Gesticulation while Explaining Complex Mathematics

Orion Junkins

February 16, 2025

1 Preliminary Remarks

This document presents an ETH Practical Work Semester Thesis completed by Orion Junkins in fall of 2024 under the supervision of Dr. Hanna Poikonen and Prof. Christian Holz.

Building on prior work by Dr. Poikonen, this paper explores the relationship between gesticulation, or usage of gestures, and functional connectivity as measured through phase synchrony in EEG data while explaining complex mathematical demonstrations. Additionally, differences between experts and novices are investigated. This document focuses on the technical methodology and identified results, but a brief introduction, discussion, and conclusion are provided so it can stand alone. A complete neuroscience paper, with a more exhaustive literature review and discussion of findings, is being drafted with Dr. Poikonen and will be submitted for publication in Spring 2025.

This document also serves as a technical reference for the code developed for this project. All source code is publicly available through GitHub. Relevant files will be referenced throughout the document so the methodology can be easily understood, reproduced, and extended. To facilitate reproduction, the notebook results_reproduction.ipynb is provided. This notebook runs all statistical tests and directly generates the tables and figures provided in the results section below. To facilitate extension, extension_tutorial.ipynb is provided. This notebook gives a brief introduction to the dataset's format and a tutorial for various utilities and helper functions for accessing specific subsets of the data. Both notebooks can be viewed as read-only documents on GitHub or opened with Google Colab to run them interactively.

In alignment with the original task proposal, various machine-learning approaches were also explored for the task of expert versus novice classification. However, due to the small sample size and lack of statistically significant group differences, these efforts were abandoned in favor of traditional ANOVA-based analysis.

2 Introduction

It has long been understood that movement in the body, beyond just what is required from the mouth and jaws for speech production, is "intimately linked to the activity of speaking" (Kendon, 1980). Motions, most notably in the arms and hands, are fundamental to how humans communicate ideas. However, studies have shown that these gestures can do more than just facilitate communication. Increased usage of gestures has been demonstrated to improve event recall (Stevanoni & Salmon, 2005), lighten cognitive load (Ping & Goldin-Meadow, 2010), improve working memory (Cook et al., 2012), and facilitate problem-solving (Broaders et al., 2007). Many studies have specifically shown enhanced performance on cognitively challenging mathematical tasks (Broaders et al., 2007; Cook et al., 2012). While the empirical evidence of these effects is strong, the underlying neural mechanisms of this effect, particularly in long and complex mathematical tasks, are not well understood.

Despite limited neuroscientific study of the relationship between gesticulation and complex mathematical tasks, many works have investigated the underlying neural mechanisms of

performing these tasks without considering gesticulation. Regardless of expertise, performing mathematical tasks recruits a "bilateral network of prefrontal, parietal, and inferior temporal regions" (Amalric & Dehaene, 2016). Amalric & Dehaene identify engagement of the visuospatial network and the multiple demand system and support their conclusions by studying hemodynamic changes in the brain with functional magnetic resonance imaging (fMRI).

To better understand complex neural processes that recruit multiple parts of the brain, researchers often study functional connectivity (Rogers et al., 2007). Functional connectivity, defined as "correlations between spatially remote neurophysiological events" gives insight into how different areas of the brain work together to achieve complex tasks (Linnman et al., 2012). Some approaches based on fMRI data give insight into this functional connectivity with very high spatial resolution (Logothetis, 2008). However, fMRI suffers from limited temporal resolution as the hemodynamic response recorded lags seconds behind the underlying neural process (Kim, Richter, & Uğurbil, 1997). Many processes in the brain are made possible through rhythmic fluctuations across a wide spectrum of frequencies (Buzsáki, 2006). Rhythms ranging from 0.5 to at least 150 Hz play a major role in governing cognitive processes, with the alpha band (8-13 Hz) considered particularly important (Başar et al., 2001; Ray et al., 2008). For studying these higher frequency interactions, fMRI is not optimal. Furthermore, because fMRI generally requires subjects to lie on their back relatively still in a large machine, fully naturalistic actions and interactions are challenging to study (Tsoi et al., 2022). Results from subjects lying still in an MRI machine may not represent how the brain functions while standing and moving normally.

To study these higher-frequency interactions and move toward a more naturalistic study design, electroencephalography (EEG) is often used. EEG records the electrical activity of the brain using multiple electrodes placed across the scalp (Britton et al., 2016). Studying high-frequency interactions in the brain is possible because EEG offers much higher temporal resolution, albeit at the cost of lower spatial resolution (Chiarion et al., 2023). Furthermore, much more naturalistic study design is possible as subjects can sit, stand, gesture, and move to some degree. While some movement is possible while recording EEG, it is very sensitive to other electrical activity generated by body motions or the surrounding environment. Thus, interpreting EEG often requires careful pre-processing of the data into meaningful metrics. A range of metrics exist for studying functional connectivity in EEG data, each presenting various benefits and drawbacks. At a high level, however, all of these methods seek to identify and quantify the degree to which two or more signals are synchronized. Higher synchronization between neural signals originating from distinct brain regions is hypothesized to correspond to high functional connectivity between those regions.

