arXiv:2304.04915v1 [cs.SD] 11 Apr 2023

AffectMachine-Classical：一种生成情感古典音乐的新型系统

Kat R. Agres*, †, 1, 2，Adyasha Dash*, 1，以及 Phoebe Chua[3]

[1]新加坡国立大学杨秀桃音乐学院，新加坡[2]新加坡国立大学音乐与健康中心，新加坡[3]新加坡国立大学增强人类实验室（DISA），新加坡

†通讯作者：Kat R. Agres，katagres@nus.edu.sg

本文介绍了一种名为 AffectMachine-Classical 的新型音乐生成系统，可实时生成情感古典音乐。AffectMachine 旨在集成到生物反馈系统（如脑机接口）中，帮助用户意识到自身动态变化的情绪状态，并最终对其进行调节。也就是说，该系统是为音乐类医疗技术开发的，旨在支持用户实时进行情绪自我调节。本文概述了该系统基于规则和概率的架构，介绍系统的主要组成部分及其创新之处。随后，我们展示了一项听众研究的结果，该研究旨在验证系统能否可靠地向听众传达目标情绪。研究结果表明，AffectMachine-Classical 能够非常有效地向听众传达不同水平的唤醒度（$R^2 = .96$），在效价（$R^2 = .90$）方面的表现也颇具说服力。未来的工作将把 AffectMachine-Classical 集成到生物反馈系统中，以发挥情感音乐的功效，提升听众的情绪幸福感。

关键词：自动音乐生成系统，算法作曲，音乐医疗科技，《情绪》调节，听众验证研究

1 引言

目前已有大量证据表明，音乐通过多种方式支持健康和幸福感，从激励身体活动，到促进心理健康以及培养社会联结（Fancourt and Finn, 2019年；MacDonaldet al., 2013）。音乐对于支持和调节情绪状态尤其有效。事实上，人们报告听音乐的主要原因之一，就是改变或增强自己的情绪（Thayer et al., 1994年；Lonsdale

*这些作者对本研究贡献相同，并共同列为第一作者。

以及 North，2011年；Saarikallio，2012年）。鉴于音乐能够促进健康和幸福感，加上机器学习和计算技术的进步，近期有人呼吁运用计算技术创作音乐，以用于医疗保健应用（Agres 等，2021年）。这类生成式音乐创作系统有望提供经济高效、非侵入性且非药物的方法，帮助人们改善情绪状态。鉴于全球心理健康危机（例如，美国近 20% 的成年人患有精神疾病；或者说， 2020[1] 中有 5290 万美国人），音乐医疗技术系统有望成为极具价值的工具，助力促进情绪健康，并更广泛地支持心理健康。

更广泛地说，在计算系统如今已被广泛用于生成令人印象深刻的自然语言文本和视觉艺术的时代，例如通过 OpenAI[2], 提供的技术，自动音乐生成系统（AMGSs；亦称算法作曲系统）的开发近期受到广泛关注，也就不足为奇了。与人类的音乐作曲和即兴创作类似，自动音乐生成系统通常旨在以有条理且符合音乐逻辑的方式创作和声、音色与节奏序列。然而，与计算机创作视觉艺术相比，这一处于计算机科学、音乐理论／作曲和计算创造力交叉领域的方向仍相对新兴。本研究不仅旨在缩小这一差距，还提出一种新型自动音乐生成系统——AffectMachine-Classical——能够生成可控的情感化音乐。
AffectMachine-Classical 提供了一种有效而灵活的方式，可实时传达情绪；该系统的开发目标是将其嵌入生物反馈系统，例如脑机接口（例如参见 Ehrlich 等人（2019年）），使其有望成为强大的工具，用于听众的情绪与情绪状态调节、康复期间增强身体活动等治疗应用，也可用于配乐设计以及为无声视频配上免版权的新音乐等商业场景。

1.1 相关工作

自动音乐生成综述不在本文的讨论范围内（有关该领域最新进展的综述和总结，请参阅 Herremans 等人（2017年）；Carnovalini 和 Rod`a（2020年）；Dash 和 Agres（2023年）），不过我们将简要总结自动音乐生成的主要方法。以往开发音乐生成系统的方法大致分为两类：基于学习的方法和基于规则的方法。尽管近期基于学习的方法日益受到关注，但它们在情感音乐生成方面仍面临若干挑战。首先，生态（或真实）的音乐作品通常具有层次结构和复调。例如，一个旋律乐句可能跨越多个小节，并涉及多个不同的乐器或声部。此外，音乐通常具有整体曲式，使作品始终保持音乐性和风格连贯性。因此，生成模型必须能够捕捉和声、节奏和时间结构，以及声部之间的相互依赖关系（Dong等人，2018年）。其次，基于学习的方法需要大量带有情绪标签的音乐数据集进行训练，而这类资源在学界仍然稀缺。风格迁移模型已取得成功

[1]美国国家心理健康研究所统计数据：https://www.nimh.nih.gov/health/statistics/mental- illness[2] https://openai.com/

作为能够从零开始生成新颖情感音乐的模型的替代方案——例如，Ramirez等人（2015）使用机器学习模型，根据期望的唤醒度和效价，对预先创作的输入音乐作品的时值和响度施加适当的表现性变换；而Hu等人（2020年）使用卷积神经网络从治疗性音乐作品中提取风格特征，并将其融入用户选择的歌曲中。然而，风格迁移模型及类似方法存在若干重要局限。首先，尽管利用预先创作的音乐大大简化了生成情感音乐的挑战，但所选音乐受版权保护。其次，这些方法通常无法支持实时交互，而这对于生物反馈系统或任何旨在实时创作音乐以调节用户情感状态的其他系统而言都是必不可少的。

与基于学习的方法相比，基于规则的方法依赖于人工设计的函数，将情感信号映射到音乐参数。因此，这类方法通过内置情感状态如何映射到音乐参数的知识，以及关于和声、节奏和时间结构的典型预期，能够避开与基于学习的方法相关的挑战。此外，基于规则的情感音乐生成系统的设计得益于大量理论与实证研究成果。近一个世纪以来，这些研究一直在探讨音乐结构的不同方面如何促进情绪表达（Gabrielsson 和 Lindstr¨om，2010年）。例如，Wallis 等人（2011年）所描述的系统主要受 Gabrielsson 和 Lindstr¨om（2001年）的研究启发，通过以尽可能连续的方式，将七个音乐参数（例如音符密度、和声调式）映射到效价或唤醒度，算法化地生成新颖的音乐。尽管为了简化，系统没有对速度、声部进行和声部间距等若干显著的音乐参数进行映射，但当效价和唤醒度参数的设置发生变化时，参与者仍能听到音乐情绪的相应变化，说明该系统已经足够有效。最近，Ehrlich 等人（2019年）开发了一个系统，该系统循环使用 I-IV-V-I 和声进行，并根据所需的效价和唤醒度水平，调整后续音符的和声调式、速度、节奏粗糙度、总体音高和相对响度。

