Mainsah, B., Fleeting, C., Balmat, T., Sellers, E., & Collins, L. (2024). bigP3BCI: An Open, Diverse and Machine Learning Ready P300-based Brain-Computer Interface Dataset (version 1.0.0).

Table 1: Summary of BCI Studies in bigP3BCI v1.0.0 Dataset

Study Name	Related Publication	No. of Subjects	ALS Study?	No. of Sessions	[Stimulus Paradigm(s)], Grid Size[§]
A	C. S. Throckmorton, D. B. Ryan, B. Hamner, K. Caves, K. A. Colwell, E. W. Sellers, and L. M. Collins, "Towards clinically acceptable BCI spellers: Preliminary results for different stimulus selection patterns and pattern recognition techniques," presented at the *4th International BCI Meeting*, Asilomar, CA, 2010.	13		1	[RC, CB, RD] 9 × 8
B	N. A. Gates, C. K. Hauser, and E. W. Sellers, "A longitudinal study of P300 brain-computer interface and progression of amyotrophic lateral sclerosis," in *International Conference on Foundations of Augmented Cognition*, 2011, pp. 475-483.	18	ALS	Var	CB, 6 × 6
C	B. O. Mainsah, K. D. Morton, L. M. Collins, E. W. Sellers, and C. S. Throckmorton, "Moving away from error-related potentials to achieve spelling correction in P300 spellers", *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 23, no. 5, pp. 737-743, 2015.	19		1	CB, 9 × 8
D	B. O. Mainsah, K. A. Colwell, L. M. Collins, and C. S. Throckmorton, "Utilizing a language model to improve online dynamic data collection in P300 spellers," *IEEE Transactions on Neural Systems and Rehabilitation Engineering,* vol. 22, no. 4, pp. 837-846, 2014.	17		1	RC, 9 × 8
E*		8		1	CB, 9 × 8
F	B. O. Mainsah, L. M. Collins, K. A. Colwell, E. W. Sellers, D. B. Ryan, K. Caves, and C. S. Throckmorton, "Increasing BCI communication rates with dynamic stopping towards more practical use: An ALS study," *Journal of Neural Engineering*, vol. 12, no. 1, p. 016013, 2015.	10	ALS	3	CB, 9 × 8
G	B. Mainsah, K. Morton, L. Collins, and C. Throckmorton, "Extending language modeling to improve dynamic data collection in ERP-based spellers." *6th International Brain-Computer Interface Conference*, Graz, Austria, 2014.	20		1	CB, 9 × 8
H	D. Kalika, L. Collins, K. Caves, and C. Throckmorton, "Fusion of P300 and eye-tracker data for spelling using BCI2000," *Journal of Neural Engineering*, vol. 14, no. 5, p. 056010, 2017.	16		1	CB, 9 × 8
I[†]		13		1	[CB, PB], 9 × 8
J	B. Mainsah, G. Reeves, L. Collins, and C. Throckmorton, "Optimizing the stimulus presentation paradigm design for the P300-based brain-computer interface using performance prediction," *Journal of Neural Engineering*, vol. 14, no. 4, p. 046025, 2017.	20		1	[RC, PB], 6 × 6
K	B. Mainsah, D. Kalika, L. Collins, S. Liu, and C. Throckmorton, "Information-based adaptive stimulus selection to optimize communication efficiency in brain-computer interfaces," *Advances in Neural Information Processing Systems*, vol. 31, 2018.	5		1 or 2	[CB, AD], 9 × 8
L	D. B. Ryan, K. A. Colwell, C. S. Throckmorton, L. M. Collins, K. Caves, and E. W. Sellers, "Evaluating brain-computer interface performance in an ALS population: Checkerboard and color paradigms," *Clinical EEG and Neuroscience*, vol. 49, no. 2, pp. 114-121, 2018.	11	ALS	1	[RC, CB, CBcol] 6 × 6

