, \text{Cohen's } d = -0.18$),呈现出显著的双通道优势效应。此外,视听不一致条件(604.15 ± 107.85)的反应时又显著长于视听一致($581.19 \pm 93.87, t = 4.46, p < 0.001, \text{Cohen's } d = 0.18$)和单视觉通道条件($577.42 \pm 83.42, t = 4.46, p < 0.001, \text{Cohen's } d = 0.22$)。情绪效价的主效应亦显著, $F(1, 68) = 61.50, p < 0.001, \eta_p^2 = 0.48$,中性情绪条件下的(635.42 ± 105.76)反应时显著长于生气情绪条件($579.54 \pm 99.54, t = 7.84, p < 0.001, \text{Cohen's } d = 0.55$)。

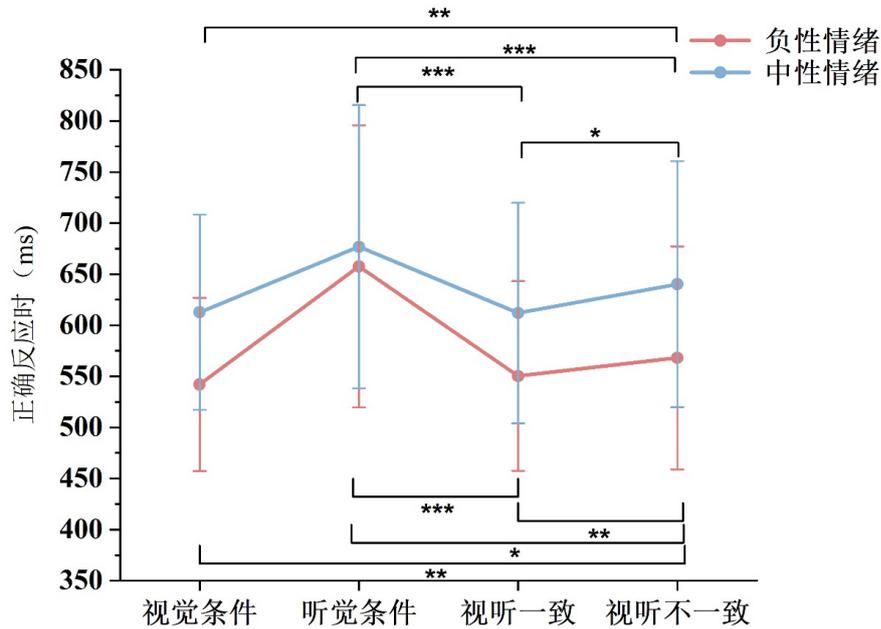
通道性质和情绪效价的交互作用显著(见图 2), $F(3, 204) = 11.16, p < 0.001, \eta_p^2 = 0.14$ 。简单效应分析显示,无论是生气情绪还是中性情绪条件下,视听一致条件(生气: 550.40 ± 92.40 ; 中性: 611.98 ± 108.05)的反应时均短于单听觉通道(生气: $657.58 \pm 138.02, t = -10.96, p < 0.001, \text{Cohen's } d = -0.89$; 中性: $676.77 \pm 138.64, t = -6.80, p < 0.001, \text{Cohen's } d = -0.53$)以及视听不一致(生气: $568.10 \pm 109.04, t = -2.81, p = 0.039, \text{Cohen's } d = -0.14$; 中性: $640.19 \pm 120.34, t = -3.84, p = 0.002, \text{Cohen's } d = -0.25$)条件下的反应时。同时,视听不一致条件下(生气: 568.10 ± 109.04 ; 中性: 640.19 ± 120.34)的反应时均长于单视觉通道(生气: $542.07 \pm 84.80, t = 4.11, p = 0.001, \text{Cohen's } d = 0.27$; 中性: $612.76 \pm 95.43, t = 3.48, p = 0.005, \text{Cohen's } d = 0.25$)以及视听一致(生气: $550.40 \pm 92.40, t = 2.81, p = 0.039, \text{Cohen's } d = 0.14$; 中性: $611.98 \pm 108.05, t = 3.84, p = 0.002, \text{Cohen's } d = -0.25$)条件下的反应时。这进一步验证了双通道整合的加工优势以及跨模态冲突带来的加工迟滞。正确反应时的描述性统计结果见表 2。