1.2 Emotion perception in music

聆听音乐通常是一种丰富的情绪与认知体验（Altenm¨uller 和 Schlaug，2012年）。许多实证研究旨在了解音乐能够表达哪些情绪（例如 Gabrielsson 和 Juslin（2003年）），以及哪些音乐因素会影响人们感知到的情绪表达（例如 Gabrielsson 和 Lindstr¨om（2010年））。音乐因素与情绪表达之间的关系相对直观，也得到了充分理解。速度和调式，以及在较小程度上包括音乐力度、音高范围、响度、节奏和发音方式在内的因素，都会影响人们对一首乐曲所表达情绪的感知。就前者而言，研究一再发现，听众往往会对一首音乐所表达的总体情绪形成一致判断，而年龄、音乐训练和性别等人口统计因素对这些判断的影响微乎其微（例如 Juslin 和 Laukka（2004年），不过也可参见 Koh 等人（2023年））。音乐的一个显著特点是，它往往无法可靠地传达细致区分的情绪（Juslin，1997年）。Sloboda（2000年）对此现象提出了一种解释：音乐在很大程度上是抽象且模糊的，

尽管音乐或许能够暗示不同程度的能量，或让人联想到某些姿态和动作，但这些情绪轮廓往往是由个人主观地具体化的。

综合来看，相关文献表明：（i）在很大程度上，音乐可以通过系统化调整来表达预期的情绪；（ii）情感音乐生成系统的效果在不同听众之间应当相当稳定。

1.3 AffectMachine-Classical

当前的音乐生成系统 AffectMachine-Classical 采用基于概率和规则的方法，实时生成情感（古典）音乐。对于音乐刺激的情绪效应，测量方法存在较大差异，从广泛使用的测量工具（如 Russell (1980) 的环形模型和日内瓦情绪音乐量表（GEMS；Zentner et al. (2008)）），到针对特定研究开发的定制方法（例如 Costa et al. (2000)；Lindstr¨om (2006)），不一而足。与许多现有的情感音乐系统研究（例如 Wallis et al. (2011)；Ehrlich et al. (2019)）一致，我们选择在 AffectMachine 中采用环形模型来表示情绪。在该模型中，情绪可视为二维空间中的点。第一维是唤醒度，表示情绪的强度、能量或“激活度”；第二维是效价，表示愉悦程度。例如，兴奋对应高唤醒度和高效价，而满足则对应低唤醒度和高效价。与其他情绪测量方法相比，环形模型具有若干优势。首先，为了向用户准确、细致地反馈其情绪状态，AffectMachine 生成的音乐理想情况下应能在整个情绪空间中平滑变化，因此，与 GEMS 等情绪分类模型相比，环形模型等情绪连续模型是自然之选。其次，让音乐特征随时间逐渐变化，有助于使音乐听起来更加自然。最后，环形模型具有较强的普适性，因此我们也可以利用以往研究成果：即使这些研究采用了较少见的情绪测量方法，也可以将其结果解读为唤醒度和效价。

AffectMachine 提供了一种模型，能够实时流畅地生成情感音乐，生成依据可以是手动输入或预先设定的唤醒度和效价值（例如，作为用于情绪调节的情感播放列表，或在情绪空间中的变化轨迹），也可以是用户的实时反馈或生理状态（例如，采集用户的脑电图活动并将其映射到唤醒度和效价）。因此，AffectMachine 提供了一种灵活而强大的方式，将（实时）情绪状态转化为声音，并影响听者的情绪状态。该系统可用于健康与保健应用，例如生成用于情绪调节的情感播放列表。此外，AffectMachine 还可集成到脑机接口（BCI）设备或其他能够提供生物反馈的系统中，通过神经反馈／生物反馈和聆听情感音乐，帮助用户实现预期的情绪状态。

本研究的主要贡献包括：(1) 设计了一种新颖的基于规则的情感音乐生成系统，用于创作非单调古典音乐；(2) 通过听众研究验证了所提出的系统能够表达不同的情绪。在本文下一节中，我们将介绍 AffectMachine-Classical 的功能（第 2 节）。随后，我们将介绍听众研究，并讨论研究结果及其启示（第 3 节），最后给出总体结论并提出未来研究方向。

（第 4 节）。

2 AffectMachine-Classical 系统描述

在本节中，我们将介绍 AffectMachine-Classical 的参数和设计。这是一种新颖的情感音乐生成系统，可生成古典风格的情感音乐。AffectMachine 的开发旨在将其嵌入脑机接口或神经反馈系统中，既能实时生成诱发情绪的音乐，也能使神经或生理信号（例如脑电图）驱动 音乐生成系统。也就是说，该系统既旨在诱发听众的情绪，也旨在向用户实时反馈其当前的情绪状态；在将 AffectMachine 嵌入脑机接口或神经反馈系统时，生成的音乐会反映（或将其转化为声音）听众的情绪状态。在本文中，我们将 AffectMachine 从任何嵌入式交互环境（例如脑机接口）中移除，并考察独立运行的 AffectMachine，重点关注 AffectMachine 生成传达预期情绪的音乐的效果。

该音乐生成系统使用 Python开发，以一系列唤醒度和效价状态作为输入，并将相应的一系列和声、节奏和音色参数编码为MIDI事件流作为输出。随后，MIDI事件流通过虚拟MIDI总线发送至数字音频工作站（数字音频工作站），以转换为声音。对于当前版本的AffectMachine，我们使用 Ableton数字音频工作站，因为它提供了丰富的乐器选择，并支持实时多轨录音。唤醒度和效价是范围 $[0, 1]$ 内的连续值，可以从传感器（例如脑电图）采样，也可以手动提供。所有音乐参数都会根据当前的唤醒度和效价值每小节更新一次。

开发基于规则的情感音乐生成系统，首先需要确定一组音乐参数和情感状态，然后设计将参数值映射到目标状态的函数。因此，谐和、节奏和音色参数的选择依据是以往关于这些参数对音乐情感表达影响的研究，并在接受过音乐作曲正规训练的音乐学院学生的协作下加以开发。

In 在以下小节中，我们将介绍 AffectMachine-Classical 系统的详细信息。 .