M[#]	X. J. Chen, D. Kalika, C. S. Throckmorton, L. M. Collins, and B. O. Mainsah, "Bayesian Adaptive Stimulus Optimization in Stimulus-driven Brain Computer Interfaces," 2024.	21		1	[CB, ADdiff], 9 × 8
N	J. Clements, E. Sellers, D. Ryan, K. Caves, L. Collins, and C. Throckmorton, "Applying dynamic data collection to improve dry electrode system performance for a P300-based brain–computer interface," *Journal of Neural Engineering*, vol. 13, no. 6, p. 066018, 2016.	8	ALS	2	CB, 6 × 6
O	G. Frye, C. Hauser, G. Townsend, and E. Sellers, "Suppressing flashes of items surrounding targets during calibration of a P300-based brain–computer interface improves performance," *Journal of Neural Engineering*, vol. 8, no. 2, p. 025024, 2011.	18		2	[CB, sCB], 9 × 8
P	D. B. Ryan, G. Frye, G. Townsend, D. Berry, S. Mesa-G, N. A. Gates, and E. W. Sellers, "Predictive spelling with a P300-based brain–computer interface: Increasing the rate of communication," *Intl. Journal of Human–Computer Interaction*, vol. 27, no. 1, pp. 69-84, 2010.	19		2	CB, 9 × 8
Q	D. Ryan, G. Townsend, N. Gates, K. Colwell, and E. Sellers, "Evaluating brain-computer interface performance using color in the P300 checkerboard speller," *Clinical Neurophysiology*, vol. 128, no. 10, pp. 2050-2057, 2017.	36		3	[CB, CBc], 9 × 8
R	M. Kellicut-Jones and E. Sellers, "P300 brain-computer interface: Comparing faces to size matched non-face stimuli," *Brain-Computer Interfaces*, vol. 5, no. 1, pp. 30-39, 2018.	20		2	CB, 9 × 8
S1	M. R. Jones and E. Sellers, "Faces, locations, and tools: a proposed two-stimulus P300 brain computer interface," *Journal of Neural Engineering*, vol. 16, no. 3, p. 036026, 2019.	10		1	CB, 9 × 8
S2	M. R. Jones and E. Sellers, "Faces, locations, and tools: a proposed two-stimulus P300 brain computer interface," *Journal of Neural Engineering*, vol. 16, no. 3, p. 036026, 2019.	24		1	CB, 9 × 8

Abbreviations: AD, Adaptive; ADdiff, Adaptive Diffuse; ALS, amyotrophic lateral sclerosis; CB, Checkerboard; CBc, Checkerboard Colour; CBs, Checkerboard with suppressed characters; No., number; PB, Performance-Based; RD, Random; RC, Row-Column; Var, variable.

[§]Grid size is specified as *number of rows* × *number of columns* in a matrix layout.

*The experiment protocol of study E is similar to that of study D. [†]The experiment protocol of study I is similar to that of study J. [#]The experiment protocol of study M is similar to that of study K.

Table 2: bigP3BCI Data Dictionary for EDF+ File Header

File Header Field	File Header Subfield	Custom File Header Subfield Label	Values
Patient identification	Patient code	[*Study label*]_[*Subject number*]	E.g., A_01, A_02, etc.
	Sex[a]		- Male - Female
	Date of birth	01-JAN-[*YYYY*]	- 01-JAN-[*YYYY*], date-shifted in years based on assumed recording start date of 01-JAN-2020. - If age is not available, set to 01-JAN-YYYY. Note: All study participants are adults (at least 18 years).
	Patient name	[*Race*]_[*Ethnicity*]_[*ALS Status*]	*Race*[a] - White - Black/African American - Asian - American Indian or Alaska Native - Native Hawaiian or Other Pacific Islander - X (if not available)
			Ethnicity[a] - Hispanic - Non-hispanic - X (if not available)
			ALS Severity[b] - NonALS - ALS_#, where # is the revised ALS Functional Rating Scale (ALSFRS-R) score ranging from 0 to 48 - ALS_X (if ALRFRS-R score is not available)
Recording identification	Start date	01-JAN-2020	All recording start dates set to the same value
	Hospital administration code of the investigation	[*Dataset name*]_[*Dataset version*]_Study[*Study label*]	*Dataset name* - bigP3BCI *Dataset version* - vX.X.X (following convention for semantic versioning) *Study label* - See **Table 4**.
			E.g., bigP3BCI_v1.0.0_StudyA
	Investigator or Technician	SE[*Session number*]	E.g., SE001, SE002, etc.
	Equipment code	[*Equipment model*]	gUSBAmp

[a]Sex, race and ethnicity categories are primarily based on definitions outlined by the NIH.

