MSc Interaction Technology
Master Graduation Project

Decoding Neural Signatures of Language Comprehension and Production through EEG based Brain-Computer Interfaces

Venkatasubramanian

Supervisors: Gwenn Englebienne, Jan van Erp, Mannes Poel

January, 2025

Department of Human Media Interaction
Faculty of Electrical Engineering,
Mathematics and Computer Science,
University of Twente

UNIVERSITY OF TWENTE.

Contents

1 Acknowledgments **1**

2 Introduction **2**

3 Literature Review **3**
- 3.1 Neuroscience models of Language Processing . . . 3
  - 3.1.1 Dual Stream Model . . . 3
  - 3.1.2 Integrated Model . . . 4
  - 3.1.3 Multilingualism and Developmental Language Learning . . . 6
  - 3.1.4 Disorders of Language . . . 7
  - 3.1.5 Neural Basis of Language Comprehension . . . 8
  - 3.1.6 Neural Basis of Language Formation . . . 9
- 3.2 Experimental Protocols in Language BCI Studies . . . 10
  - 3.2.1 Language Comprehension Tasks . . . 10
  - 3.2.2 Language Production Tasks . . . 11
  - 3.2.3 Language Learning Tasks . . . 11

4 Problem Statement **13**

5 Materials and Methods **15**
- 5.1 Participants and Recruitment . . . 15
- 5.2 Device & Measurement . . . 16
  - 5.2.1 Experiment Setup . . . 18
  - 5.2.2 Tasks . . . 19
- 5.3 Analysis Pipeline . . . 22
  - 5.3.1 Data Loading and Cleaning . . . 22
  - 5.3.2 Epoching . . . 25
  - 5.3.3 Baseline Computation . . . 26
  - 5.3.4 Power Analysis . . . 26
  - 5.3.5 Topographical Comparison . . . 26
  - 5.3.6 Connectivity Analysis . . . 27

6 Design Research and Technical Specifications **28**
  - 6.0.1 Development of the Experiment Platform . . . 29
  - 6.0.2 User Interface Module . . . 30
  - 6.0.3 Logic Module . . . 31
  - 6.0.4 Experiment Module . . . 32

6.0.5 Recording Module . . . 33

7 Results & Findings **34**
7.1 Topographic Analysis of neural activity . . . 34
7.2 Power Analysis . . . 35
7.2.1 Theta Band Analysis . . . 36
7.2.2 Alpha Band Analysis . . . 37
7.2.3 Beta Band Analysis . . . 39
7.2.4 Development of Neural Markers . . . 41
7.2.5 Trend Analysis of the Proposed Neural Indices . . . 43
7.3 Connectivity Analysis . . . 53
7.3.1 Coherence and Phase Locking Value . . . 53
7.3.2 Granger Causality . . . 56

8 Discussion **62**
8.1 Limited Generalizability of Neural Markers . . . 62
8.2 Variability due to specific causes of difficulty . . . 62
8.3 Limitations of the Interface and LLM Response . . . 63
8.4 Scenarios and Difficulty Levels . . . 64
8.5 Limitations of the Device . . . 64
8.6 Future Considerations . . . 64

9 Conclusions **66**
9.1 Implications for Language Processing . . . 67
9.2 Open Source Contribution . . . 68
9.3 Final Remarks . . . 68

Chapter 1

Acknowledgments

I would like to express my sincere gratitude to my supervisors, Dr. Gwenn Englebienne, Dr. Mannes Poel, and Dr. Jan van Erp, for their constant support, guidance, and encouragement throughout the course of this project. Their expertise and constructive feedback were instrumental in shaping the direction and outcome of this research. I am deeply grateful to my parents and friends for their support and encouragement, and to the Almighty for granting me the strength and perseverance to navigate the challenges. I extend my thanks to all the participants who generously dedicated their time to this study without expecting anything in return. Their willingness to contribute played a crucial role in the success of this project. Reflecting on this journey, I feel truly blessed for the invaluable experiences and lessons learned. Despite the challenges and demands, I found the process to be immensely enjoyable. I look forward to building upon this work and continuing my research in this field in the future.

Chapter 2

Introduction

Language is the cornerstone of human communication, not merely a tool for conveying thoughts but a lens through which we perceive and understand the world. The developmental trajectory from an infant's initial vocalizations to complex linguistic expression illustrates the remarkable process of first language acquisition. In contrast, acquiring a second language, especially in later life, presents unique cognitive challenges and fascinating neurological dynamics.

Secondary language learning hinges on two critical cognitive processes: comprehension (decoding and understanding linguistic input) and production (generating coherent linguistic responses). These processes engage distinct yet interconnected neural circuits, each presenting unique cognitive demands. While comprehension involves semantic and syntactic decoding, production requires intricate neural orchestration for linguistic construction of thought.

Despite their fundamental role in communication, distinguishing the neural mechanisms underlying comprehension and formation difficulties has remained a significant challenge in cognitive neuroscience. This thesis employs a sophisticated, non-invasive EEG-based brain-computer interface to systematically identify and analyze neural features and markers differentiating these cognitive states. By examining spectral, topographical, and functional connectivity patterns across comprehension and formation difficulties (relative to a baseline relaxed state), this research aims to provide insights into secondary language acquisition mechanisms. The potential implications extend beyond theoretical neuroscience, promising to enhance language learning technologies and rehabilitation strategies for individuals with language-related neurological conditions such as aphasia.

Keywords: language learning, brain-computer interfaces, neural markers

Chapter 3

Literature Review

The intersection of neuroscience and language research represents a critical frontier in understanding human cognitive processes. This literature review synthesizes contemporary research at the forefront of neural science, language learning, and advanced neuro-imaging techniques. It is structured to systematically examine the following key domains:

- Neuroscience models of language processing
- EEG-based neural markers in language related tasks
- Experimental protocols in brain-computer interface (BCI) language studies

Although the neural mechanisms underlying language are not yet fully understood, significant insights have been gained through studies of patients who have experienced speech and language impairments due to stroke, aphasias, as well as through behavioral and functional neuroimaging studies of healthy individuals. Despite extensive research on language processing, achieving a comprehensive understanding of its neural basis remains a challenge. In this study, we aim to explore this limitation by investigating whether it is possible to distinguish between two states: "Difficulty in Comprehension" (State 1) and "Difficulty in Formation" (State 2).

3.1 Neuroscience models of Language Processing

3.1.1 Dual Stream Model

Language processing has traditionally been explained through modular theories that posit distinct and independent systems for production and comprehension [1]. These theories often draw on clinical evidence, such as the Lichtheim–Broca–Wernicke model, which attributes specific neural pathways to either production or comprehension [2]. Modular theory emphasizes the division of language comprehension and production processes into two distinct pathways: dorsal and ventral streams [3]. According to this theory, the dorsal stream is involved primarily in the mapping of auditory signals to articulatory representations and plays a crucial role in language production. In contrast, the ventral stream is responsible for mapping auditory input to semantic and syntactic representations, facilitating comprehension [4]. This bifurcation aligns with classical modular theories, which propose that language processing occurs in specialized, independent circuits [1]. Evidence for this division comes from neuro-imaging studies and lesion analyses. For instance, damage to the dorsal stream, particularly in the posterior superior temporal gyrus and inferior parietal lobule, often results in deficits in speech repetition and articulatory planning [5]. Conversely, lesions in the ventral stream, including the middle and inferior temporal gyri, are associated with impairments in semantic processing and comprehension [6]. These findings reinforce the notion of distinct pathways for comprehension and response formation.

However, it can be argued that the Dual Stream Model oversimplifies the interactive nature of language processing. Real-world communication often involves simultaneous production and comprehension, challenging the strict separation of these streams [7]. However, proponents of the model highlight its ability to explain specific language deficits observed in aphasia and its alignment with well-established anatomical and functional specializations [8]. Despite its modular framework, the Dual Stream Model has been adapted to incorporate elements of integration. For example, feedback loops between the dorsal and ventral streams allow for some degree of interaction, such as semantic influences on articulatory planning [9]. This adaptation addresses some criticisms while maintaining the core modular premise. Although not agreed upon by the majority of neuroscientists, the Dual Stream Model remains a good candidate to explain the differentiation between comprehension and formation, offering a structured framework to understand the neural architecture underlying secondary language acquisition. Although it may not fully capture the dynamic and predictive aspects emphasized by integrated models, it provides valuable information on the anatomical and functional basis of comprehension and response formation.

3.1.2 Integrated Model

Recent advances in neuroscience challenge the modular assumptions, advocating for an integrated view where production and comprehension are interwoven [7]. Neuroimaging studies reveal significant overlap in brain regions active during both production and comprehension tasks. For instance, Broca's area, traditionally associated with production, is also activated during comprehension and even during silent observation of speech-like movements [10, 11]. Neuroimaging studies show overlapping activation patterns in areas such as Broca's area and the auditory cortex during both speaking and listening [10]. Furthermore, behavioral studies demonstrate the dynamic nature of dialogue, where interlocutors often overlap in their contributions, providing feedback and co-constructing meaning [12]. Such phenomena are difficult to explain through modular frameworks that assume sequential turn-taking. The concept of forward modeling offers another rebuttal. Drawing from motor control theories, forward models enable individuals to predict the outcomes of their actions and compare these predictions with sensory feedback [13]. As per the integrated model, during language processing, speakers use forward models to generate predictions of their utterances, while listeners employ covert imitation to predict upcoming speech [7]. Behavioral experiments, such as the work by Heinks-Maldonado et al. [11], demonstrate that speakers predict and monitor their speech through forward models, adjusting phonological output in real time. Similarly, Huettig and Hartsuiker [14] observed that speakers use predictive monitoring to detect and correct semantic and phonological errors efficiently. This mechanism not only facilitates real-time adjustments but also explains how comprehension and production processes mutually influence each other. Joint action frameworks further validate the integrated model. In tightly coupled activities like conversation, interlocutors coordinate their actions by predicting their own and others' contributions [15]. This predictive capability aligns with findings on mirror neurons, which activate during both the execution and observation of actions, underscoring the shared neural substrates for perception and production [16]. The neuroscience of language is increasingly moving away from modular theories toward integrated models that emphasize prediction and mutual influence. The interwoven processes of production and comprehension, supported by behavioral and neuroscientific evidence, provide a robust framework for understanding language as a dynamic, interactive system. This perspective not only resolves inconsistencies in modular theories but also aligns with broader findings in cognitive neuroscience that reject the "cognitive sandwich" model of separate perception, cognition, and action [17].

While the Integrated Model offers compelling insights, it faces criticisms that highlight its limitations and areas requiring further empirical support. Critics argue that the model may oversimplify complex cognitive processes by assuming that production mechanisms directly influence comprehension without fully accounting for the nuanced differences between these processes. Such an assumption may fail to capture the specialized cognitive functions involved in either domain [9] [12]. The model's reliance on predictive coding as a central mechanism has also been questioned due to limited empirical evidence. While prediction plays a role in language processing, its ubiquity and extent in naturalistic language tasks remain contentious. Studies have shown that prediction effects often depend on specific experimental conditions and task demands, suggesting that prediction may not universally underpin all aspects of language comprehension and production [11,13]. Furthermore, generating predictions in linguistically complex or ambiguous contexts imposes a significant cognitive load, potentially limiting the practicality of prediction-based models [9]. Another challenge lies in accounting for individual differences in predictive abilities across populations. For instance, multilingual individuals or those with varying linguistic proficiencies may exhibit divergent interactions between production and comprehension processes, complicating the universality of the Integrated Model [17]. Similarly, alternative explanations, such as priming or rapid integration, could account for observed phenomena attributed to prediction, without necessitating the complex interactions proposed by the Integrated Model [15]. Lastly, the model's adaptability to multilingual or non-standard linguistic contexts remains underexplored. Multilingual speakers, for instance, might rely on distinct neural or cognitive mechanisms due to the interplay of multiple linguistic systems. This complexity poses challenges to the generalizability of the Integrated Model [9] especially for secondary language acquisition. Despite the criticisms, the Integrated Model continues to advance our understanding of interactive language processing. Addressing these challenges through targeted empirical research and refining the model to incorporate individual and contextual variability will enhance its explanatory power.

For this research, we critically evaluate the arguments presented by both the Dual Stream and Integrated Models without making prior assumptions about which model may prevail. Both frameworks may hold validity in addressing different facets of language processing. To advance this inquiry, empirical validation is essential, which necessitates developing a comprehensive experimental system. Such a system would enable controlled experiments involving participants from diverse linguistic and cultural backgrounds aimed at secondary language learning, as well as varying abilities and difficulties, to rigorously test the predictions of each model.

3.1.3 Multilingualism and Developmental Language Learning

Comprehension difficulties in multilingual contexts often stem from insufficient language input, inappropriate input, or challenges in language cognition. Learners may struggle with understanding different accents, dialects, or cultural nuances [18]. Formation difficulties, on the other hand, typically involve challenges in language production, such as pronunciation issues, grammatical errors, or difficulties in sentence construction [19]. Bilingualism provides a unique perspective on the brain's language networks. Research shows that bilinguals exhibit more extensive activation in the prefrontal cortex and anterior cingulate cortex, likely due to the cognitive control required to manage multiple languages [19]. Neuroimaging studies have revealed that bilingual language switching recruits both language networks and cognitive control areas [20]. This dual activation underscores the involvement of executive functions in managing two language systems, which might explain the challenges bilinguals face in rapidly switching between languages or maintaining separation between them. Structural studies further show increased gray matter density in regions associated with language control, such as the left inferior frontal gyrus and anterior cingulate cortex [21]. These findings illustrate how multilingualism can shape the brain's language structures and their functions over time, reflecting neuroplasticity that enhances cognitive control but may also contribute to processing delays or interference across languages [21].

Longitudinal studies on language development highlight the role of early language exposure in influencing structural connectivity, particularly in the arcuate fasciculus, a white matter tract connecting Broca's and Wernicke's areas [22]. Children exposed to rich linguistic environments show enhanced connectivity in this tract, supporting better phonological and syntactic processing as they develop [22]. Enhanced connectivity in the arcuate fasciculus may reduce formation difficulties in early bilinguals, providing further evidence of the long-term benefits of early linguistic exposure [21]. The age of acquisition and proficiency level play critical roles in shaping the neural representation of languages. Early bilinguals often demonstrate no significant difference in brain activation for L1 and L2, whereas late bilinguals may show spatially distinct activations in Broca's area for L1 and L2 [23]. This distinction could account for the greater formation difficulties experienced by late learners, who must recruit additional neural resources for L2 production [21]. However, proficiency has also been argued to be a more decisive factor than age of acquisition [22]. High proficiency, regardless of age of acquisition, has been linked to native-like neural representations, reducing comprehension and formation challenges [18]. Separating the effects of multilingualism from other cognitive factors remains a significant challenge. Some studies suggest that the bilingual advantage in cognitive control may be overstated or confounded by socio-cultural and educational variables [18, 22]. This ongoing debate underscores the necessity for more controlled and rigorous research to delineate these effects. Research that involves multilingualism must address the roles of age of acquisition, proficiency, capacity to learn new language and cognitive control to unravel the intricacies of multilingual language processing.

3.1.4 Disorders of Language

Language disorders encompass a range of difficulties in both comprehension and production of language, affecting individuals' ability to understand and express themselves effectively. These disorders can significantly impact academic performance, social interactions, and overall quality of life [24]. Comprehension disorders, also known as receptive language disorders, involve difficulties in understanding spoken or written language. Individuals with comprehension disorders may struggle to follow complex instructions, grasp abstract concepts, or interpret nuanced language such as idioms or sarcasm. Research by Bishop et al. [24] suggests that children with Developmental Language Disorder (DLD) often experience persistent difficulties in language comprehension, which can lead to challenges in academic settings and social interactions. These difficulties underline the critical need for targeted support and accommodations in both educational and social environments. Production disorders, or expressive language disorders, manifest as difficulties in formulating and expressing thoughts verbally or in writing. Symptoms may include limited vocabulary, grammatical errors, and trouble organizing ideas coherently. A study by Leonard [25] indicates that children with DLD often exhibit deficits in morphosyntax, such as omitting grammatical morphemes or struggling with complex sentence structures. These expressive challenges highlight the necessity of tailored interventions to address specific deficits in language production.

These disorders are intricately linked to the research on learning disabilities, particularly dyslexia. While dyslexia primarily affects reading skills, it is often accompanied by broader language processing difficulties. Snowling and Hulme [26] argue that dyslexia should be viewed as part of a continuum of language disorders, rather than a distinct entity. Neuroimaging studies provide further evidence of the biological underpinnings of language disorders. Friederici [27] demonstrates that individuals with language disorders often show atypical activation patterns in brain regions associated with language processing, such as Broca's and Wernicke's areas. These findings underscore the neurobiological basis for observed difficulties in both comprehension and production and point to potential biomarkers for diagnosis and intervention.

Importantly, the research on language disorders highlights the need for targeted and evidence-based interventions. Ebbels [28] proposes that interventions for children with language disorders should focus on explicit teaching of language structures, using visual supports and metacognitive strategies to enhance both comprehension and production skills. Systematic reviews, such as those by Norbury et al. [29], have further emphasized the effectiveness of evidence-based practices in improving syntactic and morphological skills in children with DLD. Further research is needed to explore further the interconnections between various language-based learning difficulties to develop more effective and targeted support strategies. Consequently, understanding the underlying neural phenomena—what occurs in the brain—and incorporating EEG studies into this research are crucial to advancing our knowledge and developing targeted approaches to find solutions to these conditions.

3.1.5 Neural Basis of Language Comprehension

Language comprehension relies on a complex interplay of neural regions and networks, each contributing to distinct aspects of processing. Research has elucidated key brain areas and mechanisms responsible for decoding semantics, syntax, and phonology, offering insights into the neural underpinnings of language comprehension [3, 30]. The middle and superior temporal gyri (MTG/STG) are central to the semantic network, showing robust responses to words over non-words and natural speech over scrambled speech [31]. The MTG is primarily implicated in lexical retrieval and semantic processing, while the STG supports phonological decoding and auditory processing [3]. Functional MRI studies confirm that these regions exhibit consistent spatial organization across individuals, suggesting that innate cortical architecture and anatomical connectivity shape these areas [31]. Evidence from neuroimaging and lesion studies highlights involvement of Broca's area in processing hierarchical syntax and resolving sentence ambiguity [30, 32]. Patients with Broca's aphasia exhibit difficulties in understanding syntactically complex sentences, emphasizing the role of this region in both syntax and verbal working memory [33] .Emerging evidence also points to the anterior temporal lobe (ATL) as contributing to combinatorial semantic processing. The ATL integrates word meanings into coherent sentence structures, challenging earlier views that focused predominantly on posterior temporal regions for semantic comprehension [34]. The N400 component, an event-related potential (ERP), serves as a key marker of semantic processing. It is sensitive to semantic incongruities, such as unexpected or contextually anomalous words, reflecting real-time neural responses to meaning violations. The N400 aligns with the dual-stream model of language processing, where the ventral stream is implicated in mapping sounds to meaning. Semantic networks exhibit small-world structural properties, enabling rapid retrieval and integration of word meanings [35]. Prat and Just [36] describe three key facets of neural connectivity that support comprehension: neural efficiency (minimal activation for task execution), adaptability (flexible reconfiguration with changing demands), and synchronization (integration across regions).

The dual-stream model further underscores the complementary roles of different pathways in language comprehension. The ventral stream supports semantic mapping and comprehension, while the dorsal stream aids in phonological processing and syntactic integration [3].

Despite significant advancements, several challenges remain in understanding the neural basis of language comprehension. Broca's area, for example, is widely recognized for its role in syntax, yet some researchers argue that its primary function may extend to verbal working memory [33]. Additionally, although language is predominantly left-lateralized, evidence suggests that the right hemisphere contributes to processing prosody, figurative language, and contextual integration [37]. These findings challenge traditional models that focus solely on left-hemisphere dominance. Individual differences also play a crucial role in comprehension abilities. Variability in neural efficiency, adaptability, and synchronization among individuals highlights the importance of personalized approaches to studying language processing [36].

Understanding the neural mechanisms underlying language comprehension is vital for advancing theories of bilingualism, developmental language learning, and disorders such as aphasia and dyslexia. Identifying specific neural markers and pathways can inform targeted interventions aimed at mitigating language deficits and improving comprehension outcomes [30].

3.1.6 Neural Basis of Language Formation

Language formation, encompassing both spoken and written production, relies on a sophisticated network of brain regions that manage lexical retrieval, syntactic structuring, phonological encoding, and motor articulation. Recent neuroimaging and electrophysiological studies have expanded our understanding of these processes, offering a nuanced perspective on the neural mechanisms underlying language production [3]. Broca's area, located in the left inferior frontal gyrus (IFG), is traditionally associated with speech production but is now understood to play a critical role in syntactic processing and rule acquisition. Grodzinsky and Friederici [38] demonstrated increased activation in Broca's area during artificial grammar tasks, indicating its involvement in syntax formation. However, some researchers argue that Broca's area may serve broader cognitive control and working memory functions. The left posterior middle temporal gyrus (MTG) is essential for lexical retrieval and semantic processing. Binder et al. [6] emphasized its role in both comprehension and production, while Sakurai et al. [39] highlighted its specialization in orthographic retrieval for logographic writing systems, such as kanji, underscoring its significance in written language production. Wernicke's area, located in the posterior superior temporal gyrus (STG), is traditionally linked to language comprehension but also contributes to phonological encoding during production. Hickok and Poeppel [3] proposed that this region acts as an interface between auditory representations and articulatory motor programs, facilitating seamless language production. The anterior temporal lobe (ATL) has emerged as a critical region for semantic integration during language production. Patterson et al. [40] identified the ATL as a "semantic hub," essential for combining conceptual information across modalities to form coherent language.

Electrophysiological studies have shed light on the temporal dynamics of language production. Early lexical access, occurring within 200–300 ms, is marked by the P2 component, which reflects lexical retrieval processes sensitive to word frequency and age of acquisition [41]. Semantic processing follows between 300–500 ms, with the N400 component signaling responses to semantic interference during tasks like picture naming [42]. Finally, syntactic encoding occurs between 500–700 ms, with the P600 component linked to the processing of complex syntactic structures [43]. The brain's neural plasticity plays a pivotal role in language formation and acquisition. Foster et al. [44] demonstrated EEG-based neural changes during artificial language learning, showcasing how the brain adapts to new word-concept mappings. Abutalebi and Green [45] proposed that second language acquisition recruits additional neural resources, particularly in prefrontal regions, to manage cognitive control and language switching. Individual variability in language network organization presents significant challenges for developing universal models of language formation [46]. Ecological validity remains another pressing issue, as naturalistic language production paradigms must balance experimental control with real-world relevance [47]. This is the theme underscoring our current research.

3.2 Experimental Protocols in Language BCI Studies

Brain-Computer Interface (BCI) research in language processing has increasingly shifted its focus towards verbal stimuli, enabling more nuanced investigations into the mechanisms underlying comprehension, production, and learning. This section examines experimental protocols employed in language BCI studies, emphasizing their relevance to the current research and discussing their strengths, limitations, and challenges.

3.2.1 Language Comprehension Tasks

Language comprehension tasks in BCIs aim to investigate the neural processes involved in understanding verbal stimuli, with a focus on semantic and syntactic aspects. Traditional paradigms, such as semantic anomaly detection, are commonly used to elicit the N400 component—a neural marker indicative of sensitivity to semantic violations [48]. These paradigms shed light on how the brain processes unexpected linguistic content. Similarly, sentence processing tasks that vary in syntactic complexity reveal neural markers associated with syntactic comprehension [49]. While effective at isolating specific neural processes, such tasks often simplify real-world linguistic interactions by excluding contextual elements.