2.1 和声参数

2.1.1 调式

此前基于规则的音乐生成系统通过选择固定的和声进行（如 I-IV-V-I）来控制调式参数；少数系统还会改变和弦所取自的和声调式（例如，将每种和声调式映射到特定的效价水平，通常认为利底亚调式表达的效价最高，而洛克里亚调式或弗里吉亚调式表达的效价最低——这种逻辑更简单、也常见得多的做法，是在大调和小调之间切换。）

在AMG系统中，我们通过使用一种专门设计的概率和弦进行矩阵，以一种全新的方式控制调式；该矩阵的灵感来源于（人类创作的）古典音乐中所采用的主题与变奏曲式。音乐循环演奏一个八小节主题，

每个小节的和弦功能是固定的，但所使用的具体和弦及其概率，则由所期望的效价目标水平决定。这种方法并不直观，据我们所知，此前从未在计算机音乐生成系统中实现过。不同于以往受限于特定和声进行的系统，AMG系统极具灵活性——唯一的限制是音乐必须贯穿整个八小节主题进行。（不过请注意，大多数人类创作的音乐也遵循重复的X小节结构。）因此，这种新颖的方法能够带来更广泛的音乐可能性（以及作品更高的“趣味性”），具有明显优势。同时，通过在八小节主题中使用和弦替代来表达所期望的效价水平，并以终止式结束主题的每次循环，音乐能够实现比基于机器学习的方法中通常呈现的更高程度的音乐结构连贯性。

为了构思八小节主题，效价范围被划分为十个区域，并为每个区域编写了一组概率性和弦进行，以匹配预期的效价水平。例如，在较高的效价水平下，和弦进行采用大调创作，因为大调通常与正效价的表达相关联。随着效价降低，具有更强紧张感或不协和的和弦（例如包含减音程或小音程的和弦）出现的可能性增加。对于给定的小节（例如 1-8）和效价水平（例如0-1），有一组可选和弦，每个和弦都有一个特定的出现概率，范围为0.1到0.8。在任意给定的小节和效价水平下，通常有多个和弦（一个至五个）可供选择。

2.2 声部的音高特征

2.2.1 声部进行

声部进行指的是创作出感知上彼此独立的音乐线条（例如男高音声部、女高音声部等），并将它们组合成一首连贯作品的艺术（Huron，2001年）；它是几乎所有人类创作的复调音乐中不可或缺的组成部分。尽管声部进行十分重要，但自动生成包含多个声部或音轨的复调音乐仍是一项挑战，相关研究才刚刚开始着手应对，主要采用基于学习的生成方法（例如 Dong等人（2018年））；而这些系统中有许多要么完全没有处理声部进行，要么采用了高度简化的声部进行形式。

在 AffectMachine 系统中，我们实现了一种新颖的基于规则的音乐生成系统。该系统既借鉴了传统的声部进行规则，也运用了人类音乐家所使用的启发式规则，从而创作出具有感知上相互独立的音乐线条，并呈现出非平凡的复杂性与可变性的作品。通过将这些规则映射到不同程度的唤醒度和效价，我们还为听众识别音乐所传达的情绪提供了更多线索，并能够更细致地控制情感状态与音乐参数之间的映射关系。这是我们方法极其重要的一个方面，也是其重要优势。

AffectMachine-Classical 的开发旨在生成由四个声部组成的音乐。低音声部由弦乐组演奏，负责演奏当前和弦的根音。主旋律由女高音声部演奏，并由单簧管和马林巴担任。两个内声部均由钢琴演奏，其中男高音声部只负责演奏和声进行，女低音声部则提供和声伴奏。Instru-

相关内容将在音色参数部分中作更详细的解释。

虽然有许多原则支配着声部进行，亦即创作出在感知上彼此独立的声部（Huron，2001年），但我们选取了几条简单明了的规则，以在尽量减少令人不悦或听起来不自然的旋律线条的同时，提供足够的旋律多样性。通过声部进行逻辑确定的三个声部是男高音、女低音和女高音声部。对于主旋律，我们的首要目标是避免意外的不协和。因此，音符序列是从和弦音中随机选取的序列。对于演奏和声进行的男高音声部，我们遵循（Wallis 等人，2011）中概述的启发式方法：钢琴演奏者倾向于以尽可能接近前一个和弦的方式配置新和弦，具体体现在音程和键盘上的位置方面。我们按照公式 1 计算两个音符集合（N、N'）之间的差异度，并选择差异度最低的和弦配置，作为第一个内声部演奏的内容。

$$dissimiliarity = \sum_{i} \sum_{j} \mid N_i - N'_j \mid \forall i \in N, \forall j \in N' \tag{1}$$

对于女低音声部，我们将和弦音与级进运动规则结合起来。该规则指出，如果旋律中的下一个音符音高不同，那么音高移动应按自然音阶级进。我们将这些规则编码为转移矩阵。共有四种可能的状态或动机：-1 表示沿音阶下行一级；1 表示沿音阶上行一级；0 表示音高不变；CT 表示跳进到随机选取的和弦音（CT）。我们将唤醒度范围分为两个相等的区间，并为每个区间分别构建一个矩阵，以生成适合相应唤醒度水平的旋律。转移矩阵经过设计，使唤醒度较高时，旋律更有可能采用音阶模式，从而降低因速度和音符密度增加而导致音乐过于不协和或令人不悦的风险。

2.2.2 音高音区

近一个世纪以来，音乐心理学研究一直将音高和音区与情绪表达联系起来（Hevner，1937）；然而，自动音乐生成系统往往并未明确纳入音高。通常，较高的音高与兴奋、平静等正效价情绪相关（Collier 和 Hubbard，1998），而较低的音高则与悲伤等负效价情绪相关。

在 AffectMachine-Classical 中，最低声部的音区保持不变（为 C3）。其余声部的音区则可在由当前效价水平决定的允许范围内变化。

为了实现音高音区的变化，我们再次将效价范围划分为十个等距区域，并凭听觉调整允许音高的上下界。随着效价增加，允许音高范围的下限和上限都会逐渐提高。在最低效价区域，允许音高范围从 [C1、C5] 开始，并逐渐移动到最高效价区域的 [G3、C6]。

2.3 时间与节奏参数

2.3.1 节奏

在大多数自动生成方法中，音乐的节奏内容要么是固定的（例如使用重复模式或预先编排的节奏模板），要么音符的时值（即节奏内容）基于机器学习生成过程，而这种过程很难保证音乐的连贯性。这往往会使音乐听起来极其重复，或是对大多数听众而言相当不连贯且令人不悦。

