D3.1 – The BINGO Benchmarking Framework

BINGO

Brain Imagined-Speech Communication

Hellenic Foundation for Research & Innovation

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 14, 27/01/2025
Actual date of delivery:	Month 14, 27/01/2025
Work Package:	WP3 - Benchmarking framework
Task:	T3.1 - Data Collection Protocol
Type:	DEM
Approval Status:	final
Version:	1
Number of pages:	17
Filename:	D3.1_BINGO_BenchmarkingFramework_v1.docx

Executive Summary: This document constitutes the first version of the D3.1, which describes the project's recording protocol. In detail it includes: i) the objectives that drive overall design process, ii) the experimental setup including both the hardware and software that will realize the protocol, iii) the recording protocol by delineating the prompts, the stimuli, and the experiment's timeline, and iv) the information regarding the participants with emphasis on the inclusion and exclusion criteria.

HISTORY

Version	Date	Reason	Revised by
v0.1	01/12/2024	Table of Contents	Fotis P. Kalaganis
v0.2	10/01/2025	Content Input	Maria Kyrou
v0.3	17/01/2025	Revised Input	Fotis P. Kalaganis
v1.0	27/01/2025	Review and Approval	Spiros Nikolopoulos

AUTHOR LIST

Organization	Name	Contact Information
CERTH	Fotis 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
ALS	Amyotrophic Lateral Sclerosis
SE	Scientific Exploration
LSL	Lab Streaming Layer
PTB	PsychToolBox

Contents

1. INTRODUCTION

1.1 BACKGROUNDAND MOTIVATION

INNER SPEECH, THE SILENT ARTICULATION OF WORDS WITHIN ONE'S MIND WITHOUT EXTERNAL VOCALIZATION, PLAYS A CRUCIAL ROLE IN COGNITIVE FUNCTIONS SUCH AS THOUGHT PROCESSING, MEMORY, AND PROBLEM-SOLVING. THE ABILITY TO DECODE INNER SPEECH FROM NEURAL SIGNALS HAS SIGNIFICANT IMPLICATIONS FOR BRAIN-COMPUTER INTERFACES (BCIS), PARTICULARLY FOR INDIVIDUALS WITH SPEECH IMPAIRMENTS DUE TO CONDITIONS SUCH AS AMYOTROPHIC LATERAL SCLEROSIS (ALS), LOCKED-IN SYNDROME, OR STROKE. ELECTROENCEPHALOGRAPHY (EEG) IS A NON-INVASIVE NEUROIMAGING TECHNIQUE THAT RECORDS ELECTRICAL ACTIVITY OF THE BRAIN WITH HIGH TEMPORAL RESOLUTION, MAKING IT A PROMISING MODALITY FOR INNER SPEECH DECODING. HOWEVER, INNER SPEECH PRESENTS UNIQUE CHALLENGES FOR EEG-BASED DECODING, AS IT LACKS THE OVERT MOTOR COMPONENTS ASSOCIATED WITH SPOKEN LANGUAGE, LEADING TO WEAKER AND MORE COMPLEX NEURAL SIGNATURES. RECENT ADVANCES IN MACHINE LEARNING AND SIGNAL PROCESSING HAVE IMPROVED THE FEASIBILITY OF INNER SPEECH DECODING, BUT ACHIEVING HIGH CLASSIFICATION ACCURACY REMAINS AN ONGOING CHALLENGE [SHAH, 2022]. THIS REPORT AIMS TO ESTABLISH A ROBUST EXPERIMENTAL PROTOCOL FOR EEG ACQUISITION SPECIFICALLY TAILORED FOR INNER SPEECH DECODING, ENSURING HIGH-QUALITY DATA COLLECTION AND REPRODUCIBILITY. BY DEFINING CLEAR METHODOLOGIES FOR PARTICIPANT SELECTION, DATA ACQUISITION, AND EXPERIMENTAL DESIGN, THIS DOCUMENT CONTRIBUTES TO THE DEVELOPMENT OF MORE EFFECTIVE BCIS THAT CAN BRIDGE THE GAP BETWEEN THOUGHT AND COMMUNICATION, ULTIMATELY ENHANCING ASSISTIVE TECHNOLOGY FOR INDIVIDUALS WITH SEVERE COMMUNICATION DISORDERS. THE PROTOCOL IN THIS DOCUMENT IS THE OUTCOME OF AN ITERATIVE PROCESS THAT INVOLVES SEVERAL ADJUSTMENTS STEMMING FROM REHEARSAL EXPERIMENTAL RECORDINGS. THESE REHEARSAL EXPERIMENTS WERE CONDUCTED WITH A SMALL GROUP OF PILOT PARTICIPANTS. THESE TRIALS ALLOWED US TO ASSESS TASK CLARITY, PARTICIPANT ENGAGEMENT, AND EEG SIGNAL QUALITY UNDER REAL EXPERIMENTAL CONDITIONS. THE EFFECTIVENESS OF PREPROCESSING TECHNIQUES WAS ALSO EVALUATED. BASED ON THESE FINDINGS, REFINEMENTS WERE MADE TO TASK INSTRUCTIONS, TIMING INTERVALS, AND PREPROCESSING METHODS, ENSURING A ROBUST AND VALIDATED FRAMEWORK FOR THE FULL-SCALE EEG INNER SPEECH DECODING STUDY.

