arXiv:2601.04915v1 [cs.HC] 8 Jan 2026

OnomaCompass：在文字与图像之间切换的纹理探索界面

MIKI OKAMURA，日本筑波大学
SHUHEY KOYAMA，日本筑波大学LI JINGJING，日本筑波大学YOICHI OCHIAI，日本筑波大学

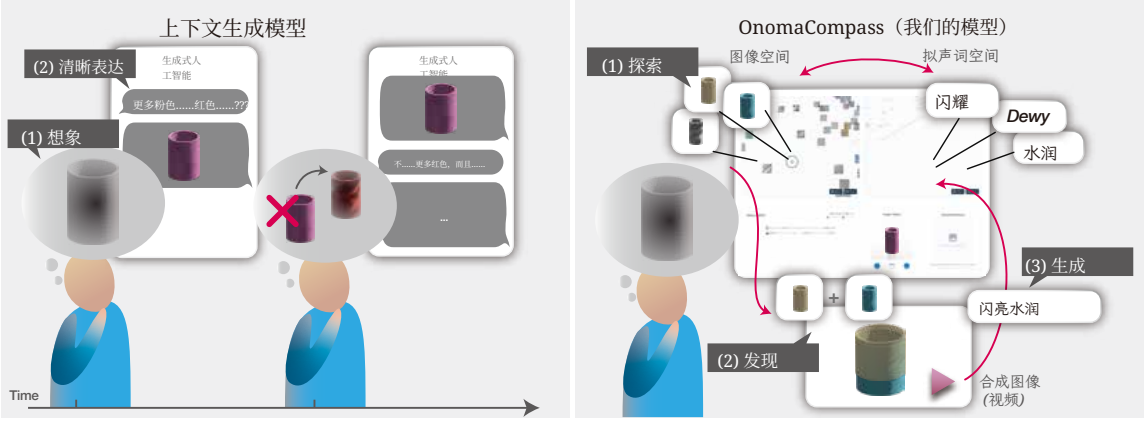

图1。以提示词为中心的生成与 OnomaCompass 的概念比较：用户无需反复将意图转化为文字（左），而是可在相互关联的纹理地图与拟声词地图之间切换（右），利用跨模态线索导航材料质感、将纹理应用于物体，并通过视频插帧和帧重新嵌入拓展探索。

人类能够细致地感知材料纹理，但要用词语表达这类身体感受，是设计构思中的认知瓶颈。我们提出了 OnomaCompass，这是一种基于网页的探索系统，将声音象征性拟声词与视觉纹理表征相连接，以支持早期阶段的材料发现。OnomaCompass 不要求用户为生成式人工智能构思精确的提示词，而是提供两个相互协调的潜空间地图——一个用于纹理图像，另一个用于拟声词——这些地图基于人工编制数据集构建，该数据集包含虚构的拟声词，以及通过 Stable Diffusion 生成的对应纹理。用户可以在两个空间中导航，触发跨模态高亮，将发现整理至图库，并通过图像编辑模型预览应用于物体的纹理。系统还支持在所选纹理之间进行视频插帧，并对提取的视频帧进行重新嵌入，从而形成涌现式探索循环。我们开展了一项包含 11 名参与者的被试内研究，将 OnomaCompass 与使用 Gemini 2.5 Flash Image（“Nano Banana”）的以提示词为中心的图像生成工作流进行了比较。OnomaCompass 显著降低了工作负荷（NASA-TLX 总体得分、心理需求、努力程度和挫折感；$p < .05$），并提高了享乐型用户体验（用户体验问卷），而可用性（系统可用性量表）则以基线为优。定性研究结果表明，OnomaCompass 能帮助用户将模糊的感官预期外化，并促进意外发现，但也揭示了空间导航方面的交互挑战。总体而言，将声音象征作为一种轻量级线索，为超越以提示词为中心的生成式人工智能的感性工学驱动的材料构思提供了一种互补方法。

CCS概念： • **以人为中心的计算** → **交互式系统与工具**； **交互范式**； 人机交互设计与评估方法； 交互技术。

作者联系方式：MIKI OKAMURA，筑波大学，日本，mikio_kamura@digitalnature.slis.tsukuba.ac.jp；SHUHEY KOYAMA，筑波大学，日本，shuhey@digitalnature.slis.tsukuba.ac.jp；LI JINGJING，筑波大学，日本，li@digitalnature.slis.tsukuba.ac.jp；YOICHI OCHIAI，筑波大学，日本，wizard@slis.tsukuba.ac.jp。

其他关键词和短语：感知增强、创造力支持工具、声音象征、潜在空间探索、拟声词

1 引言

1.1 创造力的新前沿：设计“材料质感”

数字技术的进步拓展了创作的主体范围，使其从专家扩展至普通大众，开启了人人都能成为创作者的时代 [35]。在这一更广泛的趋势中，人们对创造力的关注开始从单纯设计形态转向设计材料质感，而材料质感决定了产品给人的印象和触感。本研究明确界定了“材料质感”一词的使用范围。在三维计算机图形学和工业设计等专业领域中，“纹理”通常指应用于表面的二维图像，而“材料”则常指一组光学参数。相比之下，为了贴近日常生活中缺乏专业知识的新手所进行的意义建构，本研究采用了更广义的材料质感概念。具体而言，“材料质感”指人类感知到的整体视觉与触觉印象，它综合了物质固有的光学特性，即材料外观，以及表面几何形态；后者包括细微的不规则形状和运动，例如羽毛的聚集形态或凝胶的起伏。对于这一新兴设计领域，尤其是对于非专业的新手而言，目前尚未建立起能够引导他们从模糊的感官意象迈向具体设计概念的方法。本研究聚焦于创作过程中最前端的构思阶段。

1.2 生成式人工智能中的“语言表达障碍”

近期的文生图系统为应对这一挑战提供了有力的解决方案。然而，主流交互范式本质上依赖于用户的语言表达能力 [10]。当代与人工智能进行的创作对话通常以轮次推进：用户必须将内心的图像转化为离散的词语，即“提示词”，人工智能再对这些提示词进行解读并生成“图像” [27]。当用户拥有明确的语言词汇时，这一过程虽有效，但在处理难以用语言表达的感官图像时，便会成为瓶颈。有研究指出，这种语言表达障碍会减少意外发现的机会，并限制创作潜力。

1.3 在词语与图像之间“穿梭”：一场连接两个世界的对话

为应对这一挑战，本文提出一种新的交互范式，将人机对话从单向的指令—响应模式转变为一种对话式过程，在词语（概念）与图像（知觉）两个世界之间双向穿梭。这一构想源于认知过程：思维通过在具体知觉与抽象概念化之间不断转换而深化。所实现的系统“OnomaCompass”构建了两个不同的表征空间：一个捕捉视觉特征的图像空间 [21]，以及一个捕捉语义相关性的语言空间。通过在这两个空间之间进行交互式穿梭，用户可以借助视觉发现与概念理解，逐步将模糊的图像具体化。

1.4 拟声词：通往感官世界与人工智能视野的桥梁

作为在这些空间之间穿梭的桥梁，本研究聚焦于日语拟声词，尤其是拟态词。在更广泛的语言学语境中，这类词被称为象声词 [4]。本研究采用的方法是以拟声词为起点，指导基于人工智能的图像生成。生成的图像不仅包括材料表面的特写，也包括能够象征性地表达特定材料性质的具象主题（例如，一个

用于表达“ほくほく”的红薯）。这一观察反映出一个耐人寻味的特征：在将材料质感的语言概念视觉化时，人工智能经常会描绘具有目标材料质感的典型物体。本研究将这类具象主题视为传达材料质感的有效视觉表达，并将其纳入数据集。

1.5 提案与贡献

基于这些理念，本文提出了“OnomaCompass”，这是一种新型界面，通过探索连续的潜在空间，支持感官材料质感设计。不同于以往主要面向专家分析与控制而设计的潜在空间探索界面，该系统面向从事创意活动的新手，并回应他们难以用语言表达的感官需求。为此，我们实现了一种对话模型，用于连接不同模态的潜在空间，并支持在这些空间之间穿梭；多模态人工智能的最新进展使这一能力成为可能。本文的贡献有三点。第一，为一种新的人机交互范式提供概念验证：该范式在人机创意对话中，从离散的语言指令转向连续的空间穿梭。第二，将这一范式落实为一个具体系统，借助拟声词连接两个潜在空间，并执行动态视频插值。第三，对 11 位参与者进行的半结构化访谈表明，所提出的方法带来了性质独特的价值——即意外发现的乐趣，以及摆脱语言表达压力的轻松感；这一点对新手在发散思维阶段尤为重要。本文其余部分安排如下：第 2 节回顾相关工作，第 3 节详述研究方法，第 4 节报告评估结果，第 5 节讨论研究发现，第 6 节总结并展望未来方向。

