The Same Ultra-Rapid Parallel Brain Dynamics Underpin the Production and Perception of Speech

Amie Fairs, Amandine Michelas, Sophie Dufour, Kristof Strijkers

HAL Id: hal-03317758
https://hal.science/hal-03317758

Submitted on 7 Aug 2021

ORIGINAL ARTICLE

The Same Ultra-Rapid Parallel Brain Dynamics Underpin the Production and Perception of Speech

Amie Fairs, Amandine Michelas, Sophie Dufour and Kristof Strijkers

Aix-Marseille University & CNRS, LPL, 13100 Aix-en-Provence, France

Address correspondence to Kristof Strijkers, Centre de la Recherche Scientifique (CNRS), Aix-Marseille Université (AMU), Laboratoire Parole et Langage (LPL) & Institute for Language, Communication and the Brain (ILCB), 5 Avenue Pasteur, 13100 Aix-en-Provence, France. Email: kristof.strijkers@univ-amu.fr

Abstract

The temporal dynamics by which linguistic information becomes available is one of the key properties to understand how language is organized in the brain. An unresolved debate between different brain language models is whether words, the building blocks of language, are activated in a sequential or parallel manner. In this study, we approached this issue from a novel perspective by directly comparing the time course of word component activation in speech production versus perception. In an overt object naming task and a passive listening task, we analyzed with mixed linear models at the single-trial level the event-related brain potentials elicited by the same lexico-semantic and phonological word knowledge in the two language modalities. Results revealed that both word components manifested simultaneously as early as 75 ms after stimulus onset in production and perception; differences between the language modalities only became apparent after 300 ms of processing. The data provide evidence for ultra-rapid parallel dynamics of language processing and are interpreted within a neural assembly framework where words recruit the same integrated cell assemblies across production and perception. These word assemblies ignite early on in parallel and only later on reverberate in a behavior-specific manner.

Key words: EEG, language perception, language production, parallel processing, word assemblies

Introduction

Language behavior concerns a complex, multi-component process where different linguistic representations (semantic, lexical, syntactic, and phonological knowledge) need to be retrieved and merged together in order to produce and perceive communicative signals. Insights into "when" and "how" different linguistic knowledge becomes activated is therefore key to understanding the cortical mechanics that can sustain this unique human behavior (Pulvermüller 1999, 2018; Friederici 2002, 2011; Indefrey and Levelt 2004; Hauk 2016). Despite the time course of language processing being a longstanding question in the field, debate remains between proponents of more sequential dynamics versus those arguing for parallel retrieval of linguistic knowledge.

At least two main reasons exist for this debate: first, in terms of data, empirical evidence has been found for both sequential and parallel brain dynamics; second, it is often difficult to disentangle whether neurophysiological responses to linguistic knowledge support a sequential or parallel time course. This is in part due to the fact that language processing recruits a vast number of regions in the brain, making it hard to know whether temporal differences reflect functionally distinct mental operations or merely physical distance in the brain (and different axonal conductance velocities) but functionally parallel activations. Interestingly though, the debate concerning the temporal dynamics in brain language models exists both for speech production and speech perception, even though traditionally these modalities have been studied separately (Price 2012). This shared

debate (but separated approach) is interesting because directly contrasting the time course of language processing between the production and perception of words could circumvent the above-mentioned conceptual problems and provide novel and more explicit insights into how our brain computes language in time. This is because a sequential model predicts the reverse temporal dynamics between the production and perception of words, while a parallel model predicts the same, simultaneous onset of word components across the language modalities. Therefore, in the current study, we systematically contrasted the event-related brain potentials (ERPs) elicited at the single-trial level by lexico-semantic and phonological word knowledge in both production and perception.

In sequential models of word production (Levelt et al. 1999; Indefrey and Levelt 2004; Indefrey 2011; Hickok 2012) and perception (Friederici 2002, 2011; Scott and Johnsrude 2003; Hickok and Poeppel 2007), the idea is that different word components are functionally separated in time (and space): In production, first the semantic and lexical information of to-be-uttered words are retrieved, followed by the phonological code, which is then translated into a motor program to trigger the vocal folds for articulation. In perception, incoming acoustic speech signals are phonologically encoded and subsequently activate the lexical and semantic knowledge associated with those phonological forms. Though sequential models differ in how many processing stages are necessary to go from concept to speech or speech to concept and can incorporate cascading and interactivity between processing layers (Dell 1986; Caramazza 1997; Friederici 2011) rather than being strictly serial (Levelt et al. 1999; Friederici 2002; Indefrey and Levelt 2004), they do share the common feature that upstream processes should initiate prior to downstream processes within the hierarchical architecture (Dell and O'Seaghdha 1992; Indefrey and Levelt 2004; Friederici 2011). Neurophysiological evidence supporting such sequential dynamics show that in speech production lexico-semantic effects emerge earlier in time (roughly around 200–250 ms) than phonological and phonetic effects (roughly around 300–400 ms) during speech production tasks (Salmelin et al. 1994; Levelt et al. 1998; Van Turennout et al. 1998; Vihla et al. 2006; Laganaro et al. 2009; Sahin et al. 2009; Schuhmann et al. 2012; Valente et al. 2014; Dubarry et al. 2017). Similar sequential-like neurophysiological responses have been found in speech perception (but in the reverse direction), with phonetic and phonological effects emerging roughly within the first 200 ms of processing and lexico-semantic effects manifesting roughly around 300–400 ms after spoken word onset (Holcomb and Neville 1990; Friederici et al. 1993; Van Petten et al. 1999; Hagoort and Brown 2000; Halgren et al. 2002; O'Rourke and Holcomb 2002; Dufour et al. 2013; Winsler et al. 2018).