1.2 OBJECTIVES

By delineating the objectives defined in the technical document, BINGO's EEG recording protocol should enable the following Scientific Explorations (SEs):

- **SE1:** Enable word-level inner speech decoding from EEG recordings.
- **SE2:** Create a large data corpus that is suitable for both typical and deep learning approaches.
- **SE3:** Allow uncovering the correspondence between words of the same meaning expressed in different languages (i.e., English and Greek).
- **SE4:** Enable the possibility of creating a scalable BCI by incrementally increasing the lexicon using recordings that take place during the same recording session.

- **SE5:** Enable the possibility of creating a scalable BCI by incrementally increasing the lexicon using recordings that take place during different recording sessions.
- **SE6:** Enable the system’s assessment using recordings that take place during a single session.
- **SE7:** Enable the system’s generalization capabilities by assessing its performance on recordings that take place in different recording sessions.
- **SE8:** Enable the assessment and conceptualization of transfer learning and calibration techniques.

2 EXPERIMENTAL SETUP

2.1 EEG EQUIPMENT AND SPECIFICATIONS

To realize the BINGO recording protocol, the DSI-24 EEG headset will be employed (depicted in Fig.2.1). This solution is a wireless, wearable dry electrode system designed for high-quality brain signal acquisition in various research and clinical applications. Manufactured by Wearable Sensing, the DSI-24 offers 24-channel EEG recording with dry electrodes, eliminating the need for conductive gels and reducing setup time compared to traditional wet electrode systems. The device operates with a sampling rate of 300 Hz, providing sufficient temporal resolution for real-time brain activity monitoring. The headset features an adjustable, flexible design, ensuring a comfortable fit for diverse head shapes and sizes while maintaining good electrode contact. Electrode placement follows the 10-20 international system, covering key cortical regions relevant to cognitive and motor functions. The DSI-24 communicates wirelessly via Bluetooth, enabling mobility and reducing motion constraints, making it ideal for naturalistic and mobile EEG studies. Its built-in noise filtering and real-time impedance monitoring enhance signal quality, ensuring reliable data acquisition even in non-laboratory environments. Given these features, the DSI-24 is particularly well-suited for inner speech decoding research, where minimizing preparation time and maximizing participant comfort are essential for capturing prolonged, natural neural activity associated with silent speech processing.

Figure 2.1. The DSI-24 EEG headset manufactured by Wearable Sensing. The system features 24 channels for recording electrical activity.

2.2 ELECTRODE PLACEMENT AND CONFIGURATION

For effective inner speech decoding using EEG, careful electrode placement is crucial to capture neural activity associated with silent articulation, phonological processing, and language-related cognition. The DSI-24 EEG headset follows the 10-20 (depicted in Fig2.2) international system, which ensures standardized electrode positioning across participants while targeting key brain regions implicated in inner speech processing. Given that inner speech involves language comprehension, subvocal articulation, and working memory, electrode coverage should prioritize frontal, temporal, and parietal areas [Fan, 2016]. Specifically, electrodes over the left inferior frontal gyrus (Broca's area, e.g., F7, F3) are critical for silent speech production and phonological encoding, while electrodes near the superior temporal gyrus (Wernicke's area, e.g., T7, TP7) are essential for processing linguistic meaning and auditory imagery. Additionally, parietal electrodes (e.g., P3, P4) capture higher-level cognitive functions, such as working memory and attention, which are often engaged during inner speech tasks. Central electrodes (C3, C4, Cz) are included to monitor any residual motor activity, such as involuntary articulatory movements, which could contribute to decoding performance. Since inner speech generates lower-amplitude EEG signals compared to overt speech, maintaining optimal electrode contact and minimizing impedance is critical. The DSI-24's dry electrodes provide real-time impedance monitoring to ensure reliable signal acquisition without conductive gel. Furthermore, reference and ground electrodes are positioned at ear mastoids (A1, A2), respectively, to balance signal stability and reduce artifacts. By strategically configuring electrodes to emphasize linguistic and cognitive processing areas, this setup maximizes the chances of accurately distinguishing inner speech patterns from background neural activity, improving the feasibility of robust decoding models.

