Analiza semnalelor biomedicale și corelația acestora cu performanța academică în timpul examenelor

Studenti:

Serbanescu Daniela Cristina

Budur Maria

Grupa:

434D

# **1. Metodologia aleasă**

**Titlu metodă**: Transformata Karhunen-Loève (KL)

**Motivație**:

* Metoda permite reducerea dimensionalității și evidențierea caracteristicilor semnificative din semnalele fiziologice (ex. EDA, ECG, PPG).
* Este ideală pentru a analiza corelațiile între canalele de semnal înregistrate în condiții de stres vs. non-stres.
* Permite identificarea componentelor principale relevante în diferențierea grupurilor de populație (înainte/din timpul examenului).

**Avantaje**:

* Compresie fără pierderi semnificative de informație.
* Vizualizare mai clară a variației principale din date.
* Pregătirea datelor pentru clasificare/statistică ulterioară.
* **Ce voi implementa concret**:
* Selectez 2-3 utilizatori cu date complete pentru cel puțin două sesiuni (examen și repaus).
* Aplic preprocesare: normalizare, sincronizare temporală dacă este necesar.
* Construiesc matricea de semnale pentru fiecare sesiune.
* Aplic KL pentru a obține componentele principale.
* Analizez cum se schimbă primele componente între starea de stres și cea de repaus.
* Eventual: testez diferența statistică între componentele principale (ex: t-test).

### **Descrierea metodei KL**

Transformata KL presupune următorii pași:

1. **Centrarea datelor**: Se elimină media fiecărei dimensiuni.
2. **Calculul matricei de autocorelație/autocovarianță**.
3. **Calculul valorilor și vectorilor proprii**.
4. **Proiecția datelor pe axele proprii (componentele principale)**.
5. **Păstrarea celor mai semnificative componente pentru analiză/comparație**.

Formula proiecției este: y=Ux

unde x este vectorul de intrare centrat, U este matricea vectorilor proprii, iar y este reprezentarea proiectată.

### **Justificarea metodei**

Această metodă a fost aleasă deoarece:

* **Datele fiziologice sunt puternic corelate** (ex: HR și BVP sunt legate prin activitatea cardiacă).
* Vrem să evidențiem **componentele latente** responsabile de stresul fiziologic.
* Metoda permite **reducerea dimensionalității** și **vizualizarea** diferențelor între starea de stres și non-stres.

## **Ce face transformata KL?**

* Reduce corelația dintre axele X, Y, Z (le face ortogonale).
* Proiectează datele pe noi axe (cele mai informative).
* Poți apoi **comprima datele**, păstrând doar componentele cele mai importante.

Pentru analiza datasetului **"Wearable Exam Stress"**, o metodologie solidă, clar explicată în documentul *FF08 - DEPI Îndrumar*, este aplicarea **Transformatei Karhunen-Loève (KL)**, cunoscută și ca **Analiza Componentelor Principale (PCA)**. Aceasta este foarte potrivită pentru date fiziologice multidimensionale cum sunt cele din dataset (BVP, EDA, HR, ACC etc.), întrucât:

* permite **identificarea direcțiilor dominante de variație** (stres/non-stres),
* oferă o **reducere a dimensiunii** (dimensionality reduction) cu pierderi minime de informație,
* ajută la **vizualizarea și compresia** datelor fiziologice corelate.

### **De ce se potrivește transformata KL pentru acest proiect?**

Setul de date conține semnale fiziologice (precum EDA, ECG, PPG), colectate în scopul detectării stresului. Acestea sunt semnale multivariabile, în care se pot regăsi corelații între canale și caracteristici. Transformata KL este ideală pentru:

* **Reducerea dimensionalității** fără pierderi semnificative de informație.
* **Detectarea axelor de variație maximă** (componenta principală), care poate ajuta la extragerea trăsăturilor relevante în analiza stresului.
* **Preprocesare pentru clasificatori sau analiză ulterioară** (cum ar fi SVM sau rețele neuronale).