[b]The revised Amyotrophic Lateral Sclerosis Functional Rating Scale (ALSFRS-R) is an instrument for evaluating degree of functional impairment in individuals with ALS; the total ALSFRS-R score ranges from 0 (worst) to 48 (best).

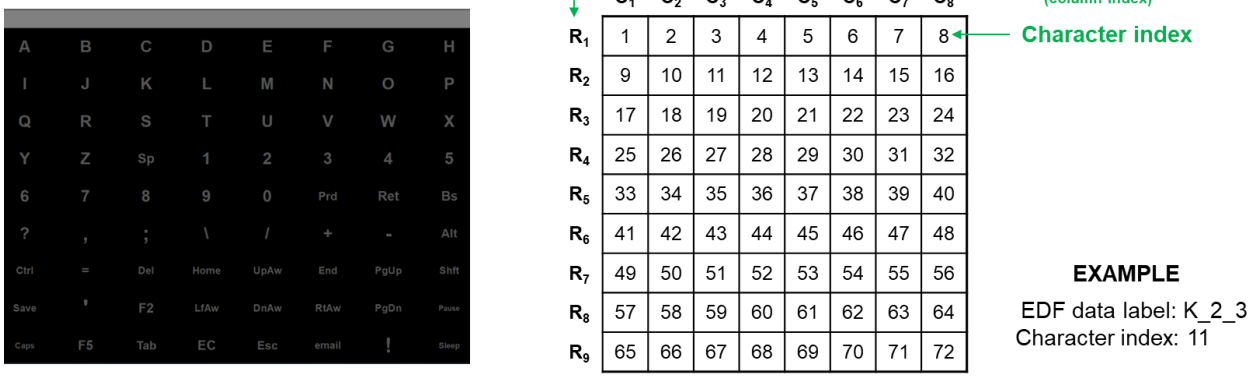
Table 3: bigP3BCI Dictionary for EDF+ Data Label

BCI Data Level*	Data		Data Label[§, #]	Values
Level 0: Biosignals	EEG signals		EEG_<*channelName*> e.g., EEG_Cz	microvolts (μV)
	Eye tracker (ET) signals	Eye data validity	ET<Left/Right>EyeValid	0: Eye tracker data invalid 1: Eye tracker data valid
		Eye gaze position	ET<Left/Right>EyeGaze<X/Y>	- Ratio between 0 and 1, location on the screen the participant is looking - $(X = 0, Y = 0)$ corresponds to the top left of the screen
		Eye position	ET<Left/Right>EyePos<X/Y>	- Ratio between 0 and 1 (eye position relative to the camera in 2D space) - $(X = 0, Y = 0)$ corresponds to the top left of the camera's view
		Eye distance	ET<Left/Right>EyeDist	mm (distance between the screen and the eyes)
		Pupil size	ET<Left/Right>PupilSize	mm
Level 1: BCI Training	Target character trial events		PhaseInSequence	Phase 0: Pre-run Phase 1: Pre-trial Phase 2: During trial Phase 3: Post-trial
	Stimulus events (active during Phase 2)		StimulusBegin	0: stimulus off 1: stimulus on
			StimulusType	0: non-target stimulus event when StimulusBegin = 1 1: target stimulus event
	Character presentation events (active during Phase 2)		<*Character*>_<*rowIndex*>_<*column Index*> E.g., For a 9 x 8 grid (R rows x C columns), K_2_3	0: character not presented 1: character presented
	Target character (active during Phase 2)		CurrentTarget	- Index of target character during a trial - $(r-1) * R + C$, where r is row index, R is number of rows and C is number of columns - E.g., index of K_2_3 in a 9 x 8 grid is 11. See Error! Reference source not found..
Level 2: BCI Feedback	Predicted target character (active during Phase 3)		Selected<Target/Row/Column>	Index of predicted target character/row/column). See Error! Reference source not found..
	Presented feedback (active during Phase 3)		DisplayResult	0: Feedback not presented 1: Feedback presented
			FakeFeedback	Index of presented target character overriding predicted target character during a character trial. See Error! Reference source not found..