为克服这一问题，AffectMachine-Classical 中的不同声部/声部/音轨采用不同的节奏逻辑，例如，一个声部采用概率节奏，而另一个声部采用编创的节奏。通过这种方式，我们的新方法在编创元素与概率元素之间找到了良好且美学上令人愉悦的平衡。

如上文所述，AffectMachine-Classical 旨在使用四个声部生成音乐。低音声部（弦乐组）和第一男高音声部（钢琴）采用固定的节奏型——每个小节的第一拍都会演奏。对于女高音声部（单簧管和马林巴），我们将唤醒度范围划分为三个区间：低（唤醒度 < 0.4）、中（$0.4 \geq$ 唤醒度 < 0.75）和高（唤醒度 > 0.75）。与调式的实现方式类似，对于给定的小节（例如 1-8）和唤醒度区间，有一组包含两种可能节奏型（或称“乐句”）的集合，二者出现的概率相同。节奏型在代码中表示为一个二进制值列表，用来指示每一拍（细分拍）是否对应一个音符激活度。

最后，对于女低音声部（钢琴），我们引入节奏粗糙度这一指标，用于衡量一段音乐的节奏有多不规则。节奏平滑、规则的音乐通常会被感知为具有更高的效价。在 AffectMachine-Classical 中，我们使用音符密度作为节奏粗糙度的替代指标（Wallis 等人，2011）。随着唤醒度升高，粗糙度降低，音符密度增加。当粗糙度为 0 时，每个小节由八个等长音符组成。然而，这通常会导致输出听起来过于密集，因为唤醒度较高时，速度也会较快。因此，我们将粗糙度的下限设为 0.3。

2.3.2 速度

速度，即每分钟的节拍数，决定了每个小节中的音符演奏得有多快。或者，也可以将速度理解为音符时值的衡量标准——速度越快，音符时值越短。在 AffectMachine-Classical 中，速度由与唤醒度的简单线性关系决定，范围从唤醒度 $= 0$ 到 200 bpm 的 60 bpm。

2.4 音色和响度参数

有两个参数会影响音色变化：（i）AffectMachine- Classical 的乐器配置；（ii）音符的力度，即演奏音符时所用的力量。

2.4.1 力度范围

与 Williams 等人（2017年）开发的算法作曲系统类似，我们将唤醒度较高的坐标映射为更明亮、更强劲的音色，这些音色是通过提高 MIDI 力度生成的。在 MIDI 中，力度的取值范围为 0–127。在

我们的系统中，唤醒度为 $= 0$ 时，允许使用的 MIDI 力度范围为 $[40, 70]$；随着唤醒度升高，该范围的下限和上限线性增加，至唤醒度为 $= 1$ 时达到 $[85, 115]$。我们使用该范围内的均匀分布来确定每个小节的力度。

$$Velocity = unif 40 + aro * 45, 70 + aro * 45 \tag{2}$$

2.4.2 力度变化

力度变化的模式会影响情感表达。例如，大幅度变化与恐惧相关，而小幅变化则可能体现愉悦度。

在使用该系统进行实验时，我们发现，频繁改变力度会导致输出听起来生硬、不连贯。为了在力度变化足够丰富与尽可能自然地融入这些变化之间取得平衡，我们限制了每个小节内允许的最大力度变化幅度。

2.4.3 乐器配置

系统使用了四种虚拟乐器（钢琴、弦乐组、单簧管和马林巴），以营造古典音乐风格。如前所述，最低声部由弦乐组演奏，两个内声部则由钢琴演奏。主旋律置于最高声部，由单簧管演奏。当效价较高（效价 > 0.8）时，会使用马林巴与单簧管齐奏，这是因为马林巴的音色听起来欢快（而且在系统实验过程中，我们发现马林巴与单簧管的音色很搭；单簧管在较高音高时听起来可能略显尖锐）。确定所有其他和声、节奏及音色参数后，便使用数字音频工作站（Ableton）中的乐器采样生成最终音频。

3 AffectMachine-Classical 听众研究

3.1 方法

为验证 AffectMachine-Classical 生成情感音乐的有效性，我们开展了一项听赏研究。首先，我们使用该系统从Russell(1980)的环形模型中的唤醒度—效价空间周围的不同位置生成了简短的音乐示例。随后，听众为每段摘录提供唤醒度与效价评分，以考察目标《情绪》（就唤醒度和效价而言）是否确实如预期般被听众感知。

3.1.1 参与者

这项聆听研究共有 26 名健康参与者（平均年龄 $= 22$ 岁，标准差 =4 岁），其中男性 11 名，女性 15 名。26 名参与者中有 12 名报告曾接受过既往音乐训练。所有参与者在提供书面同意前，均获悉有关聆听研究的口头和书面说明。该研究已获新加坡国立大学（NUS）机构审查委员会（IRB）批准。

3.1.2 刺激材料

AffectMachine-Classical 旨在创作能够覆盖整个效价-唤醒度平面的情感音乐。在验证研究中，音乐刺激根据效价-唤醒度平面上周围的 13 个不同点生成。这些点旨在代表该空间内不同的情绪状态，涵盖各象限的角落、中心位置以及空间的中性中心点。这些点为： {效价、唤醒度} = [{0,0}；{0,0；5}； {0,1}； {0；25}； {0；75}； {0；0}； {0；5}； {0；1}； {0；25}； {0；75}；{1,0}； {1,0；5}； {1,1}]。文献中已有先例，在验证音乐生成系统时，会选择效价-唤醒度平面上的这些点（Ehrlich 等人，2019）。

为兼顾系统的概率特性，我们从 13 个点中的每个点生成了 3 种不同的音乐刺激，共得到 39 段音乐片段。这样可以降低任何特定刺激中的伪影影响听众评分的风险，从而获得更稳健的结果。音乐刺激的平均时长为 23.6 秒。刺激根据 8 小节或 16 小节的乐句进行创作，以便音乐能够形成终止式。虽然可以生成时长固定的刺激，但这样往往会导致刺激结束得非常突然，进而可能影响听众对刺激的情绪反应。对于速度较快的刺激（例如唤醒度较高的片段），我们采用了 16 小节，因为 8 小节会使这些片段过于短暂。所有音乐刺激均以随机顺序呈现给每位参与者，以避免参与者之间出现顺序效应。本验证研究中使用的音乐刺激可在线获取：https: //katagres.com/AffectMachineClassical_stimuli。

