Received: 8 August 2019 | Revised: 11 October 2019 | Accepted: 13 November 2019

DOI: 10.1111/psyp.13511

ORIGINAL ARTICLE

PSYCHOPHYSIOLOGY SPR WILEY

A multivariate investigation of visual word, face, and ensemble processing: Perspectives from EEG-based decoding and feature selection

Dan Nemrodov | Shouyu Ling | Ilya Nudnou | Tyler Roberts | Jonathan S. Cant | Andy C. H. Lee | Adrian Nestor

Department of Psychology at Scarborough, University of Toronto, Toronto, Ontario, Canada

Correspondence
Adrian Nestor, Department of Psychology at Scarborough, University of Toronto, 1265 Military Trail, Toronto, ON M1C1A4, Canada.
Email: adrian.nestor@utoronto.ca

Funding information
This research was supported by the Natural Sciences and Engineering Research Council of Canada (JSC, ACHL, AN)

Abstract

Recent investigations have focused on the spatiotemporal dynamics of visual recognition by appealing to pattern analysis of EEG signals. While this work has established the ability to decode identity-level information (such as the identity of a face or of a word) from neural signals, much less is known about the precise nature of the signals that support such feats, their robustness across visual categories, or their consistency across human participants. Here, we address these questions through the use of EEG-based decoding and multivariate feature selection as applied to three visual categories: words, faces and face ensembles (i.e., crowds of faces). Specifically, we use recursive feature elimination to estimate the diagnosticity of time and frequency-based EEG features for identity-level decoding across three datasets targeting each of the three categories. We then relate feature diagnosticity across categories and across participants while, also, aiming to increase decoding performance and reliability. Our investigation shows that word and face processing are similar in their reliance on spatiotemporal information provided by occipitotemporal channels. In contrast, ensemble processing appears to also rely on central channels and exhibits a similar profile with word processing in the frequency domain. Further, we find that feature diagnosticity is stable across participants and is even capable of supporting cross-participant feature selection, as demonstrated by systematic boosts in decoding accuracy and feature reduction. Thus, our investigation sheds new light on the nature and the structure of the information underlying identity-level visual processing as well as on its generality across categories and participants.

KEYWORDS

analysis/statistical methods, content/topics, content/topics, EEG, hemispheric differences/laterality, methods, methods, visual processes

Dan Nemrodov and Shouyu Ling are contributed equally to this work.

The University of Toronto has filed a U.S. patent application that includes portions of the method for feature selection described here. Adrian Nestor and Dan Nemrodov are co-inventors on this patent.

 © 2019 Society for Psychophysiological Research

1 | INTRODUCTION

The spatiotemporal dynamics of visual recognition have elicited a significant interest in the study of perceptual processing (Bar et al., 2006; Cichy, Pantazis, & Oliva, 2014; Isik et al., 2014; King & Dehaene, 2014; Retter & Rossion, 2016). Recently, work capitalizing on the robustness and versatility of pattern analysis has targeted these dynamics by examining the multivariate structure of neural data. For instance, the time course of individual face processing has been assessed by multivariate investigations of magneto/electroencephalography data (Ambrus, Kaiser, Cichy, & Kovács, 2019; Dobs, Isik, Pantazis, & Kanwisher, 2019; Nemrodov, Niemeier, Mok, & Nestor, 2016; Vida, Nestor, Plaut, & Behrmann, 2017). While these investigations revealed critical aspects of neural dynamics, such as the onset of visual recognition, other properties of theoretical and practical significance are yet to be elucidated. Notably, the relative diagnosticity of neural features for decoding purposes as well as its generalizability across distinct visual categories and across participants remain largely unknown. Accordingly, the present work addresses these questions by appeal to identity-level decoding and cross-participant feature selection for multiple visual categories.

Specifically, here we aim to decode the identity of visual stimuli from corresponding electroencephalography (EEG) data for three distinct classes of stimuli: faces, words, and ensembles—to be clear, by ensemble processing we refer here to the ability to encode the average identity of a group of visual objects, such as that of a group of faces (for a review see Whitney & Yamanashi Leib, 2018). Importantly, decoding success within each class serves as a foundation for assessing relative feature diagnosticity, for estimating optimal subsets of features, and for relating them across categories and across participants.

Theoretically, our focus on the three categories above is grounded in previous research relying, in particular, on functional magnetic resonance imaging (fMRI). Specifically, previous work has suggested that single faces and words, despite their widely dissimilar appearance, compete for common processing resources in high-level visual cortex (Dehaene et al., 2010; Harris, Rice, Young, & Andrews, 2016; Hasson, Levy, Behrmann, Hendler, & Malach, 2002; Nestor, Behrmann, & Plaut, 2013). In contrast, face ensembles (i.e., a group of multiple faces presented simultaneously) and single faces, despite both being encompassed by the broad scope of face recognition, appear to recruit different cortical areas: ensembles show particular reliance on dorsal stream areas in parietal cortex while single faces rely considerably on ventral stream areas in occipitotemporal cortex (Im et al., 2017). Hence, the possibility that these categories also exhibit dissimilar dynamics, as reflected in the temporal and frequency domains, is an intriguing hypothesis. In particular, this presents clear interest for theories articulating the neural integration of these domains (Plaut & Behrmann, 2011).

Methodologically, we note that most EEG research in cognitive neuroscience has focused on the global amplitude of the signals (e.g., an average estimate across sensors and time windows). Accordingly, such work relies on extensive data averaging to increase the signal-to-noise ratio (SNR) and to reduce dimensionality, rendering data suitable for univariate analyses. In contrast, data-driven approaches, such as the one pursued here, exploit the multivariate structure of EEG patterns, which is often lost through averaging, while, at the same time, they commonly remove the global signal amplitude from patterns (e.g., through normalization procedures). Thus, traditional univariate approaches and novel pattern analyses tend to exploit complementary properties of the EEG signal. Hence, they can provide complementary but, hopefully, convergent insights into the nature and significance of neural data.

Of particular relevance here, recent work has found evidence for the ability of EEG pattern analyses to support the decoding of facial identity (Ambrus et al., 2019; Nemrodov, Niemeier, Patel, & Nestor, 2018), visual words (Chan, Halgren, Marinkovic, & Cash, 2011; Ling, Lee, Armstrong, & Nestor, 2019), and face ensembles (Roberts, Cant, & Nestor, 2019) suggesting a promising strategy to address cross-category commonalities in neural dynamics through multivariate analyses.

Further, in relationship to this recent work, the present investigation is motivated by the need to boost decoding accuracy and to constrain feature dimensionality. Specifically, feature selection can facilitate the detection of more subtle decoding effects, speed up data collection (e.g., by deploying fewer channels known for their diagnosticity) and reduce computational load in data analysis (e.g., by restricting analyses to smaller pools of relevant features) (Alotaiby, El-samie, Alshebeili, & Ahmad, 2015). To this aim, we turn here to recursive feature elimination (RFE), a powerful method for feature selection (Guyon, Weston, Barnhill, & Vapnik, 2002) with proven value in the analysis of neuroimaging data such as fMRI (De Martino et al., 2008; Nestor et al., 2013) and EEG (Gysels, Renevey, & Celka, 2005; Hidalgo-Muñoz et al., 2013; Yin, Wang, Liu, Zhang, & Zhang, 2017). Briefly, RFE is a wrapper method for feature selection that computes iteratively feature weights, as determined by a discriminant function such as support vector machines (SVM), and removes systematically features with the least impact on discrimination. SVM-RFE generally outperforms univariate approaches for feature selection and is comparable, if not superior, to other multivariate methods (Hidalgo-Muñoz et al., 2013; Shen, Ong, Li, Hui, & Wilder-Smith, 2007). However, SVM-RFE is computationally intensive and relatively demanding with regard to the number of observations needed. For

instance, to avoid circularity, each data set should be sufficiently large to allow its division into multiple subsets for different purposes such as training, validation and testing (De Martino et al., 2008). Here, we handle this requirement by considering jointly the data from multiple participants. In line with transfer learning approaches (Jayaram, Alamgir, Altun, Scholkopf, & Grosse-Wentrup, 2016; Yin et al., 2017), this strategy is intended to both increase the size of our data sets and to examine the robustness of feature selection across participants.

Briefly, our investigation applied the strategy above to three EEG data sets dedicated to the study of facial identity processing (Nemrodov et al., 2018), word recognition (Ling et al., 2019), and ensemble perception (Roberts et al., 2019). Our examination revealed robust patterns of feature diagnosticity within and across participants, similar profiles of spatiotemporal information guiding the decoding of single faces and words, but not of face ensembles, as well as systematic boosts in decoding accuracy accompanied by significant reductions of feature sets for each category. Thus, theoretically, our results point to the integration of perceptual mechanisms for certain visual categories from a neural dynamics perspective and, practically, serve to inform the design and methodology of EEG investigations of visual recognition.