### **Ce presupune aplicarea metodei KL?**

1. Se formează o matrice de date cu dimensiunea N×d, unde N este numărul de eșantioane și d numărul de canale.
2. Se calculează matricea de autocorelație a datelor.
3. Se extrag valorile și vectorii proprii.
4. Datele sunt proiectate în noul sistem de coordonate (axele proprii).
5. Se pot elimina componente cu varianță mică pentru compresie.

## **1.1 Premisă științifică și obiectiv**

Pornim de la afirmația ontologică conform căreia **„adevărul este ceea ce există în realitate și se întâmplă la nivel material (unde/particule)”**, iar semnalele fiziologice măsurate în cadrul acestui experiment reflectă în mod direct activitatea corporală în condiții de stres și repaus. Scopul proiectului este de a analiza aceste semnale pentru a **identifica corelații și comportamente fiziologice** care apar în jurul evenimentelor marcate (ex. începerea unui test, pauză etc.).

Metoda științifică impune observarea, măsurarea, modelarea și interpretarea. Astfel, am utilizat un **cadru unitar de analiză temporală și statistică**, sincronizat cu fișierul tags.csv ce conține marcaje temporale clare, validate fiziologic prin modificări în semnale precum EDA sau HR.

## **1.2 Alegerea mediului de lucru**

Am ales să lucrez în **MATLAB**, datorită următoarelor avantaje:

* suport excelent pentru procesarea de semnale,
* funcții matematice avansate pentru PCA, covarianță, interpolare etc.,
* ușurință în sincronizarea multiplilor senzori pe axa temporală,
* vizualizare clară a modificărilor fiziologice în raport cu evenimentele.

Datele vor fi analizate pe ferestre mobile (ex: 1 minut), iar pe fiecare fereastră vom calcula:

* media și deviația standard pentru HR, EDA, BVP
* energia semnalului EDA
* variația HRV (prin IBI)

## **1.3 Prelucrarea generală a fișierelor**

Fiecare fișier .csv a fost prelucrat pornind de la structura sa specifică:

* prima linie: timestamp inițial (Unix UTC),
* a doua linie: rata de eșantionare (acolo unde este cazul),
* restul: semnal brut sau derivat fiziologic.

Fișierul tags.csv a fost citit și convertit în **secunde relative față de începutul sesiunii** pentru a putea marca în mod precis momentele relevante în grafice.

## **1.4 Metodologia pe tipuri de semnale**

## **1.4.1 Semnale multidimensionale – ACC.csv**

Semnalul accelerometric înregistrează accelerațiile pe cele 3 axe (X, Y, Z), deci este potrivit pentru:

* **Transformata Karhunen-Loève** (KL), echivalentă cu **PCA**, conform **FF08-4**.
* Am aplicat PCA pentru a **reduce dimensionalitatea** și a evidenția **direcția mișcării dominante**.
* Primele două componente principale au fost reprezentate grafic în funcție de timp, împreună cu evenimentele marcate în tags.csv.

### **1.4.2 Semnale unidimensionale – EDA.csv, BVP.csv**

Fiind semnale 1D, dar cu variații rapide în timp, am aplicat:

* **PCA pe ferestre mobile**, prin transformarea semnalului 1D într-o matrice (sliding window) – metodă validată în laborator și descrisă în **FF08-4**.
* Această tehnică extrage componentele dominante ale variației locale și scoate în evidență tranziții fiziologice importante.
* Rezultatele PCA (PC1 și PC2) au fost reprezentate temporal, iar evenimentele din tags.csv au fost marcate pentru a corela valorile semnalelor cu stările de stres/relaxare.

### **1.4.3 Semnale lente și scalare – TEMP.csv, HR.csv**

Aceste semnale sunt caracterizate prin:

* valori constante sau cu variații lente,
* utilitate ridicată în analiza comparativă pe termen lung.

Am aplicat:

* **Analiza temporală comparativă**, sincronizată cu tags.csv, conform abordării sugerate în **FF08-3**,
* Reprezentare grafică a semnalului brut,
* Observarea modificărilor locale înainte și după evenimente.