3.1.3 实验流程

实验在安静的房间内进行，每次一名参与者，尽量减少听觉和视觉干扰。实验人员首先以口头和书面形式说明实验内容，随后参与者签署书面知情同意书，同意参加研究。在聆听研究过程中，参与者坐在电脑前，通过耳机聆听音乐刺激，音量则调整至舒适的聆听水平。

聆听任务开始前，参与者需要填写一份人口统计学问卷，其中包括年龄、以往音乐训练经历、族裔等问题。随后，参与者对自己当前的情绪状态进行评分。

音乐聆听研究以两次练习试验开始，随后以随机顺序进行39次实验试验。每听完一个刺激，参与者都需要在一种称为自我评估人偶（SAM）的视觉9点评分量表上，指出该刺激所<styleid='1'>感知到的 情绪（即他们感受到音乐所传达的情绪）。这些评分分别针对唤醒度和效价收集。简而言之，效价指情绪愉悦度的程度，而唤醒度指情绪的激活度或能量水平。对于效价，SAM量表范围从“非常不愉悦”（1）到“极其愉悦”（9）；对于唤醒度，则从“平静”（1）到“兴奋”（9）。参与者可以根据需要花费足够长的时间来完成这些评分，但每个音乐刺激只能听一次。实验总时长约为40分钟，参与者因付出的时间获得了6新加坡元（相当于4.50美元）的补偿。

3.2 结果与讨论

为了评估音乐生成系统的效能，我们分析了音乐聆听研究期间收集的用户评分。我们旨在探究：（1）系统生成的音乐能否向听众表达所需水平的效价与唤醒度；（2）感知到的效价与唤醒度是否取决于听众既往的音乐训练／知识。为此，我们在两个小节中呈现结果：（1）唤醒度与效价评分；（2）既往音乐训练对情绪评分的影响。由于样本量有限，我们不进一步分析年龄和族裔等人口统计因素。

正如在音乐情绪的听众研究中通常发现的那样，我们观察到，不同听众之间的平均效价评分和唤醒度评分存在差异。这种差异通常归因于个体在音乐偏好和训练方面的个体差异，以及听众的人口统计和文化特征（Koh等人，2023年）。为了减小听众之间的差异，我们对每位用户的感知评分进行了归一化处理（见下文公式 3和4）。此处，$Max_{Valence}$ 指可能的最大效价评分（即9），而 $Min_{valence}$ 指可能的最小效价评分（即1）。相同的 Max 和 Min 值适用于唤醒度。取值范围为 0到1 的归一化效价和归一化唤醒度评分将用于进一步分析。在本文的其余部分，归一化效价和归一化唤醒度评分将分别称为效价评分和唤醒度评分。

$$Normalized_{Valence} = \frac{Rated_{Valence}}{(Max_{Valence} – Min_{Valence})} \tag{3}$$

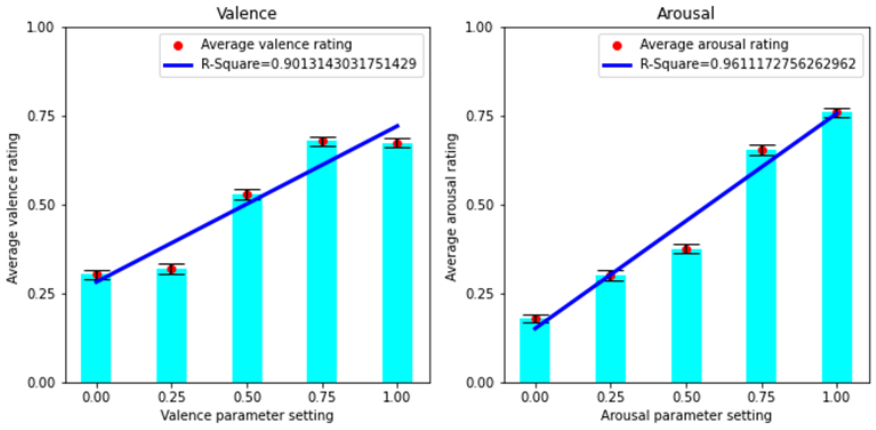
$$Normalized_{Arousal} = \frac{Rated_{Arousal}}{(Max_{Arousal} – Min_{Arousal})} \tag{4}$$

3.3 唤醒度与效价评分

为了考察 AffectMachine 是否能够通过音乐准确表达预期情绪，我们将参与者对音乐刺激的平均（归一化）情绪评分，与音乐生成过程中使用的效价或唤醒度参数设置进行了比较。例如，我们使用参数设置为 {valence, arousal} $=$ [{0,0}; {0,0.5}; {0,1}] 时生成的所有刺激的平均效价评分，来评估系统在效价设为零时的表现。图1展示了平均评分（及其标准误）的条形图。正如预期，平均效价评分和唤醒度评分均随相应参数设置呈现明显的上升趋势。就效价评分而言，我们观察到大多数评分落在 < 0.25 和 > 0.75 参数设置之间。在心理测量评分量表的中间位置，回答通常更为密集（例如，量表两端收到的回答比例较少）（Leung，2011年）。这也可能表明，听众较难区分效价参数值的极端水平。另一方面，在所有唤醒度水平上，平均唤醒度评分与相应参数值之间都呈现出更好的一致性。

为了检验平均效价和唤醒度用户评分与参数设置之间的关系，我们进行了线性回归分析（如图1所示）。效价的决定系数为 $R^2 = .90$ ($F = 27, p < 0.05$)，而 $R^2 = .96$ ($F = 74, p < 0.01$)

对于唤醒度，这证实了这两个参数在传达其所对应的情绪维度方面都非常有效。结果还显示，与效价相比，唤醒度表现出更强的线性关系（平均唤醒度评分与参数设置之间）。这一发现表明，音乐对唤醒度的表达比对效价的表达更为可靠，这一点此前也已在相关文献中得到证实（Ehrlich等人，2019年；Wallis 等人，2011年）。这些结果表明，AffectMachine-Classical 生成的音乐总体上能够向听众传达预期的效价和唤醒度水平。



(a) (b)

