The time-course of phonological encoding: insights from time-resolved MVPA

Li Calzi, G.; Meyer, A.S.; Burght, C.L. van der

Title: The time-course of phonological encoding: insights from time-resolved MVPA

Abbreviated title: Decoding speech production planning using EEG

Authors: Giulia Li Calzi[1,2*], Antje S. Meyer[2,3], Constantijn L. van der Burght[2,4]

[1]Centre for the Interdisciplinary Study of Language Evolution, University of Zurich, 8050 Zurich, Switzerland

[2]Max Planck Institute for Psycholinguistics, 6525 XD Nijmegen, The Netherlands

[3]Donders Institute for Brain, Cognition and Behaviour, Radboud University, 6525 EN Nijmegen, The Netherlands

[4]Leiden University Centre for Linguistics, Leiden University, 2311 BD Leiden, The Netherlands

[*]corresponding author

Corresponding author email address: giulia.licalzi@uzh.ch

Please cite as:

Li Calzi, G., Meyer, A. S., & van der Burght, C. L. (2025). The time-course of phonological encoding: insights from time-resolved MVPA. *The Journal of Neuroscience*.

https://doi.org/10.1523/jneurosci.0546-25.2025

Number of pages: 22

Number of figures: 4

Number of tables: 6

Number of words for abstract, introduction, and discussion: 198 (abstract), 649 (introduction), 1496 (discussion)

Conflict of interest statement: None declared.

Acknowledgments: The authors thank Birgit Knudsen for help with data collection and Geert van der Meulen for help with the figure design. The study was funded by the Max Planck Society.

Abstract

To produce a word, speakers need to decide which concept to express, select an appropriate item from the mental lexicon and spell out its phonological form. The temporal dynamics of these processes remain a subject of debate. We investigated the time course of lexical access in picture naming with electroencephalography (EEG). Thirty participants (23 female) named pictures using simple nouns. The pictures varied in conceptual category (animate or inanimate), stress pattern (first or second syllable), and the structure of the first syllable (open or closed). Using time-resolved multivariate pattern analysis (MVPA), we decoded the time course in which each dimension was available during speech preparation. The results demonstrated above-chance decoding of animacy within 100 ms after picture onset, confirming early access to conceptual information. This was followed by stress pattern and syllable structure, at around 150 and 250 ms after picture onset, respectively. These results suggest that a word's stress pattern can be retrieved before syllable structure information becomes available. An exploratory analysis demonstrated the availability of the word-initial phoneme within 100 ms after picture onset. This result hints at the possibility that during picture naming conceptual, phonological and phonetic information may be accessed rapidly and in parallel.

Significance statement

Producing spoken words is an effortless yet complex process. The mechanisms through which we retrieve and assemble the sound structure of words remain largely unknown. So far, speech production theories have mostly relied on behavioural experiments. We investigated the time course of phonological encoding during spoken word production using EEG. Results show that after rapid processing of word meaning, speakers access a word's stress pattern, followed by the composition of individual syllables. We show that subtle linguistic characteristics can be predicted before a speaker produces a word. Importantly, this study demonstrates successful application of MVPA on pre-articulation data from the widely-available method EEG, offering an accessible approach to address novel questions in speech production research.

Introduction

Current psycholinguistic theories largely agree that the retrieval of a single word from the mental lexicon minimally requires the selection of the concept to be expressed, the selection of a unique lexical item (the lemma), and the retrieval of the corresponding phonological form (Levelt et al., 1999). However, despite extensive research, the temporal dynamics of these processes remain debated (e.g., Strijkers & Costa, 2016; Indefrey, 2016). Some behavioural and neurophysiological studies point to a cascade of processes, with the onset of lemma selection preceding the onset of phonological encoding (e.g., Carota et al., 2022; Hultén et al., 2009; Laganaro et al., 2009; Maess et al., 2002; Sahin et al., 2009; van Turennout et al., 1997, 1998; for reviews, see Indefrey, 2011 and Indefrey & Levelt, 2004). Other studies found evidence for near-simultaneous retrieval of semantic and phonological knowledge during the earliest stages of speech production planning (e.g., Carota et al., 2023; Fairs et al. 2021; Feng et al., 2021; Miozzo et al., 2015; Riès et al., 2017; Strijkers et al., 2010, 2017). These latter findings can be explained by theories assuming semantic and phonological information are activated rapidly and simultaneously because, due to their co-occurrence, they have been bound into integrated cell assemblies during the course of language acquisition (e.g., Kerr et al., 2023; Pickering & Strijkers, 2024; Strijkers & Costa, 2016). After this parallel ignition of the "word web", activity reverberates in local parts of the assembly producing the distinct spatiotemporal sequences reported in previous studies (e.g., Indefrey, 2011).

In the present study we used multivariate pattern analysis (MVPA) of EEG data to study the time course of conceptual and phonological encoding during picture naming in adult native speakers of Dutch. We presented pictures representing target words (see Figure 1) that depicted animate or inanimate concepts, and varied in their phonological properties, specifically their metrical structure (first or second syllable stress) and the internal structure of the first syllable (open, CV, vs. closed, CVC). Inspired by Carota et al. (2023), we also conducted an exploratory analysis decoding the manner of articulation of the word-initial phoneme (plosive vs. fricative).

One goal of the current study was methodological, namely to establish whether EEG data collected during a speech production task allow for time-resolved MVPA, which before had only been performed on MEG data collected during picture naming (e.g., Carota et al., 2022, 2023). A second goal was to examine the time course of conceptual and phonological encoding, for which cascaded and parallel models make different predictions. The cascade view predicts that conceptual information should be decodable before metrical structure and phoneme information, whereas word-web models allow for early, simultaneous availability of all stored information about a word.

Our third and most important goal was to study the timing of the processes *within* the phonological component, which has not been done with MVPA before (and rarely with other methods, but see Schiller et al., 2003). In languages with lexical stress, such as English or Dutch, the segmental structure of words (e.g., g-ɪ-t-ɑ-ɹ) and their stress pattern (σˈσ) must be stored in the mental lexicon (or at least information on whether the word's stress pattern is regular or irregular). It is often assumed that segmental and metrical information are represented as separate tiers, which are retrieved in parallel and subsequently combined in the process of syllabification (e.g., Alderete & O'Séaghdha, 2022; Cholin et al., 2006; Goldrick & Rapp, 2007; Levelt et al., 1999). This combination, which leads to the grouping of strings of segments into syllable chunks, is *not* stored in the mental lexicon. Rather, it is context-dependent and often straddles morpheme and word boundaries (e.g., *se-lect* vs. *se-lec-ting*). Importantly, both traditional cascaded and word web models predict that such context-sensitive processes should follow the retrieval of stored conceptual and phonological information. Time-resolved MVPA of EEG data, recorded during speech planning, should shed light on the relative time courses of these processes.

