Odin Vision

A Haptic and Sound-Based Approach to Art for the Visually Impaired

"The Blind Girl" (1856) by John Everett Millais symbolizes resilience, sensory perception beyond sight, and the contrast between physical blindness and emotional or spiritual insight.

Adya Aggarwal (2022043)
Pankhuri Singh (2022348)
Pranav Jain (2022365)

Problem Statement

"Art, as a medium of expression and culture, is largely inaccessible to individuals with visual impairments. Traditional solutions like bas-reliefs or audio descriptions lack the depth, color perception, and interactivity necessary to fully experience artwork. These existing tools either sacrifice detail or place high cognitive demands on the user."

This project addresses the following challenges:

- **Lack of immersive sensory engagement** for visually impaired individuals in art.
- **High cost and complexity** of current tactile or auditory feedback systems.
- **Cognitive overload** in interpreting color and **spatial information**.
- Lack of unified systems that integrate **haptics, sound, and visual data in real time**.

Research Questions

RQ1: How can we design a cost-effective haptic feedback system?
Can basic vibration motors simulate texture in a meaningful way for visually impaired users without relying on expensive hardware?

RQ2: How can we reduce cognitive overload during art exploration?
What design strategies can ensure that users are not overwhelmed by multisensory input, and instead experience art in an intuitive and engaging way?

RQ3: Can Large Language Models enhance user interaction?
Is it possible to integrate LLMs to provide real-time, contextual guidance and adaptive learning experiences while users explore artworks?

Methodology and Modalities of Interaction

Methodology phases:

Data Collection: *Interviews, surveys, and personas informed user-centric design.*
System Development: *Integration of computer vision, haptic motors, and audio output. Modular Python code with Arduino microcontroller communication.*
Hardware Integration: *ERM motors, Arduino, overhead webcam, and headphones.*
Chatbot Integration: *Vision-Language AI for context-aware Q&A.*
Evaluation: *Accuracy metrics, SHAP explainability, user studies, and performance benchmarking.*

Methodology and Modalities of Interaction

Modalities of Interaction:

Tactile: *Vibration motors simulate texture and depth.*
Auditory: *Sound cues convey color information.*
Linguistic: *Chatbot offers real-time voice/text interaction.*

Through a well-structured methodology and carefully selected interaction modalities, our system bridges the sensory gap in art appreciation. By combining ***touch****,* ***sound****, and* ***language****, it enables visually impaired users to explore visual art in a way that is both intuitive and immersive.*

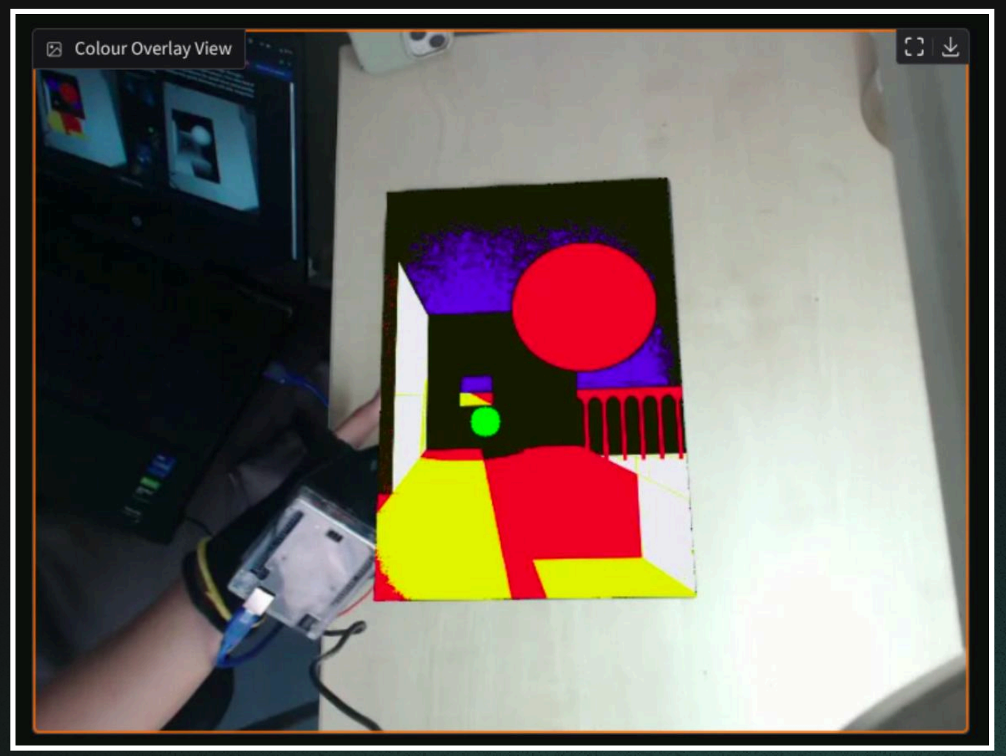
This multimodal design doesn't just improve accessibility it redefines how art can be experienced beyond sight, making inclusion a central part of artistic engagement.

