

音乐训练促进音高与时间维度 在听觉工作记忆中的交互*

周临舒¹ 张雨青¹ 蔡丹超²

(¹上海师范大学音乐学院, 上海 200234) (²上海市公共卫生临床中心, 上海 201508)

摘要 音乐训练可增强个体对时间性和非时间性结构的敏感性, 但这些结构在工作记忆中的联合影响尚不明确。本研究探讨了音高和节奏结构在听觉工作记忆中的加工方式以及音乐训练的作用。实验操控了不同长度旋律中的音高和节奏结构, 音乐家和非音乐家根据音高或节奏维度的变化做出相同-不同的判断, 同时抑制另一维度干扰。结果显示, 在音高保持任务中, 非音乐家独立加工音高和节奏结构, 而音乐家则以交互方式加工, 且交互效应与音乐素养评分呈正相关; 在节奏保持任务中, 两组被试均以独立方式加工结构, 表明音乐训练对结构整合的影响受任务类型调节。此外, 交互效应在较短序列的音高保持任务中更明显, 提示这种整合进一步受任务属性与难度的限制。这些发现支持动态注意理论, 表明音乐训练能增强个体在多维信息整合加工中的灵活性和适应性。

关键词 音乐训练, 听觉工作记忆, 时间规则, 音乐结构, 动态注意理论

分类号 B842

1 前言

人类认知系统面临的核心挑战在于如何将离散的感知维度整合为统一表征。音乐作为典型的多维信息载体, 同时涉及音高(“what”)和节奏(“when”)两种层级结构, 为探索多维信息整合机制提供了理想模型(Krumhansl, 2000)。一般而言, 音乐音高通过调性规则构建音级稳定性层级(Krumhansl, 1990), 而节奏则依赖节拍周期性形成时间预期框架(Prince, Thompson, & Schmuckler, 2009)。这种双重层级结构创造了有序的声音模式, 且有助于探索人类并行加工时间性与非时间性信息的认知能力(Fitch, 2013; Prince, Thompson, & Schmuckler, 2009)。已有研究表明, 这两类结构均显著影响大脑的加工效率。例如, 在工作记忆任务中, 遵循调性原则的音高序列相比无调性材料拥有更明显的模块化特点, 因而更易于被存储(Albouy et al., 2013; Bharucha & Krumhansl, 1983; Dowling, 1991;

Lévêque et al., 2022; Schulze et al., 2012); 而相比复杂节奏, 简单节奏拥有可预测性更高的时间框架, 有助于提升再现的精度和反应速度(Essens & Povel, 1985; Martin et al., 2005; Sakai et al., 1999)。然而, 这些发现均源自维度分离范式, 实质上将音乐解构为单一维度的听觉刺激。这种研究取向存在一定局限: 一方面, 它忽视了真实音乐认知过程中多维结构之间的耦合; 另一方面, 分离范式无法揭示核心认知机制——音高与节奏的层级结构是作为独立的模块存储, 还是通过加工资源的协调实现整合。因此, 探究两类结构在工作记忆中的加工关系模式, 不仅关乎音乐信息的存储机制本质, 也有助于探讨多维信息的整合加工。

音乐多维加工的理论建构主要涉及两类假说。动态注意理论(Jones, 1976; Jones & Boltz, 1989)强调时间结构对认知资源的全局性支配: 当节奏层级形成稳定的周期性预期时, 注意脉冲会与之同步振荡, 从而在特定时间窗口优先捕获与之协调的音高

收稿日期: 2024-09-18

* 上海市卫生健康委员会青年科研项目(20204Y0334)资助。

通信作者: 周临舒, E-mail: zhoulsh@shnu.edu.cn; 蔡丹超, E-mail: danchao.cai@outlook.com

事件。基于这一理论,音乐工作记忆存在跨维度的增益效应——当音高事件与节奏重音对齐时,工作记忆精度显著提升(Large & Jones, 1999)。与此不同,双成分模型主张音高和节奏加工相互独立:音高结构基于规则的符号化编码,而节奏加工基于时间模拟的预测机制(Povel & Essens, 1985),二者依赖于不同的脑内神经通路(Jerde et al., 2011; Schwartz & Kotz, 2013)。

动态注意理论的核心预测——时间结构对非时间加工的全局性调制——在特定任务条件下得到部分支持。例如,当要求听众整体评估音乐完成感时,非等时性节奏会削弱和声结构的影响;但若任务是局部和弦判断,这种跨维度干扰便消失(Tillmann & Lebrun-Guillaud, 2006)。这提示时间维度的支配作用可能受任务调节:整体加工促进维度整合,而局部加工依赖模块化处理。类似地,Prince 和 Schmuckler 等人(2009)发现,非层级性的时间变化仅能干扰无调性背景下的音高判断,在调性背景下则失效,表明调性结构可能缓冲时间维度的干扰,这与动态注意理论中“结构协调增强整合”的假设一致。然而,双成分模型的功能独立性假说同样获得证据支撑。当要求被试分别判断调性稳定性和节奏位置时,节奏层级受到调性结构影响,但调性判断却独立于节奏语境(Prince, Thompson, & Schmuckler, 2009)。这说明即便存在维度间影响,其作用方向与强度仍受任务需求的约束。此外,音乐句法加工的研究显示,对音高与节奏结构的早期探测彼此独立,其交互效应仅出现在晚期整合阶段,表明维度特异性加工早于策略性整合(Sun et al., 2020; Zhang et al., 2019)。

尽管这些发现深化了对音乐多维加工的理解,但仍存在一些知识缺口:已有的实验任务大多聚焦于即时知觉判断,尚不清楚音乐多维信息在工作记忆存储中的加工特性;同时,结构层级性与任务类型的协同效应尚未系统考察,音高与节奏的交互究竟是音乐认知的稳定属性,还是特定任务情境催生的附带现象?这一问题可以在选择性注意范式中通过行为分离得到直接检验。若双成分模型成立,则任务无关维度的结构变化不应影响目标维度的加工表现;反之,若动态注意理论主导,则时间结构的层级性将跨任务影响音高加工效率。

音乐训练对多维结构加工的调节作用进一步增加了该问题的复杂性。尽管一般听众通过长期曝

露于西方调性环境即可获得基础的音乐结构加工能力(Koelsch et al., 2000; Koelsch et al., 2007),专业训练可显著增强这种认知优势。音乐家在工作记忆任务中表现出调性结构加工能力的提升(Schulze et al., 2011),并且能更精准地同步复杂节奏(Chen et al., 2008)。同时,他们还表现出对音高(Jentschke & Koelsch, 2009; Koelsch et al., 2002)和节奏结构(Sun et al., 2018)的句法加工优势。然而,对音乐训练的效果评估通常基于解构的单一维度(音高或节奏),而真实音乐体验必然涉及多维结构的实时绑定。值得注意的是,当要求同时加工音高和节奏结构时,音乐家的句法整合优势反而消失(Sun et al., 2020)。这引出一个关键问题:音乐训练赋予的加工优势是否仅体现为维度特异性的加工能力强化,还是同时涉及跨维度整合能力的提升?

基于此,本研究旨在探讨音乐工作记忆中音高和节奏结构的交互机制,并探讨音乐训练的调节作用。我们招募音乐家和非音乐家完成两个独立的实验,分别考察他们对音高和节奏信息的保持。实验 1 聚焦于听觉工作记忆中的音高保持,要求被试在抑制节奏变化的条件下判断音高异同;实验 2 则逆转任务要求,考察听觉工作记忆中的节奏保持。两个实验都操控了音高结构(调性 vs. 无调性)和节奏结构(简单节奏 vs. 复杂节奏),以探讨二者的加工是否独立或交互。如果音高结构的变化影响了节奏结构的效应,或节奏结构的变化影响了音高结构的效应(即出现交互作用),则表明两者在加工过程中被整合,从而实现结构增益。相反,如果音高结构和节奏结构各自产生独立的主效应,而二者之间未表现出交互作用,则表明在当前任务条件下,音高和节奏结构的加工相对独立(Prince, 2011; Prince & Pfordresher, 2012; Tillmann & Lebrun-Guillaud, 2006)。

基于前人研究,我们假设两组被试对结构化信息(调性、简单节奏)的工作记忆优于非结构化条件(无调性、复杂节奏),但音乐家的优势幅度更大(Schulze et al., 2011)。此外,考虑到音乐训练可以提高对层级性音高和节奏结构的敏感性(Koelsch et al., 2002; Sun et al., 2018),我们预期音乐家由于其增强的信息整合技能(周临舒 等, 2017),可能会更交互地加工音高和节奏结构。最后,我们预测两个实验的结果模式会有所不同,反映任务特定的认知过程对音乐结构整合的不同影响。

2 实验 1：音高保持

实验 1 通过音高的相同-不同任务考察音乐家和非音乐家对音高信息的保持。工作记忆的容量限制了可以记住的材料的数量：序列越长(或事件越多)，工作记忆的表现越差(Cowan, 2000)。考虑到前人同类研究主要采用五至七音的序列来考察音乐工作记忆(比如 Schulze et al., 2012)，本研究选择五音和七音序列来考察序列长度的影响。因此，实验同时操控了旋律刺激的音高结构(调性 vs. 无调性)、节奏结构(简单节奏 vs. 复杂节奏)以及序列长度(五音 vs. 七音)。这一设计旨在了解音乐训练如何影响音高保持，以及不同的音乐结构和序列长度如何影响两组被试的表现。

