科学通报 2016 年 第 61 卷 第 10 期: 1099 ~ 1111
评 述
《中国科学》杂志社
SCIENCE CHINA PRESS

音乐句法的加工

马谐[①②③], 杨玉芳[①*], 张秋月[②③]

① 中国科学院心理研究所, 北京 100101;
② 云南师范大学教育科学与管理学院, 昆明 650500;
③ 云南师范大学民族教育信息化教育部重点实验室, 昆明 650500
* 联系人, E-mail: yangyf@psych.ac.cn

2015-12-02 收稿, 2016-01-12 修回, 2016-01-13 接受, 2016-02-22 网络版发表
国家自然科学基金(31470985)资助

摘要 音乐句法是指将离散的音符元素结合成层级结构的原则. 现有研究对句法加工机制产生了较多争议. "认知主义"提出整合句法需要借助基于长时记忆的图式驱动, "物理主义"则认为加工句法更多只需借助基于短时记忆的知觉驱动. 两者从理论模型、行为研究和认知神经层面论证了各自的观点. 未来研究有必要进一步探讨大尺度层级句法结构的加工能力及加工机制、分离知觉驱动与图式驱动的机制效应、增强脑电和脑成像技术的应用以及关注句法结构与情绪的关系. 对音乐句法加工的探索有助于揭示人类高级认知智慧的特性, 同时, 具有重要的临床意义.

关键词 音乐, 句法, 和声, 图式驱动, 知觉驱动

音乐句法是指将离散的音符元素结合成层级结构的原则. 建立在调性和声基础上的音高组织规则是最重要的音乐句法, 制约着听众对音乐事件的期待、建构与反馈, 并在此基础上促成音乐美学体验的产生[1,2]. Lerdahl和Jackendoff[3]在调性音乐的生成理论(generative theory of tonal music, GTTM)中, 提出了与语言句法类似的音乐结构树状模型. Rohrmeier[4]在句法生成模型(generative syntax model, GSM)中, 进一步细化了调性句法的层级生成法则.

现有研究通过不同的实验范式和不同的测量技术探讨了听众对音乐句法的加工能力及加工机制[5,6]. 证实了西方听众普遍存在着对线性句法结构感知的能力, 然而, 却对句法加工机制产生了较多分歧, "认知主义"提出音乐句法表征如同语言句法表征, 反映着声音与意义间的抽象认知结构关系, 整合句法需要借助基于长时记忆的图式驱动; "物理主义"则认为调性句法关系具有根植于音频的心理声学属性, 源于短时记忆作用下的知觉驱动机制就能做出解释[7,8]. 多年来, 两者分别从理论模型、行为研究和认知神经层面论证了各自的观点, 这些结论和方法将引导着未来研究的走向. 本文在简要论述音乐句法理论的基础上, 总结探讨句法加工能力的相关研究, 比较和论述有关音乐句法加工机制的争议. 最后, 提出对未来研究的思考与展望.

1 音乐句法的基本概念及特征

音乐是具有复杂音高结构和时间结构的声学事件, 建立在调性和声基础上的音高组织规则是西方调性音乐叙事方式中最重要的内在逻辑, 也因此成为了狭义音乐句法概念的代表[5,7,9]. 和声涉及: (ⅰ) 和弦(chord), 即由多个音按照一定的方法构成, 组织起乐曲中音高的纵向结构; (ⅱ) 和声进行(harmonic progression), 即各个和弦的连接, 组织起乐曲发展的横向结构. 句法的重要特性是, 每一个和弦或单音

引用格式: 马谐, 杨玉芳, 张秋月. 音乐句法的加工. 科学通报, 2016, 61: 1099–1111
Ma X, Yang Y F, Zhang Q Y. Music syntactic processing (in Chinese). Chin Sci Bull, 2016, 61: 1099–1111, doi: 10.1360/N972015-01340

© 2016《中国科学》杂志社

在一定的音乐背景下, 都有其特定的句法功能, 功能之间在重要程度和稳定程度上具有层级性. 例如, 给定某个调性创造出某个背景, 在这种特定的范畴下, 将有7个和弦成为构建和声关系发展的主要和弦, 被称为"调内和弦". 而在这7个和弦中, 又有不同的层级结构, 主和弦(tonic)位于最高层级, 成为乐曲发展的中心点. 属和弦(dominant)和下属和弦(subdominant)位于第2层级, 与主和弦共同成为支撑乐曲发展的主干. 其中, 又以主和弦-属和弦-主和弦的结构成为最核心及最基础的发展格式. 剩余的4个和弦位于第3层级, 在乐曲发展中处于较不重要的地位, 承担配合、修饰和扶持主干和弦的功能. 除此7个"调内和弦"之外的"调外和弦", 与特定调性背景下的主和弦成为了远关系和弦[10].

Heinrich Schenker是第一个从理论上系统性地提出音乐句法概念的人, 他指出和声的行进包括3个基本原则: (ⅰ) 层级性; (ⅱ) 省略性(有一些不重要的部分可以进行省略); (ⅲ) 循环性(在基本的格式基础上将可以嵌套出多个循环的句法单元)[5]. Schenker的学说被Lerdahl和Jackendoff[3]正式发展成为著名的调性音乐生成理论, 并进一步提出了与语言句法类似的音乐结构树状模型. 其中, 用时程还原(time-span reduction)概念指示相关事件之间的结构重要性, 从而区分出重要事件与装饰事件的层级关系; 用延长还原(prolongation reduction)概念指示某一事件相对于其他相关事件而产生的紧张或放松关系, 局部的紧张或放松运动往往嵌入在更高层级的运动之中[11~13]. GTTM模型在调性音高空间(tonal pitch space theory, TPS)理论中得到发展, 使之从纯理论分析模型实现了可计算化. 该理论认为音高是音乐感知的核心. 以和弦为例, 任意的两个和弦之间存在一定的心理距离, 这个距离可以计算. 和声距离能直接影响音乐序列中此消彼长的音高感知, 并进而产生紧张-放松的体验, 感受到音乐表达的情感意义[12,14]. Rohrmeier[4]在GTTM模型基础上, 进一步细化和阐述了调性句法的生成法则, 提出了句法生成模型. Rohrmeier根据句法抽象水平的不同, 界定了由低至高4个加工水平的句法结构. 最底一层是表面结构水平(surface-structure level), 如C-G-C; 上面一级是音程结构水平(scale-degree structure level), 如I-V-I; 再上面一层是功能结构水平(functional-structural level), 如t-d-t; 顶部最高一层是段落结构水平(phrase-structure level), 如t区域-d区域-t区域(tonic regions-dominant regions-tonic regions). 下面每一个层级都可以视为段落句法的生成规则, 生成开始于最高一级(段落句法), 在派生过程中通过次级层传输信息, 不断地产生出循环句法或者更复杂的句法序列, 层层递进向下衍生. 其在派生过程中遵循3个原则: (ⅰ) 背景依存性, 音乐事件不是孤立的, 总是由一系列相互影响的事件组成的. 每一个元素(事件)都与前边的事件、后边的事件有着不同的结构功能关系, 最简单的表征形式就是句法树, 但元素与元素之间的位置并非完全是临近的; (ⅱ) 可省略性, 句法树暗示着有一些元素即使遗漏, 仍然不影响整个音乐序列的正确性; (ⅲ) 和弦功能性, 和弦的功能取决于它在句法树中所处的位置[4,9].

音乐调性生成句法树概念主要借鉴了Chomsky[15]的语言句法树概念, 两者可以进行类比的共同点主要在于规则性和层级性. 音乐和语言的句法都包含特定的语法规则, 语言中有主语、谓语、宾语等语法功能, 西方调式音乐中有主和弦、属和弦、下属和弦等和声功能. 此外, 音乐和语言句法都是以层级关系来进行组织, 最底层的句法序列被称为线性序列句法, 语法功能元素之间具有临近依存关系. 位于高层级的句法序列被称为层级序列句法, 语法功能元素之间具有远距离依存关系, 不同单元的句法序列之间层层嵌套和循环生成, 这些共性成为人们比较音乐句法加工机制与语言句法加工机制的基础[11,16,17].

