Research Articles: Behavioral/Cognitive

The time course of language production as revealed by pattern classification of MEG sensor data

https://doi.org/10.1523/JNEUROSCI.1923-21.2022

Cite as: J. Neurosci 2022; 10.1523/JNEUROSCI.1923-21.2022
Received: 23 September 2021
Revised: 5 April 2022
Accepted: 12 April 2022

This Early Release article has been peer-reviewed and accepted, but has not been through the composition and copyediting processes. The final version may differ slightly in style or formatting and will contain links to any extended data.

Alerts: Sign up at www.jneurosci.org/alerts to receive customized email alerts when the fully formatted version of this article is published.
The time course of language production as revealed by pattern classification of MEG sensor data

Francesca Carota 1-2, Jan-Mathijs Schoffelen1-2, Robert Oostenveld2,4, Peter Indefrey1-2-3

1 Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
2 Donders Institute for Cognitive Neuroscience, Radboud University, Nijmegen, The Netherlands
3 Institut für Sprache und Information at Heinrich Heine University, Düsseldorf, Germany
4 NatMEG, Karolinska Institutet, Stockholm, Sweden

Corresponding author: Francesca Carota; francesca.carota@mpi.nl

The authors would like to thank Jessica Askamp and Kristina Baumgart for help with lab management during data collection; Peter Hagoort, David Kemmerer, Vitoria Piai, Eelke Spak, Marina Papoutsi, Pinie Zwitserlood, and two anonymous reviewers for their comments at different stages of our study.

This work was funded by a Research Fellowship of the Max Planck Gesellschaft to Peter Indefrey.
Abstract

Language production involves a complex set of computations, from conceptualisation to articulation, which are thought to engage cascading neural events in the language network. However, recent neuromagnetic evidence suggests simultaneous meaning-to-speech mapping in picture naming tasks, as indexed by early parallel activation of fronto-temporal regions to lexical semantic, phonological and articulatory information. Here we investigate the time course of word production, asking to what extent such “earliness” is a distinctive property of the associated spatiotemporal dynamics. Using magnetoencephalography (MEG), we recorded the neural signals of 34 human subjects (26 males) overtly naming 134 images from 4 semantic object categories (animals, foods, tools, clothes). Within each category, we co-varied word length, as quantified by the number of syllables contained in a word, and phonological neighbourhood density to target lexical and post-lexical phonological/phonetic processes. Multivariate pattern analyses (MVPA) searchlights in sensor space distinguished the stimulus-locked spatio-temporal responses to object categories early on, from 150-250 ms post picture onset, whereas word length was decoded in left frontotemporal sensors at 250-350 ms, followed by the latency of phonological neighbourhood density (350-450 ms). Our results suggest a progression of neural activity from posterior to anterior language regions for the semantic and phonological/phonetic computations preparing overt speech, thus supporting serial cascading models of word production.

(209 words)
Significance Statement

Current psycholinguistic models make divergent predictions on how a preverbal message is mapped onto articulatory output during the language planning. Serial models predict a cascading sequence of hierarchically organised neural computations from conceptualisation to articulation. In contrast, parallel models posit early simultaneous activation of multiple conceptual, phonological, and articulatory information in the language system. Here we asked whether such earliness is a distinctive property of the neural dynamics of word production. The combination of the millisecond precision of magnetoencephalography with multivariate pattern analyses revealed subsequent onset times for the neural events supporting semantic and phonological/phonetic operations, progressing from posterior occipito-temporal to frontal sensor areas. The findings bring new insights for refining current theories of language production.

Introduction

The brief time interval preceding word production involves complex planning processes to transform a preverbal message into speech. How an intended meaning is rapidly mapped onto articulatory programs is, however, still disputed.

Two competing views of language production emerged from earlier studies. One account posits a hierarchically organised sequence of psycholinguistic operations with distinct latencies (Levelt, 1989; Levelt et al., 1999; Levelt, 1999; Hagoort and Levelt, 2009). As estimated by relating...
chronometric data to metabolic studies of word production (Indefrey and Levelt, 2000; 2004), a
target concept (e.g., APPLE) is accessed along with related concepts like PEAR, PEACH, FRUIT
within the first 200ms post picture onset in overt naming tasks. Once lexical information is
selected, phonological codes are retrieved between 200-400ms post picture onset in posterior
middle/superior temporal cortex, and abstract segmental representations required for syllable
planning (syllabification) are accessed around 400-600ms, involving left inferior frontal activity.
The target articulatory programs are retrieved about 100ms later, as reflected in premotor activity.
Importantly, the abovementioned onset times reflect approximations only, and are influenced by
characteristics of the to-be-pronounced words, such as word length, lexical frequency and
familiarity (Hanulová et al., 2011; Indefrey, 2011), motivating a rescaling of the latencies on a
word-by-word basis (Roelofs and Shitova, 2016).