3.2.2. 错误反应时

错误反应时分析结果显示,通道性质的主效应显著, $F(3, 204) = 7.58, p < 0.001, \eta_p^2 = 0.10$ 。视听一致条件下(520.90 ± 157.45 ms)的错误反应时显著短于单听觉通道($595.82 \pm 188.22, t = -3.53, p = 0.004, \text{Cohen's } d = -0.43$)及视听不一致条件($606.04 \pm 169.76, t = -4.18, p = 0.001, \text{Cohen's } d = -0.52$)。

Table 2. Descriptive statistical table for correct reaction time**表 2.** 正确反应时的描述性统计表

实验条件	负性情绪($M \pm SD$)	中性情绪($M \pm SD$)
视觉条件	542.07 (84.80)	612.76 (95.43)
听觉条件	657.58 (138.02)	676.77 (138.64)
视听一致	550.40 (92.40)	611.98 (108.05)
视听不一致	568.10 (109.04)	640.19 (120.34)



注: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$ 。

Figure 2. Correct reaction times across different experimental conditions
图 2. 不同实验条件下的正确反应时

通道性质和情绪效价的交互作用显著, $F(3, 204) = 2.91, p = 0.043, \eta_p^2 = 0.04$ 。简单效应分析发现, 在生气情绪条件下, 视听一致条件的错误反应时(503.91 ± 239.69)显著短于单听觉通道($632.58 \pm 254.67, t = -4.02, p = 0.001, \text{Cohen's } d = -0.52$), 而在中性情绪条件下, 视听不一致条件(619.64 ± 177.44)的错误反应时显著长于单视觉通道($612.76 \pm 95.43, t = 2.91, p = 0.029, \text{Cohen's } d = 0.42$)和视听一致($549.68 \pm 155.06, t = 3.45, p = 0.006, \text{Cohen's } d = 0.45$)条件。

此外, 三重交互作用(通道性质、情绪效价和分组)呈边缘显著(见图 3), $F(3, 204) = 2.46, p = 0.073, \eta_p^2 = 0.04$ 。进一步分析显示, 低社交焦虑组在生气情绪条件下表现出显著的通道差异, 即视听一致条件的错误反应时(475.13 ± 230.54)显著短于单听觉通道($660.46 \pm 227.20, t = -4.09, p = 0.001, \text{Cohen's } d = -0.82$), 而中性情绪下未发现显著差异($ps > 0.45$)。高社交焦虑组则在中性情绪条件下出现显著的跨模态不一致效应, 即视听不一致条件的错误反应时(647.62 ± 183.00)显著长于单视觉通道($526.41 \pm 173.63, t = 3.57, p = 0.004, \text{Cohen's } d = 0.69$)及视听一致条件($540.40 \pm 219.50, t = 3.200, p = 0.013, \text{Cohen's } d = 0.54$), 而负性情绪下未发现显著的通道差异($ps > 0.69$)。错误反应时的描述性统计结果见表 3。

Table 3. Descriptive statistical table for incorrect reaction times
表 3. 错误反应时的描述性统计表

实验条件	高社交焦虑组		低社交焦虑组	
	负性情绪($M \pm SD$)	中性情绪($M \pm SD$)	负性情绪($M \pm SD$)	中性情绪($M \pm SD$)
视觉条件	558.96 (180.94)	526.41 (173.63)	539.81 (195.09)	572.96 (132.42)
听觉条件	604.70 (280.00)	596.74 (207.18)	660.46 (227.20)	521.36 (255.65)
视听一致	532.70 (248.48)	540.40 (219.50)	475.13 (230.54)	535.39 (163.32)
视听不一致	608.07 (249.89)	647.62 (183.00)	576.83 (232.00)	591.65 (169.68)

