EEG-Based Spanish Language Proficiency Classification: A Power Spectrum and Cross-Spectrum Analysis

Blaise O'Mara[1], Skyler Baumer[1], Ronald Croce, PhD[2], Wayne Smith II, PhD[1], Mauricio Pulecio, PhD[3]
University of New Hampshire Departments of:
[1]Electrical and Computer Engineering, [2]Kinesiology, and [3]Literatures, Languages, & Cultures

Introduction

The Big Questions
- What happens in the brain cortex as someone becomes more proficient in a second language (L2)?
- Can these electrophysiological changes be measured with EEG?

The Neural Efficiency Hypothesis
- Skill/task proficiency increases → brain _power_ and _resources_ used decreases.

Brain Power
- Event-Related Desynchronization and Synchronization (ERD/ERS): measures of relative power change.
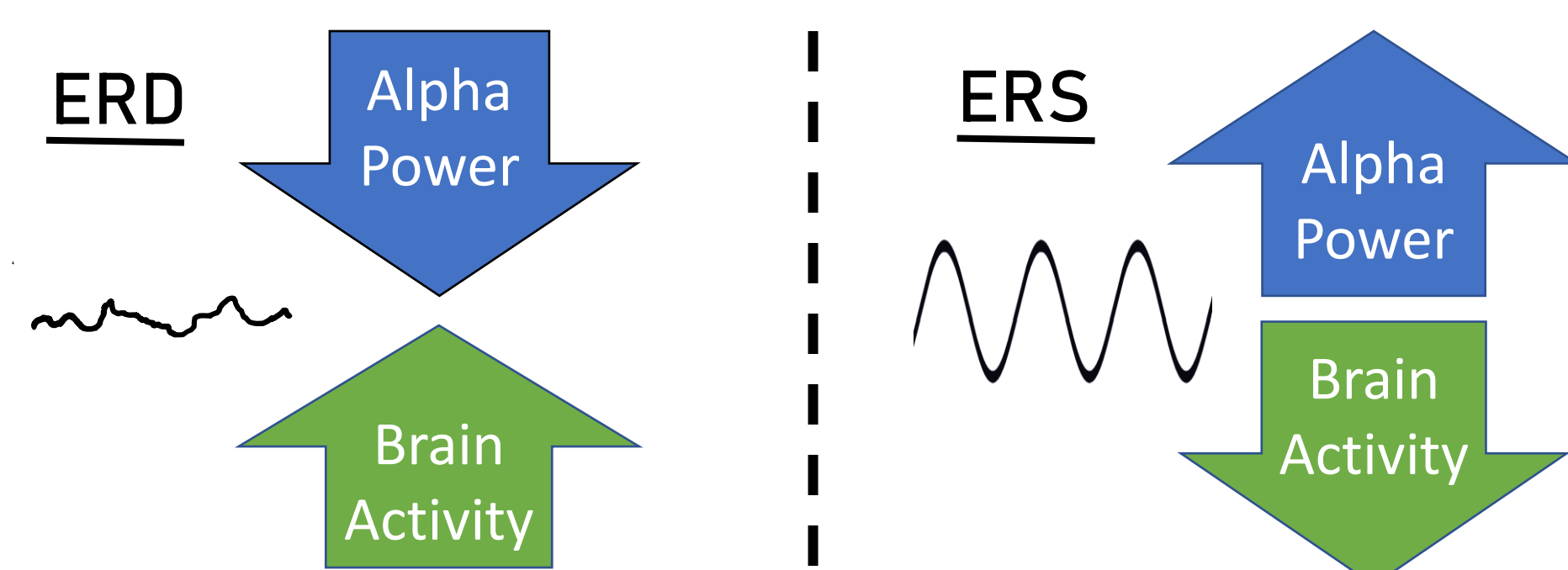

Figure 1: ERD/ERS in reference to alpha "inhibitor" frequency band.

Brain Resources
- Coherence: synchronization between two electrodes
- Indicates regions of the brain working together

Objective & Hypotheses

Objectives
1. Observe relationship between Spanish L2 proficiency and Alpha ERD, Alpha and Beta Coherence
2. Identify EEG features for a machine learning system.