Neurophysiological evidence (e.g., Salmelin et al, 1994; Levelt et al., 1998; Maess et al., 2002;
Sörös et al. 2003; Vihla et al., 2004; Hultén et al., 2009; Liljeström et al., 2009; Laaksonen et al.,
2012; for review: Munding et al., 2015), and electrophysiological studies testing for reliability of
language production tasks (e.g., Laganaro et al., 2017; Roos and Piai, 2020; Ala-Salomäki et al.,
2021), confirmed the proposed cascade of functionally and temporally dissociable neural events.

In contrast, recent MEG results from picture naming suggested early (130-160 ms post picture
presentation), near-simultaneous occipital/temporal activations to visual complexity, and left
inferior frontal and temporal activations to semantics and word form, respectively (Miozzo et al.,
2014). Another MEG study found early modulation of frontotemporal activity by word frequency
160-240 ms post picture presentation (Strijkers et al., 2017), and concurrent superior temporal
activity reflecting differences in the phonetic/articulatory properties of a word’s initial speech segment (either a labial or a coronal phoneme, like in monkey vs. donkey, respectively).

Furthermore, the place of articulation of the words’ initial phoneme activated the motor regions controlling for lip and tongue movements, suggesting early articulatory planning.

In the light of these premises, here we asked to what extent the earliness of simultaneous brain responses to multiple psycholinguistic functions is an intrinsically distinctive property of the neural dynamics of word production.

To determine the spatiotemporal specificity of semantic and phonological planning, we combined the temporal precision of magnetoencephalography (MEG) with multivariate pattern analyses (MVPA) in sensor space. Unlike univariate approaches, MVPA employs sensitive multivariate statistics to measure the covariation of signals across multiple neighbouring channels at the single trial level, and thus extracts information from distributed patterns rather than averaged amplitudes across selected channels. MVPA of MEG data also captures rich spatial information from the subtle multivariate magnetic field patterns of brain activity at sensor level, which result from small differences in orientation and angle of neighbouring dipoles, and may overlap, but still be statistically separable (Cichy and Pantazis, 2017; Stokes et al., 2015). This allows for the differentiation of the subject-specific spatial topographies over subjects (Stokes et al., 2015).

We adopted an orthogonal design co-varying word length and phonological neighbourhood density, as targeting syllabification and phonetic encoding, across four semantic object categories (animals, foods, tools, and clothes) to test:
1) the serial/hierarchical prediction of early neural indexes of conceptual preparation, but later frontal activity of syllabification (350ms) and phonetic encoding (450ms post picture onset).

2) the early/parallel prediction of early simultaneous activations to both conceptual and phonological/phonetic processes.

The present investigation aimed at probing the validity of these proposals to inform and refine current theories of language production. (646 words)

**Methods**

*Subjects:* 34 native Dutch speakers (26 male, mean age = 24 years, *sd* = 3.6) participated in the experiment after providing written informed consent. All subjects were right-handed, had normal or corrected-to-normal vision, and reported no history of neurological, developmental, or language deficits. The study was approved by the ethical board CMO Arnhem/Nijmegen, under registration number CMO2014/288.

*Materials:* Stimuli consisted of 134 images from 4 object categories including animals, foods, tools, and clothes. We used coloured realistic images from the picture database of Bank of Standardized Stimuli (BOSS) (Brodeur et al., 2014), and public domain images from the internet (e.g., Freepng.ru).
Images were selected on the basis of a list of depictable target words that are most commonly used to name the corresponding objects in Dutch. The list of object words was generated by covarying the length of target words and phonological neighbourhood density within each semantic category. Word length affects lexical (word form retrieval) and postlexical (syllabification, phonetic and articulatory encoding) processing stages in a straightforward quantitative way (Indefrey & Levelt, 2004; Indefrey, 2011; Papoutsi et al., 2009), that is, the more phonemes/syllables there are, the higher the load on the processes and the longer they take. Word form retrieval, for example, takes some 20-25 ms per phoneme (Indefrey & Levelt, 2004; Indefrey, 2011). We included phonological neighbourhood density, because it affects lexical as well as postlexical phonological processes but via different mechanisms (Harley and Brown 1998; Vitevitch 2002; Dell and Gordon, 2003): whereas the facilitative effect of phonological neighbours on word form activation seems to be mediated by their shared phonological segments (Vitevitch, 2002), the facilitative effect of phonological neighbourhood density at phonetic/articulatory processing stages seems to be due to its correlation with phonotactic probability (Vitevitch et al., 1999). Words with many neighbours tend to contain more common phonemes and phoneme sequences than words with fewer neighbours, making it easier to phonetically encode and articulate them. The effects of phonological neighbourhood density on lexical retrieval and articulation have been shown to be independent and additive (Buz & Jaeger, 2016). A previous MEG study by Miozzo and colleagues (2014) has combined word length with phonological neighborhood density into a single predictor of neural activity related to word form access. For our investigation of the relative time courses of processes related to early stages of word production (conceptual preparation) and processes related to later phonological and phonetic stages, we, therefore, found it important to ensure comparability with the Miozzo et al., (2014) study by employing the same variables to target phonological
processing. At the same time, in view of the different nature of the effects of word length and phonological neighbourhood density mentioned above, we reasoned that combining the two variables in one component might underestimate their specific effects, in particular the effect of phonotactic probability captured by phonological neighbourhood density but not by word length. We, therefore, decided to examine the two variables separately.

