SCI 6506: Design Analytics: Predicting Human Spatial Experience

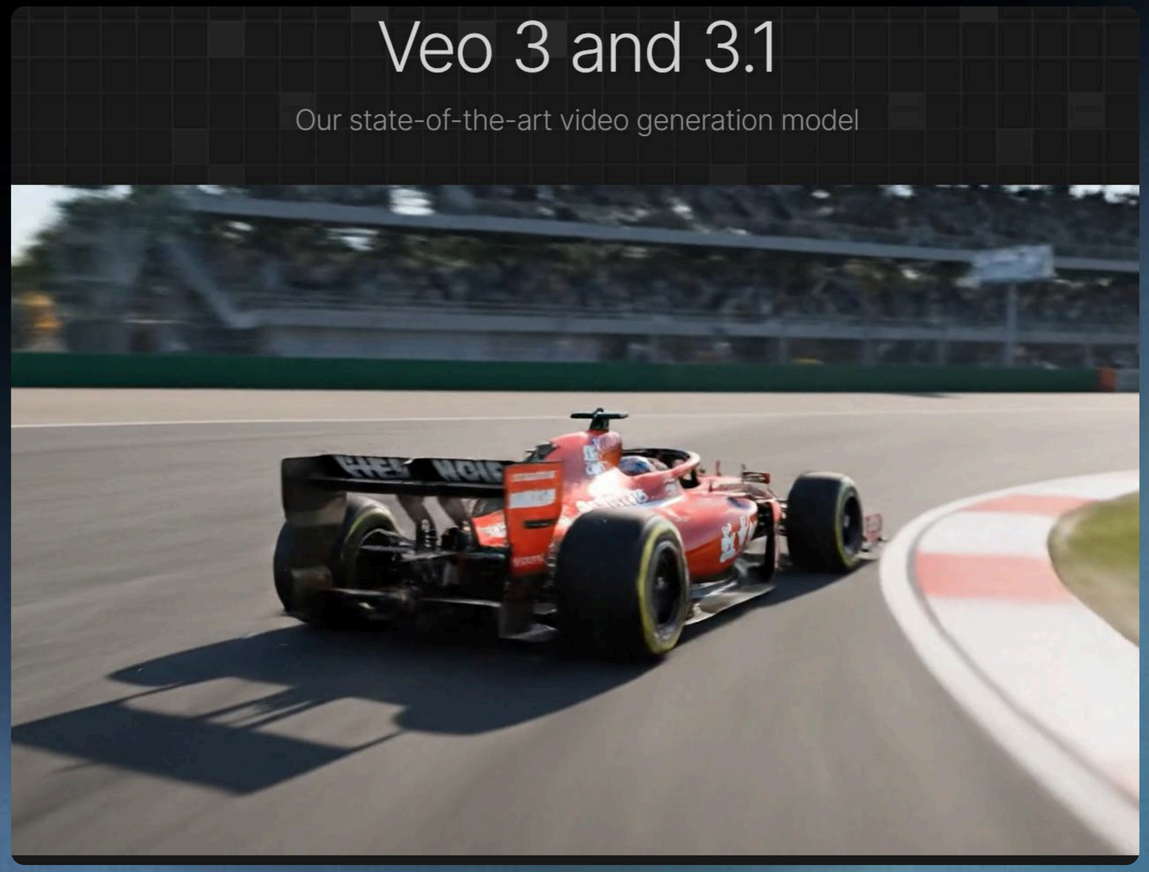
Cinematic Spaces in the AI Lens: Mapping Arousal & Perception

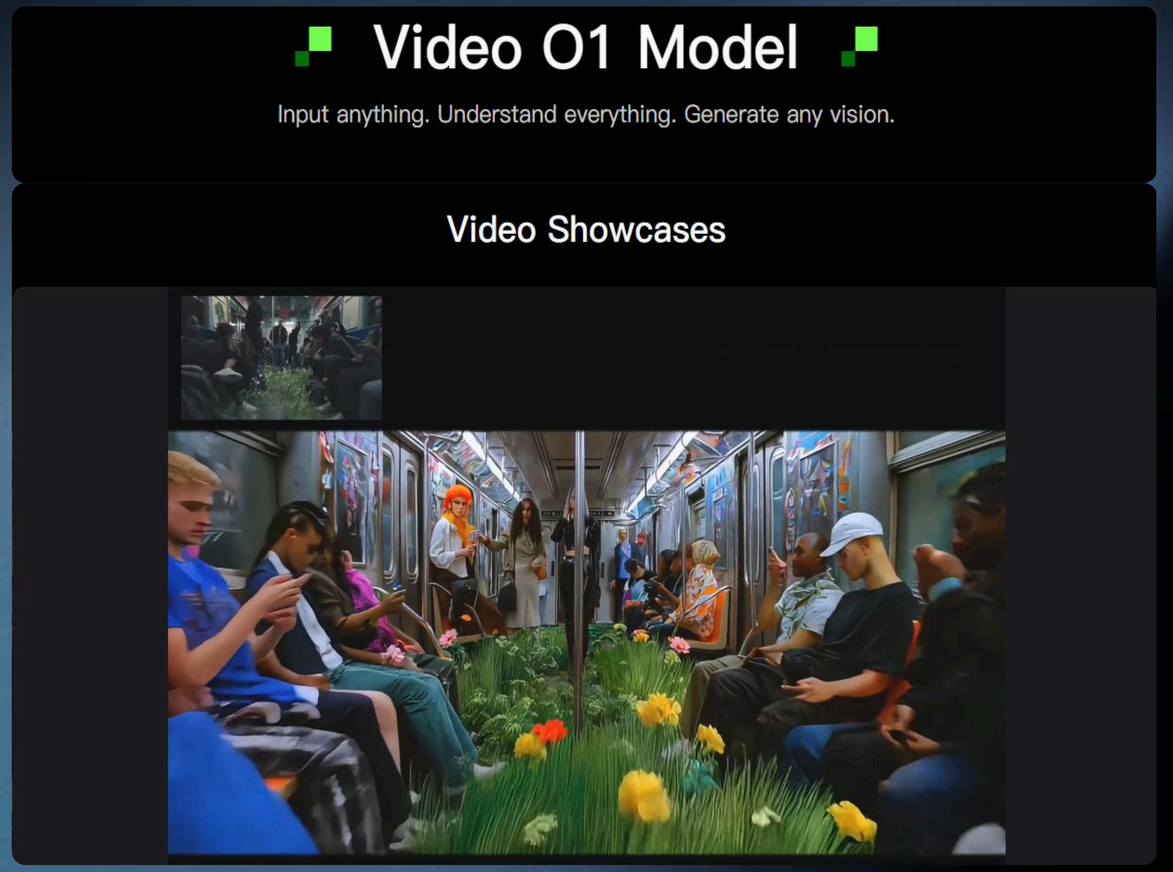
Background

Machine Vision vs. Human Perception

Aspect	Human (On-site)	Traditional CV	VLM
Perception Basis	Multi-sensory, stereoscopic, holistic	Segmentation + pre-trained weight calculation	Image-based scene understanding
Strengths	Dynamic, rich dimensionality	Structured scoring logic	Superior spatial relation & scene recognition
Limitations	High subjectivity, hard to statisticalize	Poor art museum adaptation; recognition errors	Overly general; lacks true subjectivity
Key Issue	Subjectivity affecting data stats	Generic for special spaces (e.g., art museums)	Likely due to over-generalized training data

