H.F.R.I.
Hellenic Foundation for Research & Innovation

D3.1 (v2) – The BINGO Benchmarking Framework

BINGO

Brain Imagined-Speech Communication

Greece 2.0
NATIONAL RECOVERY AND RESILIENCE PLAN

Funded by the European Union
NextGenerationEU

Funded by the European Union
NextGenerationEU

Greece 2.0
NATIONAL RECOVERY AND RESILIENCE PLAN

H.F.R.I.
Hellenic Foundation for Research & Innovation

The research project is implemented in the framework of H.F.R.I call "Basic research Financing (Horizontal support of all Sciences)" under the National Recovery and Resilience Plan "Greece 2.0" funded by the European Union – NextGenerationEU (H.F.R.I. Project Number: 15986).

Dissemination level:	Public (PU)
Contractual date of delivery:	Month 21, 27/08/2025
Actual date of delivery:	Month 23, 27/10/2025
Work Package:	WP3 - Benchmarking framework
Task:	T3.2 Data Recording
Type:	DEM
Approval Status:	final
Version:	V2.0
Number of pages:	19
Filename:	D3.1_BINGO_BenchmarkingFramework_v2.docx

Abstract: This document accompanies the recorded data that have been gathered as part of BINGO and constitutes a sequel of D3.1-v1. The collected data aim to cover two basic aspects of inner speech decoding research: i) A complete and large-scale dataset that supports a wide variety of words (also capable of acting as a generic spelling corpus), and ii) A dataset capable of driving research among the neural interconnections between inner speech in two distinct languages. In this document we present the final experimental protocols, and the appropriate quality metrics that guarantee the validity of the recorded data. This deliverable aims to serve as a document accompanying the released datasets (available at project's website).

HISTORY

Version	Date	Reason	Revised by
V1.1	11/9/2025	Table of Contents	Kostas Georgiadis
V1.2	9/10/2025	Initial Draft	Kostas Georgiadis
V2.0	23/10/2025	Final	Spiros Nikolopoulos

AUTHOR LIST

Organization	Name	Contact Information
CERTH	Fotis P. Kalaganis	fkalaganis@iti.gr
CERTH	Kostas Georgiadis	kostas.georgiadis@iti.gr
CERTH	Maria Kyrou	mariakyrou@iti.gr
CERTH	Spiros Nikolopoulos	nikolopo@iti.gr
CERTH	Ioannis Kompatsiaris	ikom@iti.gr

ABBREVIATIONS AND ACRONYMS

API	Application Programming Interface
BCI	Brain Computer Interface
EEG	ElectroEncephaloGram
LSL	Lab Streaming Layer
PPVT	Peabody Picture Vocabulary Test
ASR	Artifact Subspace Reconstruction
ZCR	Zero-crossing Rate

Contents

INTRODUCTION

This deliverable accompanies two electroencephalography (EEG) datasets collected in the context of research on inner speech decoding. Inner speech—the silent articulation of words without overt vocalization—has attracted growing interest due to its relevance for brain–computer interface (BCI) development, cognitive neuroscience, and assistive communication technologies. High-quality, well-documented EEG datasets are essential for advancing robust decoding methods and ensuring reproducibility across studies.

The first dataset comprises EEG recordings from subjects performing inner speech of letters from the NATO phonetic alphabet. This dataset is designed to support fine-grained analysis of neural patterns associated with internally articulated phonetic units under controlled experimental conditions. The second dataset focuses on inner speech of complete words, including words in English and their corresponding translations in Greek. By incorporating two languages, this dataset enables investigation of language-specific and cross-linguistic neural representations of inner speech.