Materials and methods

Participants

Thirty-eight native Dutch speakers participated in the experiment after providing written, informed consent. Five participants were excluded due to technical issues and three for miscellaneous reasons

(response accuracy below 75%, wrong EEG set-up, and bad signal quality), leaving a final sample of 30 participants (23 females, mean age = 23.4 years, *SD* = 4.2 years). All participants were right-handed, had normal or corrected-to-normal vision, and reported no history of neurologic, developmental or language deficits. Participants received €15 for participation. The study was approved by the Ethics Committee of the Radboud University Faculty of Social Sciences under registration number ECSW-2019-019.

Materials

The stimuli consisted of 48 line-drawings taken from the picture database of the Max Planck Institute for Psycholinguistics or public domain repositories. Each picture corresponded to a disyllabic Dutch noun indicating either an object, an animal, or a professional figure (see Figure 1 for a full list). All targets but the three indicated with an asterisk were selected from the 96 disyllabic items used by Schiller et al. (2003). The items were balanced across the dimensions of animacy (animate vs. inanimate), lexical stress (word-initial vs. word-final) and structure of the first syllable (open vs. closed). We chose to investigate lexical stress and syllable structure by contrasting first vs. second syllable stress and open vs. closed syllables, respectively, because these features have been used to study the processing of lexical stress and syllable structure in previous work (Schiller et al., 2003). The conceptual dimension was included to enable a comparison to the time course of conceptual processing in Carota et al. (2022). The animacy feature, specifically, was selected because the Schiller et al. (2003) stimulus set contained a selection of both animate and inanimate target words. The mean lexical frequency of each item was obtained from SUBTLEX-NL (Keuleers et al., 2010), a database that provides a standard measure of word frequency independent of corpus size. Table 1 summarises the mean Zipf frequencies per condition.

Word-initial stress				Word-final stress			
Open syllable structure		Closed syllable structure		Open syllable structure		Closed syllable structure	
Animate	**koning** (king)	Animate	**dokter** (doctor)	Animate	**konijn** (rabbit)	Animate	**soldaat** (soldier)
	vogel (bird)		**zuster** (nurse)		**piraat** (pirate)		**vampier** (vampire)
	tijger (tiger)		**varken** (pig)		**kameel** (camel)		**pastoor** (pastor)
	jager (hunter)		**vlinder** (butterfly)		**piloot** (pilot)		**kalkoen** (turkey)
	zebra (zebra)		**pinguin** (penguin)		**docent*** (teacher)		**dolfijn** (dolphin)
Inanimate	hamer (hammer)		**panter** (panther)		**matroos*** (sailor)	Inanimate	magneet (magnet)
	ketel (kettle)		**lifter** (hitchhiker)	Inanimate	fabriek (factory)		kasteel (castle)
	lepel (spoon)		**hamster*** (hamster)		raket (rocket)		balkon (balcony)
	molen (windmill)	Inanimate	tempel (temple)		toneel (stage)		portret (portrait)
	nagel (fingernail)		cirkel (circle)		gitaar (guitar)		kompas (compass)
	bezem (broom)		pleister (band-aid)		sigaar (sigar)		gordijn (curtain)
	vlieger (kite)		tractor (tractor)		kanon (cannon)		trompet (trumpet)

Figure 1: Target picture names used in the experiment divided over the phonological dimensions "lexical stress" (word initial vs. word final) and "syllable structure" (open vs. closed). The dimension "animacy" is indicated with bold or regular font. Words marked with an asterisk were not used in Schiller et al. (2003).

Table 1: Mean word frequency and standard deviation (SD) per condition

Animacy	**Lexical stress**	**Syllable structure**	**Frequency (*SD*)**
Animate	Word-initial	Open	4.25 (0.62)
		Closed	3.78 (1.03)
	Word-final	Open	3.85 (0.43)
		Closed	4.08 (0.53)
Inanimate	Word-initial	Open	3.68 (0.12)
		Closed	3.75 (0.43)
	Word-final	Open	4.07 (0.18)
		Closed	3.70 (0.38)

Procedure

Participants were tested individually while seated in front of a computer screen located inside an electrically-shielded, sound-attenuated booth. After being fitted with the EEG cap, participants took part in a practice round where they named eight monosyllabic targets not included in the experiment. During this first round, they reviewed PowerPoint slides containing line drawings of the practice targets with their designated names (e.g., the Dutch word “bloem” and the picture of a flower) and then produced these words in a short practice experiment that mirrored the timing of the experiment to familiarise themselves with the task. They then familiarised themselves with the experimental items, reviewing a PowerPoint presentation containing the 48 experimental pictures with their names. After reviewing the experimental targets, participants performed a round of self-paced naming with feedback from the experimenter to confirm that they had memorised the target items correctly.

At the beginning of each trial, the picture was presented at the centre of the screen for 450 ms followed by a blank interval and a fixation cross for 600 ms each. Stimuli were presented using Presentation (Neurobehavioral Systems). The pictures were displayed at the centre of the screen (1920 x 1080 resolution, 60 Hz refresh rate) at a size of 250 x 250 pixels on a light grey background. Vocal responses were recorded using a microphone placed in front of the participant and digitised at a sampling rate of 44.1 kHz. Every target picture was presented ten times, amounting to a total of 480 trials divided into ten blocks of 48 trials each. Since previous work has shown that naming latencies stabilise from the second repetition onwards (Corps & Meyer, 2025; van Turennout et al., 2003), the first two naming cycles were excluded from the analyses. Each block contained all 48 items in randomised order. Participants took short self-paced breaks in the middle and at the end of each block. The breaks often lasted for less than 15 seconds, and no participant paused for more than one minute. At the end of the second naming block (i.e., during the fourth break), the experimenter verbally corrected non-target productions (e.g., please use “portrait” instead of “painting”). After this,

participants received no more feedback for the remainder of the experiment. Altogether, the training and the actual experiment lasted approximately 20 minutes.

EEG recordings

Participants were instructed to focus on the centre of the screen, to avoid blinking and body movements. The EEG signal was recorded at 1000 Hz from 62 scalp electrodes using a 64-channel BrainAmp amplifier. The two remaining electrodes were placed on the left mastoid, and under the left eye. During acquisition, all channels were referenced to the right mastoid and an electrode placed on the forehead was used as the ground. Channels 1-9, 11-32 and 33-63 correspond to electrodes located on the scalp. The remaining channels were connected to electrodes placed at the left mastoid (channel 10) and below the left eye (channel 64). To ensure symmetry, channel 21 was placed at position FCz instead of position TP10.

Data analysis

As part of the picture naming paradigm, onset latencies and accuracy data were acquired. These behavioural variables mainly served as control measures, to ascertain that any EEG effects were not driven by differences in processing speed or accuracy across conditions. For animate words, we expected slightly faster naming latencies than inanimate words (e.g., Sá-Leite et al., 2021). In addition, we expected that words with first-syllable stress would be named faster than those with second-syllable stress (Schiller, 2006).

Behavioural data

The data from the first two blocks were considered warming-up trials and were excluded from the analyses. The remaining trials was manually tagged as either correct or incorrect. Incorrect trials were any trials where participants produced words that did not correspond exactly to the target name introduced during the familiarisation phase, trials including filled pauses and/or repairs, trials where no verbal response was produced, and trials where the target word was not completed by the end of

trial (i.e., within 1650 ms after picture onset). When a response straddled a trial boundary, the following trial was also excluded. If response accuracy was below 75%, the participant was excluded from further analyses (1 participant). Response latencies for correct responses were calculated offline by subtracting the time of picture onset from the time of speech onset. The onset and offset of articulation were identified semi-automatically in Praat (Boersma & Weenink, 2024) and manually corrected where needed.

