Spring 2023

EEG-based Spanish Language Proficiency Classification: An EEG Power Spectrum and Cross-Spectrum Analysis

Blaise Xavier O'Mara
University of New Hampshire, Durham

Skyler Baumer
University of New Hampshire, Durham

EEG-based Spanish Language Proficiency Classification:

An EEG Power Spectrum and Cross-Spectrum Analysis

Author(s): O'Mara, Blaise[1], Skyler Baumer[1]

Submission Date: 05/09/2023

Supporting Faculty: Ronald Croce, PhD[2], Wayne Smith, PhD[1], Mauricio Pulecio, PhD[3]

Supporting Institutions: McNair Scholars Program, Department of Computer and Electrical Engineering*, Department of Health Sciences, Department of Literatures, Languages and Cultures

University of New Hampshire

---

[1] Department of Electrical and Computer Engineering
[2] Department of Kinesiology
[3] Department of Literatures, Languages, and Cultures

Table of Contents

Table of Contents ii
ABSTRACT iv
DEFINITION OF TERMS v
Event-Related Desynchronization v
Coherence v
1 INTRODUCTION 1
1.1 Problem Statement ..... **Error! Bookmark not defined.**
1.2 Significance & Objectives ..... **Error! Bookmark not defined.**
1.3 Language Processing and the Brain 1
1.4 Predictors of Language Proficiency 1
1.5 Applications of Machine Learning with EEG 3
2 METHODS AND PROCEDURES 5
2.1 Participants 5
2.2 Electrode Sites and Regions of Interest 7
2.3 Stimulus Material and Surveys 8
2.4 EEG Data Acquisition 9
2.5 Data Processing: BrainVision Analyzer 10
2.6 Alpha ERD and Alpha and Beta Coherence ANOVA Analysis in Python 12
2.7 Training a Machine-Learning System 14
2.8 Cost Analysis 15
3 RESULTS 16
3.1 Comprehension and Engagement Questionnaires 16
3.2 Alpha ERD 3x3 ANOVA 17
3.3 Alpha Coherence 3x3 ANOVA Analysis 18
3.4 Beta Coherence 3x3 ANOVA Analysis 19
4 DISCUSSION 21
4.1 Appropriate Audio Selections for Conditions 21
4.2 Lack of Alpha ERD Significance 25
4.3 Bilateral Coherence Dominance as a Predictor for Spanish L2 Proficiency 26
4.4 Limitations and Challenges 28
5 Future Work 28

5.1 Participant Pool & Group Labels 28
5.2 EEG Data Feature Consideration 29
5.3 Designing a Machine Learning System to Classify Spanish L2 Proficiency 29
6 Conclusion 29
7 Statements 30
7.1 Availability of Data 30
7.2 Ethical Considerations: Human Participants 30
7.3 Proposed Timeline of Tasks and Milestones 32
7.4 Budget & Bill of Materials 32
8 Standards Utilized 34
9 Individual Contributions 34
9.1 Skyler Baumer 34
9.2 Blaise O'Mara 34
10 Acknowledgements 36

ABSTRACT

Second language proficiency may be predicted with electrophysiological techniques. In a machine learning application, this electrophysiological data may be used for language instructors and language students to assess their language learning. This study identifies how electroencephalogram (EEG) power spectrum and cross spectrum data of the brain cortex relates to Spanish second language (L2) proficiency of 20 Spanish language students of varying proficiency levels at the University of New Hampshire. The two metrics for assessing cortical power and processing were event-related desynchronization (ERD)—a measure of relative change in power—of the alpha (8-12 Hz) brain frequency band, and alpha and beta (13-30Hz) brain frequency band coherence—a relative measure of spectral correlation between two cortical areas, respectively. Alpha ERD and alpha and beta coherence were calculated from EEG data collected on participants of ACTFL Spanish L2 proficiency levels Novice, Intermediate, and Advance while listening to three audio conditions of varying Spanish language difficulty. Significant differences in both alpha and beta coherence were found between proficiency groups. Higher proficiency Spanish L2 students exhibited more bilateral alpha and beta coherence dominance in the frontal and central cortices while the lower proficiency Spanish L2 students demonstrated greater unilateral alpha and beta coherence between the posterior cortices and Broca and Wernicke’s Area. This suggests that higher proficiency simultaneous bilinguals utilize the frontoparietal and fronto-occipital networks for achieving language comprehension and focus.

Keywords: Spanish, language, proficiency, EEG, ERD, coherence, machine learning

DEFINITION OF TERMS

Coherence

Coherence is a relative measure of magnitude and phase change between two time-bound voltage signals. When the magnitude and phase difference between two signals remain constant, the two signals have a high coherence. This relative measure between 0-1 (1 is high coherence) establishes how linearly one of two signals may be predicted from the other. Thus, magnitude squared coherence compares the

$$Magnitude\ Squared\ Coherence = |C^{X,Y}(t,f)|^2 = \frac{|\Sigma_{k=1}^{K} X_k(t,f) Y_k^*(t,f)|^2}{\Sigma_{k=1}^{K} |X(t,f)|^2 \, \Sigma_{k=1}^{K} |Y_k(t,f)|^2}$$

Equation 1: Calculation for magnitude-squared coherence between two electrode sites. This is a normalized percentage value from 0-1 where $X_k(t,f), Y_k(t,f) = complex\ spectrum\ values\ for\ input\ x\ and\ y, k = segment\ index, t = time, f = frequency$

Event-Related Desynchronization

Event-related desynchronization (ERD) is a relative measure of power change within a frequency band of interest. In the context of electroencephalography, ERD is often used to assess relative power changes within the alpha frequency band in reference to an event-related (voltage) potential. A decrease in alpha band ERD at an electrode site is associated with a release of inhibition at that brain cortex region.

$$ERD\ (\%) = \frac{A(n) - R}{R} * 100$$

Equation 2: Calculation for Event-Related Desynchronization at an electrode site. This is a relative measure of power change, a normalized percentage value from 0-1 where $A(n) = average\ power\ at\ sample\ point\ n, N = number\ of\ sample\ points, n = 1, \ldots, N, and\ R = mean\ value\ of\ A(n)\ in\ reference\ interval.$

1 INTRODUCTION

1.1 Language Processing and the Brain

It has been well established that language production and comprehension occur within Broca's area and Wernicke's area in the left hemisphere (LH) of the brain, respectively. Corroborating this statement, both Truetler and Soros [1] and Ganushchak et. al [2] observed that L2 production is associated with activation of the left primary sensorimotor cortex, left inferior frontal gyrus, left anterior insula, and the bilateral cerebellar hemispheres. In terms of EEG data collection, these regions correspond to five electrode clusters: the medial frontal, left/right frontotemporal, and left/right posterior cortices. In examination of how L2 proficiency is related to EEG oscillations and coherence, Soares et. al [3], Bice et. al [4], and Reiterer et. al [5], [6], all chose these clusters within their experiment paradigm. When examining these clusters in the context of L2 proficiency, all electrodes should be placed according to the International 10/20 Standard.

However, each cohort differed in the number of electrodes they examined. Soares and cohort [3] used a 32-electrode cap, in which 27 electrodes were of interest. They had the greatest spatial resolution of the three cohorts, which allowed them to conduct a more thorough coherency analysis. Reiterer and cohort [5], [6] only examined 19 electrodes which connected to a 21-channel Nihon Kohden recorder. Although their setup was cumbersome, they found significant results in ERD characteristics and coherence. Bice et. al [4] used the fewest number of electrodes with the 14-channel Emotiv EPOC. The EPOC, a more affordable and mobile EEG device, has a faster setup time than the 32-electrode caps. However, its accuracy fluctuates from 61.84% to 92.26% [7]. Also, researchers need to purchase EmotivPro licensing plan ($1,068 per year) to collect any meaningful data. When selecting EEG collection equipment, the factors of precision, accuracy, cost, and convenience must be weighed. In relating L2 proficiency with EEG characteristics, data accuracy and precision are paramount. Thus, this study will record data from 64 electrodes that are placed according to the 10/20 standard. However, the main regions of interest included 14 electrodes Bice and cohort [4] explored.

1.2 Predictors of Language Proficiency

The neural efficiency hypothesis states that as proficiency increases in a skill or task, the brain uses less power to execute that skill or task [8]. The alpha band is most pertinent to the neural efficiency hypothesis. Alpha activity, seen through event-related synchronization (ERS), is associated with inhibition of brain regions that are non-essential to task execution. Conversely, alpha event-related desynchronization (ERD) is associated with activation of brain regions essential to task execution [9]. In these same regions, beta, and gamma ERS juxtapose alpha ERD [8]. Interestingly, as proficiency increases in a task, alpha ERD in the task-essential brain region(s) decreases, meaning that the brain is devoting less energy to inhibition [10]. These power characteristics, defined as coherence, detail the magnitude power of brain regions and their interspatial correlates.

Building on the neural efficiency hypothesis, L2 proficiency can be characterized by alpha ERS and ERD. The latter, alpha ERD, is often interpreted as a decrease in alpha power. The alpha band is typically broken into two frequency ranges: low alpha (Alpha1) and high alpha (Alpha2). Alpha1 (8-10 Hz) is related to general attention demands and maintaining inhibition, while Alpha2 (10-13 Hz) is related to semantic processing [3]. During semantic processing, alpha ERD is prevalent in task-relevant regions; yet proficient performers will show greater power in ERS [8]. Bice and cohort [4], in their study of bilingual classification and L2 proficiency, corroborate that greater alpha power is observed in the task-free brain activity of high-proficiency bilinguals. High-proficiency L2 learners should demonstrate greater alpha ERD in task-relevant regions and greater ERS in task-irrelevant regions.

Beta and gamma ERS and ERD also characterize L2 proficiency. Like alpha, the beta band is typically broken into two frequency ranges: Beta1 (13-18 Hz) and Beta2 (18.5-33 Hz). Gamma operates in the frequency range of 33-100 Hz [9]. While alpha ERD may be used to gauge proficiency, beta and gamma ERS may be used to do the same. Both beta and gamma work at the interface of language processing and working memory [4]. Thus, these two are active (ERS) in task-relevant regions of the brain during language processing whereas alpha is inactive (ERD). This shows that alpha synchrony is inversely related to beta and gamma synchrony. Several L2 proficiencies studies [3], [4], [5], [6] support this relationship.

Alpha and beta coherence further characterizes one's L2 proficiency. In particular, the regions of coherence may indicate an L2 learner's fluency. Coherence is a measure comparing spectral content between two or more electrode sites. By comparing frequency responses in power and phase, the neural firing and relationship between two regions of the brain may be observed. Furthermore, coherence shows which brain resources are used to accomplish a skill or task. When learning a foreign language in a traditional academic setting, not all activity associated with L2 comprehension is found within the LH. In L2 acquisition, there is a change in lateralization of language processing from the right hemisphere (RH) to the LH and subcortical, posterior regions. Observing this change among differing proficiency English L2 speakers, Reiterer and cohort [5] found that the low-proficiency L2 English speakers had significantly greater activity in the RH compared to the high-proficiency group. Bice et. al [4] supports this by showing that “greater bilingual skill is associated with less reliance on frontal structures.” The posterior and subcortical structures are the rudimentary, less cognitive-intensive areas of the brain. By the neural efficiency hypothesis, it is logical for these less power-intensive areas to command L2 language comprehension. In general, lower proficiency L2 speakers use more cognitive resources.

A similar relationship can be observed in the alpha and beta frequency bands. The individual coherence of the alpha and beta bands provides a more resolute relationship between power and varying L2 proficiency. During comprehension tasks of L2 audio stimuli, low-proficiency German English-learners had greater overall, widespread Alpha1 coherence compared to the high proficiency group [5]. Soares et. al [3] also found this widespread coherence such that L2 “proficiency positively predicted greater alpha rs-EEG coherence between frontal left and right clusters.” As for beta, the high-proficiency group’s Beta2 coherency in Reiterer et. al [5] was greater and more concentrated in the left temporal cluster—where Wernicke’s area is located. Bice et. al [4] corroborate this finding, as they observed that bilinguals had significantly greater Beta coherence between the left posterior and left fronto-temporal electrode clusters. Furthermore, both cohorts found that as an overall trend, low-proficiency L2 learners had greater alpha and beta coherence in clusters, whereas high-proficiency speakers exhibited greater coherence between cluster regions in these frequency bands. Based on this body of evidence, it is expected that higher proficiency bilingual speakers have greater alpha and beta coherence within the left frontotemporal network while lower proficiency speakers will have greater far-reaching, widespread coherence between the right and left frontal cortices.