2 | METHOD

2.1 | Experimental procedures

The data examined here were collected as part of three different studies seeking to assess the temporal dynamics of word, face, and ensemble processing (see Figure 1 for examples of stimuli).

Word processing was assessed based on the data of 14 participants (9 females, age: 20–26) who performed a one-back identity task with three-letter consonant–vowel–consonant (CVC) nouns (Ling et al., 2019). Each of 80 words was presented for 300 ms, followed by a variable 600–700 ms intertrial interval (ITI), and was repeated 96 times (i.e., 3 times per block across 32 blocks). Stimuli were centrally presented and subtended a visual angle of 4.9° × 2.9° from a distance of 80 cm.

Face processing was evaluated using the data of 13 participants (7 females, age: 18–27) who performed a go/no-go gender categorization task with 54 male faces, on no-go trials, and 10 female faces, on go trials (Nemrodov et al., 2018). Male and female face stimuli were selected and displayed in a similar manner to show frontal views of young adult individuals with a frontal gaze. All male face images used in the current analysis contained neutral expressions, were centrally presented and subtended a visual angle of 3.2° × 4.9° from a distance of 80 cm. Each face was presented for 300 ms, followed by a variable 925–1,075 ms ITI, and was repeated 64 times (i.e., twice per block across 32 blocks).

Face ensemble processing was assessed based on the data of 14 participants (10 females, age: 19–25) who performed a go/no-go gender categorization task (Roberts et al., 2019). Each of four base ensembles, presented on no-go trials and used in the current analysis, consisted of six different male faces displayed in a circular arrangement. An additional ensemble consisted of female faces on go trials. Of note, participants were asked to fixate the center of the screen, marked by a fixation cross presented prior to the stimulus, rather than fixate any individual face in an ensemble. Also, the position of any given face within an ensemble varied from trial to trial, so that any given base ensemble would vary in its overall appearance from

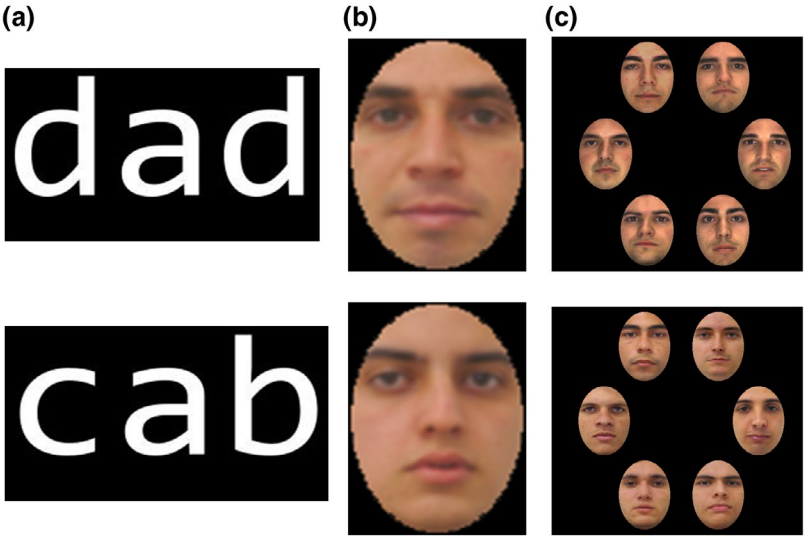

FIGURE 1 Examples of experimental stimuli: (a) words. (b) faces, and (c) face ensembles

trial-to-trial though its component faces remained the same. Each male and female face ensemble subtended an angle of 9° × 7° from a distance of 80 cm. Each ensemble was presented for 300 ms, followed by a variable 600–700 ms ITI, and was repeated 768 times (i.e., 48 times per block across 16 blocks).

All participants in the three studies were right-handed neurotypical adults with normal or corrected-to-normal vision. All participants provided informed consent and all experimental procedures were approved by the Research Ethics Board at the University of Toronto.

In all experiments, EEG data collection relied on a 64-electrode Biosemi ActiveTwo EEG recording system (Biosemi B.V., Amsterdam, Netherlands) with a sampling rate of 512 Hz. The electrodes were arranged according to the International 10/20 System. Each experiment was conducted across two sessions on different days for each participant.

2.2 | Preprocessing

Preprocessing was performed using Matlab 2017a (Mathworks, Natick, MA) and Letswave 6 (Mouraux & Iannetti, 2008). All data were digitally filtered offline (zero-phase 24 dB/octave Butterworth filter) with a bandpass of 0.1–40 Hz. Next, data were separated into epochs, from 100 ms prior to stimulus presentation until 900 ms later, and baseline corrected. Further, epochs were rereferenced to the average reference. In addition, prior to univariate ERP analysis, data were cleaned of ocular artifacts using Infomax ICA (Delorme, Sejnowski, & Makeig, 2007).

Next, for the time domain analysis, for each epoch, an interval of 50–650 ms after stimulus onset was selected and the corresponding data points were concatenated across electrodes into a single vector consisting of 19,712 features (i.e., 308 time bins × 64 electrodes). Then, each vector was normalized by z-scoring across features to remove trial-specific variation in average amplitude. Next, trials within blocks were averaged separately for each stimulus (and also, across consecutive blocks in the case of words and faces), to yield the same number of observations per stimulus and participant for each study, that is, 16 observations for decoding purposes. Further, all observations thus derived were scaled between 0 and 1, feature by feature across observations, separately for each participant, in order to eliminate the impact of differences in range and average magnitude across features.

For the frequency analysis, each1-s-long epoch (i.e., 512 bins between −100 and 900 ms) was individually tapered with a symmetric Hanning window and converted to the frequency domain using FFT. Amplitude values for 40 frequency estimates (i.e., 1–40 Hz) were concatenated across all electrodes into a single vector consisting of 2,560 features (i.e., 40 frequencies × 64 electrodes) per epoch. Then, each vector was normalized by z-scoring across features. Next, feature vectors were averaged separately for each stimulus to yield 16 observations for decoding purposes per stimulus and participant for each study, same as for the time domain analysis described above. Last, observations were scaled between 0 and 1, feature by feature across observations, separately for each participant.

To be clear, we point out that our analyses aimed to disentangle the contribution of temporal and frequency information. Accordingly, frequency features were derived once from a large temporal window rather than multiple times across more compact, distinct windows spanning an epoch. However, it is possible that appealing to time–frequency decomposition to yield a larger number of features that capture the variable structure of spectral information may be beneficial for decoding purposes and this possibility will need to be explored by future research.

2.3 | Stimulus decoding

Pairwise stimulus decoding was conducted separately for the data of each participant in each study with the aid of linear SVM ($c = 1$) and leave-one-out cross-validation. For instance, the discriminability of any two words was estimated by training a classifier on 15 observations pairs and testing the performance of the classifier on the remaining pair. Then, an estimate of classification accuracy was computed by averaging performance across all word pairs. A similar procedure was followed for face and ensemble decoding. Stimulus decoding was conducted separately in the time and frequency domains for each participant using different subsets of features as described below.

2.4 | Feature selection

The contribution of different features (i.e., electrode × time point/frequency amplitude) to decoding was estimated with the aid of SVM-RFE. This wrapper method for feature selection (Guyon et al., 2002) is instrumental in the effort to decrease the dimensionality of relevant observations, to boost SVM-based discriminability and to estimate feature diagnosticity.

Briefly, the method works by repeatedly eliminating the features that are the least diagnostic for a given type of classification—diagnosticity here was measured by a common metric, the square of the classification weights for each feature (Hanson & Halchenko, 2008). Specifically, the method proceeded as follows: (i) a linear SVM classifier was trained on a given feature set, (ii) a ranking score was computed for

all features (based on classification weights), (iii) the batch of features with the smallest ranks were eliminated, and (iv) the procedure was repeated until feature depletion. Thus, for any discrimination, the method produces a ranking of classification features from the least diagnostic (eliminated first) to the most diagnostic (eliminated last).

Of note, SVM-RFE ensures a more robust and reliable ranking than that based on a single-pass classification (e.g., by ordering the weights of a single classification model). However, it is computationally more demanding since classification is performed n times, where n is the number of feature batches in the set, instead of a single time. Here, in order to equate the number of RFE iteration steps (i.e., 64) for the two domains, each batch consisted of 308 features in the time domain and 40 features in the frequency domain for each participant.

To ensure noncircularity and to maximize, at the same time, the use of available data, RFE-based feature selection was performed across different subsets of participants. Accordingly, the method consisted of three stages, as follows: ranking, validation and testing.

First, half of the participants in each study (i.e., with odd or even numbers) were assigned to the ranking group. Feature ranking was computed for each pair of stimuli separately for each participant in this group using SVM-RFE. Estimates of feature diagnosticity were then averaged across stimulus pairs and participants. Based on average ranks, features were divided into 64 batches from the least to the most diagnostic.