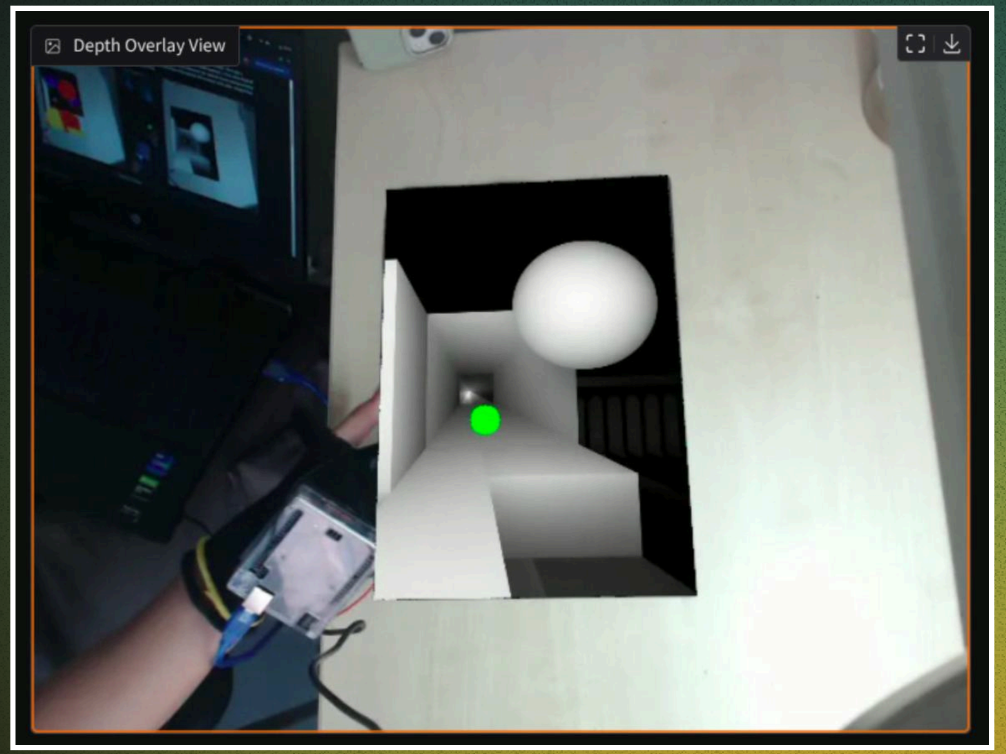
Working of the Project

This project introduces a multimodal system combining computer vision, depth perception, and real-time haptic-audio feedback to make art accessible.

Step-by-Step System Pipeline:

1. Artwork Setup & Camera Placement:
   - An artwork is placed on a flat surface.
   - A webcam above captures the image and finger movements.
2. Finger Tracking:
   - OpenCV and MediaPipe detect the fingertip (landmark 8) as the user explores the artwork.
3. Depth Estimation:
   - The image is processed using the SculptOK API to get a pixel-wise depth map.
   - This is normalized and mapped to vibration intensity using a bell-shaped function: mathematica and CopyEdit





