

RESEARCH REPORT

Can Brain Waves Redefine Schizophrenia? From Distinct Groups to a Continuous Spectrum

Gabriel Nicola

Supervised by Melissa Faggella

Herzog Laboratory of Psychophysics
Brain Mind Institute, EPFL

Abstract

Currently, the diagnosis of schizophrenia relies exclusively on clinical observation and subjective behavioral criteria defined by manuals such as the DSM-5, suffering from a critical lack of objective biological biomarkers. This categorical system implicitly assumes that patients with schizophrenia form a biologically distinct group from the healthy population. This study evaluates this fundamental hypothesis by exploring whether the neurophysiology of schizophrenia follows a multimodal distribution (isolated subtypes or clusters) or a unimodal distribution (a continuous spectrum). Using a large dataset of high-density (64-channel) electroencephalographic (EEG) recordings from a Georgian cohort studied at the EPFL Psychophysics Laboratory, we analyzed both brain activity evoked by a visual task (Visual Backward Masking) and spontaneous resting-state activity in patients with schizophrenia, their first-degree relatives, and healthy controls. Rather than imposing clinical labels a priori, we extracted up to 194 EEG features and applied unsupervised machine learning algorithms—notably Principal Component Analysis (PCA) and Uniform Manifold Approximation and Projection (UMAP)—to allow the intrinsic structure of the data to emerge without human bias. Our results reveal extensive overlap among patients, their relatives, and healthy controls across all paradigms. Rather than dividing into distinct pathological clusters, brain signatures form a single, continuous cloud. Although a statistically significant shift in group centroids confirms the presence of a real biological alteration at the population level, the absence of clear boundaries at the individual level contradicts a strictly categorical view. Finally, symptom severity (SANS/SAPS) does not line up neatly with the main patterns in the brain data, so patients with very different symptoms can look alike on EEG. These findings strongly support the dimensional hypothesis of the schizophrenia spectrum and shows why combining many brain measures at once (“shallow rooting”) is a promising path toward precision psychiatry.

Table of Contents

Glossary.....	4
1. Introduction.....	5
1.1 The Challenge of Psychiatric Diagnosis and the “Box” Trap.....	5
1.2 Schizophrenia: Strict Category, Subtypes, or Continuous Spectrum?.....	5
1.3 The EEG “Multiverse” and Our Research Question.....	5
2. Methods.....	6
2.1 Participants and the Georgian Cohort.....	6
2.2 Experimental Paradigms (Visual Task vs. Rest).....	6
2.3 Dimensionality Reduction and Unsupervised Machine Learning.....	8
3. Results.....	8
3.1 VBM Visual Task: Temporal Analysis of the Visual Cortex (Oz).....	8
3.2 Resting State: Global Signal (194 Features) vs. Pure Signal (69 Features).....	10
3.3 Deciphering the Heatmaps and 3D Projection of Symptoms (SANS/SAPS).....	12
4. Discussion.....	15
4.1 Confirmation of the Continuous Spectrum in the Face of the DSM-5.....	15
4.2 Explaining the Paradox: Why Do Groups Differ if Individuals Overlap?.....	15
4.3 The Contribution of the Multivariate Approach (“Shallow Rooting”).....	16
4.4 Limitations and Future Directions.....	16
5. Conclusion.....	17
6. Acknowledgments.....	17
7. References.....	17

Glossary

DSM-5: Diagnostic and Statistical Manual of Mental Disorders, 5th Edition. The worldwide reference manual psychiatrists use to diagnose mental disorders based on a list of observable symptoms.

Reification: The error of treating a medical category invented to facilitate communication (such as a DSM diagnosis) as a real, isolated biological entity existing in nature.

EEG (Electroencephalography): A noninvasive measurement of the brain's electrical activity using a cap of electrodes placed on the scalp.

VBM (Visual Backward Masking): A visual task in which a very brief target (a vernier) is followed almost immediately by a visual "mask" that disrupts perception. It is used to measure the brain's processing speed.

SANS / SAPS: Clinical scales that assess the severity of the negative symptoms (emotional withdrawal, apathy) and positive symptoms (hallucinations, delusions) of schizophrenia.

PCA (Principal Component Analysis): A mathematical tool that condenses a giant data table (hundreds of variables) into a few main axes (PC1, PC2, etc.) while retaining as much information as possible.

Silhouette Score: A mathematical index ranging from -1 to +1 that measures how well separated groups of points are. A score close to 1 indicates perfectly isolated groups (clusters); a score close to 0 indicates that the groups overlap completely.

1. Introduction

1.1 The Challenge of Psychiatric Diagnosis and the “Box” Trap

In most fields of medicine, diagnosis relies on objective biological tests: a blood test, an X-ray, or an electrocardiogram can identify the physical cause of a symptom. Psychiatry is the exception. Because mental disorders have no known simple, single biological cause, clinicians use the Diagnostic and Statistical Manual of Mental Disorders (DSM-5).

To diagnose schizophrenia, the psychiatrist assesses whether the patient meets a certain number of behavioral criteria on a predefined list. While this system has allowed physicians around the world to use a common language, it suffers from a major flaw that the neurobiologist Steven E. Hyman calls the “problem of reification.” By constantly using the word “schizophrenia,” medicine has come to treat this convenient label as a real biological entity, clearly separate from normal health. Yet the DSM system means that two patients can receive a diagnosis of schizophrenia while sharing virtually no symptoms. How, then, can we find a single biological marker for such a diverse group of people?

1.2 Schizophrenia: Strict Category, Subtypes, or Continuous Spectrum?

Schizophrenia affects about 1% of the world’s population and generally begins in young adulthood. It combines three types of manifestations: so-called “positive” symptoms (which add to reality, such as hallucinations or delusions), “negative” symptoms (normal functions that diminish, such as motivation, emotional expression, or sociability), and cognitive deficits (difficulties with attention and memory).