Background

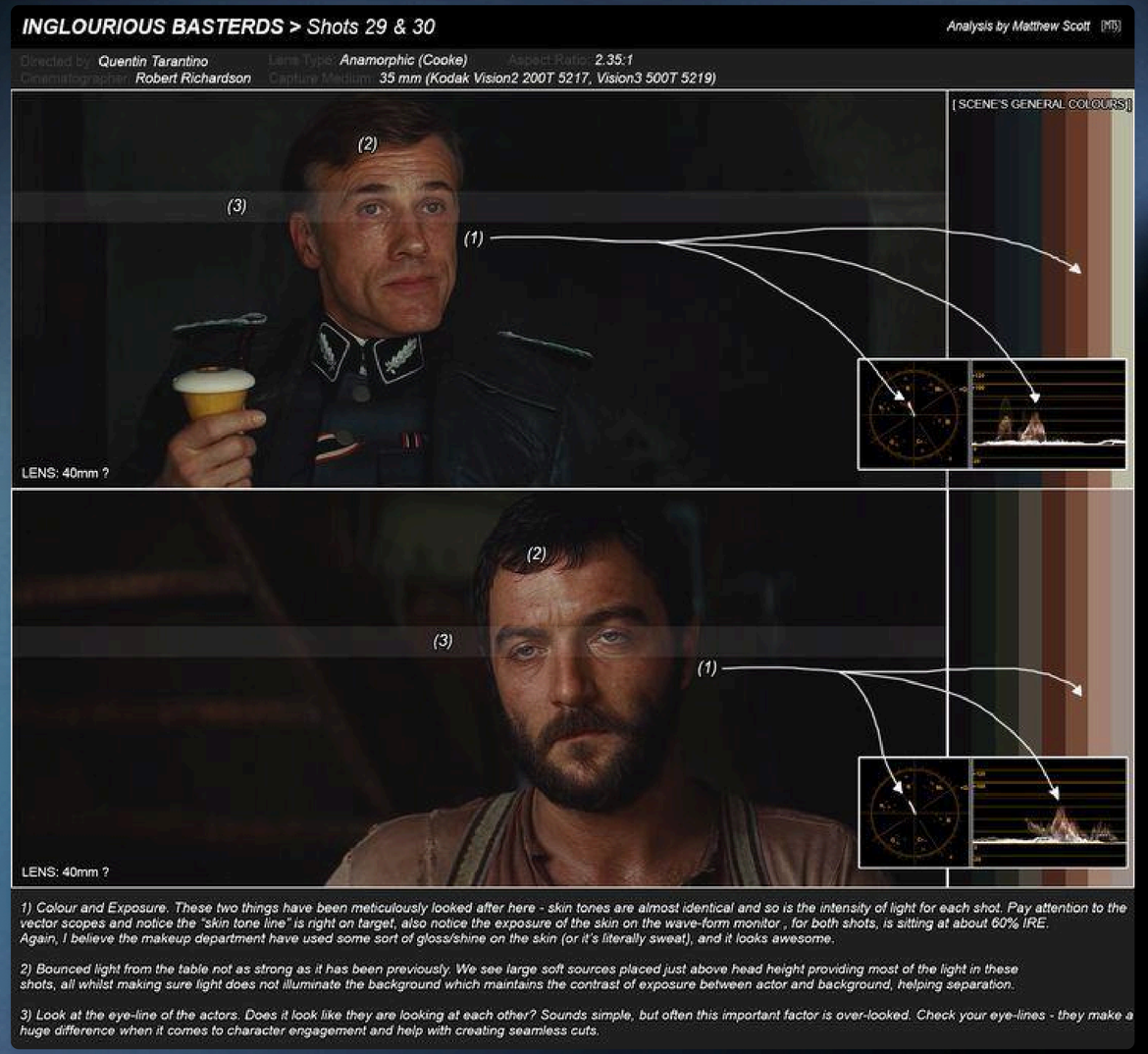




From "AI Perception" to "Perception of AI": A Paradigm Shift

Hypothesis

Spatial Elements (lighting, setting, materials) Determine Arousal & Valence



Hypothesis

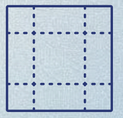
Liminal Space Threat

Research Question(s)

1. Impact (Primary)

How do core **spatial variables** (lighting, setting, texture) affect viewers' **Emotional Arousal** and **Spatial Perception** in **AI-generated spatial videos**?

2. Pattern



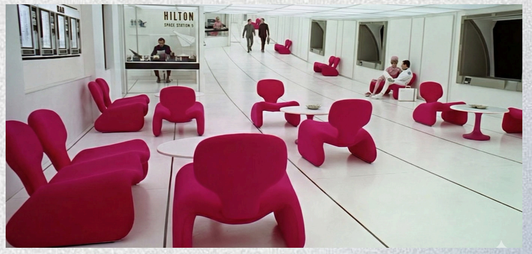
*Do cross-scenario **common perceptual patterns** exist when viewing cinematic spaces with controlled variables?*

3. Visual Attention

*Can changes in Visual Attention (Eye-tracking) explain the **physiological mechanism** of the emotional responses?*

8 Movies

Swan Song



2001: A Space Odyssey

The Grand Budapest Hotel

Happy Together

Severance

F1: The Movie

Parasite

The Shining

Theme 1
futuristic

Theme 2
dreamy

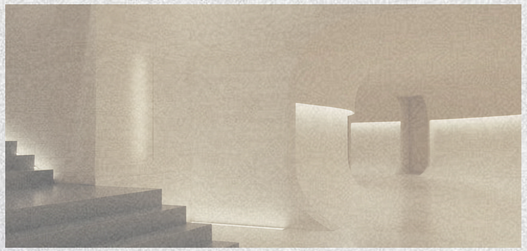
Theme 3
documentary

Theme 4
horrific

8 Movies x 4 Variables

Tools

Google Nano Banana (image editting) → Veo3 (frame to video)

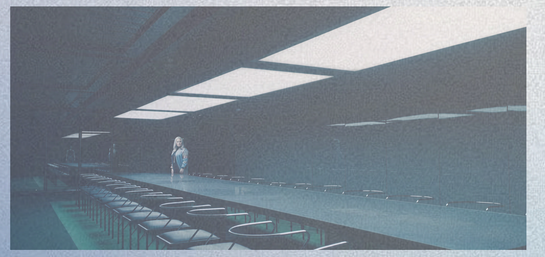


Swan Song

2001: A Space Odyssey

The Grand Budapest Hotel

Happy Together



Severance

F1: The Movie

Parasite

The Shining

Remove Subject

Setting

Lighting

Texture

A

B

C

D

8 Movies x 4 Variables = 32 Video Clips

	Swan Song	Happy Together	The Grand Budapest Hotel	2001: A Space Odyssey	F1: The Movie	Severance	Parasite	The Shining
A	 Original	 Original	 Original	 Original	 Original	 Original	 Original	 Original
B	 Setting	 Setting	 Setting	 Setting	 Setting	 Setting	 Setting	 Setting
C	 Light	 Light	 Light	 Light	 Light	 Light	 Light	 Light
D	 Material	 Material	 Material	 Material	 Material	 Material	 Material	 Material
	Movie 1	**Movie 2**	**Movie 3**	**Movie 4**	**Movie 5**	**Movie 6**	**Movie 7**	**Movie 8**