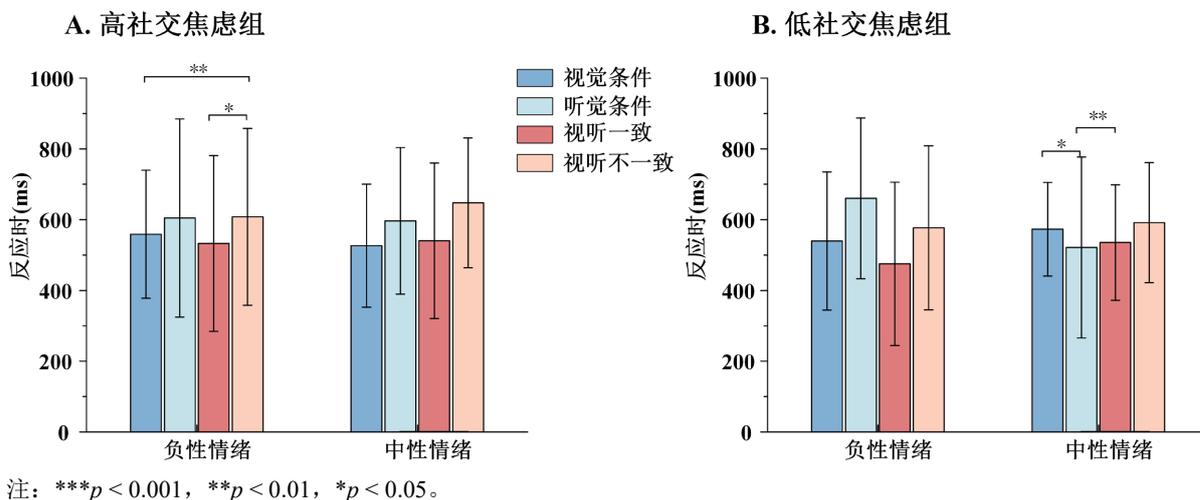


Figure 3. Incorrect reaction times in high and low social anxiety groups
图 3. 高低社交焦虑组的错误反应时

4. 讨论

本研究系统考察了不同社交焦虑水平个体的跨通道情绪加工特征, 结果揭示了显著的双通道加工优势及跨模态不一致所带来的干扰效应, 并进一步表明情绪效价和社交焦虑水平会调节上述加工模式。

4.1. 双通道情绪整合的优势效应

无论在正确率还是反应时指标上, 视听一致条件的表现均显著优于单视觉、单听觉及视听不一致条件, 呈现出典型的双通道加工优势。这表明, 当视觉与听觉情绪线索保持一致时, 个体能够更高效地提取和整合情绪特征, 从而提高识别的准确性并加快反应速度。该结果与既有研究一致, 即视听双通道信息能够显著促进情绪信息的加工效率, 表现为更快的反应和更高的正确率(Laukka et al., 2024; Schelenz et al., 2013; 张亮等, 2009; 张明等, 2022)。

进一步分析发现, 情绪效价显著调节跨通道加工模式, 表现为生气情绪条件下的跨模态整合效应更为明显。这可能是由于具有威胁性或较高情绪显著性的情绪刺激更容易触发跨感官整合机制, 从而放大双通道输入的加工优势。以往研究同样发现, 当面部表情与语音线索在恐惧或厌恶等负性情绪上保持一致时, 双通道条件相较于单通道条件可显著提高情绪识别的准确性(Collignon et al., 2008)。张明等(2022)研究亦表明, 与中性刺激相比, 视听一致呈现的负性刺激(如恐惧)能够更有效地促进跨通道整合, 并在早期加工阶段表现出更强的整合效应。来自 ERP 研究的证据进一步支持了上述观点, 相关研究发现, 负性情绪或具有高度显著性的情绪刺激(如高强度或威胁性情绪)往往在更早的时间窗内诱发更强的跨模态整合反应(Gao et al., 2018; Kokinous et al., 2015; Pan et al., 2017, 2019)。综合来看, 这些行为学与神经电生理证据共同表明, 具有威胁性的负性情绪等高情绪显著性刺激, 有助于增强跨通道情绪整合。

相比之下, 中性情绪由于缺乏突出的情绪线索, 较难自动激活跨感官整合过程, 跨模态一致性对加工的促进作用因而减弱。已有研究同样发现, 当情绪刺激的情绪强度较弱或显著性较低时, 跨通道一致性所带来的加工优势会明显下降, 甚至消失(Chen et al., 2022; Gao et al., 2018)。这表明, 情绪信号的显著性及其跨通道一致性有利于情绪信息在知觉加工阶段的整合, 从而提升跨通道情绪加工的效率。

4.2. 视听不一致出现的跨模态冲突效应

视听不一致条件在正确率上最低, 在反应时上显著延长, 构成了典型的跨模态冲突效应。这一效应

可能源于当视觉与听觉通道所传递的情绪信息在类型上不一致时, 个体在加工过程中需要处理来自不同通道的竞争性输入, 从而增加了整体的信息加工负荷。这一行为模式与以往跨模态冲突研究高度一致。已有研究指出, 不一致的视听刺激会同时激活多个情绪表征, 引发资源竞争并干扰加工过程, 进而表现为正确率下降和反应时延长(Diaconescu et al., 2011; Donohue et al., 2013; Watson et al., 2013)。在视听语音加工研究中亦发现, 视听不一致情境通常伴随更慢的反应速度(Hu et al., 2012)。总体来看, 本研究的行为结果与既有研究一致, 表明跨模态情绪不一致会显著增加加工难度, 从而引发明显的行为成本。