Word length was expressed by the number of syllables (65 short monosyllabic words, 66 long bisyllabic words, and 3 trisyllabic words). Phonological neighbourhood density was expressed by the number of words that differ in one phone from the target word. This was calculated by counting all word entries in CELEX that differ in one phone symbol from the target word, after discarding stress and syllable markers from the phonological word representation in CELEX. Words were ranked according to such difference and subdivided into four groups (lower, low, high, higher) while keeping the group size as balanced as possible: 39 lower (0-4 phones), 29 low (5-9 phones), 35 high (10-19 phones), and 31 higher (20-39 phones). The psycholinguistic properties of the 134 stimulus items are summarized in Table 1. Mean and standard deviation (SD), median and range (R, the difference between largest and smallest value of the variable) are reported, as calculated across items (1.A), and for Mono- and disyllabic words separately (1.B).

Table 1

<table>
<thead>
<tr>
<th>A.</th>
<th>All Words</th>
</tr>
</thead>
</table>

8
<table>
<thead>
<tr>
<th>Category</th>
<th>Subtlex word frequency</th>
<th>Nr. of syllables</th>
<th>Nr. of phon. neighbours</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (SD)</td>
<td>Median (R)</td>
<td>Mean (SD)</td>
</tr>
<tr>
<td>animals</td>
<td>8 (10)</td>
<td>6.1 (49.9)</td>
<td>1.5 (0.6)</td>
</tr>
<tr>
<td>clothes</td>
<td>9.8 (15.8)</td>
<td>4.2 (67.2)</td>
<td>1.5 (0.5)</td>
</tr>
<tr>
<td>foods</td>
<td>5.7 (9.2)</td>
<td>2 (39.6)</td>
<td>1.5 (0.5)</td>
</tr>
<tr>
<td>tools</td>
<td>9.4 (8.8)</td>
<td>6.8 (37.5)</td>
<td>1.6 (0.6)</td>
</tr>
<tr>
<td>all</td>
<td>8.3 (11.4)</td>
<td>4.8 (67.3)</td>
<td>1.5 (0.5)</td>
</tr>
</tbody>
</table>

### B.

#### Monosyllabic words

<table>
<thead>
<tr>
<th>Category</th>
<th>Subtlex word frequency</th>
<th>Nr. of phon. neighbours</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (SD)</td>
<td>Median (R)</td>
</tr>
<tr>
<td>animals</td>
<td>11.1 (12.8)</td>
<td>7.5 (49.4)</td>
</tr>
<tr>
<td>clothes</td>
<td>17.6 (19.6)</td>
<td>11.6 (65.3)</td>
</tr>
<tr>
<td>foods</td>
<td>8.7 (12.3)</td>
<td>2.2 (39.5)</td>
</tr>
<tr>
<td>tools</td>
<td>13.7 (10.5)</td>
<td>10.8 (35.7)</td>
</tr>
<tr>
<td>all</td>
<td>12.8 (14.4)</td>
<td>7.8 (67.1)</td>
</tr>
</tbody>
</table>

#### Bisyllabic words

<table>
<thead>
<tr>
<th>Category</th>
<th>Subtlex word frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (SD)</td>
</tr>
<tr>
<td>animals</td>
<td>5.4 (12.8)</td>
</tr>
<tr>
<td>clothes</td>
<td>2.5 (19.6)</td>
</tr>
<tr>
<td>foods</td>
<td>2.9 (12.3)</td>
</tr>
<tr>
<td>tools</td>
<td>5.1 (10.5)</td>
</tr>
<tr>
<td>all</td>
<td>3.9 (14.4)</td>
</tr>
</tbody>
</table>

|          | Mean (SD) | Median (R) | Mean (SD) | Median (R) |
| animals  | 6.1 (5.5) | 4 (21) |
| clothes  | 5.8 (5.3) | 3.5 (15) |
| foods    | 3.6 (2.8) | 3 (9) |
| tools    | 8 (5.9) | 7 (24) |
| all      | 5.8 (5.1) | 4 (26) |
Word length and phonological neighbourhood density were negatively correlated with each other $(r[132] = -0.67, p < .001)$. Word frequency could not be fully matched across conditions. Word frequencies were obtained from SubtLex (Keuleers et al., 2010), which provides a standard measure of word frequency independent of corpus size: frequency per million words with a 4-digit precision. Word frequency was negatively correlated with word length $(r[132] = -.37, p < .001)$ and positively correlated with phonological neighbourhood density $(r[132] = .32, p < .001)$.

