Check for updates

Empirical Research

EEG Decoding of Finger Numeral Configurations With Machine Learning

Roya Salehzadeh [1] , Brian Rivera [2] , Kaiwen Man [3] , Nader Jalili [1] , Firat Soylu [3]

[1] *Department of Mechanical Engineering, The University of Alabama, Tuscaloosa, AL, USA.* [2] *Department of Psychology, University of Nebraska–Lincoln, Lincoln, NE, USA.* [3] *Department of Educational Studies, The University of Alabama, Tuscaloosa, AL, USA.*

Abstract

In this study, we used multivariate decoding methods to study processing differences between canonical (montring and count) and noncanonical finger numeral configurations (FNCs). While previous research investigated these processing differences using behavioral and event-related potentials (ERP) methods, conventional univariate ERP analyses focus on specific time intervals and electrode sites and fail to capture broader scalp distribution and EEG frequency patterns. To address this issue a supervised learning classifier—support vector machines (SVM)—was used to decode ERP scalp distributions and alpha-band power for montring, counting, and noncanonical FNCs (for integers 1 to 4). The SVM was used to test whether the numerical information presented in FNCs can be decoded from the EEG data. Differences in the magnitude and timing of accuracy rates were used to compare the three types of FNCs. Overall, the algorithm was able to predict numerical information presented in FNCs beyond the random chance level accuracy, with higher rates for ERP scalp distributions than alpha-power. Montring had lower peak accuracy compared to counting and noncanonical configurations, likely due to automaticity in processing montring configurations leading to less distinct scalp distributions for the four numerical magnitudes (1 to 4). Paralleling the response time data, the peak decoding accuracy time for montring was earlier for montring (472 ms), compared to counting (577 ms) and noncanonical FNCs (604 ms). The results provide support for montring configurations being processed automatically, somewhat similar to number symbols, and provide additional insights for processing differences across different forms of FNCs. This study also highlights the strengths of decoding methods in EEG/ERP research on numerical cognition.

Keywords

finger numeral configurations, ERP, EEG, decoding, machine learning, numerical cognition

Numerical cognition, the ability to reason with numbers, is crucial for individual life and career advancement (Ritchie & Bates, 2013). Early numerical abilities form the foundation for the development of more complex mathematical skills (Siegler & Braithwaite, 2017). During development, finger-based interactions provide preliminary access to foundational mathematical constructs such as one-to-one correspondence. Fingers are used across cultures to communicate numerical information, count, and do arithmetic (Bender & Beller, 2012). Over the last few decades extensive research has been done on the intricate relationship between fingers and numbers (see Barrocas et al., 2020; Berteletti & Booth, 2016; Soylu, Lester, et al., 2018 for reviews). For example, these studies have shown the involvement of hand motor circuits in counting (Andres et al., 2007) and arithmetic tasks (Berteletti & Booth, 2015), a functional association between finger motor and number processing (Michaux et al., 2013; Soylu & Newman, 2016), a relation between finger gnosis (the ability to individuate fingers) and number skills (Rusconi et al., 2005; Soylu, Raymond, et al., 2018; Wasner et al., 2016),

and numerical magnitude skills originating from our ability to interact with physical magnitudes (Badets et al., 2007; Sato et al., 2007). These findings were interpreted from an embodied cognition perspective to argue that finger-based interactions, such as finger-counting, leave a lasting trace in the neural systems associated with numerical cognition (de Freitas & Sinclair, 2013; Nemirovsky & Ferrara, 2009; Núñez, 2012), which has implications for understanding both children's and adults' numerical skills (Badets et al., 2010; Newman & Soylu, 2014)

Finger-numeral configurations (FNCs) refer to specific ways fingers are raised to enumerate number sequences (finger-counting) and to communicate numbers to other people (finger-montring). While finger-counting has been extensively studied in the literature, in the context of both counting and arithmetic tasks, finger-montring is a relatively new concept (first defined by Di Luca & Pesenti, 2008). Finger-counting has a self-directive facilitating function, serving both counting and doing arithmetic (e.g. when a child opens their fingers in a sequence to answer 4 + 3). Finger-montring is used to communicate numerical information; for example, a child may spontaneously raise their index, middle, and ring fingers to indicate *three*, in response to a question inquiring how many strawberries they want.

Previous research has found both behavioral and neural differences in visually identifying canonical (culturally dominant) and noncanonical FNCs with adult participants (Di Luca et al., 2010; Di Luca & Pesenti, 2008; Soylu et al., 2019; van den Berg et al., 2021). Cultural norms regarding which fingers are used to represent numerosities determine which finger configurations are canonical. Finger-counting and finger-montring configurations are referred to as *canonical* since they follow culture-specific norms. In contrast, unfamiliar FNCs are said to be *noncanonical* (e.g., raising the thumb and pinky fingers to enumerate *two*). Investigating the processing differences between canonical and noncanonical finger configurations can help better understand how cultural and developmental experiences with FNCs impact the number processing network; effects that can still be observed in adulthood.

In the first study that studied processing differences across montring, counting, and noncanonical configurations, participants were asked to identify numerical information in FNCs (Di Luca & Pesenti, 2008). No behavioral performance differences were found in identifying montring and counting configurations, while the behavioral performance was lower when identifying noncanonical configurations, compared to both canonical configurations. In the same study, based on results from a masked priming experiment, Di Luca and Pesenti reported evidence for montring configurations automatically activating numerical magnitudes, while this was not the case for noncanonical configurations. In a follow-up study, Di Luca, Lefèvre, and Pesenti (2010) compared the priming effects of montring and noncanonical finger numeral configurations on naming Arabic digits or verbal numerals. They found a priming effect decreasing with the numerical distance between the prime and the target for the montring prime, but not for the noncanonical faster responses. There were faster responses when the numerical distance between the prime and the target was 0 (e.g., FNC showing 5 being the prime and Arabic number 5 or "five" being the target) than when the distance was 4 (e.g., a prime being an FNC showing 1 and Arabic number 5 or "five" being the target). These results were interpreted as montring configurations automatically triggering a numerical magnitude representation, similar to Arabic numerals, while noncanonical finger configurations need further elaboration to access numerical magnitude information. This raises the question of how the processing of montring and noncanonical configurations differ; whether in the early stages of perceptual processing (e.g., faster perceptual processing for montring) or in later semantic processing, when the numerical magnitudes associated with FNCs are represented. To answer this question, Di Luca and Pesenti (2010) conducted a visual-detection experiment, where participants identified if a montring configuration target was present among a varying number of distractor noncanonical configurations. The results showed linearly increasing response latencies, as a function of the number of distractors, providing evidence for the lack of a *pop-out effect* and lack of perceptual salience of montring configurations.

Soylu et al. (2019) used event-related potentials (ERP) to investigate processing differences among montring, counting, and noncanonical configurations and found evidence for distinct differences across the three configurations at different stages of processing. They reported higher positivity for montring in the P1/N1 range, compared to counting and noncanonical, which shows an early perceptual processing difference for montring, possibly due to modulation of attentional resources. In later stages of processing, in the P3 range, montring and counting showed no differences, but both were distinguished from noncanonical, which was explained based on strategy differences (automated access vs. counting). Soylu et al. also reported faster and more accurate recognition of montring configurations, compared to counting and noncanonical. van den Berg et al. (2021) also used ERPs to study processing differences between

montring and noncanonical configurations, but this time with an arithmetic verification task and with configurations involving both hands. The results showed both behavioral and ERP differences, especially for configurations matching with numbers 2 to 4. In addition to faster and more accurate sum verification with montring configurations, enhanced right parietal P2p and central-parietal P3 responses were reported.