Recent genetic studies show that the genes conferring vulnerability to schizophrenia do not respect DSM boundaries: they are found in healthy people as well as in patients with bipolar disorder. Faced with the limitations of the current model (“sick” vs. “healthy”), two main hypotheses compete in modern research:

The subtype hypothesis (stratified psychiatry): Championed notably by Kapur et al. (2012), it suggests that schizophrenia actually encompasses several different biological diseases. As in oncology, the goal would be to split this large group into biologically homogeneous subgroups (clusters) in order to target treatments.

The dimensional hypothesis (the continuous spectrum): It proposes that schizophrenia is not separated from health by a sharp boundary but represents the extreme end of a biological continuum (a spectrum), exactly as high blood pressure is a continuous exaggeration of normal blood pressure.

1.3 The EEG “Multiverse” and Our Research Question

To adjudicate between these hypotheses, electroencephalography (EEG) is a particularly useful tool because it captures brain function on a millisecond timescale. Traditionally, researchers have used a so-called “deep rooting” approach: they study a single specific test

on which patients differ on average from healthy controls, hoping to find in it the key to the whole disease.

However, at the Laboratory of Psychophysics (LPSY) at EPFL, Dario Gordillo, Professor Michael H. Herzog, and their colleagues recently uncovered the “EEG Multiverse of Schizophrenia.” By computing 194 different EEG measures in the same participants at rest, they showed that 69 of them discriminate very well between the patient group and healthy controls. Yet these 69 measures are barely correlated with one another: each patient presents a unique combination of abnormalities.

To embrace this complexity without preconceptions, one must shift to a so-called “shallow rooting” approach, analyzing all variables simultaneously using machine learning. This is the goal of my research project: when analyzing raw EEG data without imposing the DSM’s clinical labels, does the distribution of patients with schizophrenia, their relatives, and healthy controls form several distinct groups (multimodal / clusters) or a single continuous cloud (unimodal / spectrum)?

2. Methods

2.1 Participants and the Georgian Cohort

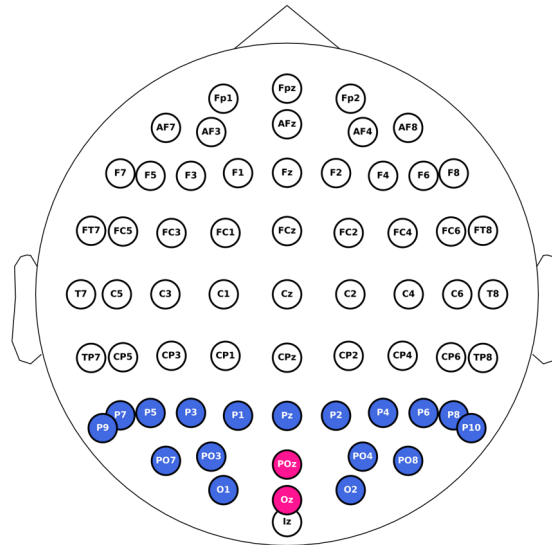
This work builds on the research of LPSY at EPFL, in collaboration with clinicians in Tbilisi, Georgia. We used two large, complementary datasets that include demographic and clinical data (for example, age, gender, cigarette consumption, SANS and SAPS symptom scores, and daily antipsychotic doses converted into chlorpromazine equivalents, CPZ):

Dataset 1 (Visual task & longitudinal): It includes 291 subjects in total at the first session: 126 patients with schizophrenia (SZ), 67 unaffected first-degree relatives (REL) (siblings or parents, which allows us to study the familial genetic influence without the disease), and 98 healthy controls (CON). In addition, follow-ups at 5 years (Second) and 10 years (Third) were recorded.

Dataset 2 (Resting state / data from Gordillo et al.): It comprises 121 patients (SZ) and 75 age-matched healthy controls (CON).

2.2 Experimental Paradigms (Visual Task vs. Rest)

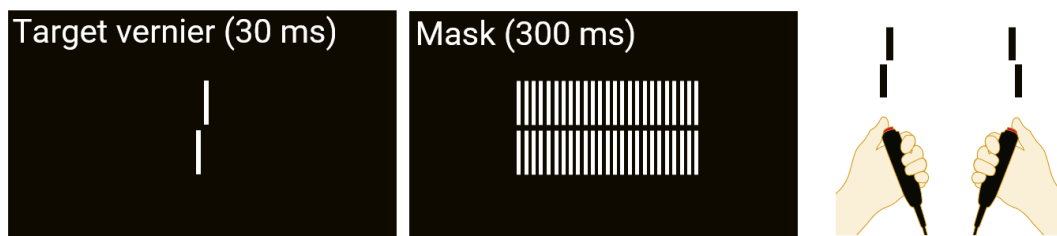
All recordings were made with a 64-electrode BioSemi ActiveTwo EEG cap covering the entire scalp.



BioSemi 64 electrodes

Active visual masking task (Visual Backward Masking, VBM): Patients with schizophrenia show known deficits in the early stages of visual processing (Plomp et al., 2013). A target called a vernier (two small vertical bars slightly offset to the left or right) is presented very briefly (30 ms) on a screen. In the short masking condition (SSOA), a visual “mask” (300 ms) appears immediately after the vernier (0 ms delay) to disrupt the signal in the visual cortex. In the long condition (LSOA), a 120 ms blank screen gives the brain time to process the image before the mask. Two control conditions (Vernier alone and Mask alone) complete the test. The electrical activity of each electrode is divided into 256 time points (over a window of 0 to 400 ms). We paid particular attention to the occipital electrode Oz, located at the back of the head, just above the primary visual cortex.