Because this variable is known to affect all stages of the word production planning (Hanulová et al., 2011; Indefrey. 2004), we assessed the related statistical effect on our variables of interest, by conducting one-way repeated measure ANOVA with Object category (animals, foods, tools, clothes), and Word Length (short, long), and Phonological Neighbourhood density (low, high) as independent variables. There was no effect of Word Frequency on Object category. As for the phonological variables, there was a main effect of Word Frequency on Word Length $(F[1,132] = 23.95, p < .001)$, with short words being more frequent $(M=12.82)$ than long words $(M=3.95)$. There was also a main effect of Word Frequency on Phonological Neighbourhood Density $(F[1,32] = 11.69, p = .001)$, with low density words being less frequent $(M=5.07)$ than high density words $(M=11.53)$.

**Experimental design:** Before starting the experiment, subjects received written and oral information about the picture naming procedure. Subjects were instructed to name aloud each object picture as fast and accurately as possible, and to speak clearly.

At the beginning of each experimental trial, a fixation cross was presented at the centre of the screen for 500 ms, followed by a blank interval for 500 ms, and the object picture was subsequently
presented for 350 ms. The 134 pictures were presented three times, once in each of 3 separate blocks, following a random order. The experiment lasted around 30 minutes including two self-paced breaks.

201 **Behavioural recordings**

Stimuli were presented using the Presentation software (Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com). The pictures were displayed at the centre of the screen at a size of 300 by 300px (1920 by 1080 screen resolution and a refresh rate of 120 Hz, delay <1ms with almost instantaneous presentation of the full screen), in a light grey background within a visual angle of 4 degrees. They were presented using a liquid crystal display video projector and back projected onto the screen using two front-silvered mirrors.

Vocal responses were captured with a microphone and recorded at 44.1 kHz using the Audacity software (https://audacityteam.org/).

Vocal responses were recorded as wav files and response latencies were determined offline using the Praat software (Boersma a Weenink, 2018).

**MEG recordings**

Subjects were seated in the MEG system in a magnetically shielded room. They were asked to sit comfortably but to keep their body and head still during the task, and to try to avoid blinking. They were instructed to look at the stimulus screen, located 40 cm in front of them, focusing on the
centre of the screen. The MEG signals were recorded using a high-density whole-head system (OMEGA 2000; CTF Systems), consisting of 275 axial gradiometer channels and 29 dedicated reference channels for environmental noise cancellation. The subject’s head was registered to the MEG sensor array using three coils placed at 3 anatomical landmarks (nasion, and left and right ear canals). The head position was continuously monitored during the MEG recordings, and readjusted during breaks if it deviated more than 9 mm from the initial position (Stolk et al., 2013). Head movements did not exceed 1.25 cm between blocks. Pairs of Ag/AgCl electrodes were used to record the horizontal and vertical electro-oculograms (EOGs), the electrocardiogram, and the surface electromyogram (EMG) from the orbicularis oris muscle (electrodes placed above the upper lip and below the lower lip) (impedance was kept lower than 15 kΩ for all electrodes). MEG, EMG, and EOG signals were analog low-pass filtered at 300 Hz, digitised at 1200 Hz, and saved to disk for offline processing.

Data Analysis

Behavioural Data

Latencies of the subjects’ verbal responses were calculated offline by subtracting the time of picture onset marked by a 10 ms auditory signal (inaudible to the participants) from the time of speech onset. Praat software (http://www.praat.org; Boersma and Weenink, 2019) was used to analyse the recorded audio signal and to semi-automatically identify the onset of beep and articulation.
Automatic silence/non-silence interval boundaries were obtained by using intensity (dB) thresholds, and the resulting onset boundaries were manually inspected and corrected where needed (most often at word-initial voiceless consonants or vowels). 83% of response trials were correctly named (identical to the target word). Incorrectly named trials (3%) and verbal disfluencies (stuttering, utterance repairs, and production of non-verbal sounds) (14%) were excluded from the analyses.

In order to assess and assure synchrony between the onsets of picture and of the auditory beep signal for subsequent MEG data analyses, the audio files were aligned with the picture onset triggers for the pre-processing of the MEG data, and with the audio channel in the MEG. For the alignment, the (very small) difference in clock speed of the computer for audio recording and the MEG acquisition computer was taken into account by estimating the delay between the presentation triggers and the delivery of the beep.

For behavioural data analysis, the effects of the variables of interest on naming latencies was assessed by conducting a one-way repeated measures ANOVA on the averaged naming latencies of each subject with Semantic Category, Word Length, and Phonological Neighborhood Size as independent variables.