1.3 Applications of Machine Learning with EEG

Machine learning techniques are increasingly being applied EEG feature data. The purpose of implementing machine learning is to classify or characterize a certain skill, task, or cognitive state with brain activity of the EEG spectrum. Popular applications include classifying mood [11], [12],

[13], yet it is sparse within classifying second language (L2) proficiency. While databases like DEAP (Dataset for Emotion Analysis using Physiological Signals) and IDEA (Intellect Database for Emotion Analysis) have been created solely for EEG-based mood classification, few have been established for EEG-based L2 proficiency.

In a review of EEG-based emotion recognition, Dadebayev et. al [7] cites 30 studies using EEG and machine learning to accomplish mood classification between the years of 2006-2019. Many studies use the well-established DEAP database [11], whereas others like Joshi and Ghongade [13] established their own IDEA database for emotion recognition. Among the studies, use of convolutional neural networks (CNN) is foundational for mood classification. The cohorts first collect substantial amounts of EEG data of a specific mood (i.e., happiness) and then use it to train CNNs to recognize features of the respective emotions [11], [13], [14], [15], [16]. Among features of interest are ERS, ERD, and coherence.

There does not yet exist a database for EEG-based L2 proficiency data. It would be novel to create an EEG-based L2 proficiency database using ERS, ERD, and coherence feature data for the purpose of L2 proficiency classification.

1.4 Significance

In the Information Era, second language (L2) acquisition is becoming crucial in global commerce, education, and international relationships. English, the most dominant language in these exchanges, is regularly taught in European and Asian countries. Most used to assess English language proficiency is the Common European Framework of Reference for Languages (CEFR). Beyond the classrooms of language instructors, CEFR is often used by cognitive neuroscientists examining the relationship between neural efficiency and L2 proficiency through electroencephalography (EEG) [5], [6], [17], [18], while the American Council for the Teaching of Foreign Language (ACTFL) [19] guidelines are used less frequently.

The neural efficiency hypothesis states that as proficiency increases in a skill or task, the brain uses less power to execute that skill or task [8]. In the context of power-spectral density (PSD), it is expected that PSD decreases during a task as proficiency increases. To observe this concept, brain activity was monitored using electroencephalography (EEG). The EEG spectrum details cortical brain activity both spatially and temporally within five major frequency bands: delta, theta, alpha, beta, and gamma. The frequencies bands range from 0.5-4 Hz, 4-8 Hz, 8-13 Hz, 13-32 Hz, and 32-100 Hz, respectively. Each of these frequencies provides information about neural efficiency. The correlation between neural efficiency and language proficiency may provide the ability to objectively measure one's language proficiency based on a physiological metric as opposed to a standard paper test [17]. However, this metric has not yet been widely implemented for language proficiency assessment. Applying the neural efficiency hypothesis to language learning not only has the potential to benefit language instructors, but also language learners.

Likewise, in the context of assessment, language proficiency classification has not been implemented using machine learning based on EEG data features. Most commonly, EEG data and machine learning implementations concern mood-arousal detection and clinical diagnoses, such as catching symptoms of early-stage Alzheimer's Disease.

1.5 Objectives

For language instructors, it is crucial to understand how and how well their students are learning. While standard paper and conversational tests can determine a student's proficiency level, EEG signal data can potentially provide information on proficiency progression based on physiological metrics. This would be a tool for language instructors as they assess their student's language acquisition. This creates an opportunity for researchers to learn how the brain acquires proficiency in verbal skills. In future work, the understanding of how the brain changes its ERD and coherence, as reflected in the EEG, to learn languages may be applied to proficiency evaluation in other skills and tasks.

The objective of this research is twofold: (1) to observe the relationship between Spanish language (SL) proficiency levels (Novice, Intermediate, and Advanced) and the Event-Related Desynchronization (ERD), Event-Related Synchronization (ERS), and coherence of the frequency bands in the EEG spectrum, and (2) to provide suggestions on how to used EEG feature data to perform Spanish L2 proficiency classification according to ACTFL guidelines with a novel machine learning application.

2 METHODS AND PROCEDURES

2.1 Participants

This study was approved to have human participants by the University of New Hampshire's (UNH) Internal Review Board. The participant pool consisted of 20 Spanish second language (L2) students attending the University of New Hampshire (UNH). They were recruited via UNH Spanish language instructors, in-class visits, flyers, and e-mail communication. These 20 students were categorized into three Spanish L2 proficiency groups: six in the NG (NG), six in the IG (IG), and eight in the AG (AG). Originally, the goal was to recruit 12 participants per proficiency group. Members in each group are differentiated by their Spanish L2 listening proficiency according to ACTFL listening proficiency guidelines of the same name categories: Novice, Intermediate, and Advanced. To accurately match the Spanish L2 proficiencies of the participants to ACTFL proficiency guidelines, Spanish instructors at UNH were consulted to assess the expected Spanish L2 proficiency of their students. It was determined that students who completed UNH Spanish 400-level, 500-level, or 600-level course(s) would correspond to ACTFL Novice, Intermediate, and Advanced (and beyond) proficiency levels (**Figure 1**), respectively.

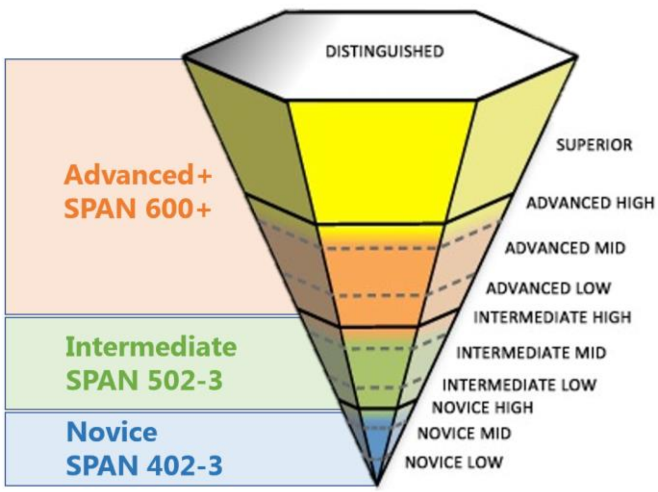


Figure 1: American Council for the Learning of Foreign Language proficiency guidelines corresponding to Spanish language course level at the University of New Hampshire.

All participants were screened to meet the inclusionary criteria regarding bilingual type, handedness, past medical history, and drug use (**Section 10.1**). To meet inclusion criteria, all participants must have been sequential bilinguals, have been right-handed, have no chronic hearing impairments, have no history of serious, chronic mental illness or disorders, have no history of brain-altering drug use (e.g., SSRIs, LSD), and have no history of issues that may interfere with normal brain function (e.g., prior intracranial surgery, history of moderate to severe traumatic brain injury, history of learning disability or attention deficit/hyperactivity disorder, or neurological diagnoses). There are two bilingual types: sequential and simultaneous. In the former category, second language(s) are acquired after having already learned their dominant, native language. As for the former category, a simultaneous bilingual acquires two (or more) languages at the same time. Sequential language acquisition is often exclusive to an academic environment starting in grade school, whereas simultaneous language acquisition occurs since birth in a home environment where two languages are often spoken interchangeably. This study aims to observe the differences in alpha ERD and alpha and beta coherence between sequential, Spanish L2 bilinguals at various proficiency levels.

Left-handed individuals were excluded from the study due to an increased percentage of atypical language lateralization in left-handed people [20]. Compared to right-handed individuals, Knecht et. al [20] observed that left-handed individuals experience 20-25% more atypical language lateralization. Alpha ERD, and alpha and beta coherence measures are spatially dependent on electrode sites; so, to ensure that resulting EEG participant data would be spatially consistent, left-handed individuals were excluded.

Health history and drug use were the last two exclusionary factors for participants. All participants had no history of traumatic brain injuries, brain lesions, multiple concussions, and chronic mental illness or disorder. They also had no history or current use of brain-altering medications or drugs. Additionally, participants did not identify with diagnosed learning disabilities nor attention deficit/hyperactivity disorders.

2.2 Electrode Sites and Regions of Interest

EEG data was recorded from 20 electrode sites (**Figure 2**). Most electrodes lied near the language centers of the brain, Broca and Wernicke’s Areas. The electrode cluster of AF3, F3, F5, F7, and FC5 represent the cortical activity of Broca’s Area, and electrodes C5 and CP5 represent the cortical activity of Wernicke’s Area. Electrodes T7, P7, and O1 represent the left temporal, left parietal, and left occipital cortices, respectively. There are also ten more electrodes located in the right hemisphere of the brain that match their left hemisphere analogues.

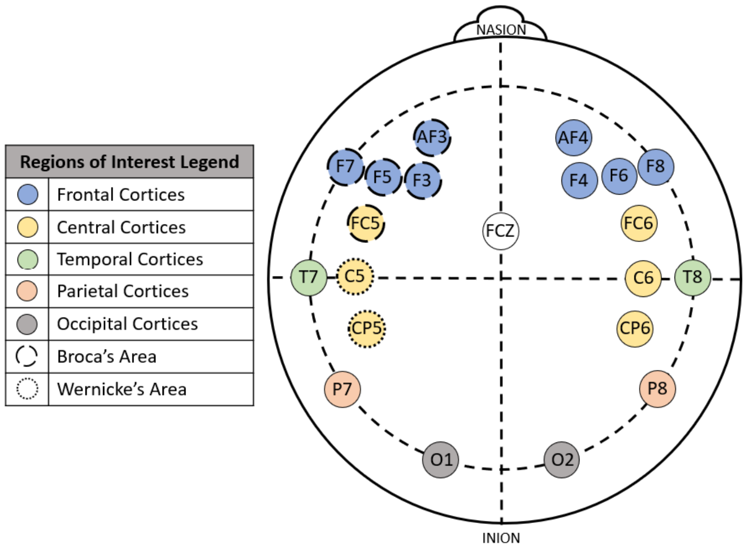


Figure 2: The 21 observed electrodes labeled and clustered by their corresponding ROIs. These 21 electrodes are spaced according to the 10/20 standard on a 64-electrode cap.

These 20 electrodes make up eight regions of interest (ROI), in which whose selection is based on Bice et. al [4] and Blagoventchenski et. al [21]. Five are in the left hemisphere and three in the right hemisphere. Electrodes AF3, F3, F5 and F7 represent the left frontal cortex. The left central-temporal cortices are represented by electrodes FC5, C5, CP5, and T7. The left posterior cortices are represented by electrodes P7 and O1. Lastly, the right frontal, central-temporal, and posterior ROIs are represented by the electrode analogues of their left hemisphere ROI counterparts.

2.3 Stimulus Material and Surveys

Participants were exposed to three different Spanish audio stimuli. The content of each audio stimuli, referred to as conditions, corresponded to ACTFL language proficiency guidelines. Specifically, the proficiency difficulty of each audio recording corresponded to ACTFL listening/viewing and listening/viewing guidelines, which feature many "can-do" statements (i.e., "I can navigate in a grocery store with little help). Conditions 1, 2 and 3 (C1, C2, and C3), featured Spanish audio with proficiency difficulties of Novice, Intermediate, and Advanced, respectively. These audio selections for conditions were selected with the help of language instructor Dr. Mauricio Pulecio of the Literatures, Languages, and Cultures Department at UNH. Each condition was approximately four minutes in length, and they were played through noise canceling headphones worn by participants.

After each condition, participants completed a comprehension and engagement questionnaire (e.g., Novice-level condition survey: **Section 10.2**). Each comprehension quiz featured four questions; so, participants may score a maximum of 12 comprehension points across all conditions. The purpose of the comprehension questionnaire is to observe (1) if the participant understood the content of the conditions, and (2) any trends between condition comprehension scores and participant proficiency groups, and (3) validate that the proficiency difficulty level of conditions are appropriately matched to their respective ACTFL proficiency category.

An engagement assessment followed each comprehension quiz. The purpose of the engagement assessment was to understand the degree to which participants were paying attention to each audio condition. Because the conditions were four minutes long, there was a chance that participants would lose focus or “space out” while listening. Losing focus for a large portion of the condition could negatively affect the consistency of Alpha ERD and coherence metrics. The biggest concern for losing focus would be the difficulty of the condition [5]. A likely scenario would be for the NG to lose focus during C3, as it is the most difficult audio selection. If they do not understand the audio content, then they may just lose attention and “space out.” So, participants were asked to indicate to which degree they agreed with the statement, “I was actively engaged with the story for the entire 4 minutes” on a Likert scale from 1 to 5, where the numbers were 1-Stongly Disagree, 2-Somewhat Disagree, 3-Neither Agree nor Disagree, 4-Somewhat Agree, and 5-Strongy Agree.