2.1 方法

2.1.1 被试

使用 G*Power 3.1 (Faul et al., 2009)计算了所需样本量。我们基于之前考察音乐训练和音乐结构加工的研究(Chen et al., 2008)的效应量进行计算。为了在显著性水平 0.05 下检测较大效应($f = 0.739$)并达到 80%的统计检验力，每组至少需要 17 名被试。为确保足够的统计效能，并考虑个体差异和实验设计的需求，最终招募了 36 名音乐家和 36 名非音乐家作为本研究的被试，以提高结果的稳健性和可推广性。音乐家的乐器训练平均时长为 13.92 年($SD = 2.97$)，每天平均练习 4.04 小时($SD = 0.91$)，其主修乐器包括钢琴、小提琴、单簧管和吉他。

非音乐家未接受过学校音乐课以外的任何课外音乐训练。正式实验前，所有被试完成了非语言智力测试(Raven 标准推理测验, SPM; Raven, 2000)和 Goldsmiths 音乐素养指数的评估(Gold-MSI; Lin et al., 2021; Müllensiefen et al., 2014)。如表 1 所示，音乐家与非音乐家在年龄、性别、偏手性、受教育年限及非语言智力上均无显著差异，但音乐家的音乐素养指数显著高于非音乐家。所有实验得到上海师范大学学术伦理与道德委员会的批准。所有被试在实验前签署了知情同意书。实验结束后，被试获得相应的报酬。

2.1.2 刺激

实验刺激由自行创作的单声部旋律组成，每条旋律持续四拍。为确保旋律长度的多样性，我们最初创作了 20 条旋律，其中包括 10 条五音序列和 10 条七音序列。我们根据以往研究对旋律的音高和节奏结构进行了操控。音高结构分为调性和无调性水平(如 Schulze et al., 2011; Schulze et al., 2012)，而节奏结构则分为简单节奏和复杂节奏水平(如 Chen et al., 2008)。图 1 展示了这些旋律刺激的示例。

具体来说，调性旋律使用 C 大调音阶中的音符(C4: 261.63 Hz, D4: 293.66 Hz, E4: 329.63 Hz, F4: 349.23 Hz, G4: 392 Hz, A4: 440 Hz, B4: 493.88 Hz, C5: 523 Hz)，并以 C 大调主和弦的音符开始，以主音(C) 或属音(G)结尾，营造出强烈的调性感。而无调性旋律则基于全音阶，从 C 音开始(C4: 261.63 Hz, D4: 293.66 Hz, E4: 329.63 Hz, F#4: 369.99 Hz, G#4:

表 1 音乐家和非音乐家被试的基本信息

变量	音乐家($n = 36$)	非音乐家($n = 36$)	t	p	Cohen's d	95% CI
年龄(岁)	23.58 (1.96)	23.25 (1.36)	0.84	0.405	0.20	[-0.46, 1.13]
性别	24 女, 12 男	24 女, 12 男	—	—	—	—
偏手性	右利手	右利手	—	—	—	—
教育年限	17.00 (1.27)	16.92 (0.91)	0.32	0.749	0.08	[-0.43, 0.60]
Raven's SPM	64.31 (5.33)	64.58 (3.82)	-0.25	0.800	-0.06	[-2.46, 1.90]
Gold-MSI	183.72 (13.66)	143.22 (12.11)	13.31	< 0.001	3.14	[34.43, 46.57]



图 1 五音序列和七音序列中音高与节奏结构组合的旋律刺激示例

415.305 Hz, A#4: 466.164 Hz, C5: 523 Hz), 该音阶既没有调性中心, 也无法形成传统的大小三和弦, 难以产生调性感。我们使用 Krumhansl 和 Schmuckler 提出的调性识别算法(见 Krumhansl, 1990)分析了旋律上下文所建立的调性中心强度。该算法计算最大的调性属性相关值(Maximum Key-Profile Correlation, MKC), 即与最佳匹配调性层级的相关系数, 用于指示最强的调性确立程度。我们分别计算了本研究所有旋律序列与 12 个大调和 12 个小调的调性属性的相关性。以音高结构(调性 vs. 无调性)和序列长度(五音 vs. 七音)作为自变量, 以 MKC 作为因变量进行两因素方差分析(ANOVA)发现: 音高结构的主效应显著(0.76 vs. 0.48, $F(1, 76) = 121.74$, $p < 0.001$, $\eta_p^2 = 0.61$), 表明调性刺激的 MKC 高于无调性刺激; 序列长度的主效应、音高结构和序列长度的交互作用均不显著($ps > 0.109$)。此外, 我们使用 MIRtoolbox (Lartillot et al., 2008)进行声学分析, 计算与调性相关的两个参数: 色谱质心 (Chromagram Centroid) 和调性清晰度 (Key Clarity)。两因素 ANOVA 结果表明: 无论是色谱质心(0.24 vs. 0.11, $F(1, 76) = 9.32$, $p = 0.003$, $\eta_p^2 = 0.11$), 还是调性清晰度(0.74 vs. 0.45, $F(1, 76) = 146.45$, $p < 0.001$, $\eta_p^2 = 0.66$), 均存在显著的音高结构主效应, 调性和无调性条件存在显著差异; 而序列长度的主效应、音高结构和序列长度的交互作用均不显著($ps > 0.083$)。这些结果验证了调性和无调性旋律在调性特征上的差异。

简单节奏序列由基本音符(如二分音符、四分音符、附点四分音符等)构成, 不含切分音, 保证了节拍的稳定。而复杂节奏序列则包含切分音, 重音落在弱拍或次拍上, 导致节奏冲突并形成不稳定的时间结构。根据 Povel 和 Essens (1985)的理论, 听者会识别出一个最佳的内部时钟间隔, 使其与给定的时间模式尽可能对齐。该时钟的适切性由 C 指标决定, 该指标用于量化单一时钟间隔维持的反证程度。C 指标的计算方式如下: $C = (W \times -ev) + (1 \times 0ev)$ 。其中, 0ev 代表与非重音事件重合的时钟刻度数, -ev 反映落在静音位置(休止)上的时钟刻度数, W 是权重常数(通常 $W = 4$; 详见 Povel & Essens, 1985)。较低的 C 值表示较少的反证, 即更强的时间规则性。在本研究中, 我们针对所有简单节奏和复杂节奏序列计算了 C 值, 采用 200 ms 作为基本时间单位(即本实验材料的最短间隔), 从而将节奏模式表征为该单位的 1、2、3 或 4 倍。随后, 我们确

定了每个序列 C 值最低的时钟作为其最佳内部时钟, 并以此作为相对节拍强度的指标。结果表明, 对于简单节奏和复杂节奏序列, 4 个时间单位的时钟是最优内部时钟。对最小 C 值进行两因素 ANOVA 发现: 节奏结构的主效应显著, 简单节奏序列的 C 值低于复杂节奏序列, 2.45 vs. 4.05, $F(1, 76) = 10.89$, $p = 0.001$, $\eta_p^2 = 0.13$; 序列长度的主效应显著, 七音序列的 C 值低于五音序列, 2.40 vs. 4.10, $F(1, 76) = 12.30$, $p < 0.001$, $\eta_p^2 = 0.14$; 但节拍类型与序列长度的交互作用不显著, $p = 0.681$ 。这些结果表明, 简单节奏序列的时间规则性比复杂节奏序列更强。

根据音高和节奏结构的四种组合条件, 我们生成了 80 个旋律刺激: 20 个调性旋律与简单节奏的组合, 20 个调性旋律与复杂节奏的组合, 20 个无调性旋律与简单节奏的组合, 以及 20 个无调性旋律与复杂节奏的组合。每组包含 10 个五音序列和 10 个七音序列, 因此共计 40 条五音旋律和 40 条七音旋律。本研究共 8 个实验条件(音高结构×节奏结构×序列长度), 每个条件包含 10 条旋律刺激。

每个旋律刺激的持续时间为 3200 ms, 速度设定为每分钟 75 拍。我们将创建的 MIDI 文件导入 Cubase 5.1 软件, 使用 YAMAHA S90ES 大钢琴音色, 并最终通过 Adobe Audition CS6 导出为 WAV 文件。所有刺激均为单声道, 采样率为 44.1 kHz, 位深度为 16 位, 并使用 Adobe Audition CS6 将音量标准化至约 68 dB SPL。

2.1.3 程序

本实验采用音高识别范式, 每个旋律刺激在每个试次中都会呈现两次。因此, 在 80 条原始音乐旋律的基础上, 创建了 160 个试次, 每个试次包含先后呈现的两条旋律。在一半(80 个)试次中, 第二条旋律的节奏不变但音高改变; 而在另一半(80 个)试次中, 第二条旋律的节奏改变但音高不变。具体做法如下: 在节奏不变但音高改变的条件下, 旋律中的一个音高会升高或降低一个半音或一个全音, 升降方向的概率均为 50%, 并确保音高调整不会改变调性结构或原本的旋律轮廓; 在节奏改变但音高不变的条件下, 调换了旋律第二或第三拍音符时值的顺序, 或将一个音符时值拆分为更小的单位再重新进行音值组合, 并确保节奏调整不会改变音符的数量或原本的节奏结构(简单节奏或复杂节奏)。为避免任务过于简单从而导致天花板效应, 音高或节奏变化不会出现在旋律的开头或结尾, 而是随机出现

