**摘要部分**

计算创造力(computational creativity) 音乐元创造(meta creation)

试图证明人造系统能够实现创造行为创作出艺术品

但评估过程往往取决于调查者**主观**而非普遍客观

本文提供理论上**更系统的对音乐元创作和计算创作的评估**、

并展示最近评估人类和机器创作的其他方法（可用）

为突出多种评测工具，三种创作系统的差别：

**纯生成、包括内外反馈、能够反射和自反射**

**1、评估创作过程、生成艺术品、对观众感官、认知、情感状态的冲击**

2、在创作系统中构建反射机制，包含人类感官、认知的建模

为创造系统提供自检测机制来促进**自我反射**过程

第一种评测（创作系统外部）

更好的理解系统的性能和冲击力

将反馈机制加入系统

在这里，我们认为理解人类的**创造力**可以帮助理解计算方法，并且了解人类如何**感知创造性**系统及其输出可以作为**反馈**融入**人工代理**中，以提供一种创造如何影响观众的感觉。

第二种评测（创作系统内部）

**解释**自己的行为并创作

没有创造性、元创造系统的反馈和反映，创造性行为不可能发生

更严格的测试允许计算云创造系统从定义上更具有创造性、并可用于证明特定方法的影响和新颖性

**一、介绍**

近年来，**音乐元创造**（mume）得到了快速发展，特别是在**计算创造力**（cc）领域。

Cc定义：

计算系统的哲学、科学和工程，通过承担特定的责任，表现出无偏见的观察者认为具有创造性的行为。

Mume出现增多，但对元创造系统或人工制品的**系统评价却很少**，尤其是在经验术语中。

**因为MuMe被认为是CC的一个分支**。在这方面的研究应遵循更大的CC社区建立的经验和评估标准，这反过来将有助于促进这两个领域的进步。

**生成**模型和**创造**模型的**差异**

生成：面向给定数据和审美导向

尽管许多生成模型已经很复杂，但它们不包含**反射或评估**的元素，因此不必将其本身视为创造性系统（参见图1），例如，生成系统**依赖人类数据**并不罕见。设计师和/或使用者负责美学事务，这导致缺乏内在的评价或自我反思能力。

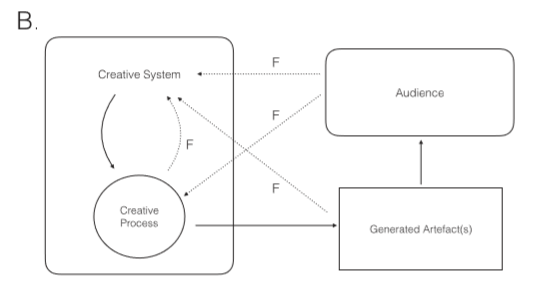
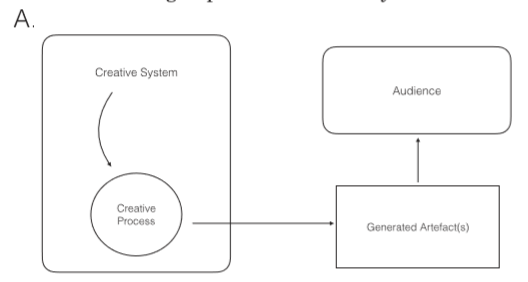
创作：自我反思

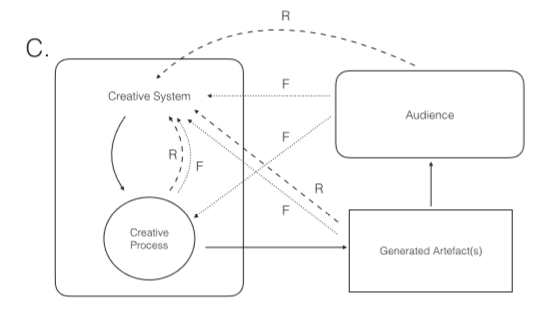
从CC的角度来考虑元创造系统可能会呈现出先前未被考虑的理论和实际应用端口，包括**基于观众如何作出反应**的模型的音乐生成。

在尝试产生一套有意义的、定量的和定性的音乐创造力测量方法时，人们可以通过研究**人类如何表现出创造性的音乐行为**来找到线索。

个体人类的创造力，以及对创造的人工制品的自我评价，取决于个体领域的专门知识、创造者所处的社会经验和社会环境，以及个体的特定目标。手头的特定受众也是一个关键因素，尤其是在计算生成音乐的背景下，因为学术和领域专家可能会关注系统及其输出的不同目标或属性

通过考虑人类创造行为和自我评价的基本过程，我们可以学习如何最终将这些反思和批判的技术构建成音乐生成系统，将创造过程中的**人文主义要素**融入音乐生成系统





为了确定人类音乐创造力和评价的基础机制，考虑来自**不同领域**的一系列经验发现，包括**行为功能学**（例如，横向和发散性思维任务）、**认知科学**（例如，知识表征和创造性行为的学习和解释机制）、**相关的理论结构**（如概念空间理论）和**神经科学**（如，在参与创造性行为的同时检查大脑反应，以评估与创造性任务有关的神经过程和大脑区域的功能组织）。

我们将讨论如何实施反馈和自我评估方法，以扩展生成模型的创造性潜力

我们假设，这一领域的最大挑战是**模拟观众/听众将感知到什么，以及他/她将如何反应**，这需要一个听众模型、反射和对此反馈和反射的动态响应。

本文所概述的方法范围决不是详尽的，但我们希望支持扩展当前用于测试计算创造力的方法（以及更严格的经验评估的理论动机），并将我们的结论扩展到mume领域。

本文的结构不仅反映了我们对**生成的事实和潜在的创造性过程的评估**，还反映了包括**反馈和反射在内的系统和系统的评估要求**。

为了提出针对这些不同目的的评估方法，我们考虑了创造性系统的三个重要方面

1. 首先是纯粹的生成部分，在这个部分中，软件系统通过一些生成过程生成一些事实，然后将这些事实呈现给观众。
2. 第二个方面是反馈，它可能包括创造性系统自身的过程或产生的输出（例如，一个mume系统可能考虑到先前产生的旋律来构成下一个旋律），或者包含观众的直接或间接的心理、认知和/或情感反应。
3. 第三个方面是通过反射进行自我评估，这可以再次基于系统自身的过程和输出，以及对这些过程和输出对听众或听众的影响的推理。

前两个方面:

可以通过实施系统外部的评估方法来解决，以评估创造性过程、生成的人工制品和听觉感知/认知/情感状态。

除了测试系统及其对受众的影响之外，这些方法还可用于在系统中构建反馈，以影响实时的创作过程。

第三个方面，反射:

可以通过将系统模型纳入其性能空间或观众某些方面的模型来解决。

因此，第4节讨论了外部评价的两个方面：

评价创造性过程的特征的方法和产生的艺术品，

测试观众对系统及其输出的感知和情感反应的方法。

然后，第5节将重点放在内部评估上:

通过提供可用于模拟受众状态的技术示例，这对于将思考构建到创造性系统中是有用的（如果不必要的话）。在深入研究这些方法之前，我们首先通过**概述创造性的理论方法**（第2节）来将我们的讨论内容具体化，并界定与MUME和CC系统共同目标相关的评估的重要性（第3节）。