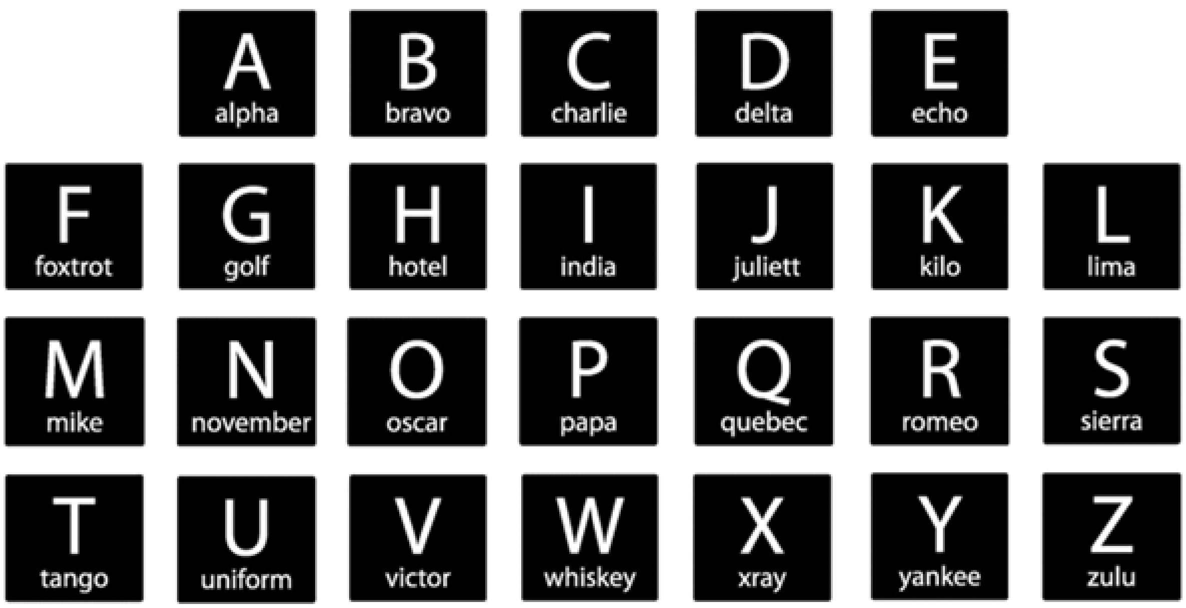
In addition to the EEG recordings, this document provides a comprehensive set of EEG quality metrics aimed at assessing signal integrity, recording reliability, and overall data validity. These metrics are included to support transparent evaluation of the datasets and to facilitate informed use by researchers for downstream analysis and model development. Together, the datasets and accompanying quality assessments constitute a valuable resource for the study of inner speech decoding and related EEG-based applications.

Both datasets can be accessed at project's website (https://bingo-project.gr/) under menu Results -> Datasets.

NATO PHONETIC ALPHABET – CLASSIFICATION TASK

OVERALL EXPERIMENTAL PROTOCOL

The inner speech task was designed using the NATO phonetic alphabet (shown in the Figure below) as a structured linguistic framework to elicit neural responses associated with silent speech production. This approach ensured a controlled and standardized set of stimuli with a diverse phonological. The NATO alphabet constitutes an exceptional opportunity for serving as the basis for inner speech decoding (using solely EEG signals) at a word level but also may serve as a generic alphabet where typing in a letter-by-letter setting can be realized.



Prior to beginning the recording session, all subjects (20 in total) were able to hear the pronunciation of the NATO words so as to minimize the variability across phonological representations. The task incorporated visual cues to prompt participants to internally articulate specific phonetic words without overt speech or subvocalization. The selection of sole visual cues during the recording process minimizes the contamination of brain signals stemming from brain areas highly involved in inner speech (e.g. Broca's area).

Each participant, attended the experimental protocol three times, each time taking place in a different day (referred to as Day1, Day2, and Day3). During the first day, words corresponding to the letters A-M were recorded (35 times per word per participant), whereas the rest (N-Z) during Day 2 (35 times per word per participant). In the third and last day, all the words were recorded again (10 times per word per participant) so as to serve as the test set.

TRIAL STRUCTURE AND TIMING

Our EEG recording protocol consists of two phases. A preparatory one, referred to as Calibration, and the main one, referred to as Inner Speech Trials.

The calibration phase is an essential preparatory step to establish a baseline neural activity profile for each participant. It consists of two conditions, each lasting 1 minute:

Eyes Open Condition: Participants are instructed to keep their eyes open while fixating on a centrally presented fixation cross ("+"). This phase helps capture baseline neural activity associated with wakeful rest and visual fixation, primarily recording alpha activity suppression in the occipital cortex.

Eyes Closed Condition: Participants are then asked to close their eyes while remaining relaxed. This phase is crucial for detecting resting-state alpha rhythms, which typically increase when the eyes are closed.