The above temporal evidence paints a mirror image between language behaviors where cortical activation of different word components seems to progress in functional delays of some 100 ms and with reversed order between speech production and perception. However, this temporal segregation between the retrieval of meaning and sounds for words has received criticisms both in language perception (Pulvermüller et al. 2009) and production (Strijkers and Costa 2011), and inconsistent evidence has been presented in both domains of language. In language perception, electrophysiological and neuromagnetic data have demonstrated that semantic, lexical, and phonological word properties can emerge very rapidly (within 200 ms of processing) and near-simultaneously (Pulvermüller et al. 2005; Näätänen et al. 2007; MacGregor et al. 2012; Shtyrov et al. 2014). Similar rapid brain indices for extracting the meaning and sounds of words have been observed in language production

tasks (Strijkers et al. 2010, 2017; Miozzo et al. 2015; Riès et al. 2017; Janssen et al. 2020; Feng et al. 2021). These results challenge the traditional sequential view on the temporal dynamics of language processing, and instead are better captured by those brain language models, which assume words are represented as unified cell assemblies where meaning and sounds ignite rapidly in parallel, both when perceiving (Pulvermüller 1999, 2018; Pulvermüller and Fadiga 2010) and producing words (Strijkers 2016; Strijkers and Costa 2016).

According to the above "Integration Models" (IMs), the neurophysiological data demonstrating sequential dynamics do not concern the first-pass (ignition) activation of language representations. Rather it is assumed to reflect second-pass (reverberatory) activation associated with linguistic operations upon the parallel retrieved words such as verbal working memory, semantic integration, grammatical inflection, and articulatory control (Pulvermüller and Fadiga 2010; Strijkers and Costa 2016; Schomers et al. 2017; Strijkers et al. 2017). In other words, IMs predict (at least) two functionally distinct time courses underlying word processing: Early on the explosion like activation (ignition) of words as a whole indexing target recognition, and afterwards more sequential reverberations linked to task- and context-specific processing. For example, in an influential intracranial study of (written) word production, Sahin et al. (2009) observed that depth electrodes in Broca's region displayed a serial response to the lexical (word frequency) and phonological (phonological inflection of plurals) information during word production planning. The authors interpreted the result as supporting sequential models of language production where lexical information is available well before phonological information. However, within the framework of IMs another interpretation is possible, namely that the later phonological effect did not reflect word form activation per se, but rather reverberatory activation to phonologically inflect the earlier activated target word (i.e., add an "s" to the activated noun form in order to correctly articulate a plural). By incorporating both parallel brain dynamics linked to the neural activation of a word and sequential brain dynamics linked to task- and context-specific linguistic operations upon that word representation, IMs can offer a unified account for the parallel versus sequential time courses as observed in response to language processing in the literature.

However, proponents of sequential brain language models remain skeptical whether such early parallel activation of lexical and phonological word representations is plausible. One frequently uttered reservation concerns the possibility that these early effects highlight sensorial and/or attentional differences between stimuli-sets rather than linguistic brain activity (Hagoort 2008; Mahon and Caramazza 2008; Indefrey 2016). Indeed, several of the studies showing that both lexical and phonological brain correlates are found within the first 200 ms of processing relied on physically different stimuli-sets (Strijkers et al. 2010, 2017; MacGregor et al. 2012; Miozzo et al. 2015). Another criticism is whether "fast" equals "parallel" brain dynamics (Brehm and Goldrick 2016; Mahon and Navarrete 2016). Recent neuromagnetic data taken to support parallel activation of lexical and phonological knowledge in speech production exemplifies this issue: In an MRI-constrained MEG study of overt object naming, Strijkers et al. (2017) observed that both word frequency, taken as a metric for lexico-semantic processing, and the place of articulation of the word-initial phoneme (i.e., labial versus coronal), taken as a metric for phonological and phonetic encoding, elicited stimuli-specific fronto-temporal source activity between 160 and 240 ms after

stimulus onset. While the authors argued these data pattern best-fitted parallel dynamics, the data may still be explained within a sequential framework where functional delays between linguistic components are less than the traditional assumed hundreds of ms (in this example: the lexico-semantic effects occurring in the beginning of the 80-ms time window and the phonological effects emerging near the end of the 80-ms time window). In other words, finding early onsets for distinct word processing components, though requiring modifications of the classical temporal assumptions in sequential models, does not necessarily exclude sequential processing dynamics per se.

