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

D2.2 (v2) – Report on the developed imagined speech decoding approaches

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, 22/10/2025
Work Package:	WP2 - Neuroengineering for EEG-based Imagined Speech Decoding
Task:	T2.4 Interconnection between languages
Type:	DEC
Approval Status:	final
Version:	V2.0
Number of pages:	20
Filename:	D2.2_Report_Developed_ImaginedSpeech_Decoding_Approaches_v2.docx

Abstract: This deliverable investigates cross-linguistic neural phenomena associated with inner speech using electroencephalography (EEG), with the aim of identifying language-invariant and language-dependent brain representations. Two complementary methodological approaches are explored: a Riemannian geometry framework that models EEG activity through covariance-based representations on symmetric positive-definite manifolds, and a spectro-temporal analysis that examines frequency- and time-resolved neural dynamics. Due to the absence of suitable publicly available EEG datasets for systematic cross-linguistic inner speech research, the proposed methodologies were applied to the BINGO cross-linguistic inner speech dataset, which was specifically collected for this purpose. The need to await the progression of the data collection process resulted in a two-month delay in the execution of this work. The analyses presented provide insight into multilingual inner speech processing while establishing a methodological foundation for future cross-language EEG studies.

HISTORY

Version	Date	Reason	Revised by
V1.1	11/09/2025	Table of Contents	Kostas Georgiadis
V1.2	13/10/2025	Initial Draft	Kostas Georgiadis
V2.0	22/10/2025	Final	Spiros Nikolopoulos

AUTHOR LIST

Organization	Name	Contact Information
CERTH	Kostas Georgiadis	kostas.georgiadis@iti.gr
CERTH	Fotis P. Kalaganis	fkalaganis@iti.gr
AUTH	Nikolaos Laskaris	laskaris@csd.auth.gr
CERTH	Spiros Nikolopoulos	nikolopo@iti.gr
CERTH	Ioannis Kompatsiaris	ikom@iti.gr

ABBREVIATIONS AND ACRONYMS

EEG	ElectroEncephaloGram
SPD	Symmetric Positive Definite
CCV	Cross-Covariance
SCM	Spatial Covariance Matrix
AIRM	Affine-Invariant Riemannian Metric
MDS	Multi-Dimensional Scaling
CWT	Continuous Wavelet Transform

Contents

INTRODUCTION

Understanding how the brain represents inner speech across different languages is a fundamental question in cognitive neuroscience, with direct implications for multilingual brain–computer interfaces and language-agnostic decoding systems. Electroencephalography (EEG) provides a non-invasive means to investigate the temporal and spectral dynamics underlying inner speech, yet the study of cross-linguistic inner speech phenomena remains largely unexplored. This is due, in part, to the complexity of inner speech itself and to the scarcity of suitable datasets that combine multilingual paradigms with high-quality EEG recordings.

The objective of this deliverable is to investigate cross-linguistic neural phenomena associated with inner speech using EEG, with a particular focus on identifying language-invariant and language-dependent representations. To this end, two complementary analytical approaches are explored. The first leverages Riemannian geometry, which models EEG signals through covariance representations and exploits the intrinsic geometry of symmetric positive-definite matrices to enable robust cross-condition and cross-language comparisons. The second approach employs spectro-temporal analysis, examining how frequency-specific and time-resolved neural dynamics contribute to inner speech representations across languages. Together, these methods provide a comprehensive framework for characterizing inner speech from both geometric and signal-processing perspectives.

At the outset of this work, no suitable publicly available EEG datasets were identified that support systematic investigation of cross-linguistic inner speech, particularly with controlled bilingual or multilingual experimental designs. As a result, the proposed methodologies were applied to the BINGO cross-linguistic inner speech dataset (detailed in D3.1 v2), which was specifically collected to address this gap. The necessity of waiting for the completion of the data collection and validation process introduced an unavoidable delay of approximately two months in the execution timeline of this deliverable. Nevertheless, the availability of the BINGO dataset enabled a rigorous and methodologically sound investigation, ensuring that the analyses presented herein are grounded in high-quality, purpose-built EEG data.

DATASET

The EEG acquisition protocol (for a detailed description of the selected lexical items, data pre-processing pipeline and the overall rationale of this experiment is presented in BINGO's deliverable D3.1-v2) is organized into two distinct stages: an initial preparatory stage, referred to as the calibration phase, followed by the primary experimental stage, referred to as the inner speech trials.

The calibration phase serves to establish an individual baseline of neural activity for each participant and is a critical step prior to the main task. It includes two conditions, each with a duration of one minute. In the eyes-open condition, participants are instructed to remain still while keeping their eyes open and focusing on a centrally displayed fixation cross ("+"). This condition captures baseline EEG activity during alert rest and visual fixation and is particularly informative for assessing alpha suppression in posterior regions. In the eyes-closed condition, participants are asked to close their eyes and relax, allowing for the observation of resting-state alpha oscillations, which typically increase in the absence of visual input.

Together, these calibration conditions facilitate the identification and characterization of common artifacts, such as ocular and muscle-related activity, as well as background noise, thereby supporting more reliable analysis during the experimental phase.