User Study Design

EXPERIMENTAL DESIGN: MITIGATING FATIGUE & HABITUATION

A. PARTICIPANTS & STIMULI

PARTICIPANTS: 20-30 University Students & Young Professionals

STIMULI: 32 VIDEO CLIPS

4 VARIATIONS			
8 CLASSIC MOVIE SEENES	Original	Lighting-altered	Set altred Materiof altred
• The Stmiing • Psycho • Psycho • Raeriencic Scaes			
• Prycke • Ligening altred • Raertins Mnaussons • Scenicties			
• Thew Rowich Sooses • Mar tinmes • Counting Snters • Meserial altred			
• The Sohoctor the xocies • The Sotwring • Mew Inby sfoines			

e.g, "The Shining" Original, Lighting-altred, etc

B. EXPERIMENTAL STRATEGY

BLOCKED RANDOMIZATION + BALANCED LATIN SQUARE

GROUPING

- Divide 32 videos into Blocks (8-12 videos/block)
- Each Block covers different scene variation per scene

Arranges playvack oder

- Eliminates Eliminates Order Effect

LATIN SQUARE

- Example: Part, A sees 'Shining' Original first, Part. B sees B sees 'Shining' Lighting-altred first

A	B	C
A	B	E
C	B	C

REST PERIODS (15-30s BLACK SCREEN) BETWEEN VIDEOS

Allows EDA levels to return to Baseline

MULTIMIODAL DATA COLLECTION

SUBJECTIVE (SELF-REPORT SURVEY)

Completed immeduately after each video

Emotional Metrics:

- **Valence** (Positive/Negative)
- **Arouisa**/Excited

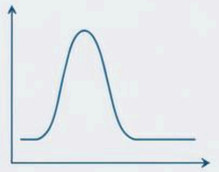
Spatial Metrics:

- Thermal feel (CoolWarm) Threat level (Safe/Threatening) Realism
- Realism Artificial (Realistic)

PHYSALOGICAL (EDA/GSR)

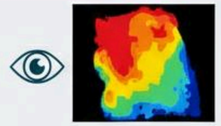
GROUPING

- Monitors Skin Condutance Response SCR peck/block) peak and recovery time



- Quantifies subescunous physslogical arousal
- Quantifies SCR slope changes during video zoom-in

BEHAVROAL (EYE-TRACKING)



Heatmaps & Fixation Duration

- Observe gaze patterns when 'set 'set' disappseres or mdeo changes
- Is gaze diffuse (seking clues) or focused specific threat points?

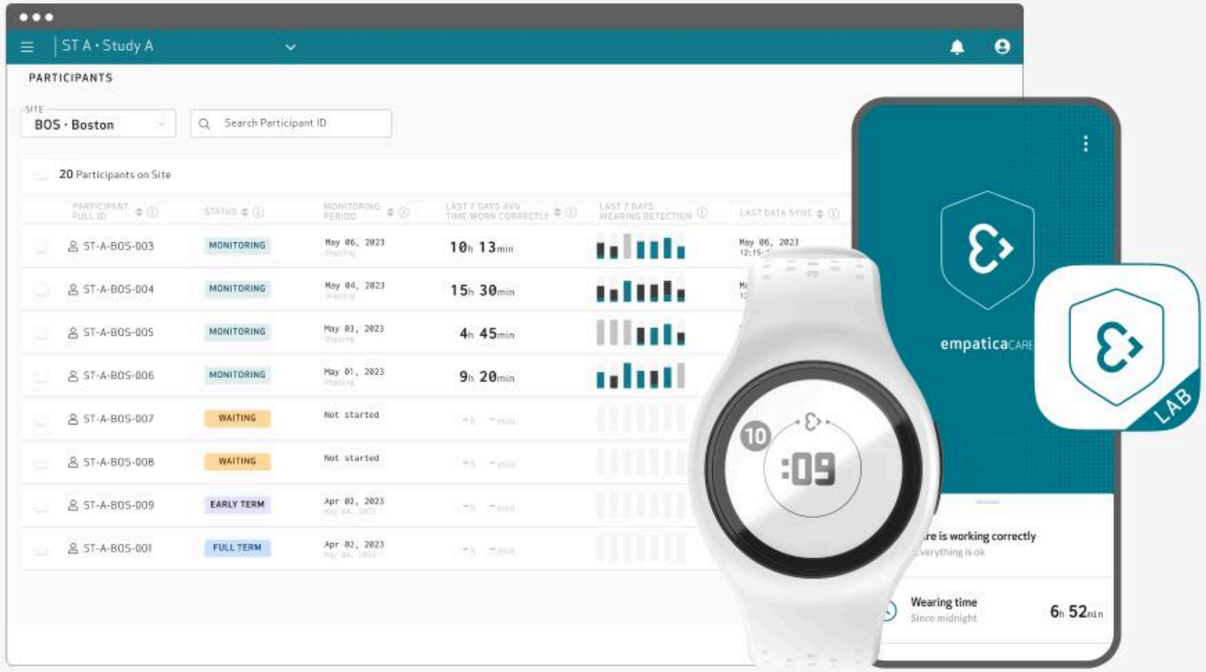
To address the challenges of **Fatigue Effect and Habituation in EDA data, we adopt a Mixed Method combined with Balanced Incomplete Block Design.*

Data Collection

- Survey (qualtrics)
- Eye-Tracking (RealEye.io)
- EDA (Empatica Plus)

qualtrics XM

RealEye



One video example

User Gender | User Age

Q1: How did the video **make you feel emotionally**? (rate from 1-9, 1 means Very Negative / Unpleasant; 5 means neutral; 9 means Very Positive / Pleasant)	Q2: How **emotionally calm or excited** did you feel while watching the video? (rate from 1-9, 1 means Very Calm / Sleepy; 5 means neutral; 9 means Very Excited / Alert)	Q3: How did the **atmosphere of the space** feel to you? (rate from 1-9, 1 means Very Cold / Detached; 5 means neutral; 9 means Very Warm / Inviting)	Q4: How **safe or threatened** did you feel in this space? (rate from 1-9, 1 means Very Safe / Relaxed; 5 means neutral; 9 means Very Threatening / Oppressive)	Q5: How **realistic** did the space appear to you? (rate from 1-9, 1 means Completely Fake / Glitchy; 5 means neutral; 9 means Very Realistic)	Q6: Did anything in the video particularly **influence how you felt**? Please describe. (Enter N/A if nothing)
1	1	1	9	7	The dark end