本研究进一步发现, 视听不一致引发的跨模态冲突效应在中性情绪条件下更为明显, 而在生气情绪条件下则相对减弱, 提示情绪效价在跨模态冲突加工中可能具有调节作用。一种解释是, 相较于中性情绪, 负性情绪刺激具有更高的情绪显著性, 在加工过程中占据更为突出的地位, 从而改变了不同通道信息之间的竞争关系。该解释在加工机制层面可与 Kokinous 等(2015)的发现相互呼应, 该研究表明, 在中性声音条件下, 即便视听信息不一致, 早期感觉加工阶段仍会发生跨模态抑制, 可能导致来自不同通道的信息同时进入后续加工, 从而增加冲突解决的负担; 而在愤怒声音条件下, 不一致信息在早期阶段即未表现出明显的跨模态抑制, 提示情绪显著性可能影响跨模态信息进入整合通道的方式, 从而调节后续的行为冲突效应。

4.3. 社交焦虑对跨模态情绪加工模式的调节

在正确率方面, 研究结果表明, 社交焦虑水平在跨模态情绪加工中发挥了一定的调节作用。在生气情绪条件下, 高、低社交焦虑组均表现出典型的双通道加工优势, 即视听一致条件下的正确率显著高于单视觉、单听觉及视听不一致条件。这一结果提示, 负性情绪刺激所具有的较高情绪显著性, 可能有助于引导加工资源更多地投入于多通道信息的整合, 从而促进跨模态情绪识别(李萍等, 2019; 张亮等, 2009)。该发现与 Yuan 等(2026)的研究结果一致, 即在负性情绪条件下, 不同社交焦虑水平的个体均可表现出明显的视听整合优势。

值得注意的是, 从效应量角度看, 高社交焦虑个体在负性情绪条件下表现出相对更强的整合效应。已有研究表明, 高社交焦虑个体通常具有较高的威胁监控水平和威胁偏向(Yang et al., 2025), 这可能使其在负性情绪情境中更倾向于对多通道线索进行深入加工, 以检测潜在威胁。该研究结果与 Heffer 等(2022)的研究一致, 其发现, 在生气情绪线索出现时, 高特质焦虑个体相较于低特质焦虑个体表现出更为显著的视听整合效应。此外, 一项 ERP 研究亦为这一观点提供了神经电生理证据。Gan 和 Li(2023)发现, 社交焦虑障碍个体在跨通道社交线索(面孔 + 声音)加工过程中, 在较早和较晚的加工阶段(如 P3 和 LPP)均表现出较健康控制组更强的跨模态整合效应, 提示高社交焦虑个体在多感官情绪信息加工中可能存在不同的整合模式。

在中性情绪条件下, 两组均表现出显著的跨模态冲突效应, 但低社交焦虑组在视听不一致条件下的正确率下降幅度更大。既有研究表明, 社交焦虑个体倾向于将模糊或情绪不明确的面孔过度解读为潜在威胁, 从而对中性表情表现出负性解释偏向(Chen, Short, & Kemps, 2020)。因此, 本研究中的中性冲突条件, 可能在高社交焦虑个体中被体验为一种“模糊威胁”或不确定情境, 而非低情绪显著性条件。这一解释与不确定性加工理论的观点相一致, 即焦虑个体在情绪效价不明确或信息模糊的情境中维持较高水平的警觉与信息监控(Carleton, 2016; Grupe & Nitschke, 2013)。在此框架下, 高社交焦虑个体可能在中性冲突条件下持续动员多通道信息监控机制, 从而在一定程度上缓冲了跨模态不一致对正确率的干扰。

在错误反应时方面, 高社交焦虑组在中性情绪条件下表现出显著的跨模态不一致效应, 视听不一致条件下错误反应时显著延长。这一结果可结合注意控制理论进行进一步解释。根据 Eysenck 等人提出的注意控制理论, 焦虑主要削弱加工效率而非加工效能, 即个体可能通过投入额外的认知资源来维持行为绩效。换言之, 在相对简单或中性情境下, 高焦虑个体可能表现出“补偿性努力”, 通过增加认知控制