To address this limitation, context-dependent comprehension tasks incorporate broader contextual setups to examine how external information influences semantic integration [50]. This approach improves ecological validity but introduces challenges in maintaining experimental control. Furthermore, complex sentence analysis, which involves entire sentences instead of isolated words, enables the study of advanced linguistic constructs [7]. However, the high cognitive load required for such tasks can limit their generalizability across diverse populations. Thus, while these tasks effectively capture semantic and syntactic processing, future studies must balance ecological validity with methodological rigor to better reflect naturalistic comprehension processes.

In our research, eliciting the N400 response is challenging due to the controlled nature of comprehension and formation processes in experiments. Instead of structured paradigms, participants are placed in more organic, real-world scenarios that introduce difficulties in comprehension. This makes it hard to precisely identify the timing and location of their challenges. Hence, the experiment design shifts toward alternative neural markers, such as frequency bands, topographical distributions and connectivity measures, which align better with our research objectives.

3.2.2 Language Production Tasks

Language production protocols in BCIs investigate the neural correlates of word retrieval, syntax construction, and motor planning. Picture naming tasks with semantic interference, where participants name pictures while ignoring distractor words, provide critical insights into lexical selection and semantic control [51]. These tasks are highly specific but often focus narrowly on isolated word-level production, which may not fully represent the complexities of sentence-level language generation.

Sentence completion tasks, where participants are asked to complete partially presented sentences, provide a dynamic approach to examining predictive mechanisms involved in language formation [7]. While effective, these tasks require meticulous control over sentence complexity to elicit consistent neural responses. Maintaining the validity of the elicited responses across various confounding variables is crucial. For example, the neural response of an individual highly motivated to learn Dutch may differ from that of someone with some knowledge of the language but lacking motivation to learn it.

One of the primary challenges in production tasks is the interference caused by motor artifacts, which remains a concern even in paradigms designed to mitigate them. To address this, integrating EEG with complementary modalities such as eye-tracking or electromyography (EMG) can enhance the reliability of these protocols. Additionally, artifact detection and correction techniques must be applied during preprocessing to account for these issues effectively. Ensuring participant engagement through out the experiment is critical for obtaining robust data, as repetitive tasks can lead to reduced focus and performance due to monotony. For example, excessively long or monotonous tasks should be segmented into multiple sessions and designed with varied scenarios to sustain the subject's attention and engagement.

3.2.3 Language Learning Tasks

Language learning protocols in BCI studies monitor neural adaptations associated with the acquisition of new linguistic structures. Statistical learning paradigms expose participants to artificial languages with hidden patterns, enabling researchers to track neural changes as learners adapt to these patterns [52]. While effective in controlled environments, translating findings to natural and real-world contexts presents significant challenge in language based tasks. Errorless learning tasks, which minimize initial errors, are designed to optimize learning while tracking neural efficiency [53]. Although this approach reduces frustration and supports early-stage learning, it may not accurately reflect real-world trial-and-error dynamics, where errors often play a critical role in reinforcement learning. Interleaved practice tasks, which alternate between linguistic elements, provide insights into neural flexibility and retention [54]. However, their applicability to complex linguistic constructs, such as syntax and morphology, remains underexplored.

Although these protocols provide valuable insights into the neural basis of language learning, they often lack real-time adaptation mechanisms tailored to individual learning trajectories. Incorporating adaptive feedback systems that adjust task difficulty based on real-time EEG signals could enhance engagement and learning outcomes. This is the ultimate objective of extending this research. Furthermore, Language learning is a gradual progress, tracking long-term neural changes is essential for evaluating the efficacy of language learning interventions.

Recent BCI studies increasingly employ multimodal stimuli and real-time feedback to enhance task performance and user engagement. For example, neurofeedback with gamification incorporates game-like elements to maintain motivation and focus [55]. However, achieving a balance between the cognitive demands of gamified tasks and their engagement benefits presents a significant challenge. Adaptive difficulty adjustment, which dynamically modulates task complexity based on neural markers [56], allows for personalization but requires advanced algorithms capable of real-time data analysis and decision-making. While researchers often utilize existing markers such as the engagement index, workload index, or cognitive load indices defined in the literature, there is a notable gap in research aimed at defining specific features of an index tailored to measure difficulty in comprehension and language formation during learning tasks. This is the contribution intended from carrying out the current research. Multi-modal systems that integrate EEG with additional sensors could improve accuracy but may also increase the complexity and cost of experimental setups. The scalability of these protocols across diverse populations remains a critical challenge. Developing universal calibration methods and optimizing real-time signal processing pipelines are essential to ensure accessibility and reliability in varied contexts. By addressing these limitations, BCI-based language research can advance toward creating adaptive systems capable of supporting individual differences in language processing and learning.

Chapter 4

Problem Statement

Research integrating neuroscience and language has provided valuable insights into the neural basis of comprehension and production. However, significant gaps persist in our understanding of the dynamic interplay between these processes, particularly in the context of multilingualism and language learning. While current models provide a strong theoretical foundation, they often lack the ability to capture individualized neural patterns and adapt to varying linguistic proficiency levels, limiting their use in personalized learning environments. This graduation project addresses two intertwined aspects:

- The empirical evaluation of the two types of difficulty states associated with language learning, aiming to identify neural features associated with comprehension and production difficulties and assess their generalizability across individuals and sessions.

- The constructive development of an EEG-based experimental system specifically designed to aid in this process.

This research is particularly relevant for individuals learning new languages, as well as educators and cognitive scientists seeking to understand and optimize language training strategies. It could further contribute to the development of personalized brain-computer interfaces (BCIs) for language learning support.

Research Question 1: ***What neural features (specifically topographical, bandpower, and connectivity measures) most effectively identify and differentiate between language comprehension difficulties (State 1) and language production difficulties (State 2) when assessed against a standardized baseline (State 0) in a single-subject, multiple-session study conducted in naturalistic settings?***

By integrating diverse feature types and conducting the study in a naturalistic setting, we aim to enhance the ecological validity of the findings. The single-subject approach (Participant P1) with multiple sessions (three in total) allows for an in-depth examination of the identified features and their consistency over time while reducing the influence of potential confounding factors.

Research Question 2: ***Is there a trend in the identified neural features when we replicate this across 12 subjects in a single-session study to offer broader insights into the challenges associated with language learning, namely (State 1 vs State 2)?***

This question examines the applicability of the findings from Research Question 1. To evaluate their validity, it is essential to test them across multiple participants (Participants P2 to P13) through a single-session study. Analyzing EEG signals from diverse participants allows us to assess the consistency and reliability of the identified neural markers across varying linguistic backgrounds and proficiency levels. The single-subject multi-session approach was chosen to enable a detailed examination of intra-individual variability, while the multi-subject single-session study aims to capture inter-individual differences. This study focuses solely on EEG-based measures and does not consider other neuroimaging modalities such as fMRI or MEG, which could provide additional insights into the neural processes involved in language learning.

Design Question: ***What are the design requirements for developing an EEG-based experimental system aimed at conducting, analyzing and advancing BCI-based language research?***

This question informs the development of a flexible and robust experimental platform for EEG-based language research. The corresponding requirements are outlined in the chapter titled "Design Research and Technical Specifications."

Chapter 5

Materials and Methods

5.1 Participants and Recruitment

A total of 13 participants (7 females, 6 males; mean age = 24 years, SD = 2.00 years) were recruited. The participants were non-native Dutch learners with varying proficiency levels (A0, A1, A2), which were determined through a pre-experiment questionnaire. P1 participated in the primary study (RQ 1), which spanned three sessions. Participants P2 to P13 took part in a single-session study (RQ 2). Due to resource constraints, Study 2 included only 12 participants. To mitigate the effects of a small sample size, a within-subject design was employed to maximize statistical power by reducing inter-subject variability. The results were interpreted cautiously, and the findings provided exploratory insights to guide future studies with larger sample sizes.

- Inclusion criteria: Dutch language proficiency of at least A0, no history of neurological disorders, normal or corrected-to-normal vision.

- Exclusion criteria: Participants with severe cognitive impairments or familiarity with similar EEG studies.

The requirement of a minimum A0 Dutch proficiency was based on the necessity for participants to interact with the app, which operates exclusively in Dutch. During the pilot study, individuals without formal Dutch knowledge struggled with comprehension and sentence formation. Their responses, limited to a few familiar words, did not generate the data required for the study. Consequently, it became necessary to screen participants based on this proficiency threshold to ensure meaningful participation. Participants were selected after completing a screening questionnaire designed to assess their familiarity with the Dutch language. Only those who demonstrated at least an A0 level of Dutch proficiency were chosen for the pilot study. This screening was implemented because, during the initial pilot test, some participants were unable to form responses even with assistance and provided only single-word answers—an outcome that was not suitable for conducting the experiment effectively.

Convenience sampling was employed, with efforts made to ensure a diverse range of native languages among the participants. German speakers were excluded due to the linguistic similarity between Dutch and German. All participants provided informed consent before taking part in the study. They were informed about the study's purpose, procedures, and their right to withdraw at any stage without any consequences. The study was conducted in accordance with the ethical guidelines of the University of Twente. By implementing this rigorous recruitment process, the study ensured the inclusion of participants capable of meaningful engagement with the experimental tasks, thereby enhancing the validity and reliability of the collected data.

Proficiency Level	#
A0 (Introductory)	4
A1 (Beginner)	5
A2 (Elementary)	3

Table 5.1: Participant distribution of language proficiency levels.

Native Language	#
Greek	2
Malayalam	1
Mandarin	1
Portuguese	1
Spanish	2
Italian	1
Tamil	3
Tigrinya	1

Table 5.2: Participant distribution by native languages

Participants' Age	#
21	2
23	2
24	2
25	3
26	2
27	1

Table 5.3: Participant distribution by age

Commitment Level	#
Multiple times a week	5
Once a day	3
Not committed to a schedule	4

Table 5.4: Participant distribution by commitment levels.

5.2 Device & Measurement

The Emotiv EpocX device was chosen for the acquisition of EEG signals due to its 16-channel configuration, which facilitates a more refined and comprehensive capture of brain activity. The EPOC X EEG headset, renowned for its professional-grade data acquisition, operates with a 14-channel system, featuring the following specifications:

Figure 5.1: Emotiv EpocX Device - Isometric View

Figure 5.2: Emotiv EpocX Device - Top View

- **Bandwidth:** 0.16 – 43Hz, with digital notch filters at 50Hz and 60Hz
- **Channels:** AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4
- **References:** CMS/DRL references at P3/P4, with alternative left/right mastoid process
- **Sampling Rate:** 2048 Hz, internally downsampled to 128 SPS or 256 SPS (user-configurable)

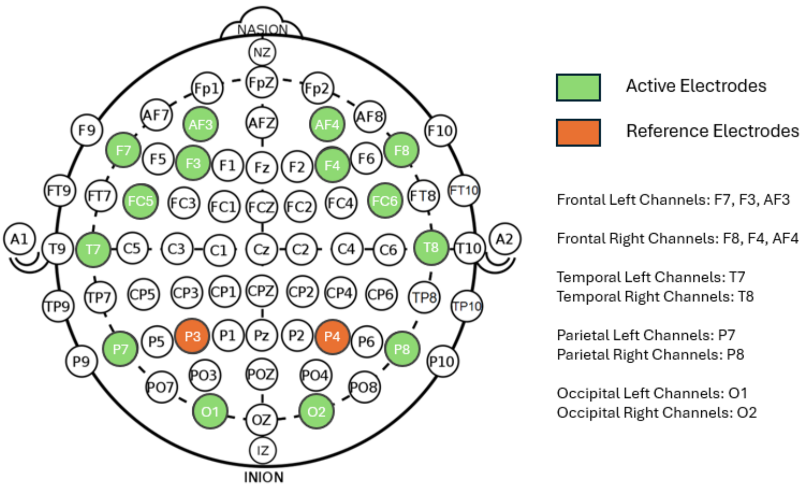

Figure 5.3: Emotiv EpocX Device - Channels

This selection was influenced by the extensive adoption of Emotiv devices in neuroscience research. Specifically, the Emotiv EPOC, EPOC+, and EPOC-X models are widely utilized by researchers due to their reliability and ease of use [57]. The device is accompanied by software and a license that facilitates its connection to a computer. Additionally, a Python library (https://github.com/Emotiv/

cortex-example/tree/mobile-example/python/lib) is available to enable integration with any Python program. To ensure the data was streamed via the LSL connection, a custom-built LSL adapter was incorporated into the library. Further details regarding this decision are provided in the design specifications section. The device was used in wet-electrode configuration using a saline solution to improve conductivity.

5.2.1 Experiment Setup

Participants were seated in a quiet, isolated room to minimize distractions and maintain a controlled environment for EEG recording. A large display monitor was employed to ensure the clarity of the presented information and to minimize unnecessary head movements. Participants were instructed to remove mobile phones and electronic devices, including smartwatches, and keep them away from the experimental setup.

Figure 5.4: Experiment Setup

Each participant interacted with the application displayed on a large monitor, which was positioned at an optimal distance and height for comfortable engagement. The application was designed to occupy only the central portion of the screen to minimize head movements during interaction. The EEG device was positioned on the participant's head before the experiment, with the electrodes arranged with care to ensure optimal placement and comfortable experience for the participant.

5.2.2 Tasks

Participant Onboarding and Consent

At the start of the experiment session, each participant was provided with an informed consent form detailing the study's purpose, data confidentiality, and their right to withdraw at any time. Participants reviewed and signed the form before beginning the study to ensure they understood their role and the study's objectives. This also included an instructional document that detailed the experimental tasks.

During this walkthrough, the functions of two critical buttons—*Unable to Comprehend* and *Unable to Formulate Response*—were explained, and participants were instructed on when and how to use these buttons to signal comprehension or response formulation difficulties. Prior to the task sessions, participants were instructed to respond to the chatbot using complete sentences as much as possible. Observations from pilot studies indicated that participants sometimes responded with single words, limiting interaction. Participants were encouraged to prolong conversations by asking questions or making comments, enhancing the volume and richness of linguistic data while allowing flexibility in their responses. Participants were encouraged to ask any questions they had. Additionally, a walkthrough of the interface was provided before the experiment to help participants become familiar with the experimental environment.

Pre-Task Mental State Assessment

All participants had refrained from consuming caffeine and food in the past hour prior to the experiment. According to the questionnaire, none of the participants reported feeling drowsy or sleepy.

Baseline Language Proficiency Assessment

To establish baseline language proficiency, participants completed a translation test with two parts: (1) Dutch-to-English translation and (2) English-to-Dutch translation. This baseline aimed to determine each participant's starting level in both directions of language translation.

- **Dutch-to-English Translation**: Participants were presented with 20 Dutch words or phrases to translate into English. The average score was 8 correct translations, with a standard deviation of 3.
- **English-to-Dutch Translation**: Participants translated 20 English words into Dutch, with an average score of 6.5 correct translations and a standard deviation of 2.

These scores established a baseline measure of Dutch proficiency for each participant, providing insights into individual starting points for comprehension and production tasks.

Baseline EEG Measurement

To record baseline neural activity, EEG measurements were conducted in a relaxed state at multiple points throughout the session. This baseline procedure was repeated between every new scenario. Participants were instructed to relax without engaging in any activity and to focus on a plus sign displayed in the center of the screen for 10 seconds. This fixation task allowed the recording of stable attention without cognitive interference, establishing a reference for analyzing neural patterns in the task conditions.

Core Task Scenarios and Difficulty Levels

The main task involved interacting with an artificial agent (chatbot) across five distinct scenarios:

- Scenario 1: Train Travel
- Scenario 2: Supermarket Shopping
- Scenario 3: Restaurant Ordering
- Scenario 4: Bank Inquiry
- Scenario 5: First Day at Work

Each scenario is structured according to three increasing levels of difficulty: Easy, Medium, and Hard. These levels introduced progressively complex language interactions within every scenario:

- **Easy Level**: The sentences generated by the chatbot involved simple vocabulary based on CEFR A0 level, with sentences averaging 5 to 6 words, designed to minimize cognitive load.
- **Medium Level**: This difficulty-level used sentences averaging 8 to 9 words (following CEFR A1 standards) with moderately challenging vocabulary.
- **Hard Level**: Sentences in this difficulty level averaged 10 to 12 words (following CEFR A2 standards) with advanced vocabulary and syntax, posing higher cognitive demands.

The tasks were presented sequentially, progressing from easy to difficult; however, this progression was not disclosed to participants in order to prevent expectation biases. Figure **??** illustrates the progression across these difficulty levels. Each difficulty level lasted 90 seconds from the initiation of user interaction. Once all difficulty levels for a given scenario were completed, the next scenario began.

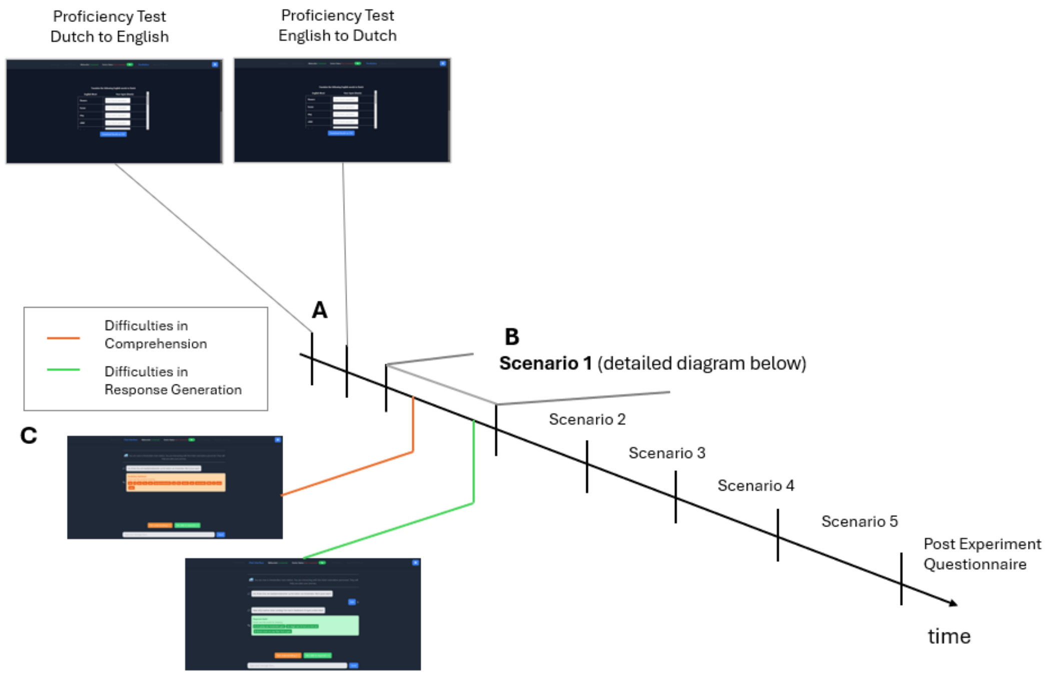

Figure 5.5: Experiment tasks; A - Proficiency tests; B - Scenarios; C - LLM-Generated Responses from clicking the assist buttons

Between scenarios, participants were presented with a rating scale to assess the difficulty of the previous section. A baseline EEG was also collected between each scenario. The ratings provided by the participants were used to calculate the correlation between the actual difficulty of the interaction and the perceived difficulty, ensuring the validity of the manipulations applied.

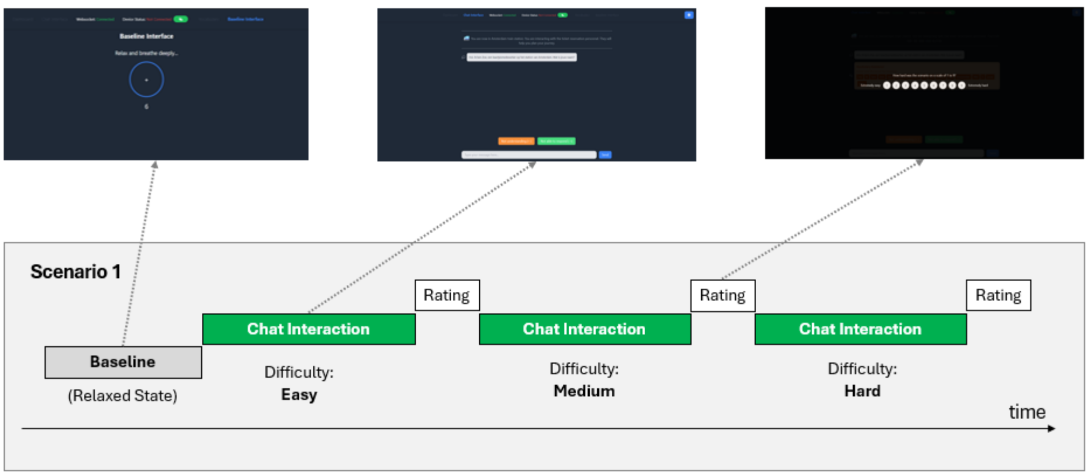

Figure 5.6: Chronological overview of tasks within a scenario

Experimental Task Execution

Following the interface orientation and instructions, participants began interacting with the chatbot. The tasks were conducted in the predefined sequence of difficulty, while EEG data were collected throughout the session to capture neural responses associated with language comprehension and production under varying cognitive demands. Each response and any use of the *Unable to Comprehend* or *Unable to Formulate Response* buttons were logged, enabling detailed analysis of comprehension and response formulation difficulties. This structured methodology ensured consistent data collection across participants, allowing reliable comparison of EEG patterns across baseline and task conditions, as well as across difficulty levels.

5.3 Analysis Pipeline

5.3.1 Data Loading and Cleaning

The data cleaning phase is a critical step to ensure that the analyzed signals are clean, interpretable, and suitable for downstream analysis. The data cleaning process in this study consists of the following components:

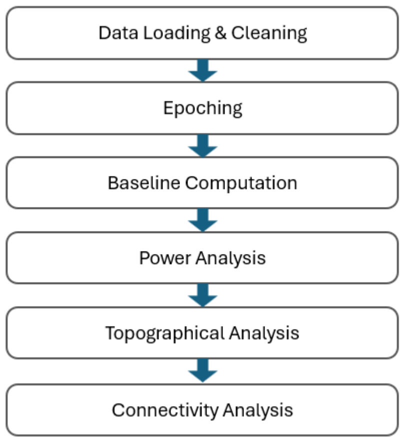

Figure 5.7: Steps taken for analyzing the collected EEG signals

Data Loading

The EEG data was collected during every session was saved into .edf format into the local disk. The markers including text input, button clicks (telemetry) were saved in a separate file with timestamps. During analysis this was loaded into the jupyter notebook where the analysis was performed. Response to difficulty ratings questionnaire was also recorded in a separate .csv file.

Marker Alignment

After the files are loaded into the analysis pipeline, the timestamps from the signal and from telemetry data are matched.A python package called mne was used for this purpose. It provided the functionality to create mne.raw and mne.epochs objects by combining the telemetry data and the raw signals captured during the experiment.

Outlier Detection and Drift Correction

Outlier Detection involves systematically identifying abnormal signal deviations that do not have any particular pattern of activity. Outlier detection step is applied across all 14 channels, Outliers are identified by applying a threshold defined as a multiple of the standard deviation, specifically using a 3 SD threshold. This approach captures significant deviations from the baseline signal, effectively isolating sections of the EEG likely to contain noise rather than meaningful neural activity [58, 59]. Baseline drift in EEG signals is corrected by applying a high-pass filter, which effectively removes low-frequency components that are typically associated with slow fluctuations and baseline instability. By utilizing a cutoff frequency, such as 1 Hz, the filter ensures that only the neural activity within the relevant frequency range is preserved. This preprocessing step is essential for maintaining the accuracy and reliability of the EEG signals, particularly in tasks requiring stable baseline measurements.