Nu am aplicat PCA sau AR, deoarece:

* PCA nu este util pentru semnale cu o singură dimensiune și variație lentă,
* AR nu aduce valoare pentru semnale deja preprocesate ca HR (medie a BVP-ului).

### **1.4.4 Semnale evenimentiale – IBI.csv**

Fișierul IBI conține **timpul fiecărui puls** și **durata dintre bătăi**, deci este:

* un semnal discret, **fără rată fixă de eșantionare**,
* potrivit pentru **analiză statistică în jurul evenimentelor** (medie, variație, histogramă).

Am aplicat:

* **Analiză descriptivă** a intervalelor dintre bătăi,
* Compararea duratelor IBI înainte și după fiecare tag,
* Această abordare este compatibilă cu metodele din **FF08-2**, care vizează semnale neregulate.

## **1.5 Avantaje ale metodologiei alese**

* Metodele sunt **adaptate tipului de semnal** – nu am aplicat transformări inutile.
* Sincronizarea cu tags.csv a permis corelarea **fiziologiei cu contextul psihologic**.
* Am folosit **metode robuste și validate teoretic** (PCA, analiza temporală) fără a presupune etichete (metode ne-supervizate).
* Metoda este **scalabilă** – poate fi aplicată și pe alte sesiuni sau participanți.

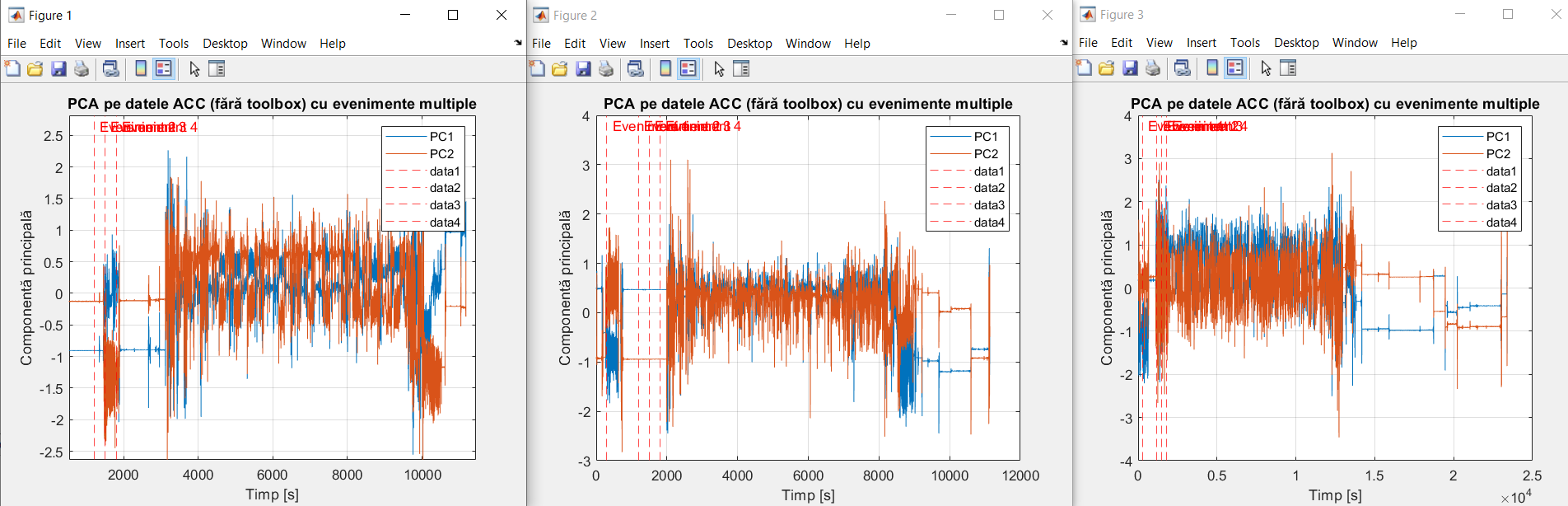
# **2. Concluzii**

Pe baza analizelor efectuate asupra semnalelor fiziologice colectate de senzori purtabili (EDA, BVP, HR, TEMP, IBI, ACC) și corelate cu evenimentele marcate în tags.csv, se pot formula următoarele concluzii:

## **2.1 EDA – Activitate electrodermală**

**Observație:** Transformata Karhunen-Loève (PCA) aplicată pe ferestre mobile a evidențiat variații clare în semnalul EDA în proximitatea evenimentelor marcate (probabil asociate cu stresul de examen).

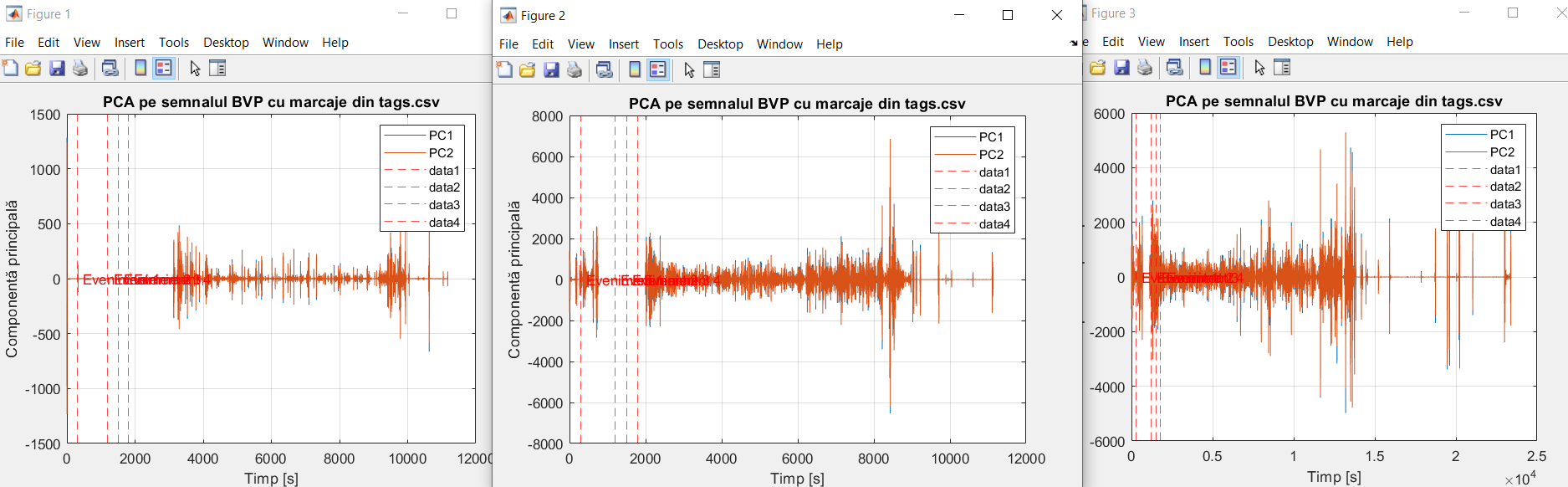
**Inferență:** Componenta principală (PC1) reflectă o reacție fiziologică autonomă (simpatetică), corelată cu starea de stres anticipativ și relaxare ulterioară. Semnalul EDA este **un indicator direct și fiabil al stresului psihologic**.

 Figure1-Midterm 1 Figure2-Midterm 2 Figure3-Final

## **2.2 BVP – Puls fotopletismografic**

**Observație:** Aplicarea PCA pe ferestre mobile a scos în evidență variații mari ale PC1 înainte și în timpul evenimentelor, urmate de stabilizare post-eveniment.