The effect of the experimental variables on response accuracy and onset latencies (log-transformed) was assessed with (generalised) linear mixed models (GLMMs) conducted using the *lme4* package (Bates et al., 2015) in *R* (R Core Team, 2024). For the fixed effects analysis, the experimental factors were sum-coded and we aimed to include a maximal random effects structure (Barr et al., 2013). In case of convergence issues, the correlation between random intercepts and slopes was removed first, then the random slope that accounted for the least amount of variance until the model converged. As a result of this procedure, the random effects structures were slightly different across models, yet each model contained the maximal random effects structure that still converged. For both dependent variables, the full model included fixed effects for animacy, lexical stress, and syllable structure. The random effects structure included by-participant and by-item random intercepts as well as a by-participant random slope for animacy and syllable structure while omitting the correlation between random intercepts and slopes for the response accuracy model. In the analysis of onset latencies, the random effects structure included a by-participant random slope for animacy, lexical stress, and syllable structure.

To assess the effect of manner of articulation, we ran further, exploratory models that included animacy, lexical stress, and syllable structure, with manner of articulation as an additional fixed effect. We chose to analyse manner of articulation in separate models, so that the main, confirmatory model contained only predictors that had been originally planned. In the exploratory model of response accuracy, the random effects structure included by-participant and by-item random intercepts as well as by-participant random slopes for all predictors, while omitting the

correlations between random intercepts and slopes. In the exploratory model of onset latencies, the random effects structure included by-item and by-participant random intercepts and by-participant random slopes for animacy, lexical stress, and manner of articulation. Statistical inference was performed according to the z-value (response accuracy) or t-value (onset latencies) associated with the fixed effects parameter estimates. Values larger than 1.96 and smaller than -1.96 were considered statistically significant.

EEG data pre-processing

The EEG data was pre-processed in MATLAB (The MathWorks Inc., 2024) using the FieldTrip toolbox (Oostenveld et al., 2011). Prior to segmentation, the continuous EEG signals were pre-processed using a discrete Fourier transform for line noise removal and bandpass filtered in the 0.1-200 Hz range to minimise temporal distortions (Tanner et al., 2015). The EEG data was later epoched into segments from -100 to 1500 ms relative to picture onset. Following the rejection of trials with non-target productions, non-EEG channels were discarded and the data was re-referenced to the average of the 62 scalp electrodes. Because MVPA is more robust to artefacts than ERP analyses (Carlson et al., 2019), standard artefact rejection (e.g., ICA, manual rejection of trials contaminated by eye blinks or muscle artefacts, etc) was not performed.

Multivariate EEG analyses

To assess whether the experimental manipulations could be decoded from the EEG signal, we performed within-participant multivariate pattern analysis (MVPA) using the Linear Discriminant Analysis (LDA) implemented in the MVPA-Light toolbox (Treder, 2020). Decoding performance was assessed using the fraction of class labels predicted correctly (i.e., classification accuracy). Trials were classified according to their animacy, lexical stress, or syllable structure. Because decoding the variable of interest constituted a binary classification task in all three variables (animate/inanimate, word-initial/word-final, open/closed), the theoretical chance accuracy was always 0.5.

In the first set of analyses, we performed (i) time-resolved MVPAs to investigate classification performance across time and (ii) searchlights across time and channels to identify which electrodes drove classification performance at different times in the trial. In the second set, we performed time x time generalisations to investigate whether the neural code supporting above-chance decoding in the first set of analyses was stable or dynamically evolving (King & Dehaene, 2014). The resulting temporal generalisation matrix represents the decoding performance of a classifier trained at one time point of the trial and later tested at all the other time points of the trial. Square generalisation matrices indicated that the neural code supporting above-chance decoding is stable (i.e., the neural representations are sustained over time), while diagonal matrices indicate that the classifier can generalise only over a transient period of time, suggesting the neural representations are dynamically evolving (for a review, see King & Dehaene, 2014).

In both sets of analyses, the data points included in a 75 ms window centred around a given time point were considered "time" neighbours of that specific time point. In the searchlights across time and channels, we also calculated the pairwise distances between electrodes using the information contained in layout.pos (see FieldTrip's acticap-64ch-standard2.mat). In the distance matrix obtained using these pairwise distances, all electrodes closer to each other than 0.0001 units were considered neighbours (approximately 2-3 neighbours per electrode). Including neighbours increases the dimensionality of the feature space, thus providing more information to the classifier.

Increasing the dimensionality of the feature space, however, also makes the classifier more susceptible to the danger of overfitting (Treder, 2020). To control for the risk of overfitting, classification performance was evaluated using *k*-fold cross-validation with $k = 5$. In this validation technique, the data is randomly split into *k* equal folds to train and evaluate the model on slightly different datasets. To achieve this, the model is trained on $k-1$ folds and tested on the remaining fold for *k* times to ensure each fold is used for validation at least once. The resulting *k* performance metrics are then averaged to obtain a stable estimate of model performance. This entire cross-validation process was repeated five times using new randomly assigned folds to reduce bias that might be

caused by the way the data was randomly partitioned into folds. This procedure allows for a robust estimate of model performance resulting from the average of the accuracies obtained in the various rounds of cross-validation (Carota et al., 2022).

Prior to training, randomly selected samples belonging to majority classes were discarded until each class was represented by an equal number of samples to avoid classification bias as a result of class imbalances caused by trial rejection. The data was then z-scored using parameters obtained solely from the training data to avoid information transfer between the training and test set (nested pre-processing). Although z-score normalisation may not always be necessary or helpful, we included it to guard against the possibility that relatively noisier channels would drive multivariate findings (Ashton et al., 2022). In the final step of the MVPA pre-processing pipeline, we binned 16 z-scored trials from the same class into a single pseudo-trial to downsample the data and simultaneously increase signal-to-noise ratio (Treder, 2020). Statistical significance of classification accuracies against chance (i.e., right-tail tests against the chance level of 0.5) was established using non-parametric permutation tests with 1000 repetitions and cluster-based correction for multiple comparisons (cluster-defining threshold = 1.96, alpha = 0.01; Maris & Oostenveld, 2007). The analysis scripts and the results of the MVPA analyses have been deposited on the Open Science Framework (https://osf.io/nvgwx/).

Results

Behavioural results

The thirty participants included in the analyses produced the target word fluently and within the allotted time frame in 89.8% of the trials included in the analysis (Table 2). In the remaining trials, they produced verbal disfluencies (3.2%), trial overlaps (2.6%), incomplete responses (2%), incorrect targets (1%) or failed to produce a response (1.4%). In the analysis of response accuracy, there was a main effect of lexical stress (β = -.472, *SE* = .151, *z* = -3.129; Table 3). On average, participants named targets with word-initial stress more accurately than targets with word-final stress (92% vs. 87.6%).