小结：

音乐元创造(MuMe)

生成模型与创作模型(**人文主义要素**)

**行为功能学**、**认知科学**、**相关的理论结构**和**神经科学**

**创作=生成+反射+评价**

**系统的反馈**

**评估创造性系统（内部、外部）**

**二、创新理论视角**

阿瑟·柯斯特勒 Arthur Koestler 1964

提出了一个人类创造力的一般理论，它引入了**“双联想”**(bisociations)的概念：从先前不相关的**思维语义矩阵**中提取的思想的混合。该理论侧重于**抽象、分类和类比**，将其作为形成异类联想的一种手段。这种方法对后来的创造力模型非常有影响，尤其是**概念融合理论：**

**概念融合**提供了一个**框架和术语**，用于讨论创造性思维过程，并概述现有概念如何结合形成一种新颖的思想融合。不难想象这种方法在计算环境中是如何在现有材料的基础上开发新的**旋律、和声或节奏**内容的。

玛格丽特·博登（Margaret Boden）在其开创性著作《创造性思维》（The Creative Mind）中指出了三种不同类型的创造力，与包含所有特定类型概念的概念空间概念相比：**组合性、探索性和转化性**。

**组合创意**原则上与上述概念融合框架相似，是将现有概念要素融合形成新概念的过程，在这种过程中，组合的概念来自不同的概念空间，形成新的概念空间。例如，在一个特定的音乐类型的概念空间中，人们可以发现新的特征组合以达到艺术目标，例如，将传统的民歌旋律与蓝调或者由巴赫创作出惯用的爵士乐变奏曲，协调结合起来，这在2006年由Tzimes和Mangina证明。

**探索性创造力**包括在现有的概念空间中发现新的概念或项目。例如，在拉丁爵士或传统的西非音乐“Tzimeas and Mangina 2009”的现有节奏图式和切分结构中，发现了一种以前未被抵消的空间节奏模式。

最后，**转换创造**涉及到改变定义概念空间本身的**结构或规则，**或者改变通过这个空间的**遍历类型**。因此，变革性的创造力导致思想或范式的转变，使一套新的可能创造的人工制品成为可能。现代技术改变了我们能够产生或实现人类表演者和人工系统之间新的创造性互动的声音类型，是这类**边界转移创造力**的候选者，Wiggins和Forth 2017。Boden还提供了**心理创造力**（p-creation）与**历史创造力**（h-creativity）之间的区别，p-creativity指的是**特定个体的创造力**，h-creativity指的是**社会层面上公认为小说(?)的创造力**，Boden 2004。

Geraint Wiggins已经在他的**创作系统框架**（CSF）Wiggins 2006a，2006b]中正式化了Boden的概念空间理论。CSF是**描述和比较创新系统**的理论框架，概念化为*计算过程*，但不排除*人类组件*（Wiggins和Four2017）。因此，它提供了一个**共享的术语**，使关于创造性系统的讨论成为可能，例如，Ritchie 2007；Maher 2010]。CSF将Boden的**探索性创造力**概念**形式化**为一个**定义明确**的贯穿于部分概念体系的过程。这个搜索过程考虑了一个规则集，它指定了体系的一个子集，对应于博登的概念空间思想。该系统必须能够在内部评估它遇到的制品，其中可能包括不符合规则的发现，规定了当前的概念空间所谓的**异常概念**，可能是或可能不是有价值的。异常概念的发现可能要求系统修改其概念空间和/或遍历机制。Boden提出的**转型创造力**在CSF中被形式化为**元层面的探索性创造力**，其中概念空间是**概念空间**的概念空间，而不是**概念或人工制品**的概念空间。理解探索性创造力和转化性创造力之间的这种关系，可以得出这样一个结论：**转化性创造力必然涉及反思**：

系统必须能够根据**外部反馈或内部评估机制**对自身进行推理。通过用CSF描述MUME系统，设计者可以将特定实现的各个方面与更广泛的创造力理论联系起来。此外，形式主义有助于精确说明假设和适当的评估标准，重要的是，**使不同系统之间的结果具有可比性**。

Boden[1998]将**评价问题**作为**创新**的一个**基本方面**提出。文中指出，创造性系统的评价有两个方面：第一，从外部看，对系统本身的评价是一种**科学和/或创造性贡献**，它可能涉及创造性产出、创造性过程，或两者兼而有之；第二，系统本身**对自身产出的评估**，作为创造过程的一部分。在前一种情况下，评估可能是隐式的，例如，在启发式搜索过程中，因此识别它可能需要微妙的分析。同样，**CSF提供了一些工具来思考创造性系统中的评估**。

Wiggins等[2015]将评估表述为四个参数的函数：**人工制品、创造者、受众和上下文。**这清楚地表明了问题的一般性有多困难。然而，对于mume来说，这些参数可以被固定下来。创建者通常是一个混合系统，包括音乐程序员和他的或她的软件，观众和背景也很好理解。因此，问题变得更容易处理，如Wiggins和Foth（2017）所示。

小结：

Boden三种**创造性**：**组合创意(A+B)、探索性创造力(A+)、转换创造(钢琴变成吉他)**

CSF**创作系统框架，描述和比较创新系统**

评估表述为四个参数的函数：**人工制品、创造者、受众和上下文。**

**三、语境化评价：MuMe和音乐CC系统的兴趣领域**

现在，我们已经考虑了检验创造力的理论观点，并根据mume系统的类型（**生成性、反馈生成性或完全反思性**）激发了对不同类型评价的需求，这有助于我们在实际应用中的讨论打下基础。因此，我们提出了一些对音乐和mume中的计算创造力有重要意义的主要领域，以便为这些领域中可能应用的评估方法提供明显的应用。**下面我们讨论这些系统的共同目标**，包括**旋律、和声和节奏内容的生成**；在观众中产生**情感反应**；以及在创作过程中考虑到**表演者或观众的互动系统**。我们在这里讨论的技术适用于或与音乐创作的这些不同层次和方面相关。

**3.1 旋律与和声的产生**

**计算生成旋律和/或和声的问题**是计算机生成音乐的古老兴趣之一，可能是因为最早的计算机无法为自己生成音频，所以它们被概念化为**乐谱生成器**。已知最早的计算机生成的乐谱是弦乐四重奏的Illiac组曲（Hiller和Isaacson，1959年，1993年），其中音乐纹理是统计生成的。以生成它的计算机命名的illiac组曲纯粹是作为一件艺术品呈现的（尽管使用的方法已经解释过）。**纯粹的艺术视角意味着科学意义上的评价是不必要的：作品就是它本身，任何形式的评价都不会改变这一点。**