2 音乐句法加工的认知机制

GTTM模型和GSM从理论角度提出了音乐句法与语言句法的类比关系. 但明显不同于语言加工的是, 听众对音乐句法的加工表现出了"跨经验性"和"跨年龄性". 无论是在早期的行为学研究中[18~22], 还是在近期的事件相关电位(event related potentia, ERP)研究中[23~28], 大多数证据支持至少在简单的线性句法上, 西方听众普遍存在着句法图式, 这种图式指导着人们形成音乐期待, 且与被试的音乐经验[24,28]和被试的年龄[29,30]并无太大的关联. 而有研究进一步发现, 普通听众不仅能够加工简单的线性句法结构, 还能够感知及整合带有嵌套关系的复杂层级句法结构[6], 虽然研究结果还存在争议[31~37]. 但已无疑说明, 听众在无需外显学习的情况下, 就能够获得较好的音乐句法加工能力, 表现出与语言句法

学习的显著区别.

音乐句法能力较为容易就可习得, 引发了研究者对音乐句法加工机制的思考. 音乐句法在多大程度上反映了声音间的抽象认知关系, 而非心理声学的关系? 主导听众形成音乐句法期待的认知加工机制是什么? 它更多是类似于加工语言的, 需要借助于抽象认知机制的自上而下的图式驱动, 或仅此只是借助于自下而上的知觉驱动就能完成? 多年来, 形成了"认知主义"和"物理主义"的争议. 认知主义的观点提出: 调性句法反映着声音与意义间的抽象认知结构关系, 人们在整合调性句法过程中, 需要启动基于文化图式的长时记忆知识. 该观点代表性理论包括有"调性和声扩散激活模型"(music and activation, MUSACT)[8,38]和"句法整合资源共享假设"(shared syntactic integration resource hypothesis, SSIRH)[39~42]. 物理主义的观点则认为: 知觉驱动机制就能能解释句法加工过程. 因为音乐的句法关系具有根植于声音的心理声学属性, 例如, 泛音关系和音调之间某些频率间隔的平滑性或粗糙性都有可能引起句法期待, 调性句法缺乏语言句法的抽象性, 元素之间的心理关系反映了声音的物理属性而非纯粹的常规结构关系. 该观点的代表理论包括周期性音高模型(the periodicity pitch model, PP), 以及在它基础上发展而成的听觉短时记忆模型(auditory short-term memory, ASTM)[7,43].

2.1 认知主义模型及实验例证

(ⅰ)认知主义观点的基本模型. Bharucha[38]将语言研究中解释语义启动效应的激活扩散模型引入音乐期待研究, 从而形成了解释音乐调性启动的"调性和声扩散激活模型". 该模型认为调性层级规则, 通过人们长期对西方调性音乐的被动阐释, 已经变得内在化. 根据Bharueha的观点, 这些内化的知识可能是以一个由单元(units)内部连接的网络起作用. 这个网络包含3个单元(音、和弦、调). 每个单元与3个大和弦和3个小和弦相连接. 当一个单元被激活, 它就会激活与其相关的单元(和弦), 再扩散到下一个单元(调), 并同时反射回来. 接下来由调自上而下地将激活传回和弦, 以此往复直到达到平衡. Patel等人[39~42]进一步提出了音乐和语言句法加工比较的理论模型——"句法整合资源共享假设", 该理论认为, 虽然音乐和语言在句法表征上具有很大差异, 但二者对句法结构进行整合的认知机制是相同的. 他假设提供句法整合资源的大脑网络是"资源网络", 音乐和语言激活表征结构并进行整合的脑区和认知资源存在重叠, 加工音乐句法如同加工语言句法一样, 是高度抽象化的认知过程.

(ⅱ)认知主义观点的实验支撑. Bharucha等人[18~21]在早期的单音启动范式中, 操控目标和弦与启动和弦分别存在共享调性的近距离和声关联和无共享调性的远距离和声关联. 实验揭示, 当目标和弦与启动和弦具有近距离和声关联时, 能够显著易化被试的任务操作, 说明启动刺激激活了被试的句法图式, 形成了句法期待. 但是, 由于西方调性音乐自身的声学特殊性, 在这个实验操控中所用的共同调性关系和弦比起无共同调性关系和弦具有更多的共同音, 既可能是通过共同声学成分关系, 也可能是通过调性关系来实现启动效应. 因为一个与之前出现过的和弦有共同组成音或泛音的和弦会比一个与之前出现的和弦无共享成分音或泛音的和弦得到更多基于知觉驱动的重复启动效应[44]. 换言之, Bharucha和Stoeckig[19]实验中在目标刺激与启动刺激具有共同调性关系情况下出现的易化加工现象, 无法确定是基于重复启动效应的知觉驱动所致, 还是基于长时记忆知识的图式驱动所致. 为此, Tekman和Bharucha[21]的研究通过操控启动时间(stimulus onset asynchrony, SOA)和操控目标和弦的方法, 试图实现知觉驱动和图式驱动的分离. 一种目标刺激与启动刺激共享音符, 但不共享调性, 另外一种目标刺激与启动刺激没有共享音符, 却共享调性, 共享音符的易化效应只发生在较短的SOA上(50 ms), 共享调性关系的易化效应则发生在较长的SOA上(等于或长于500 ms上). 研究者认为, 在较长时间范围下, 图式启动效应都优越于重复启动效应.

有力支持认知主义观点的实验证据来自于人们对音(和弦)的感知具有较强的背景依存性, 也就是人们对音(和弦)的感知需要借助于语境和结构关系, 而不是依据其内在的物理特征. Bigand和Pineau[22]的研究中创造了2组包括8个和弦的序列, 在两组和弦序列中, 最后两个音是完全等同的(物理属性完全一致), 而前边的6个和弦分别创造出两种调性背景, 当最后两个目标和弦与前边调性背景形成比较符合句法的和声走向时, 实验操作能够显著得到易化. 研究者认为, 被试对相同物理属性的音符在不同的背景

中产生了不同的感受, 说明不同的背景激活了被试不同的调性图式, 引导被试形成了不同的期待, 基于“图式驱动”的抽象认知功能在音乐句法加工中发挥着重要作用, 此研究和弦序列得到的结果在较多相关实验中得到了重复[45,46]. Bigand等人[45]在Bigand和Pineau[22]的实验基础上, 通过操控目标和弦在启动背景中出现的频率, 以直接比较重复启动与图式启动的作用. 研究结尾分别采用符合句法的主和弦和不符合句法的下属和弦, 同时, 下属和弦与前边启动背景具有更多的共享音符成分. 研究进一步操控下属和弦是否在启动背景中出现, 以增加在下属和弦结尾情况下的重复启动效应. 实验结果显示, 即使是在非常快的速度下, 基于图式产生的启动都优于基于重复启动产生的效应.