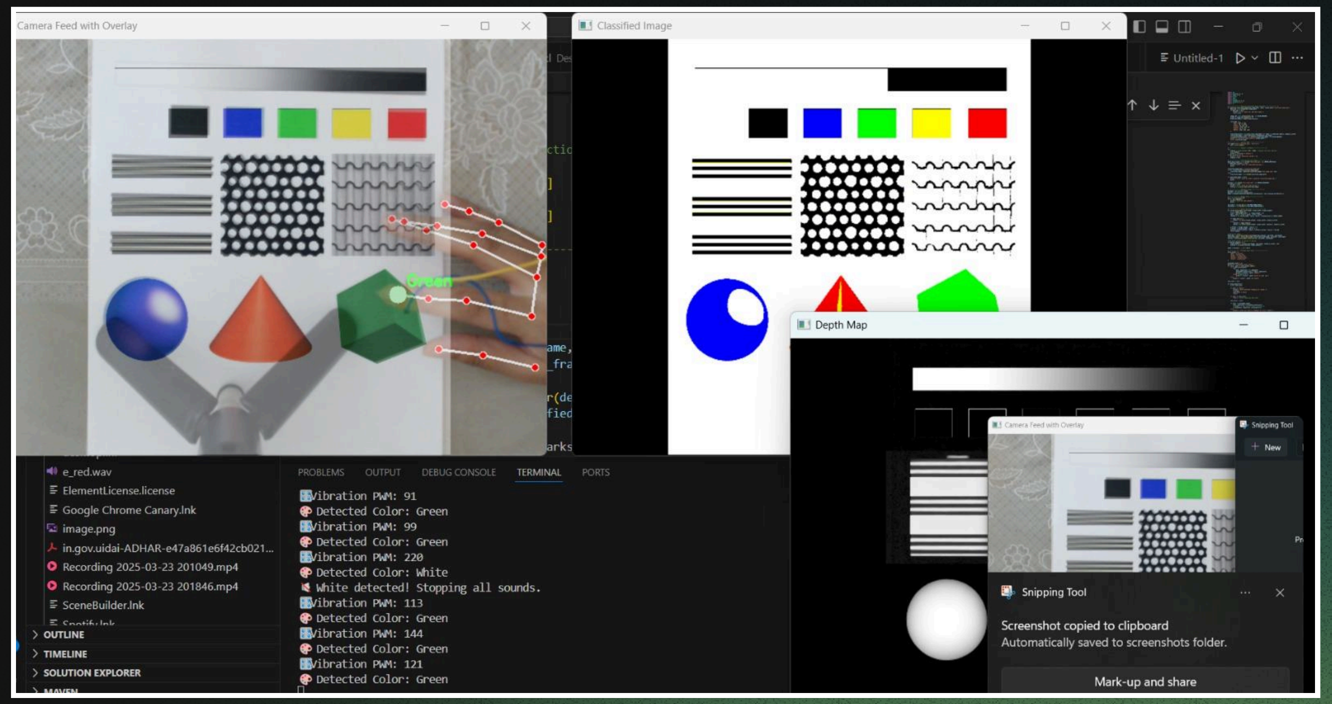
Working of the Project

4. Haptic Feedback:

- Vibration motors (ERM) are controlled via an Arduino UNO.
- The intensity changes with depth, allowing users to feel texture and structure.

5. Color Detection & Audio Mapping:

- KNN classifier identifies the color under the fingertip.
- Each color is mapped to a unique instrument sound:
  - Red → Guitar, Blue → Flute, Yellow → Piano, etc.
- Sounds are generated using the sounddevice Python library.6.



6. Integrated Feedback Loop:

- Tactile (vibration) and auditory (sound) feedback are simultaneously provided as users move their fingers.

7. Chatbot Integration:

- A vision-language pipeline (BLIP-2 + Mistral-7B) allows users to ask questions about the artwork and get contextual answers.

Dataset Description

To build and evaluate the hand-tracking component of our system, we utilized the HaGRID Sample Dataset (30K, 384p), available on Kaggle.

The HaGRID (Hand Gesture Recognition Image Dataset) is a large-scale dataset designed for gesture recognition tasks. It contains 30,000 annotated images at a resolution of 384p, featuring a wide range of real-world hand gestures captured under diverse lighting conditions, backgrounds, hand poses, and skin tones. This diversity makes it ideal for training and testing robust, inclusive computer vision models.

In our project, this dataset was used to test the performance of MediaPipe Hands, which identifies 21 3D hand landmarks in real time. Evaluation on HaGRID demonstrated the system's high precision (1.000) meaning all detected hands were true positives along with an F1 score of 0.805, reflecting strong performance in real-world conditions.