在这一领域的其他著名作品有cope（1987、1991、2005）和ebcioglu（1988、1990）。cope使用一种基于模板的方法，其中一个人创作作品的和声和旋律内容与另一个人的节奏结构相适应。它是通过所谓的“**音乐图灵测试**”（cope 2005）来评估的，在这种测试中，人们被邀请在一场公共音乐会上认可它作为**风格仿制品(stylistic pastiche)**。虽然这可能会给人们一个普遍认可的印记，但这种方法却充满了困难，因为它涉及面很广，因此它没有提供具体的信息来改进系统或理论，而且，**由于有证据表明人类在创造性的环境中可能会对计算机产生偏见(?)**[Moffat和Kelly 2006年），结果可能有偏差。

ebcioglu的CHORAL系统是“**传统人工智能**”的原型，其中**高度详细的规则系统与专门设计的智能回溯语言相结合**，以处理**高度复杂的和声搜索空间**（在本例中，以巴赫的风格）。Phon Amnuaisuk et al2006]和Hornel and Menzel（1998）也专注于巴洛克和声，Formerechoring ebcioglu（1988）则专注于搜索控制，后者则采用尖端人工神经网络技术。***巴赫之所以引起如此强烈的兴趣，部分原因在于他的风格在音乐学文献中得到了大量的研究和阐释***。因此，这种风格产生的和谐相对容易评估，至少作为一个**风格仿制品生成器**，因为有一个好的模型存在与之比较。绝大多数音乐风格并非如此，Spearce和Wiggins[2007]在考虑巴洛克合唱风格的旋律生成时更不寻常。同样，对音乐流派的**良好的音乐理论理解**使评价成为可能，并在本文献中独树一帜，通过一个版本的一致性评价技术【Amabile 1996b】，对该体系进行了具体的改进，在下文第4.2节中进行了讨论。最近，Brown等人[2015]开发了一系列**基于偏好规则和音乐认知统计模型的旋律生成算法**。

另一个理论化很好的领域是**爵士乐**，**既有旋律又有和声**。许多研究者已经研究过爵士乐旋律和和声，在后者的情况下，有一个很好的模式来工作[例如，Pachet 1990；Biles 1994，Steedman 1996Papadopoulos和Wiggins 1998；Lewis 2000；Pachet 2002]。

罗伯特·罗（Robert Rowe）2001是一位目标迥异的成功研究员。Rowe的密码系统始于20世纪80年代末，目前仍在使用中，它是一种**表现环境**，专门用于从“乐谱”中实时计算生成音乐，但由**乐器表现**驱动，通过MIDI（乐器数字接口）连接到计算机。在这部作品中，计算机生成的和声、旋律和节奏结构由**人类作曲家定义**，并根据**人类表演者的刺激而产生**。因此，很难在**机器创作**上划出一条清晰的界线，因为总体效果是人类（作曲家和表演者）和计算机的混合体。

尽管爵士乐和巴赫有例外，但强有力的音乐理论为基于风格的半客观评价提供了基础，

普遍缺乏强有力的模型来比较音乐输出，因为要完全做到这一点，需要解决许多开放的音乐感知问题，和/或涉及比通常情况下更详细的音乐理论解释：

Chuan和Chew[2011]通过参考Temperley和Sleator[1999]的谐波分析仪和信息理论措施对其谐波序列发生器进行评估，但这是例外，并非规则。另一个值得注意的例子是Burnett等人[2012]的**谐波序列发生器**，

它还提供了一个有用的论述，关于**如何区分计算机生成与人类生成**的例子，评价mume的创造力（通过重新构建有关计算机系统创造性程度的问题）。电子舞蹈音乐的产生在这方面也特别有趣，因为其风格本身是高度自动化的制作技术的结果，并为当代文化实践的评估提供了潜力，这可以通过听力测试、在线问卷调查，甚至通过秘密地将计算机生成的作文发布到公共领域，以获得定性和定量的反馈（例如，Collins 2008；Anderson等人2013）。然而，一般来说，**比较器确实存在**，它们只适用于特定的文体风格，然后以或多或少非正式的方式[例如，biles 1994:Phon Amnuaisuk等人1999；cope 2005]。这使得对乌梅的科学研究处于一个困难的位置，而不仅仅是纯技术的研究。**我们的解决方案是，超越音乐学和文体学，将目光投向支持听力和音乐认知的心理学理论**[Wiggins 2009，2012a，2012c2012d]。因此，我们声称，我们可能能够找到帮助我们评估mume的一般方法，不仅是在**文体生成**的背景下，而且在更普遍的**创造性术语**中。

**3.2** **节奏生成**

有两种解释节奏生成的称谓的方法：**音乐组织的一般时间方面**，如音程和节奏，以及**打击乐器部分的特定生成**，乐谱或音轨在音乐结构中提供有节奏的伴奏。现存的绝大多数音乐创作作品都属于前者。这两个类别中的生成可能与组合结构和/或表现性能时间有关。值得注意的是，有一个对**节奏和节奏的心理学和神经科学**的研究感兴趣的群体，他们有自己的网站；这些文献为有兴趣分享他们的创造责任的MUME从业者提供了丰富的机会。

在Mume中，**Meter和Rhythm**是非常**开放**的问题。虽然在极端的计算机性能（特别是非常快速和/或非常精确的播放）中有一定的艺术潜力，这些影响只能在音乐方面起到如此大的作用，而且一段时间以来，这些影响一直是更广泛的音乐探索的主题，例如，南沙罗的钢琴作品可以追溯到20世纪40年代末，或者更近一段时间，是20世纪70年代初穆格音序器的出现。**要使mume的发展超越这些有限的表达手段，就需要更多的理论发展**。也许，问题在于音程和节奏是人类音乐行为的基础，我们甚至没有注意到我们在做这些；这一建议得到了约翰·凯兹（John Cains）随机音乐的成功支持，在这种音乐中，很容易听到节奏，尽管，由于声音是随机的，客观上并不存在，一种被称为主观节奏的现象[Bolton 1894]。

然而，音乐理论家们对音程和节奏进行了广泛的研究。伦敦观点的前提是，Meter是一个认知时间框架，在这个框架内，Rhythm发生。

有许多例子表明，生成系统特别关注**节奏**，尽管它们不太具体地考虑**节奏创造性和心理方面的韵律感知**。

除了风格复制，Kaliakatsos-Papakostas等人[2012]研究了节奏生成领域中L-系统的质量，产生了经典配方的一种模式。。。结合时间知觉的心理学模型，L系统也被用于产生表达时间【Patricio和Honing，2014年】。

多智能体系统范式也被用来生成有节奏的序列，通常是由这些系统产生紧急行为的潜力所驱动。。。。.在一个类似游戏的场景中，因子通过尝试满足**共同的高级别目标**来生成“连贯的音乐”，从而共同开发新的音乐。

**3.3 在听众中创造情感反应**

大多数作者都认为音乐创作和表演的**主要目的**是在**听众中获得情感反应**。一些人进一步指出，**作曲家的目标是让听众自己的大脑状态与自己的大脑状态保持一致**【Bharucha等人2009年】。同样，据报道，听音乐的一个主要原因是Alteror放大了情绪[Denora2000]

音乐情感是一个极其困难的领域，因为它具有深刻的**主观性**。有一个方面似乎是大家普遍认同的，那就是人类对音乐的偏爱，这种音乐听起来像是由人类演奏的（除非非人类的特征，如极端的精确性或速度，是审美的一部分），这将是对Mume的一个富有成果的研究领域：在伦康音乐渲染大赛中，有一个传统可以借鉴。**在人类听众中实现特定的情感反应对人类专家作曲家来说是一个挑战**，更不用说在mume中使用的算法了。尽管如此，一些研究人员还是试图根据算法生成特定情绪的音乐映射，或者基于可操纵的心理唤醒和配价水平的音乐：然而，这些案例中的许多缺乏听众Wallis等人2008的经验评估。