Second, validation and testing were conducted using a leave-one-participant-out procedure. Specifically, all remaining participants except one were assigned to a validation group. The aim of validation was to estimate the optimal number of batches for classification purposes. To this end, estimates of classification accuracy were computed for each participant with decreasing number of batches (i.e., 64, 63, and 62). Validation curves were then averaged across participants to identify the feature subset that maximizes accuracy.

Third, independent estimates of classification accuracy were computed using the data of the test subject. Since the procedure was conducted, in turn, with odd and even subjects as part of the ranking group, each participant had the opportunity to provide a test data set exactly once in each experiment. Thus, RFE-optimized performance was estimated as the average of all test participant estimates in each experiment.

2.5 | Feature ranking and consistency

A first set of analyses considered the presence and the characteristics of the rankings based on the first stage of feature selection as follows.

First, to assess the presence of reliable and systematic structure in feature ranking its consistency was estimated both within and across participants for each stimulus domain. Specifically, consistency was computed within participants by correlating ranks across electrodes and time points (or frequency bins) obtained for even versus odd observations. Similarly, between-participant consistency was computed by correlating the ranks for odd observations from one participant with the even observations of every other participant. In addition, to maximize the use of available data and thus, increase the robustness of the estimates, between-participant consistency was recomputed by considering rankings based on all data of a given participant and all data of every other participant.

Correlation values were converted to z-scores using the Fisher transform and were then compared against chance using two-tailed one-sample t tests across participants, separately for each type of feature, consistency estimate, and stimulus category. Further, consistency was assessed using repeated measures ANOVAs (2 feature types: time, frequency × 3 consistency types: within, between-half, between-full) separately for each stimulus category (Greenhouse–Geisser corrections were applied wherever the assumption of sphericity was not supported by the data).

Second, in order to assess the spatial and temporal profile of the data, rankings were compared across stimulus categories and types of features. Specifically, to assess the relative contribution of different channels to decoding, rankings were averaged across time points (or frequency bins) and across participants. Then, the resulting estimates were correlated across stimulus categories. Complementary analyses proceeded in a similar manner by averaging rankings across channels while retaining temporal (or frequency) information and correlating the estimates across stimulus categories.

Next, to examine in more detail the spatial profile of the data, rankings were averaged across time/frequency bins and channels, respectively. Then, rankings were assessed using a three-way repeated-measures ANOVA (3 electrode groups: posterior, central, anterior × 2 hemispheres: left and right × 2 feature types: frequency and time) separately for each stimulus category.

2.6 | Feature selection and decoding performance

To determine the effectiveness of feature selection across stimulus categories, decoding performance was compared before and after selection using independent classification estimates (i.e., from the test data sets). Specifically, decoding accuracy was evaluated using the optimal set of features versus the entire set of features, separately for each stimulus category and feature type, using two-tailed paired t tests.

Next, to assess the possibility that different feature types may contribute complementary information for decoding

purposes, feature selection was conducted again using joint sets of both temporal and frequency features. Then, decoding accuracy using optimal subsets of mixed features was compared against accuracy estimates based on optimal subsets of homogeneous features (i.e., only temporal or frequency based).

3 | RESULTS

3.1 | Ranking consistency/validity

In order to confirm the reliability of feature rankings and their potential utility for feature selection, their consistency was assessed separately for each category of stimuli and type of feature. Specifically, average ranks were compared to each other both within and across participants via correlation and, then, the resulting estimates were compared to chance following their conversion to z scores. All correlations were significant (one-sample t tests against 0 across participants, $p < .001$; Bonferroni corrected) (see Figure 2).

Next, consistency was assessed separately for each category of stimuli via two-way repeated-measures analyses of variance (2 types of features × 3 consistency measures). For words, this analysis found significant main effects of feature type ($F(1, 13) = 48.85$, $p < .001$, $\eta_p^2 = .562$) and consistency measure ($F(1.12, 14.61) = 198$, $p < .001$, $\eta_p^2 = .654$), as well as a significant interaction ($F(1.04, 13.49) = 13.99$, $p = .002$, $\eta_p^2 = .169$). Similarly, for faces, significant effects were found for feature ($F(1, 12) = 25.07$, $p < .001$, $\eta_p^2 = .228$), consistency ($F(1.04, 12.52) = 24.85$, $p < .001$, $\eta_p^2 = .4$), and for the interaction ($F(1.04, 12.43) = 6.39$, $p = .025$, $\eta_p^2 = .085$). Finally, for ensembles, we found significant main effects of feature ($F(1, 13) = 23.73$, $p < .001$, $\eta_p^2 = .258$) and consistency ($F(1.01, 13.18) = 62.66$, $p < .001$, $\eta_p^2 = .608$), but no interaction ($F(2, 26) = 1.3$, $p = .3$, $\eta_p^2 = .015$).

Further post hoc analyses indicated that within-participant consistency estimates were, as expected, higher than their between-participant homologues, while between-participant consistency estimates were typically higher when using all data rather than split halves (see Figure 2).

Thus, feature ranking appears to capture reliable data structure within participants. More importantly, it yields similar outcomes across participants, which is critical for the purpose of cross-participant feature selection and decoding.

3.2 | Feature diagnosticity

To visualize the contribution of different features to decoding, ranks were averaged across participants, separately for each stimulus category, and displayed by channel and time point (Figure 3 top) or frequency bin (Figure 3 bottom). An examination of the results indicates that an extensive set of features contribute to decoding. However, it is also apparent that certain groups of features have less utility, such as

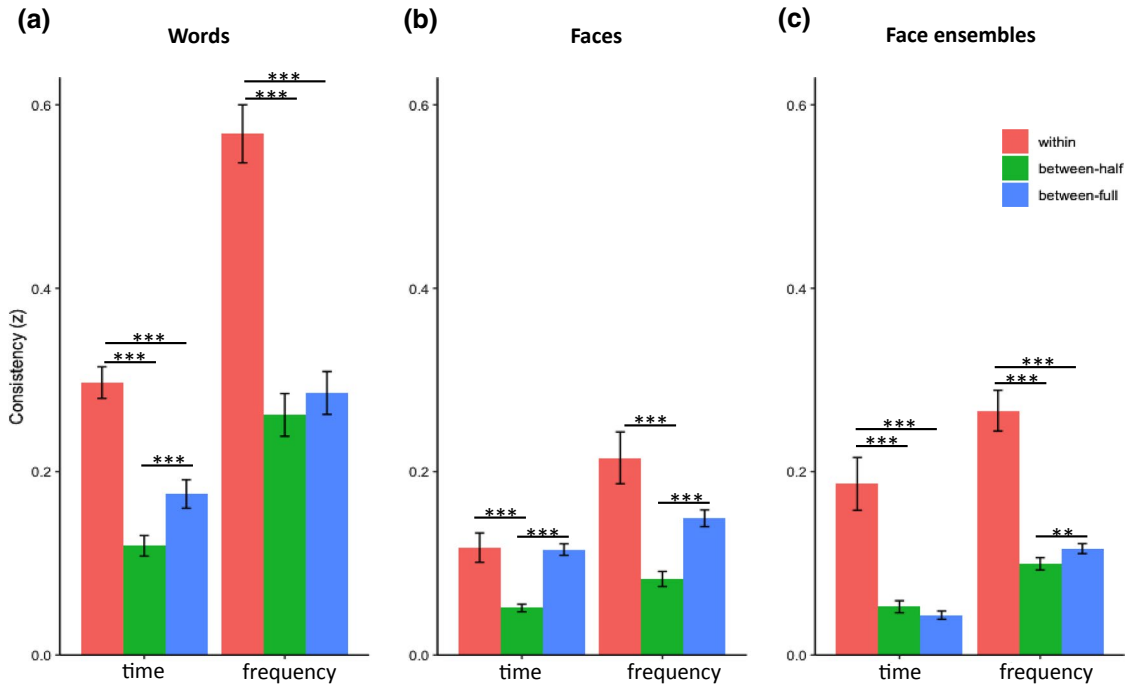

FIGURE 2 Ranking consistency within and across-participants for (a) words, (b) faces, and (c) face ensembles (**$p < .01$; ***$p < .001$; error bars show ±1*SE* across participants)