2 相关工作

本研究位于三个领域的交汇处：人机创意对话、潜在空间探索，以及拟声词的认知科学。本节回顾这些领域的既有研究，以阐明本文所要解决的问题及其方法的原创性。

2.1 创意对话中的离散性与连续性

所提出交互的核心，是从离散性转向连续性。

2.1.1 文本提示词面临的挑战：语义鸿沟与离散式试错。与当前图像生成式人工智能的交互主要通过文本提示词进行。尽管这种方式功能强大，但用户连续、感官化的意象与系统所需的离散语言表达之间存在众所周知的语义鸿沟，这一问题早已在信息检索领域得到确认 [10]。用户往往不得不反复修改提示词并重新生成结果，以弥合这一鸿沟 [27]。这一过程可能带来较高的认知负荷，尤其是对缺乏明确词汇的初学者而言，并可能打断创造性思维的流动。

为应对这些挑战，近期的系统引入了交互式支持，利用大语言模型（LLMs）或注意力可视化机制来辅助提示词工程 [6, 38]。研究者还构建了大规模提示词数据集，以帮助理解模型如何响应不同的语言输入 [40]。值得注意的是，PromptMap [2]提出了一种替代性交互方式，用户可以通过带有语义缩放功能的地图界面，自由探索大量合成提示词。尽管基于文本的辅助取得了这些进展，但对语言描述的依赖仍然是纯感官或抽象构思的一大障碍。

2.1.2 连续界面的有效性。人机交互研究表明，与离散值输入相比，连续参数操控（如动态查询）能显著提升探索性任务的表现 [3]。然而，仅仅

增加控制维度而不进行语义组织，会加重认知负荷 [8]。因此，有效的创意探索需要界面能够引导用户在经过语义结构化的潜在空间中进行直观的视觉导航，而不是单纯地暴露原始参数。

2.2 探索感知世界：潜在空间的探索

本研究延续了这样一条研究路线：为交互式探索由人工智能[12, 21]学习得到的潜在空间开发界面。

2.2.1 统一表征的基础。 近期进展使潜在空间超越了单一模态。研究者已经建立统一表征空间，利用语言这一连接元素 [11, 29, 43]对齐多种模态，例如视频、音频、深度和热数据。这类多模态空间支持更丰富的语义对齐，从而能够对不同数据类型进行综合分析 [24]。

2.2.2 可视化与操作工具。 为了在这些高维空间中进行导航，人们开发了各种可视化工具。其中包括用于映射和比较语义维度 [28, 39],的可扩展可视分析工具，以及通过降维浏览大型图像数据集或文化藏品的浏览器 [13, 31]。在视频领域，已有研究表明，将视频帧投影到可导航的潜在空间中有助于开展头脑风暴 [26]。

在操控方面，研究主要聚焦于发现这些空间中的可控方向。一些方法通过主成分分析或线性子空间，在生成对抗网络的潜在空间中识别可解释的方向（例如姿态、年龄） [16, 34]。近年来，文本引导操控技术使人们能够使用自然语言进行稳健的零样本编辑[5, 22, 23, 44]。

2.2.3 面向构思的界面。 虽然上述工具通常着眼于精确控制或分析，但人们对专为创意构思设计的界面越来越感兴趣。近期的系统通过促进视觉参考与关键词的重新组合，助力创建情绪板 [7, 37], ；也有系统允许用户浏览高层次的语义概念，以弥合隐性用户意图与生成之间的差距 [1, 25]。然而，专门以感官语言为基础、帮助新手探索材料外观的界面仍未得到充分研究。

2.3 通过拟声词连接感官

为探索这一尚未开拓的领域，本研究聚焦于日语拟声词的声音象征，尤其是拟态词。声音象征指的是词语的声音与其意义之间非任意（直觉上相联系）的关系。

2.3.1 直觉理解与认知负荷降低。这种声音与意义之间直观的联系 [30] 表明，拟声词的认知负荷较低，容易学习。“声音象征引导假说”[17] 认为，这种敏感性是生物本能赋予的，并在语言习得和词汇表征的建立过程中发挥关键作用。研究表明，声音与意义之间的映射有助于词语学习 [18],，而这种直觉性可以跨越语言障碍 [19]。这些发现有力地支持将拟声词视为适合新手使用的有效且直观的界面元素。

2.3.2 与物质性紧密相关。 拟声词与多感官知觉密切相关 [36]。在知觉视觉物质性的过程中，研究表明，拟声词能够捕捉湿润度、蓬松度和光滑度等基本感知维度 [15]。计算方法也尝试利用这种联系，例如，计算机视觉

系统通过将音素映射到纹理类别，利用声音象征性词语来表达纹理 [42]。功能性磁共振成像研究表明，在处理不同领域的日语拟声拟态词（如运动和形状）时，右侧后部上颞沟会持续激活，这反映了这些词兼具语言符号和象似符号的作用[20]。这些研究为将拟声词用于探索材料外观的任务提供了科学依据。

2.4 生成式人工智能与材料质感的创造

本研究也立足于人工智能驱动的材料质感生成这一背景。近期的系统（如 TactStyle[9] ）可根据图像提示词生成用于数字化制造的触觉表面属性。TactStyle 专注于忠实再现特定的材料性，而我们的方法则针对上游的构思阶段。通过利用拟声词作为感官线索，我们的方法使用户能够在多种纹理之间的概念空间中探索。借助生成式插值，所提出的系统促进了对这些过渡状态中涌现的新颖材料特性的发现。

3 研究方法

为评估所提出系统的有效性，我们实现了一个原型，并通过一项结合标准化问卷与半结构化访谈的被试内混合方法研究对其进行了评估。

3.1 系统设计与实现

本节介绍所提出的“OnomaCompass”系统的具体设计与实现。为实现第 1 节提出的理论框架，我们构建了一个基于网络的系统，使用户能够在视觉与语言潜在空间之间交互式地穿梭。

3.1.1 设计目标。为实现词语与图像之间穿梭这一概念，我们制定了四项设计目标。

DG1：降低认知负荷。构建一个系统，绕过传统图像生成式人工智能所要求的明确语言表达，让用户通过直观的视觉与空间操作开展创意探索。我们特别注重让没有专业知识的新手也能开始探索材料质感，而不必先掌握复杂的提示词工程。

DG2：促进意外发现。主动促成不受用户既有知识与词汇限制的意外发现。这种方法与基于预定义查询的传统搜索方法有着根本区别。系统通过空间布局将语义邻近性可视化，并呈现跨模态对应关系，从而营造一种让用户能够发现意想不到关联的环境。

DG3：实现连续探索体验。通过在连续空间布局中流畅导航（将二维地图呈现为可导航的三维视图），提供连续探索体验，而非离散的试错。交互设计旨在让用户在视觉表征与语言表征之间自然切换，同时不打断思路。