句法匹配范式与ERP测量技术的结合, 是近些年研究句法表征的主要方法, 在Koelsch等人[23,24]和Leino等人[25]的研究中, 通过在不同位置操控那波里和弦(Neapolitan), 并通过ERP技术的使用, 对认知主义的观点做出了进一步的支持. Neapolitan和弦包括两个调外音, 容易带来听觉失衡(sensory dissonance). 从句法和声关系上来说, 属于远距离和弦. 但这个和弦在某种特定的音乐风格下, 可以在下属和弦出现的位置代替其使用, 具有一定的句法合理性. Koelsch等人和Leino等人研究发现, Neapolitan和弦出现在句法违反的位置时, 比起出现在句法正确的位置时, 能够显著诱发两种重要负波成分, 分别是ERAN(the early right anterior negativity)成分和N5成分, ERAN成分被解释为和声句法违反的早期指示性成分, 一般出现在大约150~200 ms窗口[23,27], 而N5被解释为和声句法违反的晚期指示性成分, 通常与和声整合困难相关, 一般出现在大约500~800 ms窗口[23,27,28]. 此研究结果为音乐句法加工激活抽象认知机制做出了较好的佐证. 同时, 音乐专业者比非音乐专业者, 在错误使用Neapolitan和弦的位置上, 诱发了更大的ERAN, 说明在音乐句法加工过程中, 音乐专业者更受制于图式激活的期待作用, 当加工的和弦与背景违反时, 诱发了更强烈的不匹配信号. 但操控Neapolitan和弦的研究范式面临的问题是, 它本身会带来听觉失衡, 其所产生的效应容易混淆知觉机制和抽象认知机制在其中的作用. 因此, 为了更好地排除和弦自身所携带的诱发知觉期待的可能, 研究者采用了重属和弦进行研究. 重属和弦与主调和弦具有较小的和声关联. Koelsch等人[26]使用了5个和弦序列, 结尾部分分别使用了主和弦(句法正确), 上中音和弦(句法轻微违反)以及重属和弦(句法违反). 在TPS理论中, 重属和弦与主和弦的距离大于上中音和弦与主和弦的距离; 另一方面, 主和弦与背景共享的音低于重属和弦以及上中心音和弦. 研究者已通过计算, 保障了中心和弦与背景具有最小的谐波频谱重叠程度. 结果发现, 使用重属和弦及上中音和弦的情况, 均比使用主和弦的情况, 显著诱发了ERAN和N5成分波幅, 说明基于“图式驱动”的抽象认知功能在音乐句法加工中占据着更重要的作用. 此外, 研究者认为, 由于重属和弦自身不存在引发听觉失衡的问题, 在这种情况下仍然诱发了ERAN, 说明ERAN能够代表抽象句法加工过程. Choi等人[47]在实验材料中, 进一步控制了可能促使句法期待的其他无关变量(使违例音符的出现概率等同于背景中的其他音符, 使节拍不规则化以及使违例音符出现的节拍具有不可预期性等). 结果发现, 调外音符显著诱发了P2波幅, 支持了“图式驱动”的作用. 近期, Koelsch等人[6]在以往的研究基础上发现, 听众不仅能够加工线性句法结构, 还能够感知带有嵌套关系的大尺度层级句法结构, 研究采用包括有离调部分的音乐结构, 基本段落调性结构为G-C-G, 研究显示, 当出现离调部分前后之间调性关系不匹配时, 激活了更大的ERAN和N5波幅. 说明听众能够知觉到句法的违反, 鉴于包含嵌套关系的层级句法结构具有较高程度的抽象性, 听众很难依靠短时记忆的作用完成加工, 因此, 该项研究同样也为认知主义的观点提出了例证.

还有一些研究同时操控音乐句法与语言句法的同步加工, 进而论证音乐句法与语言句法加工过程的相似程度. 实验逻辑是, 如果音乐和语言句法加工具有相同的认知机制, 那么在认知资源有限的情况下, 二者的句法加工将互相干扰. Fedorenko等人[48]发现, 当音乐和语言同时产生句法整合困难时, 被试在语言理解任务中的准确率就会下降. Slevc等人[49]的研究发现, 当音乐和语言句法同时产生违反时, 被试在自控速阅读范式下的阅读速度变慢. 有研究进一步通过电生理方法考察音乐和语言句法的同步加工. 当语言句法违反伴随着不规则的终止和弦时, 诱发的语言句法早期成分ELAN(early left anterior negativity)会明显变小, 说明诱发LAN和ERAN的加工在争夺相同的神经资源[50~52]. Carrus等人[53]通过对音乐

和语言句法加工的神经元振荡进行时频分析发现，二者的神经元活动在低频段振荡中存在重叠. 这些研究在一定程度上为音乐加工机制的“抽象性”做出了侧面论证，但这类实验范式存在的显著问题是，就算两者共享了认知资源，也不意味着一定是发生在抽象认知机制层面的共享，还有可能仅是发生在知觉层面. 因此，仍然不能够较好地解释音乐句法的加工机制.

2.2 听觉短时记忆模型及实验例证

（ⅰ）听觉短时记忆模型. 以Leman为代表的物理主义观点学者，提出了“听觉短时记忆模型”(auditoryshort-term memory, ASTM). 该理论认为听众形成的句法期待是短时记忆中和声音频不断累加的结果[43,54]，因为启动效应在很大程度上取决于启动和目标刺激之间谐波频谱重叠程度. 抽象知识在音乐句法的加工过程中只是“充分非必要条件”. ASTM模型包括4个部分，分别为外围听觉系统、音高周期性分析、回声记忆和音调语境. 外围听觉系统主要是将音频信号转化为神经元触发概率来激活听觉神经中的神经元触发速率编码模式，生成虚拟音高. 音高周期性分析把神经元触发概率作为输入，把生成的总结式自相关模式作为输出. 回声记忆部分的任务是随时间对音高进行积分，并最终生成即时音高认知图像和全局音高认知图像. 音调语境计算局部图像和全局图像间的相关度，当相关度系数越高时，说明即时音高事件与全局音高事件具有越高的兼容度，在知觉上也具有更高的期待度[7,43].

（ⅱ）听觉短时记忆模型对传统认知解释的质疑. 虽然短时记忆模型在十多年前就被提出，但由于早期计算技术限制，并没有对传统的研究所用的材料进行计算. 而近期，研究者采用听觉短时记忆模型拟合了以往一些较为经典的研究. 其结果发现，大部分研究出现的“图式驱动”效应都能够被基于“知觉驱动”的短时记忆模型进行解释. 例如，短时记忆模型通过拟合Bigand和Pineau[22]的实验材料发现，句法预期较高的和声走向与长距离启动背景的兼容度显著高于句法预期较低的和声走向与背景的兼容度，解释了句法预期较高条件下的任务能够得到易化，相同范式的其他研究也得到了短时记忆模型拟合的支持[46]. 例如，在操控Neapolitan和弦的研究中，Leino等人[25]在句法正确位置情况下使用Neapolitan和弦与前边背景的兼容度显著大于在句法不正确位置情况下使用Neapolitan和弦与前边背景的兼容度，解释了在位置正确情况下诱发的ERAN显著小于在其他位置下诱发的ERAN波幅. 此外，听觉短时记忆模型还较好地拟合了Koelsch等人[55]实验中使用调外和弦作为句法违反的情况，当目标刺激符合句法期待时，同时也与背景具有更高的兼容度，也激活了相对更小的ERAN波幅.

但是，目前听觉短时记忆模型也有不能够拟合的实验结果，例如，在Bigand等人[45]的研究中，无论在什么时间窗口拟合，出现的情况总体都是和声关联度较高的情况与背景的兼容度系数低于和声关联度较低的情况下与背景的兼容度系数，但结果却相反，和声关联度较高的情况下易化了加工. 此外，对于Koelsch等人[26]的研究结果也无法做到完全解释，在较短的时间窗口，符合句法期待的背景兼容度系数显著小于不符合句法期待的背景兼容度系数，而在较长的时间窗口，符合句法的和声走向与背景的兼容度系数显著小于不符合句法的和声走向与背景的兼容度系数. 这就不能解释在较短窗口上，句法违反所诱发的ERAN效应. 当然，对于目前不能够拟合的实验例证，也并不意味着一定是基于“图式驱动”下的效应，也许仅此是由于现阶段短时记忆模型的发展还不够成熟所致[7,8].