2.4 EEG Data Acquisition

EEG data collection sessions were approximately 1-1.5 hours in length. The first 30-45 minutes was spent fitting participants with BrainVision 64-channel actiCap snap electrode caps. Although only 20 electrode sites were of interest, all 64 BrainVision actiChamp electrodes were inserted in the actiCap using the International 10-20 Standard. These active amplifier electrodes Using blunt Teflon needles, saline gel was applied between the scalp and electrodes. The aim was to reduce the impedance below 20 KΩ between the electrodes and scalp. Any electrodes that had impedance values over 55 KΩ were eliminated from data collection. These impedance values were monitored with the BrainVision Recorder software.

Following cap and electrode preparation was the 30-minute data collection phase. The EEG recording environment was dimly lit with minimal visual and auditory distraction. Participants were fitted with a set of noise canceling headphones. As depicted in **Figure 3**, there are four stages to each audio task: (1) a baseline measurement is taken, (2) an audio condition is played, (3) a comprehension and engagement questionnaire is taken, and (4) a brief break is taken. Before each baseline measurement, the participants were instructed to avoid excessive blinking, sit still and upright, refrain from moving their body, remain engaged with the audio selection for its entirety, and lastly relax and clear their mind. The baseline measurements were five seconds in length, and they were used as the reference for alpha ERD measures of their respective condition. Condition order was randomized for each participant.

The remaining approximate 15-30 minutes was used for participants to clean their scalps.

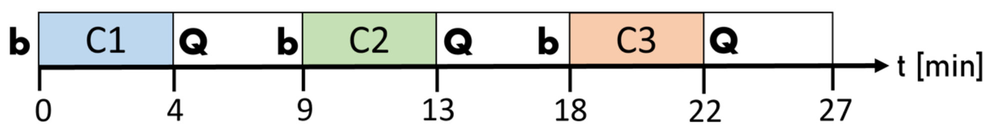


Figure 3: Data collection timeline. There are three Spanish language audio conditions (C1, C2, and C3) that are each four minutes in length. Audio conditions correspond to ACTFL difficulty guidelines. C1, C2, and C3, correspond to Novice, Intermediate, and Advanced difficulty levels, respectively.

2.5 Data Processing: BrainVision Analyzer

The resulting .vhdr, .eeg, and .vmrk files for each participant were processed in BrainVision Analyzer (BVA) version. 2.2. BVA served to prepare alpha ERD and alpha and beta coherence data for export to a Python environment. First in the data processing pipeline (**Figure 4**) was applying an 8th-order Butterworth bandpass filter with low and high cut-offs of 5 Hz and 50 Hz, respectively (**Figure 5**). A 60 Hz notch filter was also applied to the data. Next, the reference for the electrodes were changed from electrode FCz to the average of the 20 electrodes of interest. This was implemented to reduce the effects of ground localization.

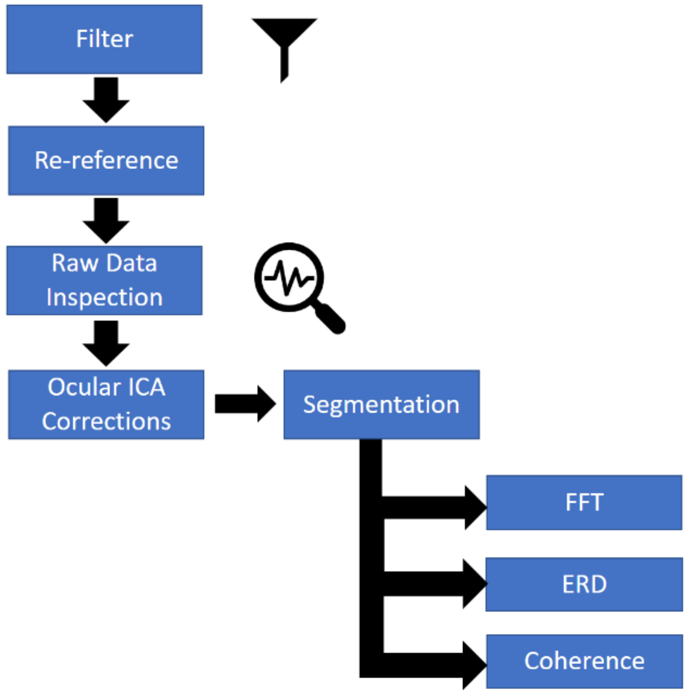


Figure 4: BrainVision Analyzer EEG data feature workflow.

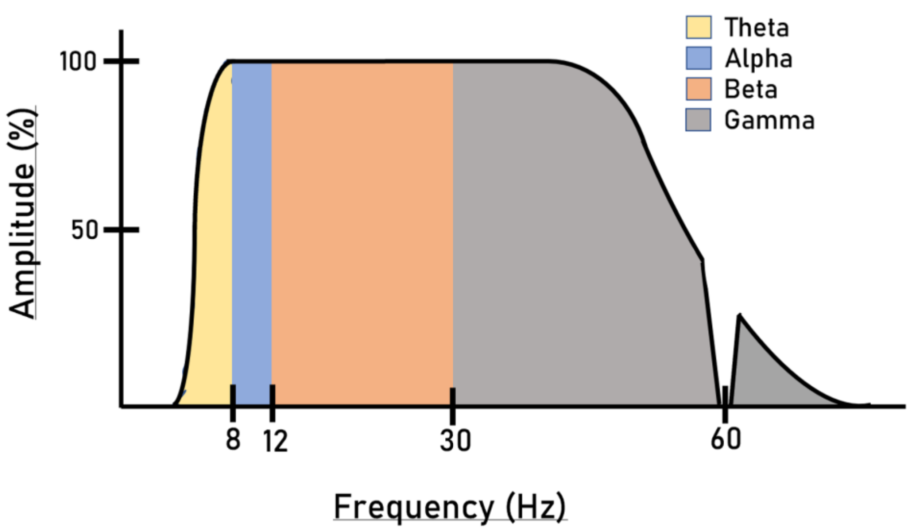


Figure 5: Digital 8th-order Butterworth bandpass filter with 60 Hz notch. Theta (3.5-7.5 Hz), Alpha (8-12 Hz), Beta (12-30 Hz), and Gamma (30-100 Hz) frequency bands are highlighted.

Artifact rejection followed re-referencing. BVA's Raw Data Inspection tool automatically detects artifacts resulting from electromyography, abnormally quick changes in voltage, electrooculography (EOG), and other environmental factors. These types of artifacts were removed with the Raw Data Inspection's semi-automatic mode, which detects artifacts but lets the user decide to remove them. EOG artifacts, resulting from blinking, were the most prevalent. Another tool, the Ocular Correction ICA (independent component analysis), was applied to remove these blinking artifacts. This was the last step of artifact rejection.

Last in the data processing pipeline was computing EEG feature data and exporting it to a Python environment. EEG feature data consisted of (1) mean alpha ERD data per condition, per electrode, per participant, and (2) both mean alpha and beta coherence per condition, per electrode pair, per participant. Mean alpha ERD was calculated using BVA's Segmentation, FFT, and ERD tools. Each condition was segmented into five-second intervals with 50% overlap. Next, the FFT of each epoch was computed. Then, the mean ERD over the course of five seconds—in reference to the five-second baseline, was calculated using all five-second epochs. Lastly, the BVA Export tool was used to simultaneously compute and export a single mean alpha ERD value, from the five-second mean alpha ERD to a text file.

Mean alpha coherence was calculated using BVA's Segmentation, FFT, and Coherence tools. Each condition was segmented into one-second epochs with 50% overlap. Following, the FFT of each epoch was computed. Then, the Coherence for each electrode pair was computed based on the FFT per epoch. Finally, the mean alpha coherence was simultaneously calculated and exported to a text file using BVA's Export tool, which allows the user to specify a frequency band of interest; 8-12 Hz for the alpha frequency band. To calculate mean beta coherence, the beta frequency band limits (13-30 Hz) were instead listed in the Export tool.

The exported text files contain mean alpha ERD values, mean alpha and beta coherence values, labels for each value in a csv format. Mean alpha ERD values are labeled by participant number, condition (C1, C2, or C3), and electrode. Mean alpha and beta coherence values are labeled by participant number, condition (C1, C2, and C3), and electrode pair. This labeling system was utilized for organizing data in a Python environment.

2.6 Alpha ERD and Alpha and Beta Coherence ANOVA Analysis in Python

3(group)x3(condition) ANOVA analysis comparing differences of mean alpha ERD, mean alpha coherence, and mean beta coherence values between groups, conditions, and across groups and conditions was computed per electrode or electrode pair. As seen in **Figure 6**, Factor A was proficiency group (NG, IG, and AG) while Factor B was the condition difficulty (C1, C2, and C3).

There were three response variables of comparison: (1) mean alpha ERD, (2) mean alpha coherence, and (3) mean beta coherence. A 3(group)x3(condition) ANOVA was computed per electrode for alpha ERD whereas a 3(group)x3(condition) ANOVA was computed per electrode pair for both mean alpha and beta coherence. In total, there were 20 3(group)x3(condition) ANOVAs comparing mean alpha ERD, and there were 231 for both mean alpha and beta coherence.

Response Variable at Electrode(s) of Interest		CONDITION			
		C1	C2	C3	Group Means
GROUP	Novice				
	Intermediate				
	Advanced				
	Condition Means				Grand Mean

Figure 6: 3(group)x3(condition) ANOVA determining significant differences between EEG feature data between proficiency groups, conditions, and the interaction between proficiency groups and conditions.

Before computing the 3(group)x3(condition) ANOVA analysis, the exported data from BVA was organized in Python DataFrame objects. Participant label numbers were replaced with an integer number proficiency label. Condition labels were also replaced with an integer label number. However, the electrode/electrode pair labels remained The column header of the DataFrame consisted of electrode/electrode pair, proficiency label, condition label, mean alpha ERD value (percentage unit from 0-1), mean alpha coherence value (correlation unit from 0-1), and mean beta coherence value (correlation unit from 0-1).

Python dictionary objects were used to organize the response variable data per electrode/electrode pair. The electrode/electrode pair label served as keys to access response variable data for computing 3(group)x3(condition) ANOVAs. Likewise, the resulting ANOVA data was stored in dictionary objects. A function was made to identify and store any

3 (group) x 3 (condition) ANOVA's that yielded significant results with p-values less than or equal to 0.5. It also identified p-values with 0.01 significance. Using these significance values, four hypothesis questions were asked:

1. Does condition difficulty depend on group proficiency level?
2. Does group proficiency level depend on condition difficulty?
3. Are the dependent variables predictors for group proficiency level?
4. Is there statistical significance between groups given these dependent variables?

To answer these four questions, two sets of hypotheses and null hypotheses will need to be examined. With respect to group proficiency level, H_{P0} and H_{P1} will be examined. For condition difficulty, H_{D0} and H_{D1} will be examined. The character μ denotes the mean value of a dependent variable while its subscripts correspond to appropriate dependent variables. For example, μ_{C1} refers to the Condition 1 mean and μ_N refers to the Novice mean.

$$H_{P0}\colon \mu_N = \mu_I = \mu_A \text{ (Null)}$$

$$H_{P1}\colon \mu_N \neq \mu_I \neq \mu_A$$

$$H_{D0}\colon \mu_{C1} = \mu_{C2} = \mu_{C3} \text{ (Null)}$$

$$H_{D1}\colon \mu_{C1} \neq \mu_{C2} \neq \mu_{C3}$$

2.7 Training a Machine-Learning System

The Machine-Learning system is a supervised model and aims to make classifications based on input parameters. The exact algorithm is still actively being determined as data is analyzed. There are several factors to consider when choosing an algorithm, each with its own advantages and disadvantages. Options include Linear Regression, Naïve Bayes, Decisions Trees, Neural Networks, and Support Vector Machines. **Figure 7** shows the general concept for the neural network achieving Spanish language proficiency classification.