These calibration steps allow for artifact identification, ensuring that eye blinks, muscle activity, and environmental noise are accounted for before the main task begins.

Following calibration, the inner speech trials begin. Each trial consists of a structured sequence of visual stimuli and silent speech production to elicit neural responses associated with covert articulation. To avoid block effects, all words were randomized and presented in a randomized sequence (different for each participant).

Each trial follows a strict timeline (we note that we consider t=0 the time that actual inner speech takes place) being the with distinct time points and stimulus presentations:

1. **t = -2.5s:** A randomly selected NATO phonetic alphabet word (e.g., "Charlie") is presented on the screen. The participants were instructed to keep the depicted word in their mind without overt speech or subvocalization.
2. **t = -1s**: The word disappears from the screen, and a countdown takes place.
3. **t = 0s**: A fixation cross ("+") appears, signaling the participant maintain focus, avoid movement-related artifacts, and to begin the inner speaking of the previously seen prompt.
4. **t = 1.5s**: The fixation cross disappears, marking the end of the trial. Participants remain in a resting state until the next trial begins.

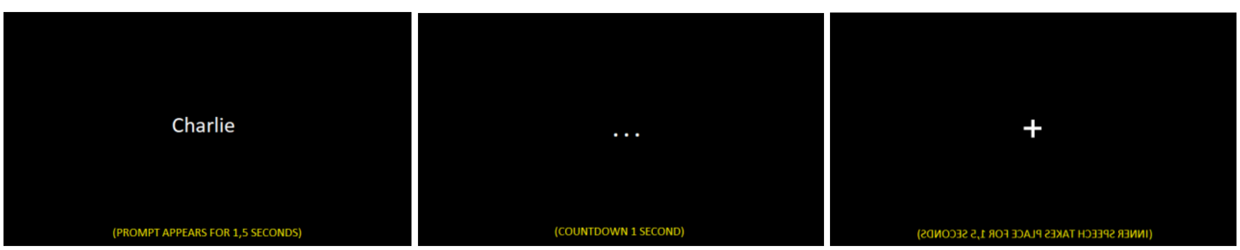


Figure 1 Illustrating in chronological order the visual cues and timings of a single trial.

CROSS-LINGUISTIC INNER SPEECH ELICITATION

RATIONALE AND TRANSITION FROM THE INITIAL DESIGN

In the initial version of the experimental protocol (described in D3.1a), cross-linguistic inner speech was to be elicited using selected items from the NATO phonetic alphabet that also possess a clear and widely recognized lexical meaning in Greek (e.g., Hotel → Ξενοδοχείο). The underlying assumption was that identical prompts at the conceptual level would allow participants to covertly articulate the same lexical item in two languages, thereby supporting a direct comparison of language-specific brain activation patterns during inner speech.

However, during the refinement of the protocol, a practical and methodological limitation emerged. Only a small subset of NATO alphabet words have a concise, unambiguous, and commonly used lexical equivalent in Greek. This substantially restricted the size of the stimulus set and limited experimental control over key psycholinguistic variables (e.g., word-length, frequency, and grammatical features). As a result, the initial approach risked introducing confounds related to uneven lexical properties and insufficient statistical power.

To address these limitations, the protocol was revised to adopt a linguistically and neuroscientifically grounded stimulus selection strategy, while preserving the overall experimental structure of the paradigm aiming to study the brain's responses and elicited patterns when the inner speech task is performed for a word of the same meaning in two different languages (i.e. Greek and English).

LINGUISTIC AND NEUROSCIENTIFIC CONSIDERATIONS GUIDING STIMULUS SELECTION

To ensure that the revised protocol would remain theoretically sound and experimentally interpretable, the transition away from NATO-alphabet-based items necessitated an explicit grounding in established models of lexical access and cross-linguistic word processing. Lexical access and word processing are regulated by several well-established factors, including age of acquisition, word-length (measured in letters, syllables, or phonemes), lexical frequency, and neighborhood density. In cross-linguistic paradigms, additional constraints arise from differences in orthographic transparency between languages. Greek is a relatively transparent (shallow) orthography with consistent grapheme–phoneme mappings, whereas English is a deep orthography with more ambiguous mappings. Consequently, matching stimuli solely on the basis of number of letters or syllables would be insufficient; phoneme-based word-length matching is more appropriate for ensuring comparable processing demands across languages.