MEG data pre-processing

Data were processed using MATLAB (Version R2021a) and the FieldTrip toolbox (Oostenveld et al., 2011). Data were epoched into segments from -100 to 1000 ms relative to picture onset. Independent Component Analysis (ICA) was used to remove ECG artifacts using the logistic
infomax ICA algorithm (Bell and Sejnowski, 1995), using the EEGLAB implementation (Delorme and Makeig, 2004; http://eeglab.org). Prior to decomposing the MEG signal into components, data were band-pass filtered in the 1-30 Hz range and down-sampled to 300 Hz. The topographies of the components were visually inspected, along with their time course for the first 40 trials, and the effect of removing the components that were identified as containing artifacts was checked. Samples contaminated by artifacts due to eye movements, muscular activity, and superconducting quantum interference device jumps were replaced by NaN (not a number) to allow excluding those samples from further analysis. For the sensor-level analysis, we calculated synthetic planar gradients using a nearest-neighbour interpolation (Bastiaansen and Knosche, 2000). Combining horizontal and vertical components of the estimated planar gradients allows for easier interpretation of the sensor-level results, as the maximal activity is typically located above the source (Hämäläinen et al., 1993).

Classification pattern analyses

Spatiotemporal multivariate pattern analysis (MVPA) was used to assess whether the experimentally manipulated stimulus features could be decoded from the MEG sensor signals. The stimulus features of interest were 1. Object category, 2. Word Length as quantified by the number of syllables, and 3. Phonological Neighbourhood Density. The object category variable coded for the four categories of animals, tools, foods, clothes, the word length variable coded for short (mono-) and long (bi-syllabic and trisyllabic) words, the phonological neighbourhood density variable was discretized into four classes of approximately equal size (smaller/small/large/larger). As such, decoding for the variables of interest constituted 4-way and binary classification tasks.
To control for low-level visual confounds, we decoded the object categories while accounting for low-level visual features of the object images in the estimation of the within-subject null-distribution of the category classification (see below). We took the following visual features into account: contrast, as measured by the intensity contrast between a pixel and its neighbours over the whole image (Corchs et al., 2016), visual complexity, quantified by Edge Density as the percentage of pixels that are edge pixels (Rosenholtz et al., 2007; Forsythe et al., 2008), and colourfulness, measured using distributional properties of pixel colour values, as being strongly correlated with human judgments (rs = .95) (Hasler & Suesstrunk, 2003). These variables were discretized into two classes of approximately equal size (low/high).

Additional post hoc analyses were conducted as a sanity-check of the data to ascertain whether an uncontrolled variable, such as lexical frequency, yielded spatiotemporal patterns of classification that were similar to the patterns resulting from the main analysis, thus confounding the results. Word frequency was included because it is highly correlated with word length and neighbourhood density.

We used a Gaussian naive Bayes classifier (Mitchell, 1987), as implemented in the MVPA-light toolbox (Treder, 2020). Despite its assumption that features are conditionally independent of each other given the class label (Bishop, 2006), this classifier has shown to be remarkably successful in a wide range of classification problems. To evaluate classification performance and to control for overfitting, repeated stratified five-fold cross-validation was employed. The data were randomly split into five equal folds, ensuring the equal presence of classes in each fold (stratification). The model was trained on four folds and validated on the fifth fold. The process was repeated five times, such that each fold was used for validation. The entire cross-validation was furthermore repeated five times with new randomly assigned folds, to reduce bias that might be caused because...
of how data was randomly partitioned into folds, and the final averaged results are reported. To avoid classification bias due to class imbalance in the class labels, random under-sampling was applied to training and test data, by discarding randomly selected samples from majority classes until each class was represented by an equal number of samples.

We quantified the decoding performance by means of classification accuracy, which is the fraction of correctly predicted class labels. The higher the classification accuracy, the better response patterns associated with the class labels can be determined.

We applied a spatiotemporal searchlight procedure and defined for each time point and channel a local searchlight, consisting of a 150 ms time window (45 data points at the downsampled sampling rate of 300 Hz), and the set of first order neighbouring synthetic planar gradient channels. On average, each channel had 7.4 neighbours (SD=1.1, range 4-10), and thus the classification at each channel-time point was based on 332 features (range 180-450). To assess whether classifier performance was above chance performance, we estimated the chance level empirically, using permutations at the single subject level. We repeated the classification testing after shuffling the class labels, and recomputed classifier performance on the shuffled class labels to obtain a distribution under the null hypothesis of exchangeability of class labels (see, e.g., Kaiser et al., 2016; Cichy et al., 2014; Isik et al., 2014). The randomisation of class labels for the number of syllables and the phonological neighbourhood density classification was constrained to account for the fact that the object category was not fully orthogonal to the other features of interest. To this end, the randomisation of class labels was performed for each object category separately.

We controlled for low-level visual confounds in the classification of semantic object categories, by constraining the within-subject randomization procedure (to obtain the subject-specific
distribution for object classification under the null hypothesis) to binned collections of stimuli, where the bins were defined according to the visual features of the images. For statistical inference, we used non-parametric cluster-based permutation tests across space and time (Maris and Oostenveld, 2007), using 2,000 permutations. The cluster-based permutation procedure employs the same spatial neighbourhood structure that was used in the classification searchlight procedure, clustering the selected samples (sensors, time points) on the basis of spatial and temporal adjacency. The test-statistic used was a group-level T-statistic against the empirical chance level using the subject-specific Z-standardised decoding accuracy scores. These Z-scores for each subject were obtained by subtracting the mean accuracy obtained from 100 randomizations from the observed accuracy, and dividing by the standard deviation across randomizations.