Prior EEG synchrony metric-based studies on mathematical processing have demonstrated that training in arithmetic strategies modifies theta (3-6 Hz) and lower alpha (8-10 Hz) amplitude synchronies of parieto-occipital electrodes (Grabner & De Smedt, 2012). Grabner and De Smedt have also shown that for arithmetic tasks where simple memory recall is insufficient and procedural reasoning strategies must be employed, parieto-occipital alpha amplitude de-synchrony occurs.

Our work seeks to bridge the gap between the empirical studies on gesticulation and mathematical tasks and the neuroscientific studies on the neural correlates of complex mathematical reasoning. Given that gesticulation enhances performance on cognitively challenging mathematical tasks (Broaders et al., 2007; Cook et al., 2012) we explore if the underlying synchrony patterns associated these tasks are modulated by free versus restricted gesticulation. Because gesticulation is primarily associated with the communication of ideas, we focus on the presentation of mathematical demonstrations. After watching a series of demonstrations (up to one minute each), subjects (both expert and novice mathematicians) are asked to explain a subset of the topics back to experimenters, both with and without free usage of gestures.

3 Materials and Methods

3.1 Participants

Thirty-four math experts (bachelor and master students in math or math-related disciplines, like physics or engineering) and thirty-five math novices (completed high-school but no university-level math studies) participated in the experiment. However, 14 participants from the group of math experts and 17 participants from the novice group were discarded from the data analysis because either their EEG data was noisy due to heavy contamination by the 50 Hz line noise, or some of the relevant data was missing due to malfunctioning EEG amplifier. Thus, in the group of math experts, there were 20 participants, and 18 participants in the novice group. The background of the participants was screened by a math questionnaire. We were not able to calculate the sample size needed as there were no prior EEG studies on change in phase synchrony during long math explanations, and therefore no information on expected variability. We relied on the previous neuroscientific literature related to expertise of mathematical processing and we collected data from a similar number of participants (Grabner & De Smedt, 2012; Amalric & Dehaene, 2016; Jeon, Kuhl, & Friederici, 2019).

The age of the participants ranged from 19 to 25 years (mean 21.2 years) among math experts and from 19 to 35 years (mean 22.9 years) among novices. We aimed for age-matched groups. After excluding subjects, the age of the expert and novice groups does not differ statistically according to 1-way ANOVA: $F(1,36)=3.37$, $P=0.075$. All participants in both groups were right-handed. No participants reported hearing loss or a history of neurological illnesses. The experiment protocol was conducted in accordance with the Declaration of Helsinki and approved by the Executive Board of ETH Zurich after a review by the ETH Zurich Ethics Commission. All participants provided written informed consent.

3.2 Stimuli and Procedure

All subjects first watched a series of 16 mathematical demonstrations up to one minute each. Of the 16 presentations, 8 presented the content in a symbolic manner. The other 8 presented the same content but in a geometric manner. Figure 1 below shows an example of a symbolic and geometric presentation for a single topic. Poikonen presents the detailed procedure and analysis of the resulting EEG data in a separate work (Poikonen et al., 2024). While this portion of the study focused on experts and novices *watching* these math demonstrations, a second set of data was collected immediately after where subjects *explained* the same mathematical topic back to experimenters.

For these demonstrations, subjects were required to present the 8 mathematical concepts they had just watched. In a pseudo-randomized order, a single slide giving a high-level overview of each topic was shown to the participant. An example of this single high-level slide appears in the middle of figure 1. The participant was instructed to explain the topic shown to an experimenter in the room. No time limit was given, and participants continued explaining the topic until they had nothing more to say. For experts, this response time ranged from 22.8 to 164.6 seconds (mean=69.6, SD=28.3). For novices, this response time ranged from 16.3 to 202.0 seconds (mean=70.6, SD=31.8).

These mathematical explanations were given under two conditions. In the first condition, subjects were allowed to use hand gestures freely. In the second condition, subjects' hands were fixed behind their backs, preventing the usage of gestures. The ordering of the two conditions was randomized for each participant.