Group 1 | Block 1 | Video 8_B

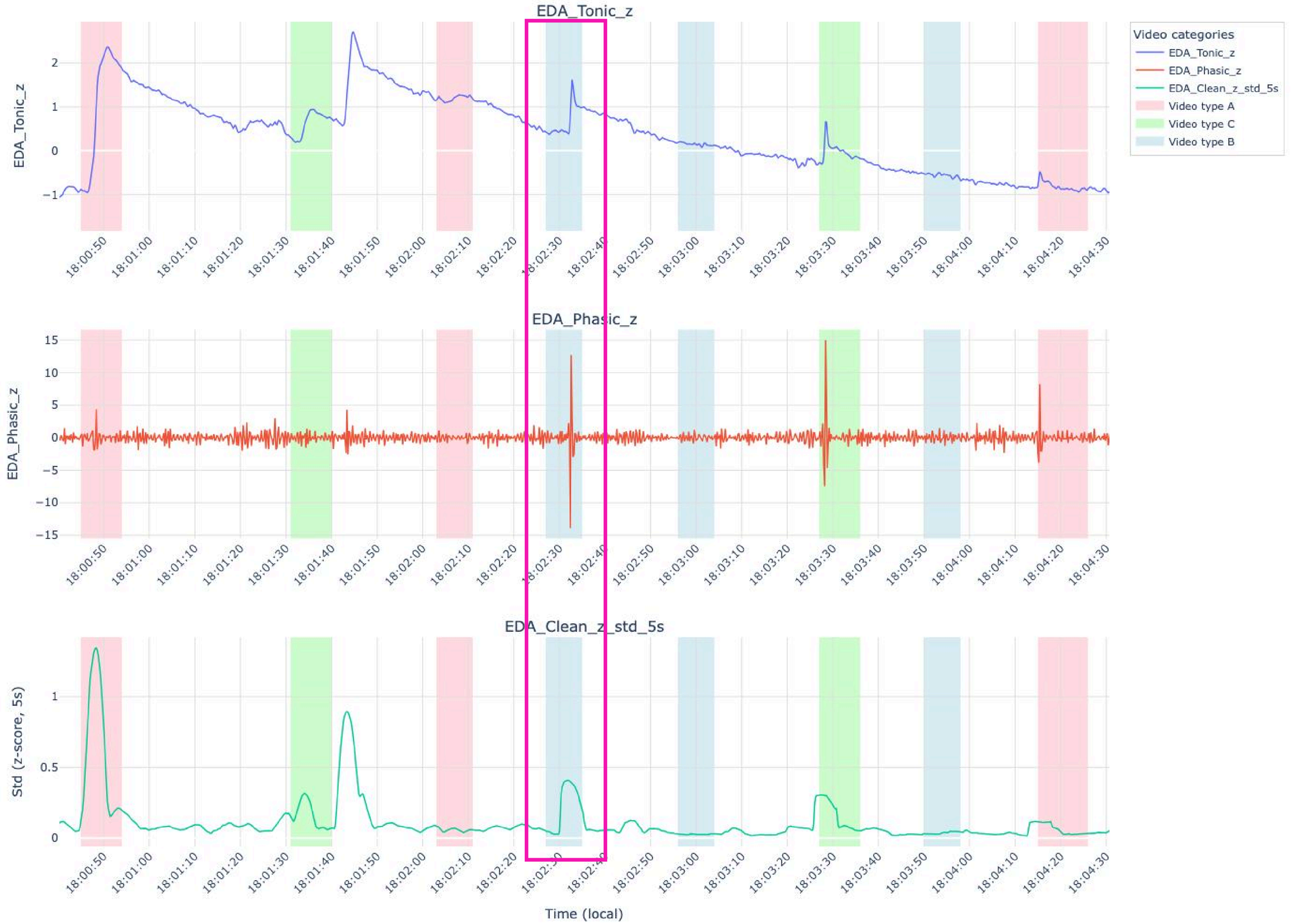
One video

raw EDA

SCL (Tonic)

SCR (Phasic)

Variability




1

Data Collection

26 groups of self-reported data

22 groups of synchronized eye-tracking data

8 groups of synchronized physiological (EDA) data

a group = **24 (8 × 3) videos*

**a group here means one participant*

~600 data points collected

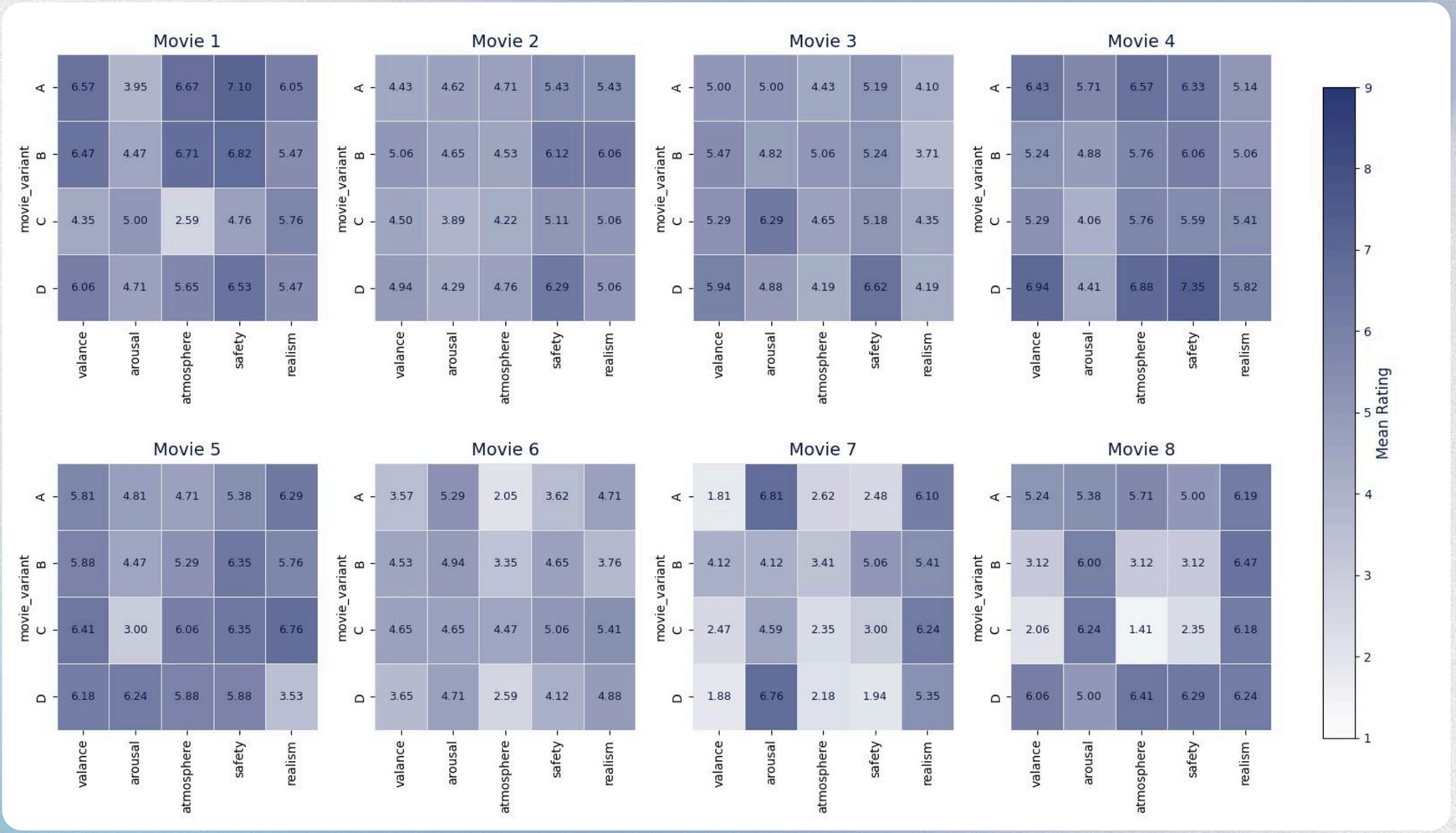
Results

Self-Reported Data

A: original
B: setting
C: lighting
D: texture

- Variant effects are consistent, but movie context amplifies differences.
- Movie 1, 4, 5, 7, 8, show larger spreads between variants → visual changes strongly shift perception.
- Movie 2, 3, 6 show narrower spreads → baseline emotion tone of the movie dominates.