After calibration, participants proceed to the inner speech trials. Each trial follows a predefined sequence of visual cues and silent word articulation designed to elicit neural activity associated with covert speech production. To minimize order and block-related effects, all lexical items are presented in a randomized order, with a unique sequence generated for each participant. More specifically, each trial (we note that trials take place after eyes open/closed calibration period) 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" or "όχημα"). 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.

RIEMANNIAN GEOMETRY

MOTIVATION

Investigating cross-language neural phenomena using EEG through the lens of Riemannian Geometry is motivated by both the complexity of multilingual brain representations and the limitations of traditional Euclidean analysis methods. EEG signals are inherently multivariate, non-stationary, and highly correlated across channels, properties that are naturally captured by covariance matrices rather than raw time-domain features. These covariance matrices reside on a Riemannian manifold of symmetric positive-definite matrices, where distances and averages are more meaningfully defined using Riemannian metrics. By operating directly on this manifold, Riemannian Geometry–based methods preserve the intrinsic structure of EEG data, enabling more robust comparisons across subjects, sessions, and, critically, languages. In the context of cross-language inner speech, such approaches offer a principled framework for disentangling language-specific neural patterns from language-invariant representations, facilitating transfer learning and cross-lingual decoding. This is particularly important for bilingual or multilingual scenarios, where neural variability induced by different phonological and semantic systems can obscure shared cognitive mechanisms. Leveraging Riemannian Geometry thus provides a mathematically grounded and empirically powerful means to study how the brain encodes inner speech across languages, advancing both neuroscientific understanding and the development of language-agnostic EEG-based brain–computer interfaces.

METHODOLOGICAL APPROACH

Let $\mathbf{X}_i \in \mathrm{R}^{E \times T}, i = 1,\ldots,n$ be a multichannel EEG segment, where E denotes the number of electrodes, T the number of time samples and n the number of available segments (or trials). Each segment (assuming zero mean signals) can also be described by the corresponding spatial covariance matrix $C_i = \frac{1}{T-1} X_i X_i^{\top} \in R^{E \times E}$, where $(\cdot)^{\top}$ denotes the transpose operator. By definition and under a sufficiently large T value to guarantee a full rank covariance matrix, spatial covariance matrices are Symmetric Positive Definite (SPD) that lie on a Riemannian manifold instead of a vector space (e.g. scalar multiplication does not hold on the SPD manifold). In the field of differential geometry, a Riemannian manifold is a real, smooth manifold endowed with an inner product on the tangent space of each point that varies smoothly from point to point.

When treating EEG data, the manifold of SPD matrices denoted by $Sym^{+}_{E} = \{\mathbf{C} \in \mathrm{R}^{E \times E} : \mathbf{x}^{\top}\mathbf{C}\mathbf{x} > 0$, for all non-zero $\mathbf{x} \in \mathrm{R}^{E}\}$, is typically studied when it is equipped with the AIRM [Conney, 2020],

$$\langle \mathbf{A}, \mathbf{B} \rangle_{\mathbf{P}} \triangleq Trace(\mathbf{P}^{-1}\mathbf{A}\mathbf{P}^{-1}\mathbf{B}) \tag{1}$$

for $\mathbf{P} \in Sym^{+}_{E}$ and $A, B \in T_E^{+}(P)$, where $T_E^{+}(P)$ denotes the tangent space of Sym^{+}_{E} at $\mathbf{P}$. Then, the following geodesic distance is induced

$$\delta(C_i, C_j) = \|logm\left(C_i^{-1/2} C_j C_i^{-1/2}\right)\|_F = \sqrt{\sum_{q=1}^{E} \quad log^2 \lambda_q} \tag{2}$$

where $logm(\cdot)$ denotes the matrix logarithm operator and λ_q the eigenvalues of the matrix $\mathbf{C}_i^{-1/2}\mathbf{C}_j\mathbf{C}_i^{-1/2}$ or equivalently of the matrix $\mathbf{C}_i^{-1}\mathbf{C}_j$. We note that these two matrices are similar (i.e., hold the same eigenvalues) and that the indices i and j can be permuted. Among the other useful properties that are discussed thoroughly in [Pennec, 2006], δ is congruent invariant for non-singular matrices **W**, i.e. $\delta(\mathbf{W}\mathbf{C}_i\mathbf{W}^\top, \mathbf{W}\mathbf{C}_j\mathbf{W}^\top) = \delta(\mathbf{C}_i, \mathbf{C}_j)$. This is an important property in EEG signal processing since it provides equivalence between the sensor and the source space [Congedo, 2017]. According to the prevailing EEG model, the recorded activity is well approximated by a linear mixture of source signals. Hence, $\mathbf{X}_i = \mathbf{M}\mathbf{S}_i$ with **M** denoting the mixing matrix and $\mathbf{S}_i$ the source signals. Then, by substituting the observed signal with the equivalent mixing of sources, one may obtain the following covariance matrix, $C_i = \frac{1}{T-1} M S_i S_i^\top M_i^\top$. Therefore, the mixing procedure in the time domain results in a congruent transformation in the corresponding covariance matrices. It becomes obvious that since δ is invariant to such transformations, the two spaces are considered equivalent. In a strict mathematical sense this is partially true (e.g., for certain forms of **M**) and this topic is thoroughly discussed in [Congedo, 2017]. Hereafter, the terms "AIRM-induced geodesic distance" or simply "geodesic distance" will be used interchangeably and will refer to Equation 2.