作为这一领域缺乏对人类严格测试的一个显著例外，一些研究人员在CC和情感计算的交叉点工作，值得注意的是，Kristine Monteith和Dan Ventura及其同事们对计算机音乐的产生和评估都进行了讨论，这些音乐旨在引发特定的情绪或诱发特定的生理反应[Monteith等人2013年]，例如呼吸和心率的变化。例如，一个系统使用代表不同情绪的音乐语料库的统计分布生成与目标情绪匹配的音乐（Monteith等人2010）。其他工作的目的是探索计算生成的声带如何引导或增强同时讲故事中的感知情感【Monteith等人2011年；Rubin和Agrawala 2014年】。**在情感反应方面，本文着重探讨了如何利用Mume系统不仅可以唤起听众的某些情绪，而且可以利用听众的主观和生理反应作为对生成系统的反馈。**

**3.4 交互系统：包括作曲家、表演者和观众反馈**

**音乐算法的行为**可以通过四种方式受到**人类互动者**的影响：通过代码或图形用户界面，**直接参与**合成或计算级的算法过程；**仪器控制，**如MIDI或OSC（开放式声音控制）；**明确反馈**，例如，人对音乐表示喜欢或不喜欢；例如，通过**内隐反馈**，通过测量人体的生理反应，如心率，来确定对音乐的唤醒。通过考虑这些人类的**智力、情感、心理和生理因素，精神和身体状态**可以被模拟并纳入音乐元创造系统中。

如第3.1节所述，在作曲实践中使用计算机作为工具是生成音乐的**最古老形式**之一。**计算机辅助作曲**的领域主要集中在**操纵和生成乐谱**上，尽管并非唯一。这个术语可以说适用于**抽象音乐和作曲思想的符号表示的处理**，这些抽象音乐和作曲思想可能包括感性或心理动机的技术。。。**开放式音乐体系**及其前身**“补丁作品”**是文学史上的一个显著例子，其长期的积极发展和在当代主要音乐作曲家中的广泛采用都是如此。。。**这些语言使作曲程序员能够更灵活地定义和控制作曲思维中涉及的计算过程，同时也可以考虑互动和表演维度，进一步冲破构图、即兴表演和现场表演之间的界限。**

在这些系统以及其他类似的系统中，可以将**反馈**明确地引入到音乐生成算法中，例如，通过屏幕控制，或通过适当的传感设备的输入进行隐式引入。因此，也许，Mume需要考虑如何在这种情况下**分担创造性的责任**【Colton和Wiggins，2012年】：而不是控制电脑，也许音乐家可以提出建议。

由于无法收集必要的响应数据，**观众响应的干扰性能**问题更大。例如，有人试图让计算性能对彩色卡片给出的观众反馈做出反应。然而，有一种趋势，这种反应平均出整个人类的主观性，因此很少得到。。。这个实验室是为音乐感知研究而设计的。如果安装观众所需的延长准备时间不受限制，这种资源可用于艺术目的，产生巨大的效果。

考虑到反馈的一般问题，提出了表示问题，以及与反馈数据相关的含义（在音乐术语中）。Livingstone等人[2005]概述了一个**“情感表现框架”**，它提供了一个**听众态度模型**和一种**用情感元数据注释作品**的方法。这项工作的目的是**通过结合预期情感表达的知识，使计算机音乐表演更现实和表达**；然而，也可以采用类似的方法，**让表演者或观众在更高层次的音乐概念化中对影响生成系统进行反馈。**

**我们得出的结论是，虽然现有的工具控制算法的方法太多了，但在mume中，其他更微妙的人类控制大多不存在，而且在它们存在的地方，它们被概念化为参数控制。在明确的计算创造环境中，似乎很少有工作能真正与计算机共享创造责任。**

小结：

共同目标：

**旋律、和声和节奏内容的生成**；在观众中产生**情感反应**；以及在创作过程中考虑到**表演者或观众的互动系统**。

1、旋律、和声：

风格仿制品

**音乐图灵测试**

**区分计算机生成与人类生成**

**超越音乐学和文体学，将目光投向支持听力和音乐认知的心理学理论**

**2、节奏生成**

**Meter和Rhythm**是非常**开放**的问题

**要使mume的发展超越这些有限的表达手段，就需要更多的理论发展**

**3、在听众中创造情感反应**

**作曲家的目标是让听众自己的大脑状态与自己的大脑状态保持一致**

**在人类听众中实现特定的情感反应对人类专家作曲家来说是一个挑战**

**在情感反应方面，本文着重探讨了如何利用Mume系统不仅可以唤起听众的某些情绪，而且可以利用听众的主观和生理反应作为对生成系统的反馈。**

**4、交互系统：包括作曲家、表演者和观众反馈**

**人类互动者**的影响：

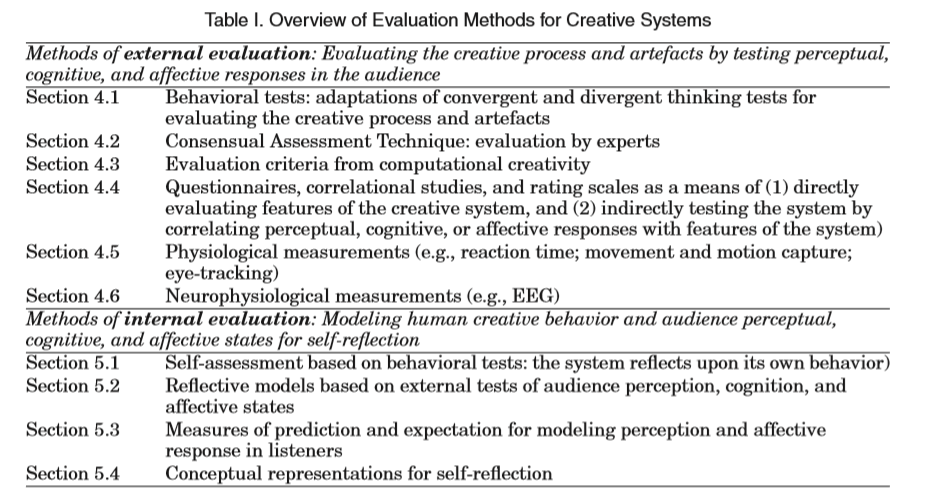
**直接参与、仪器控制、明确反馈、内隐反馈**

**“情感表现框架”**，它提供了一个**听众态度模型**和一种**用情感元数据注释作品**的方法

**在明确的计算创造环境中，似乎很少有工作能真正与计算机共享创造责任。**

**四、外部方法：通过测试观众的感知、认知和情感反应来评估创作过程和人工制品**

在本节中，我们首先讨论了**解决人类创造力**的方法，因为这些方法对CC和MUME研究具有相关性和启发性。在提供了这一背景之后，将讨论可应用于MUME研究的具体方法和应用，并参考第3节中介绍的**共同兴趣领域**。由于mume系统在历史上严重依赖于搜索者自己的主观意见来评估这些系统（这不是很严格，也不可复制），因此我们将重点放在**定量**而不是定性评估上。不可能列出所有可能使用的方法；以下突出显示的示例作为评价mume和音乐CC系统的一些**最有用方法的选择提供。**关于定性和定量方法和评估技术的更多信息，我们建议参考Mertens[2014]