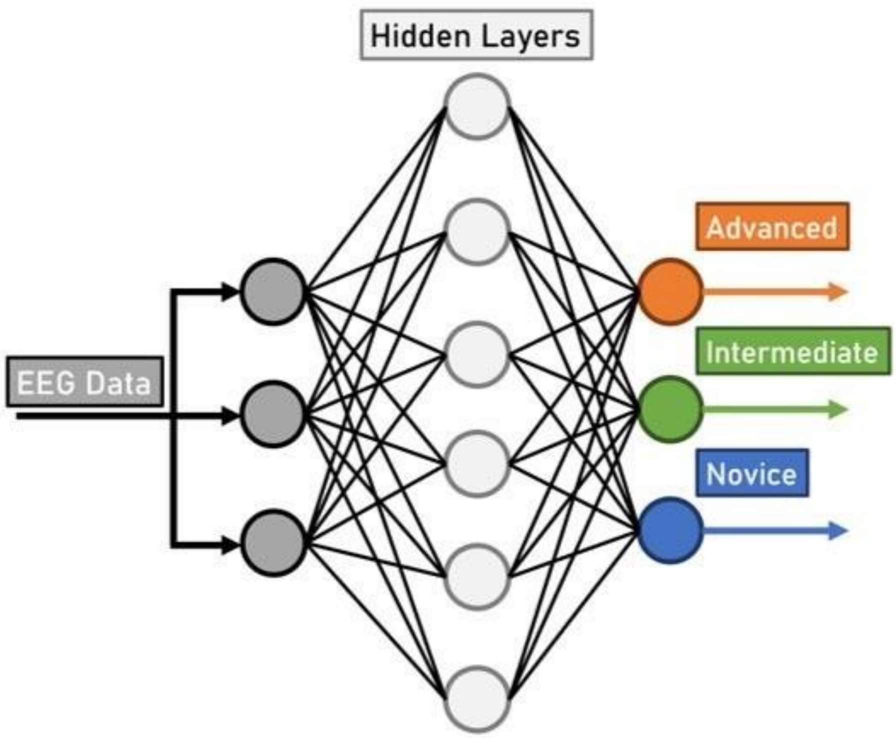


Figure 7: General idea for a machine learning system that would perform Spanish L2 proficiency classification using EEG data features.

The proper system to use depends on the nature of the resulting data, and it cannot be determined without first performing a thorough data analysis. Currently, the system is being implemented using a basic neural network. If accuracy and performance are not ideal, a different model may be used. The model is implemented using Python and appropriate libraries.

Other development tools such as Git, JIRA, Bitbucket, and Visual Studio Code are aiding in the process. These tools are used for version control and providing consistency between local development environments. Once the model is prepared, a standard 80/20 split of the collected data to train and test the algorithm. This designates 80% for training data and 20% for testing the model.

2.8 Cost Analysis

The total project cost and itemized breakdown is shown in **Table 3**. Most materials purchased were for data collection and proper equipment maintenance. The headphones, conducting gel, syringe tips, blunt Teflon needles, and disposable gloves were all used during data collection.

Similarly, the electric toothbrush heads, shampoo & conditioner, and Q-tips were used to clean equipment after each session.

The Amazon gift cards were the primary incentive for gathering participants, but the quantity should have been reduced. An overestimation of participants resulted in unused gift cards. To avoid this issue, the gift cards were purchased in batches across several months. However, the batches sizes should have been reduced more closely match the number of subjects.

3 RESULTS

3.1 Comprehension and Engagement Questionnaires

There was a linear trend between the average total comprehension score across conditions and proficiency level. Each participant could earn a maximum of four comprehension points from each condition; thus, they could earn 12 comprehension points in total. To observe the trend between proficiency level and comprehension across conditions, the total comprehension score was average within all proficiency groups. From least to greatest in average group proficiency scores, the NG scored 3.75, Intermediate 7.13, and Advanced 9.72 (**Figure 8**). **Figure 8** shows that proficiency level was directly proportional with average total comprehension score with an R-squared coefficient of 0.9943.

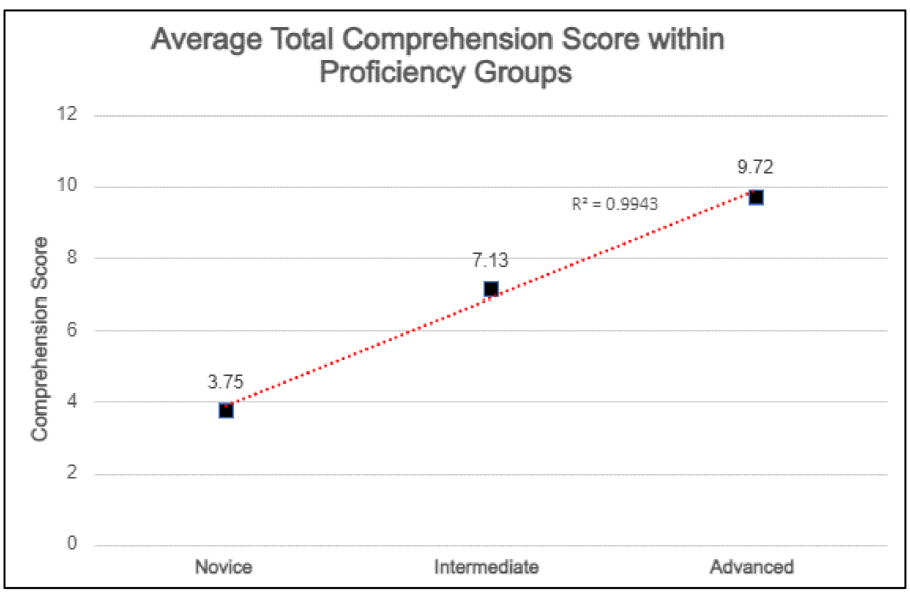


Figure 8: Trendline showing the strong, positive, linear correlation ($R^2 = 0.9943$) between average total comprehension score and proficiency group. The total comprehension score is the sum of all comprehension scores from conditions C1, C2, and C3. This total comprehension score was then averaged within groups.

Although, there was no discernable relationship between proficiency level and average total engagement score. The correlation between proficiency level and average total engagement score, $R^2 = 0.3148$, was positive yet weak. Overall, the NG showed the lowest average total engagement with a score of 12, the AG showed the second greatest with 12.75, and then lastly the greatest of 13.33 within the IG. However, all groups did report a high average total engagement score overall, as the maximum possible engagement score is 15 points.

3.2 Alpha ERD 3x3 ANOVA

Only two electrode sites showed a significant difference in alpha ERD. Electrode AF3 had significance between groups (NG, IG, AG, $p \leq 0.05$) whereas electrode T8 showed significance within groups ($p \leq 0.05$). As seen in AF3's interaction plot (**Figure 9**), the AG had the greatest amount of alpha ERD, followed by NG, and then closely followed by IG. Although no significance between groups were produced at any other electrode site, many exhibited the general trend ranking AG, NG, and then IG with the highest-to-lowest alpha ERD.

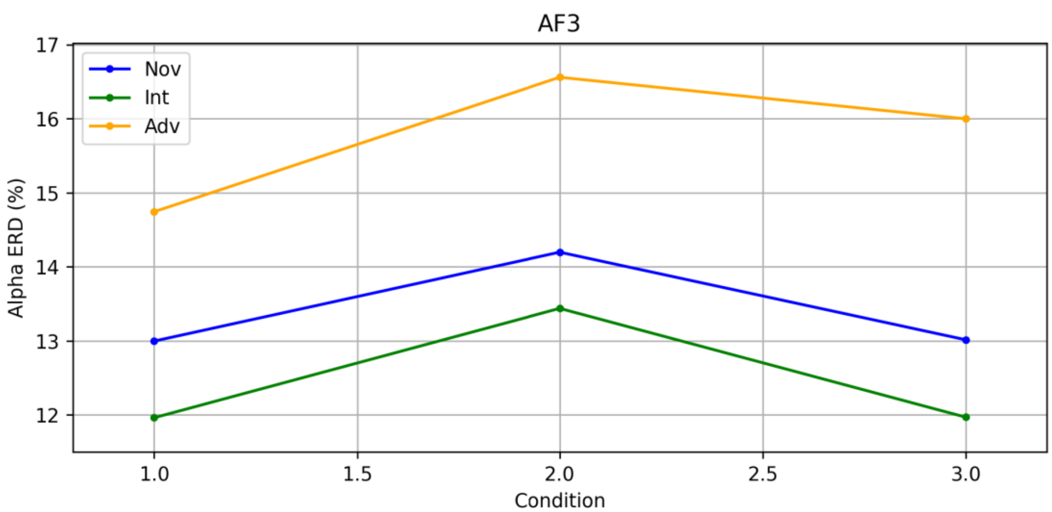


Figure 9: Interaction plot comparing the mean values of alpha ERD per condition of each proficiency group at electrode site AF3. The mean alpha ERD from least to greatest for all conditions is Intermediate, Novice, and Advanced.

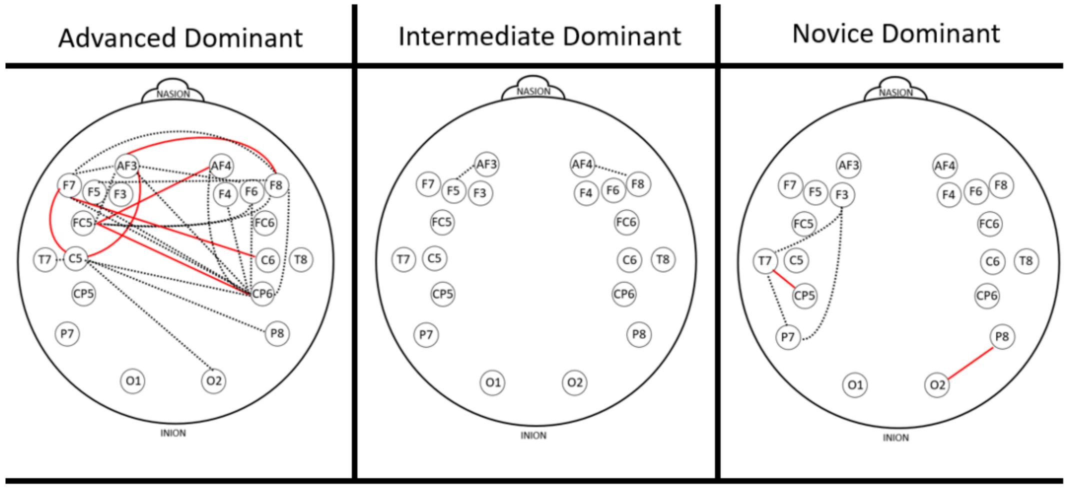
The latter significance result in T8 was rejected because the NG did not have a mean alpha ERD for any conditions. The electrode was eliminated from the NG—and thus eliminated from data analysis—due to large amounts of noise perturbing the T8 channel.

No significance between groups, within groups, nor interaction between groups and conditions were observed besides in electrode AF3.

3.3 Alpha Coherence 3x3 ANOVA Analysis

A total of 31 and 37 electrode pairs yielded significant differences in coherence between groups for the alpha and beta bands, respectively. In examining alpha coherence, seven pairs had p-values $p \leq 0.01$ while 24 pairs had p-values $p \leq 0.05$. The interaction plots highlight a general trend with the AG having the greatest coherence mean of the electrode pairs (76.67% of the time), followed by Intermediate (56.67% of the time), and then by Novice (50.00% of the time). For beta coherence, the AG also has the greatest coherence mean of the electrode pairs (62.16% of the time). However, the IG and NG had the second highest beta coherence in electrode pairs (36.14% of the time, each), and then followed by the IG for least beta coherence (56.76% of the time). In general, the AG had the greatest alpha and beta coherence means among the three conditions within electrode pairs while NG and IG had the least, respectively.

Alpha coherence between the frontal and central cortices were significantly greater in the AG than the Intermediate and NGs. In six electrode pairs the AG showed alpha coherence dominance where $p \leq 0.01$. In **Figure 10**, electrode C5, belonging to Wernicke's Area, demonstrates this AG coherence dominance with electrodes F7 and AF3, both of which belong to Broca's Area. Electrode FC5, belonging to Broca's Area, also acts as a sink for alpha coherence dominance bilaterally with electrodes AF4 and CP6. This bilateral AG alpha coherence dominance is again seen between electrodes AF3 (rim of Broca's Area) and F8. Furthermore, this bilateral alpha coherence between the frontal cortices, central cortices, and cross frontal and central cortices is supported by 14 significance pairs with $p \leq 0.05$. Among all significance pairs demonstrating AG coherence dominance, electrodes C5, CP6, and FC5 act as sinks. Additionally, electrode C5 interacted with the temporal T7 electrode and the right posterior region electrodes P8 and O2.