CLUSTER EVALUATION – SILHOUETTE SCORE

The silhouette score [Shahapure, 2020] is a quantitative metric used to evaluate the quality of clustering by measuring how well each data point fits within its assigned cluster compared to other clusters. For a given sample, the silhouette value reflects the difference between the average distance to points within the same cluster (intra-cluster similarity) and the average distance to points in the nearest neighboring cluster (inter-cluster separation), normalized to lie between −1 and 1. Values close to 1 indicate that a sample is well matched to its own cluster and clearly separated from others, values near 0 suggest overlapping clusters, and negative values imply potential misclassification. When averaged across samples or clusters, the silhouette score provides an intuitive summary of cluster compactness and separability, making it particularly useful for assessing structure in neural data embeddings.

RESULTS

EEG-BASED INVESTIGATION OF WORDS' MEANING

In this section, we present the results of our analysis examining whether words with identical semantic meaning, internally articulated in Greek and English, elicit similar neural patterns as captured by EEG. Neural representations were modeled using covariance matrices and compared using the Affine Invariant Riemannian Metric (AIRM), which enables meaningful distance computations on the manifold of symmetric positive-definite matrices while preserving their intrinsic geometric structure. Pairwise Riemannian distances were subsequently embedded into a low-dimensional space using Multidimensional Scaling (MDS) to facilitate qualitative visualization of the relationships between inner speech representations across languages. The resulting embeddings allow for intuitive inspection of whether semantically equivalent words cluster together irrespective of language. To quantitatively assess the degree of clustering and separability, the silhouette score was employed, providing a measure of intra-cluster compactness relative to inter-cluster separation. Together, the MDS visualizations and silhouette-based evaluations offer complementary qualitative and quantitative evidence regarding the presence of shared, language-invariant neural patterns associated with inner speech of semantically identical words in Greek and English.

All investigations presented below are conducted systematically across the common EEG frequency bands, namely delta, theta, alpha, beta, gamma, as well as the wideband range (1–145 Hz). For each band, the EEG signals are band-limited prior to covariance estimation, allowing the analysis to capture frequency-specific neural dynamics associated with inner speech processing. This multi-band approach enables examination of whether cross-language similarities or separability in neural representations are driven by specific oscillatory components or are distributed across the broader spectral content of the signal. By evaluating the results consistently across canonical frequency bands and the wideband representation, we aim to provide a comprehensive and nuanced characterization of language-invariant and language-dependent neural patterns in inner speech.

Table 3.1: Average Silhouette Score across all participants for each pair of the lexical items (i.e., ‘Antenna’/‘Κεραία) across the common EEG frequency bands.

Word-pair	Delta	Theta	Alpha	Beta	Gamma	WideBand
'Antenna'	-0.29	-0.39	-0.19	-0.31	-0.65	-0.47
'Apple'	0.05	-0.25	-0.25	-0.15	-0.09	-0.44
'Arrow'	-0.24	-0.27	-0.24	-0.13	-0.46	-0.46
'Belt'	-0.15	-0.24	-0.22	-0.19	-0.49	-0.45
'Button'	-0.19	-0.26	-0.14	-0.16	-0.37	-0.44
'Candle'	-0.33	-0.21	-0.27	-0.34	-0.58	-0.25
'Compass'	-0.31	-0.19	-0.29	-0.32	-0.50	-0.18
'Dice'	-0.25	-0.38	-0.28	-0.061	-0.29	-0.35
'Feather'	-0.20	-0.22	-0.23	-0.26	-0.51	-0.38
'Guitar'	-0.30	-0.22	-0.22	-0.29	-0.51	-0.16
'Pencil'	-0.21	-0.29	-0.15	-0.35	-0.53	-0.51
'Plate'	-0.29	-0.26	-0.24	-0.26	-0.46	-0.30
'Saddle'	-0.32	-0.14	-0.34	-0.42	-0.08	-0.58
'Vehicle'	-0.15	-0.25	0.03	-0.25	-0.59	-0.31
'Wheel'	-0.26	-0.19	-0.31	-0.22	-0.46	-0.35

Table 1 reports the average silhouette scores, aggregated across all participants, for each pair of semantically identical lexical items internally articulated in English and Greek (e.g., *“Antenna”/“Κεραία”*) across the common EEG frequency bands. Overall, the silhouette values are predominantly negative across words and frequency ranges, indicating substantial overlap between the neural representations of English and Greek inner speech not only for the same semantic concepts but also for discrete ones. This pattern suggests that, for most lexical items, trials corresponding to the two languages do not form well-separated clusters in the Riemannian space but instead exhibit considerable similarity across several semantic concepts. Among the examined bands, the gamma and wideband ranges tend to yield the most negative scores, pointing to increased intermixing of language-specific representations at higher frequencies and in broadband activity. In contrast, slightly higher (though still near-zero or negative) silhouette values are observed in the delta, theta, and alpha bands for some words (e.g., *“Apple”* and *“Vehicle”* in the alpha band), hinting at limited frequency-dependent variation in cross-language settings. Taken together, these results provide quantitative evidence that inner speech of semantically equivalent words in Greek and English elicits largely overlapping neural patterns across several concepts, with only modest and inconsistent differentiation across frequency bands.