听觉短时记忆模型对传统的句法加工主要受制于“图式驱动”的理论解释提出了挑战. 虽然目前听觉短时记忆模型还不能够解释一些实验例证，但却无疑引发了人们对“知觉驱动”和“图式驱动”争议的重新思考. 首先，从方法学角度来说，早期研究主要是通过控制目标刺激在启动刺激中出现的频率来实现操控知觉驱动机制的方法，目前来看已经不再适用，单纯地消除目标刺激出现并不意味着能够消除谐波频谱的重叠，能够消除“知觉驱动”在其中的作用. 而为了进一步更好地操控认知在其中的作用，应该使认知的变化完全独立于声学成分的变化，在这种情况下，出现的效应可以较好地被解释为“图式驱动”的作用. 这提示在未来的研究中，需要更多地去思考实验的操控方法. 其次，听觉短时记忆模型能够拟合出大量曾经使用“图式驱动”进行解释的效应，引发对音乐句法加工机制的重新思索. 认知主义研究者一贯主张，音乐句法加工与语言句法加工在启动高抽象认知机制上具有类比性，其最核心的证据

在于人们对输入和弦的感知有较大的“背景依存性”. 也就是, 人们对当前输入刺激的心理反应特征依赖于背景的不同而不同[40]. 而目前来看, 这种“背景依存性”可以解释为在短时记忆中的背景累加效应, 背景声频元素的不断累加, 形成了与当前输入刺激兼容程度的改变. 兼容程度已经可以较好地预测输入刺激与背景信息在句法关系上的匹配程度, 并进一步预测人们对音乐事件预期的形成过程. 听觉短时记忆模型在解释力上的普遍性, 虽然不能够说明抽象的“图式驱动”加工已不存在, 但它至少说明了“知觉驱动”在现有实验材料条件下的音乐调性句法加工中扮演着非常重要的角色. 而这一点与语言加工形成了鲜明的对比, 同时, 也为音乐为什么能够做到“跨音乐经验加工”和“跨年龄加工”做出了非常具有说服力的佐证[56].

3 音乐句法加工的脑功能定位

Maess等人[57]较早采用脑磁图(magnetoencephalography, MEG)技术, 考察了Neapolitan和弦作为终止和弦的句法违例情况所激活的脑区活动. 研究发现Neapolitan和弦与主和弦(句法正确)相比, 显著激活了额下回(inferior frontal gyrus, IFG)后部的岛盖区域(inferior part of the pars opercularis), 即布洛卡44区(Broca 44), 激活表现出右侧化倾向. Koelsch等人[58]采用功能核磁共振(functional magnetic resonance imaging, fMRI)技术, 考察了调外和弦作为终止和弦的句法违例情况所激活的脑区活动, 研究发现, 调外和弦与中心和弦(句法正确)相比, 显著激活了布洛卡区、威尼克区(Wernicke)、颞上沟(the superior temporal sulcus, STS)、哈氏回(Heschl's gyrus)以及脑岛皮质前上部(the anterior superior insular cortices). Tillmann等人[59]采用fMRI技术, 考察了调外和弦作为终止和弦的句法违例情况所激活的脑区活动, 研究发现调外和弦与中心和弦相比, 显著激活了双侧额下回区域, 额叶岛盖区域, 脑岛区域(insula). Koelsch等人[60]采用10岁儿童、没有音乐经验的成人以及有音乐经验的被试进行研究, 考察了Neapolitan和弦作为终止和弦的句法违例情况所激活的脑区活动, 研究发现Neapolitan与主和弦相比, 显著激活了额上回(IFG)、额眶皮质、前脑岛区域、腹外侧前运动皮层(ventrolateral premotor cortex)、颞上回前部和后部以及缘上回区域. 在右半球区域, 儿童和成人大体一致, 但在左半球区域, 尤其是在前额区域、缘上回区域和颞叶区域, 成人的激活水平明显高于儿童. Tillmann等人[61]采用fMRI技术, 考察了下属和弦作为终止和弦的句法违例情况所激活的脑区活动, 研究发现下属和弦与主和弦相比, 显著激活了右侧额下回区域, 此外, 眶回区域(orbital gyrus)、颞叶区域和缘上回区域(supramarginal gyrus)也有激活. Sammler等人[62]采用ERP技术, 考察了额下回受损的病患加工句法的研究, 以重属和弦结尾作为句法违例情况, 研究发现, 病患出现了显著的ERAN分布异常, 说明额下回与句法加工具有较高的关联. Wehrum等人[63]采用经过音乐训练的被试和非音乐训练的儿童作为被试, 探讨了在旋律中间出现调外音作为句法违例激活的脑区活动. 研究发现当出现明显的违例时, 显著激活了额下回, 颞上回, 缘上回区域, 但是, 经过训练的儿童在左侧脑岛皮层和左侧缘上回具有明显的激活. 相反, 当出现微弱的违反时, 发现只有经过训练的儿童才能够激活额下回区域、前运动皮层区域、前脑岛区域、颞上回、缘上回区域.

从前人的脑成像研究中可以看出, 在句法加工的过程中, 不管采用何种违例方式, 有一些区域是普遍被激活的, 首先就是额下回区域, 尤其是额下回后部的岛盖区域[57~63]. 此外, 颞上回前部[58,60]、腹侧前运动皮层[58,60]也是普遍被激活的. 额下回区域、颞上回区域及腹侧前运动皮层区域都在语言句法的加工中扮演着非常重要的角色, 额下回区域是语言加工中涉及层级句法信息加工的核心区域[64~66]. 腹侧前运动皮层区域则是加工语言线性句法的关键区域[65]. 目前, 脑成像研究的发现说明, 音乐调性句法和语言句法至少在句法加工的部分认知过程中是具有重叠性的, 这在一定程度上支持了调性句法加工需要启动抽象化认知图式[41,42]. 但是, 音乐句法与语言句法的抽象程度在多大程度上具有可比性, 仍然是存在较大争议的. 因为现有研究仅能说明句法加工涉及这些脑区, 而无法证明这些脑区是音乐和语言句法加工的必要条件, 此外, 音乐和语言句法加工存在的重叠脑区, 并没有办法说明是在基本声学加工上存在的重叠, 还是在抽象认知法则上存在的重叠. 再者, 具有重叠的神经区域并不意味着神经连接和神经激活方式上的共享[52,67]. 近期研究发现, 音乐与语言在重叠的区域内却存在着各自加工的特异性, 例

如，Sammler通过颅内脑电图(electroencephalogram, EEG)记录技术发现，音乐句法违例和语言句法违例都显著激活了双侧颞上回区域，但在半球激活的时间进程和参与的额叶、颞叶结构程度上存在较大区别. 在传统的EEG和fMRI定位技术上，一些新型分析技术，如颅内EEG记录技术(intracranial recordings)、fMRI改良范式(fMRI adaptation paradigm)、多体素模式分析技术(multi-voxel pattern analysis)等，已经被应用于考察音乐加工的具体神经连接和神经激活模式，并比较与语言及其他领域加工的共性与特异性[67~70]. 这些技术对于深化句法认知神经机制的探索具有重要意义，同时，由于句法认知机制的联系与领域间能力关系具有较高程度的相关，研究可能在一定程度上为语言障碍治疗和其他认知功能障碍治疗提供借鉴[71,72].

4 思考及展望

4.1 加工层级句法结构的研究

Jackendoff[11]和Lerdahl[12,13]在GTTM中，提出完整音乐形象的表达有赖于多个局部的句法关系运动嵌入到更高层级的运动之中，构成具有层级嵌套关系的整体性事件. 理论上认为，听众要想获得对音乐表达意义的充分理解，不仅需要加工局部的句法关系，还需要在高层级上整合句法结构之间的相互嵌套关系，理解大尺度层级句法结构. 对层级句法结构的加工一直被视为人类高级认知智慧的特性，体现着人与其他动物的本质区别. 研究者在不同领域对此问题做出了探索，证实人们能够加工带有嵌套关系的语言层级法则、数学层级法则以及动作序列层级法则[64,73]. 然而，在音乐句法领域的研究中，层级句法的加工还较少受到关注，而在不多的几项研究中，仅有Koelsch等人[6]发现了听众在远距离依存关系的句法结构违反条件下，仍然能够诱发出在线性句法违反条件下的ERAN成分和N5成分，证实听众能够感知层级句法结构，存在整合层级句法表征的能力. 虽然Koelsch等人的研究为论证音乐层级句法的加工带来了曙光，但是，由于它与以往相比[31~36]，在实验范式、音乐材料以及测量指标方面都存在较多不一致，使得较难与以往研究进行对比，研究结论也有必要继续验证. 研究者可以考虑在现有实验范式基础上，增加音乐材料的长度和句法复杂度，为句法层级结构的感知做出更好的论证.