DG4：构建涌现循环。实现这样一个系统：用户的探索行为本身会生成新的探索目标。具体而言，开发了动态内容生成与重新绘制机制，使探索范围能够超越固定数据集。

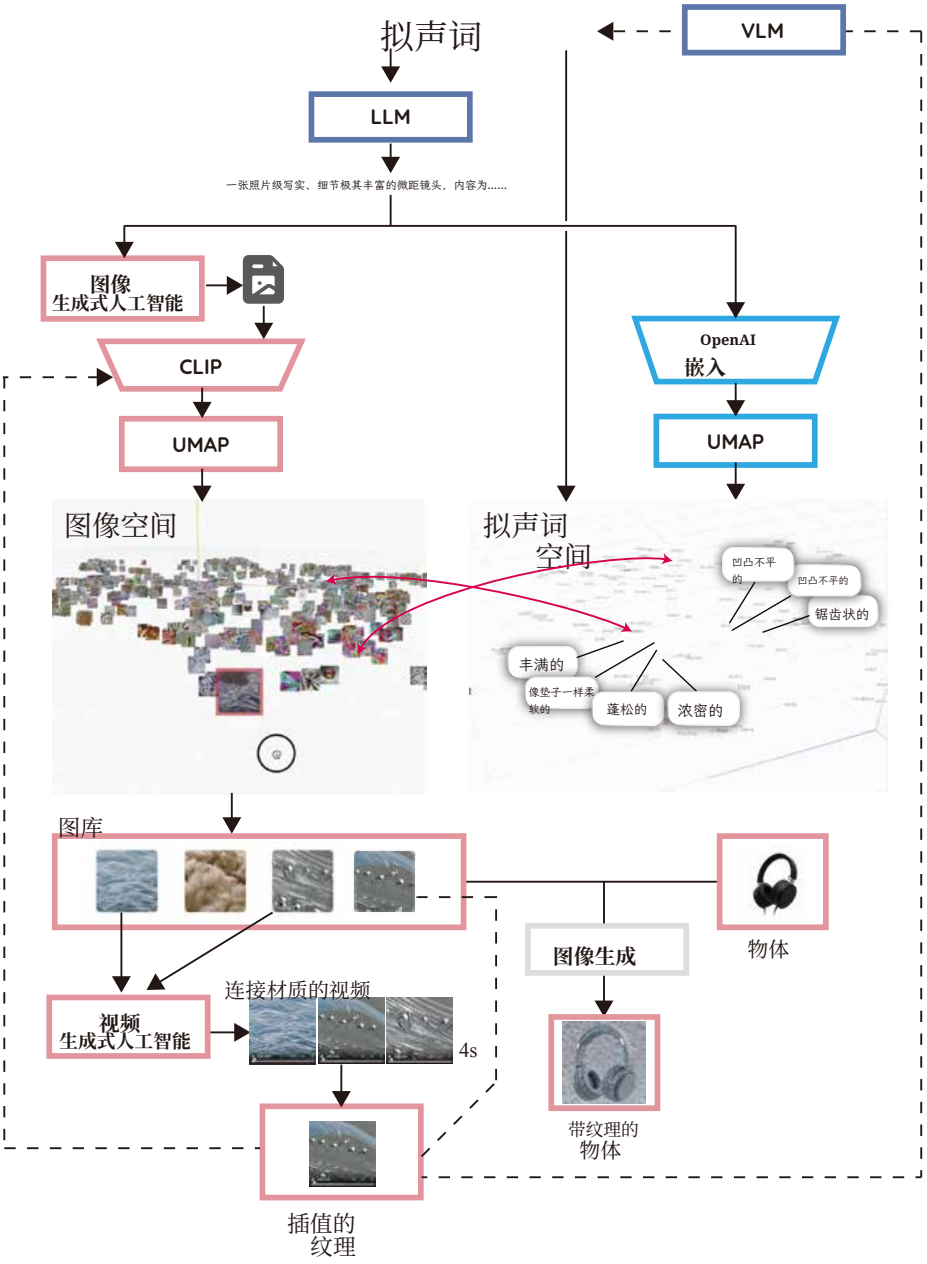

图 2. **OnomaCompass 系统概览：**图像与语言潜在空间的双重可视化及跨模态高亮、用于整理和为物体应用纹理的图库，以及一个动态视频插值与重新绘制流程，可将新生成的帧反馈到这两个空间中。

3.1.2 核心功能与界面。 为实现上述设计目标，系统实现了以下五项核心功能。

双潜在空间可视化。 系统的核心是对图像潜在空间和语言潜在空间进行并行可视化（图 3）。两个相互协调的地图并排呈现。每个地图都通过 UMAP 在二维空间中计算，并在三维场景中渲染，以便进行导航。左侧的图像空间包含 676 张纹理图像，右侧的语言空间包含 235 个拟声词；每个元素都位于三维视图中的二维坐标上（具有固定的 z 值）。系统围绕由拟声词及其视觉表达构成的两个潜在空间展开。由于系统旨在支持不受通用物体概念限制的材料质感创意探索，因此特意使用虚构的拟声词构建数据集。这样做并非将条目局限于日常熟悉的词语，而是希望通过声音象征激发想象力，从而促进创造力。首先，对于作者收集的以及来自先前项目的虚构拟声词，系统使用大语言模型（LLM）分阶段生成图像提示词。此过程包括识别拟声词唤起的材料、定义其物理特性，并将其转写为英文，最终生成用于 Stable Diffusion 的提示词。选择 Stable Diffusion 并非因为其照片级真实感处于业界领先水平，而是因为它能够提供富有艺术感且多样化的材料质感诠释。每个空间都通过 Three.js 使用 WebGL 在浏览器中渲染。用户可以使用鼠标自由调整视角：拖动以旋转，滚动滚轮以缩放，右键拖动以平移。图像以 64×64 px WebP 缩略图的形式加载，并以基准缩放比例为 baseScale = 0.075（约相当于 4.8×4.8）的精灵图显示；视角变化时，它们始终面向摄像机，从而在可导航的三维视图中呈现的二维地图内最大限度地保持辨识度。

跨模态高亮。 作为 DG2（意外发现）的核心功能，系统实现了跨模态高亮。当用户在一个空间中选择某个项目时，另一个空间中的相关项目会依据数据集构建流程预先定义的跨模态链接（即用于为每个虚构词语生成纹理的共享材质描述文本）而突出显示。换言之，跨模态关系并非通过相似度搜索在线推断得出，而是由构建流程定义，该流程将每个虚构的拟声词与其材质描述以及由此生成的纹理相连接。选择一个拟声词后，其关联的纹理图像（1 至 3 张，具体取决于生成是否成功）会突出显示，并显示在小型预览面板中，供用户从多个变体中进行选择。选择一张纹理图像后，则会突出显示生成该图像所用的唯一拟声词。这种非对称映射支持两种互补的探索方式：“这个词能唤起什么图像？”以及“什么词能描述这张图像？”

图库与纹理应用。 为了保存并利用探索过程中发现的纹理，系统实现了图库功能。用户可以将保存的图像拖放到展示的物体图像（花瓶和耳机）上，生成纹理应用预览。应用纹理时，系统使用 Gemini 2.5 Flash Image（图像生成与编辑）将纹理图像（同时呈现材料外观与表面几何形态）与物体图像合成，生成应用材质后的图像。

动态视频插值与重新绘制。 为了实现 DG4（一个涌现循环），我们开发了一套动态视频插值与重新绘制系统。用户可以在图库中选择两张图像，并生成一段呈现二者之间视觉转变的视频。视频生成使用 Luma AI Ray 1.6 进行单向插值（A→B），着重表现鲜活、流动的“材料行为”，而非简单的变形。具体而言，系统使用一段较长的提示词，从固定摄像机位置拍摄照片级真实感的超高分辨率微距镜头，其中包含局部的有机与奇幻现象（流体转变、涌现生命形式、物质相变等）。此外，

可以从生成的视频中提取任意视频帧，并通过四阶段人工智能流程将其重新定位到现有的潜在空间中：步骤 1：视觉特征提取 利用 Gemini 2.0 Flash（gemini-2.0-flash- exp）的视觉语言模型能力，分析提取的视频帧中的视觉特征，并直接生成对应的拟声词。步骤 2：概念描述生成 根据步骤 1 得到的拟声词，OpenAI o1 模型生成详细的概念描述（description），为后续的嵌入过程提供上下文语义信息。步骤 3：特征嵌入 图像特征：使用 CLIP ViT-B/32 将视频帧嵌入为一个 512 维向量。文本特征：使用 OpenAI text-embedding-3-small 将步骤 2 中的概念描述嵌入为一个 1536 维向量。步骤 4：投影到潜在空间 使用预训练的 UMAP 模型，将步骤 3 中的特征投影到现有的二维坐标空间中。系统会计算对应于纹理空间和拟声词空间的坐标，从而无需重新训练即可放置新数据。重新绘制的动态点以橙色标示，由此实现涌现循环：用户的探索行为会生成新的目标，供进一步探索。

3.1.3 实现细节。

系统架构。 OnomaCompass 采用前端中心架构，以便在持续探索过程中保持响应能力（DG3）。嵌入向量和地图坐标均预先计算并发送至客户端，因此大多数交互都可在浏览器中处理，无需反复与服务器往返通信。后端使用 Python 3.12 和 FastAPI 0.104 实现，前端则使用 React 19.1.1 和 TypeScript 5.8.3 实现。可视化通过 React Three Fiber 使用 Three.js r162 渲染。