LEGEND	
▪▪▪	$0.01 < p \leq 0.05$
▬	$p \leq 0.01$

Figure 10: Significant alpha band coherence dominance divided by proficiency group. The dotted black and the bolded red lines indicate 0.05 and 0.01 significance, respectively.

The NG showed alpha coherence dominance extending from the left temporal cortex, between the left parietal and left frontotemporal cortex, and between the right posterior and occipital cortices. Electrode T7 acted as a coherence sink for electrodes F3, CP5, and P7 (**Figure 10**). Of these pairs, only the alpha coherence between T7 and CP5 reached $p \leq 0.01$ significance while T7 pairing with F3 and P7 reached 0.05 significance. This indicates NG alpha coherence dominance between the left temporal cortex and Wernicke's Area. As for Broca's Area, NG alpha coherence dominance between Broca's Area and the left parietal cortex is demonstrated by the $p \leq 0.05$ significance paring between electrodes F3 and P7. Lastly, the $p \leq 0.01$ significance paring between electrodes P8 and O2 shows the NG's alpha coherence dominance between the right posterior and occipital cortices.

The IG showed alpha coherence dominance only in two electrode pairings occurring in the frontal cortices. Both electrode pairs, AF3 and F5, and AF4 and F8, had $p \leq 0.05$ significance. As seen in **Figure 10**, these Intermediate alpha coherence dominance pairings appear to mirror bilaterally.

3.4 Beta Coherence 3x3 ANOVA Analysis

Beta coherence dominance is more widespread in the AG than in the Intermediate and NGs. In this context, widespread coherence is defined as coherence between electrodes belonging to cortices that are not neighboring each other. Exhibiting AG beta coherence dominance, six and 16 electrode pairs reached $p \leq 0.01$ and $p \leq 0.05$ significance, respectively. These significance pairs have wide-reaching, bilaterally-crossing coherence between several ROIs, including Broca's Area and the right central-parietal cortex, Broca’s Area and the occipital cortex, and the right frontal cortex and the left occipital cortex (**Figure 11**) Electrode CP6 is a sink for nine of these 22 significant pairs. CP6, of the right central-parietal cortex, coordinates with electrodes AF4, F4, and F6 within the right frontal cortex, AF3, F5, F7, and FC5 within the left frontal cortex, C5 within the left central cortex, and O1 in the left occipital cortex. Of these significance pairs with CP6, two of them, F5 of Broca’s Area and C5 of Wernicke’s Area, reached 0.01 significance. Not only does the AG demonstrate beta significance dominance bilaterally between Broca’s Area and the right central-parietal cortex, it also has more significant pairs than both Novice and Intermediate within these ROIs.

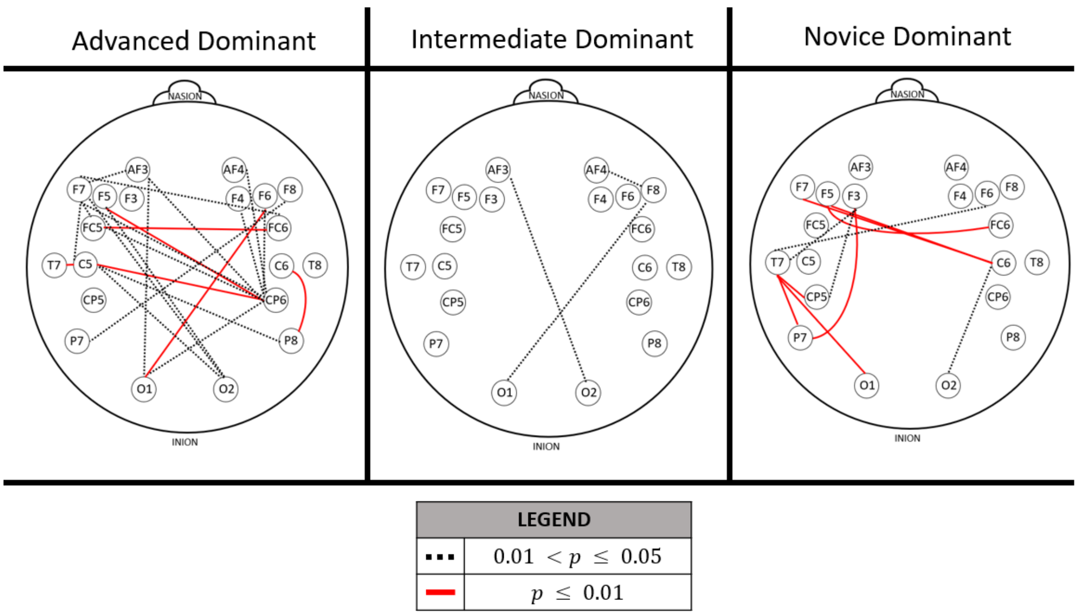


Figure 11: Significant beta band coherence dominance divided by proficiency group. The dotted black and the bolded red lines indicate 0.05 and 0.01 significance, respectively.

But it is also true that the AG demonstrated beta coherence dominance between laterally neighboring ROIs. Electrode pair F7 and C5 ($p \leq 0.05$) showed interaction between Broca and Wernicke’s Areas, C5 and T7 ($p \leq 0.01$) showed interaction between Wernicke’s Area and the left temporal cortex, and C6 and P8 ($p \leq 0.01$) showed interaction between the right central and parietal cortices. Furthermore, AG beta coherence dominance was seen within Broca’s Area, as electrode F7 has reached $p \leq 0.05$ significance with electrodes FC5 and AF3. Although the AG had beta coherence dominance between neighboring ROIs and within ROIs, its prevalence of beta coherence dominance was mainly observed bilaterally between ROIs.

The prevalence of significant NG beta coherence was demonstrated both bilaterally and unilaterally (**Figure 11**). Electrodes F7 and F5 collectively serve as the sink between Broca's Area and the right central cortex electrodes C6 and FC6. Both F7 and F5 showed $p \leq 0.01$ significance with C6, and F5 also showed $p \leq 0.01$ significance with FC6. Another bilateral pair, T7 and F6 ($p \leq 0.05$), was between the left temporal and right frontal cortices. Unilateral significant beta coherence pairings showing NG beta coherence dominance were mainly located between the temporal, central, and posterior ROIs. Electrode T7 acted as a sink for communicating with the left parietal (P7) and occipital (O1) cortices. The significance of NG beta coherence dominance reached $p \leq 0.01$ significance pairs T7 and P7, and T7 and O1. Also in the left hemisphere, there were two $p \leq 0.05$ significant NG beta coherence dominance pairs extending from electrode F3 to electrodes T7 ($p \leq 0.05$) and CP5 ($p \leq 0.05$). Additionally, there was a $p \leq 0.01$ significance pair between F3 and P7, indicating a strong beta coherency among Novices between Broca's Area and the left parietal cortex. In the right hemisphere, a 0.05 significance beta coherence pair, C6 and O2 ($p \leq 0.05$), indicate NG beta coherence dominance. In comparison to AG beta coherence dominance, NG beta coherence dominance is more prevalent in closely neighboring regions unilaterally and the frontal and central cortices bilaterally.

The IG featured small amounts of beta coherence dominance both widespread between and localized within ROIs (**Figure 11**). There were two widespread pairs with $p \leq 0.05$ significance: AF3 and O2 ($p \leq 0.05$), and F8 and O1 ($p \leq 0.05$). In comparison to other groups, only the Advanced proficiency group also showed significant beta coherence dominance between the occipital and frontal cortices. As for the localized beta coherence significance pairing, only one pair, AF4 and F8, reached a significant p-value ($p \leq 0.05$). Unlike alpha coherence pairing, Intermediate experienced the least prevalence of beta coherence dominance. Although, it was both widespread between bilateral and localized within unilateral ROIs.

4 DISCUSSION

4.1 Appropriate Audio Selections for Conditions

Score results from the comprehension questionnaire verified that the audio selections for the conditions were appropriate. The purpose of the comprehension questionnaire was to ensure that the audio selections for conditions C1, C2, and C3 correctly corresponded to Novice, Intermediate, and Advanced proficiency level difficulty, respectively. Participants could earn a maximum of four points per questionnaire; so, it was expected that the NG would score around 2-6 points, Intermediate 6-10 points, and Advanced 8-12 points. The corresponding group comprehension score averages were 3.75, 7.13, and 9.72 with a strong, positive correlation of $R^2 = 0.9943$ (**Figure 8**). This directly proportional relationship between proficiency level and comprehension score demonstrates that the condition difficulties were appropriate.

In contrast, average total engagement scores did not correlate strongly with proficiency (R^2 = 0.3148). All groups' average engagement score, which can be a maximum of 15, was between 12 and 13.33 (**Figure 12**). The purpose of the engagement questionnaire was to understand if participants were "spacing out" during a condition. Being actively attentive to a four-minute audio recording of a foreign language may prove difficult—especially if its content is beyond one's proficiency level. It appears each proficiency group had the same amount of engagement across all conditions. However, engagement with each individual recording does have strong, inversely proportional trend: as condition difficulty increases, engagement scores across proficiency groups decrease (R^2 = 0.9734).

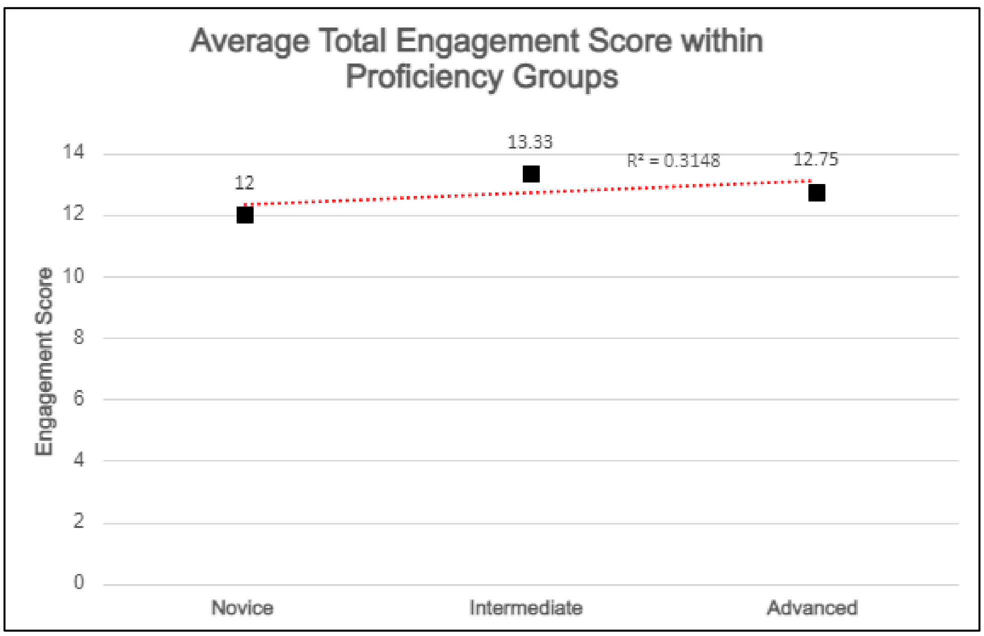


Figure 12: Trendline showing the weak, positive correlation ($R^2 = 0.3148$) between average total engagement score and proficiency group. The total engagement score is the sum of all engagement scores from conditions C1, C2, and C3. This total comprehension score was then averaged within groups.

This negative relationship between condition difficulty and average engagement score (**Figure 13**) across groups may explain the general trend of the mean alpha ERD interaction plots in **Figure 14** and **Figure 15**. **Figure 14** shows how mean alpha ERD changes with conditions C1, C2, and C3 amongst for each proficiency group. In most cases for Novice, Intermediate, and AGs, the mean alpha ERD peaks during C2. According to the neural efficiency hypothesis, the peak alpha ERD would be expected to occur at C3—the most difficult audio selection. However, alpha ERD may be less during C3 *because* the audio selection is too difficult. Based on the negative relationship between condition difficulty and mean engagement scores, participants may be "spacing out" during C3 because they do not understand everything. Instead, alpha ERD may be more indicative of Spanish L2 focus and or understanding.