在旋律中间。因此, 每个实验条件包含 20 个试次, 其中一半是相同的配对, 另一半是不同的配对。此外, 为避免被试发展出特定的反应策略, 加入了 32 个填充性质的试次, 这些试次中的两条旋律完全相同。填充试次的的数据不进入统计分析。

实验在安静的房间内进行, 程序通过电脑呈现, 音乐刺激通过耳机播放。每个试次从呈现第一条旋律开始, 该旋律持续 3200 ms, 随后进入 2000 ms 的保持阶段。接着播放第二条 3200 ms 的旋律。旋律呈现结束后是反应界面。实验任务是要求被试在忽略节奏信息的前提下, 快速且准确地判断连续呈现的两条旋律的音高是否完全相同。按键的分配在被试的左右手之间进行了平衡。每次作答后, 会有 500 ms 的空白屏幕, 之后开始下一次试次。

在正式实验前, 被试完成四个带有反馈的练习试次, 以熟悉实验程序和任务。正式实验包含四个区组, 每个区组包含 40 个试次。刺激呈现顺序为伪随机, 确保相同或不同的配对不连续出现超过三次, 且每对中的第一个旋律不会在试次中连续重复。

2.1.4 统计分析

基于信号检测理论, 我们使用标准公式 (Macmillan & Creelman, 2004) 计算了探测敏感性 [$d' = z(\text{击中率}) - z(\text{虚报率})$] 和反应偏差 [$c = -0.5 \times (z(\text{击中率}) + z(\text{虚报率}))$]。 d' 值反映了被试区分“相同”与“不同”试次的能力, 数值越高, 说明敏感性越强。 c 值反映了反应偏向, 正的 c 值表示被试倾向于选择“不同”, 而负的 c 值表示倾向于选择“相同”。击中率定义为在“不同”试次中正确判断的比例, 虚报率则是“相同”试次中错误判断的比例。对于没有虚报的情况, d' 和 c 的修正值设为 0.05, 而对于击中数达到最大值的情况, 修正值设为 0.95。此外, 为了考察记忆任务的决策速度, 我们还计算了每种条件下正确反应的平均反应时, 反应时的计算仅基于正确的试次进行分析。为了研究实验条件的影响, 我们进行了四因素重复测量 ANOVA。组别作为被试间变量, 音高结构、节奏结构和序列长度作为被试内变量。我们分别对 d' 、 c 和反应时进行了分析, 以探讨这些因素在不同指标上的主效应和交互作用。

2.2 结果和讨论

2.2.1 探测敏感性

探测敏感性的 ANOVA 结果表明, 组别的主效应显著, $F(1, 70) = 49.56, p < 0.001, \eta_p^2 = 0.42$, 音乐家的探测敏感性高于非音乐家 (1.93 vs. 0.94, 95% CI = [0.71, 1.27])。音高结构的主效应显著, $F(1, 70)$

$= 64.03, p < 0.001, \eta_p^2 = 0.48$, 对调性旋律的探测敏感性高于无调性旋律 (1.64 vs. 1.24, 95% CI = [0.29, 0.50])。节奏结构的主效应显著, $F(1, 70) = 24.36, p < 0.001, \eta_p^2 = 0.26$, 简单节奏条件下的探测敏感性高于复杂节奏条件 (1.59 vs. 1.28, 95% CI = [0.19, 0.43])。序列长度的主效应显著, $F(1, 70) = 10.70, p = 0.002, \eta_p^2 = 0.13$, 对五音序列的探测敏感性高于七音序列 (1.53 vs. 1.34, 95% CI = [0.07, 0.30])。同时, 组别和序列长度的交互作用显著, $F(1, 70) = 4.17, p = 0.045, \eta_p^2 = 0.06$ 。简单效应分析表明, 音乐家对五音序列的探测敏感性比七音序列更高, 2.08 vs. 1.78, $t(35) = 4.06, p < 0.001, \text{Cohen's } d = 0.68, 95\% \text{ CI} = [0.15, 0.45]$, 但是非音乐家的敏感性不受序列长度影响, 0.98 vs. 0.91, $t(35) = 0.81, p = 0.422$ 。此外, 组别、音高结构、节奏结构三因素交互作用边缘显著, $F(1, 70) = 3.97, p = 0.050, \eta_p^2 = 0.05$, 以及音高结构、节奏结构、序列长度三因素交互作用显著, $F(1, 70) = 8.57, p = 0.005, \eta_p^2 = 0.11$ 。以上结果验证了音乐家在音高探测敏感性方面的优势, 且音高探测敏感性受到音高结构、节奏结构和序列长度影响, 具体表现为调性旋律、简单节奏和较短序列条件下敏感性更高。

为进一步探讨组别、音高结构、节奏结构之间边缘显著的交互作用, 分别对音乐家和非音乐家的数据进行两因素 ANOVA。对于音乐家, 发现音高结构的主效应显著, $F(1, 35) = 54.50, p < 0.001, \eta_p^2 = 0.61$, 对调性旋律的敏感性高于无调性旋律 (2.21 vs. 1.65, 95% CI = [0.40, 0.70]); 节奏结构的主效应显著, $F(1, 35) = 24.32, p < 0.001, \eta_p^2 = 0.41$, 对简单节奏序列的敏感性高于复杂节奏序列 (2.09 vs. 1.77, 95% CI = [0.19, 0.46]); 同时, 二者交互作用显著, $F(1, 35) = 6.32, p = 0.017, \eta_p^2 = 0.15$ 。如图 2 所示, 简单效应分析表明, 无论在调性或无调性条件下, 音乐家对简单结构的敏感性都比复杂结构更高, 但调性条件下的效应 (2.43 vs. 1.98, $t(35) = 5.36, p < 0.001, \text{Cohen's } d = 0.89, 95\% \text{ CI} = [0.28, 0.63]$) 比无调性条件下的 (1.75 vs. 1.56, $t(35) = 2.30, p = 0.027, \text{Cohen's } d = 0.38, 95\% \text{ CI} = [0.02, 0.36]$) 更大; 相似地, 在简单或复杂节奏条件下, 音乐家对调性旋律的敏感性都比无调性旋律更高, 但简单节奏条件下的效应 (2.43 vs. 1.75, $t(35) = 8.64, p < 0.001, \text{Cohen's } d = 1.44, 95\% \text{ CI} = [0.52, 0.85]$) 比复杂节奏条件下的 (1.98 vs. 1.56, $t(35) = 4.11, p < 0.001, \text{Cohen's } d = 0.69, 95\% \text{ CI} = [0.21, 0.63]$) 更大。对于

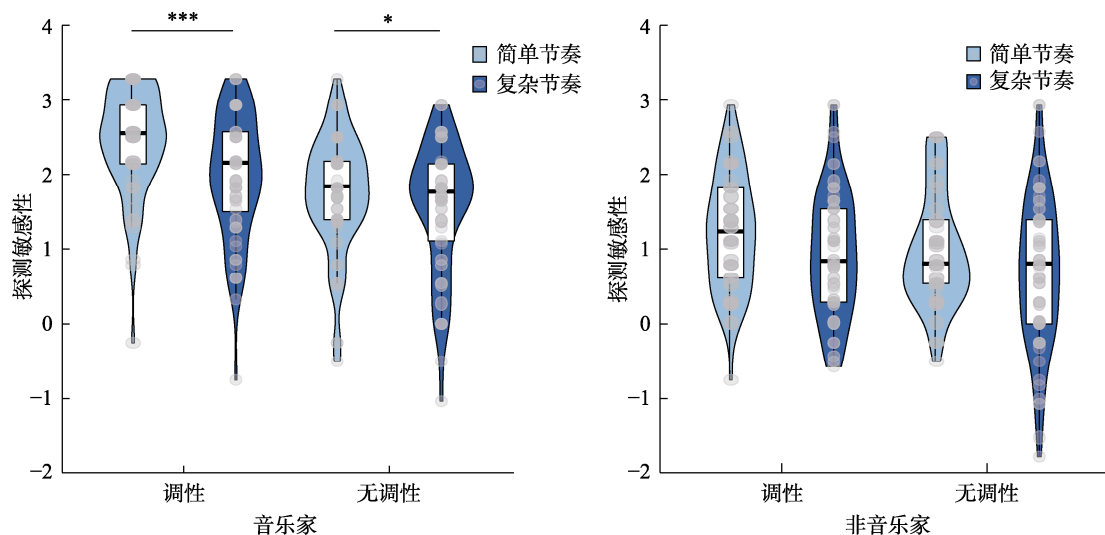


图2 实验1的探测敏感性(d')结果在不同组别、音高结构以及节奏结构条件下的表现

注：箱型图中间的水平线表示中位数，箱体长度表示四分位距，小提琴图展示了被试数据的分布情况，灰色点代表每位被试的个体数据。仅音乐家组表现出显著的交互作用，非音乐家组未表现出交互作用。 $*p < 0.05$, $***p < 0.001$