Overall, the limited number of studies on processing differences between canonical and noncanonical FNCs hint at early developmental experiences with FNCs having a lasting effect that can be observed in adulthood, and canonical, particularly montring, configurations being processed in a way that is somehow similar to number symbols. However, how canonical configurations are processed differently than noncanonical ones (e.g., which stages of processing) is still not clear. While ERPs provide the high temporal resolution needed to answer this question, use of conventional ERP analysis requires a priori determination of ERP components to be studied, which are characterized by specific types of processing, associated with a time interval and electrode site. This approach works when the focus of investigation can be narrowed down to specific ERP components (e.g., use of the N400 component in language studies on semantic incongruence), however, it is not ideal when the tested effects are distributed across multiple electrode sites and time intervals.

Machine learning methods provide alternative ways to analyze EEG, and in general neuroimaging, data that allow studying processing differences in wider temporal and spatial scales, and with higher sensitivity (Grootswagers et al., 2017). Multivariate pattern analysis (decoding) methods have been extensively applied to the decoding of neuroimaging data for brain-computer interfaces (BCI) and have recently become popular for studying brain function (Hebart & Baker, 2018). Decoding refers to the prediction of experimental conditions based on brain data. The use of decoding for real-world applications (e.g., BCI) and for studying brain function imposes different goals and expectations. While in real-world applications the goal is to maximize the prediction accuracy, the study of brain function targets identifying differences in neural processing across cognitive tasks. When decoding neuroimaging data, instead of testing for significant effects using univariate tests, a machine learning decoder is trained to capture salient features of the data that distinguish one form of processing (or experimental condition) over others. The trained decoder is then used to predict the type of processing taking place based on brain data. The predictive success of the algorithm indicates to what extent the data used includes *signal* that is unique to the process studied. When compared to traditional univariate analysis, multivariate analysis (decoding) was found to be complementary and more sensitive to distributed processing of information in the brain (Jimura & Poldrack, 2012). While decoding has become a mainstream method for the analysis of fMRI data (Hebart & Baker, 2018), it has also been applied to the analysis of EEG data. For example, Bae and Luck (Bae & Luck, 2018) decoded which of the 16 orientations of a shape was held in the working memory by considering both sustained potentials (ERP scalp distributions) and alpha-band oscillations, and similar methods were used to decode spatially distributed direction-of-motion (Bae & Luck, 2019), visual working memory in schizophrenia compared to controls (Bae et al., 2020), and to categorize individuals with attention-deficit / hyperactivity disorder and controls (Ghasemi et al., 2022). Decoding methods can help detect differences across cognitive processes that are not available through conventional univariate ERP analysis and identify features that predict behavioral performance.

In this study, we used multivariate decoding methods to study FNCs for the first time. Our overarching goal was to investigate whether sustained potentials (ERP scalp distributions) and alpha-frequency EEG signals can be decoded to predict the numerical information represented in FNCs and to compare prediction accuracy rates across different types of FNCs. The presented analysis is a follow-up to a previous study (Soylu et al., 2019), where a traditional univariate ERP analysis was conducted to characterize processing differences across montring, counting, and noncanonical FNCs. Based on results from the previous ERP study, we expected time intervals in the P1/N1 and P3 range to show higher ERP decoding accuracies, since these intervals showed larger ERP amplitude differences across the three types of FNCs. As for alpha-frequency decoding, previous studies using decoding methods had highlighted the role of alpha-band oscillations in sustained attention (Awh & Jonides, 2001) and in tracking the location of items most relevant to working memory (van Ede et al., 2017) as well as their dimensions (Rose et al., 2016). Based on these findings, we expected alpha-decoding to especially highlight perceptual processing differences in early time intervals, since in the previous study P1/N1 differences were interpreted as pointing to differences in visual attention in early perceptual processing. We report results from the implementation of a support vector machine (SVM) algorithm to decode numerical informa-

tion in FNCs, based on both ERP scalp distributions and instantaneous alpha-band power. We also present a general discussion on the implications of using multivariate decoding methods in numerical cognition studies.

Materials and Method

Data

The EEG dataset analyzed in this study was a publicly available dataset previously used in an ERP study, investigating behavioral and ERP differences in processing montring, counting, and noncanonical FNCs, using traditional univariate methods (Soylu et al., 2019). The raw EEG data and the analysis scripts for that study are publicly available in Harvard Dataverse (Soylu, 2019).

Participants

Data from thirty-eight adult participants were included in the analysis (20 female, *M* = 19.68 years, *SD* = 1.84). All participants were native English-speaking undergraduate students, with no history of neurological illness and normal or corrected-to-normal vision. Written informed consent was obtained from all participants. The research was approved by the Institutional Review Board of The University of Alabama.

Stimuli and Experimental Procedures

The stimulus set for the experiment constituted the combination of three types of FNCs—montring (M), counting (C), and noncanonical (NC)—and four numbers: one to four shown by different FNCs, separately for the left and right hands. This added up to 24 pictures of unique FNCs. Because the previous study reported no differences between the FNCs for the left and the right hands, data from both hands were pooled together, generating 12 unique categories (Soylu et al., 2019) (3 types of FNCs x 4 numbers; Figure 1).

Figure 1

12 FNCs (3 Types x 4 Numbers) and the Associated Topography (Scalp Distribution) of Instantaneous Alpha-Power (First Row) and ERPs (Second Row) for Each FNC, at 500 ms, Averaged Across Participants
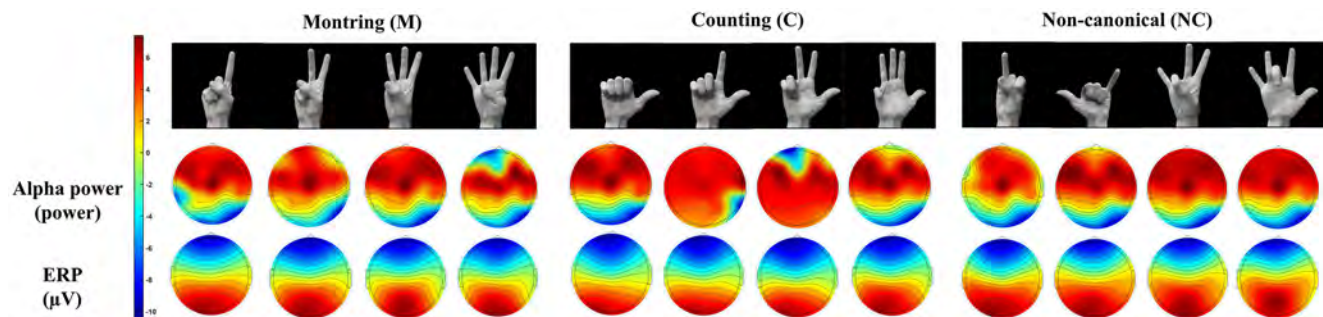
The experiment was divided into 10 blocks, with 96 trials in each block, including four sets of the 12 configurations, randomly sequenced across each block. There were 960 trials in total, with 80 trials for each FNC. In each trial, an FNC was presented to the participant for 500 ms followed by an Arabic numeral for 1000 ms (Figure 2). The task involved validating whether the Arabic numeral shown represents the same numerical magnitude as the FNC that was shown previously, by pressing one of the two buttons on a Logitech F310 game controller. The inter-trial interval was 1200 ms, with a 300 ms additional jitter (total ITI varying 1200-1500 ms).