Results

Behavioural results

Table 2 reports the average RT and naming accuracy per object category, as quantified based on subjects' actual responses during the MEG experiment. One-way repeated measures ANOVA on the average naming latencies of each subject with Object Category, Word Length, and Phonological Neighborhood Size as independent variables revealed a main effect of Object Category ($F[1, 32] = 12.6, p < .001$), with the RTs being slower for Clothes (see Table 2).

Table 2

Averaged RTs and naming accuracy values per object category.
### Table 1: Reaction Times and Accuracy for Each Category

<table>
<thead>
<tr>
<th>Category</th>
<th>Mean</th>
<th>SD</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Animals</td>
<td>831</td>
<td>90</td>
<td>84.3</td>
<td>12.9</td>
</tr>
<tr>
<td>Clothes</td>
<td>891</td>
<td>100</td>
<td>67.3</td>
<td>12.2</td>
</tr>
<tr>
<td>Foods</td>
<td>841</td>
<td>100</td>
<td>84.7</td>
<td>10.2</td>
</tr>
<tr>
<td>Tools</td>
<td>834</td>
<td>80</td>
<td>88.2</td>
<td>12.3</td>
</tr>
</tbody>
</table>

The analyses also revealed a main effect of Word Length on the RT (F(1, 32) = 7.34, p = .011), mono-syllabic words (M=838 ms, SD=98) showing a shorter RT than bi-syllabic words (M=855 ms, SD=92). Neighbourhood density also showed a main effect (F(1, 32) = 7.09, p = .012), as words with a high number of phonological neighbours showed shorter RTs (M=838 ms, SD=86) than those with a low number of phonological neighbours (M=854 ms, SD=95).

Furthermore, we inspected whether Object Categories interacted with Word Frequency. As expected, high frequency words (M=821 s, SD=85) were named faster (F(1, 32) = 50.43, p < .001) than low frequency words (M=877 s, SD=102). Importantly though, there was no interaction of Word Frequency with Object category, suggesting that the semantic effects were not linked to word frequency.

### Results from MVPA searchlight analysis on MEG data

Results from the spatiotemporal searchlight analyses further revealed specific neural dynamics associated with the key semantic and phonological variables of the present study. Decoding outcomes yielded above chance raw accuracy values of 37% for Object Category, 56% for Word
Length, and 28% for Phonological Neighbourhood Density. Although raw accuracy values were not high, they were above chance. Figure 1 and Figure 2 respectively display the topographic distributions and the time course of decoding accuracy for each contrast of interest. We found that the four object categories were significantly distinguished early on, starting in the time window 150-250 ms, in bilateral occipital, temporo-parietal, central and left frontal sensor areas, as shown in Figure 1 (top panel).

In contrast, above chance decoding accuracy for word length was first observed in a central sensor area starting in the time window 150-250 ms, and later, starting in the time window 250-350 ms, in left temporal and frontal sensors (see Figure 1, middle panel). Starting in the subsequent time window (350-450 ms), decoding accuracy for phonological neighbourhood density was significantly above chance over a more focused sub-set of left frontal sensors (Figure 1, bottom panel).

The corresponding time courses of decoding accuracy from selected left occipito-parietal and posterior temporal sensors for Object Categories, and from left frontal and adjacent anterior temporal sensors for Word Length and Phonological Neighbourhood Density are displayed in Figure 2.

The searchlight results for the key study variables are summarised as time courses of decoding accuracy (Z-standardised values) across left frontal and temporal sensors in Figure 3.

For the control variable of Word Frequency, decoding accuracy was above chance in central sensors in an early time window (Figure 4A), which was comparable to the spatiotemporal effects observed for word length. Above chance accuracy for Word Frequency across left frontal and
temporal sensors was however only seen from a later time window (450-550 ms), as displayed in Figure 4B, suggesting differences in the underlying processes.

Figure 1

Figure 1. Topographic maps displaying the effect of decoding accuracy (expressed as a T-statistic for each key variable: 1) “Object Categories (upper panel); 2) Word Length (middle panel); 3) Phonological Neighbourhood Density (bottom panel). Data points exceeding the nominal cluster-threshold at p<.05 (uncorrected) are marked with an asterisk. Note that object category information was decoded early on, starting in the time window from 150-250 ms, across a distributed set of bilateral occipito-parietal, and left temporal sensors. From the subsequent time window (250-350
ms), word length was decoded in frontotemporal sensor area, whilst phonological neighbourhood
density was decoded in left frontal sensors from 350-450 ms. Note that for phonological
neighbourhood density no data points exceeded the a priori cluster threshold for multiple
comparisons correction.