图1：（a）参数设置与平均效价评分之间的线性回归，（b）参数设置与平均唤醒度评分之间的线性回归。误差条表示标准误。

接下来，我们研究效价的感知是否会受到唤醒度参数设置变化的影响，反之，唤醒度的感知是否会受到效价参数设置变化的影响。为此，我们同时分析平均情绪评分对效价和唤醒度参数设置的依赖关系。图2通过展示作为情绪参数设置函数的插值平均效价评分（左）和唤醒度评分（右），将这种依赖关系可视化。图中的星号代表用于生成音乐刺激的效价和唤醒度平面周围的13个点。如左图所示，对于表达唤醒度值 <0。4 的片段，所感知到的效价低于实际效价参数设置（对于 $V > 0.7$）。也就是说，当唤醒度设置较低时，为表达高效价而生成的片段只能传达中等程度的效价。这在一定程度上可能是较慢速度的影响所致。然而，低效价设置下的评分与其相应的参数值一致。相比之下，无论效价参数设置如何变化，我们都观察到唤醒度参数值与唤醒度评分之间存在一致的对应关系。我们的研究复现了Wallis 等人（2011年）此前描述的一种现象——作者发现了唤醒度与效价之间不对称的“交叉”效应：感知效价与预期唤醒度相关，而感知唤醒度与预期效价并不显著相关

具有预期的效价。

[{"id": 0, "output": "为考察这些线性依赖关系，我们对效价和唤醒度参数设置（自变量）与平均效价/唤醒度评分（因变量）之间进行了多元线性回归分析。结果表明，感知到的效价评分显著受到效价（$F = 63, p < 0.001$）和唤醒度（$F = 11, p < 0.01$）参数设置的影响，这与我们在图2中观察到的结果一致。然而，感知到的唤醒度评分仅显示出对唤醒度设置（$F = 153, p < 0.001$）的显著依赖。这一观察结果与相关文献中的发现一致：情绪的唤醒度成分比效价成分更易于建模 Wallis 等人（2011年）；Yang 等人（2008年）。尽管如此，对于平均效价和唤醒度评分，所得到的R^2 数值均较高 $R^2 > 0.85$ 。这证实了，无论考察哪种情绪成分，多元线性回归分析中平均评分变异性的绝大部分都可以由效价和唤醒度设置值解释。"}]

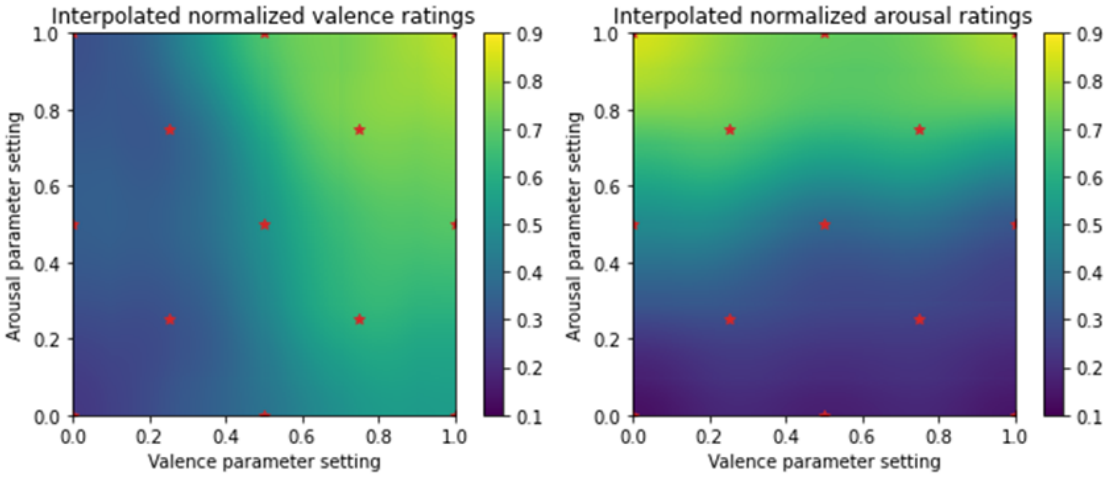


图 2：效价和唤醒度评分随效价和唤醒度参数变化的平均值（插值）。

总之，听众研究验证了 AffectMachine-Classical 能够生成表达特定情绪水平的音乐，这些水平以唤醒度和效价衡量。这证实了该系统有望应用于受益于情感音乐的场景——例如，AffectMachine-Classical 可以与生物反馈系统集成，利用用户的神经信号（或其他生理信号）驱动音乐，以反映其情绪状态。这一方向为开发更复杂的情绪调节系统带来了希望，并可应用于医疗保健领域（Agres 等，2021年）。下一节中，我们将分析参与者既往音乐训练对情绪评分的影响。

3.4 既往音乐训练对情绪评分的影响

本节将比较受过和未受过既往音乐训练的参与者所提供的用户评分。参与者在填写的人口统计问卷中说明了自己是否接受过既往音乐训练。根据参与者对“您目前会演奏乐器（包括声部）吗？”这一问题的回答，他们被分为两组——受过音乐训练组（MT）和无音乐训练组（NMT）。MT 组和 NMT 组分别有 12 名和 14 名参与者。

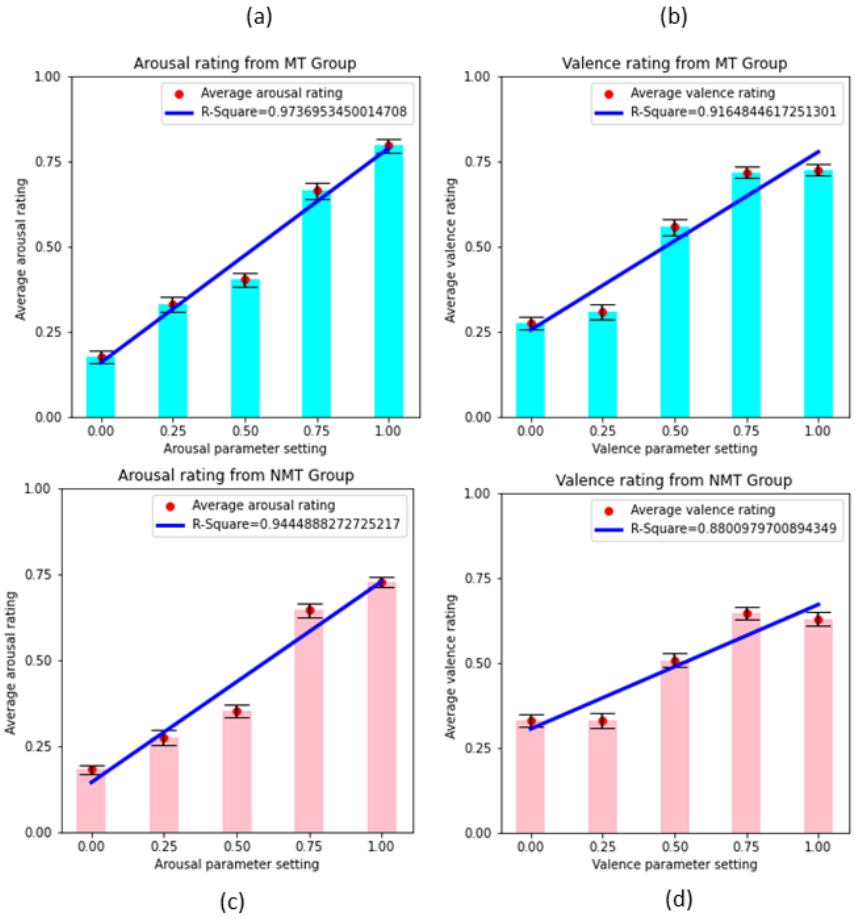


图3：（a）受过音乐训练组的平均唤醒度评分及线性回归，（b）受过音乐训练组的平均效价评分及线性回归，（c）未受过音乐训练组的平均唤醒度评分及线性回归，（d）未受过音乐训练组的平均效价评分及线性回归。误差棒表示标准误。