与信息监控以维持正确率水平, 但代价是反应时延长。因而, 本研究中高社交焦虑个体在中性冲突条件下错误反应时的延长, 可能反映了其在不确定或模糊信息情境中为维持绩效而付出的额外认知成本, 而非单纯的加工能力下降。相比之下, 低社交焦虑个体在情绪显著性较高的生气条件下表现出明显的双通道整合优势, 提示在威胁明确情境中, 多模态一致信息能够有效促进加工效率; 但在中性冲突情境下, 由于其整体警觉水平较低, 可能较少持续动员跨通道监控资源, 从而更易受到不一致线索的干扰。

需要注意的是, 在正确反应时上, 未发现社交焦虑水平对跨模态情绪加工的显著调节作用, 这与既有研究结果一致(Heffer et al., 2022)。这一发现并不意味着社交焦虑对跨模态情绪加工缺乏影响, 而更可能反映不同反应指标在揭示加工差异方面的敏感性不同。根据注意控制理论(Eysenck et al., 2007), 焦虑并不必然损害任务表现结果, 而更可能降低加工效率, 其影响往往在高冲突或加工失败情境中更为明显。与此一致, 本研究发现社交焦虑相关的跨模态不一致效应主要体现在错误反应时上, 表明错误反应时可能较正确反应时更为敏感地反映社交焦虑个体在跨模态情绪加工中的认知负荷变化。

4.4. 研究的不足之处

尽管本研究揭示了社交焦虑水平对跨模态情绪加工的调节作用, 但仍存在一定局限。首先, 本研究主要依赖行为指标(正确率与反应时)来反映跨模态加工差异, 难以直接揭示其时间动力学特征及神经机制。未来研究可结合 ERP、EEG 或 fMRI 等技术, 从神经层面进一步刻画多通道整合过程。其次, 本研究的高社交焦虑组为非临床样本, 基于量表划分而非临床诊断, 其结果在临床群体中的推广性仍需谨慎解释。鉴于临床社交焦虑者往往表现出更显著的威胁敏感性和加工偏向(Heeren & McNally, 2018), 后续研究可纳入临床样本或采用维度化评估进行验证。最后, 本研究仅采用生气与中性情绪刺激, 限制了结论的外推性。未来研究可扩展至其他情绪类型, 系统考察不同情绪维度在跨模态整合中的作用, 以构建更为全面的情绪整合模型。

4.5. 研究意义

总体而言, 本研究从跨模态信息整合的视角揭示了社交焦虑水平不仅影响多通道信息整合的效率, 还调节个体在跨模态情绪冲突条件下的加工表现。该发现为理解社交焦虑个体在真实社会情境中的知觉与加工偏差提供了新的行为证据。社交互动中, 个体需同时整合面部表情、语音及语调等多种社会线索, 而跨模态整合过程中的偏差可能是社交焦虑者产生威胁放大、误解及社交回避的重要认知机制之一。本研究通过视听整合任务揭示了社交焦虑者在情绪不一致条件下的加工差异, 为解释现实社交情境中的沟通困难提供了更具生态效度的实验依据。

此外, 本研究拓展了焦虑相关加工偏向的研究框架。以往研究多聚焦于单一通道的威胁注意或解释偏向, 而本研究强调多通道、并行信息的整合过程, 使焦虑偏向的理解更贴近真实情绪加工的复杂性。进一步地, 社交焦虑在情绪冲突条件下对跨模态整合的调节作用提示, “冲突监控 - 情绪信息处理”机制可能在焦虑个体的跨模态加工中发挥关键作用, 为未来结合 EEG/ERP、fMRI 等技术探讨其神经基础提供了理论线索。

最后, 本研究对干预实践亦具有启示意义。针对社交焦虑者在多模态社会线索整合中的潜在偏差, 现有基于注意或信息加工的干预方式可进一步拓展为多通道社会线索训练, 通过整合声音、表情与社会情境, 提高其在复杂社交环境中对情绪线索的准确加工能力, 从而减少误解并缓解焦虑。

4.6. 研究结论

本研究表明, 社交焦虑水平显著影响跨模态情绪加工。在负性情绪条件下, 高、低社交焦虑者均表现出双通道优势, 但高社交焦虑者整合效应更强, 可能与其对潜在威胁的敏感性和持续注意监控有关。

在中性情绪条件下, 高社交焦虑者在视听不一致情境下正确率下降幅度较小, 而错误反应时延长, 提示其认知负荷较高。总体而言, 本研究拓展了社交焦虑偏向理论至多通道加工情境, 为理解焦虑个体在社会互动中的认知偏差提供了行为依据, 并为未来神经机制研究和干预策略提供了参考。