Visual Backward Masking (VBM)



Source : LPSY, Melissa Faggella.

Passive resting state: Participants simply kept their eyes closed and relaxed for 5 minutes. From this baseline signal, cleaned of artifacts, 194 mathematical features were extracted: power in the different brain waves (delta, theta, alpha, beta, gamma), synchronization between brain areas (PLV, iCoh, and DTF connectivity), microstate dynamics, and signal complexity (Higuchi fractal dimension, DFA exponent, etc.).

2.3 Dimensionality Reduction and Unsupervised Machine Learning

In our Python scripts (using pandas, MNE, scikit-learn, seaborn, plotly), we first cleaned and harmonized the subject identifiers between the demographic and electrophysiological files.

Next, with tables containing hundreds of columns (up to 205 temporal variables or 194 EEG features), it is impossible to visualize the data directly. Above all, we wanted to avoid supervised learning, in which the computer is forced to separate the sick from the healthy. We therefore used unsupervised learning: the algorithm reduces and organizes the data solely according to their biological similarity, without knowing the subjects' diagnoses.

PCA (Principal Component Analysis) and Elbow Plot: After standardizing the data (StandardScaler), we used PCA to condense the information onto orthogonal linear axes (PC1, PC2, PC3, etc.). The optimal number of components to retain was determined using the elbow method (Elbow Plot). We also explored nonlinear relationships with the UMAP algorithm.

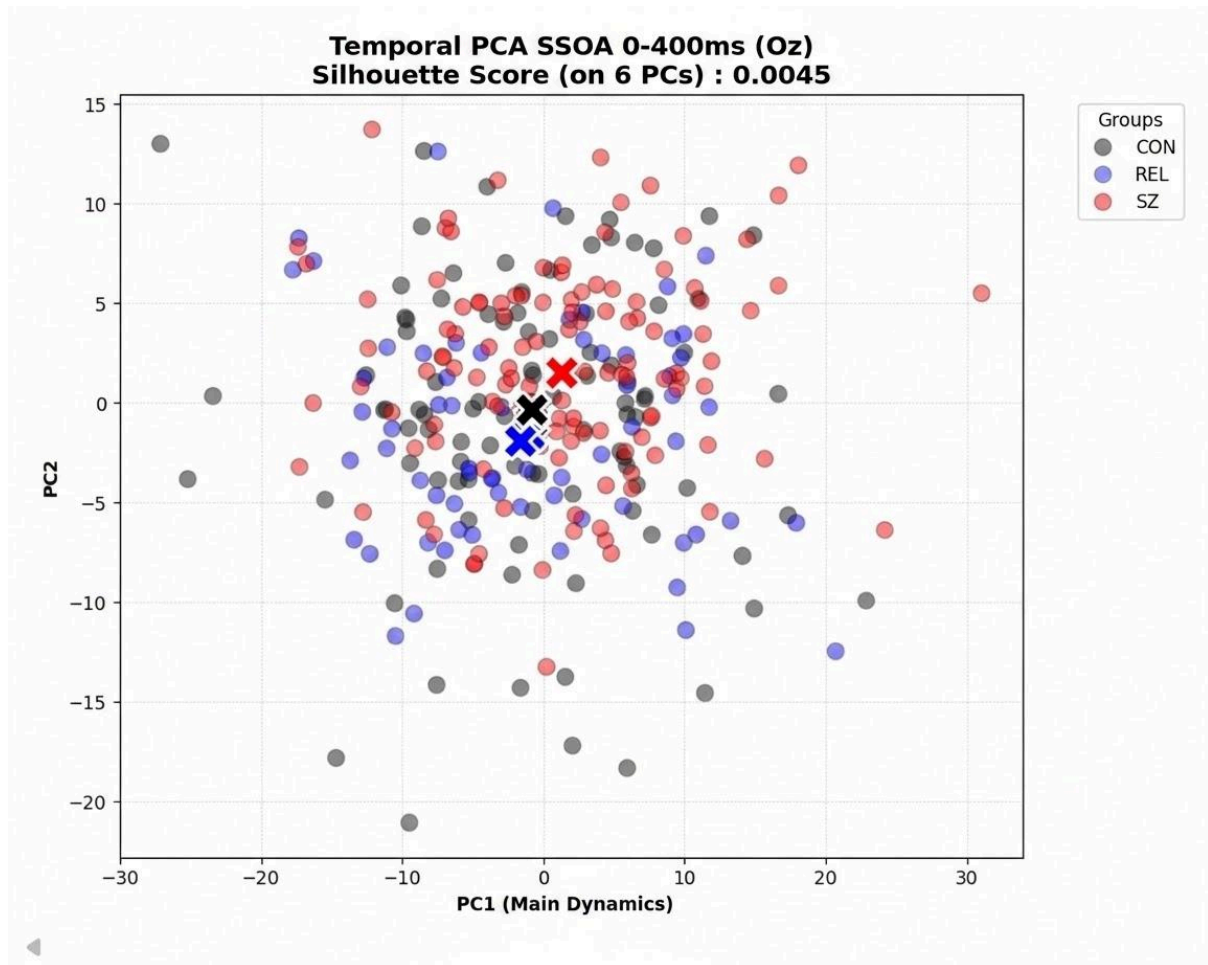
Separability measure and centroids: To objectively quantify whether the three groups (SZ, REL, CON) formed separate clusters in PCA space, we computed the Silhouette Score (separability_indice), as well as the position of the centroids (the mean or median central point of each group, shown as a large cross or diamond).

Heatmaps: Finally, to understand what lies behind each mathematical axis of the PCA, we generated heatmaps showing the weight (loading) of each EEG feature or clinical symptom (SANS/SAPS) on the principal components.