Mean Difference



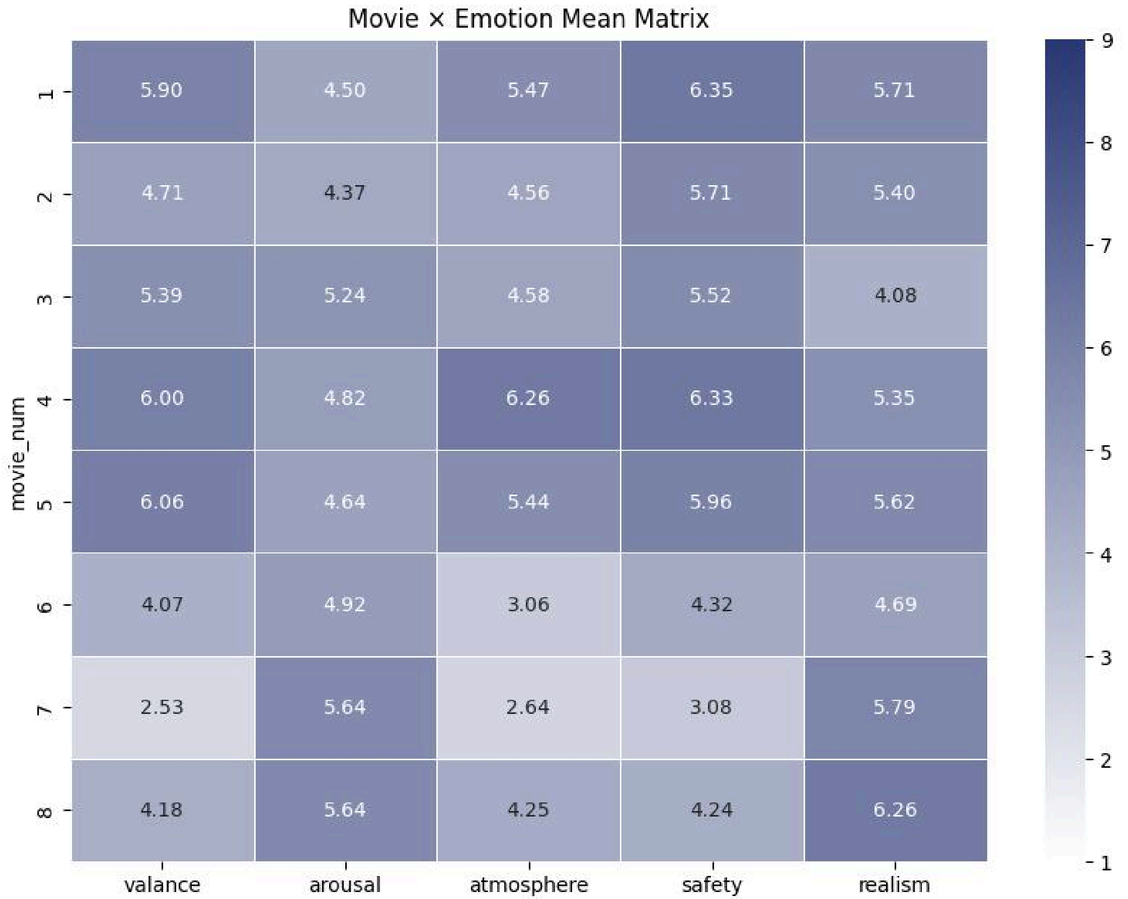
Results

Self-Reported Data

A: original
B: setting
C: lighting
D: texture

- **Movies** have distinct emotional signatures, Movie 1/4/5 highest; Movie 6/7 lowest.
- **Atmosphere** varies the most across movies; **arousal** varies the least.
- **Valence and safety** rise and fall together, forming consistent emotional patterns across scenes.