In addition to explaining the mathematical topics, data for all subjects under both conditions (unrestricted and restricted gesture usage) was collected for a baseline explanation. Specifically, subjects were asked to describe their study background. Again, no time limit was given. For experts, this response time ranged from 10.6 to 261.2 seconds (mean=122.7, SD=42.0). For

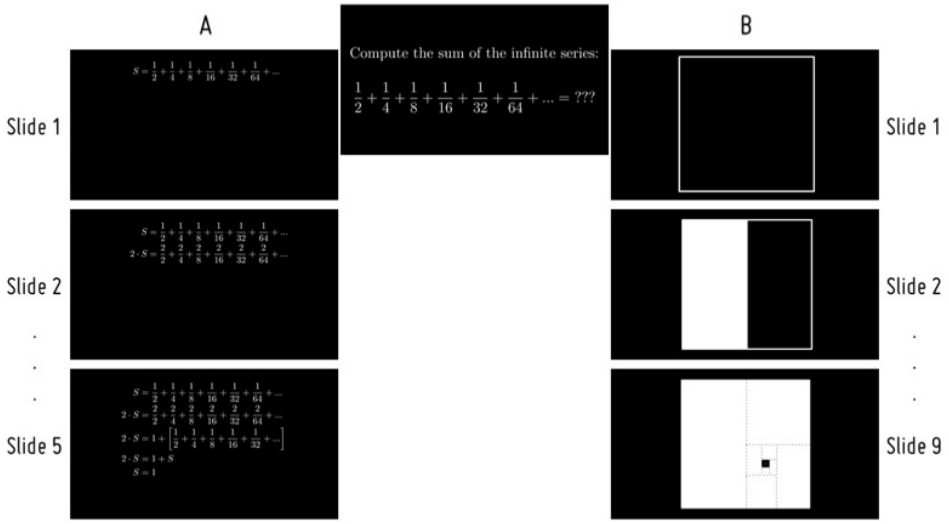

Figure 1: One out of 8 math demonstrations shown in symbolic (A) and geometric (B) form. The symbolic presentation of this specific math demonstration had in total 5 slides and the geometric presentation 9 slides. The slides 4 and 5 for the symbolic presentation are not shown in this image, nor the slides 3-8 for the geometric presentation. The first slide, the introduction of the math demonstration (image in the middle) was the same in both symbolic and geometric presentation. For the explanations, subjects were presented ONLY with the first slide, and asked to explain based on the presentations they had previously seen.

novices, this response time ranged from 49.8 to 307.2 seconds (mean=149.6, SD=66.3).

The data were recorded using Ant Neuro eego mylab electrode caps with active 128 EEG channels and four external electrodes placed below, above and on the left side of the left eye and on the right side of the right eye. The offsets of the active electrodes were kept below 30 mv at the beginning of the measurement, and the data were collected with a sampling rate of 2048 Hz. The beginning of each explanation (either a single slide or the end of the baseline question prompt) was marked with a trigger in the EEG data. The triggers were sent wirelessly via Lab Streaming Layer (LSL; RRID:SCR 017631).

3.3 Data Analyis

3.3.1 Preprocessing

The EEG data of all participants were first preprocessed using the MNE Python toolkit version 1.7 (Gramfort et al., 2013). Bad channels were identified and flagged by analyzing the signal's min, median, max, and standard deviation as well as visually inspecting signals with the MNE-interactive GUI. The number of bad channels identified ranged from 0 to 9 for novices and 0 to 8 for experts. However, of the 12 electrodes ultimately used in this study (described in section 3.3.2 below), at most 1 channel was flagged as bad for novices and 0 for experts. [`See 0_mark_bads.py for implementation`]. All recordings were resampled from 2048 to 512 Hz to facilitate faster processing. Flagged bad channels were interpolated in the downsampled recordings using the default MNE spline interpolation scheme. The average of all electrodes was set as the reference. All signals were filtered to 0.5 to 40 Hz using a bandpass finite impulse response (FIR) filter. The default "firwin" filter provided by MNE was used, which leverages the SciPy Implementation (Virtanen et al., 2020). [`See 1_preproc.py for implementation`]. Next, data were processed using independent component analysis (ICA) to remove artifacts related to eye movements and blinks. Rather than a static component count, the number of components used was set dynamically so that the final components explain 99.99% of the variance. This yielded 90.7

components on average (SD=18.9). Candidate components corresponding to eye blink artifacts were proposed automatically using the MNE "find_bads_eog" function which flags channels closely correlated with EOG channels. Although helpful as a proposal mechanism, all proposed candidates were manually verified and additional bad components were identified through visual inspection. Depending on the subject, 1 to 5 components were removed. [`See 2_select_ica.py for implementation`].

Given data preprocessed according to the steps above, specific subsets of the data were isolated. Specifically, data was split into delta (0.5 to 4.0 Hz), theta (4.0 to 8.0 Hz), low alpha (8.0 to 10.0 Hz), high alpha (10.0 to 12.0 Hz), low beta (13.0 to 20.0 Hz) and high beta (20.0 to 30.0 Hz) with bandpass filtering. As with the initial filtering, this bandpass was performed with MNE's default "firwin" Scipy implementation. [`See 3_compute_connectivity_entropy.py for implementation`].