非音乐家，发现音高结构的主效应显著， $F(1, 35) = 13.67$, $p < 0.001$, $\eta_p^2 = 0.28$ ，对调性旋律的敏感性高于无调性旋律(1.06 vs. 0.82, 95% CI = [0.11, 0.37]); 节奏结构的主效应显著， $F(1, 35) = 7.62$, $p = 0.009$, $\eta_p^2 = 0.18$ ，对简单节奏序列的敏感性高于复杂节奏序列(1.09 vs. 0.80, 95% CI = [0.08, 0.51]); 但二者的交互作用不显著， $p = 0.893$ 。由此可见，在音高探测的敏感性指标上，只有音乐家表现出音高结构与节奏结构之间的显著交互作用，说明音乐家对调性与简单节奏的整合实现了结构增益。

为了进一步确认音乐训练经验的作用，我们计算了调性和无调性条件下节奏结构效应的差值，作为音高-节奏结构交互作用的分数，并将其与被试的 Gold-MSI 音乐素养指数评分进行相关分析。如图3所示，两者之间存在显著的正相关，Pearson's $r(70) = 0.28$, $p = 0.019$, Fisher's $z = 0.28$, 95% CI = [0.05, 0.48]。这一结果表明，音乐素养越高的被试，

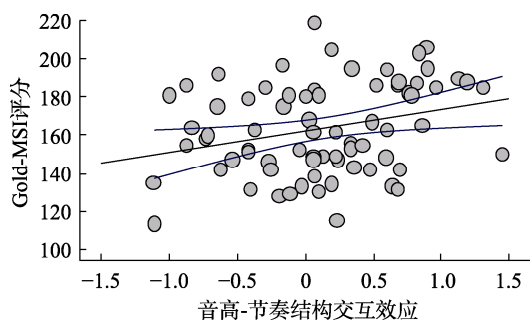


图3 实验1音高-节奏结构交互效应与音乐素养评分的相关

注：灰色点表示个体被试的数据，直线表示线性回归趋势线，两条曲线表示 95%置信区间。

其音高与节奏结构之间的交互效应越明显，反映了音乐训练与音高-节奏结构整合的关联。

此外，为了探讨音高结构、节奏结构和序列长度之间的交互作用，分别对五音和七音序列进行两因素 ANOVA。如图4所示，对于五音序列，发现音高结构的主效应显著， $F(1, 71) = 122.66$, $p < 0.001$, $\eta_p^2 = 0.63$ ，调性旋律的探测敏感性高于无调性旋律(1.79 vs. 1.27, 95% CI = [0.43, 0.62]); 且音高和节奏结构的交互作用显著， $F(1, 71) = 13.03$, $p < 0.001$, $\eta_p^2 = 0.15$ 。简单效应分析表明，调性条件下对简单节奏的敏感性高于复杂节奏，1.91 vs. 1.67, $t(71) = 3.35$, $p = 0.001$, Cohen's $d = 0.40$, 95% CI = [0.10, 0.39]，但在无调性条件下没有效应，1.22 vs. 1.31, $t(71) = -0.86$, $p = 0.392$ 。对于七音序列，发现音高结构的主效应显著， $F(1, 71) = 10.38$, $p = 0.002$, $\eta_p^2 = 0.13$ ，对调性旋律的敏感性高于无调性旋律(1.48 vs. 1.21, 95% CI = [0.10, 0.43]); 节奏结构的主效应显著， $F(1, 71) = 44.51$, $p < 0.001$, $\eta_p^2 = 0.39$ ，对简单节奏序列的敏感性高于复杂节奏序列(1.61 vs. 1.07, 95% CI = [0.38, 0.70]); 但二者的交互作用不显著， $F(1, 71) = 0.63$, $p = 0.429$ 。这一结果表明，音高与节奏结构的交互作用受到序列长度的调节，只有在较短序列条件下，对调性和简单节奏的整合可以促进音高探测敏感性。

2.2.2 反应偏差

对反应偏差的 ANOVA 结果表明，音高结构的主效应显著， $F(1, 70) = 80.09$, $p < 0.001$, $\eta_p^2 = 0.53$ ，表明调性旋律比无调性旋律诱发了更大的反应偏

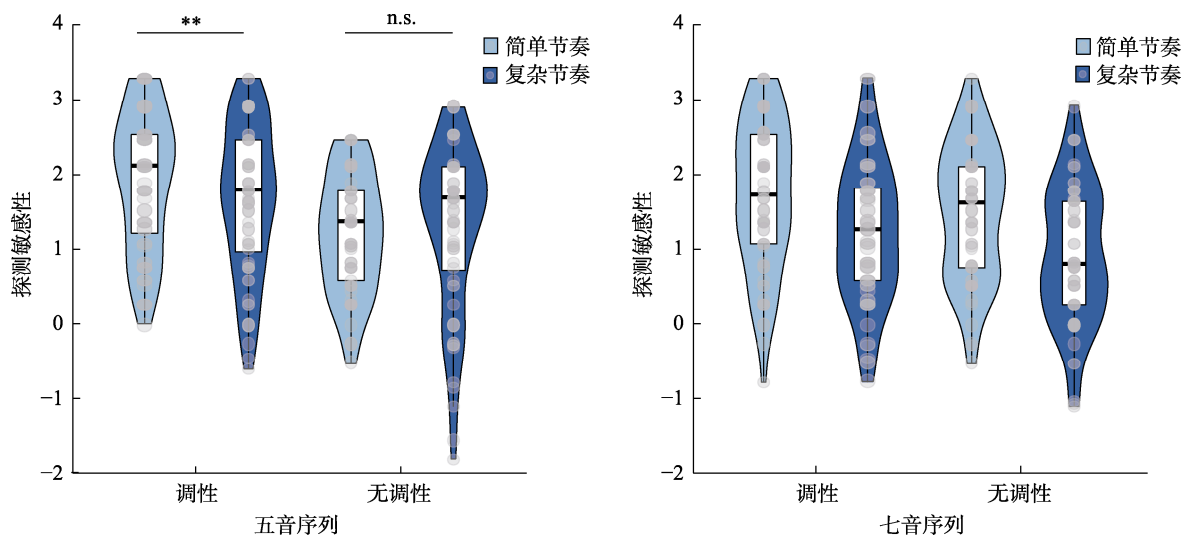


图4 实验1的探测敏感性(d')结果在不同序列长度、音高结构以及节奏结构条件下的表现

注: 箱型图中间的水平线表示中位数, 箱体长度表示四分位距, 小提琴图展示了被试数据的分布情况, 灰色点代表每位被试的个体数据。仅五音序列条件下表现出显著的交互作用, 七音序列条件下未发现交互作用。** $p < 0.01$

差(0.23 vs. 0.06, 95% CI = [0.13, 0.21])。节奏结构的主效应显著, $F(1, 70) = 122.49, p < 0.001, \eta_p^2 = 0.64$, 表明复杂节拍比简单节拍诱发了更大的反应偏差(0.29 vs. 0.004, 95% CI = [0.24, 0.34])。此外, 序列长度的主效应边缘显著, $F(1, 70) = 3.90, p = 0.052, \eta_p^2 = 0.05$, 表明七音序列倾向于比五音序列诱发更大的反应偏差(0.17 vs. 0.12)。同时, 序列长度与组别的交互作用显著, $F(1, 70) = 4.56, p = 0.036, \eta_p^2 = 0.06$, 简单效应分析表明音乐家对七音序列的反应偏差比五音序列更大(0.17 vs. 0.07, $p = 0.005$, 95% CI = [0.007, 0.20]), 但非音乐家没有这一效应(0.18 vs. 0.18, $p = 0.911$)。此外, 音高结构、节奏结构、序列长度三因素交互作用显著, $F(1, 70) = 5.12, p = 0.027, \eta_p^2 = 0.07$ 。其他主效应和交互作用不显著, $ps > 0.157$ 。以上结果表明, 调性旋律和复杂节奏诱发了更大的反应偏差, 且仅音乐家对较长序列表现出更大的反应偏差。

为进一步探讨音高结构、节奏结构和序列长度之间的交互作用, 分别对五音和七音序列进行分析。对于五音序列, 发现音高结构主效应显著, $F(1, 71) = 114.89, p < 0.001, \eta_p^2 = 0.62$, 调性比无调性诱发了更大的反应偏差(0.25 vs. -0.01, 95% CI = [0.19, 0.33]); 节奏结构主效应显著, $F(1, 71) = 75.85, p < 0.001, \eta_p^2 = 0.52$, 复杂节拍比简单节拍诱发了更大的反应偏差(0.26 vs. -0.01, 95% CI = [0.18, 0.37]); 二者交互作用不显著, $F(1, 71) = 2.88, p = 0.094$ 。对于七音序列, 发现音高结构的主效应显著, $F(1, 71) = 7.04, p = 0.010, \eta_p^2 = 0.09$, 调性比

无调性诱发了更大的反应偏差(0.22 vs. 0.13, 95% CI = [0.009, 0.16]); 节奏结构主效应显著, $F(1, 71) = 65.63, p < 0.001, \eta_p^2 = 0.48$, 复杂节拍比简单节拍诱发了更大的反应偏差(0.33 vs. 0.02, 95% CI = [0.21, 0.40]); 二者交互作用不显著, $F(1, 71) = 2.35, p = 0.130$ 。由此可见, 无论对于五音或七音序列, 音高结构和节奏结构均独立地影响对音高探测的反应偏差。