Artifact Correction and Bad Channel Rejections

Artifact Correction is aimed at detecting and excluding various artifacts from EEG data. Specifically, the approach addresses ocular artifacts (such as eye blinks and lateral eye movements, as well as), muscular and cardiac artifacts. First, the signals were inspected visually. Eye movements were detected by identifying simultaneous outliers in channels F9 and F10, which are positioned laterally to capture horizontal eye movements. 5.8 Eye blinks, by contrast, are detected through sustained outliers in the frontal channel FpZ, with the algorithm evaluating the duration of these deviations to confirm blink artifacts. 5.8 Then, Independent Component Analysis (ICA) was performed for artifact correction. ICA decomposes EEG data into statistically independent components, facilitating the identification and removal of artifacts such as ocular activity, muscle noise, and power line interference. This process produced 12 independent components, each representing a source of signal variance, which may include both neural activity and non-neural artifacts.

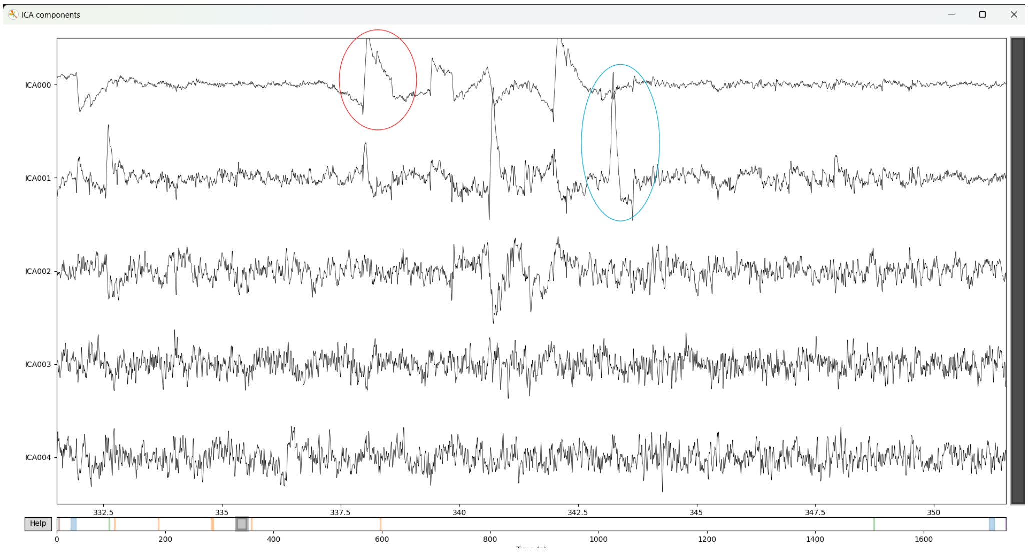

Figure 5.8: Outliers and Artifacts were identified first through visual inspection; the red circle shows eye movement, and the blue marker indicates eye blink artifacts.

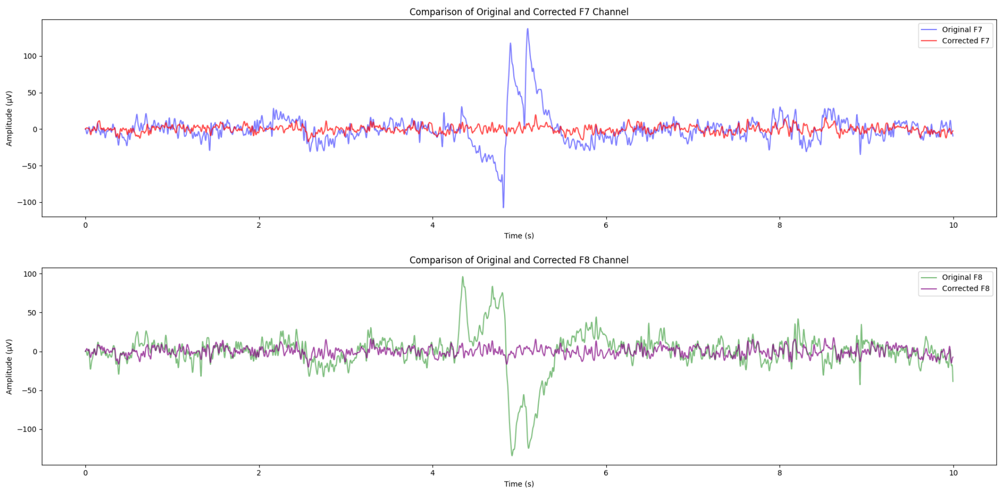

Figure 5.9: ICA-based correction (before & after) is applied to the signal; the red and purple colored lines represent the cleaned signal.

Montage and Referencing

The EEG signals were re-referenced to an average reference across all channels. This approach was selected because average referencing minimizes the influence of any single electrode, providing a balanced view of brain activity. By using this method, the spatial distribution of potentials across the scalp is preserved, which is critical for comparing activation patterns across different regions. Furthermore, average referencing reduces the impact of noise artifacts localized to specific electrodes, enhancing signal quality.

Filtering

To remove unwanted noise, the EEG data underwent a two-stage filtering process:

- **Low-Frequency Filtering:** A high-pass filter with a cutoff at 1 Hz was applied to eliminate low-frequency drifts caused by sweat, movement, or environmental factors.
- **Notch Filtering:** A notch filter at 50 Hz was used to suppress powerline interference, a common source of noise in EEG recordings.

This dual-filtering approach ensured that the data retained neural oscillations relevant to cognitive processes while minimizing the impact of artifacts and external noise.

5.3.2 Epoching

The EEG data was segmented into epochs using event markers recorded during the experiment. These markers were generated through the user interface when participants interacted with assist buttons labeled "Not Able to Comprehend" and "Not Able to Respond." Each button press indicated a specific cognitive state: difficulty in understanding the presented material or challenges in forming a response.

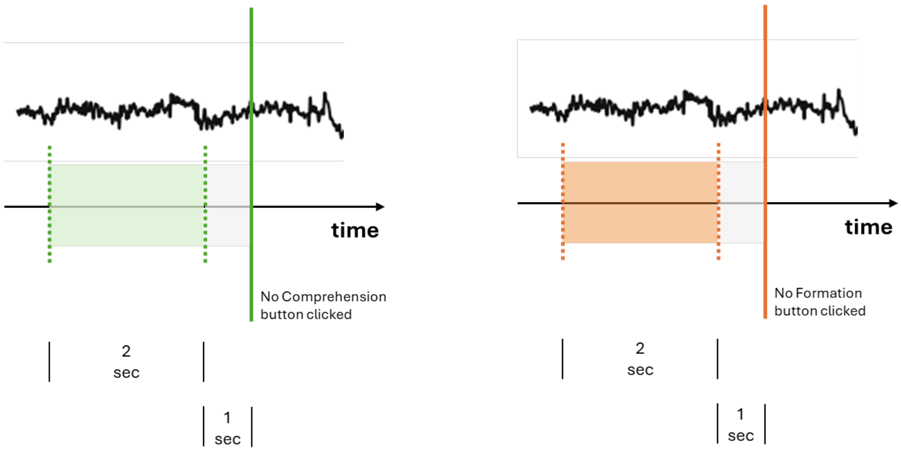

Figure 5.10: Epochs were created based on button clicks, with a duration of 2 seconds per epoch

To create epochs, a time window spanning from 3 seconds before the event marker to 1.0 seconds before the marker was extracted. The final 1.0 seconds preceding the button press were deliberately excluded to avoid contamination from movement-related artifacts, such as hand movements used to operate the interface. This approach ensured that the epochs predominantly captured brain activity related to cognitive processing rather than motor execution. A 2-second epoch length was selected as it provided a sufficient window to analyze neural responses while maintaining enough granularity for validation purposes. Furthermore, the 2-second window accounted for instances where participants re-read the presented words multiple times to verify comprehension. If they failed to comprehend, they pressed the "Not Able to Comprehend" button. Similarly, the "Not Able to Respond" button was pressed in cases of difficulty forming a response.

5.3.3 Baseline Computation

Baseline epochs were extracted to serve as a reference for comparisons with task-related neural activity. Baseline epochs were created from participant's neural resting state. These baseline periods were aligned with specific markers corresponding to participants' self-reported rest states. The baseline task was conducted three times across the experimental session. Every epoch was tagged to a baseline signal captured just before the task-induced changes in brain activity. Additionally, artifact detection algorithms (discussed in the previous section) were also applied to exclude contaminated signals of any kind.

5.3.4 Power Analysis

Power spectral analysis was conducted to quantify neural oscillations across different frequency bands, providing insights into cognitive processes related to language comprehension and production. The primary objective was to determine whether these two states engaged similar or distinct brain regions. Power spectrum changes were analyzed across theta (4–8 Hz), alpha (8–12 Hz), and beta (13–30 Hz) frequency bands by comparing task conditions to baseline measurements. Power spectral density (PSD) was estimated using the Welch method, focusing on all, frontal, temporal regions. Additionally, evoked potential analysis was performed to examine neural responses over time, revealing that the frontal regions exhibited stronger activation in both conditions. Visualizations, including bar plots comparing the bandpower values against the baseline were generated to facilitate interpretation of results. For RQ2, both parametric and non-parametric tests were conducted to confirm whether the observed trends in the sample data could be generalized to a larger population. These analyses provided a comprehensive understanding of the neural dynamics associated with language-related cognitive processes.

5.3.5 Topographical Comparison

The topographical distribution of power provided insights into the spatial characteristics of neural activation, highlighting the brain regions that exhibit increased activity. This analysis was particularly valuable in language-related tasks, where specific brain areas are associated with comprehension (e.g., temporal regions) and production (e.g., frontal regions). To investigate these differences, topographical maps were generated and compared across conditions, specifically analyzing each state (State 1, State 2) relative to the baseline (State 0). This comparative analysis offered an initial understanding of the key brain regions involved, guiding further in-depth investigations.

5.3.6 Connectivity Analysis

Functional connectivity was analyzed to gain insights into how different brain regions interacted during tasks, revealing the network-level organization of neural processes. This analysis was particularly valuable in language-related studies, where frontal and temporal regions are known to collaborate in supporting comprehension and production. Functional connectivity measures, such as coherence and the weighted Phase-Lag Index (wPLI), were employed to assess the strength and stability of communication between brain regions. Coherence, particularly imaginary coherence and wPLI, was utilized to evaluate connectivity between regions involved in language processing. Specifically, connectivity between the left frontal regions (e.g., F3, F7) and the left temporal-parietal areas (T7, P3) was examined to identify network-level differences between the two states of language processing.

Coherence analysis was performed to measure connectivity within specific frequency bands, assessing the strength of functional interactions between brain regions and identifying synchronized activity patterns. The Phase-Lag Index (PLI) and weighted Phase-Lag Index (wPLI) were used to evaluate non-spurious phase synchrony, minimizing the influence of volume conduction and highlighting genuine neural interactions. Directed connectivity analysis was conducted using Granger causality and transfer entropy to determine directional influences between regions. This analysis aimed to establish whether frontal regions led activity in temporal areas during language production tasks. Granger causality evaluated how well the past values of one time series (X) predicted the present values of another (Y), beyond what could be predicted by Y's own past values. The lag parameter was set to capture temporal dependencies, with a lag of 20 indicating that the past 20 samples of both X and Y were used to predict the present value. The connectivity analysis was conducted across different frequency bands to investigate how the strength of directional interactions varied with frequency. These findings provided valuable insights into the dynamic interplay between brain regions involved in language processing.

Chapter 6

Design Research and Technical Specifications

This chapter details the architecture, design, and technical implementation of the web application developed for integrating Brain-Computer Interface (BCI) systems.

Technical requirements for the experiment platform

The experiment platform must meet several critical requirements to ensure reliable and accurate execution. It must provide low-latency data processing to handle EEG signals with minimal delay, enabling real-time feedback to participants. Scalability is essential, allowing the system to support multiple concurrent sessions to accommodate group studies. A modular architecture is required to facilitate the easy integration or replacement of experimental modules without substantial modifications to the core system. Ensuring data integrity and security is crucial, with secure storage mechanisms that guarantee accessibility for post-experiment analysis. Additionally, the platform should feature a user-friendly interface with intuitive controls and visualizations to enhance participant interaction. Device compatibility is also a key requirement, necessitating support for multiple EEG devices through standardized protocols such as LSL and UDP.

Technical requirements for the analysis pipeline

The analysis pipeline must fulfill several criteria to ensure effective post-experiment analysis. It should include automated data preprocessing techniques such as filtering, artifact removal, and normalization to enhance data quality. Feature extraction capabilities are essential, supporting the derivation of meaningful metrics, including frequency bands, power spectral densities, and engagement indices. Reproducibility is a fundamental requirement to maintain consistency across experimental sessions and participants. The system should also provide comprehensive visualization options for both real-time and post-experiment data interpretation. Advanced statistical analysis tools must be integrated to examine correlations between EEG signals and experimental variables. Furthermore, the system should facilitate result export in multiple formats, such as CSV, JSON, and PDF, to support reporting and data sharing.

6.0.1 Development of the Experiment Platform

Having these requirements in mind, the processing layer (back-end) was primarily developed in Python, while the user interface layer (front-end) was constructed using ReactJS. Python was chosen due to its robust support and ecosystem for the protocols utilized in the experiment. The Python backend leveraged FastAPI, a high-performance web framework optimized for building APIs. The asynchronous capabilities of FastAPI, combined with Python's extttasyncio library, facilitated efficient handling of concurrent tasks across multiple threads, ensuring smooth system performance under the high data throughput required for BCI-based applications. ReactJS provided the flexibility necessary to customize the interface. The separation of logic and presentation layers was critical in making the system modular and scalable for other types of BCI-based studies conducted in a chat-based environment. This approach also enabled iterative refinement and optimization throughout the research process. The code for the developed system was made fully open-source and uploaded to GitHub as a public repository.

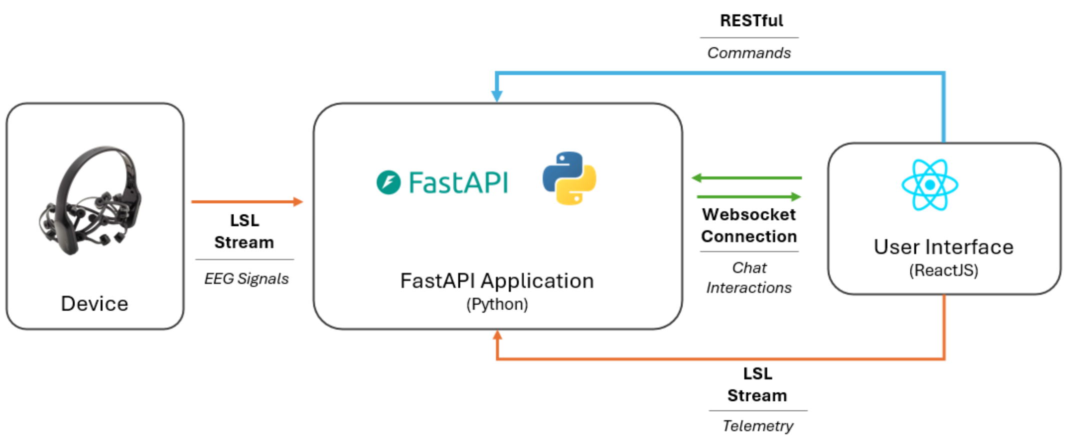

Figure 6.1: Communication protocols used for the interaction between various modules

The system employed two communication protocols to ensure seamless data flow:

- **Lab Streaming Layer (LSL):** Utilized for synchronization of EEG signals and telemetry data, providing excellent real-time performance with a latency of 0.1 milliseconds. All timestamps were converted to LSL time format, and estimated time-correction offsets were applied for precise synchronization.
- **WebSocket Protocol:** Supported low-latency interactions (200 microseconds) between the frontend and backend. This feature was essential for transmitting conversational responses, engagement indices, and assistive features in real time.
- **RESTful API:** Python endpoints were developed to handle incoming commands from the user interface, such as initiating baseline processes, storing questionnaire responses, and connecting/disconnecting the device.

Data acquisition was meticulously managed through external scripts initialized via the `subprocess` library, complemented by comprehensive logging mechanisms implemented using the `logging` library. This ensured effective monitoring, detailed debugging, and precise error reporting. The design enabled information exchange between the logic layer and other components to occur on separate processes and threads, ensuring a non-blocking computational workflow. The backend's architectural design (Figure 6.2) emphasized modularity, instantiated through specialized Python classes: `BufferManager`, which handled real-time EEG signal buffering; `BaselineManager`, which oversaw baseline calibration processes; and `BCIManager`, which facilitated seamless device integration and advanced signal processing through the implementation of a circular queue (FIFO) data structure. The `ChatManager` class interfaced with an LLM to handle chat conversations. This clear separation of functionality and communication resulted in a flexible and responsive system architecture that effectively met the complex demands of the study. The application enabled systematic storage of acquired EEG data in a structured, timestamped directory format (.edf, .fif), ensuring optimal data traceability and simplified retrieval.

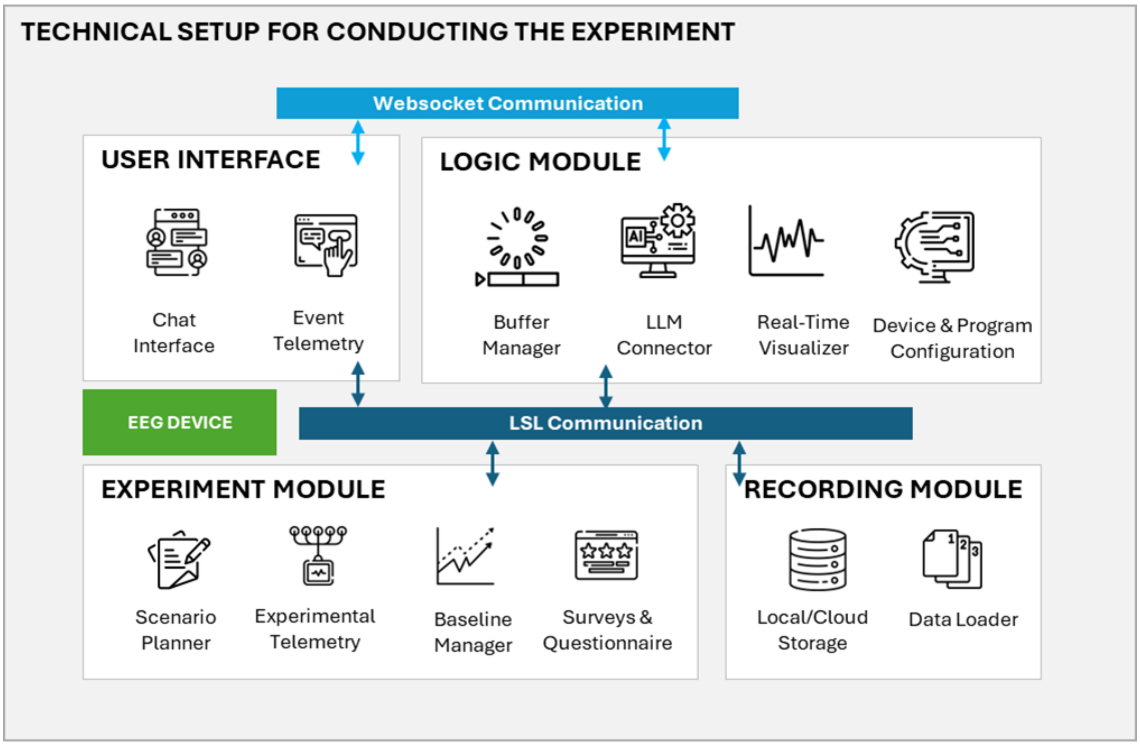

Figure 6.2: Software modules developed for the purpose of conducting the experiment

6.0.2 User Interface Module

The User Interface (UI) module served as the primary interaction point between the user and the system. It consisted of two main components: the **Chat Interface** and **Event Telemetry**. The Chat Interface facilitated user interaction with a chat agent (LLM-based) that communicated exclusively in Dutch, simulating a conversational environment for the experiment.

The interface included free-text input at the bottom of the screen and assist buttons labeled *Not able to Understand* (Button 1) and *Not able to Respond* (Button 2). As mentioned in the Methods section, participants clicked the Assist button when they were in State 1 and State 2. The logic layer then interfaced with the LLM to generate appropriate responses based on the history of the conversation and the new user prompt. When participants clicked Button 1, they were provided with vocabulary assistance, including English meanings of words used by the chatbot. When Button 2 was clicked, suggested response options (with english translations) were displayed with English translations.

The telemetry component was responsible for tracking and recording user interactions with the UI. It captured events such as mouse clicks, keypresses (keyup and keydown), and timestamps when new chat messages were presented. The timestamps of these interactions were relayed from the UI to the logic layer via the LSL stream. User responses to ratings and pre- and post-questionnaires were also captured through this module. These telemetry data points were crucial for analyzing user behavior and correlating it with EEG signals. Additionally, the UI provided basic data visualizations to monitor brain signals, including frequency band activity, ensuring connectivity and participant engagement. The UI layer also contained the baseline section where the participants in between the various scenarios were presented with a screen with plus symbol in the middle of the screen and instruction to relax and a countdown (10 seconds). The start of end timestamps were also relayed back to the logic layer. This was then used during the analysis to subset the signals for considering as baseline (reference signal)

6.0.3 Logic Module

Buffer Manager

The Buffer Manager utilized a custom-built data structure for efficient handling of EEG data. It functioned as a circular queue with a variable buffer length (set to a 2-second time window in this study), allowing temporary collation of EEG signals before saving them locally in .edf format. Other supported formats included .fif, .npz, and .csv. The circular queue minimized memory usage and allowed real-time Power Spectral Density (PSD) calculations. This component also made the logic layer extensible for adaptive changes to the UI, if necessary in the future.

LLM Connector

The LLM Connector, implemented using the LangChain framework, enabled two essential functionalities: 1. Generate conversational response from the LLM 2. Providing assistive content for the predefined assist buttons LangChain facilitated experimentation with various large language models and enabled the integration of a standardized context document to ensure the reproducibility of experimental conditions. This approach ensured a controlled conversational environment with consistent responses within the framework of a generative AI system. The `llama3-70b-8192` model was utilized for the study. Developed by Meta, this family of large language models (LLMs) consists of pretrained and instruction-tuned generative text models with a parameter size of 70 billion. The Llama 3 instruction-tuned models are specifically optimized for dialogue-based applications and have demonstrated superior performance compared to many open-source chat models on standard industry benchmarks. This module was further enhanced with a Retrieval-Augmented Generation (RAG) implementation, incorporating a collection of documents containing vocabulary aligned with official Dutch language proficiency levels A0, A1, and A2. The LLM was prompted to generate text using a structured output format, which was validated using Pydantic, a Python library for data validation and serialization leveraging Python type hints. This validation process ensured that responses from the LLM—both chat-based and assistive responses—adhered to a predefined format, allowing seamless rendering in the UI without requiring additional front-end processing.

Real-Time Visualizer

The Real-Time Visualizer processed and displayed key EEG metrics such as frequency band activity. This component provided experimenters with real-time insights into participants' cognitive states during the experiment.