Within each isolated frequency band, data was further divided temporally according to the events of interest. For all four condition combinations (baseline and demonstration, both with and without gestures), all corresponding events' start and end timestamps were isolated. For each event of interest, the first two seconds and the last one second were discarded. The remaining data was split into epochs of 5 seconds in length with 50% (2.5 seconds) overlap. For each epoch, a single connectivity score is calculated. The exact procedure for this calculation is described in section 3.3.2 below.

3.3.2 Connectivity Metric Computation

This study uses an entropy-based phase synchrony measure based on the approach proposed by Tass et al. (1998). This exact metric has been used in several prior studies (Poikonen et al., 2018a,b, 2024). Because this metric is pairwise, the number of calculations grows exponentially as the number of electrodes increases. Thus, only a relevant subset of 12 electrodes are included in the analysis: F3, Fz, F4, FCz, Cz, CP3, CP4, P1, Pz, P2, PPO1, PPO2. This subset is identical to those used by (Poikonen et al., 2024). These electrodes were chosen on both hemispheres and the central line along the fronto-parietal pathway based on the previous literature on math expertise (Grabner et al., 2007; Amalric & Dehaene, 2016; Sokolowski et al., 2017; Camilleri et al., 2018) and multiple-demand system (Duncan, 2010; Camilleri et al., 2018). Pairwise comparisons between these 12 electrodes yield 66 unique pairings.

The core calculation for this metric, detailed in algorithm 1, takes a single epoch, computes the analytic signal using the Hilbert transform, and isolates the instantaneous phase. The phase difference is computed and binned into 50 uniformly sized bins for every possible pair of electrodes in the selected subset. The final connectivity metric is computed as 1 minus the Shannon entropy of this normalized histogram.

Specifically, given the phase difference vector $\mathbf{d}_{k,m}$ for electrodes k and m and a number of bins B, the normalized entropy $\mathbf{H}_{k,m}$ can be computed as follows:

$$\text{Normalized Shannon entropy:} \quad \mathbf{H}_{k,m} = -\frac{\sum_{i=0}^{B} \mathbf{d}_{k,m}^{(i)} \cdot \log\left(\mathbf{d}_{k,m}^{(i)}\right)}{\log(B)}. \tag{1}$$

This entropy metric gives a notion of the dispersion of the phase differences. A high $\mathbf{H}_{k,m}$ indicates that the distribution of phase differences is well dispersed and lacking a clear peak. This indicates low synchrony between the signals for electrodes k and m, and thus, corresponds conceptually to low neural synchronization between those brain regions. To make the metric easier to interpret, it subtracted from 1. This inverts the measure but preserves the range of [0,1]. This synchrony metric, $\mathbf{S}_{k,m}$, will be 0 when the phase difference distribution for electrodes k and m is uniform and approach 1 as the signals approach perfect synchronization. It indicates

functional connectivity and can be defined as follows:

$$\mathbf{S}_{k,m} = 1 - \mathbf{H}_{k,m} = 1 + \frac{\sum_{i=0}^{B} \mathbf{d}_{k,m}^{(i)} \cdot \log\left(\mathbf{d}_{k,m}^{(i)}\right)}{\log(B)}. \tag{2}$$

Algorithm 1 Phase Synchrony via Normalized Entropy

Require: A single epoch of EEG data, represented as a matrix $\mathbf{X} \in \mathbb{R}^{T \times CH}$:
T: Number of time points
CH: Number of channels
Ensure: Pairwise synchrony matrix $\mathbf{S} \in \mathbb{R}^{CH \times CH}$

Initialize the synchrony matrix:

$$\mathbf{S} = \mathbf{0}_{CH \times CH}$$

Compute the analytic signal using the Hilbert transform: $\mathbf{Z} = \mathcal{H}(\mathbf{X})$, where $\mathbf{Z} \in \mathbb{C}^{T \times CH}$
Extract the instantaneous phase: $\boldsymbol{\phi} = \arg(\mathbf{Z})$, where $\boldsymbol{\phi} \in \mathbb{R}^{T \times CH}$
for each pair of electrodes (k, m) where $k < m$ **do**
    Extract phase vectors $\boldsymbol{\phi}_k \in \mathbb{R}^{T \times 1}$ and $\boldsymbol{\phi}_m \in \mathbb{R}^{T \times 1}$
    Compute the phase difference for pair (k, m):

$$\Delta\boldsymbol{\phi}_{k,m} = \boldsymbol{\phi}_k - \boldsymbol{\phi}_m, \quad \Delta\boldsymbol{\phi}_{k,m} \in \mathbb{R}^{T \times 1}$$

    Bin the phase differences into B bins over the range $[-\pi, \pi]$ to form a histogram:

$$\mathbf{d}_{k,m} = \text{Histogram}(\Delta\boldsymbol{\phi}_{k,m}, B), \quad \mathbf{d}_{k,m} \in \mathbb{R}^{B \times 1}$$