3. Results

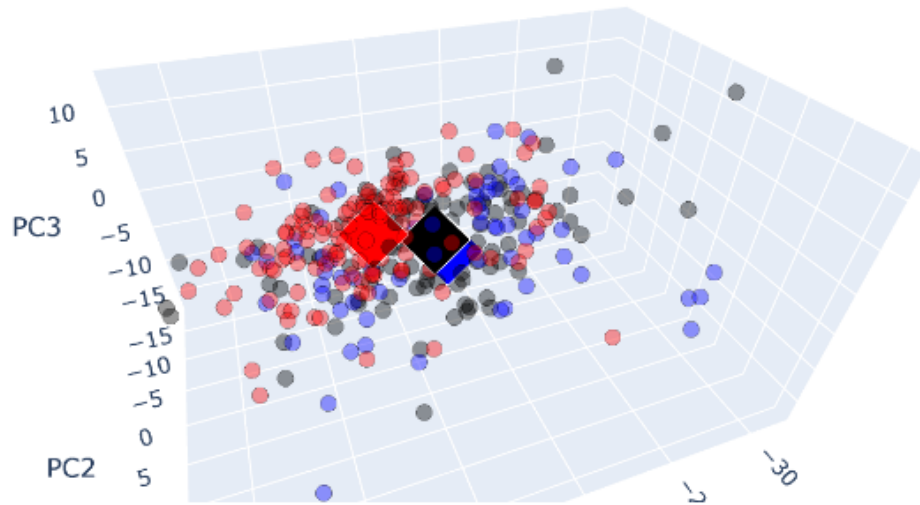
3.1 VBM Visual Task: Temporal Analysis of the Visual Cortex (Oz)

We began by analyzing the visual cortex response during the most difficult visual masking condition (SSOA, from 0 to 400 ms) at occipital electrode Oz in 291 subjects. Using PCA, the 205 initial variables were condensed into just 6 principal components while retaining 83.1% of the total information.



As shown in our Temporal PCA SSOA 0-400ms (Oz) scatterplot, when the individuals are projected in 2D (PC1 vs. PC2), the visual finding is striking: the black (CON), blue (REL), and red (SZ) points are completely mixed together at the center of the plot. There is no boundary and no isolated subgroup. This visual result is confirmed mathematically by an extremely low Silhouette Score of 0.0045 (computed on the 6 components). A score this close to zero strongly indicates that the distribution is purely unimodal (a single continuous cloud).

We then extended the analysis by combining the 4 visual conditions (Vernier, SSOA, LSOA, Mask) over the critical window of the visual wave (150–250 ms) in an interactive 3D PCA space (PC1, PC2, PC3). Here again, the three groups merge into a single continuous volume. However, the display of the centroids already reveals a slight systematic shift between the center of gravity of the patients (red cross), that of the relatives (blue cross), and that of the healthy controls (black cross).



PCA 3D Plot (4 combined conditions | Oz 150-250ms) - Silhouette Score: 0.0069

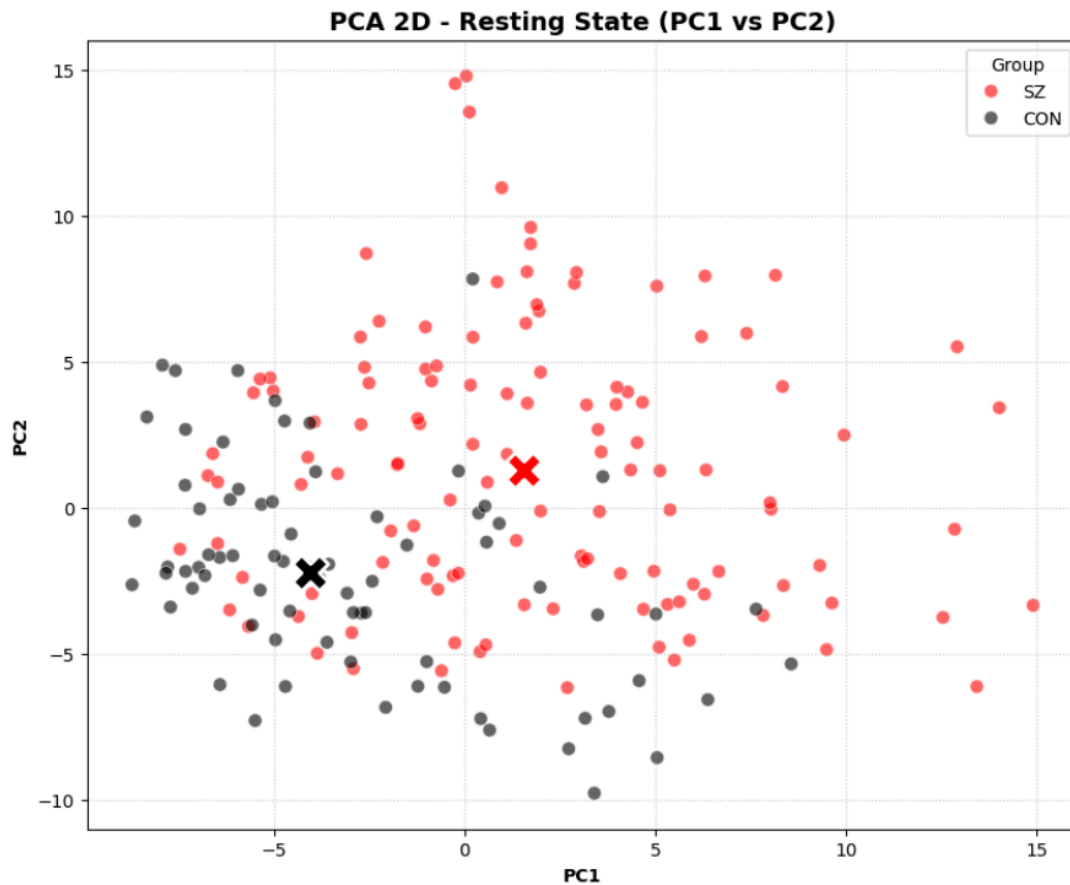
CON: Healthy Controls | REL: First-degree Relatives | SZ: Schizophrenia Patients