数据集构建与嵌入流程。 我们整理了 235 个自创拟声词。对于每个词，我们使用 OpenAI o1 分阶段生成 Stable Diffusion 提示词（拟声词 → 材料识别 → 物理特性定义 → 英文表述 → Stable Diffusion 提示词），随后为每个词生成三张纹理图像。由于偶尔会生成失败，部分词关联的纹理少于三张，因此图像总数为 676 张。这段描述文本也定义了界面中的跨模态链接，将每个自创词与由其生成的纹理集合相连，从而实现确定性的跨模态高亮。为构建两个相互协调的地图，我们计算了图像和语言嵌入，并使用 UMAP 将其投影到二维坐标中（n_邻居数=15，min_距离=0.5，度量=余弦）。图像嵌入使用 CLIP ViT-B/32 [33]计算为 512 维向量。语言嵌入则使用 text-embedding-3-small 对生成的提示词文本（即用于通过 Stable Diffusion 生成纹理的英文描述）进行处理，得到 1536 维向量。重要的是，这种语言表征是与每个自创词相关的预期材料概念的替代指标；它并未对语音象征进行明确建模。在提示词生成过程中，我们明确要求大语言模型在生成英文描述时保留并清晰表达每个词预期的声音象征性（拟态）印象。这段共享提示词文本为原型中的不同模态提供了共同表征：它用于 (i) 使用 Stable Diffusion 生成纹理图像，(ii) 通过文本嵌入计算语言地图布局，以及 (iii) 定义跨模态链接，以实现确定性高亮。

可视化与交互。生成的二维坐标以并排地图的形式呈现在可导航的三维视图中（z 值固定），以支持基于摄像机的探索。纹理图像以精灵图的形式渲染，以维持绩效；缩略图（64×64 px WebP）会逐步加载，而全分辨率图像则按需加载。用户可以使用 OrbitControls 旋转、缩放和平移；摄像机还可以自动聚焦于选定项目，以便快速在纹理地图与拟声词地图之间切换。

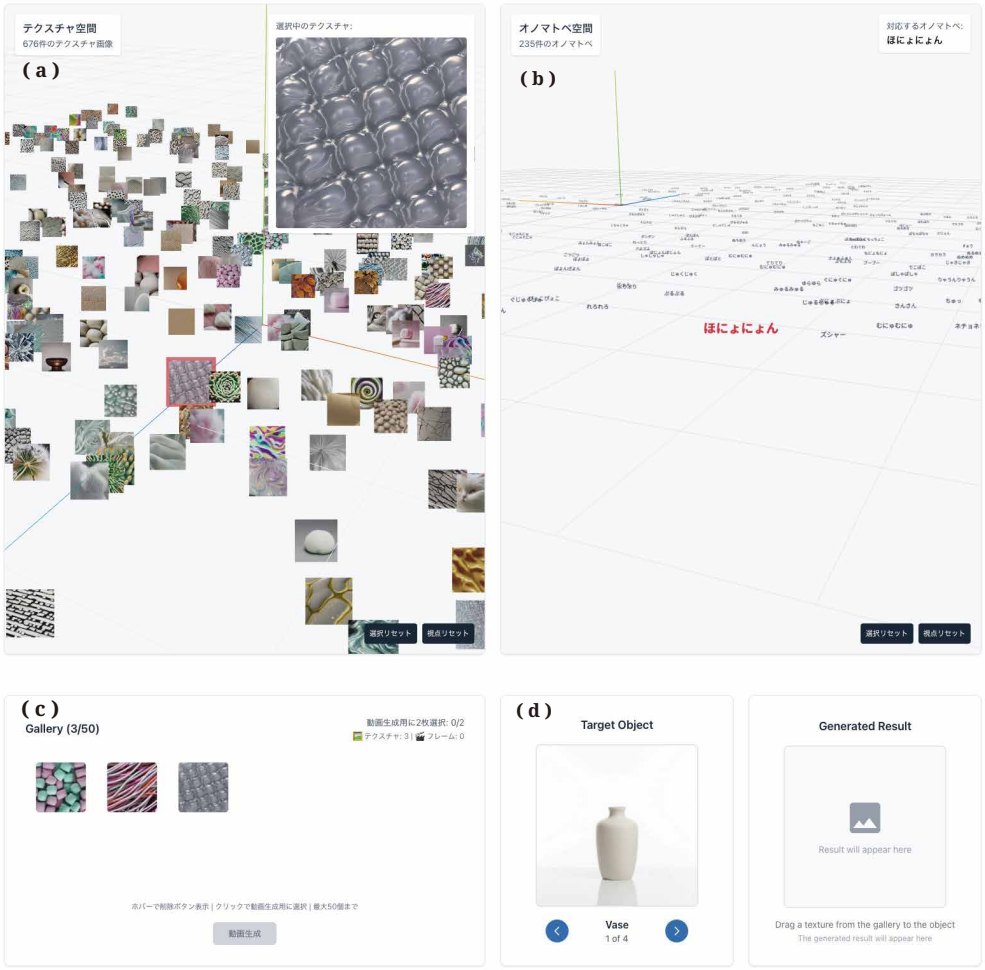

图 3. **所提出系统的用户界面**。(a) 纹理空间：根据视觉特征排列纹理图像的散点图。(b) 拟声词空间：根据大语言模型生成的提示描述的文本嵌入相似度排列拟声词的散点图（这些描述旨在体现声音象征印象；例如，“Honyonyon”）。(c) 图库与预览：用于管理所选纹理并模拟其应用于目标物体（例如花瓶）的区域。

3.2 评估研究

本研究不评估客观的“最佳纹理”结果。相反，我们关注界面如何塑造早期构思——包括工作负荷、探索行为以及感官预期的外化；在这一阶段，目标可能仍处于形成过程中。我们设计了一项以半结构化访谈为核心的混合方法评估，以阐明所提出的系统如何影响用户的创造性思维。

3.2.1 参与者。11名参与者（本科生和研究生）均为日语母语者，认同本研究的背景与目的，并参加了实验。

3.2.2 任务与流程。研究人员告知参与者，本研究设想的是“一个新手能够创造出具有理想材料质感的伪影的未来”。他们使用 OnomaCompass（OC）和基线（NB），探索并为两种目标生成材料质感：耳机和花瓶。每种条件持续15分钟：每种目标各7分钟（另加切换时间）。每种条件下目标的顺序交替安排。每种条件结束后，参与者填写问卷。

基线条件：*Nano Banana*。参与者获得了目标物体（花瓶或耳机）的基础图像，并可以选择：（i）从提供的物体图像开始，反复编辑；或（ii）从头生成纹理图像，然后生成融入该纹理的物体图像。未明确允许使用其他参考图像。参与者可以创建多个新的聊天对话线程，并可根据需要使用聊天功能进行澄清和构思，且未设置自定义系统提示词。

OnomaCompass 条件。研究人员讲解了第 3.1.2 节中描述的所有系统功能（在不同空间之间切换、图库、纹理应用、视频插帧和重新绘制），随后安排了一分钟的熟悉时间。研究采用被试内设计。为平衡顺序效应，六名参与者先使用 OnomaCompass，五名参与者先使用图像生成式人工智能。

3.2.3 数据收集与分析。完成所有任务后，研究人员进行了半结构化访谈。访谈询问了：（i）总体印象；（ii）各系统的优势与不足；（iii）材料质感探索过程的差异；（iv）令人惊讶或感兴趣的时刻；（v）在不同场景中的使用情况（快速探索与构思）；（vi）向朋友推荐的可能性；以及（vii）OnomaCompass 的改进方向。研究人员采用定性主题分析对口头数据进行编码，并对涌现的主题进行归纳和分析。

4 结果

4.1 定量评估

本节报告了比较 OnomaCompass（OC）与 Nano Banana（NB）的 UEQ、NASA-TLX、系统可用性量表和原始问卷的结果。实验采用被试内设计（$N = 11$）。参与者的评分按条件汇总后进行比较。使用 Shapiro–Wilk 检验验证正态性。对于服从正态分布的数据，采用配对 t 检验；对于不满足正态性的数据，采用 Wilcoxon 符号秩检验。显著性水平设定为 $\alpha = 0.05$。报告的 p 值均为双尾且未经校正。效应量方面，t 检验采用 Hedges’ g（针对小样本量进行了校正），Wilcoxon 检验采用 r。NASA-TLX 使用原始（未加权）平均值，并将每个条目解释为分数越高表示工作负荷越高。