    Normalize the histogram:

$$\mathbf{d}_{k,m} = \frac{\mathbf{d}_{k,m}}{\sum_{i=0}^{B} \mathbf{d}_{k,m}^{(i)}}$$

    Compute the normalized Shannon Entropy according to equation 1:

$$\mathbf{H}_{k,m} = -\frac{\sum_{i=0}^{B} \mathbf{d}_{k,m}^{(i)} \cdot \log(\mathbf{d}_{k,m}^{(i)})}{\log(B)}$$

    Compute the synchrony value according to equation 2:

$$\mathbf{S}_{k,m} = 1 - \mathbf{H}_{k,m}$$

    Symmetrically fill the synchrony matrix:

$$\mathbf{S}_{m,k} = \mathbf{S}_{k,m}$$

end for
return S

In our implementation, $B = 50$, $CH = 12$ and $T = 5$ *seconds* $*$ 512 Hz $= 2560$ time points but future work could investigate different window sizes, bin counts, and/or electrodes. Additionally, in the provided implementation, the for loop is vectorized across multiple epochs. While this sacrifices some readability in the code, it drastically reduces computation time for large datasets by enabling parallel computation. Small constants are also introduced to division and logarithm operators where necessary to ensure numerical stability.

For each of the four condition combinations (baseline and demonstration, both with and without gestures), this connectivity score is calculated and stored for every possible pair of electrodes of interest for every isolated epoch. For each subject, this results in four $E \times K \times K \times F$ arrays where E corresponds to the number of extracted epochs for that particular condition combination, K corresponds to the number of electrodes studied (fixed to 12, corresponding to the electrode set defined above), and F corresponds to the number of frequency bands studied

(fixed to 6, corresponding to the ranges defined above). Note that this metric is symmetric; thus, each $K \times K$ subset is symmetric about the diagonal. The four 4D arrays for each subject are provided in the companion GitHub repository as numpy '.npy' files to facilitate easy reproduction. [`See 3_compute_connectivity_entropy.py for implementation`].

3.3.3 Statistical Analysis

Due to high inter-subject variability, all data is normalized based on each individual's minimum and maximum baseline values. Specifically, for every subject, the minimum and maximum of all baseline data, both with and without gestures, is identified for each electrode pair at each frequency band. All data, baseline, and demonstration, is normalized by subtracting the minimum and dividing by the range. Thus, units for the following results can be interpreted as percentages of a particular subject's baseline range for that particular electrode pair in that particular frequency. Various normalization methodologies (namely Z-score and mean normalization) were experimented with, and all yielded consistent results. The min-max approach was selected for the intuitive percentage-based interpretation it enables. [`See dataset.py for implementation`].

After this normalization, the $E \times K \times K \times F$ matrix is averaged across all epochs, yielding a single $K \times K \times F$ matrix for each of the four condition combinations per subject for each of the six frequency bands. A distinct mixed ANOVA was run for each electrode pair at each frequency. All results of these mixed ANOVA tests are reported below.

4 Results

This section reports and discusses the statistically significant effects of allowing versus restricting gesticulation during cognitively challenging mathematical presentations across all subjects. While they failed to reach statistical significance thresholds, we also present observed differences in these effects between experts and novices to encourage future work with larger sample sizes.

In the baseline data, no p-values below 0.1 were observed for any electrode pair at any frequency for group, condition, or interaction.

In the demonstration data, widespread significant results were observed for condition differences (with versus without gestures) at low alpha frequencies (FDR-adjusted p-values < 0.05). This indicates significant differences between data collected with and without gestures for each subject. To isolate the most significant, all connections with FDR-adjusted p-values > 0.05 are excluded. A secondary filtering is performed using the condition η^2 effect size of the remaining connections. Any electrode pair with $\eta^2 < 0.1$ is excluded.

To identify the directionality of the relationship, the difference in mean for with gestures ("WiG") and without gestures ("NoG") is calculated. Specifically, each subject's mean WiG and NoG connectivity is calculated for all electrode pairs across all extracted epochs. To identify the directionality, a difference is calculated for each subject for each electrode pair by subtracting the NoG mean from the WiG mean. It is a directional value calculated as WiG - NoG so a positive value indicates that WiG $>$ NoG while a negative value indicates that NoG $>$ WiG. After computing this difference for all subjects individually, the overall mean and standard deviation across all subjects are calculated.

Table 1 below presents the FDR adjusted p-value, the η^2 effect size, the mean and standard deviation for the difference in per subject WiG and NoG means, and the corresponding non-FDR-adjusted p-value for the baseline condition. For a more visual exploration of these results, figure 2 below indicates all significant connections. Note that electrode positions are projected using a generic montage and do not represent the exact spatial locations. The provided channel_locs.elc file gives the precise electrode locations used in the study.