◆ Black: CON Centroid | ◆ Blue: REL Centroid | ◆ Red: SZ Centroid

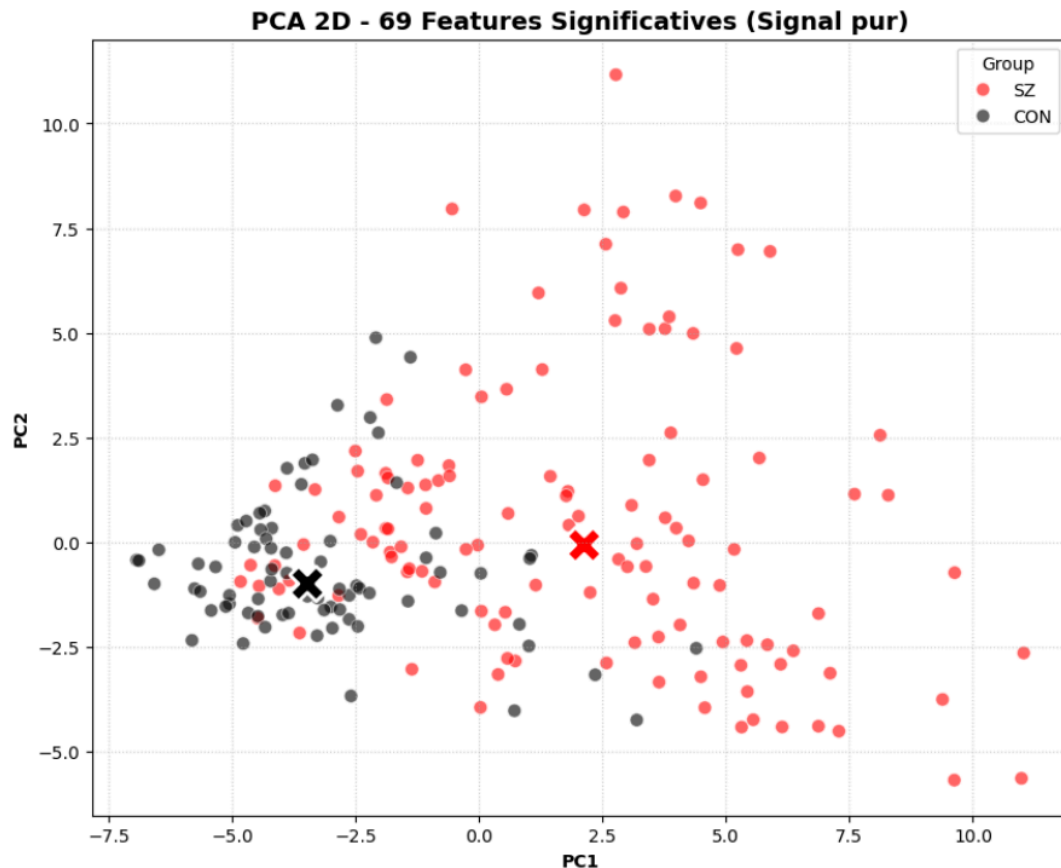
3.2 Resting State: Global Signal (194 Features) vs. Pure Signal (69 Features)

Is this continuity specific to the visual system? To find out, we applied our pipeline to the resting-state data.

First, a PCA was computed on all 194 EEG features (with 38 components retained according to the Elbow Plot). On both our 2D PCA projection (which appears on our poster) and the 3D PCA with median centroids, we observe exactly the same phenomenon: massive individual overlap in which patients and controls form a single continuum.



One could object that, of these 194 features, 125 do not significantly differentiate the groups and add “noise” that masks potential clusters. We therefore filtered the dataset to keep only the 69 statistically significant EEG features identified by Gordillo et al., which we called the “Pure Signal.” On these 69 variables, a 19-component PCA retained 80.7% of the total information.