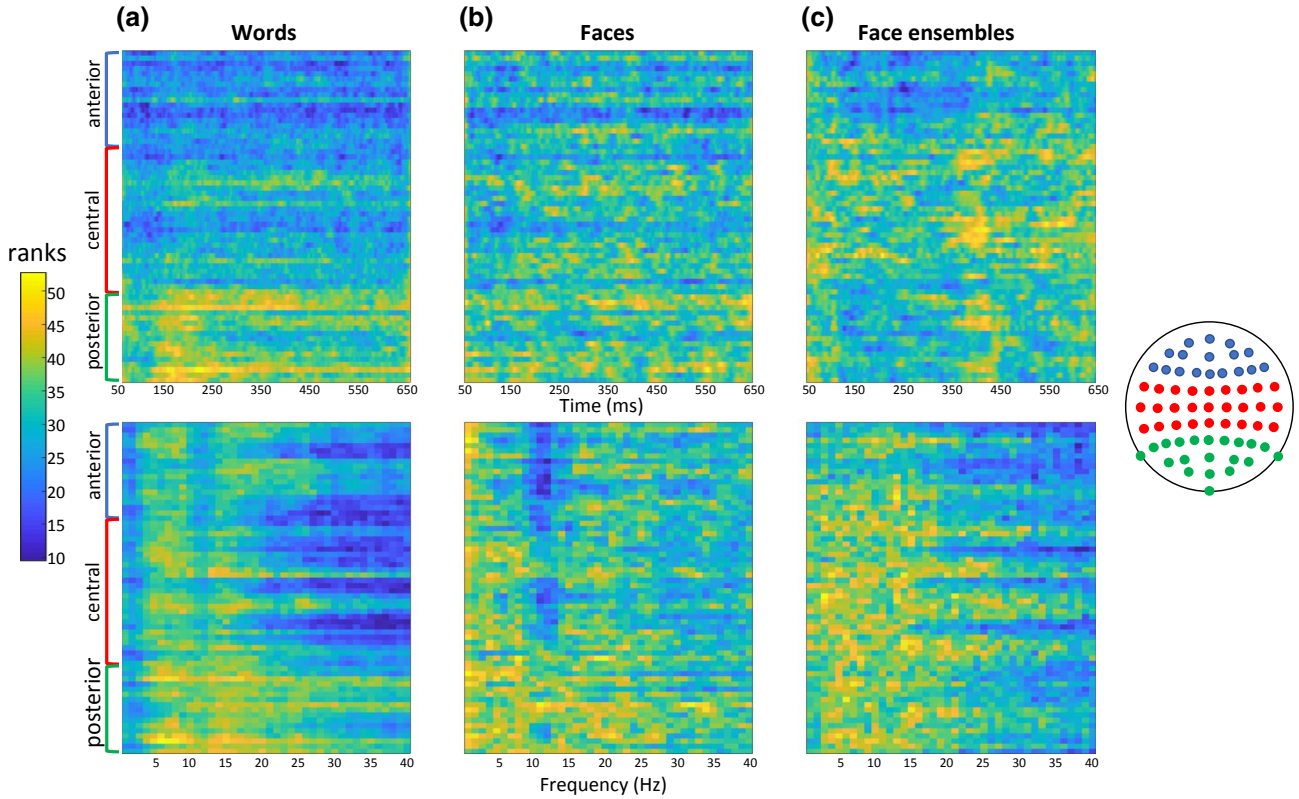

FIGURE 3 Average feature ranks across participants for (a) words, (b) faces, and (c) face ensembles in the time (top) and frequency domain (bottom)

temporal features across anterior electrodes for the purpose of word decoding. Also, rankings appear to differ across types of features. For instance, frequency features, as opposed to temporal features, do appear to provide some contribution to word decoding across anterior electrodes. Furthermore, rankings diverge across stimulus category. For instance, late temporal features across central channels appear to contribute more to ensemble decoding than to single face or word decoding.

To examine the spatial structure of diagnostic information more directly, feature ranks were averaged across time/frequency and the outcomes were displayed as topographical maps (Figure 4a). In the case of faces and words, decoding appeared to rely prominently on occipital–temporal channels while ensembles elicited preference for central and parieto-temporal channels.

To allow a more detailed and rigorous examination of the observation above, ranks were analyzed separately for stimulus category using three-way repeated-measures ANOVAs (electrode group: posterior, central, anterior × laterality: left, right × feature type: frequency and time) (see Supplementary Figure 1). These analyses revealed, in the case of words, a significant main effect of electrode group ($F(1.48, 19.24) = 53.34$, $p < .001$, $\eta_p^2 = .637$) and a significant interaction between electrode group and feature type ($F(2, 26) = 15.92$, $p < .001$, $\eta_p^2 = .184$). Post hoc comparisons showed that, for time, posterior features ranked higher than central and anterior ones ($p < .001$ for both), and central features ranked higher than anterior ones ($p = .007$). Similarly, for frequency, posterior features ranked higher than central and anterior ones ($p < .001$ for both), but central and anterior rankings were not significantly different (see Supplementary Figure 1a,d).

For faces, the analysis found a significant main effect of electrode group ($F(1.06, 12.66) = 17.66$, $p < .001$, $\eta_p^2 = .44$) and an interaction between group and laterality ($F(1.18, 14.12) = 7.00$, $p = .004$, $\eta_p^2 = .047$). Post hoc comparisons showed that the only significant laterality effect was found for posterior electrodes, where an advantage for the right hemisphere was observed ($p = .049$; see Supplementary Figure 1b,e).

For ensembles, the analysis found significant main effects of electrode group ($F(1.48, 19.24) = 16.8$, $p < .001$, $\eta_p^2 = .362$) and of laterality, with an advantage for right electrode features ($F(1, 13) = 5.78$, $p = .031$, $\eta_p^2 = .031$). Post hoc comparisons showed that posterior and central electrodes were ranked significantly higher than anterior ones ($p = .002$ and $p < .001$, respectively; see Supplementary Figure 1c,f).

These results demonstrate the differential contribution of features associated with specific channels as a function of position across the posterior–anterior axis, laterality, and feature type.

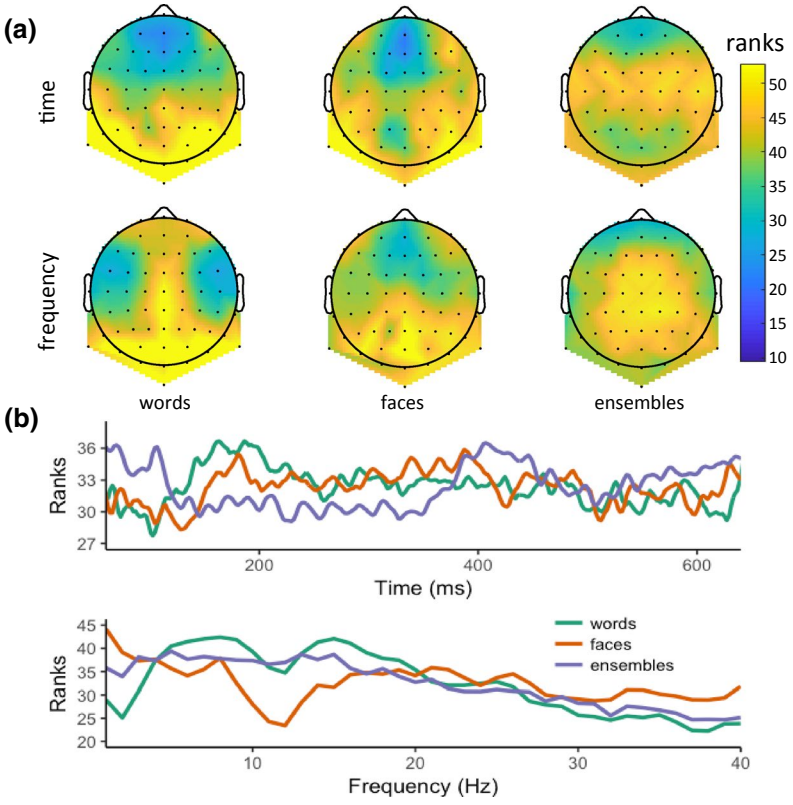

FIGURE 4 The structure of feature ranks: (a) spatial structure averaged across time and frequency along with (b) the temporal and frequency structure of feature ranks averaged across channels

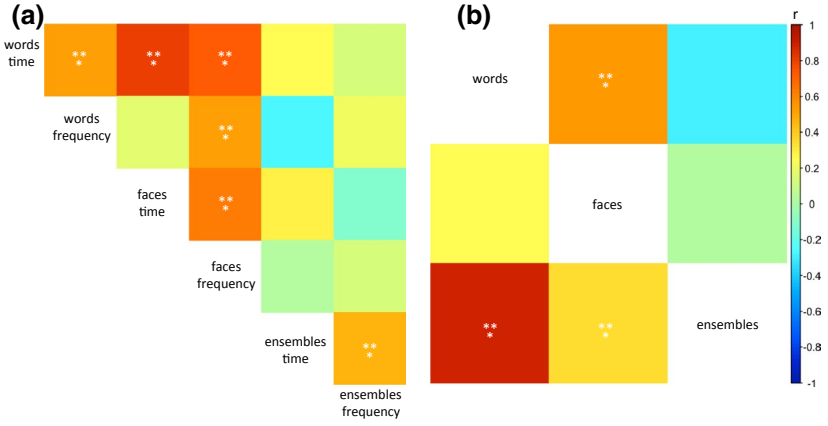

FIGURE 5 Correlations of average ranks across (a) channels, (b, top right) time, and (c, bottom left) frequency (***$p < .001$)

A complementary assessment was next conducted for time and frequency, after averaging across electrodes (Figure 4b). In the time domain, both faces and words show that ranking peaks roughly between 150 and 200 ms after stimulus onset, whereas ensemble ranking peaks later, roughly between 400 and 450 ms. In the frequency domain we note a gradual decrease in ranking after 20 Hz for all visual categories as well as a clear dip in ranking around 12 Hz, especially for faces and words. An additional examination across different frequency bands showed that both word and ensemble processing relied on theta (5–7 Hz) and alpha (8–15 Hz) bands, while face processing relied on delta (1–4 Hz) and theta bands—for a quantitative analysis of these observations see Supplementary Results and Supplementary Figure 2.