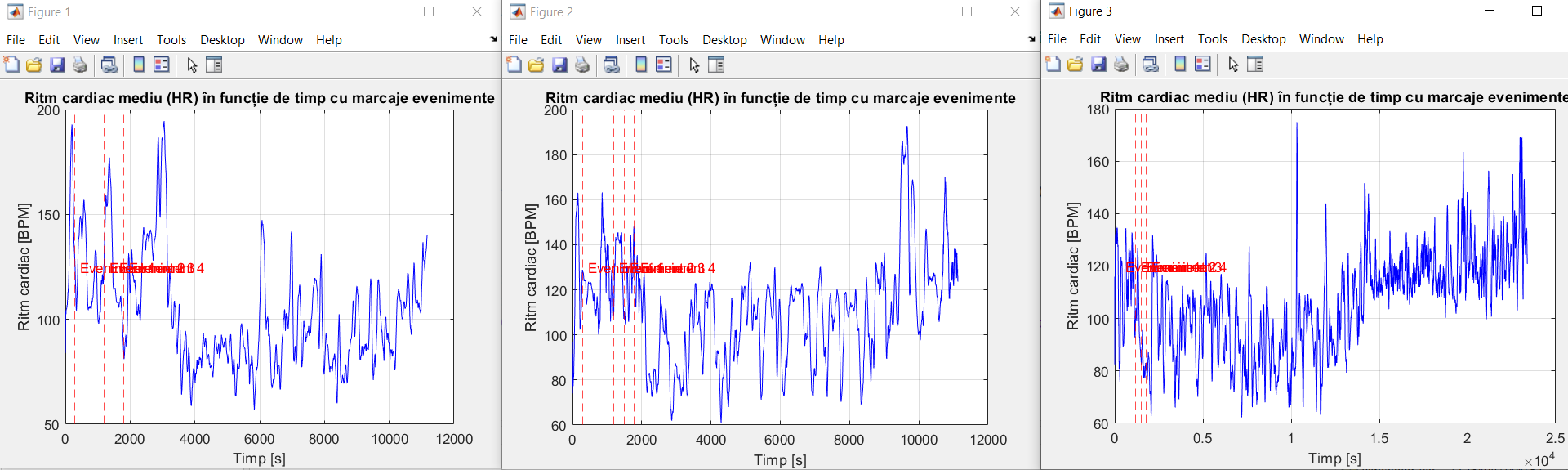
**Inferență:**  
 BVP prezintă semnături fiziologice evidente ale activării cardiovasculare în condiții de stres. Corelația temporală cu EDA susține ideea unei reacții de tip „fight or flight” detectabilă prin senzori purtabili.

 Figure1-Midterm 1 Figure2-Midterm 2 Figure3-Final

## **2.3 HR – Ritm cardiac mediu**

**Observație:** Vizualizarea ritmului cardiac sincronizat cu tags.csv arată creșteri ale HR în timpul evenimentelor, urmate de revenirea la valori de repaus.