To address these criticisms, in the current study we explored the time course of a lexico-semantic and a phonological variable both in speech production and perception. That is, we directly contrast the same brains, stimuli and, importantly, psycholinguistic manipulations across the two language modalities. Including "language modality" (i.e., production vs. perception) as variable to assess the cortical activation time course of different linguistic components has a marked advantage compared to prior spatiotemporal research in word production and perception done in isolation: It allows us to assess time course from a different perspective, which removes both the problem of physical variance and the conceptual problem of whether "fast" equals "parallel." This is because the relevant dimension for testing the hypotheses now becomes the "relative" time course of the sequence of events between the language modalities, rather than the exact, absolute value of when a linguistic component becomes active within each behavior. This feature of our design is crucial in order to reliably compare two different language modalities, which rely on two different inputs (auditory vs. visual): 1) by contrasting "effects" between modalities (and thus substracting away from the visual or auditory input), rather than modalities per se, we reduce physical variance between the modalities (the feasibility of this logic has already been demonstrated: Strijkers et al. 2011, 2015); 2) the relevant contrast for testing the hypotheses in this experiment between sequential and parallel dynamics is whether the two manipulated word components (lexico-semantics and phonology) display the same or a different temporal structure between modalities, regardless of when that temporal structure emerges within each modality. For example, if the "parallel" or "sequential" effects manifest earlier/later for one modality compared to the other, it would not alter the conclusion that the data support a "parallel" or "sequential" processing dynamic (instead, it would merely demonstrate that this "parallel or 'sequential' dynamic manifests earlier/later depending on type of input). In sum, regardless of whether a lexico-semantic and phonological manipulation both elicit fast responses in each language modality or not, what matters is if the time course associated with each of the manipulations is different in production and perception. In a sequential model it should be since lexico-semantic and phonological activation is functionally separated in the 'reverse' order between production and perception. In contrast, in IMs it should not since in both modalities the word, with its lexico-semantic and phonological information, ignites as a whole in parallel (see Fig. 1).

In order to tap into lexico-semantic and phonological processing, we manipulated two well-known variables, which have been successfully applied both in language production and perception. As a metric for lexico-semantic processing, we manipulated word frequency (see Fig. 1). The word frequency effect refers to the observation that we are faster in processing words we are well acquainted with (e.g., dog) compared to rarer words (e.g., owl), and has been taken to reflect the speed of lexical activation in

both language production (Caramazza et al. 2001; Graves et al. 2007; Strijkers et al. 2010, 2011, 2017) and perception (Sereno et al. 1998; Dahan et al. 2001; Dufour et al. 2013; Winsler et al. 2018). As a metric for phonological and phonetic processing, we manipulated word phonotactic probability, which refers to how often phonemes and sequences of phonemes, biphones, of a word co-occur[1] (see Fig. 1). Both in production (Levelt and Wheeldon 1994; Vitevitch et al. 2004; Cholin et al. 2006; Laganaro and Alario 2006) and perception (Vitevitch and Luce 1998; Vitevitch et al. 1999) it has been found that words with high phonotactic frequency (HPF) (e.g., horse) are processed more quickly than words with low phonotactic frequency (LPF) (e.g., glasses); an effect attributed to the speed of word form and/or phonetic processing in both language modalities. This conclusion is further supported by neurophysiological investigations of phonotactic frequency effects displaying differences in a time frame consistent with phonological encoding (Pylkkänen et al. 2002; MacGregor et al. 2012; Dufour et al. 2013; Hunter 2013; Bürki et al. 2015; den Hollander et al. 2019; Di Liberto et al. 2019).

In order to explore, acorss language modalities, the time course elicited by these lexico-semantic and phonological variables, respectively, we embedded them in a well-known production task, object naming, and a well-known perception task, go/no-go semantic categorization (see Fig. 1). These tasks are especially useful here since they have been successfully applied with ERPs to study language production and perception (for reviews see Osterhout and Holcomb 1995; Grainger and Holcomb 2009; Ganushchak et al. 2011), including for the word frequency and phonotactic frequency effects we are targeting in this study (Strijkers et al. 2010, 2015; Bürki et al. 2015; Dufour et al. 2017; Winsler et al. 2019). The rationale we followed was to track the ERP responses elicited by the word versus phonotactic frequency effects, and compare their time course within the same participants performing the object naming and go/no-go semantic categorization tasks (Fig. 1). With such a design, sequential models predict a specific interaction between "language modality" and the "lexico-semantic" and "phonological" manipulations where the "lexico-semantic" variable should elicit earlier ERP responses than the "phonological" variable for the language production task and the opposite ERP pattern (earlier manifestations of the "phonological" manipulation compared to the "lexico-semantic" manipulation) for the language perception task (Fig. 1). In contrast, parallel IMs predict the absence of such interaction since both in the production and perception tasks the ERP responses elicited by the "lexico-semantic" and "phonological" manipulations should emerge simultaneously (Fig. 1).

Methods

In accordance with Open Practices all data (raw and processed), code, analyses pipelines and materials are available in the project OSF repository at https://osf.io/hp2me/.