Interestingly, the results above point to distinct spatial, temporal, and frequency patterns of diagnosticity across visual domains. To quantify the dissimilarity of these patterns, feature ranks were compared to each other via Pearson correlation. First, regarding spatial information, feature ranks were averaged across time (or frequency) and were then related to each other (Figure 5a). An examination of the results points to similar spatial patterns across feature types for each stimulus domain separately. Specifically, temporal and frequency features yield similar rankings across electrodes for word decoding as well as for the other two visual domains (FDR corrected, $q < .05$). More importantly, faces and words yield similar rankings for both temporal and frequency features as well as across

feature types (i.e., words–time and faces–frequency). In contrast, ensembles did not seem to share the same spatial structure with the other two categories.

Complementary analyses were next conducted for temporal and frequency patterns, after averaging across electrodes. Consistent with the results above, faces and words were similar in the temporal domain (Figure 5b, top right) ($q < .05$). In contrast, in the frequency domain ensembles were similar to both words and faces (Figure 5b, bottom left).

Taken together, the present results point to common informational structure in the decoding of central stimuli, such as faces and words, but to divergent spatiotemporal structure in the decoding of face ensembles.

3.3 | Decoding performance

A specific aim of feature ranking and selection is to improve decoding performance. Accordingly, we examined the effectiveness of such improvements as an outcome of RFE-based feature selection. As SVM classification scales well with pattern dimensionality, the improvement in decoding accuracy is expected to be relatively small (Guyon et al., 2002). In contrast, the reduction in the number of features can be dramatic, given that the procedure, as applied here, aims to identify a minimal number of critical features during the validation stage.

An examination of validation curves (Supplementary Figure 3) shows results that agree with the expectation above. Overall, classification accuracy gradually increases as irrelevant or redundant features are eliminated and, then, it rapidly declines as critical features are eliminated. Further, classification peaks tend to occur relatively late in the feature selection process, consistent with a considerable reduction in pattern dimensionality. Yet, they provide systematically a performance boost compared to the starting point, which considers all available features.

To assess independent estimates of classification performance, decoding accuracy was considered based on the test data sets. First, performance was compared to chance both prior to and post feature selection (one-tailed one-sample t tests against 50%, Bonferroni corrected) (Figure 6a,b). This analysis revealed above-chance classification for all visual domains and types of features irrespective of feature selection ($p = .018$ for faces in the time domain when

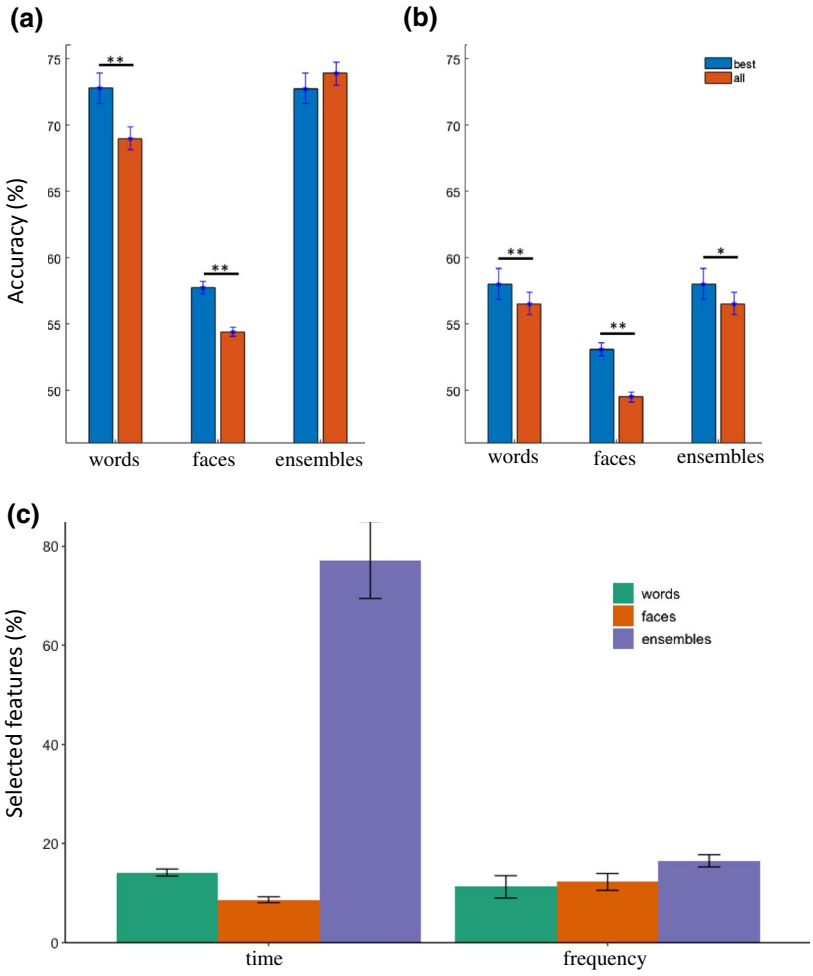

FIGURE 6 The impact of feature selection on (a, b) classification accuracy in the time and frequency domains (**$p < .01$, *$p < .10$; error bars show $\pm 1SE$ across participants) as well as on (c) feature reduction for each visual domain and type of feature (100% corresponds to 64 feature batches; error bars show ± 1 *SE* across participants)

using all features; all p's < .001 for words and ensembles) with the exception of face decoding in the frequency domain prior to feature selection (accuracy: 49.5%). Second, temporal-based decoding surpassed its frequency-based counterpart in every single case (one-tailed paired t tests, for ensembles prior to feature selection: $p = .008$; for faces prior to and after selection: $p = .002$; for all others: p's < .001, Bonferroni corrected). To be clear, this is not at odds with the higher consistency of frequency domain features relative to time domain features (Figure 2) given that our measure of consistency, feature ranking for classification purposes, is relatively independent of the amount of information regarding class separation as carried by different types of features. Third, the impact of feature selection was considered for each visual domain and type of feature (one-tailed paired t tests, Bonferroni corrected). This analysis indicated that feature selection boosted performance significantly for faces and words (p's < .001) and marginally for ensembles in the frequency domain ($p = .086$) but not in the time domain.

Next, we examined the dimensionality of optimal feature subsets. In most cases, dimensionality was reduced by more than 80% (Figure 6c) and these results were relatively consistent across participants (see Supplementary Figure 4). The only exception was provided by time-based ensemble classification—in this case we note that the validation curve (Supplementary Figure 3) shows a gradual decrease in performance over the course of feature elimination, rather than a gradual increase, and test performance does not indicate any decoding benefit relative to the starting point (Figure 6a). Thus, feature selection achieves its intended aims of boosting decoding accuracy and dimensionality reduction in a majority (but not all) of cases.

To be clear, the fact that a small proportion of features are critical for decoding purposes is not at odds with the finding that a considerable proportion of features carry diagnostic information (as shown in Figure 3). This divergence is due to the fact that RFE-based feature selection eliminates not only irrelevant features but also diagnostic features that carry redundant information relative to optimal feature subsets. In this sense, an examination of feature ranking, as carried out in Section 3.2, and the current evaluation of optimal feature subsets provide related but complementary perspectives regarding feature selection and its impact on decoding performance.

Last, we examined the possibility that temporal and frequency-based features provide complementary information for decoding purposes. However, an application of feature selection to joint sets of time and frequency features showed no benefit in decoding accuracy relative to time-based decoding for any visual category (one-tailed paired t-tests, all p's > .33). Thus, consistent with the systematic correlation of time and frequency feature ranks (Figure 5a), the two types of features appear to provide overlapping rather than complementary information.

4 | DISCUSSION

The present investigation targeted the informational structure supporting identity-level processing across three visual domains as revealed by pattern analyses of EEG data. This investigation evinced several notable outcomes, as follows.