Figure 2

The Structure of Each Trial in the Experimental Paradigm
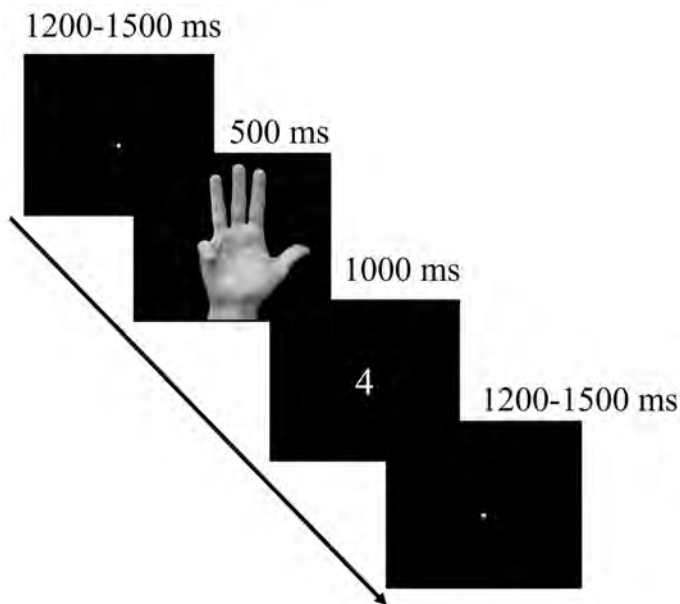
Data Acquisition

A BrainVision 32 Channel ActiChamp system, with Easy Cap recording caps using Ag/AgCl electrodes (Fp1, Fz, F7, FT9, FC5, FC1, C3, T7, TP9, CP5, CP1, Pz, P3, P7, O1, Oz, O2, P4, P8, TP10, CP6, CP2, C4, T8, FT10, FC6, FC2, F4, F8, Fp2, Cz) was used for data collection. The international 10-20 system was used for electrode locations. The electrode Cz was the recording reference electrode, and the recording sampling rate was 500 Hz.

Data Pre-Processing

A custom MATLAB script using EEGLAB (Delorme & Makeig, 2004) and ERPLAB (Lopez-Calderon & Luck, 2014) functions were used for pre-processing. The pre-processing script is publicly available at GitHub (see Supplementary Materials). The EEG data were re-referenced to the average reference (Cz was added back to the data electrodes). A 0.1 Hz (half-amplitude cutoff) high-pass and an 80 Hz (half-amplitude cutoff) low-pass IIRButterworth filter (24 dB/octave) were applied, and the data were resampled at 250 Hz.

The EEG signals were epoched between 500 ms before the stimulus onset (FNC presentation) to 1500 ms after (stimulus offset). All epochs were corrected to the 500 ms pre-stimulus baseline. A moving window peak-to-peak threshold algorithm (threshold 60 μV, window size 80 ms, window step 20 ms) was used for detecting eye blinks, and a step-like artifacts algorithm (threshold 50 μV, window size 200 ms, window step 100 ms) was used for detecting eye movements. The results of the automatic artifact detection were inspected visually. Epochs that were marked during the artifact detection step were excluded (20.92% of trials, *SD* = 21.45). In addition, only the epochs that preceded a correct response (Arabic numeral validated correctly) were included in the analysis (79.08% of trials).

After artifact detection, separate pre-processing steps were followed for the ERP-based and the alpha-based decoding. To ensure the decoding of non-overlapping signals for the two analyses, ERP decoding was limited to frequencies less than 6 Hz and alpha-band decoding was limited to frequencies between 8 and 12 Hz. For the ERP-based decoding, a 6 Hz (half-amplitude cutoff) low pass IIRButterworth filter (24 dB/octave) was applied to the epoched data. For the alpha-based decoding, an 8-12 Hz band-pass filter was used on the epoched data. Hilbert Transform was applied to improve the measurement of instantaneous alpha amplitude. The amplitude of the complex analytic signal was calculated and squared at each time point to calculate alpha power.

Decoding

The decoding approach used in this paper was adapted from Bae and Luck (2018), where they decoded which of the 16 orientations of a shape was held in the working memory by considering both sustained potentials (ERP scalp distributions) and alpha-band oscillations (see Figure 3 for a summary of the decoding process). We implemented a modified version of the publicly shared MATLAB analysis scripts associated with this study (Bae, 2017).

Figure 3

Procedure of ERP Decoding Analysis by Steven J. Luck, Gi-Yeul Bae, and Aaron M. Simmons
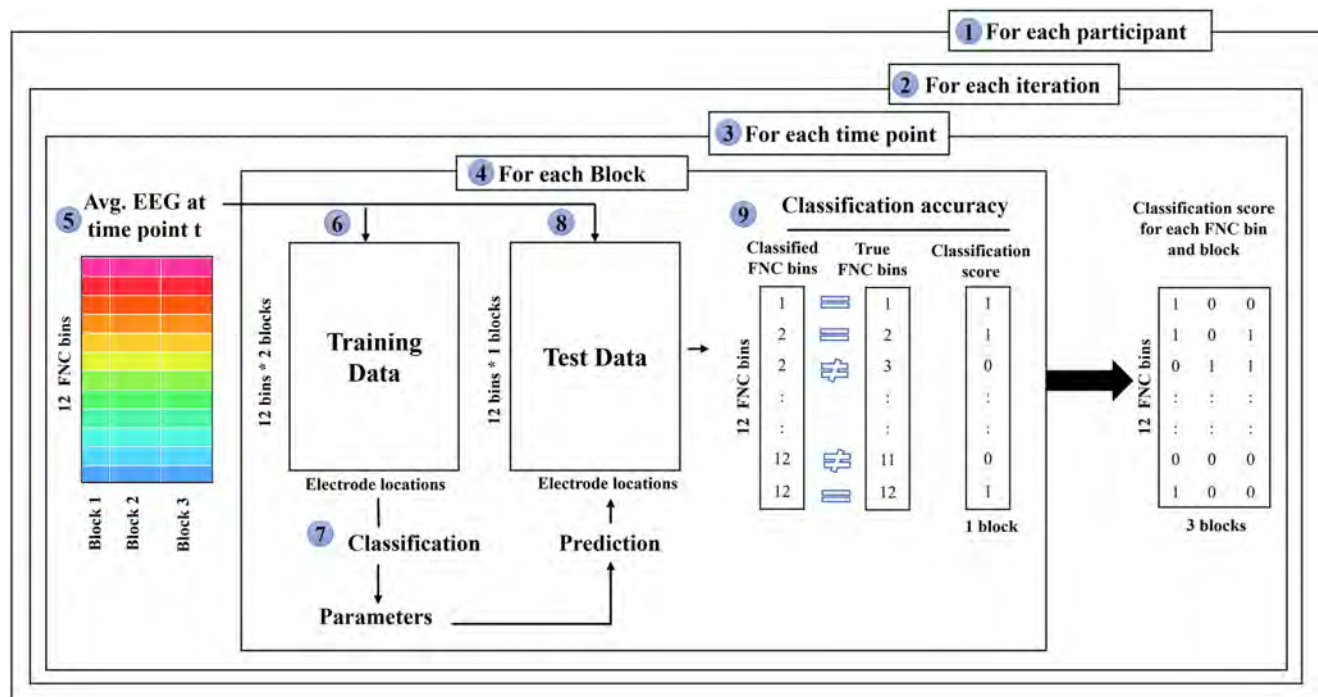
Note. The decoding started with choosing a participant in Step 1 and then shuffling the raw EEG data for that participant in Step 2 (iteration = 10). The time point to run decoding was selected at Step 3 from the interesting time range and at Step 4 threefold cross-validation was implemented at Step 5 by dividing the trials and putting them in three different blocks. ERP signals were calculated at Step 5 by taking the average of EEG data at each time point and Steps 6 to 8 were for choosing taring data, classifying the training data, and then testing the classifier on test data. Finally, in Step 9 classification score was calculated for each classification attempt.