再者，影响听众对层级句法结构整合的因素是什么？以及影响机制是什么？基于音乐与语言在语法层级结构上的相似性，那么，在语言句法加工中，影响层级语言结构句法整合的常见因素，例如，高层级结构与低层级结构之间的位置依存关系，高层级结构与低层级结构之间的语义关系，能否影响到音乐句法中的高层级和声整合呢？并进一步影响到对音乐意义和情绪的提取呢？相关研究都强调工作记忆在层级句法整合中具有非常重要的作用，那么，工作记忆在音乐层级句法的加工中，又起到什么作用呢？它如何整合基于长时记忆的远距离依存关系句法和基于短时记忆的临近依存关系句法结构[65,66]？又是如何协调“图式驱动机制”和“知觉驱动机制”的工作呢？

此外，还有必要进一步探讨加工层级关系句法的认知神经机制[74,75]. 当前语言及其他领域研究倾向于认为，大脑存在两条平行的通路处理层级句法结构(这些结构包括数学领域、动作领域、人工语法领域及自然语言领域)，一条是沿着前额皮层(prefrontal cortex, PFC))后部到前部的梯度区域(BA 47/45a和10)，负责处理层级结构中的复杂序列，当任务需要的执行控制成分越多时，它激活的部位越靠前边，它带有更多的计算性. 另一条通路位于PFC的后部(额下回后侧IFG (BA 44/45p))，它是较为自动化处理语言层级结构的特异性模块，当处理层级句法结构的自动化程度越高时，那么，对它的激活程度越高，而当处理层级句法结构自动化程度越低时，就越需要较多的执行控制，越需要调动前额皮层的前边部分[64,74,75]. 那么，对音乐层级句法结构的加工，究竟是更偏向于依赖前额皮层的前部，还是更偏向于直接依赖前额皮层后部呢？它是否存在自己的特异性模块？又是否仍然需要调动执行控制的相关区域？层级结构的加工涉及更为复杂的神经网络[64,74,75]，对它的深入讨论，有利于进一步揭示人类高级认知功能的本质特性. 同时，人们对音乐的理解，往往是以大尺度的层级结构为单位，对大尺度层级结构整合的研究，具有更广泛的生态效度及应用前景.

4.2 分离知觉驱动与图式驱动机制

听觉短时记忆模型能够拟合出大量曾经用“图式驱动”进行解释的效应，无疑引发对音乐句法加工机

制的重新思索, 但也仍然看到了听觉短时记忆模型存在的明显局限. 首先, 面对现有基于“图式驱动”解释的结果, 听觉短时记忆模型并不能够完全拟合[7]. 从拟合的情况来看, 听觉短时记忆模型对选取的时间窗口具有较大的敏感性, 在背景信息时间较长的窗口和局部信息较短窗口的情况下, 往往不能够拟合, 这说明背景信息的加长, 会使听觉短时记忆模型的解释力受到限制. 其次, 从现有研究来看, 其考察的都主要是集中于较为简单的线性句法, 听觉短时记忆模型所拟合的数据也都是基于线性句法关系之上的结果. 从考察材料的具体实验材料来看, 大部分都是时长较短的材料[37], 这些情况都无疑加大了基于知觉驱动发生的可能性. 所以, 听觉短时记忆模型所带来的对“图式驱动”的挑战更多只能是局限于现有的实验条件. 当然, 更主要的问题应该看到, “图式驱动”观点的提出并不是完全排斥“知觉驱动”的可能, 而“知觉驱动”的提出也不是完全否认基于长时记忆“图式驱动”的存在. Bigand等人的研究具有较好的借鉴意义, 实验通过操控目标和弦在启动背景中出现的频率, 以直接比较重复启动与图式启动的作用. 实验结果显示, 即使是在非常快的速度下, 基于图式产生的启动都优于基于重复启动产生的效应. Collins等人[8]的研究更为直接, 通过回归模型拟合及量化了知觉驱动和图式驱动的效应系数. 研究从近期的7项和声启动研究中选取了303段刺激材料, 采用逐步回归的方法, 依次拟合了短时记忆模型和认知模型对反应时的解释率, 并最终纳入了统一的模型中, 研究发现, 如果只是单独的短时记忆模型, 只能解释数据21%的变异率, 单独的认知模型只能解释数据22%的变异率, 认知模型比短时记忆模型更具有解释力, 而当统一纳入模型时, 两者最终能够达到26%的解释率. 研究结果充分说明, 任意单一模型都不能够解释音乐句法的加工过程, 自上而下和自下而上的加工方式都是同时存在的, 并且还会发生交互作用, 仅论证某个机制是否发生作用已经没有太大意义. 现有语言领域的研究进一步发现, 几乎所有涉及语言加工的皮层都在同时性地存储和加工信息, 只是累积的时间尺度沿着大脑的皮层区, 以层级性方式发生变化, 伴随着较短的时间窗口向较长时间窗口的发展, 皮层运动的主导性由感觉皮层区衍生至高级皮层区[76], 可见, 即使是在语言这样具有较高抽象关系的声学刺激条件下, “图式驱动”和“知觉驱动”也都是同时存在的, 而每一次对新信息加工的心理反应, 都一定是基于局部的整合关系和全局整合关系共同进行的, 也都一定是依靠短时记忆和长时记忆共同完成的[76]. Collins等人进一步指出感知-认知机制的启动是一个连续的过程, 而不是绝对两级化的过程. 随着句法复杂程度和时间窗口的不同, 图式驱动和知觉驱动彼此间的作用大小及关系将会发生调整. 所以, 未来研究中应该着眼于在不同的句法层级水平和不同的时间窗口条件下进行讨论, 同时, 应该考虑应用实验操控方法和数据统计技术, 分离不同抽象层级水平认知机制的启动效应, 比较各自的作用权重, 并进一步探讨彼此可能存在的交互作用.

4.3 脑电和脑成像技术的进一步应用

目前, 虽然脑电和脑成像技术已经广泛地应用于音乐句法加工的研究中, 但由于实验设计和分析技术的限制, 使得这些技术本身的价值还远没有发挥出来. 对于ERP技术来说, 首先由于目前实验操控技术上的限制, 使得一些重要的成分所代表的认知加工机制仍然带有较大的模糊性. 例如, 早期的句法整合成分指标ERAN, 它在多大程度上能够作为句法整合的特异性指标, 它又在多大程度上能够代表句法加工过程的“图式驱动”过程, 它与代表引发冲突的早期指标MMN(mismatch negativity)究竟有什么区别[27,77]? 晚期窗口的句法整合成分指标N5也存在争议性, 它除了可能代表句法的冲突外[23,24], 它还可能代表音乐语义的冲突[78], 也可能代表更广泛意义上的整合及修复[6,28]. 今后研究需要进一步借助于实验操控来实现分离, 同时, 还需要借助于溯源分析[77], 才能更好地探明这些成分的解释功能.

其次, 有必要进一步利用好EEG和ERP的分析技术, 挖掘数据的价值. 从分析方法的角度来说, 现有研究还主要集中于时域分析, 应该向频域分析发展, 并建立起时域指标与频域指标之间的关系. 既可以分析同一频段内部能量的变化或不同频段之间能量的相互关系, 还可以分析同一频段内部的相位变化或不同频段之间相位的相互关系, 以及一个频段的能量和另一频段的相位之间的关系. 时频分析的应用有助于通过观察能量的变化, 揭示所激活的认知功能及神经机制[79]. 现有语言研究领域普遍认为: 低频波段(1~10 Hz)反映了非特异加工过程, 如感觉、注意、记忆等基本的认知任务; 较高频段(11~31 Hz)

反映了语言特异性加工过程, 如句法加工和语义加工[80,81]. 那么, 在音乐句法加工中, 反映其加工过程的波段分别是什么呢? 它们是否与语言加工具有相似性呢? Carrus等人[53]首次把时频分析技术用于音乐句法加工研究, 研究发现, 句法违反显著诱发了delta波段和theta波段在晚期窗口上的增能量效应(350~700 ms). 这与语言研究中在数、性、人称等形态句法违反表现出了一些相同特征[79,82], 但由于Carrus等人的研究仅涉及低频波段(delta和theta), 没有办法在句法特异性指标上做出更多比较(如 α, β和γ), 今后研究需要更多关注较高频波段和高频波段在句法加工中的能量变化, 因为它们与句法加工可能有着更特异性的联结[81,83]. 同时, 可以在时频分析基础上, 通过EEG相干技术进一步讨论不同脑区之间能量变化的相干模式, 更为全面地揭示脑区之间的信息传递过程[80].