To begin with, we found systematic variation of feature diagnosticity in the temporal and the frequency domains for every visual category. Importantly, such variation was robust both within and across participants despite the expected variability of the EEG signal (Thigpen, Kappenman, & Keil, 2017). This confirmed the presence of informative structure in our data, motivated our methodological approach and, thus, argued for the feasibility of our research objectives.

Notably, feature ranking revealed the reliance of different categories on common information. Specifically, single face and word decoding showed particular reliance, spatially, on bilateral occipitotemporal channels and, temporally, on a broad interval spanning around 150–450 ms after stimulus onset. In contrast, ensemble processing exhibited reliance on central channels and on an interval between 350 and 450 ms. Interestingly, single faces and ensembles, but not words, showed higher reliance on right compared with left channels. However, single faces and ensembles diverged in the frequency domain in which ensembles showed, similarly to words, higher reliance than faces on alpha band features.

The results above are consistent with intermingling but distinct neural resources underlying the processing of different categories (Plaut & Behrmann, 2011). For single faces and words, such intermingling has been previously suggested by fMRI (Mei et al., 2010; Nestor et al., 2013) and neuropsychological studies (Albonico & Barton, 2017; Gerlach, Marstrand, Starrfelt, & Gade, 2014; Sigurdardottir, Ívarsson, Kristinsdóttir, & Kristjánsson, 2015, but see Starrfelt, Klargaard, Petersen, & Gerlach, 2018). Moreover, prior EEG work has documented the reliance of these two categories on occipitotemporal channels around 170 ms as reflected, for instance, by the importance of the N170 ERP component for faces (Bentin, Allison, Puce, Perez, & McCarthy, 1996; Ince et al., 2016; Rossion, Joyce, Cottrell, & Tarr, 2003), words (Bentin, Mouchetant-Rostaing, Giard, Echallier, & Pernier, 1999; Maurer, Zevin, & McCandliss, 2008), and, more generally, for visual expertise (Tanaka & Curran, 2001). The similar spatiotemporal profile of diagnostic information, as revealed here by multivariate analyses, is broadly consistent with this research. Specifically, we find that feature diagnosticity for within-category discrimination ramps up around 170 ms for both faces and words especially at occipitotemporal channels.

Concurrently though, we find that faces and words diverge in significant ways. A right-side advantage is found for posterior channels for faces, consistent with the hemispheric asymmetry in face perception (Mercure, Dick, Halit, Kaufman, & Johnson, 2008; Rossion et al., 2003). The homologous left-side advantage, suggested by the importance of left hemispheric resources for language, was not found for words though. While the spatial resolution of EEG precludes definitive conclusions here, we note some convergence with the idea of complementary information in visual word recognition across the two hemispheres (Barton et al., 2010)—see also Laszlo and Federmeier (2014) regarding the sensitivity of word recognition dynamics to context and task. Further, we find differential reliance on frequency bands for the two categories, hinting at different modes of processing.

Regarding ensembles, while extensive behavioral work has documented our ability to extract summary representations from groups of faces (de Fockert & Wolfenstein, 2009; Haberman, Brady, & Alvarez, 2015; Neumann, Schweinberger, & Burton, 2013; Sama, Nestor, & Cant, 2019; Yamanashi Leib et al., 2014), little is known about their neural underpinnings. Interestingly, recent fMRI work (Im et al., 2017) found that face ensembles exhibit marked reliance on dorsal stream areas along with a right hemispheric advantage. Our results agree with these findings by showing an extension of feature diagnosticity beyond posterior to central channels along with a significant advantage for right channels. Further, diagnosticity does not ramp up around 170 ms, as for words and single faces, but instead peaks later within a 350–450 ms interval. This temporal profile presumably reflects a slower process of information accumulation underlying the derivation of a summary representation (Haberman, Harp, & Whitney, 2009; Roberts et al., 2019). Arguably, this demonstrates how face ensemble and single-face processing diverge systematically in their spatiotemporal dynamics in response to distinct perceptual demands.

Interestingly though, we find that words and ensembles are similar in one respect: their reliance on alpha band features relative to single faces. Alpha activity has previously been related to top-down processes recruited by attention (Klimesch, 2012; Ward, 2003) and semantic knowledge, as required, for instance, by famous face recognition (Zion-Golumbic, Kutas, & Bentin, 2010). Our results largely agree with this hypothesis. Ensemble perception involves a complex process of information integration dependent on specific attentional resources (Cohen, Dennett, & Kanwisher, 2016; Jackson-Nielsen, Cohen, & Pitts, 2017, but see Alvarez & Oliva, 2009), while word recognition relies on a language processing network that involves top-down phonological and semantic information (Price & Devlin, 2011). In contrast, unfamiliar face perception may reflect primarily automatic bottom-up processing (Dobs et al., 2019; Megreya & Burton, 2006; Natu & O'Toole, 2011) less prominent in the alpha band.

Next, methodologically, our investigation demonstrates the benefits of RFE-based feature selection. First, we find a systematic boost in decoding accuracy across stimulus categories and feature types. This boost is particularly notable for single-face frequency-based decoding, which is at chance level prior to feature selection, and underscores the ability of feature selection to facilitate the detection of smaller, more subtle effects. Second, we find that, in most cases, the procedure yields a drastic reduction in feature size of over 80%. Notably, these results are robust across participants demonstrating that smaller pools of features do not owe their success to capturing participant and category-specific idiosyncrasies of the EEG signal. An interesting exception is that of ensemble decoding in the temporal domain. Here, accuracy does not improve and only a minority of features are removed. This result suggests a widely distributed spatiotemporal structure and, likely, an extensive cortical network facilitating ensemble processing. Thus, the benefit of feature selection is likely to scale, reasonably, with the reliance more local, less dispersed neural resources.

Importantly, the similarity between face and word processing noted above has its own practical implications. Specifically, both categories appear to rely on a largely overlapping set of 12–16 channels robust in their decoding performance across participants. This suggests that the same electrode montage can capture visual processing for different categories and across different participants. Accordingly, this could facilitate the use of brain–computer interfaces (BCI) targeting visual recognition, by constraining the number of channels and, also, by enhancing their versatility (e.g., as related to visual recognition of multiple categories of interest).

A potential limitation our investigation is the use of different tasks across experiments for words and face stimuli. Indeed, task differences are known to modulate, for instance, the amplitude of ERP components in face processing (Goffaux, Jemel, Jacques, Rossion, & Schyns, 2003; Rousselet, Gaspar, Wieczorek, & Pernet, 2011). Our analyses target the multivariate EEG structure rather than the amplitude of ERP components. Yet, it is possible that task differences across faces and words also impact this structure and limit our ability to capture cross-category similarities. Despite this possibility, we note here the marked similarity in the spatiotemporal profile of single face and word decoding. However, future investigations will have to elucidate the precise impact of such tasks on EEG patterns.

Regarding the performance of different feature types, we note that temporal domain features surpassed systematically frequency domain features. However, we also found that frequency domain features performed above chance for every category, pointing to their potential utility in designs that involve more variable temporal structure, such as in memory and/or visual imagery tasks that preclude locking the EEG signal to stimulus onset.

Another point to consider here is that decoding did not appeal to more elaborate, neuroengineered features (e.g., estimates of signal complexity and coherence) (Jenke, Peer, & Buss, 2014; Lotte & Congedo, 2016) or to those derived by convolutional neural networks (Schirrmeister et al., 2017; Sturm, Lapuschkin, Samek, & Müller, 2016). Admittedly, the use of more elaborate features may further boost decoding accuracy and, also, close the gap between the performance of time- and frequency-based information.

Importantly, from a practical standpoint, feature rankings showed consistency in both within- and cross-participant analyses. As expected, cross-participant consistency was lower than its within-participant counterpart. This illustrates a main challenge for transfer learning endeavors to exploit commonalities in the data of different participants despite marked idiosyncratic differences (Jayaram et al., 2016; Yin et al., 2017). Importantly here, our investigation confirms significant similarity across participants in the domains of face and word perception. Hence, the present results could serve as a stepping stone for future efforts to optimize the extraction of cross-participant information and to maximize its utility in these domains (e.g., by modeling the space of individual variability).

In conclusion, our study identifies robust and systematic patterns of spatiotemporal and frequency-based information in the EEG signal. Notably, such patterns reflect visual processing across multiple categories that is robust across participants. These results serve to further clarify the neural dynamics of visual recognition and, importantly, to guide EEG methodology and design.