Movie × Emotion Mean Matrix



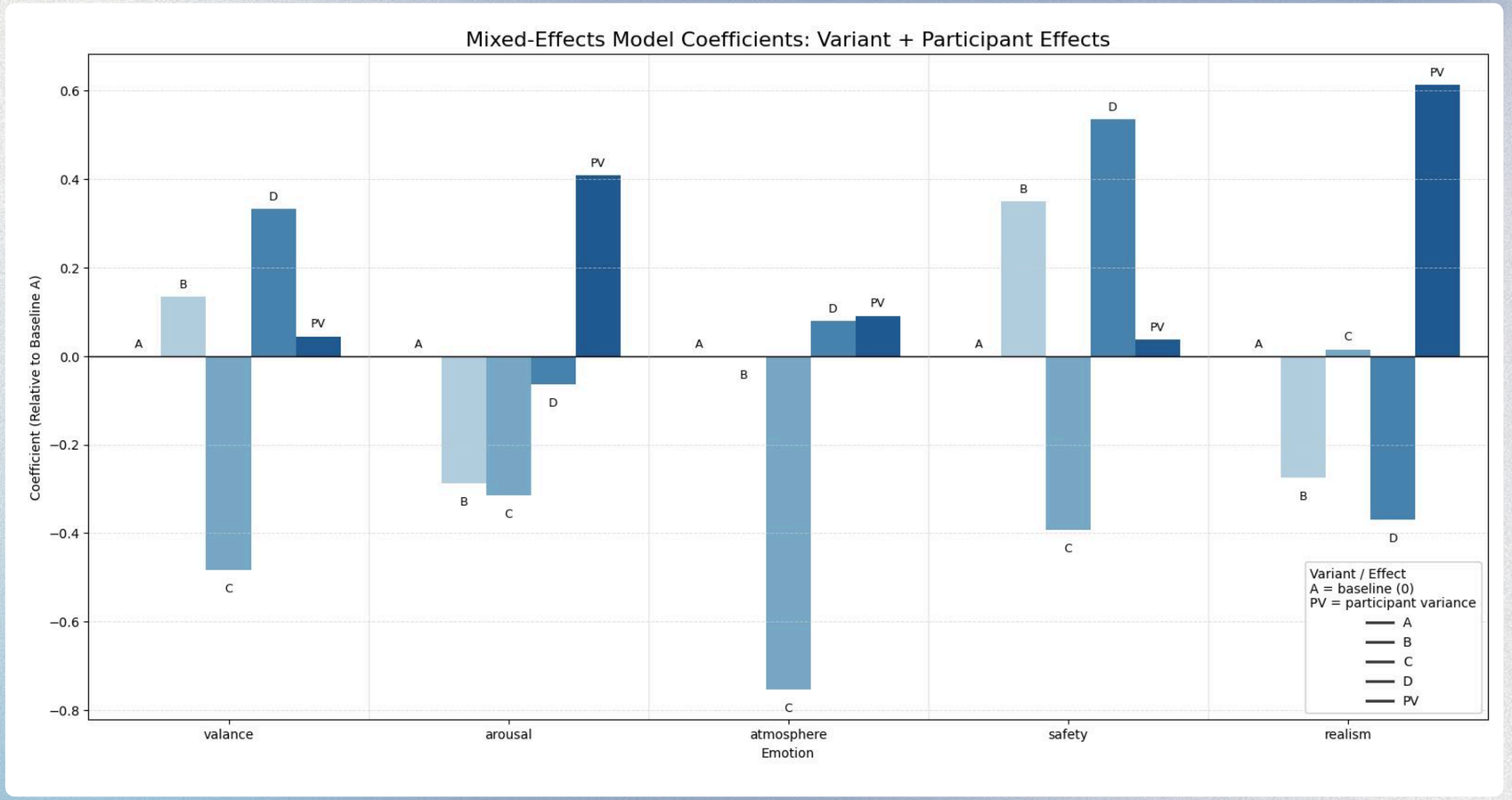
Results

Self-Reported Data

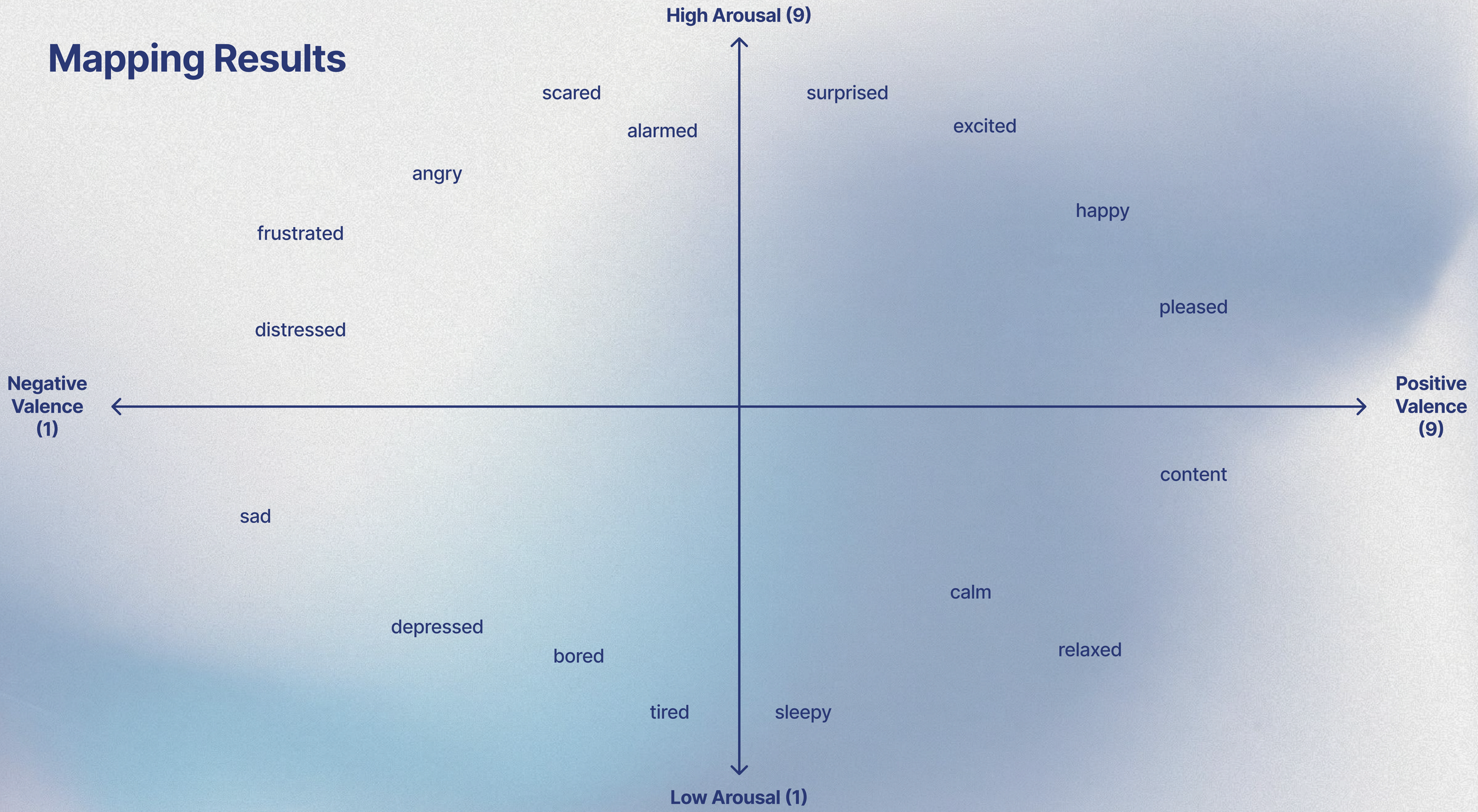
A: original
B: setting
C: lighting
D: texture

- B and D generally increase positive emotion ratings.
- C consistently reduces valence, atmosphere, and safety.
- **Lighting** is the most disruptive factor; **texture** is the most enhancing.
- Participant variance (PV) is the largest contributor → strong individual differences.

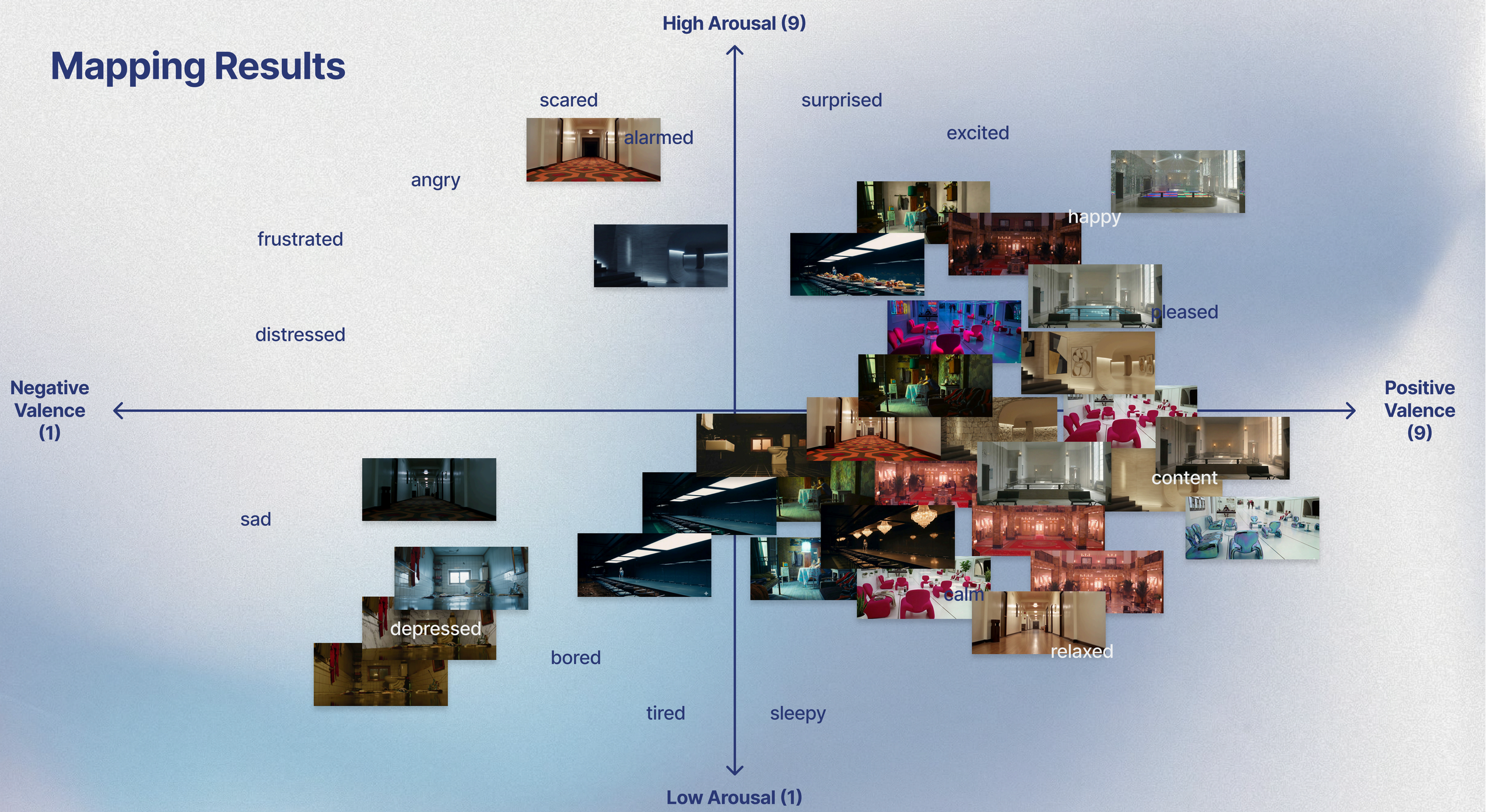
Mixed-Effect (Variant vs. Emotion)