The proposed decoding method by Bae and Luck (2018) (Figure 3) was a participant-based approach, so decoding starts with choosing a participant (Step 1). Then the trials for each participant were shuffled (Step 2) and after choosing a time point in the interesting time range of decoding (Step 3), the threefold cross-validation method was used (Step 4) where the shuffled data were randomly divided into three blocks and ERPs were calculated for each block (Step 5). Two of these blocks (2/3 of trials) were used for training (Step 6) and the remaining block (1/3 of trials) was used for testing the classifier (Step 8). This process was repeated three times until all blocks were used as the testing block. A combination of SVM and error-correcting output codes (ECOC) was used for the ERP classification (Step 9). The ECOC model solves multi-class categorization problems by combining results from multiple binary classifiers. In this study, there were 12 unique FNCs in the stimulus set; four numbers (1, 2, 3, and 4) for each of the three FNC categories (montring, counting, and noncanonical). These 12 FNCs constituted the classes for the decoding process (a total of 12 classes). To implement a threefold cross-validation, a matrix of 3 groups of 12 finger configurations and 32 electrodes was created for each time point. The ECOC model took two training data sets with known finger configuration labels and trained 12 SVMs. The decoding method was one vs. all; each class was separated from the other 11 classes at the current time point through a binary classification. Next, the set of 12 trained ECOC models was used to make predictions using the MATLAB

predict function for the unlabeled FNCs used for testing. The *predict* function minimizes the average binary loss over the 12 SVMs to predict a label for the test data set. Finally, real labels of FNCs were compared to the predicted labels to compute the classification score (Step 9). The classification score for each participant at the given time point was a 2D matrix with a dimension of the number of FNCs * number of cross-validation blocks.

In this study, for each participant, the shuffling EEG data at Step 2 was repeated 10 times which means ten iterations, and for each iteration, decoding was done separately for 100-time points at the time range of -500 ms to 1500 ms (data points were selected with a frequency of 20 Hz). Therefore, after completing all the iterations and cross-validation procedure, the classification score for each participant was a 4D matrix with a dimension of the number of iterations * number of time points * number of cross-validation * number of classes. Also, the total number of decoding attempts for each participant was the product value of number of iterations * number of time points*number of cross-validation * number of classes (i.e., 10*100*3*12). Accuracy rates for each time point, separately for each participant were calculated and then the point-by-point accuracy rates were then averaged across the participants to get the final decoding accuracy rates for each time point.

The decoding steps for the two types of analyses, ERP-based and alpha-based, were identical, except for the nature of the time series data; instantaneous alpha-power for the alpha-based analysis and EEG amplitudes for the ERP-based analysis.

Results

Alpha-Based and ERP-Based Decoding

Since there were 12 classes the chance level accuracy was 8.3% (1/12), meaning that if the signal analyzed (alpha-power or ERP scalp distributions) contained no identifying information about FNC categories, the decoding accuracy would be expected to be around 8.3%. For both analyses, the decoding accuracy went above the chance level, but higher for the ERP-based analysis (Figure 4).

Figure 4

Mean Accuracy of Alpha-Based Decoding and ERP-Based Decoding Averaged Across All 12 FNCs
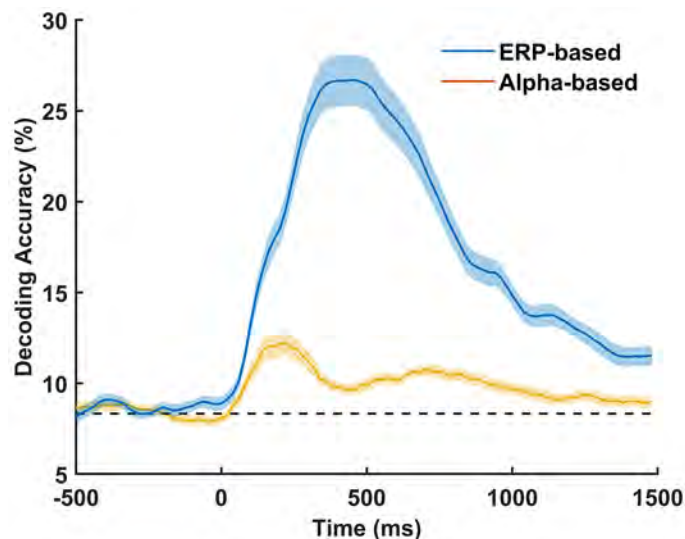
Note. The black horizontal line indicates the chance-level performance (1/12 ≈ 0.083≈ 8.3%). The shaded areas indicate ± 1 SEM.

We averaged the decoding accuracies across all 12 FNCs to compare the performance of the two decoding analyses. The decoding accuracy peaked at 220 ms with a 12.3% accuracy rate for the alpha-based decoding. The ERP-based decoding

was robust and showed extensive time windows where accuracy was greater than chance. The maximum decoding accuracy across all time points for the ERP-based decoding was 26.7% (at 490 ms), which is more than threefold the chance level accuracy. This is a noticeable result and implies that the decoding approach applies to the domain studied. The results indicated that both ERP-based and alpha-based decoding provided above-chance level accuracy for decoding FNCs. However, overall, the ERP-based accuracy was higher than the alpha-based accuracy; therefore, we implemented only the ERP-based method for the category-specific (montring vs. counting vs. noncanonical) decoding.

An addition to the decoding accuracy, confusion matrices provided for both alpha-based and ERP-based decoding ERP-based decoding (Figure 5). These confusion matrices were obtained by comparing true classes and predicted classes across all participants and all decoding attempts for each participant (i.e., 38 participants and 36000 decoding attempts for each participant). So, the confusion matrices are not obtained for a specific time point and cover all the time ranges. In the confusion matrices, the vertical axis represents true classes/stimuli, and the horizontal axis represents the predicted classes. Each cell value has been normalized by the number of observations that has the same true class. The values in the diagonal of the matrix represent the probability of correct prediction for the class (i.e., true positive rate). The values outside of the diagonal in each row represent the false positive rate when other numbers were incorrectly classified as the true class in that row. The values outside of the diagonal in each column represent a false negative rate where other classes were misclassified as that class. The results of the confusion matrices revealed that noncanonical FNC number four (NC4) had the highest true positive rate of 19.1% for ERP-based decoding, however, counting FNC number one (C1) had the highest true positive rate of 10.8% for alpha-based decoding. According to the results, counting FNC number three (C3) and montring FNC number two (M2) had the lowest true positive rate of 8.6% and 12.5% for alpha-based and ERP-based decoding; respectively. For all 12 FNCs, the ERP-based decoding had higher accuracy, compared to the alpha-based decoding.