2.2.3 反应时

反应时的ANOVA结果表明, 组别主效应显著, $F(1, 70) = 15.50, p < 0.001, \eta_p^2 = 0.18$, 音乐家的反应速度比非音乐家快(430 ms vs. 583 ms, 95% CI = [-233.10, -72.79]); 音高结构的主效应边缘显著, $F(1, 70) = 3.99, p = 0.050, \eta_p^2 = 0.05$, 对调性旋律的反应速度倾向于比无调性旋律的更快(498 ms vs. 514 ms, 95% CI = [-32.30, -0.03]); 节奏结构的主效应不显著, $p = 0.068$ 。此外, 组别、节奏结构、序列长度三因素交互作用显著, $F(1, 70) = 4.47, p = 0.038, \eta_p^2 = 0.06$ 。为了进一步探讨这一交互作用, 分别分析了简单节奏和复杂节奏条件下的数据。对于简单节奏序列, 发现组别的主效应显著, $F(1, 70) = 14.78, p < 0.001, \eta_p^2 = 0.17$, 音乐家的反应比非音乐家快(425 ms vs. 572 ms, 95% CI = [-226.22, -67.86]), 且组别和序列长度的交互作用显著, $F(1, 70) = 4.14, p = 0.046, \eta_p^2 = 0.06$ 。简单效应分析表明, 音乐家对五音序列的反应比七音序列更快, 410 ms vs. 440 ms, $t(35) = -2.56, p = 0.015$, Cohen's $d = -0.43$, 95% CI = [-53.63, -6.43], 但非音乐家没有这一效应, 586 ms vs. 558 ms, $t(35) = 1.05, p =$

0.299。对于复杂节奏序列,发现组别的主效应显著, $F(1, 70) = 15.28, p < 0.001, \eta_p^2 = 0.18$, 音乐家的反应比非音乐家快(435 ms vs. 593 ms, 95% CI = [-243.62, -74.06]); 但是序列长度主效应以及组别和序列长度的交互作用不显著, $p_s > 0.771$ 。结果表明, 音乐家的反应速度显著快于非音乐家, 尤其在简单节奏和较短序列条件下, 这一优势更明显。

实验 1 的结果表明, 音高结构与节奏结构影响了非音乐家对音高保持的各个行为指标, 但二者之间不存在交互, 即音高结构不影响节奏结构的效应, 反之亦然, 表明非音乐家在音高工作记忆中相对独立地加工两种结构。对于音乐家来说, 音高结构与节奏结构影响他们对音高保持的表现, 并且两种结构的效应在探测敏感性指标上交互产生影响, 表明音乐家在音高工作记忆中整合了音高和节奏结构。同时, 这一交互效应与被试的音乐素养评分呈正相关, 表明音乐素养越高的个体在加工音高和节奏结构时交互效应越明显, 进一步验证了音乐训练在音高-节奏结构交互中的作用。此外, 相较于七音序列, 交互作用在五音序列条件下更明显, 表明音高-节奏结构交互可能受到任务难度的调节。

3 实验 2: 节奏保持

实验 2 通过节奏识别任务, 考察了音乐家和非音乐家对节奏的工作记忆。与实验 1 类似, 本实验操控了音高结构、节奏结构和序列长度。

3.1 方法

3.1.1 被试

与实验 1 相同。

3.1.2 刺激、程序和统计分析

刺激、程序和统计分析与实验 1 相同。唯一的区别是实验任务。本实验考察节奏保持, 要求被试在忽略音高信息的前提下, 快速且准确地判断连续呈现的两条旋律的节奏是否完全相同。为控制顺序效应, 实验顺序在两组被试中都进行了平衡。

3.2 结果和讨论

3.2.1 探测敏感性

探测敏感性的 ANOVA 结果表明, 组别主效应显著, $F(1, 70) = 41.99, p < 0.001, \eta_p^2 = 0.38$, 音乐家的探测敏感性高于非音乐家(2.08 vs. 1.18, 95% CI = [0.62, 1.17]); 节奏结构的主效应显著, $F(1, 70) = 180.77, p < 0.001, \eta_p^2 = 0.72$, 对简单节奏的探测敏感性高于复杂节奏(1.94 vs. 1.33, 95% CI = [0.51, 0.70]); 序列长度的主效应显著, $F(1, 70) = 31.85, p < 0.001, \eta_p^2 = 0.31$, 对五音序列的探测敏感性高于七音序列(1.79 vs. 1.48, 95% CI = [0.20, 0.42])。组别与音高结构之间的交互作用显著, $F(1, 70) = 6.98, p = 0.010, \eta_p^2 = 0.09$ 。简单效应分析表明, 音乐家在调性条件下的敏感性比无调性条件更高, 2.15 vs. 2.01, $t(35) = 2.18, p = 0.036$, Cohen's $d = 0.36$, 95% CI = [0.01, 0.28], 但非音乐家没有这一效应, 1.15 vs. 1.22, $t(35) = -1.50, p = 0.142$ (见图 5A)。同时, 音高结构和序列长度之间的交互作用显著, $F(1, 70) = 7.18, p = 0.009, \eta_p^2 = 0.09$, 表明音高结构在七音条件下有影响, 1.55 vs. 1.41, $t(71) = 2.29, p = 0.025$, Cohen's $d = 0.27$, 95% CI = [0.02, 0.27], 但它在五音条件下没有影响, 1.75 vs. 1.82, $t(71) = -1.33, p = 0.188$ 。此外, 组别、节奏结构和序列长度存在显著

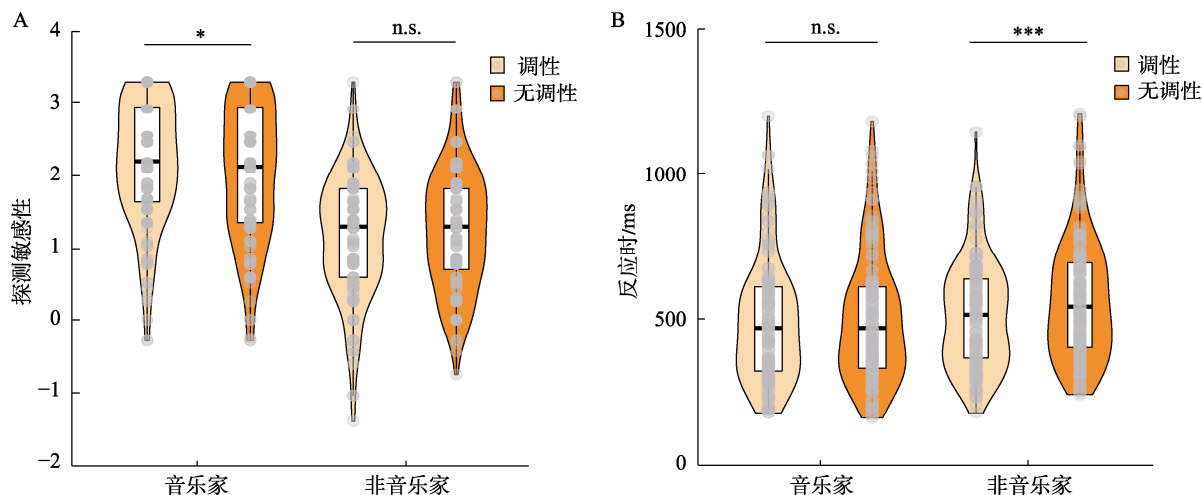


图 5 实验 2 的音乐家和非音乐家在调性和无调性条件下的(A)探测敏感性和(B)反应时

注: 箱型图中间的水平线表示中位数, 箱体长度表示四分位距, 小提琴图展示了被试数据的分布情况, 灰色点代表每位被试的个体数据。* $p < 0.05$, *** $p < 0.001$

的三因素交互作用, $F(1, 70) = 11.14, p = 0.001, \eta_p^2 = 0.14$ 。这些结果说明, 音乐家在节奏探测敏感性方面具有优势; 节奏结构和序列长度对节奏探测敏感性有显著影响, 而音高结构的影响仅限于音乐家或较长序列条件下。

为进一步了解组别、节奏结构和序列长度的交互作用, 分别对音乐家和非音乐家进行两因素ANOVA。对于音乐家, 发现节奏结构主效应显著, $F(1, 35) = 241.33, p < 0.001, \eta_p^2 = 0.87$, 表明对简单节奏的敏感性高于复杂节奏(2.44 vs. 1.71, 95% CI = [0.64, 0.83]); 序列长度的主效应显著, $F(1, 35) = 21.31, p < 0.001, \eta_p^2 = 0.38$, 表明对五音序列的敏感性高于七音序列(2.28 vs. 1.88, 95% CI = [0.22, 0.57])。同时, 节奏结构与序列长度之间的交互作用显著, $F(1, 35) = 6.37, p = 0.016, \eta_p^2 = 0.15$, 表明复杂节奏条件下对五音序列的敏感性比七音序列高, 2.01 vs. 1.42, $t(35) = 5.07, p < 0.001$, Cohen's $d = 0.85$, 95% CI = [0.35, 0.82], 但是简单节奏条件下没有这一效应, 2.54 vs. 2.34, $t(35) = 1.75, p = 0.089$ 。对于非音乐家, 发现节奏结构主效应显著, $F(1, 35) = 39.48, p < 0.001, \eta_p^2 = 0.53$, 表明对简单节奏的敏感性高于复杂节奏(1.43 vs. 0.94, 95% CI = [0.33, 0.64]); 序列长度的主效应显著, $F(1, 35) = 10.66, p < 0.001, \eta_p^2 = 0.23$, 表明对五音序列的敏感性高于七音序列(1.30 vs. 1.07, 95% CI = [0.09, 0.36])。同时, 节奏结构与序列长度之间的交互作用显著, $F(1, 35) = 4.77, p = 0.036, \eta_p^2 = 0.12$, 表明简单节奏条件下对五音序列的敏感性比七音序列高, 1.61 vs. 1.24, $t(35) = 4.04, p < 0.001$, Cohen's $d = 0.67$, 95% CI = [0.18, 0.55], 但是复杂节奏条件下没有这一效应, 0.98 vs. 0.90, $t(35) = 0.81, p = 0.423$ 。可见, 音乐家对节奏的探测敏感性仅在复杂节奏条件下受序列长度影响, 而非音乐家仅在简单节奏下受序列长度影响。