Participants

26 participants took part in the experiment (mean age = 21, SD = 3, 20 female). All were recruited through Aix-Marseille

---

1 Note that phonotactic frequency and phonotactic probability are often used interchangeably in the literature (Vitevitch et al. 1999); here, we prefer the term phonotactic frequency because our measure focusses on the summed biphone frequency, and not segment probability (a measure sometimes used in combination with biphone frequency for phonotactic probability).
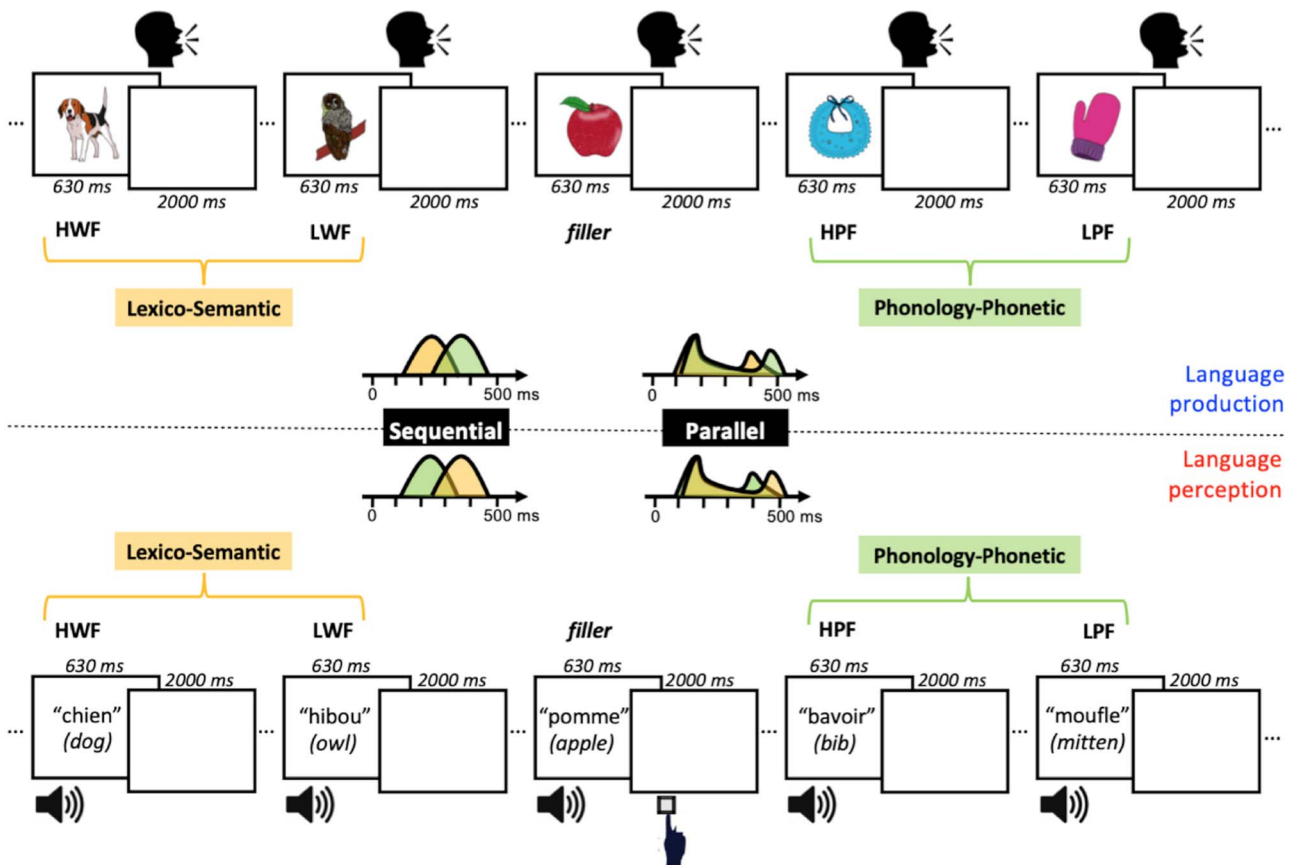
Figure 1. Schematic representation of the experimental design, critical stimuli and hypotheses. In the upper panel, an example stream of trial presentations is depicted for language production, where participants have to overtly name the presented objects as quickly as possible. In the lower panel, the same example stream of trial presentations is depicted for language perception, where participants listen to spoken words and have to press a response button when a spoken word belongs to the semantic category "food" (go-trials only on filler words; 10% of the trials). For both tasks, examples of critical stimuli are depicted, where the contrast between HWF and LWF taps into lexico-semantic processing (yellow), and the contrast between HPF and LPF taps into phonological-phonetic processing (green). Note that in both tasks, the exact same stimuli are used. In the middle, the temporal predictions for these two variables are schematically depicted where according to sequential models lexico-semantic information (yellow) should be available well before phonological-phonetic information (green) in language production, and the reverse temporal pattern is predicted in language perception. In contrast, according to parallel models lexico-semantic and phonological-phonetic word, knowledge should rapidly activate simultaneously both in language production and perception, and only later in the course of processing task-specific differences may emerge between the language modalities.

