Vol: 2, Issue: 2, 2024
Pages:172-178
Received: 24 December 2024
Accepted: 15 January 2025

CURRENT TRENDS IN COMPUTING
Research Article
DOI: 10.71074/CTC.1606104

DATASET OF EASY SCREEN P300 SPELLER BRAINCOMPUTER INTERFACE DESIGN

ABDULLAH BILAL AYGUN[1*], AHMET RESIT KAVSAOGLU[2]

[1]*Biomedical Engineering Department, Karabük University, 78050, Karabük, Türkiye*

ABSTRACT. **A dataset has been created to support advancements in brain-computer interface (BCI) research, particularly focusing on P300 speller systems and electroencephalography (EEG) signal analysis. This dataset provides detailed EEG recordings from 30 healthy participants during offline analysis, online character recognition, and word-writing tasks. A 16-channel Brain Products V-Amp device was utilized, and data were collected via a 7×7 visual stimulus matrix designed to evoke reliable P300 responses, with stimuli presented in randomized sequences. The dataset comprises raw EEG signals, binary labels, and stimulus timing information structured to facilitate the development of innovative BCI algorithms and real-time applications. This open-access resource enables novel approaches to EEG signal classification and supports the design of adaptive P300 speller interfaces, offering a foundation for advancing assistive technologies and neuroscience research.**

1. INTRODUCTION

The P300 speller paradigm represents a prominent approach in brain-computer interface (BCI) research, offering a reliable communication method for individuals with severe motor disabilities. This study emphasizes providing a detailed dataset derived from electroencephalography (EEG) recordings, enabling the development and evaluation of novel methodologies and algorithms for BCI applications. The Brain Products V-Amp 16 Channel EEG system has structured data into three main tasks: offline analysis, online character recognition, and word-writing experiments. This dataset builds upon earlier work presented by Aygun and Kavsaoglu (2022) in in [1] and is made openly accessible to encourage replication efforts and collaborative research. By structuring the dataset with detailed annotations and compatibility with widely-used analysis platforms such as MATLAB, this resource is intended to support innovation in BCI system design and signal processing.

2. EXPERIMENTAL DESIGN

EEG data have been collected to address three distinct objectives: offline analysis, online monitoring, and word-writing tasks. Each session involved a stimulus matrix with 7 rows and 7 columns, where the stimuli flashed in a randomized order. Each stimulus was presented 15 times, consisting of a 100ms ”on”

E-mail address: abilalaygun@karabuk.edu.tr (*), kavsaoglu@karabuk.edu.tr.
Key words and phrases. Brain-computer interface (BCI), P300 speller, EEG..

period followed by a 75ms ”off” period. Data were recorded for 700 milliseconds after the initiation of each flash, resulting in 210 individual stimulus events per session. Each session was treated as an independent trial.

During the offline analysis phase, EEG data were recorded from 30 participants across 25 sessions each, except for one participant who completed 24 sessions. Participants were instructed to focus on different regions of the visual stimulus matrix, as depicted in Figure 1, ensuring that any part of the matrix could serve as the source of the stimulus. This approach enabled robust and comprehensive data acquisition for developing a classification model.

FIGURE 1. **Focused offline characters in [1].**

For the online analysis, EEG data were recorded in a single session for each character recognition task. In the online analysis, 30 characters were focused on. The online analysis data includes 30 sessions and the corresponding label data.

In the word-writing application, participants could complete the entire word by matching one of the characters from E1 to E20 with the corresponding index number of the desired word after detecting the first character. These characters, as depicted in Figure 2, were displayed in a sequence ranging from 1 to 20 on the side of the interface. This approach enabled participants to write long words by detecting only two characters. The data were structured accordingly. The data were structured accordingly. Table 1 provides the list of characters that each participant focused on during the task. S1–S30 represents the numbering of participants. The session numbers are also specified in the table.

Participants engaged in the online sessions by focusing on their desired characters in specific areas of the stimulus matrix, such as the lower right, lower left, upper right, upper right, upper left, and center regions.