4.1.1 核心价值：NASA-TLX。总体工作负荷（原始平均值）在 OC 条件下为 34.46（$SD = 24.23$），在 NB 条件下为 45.55（$SD = 25.56$）。OC 条件显著较低 [$t(10) = -2.686, p = 0.023, g = -0.74$]。在各子量表中也观察到一致的差异。具体而言，心理需求 [38.36 对比 58.00，$t(10) = -2.919, p = 0.015, g = -0.80$]；努力程度 [25.09 对比 49.27，$t(10) = -2.269, p = 0.047, g = -0.63$]；挫折感 [23.46 对比 56.82，$t(10) = -4.420, p = 0.001, g = -1.22$]，在 OC 条件下均显著较低。身体需求未观察到显著差异 [$t(10)$ =1.518。同样，时间需求 [$t(10) = 0.403, p = 0.695$] 和绩效 [$t(10) = -1.514, p = 0.161$]也未发现显著差异。这些结果表明，所提出的系统显著减少了认知资源负担（心理需求、努力程度）和情绪负荷（挫折感）。与此同时，与身体操作和时间

压力相关的方面在两种条件下似乎相当。这一发现与第 4.2 节中的定性观察（“低词汇负担”和“OC 很有趣”）一致。

4.1.2 用户体验问卷：享乐价值的优势。关于用户体验问卷的综合指标（-3 至 +3），OC 的享乐质量显著更高（ [2.000 vs. -0.091， $t(10) = 4.416$， $p = 0.001$， $g = 1.22$]）。OC 的总体质量也显著更高（ [1.578 vs. 0.694，$t(10) = 2.847$， $p = 0.017$， $g = 0.78$]）。相反，实用质量未观察到显著差异（ [1.159 vs. 1.477，Wilcoxon：$Z = -0.533$， $p = 0.618$， $r = 0.11$]）。换言之，所提出的系统显著提升了愉悦感、新颖性等享乐层面，以及整体体验质量，同时没有损害其实用性。

4.1.3 系统可用性量表：一般可用性的权衡。 NB 的系统可用性量表得分显著较高： [75.68（$SD = 12.25$），而 OC 为 67.73（$SD = 17.69$）， $t(10) = -2.316$， $p = 0.043$， $g = -0.64$]。传统的基于聊天的图像生成在一般可用性方面似乎更胜一筹。这一结果可以解释为，与该原型特有的操作（如三维空间导航）相关的学习成本所产生的影响。

4.1.4 原始问卷：对发散思维的支持与意外发现。 在按 1–7 分量表评分的 12 个项目中，以下五项存在显著差异（未经校正的探索性分析）。能够探索多种纹理的不同变体（Q3）： OC ＞，NB [6.73 对比 4.00， $t(10) = 4.404$， $p = 0.001$， $g = 1.21$]不知道接下来该做什么而陷入困境（反向题，Q6）： NB ＞，OC [4.00 对比 1.82，Wilcoxon： $Z = -2.934$， $p = 0.002$， $r = 0.63$]心理负荷较低，疲劳感轻微（Q7）： OC ＞，NB [5.46 对比 3.73， $t(10) = 3.012$， $p = 0.013$， $g = 0.83$]整个过程富有创造性且有趣，而非充满技术性（Q8）： OC ＞，NB [6.27 对比 4.82，Wilcoxon： $Z = -2.667$， $p = 0.008$， $r = 0.57$]出现了超出预期的意外发现（Q10）： OC ＞，NB [6.64 对比 4.55， $t(10) = 3.348$， $p = 0.007$， $g = 0.92$]其他项目（如任务完成感、愿景清晰度、重复使用意愿）未达到统计显著性。不过，许多项目呈现出有利于 OC 的趋势，效应量为中到大。总体而言，标准量表（NASA-TLX 和用户体验问卷）与原始问卷项目均证实，所提出的系统在“避免陷入困境”“低心理负荷”“意外发现”“创造性乐趣”以及“变体探索”方面具有优势。

4.2 OnomaCompass 的定性评估

4.2.1 价值核心：基于选择的探索与发现。 参与者主要将 OnomaCompass 视为一个浏览和发现的环境，尤其是在他们脑海中没有明确用语言描述的目标纹理时。几位参与者将其与基于提示词的生成进行了对比，强调系统中有示例：“使用 [Nano Banana],时没有示例，所以很多事情都得由我自己决定。这个系统里有很多与词语对应的示例，我可以一边看一边选，感觉轻松多了。”参与者将自己的探索描述为“点击拟声词并浏览图像，看看会发生什么”，类似于“从海量壁纸目录中挑选”。这种基于选择的交互降低了迈出第一步的门槛，也让他们有机会意外发现超出自身既有词汇和知识范围的内容。

4.2.2 拟声词作为重新诠释与意义拓展的线索。 除了“发现纹理”之外，参与者还重点提到了拟声词与图像的组合如何改变他们对表达本身的理解。一位参与者描述了自己通过*shakishaki* （“爽脆／嘎吱作响”）获得的一次发现：虽然他们通常会将这个词与咬蔬菜时的声音 联系起来，但看到一张具体描绘蔬菜被咬过的表面的图像后，这种联想有了新的现实依据——“看到它之前，我从未想象过这个词所对应的视觉场景。”另一位参与者谈到了*zowazowa* （“毛骨悚然／麻刺感”），并指出，单独看所选图像并不会让人想到这个词，但

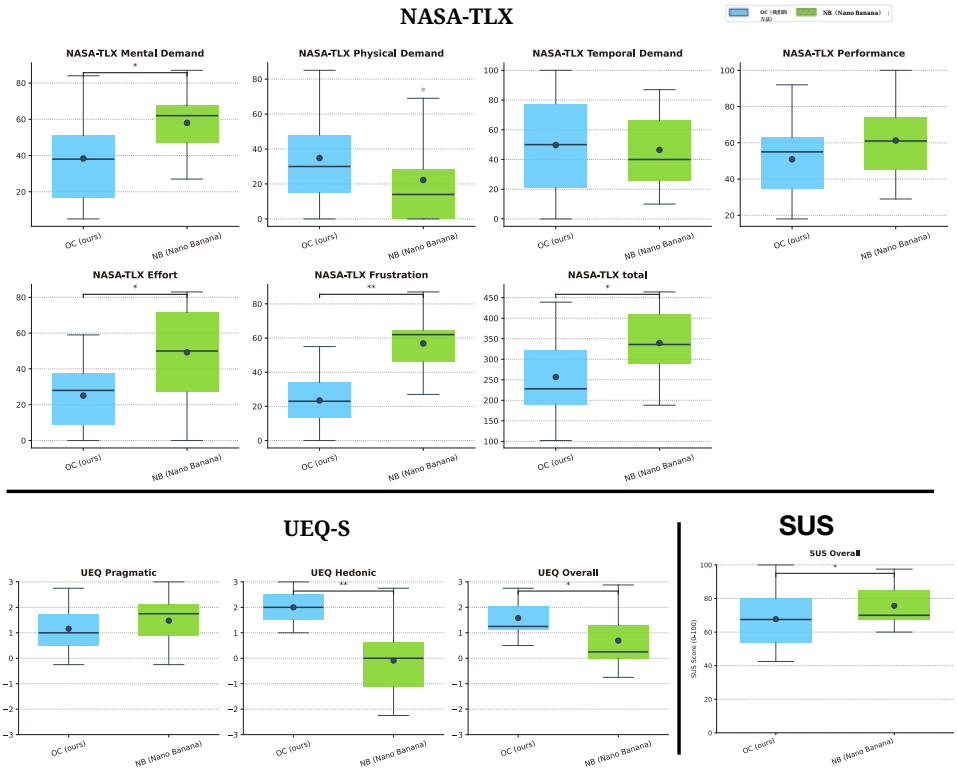

图 4. **使用标准指标进行的比较评估结果**。（a）用户体验问卷、（b）NASA-TLX（主观工作负荷）和（c）系统可用性量表的得分。在箱线图中，中线表示中位数，圆点表示平均值，箱体表示四分位距（IQR），须线表示 1.5 × IQR 范围内的取值。（OC：提出的系统 [OnomaCompass], NB：传统方法 [Nano Banana], * p ＜.05，** p ＜.01）

两者的组合却令人信服——“光看那张图，我不会想到*zowazowa*，但把它们放在一起后，就显然说得通了。”这些反馈表明，在 OnomaCompass 中，拟声词的作用与其说是一个需要精确“匹配”的标签，不如说是一种促使用户重新诠释并拓展其描述空间的线索。