Code Evaluation & Explainability

Explainability:

- To ensure the model's decisions are transparent and understandable, the system uses SHAP (Shapley Additive exPlanations)—a game-theory-based framework for interpreting machine learning models.
- In this project, SHAP was applied to the KNN color classification model to evaluate how much each input feature—Red, Green, and Blue channel intensities—contributed to a specific prediction.
- Using SHAP improves the system's interpretability, aligning with HCI and ethical AI principles by making AI behavior understandable and predictable.

Code Evaluation & Explainability

- SHAP analysis revealed that the KNN classifier makes balanced decisions by considering all three RGB channels, rather than depending on a single dominant color.
- For example, the model was able to correctly classify a color as Green even when the green channel had a negative contribution—compensated by red and blue inputs. Similarly, Blue predictions were made despite low SHAP scores for blue, showing context-aware behavior.
- These insights prove that the model has learned nuanced color relationships and isn't relying on rigid thresholds. This interpretability ensures transparency and builds trust, especially in accessibility-focused applications where users rely on consistent and meaningful feedback.

Model Performance

MediaPipe Hand Tracking

0.805	0.674	1.000
F1 Score	*Recall*	*Precision*

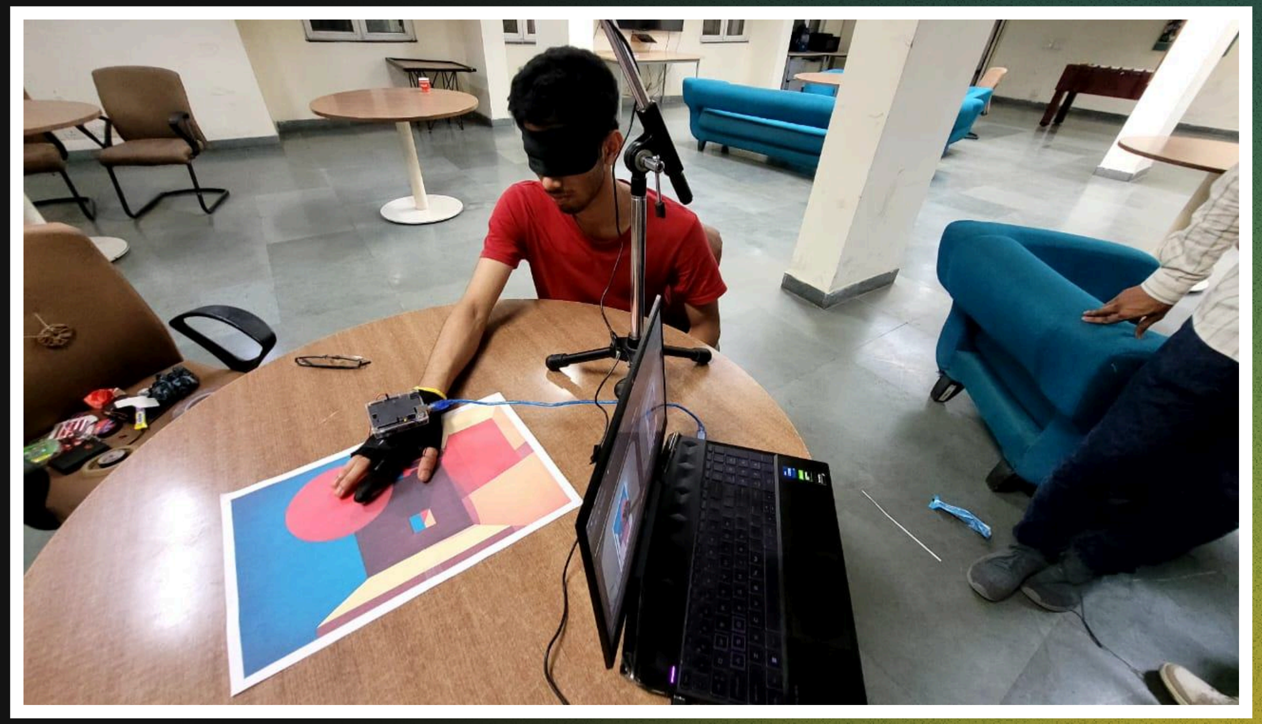
KNN for Color

0.9715	97.2%
F1 Score	*Accuracy*

Chatbot

0.32
BLEU Score

User Evaluation & HCI Principles



User Study Overview:

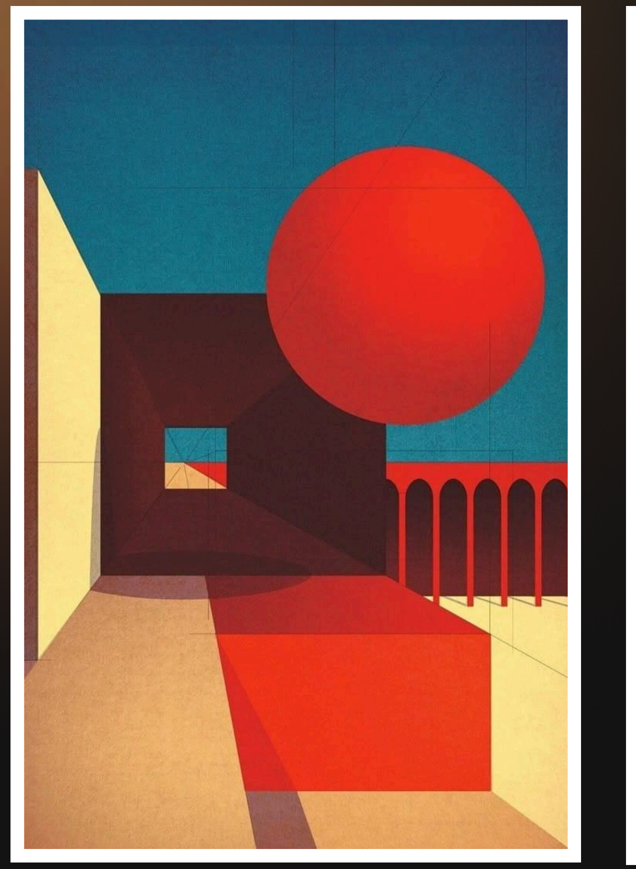
- *15 blindfolded participants interacted with the system and were asked to recreate the painting.*
- *Most could identify shapes and colors, but faced challenges in composition.*

Feedback Highlights:

- *72% of users were interested in real-world use.*
- *63% found vibration feedback somewhat clear.*
- *45% found auditory cues helpful.*

User Evaluation & HCI Principles

Artwork used



Recreated drawings

- Participants filled out a Google Form before and after the session to capture:
- Prior art experience
- Ability to recognize shapes/colors non-visually
- Clarity of vibration/audio cues

User Evaluation & HCI Principles

HCI Principles Incorporated:

- *Affordance: Rubber-dot textures on gloves provided intuitive tactile cues.*
- *Cognitive Load Reduction: Simplified color-sonification mapping reduced confusion.*
- *Spatial Awareness: Depth-based vibrations aided object recognition.*
- *Inclusivity: Diverse datasets (HaGRID) ensured robustness across users.*
- *Multimodal Interaction: Voice interfaces (STT + TTS) enhanced accessibility.*
- *System Feedback: Real-time error notifications and system status indicators improved trust and usability.*
- *User-Centered Design: Feedback loops influenced ongoing system iterations.*

Conclusion & Future Scope

Our project began with a clear goal: to create a low-cost, accessible haptic glove that enables individuals with visual impairments to experience visual art through touch and sound. By using affordable components like ERM motors and an Arduino microcontroller, we were able to develop a working prototype that translates depth and color into multisensory feedback. This budget-conscious approach makes the system suitable for deployment in resource-constrained settings such as schools, community centers, and public exhibitions.

One of the most critical next steps is to engage directly with visually impaired individuals. While our current testing relied on blindfolded sighted participants, real users with visual impairments bring unique perspectives that cannot be fully simulated. By conducting studies and co-design sessions with these users, we aim to gather authentic feedback that can inform more meaningful design improvements

This project was developed with the goal of contributing to CHI by exploring how a low-cost haptic glove can enhance art accessibility for individuals with visual impairments. Our focus was on creating an affordable, inclusive solution that aligns with CHI's values of human-centered design and accessible technology.