

### A. Four Methods for Fatigue Labeling

We initially adjusted the baseline and performed min-max normalization on four indicators: subjective data (KSS), AX-CPT reaction time (RT), AX-CPT accuracy (ACC), and button press frequency (PF).

#### 1. Overall time method (O+T):

Using the overall time method to determine fatigue, we first performed an analysis of variance (ANOVA) on the indicators (KSS, RT, ACC, PF). The results are shown in Table 7. For indicators that showed statistical significance in the ANOVA (KSS, ACC, PF), we conducted further post-hoc analyses. Our assumption was that all participants were relatively alert at the start of the experiment, as determined by the experimental design. Fatigue onset was identified when significant differences appeared in the relevant indicators compared to the initial stage of the experiment (AX-CPT1). Moments following the critical point were labeled as fatigued, while moments before were labeled as non-fatigued. The specific labeling results are shown in Table 8.

**Table 7**

ANOVA results based on overall time method.

	F	df1	df2	P
KSS	46.835	8	108	***
RT	0.437	8	109	
ACC	2.108	8	109	*
PF	18.634	8	109	***

\* $p < .05$ , \*\*\* $p < .001$

#### 2. Objective and subjective data method (O+D):

For the O+D method, we considered both subjective and objective indicators for all participants and used K-means clustering. Since K-means is an unsupervised learning method, the definition of labels is essentially random. To ensure the validity of the clustering

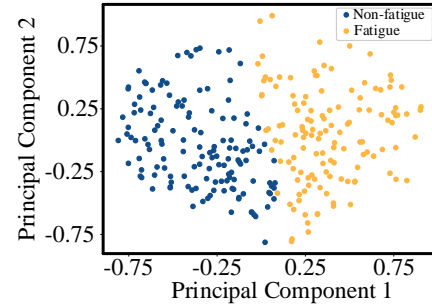
results, we first visualized the data to confirm that the distribution of clusters correlated with the expected fatigue states. We then defined clusters exhibiting high-performance behavior as non-fatigued, while other clusters were defined as fatigued. The clustering results are shown in Fig. 12.

**Table 8**

Multidimensional Fatigue Labeling Based on Overall Time Method

Stage	KSS Prob	ACC Prob	PF Prob	Classify
AXCPT1	—	—	—	0
AXCPT2				0
AXCPT3	***			0
AXCPT4	***			0
AXCPT5	***			0
AXCPT6	***			0
AXCPT7	***			0
AXCPT8	***		*	1
AXCPT9	***		***	1

\* $p < .05$ , \*\*\* $p < .001$



**Fig. 12:** PCA Visualization of Labels Using the O+D Method

#### 3. Individual time method (I+T):

For the I+T method, we used the cumulative sum method to analyze individual indicator data (KSS, RT, ACC, PF) to determine the fatigue onset moment for each participant. The formula for cumulative deviation (CD) is as follows:

$$S_k = \sum_{i=1}^k \left( x_i - \frac{1}{n} \sum_{i=1}^n x_i \right) \quad (9)$$

where  $S_k$  represents the cumulative deviation from the first data point to the  $k$ -th data point,  $x_i$  represents the  $i$ -th data point, and  $n$  is the total number of data points. The fatigue change point is determined at the position where the cumulative deviation reaches the maximum absolute value:

$$k^* = \arg \max_k |S_k| \quad (10)$$

where  $k^*$  represents the index of the indicator with the maximum absolute cumulative deviation. To evaluate fatigue states more comprehensively, we combined the average change point indicators of the four different measures (KSS, RT, ACC, PF):

$$\bar{k} = \frac{1}{4} \sum_{j=1}^4 k_j^* \quad (11)$$

where  $k_j^*$  is the index of the change point detected for each measure. The label array is defined as follows:

$$L_i = \begin{cases} 0 & \text{if } i < \text{round}(\bar{k}) \\ 1 & \text{if } i \geq \text{round}(\bar{k}) \end{cases} \quad (12)$$

where  $L_i$  indicates whether the participant is in a fatigued state at time  $i$ . Specifically, if  $L_i = 1$ , it indicates fatigue; if  $L_i = 0$ , it indicates non-fatigue.

#### 4. Individual subjective and objective data method (I+D)

For the I+D method, we used a similar approach to the O+D method but conducted clustering analysis on each participant's indicators individually. The steps are as follows:

1. For each participant, we used K-means clustering on the data, with each cluster representing a potential fatigue state.

2. After clustering, we assigned a label to each participant based on the clustering results. Generally, clusters with high-performance

behavior were assumed to represent non-fatigued states, while other clusters represented fatigued states.