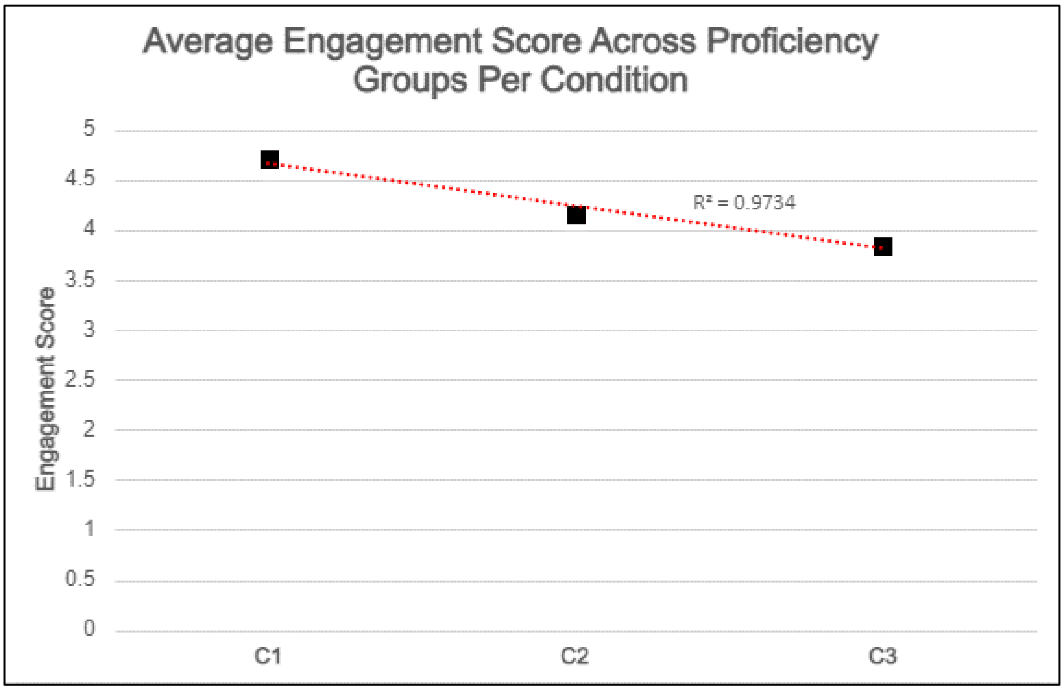


Figure 13: Trendline showing the strong, negative, linear correlation ($R^2 = 0.9734$) between average engagement score and condition difficulty. The average engagement score is across all proficiency groups.

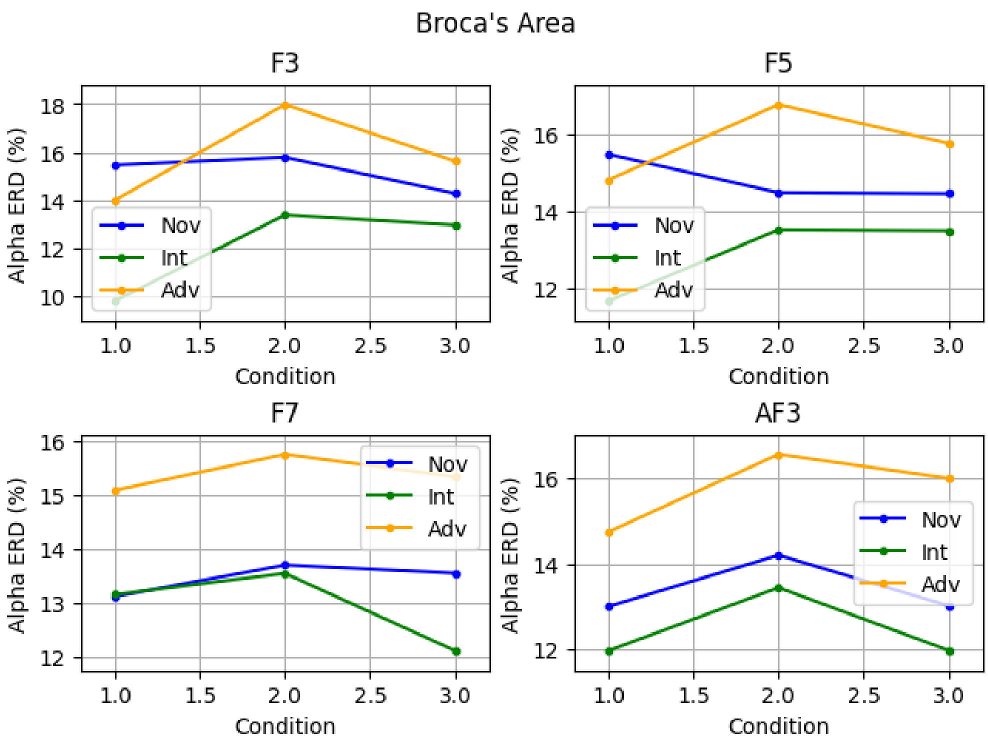


Figure 14: Interaction plots comparing the mean values of alpha ERD per condition of each proficiency group at electrode sites within the Broca's Area ROI. The mean alpha ERD from least to greatest for all conditions is Intermediate, Novice, and Advanced. The mean alpha ERD generally peaks during C2 among proficiency groups.

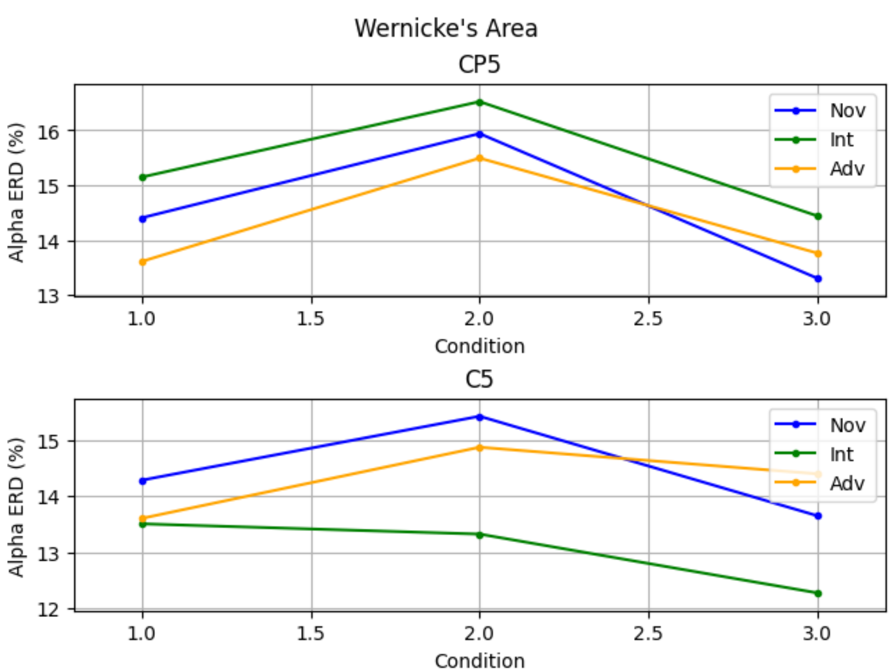


Figure 15: Interaction plots comparing the mean values of alpha ERD per condition of each proficiency group at electrode sites within the Broca's Area ROI. The mean alpha ERD generally peaks during C2 among proficiency groups.

4.2 Lack of Alpha ERD Significance

Of 20 electrodes of interest, only electrode AF3 had a significant difference of alpha ERD between proficiency groups ($p \leq 0.01$). The trend with this significance showed that the Advance group was dominant in alpha ERD, followed by the Novice, and then by Intermediate. However, as seen in the Box and Whisker plots of **Figure 16**, the AG had high-value alpha ERD outliers in conditions C1, C2, and C3. Therefore, the mean alpha ERD per condition within the AG is liable to be positively skewed. Furthermore, the Box and Whisker plots in **Figure 16** is an extreme example highlighting a general trend across most electrode sites. At most sites, both the Novice and AGs had high-value alpha ERD outliers in one or more conditions. On the contrary, the IG had more uniform distributions and some low-value alpha ERD outliers in one or more conditions. These outliers may be influencing the observed trend at electrode AF3 showing AG alpha ERD dominance, followed by Novice, and then Intermediate. Thus, due to the small sample sizes and outliers, it is unlikely that there exists actual alpha ERD significance between groups at electrode AF3.

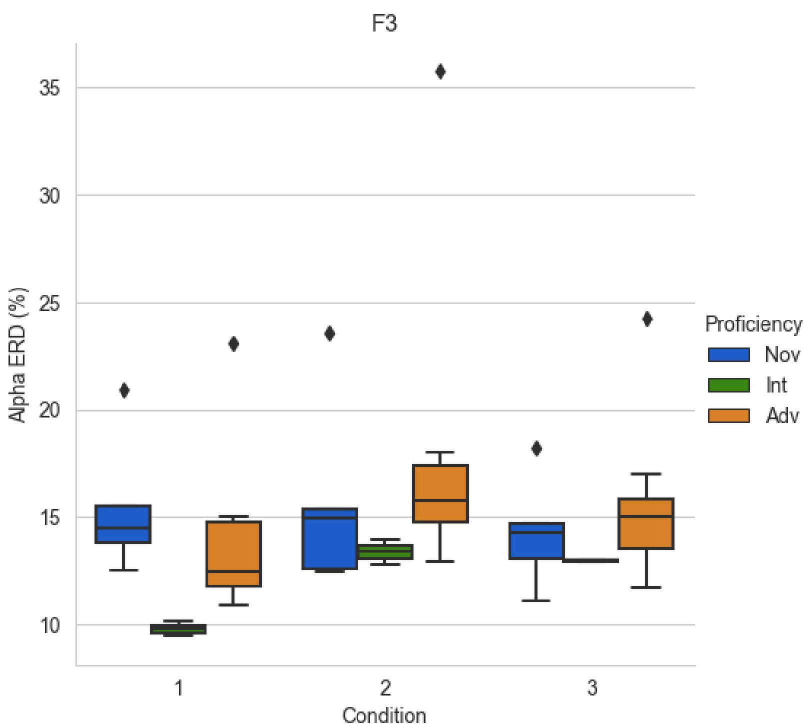


Figure 16: Box and whisker plot comparing distributions of alpha ERD among conditions C1, C2, and C3 between proficiency groups at electrode site F3. This figure highlights a general trend noticed across other electrode sites: both the Novice and AGs had high-value alpha ERD outliers (indicated by diamond) within conditions whereas the distributions for the IG were more uniform. These outliers may be positively skewing the mean alpha ERD in Novice and AGs.

Alpha ERD significance between groups at electrode sites may be better determined if outliers were eliminated from the 3x3 ANOVA calculation. However, due to the small population size within the three groups (six Novice, six Intermediate, eight Advanced), these outliers should not be excluded. If the sample size were larger—if each group had 15 participants—then it would be acceptable to exclude these outliers.

4.3 Bilateral Coherence Dominance as a Predictor for Spanish L2 Proficiency

Although the coherence significance results are contrary to the neural efficiency hypothesis, language proficiency may be better identified in the context of the frontoparietal network. The frontoparietal network serves as a hub for driving cognitive focus [22], [23]. Furthermore, this network may be accessed globally, and it is not necessarily relegated to one function [24]. But the development of the frontoparietal network does often occur in parallel with the language centers Broca and Wernicke's Area in developing children [25]. In fact, children are shown to rely on networks connecting the frontal and temporal cortices whereas adults show a shift to rely on the frontal and parietal networks [25]. In this study, because the Spanish L2 are in college and closer in adult age, they would most likely rely more on the frontal and parietal cortices. Other fMIR have identified that the parietal cortices (the inferior parietal lobule) communicates with frontal language centers like Broca's Area along the superior longitudinal fasciculus pathways [26]. Also, Hämäläinen et al. [27] found that late-age sequential bilinguals relied more on the fronto-occipital fasciculus network. Thus, bilateral and wide-reaching alpha and beta coherence dominance in the AG may be more indicative of greater focus than proficiency.

Alpha coherence dominance stemming from the left temporal and parietal cortices and the right parietal and occipital cortices distinguish the NG from the AG. As seen in **Figure 10**, the NG relies more on the posterior and central cortices while the AG relies more on the central and frontal cortices. In the NG, the alpha coherence dominance between Broca's Area and the left parietal cortex suggests greater focus between language and posterior ROIs. The NG alpha coherence dominance between the right parietal and occipital cortices also suggest this, too. Specific to language function, the alpha coherence dominance between T7 and CP5 indicate that more areas within the left temporal cortex (including Wernicke's Area) is working together toward language comprehension.