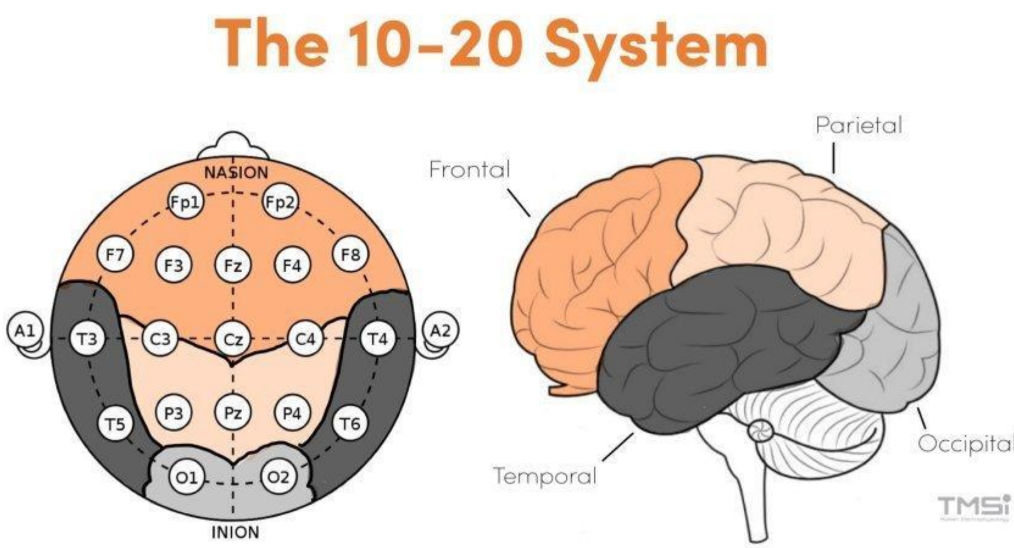

Figure 2.2. The EEG electrode placement according to the international 10-20 system. Image source: https://www.tmsi.artinis.com/blog/the-10-20-system-for-eeg

2.2 PROTOCOL DESIGN SOFTWARE

In to order to design BINGO's EEG recording protocol, a combination of Psychotoolbox (PTB; or its equivalent implementation PsychoPy Toolbox in python) [Kleine, 2007; Peirce, 2022] and Lab Streaming Layer (LSL) [Kothe, 2024] will be employed. The PTB is a widely used toolbox designed for precise stimulus presentation and response timing in neuroscience and psychology experiments. It provides highly accurate visual and auditory stimulus delivery, making it ideal for presenting the prompts during the inner speech task while ensuring precise control over timing parameters. For EEG data synchronization, the experiment will integrate Lab Streaming Layer (LSL), a real-time data transmission framework that ensures accurate timestamping and synchronization of multiple data streams. LSL will be used to send event markers from Psychtoolbox to the EEG recording system, allowing for precise alignment of neural signals with stimulus onset. This synchronization is crucial for inner speech decoding, as it ensures that EEG data segments correspond accurately to the cognitive processing of each presented word. By combining Psychtoolbox for stimulus presentation and LSL for real-time event synchronization, the experiment will achieve millisecond-level precision in data acquisition, reducing timing jitter and improving the reliability of inner speech EEG analysis.

2.3 PREPROCESSING

To ensure high-quality data for inner speech decoding, EEG signals will undergo a series of preprocessing steps aimed at removing artifacts, enhancing signal clarity, and preparing the data for analysis. The preprocessing pipeline will begin with bandpass filtering (typically 0.5–45 Hz) to remove low-frequency drift and high-frequency noise, including muscle and electrical interference. Next, artifact rejection and correction will be performed to eliminate contamination from eye blinks, muscle activity, and motion artifacts. Wavelet-based Independent Component Analysis (wICA) [Inuso, 2007] will be applied to identify and remove time-frequency components associated with ocular and muscular artifacts. Bad channels, identified through high impedance values or excessive noise, will be either interpolated or excluded from further analysis. Finally, data will be segmented into epochs, time-locked to the stimulus onset markers provided by the Lab Streaming Layer (LSL), and annotated accordingly in order to ensure precise alignment between EEG signals and meaningful inner speech processing events.