Figure 5

Percentages of Correct (Blue) and Incorrect (White) Classifications for Each of the 12 FNCs, for Both Alpha-Based (Left) and the ERP-Based (Right) Analyses
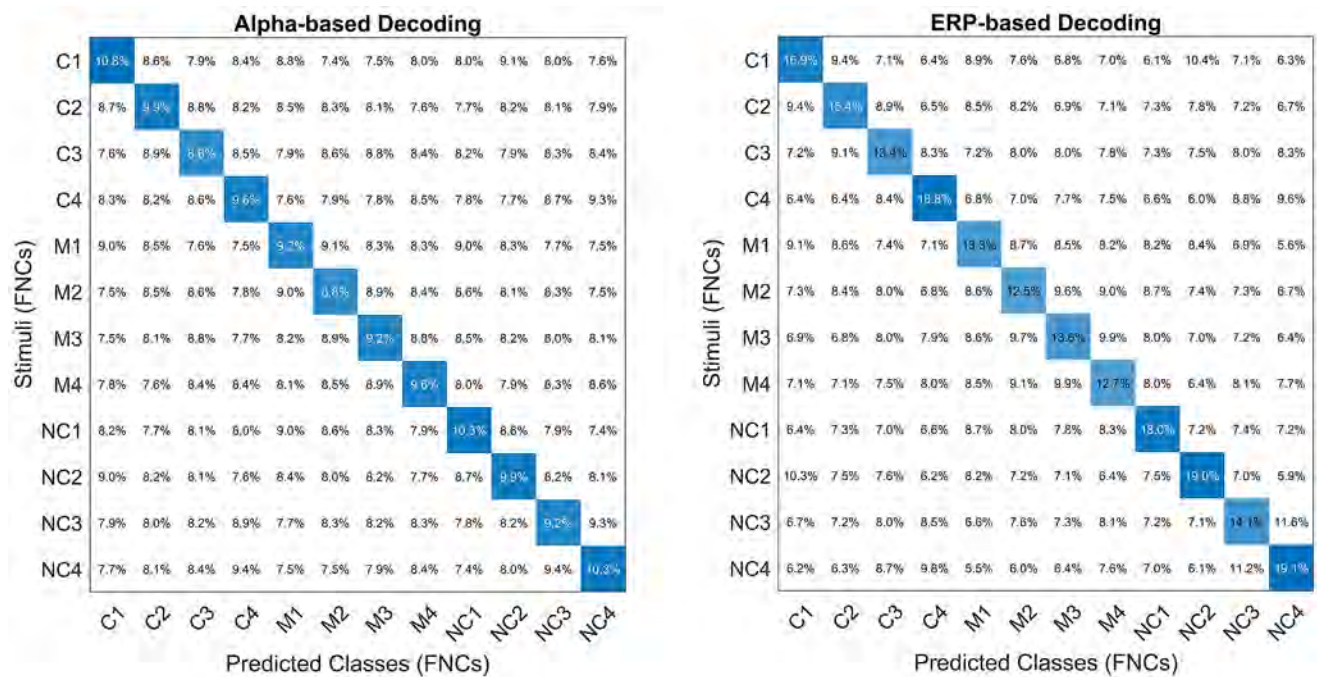
Category-Specific ERP-Based Decoding

The decoding of numerical information contained in FNCs showed higher success for the ERP-based decoding, compared to the alpha-based one. This result parallels what was reported by Bae and Luck (2018, 2019), where ERP-based decoding outperformed alpha-based in two decoding studies. Since our focus in this study was to investigate the processing differences across the three different types of FNCs, we conducted a new set of ERP-based decoding analyses for montring, counting, and noncanonical configurations separately. The decoding procedures were identical to the previous decoding analysis that considered all 12 FNCs together. The only difference was that there were only four classes (FNCs associated with numbers 1, 2, 3, and 4) defined in the decoding step, separately for each of the three types of FNCs. Therefore, the chance level accuracy was 25% (1/4), implying that if the scalp distribution contains no information about the FNCs, the decoding accuracy should be around 0.25.

The average decoding accuracy across all four numbers was calculated separately for each type of FNC (M, C, NC). Then, the average accuracy data over time was smoothed with a five-point moving window Gaussian filter to improve the signal-to-noise ratio (Figure 6).

Figure 6

Mean Accuracy of ERP-Based Decoding for Montring (M), Counting (C), and Noncanonical (NC) Configurations
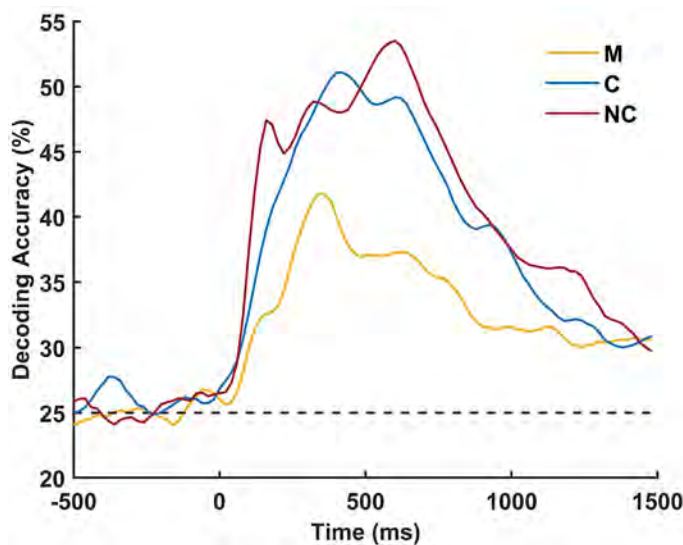
Note. The black horizontal line indicates the chance-level performance (0.25 = 1/4).

Averaged accuracy over time, peak accuracy, and time of peak accuracy in the 100 to 1000 ms range were calculated as dependent measures (Table 1). Three one-way ANOVAs were conducted to test for average accuracy over time, peak accuracy, and time of peak accuracy differences across the M, C, and NC configurations.

Table 1

Average Accuracy, Peak Accuracy, and Time of Peak Accuracy for the ERP-Based Analysis

FNC	Average accuracy (%) M (SD)	Peak accuracy (%) M (SD)	Peak accuracy time (ms) M (SD)
Montring (M)	36.5 (6.6)	47.2 (8.9)	445.789 (186.629)
Counting (C)	44.3 (9.6)	57.7 (11.4)	506.316 (184.192)
Noncanonical (NC)	46.3 (10.9)	60.4 (12.6)	515.263 (194.377)

The one-way ANOVAs were significant for average accuracy and peak accuracy, but not for peak accuracy time (Table 2). Post-hoc pairwise *t*-tests for significant ANOVAs showed that both average accuracy and peak accuracy was significantly lower for montring compared to counting and noncanonical configurations, while the counting and noncanonical configurations did not differ (Table 3). The peak accuracy time was earliest for montring, followed by counting, and then noncanonical, however, these differences were not significant.

Table 2

One-Way ANOVA Results for Average Accuracy, Peak Accuracy, and Time of Peak Accuracy for ERP-Based Analysis

| Dependent Var.	df	Mean square	F	p	η2
Average accuracy	2	0.123	24.339	< .001	0.397
Peak accuracy	2	0.186	27.330	< .001	0.425
Peak accuracy time	2	54277.193	1.492	0.232	0.039

Table 3

Post-Hoc Comparisons for Average Accuracy, Peak Accuracy, and Time of Peak Accuracy for the ERP-Based Analysis

Post-Hoc Tests		Mean difference	SE	t	Cohen's d	p_{holm}
Average accuracy						
M	C	-0.087	0.016	-5.319	-0.943	< .001
M	NC	-0.107	0.016	-6.570	-1.165	< .001
C	NC	-0.020	0.016	-1.251	-0.222	0.215
Peak accuracy						
M	C	-0.106	0.019	-5.582	-0.951	< .001
M	NC	-0.132	0.019	-6.989	-1.191	< .001
C	NC	-0.027	0.019	-1.408	-0.240	0.163

In addition, confusion matrices for all category-specific ERP-based decoding (i.e., M, C, and NC) were provided to investigate any finger-specific difference in decodability (see Figure 7). As mentioned before, the provided confusion matrices were not time specific and were extracted by comparing true classes and predicted classes across all participants and all decoding attempts for each category. Also, each cell value has been normalized by the number of observations that has the same true class. The decoding matrices show that montring number one (M1) has the highest true positive rate (33.5%) for M. Counting FNC for number four (C4) also has the highest true positive rate (40.7%) for C, while noncanonical FNC for number two (NC2) had the highest true positive rate (41.2%) for the NC.