The mean speech onset latency across conditions was 636 ms after picture onset (*SD* = 54 ms). The average word duration was 530 ms (*SD* = 61 ms). There was a main effect of animacy (β = .068, *SE* = .026, *t* = 2.615; Table 4) on speech onset latencies: on average, participants produced inanimate targets 33 ms earlier than animate ones (620 ms vs. 653 ms after picture onset). Finally, the additional analysis on manner of articulation revealed no significant effect of manner of articulation on response accuracy. The onset latency analysis showed that targets with word-initial fricatives were produced faster than targets starting with plosives (606 ms vs. 653 ms after picture onset; β = -.079, *SE* = .027, *t* = -2.916). The model results from the analyses that included manner of articulation are presented in Tables S1 and S2.

Table 2: Accuracies and onset latencies per category.

Animacy	**Lexical stress**	**Syllable structure**	**Accuracy (%)**	**Speech onset (*SD*)**
animate	word-initial stress	open	90.3	644 ms (137 ms)
		closed	92.0	646 ms (128 ms)
	word-final stress	open	80.6	688 ms (144 ms)
		closed	88.5	634 ms (129 ms)
inanimate	word-initial stress	open	93.0	608 ms (126 ms)
		closed	90.9	618 ms (130 ms)
	word-final stress	open	89.9	622 ms (124 ms)
		closed	90.1	623 ms (114 ms)

Table 3: Fixed effects from generalised linear mixed effects model of response accuracy.

Term	**Estimate**	***SE***	***z*-value**
Intercept	2.972	0.189	15.730
Animacy	-0.317	0.171	-1.850
Lexical stress	-0.472	0.151	-3.129
Syllable structure	-0.290	0.152	-1.910

Table 4: Fixed effects from linear mixed effects models of response latencies.

Term	**Estimate**	***SE***	***t*-value**
Intercept	-0.522	0.037	-14.246
Animacy	0.068	0.026	2.615
Lexical stress	0.028	0.025	1.101
Syllable structure	0.012	0.024	0.480

Multivariate pattern analysis (MVPA)

Time-resolved MVPAs

The classifier distinguished between animate and inanimate targets significantly above chance from 78 ms after picture onset onward (Figure 2A). At around 150 ms post stimulus, classification accuracy increased from below 60% to over 70%. Between 600 and 900 ms post picture onset, decoding accuracies gradually decreased before plateauing at around 60% for the remainder of the trial. The lexical stress (Figure 2B) and syllable structure (Figure 2C) of the target word started to be differentiated significantly above chance at 158 ms and 253 ms after picture onset, respectively. The decoding of lexical stress peaked around the average speech onset (from 55% to 75% between 400-700 ms post stimulus) and remained above chance throughout word production as well as after the average speech offset. The decoding accuracy of syllable structure remained above chance throughout word production (at approximately 65% accuracy). Differently from lexical stress, classification of syllable structure did not show a peak at the average speech onset nor remained stable after the average speech offset. Lastly, in the exploratory analysis, the manner of articulation of the first phoneme was differentiated significantly above chance between 91 and 139 ms after picture onset as well as between 428-1151 ms and 1219-1313 ms post stimulus (Figure 2D).

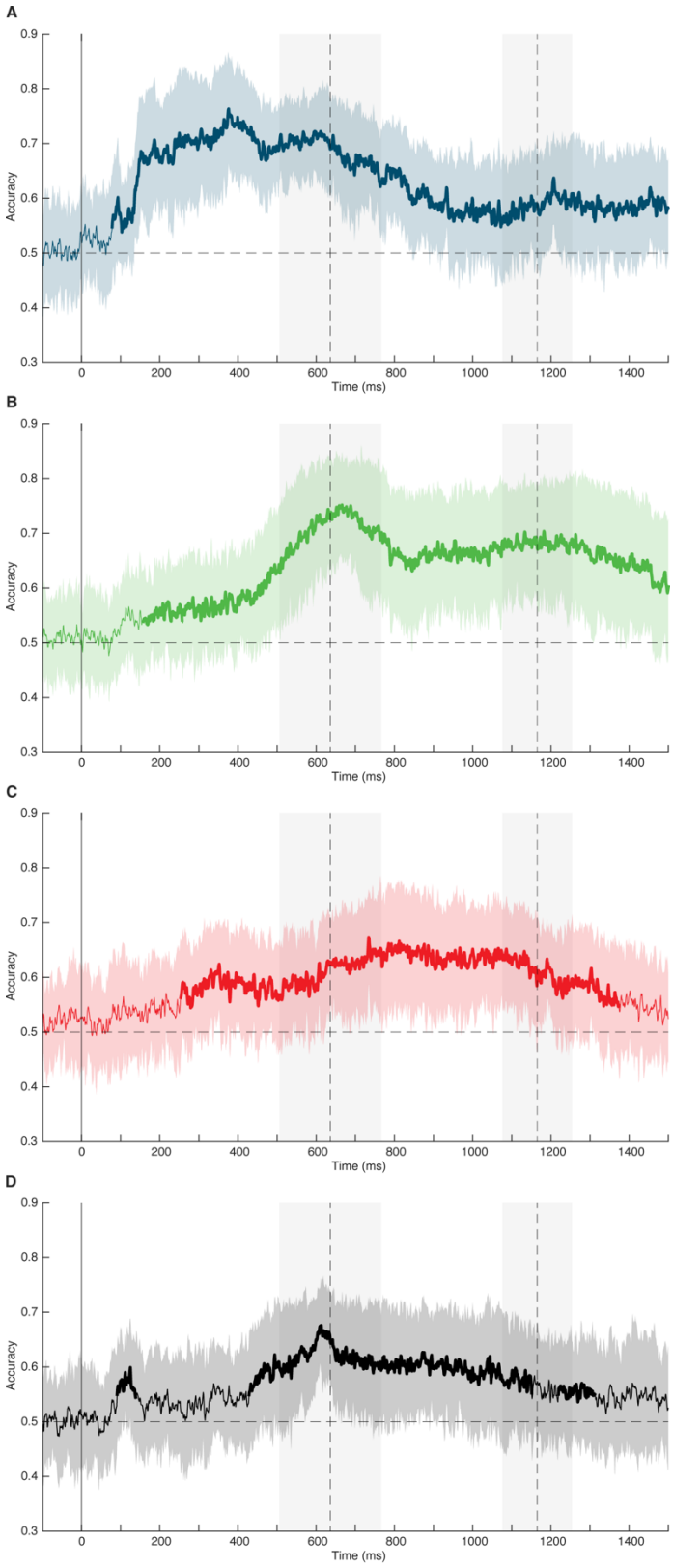

Figure 2: Time courses of decoding accuracy for the experimental variables (A) animacy, (B) lexical stress, (C) syllable structure, and (D) manner of articulation. Lines depict mean classification accuracy across participants, with bold marking above-chance level decoding accuracy, and the shaded areas

correspond to the standard error. The dashed horizontal line indicates chance-level decoding accuracy (0.5). Vertical lines indicate picture onset. Dashed vertical lines indicate the average onset and offset of speech, while shaded areas represent ±1 standard deviations around these onset and offset times.