3 EXPERIMENTAL PROTOCOL

3.1 PRE-EXPERIMENTAL PREPARATIONS

Before EEG data acquisition begins, several essential pre-experimental preparations will be conducted to ensure high-quality signal recording, participant comfort, and the reliability of inner speech decoding. These preparations will include participant briefing, physical preparation, environmental setup, and device calibration.

- **Participant Briefing and Screening:** Participants will be informed about the experiment's objectives, procedures, and expected duration. A consent form will be obtained, and participants will be screened for eligibility to ensure they meet inclusion criteria (e.g., no neurological disorders, normal or corrected vision, and fluency in the target languages). Participants will also be asked to refrain from consuming caffeine or other stimulants that may affect neural activity.
- **Physical and Physiological Preparation:** Since EEG signals are susceptible to motion artifacts and muscle activity, participants will be advised to remain as still as possible during the experiment. They will be asked to remove metallic accessories, glasses (if not needed), or hair products that could interfere with electrode contact. Long hair may need to be tied back to ensure proper electrode placement.
- **Environmental Setup:** The experiment will be conducted in a quiet, and low-light place to minimize external noise and distractions. Participants will be seated in a comfortable, upright position to maintain a relaxed state throughout the experiment.
- **DSI-24 EEG Device Calibration:** The DSI-24 EEG headset will be adjusted to fit securely on the participant's head while ensuring even pressure on all electrodes. Real-time impedance monitoring will be performed to verify optimal electrode contact, and adjustments will be made if any electrodes show high impedance. A baseline recording will be taken with the participant at rest (eyes open and closed) to assess signal quality and check for excessive noise, muscle artifacts, or blinking-related activity.

By meticulously preparing participants, the environment, and the EEG system, these pre-experimental steps will help optimize signal acquisition, reduce artifacts, and ensure the reliability of the inner speech decoding study.

3.2 TASK DESIGN AND STIMULI

The inner speech task will be designed using the NATO phonetic alphabet (shown in Fig 3.1.) as a structured linguistic framework to elicit neural responses associated with silent speech production. This approach will ensure a controlled and standardized set of stimuli with a diverse phonological representation while allowing for inter-language explorations (by selecting particular words with meaning when translated to also be articulated covertly in Greek). The NATO alphabet constitutes an exceptional opportunity for serving as the basis of our approach given that it enables inner speech decoding at a word level (**SE1**) but also may serve as a generic alphabet where typing in a letter-by-letter setting can be realized.