Hypotheses
1. Greater alpha ERD in lower proficiency learners, and
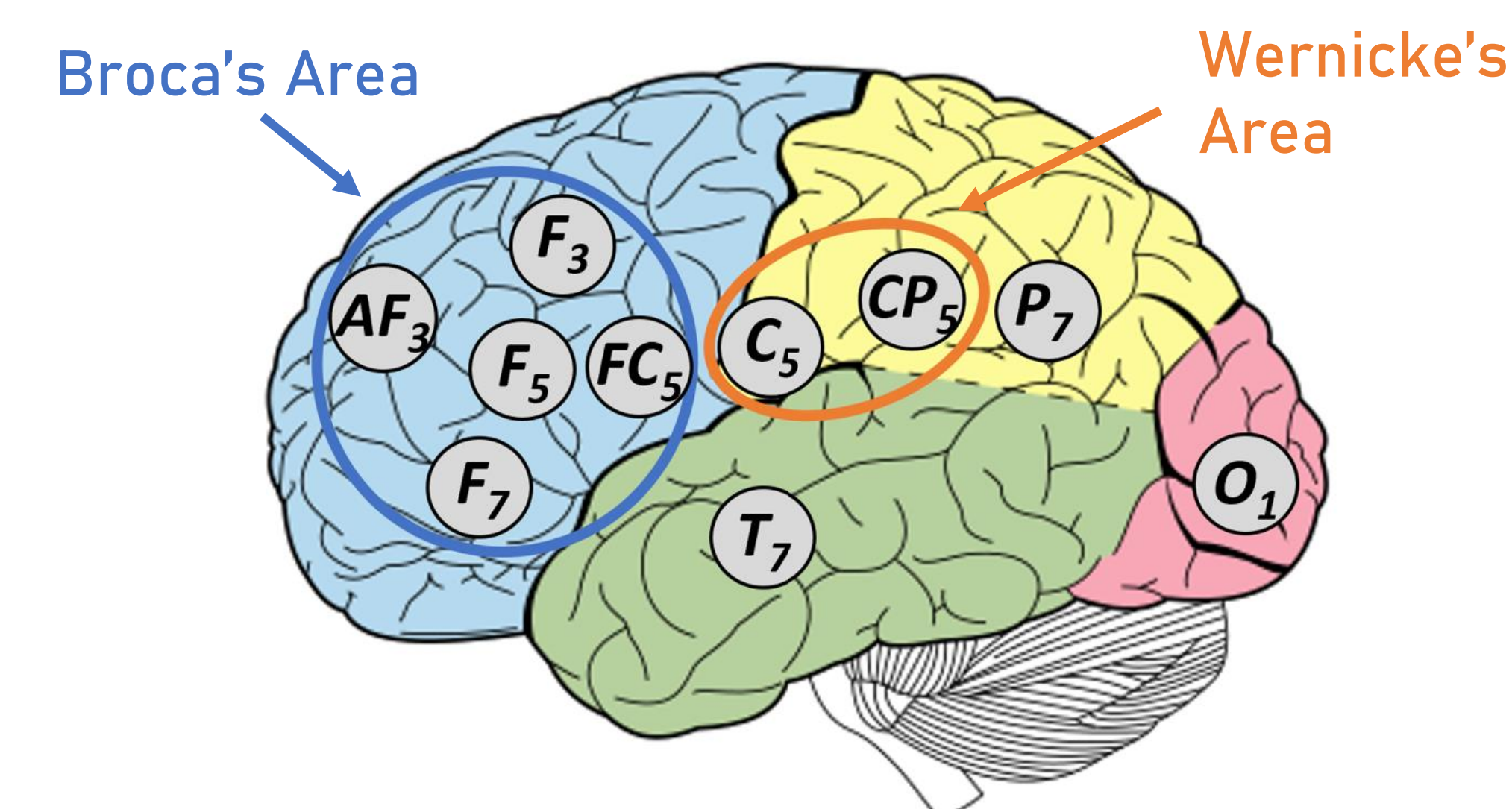2. More widespread coherence in lower proficiency groups

Regions of Interest

Figure 2: 10 of 20 Electrodes sites of interest .