The aforementioned become evident by the following indicative image which visualizes the Riemannian Space of the single trials for a random participant (i.e., S06) using wideband EEG signals.

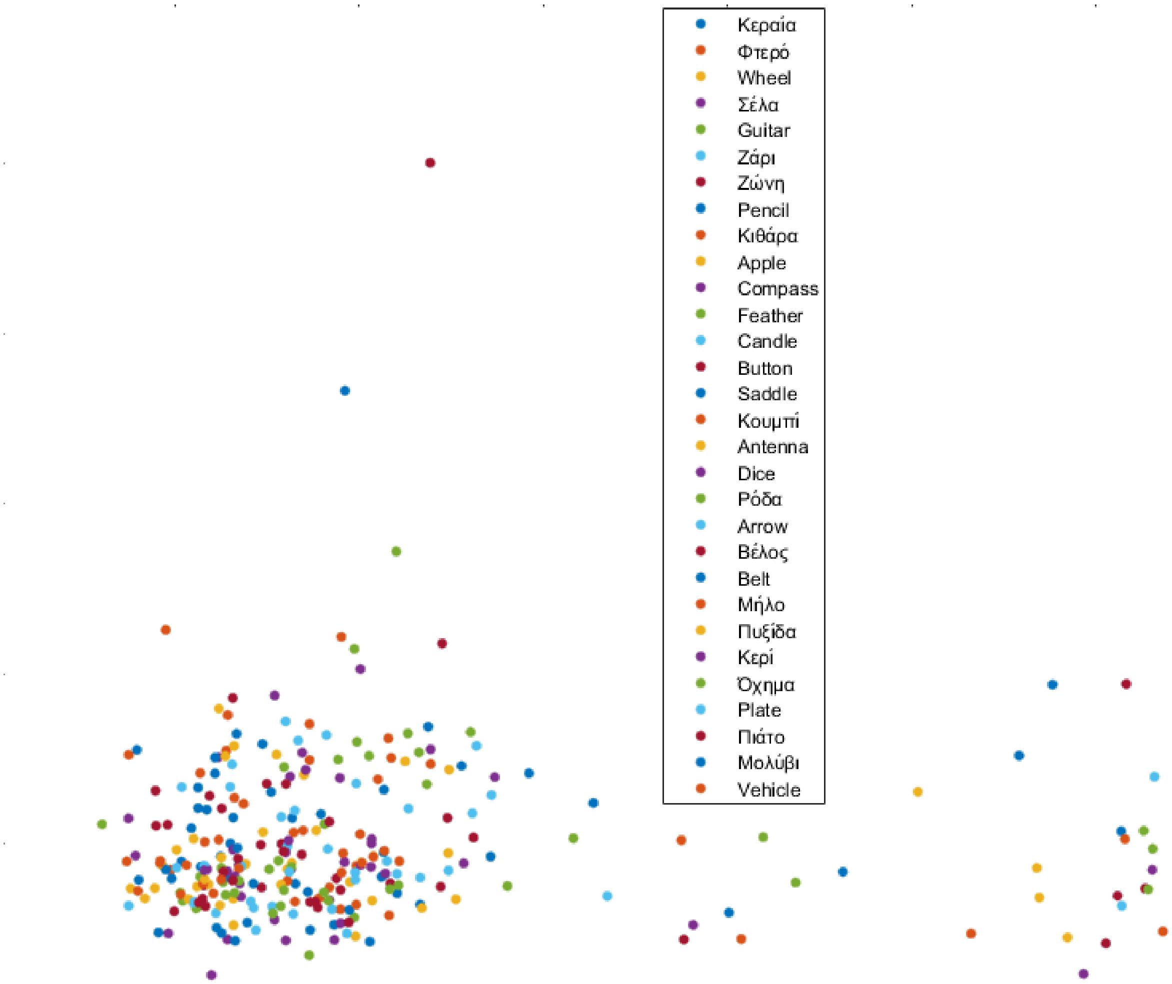

Figure 3.1: 2-dimensional representation for the wide-band single trial responses of S06 using MDS under the Affine Invariant Riemannian Metric.