University and paid €20 for participation. All participants were right handed and reported no hearing, speech or neurological disorders, and had normal or corrected-to-normal vision. Two participants were rejected due to equipment failure, and five participants were rejected due to exclusion of a large number of error trials (defined more below), leaving a sample of 19 participants for analysis. The experiment was granted ethical approval by Aix-Marseille Université "*Comité Protection des Personnes*"—"*Committee for the Protection of Individuals*": CPP 2017-A03614–49.

Design & Materials

There were two within-participant manipulations of interest, tapping into the lexico-semantic and phonological/phonetic levels of linguistic processing (see Fig. 1). The lexical level manipulation was word frequency of the target name (either high frequency or low frequency; mean high log frequency = 1.535 (SD = 0.39), mean low log frequency = 0.5 (SD = 0.34)). The phonological/phonetic manipulation was phonotactic frequency of the target name (i.e., summed biphone frequency of words;

high frequency = 6836.56 (SD = 2111.6), low frequency = 2052.50 (SD = 1070.2); and phoneme frequency: high frequency: 254.138 (SD = 61.4), low frequency: 143.122 (SD = 52.9)). Word frequency and phonotactic frequency values were retrieved from the Lexique (New et al. 2004) database. The word and phonotactic manipulations were orthogonal, with 220 target items overall. There were thus 110 high word frequency (HWF) and 110 low word frequency (LWF) items, and 110 HPF and 110 LPF items (55 items per cell). All stimuli were matched on h-index (number of alternate names given to the image), visual complexity, length of the word in phonemes, bigram frequency, number of phonological neighbors, and duration. Stimuli were significantly different with all P's < 0.001 on the crucial word and phonotactic variables: for word frequency (LPF), $t(54) = -13.85$; for word frequency (HPF), $t(54) = -15.5$; for phonotactic frequency (LWF), biphone: $t(54) = -15.23$, phoneme: $t(54) = -10.24$; and for phonotactic frequency (HWF), biphone: $t(54) = -14.6$, phoneme: $t(54) = -9.98$. (All other t-test results show $|t| < 2$; see the stimuli list on the online repository, https://osf.io/hp2me/, for all stimuli, their values on these measures, and the t-test results).

The same target items were used in both the production and perception task, i.e., the stimuli were identical in both tasks.

Participants carried out picture naming for the production task and an auditory semantic categorization task for the perception task (a go/no-go task) (see Fig. 1). Colored line drawings were retrieved from the MultiPic database (Duñabeitia et al. 2018) and presented in the center of the screen. The auditory stimuli were recorded by a female native French speaker in the soundproof chamber of the *Laboratoire Parole et Langage* of Aix-Marseille University using an RME fireface UC audio interface and a headset Cardioid Condenser Microphone (AKG C520) at a sampling rate of 44 100 Hz. The tokens were segmented and then normalized in intensity to a level of 60 dB using Praat software (Boersma and Weenink 2020). For the semantic categorization task (perception), participants only responded (gave a "go" response) on filler trials (10% of the trials), which were not analyzed (all critical trials had a no-go response). Filler trials consisted of food items (with similar, medium-range values for word and phonotactic frequency), and participants were instructed to press a button if the word they heard was a type of food. There were 25 filler trials leading to 245 trials in each task, with stimuli presented once per task. For the picture naming task (production), participants overtly uttered as quickly and accurately as possible the picture name. Five additional practice trials were present at the beginning of each task, and their data were not recorded.

The stimuli were presented in fully randomized lists per participant and per task. Task order was counterbalanced across participants. The experiment was presented using E-Prime (version 2.0) on a standard computer monitor. Naming responses were collected with a microphone placed in front of the participants for the picture naming task and participants heard the auditory stimuli through non-conductive headphones, and their button press responses on go trials were recorded by a response button box.

Procedure

Participants first were informed about the study and gave informed consent. Electrodes were fitted to participants (see below) before participants began the experiment proper. Participants were familiarized with the stimuli before the experiment began. For production this entailed seeing all the pictures with their correct name written below, and to be matched in exposure, for perception, this entailed hearing all the stimuli names. Participants began with either the production or the perception task (counterbalanced across participants). During each task participants were able to take a short break (around 10s) every 60 trials, and a longer break (a few minutes) between tasks. After the experiment participants were fully debriefed.

Experiment instructions were written presented in white on a black background and centered in the middle of the screen. Production task trials began with a fixation cross (500 ms), followed by a blank screen (500 ms), picture presentation (630 ms), and ended with an inter-trial interval blank screen (2000 ms). Perception task trials followed the same structure and began with a fixation cross (500 ms), followed by a blank screen (500 ms) before auditory stimulus presentation (approximately 630 ms; the average duration of the spoken words), and ended with a blank screen (2000 ms) (see Fig. 1). The whole experiment lasted 1 hour and 45 minutes with 45 minutes for the experiment proper (including breaks).