Methodology

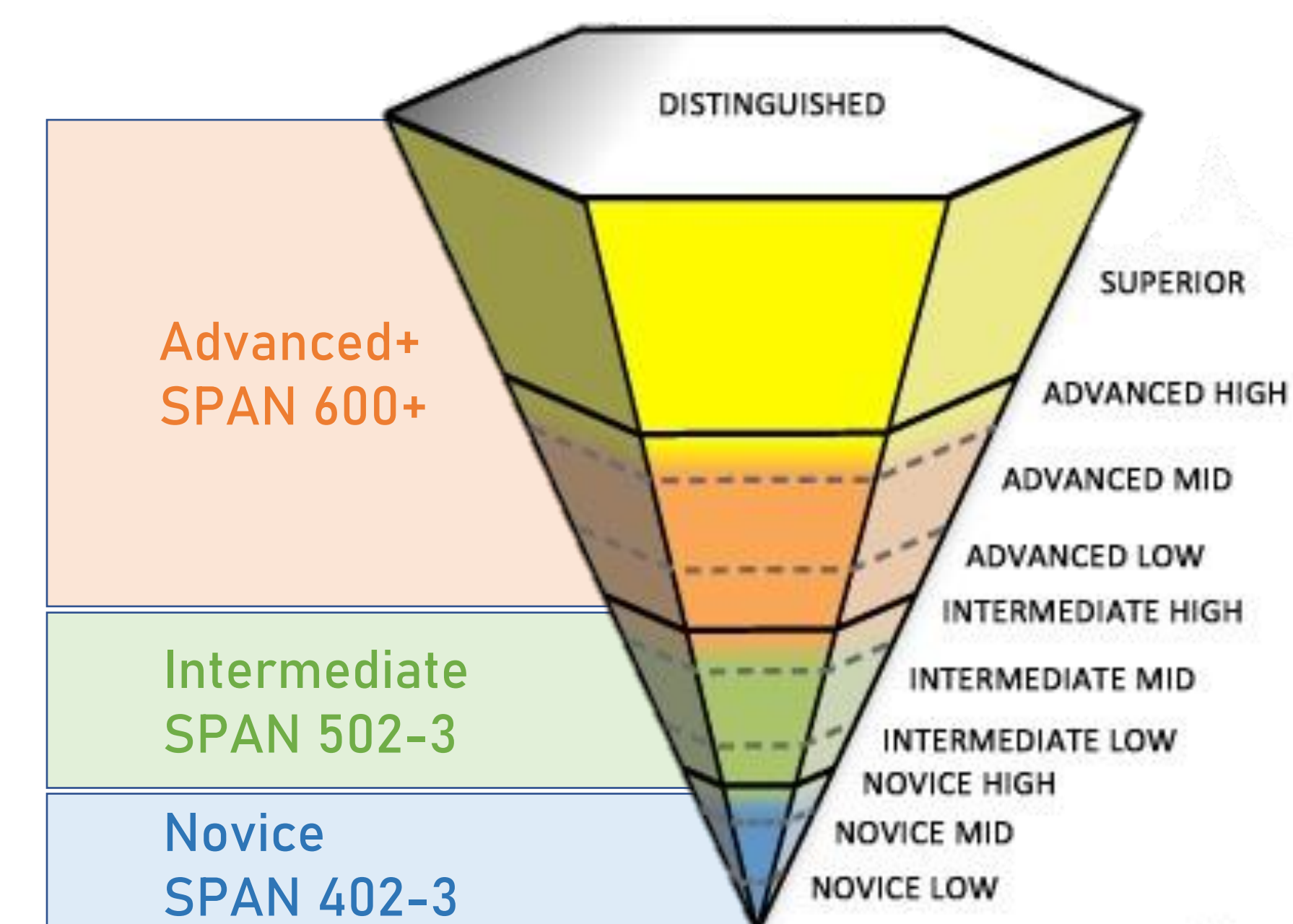### Spanish L2 Proficiency Classification

Figure 3: Language proficiency categories procured by the American Council for the Learning of Foreign Languages (ACTFL). Subjects were chosen from the corresponding courses.

Data Collection
- 3 audio stimuli
- 20 actiCAP electrodes
- 1kHz sampling rate

b – baseline measurement
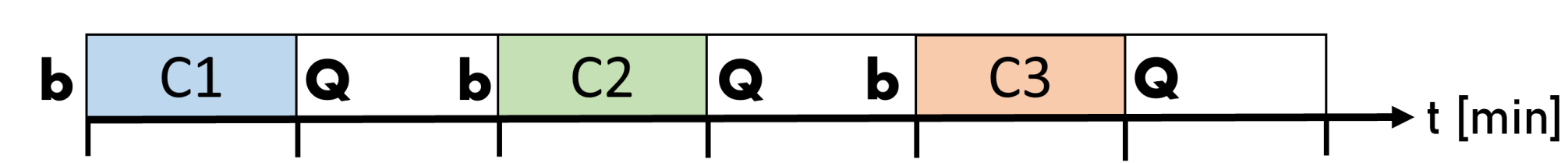**Q** – comprehension & engagement assessment

Figure 4: Timeline of audio condition stimuli during data recording.