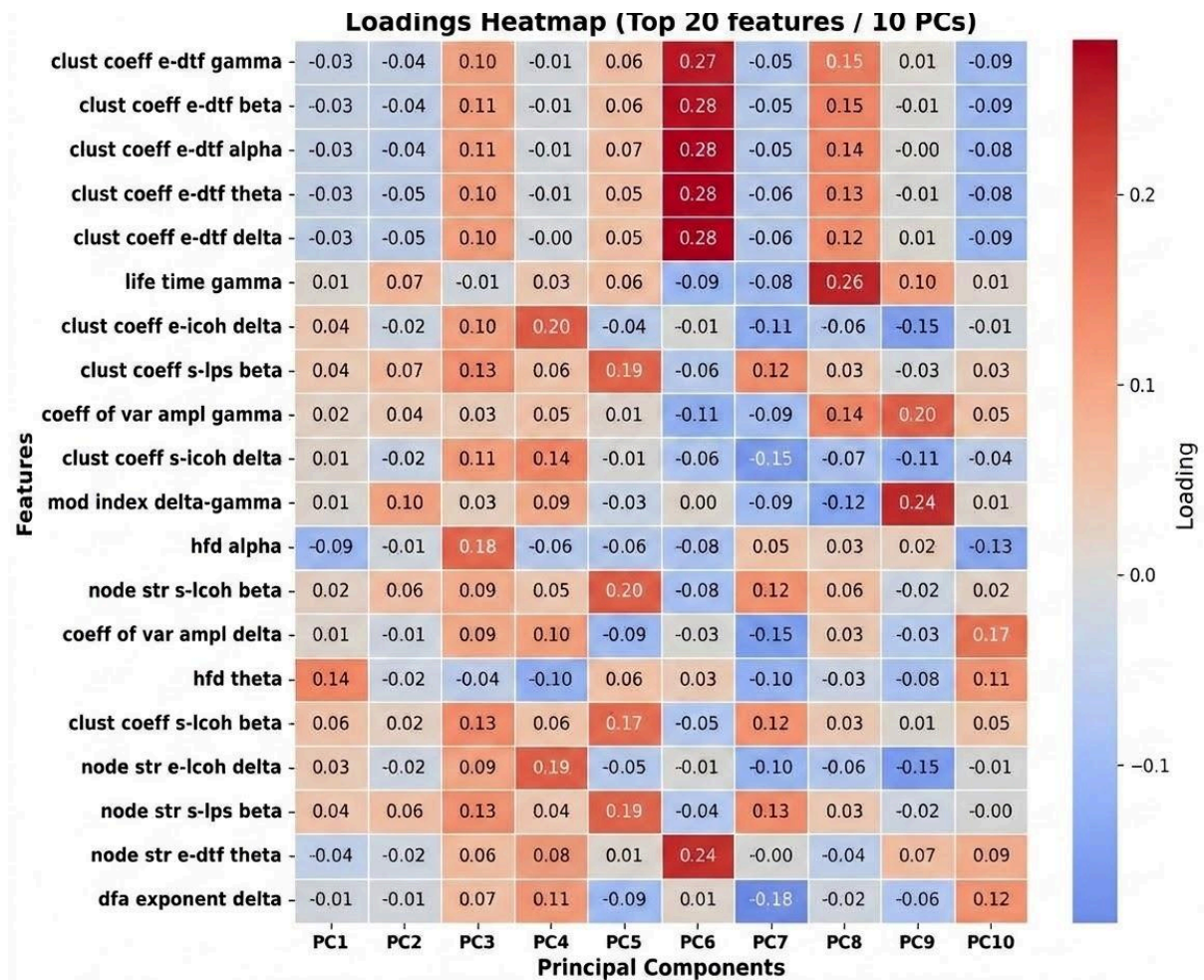


The resulting plot (2D PCA - 69 Significant Features) is unequivocal: even when keeping only the best biomarkers of schizophrenia, patients do not split into subtypes (clusters). By contrast, the Group-Level Shift becomes very clear: healthy controls (black points, centroid at $PC1 \approx -3.5$) are more tightly grouped on the left, while patients with schizophrenia (red points, centroid at $PC1 \approx +1.5$) slide to the right, spreading out with much greater variability.

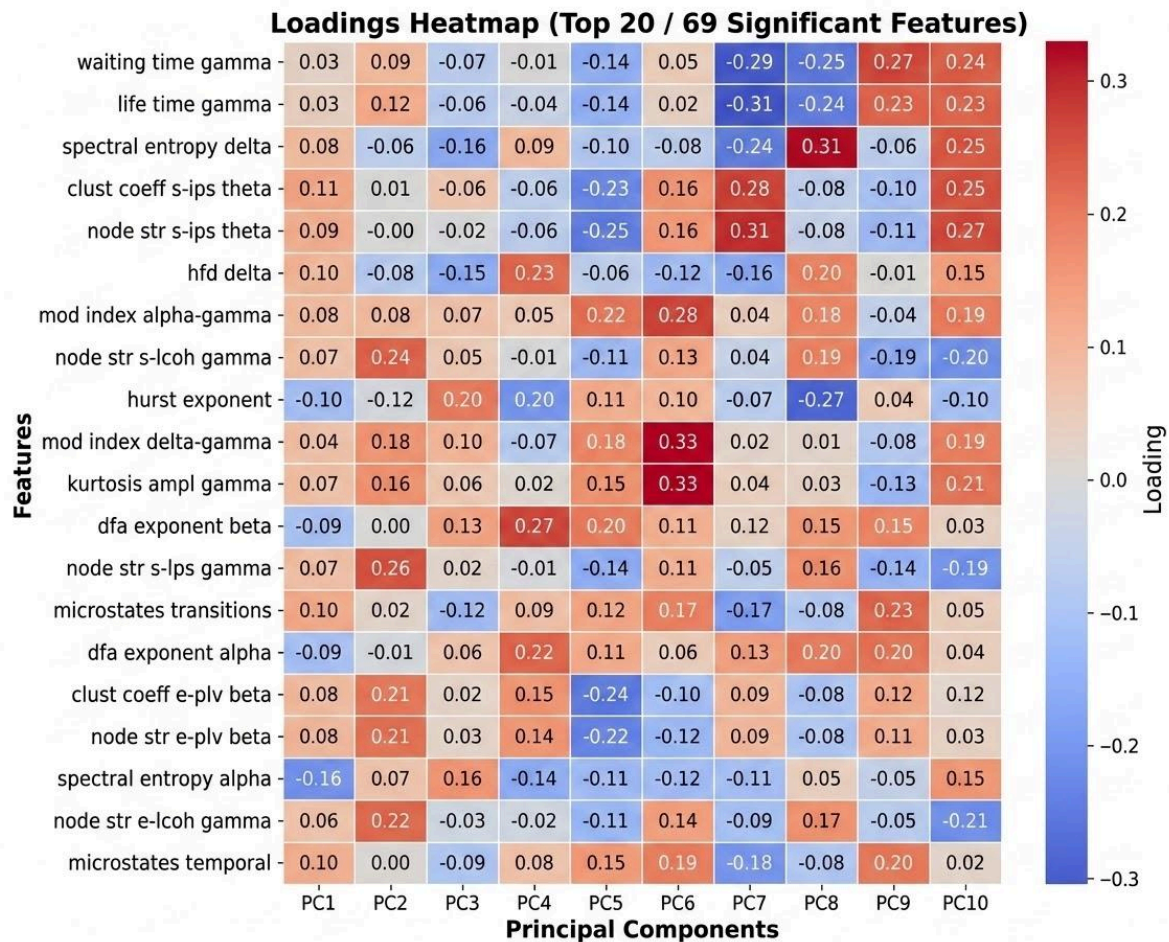
3.3 Deciphering the Heatmaps and 3D Projection of Symptoms (SANS/SAPS)