The Percentage of Accurate Classifications for Each of the 4 FNCs Is Marked With Blue and Incorrect Ones With White, Separately for the ERP-Based Analyses of M, C, and NC
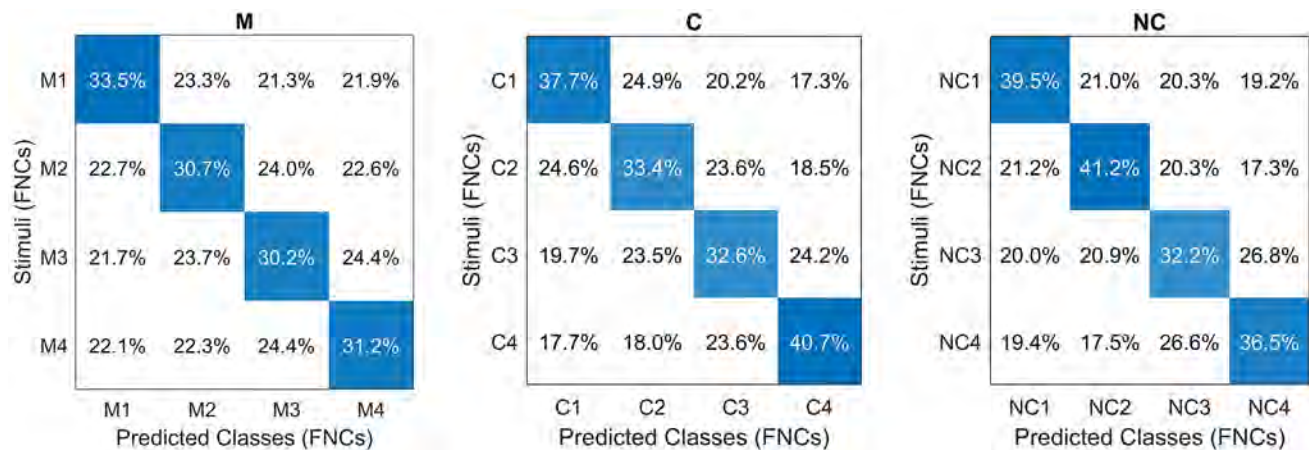
Note. The columns on the right of each figure show the percentage of accurate (blue) and inaccurate (white) classifications for each FNC.

Discussion

The present study investigated processing differences across three types of FNCs (montring, counting, and noncanonical) using a novel decoding approach. We first tested whether scalp-recorded EEG signals contain decodable information about the numerical information contained in FNCs, and which of the two measures, instantaneous alpha-power vs. sustained EEG potentials (ERP scalp distributions), provided higher decoding accuracies. The results demonstrated above-chance level decoding accuracies for both analyses, however, ERP-based analysis showed higher accuracies across all FNCs, compared to alpha-based. Secondly, ERP-based decoding was used to investigate processing differences across the three types of FNCs.

A comparison of the results reported here with the results reported in a previous study (Soylu et al., 2019), employing traditional univariate ERP analyses with the same data set, shows additional insights that can be acquired with the decoding method. Alpha oscillations are associated with spatial attention in early perceptual processing (Bae & Luck, 2018; Worden et al., 2000). Even though the decoding accuracy based on alpha-power was relatively low compared to ERP scalp distributions, alpha-based decoding results showed a relatively early peak accuracy of up to 12.0% at 220ms. This was slightly above the chance level (~ 8.3%). This might be interpreted as visual differences across FNCs impacting alpha-band activity in perceptual processing. In the previous ERP study, Soylu et al. (2019) reported differences in the P1/N1 range (100-210 ms) across FNCs, which were argued to be related to visuospatial attention. Alpha-power decoding accuracy peaking around the time interval with P1-N1 possibly hints at modulation of the same neural sources for both effects, especially given that previously alpha oscillations were found to be associated with the P1-N1 complex (Klimesch et al., 2004; Sauseng et al., 2005). However, we are cautious about this suggestion since alpha-power encoding accuracy was relatively low and did not constitute the main focus of our analysis.

A remarkable finding is how the ERP-based decoding analysis informed the timing of when the bulk of the processing associated with the retrieval of numerical information from FNCs happens. The peak decoding accuracy times for both counting and noncanonical FNCs were post-500 ms (577 and 604 ms respectively), while it was pre-500 ms only for montring (472 ms). The accuracy rates at these peaks were more than double the chance level for counting and noncanonical (47.2%, 57.7%, and 60.04% percent respectively), showing that EEG scalp distributions were successfully decoded around these time points to predict the numerical information included in the FNCs. In the previous study, Soylu et al. (2019) limited the ERP analysis to 500 ms after the FNCs were shown to the participants. This was because at 500 ms the Arabic numerals were presented for validation. Soylu et al. measured mean amplitudes between 100−150 ms, 150−210 ms, and 250−500 ms (for the P1, N1, and P3 components, respectively) to characterize processing differences

between canonical and noncanonical FNCs. In a different ERP study (van den Berg et al., 2021), ERP differences between canonical and noncanonical FNCs were investigated in the 220-310 ms and 280-550 ms windows for the P2p and P3 components. These exemplify how researchers have to make a priori assumptions about the crucial time intervals when hypothesized effects are expected to happen. In this case, the EEG signal provided the most distinguishable information about the numerical information carried in FNCs in the post-500 ms window when the FNC stimulus was no longer available to the participant. The observed effect either shows that there is a delayed working memory process taking place to retrieve the numerical magnitude for the FNC presented or the signal associated with the processing of the validation Arabic numeral (not the FNC) causes the post-500 ms peak. The latter explanation is unlikely, given that around the 500-650 ms period the Arabic numeral stimulus is still in the early perceptual stages of processing (when the numerical magnitude is yet to be processed) and, therefore, the peak decoding accuracies in the post-500 ms accuracies likely reflect a working memory process associated with the retrieval of the numerical magnitude from the FNC presented in the 0-500 ms period.

The comparison of results from the decoding analysis and previous ERP studies on FNCs shows that, in general, decoding analyses can be instrumental in identifying stages of processing that best characterize differences across the conditions studied. This can especially be helpful in areas of numerical cognition research that have not been extensively studied using ERP methods, where researchers cannot rely on previous research to determine ERP components of interest. It should be noted that similar issues with traditional ERP analysis—where time intervals and electrode sites associated with specific ERP components need to be identified a priori—were previously discussed and mass univariate methods were proposed as an alternative (Groppe et al., 2011a, 2011b). Mass univariate analysis is powerful in that it does not assume a normal distribution in the data and instead uses a distribution derived from permuting the data. In addition, it allows the comparison of two conditions across a wider range of time intervals and electrode sites, while effectively controlling for multiple comparisons. Multivariate decoding analyses have the same advantages but rely on prediction accuracy rather than the magnitude of t-scores in identifying where in the time series data two conditions show differences. These alternative approaches can be complementary to or replace traditional univariate approaches.