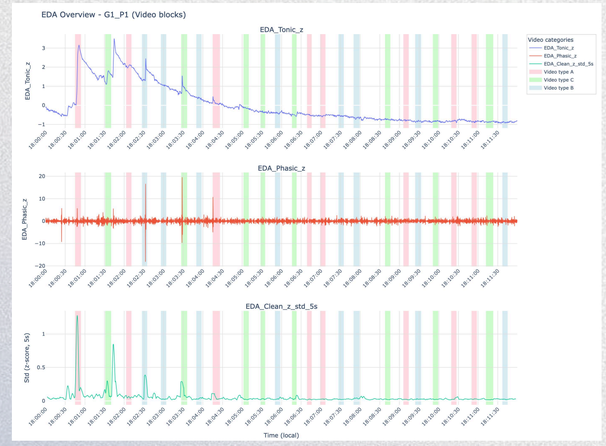
Mapping Results

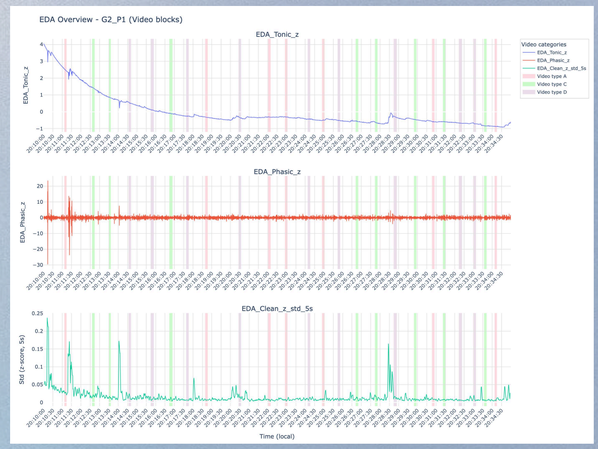


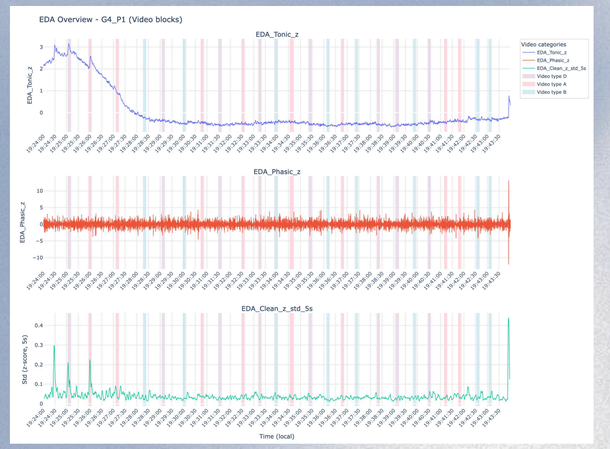
Mapping Results

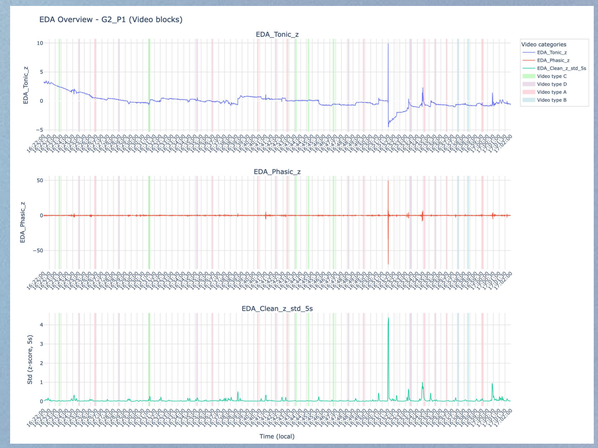


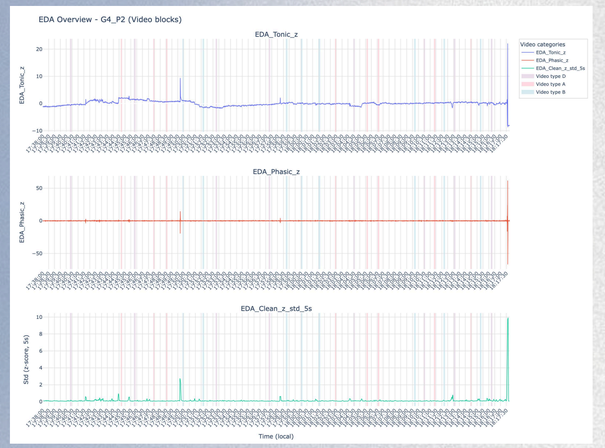
EDA plots of 8 participants







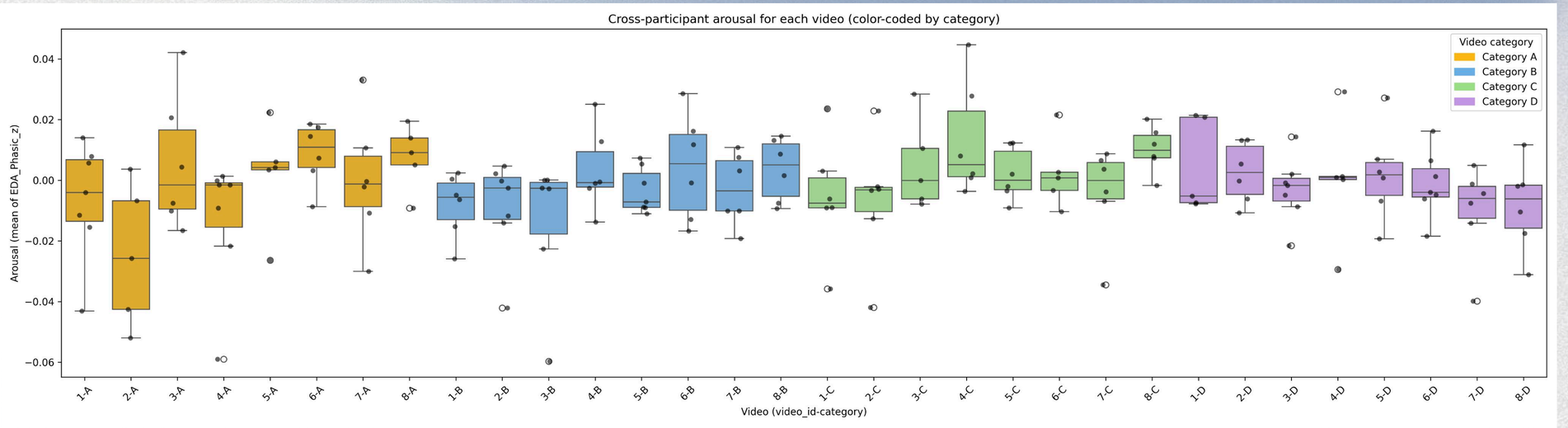




Results

EDA (SCL)