基金项目

本研究得到浙江省哲学社会科学项目(23NDJC222YB)、教育部人文社科项目(23YJCZH226)以及国家自然科学基金项目(32500961)的资助。

参考文献

- 李萍, 张明明, 李帅霞, 张火垠, 罗文波(2019). 面孔表情和声音情绪信息整合加工的脑机制. *心理科学进展*, 27(7), 1205-1214.
- 彭纯子, 龚耀先, 朱熊兆(2004). 交往焦虑量表的信效度及其在中国大学生中的适用性. *中国心理卫生杂志*, 18(1), 39-41.
- 宋素涛, 李爽, 赵诗梦, 肖观来, 张锦秀, 郑元杰(2023). 社交焦虑者对面孔表情存在持续性注意偏向: 来自 N2pc 的证据. *中国临床心理学杂志*, 31(2), 267-273.
- 张亮, 孙向红, 张侃(2009). 情绪信息的多通道整合. *心理科学进展*, 17(6), 1133-1138.
- 张明, 王婷婷, 吴晓刚, 张月娥, 王爱君(2022). 面孔表情和声音情绪信息整合对返回抑制的影响. *心理学报*, 54(4), 331-342.
- Belin, P., Fillion-Bilodeau, S., & Gosselin, F. (2008). The Montreal Affective Voices: A Validated Set of Nonverbal Affect Bursts for Research on Auditory Affective Processing. *Behavior Research Methods*, 40, 531-539. <https://doi.org/10.3758/brm.40.2.531>
- Carleton, R. N. (2016). Fear of the Unknown: One Fear to Rule Them All? *Journal of Anxiety Disorders*, 41, 5-21. <https://doi.org/10.1016/j.janxdis.2016.03.011>
- Chen, J., Short, M., & Kemps, E. (2020). Interpretation Bias in Social Anxiety: A Systematic Review and Meta-Analysis. *Journal of Affective Disorders*, 276, 1119-1130. <https://doi.org/10.1016/j.jad.2020.07.121>
- Chen, M., Zhao, S., Yu, J., Leng, X., Zhai, M., Feng, C. et al. (2022). Audiovisual Emotional Congruency Modulates the Stimulus-Driven Cross-Modal Spread of Attention. *Brain Sciences*, 12, Article 1229. <https://doi.org/10.3390/brainsci12091229>
- Collignon, O., Girard, S., Gosselin, F., Roy, S., Saint-Amour, D., Lassonde, M. et al. (2008). Audio-Visual Integration of Emotion Expression. *Brain Research*, 1242, 126-135. <https://doi.org/10.1016/j.brainres.2008.04.023>
- Diaconescu, A. O., Alain, C., & McIntosh, A. R. (2011). The Co-Occurrence of Multisensory Facilitation and Cross-Modal Conflict in the Human Brain. *Journal of Neurophysiology*, 106, 2896-2909. <https://doi.org/10.1152/jn.00303.2011>
- Dolan, R. J., Morris, J. S., & de Gelder, B. (2001). Crossmodal Binding of Fear in Voice and Face. *Proceedings of the National Academy of Sciences*, 98, 10006-10010. <https://doi.org/10.1073/pnas.171288598>
- Donohue, S. E., Appelbaum, L. G., Park, C. J., Roberts, K. C., & Woldorff, M. G. (2013). Cross-Modal Stimulus Conflict: The Behavioral Effects of Stimulus Input Timing in a Visual-Auditory Stroop Task. *PLOS ONE*, 8, e62802. <https://doi.org/10.1371/journal.pone.0062802>
- Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and Cognitive Performance: Attentional Control Theory. *Emotion*, 7, 336-353. <https://doi.org/10.1037/1528-3542.7.2.336>
- Filippi, P. (2016). Emotional and Interactional Prosody across Animal Communication Systems: A Comparative Approach to the Emergence of Language. *Frontiers in Psychology*, 7, Article 1393. <https://doi.org/10.3389/fpsyg.2016.01393>
- Gan, S., & Li, W. (2023). Aberrant Neural Correlates of Multisensory Processing of Audiovisual Social Cues Related to Social Anxiety: An Electrophysiological Study. *Frontiers in Psychiatry*, 14, Article 1020812. <https://doi.org/10.3389/fpsyg.2023.1020812>
- Gao, C., Wedell, D. H., Green, J. J., Jia, X., Mao, X., Guo, C. et al. (2018). Temporal Dynamics of Audiovisual Affective Processing. *Biological Psychology*, 139, 59-72. <https://doi.org/10.1016/j.biopsycho.2018.10.001>
- Grupe, D. W., & Nitschke, J. B. (2013). Uncertainty and Anticipation in Anxiety: An Integrated Neurobiological and Psychological Perspective. *Nature Reviews Neuroscience*, 14, 488-501. <https://doi.org/10.1038/nrn3524>
- Heeren, A., & McNally, R. J. (2018). Social Anxiety Disorder as a Densely Interconnected Network of Fear and Avoidance for Social Situations. *Cognitive Therapy and Research*, 42, 103-113. <https://doi.org/10.1007/s10608-017-9876-3>