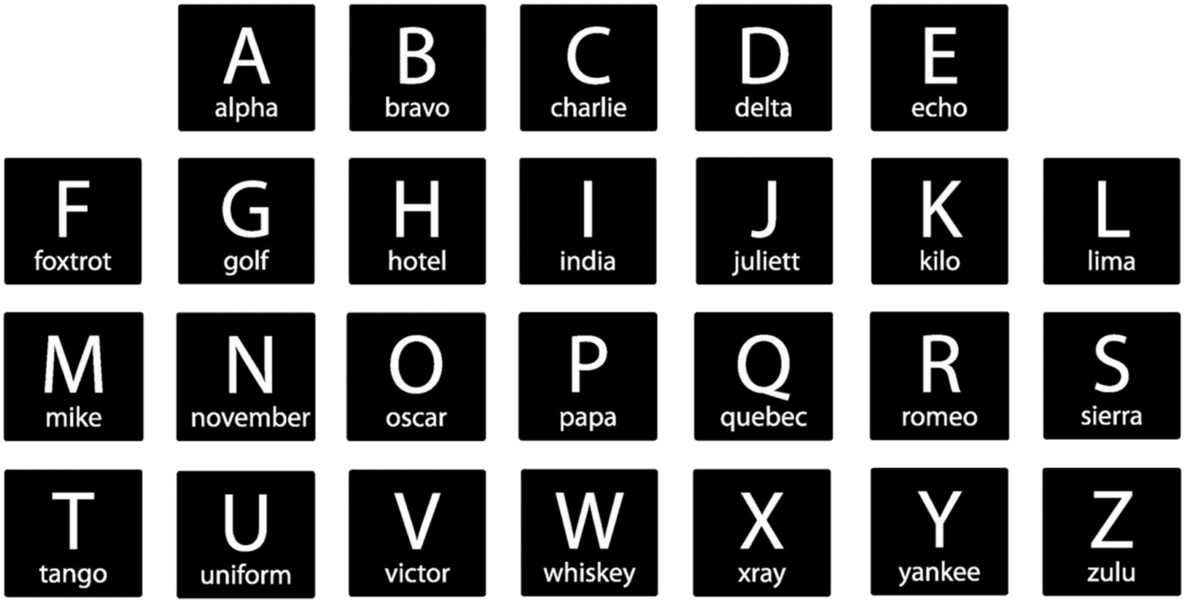

Figure 3.1. The NATO phonetic alphabet where each letter is associated with a phonetically distinctive word.

Prior to beginning the recording session, all subjects will be able to hear the current pronunciation of the NATO words so as to minimize the variability across phonological representations. The task will incorporate visual cues to prompt participants to internally articulate specific phonetic words without overt speech or subvocalization. The selection of visual cues only during the recording process minimizes the contamination of brain signals stemming from brain areas highly involved in inner speech (e.g. Broca's area). Particular words (i.e., those having a concise meaning in Greek such as "Hotel" -> "Ξενοδοχείο") of the NATO alphabet will be selected and the participants will be prompted to also covertly articulate these words in Greek so as to support **SE3**.

The foreseen experimental procedure targets at approximately 30 participants with 50 trials for each word in English and 10 words for each Greek word per participant (**SE2**). In order to enable **SE4-8**, for each participant the EEG recording procedure will be conducted in three different sessions (i.e., days) as shown in the following table.

DAY 1	DAY 2	DAY 3
Words A-M (35 trials per word per participant)	Words N-Z (35 trials per word per participant)	Words A-Z (15 trials per word per participant)
Greek words corresponding to A-M words (10 trials per word per participant)	Greek words corresponding to N-Z words (10 trials per word per participant)	

3.3 TRIAL STRUCTURE AND TIMING

BINGO's 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 2 minutes:

- **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.
Each trial follows a strict timeline with distinct time points and stimulus presentations:

1. **t = 0s:** A randomly selected NATO phonetic alphabet word (e.g., "Charlie") is presented on the screen. The participant will be instructed to keep the depicted word in their mind without overt speech or subvocalization.
2. **t = 2s**: The word disappears from the screen, removing external visual input.
3. **t = 3s**: 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 = 5s**: The fixation cross disappears, marking the end of the trial. Participants remain in a resting state until the next trial begins.

Each participant will complete several trials per word (as indicated in the Table above), with the words appearing in a randomized order to prevent block effects (e.g., anticipation or adaptation effects).

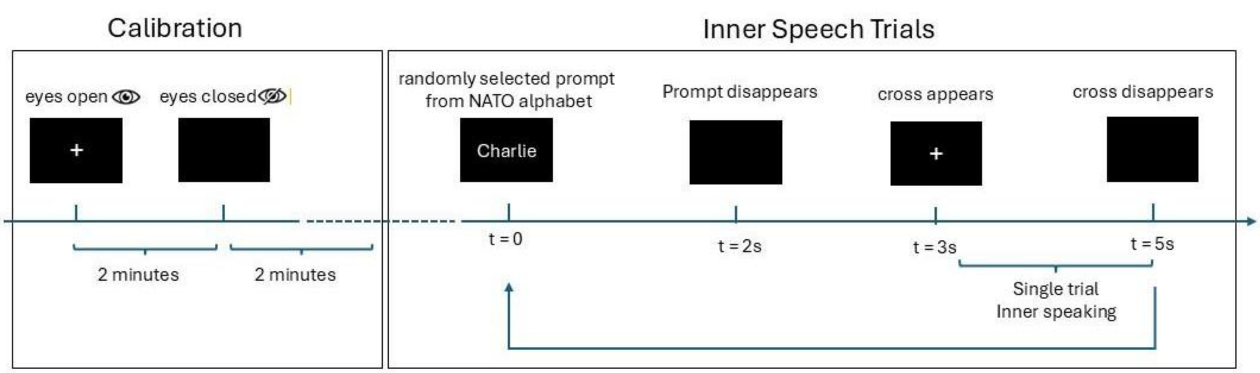

Figure 3.2. The experimental protocol timeline for EEG acquisition. The protocol consists of two phases: Calibration and Inner Speech Trials, both designed to optimize EEG data collection for subsequent analysis.