*****BCI data levels for data storage and sharing as defined by the IEEE P2731 Working Group Standard for a Unified Terminology for Brain-Computer Interfaces.

§ <···> indicates a substring within the angle brackets. <*option*> in italics indicates a variable substring within the angle brackets. <option1/option2/…/optionN> in solid indicates a variable substring from the fixed set {option1, option2,…, option}. E.g., ET<Left/Right>PupilSize indicates two options, ETLeftPupilSize and ETRightPupilSize, for the pupil size label of the eye tracker (ET).

Most labels for the EDF+ data records are derived from parameter definitions in the *P3SpellerTask* and *EyeTrackerLogger* modules in BCI2000.



	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8
R_1	1	2	3	4	5	6	7	8
R_2	9	10	11	12	13	14	15	16
R_3	17	18	19	20	21	22	23	24
R_4	25	26	27	28	29	30	31	32
R_5	33	34	35	36	37	38	39	40
R_6	41	42	43	44	45	46	47	48
R_7	49	50	51	52	53	54	55	56
R_8	57	58	59	60	61	62	63	64
R_9	65	66	67	68	69	70	71	72

Figure 1: BCI Character Encoding. Example P300 Speller interface showing character definitions (left) and corresponding character index values (right) in *CurrentTarget*, *SelectedTarget* and *FakeFeedback* data records of EDF+ file.

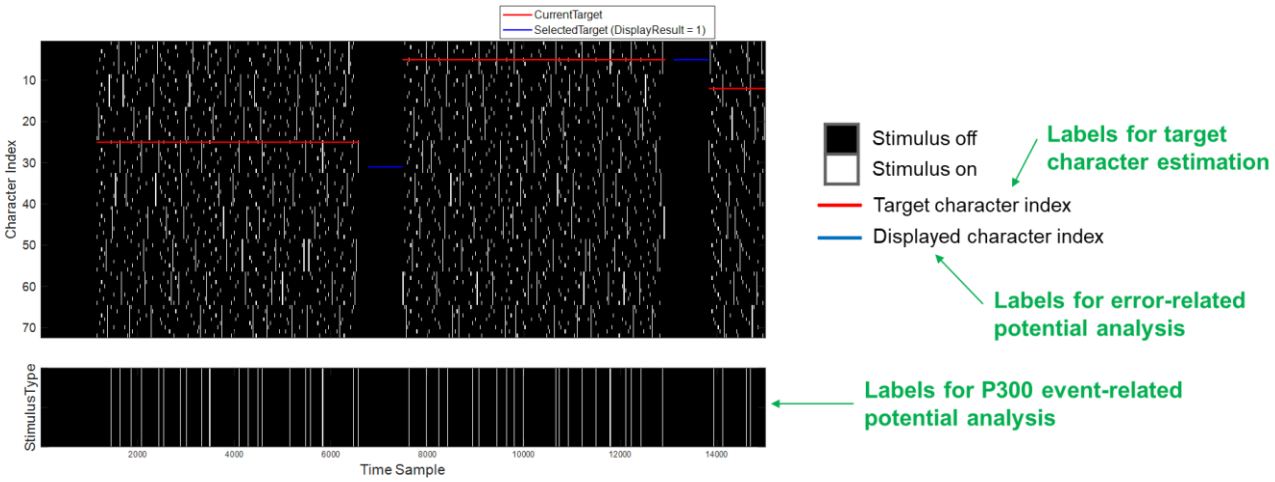


Figure 2: Example Schedule of BCI Events for a 9 x 8 BCI Speller. The stimulus presentation schedule (top panel), target characters and presented characters (top panel) and stimulus type (bottom panel), extracted from an EDF+ file are shown. Data stream of interest to obtain labels for machine learning tasks are indicated. Note that if *FakeFeedback* data stream is non-zero, the BCI *SelectedTarget* is overridden and the displayed character corresponds to the *FakeFeedback* character index value.