Geometric Evidence of Neural Identity: A Distance-Based Analysis of Emotional EEG

DREAMER Data Analysis Project

December 8, 2025

Abstract

Research Question: Can we identify individuals from their EEG responses during emotional stimulation? **Methods:** We addressed this question by analyzing the geometric structure of the EEG feature space in the DREAMER dataset (23 subjects, 18 videos). Deviating from opaque machine learning models, we introduced a transparent **Distance Discrimination Method (DDM)**—a 1-Nearest-Neighbor classifier based on Euclidean distance. We compared two feature extraction paradigms: (1) **Reactivity** (Δ = Stimulus − Baseline) and (2) **Raw State Concatenation** ($[\text{Stimulus}, \text{Baseline}]$). To capture the most distinct geometric dimensions, we selected the Top-20 features based on highest variance. **Results:** The analysis revealed a fundamental geometric distinction. The **Concatenation** method achieved a subject identification accuracy of **64.5%** (Chance: 4.3%), with a Separation Ratio of **1.32**, indicating distinct subject clusters. In contrast, the **Reactivity** method reached only **31.6%** accuracy with a Separation Ratio of **1.03** (weak clustering). Furthermore, "Video Identification" accuracy was negligible ($< 1\%$) for the Concatenation method, confirming that the signal variance is driven almost exclusively by the individual (Trait), not the stimulus (State). **Conclusion:** **Yes, individuals can be identified.** The EEG signal contains a persistent "Neural Fingerprint" encoded in its raw geometry. Standard reactivity methods inadvertently destroy this structure, whereas raw state concatenation preserves it.

1 Introduction

A central debate in affective computing concerns the separability of "State" (Emotion) and "Trait" (Identity). If EEG signals are dominated by emotional responses, then identifying a subject across different emotional stimuli should be difficult due to high intra-subject variance. Conversely, if the signal is dominated by stable individual traits, identification should be robust.

This study investigates the research question: **"Can we identify individuals from their EEG responses during emotional stimulation?"**

To answer this, we move beyond "black box" non-linear classifiers and instead analyze the **geometry** of the feature space. We employ a **Distance Discrimination Method (DDM)** to test the hypothesis that trials from the same subject naturally cluster together in Euclidean space, regardless of the emotional video being watched.

2 Methods

2.1 Data Processing

We analyzed the DREAMER dataset (23 subjects, 18 videos, $N = 414$ trials). Raw EEG was filtered (1–50 Hz) and features were extracted using two competing pipelines:

1. **Reactivity (Δ):** Calculated as $X_{stimulus} - X_{baseline}$. This method assumes the baseline is noise to be subtracted.

2. **Concatenation:** Calculated as vector $[X_{baseline}, X_{stimulus}]$. This method preserves the absolute power values.

2.2 Feature Selection

To ensure the analysis focused on the most information-rich dimensions, we performed feature selection to retain only the **Top 20 Features** for each pipeline. **Selection Criteria:** Features were selected based on **Highest Variance** across the dataset. This approach prioritizes dimensions that spread the data points out the most, maximizing the potential for geometric separation (clustering) in Euclidean space.

2.3 Distance Discrimination Method (DDM)

The DDM is a geometric classifier using a **Leave-One-Out 1-Nearest-Neighbor (1-NN)** paradigm. For every trial T_i, we calculate its Euclidean distance to all other 413 trials and identify the nearest neighbor T_{nn}.

- **Subject Accuracy:** Success if $Label(T_{nn}) == Label(T_i)$ (Same Subject).
- **Video Accuracy:** Success if $Video(T_{nn}) == Video(T_i)$ (Same Video).
- **Separation Ratio:** $\frac{\text{Avg Inter-Class Distance}}{\text{Avg Intra-Class Distance}}$. A ratio > 1.0 indicates clustering.

3 Results

3.1 Identification Performance

The geometric analysis provides a definitive answer to the research question. As shown in Table 1, the choice of feature extraction fundamentally alters the signal geometry.

Table 1: DDM Classification Performance (Top 20 Features by Variance)

Method	Subject Acc.	Separation Ratio	Video Acc.	Chance Level
Reactivity (Δ)	31.64%	1.03	3.86%	4.3% / 5.6%
Concatenation	**64.49%**	**1.32**	**0.72%**	4.3% / 5.6%

The **Concatenation** approach achieves **64.5%** accuracy using simple Euclidean distances, exceeding the chance level by a factor of 15. The Separation Ratio of **1.32** confirms that subjects form distinct, compact clusters. In contrast, Reactivity features form a diffuse cloud (Ratio ≈ 1.0) with only 31.6% accuracy.

3.2 The Source of the Match

To confirm that the identification is driven by **Identity** and not confounded by the **Stimulus**, we analyzed the specific relationship between each trial and its nearest neighbor.

Table 2: Nearest Neighbor Breakdown (Concatenation Method)

Neighbor Relationship	Count (N=414)	Interpretation
Same Subject, Diff. Video	**267 (64.5%)**	**Robust Identity**
Different Subject, Same Video	3 (0.7%)	No Video Signal
Different Subject, Diff. Video	144 (34.8%)	Noise / Error

Table 2 illustrates the "Video Blindness" of the EEG geometry. The nearest neighbor to a trial is almost never a different subject watching the same video (0.7%, which is far below chance level). Instead, it is the same subject watching a completely different video (64.5%). This proves that the "Neural Fingerprint" persists across varying emotional states.

4 Discussion

4.1 Geometric Structure of Identity

The DDM results demonstrate that the EEG feature space is naturally structured by identity. The Concatenation method's success (64.5%) using simple Euclidean distance proves that the **raw geometry** of the signal is distinct for each individual. No complex learning is required to separate subjects; they exist as isolated clusters in the high-variance feature space.

4.2 The Failure of Reactivity

The Reactivity method showed a Separation Ratio of **1.03**, barely distinguishable from random noise (1.0). By subtracting the baseline, we effectively collapsed the distinct clusters of each subject into a single distribution centered at zero. This confirms that the biometric information is encoded in the *absolute magnitude* of the power spectrum (trait offset), not in the relative change (state modulation).

5 Conclusion

This study conclusively answers the research question: **Yes, we can identify individuals from emotional EEG.** The raw EEG signal contains a robust geometric structure unique to each person. Using a simple distance-based method, we achieved **64.5% identification accuracy** (vs 4% chance), while the emotional stimulus itself was statistically invisible. We conclude that "who you are" (Trait) is a far stronger determinant of your brainwaves than "what you feel" (State).