- Heffer, N., Gradidge, M., Karl, A., Ashwin, C., & Petrini, K. (2022). High Trait Anxiety Enhances Optimal Integration of Auditory and Visual Threat Cues. *Journal of Behavior Therapy and Experimental Psychiatry*, 74, Article ID: 101693. <https://doi.org/10.1016/j.jbtep.2021.101693>
- Hu, Z., Zhang, R., Zhang, Q., Liu, Q., & Li, H. (2012). Neural Correlates of Audiovisual Integration of Semantic Category Information. *Brain and Language*, 121, 70-75. <https://doi.org/10.1016/j.bandl.2012.01.002>
- Joassin, F., Maurage, P., Bruyer, R., Crommelinck, M., & Campanella, S. (2004). When Audition Alters Vision: An Event-Related Potential Study of the Cross-Modal Interactions between Faces and Voices. *Neuroscience Letters*, 369, 132-137. <https://doi.org/10.1016/j.neulet.2004.07.067>
- Klasen, M., Chen, Y., & Mathiak, K. (2012). Multisensory Emotions: Perception, Combination and Underlying Neural Processes. *Reviews in the Neurosciences*, 23, 381-392. <https://doi.org/10.1515/revneuro-2012-0040>
- Klasen, M., Kenworthy, C. A., Mathiak, K. A., Kircher, T. T. J., & Mathiak, K. (2011). Supramodal Representation of Emotions. *The Journal of Neuroscience*, 31, 13635-13643. <https://doi.org/10.1523/jneurosci.2833-11.2011>
- Koizumi, A., Tanaka, A., Imai, H., Hiramatsu, S., Hiramoto, E., Sato, T. et al. (2011). The Effects of Anxiety on the Interpretation of Emotion in the Face-Voice Pairs. *Experimental Brain Research*, 213, 275-282. <https://doi.org/10.1007/s00221-011-2668-1>
- Kokinous, J., Kotz, S. A., Tavano, A., & Schröger, E. (2015). The Role of Emotion in Dynamic Audiovisual Integration of Faces and Voices. *Social Cognitive and Affective Neuroscience*, 10, 713-720. <https://doi.org/10.1093/scan/nsu105>
- Kreifelts, B., Ethofer, T., Wiegand, A., Brück, C., Wächter, S., Erb, M. et al. (2020). The Neural Correlates of Face-Voice-Integration in Social Anxiety Disorder. *Frontiers in Psychiatry*, 11, Article 657. <https://doi.org/10.3389/fpsy.2020.00657>
- Laukka, P., Månsson, K. N. T., Cortes, D. S., Manzouri, A., Frick, A., Fredborg, W. et al. (2024). Neural Correlates of Individual Differences in Multimodal Emotion Recognition Ability. *Cortex*, 175, 1-11. <https://doi.org/10.1016/j.cortex.2024.03.009>
- Leary, M. R. (1983). Social Anxiousness: The Construct and Its Measurement. *Journal of Personality Assessment*, 47, 66-75. https://doi.org/10.1207/s15327752jpa4701_8
- Li, S., Li, Y., Li, R., Li, Y., Wu, J., Yang, W. et al. (2024). The Modulatory Effect of Exogenous Orienting on Audiovisual Emotional Integration: An ERP Study. *Journal of Integrative Neuroscience*, 23, Article No. 210. <https://doi.org/10.31083/j.jin2311210>
- Liu, P., Sutherland, M., & Pollick, F. E. (2021). Incongruence Effects in Cross-Modal Emotional Processing in Autistic Traits: An fMRI Study. *Neuropsychologia*, 161, Article 107997. <https://doi.org/10.1016/j.neuropsychologia.2021.107997>
- Müller, V. I., Habel, U., Derntl, B., Schneider, F., Zilles, K., Turetsky, B. I. et al. (2011). Incongruence Effects in Crossmodal Emotional Integration. *NeuroImage*, 54, 2257-2266. <https://doi.org/10.1016/j.neuroimage.2010.10.047>