EEG-BASED INVESTIGATION OF LANGUAGE

In this section, we present results assessing whether internally articulated Greek and English words form distinct neural representations that can be reliably separated using EEG. As in the previous analysis, trial-wise covariance matrices were compared using the Affine Invariant Riemannian Metric (AIRM), enabling robust distance estimation on the Riemannian manifold of symmetric positive-definite matrices. These pairwise distances were projected into a low-dimensional space via Multidimensional Scaling (MDS) to provide a visual assessment of language-specific structure in the data. The MDS embeddings were examined to determine whether trials corresponding to Greek and English inner speech occupy separable regions of the representational space. To complement the qualitative visualization, silhouette scores were computed using language labels to quantitatively evaluate cluster separability, with higher scores indicating stronger within-language cohesion and clearer between-language separation. This combined geometric, visual, and quantitative analysis provides insight into the extent to which inner speech in Greek and English gives rise to distinguishable neural patterns, shedding light on language-dependent encoding mechanisms in the brain.

As in the pervious subsection, all analyses presented are performed across standard EEG frequency ranges, including the delta, theta, alpha, beta, and gamma bands, in addition to a broadband representation spanning 1–145 Hz. Prior to covariance computation, the EEG data are filtered to isolate each frequency band, enabling the investigation of neural activity specific to distinct oscillatory processes involved in inner speech. This band-wise analysis allows us to determine whether observed cross-language similarities or distinctions in neural patterns emerge from particular frequency components or reflect more global spectral characteristics.

Table 3.2: Average Silhouette Score across all participants for Greek and English words (i.e., each lexical item is assigned to its language cluster) across the common EEG frequency bands.

Freq. Band	**English**	**Greek**
Delta	0.08	-0.08
Theta	0.11	-0.11
Alpha	0.004	-0.007
Beta	0.06	-0.06
Gamma	0.13	-0.13
Wide Band (1-145)	-0.03	0.03

Table 2 summarizes the average silhouette scores across all participants when lexical items are grouped according to language, with English and Greek words assigned to separate clusters, across the common EEG frequency bands. In contrast to the word-pair analysis, the results reveal a modest but consistent tendency toward language-based separability in several frequency ranges. Positive silhouette scores for the English cluster in the delta, theta, beta, and gamma bands (ranging from 0.06 to 0.13) indicate mild within-language cohesion and some degree of separation from Greek words, while the corresponding negative values for the Greek cluster reflect the complementary nature of the silhouette metric in a two-cluster setting. The gamma band exhibits the highest absolute silhouette magnitude, suggesting that higher-frequency activity may contribute more strongly to language-specific differentiation during inner speech. Conversely, the alpha band shows near-zero scores for both languages, implying substantial overlap and minimal language-dependent structure in this frequency range. In the wideband condition, silhouette values are close to zero for both clusters, indicating limited overall separability when broadband information is considered. Overall, these findings suggest that while inner speech representations in Greek and English are largely overlapping, subtle language-related distinctions emerge in specific frequency bands, particularly at higher frequencies.

In the image bellow we provide an indicative image which visualizes the Riemannian Space of the single trials for a random participant (i.e., S06) using EEG signals filtered in gamma frequency band (signals grouped according to language).