3.2.2 反应偏差

反应偏差的ANOVA结果表明, 组别主效应显著, $F(1, 70) = 27.34, p < 0.001, \eta_p^2 = 0.28$, 表明非音乐家比音乐家的反应偏差更大(0.54 vs. 0.26, 95% CI = [0.17, 0.39]), 更倾向于判断为“不同”。节奏结构的主效应显著, $F(1, 70) = 24.41, p < 0.001, \eta_p^2 = 0.26$, 表明复杂节奏比简单节奏诱发了更大的反应偏差(0.47 vs. 0.33, 95% CI = [0.08, 0.20])。此外, 序列长度的主效应显著, $F(1, 70) = 13.03, p < 0.001, \eta_p^2 = 0.16$, 表明七音序列比五音序列诱发了更大的

反应偏差(0.45 vs. 0.35, 95% CI = [0.04, 0.15])。其他主效应和交互作用不显著, $ps > 0.067$ 。这些结果表明, 非音乐家、复杂节拍和较长序列条件均诱发了更大的反应偏差。

3.2.3 反应时

反应时的ANOVA结果表明, 音高结构的主效应显著, $F(1, 70) = 15.50, p < 0.001, \eta_p^2 = 0.18$, 对调性旋律的反应速度比无调性旋律的更快(506 ms vs. 532 ms, 95% CI = [-42.37, -10.87]); 节奏结构的主效应显著, $F(1, 70) = 56.90, p < 0.001, \eta_p^2 = 0.45$, 对简单节奏的反应速度比复杂节奏的更快(485 ms vs. 553 ms, 95% CI = [-86.39, -50.47]); 且组别和音高结构的交互作用显著, $F(1, 70) = 6.90, p = 0.011, \eta_p^2 = 0.09$, 表明非音乐家对调性旋律的反应速度比无调性旋律的更快, 519 ms vs. 565 ms, $t(35) = -4.01, p < 0.001$, Cohen's $d = -0.67$, 95% CI = [-70.14, -22.96], 但音乐家对调性和无调性旋律的反应速度没有差异, 493 ms vs. 500 ms, $t(35) = -0.69, p = 0.498$ (见图5B)。音高结构和序列长度的交互作用显著, $F(1, 70) = 4.65, p = 0.034, \eta_p^2 = 0.06$, 表明相比于无调性条件, 调性结构促进了对七音序列的反应时, 495 ms vs. 536 ms, $t(71) = -3.58, p < 0.001$, Cohen's $d = 0.42$, 95% CI = [-65.12, -18.50], 但是对五音序列没有影响, 517 ms vs. 529 ms, $t(71) = -1.23, p = 0.221$ 。此外, 节奏结构和序列长度的交互作用显著, $F(1, 70) = 9.92, p = 0.002, \eta_p^2 = 0.12$, 表现为五音序列的节奏结构效应(478 ms vs. 568 ms, $t(71) = -7.96, p < 0.001$, Cohen's $d = -0.94$, 95% CI = [-112.48, -67.42])比七音序列(492 ms vs. 539 ms, $t(71) = -4.15, p < 0.001$, Cohen's $d = -0.49$, 95% CI = [-69.48, -24.36])更大。以上结果说明, 调性旋律和简单节奏能加快节奏判断的反应速度, 虽然无调性旋律会延迟非音乐家的反应时间。

实验2的主要结果表明, 无论是音乐家还是非音乐家, 在节奏保持任务中, 音高结构和节奏结构之间没有显著的交互作用。这一结果与实验1形成对比, 表明在节奏保持任务中, 音高和节奏结构的效应是相对独立的。并且, 与音乐家相比, 非音乐家在调性和无调性条件下没有表现出敏感性的差异。尽管如此, 调性仍然为节奏保持提供了一定程度的认知支持, 使非音乐家在反应速度上受益。此外, 序列长度影响音乐家对复杂节奏序列的敏感性, 他们对简单节奏序列的判断不受到序列长度影响; 而对于非音乐家, 序列长度仅影响他们对简单节奏

序列的敏感性。这一结果表明,序列长度的影响主要体现在中等难度的节奏再认任务。当任务过于简单(如音乐家对简单节奏的再认)或过于困难(如非音乐家对复杂节奏的再认)时,被试的表现更多受到已有音乐经验或加工能力的限制,而非刺激长度本身的影响。

4 综合讨论

音高和节奏的结构化组织及其认知加工是人类音乐和语言能力的基础。本研究旨在探讨音高和节奏结构加工在听觉工作记忆中的关系模式,以及任务要求和音乐训练如何影响这种加工模式。具体来说,本研究主要关注被试在调性条件下的表现是否比无调性条件更好(音高结构效应),以及在简单节拍条件下的表现是否比复杂节拍条件更好(节奏结构效应);同时考察这两种效应是相互影响(表现为交互作用)还是独立发挥作用。此外,我们重点关注二者加工关系的情境调节作用,具体分析任务之间的差异,以及音乐专业背景的影响。结果显示,在音高保持任务中,非音乐家表现出对音高和节奏结构相对独立的加工;而音乐家以交互的方式加工音高和节奏结构,并且敏感性指标的交互效应与被试的音乐素养得分正相关。然而,在节奏保持任务中,虽然两组被试都加工了音高和节奏结构,但这两种结构的效应均相对独立。这些发现表明,音乐训练选择性地调节了音高和节奏结构在听觉工作记忆中的互动。

在探讨音高和节奏结构的互动模式之前,首先可以比较这两类结构对工作记忆任务表现的总体影响。综合两个记忆任务的结果来看,节奏结构对两类记忆均产生显著影响,表现为简单的、稳定的节拍有利于提高探测敏感性,并减小反应偏差,即便任务要求被试忽略时间方面的信息。这一结果强调了节奏结构在音乐工作记忆中的重要作用。相较而言,音高结构没有为节奏工作记忆提供类似的认知支持,表现为调性和无调性条件下对节奏变化的探测敏感性没有差别。在听觉工作记忆中,节奏的规律性可能提供了一个清晰的时间框架,帮助听众预测音乐事件的发生时间,减少认知负荷,并有助于认知资源的有效分配。这种时间规律性有助于将音乐序列划分为易于加工的片段,保持听众的注意力并促进对音乐进程的追踪。因此,本研究的结果支持了动态注意理论的观点。根据该理论,注意本质上是基于时间维度的,大脑的内源性振荡器可能

与外部的节奏信号同步,有利于将时间性的注意定向到预测的时间点,从而增强预测加工(Jones, 1976; Jones & Boltz, 1989)。相较之下,虽然音高是影响旋律复杂性的主要因素(Prince & Pfondresher, 2012),但它缺乏节奏提供的时间线索,使其在组织音乐信息方面可能不如节奏有效。尽管音高变化可以突出乐句的边界(Zhang et al., 2016),但它并不像节奏模式那样天然形成清晰的片段(Yang et al., 2022),可能导致记忆过程中的结构化程度较低。此外,节奏常常涉及身体运动,如打拍子或跳舞,这会调动运动系统,通过跨模态整合增强记忆(Chen et al., 2008)。音高通常不会以同样的方式调动运动系统,因此提供的多感官增强记忆的机会较少。这些因素可能使得节奏(“when”)的规律性在听觉工作记忆中比音高(“what”)的规律性更具影响力。

本研究验证了非音乐家可以加工音高和节奏结构,但是,音高结构的效应并未受到节奏结构的调节,反之亦然,表明非音乐家以相对独立的方式加工两种结构。该发现与双通路神经结构的假设一致(Schwartz & Kotz, 2013),表明时间维度和非时间维度的信息加工可能依赖于不同的神经通路。Prince (2011)的研究也表明,当非音乐家专注于音高或时间信息,同时忽略另一个维度时,调性和节拍对旋律悦耳度评分的影响是累加的,二者之间没有交互作用。由于本研究的任务明确要求聚焦于单一维度,同时忽略或抑制另一维度,这种独立的结构加工可能受到被试加工策略的影响。非音乐家可能通过选择性注意有效地将认知资源集中于特定的信息维度,进而削弱了两种结构之间的整合加工,反映出加工策略在调控听觉信息处理方式中的作用。