对于脑成像技术来说(fMRI/PET/MEG), 目前应用于音乐句法研究的总体数量还较少, 脑成像技术的应用范围有待扩展. 此外, 除了加强传统的定位研究以外, 还需要关注脑区的功能连接. 早期的脑科学的研究重点是确定特殊功能脑区, 而当对大脑有更深层次的了解时, 研究者发现大脑不仅具有数量众多的功能脑区, 而且这些脑区在功能上还相互连接. 没有脑连接图是不太可能理解脑的工作机制, 脑科学的研究正在从定位特殊功能脑区逐渐转变到研究脑功能连接[84]. 目前, 脑功能连接的研究已广泛地应用于其他领域(语言领域、精神疾病领域和情绪领域). 在语言领域, 一些研究开始探讨脑区之间的功能连接对句法加工的影响[85], 并尝试找到连接脑区的结构基础[74]. 句法加工网络的研究在定位信息的基础上, 让人们更清晰地了解了不同脑区是如何相互协调的, 甚至信息是如何在不同脑区之间流动的, 这为探讨句法加工脑机制提供了一个新的视角.

4.4 关注句法结构与情绪的关系

音乐认知与情绪之间存在着密切的交互作用, 是人类产生复杂行为从而适应环境的基础. Meyer从认知角度解释了句法结构诱发音乐情绪发生的过程, 认为音乐刺激与个体脑中音乐旋律表征之间的匹配程度决定了音乐情绪发生的一般化过程[2]. 音乐表征会在音乐活动中支配着参与者的音乐认知, 使个体对音乐事件的发展产生自动化的预期和期待, 当人们的期待得以“实现”时, 一切就稳定地“持续下去”, 而当人们的期待“不能得到满足”, 即音乐事件的进行与期待相冲突时, 个体就会产生某些紧张和不适的负性情绪. 音乐句法图式作为引发聆听者对音乐事件的发展形成期待的最重要因素, 也成为了诱发音乐情绪的最重要因素[86,87]. Jackendoff[11]和Lerdahl[12,13]在GTTM模型中, 通过延长还原(prolongation reduction)概念提出, 紧张-放松的关系是具有层级性的, 局部的紧张或放松运动往往嵌入在更高层级的运动之中. Koelsch[87]进一步指出, 在不同的句法层级上, 诱发的情绪反应可能是不同的, 诱发情绪的机制也可能是不同的. 例如, 在最低的句法层级上, 简单的不和谐音, 或者音色的偏离就可能诱发情绪[88], 而在高句法层级上, 需要启动与音乐文化相关的高度抽象知识才能够诱发情绪. 延长-还原句法树的提出, 为音乐句法诱发情绪做出了完整的理论构想. 遗憾的是, 现有研究才刚刚起步, 仅集中于讨论最低水平的线性句法结构与情绪的关系[55,89]. 较高层级的句法结构诱发出的情绪反应将表现出什么模式? 诱发情绪过程中将激活什么样的脑机制? 高层级句法结构和低层级句法结构在诱发情绪过程中是如何实现整合? 又是如何相互作用? 这些都是近期有必要开展的研究.

此外, 研究还有必要关注调性句法诱发情绪的生理成分. 一些音乐美学研究者认为: 音乐情绪的产生源于期待的实现与否, 而主导期待的音乐调性句法是在音乐文化中逐渐演变形成的符号规则, 它所激活的音乐情绪是纯认知的美学体验, 并不具有一般化情绪所携带的适应功能. 因此, 也就不存在具身性的“生理唤醒”, 而即使有生理变化, 也是领域特殊性的[90]. 另外一些研究者认为, 个体对声音特质的敏感是与进化相关的适应功能, 因此, 音乐句法结构对情绪的激活与其他情绪刺激物下的生理反应模式相似[90~92]. 对这个问题的争论同样应该在不同层级的句法结构框架下进行. 因为伴随着句法结构层级的不同, 其抽象程度和复杂程度都不同, 诱发情绪的机制不相同, 所产生的情绪反应模式也不相同. 同时, 对于考察音乐句法情绪的反应成分, 应该更多地监测生理指标, 观察调性句法诱发下的情绪反应中, 是否一定具有生理变化? 如果有, 生理指标与主观评价指标之间是否会呈现出某种关联? 这种关联与其他情绪刺激物下的共性与区别是什么? 进而揭示

音乐句法诱发的情绪反应, 具有领域一般化特征还是领域特殊性特征. 再者, 未来研究还有必要讨论情绪如何反作用于句法认知乃至更广泛意义上的其他音乐认知过程? 音乐认知加工脑区与情绪加工脑区之间的双向连接和动态交互如何形成音乐认知与情绪交互作用的脑机制? 音乐认知与情绪交互作用如何导致音乐治疗有效性的产生? 这些问题的讨论, 对于基础研究和临床应用都具有重要意义[87].

5 结语

音乐句法作为音乐结构的重要特征已经得到了大量的研究. 然而, 现有研究对于解释人们究竟如何加工"音乐句法树", 仅是刚开启了一个枝芽. 现有研究存在的分歧将成为无数个新议题的生长点, 等待着研究者站在宏观的脉络上做出承前启后的关注与思索. 希望本研究能够为进一步揭示音乐句法的加工机制和促进音乐句法认知的临床应用提供借鉴.

参考文献

1 Koelsch S. Brain correlates of music-evoked emotions. Nat Rev Neurosci, 2014, 15: 170–180
2 Tillmann B, Poulin-Charronnat B, Bigand E. The role of expectation in music: From the score to emotions and the brain. Wiley Interdiscip Rev Cogn Sci, 2014, 5: 105–113
3 Lerdahl F, Jackendoff R S. A generative theory of tonal music. Cambridge, Mass: MIT Press, 1983
4 Rohrmeier M. Towards a generative syntax of tonal harmony. J Math Music, 2011, 5: 35–53
5 Koelsch S. Toward a neural basis of music perception—a review and updated model. Front Psychol, 2011, 2: 110
6 Koelsch S, Rohrmeier M, Torrecuso R, et al. Processing of hierarchical syntactic structure in music. Proc Natl Acad Sci USA, 2013, 110: 15443–15448
7 Bigand E, Delbé C, Poulin-Charronnat B, et al. Empirical evidence for musical syntax processing? Computer simulations reveal the contribution of auditory short-term memory. Front Syst Neurosci, 2014, 8: 94
8 Collins T, Tillmann B, Barrett F S, et al. A combined model of sensory and cognitive representations underlying tonal expectations in music: From audio signals to behavior. Psychol Rev, 2014, 121: 33–65
9 Rohrmeier M A, Koelsch S. Predictive information processing in music cognition. A critical review. Int J Psychophysiol, 2012, 83: 164–175
10 Krumhansl C L, Cuddy L L. A theory of tonal hierarchies in music. Music percept. New York: Oxford University Press, 2010. 51–87
11 Jackendoff R. Parallels and nonparallels between language and music. Music Percept, 2009, 26: 195–204
12 Lerdahl F. Modeling tonal tension. Music Percept, 2007, 24: 329–366
13 Lerdahl F. Concepts and representations of musical hierarchies. Music Percept, 2015, 33: 83–95
14 Lerdahl F. Tonal Pitch Space. New York: Oxford University Press, 2001
15 Chomsky N. Aspects of the Theory of Syntax. Cambridge, Massachusetts: MIT Press, 2014
16 Rohrmeier M, Dienes Z, Guo X, et al. Implicit learning and recursion. Learn Recur, 2013, 67: 1–24
17 Tillmann B. Music and language perception: Expectations, structural integration, and cognitive sequencing. Top Cogn Sci, 2012, 4: 568–584
18 Bharucha J J, Stoeckig K. Reaction time and musical expectancy: Priming of chords. J Exp Psychol Hum Percept Perform, 1986, 12: 403–410
19 Bharucha J J, Stoeckig K. Priming of chords: Spreading activation or overlapping frequency spectra? Percept Psychophys, 1987, 41: 519–524
20 Tekman H G, Bharucha J J. Time course of chord priming. Percept Psychophys, 1992, 51: 33–39
21 Tekman H G, Bharucha J J. Implicit knowledge versus psychoacoustic similarity in priming of chords. J Exp Psychol Hum Percept Perform, 1998, 24: 252–260
22 Bigand E, Pineau M. Global context effects on musical expectancy. Percept Psychophys, 1997, 59: 1098–1107
23 Koelsch S, Gunter T, Friederici A D, et al. Brain indices of music processing: "Nonmusicians" are musical. J Cogn Neurosci, 2000, 12: 520–541
24 Koelsch S, Schmidt B H, Kansok J. Effects of musical expertise on the early right anterior negativity: An event-related brain potential study. Psychophysiol, 2002, 39: 657–663
25 Leino S, Brattico E, Tervaniemi M, et al. Representation of harmony rules in the human brain: Further evidence from event-related potentials. Brain Res, 2007, 1142: 169–177
26 Koelsch S, Jentschke S, Sammler D, et al. Untangling syntactic and sensory processing: An ERP study of music perception. Psychophysiology, 2007, 44: 476–490
27 Koelsch S. Music-syntactic processing and auditory memory: Similarities and differences between ERAN and MMN. Psychophysiology, 2009, 46: 179–190