The results also informed processing differences across the three types of FNCs (montring, counting, and noncanonical). Average decoding accuracy and peak accuracy were significantly lower for montring, while counting and noncanonical did not differ. The behavioral data, reported in the previous ERP study (Soylu et al., 2019), showed a similar pattern, where montring configurations were identified faster and more accurately, compared to counting and noncanonical configurations. Montring also showed higher P1/N1 amplitude than counting and noncanonical FNCs, but similar P3 amplitude to counting. The decoding results imply that there is less distinguishable processing taking place, across the four numbers, for montring, compared to counting and noncanonical. This can be explained as the processing of montring configurations being more automatic and less effortful, which was also argued for in previous studies (Di Luca & Pesenti, 2010; Soylu et al., 2019; Van den Berg et al., 2022). The common argument across these studies was that montring configurations are processed similarly to Arabic numerals, with automatized access to numerical magnitudes. This automaticity leads to less distinct ERP scalp distributions for the montring FNCs showing four numbers (1, 2, 3, 4), making it harder for the SVM to categorize the FNCs and leading to lower decoding accuracy. In contrast, the processing of counting and noncanonical FNCs are less automatic and require more effortful processing, leading to more distinct ERP signatures and making it easier for the algorithm to categorize the FNCs, which results in higher rates of decoding accuracy. This explanation is corroborated by the behavioral results showing faster response times and higher accuracy for montring. Even though peak accuracy time differences were not significant, montring accuracy peaked more than 100 ms before counting (472 ms), compared to counting (577 ms) and noncanonical (604 ms), also corroborating the automaticity explanation.

The confusion matrix results provided insightful information about the decodability of each FNC compared to all other FNCs and within their categories. These results follow the general pattern in the study where ERP-based decoding offers higher true positive rates than the alpha-based decoding results. While these differences in the true positive rate may be driven by processing salient features of that FNC (e.g., NC2 having a higher true positive rate because of its odd non-numerical gestalt), the percentages of correct classifications are somewhat comparable across all FNCs and within categories.

Overall, these results show the advantage of decoding methods in elucidating the processing of internalized representations of FNCs and in numerical cognition studies in general. Compared to traditional ERP analysis, ERP decoding allows the comparison of processing differences across wider time windows and the entire scalp distribution. These findings show the versatility of ERP decoding methods and point to their relevance for numerical cognition studies.

Conclusion

This study aimed to investigate the neural processing of FNCs using a multivariate decoding method and compare this method with conventional univariate ERP analysis. A support vector machine (SVM) was used as the decoding and classification algorithm. Both ERP scalp distributions and alpha-band power were used for decoding FNCs. The results showed higher accuracy rates during the entire time interval when ERP scalp distributions were used. The results also showed that ERP scalp distribution is best at distinguishing task categories in the early stages of perceptual processing, paralleling previous results.

The results informed processing differences across the three FNCs; montring, counting, and noncanonical. Counting and noncanonical FNCs showed higher averaged and peak accuracy rates, compared to montring. The comparison of the decoding results with the behavioral and conventional ERP analysis results reveals a complementary picture, where recognition of montring FNCs are found to be more automatic and less effortful. Montring configurations were identified faster and more accurately, corroborating the automaticity explanation. Faster and more automatized access to numerical information for montring configurations leads to less distinct ERP scalp distribution signatures, leading to lower decoding accuracy rates.

The decoding analysis better complemented the behavioral results compared to conventional ERP analysis, pointing to the advantages of decoding over conventional ERP analysis. In particular, a priori selection of time intervals and electrode sites associated with specific ERP components constitute challenges with novel domains of research, not previously extensively studied with conventional ERP methods. The decoding approach considers wider aspects of data—the entire scalp distribution instead of specific electrode sites, and wider time windows—when characterizing differences across task conditions.

Supplementary Materials

The Supplementary Materials contain a custom MATLAB script using EEGLAB (Delorme & Makeig, 2004) and ERPLAB (Lopez-Calderon & Luck, 2014) functions for data pre-processing (for access see Index of Supplementary Materials below).

Index of Supplementary Materials

Salehzadeh, R., Rivera, B., Man, K., Jalili, N., & Soylu, F. (2023). *Supplementary materials to "EEG decoding of finger numeral configurations with machine learning"* [Code]. GitHub. https://github.com/firatsoylu/ERP_Decoding_Preprocessing/blob/d5c58312abb1430c15a875bacef144dc18d7af9b/A0_ERP_Preprocessing.m

References

Andres, M., Seron, X., & Olivier, E. (2007). Contribution of hand motor circuits to counting. *Journal of Cognitive Neuroscience, 19*(4), 563–576. https://doi.org/10.1162/jocn.2007.19.4.563

Awh, E., & Jonides, J. (2001). Overlapping mechanisms of attention and spatial working memory. *Trends in Cognitive Sciences, 5*, 119–126. https://doi.org/10.1016/S1364-6613(00)01593-X

Badets, A., Andres, M., Di Luca, S., & Pesenti, M. (2007). Number magnitude potentiates action judgements. *Experimental Brain Research, 180*(3), 525–534. https://doi.org/10.1007/s00221-007-0870-y

Badets, A., Pesenti, M., & Olivier, E. (2010). Response-effect compatibility of finger-numeral configurations in arithmetical context. *The Quarterly Journal of Experimental Psychology, 63*(1), 16–22. https://doi.org/10.1080/17470210903134385

Bae, G.-Y. (2017). *Dissociable decoding of working memory and spatial attention from EEG oscillations and sustained potentials* [Analysis scripts and data]. OSF. https://osf.io/bpexa/

Bae, G.-Y., Leonard, C. J., Hahn, B., Gold, J. M., & Luck, S. J. (2020). Assessing the information content of ERP signals in schizophrenia using multivariate decoding methods. *NeuroImage: Clinical, 25*, Article 102179. https://doi.org/10.1016/j.nicl.2020.102179

Bae, G.-Y., & Luck, S. J. (2018). Dissociable decoding of spatial attention and working memory from EEG oscillations and sustained potentials. *The Journal of Neuroscience, 38*(2), 409–422. https://doi.org/10.1523/JNEUROSCI.2860-17.2017

Bae, G.-Y., & Luck, S. J. (2019). Decoding motion direction using the topography of sustained ERPs and alpha oscillations. *NeuroImage, 184*(August 2018), 242–255. https://doi.org/10.1016/j.neuroimage.2018.09.029

Barrocas, R., Roesch, S., Gawrilow, C., & Moeller, K. (2020). Putting a finger on numerical development – Reviewing the contributions of kindergarten finger gnosis and fine motor skills to numerical abilities. *Frontiers in Psychology, 11*, Article 1012. https://doi.org/10.3389/fpsyg.2020.01012

Bender, A., & Beller, S. (2012). Nature and culture of finger counting: Diversity and representational effects of an embodied cognitive tool. *Cognition, 124*(2), 156–182. https://doi.org/10.1016/j.cognition.2012.05.005

Berteletti, I., & Booth, J. R. (2015). Perceiving fingers in single-digit arithmetic problems. *Frontiers in Psychology, 6*(MAR), Article 226. https://doi.org/10.3389/fpsyg.2015.00226

Berteletti, I., & Booth, J. R. (2016). Finger representation and finger-based strategies in the acquisition of number meaning and arithmetic. In D. Berch, D. C. Geary, & K. Koepke (Eds.), *Development of mathematical cognition* (pp. 109–139). Elsevier. https://doi.org/10.1016/B978-0-12-801871-2.00005-8

de Freitas, E., & Sinclair, N. (2013). New materialist ontologies in mathematics education: The body in/of mathematics. *Educational Studies in Mathematics, 83*(3), 453–470. https://doi.org/10.1007/s10649-012-9465-z

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods, 134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Di Luca, S., Lefèvre, N., & Pesenti, M. (2010). Place and summation coding for canonical and non-canonical finger numeral representations. *Cognition, 117*(1), 95–100. https://doi.org/10.1016/j.cognition.2010.06.008

Di Luca, S., & Pesenti, M. (2008). Masked priming effect with canonical finger numeral configurations. *Experimental Brain Research, 185*(1), 27–39. https://doi.org/10.1007/s00221-007-1132-8

Di Luca, S., & Pesenti, M. (2010). Absence of low-level visual difference between canonical and noncanonical finger-numeral configurations. *Experimental Psychology, 57*(3), 202–207. https://doi.org/10.1027/1618-3169/a000025