Searchlights across time and channels

In the interest of completeness, we report the results of the searchlights across time and channels but, considering the limited spatial resolution of EEG, our discussion will focus on the analyses in the time domain. Above-chance decoding of animacy was driven by electrodes located in occipital and right centro-parietal areas during the earliest stages of speech production planning (0-100 ms post picture onset) and by all the electrodes considered in the analysis in subsequent time windows (Figure 3A). During the earliest stages of speech production planning (0-100 ms after picture onset), the classification of lexical stress was driven by one left frontal and several left occipito-parietal electrodes (Figure 3B). Between 100 and 300 ms post stimulus, classification performance was driven by electrodes located bilaterally in occipito-parietal and frontal areas. Between 300-500 ms after picture onset, classification of lexical stress was driven by the vast majority of the electrodes considered in the analysis. During the earliest stages of speech production planning (0-100 ms post stimulus), classification of syllable structure was driven by an electrode located in right fronto-temporal areas (Figure 3C). At later time windows, classification performance was driven by electrodes located in centro-parietal and fronto-temporal regions. Above-chance decoding of manner of articulation was driven by an electrode located in occipital areas between 0 and 100 ms after picture onset, by electrodes located in occipito-parietal areas between 100 and 300 ms post stimulus, and by electrodes located in occipito-parietal and left fronto-temporal areas between 300 and 500 ms after picture onset (Figure 3D).

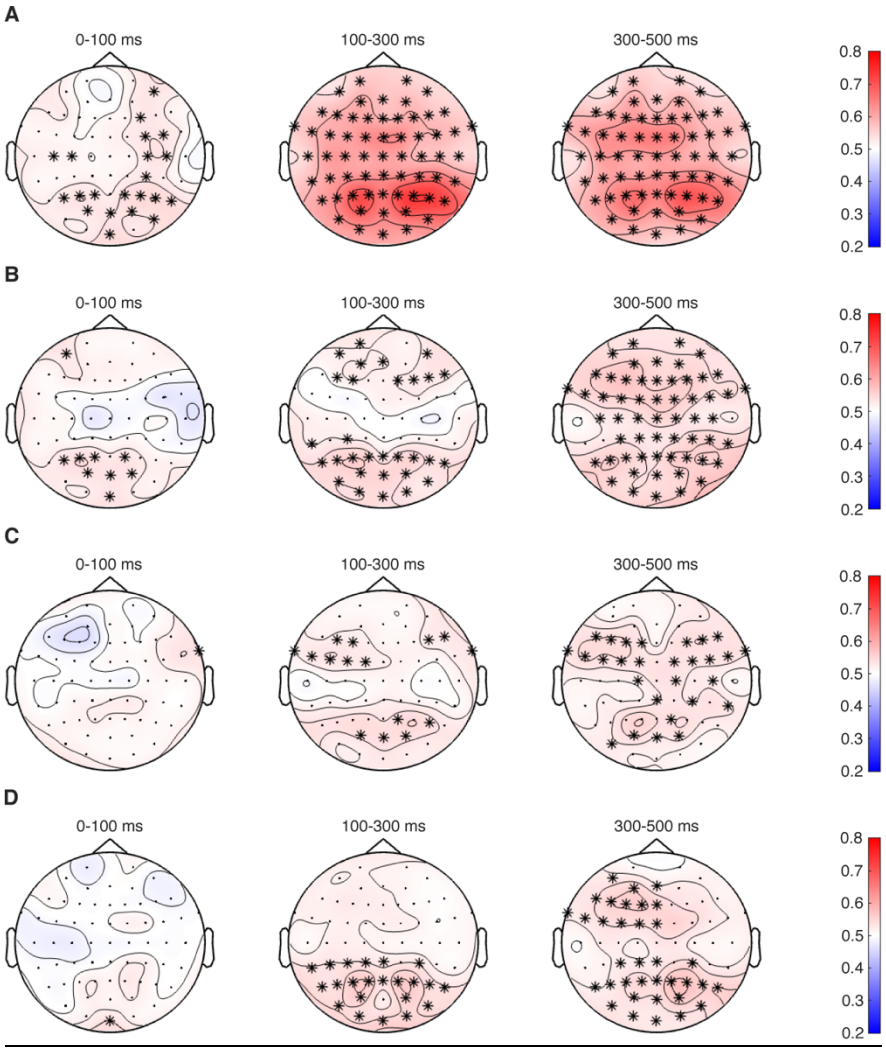

Figure 3: Topographic maps displaying the electrodes driving decoding accuracy for the experimental variables (A) animacy, (B) lexical stress, (C) syllable structure, and (D) manner of articulation across three time windows (0-100 ms, 100-300 ms, and 300-500 ms after picture onset). The colour bar indicates decoding accuracy (red: above-chance, blue: below-chance). Statistically significant points are marked with an asterisk.

Time x time generalisations

During the earliest stages of speech production planning (0-150 ms post stimulus), the temporal generalisation matrix of the animacy classifier (Figure 4A) was extremely narrow, suggesting the neural code supporting above-chance decoding in this time window is dynamically evolving (King & Dahene, 2014). Between 150 and 900 ms after picture onset, the animacy classifier generalised across

time windows of approximately 300 ms, suggesting the neural representation of animacy stabilises around the same time as decoding accuracies showed a steep increase (Figure 2A). From 900 ms post stimulus onwards, the temporal generalisation matrix revealed widespread generalisation across a time window approximately twice as large as the ones observed between 150 and 900 ms post picture onset. The transition from diagonal to sustained indicates that the neural code supporting above-chance classification during this time window is different from the one observed between 150 and 900 ms after picture onset. This pattern appears to be confirmed by the lower decoding accuracies observed from 900 ms after picture onset onwards (Figure 2A).

The temporal generalisation of lexical stress revealed three distinct areas of generalisation (Figure 4B). The first square matrix emerged between 200 and 600 ms after picture onset, suggesting that the neural code supporting above-chance decoding during this time window is stable. The second generalisation window spanned from 400 to 800 ms post stimulus, partially overlapping with the first generalisation window, but extending further in time. This transitional phase marks a shift in the neural representation of lexical stress, as corroborated by the gradual increase in decoding accuracies starting at 400 ms after picture onset (Figure 2B). This shift may reflect the transition from speech planning to the execution of actual speech, with the earliest speech onset occurring 396 ms post picture onset. The third and most extensive generalisation window occurred between 1000 and 1500 ms post picture onset.

The temporal generalisation of syllable structure revealed two distinct patterns of neural activity (Figure 4C). Between 200 and 600 ms after picture onset, the generalisation matrix was predominantly diagonal, suggesting that the neural code supporting above-chance decoding during this time window was dynamically evolving. At 600 ms post picture onset, the generalisation matrix transitioned from diagonal to sustained. This shift coincides with the average speech onset and may reflect the transition from the neural processes associated with syllabification to those supporting overt speech production.