ORCID

Adrian Nestor https://orcid.org/0000-0002-2250-8759

REFERENCES

Albonico, A., & Barton, J. J. S. (2017). Face perception in pure alexia: Complementary contributions of the left fusiform gyrus to facial identity and facial speech processing. *Cortex*, *96*, 59–72. https://doi.org/10.1016/j.cortex.2017.08.029

Alotaiby, T., El-samie, F. E. A. A., Alshebeili, S. A., & Ahmad, I. (2015). A review of channel selection algorithms for EEG signal processing. *EURASIP Journal on Advances in Signal Processing*, *1*, 66. https://doi.org/10.1186/s13634-015-0251-9

Alvarez, G. A., & Oliva, A. (2009). Spatial ensemble statistics are efficient codes that can be represented with reduced attention. *Proceedings of the National Academy of Sciences*, *106*(18), 7345–7350. https://doi.org/10.1073/pnas.0808981106

Ambrus, G. G., Kaiser, D., Cichy, R. M., & Kovács, G. (2019). The neural dynamics of familiar face recognition. *Cerebral Cortex*. Advance Online Publication. https://doi.org/10.1093/cercor/bhz010

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., … Halgren, E. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(2), 449–454. https://doi.org/10.1073/pnas.0507062103

Barton, J. J. S., Sekunova, A., Sheldon, C., Johnston, S., Iaria, G., & Scheel, M. (2010). Reading words, seeing style: The neuropsychology of word, font and handwriting perception. *Neuropsychologia*, *48*(13), 3868–3877. https://doi.org/10.1167/10.7.1012

Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological studies of face perception in humans. *Journal of Cognitive Neuroscience*, *8*(6), 551–565. https://doi.org/10.1162/jocn.1996.8.6.551

Bentin, S., Mouchetant-Rostaing, Y., Giard, M. H., Echallier, J. F., & Pernier, J. (1999). ERP manifestations of processing printed words at different psycholinguistic levels: Time course and scalp distribution. *Journal of Cognitive Neuroscience*, *11*(3), 235–260. https://doi.org/10.1162/089892999563373

Chan, A. M., Halgren, E., Marinkovic, K., & Cash, S. S. (2011). Decoding word and category-specific spatiotemporal representations from MEG and EEG. *NeuroImage*, *54*(4), 3028–3039. https://doi.org/10.1016/j.neuroimage.2010.10.073

Cichy, R. M., Pantazis, D., & Oliva, A. (2014). Resolving human object recognition in space and time. *Nature Neuroscience*, *17*(3), 455–462. https://doi.org/10.1038/nn.3635

Cohen, M. A., Dennett, D. C., & Kanwisher, N. (2016). Ensemble perception, summary statistics, and perceptual awareness: A response. *Trends in Cognitive Sciences*, *20*(9), 643–644. https://doi.org/10.1016/j.tics.2016.06.007

de Fockert, J., & Wolfenstein, C. (2009). Rapid extraction of mean identity from sets of faces. *The Quarterly Journal of Experimental Psychology*, *62*, 1716–1722. https://doi.org/10.1080/17470210902811249

De Martino, F., Valente, G., Staeren, N., Ashburner, J., Goebel, R., & Formisano, E. (2008). Combining multivariate voxel selection and support vector machines for mapping and classification of fMRI spatial patterns. *NeuroImage*, *43*(1), 44–58. https://doi.org/10.1016/j.neuroimage.2008.06.037

Dehaene, S., Pegado, F., Braga, L. W., Ventura, P., Filho, G. N., Jobert, A., … Cohen, L. (2010). How learning to read changes the cortical networks for vision and language. *Science*, *330*(6009), 1359–1364. https://doi.org/10.1126/science.1194140

Delorme, A., Sejnowski, T., & Makeig, S. (2007). Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. *NeuroImage*, *34*(4), 1443–1449. https://doi.org/10.1016/j.neuroimage.2006.11.004

Dobs, K., Isik, L., Pantazis, D., & Kanwisher, N. (2019). How face perception unfolds over time. *Nature Communications*, *10*(1), 1258. https://doi.org/10.1038/s41467-019-09239-1

Gerlach, C., Marstrand, L., Starrfelt, R., & Gade, A. (2014). No strong evidence for lateralisation of word reading and face recognition deficits following posterior brain injury. *Journal of Cognitive Psychology*, *26*(5), 550–558. https://doi.org/10.1080/20445911.2014.928713

Goffaux, V., Jemel, B., Jacques, C., Rossion, B., & Schyns, P. G. (2003). ERP evidence for task modulations on face perceptual processing at different spatial scales. *Cognitive Science*, *27*(2), 313–325. https://doi.org/10.1016/S0364-0213(03)00002-8

Guyon, I., Weston, J., Barnhill, S., & Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. *Machine Learning*, *46*(1/3), 389–422. https://doi.org/10.1023/A:1012487302797

Gysels, E., Renevey, P., & Celka, P. (2005). SVM-based recursive feature elimination to compare phase synchronization computed from broadband and narrowband EEG signals in Brain-Computer Interfaces. *Signal Processing*, *85*(11), 2178–2189. https://doi.org/10.1016/j.sigpro.2005.07.008

Haberman, J., Brady, T. F., & Alvarez, G. A. (2015). Individual differences in ensemble perception reveal multiple, independent levels of ensemble representation. *Journal of Experimental Psychology: General*, *144*(2), 432. https://doi.org/10.1037/xge0000053

Haberman, J., Harp, T., & Whitney, D. (2009). Averaging facial expression over time. *Journal of Vision*, *9*(11), 1. https://doi.org/10.1167/9.11.1

Hanson, S. J., & Halchenko, Y. O. (2008). Brain reading using full brain support vector machines for object recognition: There is no "face" identification area. *Neural Computation*, *20*(2), 486–503. https://doi.org/10.1162/neco.2007.09-06-340

Harris, R. J., Rice, G. E., Young, A. W., & Andrews, T. J. (2016). Distinct but overlapping patterns of response to words and faces in the fusiform gyrus. *Cerebral Cortex*, *26*(7), 3161–3168. https://doi.org/10.1093/cercor/bhv147

Hasson, U., Levy, I., Behrmann, M., Hendler, T., & Malach, R. (2002). Eccentricity bias as an organizing principle for human high-order object areas. *Neuron*, *34*(3), 479–490. https://doi.org/10.1016/S0896-6273(02)00662-1

Hidalgo-Muñoz, A. R., López, M. M., Santos, I. M., Pereira, A. T., Vázquez-Marrufo, M., Galvao-Carmona, A., & Tomé, A. M. (2013). Application of SVM-RFE on EEG signals for detecting the most relevant scalp regions linked to affective valence processing. *Expert Systems with Applications*, *40*(6), 2102–2108. https://doi.org/10.1016/j.eswa.2012.10.013

Im, H., Albohn, D. N., Steiner, T. G., Cushing, C. A., Adams Jr., R. B., & Kveraga, K. (2017). Differential hemispheric and visual stream contributions to ensemble coding of crowd emotion. *Nature Human Behaviour*, *1*, 828–842. https://doi.org/10.1038/s41562-017-0225-z

Ince, R. A. A., Jaworska, K., Gross, J., Panzeri, S., van Rijsbergen, N. J., Rousselet, G. A., & Schyns, P. G. (2016). The deceptively simple N170 reflects network information processing mechanisms involving visual feature coding and transfer across hemispheres. *Cerebral Cortex*, *26*(11), 4123–4135. https://doi.org/10.1093/cercor/bhw196

Isik, L., Meyers, E. M., Leibo, J. Z., Poggio, T. A., Baldauf, D., & Desimone, R. (2014). The dynamics of invariant object recognition in the human visual system. *Journal of Neurophysiology*, *111*(1), 91–102. https://doi.org/10.1152/jn.00394.2013

Jackson-Nielsen, M., Cohen, M. A., & Pitts, M. A. (2017). Perception of ensemble statistics requires attention. *Consciousness and Cognition*, *48*, 149–160. https://doi.org/10.1016/j.concog.2016.11.007

Jayaram, V., Alamgir, M., Altun, Y., Scholkopf, B., & Grosse-Wentrup, M. (2016). Transfer learning in brain-computer interfaces. *IEEE Computational Intelligence Magazine*, *11*(1), 20–31. https://doi.org/10.1109/MCI.2015.2501545

Jenke, R., Peer, A., & Buss, M. (2014). Feature extraction and selection for emotion recognition from EEG. *IEEE Transactions on Affective Computing*, *5*(3), 327–339. https://doi.org/10.1109/TAFFC.2014.2339834

King, J.-R., & Dehaene, S. (2014). Characterizing the dynamics of mental representations: The temporal generalization method. *Trends in Cognitive Sciences*, *18*(4), 203–210. https://doi.org/10.1016/j.tics.2014.01.002

Klimesch, W. (2012). Alpha-band oscillations, attention, and controlled access to stored information. *Trends in Cognitive Sciences*, *16*(12), 606–617. https://doi.org/10.1016/j.tics.2012.10.007

Laszlo, S., & Federmeier, K. D. (2014). Never seem to find the time: Evaluating the physiological time course of visual word recognition with regression analysis of single-item event-related potentials. *Language, Cognition and Neuroscience*, *29*(5), 642–661. https://doi.org/10.1080/01690965.2013.866259