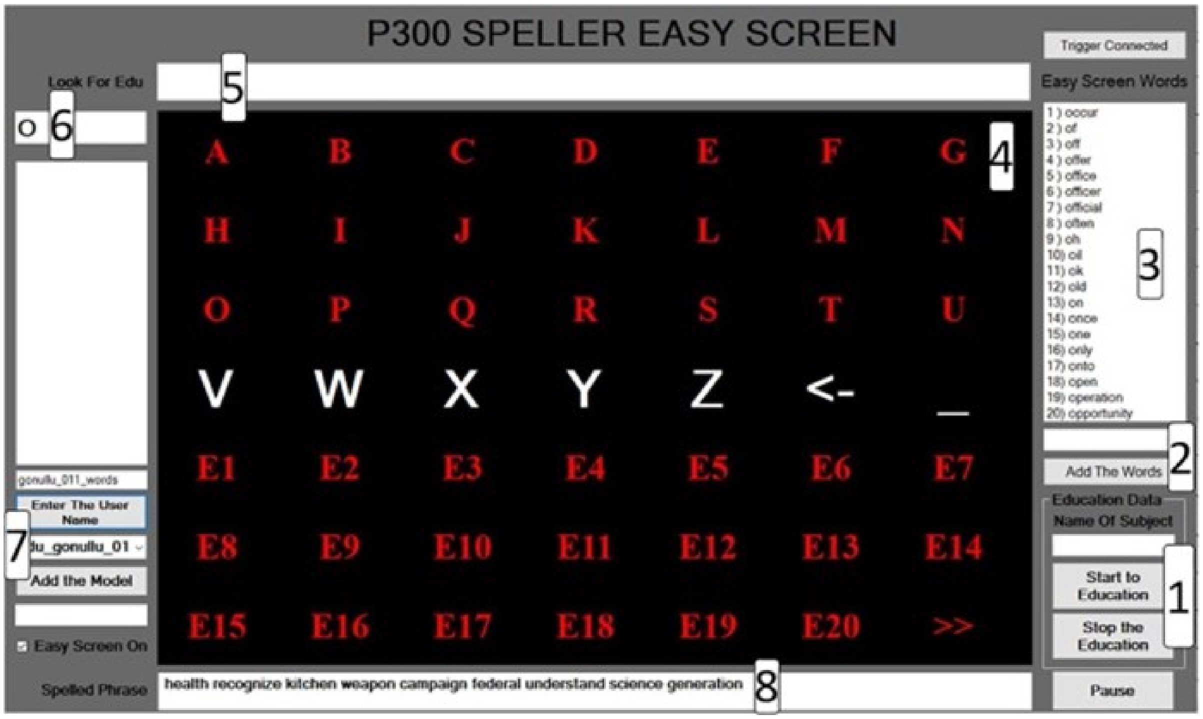

FIGURE 2. **These characters, as depicted in Figure 2, were displayed in a sequence ranging from 1 to 20 on the side of the interface [1].**

3. Dataset Description

The dataset is structured into three main categories:

- Offline Data: Includes raw EEG signals, binary labels indicating the presence or absence of P300 responses, and the timing of stimulus flashes.
- Online Characters: Contains single-session data specifically for individual character recognition experiments.
- Online Words: Comprises multi-session data for word-writing tasks, segmented into trials corresponding to each stimulus presentation.

All data files are provided in '.mat' format, ensuring compatibility with MATLAB and various other analysis platforms. Comprehensive annotations are included with each file to facilitate straightforward data interpretation and usage. Details of preprocessing steps and file naming conventions are also provided to enhance transparency.

This application was tested on 30 participants, 11 healthy females, and 19 healthy males, as shown in Table 2. Among the participants, 11 had mild farsightedness and wore glasses, while the remaining 19 participants had no vision impairments. Based on the information provided by the participants, no chronic, mental, or psychological disorders were identified.