346 Finally, an exploratory analysis showed that the classifier decoding manner of articulation of
347 the first phoneme generalised significantly above chance from 300 ms after picture onset (Figure 4D).
348 The generalisation matrix was predominantly diagonal, suggesting that the neural code supporting
349 above-chance decoding in this time window was dynamically evolving.

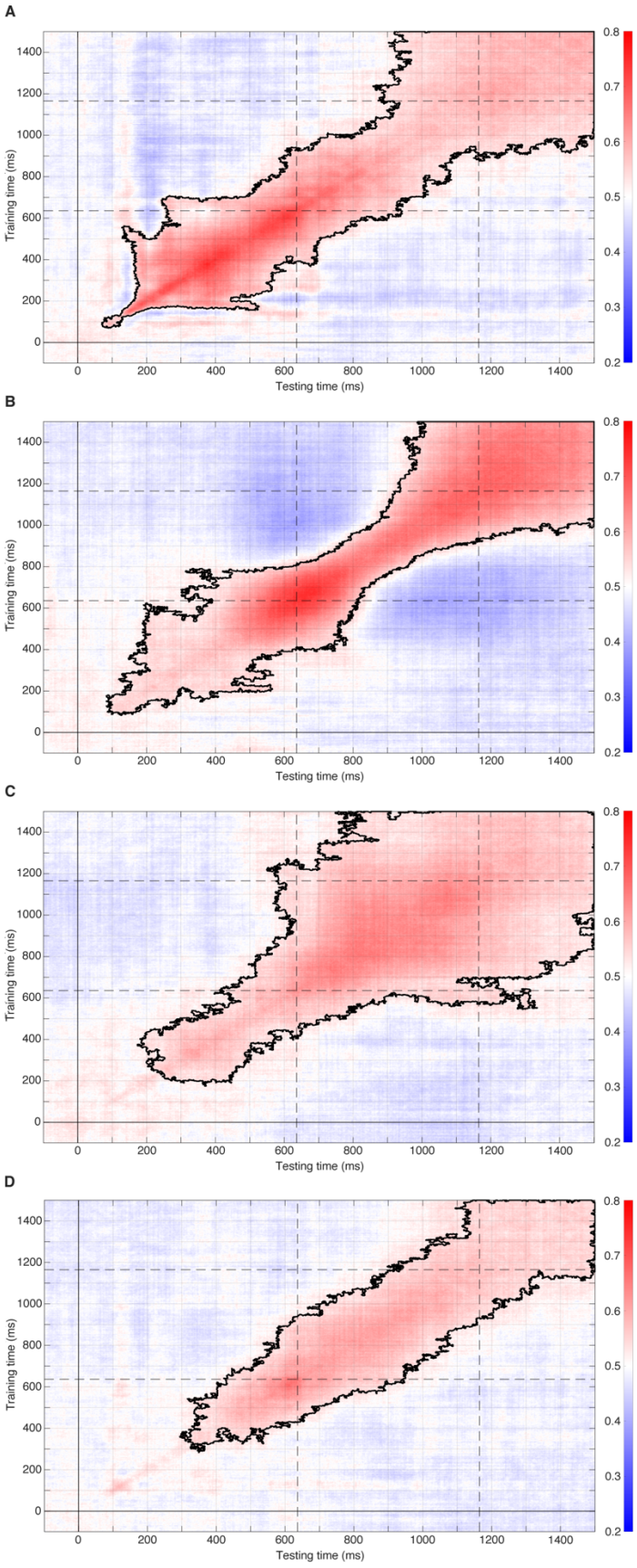

Figure 4: Generalisations across time for the experimental variables (A) animacy, (B) lexical stress, (C) syllable structure, and (D) manner of articulation. Vertical lines indicate picture onset. Dashed

lines indicate the average onset and offset of speech. The color bar indicates decoding accuracy (red: above-chance, blue: below-chance). Accuracy values are masked by statistical significance (significant decoding accuracies within the bold line). Temporal generalisation matrices represent the decoding performance of a classifier trained at one time point of the trial and later tested at all the other time points of the trial. Square generalisation matrices indicated that the neural code supporting above-chance decoding is stable (i.e., the neural representations are sustained over time), while diagonal matrices indicate that the classifier can generalise only over a transient period of time, suggesting the neural representations are dynamically evolving (King & Dehaene, 2014).

Discussion

We investigated the temporal dynamics of lexical retrieval in picture naming using multivariate pattern analysis (MVPA) of EEG data. We had three main aims. The first and second were to examine the application of EEG-based MVPA for speech production research and to use this method to compare the time course of conceptual and phonological processing. Our third and most important goal was to use EEG-MVPA to investigate the timing of two processes within the phonological domain: the retrieval of a word's stress pattern and syllable structure. Conceptual preparation for picture naming was indexed by an animacy distinction, while the phonological processes of metrical encoding and syllabification were linked to the manipulation of lexical stress and syllable (CV) structure, respectively. Inspired by recent work, we also conducted an exploratory analysis decoding the phonetic characteristics of the word-initial phoneme.

Concerning the first goal, our results clearly demonstrate that time-resolved MVPA of EEG data can be used to obtain information about the time course of lexical access in speech planning. This is important because purely behavioural methods do not suffice to study the time course of the fast processes occurring during speech planning, and because researchers generally have easier access to EEG than MEG facilities. Thus, we see EEG-MVPA as a promising tool for language production research

and an exciting addition to the methodological repertoire available in the field (van der Burght et al., 2023).

Our second goal was to compare the time course of conceptual and phonological processing during word production planning. Comparing the decoding accuracy of the conceptual to both phonological variables (metrical stress and syllable structure), the results showed that conceptual preceded phonological processing. Specifically, classification of animacy demonstrated two early peaks, first around 75-100 ms after picture onset and then again at 150 ms post stimulus onset. Further support for this two-stage process comes from the temporal generalisation matrix of the animacy classifier. Temporal generalisation matrices represent the decoding performance of a classifier trained at one time point of the trial and later tested at all the other time points of the trial. Square areas of above-chance decoding accuracy indicate that the neural code supporting above-chance decoding is stable (i.e., the neural representations are sustained over time). In contrast, diagonal matrices indicate that the classifier can generalise only over a transient period of time, suggesting that the neural representations are dynamically evolving (King & Dehaene, 2014). In the animacy classifier, generalisability became statistically significant between 75 and 150 ms after picture onset. During this period, however, the classifier could not generalise to other time points of the speech preparation time course (i.e., the matrix is diagonal), suggesting that classification performance was driven by multiple neural representations. At around 150 ms after picture onset, the temporal generalisation matrix sharply transitions from diagonal to sustained, likely reflecting the selection of the specific lexical item associated with the input. In sum, the processing animacy occurred dynamically, with its onset preceding the earliest decodability of either phonological variable.