值得一提的是，最近的一些研究已经致力于**构建体现人类感知和认知方面**的系统，例如**分块和听觉流分离**（参见Maxwell[2014]以获得这种综合方法的一个令人信服的例子）。尽管这种方法可能有助于评估观众的反应，我们的目标**不一定**是提倡**类人系统**（或指定任何特定的mume系统设置），而是研究人类的**感知、认知和情感反应**，作为评估系统创造性输出影响的一种手段。有关本节所讨论方法的概述以及随后有关内部评估方法的章节，请参阅表一。

重要的是，除了为研究人员测试他们的创新系统提供一种方法外，下面的方法还**可以被修改以在系统内提供反馈**（在图1 b/c中突出显示）。反馈循环的实现只需要系统对收集到的评估数据进行合并或响应。例如，来自观众的实时响应数据（在线听众评分、生理测量、动作捕捉数据等）可用于更新或影响系统内的创造性处理。基于响应数据的标准也可用于内部奖励系统，例如，是否惩罚对受众国家产生预期影响的不良创作或动态加权行为，CC中有这种方法的先例[例如，Saunders 2012]

**4.1 行为测试与人类创造力评估**

因为我们正在处理对创造力的评估，而创造力是一种人的构造，所以评估创造性过程和输出的技术应该适用于人和机器。鉴于人类测试创造力的历史悠久，我们首先**对创造力的行为测试**进行概述。这些技术可能被研究者用来测试他们的MUME系统的创造力。

创造性思维的行为测试是相当常见的，因为评估创造力在广泛的教育和专业背景中都是相关的。一套行为测试（称为“电池”）也常用于儿童入学考试、鉴定有天赋的学生、或安排到高级班或专业艺术学校。大多数行为电池测试**发散思维、收敛思维和艺术能力或依靠自我评估**。。。。所有这些评估方法都可以**应用于计算创新系统的开发或改进**。

最常见的创造力行为组合是托伦斯创造性思维测试（TTCT）【托伦斯1998年】（1974年开发，1974年、1984年、1990年和1998年重新获得认可），该测试被用于通过成人对婴儿进行测试。原始TTCT的评估方法基于Guilford（1959）模型的特点，包括**流畅性、灵活性、独创性和精化。**

重新规范时，增加了两项新措施，并消除了灵活性，因为发现灵活性与流畅性高度相关[Kim2006]。这就产生了**流畅性、独创性、精雕细琢、标题抽象性和抗过早闭合**五种测量方式。

其他各种不同思维的测试，其中一些是TTCT的前身，也经常被使用。这些测试中的大多数都是检查**个人对可能的解决方案的阐述，测量细节、数量、新颖性和各种想法**。

**聚合思维任务**以稍微不同的方式处理创造力。为了成功地完成这些任务，需要创造性思维来重新评估一个问题，确定一个有洞察力的解决问题的观点，或者改变策略来寻求正确的解决方案。就mume系统而言，聚合任务可能类似于将语音系统与表演者同步（通过使用基于节奏提示、音调信息和协调转换的各种策略）；

自我评估测试要求受试者考虑自己的创造力，下文第5.2节讨论了在系统内实施自我反思技术的情况。最后，如上所述，艺术测试通常旨在通过聘请专家意见来评估艺术品的创造力。因此，一致同意的评估技术可以被视为这种测试的一种实现；我们在下面概述这种方法。

**4.2 共识评估技术**

CSikszentmihalyi（1999）提出，创造性需要一个过程，产生新的想法或人工制品，被其他人认为是创造性的。可能最广泛应用的外部确认方法被称为**共识评估技术**（CAT）【Amabile 1982，1996a】，其中**相关领域的专家小组对艺术品、理论、产品或性能的创造力进行判断**。这种方法结合了**各种判据的批评**，而不是**基于任何一种特定的创造性理论**[Amabile 1982]。这种方法背后的论点是，评判创造力的最佳参考来自那些对该领域有深入了解的人，因为批评者可能有不同的意见，这些评估是由判据共同进行的。这就是说，这项技术已经被证明是合理的**可靠性和可重复性**跨届会议[拜尔和麦考尔2009年]。然而，CAT的一个局限性是，**不考虑创造艺术品产生的过程**。

值得在CAT和（不准确的）所谓的图灵测试评估方法之间进行比较。对后一种方法有广泛的解释。例如，cope 2005]声称他的系统EMI（音乐智能实验）所产生的音乐在被人类观众接收方面（尽管不是直接的猜测游戏）取得了成功。但是，还有一些要点需要说明。

首先，在Alan Turing[1950]的思想实验中讨论的模仿**游戏**，要求玩家直接比较隐藏的人和隐藏的计算机，在一个任务中，所有健康的人都会有很高的表现水平：一般对话。无论出于什么原因，音乐创作都不是这样一项任务；**这使得大多数玩家没有能力对“游戏”做出明智的判断**。第二，在图灵的模拟游戏中，**隐藏的人类试图欺骗玩家做出错误的选择**；据我们所知，这在这种研究中没有尝试过，而且还不清楚实验者会怎么做。

另一方面，CAT需要专家进行比较，首先解决这些差异。因为它只专注于一个**创造性的系统**，它没有解决**欺骗的企图**。相反，更有用的是，在科学的背景下，它需要**定性的反馈**，可以用来增强所讨论的创造性系统，皮尔斯和威金斯[2007]所做。

**4.3 计算创造力的扩展**

第4.1节中列出的一些行为方法和策略已被用于评估计算创造力。例如，如上所述，一个版本的CAT已经成功地用于评估生成音乐的创造力【Pearce和Wiggins，2007年】，这一方法也**有助于消除判据对人工系统生成的人工制品的偏见**。

当无法进行专家评估时，非专家（以及研究人员）可以使用一套标准来评估潜在创造性系统的输出。在CC领域，Ritchie[2007]提出了一套**类似于人类创造力评估中使用的评估指标的标准**。他的经验标准基于**新颖性**（不典型性、创新性或阶级成员）和**质量**（价值等级），允许创造性人工制品主观判断的变化，以及用于定义创造力本身的标准的变化[Ritchie 2001页]。请注意，他也是**关注生成的人工制品**，而不是**评估创建该人工制品的内部过程**。

**一个评估创造力的框架，既关注生成的人工制品，也关注系统的创造性行为**，是科尔顿的创造性三脚架[Colton 2008]。科尔顿强调，**创造性过程的知识**不仅影响观察者对**创造性的判断、人类创造力的价值判断**，而且还影响**计算系统的上下文**。有了这些知识，**创造性的三脚架**被作为**描述和评估创造性系统行为**的一种技术提供，它是基于**评估技能**（技术能力）、**欣赏**（在领域中有价值）和**想象**（超越了过去的适当新颖性）【Colton 2008】。科尔顿认为，如果这个系统被认为是**熟练**的、**有鉴赏力**的和**富有想象力**的，那么这个软件应该被认为是有创造力的。科尔顿和他的同事们还致力于通过提供两个描述模型作为起点，开始将计算创造力理论形式化：一个模型侧重于**创造性生成**的行为，称为FACE（**框架、美学、概念、表达**）(Framing, Aesthetic, Concept, Expression)。另一个**测试创造性系统可能对观察者产生的影响**，称为IDEA（**迭代开发-执行-欣赏**）(Iterative Development-Execution-Appreciation)【Colton等人2011年]。FACE框架可以用来**描述mume模型中旋律/和声产生和节奏产生的基本过程**，IDEA方法可以用来**融合观众的背景知识和偏好**。