Furthermore, any morphophonological overlap between lexical items can induce priming effects, which constitute a confounding factor in lexical processing experiments. Semantic ambiguity, animacy, and morphosyntactic features (including language-specific uninterpretable features such as grammatical gender in Greek) can also modulate processing cost and neural activation patterns. These considerations motivated a tightly controlled selection of lexical items that minimizes overlap and balances relevant linguistic features across languages.

FINALIZED STIMULUS SET AND LEXICAL CONTROLS

Considering the above, for the purposes of the present study, we consider the factors of word-length, frequency and features involved. Specifically, we identified 15 words in English and Greek controlling for the following lexical features [+Nouns], [-Human], [-Animate], [+Concrete], [-Ambiguous], [+Countable], [+Singular]. The trials selected are up to three syllables long and are matched in word-length measured in N of phonemes as there are differences with regard to the orthographic transparency of the two languages (MGR = 4.73, SDGR = 0.79; MENG = 5.00, SDENG = 0.84; t(28) = -.888, p =.382). With regard to frequency, we opted for everyday lexical items consulting the items included in standardized test for lexical ability screening; Peabody Picture Vocabulary Test – PPVT (Dunn & Dunn, 2007) and Renfrew Word Finding Vocabulary Test (Renfrew, 2001). With regard to grammatical gender in Greek, masculine nouns were excluded since the lexical features and word-length criteria for inclusion were not met; the finalized trial set includes nine (9) neuter and six (6) feminine nouns. The list of items per language appears in Table 1 below:

Greek Word	IPA phonetic transcription	Gender	N of letters	N of Syllables	N of phonemes	English Word	IPA phonetic transcription	N of letters	N of Syllables	N of phonemes
κερί	kɛri	neut	4	2	4	candle	ˈkændəl	6	2	6
βέλος	ˈvelos	neut	4	2	5	arrow	ˈæroʊ	5	2	4
κιθάρα	kiˈθara	fem	6	3	6	guitar	gɪˈtɑːr	6	2	5
κεραία	kɛˈre.a	fem	6	3	5	antenna	ænˈtɛnə	7	3	6
πυξίδα	piˈksi.ða	fem	6	3	6	compass	ˈkɒmpæs	7	2	6
μήλο	ˈmi.lo	neut	4	2	4	apple	ˈæpəl	5	2	4
όχημα	ˈo.xi.ma	neut	5	3	5	vehicle	ˈviːɪkəl	7	3	6
σέλα	ˈsela	fem	4	2	4	saddle	ˈsædəl	6	2	5
ρόδα	ˈroða	fem	4	2	4	wheel	ˈhwiːl	5	1	4
πιάτο	ˈpja.to	neut	5	2	5	plate	ˈpleɪt	5	1	5
ζώνη	ˈzo.ni	fem	4	2	4	belt	ˈbɛlt	4	1	4
φτερό	fteˈro	neut	5	2	5	feather	ˈfɛðər	7	2	5
ζάρι	ˈza.ri	neut	4	2	4	dice	ˈdaɪs	4	1	4
μολύβι	moˈli.vi	neut	6	3	6	pencil	ˈpɛnsəl	6	2	6
κουμπί	kuˈbi	neut	6	2	4	button	ˈbʌtən	6	2	5

EXPERIMENTAL PROTOCOL TIMELINE

The experimental protocol adopted for the cross-linguistic scenario, follows the same underlying rationale as the one originally implemented for the NATO alphabet paradigm (see D3.1a), ensuring continuity and comparability across experimental conditions. In particular, the protocol is designed to elicit inner speech in a controlled and reproducible manner.

During the experiment, participants are visually presented with a single lexical item per trial and are instructed to covertly articulate the word in the target language. Each trial is structured as a temporally well-defined sequence of visual cues and silent speech production phases, explicitly designed to isolate neural activity associated with covert articulation. More specifically, each trial (we note that trials take place after eyes open/closed calibration period; as in NATO experimental protocol) adheres to a strict timeline with clearly defined time points and stimulus presentations:

- t = -2.5 s: A randomly selected lexical item is presented on the screen (e.g., "vehicle"). Participants are instructed to attend to the word and retain it in working memory, without producing any overt speech or engaging in subvocal rehearsal.
- t = -1 s: The lexical item is removed from the screen, eliminating external visual input while maintaining the internal representation of the word. A countdown takes place to prepare the participant for the inner speech.
- t = 0 s: A fixation cross ("+") appears, signaling participants to maintain visual focus, remain still to avoid movement-related artifacts, and to initiate the inner speech of the previously presented lexical item.
- t = 1.5 s: The fixation cross disappears, marking the end of the trial. Participants then remain in a passive resting state until the onset of the next trial.