Our third goal was to examine the relative time course of metrical encoding and syllabification. As explained in the Introduction, Levelt and colleagues (1999) proposed that syllabification should follow metrical retrieval. We found that the decodability of lexical stress became statistically significant at around 160 ms, shortly after the decodability of conceptual information but before the

decodability of syllable structure at approximately 250 ms. In the temporal generalisation matrix, the decodability of lexical stress transitioned from diagonal to sustained around the time the decodability of syllable structure became statistically significant (i.e., around 200 ms after picture onset). This result indicates that the neural representation supporting metrical encoding stabilises around the onset of above-chance decoding of the structure of the first syllable, suggesting that information regarding a word's lexical stress is available before the onset of syllabification. The results support the notion that metrical encoding precedes syllabification as predicted by Levelt et al. The time course of metrical encoding and syllabification had, to our knowledge, only been directly assessed in an earlier event-related potential (ERP) study using a complex, metalinguistic judgement task (Schiller et al., 2003). The authors found no evidence for sequentiality of metrical encoding and syllabification, but, as they note, their paradigm might not have been sensitive enough to detect small time differences in the onset of the two processes. As a further explanation for the discrepancy between the Schiller et al. results and ours, time-resolved MVPA may offer increased sensitivity as compared to ERPs (Carlson et al., 2019).

Our results invite further research along a number of lines. For instance, the two stress patterns used in the current study differed in frequency: the majority of Dutch disyllabic nouns have word-initial stress, and so, effectively, the stress manipulation of 1st vs 2nd syllable stress simultaneously varied regular (default) vs. irregular stress patterns. It is therefore unclear whether the decoder decoded the actual stress patterns or frequent vs. infrequent. Alternatively, decoding accuracy could reflect discrimination between a stress pattern that can be derived by rule (as the default pattern) vs. one that needs to be stored lexically (for a discussion of the storage vs. computation of Dutch stress patterns, see Schiller 2006).

Finally, inspired by recent results (e.g., Carota et al., 2023; Fairs et al., 2021; Strijkers et al., 2017), we also conducted exploratory analyses to examine whether we could decode phonetic properties of the initial phoneme of the target words (plosive vs. fricative). We found that the decodability of word-initial phonetic information peaked at around 125 ms after picture onset, returned to chance-level accuracy until 420 ms post stimulus, and then peaked again around the

average speech onset. The late effect can be easily explained within cascaded models of lexical access, which predict that phonetic information should become available only after phonological encoding (e.g., Indefrey, 2011). The early effect, however, does not sit well with these models. Yet, similar patterns of results have recently been reported and have been seen as supporting "word web" models (e.g., Kerr et al., 2023; Pickering & Strijkers, 2024; Strijkers & Costa, 2016). Specifically, a previous MEG study showed that word frequency and the place of articulation of the word-initial phoneme (labial vs. coronal) elicited stimulus-specific fronto-temporal activity between 160 and 240 ms after picture onset (Strijkers et al., 2017). Similarly, Fairs et al. (2021) found that lexical and phonetic variables (lexical frequency and phonotactic frequency, respectively) elicited early, parallel ERP modulations. In an exploratory analysis, Carota et al. (2023) showed that manner of articulation could be decoded from MEG source data in early time windows.

However, note that our results concerning the early decoding of phonetic information stem from post hoc analyses and therefore need to be replicated before any firm conclusions are drawn. There are two additional reasons that warrant a replication of this result. First, differently from the planned contrasts, the significant classification accuracy of the phonemic contrast was not corroborated by a significant temporal generalisation matrix during the earliest stages of speech production planning, further necessitating a replication attempt (although we note that the early decodability of manner of articulation was rather short, approximately 50 ms, and the training dataset was smaller, 240 trials vs. 384 in the other classifiers, suggesting the early effect observed in the time-resolved MVPA might simply be too weak or short-lived to form a statistically significant cluster in the time x time plot; Maris & Oostenveld, 2007). Second, to ensure that naming was accurate and to achieve sufficient power, our training and experiment involved a considerable number of repetitions of each target. It is conceivable that after a certain number of repetitions the response was automatically evoked, and this may have contributed to the early phonetic effect.

To summarise, our results suggest that syllabification is initiated after the onset of conceptual and metrical encoding. Both cascading and parallel activation models can account for this result. As

discussed in the Introduction, Levelt et al. argued that syllabification of a string of segments is a context-dependent operation and is therefore considered to be computed online after metrical retrieval (i.e., not stored in the mental lexicon). Parallel activations models, too, suggest that context-dependent processes are computed after the initial activation of lexical information (Pickering & Strijkers, 2024). Our exploratory analysis, on the other hand, tentatively suggests that some phonetic information about picture names may become activated rapidly after picture onset. This result, while in need of replication, hints at the possibility that conceptual and phonetic information might be accessed in parallel during the initial activation of lexical information as suggested by parallel activations models (e.g., Pickering & Strijkers, 2024; Strijkers & Costa, 2016).

In conclusion, our findings demonstrate that EEG-MVPA can decode distinct conceptual and phonological processes during speech production planning. Our results showed that, following conceptual information (i.e., animacy), a word's stress pattern can be retrieved before a complete representation of the syllable structure is computed. Finally, we obtained preliminary evidence suggesting that some phonetic information about picture names may become rapidly available, in parallel with conceptual information.

References

Alderete, J., & O'Séaghdha, P. G. (2022). Language generality in phonological encoding: moving beyond Indo-European languages. *Language and Linguistics Compass, 16*(7), e12469.

Ashton, K., Zinszer, B. D., Cichy, R. M., Nelson III, C. A., Aslin, R. N., & Bayet, L. (2022). Time-resolved multivariate pattern analysis of infant EEG data: a practical tutorial. *Developmental Cognitive Neuroscience, 54*, 101094.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: keep it maximal. *Journal of Memory and Language, 68*(3), 255-278.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1-48.

Boersma, P., & Weenink, D. (2024). Praat: doing phonetics by computer [Computer program]. Retrieved from www.praat.org.

Carlson, T. A., Grootswagers, T., & Robinson, A. K. (2020). An introduction to time-resolved decoding analysis for M/EEG. In D. Poeppel, G. R. Mangun, & M. S. Gazzaniga (Eds.), The Cognitive Neurosciences (6th ed., pp. 679-690). MIT Press.

Carota, F., Schoffelen, J. M., Oostenveld, R., & Indefrey, P. (2022). The time course of language production as revealed by pattern classification of MEG sensor data. *Journal of Neuroscience, 42*(29), 5745-5754.

Carota, F., Schoffelen, J. M., Oostenveld, R., & Indefrey, P. (2023). Parallel or sequential? Decoding conceptual and phonological/phonetic information from MEG signals during language production. *Cognitive Neuropsychology, 40*(5-6), 298-317.

Cholin, J., Levelt, W. J., & Schiller, N. O. (2006). Effects of syllable frequency in speech production. *Cognition, 99*(2), 205-235.

Corps, R. E., & Meyer, A. S. (2025). Multiple repetitions lead to the long-term elimination of the word frequency effect. Journal of Experimental Psychology: Learning, Memory, and Cognition. Advance online publication.

Fairs, A., Michelas, A., Dufour, S., & Strijkers, K. (2021). The same ultra-rapid parallel brain dynamics underpin the production and perception of speech. *Cerebral Cortex Communications, 2*, 1-17.

Feng, C., Damian, M. F., & Qu, Q. (2021). Parallel processing of semantics and phonology in spoken production: evidence from blocked cyclic picture naming and EEG. *Journal of Cognitive Neuroscience, 33*(4), 725-738.