**4.4 问卷、相关研究和评分量表**

**调查和问卷调查**是研究者检验听者**对系统或系统输出的主观反应**的一种有价值的方法。这种类型的评估可以基于对事实信息的收集（例如，收集关于听众年龄或音乐训练年限的数据，这样系统就可以对当前听众做出反应并为其特定的表现）或听众的主观反应。

向听众提出的问题可以被选择来解决特定的感兴趣的研究问题，也可以基于上面的行为和心理测量测试，或者勾勒出一套创意系统的关键组成部分。(Jordanous [2012]) 因此，使用**离散响应标准**（例如，要求听者选择多个选项中的一个，如选择多个问题）或**评分量表**进行调查可能更容易实现，以便向创意系统提供实时反馈。讨论现场表演环境中的调查和创意输出评估

或者，对定量问题的回答非常适合用来检验所产生的人工制品的特性和它们对听众的影响之间的相关性。相关研究使用**统计分析来评估两个变量之间的关系**。关系是以方向（正相关或负相关）和强度（例如，0和1之间的相关系数）来衡量的。尽管相关研究**不能提供因果关系的证据**，但它们可以成为一个有价值的指标，并**激发能够检验因果关系的实证研究**。计算创造力中一种有用的技术是将创造性系统或创造性人工制品的特性与观察者的行为反应联系起来。。。虽然这些都是主观测试，但如果使用得当，评分量表可以为参与者的判断提供非常有力和一致的**衡量标准**。

对于关注观察者评级的个体主观性的研究人员，可以使用第4.2节中描述的团体级方法，如**一致性评估技术**(CAT)[Amabile 1982]。通过测量一组评估的同意程度，个体主体性问题得以解决。特别是，CAT还采用了**可靠性**（评分者之间一致性）和**有效性**的衡量标准，以确保调查结果的**可靠性和可复制性**。基于这些原因，MUME研究人员可能会发现**平均反应**（在一组观察者中，尤其是专家中）对于系统和可靠地评估他们的系统非常有帮助。

**4.5 测量运动和生理反应**

生理测量可以用来捕捉心理和情绪状态的物理表现，从而使这些方法与MUME的表现系统非常相关。常用的测量方法有心率、血容量、脉搏、呼吸和皮肤电反应。另一种有用的方法是肌电图，它测量与微笑和皱眉相关的肌肉的小动作。**这些技术已被用于评估感知的紧张、刺激强度或在一系列领域中的诱发降级，并且最近也被应用于测量音乐聆听过程中观众的生理和情感反应**[Egermann等人2013年]。.. 使创造性的系统能够动态地学习、建模并对受众做出反应，最终会导致系统产生自我反射，这当然会影响计算系统的运行方式。

**4.5.1 情感捕捉**

动作捕捉是另一种**测量情绪、唤醒和具体认知状态**实时指标的方法。在这种技术中，运动传感器被放置在参与者的身体上，在执行任务时实时记录运动。

例如，这项技术已被用于音乐表演的研究，以将表演者的动作（有时在特定的音乐片段中）与观众的感知情感相关联。在mume表演环境中，研究人员可以使用这种方法来捕捉表演者或观众的动作（同样，表示情感状态），并检查不同类型或生成音乐部分之间的动作是否不同。

**4.5.2 视线追踪**

眼睛跟踪是另一种间接测量在线、实时注意、感知和认知方面的技术。眼睛跟踪可以作为**注意力、信息处理和决策以及探索性行为（通过测量图像或视觉场景周围的步态）**的一种度量。这种方法的好处是**不需要公开的响应**；当收集数据时，参与者可以以相对不受干扰的方式执行手头的任务。眼睛跟踪也可能有助于多模MUME表演，以评估**视觉注意力和观众参与度**。

**4.5.3 反应时间**

反应时间（RT）可作为**感知或认知过程的间接测量**，而不是要求参与者直接做出判断或评级。RT是指刺激出现与随后的行为反应（如按键）之间的时间。在设定MuMe的表演时，**当一个节奏或旋律模式改变时，可能会要求一个听者作出反应**，以告知研究者听者的状态。或者，如果一个共同创造者或音乐家正在与mume系统合作，那么结合rt可以产生一个更健壮的模型，以适应共同创造者的反应。

**4.6 用脑电图测量神经反应**

对于那些有兴趣将**参与者或听者的神经活动**纳入他们的系统或表现的MUME研究人员来说，**脑电图**（EEG）可能是最可行的技术。脑电图测量大脑中的电活动，就像在头皮上测量的那样。由于采用了细粒度时间测量，以及现代可负担和可运输的脑电图帽（如表情系统）的可用性，**脑电描记术非常适合音乐研究**，在现场音乐生成表演和脑计算机音乐接口系统的背景下与表演者一起使用【Grierson和Kiefer，2011年，2014年；Miranda，2006年；Rosenboom，1990年】。

最常见的脑电图技术之一是事件相关电位（ERP）分析，它测量事件（即实验刺激）后立即发生的电活动（以μV为单位的振幅和ms为单位的潜伏期）。…除了刺激的可预测性外，N1的振幅也能反映出集中注意力。

在音乐生成和创造力研究方面，ERP组件可用于**评估受众或合作者的期望机制和语义处理**。…由于分割能力可以重新影响获得的内隐知识的数量，以及个体的训练/经验，因此该指标可用于将计算分割和统计机制与人类学习和惊喜的神经特征进行比较。

研究人员还经常**检查大脑皮层广袤区域的时空活动**。在时频分析中，全球振荡活动是根据不同的活动频带来评估的。除了测量这些振荡带作为mume系统的实时反馈之外，振荡活动还被用来指导系统和共同创造者之间的创造性交互。

在创造力的神经科学中，一种方法是在**创造性与非创造性任务之间检查振荡带活动**。一个假设是α活动支持发散性思维（Martindale等人1990年），因此，当个人执行创造性任务时，应该有更多的α和更少的β波段能量。尽管这一假设得到了合理的支持[Jau\_Sovec 2000；Martindale等人1984]，但应注意的是，一些研究并未发现创造性活动与非创造性活动的α能力有这种提升（参见Dietrich和Kanso[2010]进行回顾）。总之，**振荡活动可以在实时表演期间与音乐家或共同创作者一起使用**【Rosenboom 1990；Miranda 2006】**或捕捉观众的情感状态**【Egermann等人2013】**并将其纳入音乐创作**。

**4.7 特殊方法的好处和局限性**

外部方法的选择可能取决于几个因素，包括研究者对创新系统的**目标、系统的架构，甚至预算和时间限制**。因此，考虑上述各种方法的优缺点是很有用的。对于创造性行为评估的替代方法（例如托伦斯测试）可以很容易地进行调整，以询问听众关于他们对创造性系统的感知和情感反应的直接问题。事实上，**调查问卷数据**对于mume研究人员来说是一种有吸引力的资源，因为这些数据相对**容易获取和分析**，而且**不需要昂贵的专用设备来收集数据**。如上所述，将创造性输出的特性与听众的反应联系起来，既可以帮助研究人员了解什么是对他们的系统有效的，什么不是对他们的系统有效的，也可以作为听众对系统的反馈。然而，收集行为反应（尤其是非专家的反应）的一个潜在缺点是，**要求观察者对他们的观点进行评级的行为实际上会改变观察者的观点**[Schwarz999]。由于专家的反应偏差可能会降低，因此研究人员可能会寻求获得专家对使用猫的创造性人工制品的评估，尽管这可能需要大量的时间和资金。从计算创意中提取的方法，如**创意三脚架**【Colton 2008，见第4.3节】，也可能特别有用，因为这些方法是专门为测试计算创意工艺品而设计的。