Each participant completes ten (10) repetitions per lexical item, with trial order fully randomized across the session. This randomization strategy mitigates potential order, anticipation, and adaptation effects that could otherwise bias neural responses.

Importantly, the protocol is implemented identically across languages with respect to trial structure, timing, cue presentation, and task instructions. This strict procedural equivalence ensures that any observed differences in neural activation patterns can be attributed primarily to language-specific inner speech processes, rather than to variations in task demands or experimental design.

DATA QUALITY ASSESSMENT

PRE-PROCESSING AND DENOISING

Both EEG datasets that were acquired during inner speech experiments and processed using a common, standardized preprocessing pipeline to ensure consistency, data quality, and suitability for inner speech decoding analyses. Although the datasets differ in experimental paradigms and linguistic content, identical preprocessing procedures were applied to both in order to facilitate comparability across datasets and downstream analyses. Both EEG datasets underwent the following preprocessing steps:

- **Bandpass Filtering (1–145 Hz):** The continuous EEG signals were bandpass filtered between 1 Hz and 145 Hz to remove slow drifts and very high-frequency noise while preserving neural activity relevant to cognitive, linguistic, and sensorimotor processes associated with inner speech. This frequency range captures canonical EEG rhythms as well as higher-frequency components that may contribute to decoding performance.
- **Notch Filtering at 50 Hz and Harmonics:** To mitigate power line interference, notch filters were applied at 50 Hz and its harmonics. This step effectively suppresses electrical noise originating from the power supply without affecting adjacent neural frequency bands.
- **Artifact Subspace Reconstruction (ASR):** Artifact Subspace Reconstruction was employed to identify and attenuate large-amplitude, non-stationary artifacts such as those caused by abrupt movements, electrode pops, or transient environmental disturbances. ASR operates by reconstructing corrupted signal segments based on a clean data subspace, thereby preserving neural information while improving overall signal quality.
- **FORCe-Based Artifact Removal:** Further artifact correction was performed using the FORCe method to remove ocular (e.g., eye blinks and saccades) and muscle-related (EMG) activity. This approach enhances the separation of neural signals from physiological artifacts that commonly contaminate EEG recordings during cognitive tasks, including inner speech.
- **Trial-Based Segmentation:** Following artifact correction, the continuous EEG data were segmented into trials according to the experimental protocol of each dataset. Each trial corresponds to a single inner speech event, aligned to task-specific markers, enabling trial-level analysis, feature extraction, and classification.

QUALITY METRICS

The quality and reliability of the two EEG datasets are supported by a comprehensive set of signal quality metrics computed after preprocessing, demonstrating a high level of data integrity suitable for inner speech decoding. Across participants and sessions, the proportion of retained data following artifact correction exceeded 90% of the original recording duration, indicating that the applied ASR and FORCe procedures effectively removed non-neural artifacts without excessive data loss. The mean channel-wise signal-to-noise ratio (SNR), computed as the ratio between task-related signal power and baseline noise, was consistently above 12 dB, reflecting robust neural signal preservation. Power spectral density (PSD) analyses revealed well-defined canonical EEG rhythms, including alpha (8–12 Hz) and beta (13–30 Hz) bands, with no residual peaks at 50 Hz or its harmonics, confirming the effectiveness of line-noise suppression. Additionally, the percentage of rejected or interpolated channels per recording was below 5%, well within accepted EEG quality standards. Trial-level inspection showed low variance inflation and stable amplitude distributions, with fewer than 7% of trials marked as outliers due to residual artifacts. Together, these metrics indicate high recording fidelity, effective artifact mitigation, and consistent signal characteristics across trials, subjects, and datasets, thereby ensuring the validity, credibility, and suitability of the data for advanced inner speech decoding and machine learning analyses.