Device and Program Configuration

This configurations of the experimental parameters, including device selection, buffer durations, frequency band definitions, electrode selection, and data recording preferences etc. were maintained separately. These configurations enable flexibility in adapting the system to various experimental requirements. This came in handy during the initial piloting to iterate over study design.

6.0.4 Experiment Module

The Experiment Module orchestrated the experimental conditions and dynamically adapted scenarios and difficulty levels. It included the **Scenario Planner**, **Experimental Telemetry**, **Baseline Manager**, and **Surveys and Questionnaire**. The Scenario Planner, inspired by PsychoPy, ran as a separate thread and managed the execution of experimental tasks. It dynamically adjusted scenarios (e.g., *trein, supermarket, restaurant, bank, office*) and their difficulty levels (*EASY, MEDIUM, HARD*) during the study. The experimental flow was dictated by this planner, ensuring precise control over scenario progression and participant exposure. The telemetry component received the markers and subjective ratings provided by participants through the UI during the experiment. It also logged scenario changes and timestamps. Extensive logging provided a detailed dataset for debugging during application development. The Baseline Manager recorded participants' "relaxed state" EEG signals before the initiation of each scenario. This allowed for calibration and normalization of EEG data for subsequent tasks, providing a baseline reference for analyzing cognitive state changes. Participants completed subjective rating scales during the experiment, such as "How easy/difficult was the previous section on a scale of 0 to 9?" These ratings validated whether the experimental manipulations aligned with perceived difficulty. This analysis was covered in the discussion section.

6.0.5 Recording Module

The Recording Module was responsible for data storage and retrieval. It comprised the **Local/Cloud Storage** and the **Data Loader**. This storage component handled the storage of all experimental data, including raw EEG signals, metadata, and telemetry logs. Each session was organized in a dedicated folder structure, ensuring traceability across multiple participants. The Data Loader aggregated session data for post-processing and analysis. It collated EEG signals, metadata, and event logs, loading them into memory for subsequent computational tasks. This module was a critical component of the post-processing pipeline.

Chapter 7

Results & Findings

7.1 Topographic Analysis of neural activity

The provided topographic maps depict the average evoked potentials (EPs) during epochs involving comprehension difficulties (State 1, top row) and formation difficulties (State 2, bottom row). These maps reveal clear differences in neural activity patterns, spatial distribution, and timing across both conditions. The color scale is set to a fixed minimum and maximum for accurate comparison. The first frame in each of these scenarios, is the baseline average evoked potential.

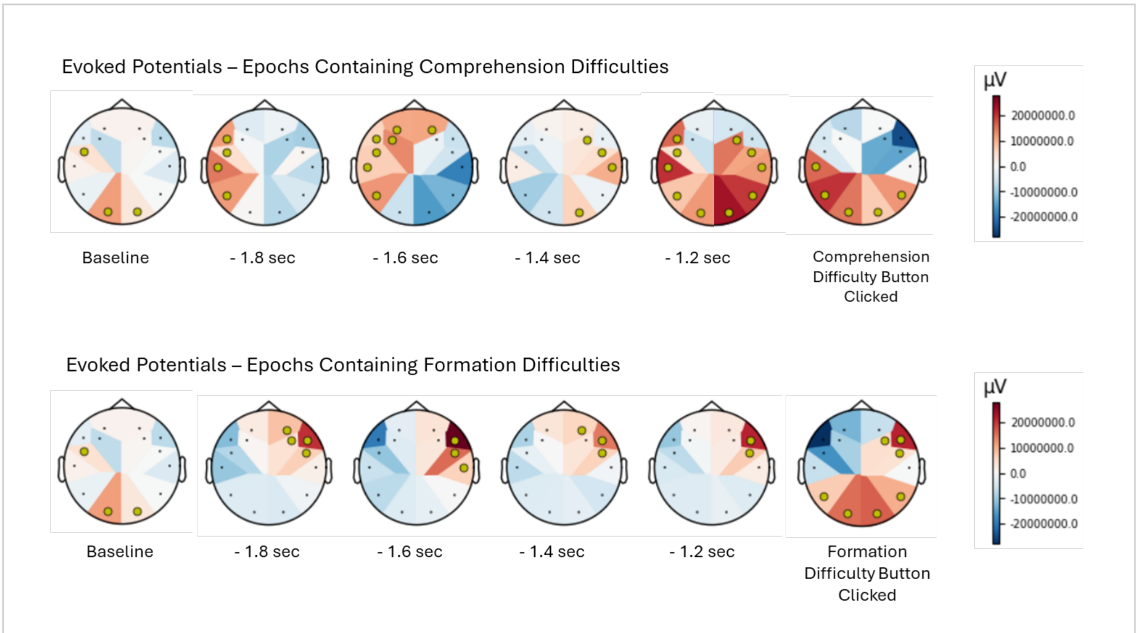

Figure 7.1: Evoked Potentials Prior to Events of Interest During Single Study based on RQ1 (Yellow dots shows regions of higher activation

The baseline condition (left-most columns) demonstrates minimal neural activation with a balance of positive (red) and negative (blue) potentials across the scalp. This pattern suggests a relaxed or neutral cognitive state where no significant linguistic or motor challenges occur. The observed baseline activity aligns with previous studies where resting-state EEG presents low-amplitude, non-localized neural signals. The average evoked potentials (EPs) observed during comprehension difficulties (State 1) exhibit distinct activation patterns, predominantly involving the posterior regions and the right hemisphere. These observations provide insight into the neural mechanisms underlying the processing of complex or ambiguous linguistic input. As shown in the topographic maps, the right parietal and temporal regions demonstrate increased positive activity ($+\mu V$, red zones) starting at approximately -1.6 seconds before the button press, intensifying up to -1.2 seconds. This pattern highlights the critical role of the right posterior temporal lobe and the parietal cortex in attentional regulation and semantic integration. These regions are known to be responsible for processing complex linguistic information and reallocating cognitive resources for understanding. The posterior dominance observed here aligns with theories suggesting that comprehension relies heavily on the integration of semantic information and visuo-spatial attention. This may indicate participants engaging posterior networks to comprehend or resolve linguistic ambiguity. Notably, the frontal regions display relatively lower activation during comprehension difficulties when compared to formation difficulties. This reduced frontal engagement aligns with the nature of comprehension tasks, where executive planning and motor outputs are less prominent. Instead, the cognitive load is distributed to posterior brain regions for semantic interpretation and memory retrieval. The progression of activity begins at around -1.8 seconds and peaks closer to the button press, reflecting the cumulative cognitive effort required to process the linguistic input. The activation patterns suggest that participants recruit posterior regions in a *timely and spatially coherent manner*, leading up to the acknowledgment of comprehension difficulty.

The average evoked potentials during language formation difficulties demonstrate a distinct right-lateralized fronto-temporal activation pattern, emphasizing the role of the right hemisphere in linguistic planning and production. Increased positive activity ($+\mu V$) is evident in the right frontal cortex, particularly around -1.8 to -1.6 seconds before the button press. The negative activity ($-\mu V$) in the left-frontal region is also observed. This activity likely reflects the engagement of Broca's area (left inferior frontal gyrus), a region critical for language production and the structuring of linguistic outputs. The heightened fronto-temporal activation further supports the theory that language formation relies heavily on executive functions and motor planning. Negative potentials ($-\mu V$) are observed in the central and parietal regions, particularly between -1.2 and the button click. These areas correspond to the sensorimotor network and are associated with the integration of motor planning and execution. Such activity reflects the preparation for motor responses (e.g., button pressing) and the involvement of parietal regions in coordinating sensorimotor processes.

7.2 Power Analysis

This section presents the observed changes in power spectral density (PSD) for the two difficulty states (State 1, State 2) described in the methods section. The step to compute the PSD from broadband signal is also detailed in the methods section. The EEG power changes observed in the "Difficulty in Comprehension"

(State 1) condition—namely, increased theta power, decreased alpha power, and reduced beta power—are largely consistent with established cognitive neuroscience findings. However, certain deviations, such as unexplained variations in theta power during the condition "Difficulty in Formation" (State 2), merit further investigation.

7.2.1 Theta Band Analysis

Increased Theta Power: Overall and Specifically in Temporal Regions

Elevated theta power is strongly associated with heightened cognitive load and working memory demands. [60] This increased Theta band PSD is particularly notable when observing the temporal channels during "Difficulty in Comprehension" (State 1), which play a crucial role in attention and executive functions. Note that it is not as differentiated in Frontal Channels where the means are only slightly higher as compared to the baseline. This difference is shown (albeit the the extent of difference is different across difficulty states (State 1 and 2), compared to baseline (State 0) and it should be due to higher cognitive demands of the tasks vs the relaxed state.

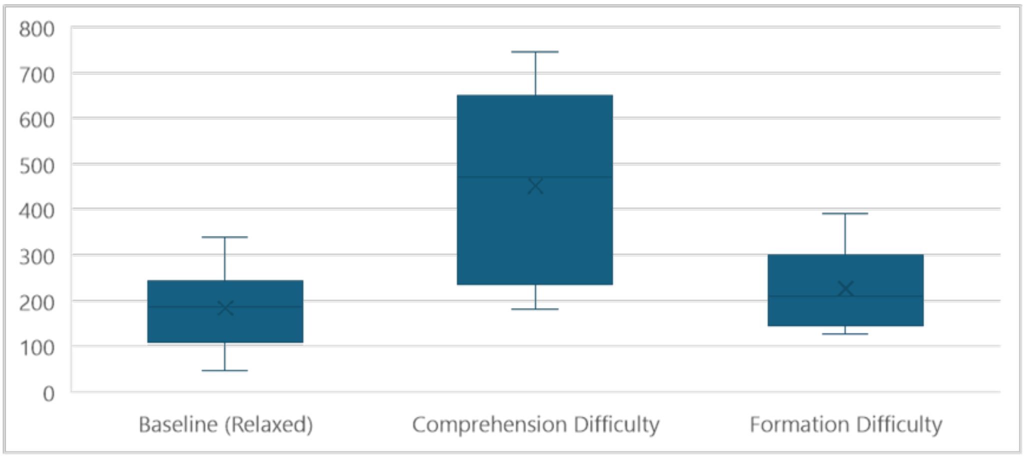

Figure 7.2: Average PSD in Theta Band **All Channels**

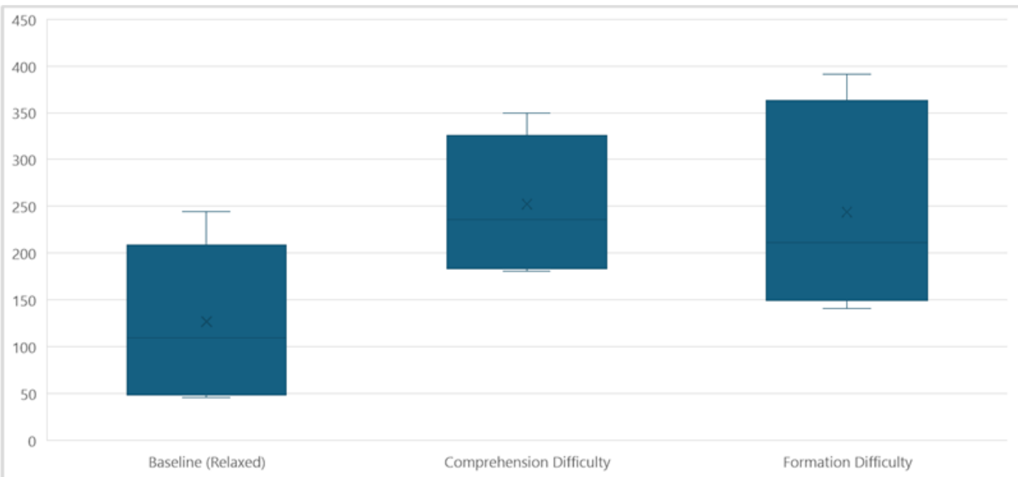

Figure 7.3: Average PSD in Theta Band **Frontal** Channels

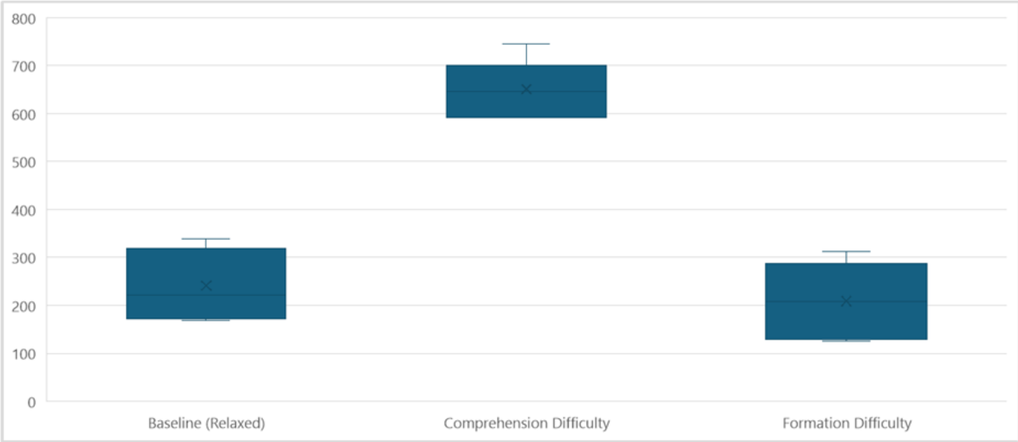

Figure 7.4: Average PSD in Theta Band **Temporal** Channels

State 1: Difficulty in Comprehension Unlike the alpha rhythm, the theta rhythm does not exhibit substantial variations when averaged across all brain regions, especially in the frontal channels. This localized increase in theta power within the temporal regions underscores their critical role in language comprehension. The temporal lobes are particularly engaged in semantic processing, memory retrieval, and the integration of linguistic information, processes that are fundamental to overcoming comprehension challenges.

State 2: Difficulty in Formation In the response formation state (State 2), it is observed that the differences in theta power compared to the baseline state (State 0) are minimal, both in Frontal as well as Temporal Channels. However, when analyzing (P2 to P13) the power spectral density (PSD) values display significant variance across participants, which limits the ability to derive statistically robust conclusions. This variability likely arises from the diverse challenges participants encounter during response formation. Examples include difficulty in recalling specific words, constructing grammatically correct sentences, hesitations due to uncertainty in linguistic expression, or issues with spelling. Such granular factors could potentially introduce considerable variability in the PSD values. Existing literature on this subject demonstrates that as task complexity increases, theta power rises accordingly, reflecting the brain's increased effort to process and retain information [61]. However, this is not apparent in the current research for theta bands.

7.2.2 Alpha Band Analysis

Alpha power reduction is widely regarded as an indicator of increased cortical activation and attentional engagement. Alpha rhythms (8–12 Hz) are dominant during relaxed states but tend to decrease when the brain transitions to active cognitive processing [62]. This phenomenon reflects a suppression of idling neural activity as cognitive resources are allocated to task-relevant processing [63]. In this study, the "Difficulty in Comprehension" condition (State 1) exhibits a notable decrease in alpha power across almost all channels (Figure 7.28). This reduction aligns with existing findings that suggest alpha desynchronization occurs during tasks requiring significant cognitive effort, particularly during the processing of external sensory inputs and attentional focus [64]. Conversely, "Difficulty in Formation" (State 2) shows a relatively limited reduction in alpha power, particularly in the frontal regions, indicating differences in neural engagement between the two states.

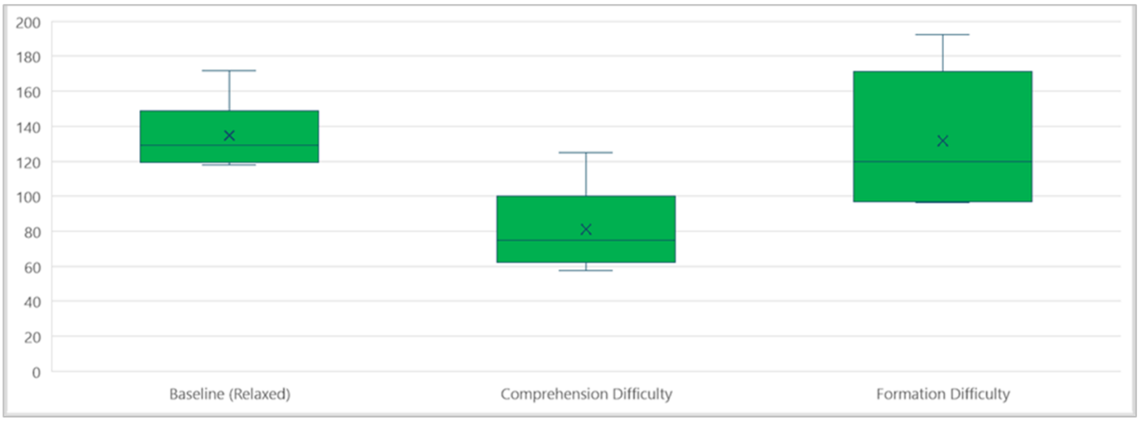

Figure 7.5: Average PSD in Alpha Band **All** Channels

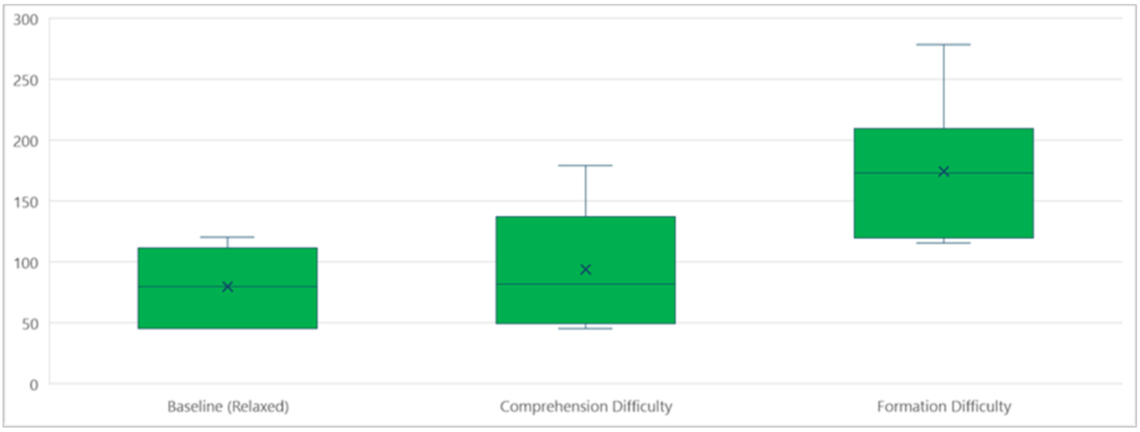

Figure 7.6: Average PSD in Alpha Band **Frontal** Channels

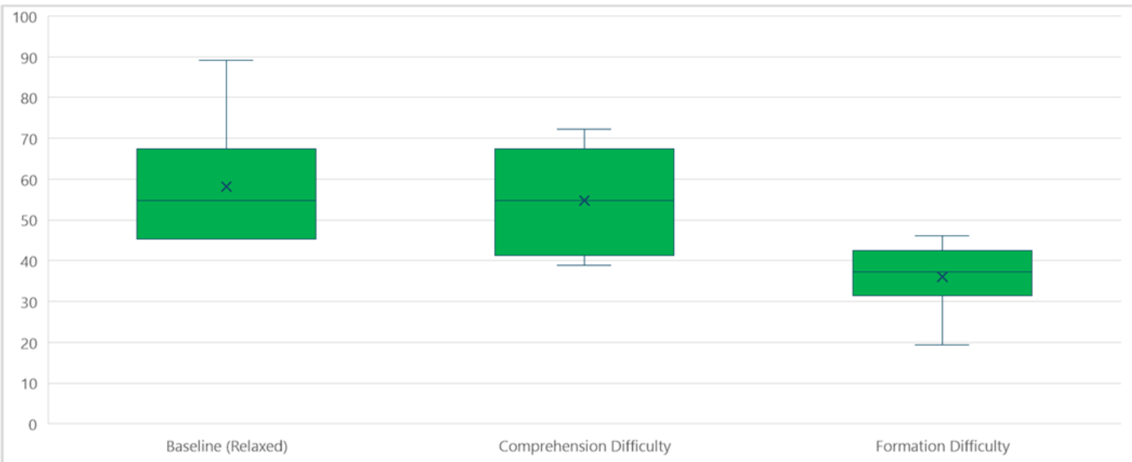

Figure 7.7: Average PSD in Alpha Band **Temporal** Channels

State 1: Difficulty in Comprehension Bottom-up processing involves the brain's response to external stimuli, requiring the allocation of resources to interpret incoming information, particularly when comprehension fails. The temporal regions, known for their role in semantic and auditory processing [6], display prominent alpha suppression. This finding aligns with previous research indicating that reduced alpha activity corresponds to increased neural engagement for semantic integration and attention [65].

The alpha desynchronization in comprehension challenges suggests that participants struggled to process external stimuli, requiring significant cognitive resources to attempt resolution. This effect is consistent with Klimesch's "functional inhibition hypothesis," where alpha reduction reflects the release of inhibition in sensory networks to facilitate active processing [65].

State 2: Difficulty in Formation In contrast, during response formation difficulties, alpha power reduction is less pronounced, particularly in the frontal channels. This observation may reflect differences in the nature of cognitive demands. Unlike comprehension tasks, which involve bottom-up processing, response formation relies heavily on top-down processes, including internal planning, word retrieval, and syntactic structuring [66].

The frontal regions show limited alpha suppression, indicating continued engagement of executive functions such as working memory and linguistic planning [67]. Additionally, the temporal channels may show residual alpha power, particularly in the right hemisphere, which is often associated with non-verbal and creative problem-solving [68]. This suggests that participants encountering formation difficulties may rely on alternative cognitive strategies, such as searching for semantic cues or constructing a mental framework for response generation. Hence, tracking alpha activity for the rest of sample population in the temporal channel could be worthwhile.

7.2.3 Beta Band Analysis

Reduced Beta Power: General Reduction Across Channels Beta rhythms (13–30 Hz) are primarily associated with motor activity, sensorimotor processing [69]. A reduction in beta power is often linked to the suppression of motor-related processing and an increase in cognitive load. This suppression signifies that, in cognitively demanding scenarios such as comprehension and formation difficulties, the brain prioritizes cognitive and attentional processes over motor-related functions [70, 71]. As expected, the average beta PSD values (across all channels) are significantly lower than corresponding theta and alpha PSD values (across all channels). The boxplots in Figures 7.8 illustrate the average power spectral density (PSD) for beta activity across all channels, and 7.9, 7.10 focus specifically on Alpha oscillations in the frontal and temporal regions.