Goldrick, M., & Rapp, B. (2007). Lexical and post-lexical phonological representations in spoken production. *Cognition, 102*(2), 219-260.

Hultén, A., Vihla, M., Laine, M., & Salmelin, R. (2009). Accessing newly learned names and meanings in the native language. *Human Brain Mapping, 30*(3), 976-989.

King, J.-R., & Dehaene, S. (2014). Characterizing the dynamics of mental representations: the temporal generalization method. *Trends in Cognitive Sciences, 18*(4), 203-210.

Indefrey, P. (2011). The spatial and temporal signatures of word production components: a critical update. *Frontiers in Psychology, 2*, 255.

Indefrey, P. (2016). On putative shortcomings and dangerous future avenues: response to Strijkers & Costa. *Language, Cognition and Neuroscience, 31*(4), 517-520.

Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition, 92*(1-2), 101-144.

Kerr, E., Ivanova, B., & Strijkers, K. (2023). Lexical access in speech production: psycho- and neurolinguistics perspectives on the spatiotemporal dynamics. In R. Hartsuiker & K. Strijkers (Eds.), *Current Issues in the Psychology of Language: Language Production* (pp. 32-65). Routledge.

Keuleers, E., Brysbaert, M., & New, B. (2010). SUBTLEX-NL: a new measure for Dutch word frequency based on film subtitles. *Behavior Research Methods, 42*(3), 643-650.

Laganaro, M., Morand, S., & Schnider, A. (2009). Time course of evoked-potential changes in different forms of anomia in aphasia. *Journal of Cognitive Neuroscience, 21*(8), 1499-1510.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*(1), 1-38.

Maess, B., Friederici, A. D., Damian, M., Meyer, A. S., & Levelt, W. J. M. (2002). Semantic category interference in overt picture naming: sharpening current density localization by PCA. *Journal of Cognitive Neuroscience, 14*(3), 455-462.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG and MEG-data. *Journal of Neuroscience Methods, 164*(1), 177-190.

Miozzo, M., Pulvermüller, F., & Hauk, O. (2015). Early parallel activation of semantics and phonology in picture naming: evidence from a multiple linear regression MEG study. *Cerebral Cortex, 25*, 3343-3355.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience, 2011*, 156869.

Pickering, M. J., & Strijkers, K. (2024). Language production and prediction in a parallel activation model. *Topics in Cognitive Science*, 1-12.

R Core Team. (2024). R: A Language and Environment for Statistical Computing [Computer software]. Retrieved from www.r-project.org

Riès, S. K., Dhillon, R. K., Clarke, A., King-Stephens, D., Laxer, K. D., Weber, P. B., Kuperman, R. A., Auguste, K. I., Brunner, P., Schalk, G., Lin, J. J., Parvizi, J., Crone, N. E., Dronkers, N. F., & Knight, R.T. (2017). Spatiotemporal dynamics of word retrieval in speech production revealed by cortical high-frequency band activity. *Proceedings of the National Academy of Sciences, 114*, E4530-E4538.

Sahin, N. T., Pinker, S., Cash, S. S., Schomer, D., & Halgren, E. (2009). Sequential processing of lexical, grammatical, and phonological information within Broca's area. *Science, 326*(5951), 445-449.

Sá-Leite, A. R., Haro, J., Comesaña, M., & Fraga, I. (2021). Of beavers and tables: the role of animacy in the processing of grammatical gender within a picture-word interference task. *Frontiers in Psychology, 12*, 661175.

Schiller, N. O. (2006). Lexical stress encoding in single word production estimated by event-related brain potentials. *Brain Research, 1112*(1), 201-212.

Schiller, N. O., Fikkert, P., & Levelt, C. C. (2004). Stress priming in picture naming: an SOA study. *Brain and Language, 90*(1-3), 231-240.

Schiller, N. O., Bles, M., & Jansma, B. M. (2003). Tracking the time course of phonological encoding in speech production: an event-related brain potential study. *Cognitive Brain Research, 17*(3), 819-831.

Strijkers, K., & Costa, A. (2016). The cortical dynamics of speaking: present shortcomings and future avenues. *Language, Cognition and Neuroscience, 31*(4), 484-503.

Strijkers, K., & Costa, A. (2011). Riding the lexical speedway: a critical review on the time course of lexical selection in speech production. *Frontiers in Psychology, 2*, 356.

Strijkers, K., Costa, A., & Thierry, G. (2010). Tracking lexical access in speech production: electrophysiological correlates of word frequency and cognate effects. *Cerebral Cortex, 20*(4), 912-928.

Strijkers, K., Costa, A., & Pulvermüller, F. (2017). The cortical dynamics of speaking: lexical and phonological knowledge simultaneously recruit the frontal and temporal cortex within 200 ms. *NeuroImage, 163*, 206-219.

Tanner, D., Morgan-Short, K., & Luck, S. J. (2015). How inappropriate highpass filters can produce artifactual effects and incorrect conclusions in ERP studies of language and cognition. *Psychophysiology, 52*(8), 997-1009.

The MathWorks Inc. (2024). MATLAB version 24.2.0 (R2024b) [Computer software]. Retrieved from www.mathworks.com

Treder, M. S. (2020). MVPA-Light: a classification and regression toolbox for multi-dimensional data. *Frontiers in Neuroscience, 14*, 289.

van der Burght, C. L., Friederici, A. D., Maran, M., Papitto, G., Pyatigorskaya, E., Schroën, J. A. M., Trettenbrein, P. C., & Zaccarella, E. (2023). Cleaning up the brickyard: how theory and methodology shape experiments in cognitive neuroscience of language. *Journal of Cognitive Neuroscience, 35*(12), 1-22.

van Turennout, M., Hagoort, P., & Brown, C. M. (1997). Electrophysiological evidence on the time course of semantic and phonological processes in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*(4), 787-806.

van Turennout, M., Hagoort, P., & Brown, C. M. (1998). Brain activity during speaking: from syntax to phonology in 40 milliseconds. *Science, 280*(5363), 572-574.

van Turennout, M., Bielamowicz, L., & Martin, A. (2003). Modulation of neural activity during object naming: effects of time and practice. *Cerebral Cortex, 13*(4), 381-391.

586 Table S1: Fixed effects from generalised linear mixed effects model of response accuracy with manner

587 of articulation as a fixed effect

Term	Estimate	*SE*	*z*-value
Intercept	2.790	0.219	12.772
Animacy	-0.157	0.191	-0.824
Lexical stress	-0.423	0.171	-2.480
Syllable structure	-0.217	0.174	-1.251
Manner of articulation	0.275	0.177	1.555

588

589 Table S2: Fixed effects from linear mixed effects models of response latencies with manner of

590 articulation as a fixed effect

Term	Estimate	*SE*	*t*-value
Intercept	-0.480	0.038	-12.550
Animacy	0.048	0.026	1.826
Lexical stress	0.019	0.026	0.720
Syllable structure	0.011	0.025	0.430
Manner of articulation	-0.079	0.027	-2.916

591