与非音乐家不同,音乐家对音高和节奏结构的加工表现出交互作用,虽然这一交互仅体现在音高保持任务中。这说明,时间和非时间维度的交互作用受到任务要求和音乐训练经验的调节。根据动态注意理论(Jones, 1976; Jones & Boltz, 1989),音高与时间结构可以共同作用,形成一个联合的重音结构,从而优化注意资源的分配,并促进对非时间信息(如音高)的加工。我们的结果不仅支持了动态注意理论关于音高和时间交互的理论预测,还揭示了这种交互可能依赖于任务类型和个体的音乐经验,进一步丰富了该理论在音乐认知领域的应用。对交互作用进一步的简单效应分析显示,简单节奏会显著增强调性的效应,而调性的旋律则会进一步增强节奏结构的效应。这表明音乐训练可能提升了音乐

家的结构整合能力,使他们能高效地同时加工音高和节奏结构;即便工作记忆任务引导了对单一信息维度的关注,音乐家也能有效利用双重信息来优化任务表现。

尽管如此,音乐家在节奏保持任务中表现出对结构的独立加工。这可能是由于节奏保持任务本身较易完成,音乐家仅凭对结构的独立加工便能顺利完成。在旋律中,音高变化一般多于时值变化,增加了音高保持的认知负荷。因此,相较于认知负荷较低的节奏保持任务,音乐家在音高保持任务中对音高与节奏的结构整合更加明显,表明中等的任务难度促进了音乐家对两种结构的交互加工。然而,当认知负荷(序列长度)进一步增加时,所有被试,包括音乐家和非音乐家的数据,均显示出独立加工的趋势,以减轻认知负荷。这表明,音乐训练赋予了音乐家灵活的策略调节能力,使他们能够根据任务需求,灵活选择交互或独立加工的策略,提示音乐训练对复杂听觉信息加工的灵活性具有显著影响。相比之下,非音乐家在工作记忆任务中始终倾向于采用独立加工策略,可能是由于音高和节奏工作记忆任务对非音乐家均具有相当的挑战性。由此来看,交互加工可能是一种补偿策略,而独立加工或许是面对困难任务时的优先选择。

此外,本研究的发现可以在音高与时间维度的两阶段加工模型框架内得到解释。根据 Thompson 等人(2001)的理论,加工过程分为两个阶段:早期阶段对旋律特征(如音高和持续时间)进行独立编码,而后期阶段将这些特征整合为一个连贯的整体。脑电研究对这一假设提供了支持,尤其是在音乐句法加工方面(Sun et al., 2020; Zhang et al., 2019)。基于这一模型, Tillmann 和 Lebrun-Guillaud (2006)进一步强调了音高和时间相互作用的任务依赖性。他们指出,在需要局部判断的任务中,独立加工占主导,而在全局评估的任务中,整合加工更为突出。在我们的工作记忆任务中,被试需根据某一维度的变化作出“相同”或“不同”的判断,同时抑制另一维度的干扰。这种显性抑制的任务设计可能促使被试更多关注旋律的局部特征,从而加强了音高和时间维度的独立加工,而非交互加工的倾向。这一结果与 Tillmann 和 Lebrun-Guillaud (2006)的观点一致,即涉及局部判断的任务更可能引发特征的独立加工。此外,这些发现提示,音高结构与节奏结构的交互作用在本质上是动态的——这一交互作用受到任务设计、加工需求以及音乐维度在任务中被强调或

抑制程度的共同影响。

除了重点关注的结果,本研究还有一些次要发现。总体来看,音乐家在敏感性和反应时间方面优于非音乐家,并表现出较小的反应偏差。这与先前研究一致(例如, Chen et al., 2008; Schulze et al., 2011; Schulze et al., 2012; Sun et al., 2018),表明音乐家能够有效利用音高和节奏结构来提高听觉加工的表现。与音乐家相比,非音乐家在节奏任务的敏感性指标上未呈现出音高结构效应。这与 Schulze 等人(2011)的研究结果一致,该研究未观察到非音乐家在调性和无调性条件间的工作记忆行为差异。但是有其他研究(如, Albouy et al., 2013; Lévêque et al., 2022; Schulze et al., 2012)观察到非音乐家的行为敏感性受到调性结构的影响。这种不一致可能源于实验设计的差异:我们的研究同时结合了音高和节奏的变化,而之前的研究通常采用时值单一、没有节奏变化的序列,这种设计可能简化了任务,使被试能够专注于音高结构。

5 总结

综上所述,本研究主要探讨了音高和节奏结构在听觉工作记忆中的加工关系,以及这种关系如何受到任务和音乐训练的影响。结果显示,非音乐家在工作记忆任务中始终表现出对音高和节奏结构的独立加工模式。节奏规律性作为时间框架帮助非音乐家提升了任务表现,而音高结构未能发挥同样的作用,提示了时间性规律在听觉工作记忆中可能比非时间性规律更易被认知系统利用。相比之下,音乐家能够灵活调节音高与节奏结构的交互或独立加工方式,反映出音乐训练赋予了他们在不同任务需求下调整加工策略的能力。未来研究可进一步探讨音高与节奏结构在不同任务要求和认知负荷下的相互作用,并深入考察不同类型的音乐训练在其中的作用。

参 考 文 献

- Albouy, P., Schulze, K., Caclin, A., & Tillmann, B. (2013). Does tonality boost short-term memory in congenital amusia? *Brain Research*, 1537, 224–232. <https://doi.org/10.1016/j.brainres.2013.09.003>
- Bharucha, J., & Krumhansl, C. L. (1983). The representation of harmonic structure in music: Hierarchies of stability as a function of context. *Cognition*, 13(1), 63–102. [https://doi.org/10.1016/0010-0277\(83\)90003-3](https://doi.org/10.1016/0010-0277(83)90003-3)
- Chen, J. L., Penhune, V. B., & Zatorre, R. J. (2008). Moving on time: Brain network for auditory-motor synchronization is modulated by rhythm complexity and musical training.