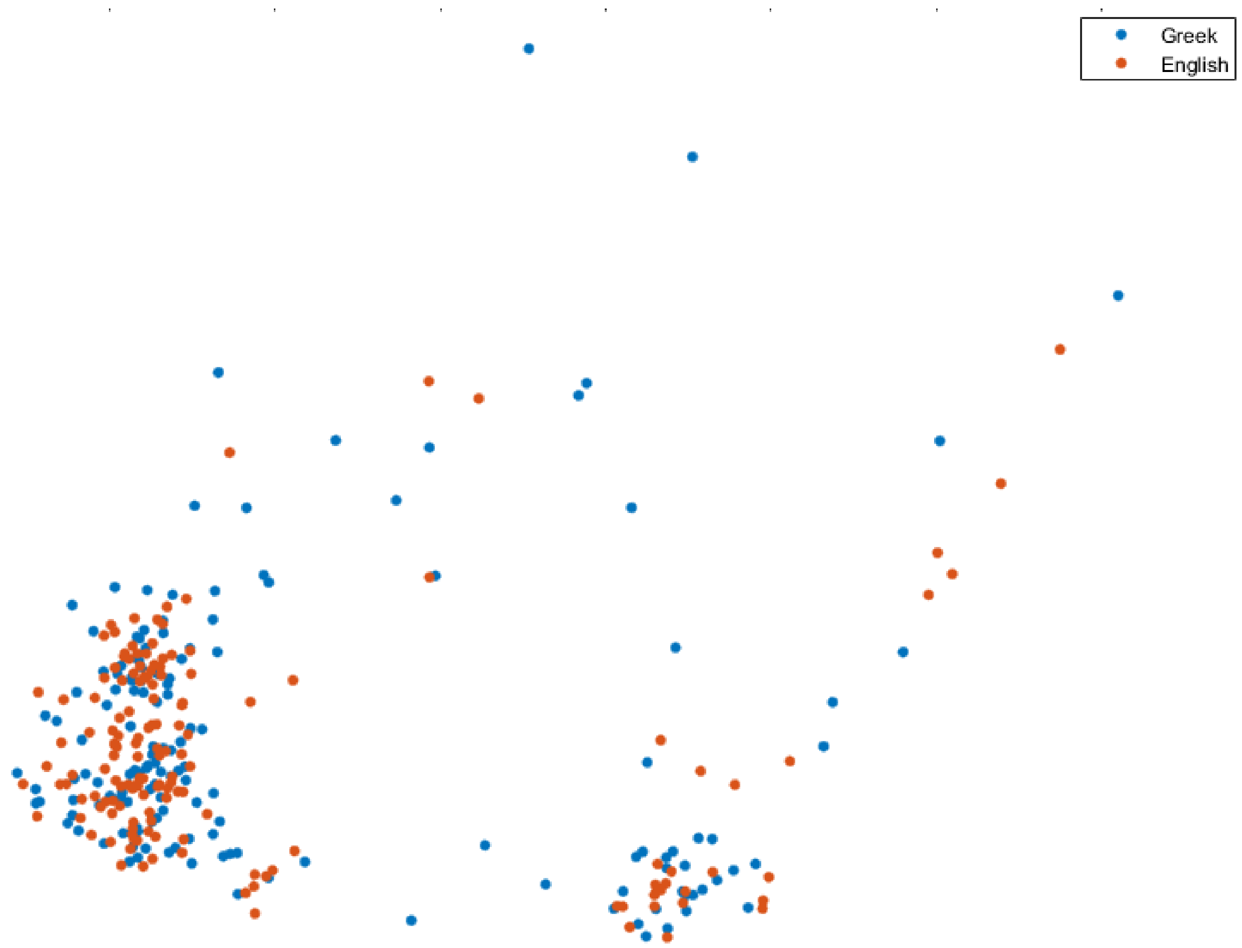

Figure 3.2: 2-dimensional representation for the gamma filtered single trial responses of S06 using MDS under the Affine Invariant Riemannian Metric when signals are colored according to language (Greek vs English).

SPECTRO-TEMPORAL NEURAL SIMILATIRY AND DIVERGENCE IN GREEK–ENGLISH WORD PAIRS

MOTIVATION

Inner speech is widely assumed to rely on abstract, language-independent representations that precede overt articulation. However, direct empirical evidence supporting this assumption, particularly at the level of individual lexical items across languages, remains limited. Bilingual inner speech provides a unique opportunity to address this question: if two words share the same meaning but differ in their phonological and articulatory realization, to what extent do their neural representations overlap during covert production?

To investigate this, we examined pairs of semantically equivalent words in Greek and English (e.g., *dice–ζάρι*, *arrow-βέλος*). This analysis was enabled by the newly acquired dataset generated within the BINGO project, comprising EEG recordings associated with English and Greek words of equivalent semantic meaning (see Deliverable D3.1 v2 for a detailed description). These word pairs allow a controlled dissociation between shared conceptual content and language-specific phonological form, making them ideal for probing whether inner speech is primarily driven by meaning or by language-dependent encoding mechanisms.

Our central hypothesis was that inner speech representations would be largely similar across languages, reflecting shared conceptual and motor planning processes, with selective divergences emerging for words that differ more strongly in phonological structure or articulatory complexity.

METHODOLOGY: PAIRWISE TIME – FREQUENCY DISCRIMINATION OF INNER SPEECH

EEG REPRESENTATION AND TIME–FREQUENCY ANALYSIS

EEG data were recorded during silent articulation of visually presented words. For each trial, the EEG signal from sensor $s$was transformed into the time–frequency domain using a continuous wavelet transform (CWT) [Herrmann, 2005]:

$$W_s(f,t,k) = |\ \mathrm{CWT}(z(x_s(t,k)))\ |$$

where $f$denotes frequency, $t$time, $k$trial index, and $z(\cdot)$indicates z-score normalization. The analysis focused on frequencies between 1–45 Hz and a post-stimulus window of 2.5–4 s, selected to emphasize sustained inner speech generation.

CROSS-LANGUAGE WORD-PAIR DISCRIMINATION

For each Greek–English word pair c, a binary discrimination analysis was performed by contrasting trials corresponding to the English word against trials of its Greek counterpart. At each sensor and time–frequency bin, trials were vectorized and evaluated using a non-parametric Wilcoxon rank-sum–based feature ranking. This yielded a discriminability score:

$$Z_s^{(c)}(f,t)$$

which quantifies the statistical separability between the two languages for word pair $c$at sensor s.

SENSOR-LEVEL WORD AND FREQUENCY DISCRIMINABILITY METRICS

To obtain interpretable sensor-level summaries, discriminative maps were aggregated in two complementary ways.

Word-Level Discriminability: For each sensor and class (word, language set, or word pair), discriminability scores were averaged across frequencies and across the selected temporal window (0 - 1.5 s). The maximum discriminability value across classes was retained:

$$\text{WordScore}(s,c) = \max_{c}\left(\frac{1}{T}\sum_{t\in[0,\ 1.5]}\frac{1}{F}\sum_{f} Z_s^{(c)}\,(f,t)\right)$$

This metric quantifies the contribution of each sensor to distinguishing specific lexical items, language sets, or cross-linguistic word pairs during inner speech.