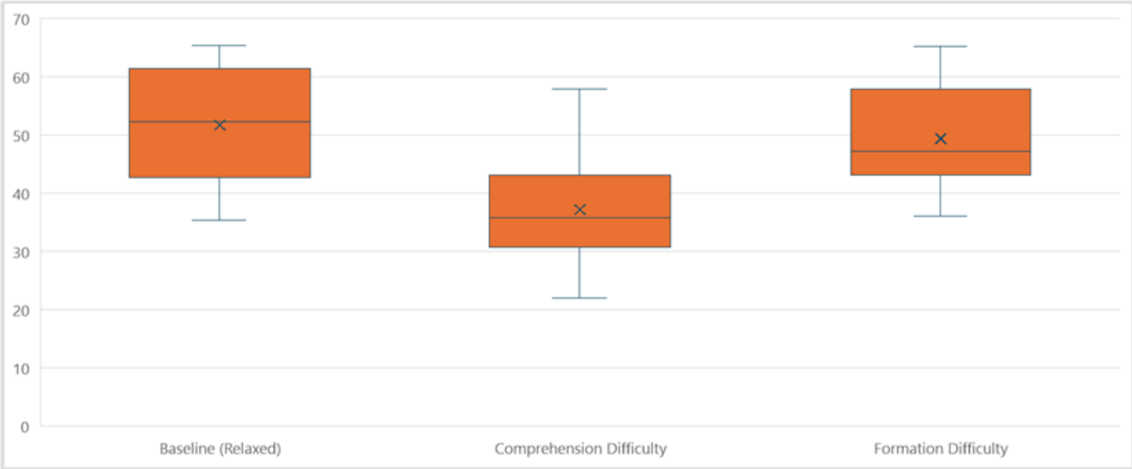

Figure 7.8: Average PSD in Beta Band **All** Channels

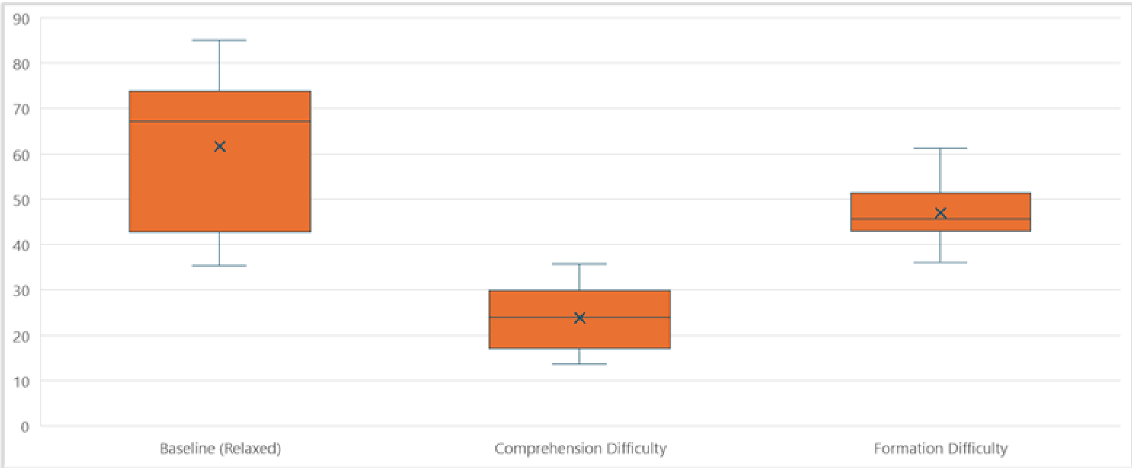

Figure 7.9: Average PSD in Beta Band **Frontal** Channels

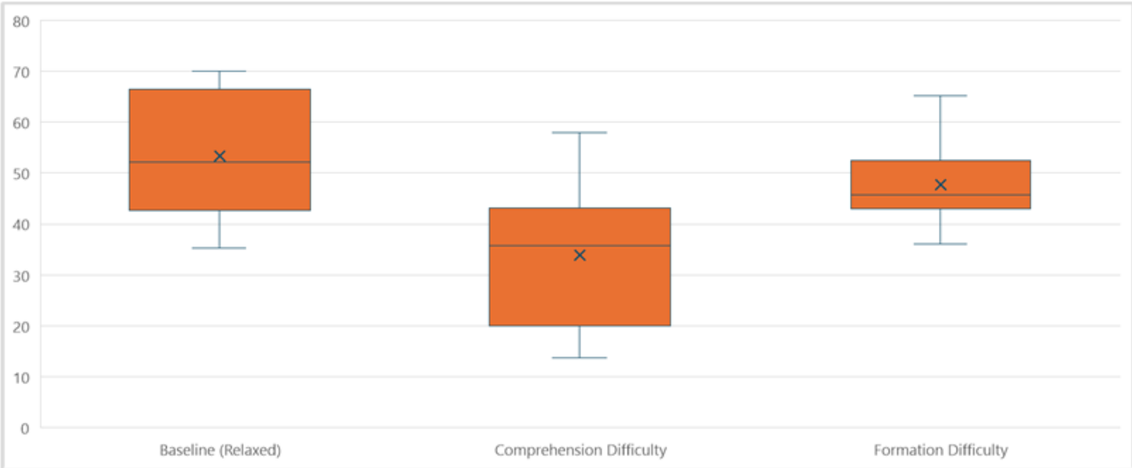

Figure 7.10: Average PSD in Beta Band **Temporal** Channels

State 1: Difficulty in Comprehension During comprehension difficulties, a slight reduction in beta power is observed across all channels (Figure 7.8). This reduction is particularly prominent in the frontal channels (Figure **??**). The frontal suppression highlights reduced motor-related engagement and increased reliance on higher-order executive functions for semantic integration and attention regulation [72]. As comprehension tasks place a heavier load on bottom-up processing of external inputs, beta suppression reflects the brain's shift from motor networks to cognitive processing pathways.

The temporal channels, which are critical for auditory and verbal processing and semantic comprehension, show a moderate reduction in beta power (Figure 7.10). This aligns with prior studies indicating beta suppression during auditory language tasks requiring attentional focus [73]. Overall, beta suppression during comprehension difficulties demonstrates the brain's effort to allocate resources to semantic and linguistic interpretation while minimizing unnecessary sensorimotor engagement.

State 2: Difficulty in Formation During response formation difficulties, beta suppression is less pronounced across channels compared to comprehension tasks, particularly in the frontal regions (Figure 7.9). The relatively limited reduction in frontal beta power suggests that motor-related processes—such as word retrieval, sentence planning, and the preparation for articulation—remain partially active. This reflects the brain's concurrent reliance on cognitive and sensorimotor pathways during the construction of linguistic responses [**?**]. Contrary to expectations, temporal regions (Figure 7.10) do not show significant changes in beta power. The temporal lobes, particularly regions linked to Wernicke's area in the posterior superior temporal gyrus, are typically engaged in semantic retrieval [6]. However, the absence of substantial beta suppression here may indicate that the task demands of response formation rely more heavily on executive control and motor processes managed by the frontal regions. This pattern aligns with findings that beta activity in temporal regions may be less sensitive to motor planning tasks compared to cognitive challenges involved in comprehension [73].

This observed trend is consistent with the results from the topographical analysis discussed in the previous section. Beta band suppression is often region-specific, and while comprehension tasks activate posterior temporal areas to resolve meaning from external stimuli, response formation tasks prioritize the integration of motor planning processes in the frontal cortex [74]. This distinction highlights the brain's task-specific neural resource allocation as it switches from semantic processing during comprehension difficulties to executive and motor engagement during response formation challenges.

7.2.4 Development of Neural Markers

The development of quantitative neural indices provides a systematic and objective method for detecting and differentiating cognitive challenges (State 1 and State 2) as compared to baseline (State 0). Based on the previous sections, where power spectral density (PSD) changes in theta, beta, and alpha bands were analyzed, two novel indices are proposed: the Inhibited Language Comprehension Index (ILCI) for State

1 (comprehension difficulty) and the Inhibited Language Formation Index (ILFI) for State 2 (formation difficulty). These indices aim to address Research Question 1 (RQ1) by providing neural markers for comprehension and response formulation challenges, enabling a possibility to classify the two states and interpretation of EEG signals.

Neural Index for Quantifying Comprehension Difficulty

The "Difficulty in Comprehension" condition (State 1) is characterized by increased theta power in the frontal regions and reduced beta power in the frontal regions. Based on the observations from literature and the analysis performed above, the Inhibited Comprehension Index (ILCI) is proposed to quantify comprehension challenges. The ILCI is defined as the ratio of theta power in the temporal channels to the beta power in the frontal channels:

$$\textbf{Inhibited Language Comprehension Index (ILCI)} = \frac{\text{Theta}_{\text{Temporal}}}{\text{Beta}_{\text{Frontal}}} \tag{7.1}$$

A higher ILCI value indicates greater cognitive strain and comprehension impairment, as theta increases and beta decreases. This metric effectively captures the shift from motor processing to cognitive processing that occurs during comprehension difficulties. The temporal dominance of theta power underscores the role of semantic retrieval and integration, while the frontal beta suppression reflects reduced motor-related engagement, aligning with findings from previous analyses.

Neural Index for Quantifying Formation Difficulty

The "Difficulty in Formation" condition (State 2) presents a distinct neural signature characterized by moderate frontal theta power indicating cognitive strain and active executive processes and elevated alpha power in the temporal regions, particularly in response to linguistic planning and retrieval processes. The presence of frontal theta power suggests continued reliance on executive networks for top-down processes such as word retrieval and motor planning [66, 74]. Simultaneously, alpha power in the temporal regions highlights semantic processing demands and the brain's attempt to resolve linguistic construction challenges [65, 73].

To quantify response formation difficulties, the Inhibited Response Formation Index (ILFI) is proposed. The ILFI is defined as the sum of theta power in the frontal channels and alpha power in the temporal channels:

$$\textbf{Inhibited Language Formation Index (ILFI)} = \frac{\text{Theta}_{\text{Frontal}}}{\text{Alpha}_{\text{Temporal}}} \tag{7.2}$$

A higher ILFI value corresponds to greater cognitive effort and motor engagement during response generation. The inclusion of temporal alpha power aligns with evidence showing that alpha activity reflects the cognitive inhibition of competing processes and semantic reconstruction during language formation tasks [62, 65]. Simultaneously, the frontal theta component reflects ongoing top-down executive control required for linguistic planning and motor preparation.This index captures the dual involvement of cognitive and semantic processes during formation difficulties. It offers a quantitative neural marker for differentiating State 2 from both the baseline (relaxed) state and comprehension difficulties (State 1).

7.2.5 Trend Analysis of the Proposed Neural Indices

In this section, the validity of the proposed neural indices, namely the **Inhibited Comprehension Index (ILCI)** and the **Inhibited Language Formation Index (ILFI)**, is assessed across participants P2 to P13. The indices are evaluated using EEG signals recorded from these participants and compared to the baseline conditions. The analysis examines the robustness of these indices as neural markers for cognitive difficulty states, using basic statistical tests.

How does ILCI Vary Across Participants? The analysis begins by evaluating the components of the ILCI across participants. Specifically, the **theta power in temporal regions** and the **beta power in frontal regions** are compared to their baseline counterparts.

Figure 7.11 illustrates the temporal theta component during the comprehension difficulty condition (State 1) for all participants. For most participants (except P5 and P7), the theta power in the temporal regions is **equal to or greater than baseline**. This suggests that increased theta activity is a consistent marker of comprehension challenges. However, the variance in theta power across participants is relatively high. This variability may arise due to individual differences or confounding variables, such as varying levels of task engagement, fatigue, or external distractions. These confounding factors will be further addressed in the discussion section.

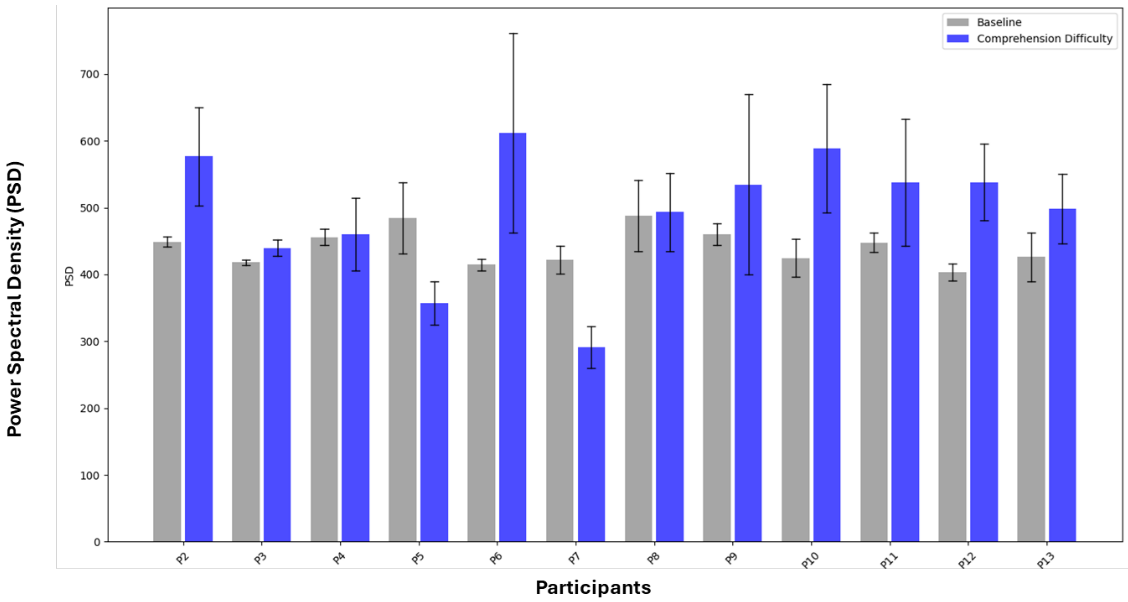

Figure 7.11: Theta Power in Temporal Regions for All Participants During Comprehension Difficulty

When analyzing the beta component in the frontal regions, as shown in Figure 7.12, the results indicate a **reduction in beta power** compared to the baseline for most participants (except P6, P11, P12, and P13). This supports the hypothesis that beta suppression in frontal channels could be indicative of cognitive reallocation away from motor-related processes to support semantic and attentional demands during comprehension difficulties.

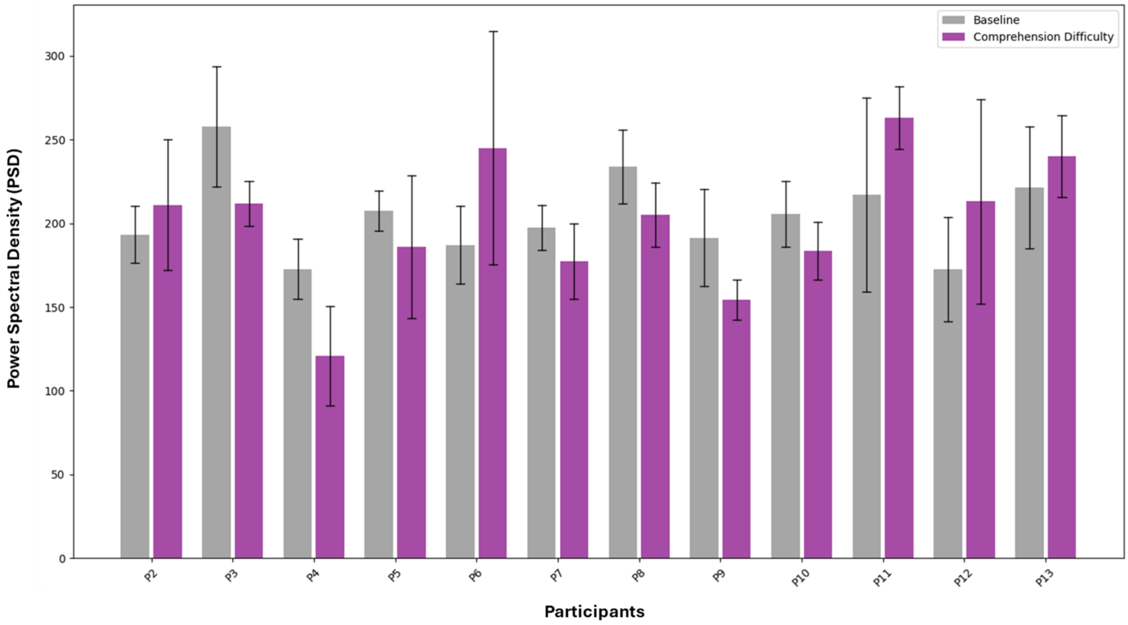

Figure 7.12: Beta Power in Frontal Regions for All Participants During Comprehension Difficulty

To further validate the consistency of these observations, a direct comparison of the theta temporal and beta frontal PSD components across all participants is shown in Figure 7.13. This figure highlights the relative increase in theta power and the suppression of beta power, reinforcing the relevance of these components for quantifying comprehension challenges.

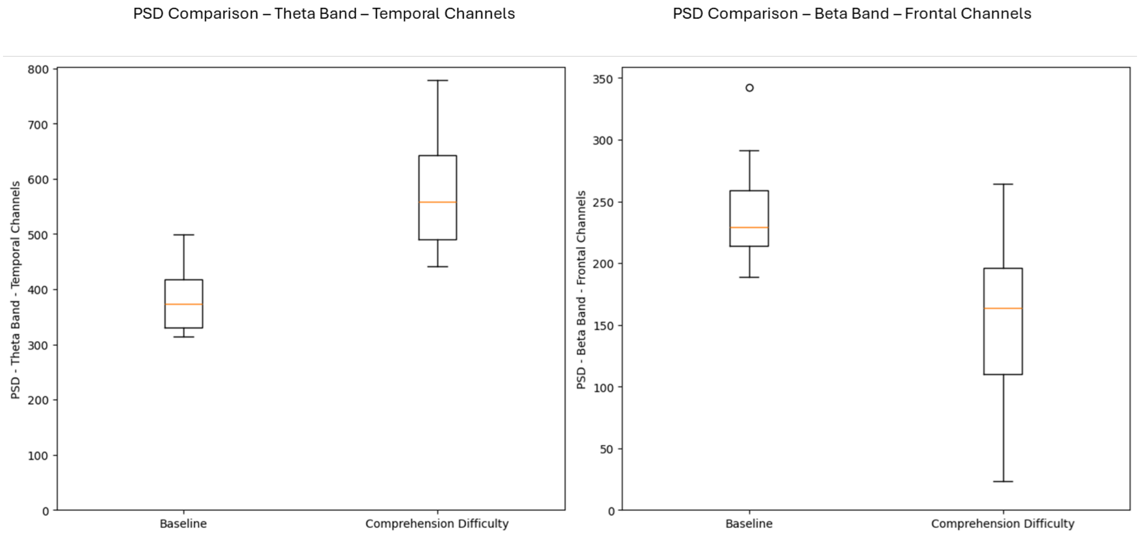

Figure 7.13: PSD Comparison of Theta Temporal and Beta Frontal Components During Comprehension Difficulty (P2 to P13)

How does ILCI Vary Across Conditions?

ILCI Validation for State 1: Comprehension Difficulty

The Inhibited Comprehension Index (ILCI) was computed for all participants during comprehension difficulty (State 1). Figure 7.14 presents the Z-scores of ILCI values relative to the baseline. For 9 out of 12 participants, the ILCI values during comprehension difficulty are at least 1.09 standard deviations (SD) higher than the baseline ILCI, with an average Z-score of 1.09. This signals that this marker could potentially be explored further as a neural signature of comprehension difficulties.

Participant	Baseline ILCI	Condition ILCI	Z-Score	Condition ILCI > Baseline
P2	2.34	2.77	1.82	TRUE
P3	1.64	2.08	2.22	TRUE
P4	2.66	3.92	5.85	TRUE
P5	2.33	1.97	-1.84	FALSE
P6	2.24	2.61	1.61	TRUE
P7	2.14	1.65	-8.77	FALSE
P8	2.10	2.42	1.01	TRUE
P9	2.45	3.50	2.47	TRUE
P10	2.07	3.26	7.42	TRUE
P11	2.17	2.04	-0.23	FALSE
P12	2.39	2.65	0.63	TRUE
P13	1.94	2.09	0.93	TRUE

Figure 7.14: Z-Scores of ILCI for All Participants During Comprehension Difficulty

From 7.14, it is evident that there is a possibility of the Condition ILCI (State 1) being higher than the Baseline ILCI (State 0). To investigate this further, we first examine the Q-Q plot to visually inspect the sample distribution for normality.

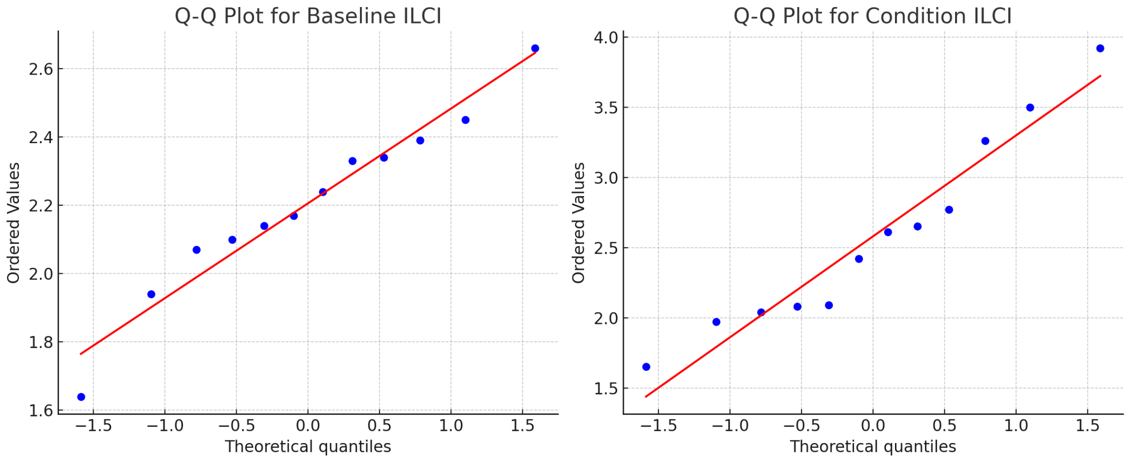

Figure 7.15: ILCI - Comprehension Difficulty (State 1)

The Q-Q plot suggests that the data appears to be normally distributed. To statistically validate this observation, the Shapiro-Wilk test is conducted with the following hypotheses:

- $\mathbf{H}_0$: The data follows a normal distribution.
- $\mathbf{H}_1$: The data does not follow a normal distribution.

The results of the Shapiro-Wilk test for both State 0 and State 1 are as follows: **Baseline (State 0 ILCI):** Test Statistic: 0.9745; p-value: 0.9516 (Since $p > 0.05$, the data is likely normal) **Condition (State 1 ILCI):** Test Statistic: 0.9365; p-value: 0.4544 (Since $p > 0.05$, the data is likely normal). These results suggest that the differences in ILCI values in State 0 and State 1 may follow a normal distribution. However, it is crucial to note that failing to reject the null hypothesis ($p > 0.05$) does not confirm normality; it only indicates insufficient evidence to reject the assumption of normality. To further analyze the data, both parametric and non-parametric tests were performed.

Parametric Test: Paired t-Test T-Statistic: -2.2911; p-value: 0.0427. Since the p-value is less than 0.05, the null hypothesis is rejected, indicating that the ILCI values during comprehension difficulty (State 1) are significantly higher than those observed during the baseline condition (State 0). The negative T-statistic suggests that the baseline values are generally lower than the condition values.

Non-parametric Test: Wilcoxon Signed-Rank Test Wilcoxon Statistic: 15.0; p-value: 0.0640. Since the p-value (0.0640) is greater than 0.05, we fail to reject the null hypothesis. This suggests that there is no statistically significant difference between the baseline and condition ILCI values based on the Wilcoxon test. While the paired t-test indicated significance, the Wilcoxon test, which is more robust to non-normal distributions, does not confirm a significant effect at the 5% significance level.

The results indicate that although there may be a trend, the evidence is not strong enough to conclude a significant difference without assuming normality.

—

ILCI Validation for State 2: Formation Difficulty

The ILCI was also applied to the formation difficulty condition (State 2) to determine whether it accurately identifies neural changes during this state. Figure 7.16 shows the Z-scores of ILCI relative to the baseline. In contrast to comprehension difficulty, the results for State 2 show that the ILCI fails to differentiate the formation difficulty condition from the baseline for **6 out of 12 participants**, with an average Z-score of **-0.94**. Even among participants where the condition holds, the Z-scores are close to baseline values.

