Multimodal Gaze-Adaptive Content Summarizer

Fusing Eye-Tracking and Heart Rate to Infer Reader State

Students: Filip David-Liviu
Deaconu Victor
Croitor David

Concept & Workflow

The Interface

The application renders text into discrete, indexed paragraphs on the client side. This structure allowing us to isolate exactly where the user is struggling.

Monitoring

As the user reads Paragraph #3, the system gathers gaze fixations and heart rate data specific to that window of time

Goal

Instead of summarizing the whole text, we target the specific paragraph triggering the stress response.

Defining the Multimodal Signatures

Confusion

Marked by rapid re-reading of specific sections and a high regression rate in the gaze pattern, accompanied by a rising heart rate as increased cognitive load elevates metabolic demand.

Frustration

Reflected in frequent scanning of interface elements paired with erratic clicking behavior, while the heart shows a sudden spike in rate consistent with an acute stress response.

Sleepiness

Appears through prolonged, drifting fixations that indicate zoning out, along with a low or gradually declining heart rate that signals reduced physiological arousal.

Research Questions & Hypotheses

Multimodal Efficacy

Does the fusion of physiological arousal data (Heart Rate) with behavioral gaze data significantly improve the classification accuracy of “Frustration” or “Confusion” compared to Gaze data alone?

Intervention Impact

Does the presentation of LLM-simplified text lead to a measurable physiological recovery (Heart Rate reduction) in users experiencing high cognitive load?

Low-Fidelity Feasibility

Can consumer-grade webcam eye-tracking achieve sufficient spatial resolution to reliably distinguish inter-paragraph revisits (indicators of confusion) from vertical scanning behavior (skimming)?

System Architecture: The Loop

1. SENSE

Client-Side
Browser captures Gaze (x, y) (Webcam) and Heart Rate (Polar Verity via Bluetooth)

2. ANALYZE

Server-Side
Python script performs Data Fusion. Synchronization of timestamps and Random Forest Classification.

3. RESPOND

Gen AI
Gemini API generates adaptive summaries for high-stress/confusion segments.

Stage 1: SENSE

Data Capture Strategy

We utilize a **Polar Verity Sense** device in order to track the users heart rate, and **WebGazer.js** to turn a standard webcam into an eye-tracker.

- **Timestamp:** Precise timing
- **Gaze Coordinates:** Predicted (x, y) on screen
- **AOI ID:** The specific paragraph being read
- **Interaction Flags:** Clicks & Scrolls (Context)
- **Heart Rate (bpm):** Real-time heart rate sample from Polar Verity sensor

Training Strategy

Challenge: Collecting thousands of real-world samples is impractical for a prototype. Instead, we use transfer learning based on high-quality public research datasets.

Solution: We pre-train on the MAHNOB-HCI dataset, which provides synchronized Eye-Tracking and ECG signals. We map its "Arousal/Valence" labels to our target states.

Deployment: These features train a Random Forest classifier. The model is exported to the Python server, enabling immediate inference without requiring user-specific data collection.

Stage 2: ANALYZE

The Brain

We employ a **Random Forest Multiclass Classifier**. This model is ideal for handling the noisy nature of webcam-based eye-tracking data.

Logic & Processing

Input: A Feature Vector representing a 10-second sliding window (Mean Fixation Duration, Re-read Frequency, Mean Heart Rate, Heart Rate Delta).

Output: Integer Class (1=Interest, 2=Confusion, 3=Frustration, 4=Sleepiness).

Why Offline? By moving heavy processing to the server, we prevent browser lag.

Stage 3: RESPOND (Adaptive Summarization)

The Intervention

When the model detects **Confusion** or **Frustration** on a specific paragraph, the system triggers the **Gemini API**.

Prompt Strategy

> "Based on a reader's state of CONFUSION, explain the following text more simply. Reduce Flesch-Kincaid grade level to 10th grade."

References & Technical Foundation

The Sensing Layer:

- **Papoutsaki, A., et al. (2016).** “WebGazer: Scalable Webcam Eye Tracking Using User Interactions.”
  - Validates the feasibility of browser-based tracking without external hardware.

Training Datasets:

- **Soleymani, M., et al. (2012).** “A Multimodal Database for Affect Recognition and Implicit Tagging.”
  - Validates the availability of synchronized eye-tracking and physiological signals for affect detection.

Behavioral Validations:

- ***Confusion:* Rayner, K. (1998).**“Eye movements in reading and information processing: 20 years of research.”
  - Validates "regression rates" as a direct index of cognitive processing difficulty.
- ***Frustration:* Scheirer, J., et al. (2002).** “Frustrating the User on Purpose: A Step Toward Building an Affective Computer.”
  - Validates the correlation between interface failure, erratic behavior, and acute physiological stress.

Conclusion

A fully integrated multimodal system that **Senses** struggle via gaze and heart rate, **Quantifies** it via Machine Learning, and **Resolves** it via GenAI adaptation.