Ghasemi, E., Ebrahimi, M., & Ebrahimie, E. (2022). Machine learning models effectively distinguish attention-deficit/hyperactivity disorder using event-related potentials. *Cognitive Neurodynamics, 16*, 1335–1349. https://doi.org/10.1007/s11571-021-09746-2

Grootswagers, T., Wardle, S. G., & Carlson, T. A. (2017). Decoding dynamic brain patterns from evoked responses: A tutorial on multivariate pattern analysis applied to time series neuroimaging data. *Journal of Cognitive Neuroscience, 29*(4), 677–697. https://doi.org/10.1162/jocn_a_01068

Groppe, D. M., Urbach, T. P., & Kutas, M. (2011a). Mass univariate analysis of event-related brain potentials/fields I: A critical tutorial review. *Psychophysiology, 48*(12), 1711–1725. https://doi.org/10.1111/j.1469-8986.2011.01273.x

Groppe, D. M., Urbach, T. P., & Kutas, M. (2011b). Mass univariate analysis of event-related brain potentials/fields II: Simulation studies. *Psychophysiology, 48*(12), 1726–1737. https://doi.org/10.1111/j.1469-8986.2011.01272.x

Hebart, M. N., & Baker, C. I. (2018). Deconstructing multivariate decoding for the study of brain function. *NeuroImage, 180*, 4–18. https://doi.org/10.1016/j.neuroimage.2017.08.005

Jimura, K., & Poldrack, R. A. (2012). Analyses of regional-average activation and multivoxel pattern information tell complementary stories. *Neuropsychologia, 50*(4), 544–552. https://doi.org/10.1016/j.neuropsychologia.2011.11.007

Klimesch, W., Schack, B., Schabus, M., Doppelmayr, M., Gruber, W., & Sauseng, P. (2004). Phase-locked alpha and theta oscillations generate the P1–N1 complex and are related to memory performance. *Cognitive Brain Research, 19*(3), 302–316. https://doi.org/10.1016/j.cogbrainres.2003.11.016

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: An open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience, 8*, Article 213. https://doi.org/10.3389/fnhum.2014.00213

Michaux, N., Masson, N., Pesenti, M., & Andres, M. (2013). Selective interference of finger movements on basic addition and subtraction problem solving. *Experimental Psychology, 60*(3), 197–205. https://doi.org/10.1027/1618-3169/a000188

Nemirovsky, R., & Ferrara, F. (2009). Mathematical imagination and embodied cognition. *Educational Studies in Mathematics, 70*(2), 159–174. https://doi.org/10.1007/s10649-008-9150-4

Newman, S. D., & Soylu, F. (2014). The impact of finger counting habits on arithmetic in adults and children. *Psychological Research, 78*(4), 549–556. https://doi.org/10.1007/s00426-013-0505-9

Núñez, R. (2012). On the science of embodied cognition in the 2010s: Research questions, appropriate reductionism, and testable explanations. *Journal of Learning Sciences, 21*, 324–336. https://doi.org/10.1080/10508406.2011.614325

Ritchie, S. J., & Bates, T. C. (2013). Enduring links from childhood mathematics and reading achievement to adult socioeconomic status. *Psychological Science, 24*(7), 1301–1308. https://doi.org/10.1177/0956797612466268

Rose, N. S., LaRocque, J. J., Riggall, A. C., Gosseries, O., Starrett, M. J., Meyering, E. E., & Postle, B. R. (2016). Reactivation of latent working memories with transcranial magnetic stimulation. *Science, 354*, 1136–1139. https://doi.org/10.1126/science.aah7011

Rusconi, E., Walsh, V., & Butterworth, B. (2005). Dexterity with numbers: RTMS over left angular gyrus disrupts finger gnosis and number processing. *Neuropsychologia, 43*, 1609–1624. https://doi.org/10.1016/j.neuropsychologia.2005.01.009

Sato, M., Cattaneo, L., Rizzolatti, G., & Gallese, V. (2007). Numbers within our hands: Modulation of corticospinal excitability of hand muscles during numerical judgment. *Journal of Cognitive Neuroscience, 19*(4), 684–693. https://doi.org/10.1162/jocn.2007.19.4.684

Sauseng, P., Klimesch, W., Stadler, W., Schabus, M., Doppelmayr, M., Hanslmayr, S., Gruber, W. R., & Birbaumer, N. (2005). A shift of visual spatial attention is selectively associated with human EEG alpha activity. *European Journal of Neuroscience, 22*(11), 2917–2926. https://doi.org/10.1111/j.1460-9568.2005.04482.x

Siegler, R. S., & Braithwaite, D. W. (2017). Numerical development. *Annual Review of Psychology, 68*, 187–213. https://doi.org/10.1146/annurev-psych-010416-044101

Soylu, F. (2019). Public dataset: ERP differences in processing canonical and noncanonical finger-numeral configurations. *Harvard Dataverse*. https://doi.org/10.7910/DVN/BNNSRG

Soylu, F., Lester, F. K., Jr., & Newman, S. D. (2018). You can count on your fingers: The role of fingers in early mathematical development. *Journal of Numerical Cognition, 4*(1), 107–135. https://doi.org/10.5964/jnc.v4i1.85

Soylu, F., & Newman, S. (2016). Anatomically ordered tapping interferes more with one-digit addition than two-digit addition: A dual-task fMRI study. *Cognitive Processing, 17*(1), 67–77. https://doi.org/10.1007/s10339-015-0737-2

Soylu, F., Raymond, D., Gutierrez, A., & Newman, S. (2018). The differential relationship between finger gnosis, and addition and subtraction: An fMRI study. *Journal of Numerical Cognition, 3*(3), 694–715. https://doi.org/10.5964/jnc.v3i3.102

Soylu, F., Rivera, B., Anchan, M., & Shannon, N. (2019). ERP differences in processing canonical and noncanonical finger-numeral configurations. *Neuroscience Letters, 705*, 74–79. https://doi.org/10.1016/j.neulet.2019.04.032

van den Berg, F. C. G., de Weerd, P., & Jonkman, L. M. (2021). Electrophysiological evidence for internalized representations of canonical finger-number gestures and their facilitating effects on adults' math verification performance. *Scientific Reports, 11*(1), Article 11776. https://doi.org/10.1038/s41598-021-91303-2

Van den Berg, F. C. G., De Weerd, P., & Jonkman, L. M. (2022). Canonical finger-numeral configurations facilitate the processing of Arabic numerals in adults: An event-related potential study. *Neuropsychologia, 170*, Article 108214. https://doi.org/10.1016/j.neuropsychologia.2022.108214

van Ede, F., Niklaus, M., Nobre, A. C. (2017). Temporal expectations guide dynamic prioritization in visual working memory through attenuated α oscillations. *Journal of Neuroscience, 37*(2), 437–445. https://doi.org/10.1523/JNEUROSCI.2272-16.2016

Wasner, M., Nuerk, H.-C., Martignon, L., Roesch, S., & Moeller, K. (2016). Finger gnosis predicts a unique but small part of variance in initial arithmetic performance. *Journal of Experimental Child Psychology, 146*, 1–16. https://doi.org/10.1016/j.jecp.2016.01.006

Worden, M. S., Foxe, J. J., Wang, N., & Simpson, G. V. (2000). Anticipatory biasing of visuospatial attention indexed by retinotopically
   specific α-bank electroencephalography increases over occipital cortex. *The Journal of Neuroscience, 20*(6), RC63–RC63.
   https://doi.org/10.1523/JNEUROSCI.20-06-j0002.2000