收集行为反应的另一种方法是**通过感知和认知的间接测量来评估创造力**，例如**眼睛跟踪和反应时间测量**。这些方法可以避开偏差问题，在专家评估不可行的情况下也很有用。在易用性方面的缺点是，为了实现这些生理方法，必须使用**专用设备以及用于数据分析的专用软件。**

**五、内部评价：为自我反省建立人类创造行为和观众感知、认知和情感状态的模型**

**人类感知、认知和情感状态的建模方法**及其相关的评估方法可以为开发具有自我评估能力的更复杂的创造性系统提供信息。在某种程度上，**自我评价可能被视为创造性行为的必要组成部分**。人类创造者对人工制品、创造过程或情感反应作出判断，不仅从他们自己的知识和视角，而且从他们的实践或作品如何获得的理解，这种内部评估机制可能只是**提供直接反馈的创意过程，导致艺术品的精致**。在更复杂的层面上，**自我评价可以形成反思的基础**，整个创作过程本身就成为考虑的主题。**自我反思因子**(Self-reflective agents)可能能够表现出与博登所描述的转型创造力相对应的创造性行为，根据CSF（第2节），这种创新被重新定义为**“元层面的探索性创造力”**[威金斯2006a，p.6]

反射可以基于上述方法，例如，系统可以使用基于脑电图数据收集的观众情感反应模型。但请注意，数据本身不足以进行反射：**系统必须具有对数据进行推理的方法**。或者，系统除了创造性处理之外，还可以利用观众的某些方面的模型（及其背景/背景，例如目标观众的音乐体验）来反映其创造性行为的影响。在CC中对反射主题的全面调查超出了目前的范围；然而，我们在下面讨论一些人类感知和认知建模的突出技术，这些技术可用于**使创造性系统将受众的知识纳入他们的推理和元推理过程**。我们首先讨论了如何将第4节中介绍的方法用作创造性系统中自我反思的基础。

**5.1 基于行为测试的自我评估**

上面第4.1节讨论了行为测试，如创造性思维的Torance测试，如何进行调整，使研究人员能够从外部测试他们系统的创造力。**发散性和收敛性思维测试**也可以作为自我反思的内部测试。考虑一个系统，它采用了一个内部的度量模型，比如独创性和精化，或者使用创造性力量的子尺度，参见Torrance 1998]比如情感表达。如果系统有一套这样的自我评估指标，以及通过这些标准对自身进行评估的方法，那么系统就能够将自己的行为和输出定位在一个旨在**实现艺术性和成功性**的环境中。

自我评估测试提供了另一种从人类行为测试中获得自我反思的清晰方法。**这些测试要求考生反思自己的创造力，或报告自传体信息，例如参与者参与艺术追求**（例如，参加视觉艺术课或写短篇小说的经验）。自我评估可以基于创造性过程的现有概念框架，例如Wallas[1926]提出的概念框架，包括**准备、孵化、启发和验证**的四个阶段。许多正式的自我评估测试也是基于托伦斯的工作。例如，卡泰纳·托兰斯创意感知目录卡泰纳和托兰斯1976）。CC领域的一些工作旨在促使创新系统审查和批评自己的产出【Colton等人，2014年】，但这还没有成为mume的标准做法。

**5.2 基于观众感知、认知和情感状态外部测试的反射模型**

如果创意系统对数据产生影响，可以从**感知和生理外部评价方法的数据**作为**内部重建的基础**。如果一个系统试图在听者中调用某种特定类型的情感反应，那么它应该拥有某种**人类情感反应的内部模型**（例如，**基于Valence和Arousal的二维模型**）。然后，一旦收集了来自**人类的反馈**（例如，记录了问卷评分），并就情感反应的内部模型进行分析（例如，评分量表上的平均分数表示受众的特定情感状态），系统可能会重新判断其目标是否为符合要求。不管收到的外部反馈是什么类型的，如果系统被赋予了一种**预测听者可能如何响应**的方法，那么它可以将听者的**实际响应与其预测进行比较**。

与**表现后反思(post-performance reﬂection)**不同，生理反应的实时测量可用于**持续、在线自我评估和反思**。在同一条线上，创意因子可以使用**眼睛跟踪测量**、**反应时间数据**、**心率数据和连续脑电图**来评估系统的预期目标是否已实现，以及是否（如何）应**更新创意过程**以支持未来的预期结果。

**5.3 感知和情感反应模型的预测和期望度量**

音乐和语言等序列的时间处理严重依赖于**预期和预测机制**。我们的大脑在我们学习和预测世界的环境中获取统计关系和模式【Saffran等人，1996年；Creel等人，2004年；Lew Williams和Saffran，2012年】。**因为期望对于学习是至关重要的**，并且影响进入**感性意识**的东西以及记忆中编码的东西，因此它也可能被认为是人类创造力的组成部分[Wiggins 2012b]。***听演讲时，人们不会被动地“听”单词；相反，听者会对下一步会发生什么做出含蓄的预测。***将人类认知的这一观点与计算创造力联系起来，研究人员可以考虑在他们的mume系统中使用类似类型的预测-动作反馈回路[cf.Brown和Gifford 2013]。因为这种机制使人类能够学习一个**统计框架和联系网络**（类比、脚本、模式等），指导**感知、认知和创造**。

**正确的预测是一个有价值的进化特征，因为我们不能总是依靠直接的经验来形成预测**（例如，当遇到一个不熟悉的大型猫科动物的成员时），理论家认为，人类已经对预测的确认或否定产生了情感反应，即使在音乐等领域，也没有真正的威胁【Huron和Margulis，2010年；Huron，2006年】。通过这种对预测的情感反应，我们对我们在世界上遇到的事情产生了偏好。。。**将听者期望模型纳入生成过程可能对mume系统的表现力和创造力产生强大的影响**。能够根据预期预测听众对新材料的可能反应，使系统能够明确地模拟有关特定情感反应的激发的更高层次的合成意图；例如，Huron[2006]和Egermann等人[2013]提供了潜在的灵感。

对于诸如**求新、创造和美学**等现象，**享乐偏好**通常可以用Berlyne[1970]所描述的倒U关系来描述。享乐函数规定，随着刺激强度或复杂性的增加，偏好或快乐会增加，直到WUNT曲线的顶部，之后，复杂性的增加会导致偏好的减少。从本质上讲，理论假定，**非常可预测的刺激被认为是无聊的，而极其复杂或强烈的刺激被认为是不可接近或过度刺激的。**因此，曲线的中心产生**最佳复杂性**的**“最佳点”**，这通常是刺激复杂性/强度的函数，也是观察者的背景和经验。该理论已应用于一系列领域，包括**音乐感知和创造力研究**[例如，Steckand Machotka 1975；Martindale等人1990；Saunders and Gero 2001；North and Hargreaves 2008]。