Aside from the significant results in low alpha, one other significant FDR-adjusted p-value existed. A single isolated significant p-value of 0.03 occurred for the condition comparison for the electrode pair P1/PPO1 in the high beta frequency band. Because of the isolated nature of this result and the close spatial proximity of the two electrodes, this is regarded as an outlier likely caused by volume conduction and not explored further.

No other FDR-adjusted p-values reached the 0.05 significance threshold. However, seven nearly significant p-values in the range of 0.055 to 0.98 appeared for the group comparison at high alpha and low beta frequencies. Specifically, Fz/PPO2 (p=0.084), and FZ/P2 (p=0.084) in high alpha and Fz/PPO2 (p=0.098), Fz/PPO1 (p=0.098), F3/PPO1(p=0.09), FCz/PPO2 (p=0.098) and FCz/PPO1 (p=0.055) in low beta. These sparse results are highlighted in the discussion as potential grounds for a more targeted, larger sample size follow-up study, but otherwise, they are not explored further.

Table 1: Condition Results for the Low Alpha Frequency band

Electrode Pair	P-Value in Explanation (FDR Adj.)	η^2 Effect Size	Explanation WiG-NoG Mean (SD)	P-Value in Baseline (not FDR Adj.)
Pz, Cz	0.009	0.409	-0.038 (0.046)	0.263
CP4, Fz	0.03	0.299	-0.034 (0.052)	0.744
P2, Cz	0.03	0.295	-0.031 (0.049)	0.271
Pz, CP4	0.03	0.299	-0.041 (0.066)	1.0
P1, CP4	0.03	0.288	-0.035 (0.055)	0.356
CP3, Fz	0.031	0.257	-0.035 (0.059)	0.93
CP3, F4	0.031	0.252	-0.038 (0.066)	0.656
PPO2, Cz	0.031	0.255	-0.028 (0.052)	0.862
P2, CP4	0.031	0.266	-0.04 (0.068)	0.685
Pz, Fz	0.031	0.271	-0.036 (0.063)	0.316
Pz, F4	0.033	0.242	-0.035 (0.062)	0.709
PPO1, P1	0.033	0.239	-0.051 (0.096)	0.628
PPO2, P2	0.033	0.233	-0.057 (0.104)	0.918
CP4, FCz	0.033	0.229	-0.027 (0.05)	0.849
CP4, F3	0.033	0.23	-0.031 (0.058)	0.923
PPO1, F3	0.035	0.223	-0.03 (0.059)	0.895
P1, F4	0.037	0.214	-0.033 (0.063)	0.792
PPO1, Cz	0.037	0.216	-0.031 (0.061)	0.482
P1, CP3	0.039	0.209	-0.044 (0.086)	0.567
CP4, CP3	0.039	0.198	-0.025 (0.052)	0.714
PPO1, CP3	0.039	0.204	-0.044 (0.087)	0.931
PPO2, CP4	0.039	0.199	-0.037 (0.075)	0.479
CP3, F3	0.039	0.203	-0.034 (0.067)	0.992
Pz, CP3	0.045	0.189	-0.037 (0.077)	0.877

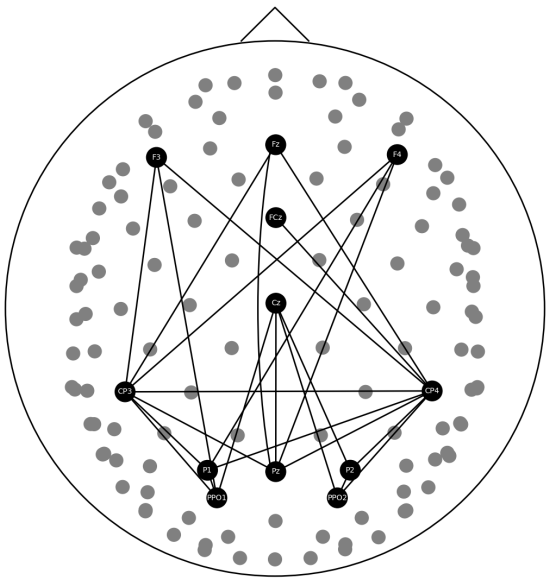

Figure 2: All connections with $p < 0.05$ and $\eta^2 > 0.1$

5 Discussion

5.1 Low Alpha De-synchronization

The results indicate that free usage of gestures correlates with decreased low alpha synchrony across a widespread frontal-parietal network in both experts and novices. Crucially, this relationship only exists when explaining complex mathematics and *not* when giving a baseline explanation about one's study background.

Prior literature has indicated that cognitively challenging mathematical tasks require widespread participation from fronto-parietal networks (Amalric & Dehaene, 2016) and these kinds of tasks correlate with parietal low alpha de-synchronization (Grabner & De Smedt, 2012). These studies indicate that alpha de-synchronization, in particular low alpha desynchronization, is correlated with completing complex mathematical tasks. Thus, if low alpha de-synchronization is inhibited, complex mathematical processing may also be inhibited; if alpha de-synchronization is promoted, performance on complex mathematical tasks may be improved.