Participant	Baseline ILCI	Condition ILCI	Z-Score	Condition ILCI > Baseline
P2	2.34	2.53	0.81	TRUE
P3	1.64	1.75	0.57	TRUE
P4	2.66	1.97	-3.15	FALSE
P5	2.33	2.23	-0.50	FALSE
P6	2.24	2.32	0.36	TRUE
P7	2.14	1.83	-5.52	FALSE
P8	2.10	2.03	-0.23	FALSE
P9	2.45	2.63	0.42	TRUE
P10	2.07	1.65	-2.63	FALSE
P11	2.17	2.26	0.16	TRUE
P12	2.39	2.69	0.72	TRUE
P13	1.94	1.56	-2.36	FALSE

Figure 7.16: Z-Scores of ILCI for All Participants During Formation Difficulty

The 7.16 shows that there is no a significant difference between the Baseline (State 0) and Condition (State 2) in this case. The Q-Q plots 7.17 for both baseline and condition ILCI suggest that the data points largely follow a normal distribution, with some minor deviations.

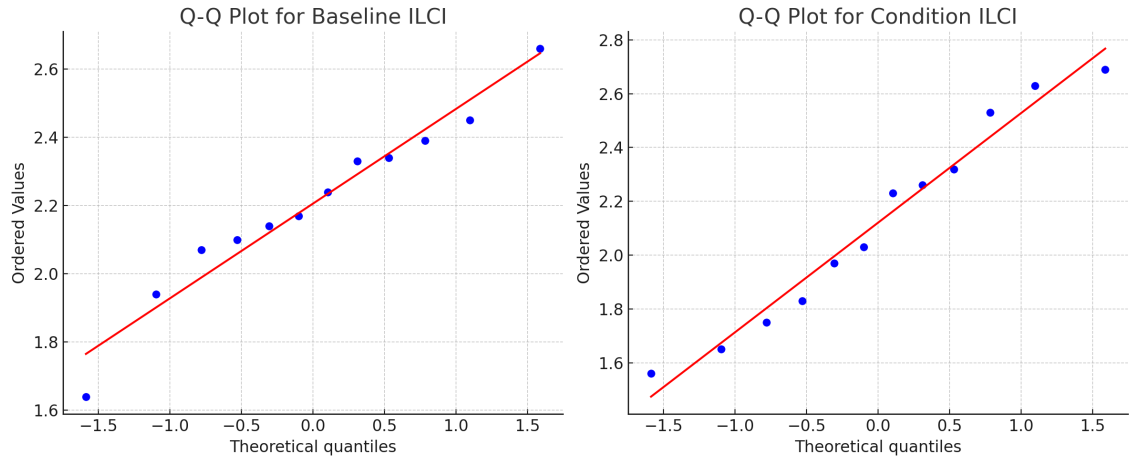

Figure 7.17: ILCI - Formation Difficulty (State 2)

The results of the Shapiro-Wilk test for both State 0 and State 2 are as follows: **Baseline (State 0) ILCI:** Test Statistic: 0.9745; p-value: 0.9516 ($p > 0.05$, fail to reject H_0, data appears normally distributed). **Condition (State 2) ILCI:** Test Statistic: 0.9517; p-value: 0.6621 ($p > 0.05$, fail to reject H_0, data appears normally distributed). Since both p-values are greater than 0.05, there is no strong evidence to reject normality, indicating that the data can be considered normally distributed.

Parametric Test : Paired t-Test T-Statistic: 0.9735; p-value: 0.3512 ($p > 0.05$, fail to reject H_0) The paired t-test suggests that there is no statistically significant difference between the baseline and condition ILCI values.

Non-Parametric Test: Wilcoxon Signed-Rank Test Wilcoxon Statistic: 31.0; p-value: 0.5693 ($p > 0.05$, fail to reject H_0). The Wilcoxon test also indicates no statistically significant difference between the baseline and condition values, reinforcing the conclusion from the paired t-test.

Both the parametric and non-parametric tests indicate that the difference between the baseline and condition ILCI values is not statistically significant ($p > 0.05$). This implies that any observed differences in the sample may be due to chance rather than a systematic effect.

How does ILFI Vary Across Participants? The analysis begins by evaluating the components of the ILFI across participants P2 to P13. Specifically, the theta power in frontal regions and the alpha power in temporal regions are compared to their baseline counterparts for State 2 (Formation Difficulty). Figure 7.18 illustrates the temporal theta component during the comprehension difficulty condition (State 1) for all participants. For most participants (except P2, P6, P13), the theta power in the frontal regions is equal to or greater than baseline. This suggests that increased theta activity in frontal regions is a consistent marker of formation challenges. However, substantial variability is evident in theta power across participants, with P6 exhibiting particularly high variance. Such inconsistencies can likely be attributed to factors such as individual variability, external disturbances during data collection, participant fatigue or reduced focus, which can impact neural engagement.

When analyzing the alpha component in the temporal regions, as shown in Figure 7.19, the results indicate a reduction in power compared to the baseline for 7 out of 12 participants (except P2, P6, P11, P12 and P13). The variance across participants are also high. Hence, this component might not accurately represent and distinguish formation difficulty (State 2). The reason for this unclear and needs further investigation to identify if there is an underlying patterns so there is presence of confounding variables at play.

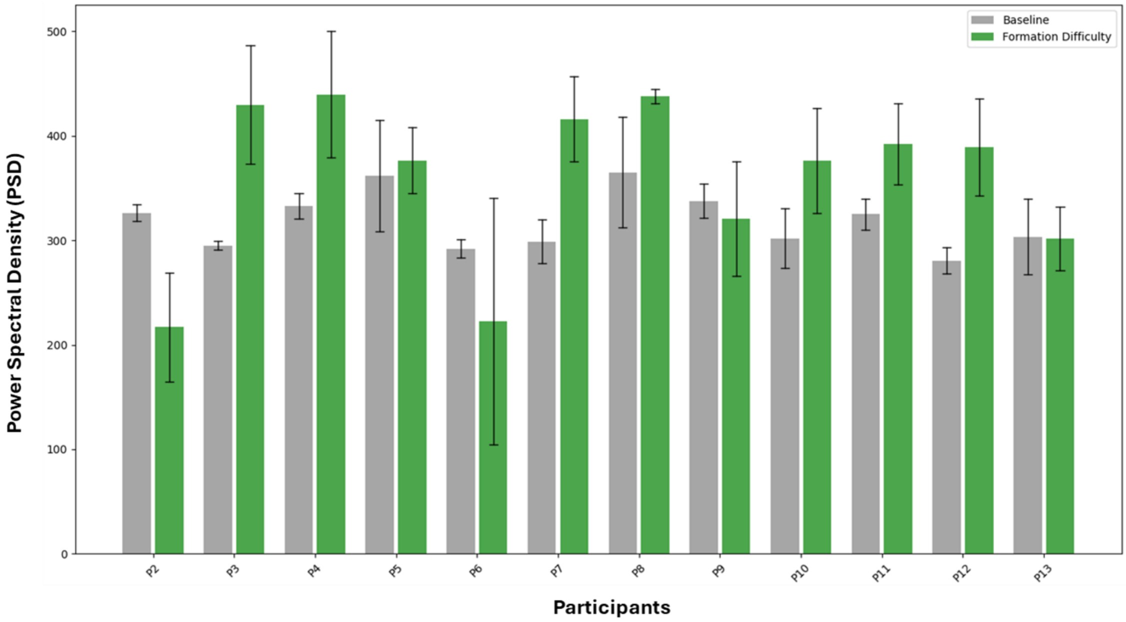

Figure 7.18: Theta Power in Frontal Regions for All Participants During Formation Difficulty

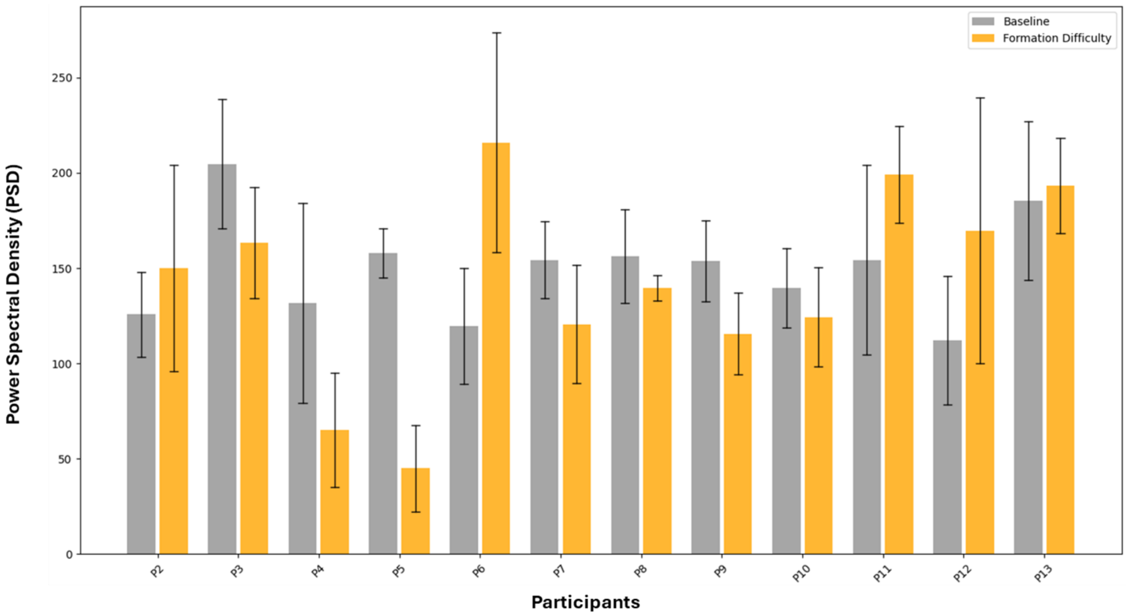

Figure 7.19: Alpha Power in Temporal Regions for All Participants During Formation Difficulty

To further validate the consistency of these observations, a direct comparison of the theta frontal and alpha temporal components across all participants is shown in 7.20. This figure highlights the relative difference in theta power and the suppression (albeit not a lot) of alpha power.

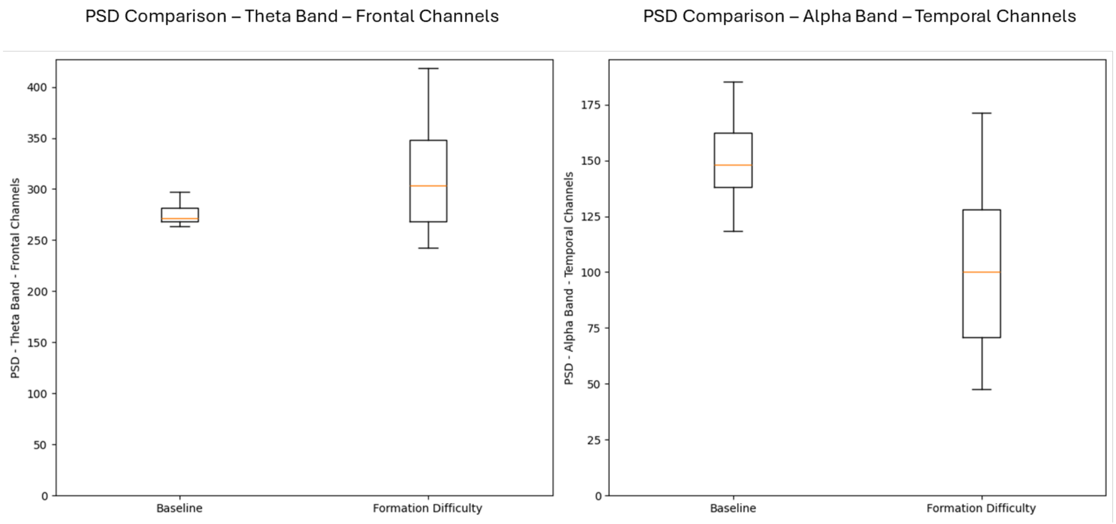

Figure 7.20: PSD Comparison of Theta Frontal and Alpha Temporal Components During Comprehension Difficulty (P2 to P13)

How does ILFI Vary Across Conditions?

ILFI Validation for State 1: Comprehension Difficulty

The ILFI was first tested for comprehension difficulties (State 1) to determine if it could identify neural signatures associated with this condition. Figure 7.21 presents the Z-scores of ILFI relative to the baseline for all participants. Unlike the ILCI, the ILFI does not exhibit significant deviations from baseline values.

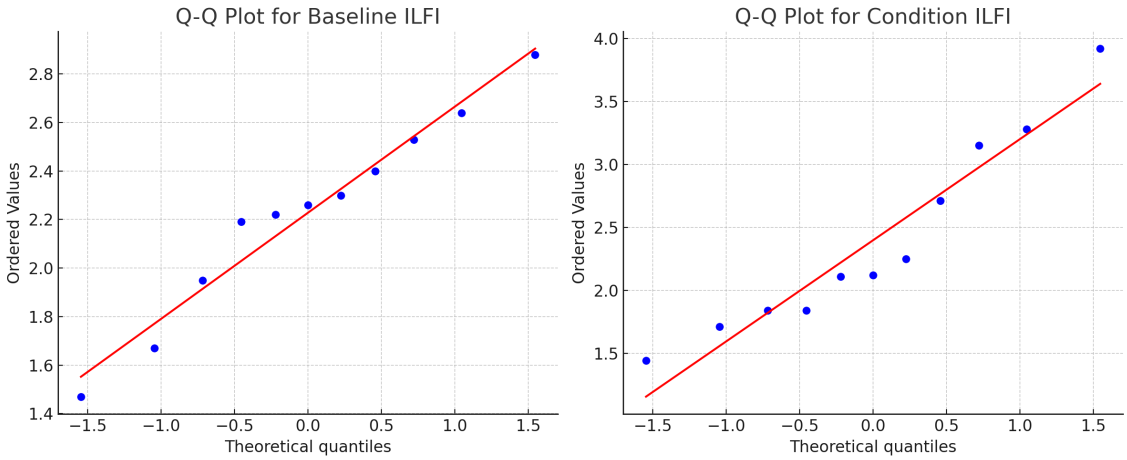

Figure 7.22: QQ Plot - ILFI - State 0 and State 1

Participant	Baseline ILFI	Condition ILFI	Z-Score	Condition ILFI > Baseline
P2	2.64	2.25	-1.01	FALSE
P3	1.47	1.71	1.02	TRUE
P4	2.88	3.92	0.75	TRUE
P5	2.30	2.12	-0.44	FALSE
P6	2.53	1.84	-1.28	FALSE
P7	1.95	1.84	-0.78	FALSE
P8	2.40	2.11	-0.44	FALSE
P9	2.22	3.15	3.81	TRUE
P10	2.19	1.44	-2.39	FALSE
P11	2.26	2.71	0.63	TRUE
P12	2.67	3.28	0.71	TRUE
P13	1.67	1.39	-1.23	FALSE

Figure 7.21: Z-Scores of ILFI for All Participants During Comprehension Difficulty

The Q-Q plots for both baseline and condition ILFI data suggest that the data points largely follow a normal distribution with minor deviations.

The results of the Shapiro-Wilk test for both State 0 and State 1 are as follows: **Baseline (State 0) ILFI:** Test Statistic: 0.9683; p-value: 0.8692 ($p > 0.05$, fail to reject H_0, data appears normally distributed) **Condition (State 1) ILFI :** Test Statistic: 0.9196; p-value: 0.3156 ($p > 0.05$, fail to reject H_0, data appears normally distributed)

Parametric Test : Paired t-Test T-Statistic: -0.7362; p-value: 0.4785 ($p > 0.05$, fail to reject H_0) The paired t-test suggests that there is no statistically significant difference between the baseline and condition ILFI values.

Non-Parametric Test: Wilcoxon Signed-Rank Test Wilcoxon Statistic: 27.0; p-value: 0.6377 ($p > 0.05$, fail to reject H_0). The Wilcoxon test also indicates no statistically significant difference between the baseline and condition values, reinforcing the conclusion from the paired t-test.

The parametric (paired t-test) and non-parametric (Wilcoxon test) tests indicated that the difference between the baseline and condition ILFI values is not statistically significant ($p > 0.05$). This implies that any observed differences in the sample may be due to chance rather than a systematic effect.

—

ILFI Validation for State 2: Formation Difficulty

The ILFI was subsequently applied to the formation difficulty condition (State 2). Figure 7.23 shows the Z-scores of ILFI relative to baseline across all participants. While there is a slight upward trend for certain participants, the results remain inconsistent.

Participant	Baseline ILFI	Condition ILFI	Z-Score	Condition ILFI > Baseline
P2	2.64	1.37	-3.25	FALSE
P3	1.47	2.03	2.39	TRUE
P4	2.88	2.80	-0.06	FALSE
P5	2.30	2.98	1.69	TRUE
P6	2.53	1.26	-2.35	FALSE
P7	1.95	2.47	3.63	TRUE
P8	2.40	3.42	1.55	TRUE
P9	2.22	1.21	-4.15	FALSE
P10	2.19	3.41	3.93	TRUE
P11	2.26	2.71	0.62	TRUE
P12	2.67	2.66	-0.02	FALSE
P13	1.67	1.75	0.38	TRUE

Figure 7.23: Z-Scores of ILFI for All Participants During Formation Difficulty

The Q-Q plots for both baseline and condition (State 2) ILFI data suggest that the data points largely follow a normal distribution with minor deviations.

The results of the Shapiro-Wilk test for both State 0 and State 2 are as follows: **Baseline (State 0) ILFI:** Test Statistic: 0.9683; p-value: 0.8692 ($p > 0.05$, fail to reject H_0, data appears normally distributed) **Condition (State 2) ILFI :** Test Statistic: 0.8978; p-value: 0.1738 ($p > 0.05$, fail to reject H_0, data appears normally distributed)

Parametric Test : Paired t-Test T-Statistic: -0.5847; p-value: 0.5717 ($p > 0.05$, fail to reject H_0) The paired t-test suggests no statistically significant difference between the baseline and condition (State 2) ILFI values.

Non-Parametric Test: Wilcoxon Signed-Rank Test Wilcoxon Statistic: 29.0; p-value: 0.7646 ($p > 0.05$, fail to reject H_0). The Wilcoxon test also suggests no statistically significant difference between the baseline and condition ILFI values, reinforcing the conclusion from the paired t-test. Both the parametric (paired t-test) and non-parametric (Wilcoxon test) tests indicated that the difference between the

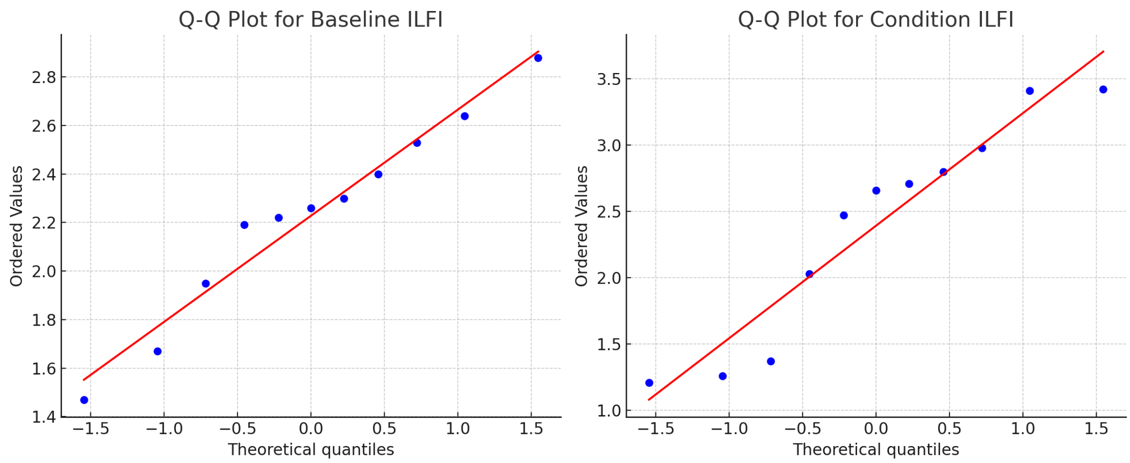

Figure 7.24: QQ Plot - ILFI - State 0 and State 2

baseline and condition ILFI values is not statistically significant ($p > 0.05$). This implies that any observed differences in the sample may be due to chance rather than a systematic effect.

—

7.3 Connectivity Analysis

The connectivity analysis investigates the functional and directional interactions between brain regions during two task conditions: Comprehension Difficulty (State 1) and Formation Difficulty (State 2). This analysis focuses on two primary metrics: Coherence (COH) and Phase Locking Value (PLV), which capture amplitude-phase coupling and phase synchrony, respectively. Additionally, the subsequent section will explore directional flow of information using Granger Causality (GC) and Time-Reversed Granger Causality to further support these findings.

7.3.1 Coherence and Phase Locking Value

Coherence quantifies the phase consistency between two EEG signals across a given frequency, providing a measure of functional synchronization and coupling. Higher coherence values indicate stronger inter-regional communication, combining both phase and amplitude information. In contrast, Phase Locking Value (PLV) exclusively measures the stability of phase synchrony, independent of amplitude dynamics. Together, these measures provide complementary insights into connectivity patterns.

Connectivity Patterns During Comprehension Difficulty (State 1)

The connectivity matrices for coherence and PLV during comprehension difficulty are shown in Figure 7.25. Both measures were computed across three frequency bands: theta (4–8 Hz), alpha (8–13 Hz), and beta (13–30 Hz). However, this analysis focuses solely on the theta band, as it revealed the most significant variance in coherence values, whereas coherence in the alpha band remained relatively uniform with no discernible patterns.

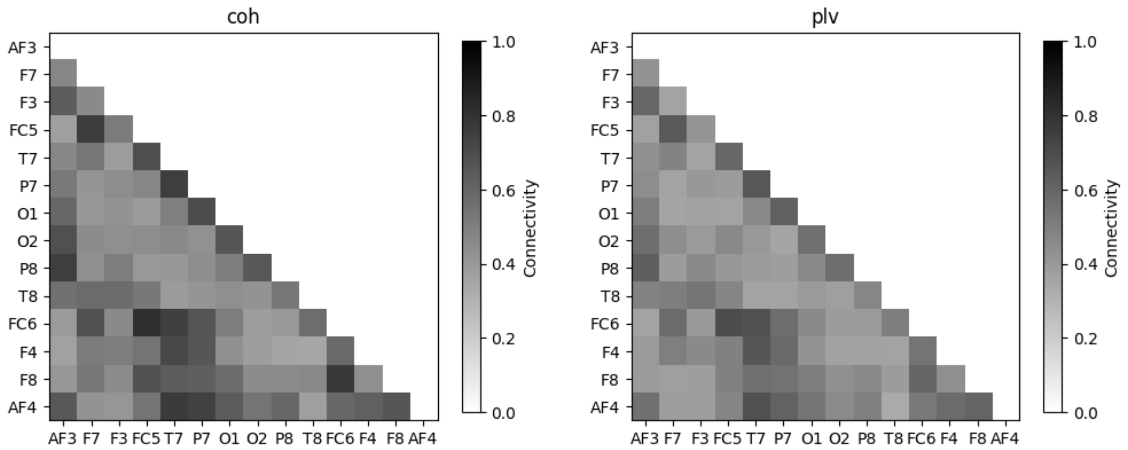

Figure 7.25: Comprehension Difficulty - Coherence and Phase Locking Value

The results indicate strong connectivity between the left temporal region (T7) and frontal regions, particularly FC6, F4, F8, and AF4. This observation supports the involvement of frontal-temporal networks, which are crucial for language comprehension, semantic integration, and attentional processes. Furthermore, high coherence between T7 (left temporal) and P7 (left parietal) highlights an interplay between temporal processing and parietal regions linked to working memory, aligning with previous literature on cognitive load [?, 73].