Figure 2

Figure 2. Time courses of decoding accuracy for the key study variables (left panels), peaking at
subsequent times (as marked by vertical bars). The horizontal dashed bars in yellow indicate the a priori cluster threshold for uncorrected $p<0.05$ for decoding accuracy. Data are extracted from left occipito-parietal and posterior temporal sensors areas for Object category (upper panel), from left frontal and adjacent anterior temporal sensors for both Word Length (middle panel) and Phonological Neighbourhood Density (bottom panel), as shown in the topographic views on the right. The sensor selection was based on a priori hypotheses on the emergence of semantic information in posterior occipito-parietal and temporal regions, and of phonological processes in left frontal regions, as proposed by the Indefrey and Levelt (2004) model.

**Figure 3**

**Figure 3.** Overview of the searchlight results displayed as time course of decoding accuracy across left frontal and temporal sensors (as displayed in the topographic view on the right of the figure) for the key study variables “Object category” (upper panel); “Word Length” (middle panel);
“Phonological Neighbourhood Density” (bottom panel). For each contrast of interest, the colour bar indicates the Z-standardised decoding accuracy values, from zero values (in green), to increasingly positive Z-scores (up to dark red).

Figure 4

A. Topographic maps displaying the effect of decoding accuracy (T-statistic) for Word Frequency. For each contrast of interest, the colour bar indicates the Z-standardised decoding accuracy values from zero values (in green), to positive Z-scores (up to dark red). Note the early activity arising from central sensors for the Word Frequency condition, in a time window which
was comparable to the effects in central sensors found for word length. B. Overview of the searchlight results displayed as time-course of decoding accuracy across left frontal and temporal sensors (as displayed in the topographic view on the right of the figure) for Word Frequency.

Discussion

Employing MVPA analyses of the MEG data, we investigated the time course of neural activity underlying semantic and phonological access in object naming. Conceptual preparation was cued by categorical distinction of the stimulus objects, whilst syllabification and phonetic encoding were respectively linked to the manipulation of word length and phonological neighbourhood density. Our method revealed discriminability of object categories in occipito-parietal and temporal sensors early on (150-250ms). Word length was decoded from the subsequent time window (250-350ms) in left frontotemporal sensors, followed by phonological neighbourhood density (350-450ms).

Regarding the competing serial/hierarchical and early/parallel hypotheses, based on the target variables of the present study, our results speak in favour of a cascade of neural events with sequential onset times for the conceptual, phonological and phonetic processes that enable overt object naming.
The observed early differentiation of object categories during conceptual preparation is consistent with established knowledge that brain activity in the ventral visual pathway rapidly evolves from the visual perception and recognition of object images in posterior occipital and parietal cortex to the perceptual feature conjunctions enabling the representation of conceptual meaning in inferior temporal and fusiform cortex, as identified by previous MEG work (Clarke et al., 2013). In relation to the sketched continuum from perception to conceptualisation in the visual regions, the initial activity sparking from occipital sensors to parietal and posterior temporal sensors was expected to reflect visual perceptual processes from which categorical object representations emerge during the early stages of object meaning identification (Carlson et al., 2014). Recent studies have shown that perceptual similarity predicts object categories more reliably than semantic category at early stages of visual object comprehension (Proklova et al., 2019; Contini et al., 2020). The perceptual properties of the visually presented objects can then obviously confound the decoded activity in language tasks (Simanova et al., 2015; Vindiola and Wolmetz, 2011). As our data were controlled for basic visual features during classification of object categories, we found no decoding effects in the semantic object category condition in the first time window (0-150 ms). We interpret the results from the second time window onwards as being predominantly driven by semantic processes linked to object meaning identification and categorisation, consistent with the prediction from the Indefrey & Levelt (2004) model regarding conceptual preparation. The present study could thus effectively decode semantic object category information about objects at sensor level independently of the visual modality (Haxby et al., 2001; Cox and Savoy, 2003; Kamitani and Tong, 2005; Reddy and Kanwisher, 2007) with which an object is presented, in line with previous MVPA studies from whole-head high-density MEG signals (Nieuwenhuijzen et al., 2013; Simanova et al., 2010). The replicability of decoding of object categories in the present study also
corroborates earlier evidence that patterns in sensor space carry information about object categories with latencies around 150-250 ms post picture presentation (see Contini et al., 2017 for review).

Our results determined the relative earliness of conceptually driven preparatory activity with respect to the later latencies of phonological processes during word production planning, confirming MVPA pattern-classification of MEG data as being a reliable tool to investigate the time course of the neural dynamics in language tasks such as overt naming.