TABLE 1. **The table indicates which character was focused on during each session**

Session Num	**S1**	**S2**	**S3**	**S4**	**S5**	**S6**	**S7**	**S8**	**S9**	**S10**
1	A	B	R	A	B	A	H	A	A	D
2	E16	E17	E14	E16	E4	E20	E11	E16	E7	E2
3	E	E	T	B	D	D	R	S	S	B
4	E9	E16	E10	E3	E2	E10	E18	E6	E14	E16
5	G	W	E	C	A	E	K	D	D	F
6	E13	E10	E18	E13	E20	E6	E6	E15	E3	E7
7	I	T	K	P	S	S	W	F	B	R
8	E20	E9	E6	E7	E8	E11	E10	E3	E17	E8
9	M	H	P	T	C	B	C	E	F	H
10	E10	E11	E18	E10	E1	E15	E3	E16	E8	E18
11	W	S	M	M	Z	C	F	R	R	L
12	E4	E6	E20	E5	E3	E6	E11	E7	E9	E14
13	L	K	A	L	W	M	U	N	H	O
14	E2	E8	E2	E9	E5	E4	E2	E8	E19	E5
15	S	A	C	D	E	R	S	L	L	M
16	E8	E7	E19	E20	E11	E6	E7	E12	E15	E9
17	D	F	D	E	R	T	G	B	O	A
18	E13	E3	E13	E1	E3	E9	E5	E2	E6	E6
19	P	P	S	O	K	K	O	H	M	S
20	E3	E20	E1	E4	E15	E8	E20	E10	E10	E14
Session Num	**S11**	**S12**	**S13**	**S14**	**S15**	**S16**	**S17**	**S18**	**S19**	**S20**
1	H	U	K	H	E	G	A	T	S	T
2	E8	E8	E6	E19	E4	E5	E7	E7	E11	E12
3	B	F	S	V	M	T	R	H	A	B
4	E12	E16	E7	E6	E10	E17	E4	E12	E8	E6
5	D	T	B	A	R	A	T	P	P	M
6	E5	E10	E11	E16	E15	E11	E10	E13	E20	E8
7	P	N	O	S	S	R	N	C	E	K
8	E18	E4	E14	E11	E13	E18	E14	E2	E4	E1
9	T	O	R	T	G	D	C	F	R	O
10	E9	E20	E18	E7	E5	E2	E9	E11	E2	E3
11	A	H	U	U	A	K	P	K	T	F
12	E16	E5	E2	E2	E16	E8	E9	E10	E11	E6
13	N	L	W	P	U	F	G	E	B	D
14	E14	E12	E10	E9	E2	E6	E12	E6	E19	E12
15	W	R	E	F	H	L	O	O	V	N
16	E10	E18	E16	E13	E11	E9	E20	E19	E7	E8
17	K	S	T	W	Z	S	J	N	D	L
18	E8	E15	E9	E18	E9	E15	E3	E1	E4	E11
19	G	V	N	M	L	V	W	S	R	V
20	E13	E6	E8	E10	E12	E4	E16	E6	E6	E8
Session Num	**S21**	**S22**	**S23**	**S24**	**S25**	**S26**	**S27**	**S28**	**S29**	**S30**
1	T	B	S	P	I	T	B	B	A	A
2	E5	E6	E12	E15	E17	E17	E6	E8	E4	E14
3	V	M	N	C	R	S	K	D	B	D
4	E3	E4	E16	E8	E7	E12	E5	E14	E9	E1
5	D	S	O	L	S	P	M	G	E	Y
6	E5	E14	E7	E7	E6	E11	E12	E9	E3	E4
7	A	F	F	M	Q	V	N	P	H	K
8	E2	E5	E6	E13	E3	E7	E9	E6	E1	E7
9	U	V	H	Z	V	Z	O	Y	K	P
10	E9	E2	E11	E11	E2	E5	E12	E7	E10	E10
11	L	H	E	I	G	H	R	H	C	L
12	E17	E11	E14	E2	E13	E13	E8	E3	E5	E9
13	G	Z	P	N	O	A	S	K	T	N
14	E13	E10	E5	E14	E5	E15	E12	E8	E12	E8
15	H	H	A	T	M	B	V	U	G	O
16	E8	E16	E11	E9	E8	E11	E5	E4	E8	E19
17	P	E	R	O	A	J	Y	Z	S	J
18	E13	E18	E9	E3	E4	E2	E8	E1	E15	E3
19	K	A	M	D	L	N	T	N	R	U
20	E6	E17	E6	E16	E9	E10	E15	E12	E16	E10