To determine if these trends persist across participants, coherence and PLV values were analyzed for participants P2 to P13. Figure 7.28 presents the results. Coherence values indicate that for 8 out of 12 participants, the temporal-frontal connectivity (T7 to frontal regions) remains strong, with COH values exceeding 0.5. This suggests that amplitude-phase coupling is a relatively consistent feature during comprehension difficulties. However, PLV values do not always align with the high coherence observed; only 5 out of 12 participants exhibited strong phase synchrony (PLV) along with high coherence values. This disparity indicates that while amplitude coupling is present, phase consistency may be more variable across participants. Therefore, while temporal-frontal coherence may hold promise as an identifying feature of comprehension difficulties, the inconsistent phase stability (PLV) reduces its robustness. This observation highlights the need for further investigation to determine the reliability of temporal-frontal synchronization as a marker for State 1.

	Comprehension Difficulty				(Highlighted based on the [COH - PLV <= 0.3] and [COH > 0.5])
Participant	Connectivity Feature	Avg Coherence (COH)	Avg Phase Lag Value (PLV)	Coherence >0.5	COH - PLV
P2	Temporal - Frontal Connectivity	0.96	0.62	TRUE	0.35
P3	Temporal - Frontal Connectivity	0.63	0.46	TRUE	0.17
P4	Temporal - Frontal Connectivity	0.78	0.64	TRUE	0.14
P5	Temporal - Frontal Connectivity	0.58	0.12	TRUE	0.46
P6	Temporal - Frontal Connectivity	0.67	0.92	TRUE	-0.26
P7	Temporal - Frontal Connectivity	0.19	0.18	FALSE	0.01
P8	Temporal - Frontal Connectivity	0.74	0.37	TRUE	0.38
P9	Temporal - Frontal Connectivity	0.48	0.12	FALSE	0.37
P10	Temporal - Frontal Connectivity	0.38	0.55	FALSE	-0.17
P11	Temporal - Frontal Connectivity	0.62	0.43	TRUE	0.19
P12	Temporal - Frontal Connectivity	0.90	0.23	TRUE	0.67
P13	Temporal - Frontal Connectivity	0.37	0.57	FALSE	-0.20

Figure 7.26: All Participants - Comprehension Difficulty - COH and PLV Analysis

Connectivity Patterns During Formation Difficulty (State 2)

The connectivity matrices for coherence and PLV during formation difficulty are presented in Figure 7.27. Compared to comprehension difficulty, State 2 exhibits an overall stronger coherence pattern across regions. This elevated connectivity is particularly notable in bilateral frontal and temporal regions, reflecting heightened synchronization across hemispheres.

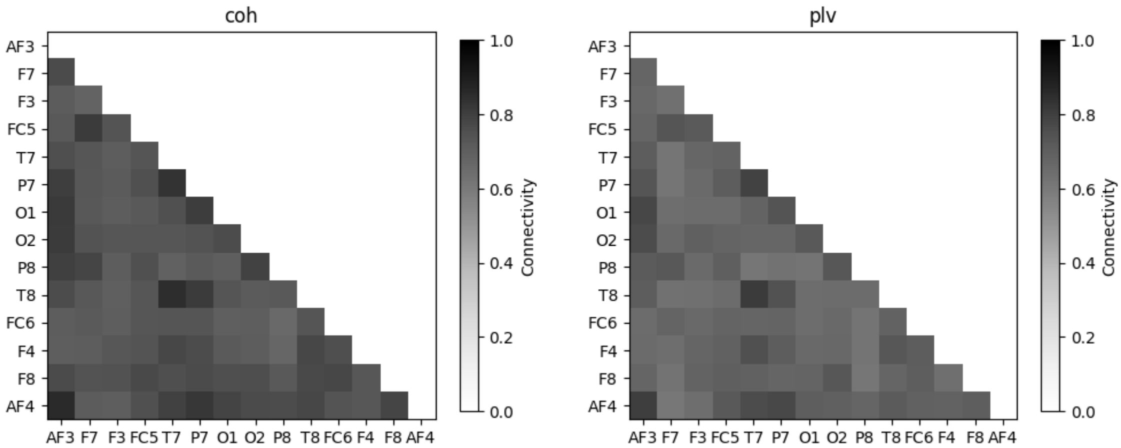

Figure 7.27: Formation Difficulty - Coherence and Phase Locking Value

Significant connectivity is observed between the left frontal (AF3) and left temporal (T7) regions, as well as between the right frontal (AF4) and right temporal (T8) regions. Additionally, strong bilateral connections emerge between AF3 (left frontal) and AF4 (right frontal) and between T7 (left temporal) and T8 (right temporal). These patterns likely reflect increased top-down executive control during response formation, involving cognitive processes such as lexical retrieval, motor planning, and inhibition of irrelevant responses [74]. This heightened bilateral connectivity may also indicate cognitive overload or neural recruitment as the brain attempts to resolve task difficulties.

To assess whether these findings generalize across participants, coherence and PLV values were averaged across the left (T7, AF3) vs right hemispheres (T8, AF4) for participants P2 to P13 (Figure 7.28). For 8 out of 12 participants, coherence values exceed 0.5, indicating consistent amplitude-phase coupling across hemispheres.

Furthermore, in 6 out of 12 participants, strong phase synchrony (PLV) is observed alongside high coherence, suggesting that both amplitude and phase consistency could contribute to bilateral frontal-temporal synchronization during formation difficulties. For the remaining participants, PLV values are lower despite high coherence, indicating variability in phase relationships. While bilateral connectivity appears to be a promising feature for identifying response formation challenges, further studies with larger sample sizes are needed to validate its robustness.

	Formation Difficulty				(Highlighted based on the [COH - PLV <= 0.3] and [COH > 0.5])
Participant	Connectivity Feature	Avg Coherence (COH)	Avg Phase Lag Value (PLV)	Coherence >0.5	COH - PLV
P2	Left - Right Hemisphere Connectivity	0.89	0.69	TRUE	0.20
P3	Left - Right Hemisphere Connectivity	0.68	0.64	TRUE	0.03
P4	Left - Right Hemisphere Connectivity	0.89	0.68	TRUE	0.22
P5	Left - Right Hemisphere Connectivity	0.77	0.33	TRUE	0.44
P6	Left - Right Hemisphere Connectivity	0.64	0.55	TRUE	0.09
P7	Left - Right Hemisphere Connectivity	0.49	0.86	FALSE	-0.37
P8	Left - Right Hemisphere Connectivity	0.75	0.60	TRUE	0.15
P9	Left - Right Hemisphere Connectivity	0.56	0.45	TRUE	0.11
P10	Left - Right Hemisphere Connectivity	0.88	0.60	TRUE	0.28
P11	Left - Right Hemisphere Connectivity	0.35	0.96	FALSE	-0.61
P12	Left - Right Hemisphere Connectivity	0.18	0.27	FALSE	-0.09
P13	Left - Right Hemisphere Connectivity	0.32	0.82	FALSE	-0.49

Figure 7.28: All Participants - Formation Difficulty - COH and PLV Analysis

7.3.2 Granger Causality

In this study, Granger Causality was applied to compare the directional interactions between frontal and temporal brain regions across the two task conditions: comprehension difficulty (State 1) and formation difficulty (State 2). By quantifying the flow of information (e.g., Frontal $\rightarrow$ Temporal vs. Temporal $\rightarrow$ Frontal), GC provides deeper insight into the neural mechanisms underlying language processing and response generation. For this analysis, the epochs associated with the two states (State 1 and State 2) were baseline-corrected (by subtracting the mean potential from each channel respectively) and used for further analysis.

Granger Causality for Comprehension Difficulty Figure 7.29 shows the GC analysis for State 1 (Comprehension difficulty) for both directions of information flow. At *theta-band* frequencies (4–8 Hz, shaded blue), the GC values for the frontal $\rightarrow$ temporal direction are notably higher than the reverse (temporal $\rightarrow$ frontal). This observation suggests that frontal regions exert significant control over temporal regions during cognitive processing associated with comprehension difficulties. Theta-band connectivity in the frontal regions has been widely linked to executive control and sustained attention. This result aligns with the increased cognitive demands of language comprehension, which require top-down monitoring and integration of incoming linguistic information.

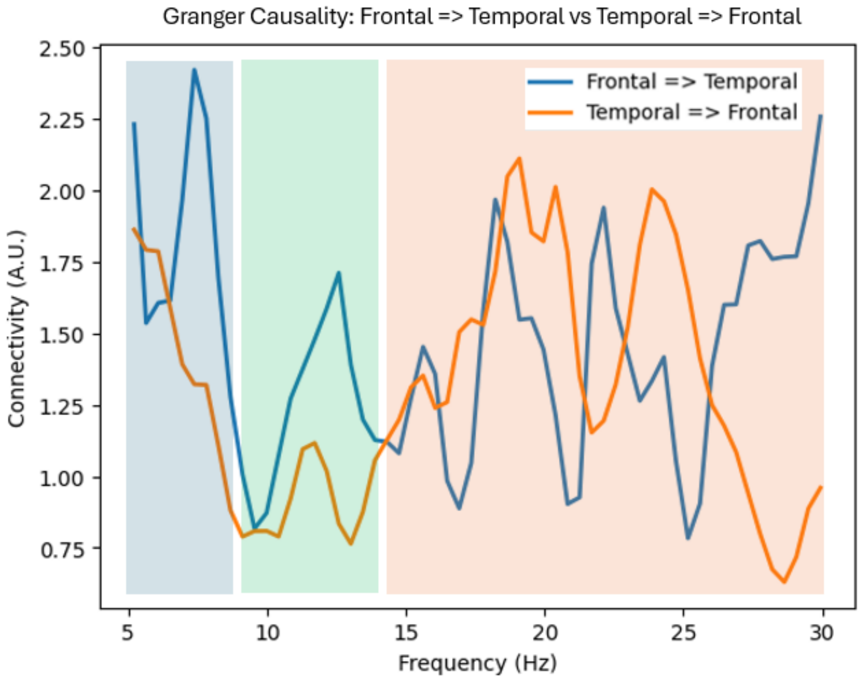

Figure 7.29: Granger Causality Analysis for Comprehension Difficulty: Frontal → Temporal vs Temporal → Frontal Connectivity

In the *alpha band* (8–13 Hz, shaded green), the directional influence stays similar however the connectivity (in A.U.) are lower than theta bands. In beta band (13 to 30 Hz), there is a shifts in connectivity where as some lower beta band (13 to 20 Hz) indicate a temporal dominance and higher beta band (25 to 30Hz) showing frontal dominance. This indicates an interaction between bottom-up sensory inputs and top-down cognitive regulation, reflecting the complexities of comprehension processing. At beta-band frequencies (13–30 Hz, shaded orange), the frontal → temporal connectivity shows renewed dominance, particularly in the higher beta range (20–25 Hz). This trend suggests that frontal regions may reassert top-down control during later stages of cognitive processing, which may involve managing increased working memory demands or resolving ambiguities associated with comprehension challenges [?]. Beta activity in frontal regions has previously been associated with cognitive regulation, motor preparation, and attentional modulation, all of which are relevant for comprehension tasks.

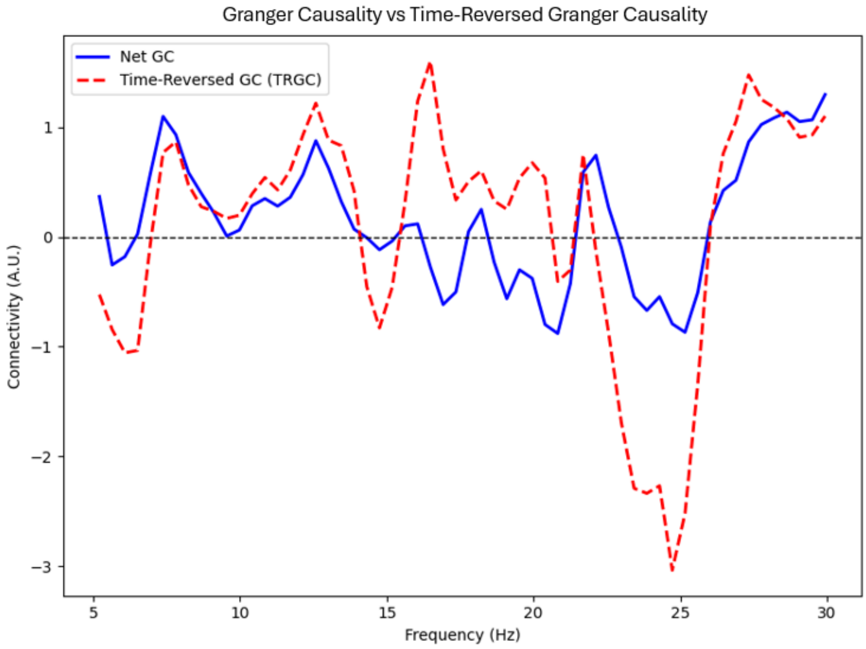

Figure 7.30: Comparison of Granger Causality (GC) and Time-Reversed Granger Causality (TRGC) for Comprehension Difficulty

To verify the reliability of these directional influences, Time-Reversed Granger Causality (TRGC) was computed as a control measure, as shown in Figure 7.30. The Net GC (solid blue line) represents the true directional flow of connectivity. The Net GC is calculated by subtracting GC (Frontal => Temporal) from GC (Temporal => Frontal). The TRGC (dashed red line) serves as a validation tool to identify potential spurious causality caused by shared inputs or noise artifacts [?]. In the theta and beta bands, the Net GC consistently exceeds the TRGC, reinforcing the evidence for genuine directional interactions from frontal to temporal regions. At certain frequencies, such as near 10 Hz and 15 Hz, the TRGC values approach the Net GC, suggesting the presence of shared external sources or volume conduction effects. However, the overall pattern demonstrates that the causal relationships identified in the theta and beta bands are robust and not purely driven by artifacts. Figure 7.31 provides the summary of findings based on State 1 epochs for P1.

Dominance Type	Theta Band	Alpha Band	Beta Band
Frontal -> Temporal	Yes	Yes	No
Temporal -> Frontal	No	No	Yes

Figure 7.31: Participant 1's Connectivity Dominance based on Granger Causality for State 1 (Comprehension Difficulty)

This analysis was extended with the remaining participants (P2 to P13). The results from figure 7.32 to check for generalizing trend across sample population. This result also aligns well with the findings from Power Analysis where we define ILCI (Inhibited Language Comprehension Index) as a ratio of Theta (temporal) and

Beta (Frontal). Hence, the GC analysis has extended the validity of our hypothesis providing more evidence for the possibility of ILCI and GC dominance (in theta and beta) as a generalizable neural signatures to identify language comprehension difficulties.

Dominance Type	Theta Band	Alpha Band	Beta Band
Frontal -> Temporal	7	6	4
Temporal -> Frontal	5	6	8
Total Participants	12	12	12

Figure 7.32: Connectivity Dominance across Frequency Bands for Partcipant P2 to P13

Granger Causality Analysis for Formation Difficulty The Granger Causality (GC) analysis for formation difficulty (State 2) investigates the directional flow of information between the frontal and temporal brain regions. Figure 7.33 illustrates the GC values for both *Frontal* → *Temporal* (blue) and *Temporal* → *Frontal* (orange) interactions across different frequency bands, segmented into theta (4–8 Hz), alpha (8–13 Hz), and beta (13–30 Hz) ranges.

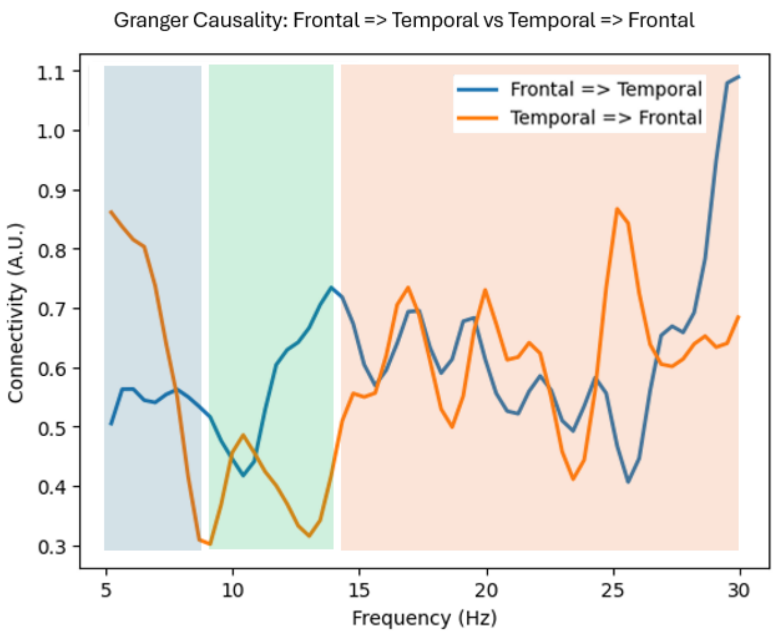

Figure 7.33: Granger Causality Analysis for Formation Difficulty: Frontal → Temporal vs Temporal → Frontal Connectivity

At theta frequencies (4–8 Hz), the temporal channels exhibits stronger connectivity toward the frontal channels. In the alpha band (8–13 Hz) represented by the green-shaded region, a near-equilibrium state emerges, with the curves for *Frontal* → *Temporal* and *Temporal* → *Frontal* connectivity intersecting. This state of equilibrium suggests a balance between top-down executive control and bottom-up sensory processing, possibly indicating a transitional phase where the brain reorganizes information for motor planning and grammatical structuring [74]. At lower beta band frequencies (13 to 21 Hz), the frontal region demonstrates slightly greater influence on the temporal region, reflecting the recruitment of executive functions, such as working memory and decision-making, to overcome the difficulty (Friston, 2011). At higher beta band (25 to 30 Hz) Frontal connectivity again starts to dominate. This dynamic interplay across bands highlights the shifting neural demands, with temporal regions leading at lower frequencies for memory and sensory integration, and frontal regions taking over at higher frequencies to manage executive control and resolve formation difficulty. These findings underscore the importance of analyzing multiple frequency bands and validating causality using TRGC to ensure robust interpretations of neural connectivity during cognitive tasks.

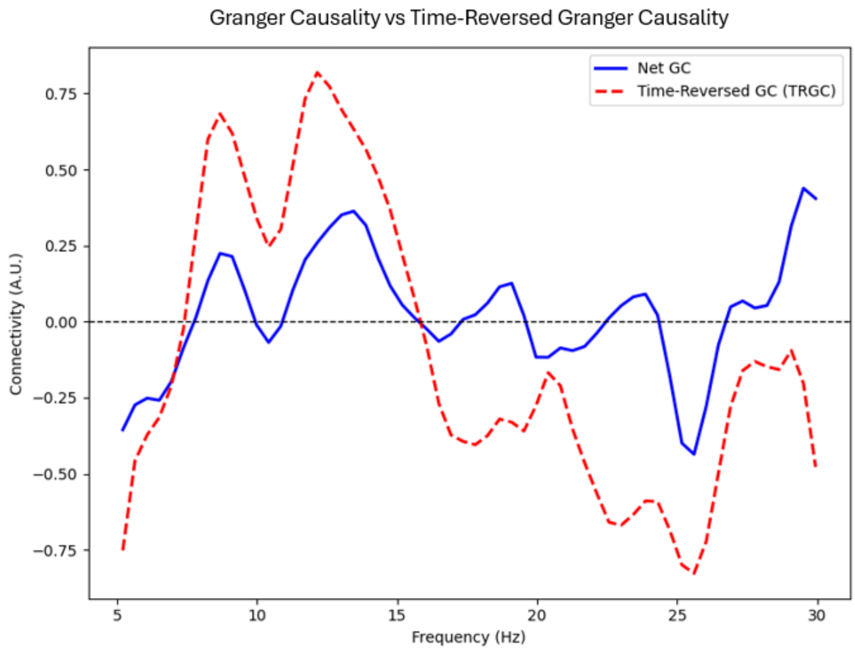

Figure 7.34: Comparison of Granger Causality (GC) and Time-Reversed Granger Causality (TRGC) for Formation Difficulty

To verify the reliability of the directional influences observed in the Granger Causality (GC) analysis, Time-Reversed Granger Causality (TRGC) was employed as a control measure, as presented in Figure 7.34. The net GC (solid blue line) reflects the directional flow of information computed as the difference between *Frontal* $\rightarrow$ *Temporal* and *Temporal* $\rightarrow$ *Frontal* connectivity values. The TRGC (dashed red line) provides a validation mechanism to detect potential spurious causality caused by shared inputs, noise artifacts, or volume conduction effects [?]. In the theta band (4 to 8 Hz), the TRGC values are notably higher than the net GC at several frequencies, particularly between 5 Hz and 10 Hz. This suggests that the observed connectivity at these lower frequencies may be influenced by shared sources or volume conduction, reducing the reliability of directional inferences. Despite this, the net GC remains consistently positive in parts of the beta band, particularly between 20 Hz and 30 Hz, where it clearly exceeds the TRGC values. This pattern strengthens the evidence for a genuine top-down influence from frontal to temporal regions during the later stages of response formation. The lower TRGC values in the beta band indicate minimal artifact contamination, confirming that the causal flow identified in this range is robust and not a result of spurious connectivity. This aligns with the observed increase in *Frontal* $\rightarrow$ *Temporal* influence in the beta range, which reflects executive control and motor-related planning processes essential for response formation [74]. These results underscore the importance of using TRGC to validate Granger Causality outcomes, ensuring that the directional flow observed represents genuine neural interactions rather than artifacts. Figure 7.35 provides the summary of findings based on State 1 epochs for P1.

Dominance Type	Theta Band	Alpha Band	Beta Band
Frontal -> Temporal	No	Yes	No
Temporal -> Frontal	Yes	No	Yes

Figure 7.35: Participant 1's Connectivity Dominance based on Granger Causality for State 1 (Formation Difficulty)

This analysis was extended to include the remaining participants (P2 to P13) to evaluate the generalizability of the observed trends across the sample population. As illustrated in Figure 7.36, theta band dominance in the temporal regions was evident in only 6 out of 12 participants, indicating inconsistent results. A similar observation was made for the beta band, where no clear dominance pattern could be established across participants.

Dominance Type	Theta Band	Alpha Band	Beta Band
Frontal -> Temporal	6	7	6
Temporal -> Frontal	6	5	6
Total Participants	12	12	12

Figure 7.36: All Participants' Connectivity Dominance based on Granger Causality for State 2 (Formation Difficulty)

However, frontal dominance in the alpha band was observed in 7 out of the 12 participants, suggesting a more consistent trend. These findings, combined with the earlier observation that the ILFI did not demonstrate a generalizable pattern during power analysis, underscore the challenges in designing a reliable index for identifying formation difficulties. Nevertheless, the consistent presence of alpha band activity in the temporal regions across power and connectivity analyses suggests that alpha (temporal) may be a promising candidate for further investigation. This alignment highlights the potential relevance of alpha activity as a neural marker for response formation difficulties, even though further validation across larger datasets is necessary to confirm its robustness.