An increase in proficiency level is associated with more unilateral and bilateral alpha coherence dominance between frontal and central cortices. An example of unilateral AG alpha coherence dominance among the frontal and central cortices is with C5 electrode of Wernicke’s Area working with the F7 and AF3 electrodes of Broca’s Area. The IG also shows unilateral alpha coherence dominance in both the right and left frontal cortices. Bilateral AG alpha coherence is demonstrated between electrode CP6 of the right central cortex and electrodes FC5, F7, F5, and AF3 of Broca’s area. This suggests that as Spanish L2 language proficiency increases, alpha coherence dominance shifts from the posterior and temporal brain regions to the central and frontal ones.

Higher Spanish L2 proficiency is associated with more widespread beta coherence dominance. As detailed in **Section 4.4**, the AG has more significant beta coherence dominance pairs between the frontal cortices and both the parietal and occipital cortices than the Intermediate and NGs. Furthermore, the AG features more bilateral beta coherence, with electrodes CP6 and C5 functioning as the main sinks for these connections. Interestingly, the AG shows beta coherence dominance between Wernicke’s area (electrode C5) and the right central (electrode CP6), parietal (electrode P8), and occipital (electrode O2) cortices. Also, beta coherence between Broca’s Area (electrodes F5 and F7) and the right occipital (O2) cortex demonstrates AG dominance.

NG beta coherence dominance shows that beta coherence among low proficiency Spanish L2 speakers mainly extends from and occurs within the left hemisphere. Of 11 significant pairs, only one occurs between two electrodes (O2 and C6) in the right hemisphere (**Figure 11**). Bilaterally, electrodes F7 and F5 of Broca’s Area serve as sinks for the right frontal and central cortices (electrodes F6, FC6, C6). But what most distinguishes the NG from the Intermediate and AGs is the beta coherence dominance extended from the left temporal cortex (electrode T7) to Wernicke’s Area (electrode CP5), and both the left parietal (electrode P7) and occipital (electrode O1) cortices. Additionally, NG beta coherence dominance is seen between Broca’s Area (electrode F3) and the left parietal (electrode P7) cortex. Among Novice participants, Broca and Wernicke’s Areas show more beta coherence dominance within the left hemisphere, and more specifically with the posterior ROIs. This suggests that as Spanish L2 language proficiency increases, beta coherence dominance transitions from being unilaterally localized in the language centers of the left hemisphere to extending more widespread bilaterally.

4.4 Limitations and Challenges

There were several limitations in this study. The greatest limitation was the size of the participant pool. In total, there were 20 participants who were learning sequentially learning Spanish as a second language. Split into three proficiency groups, the population sizes of each group were small: six in NG, six in IG, and eight in AG. Because of the small population sizes, it would not have been appropriate to replace outliers in with the median of the group. Also, the small group size indicate that these significant results may not be replicable.

Data had to be excluded from the study due to noise. EEG is inherently noisy. It is susceptible to electrooculogram (e.g., blinking) and electromyogram (e.g., moving face muscles) artifacts. These types of artifacts pervade almost every instance of EEG data collection.

Computing the grand-average alpha ERD per condition may have limited its ability to distinguish Spanish L2 proficiency. Instead of viewing alpha ERD as a general cognitive state over the course of a four-minute recording, it could be viewed dynamically. Change in alpha power over time may be a better data feature than a grand-averaged mean.

5 FUTURE WORK

5.1 Participant Pool & Group Labels

A larger participant pool should be used to create a machine learning system that classifies L2 proficiency based on EEG data features. To find significant, repeatable results, it is recommended to have at least 15 participants per proficiency group (NG, IG, and AG). But, to make language proficiency classifications, a machine learning model would most likely require a much larger sample size than 15 per group. EEG data varies widely on an individual level. By having a larger sample size, variations in the EEG data has a greater chance of being smoothed by averaging.

Significant differences in alpha and beta coherence were seen between the AG and NG. But the amount of alpha and beta coherence dominance pairs in the IG was small and difficult to interpret. A future study may find more distinguishable differences between proficiency groups if the number of groups were reduced to two. It may be beneficial to first identify differences between two extreme differences in proficiency; then, further work could make smaller distinctions between groups.

5.2 EEG Data Feature Consideration

There were three overarching EEG data features: alpha ERD per electrode, alpha coherence per electrode pair, and beta coherence per electrode pair. Per each electrode, there was an alpha ERD value for each of the three conditions. So, averaged for each proficiency group, there were 60 alpha ERD values labeled by audio condition (20 per condition). As for alpha coherence, there are 190 electrode pairs. Recording an average alpha coherence value among proficiency groups per condition, there are a total of 570 alpha coherence values labeled by audio condition (190 per audio condition). 570 beta coherence values labeled by condition, too. Per subject, there are 1200 values with unique labels: electrode or electrode pair, condition, and the target label of proficiency group. These totals illustrate how the number of EEG feature values may multiply based on the number of electrodes, conditions, and groups of interest.

But all these values are static averages. Each alpha ERD and alpha and beta coherence value is the resulting average over the course of a four-minute audio condition. It may be fruitful to examine these values as they dynamically change with time. Specifically, observing the change in alpha power would indicate changes in focus or attention to the audio selection within conditions. So, focus or attention may be a factor predicting Spanish L2 proficiency.

5.3 Designing a Machine Learning System to Classify Spanish L2 Proficiency

A machine learning system that classifies Spanish L2 proficiency should consider using EEG data features should consider incorporating alpha ERD and especially alpha and beta coherence data. Detailed in **Section 6.2**, using time-dynamic features may prove more fruitful. Instead of processing this data in BVA, future researchers should consider processing EEG data using the MNE Python open-source library. Its intended use is for processing, visualizing, and analyzing human neurophysiological data including EEG. This provides robust algorithms and models for EEG machine learning applications.

6 CONCLUSION

This body of research demonstrates that Spanish L2 proficiency may be predicted based on patterns in alpha and beta coherence dominance. Higher proficiency Spanish students showed significantly greater coherence bilaterally between the left and right hemispheres of the brain. Specifically, the higher proficiency students showed alpha coherence dominance between language centers in the left hemisphere and the right frontal, central, and posterior cortices. A similar trend is seen within the beta band, for higher proficiency Spanish L2 students showed greater beta coherence bilaterally between language centers and the frontal, central, and posterior cortices. In both the alpha and beta bands, the lower proficiency Spanish L2 students exhibited alpha and beta coherence dominance localized unilaterally. Localized, unilateral alpha and beta coherence was most prevalent between the language centers and the left temporal and posterior cortices of the brain. The delineation of widespread, bilateral coherence between higher and low proficiency students suggests that the higher proficiency Spanish L2 students relied on or utilized more of the frontoparietal network in language comprehension.

A machine learning system incorporating EEG coherence data features may be used to classify Spanish L2 proficiency. Language acquisition is paramount in today’s international collaboration within industry, education, and commerce. This work suggests that a machine learning system that classifies second language proficiency may be used as a language assessment tool for language students, language instructors, and international professionals.

7 STATEMENTS

7.1 Availability of Data

Data used in this study—without participant labeling—study may be procured for anyone upon request by contacting the e-mail of Blaise O’Mara.

7.2 Ethical Considerations: Human Participants

This study numbered IRB-FY2022-276 was approved by the UNH Institutional Review Board.

When dealing with human subjects, strict ethical conformance is required. To protect participants’ privacy, several procedures were implemented. All subjects were given aliases. These aliases were used in data collection, processing, and analysis. There was only one master file that contains the correlation between subject names and their alias. Additionally, any files with private information were kept in a locked chest. Principal investigators and advisors are the only people with the code. Documents containing personal information were properly stored in the chest at all times and were not seen by anyone outside the project.

The eligibility form screened subjects for various health conditions that must be kept private. To avoid asking potential participants to divulge personal information, the eligibility form was sent electronically before they came to the laboratory. It was explicitly stated that each subject should thoroughly read the form and only continue the process if they meet all the criteria.

7.3 Proposed Timeline of Tasks and Milestones

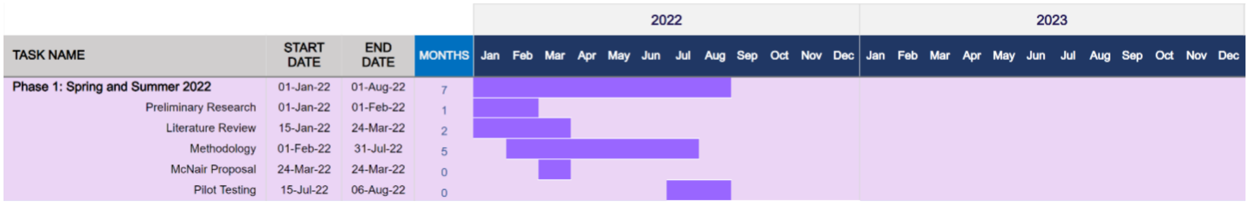


Figure 17: Proposed Phase 1 timeline.

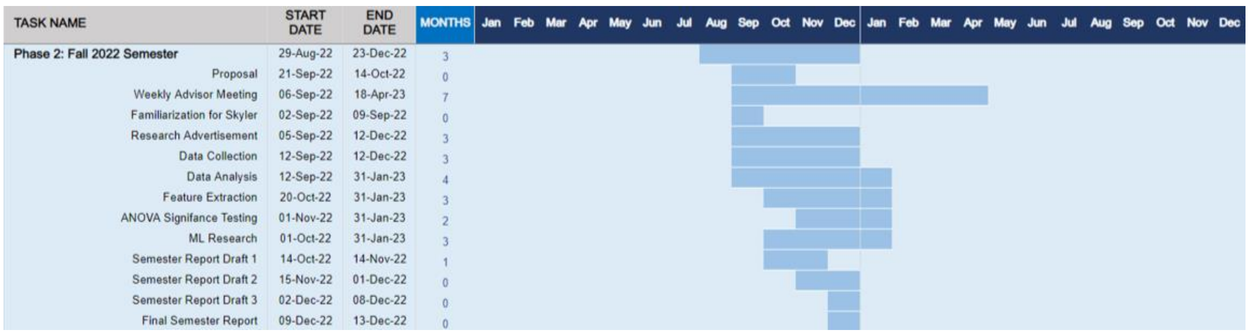


Figure 18: Proposed Phase 2 timeline.

Figure 19: Proposed Phase 3 timeline.

7.4 Budget & Bill of Materials

Budgets	
McNair Funds	$710.00
ECE Funds	$1,080.00
Senior Project Funds	$0.00
Kinesiology Funds	$706.72
TOTAL	***$2,496.72***

Table 1: Funding sources.

Item/Description	Quantity	Unit Price	Cost	Funding Source
$15 Amazon Gift Card	42	$15.00	$630.00	McNair
$30 Amazon Gift Card	37	$30.00	$1,110.00	ECE Department ($1,080), McNair ($30)
Headphones	1	$50.00	$50.00	McNair
Conducting Gel	5	$85.00	$425.00	Kinesiology Department
Syringe Tips	3	$15.00	$45.00	Kinesiology Department
Blunt Teflon Needle	3	$65.00	$195.00	Kinesiology Department
Electric Toothbrush Heads	1	$9.99	$9.99	Kinesiology Department
Disposable Gloves	1	$6.28	$6.28	Kinesiology Department
Shampoo & Conditioner	1	$18.35	$18.35	Kinesiology Department
Q-tips	1	$7.10	$7.10	Kinesiology Department
TOTAL			***$2,496.72***	

Table 2: Complete budget detailing items and quantities.

8 STANDARDS UTILIZED

To define Second-Language Proficiency groups, standards set forth by the American Council on the Teaching of Foreign Languages (ACTFL) and the Common European Framework of Reference for Languages (CEFR) were referenced. The classifications of proficiency were based on ACTFL guidelines and provided a framework for

Collection of EEG data followed the 10/20 standard, which refers to the distances between adjacent electrodes. This indicates that each electrode is either 10% or 20% of the total distance of the skull. This standard is utilized to ensure the electrodes are placed on the proper location of the scalp. Due to the dependence of EEG data and the location of the scalp, this standard was crucial.

9 INDIVIDUAL CONTRIBUTIONS

9.1 Skyler Baumer

Skyler Baumer, co-principal investigator, contributed to all aspects of the project that took place during the Fall and Spring semesters. Skyler joined the project in the beginning of September and was quickly brought up to speed by Blaise. Soon after, the responsibilities were split evenly between him and Blaise. Skyler and Blaise evenly shared the task of collecting and processing data in BVA. During the January break, Skyler focused on setting up the environment for analysis and machine learning implementation. He prepared a Bitbucket repository, a JIRA page, and helped Blaise learn how to use all the development tools. He also developed the preliminary script that parsed the data files and sorted them in a meaningful way. Skyler has contributed most prevalently to the programming aspects of the project.