Ling, S., Lee, A. C. H., Armstrong, B. C., & Nestor, A. (2019). EEG-based visual word decoding, feature derivation and image reconstruction. *Human Brain Mapping*, *40*(17), 5056–5068. https://doi.org/10.1002/hbm.24757

Lotte, F., & Congedo, M. (2016). EEG feature extraction. In M. Clerc, L. Bougrain & F. Lotte (Eds.), *Brain-computer interfaces 1: Foundations and methods* (pp. 127–143). Wiley Online Library. https://doi.org/10.1002/9781119144977.ch7

Maurer, U., Zevin, J. D., & McCandliss, B. D. (2008). Left-lateralized N170 effects of visual expertise in reading: Evidence from Japanese syllabic and logographic scripts. *Journal of Cognitive Neuroscience*, *20*(10), 1878–1891. https://doi.org/10.1162/jocn.2008.20125

Megreya, A. M., & Burton, A. M. (2006). Unfamiliar faces are not faces: Evidence from a matching task. *Memory & Cognition*, *34*(4), 865–876. https://doi.org/10.3758/BF03193433

Mei, L., Xue, G., Chen, C., Xue, F., Zhang, M., & Dong, Q. (2010). The "visual word form area" is involved in successful memory encoding of both words and faces. *NeuroImage*, *52*(1), 371–378. https://doi.org/10.1016/j.neuroimage.2010.03.067

Mercure, E., Dick, F., Halit, H., Kaufman, J., & Johnson, M. H. (2008). Differential lateralization for words and faces: Category or psychophysics? *Journal of Cognitive Neuroscience*, *20*(11), 2070–2087. https://doi.org/10.1162/jocn.2008.20137

Mouraux, A., & Iannetti, G. D. (2008). Across-trial averaging of event-related EEG responses and beyond. *Magnetic Resonance Imaging*, *26*(7), 1041–1054. https://doi.org/10.1016/j.mri.2008.01.011

Natu, V. S., & O'Toole, A. J. (2011). The neural processing of familiar and unfamiliar faces: A review and synopsis. *British Journal of Psychology*, *102*(4), 726–747. https://doi.org/10.1111/j.2044-8295.2011.02053.x

Nemrodov, D., Niemeier, M., Mok, J. N. Y., & Nestor, A. (2016). The time course of individual face recognition: A pattern analysis of ERP signals. *NeuroImage*, *132*, 469–476. https://doi.org/10.1016/j.neuroimage.2016.03.006

Nemrodov, D., Niemeier, M., Patel, A., & Nestor, A. (2018). The neural dynamics of facial identity processing: Insights from EEG-based pattern analysis and image reconstruction. *Eneuro*, ENEURO.0358-17.2018. https://doi.org/10.1523/ENEURO.0358-17.2018

Nestor, A., Behrmann, M., & Plaut, D. C. D. C. (2013). The neural basis of visual word form processing: A multivariate investigation. *Cerebral Cortex*, *23*(7), 1673–1684. https://doi.org/10.1093/cercor/bhs158

Neumann, M. F., Schweinberger, S. R., & Burton, A. M. (2013). Viewers extract mean and individual identity from sets of famous faces. *Cognition*, *128*(1), 56–63. https://doi.org/10.1016/j.cognition.2013.03.006

Plaut, D. C., & Behrmann, M. (2011). Complementary neural representations for faces and words: A computational exploration. *Cognitive Neuropsychology*, *28*(3–4), 251–275. https://doi.org/10.1080/02643294.2011.609812

Price, C. J., & Devlin, J. T. (2011). The Interactive Account of ventral occipitotemporal contributions to reading. *Trends in Cognitive Sciences*, *15*(6), 246–253. https://doi.org/10.1016/j.tics.2011.04.001

Retter, T. L., & Rossion, B. (2016). Uncovering the neural magnitude and spatio-temporal dynamics of natural image categorization in a fast visual stream. *Neuropsychologia*, *91*, 9–28. https://doi.org/10.1016/j.neuropsychologia.2016.07.028

Roberts, T., Cant, J. S., & Nestor, A. (2019). Elucidating the neural representation and the temporal dynamics of face ensembles. *Journal of Neuroscience*, *39*(39), 7737–7747. https://doi.org/10.1523/JNEUROSCI.0471-19.2019

Rossion, B., Joyce, C. A., Cottrell, G. W., & Tarr, M. J. (2003). Early lateralization and orientation tuning for face, word, and object processing in the visual cortex. *NeuroImage*, *20*(3), 1609–1624. https://doi.org/10.1016/j.neuroimage.2003.07.010

Rousselet, G. A., Gaspar, C. M., Wieczorek, K. P., & Pernet, C. R. (2011). Modelling single-trial ERP reveals modulation of bottom-up face visual processing by top-down task constraints (in some subjects). *Frontiers in Psychology*, *2*, 137. https://doi.org/10.3389/fpsyg.2011.00137

Sama, M. A., Nestor, A., & Cant, J. S. (2019). Independence of viewpoint and identity in face ensemble processing. *Journal of Vision*, *19*(5): 2, 1–17. https://doi.org/10.1167/19.5.2

Schirrmeister, R. T., Springenberg, J. T., Fiederer, L. D. J., Glasstetter, M., Eggensperger, K., Tangermann, M., … Ball, T. (2017). Deep learning with convolutional neural networks for EEG decoding and visualization. *Human Brain Mapping*, *38*(11), 5391–5420. https://doi.org/10.1002/hbm.23730

Shen, K. Q., Ong, C. J., Li, X. P., Hui, Z., & Wilder-Smith, E. P. V. (2007). A feature selection method for multilevel mental fatigue EEG classification. *IEEE Transactions on Biomedical Engineering*, *54*(7), 1231–1237. https://doi.org/10.1109/TBME.2007.890733

Sigurdardottir, H. M. H. M., Ívarsson, E., Kristinsdóttir, K., & Kristjánsson, Á. (2015). Impaired recognition of faces and objects in dyslexia: Evidence for ventral stream dysfunction? *Neuropsychology*, *29*(5), 739–750. https://doi.org/10.1037/neu0000188

Starrfelt, R., Klargaard, S. K., Petersen, A., & Gerlach, C. (2018). Reading in developmental prosopagnosia: Evidence for a dissociation between word and face recognition. *Neuropsychology*, *32*(2), 138–147. https://doi.org/10.1037/neu0000428

Sturm, I., Lapuschkin, S., Samek, W., & Müller, K. R. (2016). Interpretable deep neural networks for single-trial EEG classification. *Journal of Neuroscience Methods*, *274*, 141–145. https://doi.org/10.1016/j.jneumeth.2016.10.008

Tanaka, J. W., & Curran, T. (2001). A neural basis for expert object recognition. *Psychological Science*, *12*(1), 43–47. https://doi.org/10.1111/1467-9280.00308

Thigpen, N. N., Kappenman, E. S., & Keil, A. (2017). Assessing the internal consistency of the event-related potential: An example analysis. *Psychophysiology*, *54*(1), 123–138. https://doi.org/10.1111/psyp.12629

Vida, M. D., Nestor, A., Plaut, D. C., & Behrmann, M. (2017). Spatiotemporal dynamics of similarity-based neural representations of facial identity. *Proceedings of the National Academy of Sciences of the United States of America*, *114*, 388–393. https://doi.org/10.1073/pnas.1614763114

Ward, L. M. (2003). Synchronous neural oscillations and cognitive processes. *Trends in Cognitive Sciences*, *7*(12), 553–559. https://doi.org/10.1016/j.tics.2003.10.012

Whitney, D., & Yamanashi Leib, A. (2018). Ensemble perception. *Annual Review of Psychology*, *69*(1), 105–129. https://doi.org/10.1146/annurev-psych-010416-044232

Yamanashi Leib, A., Fischer, J., Liu, Y., Qiu, S., Robertson, L., & Whitney, D. (2014). Ensemble crowd perception: A viewpoint-invariant mechanism to represent average crowd identity. *Journal of Vision*, *13*(9), 424–424. https://doi.org/10.1167/14.8.26

Yin, Z., Wang, Y., Liu, L., Zhang, W., & Zhang, J. (2017). Cross-subject EEG feature selection for emotion recognition using transfer recursive feature elimination. *Frontiers in Neurorobotics*, *11*(APR), 1–16. https://doi.org/10.3389/fnbot.2017.00019

Zion-Golumbic, E., Kutas, M., & Bentin, S. (2010). Neural dynamics associated with semantic and episodic memory for faces: Evidence from multiple frequency bands. *Journal of Cognitive Neuroscience*, *22*(2), 263–277. https://doi.org/10.1162/jocn.2009.21251

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Nemrodov D, Ling S, Nudnou I, et al. A multivariate investigation of visual word, face, and ensemble processing: Perspectives from EEG-based decoding and feature selection. *Psychophysiology*. 2020;57:e13511. https://doi.org/10.1111/psyp.13511