Frequency-Level Discriminability: To identify informative spectral components, discriminability scores were averaged across time and classes, and the maximum value per frequency band was retained:

This measure highlights sensor-dependent frequency bands that are most sensitive to language and word-level distinctions in covert speech.

VISUALIZATION AND INTERPRETATION

The resulting discriminability metrics were visualized as sensor-by-class and sensor-by-frequency heatmaps. These visualizations provide insight into:

1. Spatial patterns of EEG sensors contributing to inner speech discrimination.
2. Spectral signatures differentiating languages and semantically equivalent cross-linguistic word representations.
3. The extent to which inner speech representations are shared or distinct across languages.

This analysis was conducted independently for each subject, and the resulting discriminability maps were subsequently averaged across the 20 subjects to obtain group-level representations.

RESULTS AND ANALYSIS: SHARED REPRESENTATIONS WITH SELECTIVE LANGUAGE-SPECIFIC MODULATION

LEXICAL-LEVEL CROSS-LANGUAGE SIMILARITY

Figure 4.1a summarizes sensor-level discriminability scores for all Greek–English word pairs, averaged across subjects. Lower values indicate stronger similarity between the two languages, whereas higher values reflect increased divergence.

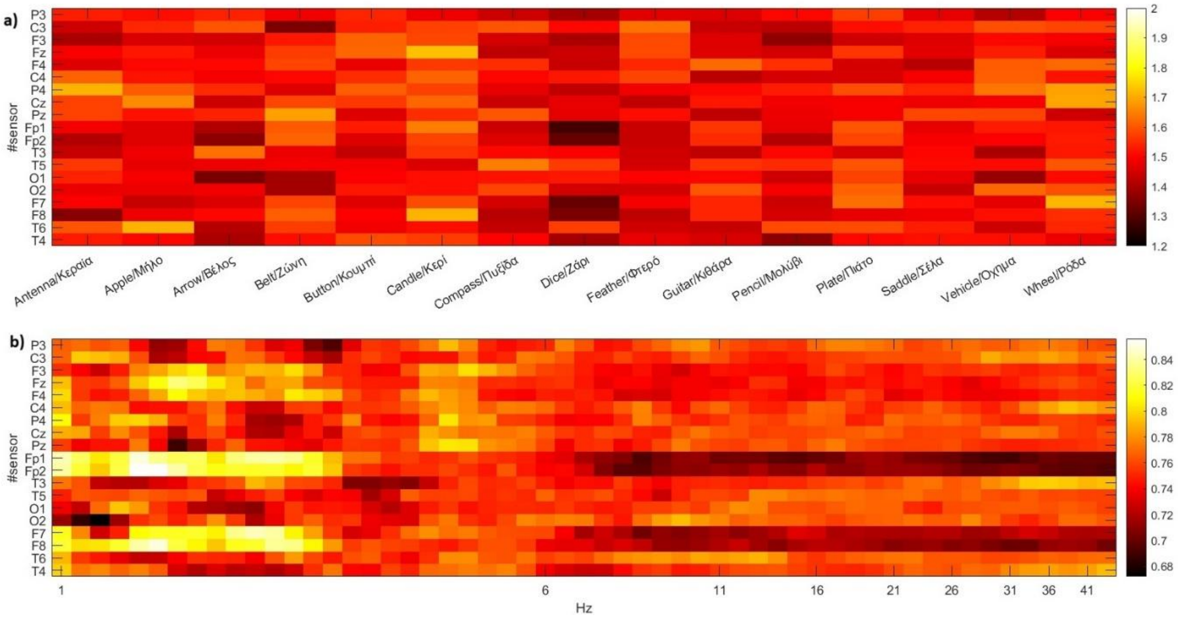

Figure 4.3: a) Pairwise word discrimination for all the available cross-linguistic word pairs averaged across subjects. b) Frequency related discrimination for each sensor averaged for all subjects and pair of words.

The dominant pattern in Figure 4.1a is one of widespread cross-language similarity: most word pairs exhibit low discriminability across nearly all sensors. This suggests that, for the majority of lexical items, inner speech representations are largely language-invariant, reflecting shared conceptual and motor planning processes rather than language-specific phonological form. This effect is spatially widespread and not restricted to a specific scalp region, suggesting that language-independent neural mechanisms dominate inner speech generation at the lexical–semantic level.

This finding aligns well with theoretical models of inner speech that posit a shared conceptual and articulatory planning stage preceding phonological encoding [Correja, 2015]. Since both words in each pair share the same meaning and communicative intent, the observed similarity likely reflects concept-driven neural activation, rather than language-specific surface forms.

In particular, word pairs such as *apple (μήλο)*, *button (κουμπί)*, *plate (πιάτο)*, *vehicle (όχημα)*, and *wheel (ρόδα)* show consistently low discriminability across frontal, central, and parietal sensors. These items correspond to highly concrete, visually grounded, and conceptually stable objects, for which semantic representations are expected to be robust and language-invariant [Li, 2023]. The neural similarity suggests that inner speech for these items may rely more heavily on shared conceptual imagery and motor planning than on language-specific phonological encoding.