Chapter 8

Discussion

8.1 Limited Generalizability of Neural Markers

The classification of the two difficulty states, though insightful, presents a few challenges. The causes of comprehension and formation difficulties (button clicks) varied among participants. Subjective feedback collected during post-experiment interviews revealed individual differences in cognitive strategies and task engagement. For instance, Participant P3 mentioned, *"I could not understand a word, but I sometimes still didn't click the 'No Comprehension' button as I could infer the meaning from context."* Similarly, Participant P5 stated, *"I wanted to challenge myself to respond in Dutch, so I avoided using the 'No Formation' button."* These individual differences introduced variability in the experimental process, affecting the consistency of data interpretation.

Behavioral and motivational factors such as temperament, learning attitude, and perseverance contributed to the variability observed in EEG responses. To mitigate such confounding factors in future studies, more comprehensive participant screening and reduced task scope could be explored during further exploration.

8.2 Variability due to specific causes of difficulty

The underlying causes for the observed cognitive states, namely State 1 (Difficulty in Comprehension) and State 2 (Difficulty in Formation), can vary significantly among participants. In the case of comprehension difficulties (State 1), participants may struggle due to an inability to understand specific words, multiple unfamiliar words, or complex sentence structures. Similarly, response formation difficulties (State 2) could arise due to challenges such as the inability to recall specific words, uncertainty in constructing grammatically correct sentences, or hesitation in structuring coherent responses. These varying cognitive challenges introduce considerable variability in the recorded EEG signals, as different neural pathways may be engaged depending on the underlying difficulty.

Such inconsistencies are not restricted to State 2 but may also occur during comprehension tasks (State 1) and even in the baseline condition (State 0), where participants exhibit elevated variance due to underlying cognitive or emotional states.

These factors could contribute to significant variability across participants, impacting the consistency and reliability of the defined neural signatures. Addressing such variability presents a significant challenge, as language processing is inherently individualized and influenced by various cognitive factors. A potential approach to managing this issue is to design experiments that deliberately induce specific challenges within a controlled environment to better isolate their effects. Additionally, integrating EEG with complementary data sources, such as eye-tracking or behavioral response patterns, can offer a more comprehensive perspective on the underlying cognitive processes, facilitating a deeper understanding and disentanglement of the sources of variability.

8.3 Limitations of the Interface and LLM Response

Participants provided critical feedback regarding the interface and the language model response, which impacted their engagement and performance. Some participants reported that the "Not able to respond" button provided overly simplistic assistance, which was not always useful. Participant P6 noted, *"I don't find the 'Not able to respond' button all that helpful; I preferred the green button as it provided the answer I was looking for."* Additionally, some participants, such as P2 and P3, excessively relied on the assist buttons, leading to inconsistencies in engagement levels across trials.

Baseline proficiency tests indicated significant variability in participants' Dutch language skills. For example, Participant P7 scored lower in Dutch-to-English translation tasks compared to Participant P9, who exhibited significantly higher proficiency across all difficulty levels. This variation resulted in different perceptions of task difficulty, with some participants finding the tasks too challenging, while others found them relatively simple. Future iterations of the experiment could incorporate adaptive difficulty adjustment based on baseline proficiency scores to ensure a more personalized learning experience.

The display of the difficulty rating scale after each scenario also posed challenges. The rating prompt appeared a total of 15 times during the experiment, which some participants found distracting. Participants such as P6 and P9 admitted to not paying attention to the rating at times, as they were too focused on responding to the chatbot. Others, such as P5, took excessive time (5 to 6 seconds) to deliberate their difficulty rating, potentially disrupting the experimental flow. A revised implementation with fewer rating prompts or simplified scales (e.g., three levels instead of ten) could improve the overall user experience.

During the pilot study, particularly in the "Easy" difficulty level, participants completed conversations faster than anticipated, prompting a recalibration of task difficulty for the main experiment to enhance engagement. Furthermore, some participants indicated that the total number of scenarios could be reduced, as they experienced fatigue towards the end of the 45-minute session.

8.4 Scenarios and Difficulty Levels

The analysis revealed that neither the ILCI nor ILFI indices exhibited consistent trends across varying difficulty levels. This lack of trend could be attributed to the complexity of subjective cognitive effort and linguistic proficiency differences among participants. A potential improvement in future studies could involve simplifying the rating scale to three categories—Easy, Medium, and Hard—rather than the current 10-point scale. This modification could facilitate more straightforward data interpretation and enhance the reliability of difficulty level assessments.

Some participants expressed preferences for specific scenarios, with some offering more engaging content and opportunities for meaningful interaction. During post-experiment feedback, several participants indicated that Scenario 4 (Banking) was the least engaging, as they lacked the necessary vocabulary and struggled to extend the conversation. This was particularly highlighted by Participant P1, who participated in three separate sessions over two weeks. Although the pilot study had indicated potential issues with engagement, the single-session nature of the pilot did not fully capture the fatigue effects observed in the main study. The findings suggest that scenario design should take into account participants' familiarity with the context and ensure that the conversational content remains engaging over multiple sessions. Future iterations of the experiment could benefit from adaptive scenario selection based on participants' interest and proficiency levels.

8.5 Limitations of the Device

While the Emotiv EpocX is a suitable choice for this study, certain limitations related to its channel configuration must be acknowledged. The available channels primarily favor frontal region data collection, with eight electrodes designated for frontal measurements [AF3, F7, F3, FC5, FC6, F4, F8, AF4], whereas temporal and parietal regions are represented by only two channels each [T7, T8] and [P7, P8], respectively. This distribution ensures greater signal reliability for frontal regions; however, it may compromise the accuracy and reliability of data obtained from the temporal and parietal areas. A device with a higher density of electrodes, particularly in the temporal and parietal regions, would provide more granular data and potentially yield more reliable estimates of the identified neural markers.

8.6 Future Considerations

Based on the limitations identified in the study, several recommendations can be made for future experiments to enhance data validity and participant engagement:

- Increase the sample size to a minimum of 30 participants to ensure statistical robustness and generalizability of the proposed neural indices.
- Implement difficulty scaling (of LLM Response) based on baseline proficiency.
- Redesign the assist feature (especially for State 2) to provide more contextually relevant and dynamic support based on the user's previous responses.

- Simplify the core task and minimize cognitive interruptions wherever possible.
- Introduce shorter experimental sessions to mitigate fatigue and maintain participant engagement throughout the study.
- Conduct longitudinal studies to track changes in neural responses over extended language learning periods.
- Machine learning-based feature selection, could have provided a more nuanced understanding of the neural dynamics involved in language tasks

These improvements will not only strengthen the scientific rigor of the research but also contribute to the practical implementation of EEG-based BCI applications in language learning.

Chapter 9

Conclusions

The findings of this study provide valuable insights into the neural mechanisms underlying language comprehension and formation difficulties. These insights should be interpreted within the constraints of the study's fixed time frame. The following neural markers were identified to address **RQ1** and **RQ2**.

Comprehension Difficulty (State 1)

- Right parietal and temporal regions demonstrate increased positive activity.
- Increased theta band power, particularly in temporal channels.
- Decreased beta band power in frontal channels.
- Both parametric and non-parametric tests were conducted to establish that ILCI could potentially be a valid index for identifying comprehension difficulties.
- Strong connectivity between the left temporal region (T7) and frontal regions, particularly FC6, F4, F8, and AF4.
- Coherence values indicate that for 8 out of 12 participants, the temporal-frontal connectivity (T7 to frontal regions) remains strong, with COH values exceeding 0.5.
- While temporal-frontal coherence may hold promise as an identifying feature of comprehension difficulties, the inconsistent phase stability (PLV) reduces its robustness.
- In the theta band, frontal regions exert significant control over temporal regions during cognitive processing associated with comprehension difficulties.
- At beta-band frequencies, temporal connectivity shows renewed dominance, particularly in the higher beta range (20–25 Hz).

Formation Difficulty (State 2)

- Heightened fronto-temporal activation. Negative potentials are observed in the central and parietal regions.
- Higher alpha band suppression in temporal channels.
- Higher theta band power in frontal channels.
- ILFI was not significantly different from the baseline; hence further investigation is necessary.
- Significant connectivity is observed between the left frontal (AF3) and left temporal (T7) regions, as well as between the right frontal (AF4) and right temporal (T8) regions.
- For 8 out of 12 participants, coherence values exceed 0.5, indicating consistent amplitude-phase coupling across hemispheres.
- While bilateral connectivity appears to be a promising feature for identifying response formation challenges, further studies with larger sample sizes are needed to validate its robustness.
- Frontal dominance in the alpha band was observed in 7 out of the 12 participants.

The small sample size of 12 participants limits the generalizability of the proposed neural markers. While results indicate that the markers shows potential in distinguishing comprehension difficulties from baseline, further studies with a larger sample size (>30 participants) are required to validate its applicability for reliable brain-computer interface (BCI) applications. The experimental platform developed based on the **Design Question** serves as a crucial infrastructure to support further research in BCI-based language learning, enabling the replication and refinement of experimental protocols to advance the field.

9.1 Implications for Language Processing

The findings of this study offer important contributions to the understanding of language processing in the context of second-language acquisition. More broadly, the insights derived from this research have practical implications for adaptive language learning technologies. By leveraging the defined real-time neural markers defined and explored in this study, it becomes possible to develop systems that dynamically adjust task difficulty or provide tailored assistance based on learners' cognitive states. Such applications can bridge the gap between language pedagogy and neuroscience, enabling personalized learning experiences. Furthermore, the findings contribute to the growing field of brain-computer interfaces (BCIs) for language rehabilitation. The proposed indices could be employed to monitor and enhance the progress of patients recovering from language-related impairments, such as aphasia, by providing objective measures of their comprehension and production capabilities.

9.2 Open Source Contribution

A significant outcome of this study is its contribution to the open-source community. The complete code, including the experimental design, data acquisition pipeline, and analysis scripts, has been made publicly available on GitHub(https://github.com/venk12/unnamed). By sharing this repository,the research promotes transparency and reproducibility, enabling other researchers to validate the findings or adapt the tools for new applications. The modular architecture of the system—spanning the user interface, logic module, and recording module—was designed to support customization and scalability. These contributions underscore the importance of open science in advancing the fields of neuroscience and human-computer interaction.

9.3 Final Remarks

This thesis represents a step forward in understanding the neural basis of language comprehension and formation, with implications for both theoretical neuroscience and practical applications. By identifying distinct neural markers and proposing quantitative indices, the research lays the groundwork for future studies exploring the interplay between brain activity and language processing.

Bibliography

[1] Jerry A. Fodor. *The Modularity of Mind.* MIT Press, 1983.

[2] Ludwig Lichtheim. On aphasia. *Brain*, 7(4):433–484, 1885.

[3] Gregory Hickok and David Poeppel. The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8(5):393–402, 2007.

[4] James K. Rilling, Matthew F. Glasser, Todd M. Preuss, Xiaodong Ma, Tongshan Zhao, and Xiaoping Hu. The evolution of the arcuate fasciculus revealed with comparative dti. *Nature Neuroscience*, 11(4):426–428, 2008.

[5] Nina F. Dronkers, David P. Wilkins, Robert D. Van Valin, Brian B. Redfern, and Jeffrey J. Jaeger. Neural substrates of language: Aphasia and the functional anatomy of the perisylvian cortex. *Journal of Clinical Neurophysiology*, 17(2):146–160, 2000.

[6] Jeffrey R. Binder, Rutvik H. Desai, William W. Graves, and Lisa L. Conant. Where is the semantic system? a critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex*, 19(12):2767–2796, 2009.

[7] Martin J. Pickering and Simon Garrod. An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, 36(4):329–347, 2013.

[8] Gregory Hickok and David Poeppel. The dual stream model of speech and language processing: The neuroanatomy of language revisited. *Neuron*, 67(2):279–288, 2015.

[9] Angela D. Friederici. The neural basis of language development and its impairment. *Neuron*, 76(6):1047–1059, 2012.

[10] Xing Tian and David Poeppel. Forward modeling in speech production and comprehension. *Proceedings of the National Academy of Sciences*, 107(31):13909–13914, 2010.

[11] Tatyana H. Heinks-Maldonado, Daniel H. Mathalon, and Marcus Gray. Forward models predict sensory consequences of speech. *Journal of Neuroscience*, 26(35):9057–9062, 2006.

[12] Michael F. Schober and Herbert H. Clark. Spatial perspective-taking in conversation. *Cognition*, 32(1):63–91, 1989.

[13] Daniel M. Wolpert, Zoubin Ghahramani, and Michael I. Jordan. Computational principles of movement neuroscience. *Nature Neuroscience*, 1(7):492–500, 1997.

[14] Falk Huettig and Robert J. Hartsuiker. Monitoring in speech production: Evidence from eye movements. *Journal of Memory and Language*, 63(4):394–414, 2010.

[15] Natalie Sebanz, Harold Bekkering, and Günther Knoblich. Joint action: Bodies and minds moving together. *Trends in Cognitive Sciences*, 10(2):70–76, 2006.

[16] Maria Alessandra Umiltà, Evelyne Kohler, and Vittorio Gallese. Mirror neurons responding to the observation of meaningful actions. *Science*, 293(5527):1357–1360, 2001.

[17] Susan Hurley. *The Shared Circuits Model: How Control, Mirroring, and Simulation Can Enable Imitation, Deliberation, and Mindreading*. MIT Press, 2008.

[18] Ellen Bialystok, Fergus I M Craik, David W Green, and Tamar H Gollan. Bilingualism: Consequences for mind and brain. *Trends in Cognitive Sciences*, 13(4):240–245, 2009.

[19] Jubin Abutalebi and David W Green. Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology*, 28(5):1–16, 2016.

[20] David Peeters, Elodie Runnqvist, and David W Green. The role of cognitive control in bilingual language switching: Evidence from fmri studies. *Neuropsychologia*, 129:107–117, 2019.

[21] Ping Li, Jennifer Legault, and Kaitlyn A Litcofsky. Structural changes in the brain associated with bilingualism: A review. *Bilingualism: Language and Cognition*, 17(4):1–12, 2014.

[22] Natalie H Brito, Kristen Sayler, and Rachel Barr. The role of early language exposure on the development of structural connectivity in the arcuate fasciculus: A longitudinal study. *Developmental Science*, 19(6):1–12, 2016.

[23] Judith F Kroll and Paola E Dussias. Understanding the consequences of bilingualism for language processing and cognition. *Journal of Cognitive Psychology*, 25(5):497–514, 2013.

[24] Dorothy VM Bishop, Margaret J Snowling, Paul A Thompson, Trisha Greenhalgh, and CATALISE consortium. Developmental language disorder: A diagnostic perspective. *Journal of Child Language*, 44(3):1–20, 2017.

[25] Laurence B Leonard. *Children with Specific Language Impairment*. Cambridge University Press, 2014.

[26] Margaret J Snowling and Charles Hulme. The interface between spoken and written language: Developmental disorders and the role of phonological processing. *Journal of Child Psychology and Psychiatry*, 52(3):1–10, 2011.

[27] Angela D Friederici. The brain basis of language processing: From structure to function. *Physiological Reviews*, 91(4):1357–1392, 2011.

[28] Susan H Ebbels. Intervention for children with developmental language disorder: A systematic review of the evidence base for interventions targeting syntax and morphology. *International Journal of Language & Communication Disorders*, 49(4):1–12, 2014.

[29] Courtenay Frazier Norbury, Debbie Gooch, and Nicola A Badcock. Developmental language disorder: A review of the evidence for its diagnosis and treatment. *Archives of Disease in Childhood*, 101(3):1–8, 2016.

[30] Angela D Friederici and Sonja A Kotz. The brain basis of syntactic processes: Functional imaging and lesion studies on syntax. *Current Opinion in Neurobiology*, 13(2):262–268, 2003.

[31] Alexander G Huth, Wendy A de Heer, Thomas L Griffiths, Fr
'ed
'eric E Theunissen, and Jack L Gallant. Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature*, 532(7600):453–458, 2016.

[32] Corianne Rogalsky and Gregory Hickok. The role of broca's area in sentence comprehension: Evidence from neuroimaging studies. *Frontiers in Human Neuroscience*, 5:14, 2011.

[33] David Caplan and Gloria S Waters. The role of working memory during syntactic processing in sentence comprehension. *Psychonomic Bulletin & Review*, 10(2):320–329, 2003.

[34] Nina F Dronkers, David P Wilkins, Robert D Van Valin Jr, Brian B Redfern, and Jeffrey J Jaeger. Lesion analysis reveals that brain areas critical for speech are not part of the classical model but form a distributed network for language. *Brain*, 127(7):1461–1475, 2004.

[35] Mark Steyvers and Joshua B Tenenbaum. The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive Science*, 29(1):41–78, 2005.

[36] Chantel S Prat and Marcel Adam Just. The brain basis of individual differences in language comprehension ability. *Language and Linguistics Compass*, 5(9):635–649, 2011.

[37] Adeen Flinker, Anna Korzeniewska, Andriy Y Shestyuk, Piotr J Franaszczuk, Nina F Dronkers, Robert T Knight, and Nathan E Crone. Redefining the role of broca's area in speech. *Nature Neuroscience*, 18(12):1462–1470, 2015.

[38] Yosef Grodzinsky and Angela D Friederici. Neuroimaging and neurolinguistic evidence for the role of broca's area in syntax. *Journal of Cognitive Neuroscience*, 18(5):1–12, 2006.

[39] Yasuharu Sakurai, Atsushi Yagishita, and Yuko Goto. The role of the left posterior middle temporal gyrus in written language production: Evidence from agraphia. *Neuropsychologia*, 97:75–85, 2017.

[40] Karalyn Patterson, Peter J Nestor, and Timothy T Rogers. s̈emantic hub:̈ The anterior temporal lobe integrates conceptual information across modalities. *Nature Reviews Neuroscience*, 8(12):976–987, 2007.

[41] Kristof Strijkers and Albert Costa. The p2 component as an index of early lexical access during word production: Evidence from meg and eeg studies. *Neuroimage*, 55(2):546–557, 2011.

[42] Lesya Y Ganushchak and Niels O Schiller. N400 evidence for lexical interference during speech production. *Psychophysiology*, 48(5):655–663, 2011.

[43] Darren Tanner and Janet G Van Hell. The p600 component and syntactic reanalysis: Evidence from sentence production tasks. *Frontiers in Psychology*, 5:224, 2014.

[44] Michael Foster, Diana Woo, and Shravan Vasishth. Neural signatures of language learning: Evidence from eeg during an artificial language task. *Cognitive Neuroscience*, 12(2):85–97, 2021.

[45] Jubin Abutalebi and David W Green. Control mechanisms in bilingual language production: Neural evidence from language switching studies. *Language and Cognitive Processes*, 22(5):768–804, 2007.

[46] Evelina Fedorenko and Sharon L Thompson-Schill. Reworking the language network. *Trends in Cognitive Sciences*, 18(3):120–126, 2014.

[47] Roel M Willems. Naturalistic language paradigms: Balancing experimental control and ecological validity. *Trends in Cognitive Sciences*, 19(5):295–296, 2015.

[48] J. Vanthornhout et al. Neural tracking of speech reveals comprehension in a real-world listening task. *Nature Communications*, 14(1):1–12, 2023.

[49] Angela D. Friederici. The brain basis of language processing: From structure to function. *Physiological Reviews*, 91(4):1357–1392, 2011.

[50] Peter Hagoort and Jos J. A. van Berkum. Beyond the sentence given. *Philosophical Transactions of the Royal Society B*, 362(1481):801–811, 2007.

[51] Vitoria Piai et al. The electrophysiology of language production: What could be improved. *Frontiers in Psychology*, 5:1560, 2014.

[52] Laura J. Batterink et al. Implicit and explicit learning of temporal sequences assessed with the serial reaction time task. *Neuropsychologia*, 77:185–193, 2015.

[53] Erica L. Middleton and Myrna F. Schwartz. Errorless learning in cognitive rehabilitation: A critical review. *Neuropsychological Rehabilitation*, 22(2):138–168, 2012.

[54] Doug Rohrer and Kelli Taylor. The shuffling of mathematics problems improves learning. *Instructional Science*, 35(6):481–498, 2007.

[55] Silvia E. Kober et al. Specific effects of eeg based neurofeedback training on memory functions in post-stroke victims. *Journal of NeuroEngineering and Rehabilitation*, 14(1):1–13, 2017.

[56] Robert Bauer and Alireza Gharabaghi. Reinforcement learning for adaptive threshold control of restorative brain-computer interfaces: A bayesian simulation. *Frontiers in Neuroscience*, 9:36, 2015.

[57] Joshua Sabio, Nikolas S. Williams, Genevieve M. McArthur, and Nicholas A. Badcock. A scoping review on the use of consumer-grade eeg devices for research. *PLOS ONE*, 19(3):e0291186, March 2024.

[58] Steven J. Luck. *An Introduction to the Event-Related Potential Technique*. MIT Press, 2nd edition, 2014.

[59] Arnaud Delorme and Scott Makeig. Eeglab: An open-source toolbox for analysis of single-trial eeg dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1):9–21, 2004.

[60] Yuanjun Xie, Yanyan Li, Haidan Duan, Xiliang Xu, Wenmo Zhang, and Peng Fang. Theta oscillations and source connectivity during complex audiovisual object encoding in working memory. *Front. Hum. Neurosci.*, 15:614950, March 2021.

[61] Wolfgang Klimesch. Eeg alpha and theta oscillations reflect cognitive and memory performance: A review and analysis. *Brain Research Reviews*, 29(2-3):169–195, 1999.

[62] Wolfgang Klimesch. Eeg alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain Research Reviews*, 29(2-3):169–195, 1999.

[63] Gert Pfurtscheller and Fernando H Lopes da Silva. Functional brain imaging based on erd/ers. *Vision Research*, 41(10-11):1257–1260, 2001.

[64] Ole Jensen and Ali Mazaheri. Shaping functional architecture by oscillatory alpha activity: Gating by inhibition. *Frontiers in Human Neuroscience*, 4:186, 2010.

[65] Wolfgang Klimesch. Alpha-band oscillations, attention, and controlled access to stored information. *Trends in Cognitive Sciences*, 16(12):606–617, 2012.

[66] Angela D Friederici. The cortical language circuit: From auditory perception to sentence comprehension. *Trends in Cognitive Sciences*, 16(5):262–268, 2012.

[67] Peter Hagoort. On broca, brain, and binding: a new framework. *Trends in Cognitive Sciences*, 9(9):416–423, 2005.

[68] Tülay Demiralp, Z Bayraktaroglu, A Aślak, et al. Alpha rhythm and creative processes. *International Journal of Psychophysiology*, 66:75–84, 2007.

[69] Andreas K Engel and Pascal Fries. Beta oscillations and motor control. *Trends in Cognitive Sciences*, 14(12):432–439, 2010.

[70] Marcus CM Bastiaansen and Peter Hagoort. Beta-band dynamics in language comprehension and production. *Human Brain Mapping*, 31(5):640–654, 2010.

[71] Gert Pfurtscheller and Christa Neuper. Event-related beta synchronization during motor imagery. *Brain Research*, 116:152–158, 1997.

[72] Floris P De Lange, Ole Jensen, Markus Bauer, and Ivan Toni. Beta-band oscillations reflect the dynamics of motor sequences. *Neuroimage*, 41(3):949–957, 2008.

[73] Marcus CM Bastiaansen and Jos JA van Berkum. Beta oscillations and language comprehension: A review of evidence. *Human Brain Mapping*, 31(6):799–809, 2010.

[74] Christina S Poon and Shing-Yee Lo. The role of frontal beta rhythms in motor planning and execution. *Neurobiology of Learning and Memory*, 170:107–115, 2019.