图3呈现了受过音乐训练组和未受过音乐训练组在情绪参数值不同水平下所对应的平均情绪评分。如图表所示，对于两个组而言，与效价相比，唤醒度的平均情绪评分与参数设定值之间表现出更强的一致性。如上文所述，较低（< 0.25）和较高（> 0.75）的平均效价评分呈现出饱和效应。

受过音乐训练组和未受过音乐训练组的参数设置值。图3还显示了所有案例的线性回归拟合结果。对 R^2 值的比较表明，与未受过音乐训练组相比，受过音乐训练组的情绪评分与参数设置之间存在更强的关系（受过音乐训练组的R^2 效价为.91, $F = 33, p = 0.01$； R^2 唤醒度为.97, $F = 111, p < 0.01$；未受过音乐训练组的R^2 效价为.88, $F = 22, p < 0.05$； R^2 唤醒度为.94, $F = 51, p < 0.01$），且效价和唤醒度均是如此。因此，我们可以推断，既往音乐训练/知识使受过音乐训练组能够更准确地识别情感音乐中目标表现力的水平。这一发现与先前的研究结果一致，后者强调音乐专长可能会影响对情感音乐中情绪的感知（Di Mauro 等人，2018年；Koh 等人，2023年）。然而，无论是否接受过音乐训练，所有参与者都能够可靠地感知音乐中的情绪表达，这一点从两个听者组中观察到的较高 R^2值（ > 0.85）可以明显看出。

此外，我们还进行了多元线性回归，以考察两个情绪维度中的参数设置对个体感知情绪评分的影响。对于所有情境，即两个组别在两个情绪维度上的情况，我们都获得了较高的 R^2 值（$R^2 > 0.8$）。此外，我们观察到，对于受过音乐训练组和未受过音乐训练组，感知效价均显著受到效价（$F = 89, p < 0.001$ 对于受过音乐训练组，以及 $F = 38, p < 0.001$ 对于未受过音乐训练组）和唤醒度（$F = 5.8, p < 0.05$ 对于受过音乐训练组，以及 $F = 15, p < 0.01$ 对于未受过音乐训练组）参数设置的影响。然而，在两个组别中，感知唤醒度评分仅受到唤醒度设置（$F = 216, p < 0.001$ 对于受过音乐训练组，以及 $F = 110, p < 0.001$ 对于未受过音乐训练组）的影响，而不受效价设置的影响。这些发现与我们对所有参与者（未进行任何分组）所观察到的结果相似。

4 总讨论

本文介绍了一种用于生成情感古典音乐的新型计算系统，名为AffectMachine-Classical。该系统提供了一种概率规则算法，能够灵活地实时生成情感音乐。AffectMachine的行为本质上类似于半结构化音乐即兴演奏，这与人类爵士乐演奏者的演奏方式相似：他们可能会遵循引谱所概述的基本旋律，同时即兴创作重新和声编配、和弦排列以及适当的伴奏。就我们所知，这是首个采用这种方法的情感音乐生成系统。这种方法的一个主要优势是，系统生成的音乐表现出很强的连贯性，同时具有最低限度的自相似性；对于需要较长篇幅音乐的研究及其他情境而言，这一特性可能十分有价值。AffectMachine 的开发旨在嵌入脑机接口系统中，这种方法结合神经反馈与情感音乐生成，帮助听众达到目标情绪状态。本文报告的听众研究旨在验证该系统生成情感音乐的有效性。

听众研究的结果令人鼓舞，表明唤醒度参数设置与平均唤醒度评分（$R^2 = .96$）之间，以及效价设置与平均效价评分（$R^2 = .90$）之间存在很强的关系。与唤醒度相比，目标情绪与感知-到的情绪之间的对应关系在效价方面更为温和，这与文献中的既有发现一致（例如，Ehrlich等（2019年）；Wallis等（2011年））。从听众研究的结果来看，显然该系统能够表达所需的情绪信息，并因此有望被用作情感引导工具，以调节/传递

参与者的情绪状态。我们想在此指出，尽管听众在既往音乐训练、个人与文化偏好以及人口统计特征方面存在差异，但有充分证据表明，听众确实如预期般感知到了系统所要表达的情绪，这使得 Affect Machine-Classical 成为一种非常有前景的工具，可用于创作能够可靠传达情绪的音乐。

关于未来的发展方向，如上文所述，AffectMachine 将嵌入生物反馈系统，例如脑机接口（类似于 Ehrlich 等人，2019），以支持听众进行情绪自我调节。此外，我们的系统还可用于健康促进应用，例如生成用于情绪调节的情感音乐“播放列表”。具体而言，借助灵活的音乐生成系统，用户可以预先设定一条“情绪轨迹”（例如，穿过情绪空间的路径，如二维效价—唤醒度空间），以界定其在聆听过程中音乐所具有的情绪特质。例如，如果用户希望聆听 10 分钟的音乐，帮助自己从抑郁的情绪状态转变为愉悦的情绪状态，他可以指定一条在所设定时长内从负向唤醒度／效价过渡到正向唤醒度／效价的情绪轨迹，系统则会据此创作定制化的情感音乐。这是一个典型例子，展示了我们的系统如何用于创作高度个性化的情感音乐，以应用于提升幸福感。

利益冲突声明

作者声明，本研究是在不存在任何可能被视为潜在利益冲突的商业或财务关系的情况下开展的。

作者贡献

KA 和 AD 主导了这项研究。KA 发起并监督了该项目。AD 主导了听众研究的数据收集和数据分析。PC 在 KA 的监督下主导了系统开发，并撰写了系统说明。所有作者均参与了论文撰写。