Time-domain quality metrics further support signal integrity. The maximum signal gradient, used to detect abrupt amplitude changes indicative of residual artifacts, remained within anticipated values for over 95% of trials, suggesting smooth temporal dynamics consistent with neural activity. The zero-crossing rate (ZCR), an indicator of excessive high-frequency noise or EMG contamination, fell within the expected EEG range, confirming successful suppression of muscle-related artifacts. Statistical characterization of the signals showed kurtosis values predominantly between 2 and 5 across channels and trials, indicating near-Gaussian amplitude distributions and the absence of heavy-tailed artifact-driven distortions. Additionally, fewer than 3% of trials were identified as outliers based on combined amplitude, gradient, and kurtosis criteria. Collectively, these metrics demonstrate stable signal characteristics, low artifact contamination, and high overall recording fidelity, ensuring that both datasets provide valid, credible, and high-quality EEG data for inner speech decoding research.

DATASET DESCRIPTION

DATA STRUCTURE

Both datasets are provided in as .mat file, structured as a nested MATLAB struct with the following hierarchy. The figure below illustrates the hierarchy for the NATO alphabet-based dataset (which also applies for the cross-linguistic one)

```
Data
├── Info
│   ├── Chanlocs                  % EEG channel locations and metadata
│   └── fs                        % Sampling frequency (Hz)
├── S1                            % Subject-specific data (Subject ID ranging between 1 and 20)
│   ├── Day1                      % Data corresponding to the first recording session (i.e. Day1)
│   │   ├── Calibration
│   │   │   ├── eyesClosedRestingEEG   % resting state EEG, eyes closed (channels × samples)
│   │   │   └── eyesOpenRestingEEG     % resting state EEG, eyes open (channels × samples)
│   │   └── Train
│   │       ├── trials                 % trials × channels × samples
│   │       ├── labels                 % class label per trial
│   │       └── time                   % timing information per trial, common across trials
│   └── Day2                      % Data corresponding to the second recording session (i.e. Day2)
│       ├── Calibration
│       │   ├── eyesClosedRestingEEG
│       │   └── eyesOpenRestingEEG
│       └── Train
│           ├── trials
│           ├── labels
│           └── time
├── S2
│   ├── Day1
│   └── Day2
├── ...
└── S20
    ├── Day1
    └── Day2
```

It should be noted that the cross-linguistic dataset is also scattered across different days and therefore this hierarchy is omitted.

DATA ACCESS

The following section provides example code demonstrating how to load and access both EEG datasets using either Python or MATLAB. These examples are intended to illustrate the basic steps required to import the preprocessed, trial-segmented data into commonly used analysis environments and to facilitate rapid inspection and downstream processing. The sample scripts assume that the datasets are accessible into the script’s directory. Users may adapt the provided examples to match their local directory structure, preferred software packages, or specific analysis workflows, such as feature extraction, visualization, or machine learning-based inner speech decoding.

MATLAB Access Example

```
% Load training set

load('train.mat');  % loads variable 'Data'

% Access global metadata

chanlocs = Data.Info.Chanlocs;

fs = Data.Info.fs;

% Example: Access Subject 3, Day 1 training data

S3_D1_trials = Data.S3.Day1.Train.trials;

S3_D1_labels = Data.S3.Day1.Train.labels;

S3_D1_time   = Data.S3.Day1.Train.time;

% Isolate activity corresponding to inner speech

S3_D1_trials_inner_speech = S3_D1_trials(:,:,S3_D1_time>0);

% Access resting-state calibration data for Day1

S3_D1_EyesClosed = Data.S3.Day1.Calibration.eyesClosedRestingEEG;

S3_D1_EyesOpen   = Data.S3.Day1.Calibration.eyesOpenRestingEEG;
```