4.2.3 通过分支式探索发挥思维辅助工具的作用。OnomaCompass 在构思初期尤其能发挥思维辅助工具的作用，此时目标仍在逐步形成。参与者表示，它“弥补了词汇量有限的问题”，而可视化地图让他们能够通过比较继续探索，而不必立刻确定某一种语言描述。一位参与者对比了两种系统带来的认知感受：“在聊天中，人工智能一回复，我就觉得自己必须马上回应。用这个系统时，我感觉自己像是套着游泳圈，平静地漂浮在一片大海里。”另一位参与者指出，探索纹理图像空间很高效，因为“视觉信息一眼就更容易判断”，而且它还支持一种不同的构思方式：“它让我想到‘把 A 和 B 结合起来也许会很有意思！’”

4.2.4 “中间地带”的涌现循环与组合式构思。参与者尤其看重插值与重新嵌入的循环，将其视为一种创作行为，而非单纯的检索。一位参与者将其比作组合式游戏：

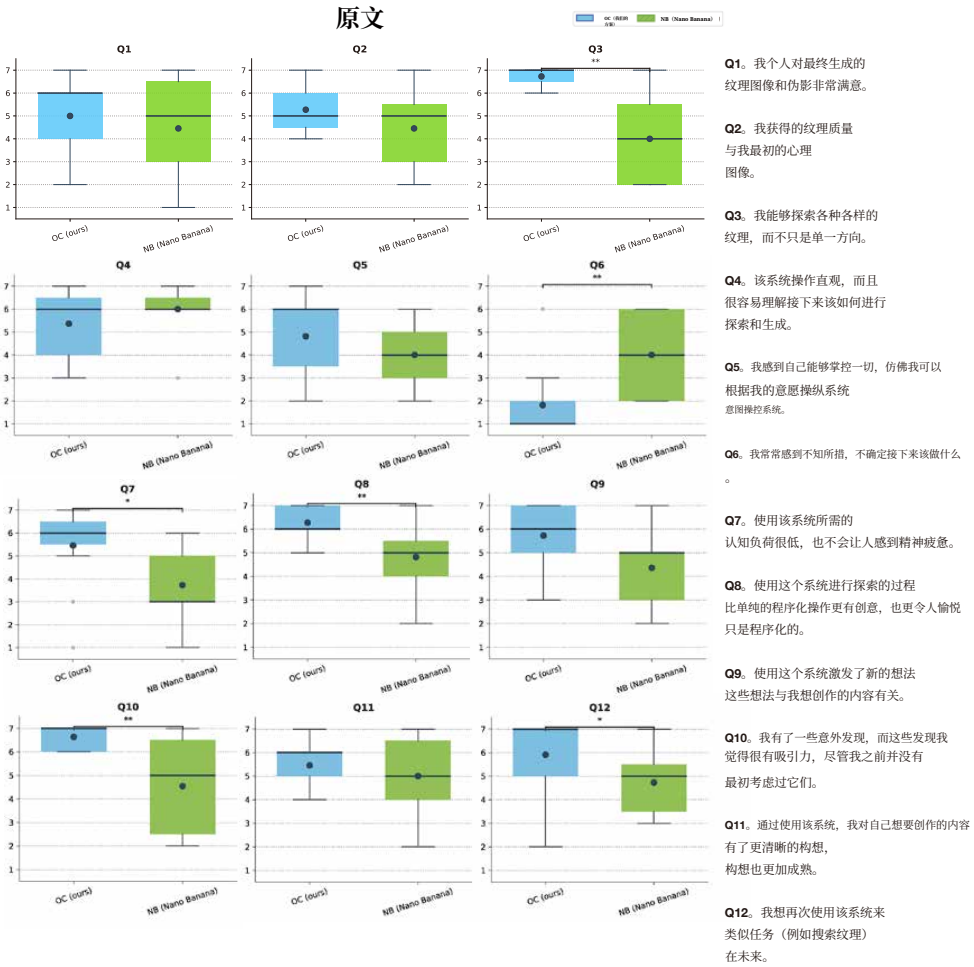

图 5. **原始问卷（Q1–Q12）的评估结果**。 关于创造力支持与探索体验的 12 个条目（7 点李克特量表）的回答分布。OC（提议的系统）获得了显著更高的评分，尤其是在多样化探索（Q3）、避免陷入困境（Q6）、创作过程（Q8）和意外发现（Q10）方面。（* p <.05, ** p < .01）

“帧插值很有趣——它不只是选择，还包含创作的成分。就像《勇者斗恶龙 怪兽篇》中的培育过程，探索会出现什么，并填满图鉴。”值得注意的是，参与者指出，语言上的“中间地带”很难直接清晰表达：“很难说出‘蓬松的’和‘锯齿状的’之间的拟声词。但对图像进行插值似乎是可行的；之后看看人工智能会为其赋予什么拟声词，也很有意思。”这表明，系统的价值在于让用户能够通过图像与词语之间的双向循环，探索材料性在概念上的中间地带。

4.2.5 可用性挑战。相反，许多可用性问题削弱了系统的价值。最突出的问题与三维空间中的交互有关：“不熟悉三维操作时很难使用”，

“感觉上是在进行平面探索，却要求进行三维操控”，以及“它与我通常使用的三维工具不同”。参与者还报告了交互可靠性方面的问题，例如“图像相互重叠，导致误选了不想选的项目”。一些参与者希望访问模式更加直观易懂（例如，二维/2.5D导航、用于查找拟声词的列表/搜索功能，以及对空间布局所编码内容的更清晰说明）。

4.3 基线的定性评估（NanoBanana）

4.3.1 可控性与输出质量的价值。 Nano Banana 的主要优势在于，当参与者能够清楚地用语言表达目标时，它具有很强的可控性。许多参与者称赞它能够指定细节（“可以精细控制哪些部分使用哪种纹理”），也称赞它能够遵循要求（“它忠实地呈现了所要求的内容”）。参与者也认可其逼真度和整体图像质量。

4.3.2 对话与“天马行空”的提示词带来的乐趣。 参与者也乐于进行有趣的互动，例如故意使用不同寻常的提示词，并欣赏人工智能出人意料的补充内容（例如，“花瓶变成布丁时，还出现了一把勺子”）。有些参与者会输入远离材质描述的词语来探索新奇效果，期待得到超越自身想象的“额外惊喜”。

4.3.3 语言表达障碍、缺乏示例与线性探索。 然而，参与者反复指出，语言表达的负担是核心挑战。他们提到，生成和编辑纹理需要仔细编写提示词，而且“在生成之前很难想象结果会是什么样”，因此意图与结果之间存在落差。几位参与者还表示，NB 一开始提供的具体示例很少，让人难以确定方向：“没有示例，所以很多事情都得由我自己决定。”还有一些参与者认为，聊天工作流在结构上是线性的：“因为它不是一种可以从多个选项中挑选的形式，所以整个过程会变成一条故事线——从第一次尝试开始一步步推进。”这种线性有时会使参与者在遇到困难时依赖熟悉的概念；有些人则觉得，基于提示词的生成“只能表达我脑海中已有的东西”，因此很难突破自身词汇的局限，探索未知的纹理。

5 讨论

5.1 发散思维与聚合思维：支持创造力的两种模型

访谈有力地表明，OnomaCompass 与基于提示词的图像生成式人工智能具有互补性，分别支持创作过程的不同阶段。Guilford 关于发散思维与聚合思维的概念恰当地解释了这种差异 [14]。OnomaCompass 在发散思维阶段尤其有效，这一阶段从没有明确目标的状态出发拓展想法。相比之下，对于图像生成式人工智能而言，“一开始很难决定自己喜欢什么”；参与者则形容 OnomaCompass 能提供“选择”，且“任务负担很小”。相反，图像生成式人工智能在聚合思维阶段表现出色：在这一阶段，特定目标会逐步得到精准实现。

5.2 开放式构思中的线性对话与分支式探索

我们的定性数据表明，基于提示的聊天生成与基于地图的探索在结构上有所不同。参与者认为，NB 需要不断作出语言上的承诺，并沿着单一轨迹（“一个故事线”）展开；而 OnomaCompass 则通过同时呈现多种备选方案，支持分支式探索，并让选择可撤回且成本低廉。这种差异对于开放式构思尤为重要，因为目标纹理往往尚未明确，偏好也常通过比较而非直接指示逐渐形成。