The results above indicate a correlation between gesticulation and low alpha de-synchronization. While insufficient to form a causal link, these results may point to the underlying neural mechanism for the performance-enhancing effects of gesticulation on complex mathematical tasks (Broaders et al., 2007; Cook et al., 2012). Future work should seek to understand if gesticulation is causing and/or facilitating the low alpha de-synchronization observed in our results and if this, in turn, is what causes and/or facilitates the performance increases attributed to increased gesticulation.

5.2 Lack of Significant Group Differences

The correlation described in section 5.1 was identified across both experts and novices, and the mixed ANOVA identified no significant group differences. Some insignificant differences were present at high alpha and low beta frequencies. While these results did not reach a threshold of significance, they were relatively clustered in a small number of frontal-parietal connections and only appeared when subjects explained complex mathematics topics. The failure to reach significance may be an artifact of the relatively small sample size, and thus, group differences in this context cannot be fully ruled out. Future work should pursue a similar study design with a larger sample size to explore whether group differences exist in this context.

6 Conclusion

This exploratory work investigates functional connectivity in math experts and novices while explaining mathematical demonstrations with and without free usage of gestures. Significant decreases in low alpha synchronization are observed when gestures are allowed versus restricted across all subjects. These differences are present across a widespread fronto-parietal network. These results suggest a possible neuroscientific explanation for the positive relationship between cognitive performance and gesticulation observed by other studies.

However, this work is limited. While the relationship between gesticulation and alpha synchrony identified is significant, the presentations given did not have a clear performance metric associated with them. Thus, it is unknown if the improved performance prior work has associated with gesticulation was present in this context; our results empirically demonstrate a correlation between gesticulation and alpha de-synchronization, but connections to performance can only be inferred based on related literature. Future studies should investigate the relationship between gesticulation and performance at cognitively challenging tasks to explore if a causal link can be established.

References

Amalric, M., & Dehaene, S. (2016). Origins of the brain networks for advanced mathematics in expert mathematicians. *Proceedings of the National Academy of Sciences*, 113(18), 4909–4917. https://doi.org/10.1073/pnas.1603205113

Amalric, M., & Dehaene, S. (2018). Cortical circuits for mathematical knowledge: Evidence for a major subdivision within the brain's semantic networks. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(20160515). https://doi.org/10.1098/rstb.2016.0515

Amalric, M., & Dehaene, S. (2019). A distinct cortical network for mathematical knowledge in the human brain. *NeuroImage*, 189, 19–31. https://doi.org/10.1016/j.neuroimage.2019.01.001

Başar, E., Başar-Eroglu, C., Karakaş, S., & Schürmann, M. (2001). Gamma, alpha, delta, and theta oscillations govern cognitive processes. *International Journal of Psychophysiology*, 39(2-3), 241–248. https://doi.org/10.1016/s0167-8760(00)00145-8

Britton, J. W., Frey, L. C., Hopp, J. L., et al. (2016). Electroencephalography (EEG): An introductory text and atlas of normal and abnormal findings in adults, children, and infants. In E. K. St. Louis & L. C. Frey (Eds.), *American Epilepsy Society*. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK390346/

Broaders, S. C., Cook, S. W., Mitchell, Z., & Goldin-Meadow, S. (2007). Making children gesture brings out implicit knowledge and leads to learning. *Journal of Experimental Psychology: General*, 136(4), 539–550. https://doi.org/10.1037/0096-3445.136.4.539

Buzsáki, G. (2006). *Rhythms of the brain*. New York, NY: Oxford University Press.

Camilleri, J. A., Müller, V. I., Fox, P., Laird, A. R., Hoffstaedter, F., Kalenscher, T., Eickhoff, S. B. (2018). Definition and characterization of an extended multiple-demand network. *NeuroImage*, 165, 138–147. https://doi.org/10.1016/j.neuroimage.2017.10.020

Chiarion, G., Sparacino, L., Antonacci, Y., Faes, L., & Mesin, L. (2023). Connectivity analysis in EEG data: A tutorial review of the state of the art and emerging trends. *Bioengineering*, 10(3), 372. https://doi.org/10.3390/bioengineering10030372

Cook, S. W., Yip, T. K., & Goldin-Meadow, S. (2012). Gestures, but not meaningless movements, lighten working memory load when explaining math. *Language and Cognitive Processes*, 27(4), 594–610. https://doi.org/10.1080/01690965.2011.567074

Duncan, J. (2010). The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. *Trends in Cognitive Sciences*, 14(4), 172–179. https://doi.org/10.1016/j.tics.2010.01.004