Python Code

```
import numpy as np
import scipy.io as sio

# Load training set
file = "train.mat"
train = sio.loadmat(file, simplify_cells=True) # loads variable 'Data'

# Access global metadata
fs = train['Data']['Info']['fs']
chanlocs = train['Data']['Info']['Chanlocs']

# Example: Access Subject 3, Day 1 training data
S3_D1_trials = train['Data']['S3']['Day1']['Train']['trials']
S3_D1_labels = train['Data']['S3']['Day1']['Train']['labels']
S3_D1_time = train['Data']['S3']['Day1']['Train']['time']

# Isolate activity corresponding to inner speech

S3_D1_trials_inner_speech = S3_D1_trials[:, :, S3_D1_time > 0]


# Access resting-state calibration data for Day1
S3_D1_EyesClosed = train['Data']['S3']['Day1']['Calibration']['eyesClosedRestingEEG']
S3_D1_EyesOpen   = train['Data']['S3']['Day1']['Calibration']['eyesOpenRestingEEG']
```

REFERENCES

Baayen, R. H., Milin, P., & Ramscar, M. (2016). Frequency in lexical processing. Aphasiology, 30(11), 1174-1220.

Barton, J. J., Hanif, H. M., Eklinder Björnström, L., & Hills, C. (2014). The word-length effect in reading: A review. Cognitive neuropsychology, 31(5-6), 378-412.

Borleffs, E., Maassen, B. A., Lyytinen, H., & Zwarts, F. (2017). Measuring orthographic transparency and morphological-syllabic complexity in alphabetic orthographies: a narrative review. Reading and writing, 30, 1617-1638.

Brysbaert, M., Mandera, P., & Keuleers, E. (2018). The word frequency effect in word processing: An updated review. Current directions in psychological science, 27(1), 45-50.

Dunn, L. M., & Dunn, D. M. (2007). Peabody Picture Vocabulary Test-IV. Circle Pines, MN: American Guidance Service.

Fahimi Hnazaee, M., Khachatryan, E., & Van Hulle, M. M. (2018). Semantic features reveal different networks during word processing: An EEG source localization study. Frontiers in human neuroscience, 12, 503.

Grohmann, K. & E. Leivada (2025). The Cambridge Handbook of Minimalism, Cambridge University Press.

Hauk, O., & Pulvermüller, F. (2004). Effects of word length and frequency on the human event-related potential. Clinical Neurophysiology, 115(5), 1090-1103.

Hendrix, P., & Sun, C. C. (2021). A word or two about nonwords: Frequency, semantic neighborhood density, and orthography-to-semantics consistency effects for nonwords in the lexical decision task. Journal of Experimental Psychology: Learning, Memory, and Cognition, 47(1), 157.

Ibrahim, R. (2006). Morpho-phonemic similarity within and between languages: A factor to be considered in processing Arabic and Hebrew. Reading and Writing, 19, 563-586.

Iniesta, A., Bajo, M. T., Rivera, M., & Paolieri, D. (2023). Transfer effects from language processing to visual attention dynamics: The impact of orthographic transparency. British Journal of Psychology, 114(1), 86-111.

Johnston, R. A., & Barry, C. (2006). Age of acquisition and lexical processing. Visual cognition, 13(7-8), 789-845.

Juhasz, B. J. (2005). Age-of-acquisition effects in word and picture identification. Psychological bulletin, 131(5), 684.

Kaltsa, M., & Papadopoulou, D. (2024). The processing of lexical ambiguity: Evidence from child and adult Greek. Journal of Psycholinguistic Research, 53(1), 16.

Kaltsa, M., Tsimpli, I. M., Marinis, T., & Stavrou, M. (2016). Processing coordinate subject-verb agreement in L1 and L2 Greek. Frontiers in psychology, 7, 648.

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. Behavior research methods, 44, 978-990.

Müller, O., Duñabeitia, J. A., & Carreiras, M. (2010). Orthographic and associative neighborhood density effects: What is shared, what is different?. Psychophysiology, 47(3), 455-466.

Pollatsek, A., & Rayner, K. (2012). Eye movements and lexical access in reading. In Comprehension processes in reading (pp. 143-164). Routledge.

Renfrew, C. E. (2001). Word Finding Vocabulary Test. 4th Edition. Bicester, UK: Speechmark.

Vainio, S., Pajunen, A., & Hyönä, J. (2014). L1 and L2 word recognition in Finnish: Examining L1 effects on L2 processing of morphological complexity and morphophonological transparency. Studies in second language acquisition, 36(1), 133-162.

Wang, J., Liang, L., & Chen, B. (2023). The age of acquisition effect in processing second language words and its relationship with the age of acquisition of the first language. Language and Cognition, 15(2), 355-376.