Evaluating the Impact of Sketch and Voice Inputs on Divergent Thinking in Gen AI

Human-Computer Interaction

Faculty: Novia Nurain | Section: B

GROUP NAME

Doctor Doom

GROUP MEMBER

011221078, 011221079, 011221080, 011221167, 011221435, 011221498

Date of Submission: 3rd December, 2025

1. Motivation

Current Generative AI interactions are dominated by the "Vocabulary Bottleneck" While state-of-the-art models like Nano Banana Pro (Gemini 3 Pro Image) can render photorealistic pixels, users are currently forced to translate complex, non-verbal visual ideas into linear text strings.

Recent HCI research (late 2024–2025) suggests this text-dependency forces users into "Convergent Thinking" (narrowing down to what they can describe verbally) rather than "Divergent Thinking" (exploring what they can imagine visually). This reliance on "Prompt Engineering" often results in Design Fixation, where users anchor on the model's default outputs rather than their own original ideas.

The release of Nano Banana Pro introduces a "Reasoning Core" capable of understanding "doodles" (user sketches) and complex context simultaneously. This creates a unique opportunity to test if Multimodal Interfaces (sketching a shape while speaking an intent) can break the fixation effects of text-only prompting and restore user agency in the creative process.

2. Research Question (RQ)

Does a Multimodal Interface significantly increase the fluency and originality of design concepts compared to a Text-Only Interface during the early ideation phase and how does this affect the user's Sense of Agency?

3. Methodology (Tentative)

Study Design: This study utilizes a Within-Subjects Controlled Experiment. Each participant will perform a creative task under two distinct conditions to allow for direct comparison of their experience.

Apparatus:

- **Condition A (Control - Text-Only):** A standard chat-based interface utilizing the Nano Banana Pro model. Users interact solely via typed text prompts.
- **Condition B (Experimental - Multimodal):** A prototype "Co-Creative Canvas" allowing users to draw rough shapes and issue voice commands.

  Task: Participants will be asked to: "Design a concept vehicle for a Martian colony." This task was selected because it lacks real-world references, forcing participants to rely on imagination rather than memory.

Procedure:

1. **Pre-Task:** Demographic survey and baseline creativity assessment.
2. **Session 1 (15 min):** Ideation using **Condition A** (Text-Only).

3. **Session 2 (15 min):** Ideation using **Condition B** (Multimodal)
   Post-Task: Participants rate each tool using the Creativity Support Index (CSI).
4. **Debrief:** Semi-structured interview regarding "Sense of Ownership."

Metrics:

- **Quantitative:**
  - Fluency: Count of distinct design ideas generated per minute.
  - CSI Scores: specifically the "Expressiveness" and "Collaboration" sub-scales.
- **Qualitative:**
  - Blind Expert Rating: Three independent senior designers will rate the final image sets on a 1–7 scale for "Novelty" and "Aesthetic Quality."

4. Participants

- **Sample Size:** N=24 participants.
- **Demographics:**
  - **Group 1:** 12 Novice Designers (Graphic Designers) .
  - **Group 2:** 12 Non-Designers (General population).
- **Recruitment Criteria:** Must have normal color vision; no prior professional experience with Nano Banana Pro tools (to avoid bias).

5. Expected Outcomes

- **Hypothesis 1 (Expressiveness):** Users will rate the Multimodal condition significantly higher on the **CSI "Expressiveness" scale**, confirming that direct manipulation feels more natural than prompting.
- **Hypothesis 2 (Diversity):** Text-only outputs will likely cluster around visual stereotypes , while Multimodal outputs will show higher **geometric diversity** (unusual shapes derived from user sketches).
- **Hypothesis 3 (Agency):** Users will report a higher "Sense of Ownership" in the Multimodal condition, reducing the feeling of "Impostor Syndrome" often associated with AI art.

6. Contributions

We aim to provide:

1. **Empirical Evidence:** Quantitative data proving whether "Prompt Engineering" is an inferior interaction model for visual tasks compared to multimodal manipulation.
2. **Design Guidelines:** A set of heuristics for building future **"Co-Creative Canvases"** (e.g., best practices for mapping voice inputs to texture generation vs. geometry).
3. **Technical Validation:** Testing the limits of the **Nano Banana Pro Reasoning Core** in interpreting messy, imperfect human inputs.

7. Significance

- **HCI Researchers:** This work contributes to the growing field of Human-Computer Co-Creativity (HCC), moving the field beyond the "Chatbot" paradigm.
- **Software Vendors (Adobe/Figma/Canva):** The findings will validate the business case for developing "Generative Canvas" tools over simple text-input boxes.

References

1. **Carroll, E. A., & Latulipe, C. (2009).** Quantifying the creativity support of digital tools through the creativity support index. ACM Transactions on Computer-Human Interaction (TOCHI). ACM Digital Library

2. **Google Cloud. (2025).** Nano Banana Pro (Gemini 3 Pro Image) available for enterprise. Google Cloud Blog

3. **Q3 Technologies. (2025).** Top Multimodal AI Trends in 2025. Q3 Tech Blog. Q3 Technologies

4. **Ideum. (2024).** Before We Refine, We Explore: Five Benefits of Divergent Thinking During Concept Design. Ideum Journal
5. **Liu, V., et al. (2024).** Beyond Text-to-Image: Multimodal Prompts to Explore Generative AI. ResearchGate