TABLE 2. **The characteristics of the participants [1]**

	Age	Height	Weight	Glasses Users
Mean	34.47 ± 9.69	172.2 ± 8.75	75.43 ± 17.62	11
Range	(18 – 61)	(160 – 198)	(48 – 137)	

The experiments were conducted over 20 days in a quiet environment, involving two individuals at a time: one participant and one experiment assistant. Visual stimuli were presented to the participants on a computer screen placed 1 meter away under moderate brightness and daylight conditions. The EEG device was also connected to the same computer, which displayed the stimuli to record brain signals. Table 1 provides the characteristics of the participants.

Brain signals were recorded using a Brain Products V-Amp 16 Channel EEG device (V-Amp, Brain Products GmbH, Gilching, Germany). Electrodes were positioned according to the International 10-20 system, as shown in Figure 2, and signals were recorded from 16 channels. The signals were filtered using a 1-12 Hz Butterworth band-pass filter and a 50 Hz Notch filter. The EEG data was digitized at 2 kHz and subjected only to filtering, with no further processing applied. The dataset is presented without downsampling to retain its original resolution.

4. APPLICATIONS AND FUTURE DIRECTIONS

Applications and Future Directions This dataset offers a comprehensive foundation for advancing brain-computer interface (BCI) technology. Key potential applications include:

- Development of adaptive speller interfaces to enable efficient and user-friendly communication systems.
- Exploration of novel machine learning algorithms, facilitating innovative approaches to EEG signal classification.
- Validation of real-time BCI systems to support the development of assistive communication technologies.

This resource promotes reproducibility, fosters innovation, and facilitates collaboration within the neuroscience community by providing open access to a well-annotated and structured dataset. Researchers are encouraged to leverage the dataset to explore new directions in BCI system design and signal processing.

5. LIMITATIONS

While this dataset offers significant potential for advancing BCI research, several limitations should be considered. The participant pool primarily comprises healthy individuals, limiting its applicability to clinical populations. Furthermore, environmental conditions, such as lighting and potential distractions during experiments, were controlled but may still influence the EEG signals. Variability in participant focus and attention could also introduce noise into the data. Future studies could address these limitations