“平静地漂浮”这一隐喻进一步表明，聊天的轮流对话特性可能会带来一种微妙的回应压力，而以浏览为导向的界面则更能维持从容探索的心态。不过，由于无法理解距离和位置的含义，参与者有时会把地图仅仅视为一组项目。这凸显了一项关键的设计要求：潜在空间界面必须让映射易于理解，并提供明确的操作方式。

5.3 构建“心理地图”的工具

我们的研究发现可以通过新手与经验丰富的从业者在探索材料概念时的差异来解读。通过反复接触与实践，人们可能会积累个人的材料参照库——将触感印象、视觉示例和描述性词语联系起来——并以此形成一张用于明确意图的内在“心理地图”。当用户已经知道如何在这张地图中明确自己的意图时，基于提示词的图像生成会很有效。相比之下，我们的参与者对材料质感的相关词汇储备有限，因此常常能从 OnomaCompass 这一地图构建辅助工具中受益：相互关联的纹理地图与拟声词地图为意义建构提供了外部结构， [32], 支持比较、命名，并逐步优化预期。从这一角度看，OnomaCompass 的价值或许不在于高效地找到单一目标纹理，而在于提供认知支架， [41] 帮助用户区分材料印象，并构建一套可供未来构思重复使用的描述词（包括声音象征性描述词）。

5.4 拟声词：一座模糊却富有成效的桥梁

我们的定性数据也表明，拟声词发挥着细腻而多元的作用：它们并非只是对应某一种视觉纹理的标签，而是能够唤起感官意义多个层面的线索（声音、触感和情境意象）。重要的是，即使图像本身并不会让参与者联想到某个词，他们有时仍会认为两者之间的关联令人信服，这说明其价值在于重新诠释，而非准确的“翻译”。这为设计带来了新的机会：与其将模糊性视为噪声，不如让界面将声音象征性表达转化为促成意义建构的提示词——通过展示扎根于实际的示例，帮助用户清晰表达并完善对材料的预期。

5.5 探索体验与涌现式玩乐的价值

最重要的发现是，参与者认为探索体验本身就有价值。图像、视频和拟声词构成的循环促成了新的发现，并带来一种超越单纯搜索工具的体验，积极激发了创造力。一位参与者将组合两种材料性以创造新事物的过程比作一个热门游戏系列：“在 DragonQuest Monsters 中，把不同物种组合起来看看会出现什么，令人兴奋，就像填满图鉴一样有趣。”这一比喻表明，该系统发挥了涌现式玩乐空间的作用，用户可以主动组合各种元素，发现未知结果。这体现了工具作为认知人工物发挥作用、拓展用户思维的一个例子。

5.6 局限性与未来方向

本研究存在一些局限。首先，原型的可用性尚不成熟，可能使参与者无法充分体验语义空间探索的核心价值。其次，空间布局的原则未能向用户充分说明。许多人表示不理解“映射的依据”，这可能妨碍了更深入的探索。未来的工作应大幅提升可用性，并更细致地研究

空间布局原则如何影响探索行为和创造力。第三，针对参与者希望“兼顾两者”的需求，设计一种无缝整合发散式探索与收敛式生成的混合创意环境，似乎是一个颇具前景的方向。此外，我们的数据集包含虚构的拟态词，以激发想象力并避免依赖熟悉的类别。尽管参与者认为这些词有助于发现且颇为有趣，但我们并不认为这些特定联想能够推广至更广泛的人群或其他语言。未来的工作应评估跨用户一致性和跨文化迁移，并在受控条件下比较虚构与常规象声词。

6 结论

本文介绍了*OnomaCompass*，这是一种面向早期材料性构思的基于网络的 AI 界面，使用户能够在纹理图像空间与拟声词空间之间切换。该系统不依赖提示词编写，而是将声音象征线索外化为可供探索的结构，具体包括：(i) 双重潜在空间可视化与跨模态高亮；(ii) 整理与纹理应用预览；以及 (iii) 涌现循环，在不同纹理之间生成插值的视频，并将提取的视频帧重新嵌入两个空间。

一项被试内研究（$N = 11$）比较了 OnomaCompass 与基于聊天的提示词驱动图像生成工作流（NB），结果显示出一致的权衡。OnomaCompass 显著降低了主观工作负荷（NASA-TLX 总分、心理需求、努力程度和挫折感），并提升了享乐型用户体验（UEQ），表明在空间中切换能够减轻语言表达负担，让探索性构思更加愉悦。与此同时，基于聊天的基线在可用性（系统可用性量表）方面表现更佳，参与者也报告了具体的交互问题（例如三维导航困难和选择错误），这凸显出以探索为导向的界面能否发挥优势，取决于交互设计的成熟度。

综合定量与定性研究结果，我们认为，OnomaCompass 最适合被视为一种发散思维支架，以及帮助新手构建材料质感“心理地图”的工具；而当用户已经能够明确表达意图时，基于提示词的生成则擅长进行收敛式完善。因此，系统的价值不在于取代文生图提示词，而在于让材料概念之间的“中间地带”变得可供探索，并通过涌现式探索循环促成意想不到的关联。

未来的工作将：(1) 重新设计交互方式，实现二维/2.5D导航和更可靠的选择；(2) 更清晰地传达跨模态对应关系的映射依据与不确定性；(3) 开发一种混合工作流，将发散式潜在空间探索与收敛式基于提示词的生成与编辑紧密结合。更广泛地说，这项工作为协同创作型 AI 界面提出了一种发展方向：从一次性指令转向互惠式探索，让用户将 AI 生成的表征用作可导航的认知支架，以探索非语言、感性驱动的领域。

参考文献

[1] Ahmed M Abuzuraiq 和 Halil Erhan。2025年。以感知为导向的设计：对生成式设计中增强设计空间探索的批判性综述与提案。《国际建筑计算期刊》23，1（2025年），5–26。[2] Krzysztof Adamkiewicz、Paweł W Woźniak、Julia Dominiak、Andrzej Romanowski、Jakob Karolus 和 Stanislav Frolov。2025年。PromptMap：一种用于人工智能图像生成的替代性交互方式。载于国际智能用户界面会议论文集。1162–1176。[3] Christopher Ahlberg 和 Ben Shneiderman。1994年。视觉信息检索：将动态查询筛选器与星域显示紧密结合。载于*SIGCHI* 计算机系统中的人为因素会议论文集。313–317。[4] K Akita 和 Pardeshi P（编）。2019年。象声词、拟态词与表现性词语。16（2019年）。[5] Ahmet Canberk Baykal、Abdul Basit Anees、Duygu Ceylan、Erkut Erdem、Aykut Erdem 和 Deniz Yuret。2023年。CLIP 引导的 StyleGAN 反演，用于文本驱动的真实图像编辑。《*ACM* 图形学汇刊》42，5（2023年），1–18。