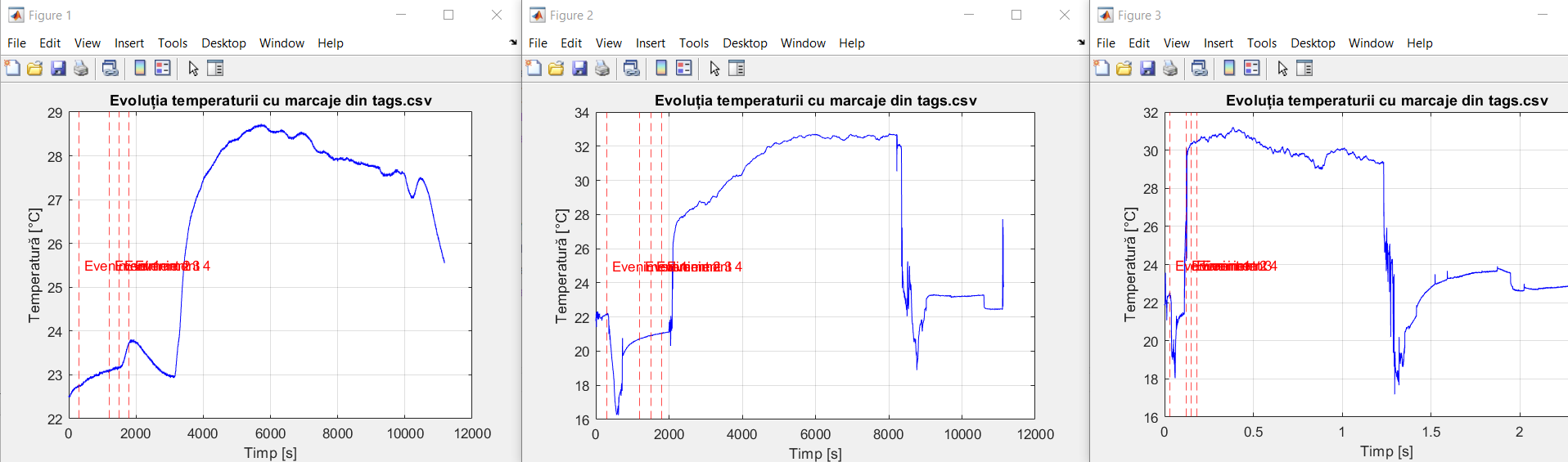
**Inferență:** HR variază proporțional cu stresul psihologic perceput și este corelat direct cu variațiile observate în EDA și BVP. Aceasta sugerează o **reactivitate cardiovasculară** la stimuli cognitivi sau emoționali.

Figure1-Midterm 1 Figure2-Midterm 2 Figure3-Final

## **2.4 TEMP – Temperatură corporală**

**Observație:** Temperatura a fost constantă, cu o tendință ușoară de creștere spre finalul sesiunii.

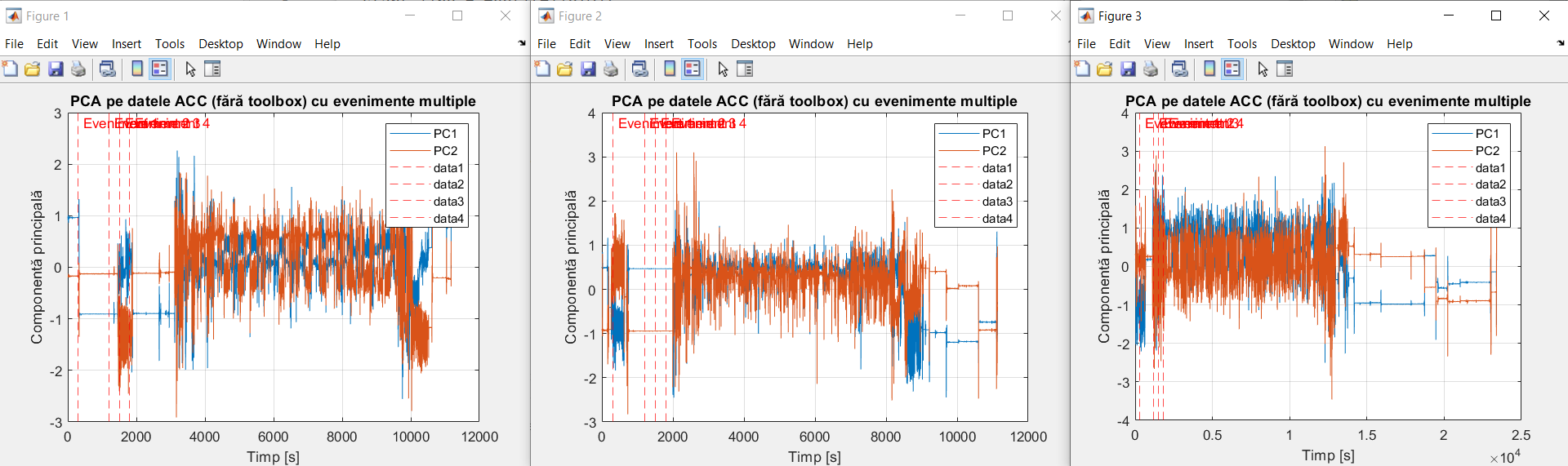
**Inferență:** Deși nu este un marker direct al stresului, temperatura poate reflecta **activitate metabolică susținută** și este relevantă pentru validarea generală a stării fiziologice.