EEG Acquisition

EEG was recorded with 64 Biosemi active electrodes with standard 10/20 positioning. Four additional Ag/Ag-Cl electrodes were placed for eye movements (one above and one below the left eye to monitor for vertical eye movements and blinks, two electrodes on the outer canthi of each eye to monitor for horizontal eye movements), and two electrodes were placed on the left and right mastoids for referencing. Electrode impedances were kept below 5 microVolts. The EEG signal was continuously recorded at 512 Hz with an online bandpass filter of 0.03 Hz—40 Hz using BrainVision software.

Behavioral Data Preprocessing

Speech latencies were automatically recorded in E-Prime using a voicekey. Each trial was manually checked (online) to determine if an error was made; errors included production of an incorrect target name, hesitations, and other disfluencies (such as sneezes and coughs). All errors were rejected (14%). In the perception task, correct go-trials (filler items) corresponded to more than 90% performance in all participants.

EEG Preprocessing

EEG data, for the production and perception task separately, were preprocessed using FieldTrip (Oostenveld et al. 2011; version 20 190 716) in Matlab (version R2019a). The raw EEG signal was first filtered with a Hamming windowed sinc filter (one pass, zero phase, cutoff -6 dB, maximum passband deviation 0.0022 Hz, stopband attenuation -53 dB), with a high pass of 0.1 Hz (transition width 0.2 Hz, stopband 0 Hz, passband 0.2–256 Hz) and low pass of 40 Hz (transition width 10 Hz, passband 0–35 Hz, stopband 45–256 Hz), re-referenced to the average of the mastoids, and epoched from -100 to 550 ms (0 ms at target onset). Error data were removed (as explained in *Behavioral data preprocessing*), along with filler trials. Data from each channel were then visually inspected, and channels, which were bad were removed and interpolated (using the "weighted" method). On average, 1.96 channels (SD = 1.9, range = 0 to 6) were interpolated for production, and 1.88 (SD = 2.1, range = 0 to 6) for perception. Trials with signal above 150 mV were also discarded. We then ran an ICA decomposition (with 40 components) to remove components relating to eye movements using the "fastica" method, per participant. Single channels for the horizontal and vertical eye movement channels were created, and each ICA component was correlated with these two channels. All components were visually inspected by plotting their topographies and time courses, and in conjunction with the correlation results artefactual ICA components were rejected before unmixing. Trials were then visually inspected to remove any trials containing excessive noise, which had not been captured. If more than 60% trials per condition for each participant were removed during the preprocessing procedure then that participant was rejected from further analysis (five participants were rejected). From the remaining participants, 74% data was retained (i.e., the 14% of trials associated with speech errors as indicated above + an additional 12% of bad EEG signal due to eye movements, etc.).

Behavioral Data Analysis

Production latencies corresponding to the cleaned EEG data were analyzed. Due to a voice key recording error, latencies for some cleaned trials did not have a latency recorded. After removing these trials, 78% of the production data remained for analysis.

Data were analyzed with linear mixed effects models (lme4 package; Bates et al. 2015) in R (R core team 2020; version 4.0.2).
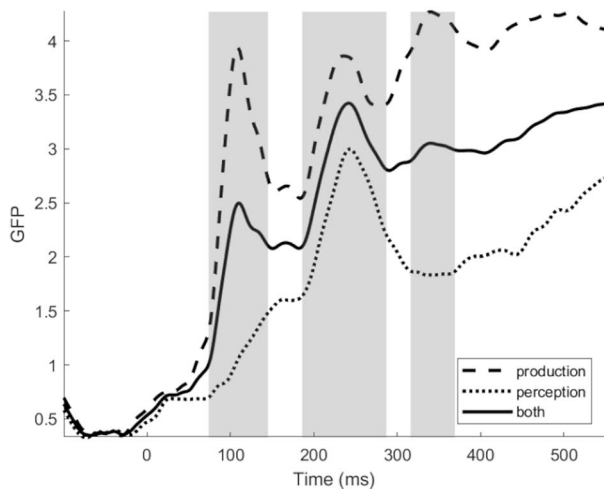
Figure 2. GFP to the production and perception tasks in the experiment. GFP was calculated across all electrodes, conditions, and participants in production and perception, and the average between the modalities was taken to assess the same time windows of analyses for production and perception (centered around the maximal peaks). In this manner, three time windows were objectively determined: 74–145, 186–287 and 316–369 ms.

Trial was entered as a control variable, and the fixed effects modeled were "word frequency," "phonotactic frequency," and their interaction. Categorical predictors (the frequency variables) were sum to zero coded, and the continuous predictor (trial) was centered and scaled. Production latency was log-transformed before analysis to reduce skew. The maximal random effects structure that converged contained random intercepts for participant and item (the picture presented), and a random slope of word frequency by participant.

EEG Analysis

In order to determine time windows (TWs) of analysis in an objective manner (i.e., no "visual assessment" of TWs), the global field power (GFP) of the baselined EEG signal was calculated (baseline of −100 to 0 ms). GFP is a measurement of the strength of the signal measured across all electrodes, conditions, and participants. GFP was calculated per participant per task using FieldTrip in Matlab. GFP was then averaged across participant and task to test for peaks in the GFP signal. GFP peaks were tested for using iPeak (O'Haver 2020) in Matlab. The reason to explore GFP peaks across tasks (i.e., across production and perception) was to obtain the same TWs for analyses across the modalities. Using this procedure, three peaks were found at 109, 242, and 340 ms (parameters amplitude threshold = 2.15, slope threshold = 0.002, smooth width = 3, fit width = 6) (see Fig. 2).