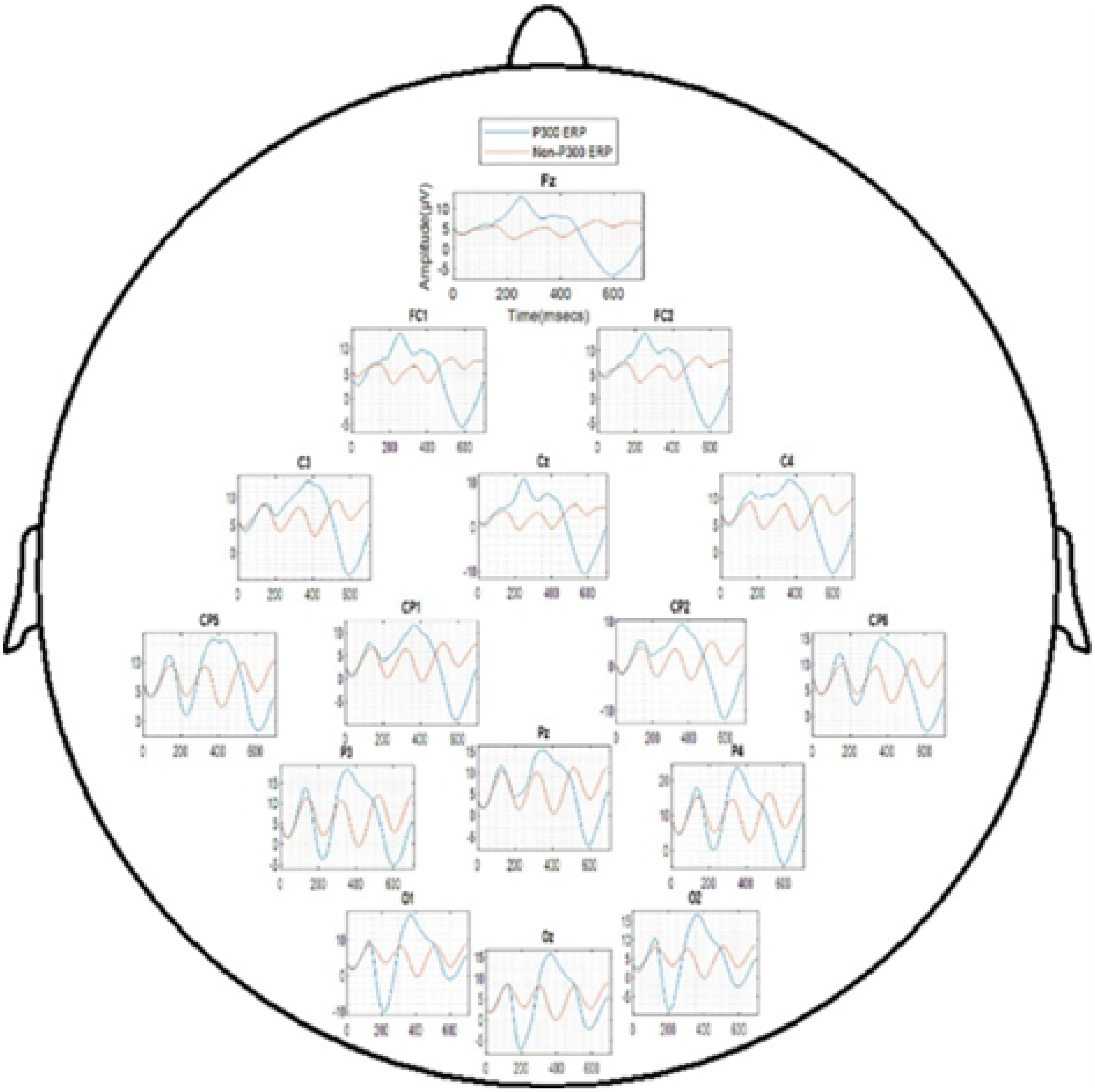

FIGURE 3. **The figure illustrates the channels from which the data were recorded [1].**

by incorporating clinical participants, expanding the dataset under varied environmental conditions, and improving signal acquisition protocols.

6. CONCLUSION

This dataset serves as a bridge between theoretical research and practical applications in BCI design. Making this data publicly available aims to promote open science initiatives and stimulate innovations in assistive technology. Researchers are encouraged to utilize this resource to enhance reproducibility and foster collaboration within the neuroscience community. Future studies are recommended to build upon this dataset by exploring novel adaptive algorithms, incorporating clinical populations, and validating real-time BCI applications in diverse settings.

DECLARATIONS

- **Contribution Rate Statement:** All authors have contributed equally.
- **Conflict of Interest:** The author declares no conflict of interest.
- **Data Availability:** The dataset is hosted on the Zenodo platform to ensure support scientific development and can be accessed using the link https://doi.org/10.5281/zenodo.13861638. It is distributed under the Creative Commons Attribution 4.0 license, permitting unrestricted use provided proper citation is given. Researchers are encouraged to utilize the raw EEG signals, stimulus timing files, and labels to develop machine learning models or investigate novel brain-computer interface (BCI) designs. Additional documentation is provided alongside the dataset to guide users in its application.
- **Statement of Support and Acknowledgment:** None

REFERENCES

[1] A. B. Aygun, A. R. Kavsaoglu, An innovative P300 speller brain–computer interface design: Easy screen, Biomedical Signal Processing and Control, Volume 75, 2022, 103593, https://doi.org/10.1016/j.bspc.2022.103593.