Individual-Level Robustness and Selective Deviations: Importantly, the analysis underlying Figure 4.1a was also performed at the individual-subject level, revealing that the observed group-level trends are highly consistent across participants. Most subjects exhibited low discriminability for the same word pairs, confirming that the similarity is not an artifact of averaging.

At the same time, individual analyses revealed subject-specific divergences. For example, the word pair *antenna* (κεραία) showed increased discriminability in subject S7, particularly over fronto-central sensors. Such effects were not uniform across subjects, indicating that language-specific inner speech representations can emerge idiosyncratically, potentially due to differences in bilingual proficiency, articulatory strategies, or imagery-based encoding.

Word Pairs with Systematic Cross-Language Divergence: A limited set of word pairs, including *arrow (βέλος)*, *dice (ζάρι)*, *compass (πυξίδα)*, and *guitar (κιθάρα)*, exhibited consistently higher discriminability across subjects and sensors (Figure 4.1a). These differences were most pronounced over fronto-central and central electrodes, implicating brain regions involved in speech motor planning and articulatory rehearsal.

These findings suggest that, while semantic representations remain shared, phonological and articulatory differences between Greek and English introduce measurable neural divergence for words with more complex or language-specific phonetic structure.

Oscillatory Signatures of Inner Speech: Figure 4.1b depicts frequency-specific discriminability scores averaged across word pairs and subjects, highlighting the oscillatory components most relevant to inner speech.

Discriminative information is concentrated in a restricted set of frequency bands:

- Theta band (4–7 Hz) over frontal and fronto-central sensors, associated with working memory, lexical selection, and internal speech rehearsal.
- Alpha band (8–12 Hz) over central and parietal sensors, reflecting internally oriented attention and suppression of overt sensory–motor output.
- Low beta band (13–20 Hz) localized over central electrodes (C3, Cz, C4), consistent with covert engagement of speech motor planning networks.

INTERPRETATION AND RELEVANCE

Taken together, the spectral results indicate that inner speech is supported by a compact and interpretable oscillatory network, dominated by frontal theta, central–parietal alpha, and focal central beta activity. The absence of strong high-frequency effects suggests that inner speech decoding in EEG is primarily driven by low- and mid-frequency oscillations, which encode cognitive control, attentional gating, and covert motor planning.

Summarizing, The Greek–English word-pair analysis demonstrates that inner speech is characterized by strong cross-linguistic neural similarity, both at the group and individual levels. Language-specific differences emerge selectively and are primarily driven by phonological and articulatory factors, manifesting in fronto-central motor-related regions and low- to mid-frequency oscillations.

These findings support a model of inner speech in which semantic and conceptual representations dominate, while language modulates neural activity at later, motor-oriented stages. The consistency across subjects, combined with selective individual deviations, highlights inner speech as a stable yet flexible neural process.

CONCLUSION

This deliverable investigated cross-linguistic phenomena of inner speech using EEG through complementary Riemannian geometry–based and spectro-temporal analysis approaches. While the applied methodologies provided a structured and principled framework for examining language-dependent and language-invariant neural representations, the results did not reveal strong or consistent evidence of robust cross-linguistic differentiation or convergence across the examined frequency bands and analytical perspectives. Overall, the observed patterns suggest substantial overlap in neural representations of inner speech across languages, with only modest and variable effects that do not allow for definitive conclusions.

These findings highlight both the inherent complexity of inner speech processes and the challenges associated with detecting subtle cross-language neural signatures using non-invasive EEG recordings. Consequently, further work is required to better characterize cross-linguistic inner speech representations. Future investigations may benefit from larger participant cohorts, expanded lexical sets, refined experimental paradigms, and the integration of additional analytical methods or multimodal neuroimaging data. Such efforts are expected to improve sensitivity and robustness, ultimately advancing our understanding of how inner speech is encoded across languages in the human brain.

REFERENCES

Congedo, M., Barachant, A., & Bhatia, R. (2017). Riemannian geometry for EEG-based brain-computer interfaces: A primer and a review. *Brain-Computer Interfaces, 4*(3), 155–174.

Correia, J. M., Jansma, B., Hausfeld, L., Kikkert, S., & Bonte, M. (2015). EEG decoding of spoken words in bilingual listeners: from words to language invariant semantic-conceptual representations. Frontiers in psychology, 6, 71.

Herrmann, C. S., Grigutsch, M., & Busch, N. A. (2005). 11 EEG oscillations and wavelet analysis. Event-related potentials: A methods handbook, 229.

Li, H., Cao, Y., Chen, C., Liu, X., Zhang, S., & Mei, L. (2023). The depth of semantic processing modulates cross-language pattern similarity in Chinese–English bilinguals. Human brain mapping, 44(5), 2085-2098.

Pennec, X., Fillard, P., & Ayache, N. (2006). A Riemannian framework for tensor computing. *International Journal of Computer Vision, 66*(1), 41–66.

Shahapure, K. R., & Nicholas, C. (2020, October). Cluster quality analysis using silhouette score. In *2020 IEEE 7th international conference on data science and advanced analytics (DSAA)* (pp. 747-748). IEEE.