28 Poulin-Charronnat B, Bigand E, Koelsch S. Processing of musical syntax tonic versus subdominant: An event-related potential study. J Cogn Neurosci, 2006, 18: 1545–1554

29 Jentschke S, Friederici A D, Koelsch S. Neural correlates of music-syntactic processing in two-year old children. Dev Cogn Neurosci, 2014, 9: 200–208

30 Koelsch S, Grossmann T, Gunter T C, et al. Children processing music: Electric brain responses reveal musical competence and gender differences. J Cogn Neurosci, 2003, 15: 683–693

31 Konečni V J. Elusive effects of Artists' "Messages". Adv Psychol, 1984, 19: 71–93

32 Gotlieb H, Konečni V J. The effects of instrumentation, playing style, and structure in the Goldberg Variations by Johann Sebastian Bach. Music Percept, 1985, 3: 87–101

33 Tillmann B, Bigand E, Madurell F. Local versus global processing of harmonic cadences in the solution of musical puzzles. Psychol Res, 1998, 61: 157–174

34 Knösche T R, Neuhaus C, Haueisen J, et al. Perception of phrase structure in music. Hum Brain Mapp, 2005, 24: 259–273

35 Lalitte P, Bigand E. Music in the moment? Revisiting the effect of large scale structures. Percept Mot Skills, 2006, 103: 811–828

36 Tymoczko D. A Geometry of Music: Harmony and Counterpoint in the Extended Common Practice. New York: Oxford University Press, 2011

37 Farbood M M, Heeger D J, Marcus G, et al. The neural processing of hierarchical structure in music and speech at different timescales. Front Neurosci, 2015, 9: 157

38 Bharucha J J. Music cognition and perceptual facilitation: A connectionist framework. Music Percept, 1987, 5: 1–30

39 Patel A D, Gibson E, Ratner J, et al. Processing syntactic relations in language and music: An event-related potential study. J Cogn Neurosci, 1998, 10: 717–733

40 Patel A D. Music, Language, and the Brain. New York: Oxford University Press, 2007

41 Patel A D. Language, music, and the brain: A resource-sharing framework. Lang Music Cogn Syst, 2002, 4: 204–223

42 Patel A D. Sharing and nonsharing of brain resources for language and music. In: Michael A A, ed. Lang, Music, Brain: A Mysterious Relationship. Cambridge, Massachusetts: MIT Press, 2013

43 Leman M. An auditory model of the role of short-term memory in probe-tone ratings. Music Percept, 2000, 17: 481–509

44 Bigand E, Tillmann B, Poulin-Charronnat B, et al. Repetition priming: Is music special? Quart J Exp Psychol, 2005, 58: 1347–1375

45 Bigand E, Poulin B, Tillmann B, et al. Sensory versus cognitive components in harmonic priming. J Exp Psychol Hum Percept Perform, 2003, 29: 159–171

46 Regnault P, Bigand E, Besson M. Different brain mechanisms mediate sensitivity to sensory consonance and harmonic context: Evidence from auditory event-related brain potentials. J Cogn Neurosci, 2001, 13: 241–255

47 Choi I, Bharadwaj H M, Bressler S, et al. Automatic processing of abstract musical tonality. Front Hum Neurosci, 2014, 8: 1–10

48 Fedorenko E, Patel A, Casasanto D, et al. Structural integration in language and music: Evidence for a shared system. Mem Cogn, 2009, 37: 1–9

49 Slevc L R, Rosenberg J C, Patel A D. Making psycholinguistics musical: Self-paced reading time evidence for shared processing of linguistic and musical syntax. Psychon Bull Rev, 2009, 16: 374–381

50 Koelsch S, Gunter T C, Wittfoth M, et al. Interaction between syntax processing in language and in music: An ERP study. J Cogn Neurosci, 2005, 17: 1565–1577

51 Sammler D, Koelsch S, Ball T, et al. Co-localizing linguistic and musical syntax with intracranial EEG. NeuroImage, 2013, 64: 134–146

52 Peretz I, Vuvan D, Lagrois M É, et al. Neural overlap in processing music and speech. Philos Trans R Soc Lond B Biol Sci, 2015, 370: 1–8

53 Carrus E, Koelsch S, Bhattacharya J. Shadows of music-language interaction on low frequency brain oscillatory patterns. Brain Lang, 2011, 119: 50–57

54 Huron D, Parncutt R. An improved model of tonality perception incorporating pitch salience and echoic memory. Psychomusicology, 1993, 12: 154–171

55 Koelsch S, Kilches S, Steinbeis N, et al. Effects of unexpected chords and of performer's expression on brain responses and electrodermal activity. PLoS One, 2008, 3: e2631

56 Perani D, Saccuman M C, Scifo P, et al. Functional specializations for music processing in the human newborn brain. Proc Natl Acad Sci USA, 2010, 107: 4758–4763

57 Maess B, Koelsch S, Gunter T C, et al. Musical syntax is processed in Broca's area: An MEG study. Nat Neurosci, 2001, 4: 540–545

58 Koelsch S, Gunter T C, v Cramon D Y, et al. Bach speaks: A cortical "language-network" serves the processing of music. NeuroImage, 2002, 17: 956–966

59 Tillmann B, Janata P, Bharucha J J. Activation of the inferior frontal cortex in musical priming. Brain Res Cogn Brain Res, 2003, 16: 145–161

60 Koelsch S, Fritz T, Schulze K, et al. Adults and children processing music: An fMRI study. NeuroImage, 2005, 25: 1068–1076

61 Tillmann B, Koelsch S, Escoffier N, et al. Cognitive priming in sung and instrumental music: Activation of inferior frontal cortex. NeuroImage, 2006, 31: 1771–1782