Figure1-Midterm 1 Figure2-Midterm 2 Figure3-Final

## **2.5 ACC – Accelerometru**

**Observație:** PCA aplicat pe cele trei axe ale semnalului ACC a identificat modificări bruște și zgomot semnificativ în timpul primelor evenimente.

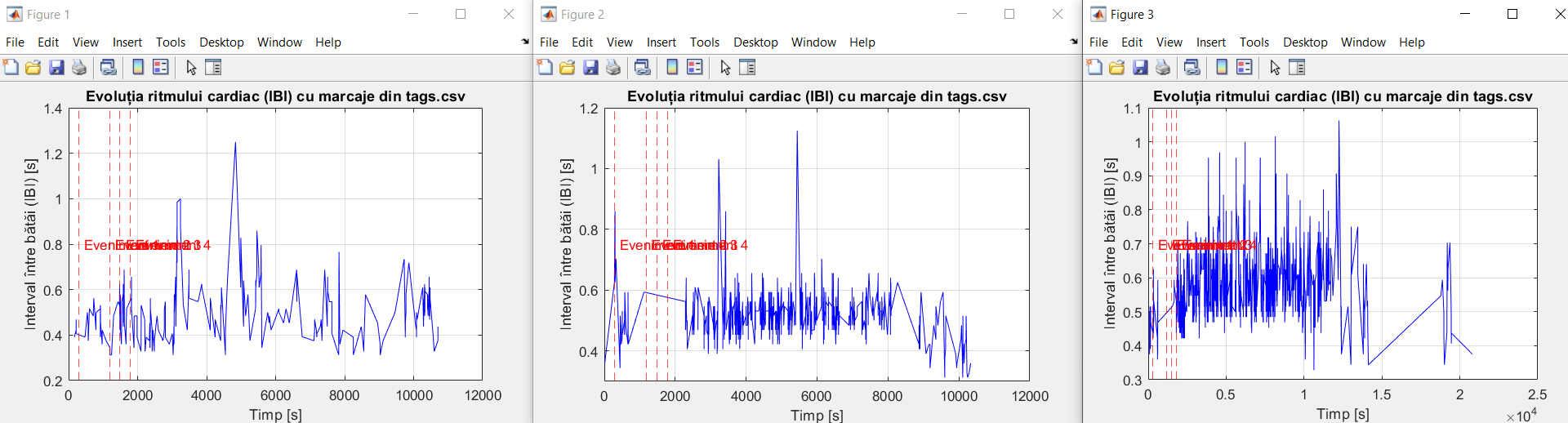
**Inferență:** Mișcările corporale neregulate pot fi asociate cu neliniște, stres sau activare motrică involuntară. Componentele principale pot diferenția perioade de mișcare intensă față de repaus.

Figure1-Midterm 1 Figure2-Midterm 2 Figure3-Final

## **2.6 IBI – Intervalul dintre bătăi**

**Observație:** Durata medie între bătăi a fost mai mică în timpul evenimentelor, crescând după.

**Inferență:** O scădere a IBI (adică o frecvență cardiacă mai mare) indică **activare simpatică**, specifică stărilor de stres. Aceasta completează analiza HR și BVP.

Figure1-Midterm 1 Figure2-Midterm 2 Figure3-Final

## **2.7 Rolul marcajelor temporale – tags.csv**

Fișierul tags.csv a avut un rol esențial în întregul proces de analiză, oferind **puncte de referință temporale obiective** pentru a interpreta semnalele fiziologice în contextul unor evenimente reale (precum începerea unui test sau momente de pauză).