[6] Stephen Brade、Bryan Wang、Mauricio Sousa、Sageev Oore 和 Tovi Grossman。2023年。Promptify：通过与大语言模型进行交互式提示词探索来生成文本图像。载于第36届ACM用户界面软件与技术研讨会论文集。1–14。[7] DaEun Choi、Sumin Hong、Jeongeon Park、John Joon Young Chung 和 Juho Kim。2024年。Creativeconnect：利用生成式人工智能支持平面设计构思中的参考素材重组。载于2024年计算机系统中人为因素CHI会议论文集。1–25。[8] Hai Dang、Lukas Mecke 和 Daniel Buschek。2022年。GANSlider：用户如何通过多个滑块控制图像生成模型，以及有无前馈信息时的差异。载于2022年CHI计算机系统人因大会论文集。1–15。[9] Faraz Faruqi、Maxine Perroni-Scharf、Jaskaran Singh Walia、Yunyi Zhu、Shuyue Feng、Donald Degraen 和 Stefanie Mueller。2025年。TactStyle：利用生成式人工智能生成用于数字化制造的触觉纹理。载于2025年CHI计算机系统人因大会论文集。1–16。[10] G W Furnas、T K Landauer、L M Gomez 和 S T Dumais。1987年。人机沟通中的词汇问题。《ACM通讯》30，11（1987年11月），964–971。[11] Rohit Girdhar、Alaaeldin El-Nouby、Zhuang Liu、Mannat Singh、Kalyan Vasudev Alwala、Armand Joulin 和 Ishan Misra。2023年。ImageBind：一个将万物绑定在一起的嵌入空间。载于IEEE/CVF计算机视觉与模式识别会议论文集。15180–15190。[12] Ian J Goodfellow、Jean Pouget-Abadie、Mehdi Mirza、Bing Xu、David Warde-Farley、Sherjil Ozair、Aaron Courville 和 Yoshua Bengio。2014年。生成对抗网络。载于第28届神经信息处理系统国际会议论文集——第2卷（NIPS'14，第2卷）。麻省理工学院出版社，剑桥，马萨诸塞州，美国，2672–2680。[13] Florian Grötschla、Luca A Lanzendörfer、Marco Calzavara 和 Roger Wattenhofer。2024年。AEye：图像数据集可视化工具。载于2024年IEEE可视化与视觉分析会议（VIS）。IEEE，281–285。[14] J P Guilford。1967年。《人类智力的本质》。麦格劳-希尔，纽约。[15] Mitsuhiko Hanada。2016年。运用日语拟声词探索视觉材料感知中的感知维度。《知觉》45，5（2016年），568–587。[16] Erik Härkönen、Aaron Hertzmann、Jaakko Lehtinen 和 Sylvain Paris。2020年。GANSpace：发现可解释的GAN控制。《神经信息处理系统进展》33（2020年），9841–9850。[17] Mutsumi Imai 和 Sotaro Kita。2014年。语言习得与语言演化的声音象征引导假说。《英国皇家学会哲学汇刊B辑：生物科学》369，1651（2014年），20130298。[18] Mutsumi Imai、Sotaro Kita、Miho Nagumo 和 Hiroyuki Okada。2008年。声音象征促进早期动词习得。《认知》109，1（2008年10月），54–65。[19] Noriko Iwasaki、David P Vinson 和 Gabriella Vigliocco。2007年。英语使用者对 gera-gera 和 yota-yota 了解多少？关于表示笑与行走的拟态词的跨语言研究。（2007年）。[20] Junko Kanero、Mutsumi Imai、Jiro Okuda、Hiroyuki Okada 和 Tetsuya Matsuda。2014年。声音象征在大脑中的加工方式：一项关于日语拟态词的研究。《公共科学图书馆·综合》9，5（2014年），e97905。[21] Tero Karras、Samuli Laine 和 Timo Aila。2019年。用于生成对抗网络的基于风格的生成器架构。载于2019年IEEE/CVF计算机视觉与模式识别会议（CVPR）。IEEE。[22] Gwanghyun Kim、Taesung Kwon 和 Jong Chul Ye。2022年。DiffusionCLIP：用于鲁棒图像操控的文本引导扩散模型。载于IEEE/CVF计算机视觉与模式识别会议论文集。2426–2435。[23] Umut Kocasari、Alara Dirik、Mert Tiftikci 和 Pinar Yanardag。2022年。StyleMC：基于多通道的快速文本引导图像生成与操控。载于IEEE/CVF计算机视觉应用冬季会议论文集。895–904。[24] Bum Chul Kwon、Samuel Friedman、Kai Xu、Steven A Lubitz、Anthony Philippakis、Puneet Batra、Patrick T Ellinor 和 Kenney Ng。2023年。潜在空间探索器：多模态潜在空间探索的可视分析。arXiv预印本 arXiv:2312.00857（2023年）。[25] Daniel Lee、Nikhil Sharma、Donghoon Shin、DaEun Choi、Harsh Sharma、Jeonghwan Kim 和 Heng Ji。2025年。ThematicPlane：连接图像生成中的隐性用户意图与潜在空间。载于第38届ACM用户界面软件与技术研讨会增补论文集。1–3。[26] David Chuan-En Lin、Fabian Caba Heilbron、Joon-Young Lee、Oliver Wang 和 Nikolas Martelaro。2024年。VideoMap：在潜在空间中支持视频探索、头脑风暴与原型设计。载于第16届创造力与认知会议论文集。311–327。[27] Vivian Liu 和 Lydia B Chilton。2022年。文本生成图像生成模型提示词工程设计指南。载于CHI计算机系统人因大会。ACM，纽约，美国。[28] Yang Liu、Eunice Jun、Qisheng Li 和 Jeffrey Heer。2019年。潜在空间制图：向量空间嵌入的可视分析。载于《计算机图形论坛》第38卷。Wiley在线图书馆，67–78。[29] Yuanhuiyi Lyu、Xu Zheng、Jiazhou Zhou 和 Lin Wang。2024年。UniBind：大语言模型增强的统一且均衡的表示空间，将万物绑定在一起。载于IEEE/CVF计算机视觉与模式识别会议论文集。26752–26762。[30] Daphne Maurer、Thanujeni Pathman 和 Catherine J Mondloch。2006年。布巴的形状：幼儿与成年人中声音与形状的对应关系。《发展科学》9，3（2006年5月），316–322。[31] Tillmann Ohm、Mar Canet Sola、Andres Karjus 和 Maximilian Schich。2023年。Collection Space Navigator：用于多维数据集的交互式可视化界面。载于第16届视觉信息传播与交互国际研讨会论文集。1–5。[32] Peter Pirolli 和 Stuart Card。2005年。通过认知任务分析识别出的意义建构过程及分析师技术的杠杆点。

[33] Alec Radford、Jong Wook Kim、Chris Hallacy、Aditya Ramesh、Gabriel Goh、Sandhini Agarwal、Girish Sastry、Amanda Askell、Pamela Mishkin、Jack Clark、Gretchen Krueger 和 Ilya Sutskever。2021年。通过自然语言监督学习可迁移视觉模型。*arXiv [cs.CV]*（2021年2月）。[34] Yujun Shen、Ceyuan Yang、Xiaoou Tang 和 Bolei Zhou。2020年。InterfaceGAN：解读生成对抗网络学习到的解耦人脸表征。*IEEE*模式分析与机器智能汇刊 44卷，4期（2020年），2004–2018页。[35] Ben Shneiderman。2007年。创造力支持工具：加速发现与创新。*Commun. ACM* 50卷，12期（2007年12月），20–32页。[36] Charles Spence。2011年。跨模态对应关系：教程式综述。注意、知觉与心理物理学 73卷，4期（2011年5月），971–995页。[37] Qian Wan 和 Zhicong Lu。2023年。Gancollage：一种生成对抗网络驱动的数字情绪板，助力创造力支持中的构思。载于*2023*年*ACM*交互系统设计会议论文集，136–146页。[38] Zhijie Wang、Yuheng Huang、Da Song、Lei Ma 和 Tianyi Zhang。2024年。Promptcharm：通过多模态提示与优化实现文本生成图像。载于*2024*年计算机系统中人为因素*CHI*会议论文集，1–21页。[39] Zijie J Wang、Fred Hohman 和 Duen Horng Chau。2023年。Wizmap：用于探索大型机器学习嵌入的可扩展交互式可视化。*arXiv*预印本 *arXiv:2306.09328*（2023年）。[40] Zijie J Wang、Evan Montoya、David Munechika、Haoyang Yang、Benjamin Hoover 和 Duen Horng Chau。2023年。DiffusionDB：面向文本生成图像生成模型的大规模提示词图库数据集。载于第*61*届计算语言学协会年会论文集（第*1*卷：长篇论文），893–911页。[41] David Wood、Jerome S Bruner 和 Gail Ross。1976年。辅导在问题解决中的作用。《儿童心理学与精神病学杂志》17卷，2期（1976年），89–100页。[42] Koichi Yamagata、Jinhwan Kwon、Takuya Kawashima、Wataru Shimoda 和 Maki Sakamoto。2021年。利用拟声象征词表达纹理的计算机视觉系统。《心理学前沿》12卷（2021年），654779。[43] Bin Zhu、Bin Lin、Munan Ning、Yang Yan、Jiaxi Cui、HongFa Wang、Yatian Pang、Wenhao Jiang、Junwu Zhang、Zongwei Li 等。2023年。LanguageBind：通过基于语言的语义对齐，将视频-语言预训练扩展至n模态。*arXiv*预印本 *arXiv:2310.01852*（2023年）。[44] Yiming Zhu、Hongyu Liu、Yibing Song、Ziyang Yuan、Xintong Han、Chun Yuan、Qifeng Chen 和 Jue Wang。2022年。一个模型编辑所有内容：通过语义调制实现自由形式文本驱动的图像操控。神经信息处理系统进展35卷（2022年），25146–25159页。