Data Processing
- Brain Vision Analyzer Software
- Bandpass filtering
- Artifact Rejection
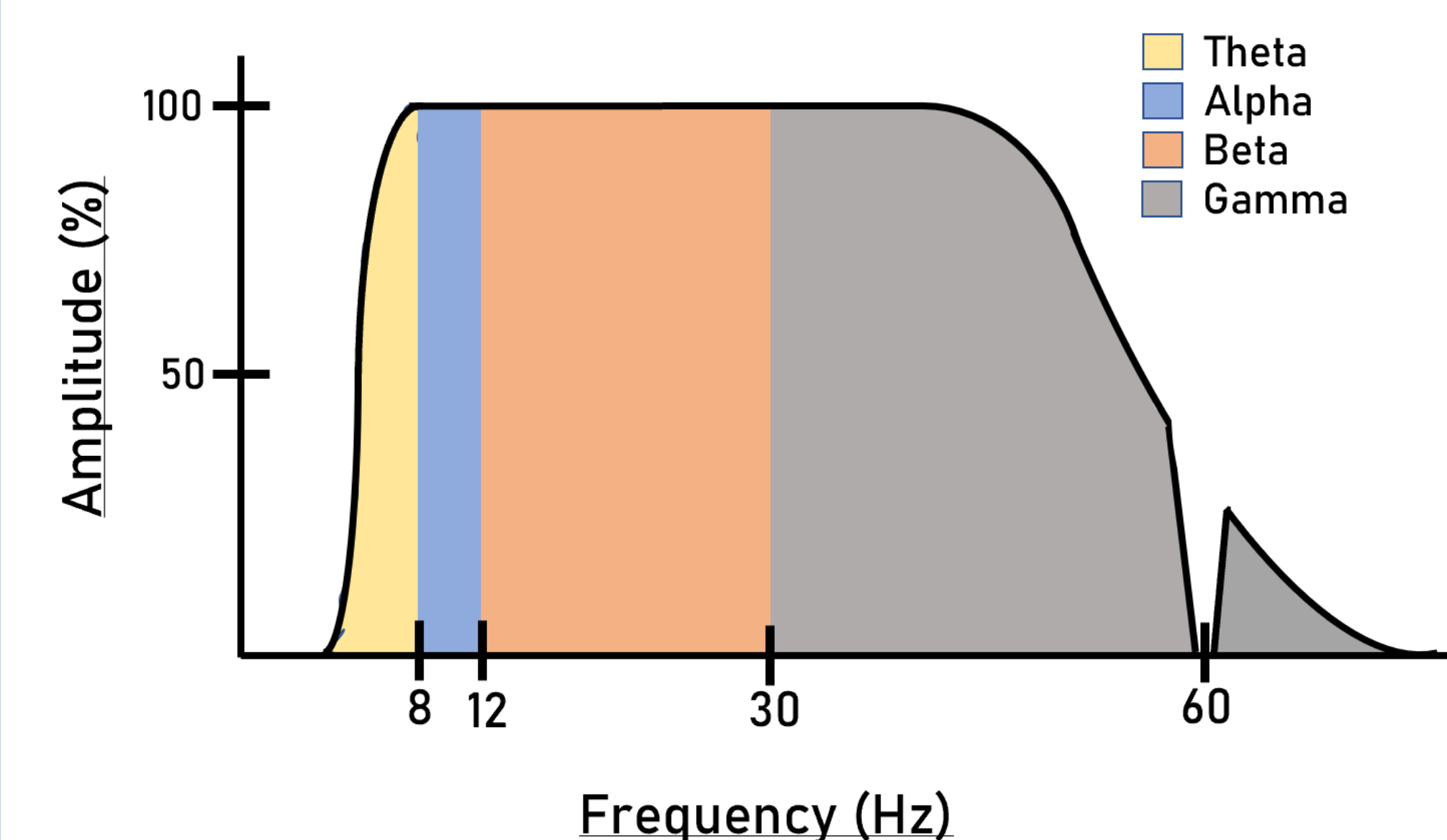- FFT, ERD, Coherence
  - 5s epochs with 50% overlap

Figure 5: Digital 8th-order Butterworth bandpass filter with brain frequency bands of interest.