资助

本研究得到新加坡 RIE2020 高级制造与工程（AME）计划基金（编号：A20G8b0102）的资助。

Acknowledgments

我们谨此感谢作曲家 Cliff Tan 在开发 AffectMachine- Classical 过程中提出音乐创意、协助完善音乐参数，以及协助实施数字音频工作站。我们还要感谢作曲家 Wen Liang Lim 在本项目早期阶段为项目构思作出的贡献。

References

Agres, K. R., Schaefer, R. S., Volk, A., van Hooren, S., Holzapfel, A., Dalla Bella, S., 等（2021年）。音乐、计算机科学与健康：音乐技术在医疗保健和幸福感领域当前及未来作用的路线图。*Music & Science* 4, 2059204321997709 Altenm¨uller, E. 和 Schlaug, G.（2012年）。音乐、大脑与健康：探索音乐健康影响的生物学基础。音乐、健康与福祉，12–24 Bradley, M. M. 和 Lang, P. J.（1994年）。测量情绪：自我评估人形图与语义差异法。*Journal of Behavior Therapy and Experimental Psychiatry*25, 49–59 Carnovalini, F. 和 Rod`a, A.（2020年）。计算创造力与音乐生成系统：艺术现状导论。*Frontiers in Artificial Intelligence* 3, 14 Collier, W. G. 和 Hubbard, T. L.（1998年）。对音高和速度各不相同的听觉刺激的幸福、明亮度、速度和速度变化的判断。*Psychomusicology: AJournal of Research in Music Cognition* 17, 36 Costa, M., Ricci Bitti, P. E. 和 Bonfiglioli, L.（2000年）。和声音乐音程的心理内涵。*Psychology of Music* 28, 4–22 Dash, A. 和 Agres, K. R.（2023年）。基于人工智能的情感音乐生成系统：方法与挑战综述。*arXiv preprint arXiv:2301.06890*Di Mauro, M., Toffalini, E., Grassi, M. 和 Petrini, K.（2018年）。长期音乐训练对从鼓乐即兴演奏中感知情绪的影响。*Frontiers in Psychology*9, 2168 Dong, H.-W., Hsiao, W.-Y., Yang, L.-C. 和 Yang, Y.-H.（2018年）。Musegan：用于符号音乐生成与伴奏的多轨序列生成对抗网络。载于*AAAI*人工智能大会论文集。第32卷 Ehrlich, S. K., Agres, K. R., Guan, C. 和 Cheng, G.（2019年）。一种用于情绪调节的闭环音乐脑机接口。*PLOS ONE* 14, e0213516 Fancourt, D. 和 Finn, S.（2019年）。艺术在改善健康与幸福感方面的作用：证据何在？一项范围综述（世界卫生组织欧洲区域办事处）Gabrielsson, A. 和 Juslin, P. N.（2003年）。音乐中的情绪表达。（牛津大学出版社）Gabrielsson, A. 和 Lindstr¨om, E.（2001年）。音乐结构对情绪表达的影响。Gabrielsson, A. 和 Lindstr¨om, E.（2010年）。音乐情感表达中结构的作用。《音乐与情绪手册：理论、研究与应用》367400, 367–44

Herremans, D., Chuan, C.-H., and Chew, E.（2017）。音乐生成系统的功能分类法。*ACM Computing Surveys (CSUR)* 50，1–30 Hevner, K.（1937）。音乐中音高和速度的情感价值。*The American Journalof Psychology* 49，621–630 Hu, Z., Liu, Y., Chen, G., Zhong, S.-h., and Zhang, A.（2020）。让你喜爱的音乐疗愈身心：用于减轻焦虑的音乐风格迁移。载于第*28*届 *ACM* 国际多媒体会议论文集。1189–1197 Huron, D.（2001）。音调与声部：从感知原理推导声部进行规则。*Music Perception* 19，1–64 Juslin, P. N.（1997）。关于音乐演奏中感知到的表达的研究结果，能否推广到不同的反应形式? *Psychomusicology: A Journal of Research inMusic Cognition* 16，77 Juslin, P. N. and Laukka, P.（2004）。音乐情绪的表达、感知与诱发：日常聆听的综述与问卷研究。*Journal of newmusic research* 33，217–238 Koh, E. Y., Cheuk, K. W., Heung, K. Y., Agres, K. R., and Herremans, D.（2023）。Merp：一个包含情绪评分和评分者个人资料信息的音乐数据集。*Sensors* 23，382 Leung, S.-O.（2011）。对4点、5点、6点和11点李克特量表的心理测量属性及正态性的比较。*Journal of social service research* 37，412–421 Lindstr¨om, E.（2006）。旋律组织对音乐中感知结构与情绪表达的影响。*Musicae Scientiae* 10，85–117 Lonsdale, A. J. and North, A. C.（2011）。我们为什么听音乐？用途与满足分析。*British journal of psychology* 102，108–134 MacDonald, R., Kreutz, G., and Mitchell, L.（2013）。音乐、健康与福祉（牛津大学出版社）Ramirez, R., Palencia-Lefler, M., Giraldo, S., and Vamvakousis, Z.（2015）。用于治疗老年人抑郁症的音乐神经反馈。*Frontiers in neuroscience* ，354 Russell, J. A.（1980）。情感的环形模型。*Journal of personality and socialpsychology* 39，1161 Saarikallio, S.（2012）。简版情绪调节音乐量表（b-mmr）的开发与验证。*Music perception: an interdisciplinary journal* 30，97–105 Sloboda, J. A.（2000）。音乐演奏与情绪：议题与发展。*Music,mind, and science* ，220–238 Thayer, R. E., Newman, J. R., and McClain, T. M.（1994年）。情绪的自我调节：改变不良情绪、提升精力和减轻紧张感的策略。*Journal of personalityand social psychology* 67，910

Wallis, I., Ingalls, T., Campana,E.和 Goodman,J.（2011年）。一种由期望效价和唤醒度控制的基于规则的生成式音乐系统。载于第八届国际声音与音乐计算会议（*SMC*）论文集。156–157 Williams, D., Kirke, A., Miranda,E., Daly,I.,Hwang,F., Weaver,J.等（2017年）。情感智能音乐创作系统中音乐特征集的情感校准。*ACM*《应用感知汇刊》（*TAP*）14，1–13 Yang, Y.-H.,Lin, Y.-C., Su,Y.-F.和 Chen, H. H.（2008年）。一种音乐情感识别的回归方法。*IEEE*《音频、语音与语言处理汇刊》16，448–457 Zentner, M., Grandjean,D.和 Scherer,K. R.（2008年）。音乐声引发的情绪：表征、分类与测量。《情绪》8，494