4 PARTICIPANTS

4.1 INCLUSION AND EXCLUSION CRITERIA

To ensure the reliability of EEG data and the validity of inner speech decoding, specific inclusion and exclusion criteria will be applied when selecting participants for the study.

Inclusion Criteria - Participants must meet the following conditions to be eligible for the experiment:

- **Age range:** Typically, between 18 and 50 years old, to ensure a homogeneous sample with stable cognitive and neural processing.
- **Right-handed dominance:** Since language processing is often lateralized to the left hemisphere in right-handed individuals, right-handed participants will be preferred for consistency.
- **Fluency in English and Greek:** Participants must have proficiency in both languages to allow for inter-language inner speech exploration using NATO phonetic alphabet words.
- **Normal or corrected-to-normal vision:** Since the task involves visual stimuli, participants must be able to clearly perceive the presented words.
- **No history of neurological or psychiatric disorders:** To minimize confounding factors related to atypical brain activity, participants must report no history of conditions such as epilepsy, stroke, traumatic brain injury, or schizophrenia.
- **No current use of psychoactive medication:** Participants should not be taking medications that could alter cognitive or neural processing, such as antidepressants, antipsychotics, or stimulants.

Exclusion Criteria - Participants will be excluded if they meet any of the following conditions:

- **Presence of metallic implants or pacemakers:** Since EEG is sensitive to electromagnetic interference, individuals with such implants will not be included.
- **Significant hair or scalp conditions:** Excessive hair volume, dreadlocks, or scalp conditions that interfere with electrode contact may lead to poor signal quality and will be grounds for exclusion.
- **Excessive blinking or involuntary movements:** Since EEG recordings are sensitive to artifacts, participants with conditions such as tics, tremors, or excessive eye blinking may be excluded.
- **Subvocalization tendencies:** If a participant habitually moves their lips or tongue during silent reading or inner speech, they may introduce unintended motor artifacts, leading to unreliable data.
- **Severe anxiety or discomfort with EEG procedures:** If a participant expresses distress or discomfort with wearing the EEG headset, they will be excluded to ensure a stress-free recording environment.

By implementing these inclusion and exclusion criteria, the study will ensure a homogeneous participant pool, minimize noise in EEG signals, and enhance the robustness of inner speech decoding analyses.

4.2 ETHICAL CONSIDERATIONS, GENDER BIASES AND CONSENT

This study will be conducted in strict accordance with ethical guidelines for human research, ensuring the safety, privacy, and voluntary participation of all individuals. Before participation, each individual will receive a detailed informed consent form, outlining the study's objectives, procedures, potential risks, and the right to withdraw at any time without consequences. Special attention will be given to gender inclusivity, ensuring that recruitment strategies promote a balanced representation of male, female, and non-binary individuals to prevent gender biases in the dataset and enhance the generalizability of inner speech decoding models. Additionally, data confidentiality will be strictly maintained, with all EEG recordings and personal information stored securely and anonymized for analysis. To mitigate potential stress or discomfort, participants will be allowed to take breaks and will be monitored for any signs of fatigue during EEG recording. By adhering to these ethical principles, the study will ensure fairness, inclusivity, and respect for all participants while maintaining the highest standards of scientific integrity. Detailed information about the ethical considerations and approval by the Organization's Ethics Committee can be found in "D1.3 – Ethics Approval".

REFERENCES

Fan, L., Li, H., Zhuo, J., Zhang, Y., Wang, J., Chen, L., ... & Jiang, T. (2016). The human brainnetome atlas: a new brain atlas based on connectional architecture. *Cerebral cortex*, *26*(8), 3508-3526.

Inuso, G., La Foresta, F., Mammone, N., & Morabito, F. C. (2007, August). Wavelet-ICA methodology for efficient artifact removal from Electroencephalographic recordings. In *2007 international joint conference on neural networks* (pp. 1524-1529). IEEE.

Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3?.

Kothe, C., Shirazi, S. Y., Stenner, T., Medine, D., Boulay, C., Grivich, M. I., ... & Makeig, S. (2024). The lab streaming layer for synchronized multimodal recording. *BioRxiv*, 2024-02.

Peirce, J., MacAskill, M., & Hirst, R. (2022). Building experiments in PsychoPy.

Shah, U., Alzubaidi, M., Mohsen, F., Abd-Alrazaq, A., Alam, T., & Househ, M. (2022). The role of artificial intelligence in decoding speech from EEG signals: a scoping review. *Sensors*, *22*(18), 6975.