- Journal of Cognitive Neuroscience*, 20(2), 226–239. <https://doi.org/10.1162/jocn.2008.20018>
- Cowan, N. (2000). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioural and Brain Sciences*, 24(1), 87–185. <https://doi.org/10.1017/s0140525x01003922>
- Dowling, W. J. (1991). Tonal strength and melody recognition after long and short delays. *Perception & Psychophysics*, 50(4), 305–313. <https://doi.org/10.3758/bf03212222>
- Essens, P. J., & Povel, D. J. (1985). Metrical and nonmetrical representations of temporal patterns. *Perception & Psychophysics*, 37(1), 1–7. <https://doi.org/10.3758/BF03207132>
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41(4), 1149–1160. <https://doi.org/10.3758/BRM.41.4.1149>
- Fitch, W. T. (2013). Rhythmic cognition in humans and animals: Distinguishing meter and pulse perception. *Frontiers in Systems Neuroscience*, 7, Article 68. <https://doi.org/10.3389/fnsys.2013.00068>
- Jentschke, S., & Koelsch, S. (2009). Musical training modulates the development of syntax processing in children. *NeuroImage*, 47(2), 735–744. <https://doi.org/10.1016/j.neuroimage.2009.04.090>
- Jerde, T. A., Childs, S. K., Handy, S. T., Nagode, J. C., & Pardo, J. V. (2011). Dissociable systems of working memory for rhythm and melody. *NeuroImage*, 57(4), 1572–1579. <https://doi.org/10.1016/j.neuroimage.2011.05.061>
- Jones, M. R., & Boltz, M. (1989). Dynamic attending and responses to time. *Psychological Review*, 96(3), 459–491. <https://doi.org/10.1037/0033-295X.96.3.459>
- Jones, M. R. (1976). Time, our lost dimension: Toward a new theory of perception, attention, and memory. *Psychological Review*, 83(5), 323–355. <https://doi.org/10.1037/0033-295X.83.5.323>
- Koelsch, S., Gunter, T. C., Friederici, A. D., & Schröger, E. (2000). Brain indices of music processing: “Nonmusicians” are musical. *Journal of Cognitive Neuroscience*, 12(3), 520–541. <https://doi.org/10.1162/089892900562183>
- Koelsch, S., Jentschke, S., Sammler, D., & Mietschen, D. (2007). Untangling syntactic and sensory processing: An ERP study of music perception. *Psychophysiology*, 44(3), 476–490. <https://doi.org/10.1111/j.1469-8986.2007.00517.x>
- Koelsch, S., Schmidt, B. H., & Kansok, J. (2002). Effects of musical expertise on the early right anterior negativity: An event-related brain potential study. *Psychophysiology*, 39(5), 657–663. <https://doi.org/10.1111/1469-8986.3950657>
- Krumhansl, C. L. (1990). *Cognitive foundations of musical pitch*. New York: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195148367.001.0001>
- Krumhansl, C. L. (2000). Rhythm and pitch in music cognition. *Psychological Bulletin*, 126(1), 159–179. <https://doi.org/10.1037/0033-2909.126.1.159>
- Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review*, 106(1), 119–159. <https://doi.org/10.1037/0033-295X.106.1.119>
- Lartillot, O., Toiviainen, P., & Eerola, T. (2008). A matlab toolbox for music information retrieval. In C. Preisach, H. Burkhardt, L. Schmidt-Thieme, R. Decker (Eds.), *Data analysis, machine learning and applications* (pp. 261–268). Springer-Verlag.
- Lévêque, Y., Lalitte, P., Fornoni, L., Pralus, A., Albouy, P., Bouchet, P., ... Tillmann, B. (2022). Tonal structures benefit short-term memory for real music: Evidence from non-musicians and individuals with congenital amusia. *Brain and Cognition*, 161, Article 105881. <https://doi.org/10.1016/j.bandc.2022.105881>
- Lin, H. R., Kopiez, R., Müllensiefen, D., & Wolf, A. (2021). The Chinese version of the Gold-MSI: Adaptation and validation of an inventory for the measurement of musical sophistication in a Taiwanese sample. *Musicae Scientiae*, 25(2), 226–251. <https://doi.org/10.1177/1029864919871987>
- Macmillan, N. A., & Creelman, C. D. (2004). *Detection theory: A user's guide* (2nd ed.). London: Lawrence Erlbaum Associates.
- Martin, T., Egly, R., Houck, J. M., Bish, J. P., Barrera, B. D., Lee, D. C., & Tesche, C. D. (2005). Chronometric evidence for entrained attention. *Perception & Psychophysics*, 67(1), 168–184. <https://doi.org/10.3758/BF03195020>
- Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The musicality of non-musicians: An index for assessing musical sophistication in the general population. *PLOS ONE*, 9(2), Article e89642. <https://doi.org/10.1371/journal.pone.0089642>
- Povel, D.-J., & Essens, P. (1985). Perception of temporal patterns. *Music Perception*, 2(4), 411–440. <https://doi.org/10.2307/40285311>
- Prince, J. B. (2011). The integration of stimulus dimensions in the perception of music. *Quarterly Journal of Experimental Psychology*, 64(11), 2125–2152. <https://doi.org/10.1080/17470218.2011.573080>
- Prince, J. B., & Pfordresher, P. Q. (2012). The role of pitch and temporal diversity in the perception and production of musical sequences. *Acta Psychologica*, 141(2), 184–198. <https://doi.org/10.1016/j.actpsy.2012.07.013>
- Prince, J. B., Schmuckler, M. A., & Thompson, W. F. (2009). The effect of task and pitch structure on pitch-time interactions in music. *Memory & Cognition*, 37(3), 368–381. <https://doi.org/10.3758/MC.37.3.368>
- Prince, J. B., Thompson, W. F., & Schmuckler, M. A. (2009). Pitch and time, tonality and meter: How do musical dimensions combine? *Journal of Experimental Psychology: Human Perception and Performance*, 35(5), 1598–1617. <https://doi.org/10.1037/a0016456>
- Raven, J. (2000). The Raven's progressive matrices: Change and stability over culture and time. *Cognitive Psychology*, 41(1), 1–48. <https://doi.org/10.1006/cogp.1999.0735>
- Sakai, K., Hikosaka, O., Miyauchi, S., Takino, R., Tamada, T., Iwata, N. K., & Nielsen, M. (1999). Neural representation of a rhythm depends on its interval ratio. *Journal of Neuroscience*, 19(22), 10074–10081. <https://doi.org/10.1523/JNEUROSCI.19-22-10074.1999>
- Schulze, K., Dowling, W. J., & Tillmann, B. (2012). Working memory for tonal and atonal sequences during a forward and a backward recognition task. *Music Perception*, 29(3), 255–267. <https://doi.org/10.1525/mp.2012.29.3.255>
- Schulze, K., Müller, K., & Koelsch, S. (2011). Neural correlates of strategy use during auditory working memory in musicians and non-musicians. *European Journal of Neuroscience*, 33(1), 189–196. <https://doi.org/10.1111/j.1460-9568.2010.07470.x>
- Schwartz, M., & Kotz, S. A. (2013). A dual-pathway neural architecture for specific temporal prediction. *Neuroscience & Biobehavioral Reviews*, 37(10), 2587–2596. <https://doi.org/10.1016/j.neubiorev.2013.08.005>

- Sun, L., Liu, F., Zhou, L., & Jiang, C. (2018). Musical training modulates the early but not the late stage of rhythmic syntactic processing. *Psychophysiology*, 55(2), Article e12983. <https://doi.org/10.1111/psyp.12983>
- Sun, L., Thompson, W. F., Liu, F., Zhou, L., & Jiang, C. (2020). The human brain processes hierarchical structures of meter and harmony differently: Evidence from musicians and nonmusicians. *Psychophysiology*, 57(9) Article e13598. <https://doi.org/10.1111/psyp.13598>
- Tillmann, B., & Lebrun-Guillaud, G. (2006). Influence of tonal and temporal expectations on chord processing and on completion judgments of chord sequences. *Psychological Research*, 70(5), 345–358. <https://doi.org/10.1007/s00426-005-0222-0>
- Thompson, W. F., Hall, M. D., & Pressing, J. (2001). Illusory conjunctions of pitch and duration in unfamiliar tone sequences. *Journal of Experimental Psychology: Human Perception and Performance*, 27(1), 128–140. <https://doi.org/10.1037//0096-1523.27.1.128>
- Yang, X., Shen, X., Zhang, Q., Wang, C., Zhou, L., & Chen, Y. (2022). Music training is associated with better clause segmentation during spoken language processing. *Psychonomic Bulletin & Review*, 29(4), 1472–1479. <https://doi.org/10.3758/s13423-022-02076-2>
- Zhang, J., Che, X., & Yang, Y. (2019). Event-related brain potentials suggest a late interaction of pitch and time in music perception. *Neuropsychologia*, 132, 107118. <https://doi.org/10.1016/j.neuropsychologia.2019.107118>
- Zhang, J., Jiang, C., Zhou, L., & Yang, Y. (2016). Perception of hierarchical boundaries in music and its modulation by expertise. *Neuropsychologia*, 91, 490–498. <https://doi.org/10.1016/j.neuropsychologia.2016.09.013>
- Zhou, L., Zhao, H., & Jiang, C. (2017). Neural plasticity to musical performance training: A meta-analysis study. *Advances in Psychological Science*, 25(11), 1877–1887. <https://doi.org/10.3724/SP.J.1042.2017.01877>
- [周临舒, 赵怀阳, 蒋存梅. (2017). 音乐表演训练对神经可塑性的影响: 元分析研究. *心理科学进展*, 25(11), 1877–1887.]

Musical training enhances the interaction between pitch and time dimensions in auditory working memory

ZHOU Linshu¹, ZHANG Yuqing¹, CAI Dan-Chao²

(¹ Music College, Shanghai Normal University, Shanghai 200234, China)

(² Shanghai Public Health Clinical Center, Fudan University, Shanghai 201508, China)

Abstract

Musical training enhances sensitivity to both temporal (e.g., rhythm) and non-temporal (e.g., pitch) structures during auditory processing. However, little is known about the joint processing of these structures in auditory working memory and the influence of musical expertise. This study explored whether pitch and rhythmic structures are processed interactively or independently in auditory working memory, and how musical expertise modulates this relationship.

Two experiments were conducted with 36 musicians and 36 nonmusicians. Experiment 1 involved a pitch maintenance task, where participants judged whether two pitch sequences were the same or different while ignoring rhythmic variations, which could act as background interference. The sequences varied in pitch structure (tonal vs. atonal), rhythmic structure (simple vs. complex), and length (five-note vs. seven-note). Experiment 2 examined rhythm maintenance, where participants judged whether two rhythmic sequences were the same or different while ignoring pitch information. While the task was conceptually the reverse of Experiment 1—requiring attention to rhythm rather than pitch—the design similarly manipulated pitch structure, rhythmic structure, and sequence length to compare how participants processed rhythmic information when suppressing irrelevant pitch content.

The results of Experiment 1 indicated that nonmusicians tended to process pitch and rhythmic structures independently during the pitch maintenance task. In contrast, musicians exhibited a significant interaction between pitch and rhythm, demonstrating their ability to integrate these structures for more effective processing. A correlation analysis further confirmed the role of musical training in this integration; sensitivity to the interaction of pitch and rhythm was positively correlated with participants' musical sophistication scores. This suggests that individuals with higher musical expertise exhibit more pronounced integration of pitch and rhythmic information, reflecting enhanced multidimensional auditory processing abilities in musicians. In Experiment 2, however, both musicians and nonmusicians processed pitch and rhythm independently during the rhythm maintenance task, with no significant interaction. This contrasts with the findings of Experiment 1,

indicating that pitch and rhythmic structures are processed separately during rhythm maintenance. Furthermore, nonmusicians showed no difference in sensitivity between tonal and atonal conditions in the rhythm maintenance task, although they responded faster in the tonal condition than in the atonal condition.

These results support dynamic attending theory, which posits that attentional resources align with temporal events, thus facilitating the processing of non-temporal information, such as pitch, when in synchrony with rhythmic structure. While nonmusicians relied more on independent processing strategies across both tasks, musicians showed enhanced cognitive flexibility, modulating their processing strategies in response to task complexity. This suggests that musical training not only improves auditory sensitivity but also allows individuals to adaptively switch between processing strategies based on task demands. Future research may further explore how different types of musical training affect the interaction between temporal and non-temporal auditory information under varying cognitive loads.

Keywords musical training, auditory working memory, temporal regularity, musical structure, dynamic attending theory