Full-width windows around these peaks were then manually identified. To do this, using the iPeak output per peak, we manually stepped backwards and forwards until reaching the first time point in the backwards or forwards trough. Carrying out the procedure in this way enables asymmetrical widths on either side of the peak, rather than taking a general average of the whole peak window where the peak value may not be centered. This procedure identified the three TWs of interest as TW1: 74—145 ms, TW2: 186—287 ms, and TW3: 316—369 ms (see Fig. 2). The clean EEG data was then averaged per participant per task per electrode in each of the three TWs, and exported for further analysis.

The data from each TW were tested in R using linear regression and linear mixed effects models (lme4). Channels were recoded by their lateral placement into left (Fp1, AF3, F3, FC3, C3, CP3, P3, PO3, F5, FC5, C5, CP5, P5, AF7, F7, FT7, T7, TP7, P7, PO7, P9, O1), mid (F2, FC2, C2, CP2, P2, F1, FC1, C1, CP1, P1, Iz, Oz, POz, CPz, FPz, AFz, Fz, FCz, Cz), and right (AF4, F4, F6, FC4, FC6, C4, C6, CP4, CP6, P4, P6, PO4, AF8, F8, FT8, T8, TP8, P8, PO8, P10, O2), and recoded by their frontal coronal plane placement into anterior (Fp1, AF7, AF3, F1, F3, F5, F7, Fpz, Fp2, AF8, AF4, AFz, Fz, F2, F4, F6, F8), central (FT7, FC5, FC3, FC1, FT8, FC6, FC4, FC2, FCz, TP7, CP5, CP3, CP1, CPz, TP8, CP6, CP4, CP2, T7, C5, C3, C1, Cz, C2, C4, C6, T8), and posterior (P1, P3, P5, P7, P9, PO7, PO3, POz, Pz, P2, P4, P6, P8, P10, PO8, PO4, O1, O2, Oz, Iz).

For the mixed effects models (and linear regression models), all categorical predictors were sum to zero coded, and all continuous predictors were centered and scaled. One model per TW was first run with the main effect predictors word frequency (HWF vs. LWF), phonotactic frequency (high vs. low), language modality (production vs. perception), channel laterality (left vs. mid vs. right), and channel anteriority (anterior vs. central vs. posterior)[2]. There were interactions between the channel placement variables, modality, and each frequency variable. Control predictors included trial number, session order (whether participants carried out the production or the perception task first), average microvoltage in the baseline period (−100 to 0 ms), and average microvoltage of previous TWs (e.g., in models investigating TW 2, the average microvoltage of TW 1 (which was the dependent variable in a TW analysis) was entered into the model as a covariate. We added these to account for independency of the data. Due to the size of this model and in order to interpret the interactions (see the section Results), we then ran models separately for the production and perception data per TW. The control variables were the same as above, and the predictors were word frequency, phonotactic frequency, channel laterality, channel anteriority, and interactions between each frequency variable and the two channel variables. The maximum random structure, which would converge for all models contained random intercepts for participant and item, and no random slopes. For the linear regression, the models contained covariates of trial number, session order, average microvoltage in the baseline period, and average microvoltage of any previous TW. The main effect predictors were word frequency, phonotactic frequency, anteriority, laterality, and interactions between each frequency variable and the channel placement variables. This structure was thus the same as for the mixed models but without a random structure.

For all models, t greater than $|2|$ was taken as a marker of significance and confidence intervals were generated using the confint function with the profile method in the lme4 package. P values were generated using the ANOVA function with type III tests selected in the Car package (Fox and Weisberg 2019). Post hoc comparisons were calculated using the emmeans package (Lenth 2020) with multiple comparisons corrected for by Tukey's HSD. To keep the results section focused, only results from predictors of interest and significant effects are reported, but the full results of all models can be found in the online repository.

---

2 Model in R syntax for ERP analyses (example here for TW2): Mean_uV_TW2 ∼ Trial + Mean_uV_TW1 + SessionOrder + Baseline + Modality*WordFrequency*Laterality*Anteriority + Modality*PhonotacticFrequency*Laterality*Anteriority + (1 | Participant) + (1 | Item)

Table 1. Average production latencies and standard deviation per condition

Word frequency	Phonotactic frequency	Mean RT	SD
High	High	968.18	369.9
High	Low	961.23	366.4
Low	High	1029.8	411.4
Low	Low	1035.73	390.3