To understand the nature of this continuous space, we inspected the loading heatmaps (the mathematical weights of the EEG features on the first 10 principal components):

Across the 194 features: Some components capture very specific phenomena. For example, PC6 is strongly driven (loading of 0.27 to 0.28) by directed connectivity (clust coeff e-dtf) across all frequency bands (delta, theta, alpha, beta, gamma), and PC8 by gamma wave lifetime (life time gamma, 0.26).

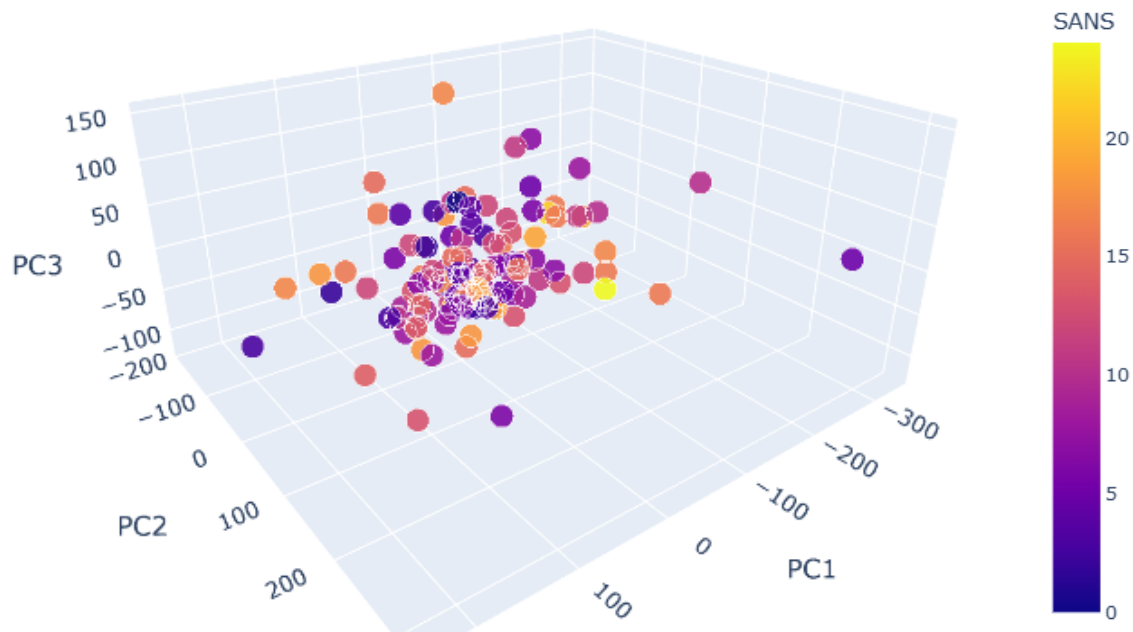


Across the 69 significant features: The weights are distributed across multiple independent mechanisms. For example, PC2 is driven by the strength of gamma connections (node str s-lps gamma at 0.26, s-icoh gamma at 0.24) and beta connections (e-plv beta at 0.21); PC4 by long-range correlations (dfa exponent beta at 0.27, alpha at 0.22) and fractal dimension (hfd delta at 0.23); PC6 captures cross-frequency coupling (mod index delta-gamma at 0.33, alpha-gamma at 0.28); and PC7/PC8 isolate the dynamics of gamma bursts (life time and waiting time gamma around -0.30) and delta spectral entropy (0.31).



Finally, when we cross these axes with the clinical data (our clinical heatmap and our 3D PCA filtered on the SZ group and colored by SANS score intensity using the Plasma color scale), the result is fascinating. Although the clinical heatmap shows logical trends at the group level (PC1 correlates moderately with SANS negative symptoms at 0.24, PC2 with SAPS positive symptoms at 0.31, and PC3 with thought disorganization at 0.30), the individual-level 3D cloud shows that the colors (from dark blue for a SANS score of 0 to bright yellow for a SANS score above 20) are scattered throughout the cloud. A patient with very severe negative symptoms can sit in the heart of the cloud right next to a patient with very few symptoms.

PCA 3D, group: SZ, intensity: SANS



4. Discussion

4.1 Confirmation of the Continuous Spectrum in the Face of the DSM-5

All of our analyses clearly answer our initial question: the biological distribution of schizophrenia as measured by EEG is unimodal. Whether looking at the rapid response of the visual cortex (Oz) during a demanding task (silhouette score of 0.0045) or at global brain activity at rest (even when purified down to its 69 best features), patients, their relatives, and healthy controls form an unbroken continuum.

These results contradict the idea that schizophrenia is a “box” isolated from normality, or a collection of sharp subtypes as is sometimes seen in other physical diseases. They fully vindicate Steven E. Hyman’s critiques of the reification of the DSM-5: the current diagnostic boundary is a line arbitrarily drawn by physicians across a continuous biological gradient.