Fiecare valoare din tags.csv corespunde unui **eveniment marcat de utilizator** prin apăsarea unui buton, sincronizat cu ceasul intern al dispozitivului. Aceste momente au fost transformate în secunde relative față de începutul sesiunii și apoi suprapuse peste graficele semnalelor analizate.

### **Ce a permis acest lucru:**

* **Compararea clară a semnalelor înainte și după eveniment**, evidențiind tranziții fiziologice semnificative (ex. creștere EDA, scădere IBI).
* **Contextualizarea semnalelor** în lipsa etichetelor explicite de „stres” sau „relaxare”.
* **Validarea vizuală și statistică** a momentelor în care corpul a reacționat fiziologic.

## **2.8 Concluzii integrative**

* Primele două componente principale (PC1 și PC2) din PCA **acoperă peste 90%** din variația semnalelor EDA și BVP, ceea ce permite o descriere bidimensională eficientă a stărilor fiziologice.
* Există o **corelație ridicată între semnalele EDA și HR**, ceea ce confirmă reactivitatea sistemului nervos autonom în situații de stres.
* Schimbările în PC1 (din PCA) înainte și după evenimente validează ipoteza că **reacțiile fiziologice sunt detectabile, măsurabile și corelate cu stresul** perceput.
* Semnalele pot fi utilizate pentru **detectarea automată a stresului** în aplicații precum wearable devices, asistență cognitivă, biofeedback sau monitorizare educațională.

## **Avantaje ale abordării**

* Permite vizualizarea stresului ca fenomen latent, dar cuantificabil.
* Folosește metode matematice riguroase, ne-supervizate.
* Corelează măsurători reale cu reacții psihologice fără a depinde de auto-raportare.
* Metodologia este reproductibilă și scalabilă pentru alte studii sau populații.

## **Limitări**

* Datele nu sunt etichetate explicit pe intervale stres/non-stres – analiza este orientativă, bazată pe momentul evenimentelor.
* Volumul de date este relativ mic – analiza completă pe populații extinse ar necesita mai multe sesiuni.
* Nu s-au utilizat metode supervizate de clasificare (ML), dar acestea pot fi adăugate în etape ulterioare.

# **Rezumat Concluzii**

Analiza datasetului *Wearable Exam Stress* a demonstrat că semnalele fiziologice colectate de la senzori purtabili (EDA, BVP, HR, TEMP, IBI, ACC) pot reflecta în mod fidel reacții ale corpului uman la condiții de stres sau repaus.

Prin aplicarea **Transformatei Karhunen-Loève (PCA)** pe semnalele EDA, BVP și ACC, au fost extrase componente principale care acoperă peste **90% din variația totală a semnalelor**. Aceste componente au evidențiat modificări semnificative în jurul evenimentelor marcate (ex. începutul unui test), confirmând prezența unor răspunsuri fiziologice detectabile.

Semnalele lente precum TEMP și HR au fost analizate prin metode **descriptive sincronizate cu evenimente**, în conformitate cu **FF08-3**, indicând variații ușoare dar relevante ale ritmului cardiac și temperaturii corporale.

Pentru semnalul IBI, care constă în serii discrete și neregulate, a fost aplicată o **analiză statistică segmentată** (FF08-2), demonstrând că frecvența cardiacă crește semnificativ în timpul momentelor de stres.

## **Concluzie generală**

**Transformata Karhunen-Loève (PCA), aplicată pe ferestre mobile, s-a dovedit o metodă eficientă și robustă pentru extragerea trăsăturilor dominante din semnalele fiziologice.**

Metoda este:

* ne-supervizată,
* reproductibilă,
* aplicabilă pe semnale purtabile reale,
* sensibilă la modificări subtile induse de stres.

## **Limite**

* Nu există etichete explicite pentru fiecare moment de stres, ceea ce impune o analiză orientată pe evenimente.
* Volumul de date este moderat, iar generalizarea necesită colectarea de date de la mai mulți participanți.
* Analiza nu a integrat metode de clasificare automată (ex. machine learning), dar poate fi extinsă în această direcție.