Data Analysis
- Parsed and organized using Python DataFrames
  - Git, JIRA/Bitbucket
- Visualization
- 3X3 ANOVA
  - Proficiency, Condition, Proficiency x Condition

Coherence Results

LEGEND	
⋯	$0.01 < p \le 0.05$
—	$p \le 0.01$

KEY FINDINGS
- Significant pairs: 31 (Alpha), 41 (Beta)
  - Significant between Proficiency groups
- Novice shows greater unilateral coherence
- Advanced+ shows greater bilateral coherence
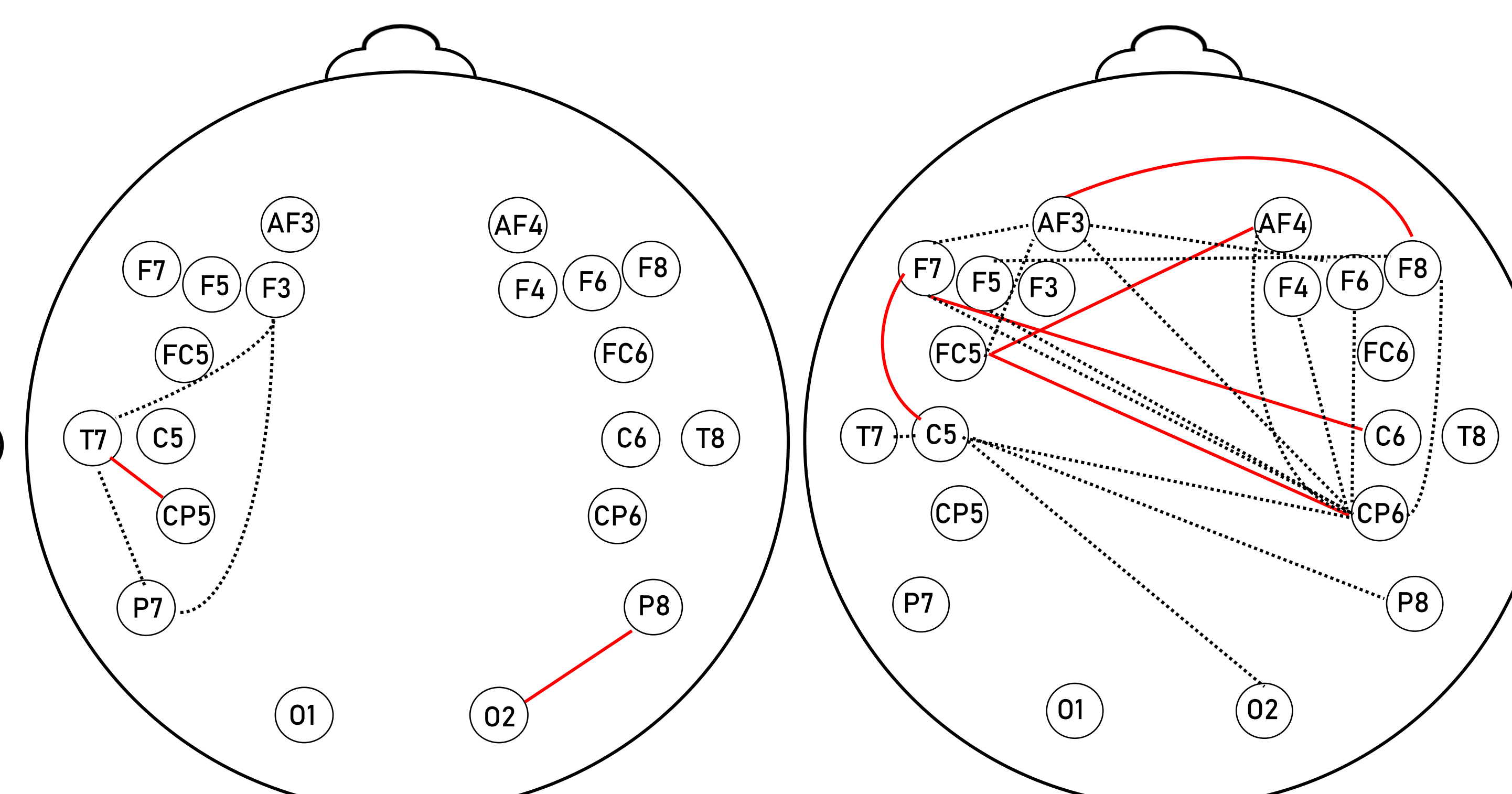
Figure 6: Electrode pairs with significance and Novice Coherence Dominance

Figure 7: Electrode pairs with significance and Advanced+ Coherence Dominance

ERD Results

Alpha ERD
- Two electrodes with significance
  - AF3 ($P < .05$ across Proficiency groups)
  - T8 ($P < .05$ across Conditions)
    - Rejected due to noise in Novice data