A: original
B: setting
C: lighting
D: texture



- Similar to self-reported data, we found that the original scenes (1-A to 8-A) are the ones that trigger most variant arousal effects. While the AI-modified scenes (B/C/D) are weakening this diversity, potentially leading to a **homogenization** of viewers' emotional responses.
- *More data is needed for verification.*

Results

Eye-Tracking Data

Visual anchoring as a mediator between spatial video and arousal?

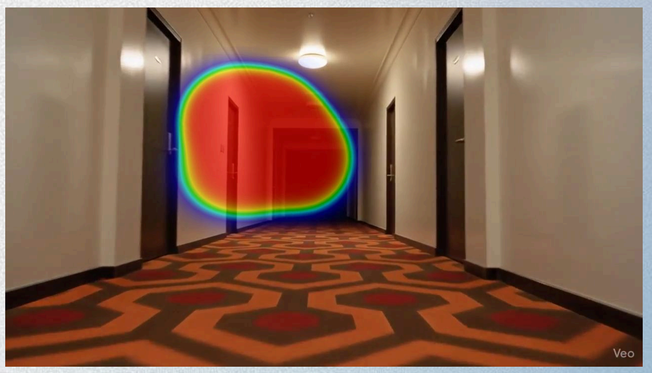
- **The Visual Anchoring (Structural)**
  - In furnished environments, objects serve as **"semantic anchors"** that stabilize gaze.
  - Removal of furniture forces the eye into a continuous search for **spatial definition**.
- **The Optic Flow Consistency (Dynamic)**
  - During a "Slow Zoom-in," the eye naturally seeks a **stable "Focus of Expansion"** (center point).
  - In minimal/empty AI spaces, **the lack of a central focal point** creates a conflict between the forward motion and the eye's inability to lock on.
- **The Cinematic Compression (Compositional)**
  - Directors use **lens compression and depth-of-field** to flatten space and dictate user focus to specific zones (Forced Perspective).
  - Altering materials or lighting **breaks this intentional "focal guidance,"** expanding the user's region of interest (ROI) unintentionally.
  - The eye attempts to re-calculate depth in a scene that has **lost its compositional hierarchy**.

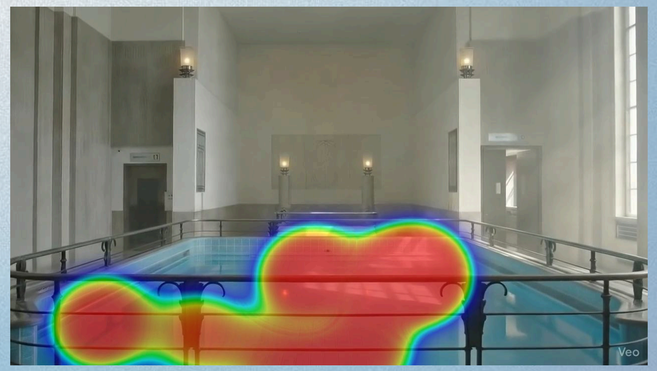
Results

Eye-Tracking Data

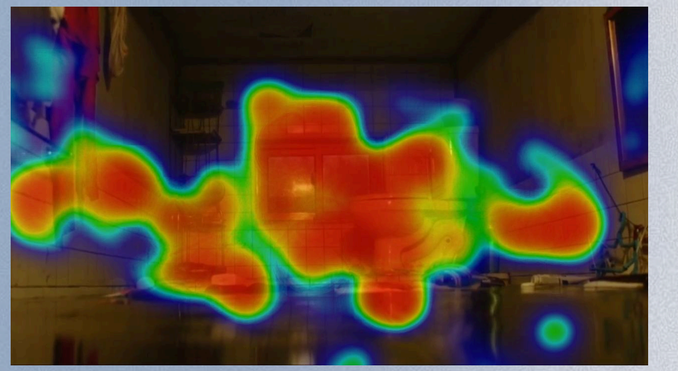
Two types of visual anchoring patterns

focus





scanning/browsing



Conclusions

“Spatial variables shape viewer’s emotion, but effects are scene-dependent.”

- Core spatial factors (lighting, setting, texture) significantly influence emotional perception, however, their impact depends on **the original movie’s emotional baseline.**

“Lighting disrupts viewer’s emotion most, while texture enhances it.”

- Across all movies, **lighting changes consistently reduce valence, atmosphere, and perceived safety**, making it the most disruptive variable. In contrast, material/texture modifications tend to enhance positive emotional responses.

Conclusions

“Stable emotional structures exist across movies.”

- Despite stylistic differences, **valence and safety strongly co-vary across scenes**, forming consistent emotional patterns.
- Arousal is comparatively stable and less sensitive to spatial modifications than other emotional dimensions.

“Visual attention mediates spatial-emotional effects”

- Eye-tracking suggests that visual anchoring plays a key mediating role: **spaces with clear semantic or structural anchors stabilize gaze and emotion**, while anchor-less AI-modified spaces induce exploratory scanning and emotional variability.

Limitations

1. Strong baseline coherence of cinematic scenes

- The selected movie scenes are highly curated and emotionally coherent by design. As a result, any spatial modifications may disrupt this already **“optimized” harmony**, leading to lower perception scores regardless of the direction of change.

2. Inconsistent control within spatial variables

- Although variables were categorized as lighting, setting, or texture, the specific **level and direction of change** was not consistent across scenes (e.g., lighting becoming darker in some cases and brighter in others). This inconsistency reduces result reliability and limits direct comparability across conditions.

Limitations

3. Cinematic narrative and familiarity bias

- Movie scenes inherently carry narrative context, and **prior exposure** to the films cannot be ruled out. Both factors may introduce **personal bias** and **memory-driven** emotional responses beyond spatial perception alone.

4. Limited control over camera motion

- **AI-generated videos** still exhibit stochastic variation in camera movement and framing, making it difficult to fully standardize motion trajectories across spatial conditions.

Next Steps

1. Shift to everyday video baselines

- Future studies will use **casual, vlog-style recordings** of everyday spaces as baselines, reducing cinematic bias and narrative influence. AI will then be applied to selectively modify spatial elements for controlled comparison.

2. Structured variable grouping and analysis

- Spatial modifications will be explicitly grouped by direction and magnitude (e.g., darker vs. brighter lighting), enabling more reliable **within-group and cross-group** analysis.

3. Deeper analysis of visual attention mechanisms

- We will further investigate **visual anchoring and attention patterns** using eye-tracking data to understand how gaze stability, exploration, and focal structure mediate emotional responses to spatial changes.

SCI6506: Design Analytics: Predicting Human Spatial Experience

Cinematic Spaces in the AI Lens: Mapping Arousal & Perception

THANK YOU