寻求将期望构建到其系统中的明确技术的研究人员可能希望求助于**信息理论**，该理论为实现上述期望和预测措施提供了工具。信息理论衡量的是传输信号中包含的信息量，并已广泛应用于天文学和语言学等领域。。。随着时间的推移，整个序列的信息理论特性对感知和记忆产生了动态影响，复杂的刺激通常会增加识别记忆性能差的影响。如上所述，期望机制在**听众对音乐的感知和体验中**发挥着重要作用。通过信息论测量和行为测试，MUME研究人员可以**量化听者的期望值**，然后将这些知识**应用到生成系统**中，以影响听者对旋律和节奏的感知，创建一个期望的情感反应，并模拟听众的心理状态。

没有提到尤尔根·施密杜伯（Jurgen Schmidhuber）的作品，特别是他关于创造力、乐趣和内在动机的形式理论，创新认知的信息论方法的总结，就不完整[Schmidhuber 2010]。这涉及到一个重要的问题，**即为什么人类喜欢有创造力**，即使这样做对他们没有任何生物学上的好处。享乐奖赏假设是从**成功的提炼**中产生的，提供了“美”的概念，尽管在我们的选择中，将美定义为可压缩性过于简单。Biedeman and Vessel[2006]提供了一种**内分泌机制**，通过这种机制，施密杜伯的理论可能得以实现。

总结**期望机制**对计算创造力的相关性**：****预测有助于大脑学习和编码一个领域的信息，而（违背）期望则与美学和影响有关。**这可以用信息理论进行建模，至少部分是这样。创造力通常涉及到发现问题或任务的新解决方案；它是探索（确定新的区域或路径）或扩展已学的心理表征空间的行为。**心理表征或概念空间**的网络可以被认为是一个复杂的先验分布，具有**统计上定义的共现性和相关特征**。在这个框架中，探索性创造力涉及到根**据已知的概率分布或概念表示生成新的想法**，而对创造艺术品（价值判断）的偏好**通常属于熟悉、新颖或可预测和复杂的首选范围内。**

**5.4 概念表示**

考虑到概念融合、关联处理以及计算创造力领域内现有想法或实例的新颖组合的重要性，有助于构建专门解决概念本身结构和性质的构思框架。虽然有几种方法可以解决这一点[Barsalou 2008；Barsalou等人2003年；Bareiss 2014年），我们将重点放在Peter G¨ardenfors的概念空间理论[G¨ardenfors 2000]上，因为它提供了一个**准形式(quasi-formal framework)的框架**，其中有明确的数学原理来支持**认知建模**。

G¨ardenfors[2000]提出了概念空间理论，作为解决认知过程建模和理解问题的代表性工具。作为一种表示，它位于一个特定的抽象层次上。他认为概念结构应该用他所说的概念层次上的几何学来表示。…因此，**概念之间的相似性是通过多维空间中点或区域之间的距离来隐式表示的**，其方式类似于Shepard提出的相似性空间视图[Shepard 1962A，1962B]。

例如，颜色的概念空间由色调、饱和度和亮度三个维度定义，所有这些都可以用几何图形来描述。…

人们可以将概念空间理论应用于**创造性生成的基础过程**，以及在创作过程中对观众或表演者的状态进行**建模和合并**。…然而，除此之外，几何学提供了一种特别直观的解释和推理手段，人类和机器都可以使用。理解可以**根据邻近度、空间区域、轨迹和转换以及几何结构**固有的其他特性进行建模[参见CSF，Wiggins 2006b]。在实际的音乐术语中，潜在的任务可以包括构建高层次概念的概念空间表示，如**类型或情绪**，它可以作为混合不同风格的基础，或者作为一个系统的组成部分，试图**产生特定的情感反应**。这些概念结构可以支持由系统设计者或系统本身对生成过程进行修改，以便能够探索以前未知的概念区域，同时重要的是保持将新发现与已知的音乐表现形式联系起来的能力。

概念空间理论在实际应用中面临的**重大挑战**是：首先，**构建这种具有几何意义的空间**；其次，**建立它们的感知有效性**。对这个主题的详细讨论超出了本综述的范围，它本身是认知科学、心理学和机器学习领域的一个主要研究领域。此外，作为一种表示理论，**概念空间理论可以与已建立的建模技术（如统计模型或深层神经网络）相结合**，以增强系统开始处理其操作的数据的意义问题的能力。

关于创造力和概念空间，可以形成一个可检验的假设，即与改变空间几何的新实例（如转化创造力）相比，位于概念空间内的新思想/实例对处理（尽管也可能被视为创造性较低）的认知要求较低。将**观众或共同创作者的概念空间模型结合**起来，可以为mume系统提供一种**自我反思的方法和思想理论**，从而增强所生成音乐的适宜性和价值。这可能被认为是一种以**人类为中心**的创造力方法，但同样，因为音乐（和音乐创造力）是一种人类构造，**人类对音乐概念化的方式可以为人工智能和机器创造力提供重要的见解。**

一些系统已经包含了从概念空间中获得灵感的方法，用于创造性生成，或者作为听众感知的模型。2014年]。…因为一个人工系统可能会产生无数可能的例子，许多不可避免的会是质量差。因此，上述系统的一个关键组成部分是通过**将系统的探索行为限定到更高价值的空间区域**来拒绝候选故事；换句话说，**概念空间的探索受到质量维度的限制**，如情感、兴趣和紧张【Le’on和Gerv’as 2010】。这项技术可以应用于mume系统，也可以用于探索有价值的新旋律、和声和节奏材料。

**六、总结和结论**

本文概述了可用于或适用于**计算创造力**研究的**理论和经验方法**，特别是针对**mume群体**。在最广泛的层面上，我们从两个角度讨论了**评估问题**。首先，从***科学的角度***，我们讨论了一系列**提供客观评价创造力的方法**。这些方法可以被研究者用来回答关于人工创造系统产生的行为或人工制品的问题。我们将这种形式的评估称为**外部评估**，*因为判断或测量的来源是来自系统本身的外部*。第二种观点是，我们从***创造性系统本身****及其****内部自我评价能力***的角度来考虑评价。

在讨论内外部评价方法的同时，我们强调了与创造力和计算创造力理论的联系，以及更广泛的主题，即如何**对创造性人工制品和产生它们的过程作出判断**，从根本上与**系统设计问题**有关。我们提供了一个简单的分类法（如图1所示），根据反馈和/或反射与系统生成能力的集成程度区分不同的系统架构。关于将内外部评价方法与创新体系目标相结合的理论论证，我们将讨论范围限定在Mume感兴趣的四个主要领域，包**括旋律和和声的产生**、**在听众中产生情感反应的节奏的产生，以及结合表演者和观众反馈**。

总之，本文提倡科学的方法研究创造力和发展的Mume系统。精确的评估方法，连同明确陈述的假设，是任何科学或科学上一致的学科进步的基本基石。我们试图证明将评估中心放在舞台上的好处，因为它不仅为个人研究人员提供了好处，也为整个社区提供了好处。此外，创造性评价的主题更普遍地为更先进的创造性系统的发展提供了见解，在这种系统中，**评价能力**成为一个基本组成部分。