Discussion

Limitations & Challenges
- Data collection timeline
- Small sample size
  - Eligibility requirements decreased subject pool
- Uneven groups
- Faulty electrodes

Future Work
- Larger sample sizes should be examined
- Examine correlation of metrics
- Machine Learning model
- Examine trends in Alpha ERD as function of time

Conclusion
- EEG data could be used to classify L2 Spanish proficiency
- Alpha & Beta Coherence provide significant distinctions between proficiency groups
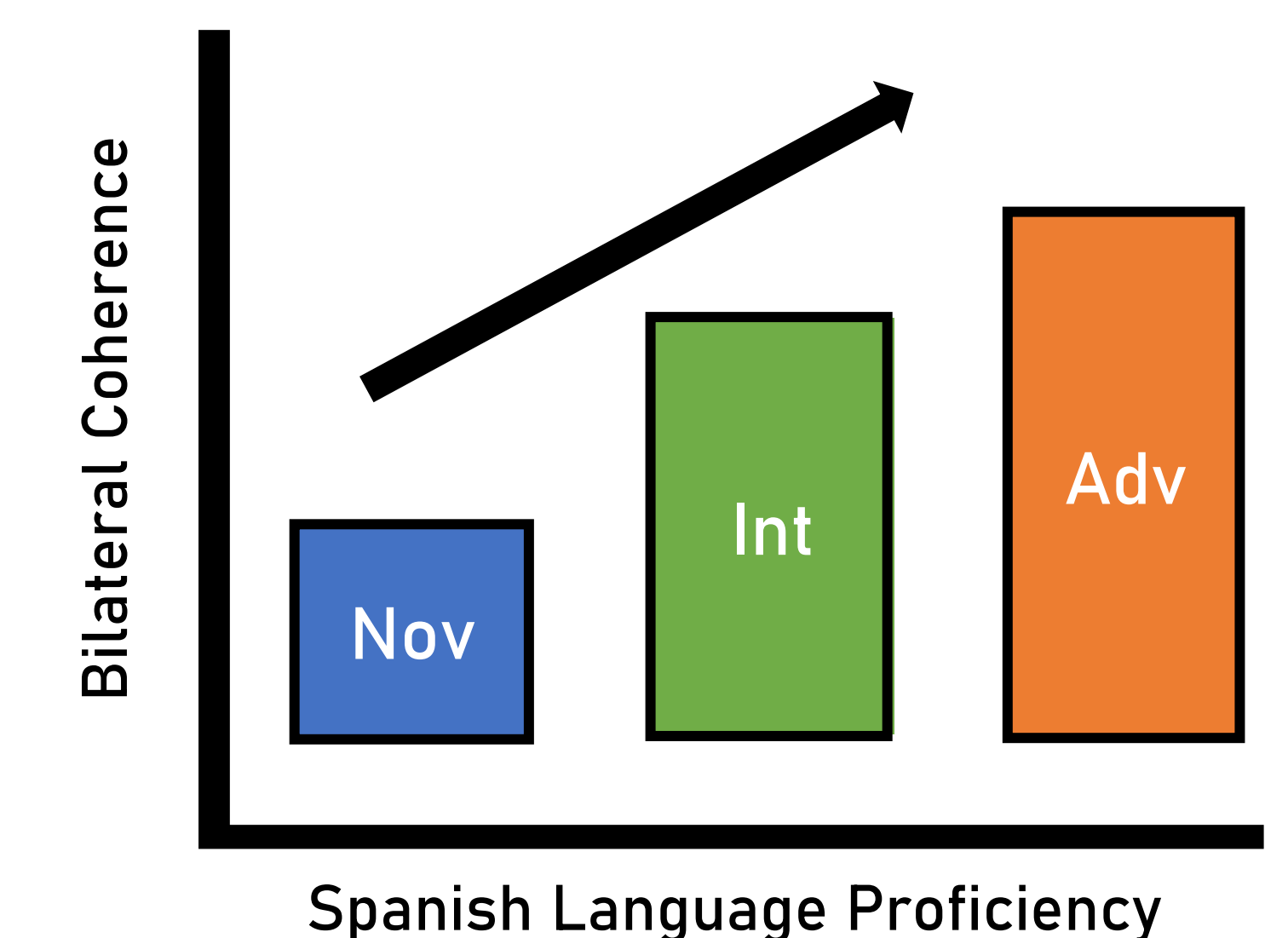

Figure 8: Positive trend between bilateral coherence and proficiency level.

Acknowledgements