62 Sammler D, Koelsch S, Friederici A D. Are left fronto-temporal brain areas a prerequisite for normal music-syntactic processing? Cortex, 2011, 47: 659–673
63 Wehrum S, Degé F, Ott U, et al. Can you hear a difference? Neuronal correlates of melodic deviance processing in children. Brain Res, 2011, 1402: 80–92
64 Friederici A D, Bahlmann J, Friedrich R, et al. The neural basis of recursion and complex syntactic hierarchy. Biolinguistics, 2011, 5: 87–104
65 Makuuchi M, Bahlmann J, Anwander A, et al. Segregating the core computational faculty of human language from working memory. Proc Natl Acad Sci USA, 2009, 106: 8362–8367
66 Makuuchi M, Friederici A D. Hierarchical functional connectivity between the core language system and the working memory system. Cortex, 2013, 49: 2416–2423
67 Perruchet P, Poulin-Charronnat B. Challenging prior evidence for a shared syntactic processor for language and music. Psychon Bull Rev, 2013, 20: 310–317
68 Alho K, Rinne T, Herron T J, et al. Stimulus-dependent activations and attention-related modulations in the auditory cortex: A meta-analysis of fMRI studies. Hearing Res, 2014, 307: 29–41
69 Angulo-Perkins A, Aubé W, Peretz I, et al. Music listening engages specific cortical regions within the temporal lobes: Differences between musicians and non-musicians. Cortex, 2014, 59: 126–137
70 Asano R, Boeckx C. Syntax in language and music: What is the right level of comparison? Front Psychol, 2015, 6: 942
71 Oechslin M S, van De Ville D, Lazeyras F, et al. Degree of musical expertise modulates higher order brain functioning. Cereb Cortex, 2013, 23: 2213–2224
72 Patel A D. Can nonlinguistic musical training change the way the brain processes speech? The expanded OPERA hypothesis. Hear Res, 2014, 308: 98–108
73 Jeon H A. Hierarchical processing in the prefrontal cortex in a variety of cognitive domains. Front Syst Neurosci, 2014, 8: 223
74 Friederici A D, Singer W. Grounding language processing on basic neurophysiological principles. Trends Cogn Sci, 2015, 19: 329–338
75 Jeon H A, Friederici A D. Degree of automaticity and the prefrontal cortex. Trends Cogn Sci, 2015, 19: 244–250
76 Hasson U, Chen J, Honey C J. Hierarchical process memory: Memory as an integral component of information processing. Trends Cogn Sci, 2015, 19: 304–313
77 Paavilainen P. The mismatch-negativity (MMN) component of the auditory event-related potential to violations of abstract regularities: A review. Int J Psychophysiol, 2013, 88: 109–123
78 Steinbeis N, Koelsch S. Shared neural resources between music and language indicate semantic processing of musical tension-resolution patterns. Cereb Cortex, 2008, 18: 1169–1178
79 Cai L, Zhang Y X. EEG time-frequency analysis of syntactic and semantic processing in sentence comprehension (in Chinese). Adv Psychol Sci, 2014, 22: 1112–1121 [蔡林, 张亚旭. 句子理解过程中句法与语义加工的 EEG 时频分析. 心理科学进展, 2014, 22: 1112–1121]
80 Wang L, Zhang Q F, Yang Y F. The application of EEG coherence analysis in language comprehension (in Chinese). Adv Psychol Sci, 2007, 15: 865–871 [王琳, 张清芳, 杨玉芳. EEG 相干分析在语言理解研究中的应用. 心理科学进展, 2007, 15: 865–871]
81 Fontolan L, Morillon B, Liegeois-Chauvel C, et al. The contribution of frequency-specific activity to hierarchical information processing in the human auditory cortex. Nat Commun, 2014, 5: 4694
82 Lewis A G, Bastiaansen M. A predictive coding framework for rapid neural dynamics during sentence-level language comprehension. Cortex, 2015, 68: 155–168
83 Doelling K B, Poeppel D. Cortical entrainment to music and its modulation by expertise. Proc Natl Acad Sci USA, 2015, 112: E6233–E6242
84 Fonteneau E, Bozic M, Marslen-Wilson W D. Brain network connectivity during language comprehension: Interacting linguistic and perceptual subsystems. Cereb Cortex, 2015, 25: 3962–3976
85 den Ouden D B, Saur D, Mader W, et al. Network modulation during complex syntactic processing. NeuroImage, 2012, 59: 815–823
86 Huron D B. Sweet anticipation: Music and the psychology of expectation. J Aesthet Art Critic, 2006, 65: 333–335
87 Koelsch S. Music-evoked emotions: Principles, brain correlates, and implications for therapy. Ann N Y Acad Sci, 2015, 1337: 193–201
88 Bidelman G M, Krishnan A. Neural correlates of consonance, dissonance, and the hierarchy of musical pitch in the human brainstem. J Neurosci, 2009, 29: 13165–13171
89 Lehne M, Rohrmeier M, Koelsch S. Tension-related activity in the orbitofrontal cortex and amygdala: An fMRI study with music. Soc Cogn Affect Neurosci, 2014, 9: 1515–1523
90 Juslin P N. Music and emotion: Seven questions, seven answers. In: Deliège I, Davidson J, Sloboda J A, eds. Music and the Mind: Essays in Honour of John Sloboda. New York: Oxford University Press, 2011. 113–135
91 Hill W T, Palmer J A. Affective responses to music without recognition: Beyond the cognitivist hypothesis. New Sch Psychol Bull, 2014, 11: 42–49
92 Ma X, Bai X J , Tao Y. Music and emotional induced model (in Chinese). Adv Psychol Sci, 2013, 21: 643–665 [马谐, 白学军, 陶云. 音乐与情绪诱发的机制模型. 心理科学进展, 2013, 21: 643–665]

Music syntactic processing

MA Xie[1,2,3], YANG YuFang[1] & ZHANG QiuYue[2,3]

[1] *Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China;*
[2] *College of Educational Science and Management, Yunnan Normal University, Kunming 650500, China;*
[3] *Key Laboratory of Educational Informatization for Nationalities, Yunnan Normal University, Kunming 650500, China*

Music Syntax refers to the principle of combining discrete key elements into hierarchical system. Organization of pitches based on tonality harmony is the most important music syntax which affects the expectation, the construction and the feedback on the music events of the audience from which aesthetic musical experience is finally produced. Lerdahl and Jackendoff proposed a musical tree-structure that parallel to linguistic syntax in their generative theory of tonal music (GTTM). And Rohrmeier further specified hierarchical generativerules of pitch syntax in the generative syntax model (GSM). Both the GTTM and the GSM put forward the analogy between music syntax and linguistic syntax. But what is obviously different from language processing is that the processing of music syntax of the audience reveals a cross-empiricalness effects. Whether in early studies of behavioristics or in recent studies of EPR, most evidences support that there is a syntax diagram guiding people's musical expectations without much association with their musical experiences among western audience, at least in the sense of simple linear syntax. Therefore, the acquisition of music syntax under the non-explicit conditions arouses researchers' thinking on music syntax processing mechanism. Current studies have not reached a consensus on processing of syntactic structure, and two antagonistic theories "Cognitivism" and "Physicalism" have generated. "Cognitivism" put forward that since pitch syntax reflects abstract cognitive structural relations between sounds and meanings, syntactic integration requires the help of scheme drive of long-term memory, of which MUSACT and the shared syntactic integration resource hypothesis (SSIRH) are the representative theories. While "Physicalism" considered that it only needs the help of short-term memory in perception drive because pitch syntax has a property of psychoacoustics derived from voice frequency. And the periodicity pitch model (PP) and its derivate, the auditory short-term memory (ASTM), are the representative theories. Both theories demonstrated their viewpoints respectively from the angle of theoretical model, behavioral studies and cognitive neuroscience.

It is necessary for future study to move on in the following aspects: (i) Investigations on the processing capability and mechanism of large scale nested syntactic hierarchical structure and it has to focus on what factors and mechanism affect the integration of syntactic hierarchical structure of the audience. It is especially essential to discuss the effect of working memory on how it works to integrate scheme drive mechanism and perception drive mechanism. (ii) Discussions on the controversy of scheme drive and perception drive should be conducted under conditions of different hierarchical syntactic levels and time windows. Meanwhile, experiments and statistics can be used to separate the mechanism effect of perception drive and scheme drive and to study the possible interaction. (iii) The application of EEG and FMRI can be further implemented to disclose functions of ERP by integrating statistics of time-domain analysis and frequency-domain analysis. Besides, FMRI, PET and MEG need to be extended from orientation studies to dynamic connection mode of brain function. (iv) Discussions on the relationship of syntax and emotions should also be carried out under conditions of different hierarchical syntax structure levels under which the effect and mechanism of induced emotion may be different.

All in all, the exploration of music syntactic processing could be used to reveal the characteristics of human advanced cognitive wisdom. Meanwhile, it has an important clinical value. It is expected our study can further reveal the mechanism of music syntax and promote the understanding of it, which can be helpful for clinical practices.

music, syntactic, harmony, schema-drive, perception-drive

doi: 10.1360/N972015-01340