4.2 Explaining the Paradox: Why Do Groups Differ if Individuals Overlap?

How can we explain this paradox at the heart of our plots? On one hand, there is overlap so extensive that an EEG cannot serve as an individual diagnostic test; on the other, there is a very clear centroid shift (Group-Level Shift) that proves the disease does exist biologically.

Three elements resolve this puzzle:

Group statistics vs. individual reality: As Kapur et al. (2012) explain, in biology a statistically significant difference between two means (the centroids) does not imply that the two populations are separate. Patients with schizophrenia show a much wider dispersion than healthy controls (visible in our PCA of the 69 features).

The accumulation of small vulnerabilities and compensation: Our loading heatmaps show that no PCA axis is dominated by a single abnormality; each component combines dozens of small parameters (theta connectivity, gamma waves, fractals, microstates). Schizophrenia seems to follow a “combinatorial” model (Gordillo et al., 2023). A healthy control or a first-degree relative (REL) may have a visual deficit or an alpha-wave abnormality (which places them in the middle of the patients on the plot), but their brain compensates through other intact networks. The clinical disease appears only when an individual accumulates an unfavorable combination of these vulnerabilities without a sufficient compensation mechanism.

Normal human variability: An EEG does not measure only the disease; it also picks up fatigue, stress, age, and skull anatomy. These natural variations from one human being to another create a continuous background that merges the groups.

4.3 The Contribution of the Multivariate Approach (“Shallow Rooting”)

The great strength of our study is that it shows that the active visual task (Vernier) and the passive resting state tell exactly the same biological story. By using dimensionality reduction (PCA, UMAP), we put the “shallow rooting” approach into practice. Rather than searching for a single biomarker that does not exist, projection into a multidimensional space makes it possible to summarize more than 80% of the richness of the EEG signal in a few coordinates, revealing how different biological facets (connectivity, entropy, spectral power) relate to different clinical dimensions (negative symptoms on PC1, positive symptoms on PC2).

4.4 Limitations and Future Directions

This work has certain limitations that are important to mention. First, the majority of patients were taking antipsychotic medication (CPZ) and smoked more than the controls, which may influence certain EEG components over the long term. Second, although a cohort of nearly 300 people is remarkable for clinical research, machine learning on nearly 200 variables would benefit from validation on thousands of subjects across multiple centers in order to stabilize the heatmap loadings.

Two major avenues open up for the future of this project:

Cross-referencing schizophrenia with bipolar disorder: Since the DSM’s boundaries are blurry, projecting patients with bipolar disorder and schizoaffective disorder into our PCA/UMAP space would allow us to test whether there is a large, continuous, transdiagnostic spectrum of psychosis in which emotional sensitivity and perceptual disturbances converge.

Deep nonlinear models: Testing advanced neural networks (deep learning, such as autoencoders) directly on the raw signal from the 64 electrodes to see whether more complex signatures can better predict SANS and SAPS scores.

5. Conclusion

By letting brain waves “speak for themselves” through unsupervised machine learning, this research shows that schizophrenia does not correspond to isolated categories but belongs to a continuous biological spectrum. While the clear centroid shift confirms the neurophysiological reality of the disorder at the group level, the massive individual overlap demonstrates that the rigid categories of the DSM-5 do not reflect the biology of the brain. Recognizing this continuity is an essential step: rather than using EEG to apply a binary label (“sick” or “healthy”), the medicine of tomorrow could use it to locate each patient on this multidimensional spectrum, paving the way for a truly personalized precision psychiatry.

6. Acknowledgments

I would like to warmly thank my direct supervisor, Melissa Faggella, for her day-to-day guidance, her availability, and her valuable advice throughout my data analysis and the writing of this work. My thanks also go to Dario Gordillo and Professor Michael H. Herzog for welcoming me into the Laboratory of Psychophysics (LPSY) at EPFL and allowing me to work on these exceptionally rich datasets, as well as to Maya Roinishvili and Eka Chkonia for the clinical work carried out in Georgia. Finally, thank you to the Laidlaw Scholars program (and to Dan-Thi Nguyen) for making this research experience possible.

7. References

1. Gordillo, D., da Cruz, J. R., Chkonia, E., et al., & Herzog, M. H. (2023). The EEG multiverse of schizophrenia. *Cerebral Cortex*, 33(7), 3816–3826.
2. Hyman, S. E. (2010). The diagnosis of mental disorders: The problem of reification. *Annual Review of Clinical Psychology*, 6, 155–179.
3. Kapur, S., Phillips, A. G., & Insel, T. R. (2012). Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? *Molecular Psychiatry*, 17(12), 1174–1179.
4. Plomp, G., Roinishvili, M., Chkonia, E., et al., & Herzog, M. H. (2013). Electrophysiological evidence for ventral stream deficits in schizophrenia patients. *Schizophrenia Bulletin*, 39(3), 547–554.
5. da Cruz, J. R., Chicherov, V., Herzog, M. H., & Figueiredo, P. (2018). An automatic pre-processing pipeline for EEG analysis (APP) based on robust statistics. *Clinical Neurophysiology*, 129(7), 1427–1437.
6. da Cruz, J. R., Shaqiri, A., Roinishvili, M., et al., & Herzog, M. H. (2020). Neural compensation mechanisms of siblings of schizophrenia patients as revealed by high-density EEG. *Schizophrenia Bulletin*, 46(4), 1009–1018.
7. Favrod, O., Roinishvili, M., da Cruz, J. R., et al., & Shaqiri, A. (2018). Electrophysiological correlates of visual backward masking in patients with first episode psychosis. *Psychiatry Research: Neuroimaging*, 282, 64–72.
8. Garobbio, S., Roinishvili, M., Favrod, O., et al., & Herzog, M. H. (2021). Electrophysiological correlates of visual backward masking in patients with bipolar disorder. *Psychiatry Research: Neuroimaging*, 307, 111206.