9.2 Blaise O’Mara

Blaise O’Mara, co-principal investigator, must perform additional work to merit the additional 2 credits awarded to honors students. To earn these Honors credit hours, Blaise will be writing and submitting a thesis summary of this research to both the UNH Honors and McNair Scholars Programs. Furthermore, the quality of the thesis will be to publishable standards. For the McNair Scholars Program, the research article must be submitted by March 31st, 2023 to the UNH Inquiry Journal.

Working with Dr. Croce and Dr. Smith, Blaise aims to publish in two separate journals: one in the field of neuroscience, and the other in the field of machine learning. The purpose of the research is (1) to understand how Spanish language proficiency may be predicted with EEG neural correlates, and (2) to build a machine learning algorithm that automatically performs this Spanish language proficiency classification based on EEG feature data. So, the findings of the first aim will be published in a neuroscience journal, whereas the results of the second aim will be published in a machine learning journal. Which two journals for submission are not certain; however, prospective journals for neuroscience are Frontiers in Language Sciences or Neuroinformatics. As for machine learning, prospective journals include IEEE Transactions on Pattern Analysis and Machine Intelligence, and Elsevier Neuroscience Informatics or Machine Learning with Applications.

10 SUPPORTING DOCUMENTS

10.1 Screening Form

IRB-FY2022-276

EEG-based Second Language Proficiency: A Database for ACTFL Guidelines

Participant Demographic and Screening Form

Study ID#: ______________________ Date: ____________

Sex: M _____ F_____

Date of birth: ____________________

Are you a Sequential Bilingual?

YES____ NO____

Spanish Language Education (# of years at UNH):______________________

Highest Completed UNH Spanish Course Number: ______________________

Do you have…	**YES**	**NO**
Hearing problems?		
If yes, can they be corrected (e.g., with hearing aid)?		
History of any of the following: • Head trauma such as stroke or brain lesions (including multiple confirmed concussions or moderate to severe traumatic brain injury ["TBI"]), prior intracranial surgery, learning disability or attention deficit/hyperactivity disorder, neurological diagnoses, or other issues that may interfere with normal brain function; • Serious, chronic mental illness or disorder; • Brain-altering drug use (e.g., LSD, cannabis, anti-depressants).		
Learning disability or attention deficit/hyperactivity disorder?		

Electrode Cap Measurements		
Head Diameter [cm]	**Nasal Dorsum-to-Occiput [cm]**	**Ear-to-Ear [cm]**

IRB-FY2022-276

Please initial here __________ as indication that you have received a copy of the consent form.

______________________________ Signature of Research Participant ____________________ Date

______________________________ Signature of Investigator ____________________ Date

10.2 Novice-Level Comprehension and Engagement Questionnaire

Comprehension & Engagement Questionnaire

The purpose of this survey is to assess your comprehension and engagement with the last audio selection.

Participant ID: ________________________

Proficiency Level (circle one): Nov / Int / Adv

Comprehension Questions

Please provide no more than one sentence for your answers to the following questions.

1. Who is the main character of the story?

2. What are some traits of Pollito Tito?

3. What was Pollito Tito doing in the morning?

4. Why was everyone worried?

Engagement Questions

For each of the following statements, please circle the response that best characterizes how you feel about the statement.

1. Strongly Disagree
2. Somewhat Disagree
3. Neither Agree nor Disagree
4. Somewhat Agree
5. Strongly Agree

	Strongly Disagree	**Somewhat Disagree**	**Neither Agree nor Disagree**	**Somewhat Agree**	**Strongly Agree**
I was actively engaged with the story for the entire 4 minutes.	1	2	3	4	5
I found the story difficult to comprehend.	1	2	3	4	5
I was interested in the story.	1	2	3	4	5
I think others at my Spanish proficiency level should be able to understand the main ideas of the story.	1	2	3	4	5

11 ACKNOWLEDGEMENTS

The UNH McNair Scholars Program, UNH Electrical and Computer Engineering Department, and UNH Kinesiology Department all played a significant role in supporting this project. All funding and equipment came from these departments.

Dr. Pulecio, Dr. Cashman, Dr. Munoz Pina, Dr. Lee, and Dr. Frost from the UNH Spanish Department were instrumental to the design and advertisement of the study. Dr. Pulecio helped in selecting audio stimuli, map proficiency groups within the UNH curriculum, and encouraging students to participate in the study. Dr. Munoz Pina, Dr. Lee, and Dr. Frost allowed the PIs to speak during their lectures in order to advertise the study. Dr. Cashman assisted with contacting professors within the department and helped us determine which classes would be best to advertise. Finally, Mrs. Babbin, the administrator of the Languages, Literature, and Cultures Department, advertised the study by sending e-mails to all current Spanish language students. All were incredibly enthusiastic about the study and undoubtedly improved the number of subjects.

The UNH students that participated should be recognized for their irreplaceable role in the project. They provided their time, energy, and enthusiasm.

Lastly, Dr. Croce and Dr. Smith are recognized for their roles as superb mentors. They were instrumental in designing the experiment paradigm. Each provided numerous hours in counselling, troubleshooting, and teaching.

Immense appreciation is extended to all who helped in this study.

12 REFERENCES

[1] M. Treutler and P. Sörös, “Functional MRI of Native and Non-native Speech Sound Production in Sequential German-English Bilinguals,” *Frontiers in Human Neuroscience*, vol. 15, 2021, doi: 10.3389/fnhum.2021.683277.

[2] L. Y. Ganushchak, I. K. Christoffels, and N. O. Schiller, “The Use of Electroencephalography in Language Production Research: A Review,” *Frontiers in Psychology*, vol. 2, 2011, doi: 10.3389/fpsyg.2011.00208.

[3] S. M. Pereira Soares, M. Kubota, E. Rossi, and J. Rothman, “Determinants of bilingualism predict dynamic changes in resting state EEG oscillations,” *Brain and Language*, vol. 223, p. 105030, 2021, doi: 10.1016/j.bandl.2021.105030.

[4] K. Bice, B. L. Yamasaki, and C. S. Prat, “Bilingual Language Experience Shapes Resting-State Brain Rhythms,” *Neurobiology of Language*, vol. 1, no. 3, pp. 288–318, 2020, doi: 10.1162/nol_a_00014.

[5] S. Reiterer, C. Hemmelmann, P. Rappelsberger, and M. L. Berger, “Characteristic functional networks in high- versus low-proficiency second language speakers detected also during native language processing: An explorative EEG coherence study in 6 frequency bands,” *Cognitive Brain Research*, vol. 25, no. 2, pp. 566–578, 2005, doi: 10.1016/j.cogbrainres.2005.08.010.

[6] S. Reiterer, E. Pereda, and J. Bhattacharya, “Measuring second language proficiency with EEG synchronization: how functional cortical networks and hemispheric involvement differ as a function of proficiency level in second language speakers,” *Second Language Research*, vol. 25, no. 1, pp. 77–106, 2009, doi: 10.1177/0267658308098997.

[7] D. Dadebayev, W. W. Goh, and E. X. Tan, “EEG-based emotion recognition: Review of commercial EEG devices and machine learning techniques,” *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 7, 2021, doi: 10.1016/j.jksuci.2021.03.009.

[8] W. Klimesch, P. Sauseng, and S. Hanslmayr, "EEG alpha oscillations: The inhibition–timing hypothesis," *Brain Research Reviews*, vol. 53, no. 1, pp. 63–88, 2007, doi: 10.1016/j.brainresrev.2006.06.003.

[9] O. Jensen and A. Mazaheri, "Shaping Functional Architecture by Oscillatory Alpha Activity: Gating by Inhibition," *Frontiers in Human Neuroscience*, vol. 4, 2010, doi: 10.3389/fnhum.2010.00186.

[10] R. H. Grabner, A. Fink, A. Stipacek, C. Neuper, and A. C. Neubauer, "Intelligence and working memory systems: evidence of neural efficiency in alpha band ERD," *Cognitive Brain Research*, vol. 20, no. 2, pp. 212–225, 2004, doi: 10.1016/j.cogbrainres.2004.02.010.

[11] A. Topic and M. Russo, "Emotion recognition based on EEG feature maps through deep learning network," *Engineering Science and Technology, an International Journal*, 2021, doi: 10.1016/j.jestch.2021.03.012.

[12] A. Dura and A. Wosiak, "EEG channel selection strategy for deep learning in emotion recognition," *Procedia Computer Science*, vol. 192, pp. 2789–2796, 2021, doi: 10.1016/j.procs.2021.09.049.

[13] V. M. Joshi and R. B. Ghongade, "IDEA: Intellect database for emotion analysis using EEG signal," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 7, pp. 4433–4447, 2022, doi: 10.1016/j.jksuci.2020.10.007.

[14] Z. He, Y. Zhong, and J. Pan, "An adversarial discriminative temporal convolutional network for EEG-based cross-domain emotion recognition," *Computers in Biology and Medicine*, p. 105048, 2021, doi: 10.1016/j.compbiomed.2021.105048.

[15] J. Hu, C. Wang, Q. Jia, Q. Bu, R. Sutcliffe, and J. Feng, "ScalingNet: Extracting features from raw EEG data for emotion recognition," *Neurocomputing*, vol. 463, pp. 177–184, 2021, doi: 10.1016/j.neucom.2021.08.018.

[16] T. Xu, Y. Zhou, Z. Wang, and Y. Peng, "Learning Emotions EEG-based Recognition and Brain Activity: A Survey Study on BCI for Intelligent Tutoring System," *Procedia Computer Science*, vol. 130, pp. 376–382, 2018, doi: 10.1016/j.procs.2018.04.056.

[17] A. S. Ihara *et al.*, "Prediction of Second Language Proficiency Based on Electroencephalographic Signals Measured While Listening to Natural Speech," *Frontiers in Human Neuroscience*, vol. 15, p. 665809, 2021, doi: 10.3389/fnhum.2021.665809.

[18] A. C. Brito, "Effects of Language Immersion versus Classroom Exposure on Advanced French Learners: An ERP Study," *Pursuit - The Journal of Undergraduate Research at The University of Tennessee*, vol. 8, no. 1, 2017, [Online]. Available: https://trace.tennessee.edu/pursuit/vol8/iss1/4/

[19] "ACTFL Proficiency Guidelines 2012 | ACTFL," *www.actfl.org*, 2012. https://www.actfl.org/resources/actfl-proficiency-guidelines-2012

[20] S. Knecht *et al.*, "Language lateralization in healthy right-handers," *Brain*, vol. 123, no. 1, pp. 74–81, 2000, doi: 10.1093/brain/123.1.74.

[21] E. Blagovechtchenski, D. Gnedykh, D. Kurmakaeva, N. Mkrtychian, S. Kostromina, and Y. Shtyrov, "Transcranial Direct Current Stimulation (tDCS) of Wernicke's and Broca's Areas in Studies of Language Learning and Word Acquisition," *Journal of Visualized Experiments*, no. 149, 2019, doi: 10.3791/59159.

[22] S. Marek and N. U. F. Dosenbach, "The frontoparietal network: function, electrophysiology, and importance of individual precision mapping," *Dialogues in clinical neuroscience*, vol. 20, no. 2, pp. 133–140, 2018.

[23] I. Hertrich, S. Dietrich, and H. Ackermann, "The Margins of the Language Network in the Brain," *Frontiers in Communication*, vol. 5, 2020, doi: 10.3389/fcomm.2020.519955.

[24] J. D. Power, B. L. Schlaggar, C. N. Lessov-Schlaggar, and S. E. Petersen, "Evidence for Hubs in Human Functional Brain Networks," *Neuron*, vol. 79, no. 4, pp. 798–813, 2013, doi: 10.1016/j.neuron.2013.07.035.

[25] R. de Diego-Balaguer, A. Martinez-Alvarez, and F. Pons, "Temporal Attention as a Scaffold for Language Development," *Frontiers in Psychology*, vol. 7, 2016, doi: 10.3389/fpsyg.2016.00044.

[26] K. Sander *et al.*, "Frontoparietal Anatomical Connectivity Predicts Second Language Learning Success," *Cerebral Cortex*, vol. 32, no. 12, pp. 2602–2610, 2021, doi: 10.1093/cercor/bhab367.

[27] S. Hämäläinen, V. Sairanen, A. Leminen, and M. Lehtonen, "Bilingualism modulates the white matter structure of language-related pathways," *NeuroImage*, vol. 152, pp. 249–257, May 2017, doi: 10.1016/j.neuroimage.2017.02.081.