- Pan, F., Zhang, L., Ou, Y., & Zhang, X. (2019). The Audio-Visual Integration Effect on Music Emotion: Behavioral and Physiological Evidence. *PLOS ONE*, 14, e0217040. <https://doi.org/10.1371/journal.pone.0217040>
- Pan, Z., Liu, X., Luo, Y., & Chen, X. (2017). Emotional Intensity Modulates the Integration of Bimodal Angry Expressions: ERP Evidence. *Frontiers in Neuroscience*, 11, Article 349. <https://doi.org/10.3389/fnins.2017.00349>
- Peschard, V., & Philippot, P. (2017). Overestimation of Threat from Neutral Faces and Voices in Social Anxiety. *Journal of Behavior Therapy and Experimental Psychiatry*, 57, 206-211. <https://doi.org/10.1016/j.jbtep.2017.06.003>
- Proverbio, A. M., & De Benedetto, F. (2018). Auditory Enhancement of Visual Memory Encoding Is Driven by Emotional Content of the Auditory Material and Mediated by Superior Frontal Cortex. *Biological Psychology*, 132, 164-175. <https://doi.org/10.1016/j.biopsycho.2017.12.003>
- Proverbio, A. M., Camporeale, E., & Brusa, A. (2020). Multimodal Recognition of Emotions in Music and Facial Expressions. *Frontiers in Human Neuroscience*, 14, Article 32. <https://doi.org/10.3389/fnhum.2020.00032>
- Robins, D. L., Hunyadi, E., & Schultz, R. T. (2009). Superior Temporal Activation in Response to Dynamic Audio-Visual Emotional Cues. *Brain and Cognition*, 69, 269-278. <https://doi.org/10.1016/j.bandc.2008.08.007>
- Schelenz, P. D., Klasen, M., Reese, B., Regenbogen, C., Wolf, D., Kato, h. Y. et al. (2013). Multisensory Integration of Dynamic Emotional Faces and Voices: Method for Simultaneous EEG-fMRI Measurements. *Frontiers in Human Neuroscience*, 7, Article 729. <https://doi.org/10.3389/fnhum.2013.00729>
- Schirmer, A., & Adolphs, R. (2017). Emotion Perception from Face, Voice, and Touch: Comparisons and Convergence. *Trends in Cognitive Sciences*, 21, 216-228. <https://doi.org/10.1016/j.tics.2017.01.001>
- Schofield, C. A., Johnson, A. L., Inhoff, A. W., & Coles, M. E. (2012). Social Anxiety and Difficulty Disengaging Threat: Evidence from Eye-Tracking. *Cognition & Emotion*, 26, 300-311. <https://doi.org/10.1080/02699931.2011.602050>
- Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A. et al. (2009). The NimStim Set of Facial Expressions: Judgments from Untrained Research Participants. *Psychiatry Research*, 168, 242-249. <https://doi.org/10.1016/j.psychres.2008.05.006>

- Watson, R., Latinus, M., Noguchi, T., Garrod, O., Crabbe, F., & Belin, P. (2013). Dissociating Task Difficulty from Incongruence in Face-Voice Emotion Integration. *Frontiers in Human Neuroscience*, 7, Article 744. <https://doi.org/10.3389/fnhum.2013.00744>
- Yang, J., Si, F., & Cao, J. (2025). Electrophysiological Evidence of Processing Social Threat Words in Social Anxiety Participants. *BMC Psychiatry*, 25, Article No. 821. <https://doi.org/10.1186/s12888-025-07257-6>
- Yeh, P., Geangu, E., & Reid, V. (2016). Coherent Emotional Perception from Body Expressions and the Voice. *Neuropsychologia*, 91, 99-108. <https://doi.org/10.1016/j.neuropsychologia.2016.07.038>
- Yuan, J., Liu, X., Chen, X., Li, M., Liu, Z., & Li, B. (2026). Audiovisual Integration of Faces and Voices in Chinese Individuals with Social Anxiety. *Behavioural Brain Research*, 497, Article ID: 115905. <https://doi.org/10.1016/j.bbr.2025.115905>