A second result of importance of our present work concerns the time course of neural activity associated with phonological information, as word length was decoded from left frontotemporal sensors, from 250-350 ms, whilst phonological neighbourhood density from left frontal and temporoparietal sensors was decoded 160 ms later, from the time window 350-450 ms. In line with the Levelt and Indefrey’s model, these phonological operations started immediately after the onset time expected for phonological code retrieval. Also, their spatial distribution confirms consolidated evidence for a role of frontal regions in phonological processes (Carreiras et al., 2006; Fiebach et al., 2002; Fiez et al., 1999; Graves et al., 2007; Hauk et al., 2004; Wilson et al., 2009; Papoutsi et al., 2009), and a role of temporal regions in mapping auditory and motor representations (Hickok and Poeppel, 2007; Hickok et al., 2003), and in phonological/phonetic encoding (de Zubicaray et al., 2002; Bles and Jasma, 2008; Wilson et al., 2009; Gorno-Tempini et al., 2004, 2008; Henry and Gorno-Tempini, 2010; Wilson et al., 2010), both in production and in perception (e.g., DeWitt and Rauschecker, 2012; Hickok and Poeppel, 2007; Indefrey, 2011; Indefrey and Levelt, 2004; Obleser et al., 2003).
Interestingly, later left temporo-parietal activity was specific to neighbourhood density (around 550 ms post picture onset), and may be due to self-monitoring processes supported by the corresponding brain regions (e.g., McGuire et al., 1996; Hirano et al., 1997; Shergill et al., 2002; Tourville et al., 2008). Earlier neuroimaging studies have shown a specific modulation of activity in superior temporal cortex by words with high neighbourhood density (Okada and Hickok, 2006). Words with a phonological neighbour may reduce neural activation in a network of superior temporal, inferior parietal, inferior frontal, and precentral regions, reflecting a facilitatory effect in lexical selection, phonological and articulatory planning, due to the overlap between the phonological representation of the target word and of its competitor neighbour (Peramunage et al., 2010; see also Harley and Brown, 1998).

Overall, spatiotemporal properties of the neural activity in our data diverge from parallel models suggesting early concurrent activity for conceptual and phonological information due to the ignition of cells in multiple regions for multiple functions (Miozzo et al., 2014; Strijkers et al., 2017). Rather, the present results provide evidence that conceptual preparation and phonological/phonetic encoding trigger activity in the language systems at subsequent time intervals. Our findings therefore support hierarchical models with temporal separation of onsets at the semantic and phonological planning stages.

In the present results, transient activity in central sensors became manifest in the topographic distribution of neuromagnetic activity for word length 150-250 ms. Central sensors notably pick up activity in medial frontal regions, such as (pre-) supplementary motor area (SMA), and/or cingulate cortex, which earlier studies have reported for similar tasks. For instance, the SMA plays diverse roles in lexical selection, phonological sequencing, articulatory (Dronkers, 1996; Jurgens, 2009; Tremblay and Gracco, 2009), and self-monitoring (e.g., Möller et al., 2007), along an
anterior-to-posterior anatomical gradient (Alario et al., 2006). However, the existing literature provides inconsistent evidence for the time course of activity and functions of this region in overt naming. In a previous metabolic study, Wilson and colleagues (2009) found activity modulation of SMA and premotor regions as a function of word length, and speculatively proposed it to reflect access to articulatory scores. Another picture naming study using MEG identified a pattern of connectivity linking medial frontal regions, supramarginal gyrus, and posterior middle temporal cortex around 300 ms post picture presentation (Liljeström et al., 2015), and suggested a role of the SMA in early control processes for articulatory planning. At similar latencies (300 ms post picture onset), electrophysiological work reported SMA activity linked to vocal onset as an index of motor preparation in speech (Riès et al., 2013).

With respect to these earlier studies, the effect in central sensors in our data occurred even earlier, when previous studies found early phonological (Miozzo et al., 2014) and articulatory activations (Strijkers et al., 2017). This issue deserves further investigation. Still, in the context of our present study, we reasoned that, if the central sensor activity was linked to purely phonological or phonetic processes, we would observe a similar early effect also for the other phonological variable we employed here, phonological neighbourhood density, which was however not the case. Since we carefully manipulated our stimulus set for our dimensions of interest, but we could not control for lexical frequency, for which post hoc analyses indicated early effects in central sensors, we interpret these early dynamics as reflecting differences in intercorrelated lexical properties, such as word length and lexical frequency (short high frequency words vs. long low frequency words), a variable known to influence behavioural performance in picture naming, and related neural activity exactly in middle frontal regions and cingulate cortex in both young and old adults (Gertel et al., 2020).
Conclusion

Using MVPA searchlights in sensor space, the present MEG study demonstrated that MEG sensor patterns carry information about both semantic object categories and phonological processes during word production planning. The results delineated a spatio-temporal dynamics of semantic access, syllabification and phonetic encoding characterised by modulation of activity in posterior occipito-parietal and temporal sensors by the different conceptual categories early on, and in left frontal and temporal sensor areas by the two phonological variables at later time windows. The time course of the neural events linked to these core word planning operations largely confirms a progression of neural activity from posterior to anterior regions of the language network along the axis from conceptualisation to phonological/phonetic operations. Our results support serial cascading models of word production.

Our approach combining MEG and single-trial MVPA spatio-temporal searchlights offers a promising method for investigating the spatio-temporal properties of language planning in order to evaluate and refine current models of language production.

(1487 words – up to 1500)
References


References