Grabner, R. H., Ansari, D., Reishofer, G., Stern, E., Ebner, F., & Neuper, C. (2007). Individual differences in mathematical competence predict parietal brain activation during mental calculation. *NeuroImage*, 38(2), 346–356. https://doi.org/10.1016/j.neuroimage.2007.07.041

Grabner, R. H., & De Smedt, B. (2012). Oscillatory EEG correlates of arithmetic strategies: A training study. *Frontiers in Psychology*, 3, 428. https://doi.org/10.3389/fpsyg.2012.00428

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L., & Hämäläinen, M. S. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience,* 7(267), 1–13. https://doi.org/10.3389/fnins.2013.00267

Jeon, H.-A., Kuhl, U., & Friederici, A. D. (2019). Mathematical expertise modulates the architecture of dorsal and cortico-thalamic white matter tracts. *Scientific Reports*, 9(1), 6825. https://doi.org/10.1038/s41598-019-43400-6

Kendon, A. (1980). Gesticulation and speech: Two aspects of the process of utterance. In M. R. Key (Ed.), *The Relationship of Verbal and Nonverbal Communication* (pp. 207–228). Berlin, Germany: De Gruyter Mouton. https://doi.org/10.1515/9783110813098.207

Kim, S. G., Richter, W., & Uğurbil, K. (1997). Limitations of temporal resolution in functional MRI. *Magnetic Resonance in Medicine*, 37(4), 631–636. https://doi.org/10.1002/mrm.1910370427

Linnman, C., Moulton, E. A., Barmettler, G., Becerra, L., & Borsook, D. (2012). Neuroimaging of the periaqueductal gray: State of the field. *NeuroImage*, 60(1), 505–522. https://doi.org/10.1016/j.neuroimage.2011.11.095

Logothetis, N. K. (2008). What we can do and what we cannot do with fMRI. *Nature*, 453(7197), 869–878. https://doi.org/10.1038/nature06976

Ping, R., & Goldin-Meadow, S. (2010). Gesturing saves cognitive resources when talking about nonpresent objects. *Cognitive Science*, 34(4), 602–619. https://doi.org/10.1111/j.1551-6709.2010.01102.x

Poikonen, H., Toiviainen, P., & Tervaniemi, M. (2018a). Naturalistic music and dance: Cortical phase synchrony in musicians and dancers. *PLoS ONE, 13*(4), e0196065. https://doi.org/10.1371/journal.pone.0196065

Poikonen, H., Toiviainen, P., & Tervaniemi, M. (2018b). Dance on cortex: Enhanced theta synchrony in experts when watching a dance piece. *European Journal of Neuroscience, 47*(5), 433–445. https://doi.org/10.1111/ejn.13838

Poikonen, H., Tobler, S., Trninić, D., Formaz, C., Gashaj, V., & Kapur, M. (2024). Math on cortex-enhanced delta phase synchrony in math experts during long and complex math demonstrations. *Cerebral Cortex*, 34(2), bhae025. https://doi.org/10.1093/cercor/bhae025

Ray, S., Niebur, E., Hsiao, S. S., Sinai, A., & Crone, N. E. (2008). High-frequency gamma activity (80–150 Hz) is increased in human cortex during selective attention. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, *119*(1), 116–133. https://doi.org/10.1016/j.clinph.2007.09.136

Rogers, B. P., Morgan, V. L., Newton, A. T., & Gore, J. C. (2007). Assessing functional connectivity in the human brain by fMRI. *Magnetic Resonance Imaging*, 25(10), 1347–1357. https://doi.org/10.1016/j.mri.2007.03.007

Sokolowski, H. M., Fias, W., Mousa, A., & Ansari, D. (2017). Common and distinct brain regions in both parietal and frontal cortex support symbolic and nonsymbolic number processing in humans: A functional neuroimaging meta-analysis. *NeuroImage*, 146, 376–394. https://doi.org/10.1016/j.neuroimage.2016.10.028

Stevanoni, E., & Salmon, K. (2005). Giving memory a hand: Instructing children to gesture enhances their event recall. *Journal of Nonverbal Behavior*, 29(4), 217–233. https://doi.org/10.1007/s10919-005-7721-y

Tass, P., Rosenblum, M. G., Weule, J., Kurths, J., Pikovsky, A., Volkmann, J., Schnitzler, A., & Freund, H.-J. (1998). Detection of n:m phase locking from noisy data: Application to magnetoencephalography. *Physical Review Letters*, 81(15), 3291–3294. https://doi.org/10.1103/PhysRevLett.81.3291

Tsoi, L., Burns, S. M., Falk, E. B., & Tamir, D. I. (2022). The promises and pitfalls of functional magnetic resonance imaging hyperscanning for social interaction research. *Social and Personality Psychology Compass*, 16(10), e12707. https://doi.org/10.1111/spc3.12707

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., & SciPy 1.0 Contributors. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*, 261–272. https://doi.org/10.1038/s41592-019-0686-2