Results

Production Latencies

Production latencies are listed by frequency in Table 1. HWF items were responded to descriptively faster than LWF items (Mhigh = 964.81, SD = 368.1; Mlow = 1032.67, SD = 401.2), whereas there was no difference between HPF and LPF items (Mhigh = 997.67, SD = 391.4; Mlow = 996.89, SD = 379.7). The mixed effects models analysis confirmed a significant effect of word frequency (estimate = −0.03, SE = 0.01, t = −2.71, CI = [−0.056, −0.009]), but no significant effect of phonotactic frequency (estimate = −0.0008, SE = 0.01, t = −0.06, CI = [−0.024, 0.022]), nor a significant interaction between the two (estimate = 0.004, SE = 0.01, t = 0.33, CI = [−0.019, 0.027]). Note that the absence of significant phonotactic frequency effects in the naming latencies is not that surprising: In contrast to the word frequency effect, the behavioral phonotactic frequency effect in production is not robust (with several studies finding it in some, but not all experiments), and most studies displays small effect sizes of roughly 5–10 ms (Cholin et al. 2006; Fairs and Strijkers 2021; for an overview see Croot et al. 2017) (note though that in contrast to the behavioral effect, the ERP effect in response to phonotactic frequency—in published work—does seem robust, and is the most relevant for our purposes).

EEG Data: Both Modalities

Mixed Models Analysis: TW1 74 to 145 ms. In TW 1, there was a main effect of modality ($X^2(1,19) > 100$, P < 0.001), with lower microvolt recordings in the production task than the perception task. There were no main effects of word frequency nor phonotactic frequency, but phonotactic frequency significantly interacted with modality ($X^2(1,19) > 100$, P < 0.001). Phonotactic frequency interacted with the channel variable laterality ($X^2(2,19) = 12.915$, P = 0.002). The three-way interaction between modality, word frequency, and both channel variables was significant ($X^2(2,19) = 11.724$, P = 0.003; $X^2(2,19) = 10.746$, P = 0.005), suggesting that the word frequency effect differed in scalp location between modalities. The interaction between modality, phonotactic frequency, and laterality was also significant ($X^2(2,19) = 9.812$, P = 0.007), suggesting that the phonotactic frequency effect was distributed differently across the scalp in each modality.

In sum, we observed that modality interacted with the word frequency and phonotactic variables; in order to understand the nature of these interactions, separate analyses per modality were performed (see below: *EEG data by modality*).

Mixed Models Analysis: TW2 186 to 287 ms. For TW 2, very similar results were found to TW 1. Again, there was a main effect of modality ($X^2(1,19) > 100$, P < 0.001), and no main effects of word or phonotactic frequency. The word frequency by modality interaction was significant ($X^2(1,19) = 18.549$, P < 0.001), and the phonotactic frequency by modality interaction was also

significant ($X^2(1,19) > 100$, P < 0.001). The word frequency by anteriority interaction was significant ($X^2(2,19) = 8.014$, P = 0.020), and the phonotactic frequency by anteriority interaction was significant ($X^2(2,19) = 24.756$, P < 0.001). The three-way interaction between word frequency, modality, and anteriority was significant ($X^2(2,19) = 30.465$, P < 0.001), and the three-way interactions between phonotactic frequency, modality, and both channel variables were also significant ($X^2(2,19) = 28.579$, P < 0.001; $X^2(2,19) = 32.350$, P < 0.001). As for TW1, these interactions were further assessed with separate analyses per modality (see below: EEG data by modality).

Mixed Models Analysis: TW3 316 to 369 ms. In TW3, the same main effects patterns were found: main effect of modality ($X^2(1,19) > 100$, P < 0.001) and no main effects of word and phonotactic frequency. There was a word frequency by modality interaction ($X^2(1,19) = 13.442$, P < 0.001), and word frequency by anteriority interaction ($X^2(2,19) = 21.817$, P < 0.001). The three-way interaction between word frequency, modality, and anteriority was significant ($X^2(2,19) = 30.021$, P < 0.001). Phonotactic frequency interacted with modality ($X^2(1,19) = 16;712$, P < 0.001) and was involved in a three-way interaction with modality and anteriority ($X^2(2,19) = 76.734$, P = 0.022). Separate analyses per modality were performed in order to assess the nature of these interactions (see below: EEG data by modality).

EEG Data: By Modality

Mixed Models Analysis: TW1 74 to 145 ms. Production Main effects of word frequency and phonotactic frequency were not significant. Word frequency, however, interacted with laterality ($X^2(2,19) = 11.90$, P = 0.003) (see Figs 3 and 4). Post hoc *t*-tests revealed that this interaction was driven by a significant increase from right to left electrodes for low-frequency items (P < 0.001, M = −1.92 to −1.76 (SEs = 0.04), diff = 0.16 mV), which was absent for high-frequency items (P = 0.799, M = −1.79 to −1.69 (SEs = 0.04), diff = 0.10 mV). There was also a significant interaction between phonotactic frequency and laterality ($X^2(2,19) = 6.974$, P = 0.031) (see Figs 3 and 4). Post hoc tests showed this interaction was driven by a marginally significant difference between high- and low-frequency items at left electrodes (P = 0.056, Mhigh = −1.57 (SE = 0.04), Mlow = −1.98 (SE = 0.04)), and a significant increase from right to left electrodes for high-frequency items (P < 0.001, Mright = −1.7 (SE = 0.04), diff = 0.13 mV), which was absent for low-frequency items (P = 0.998, Mright = −1.9 (SE = 0.04), diff = 0.08 mV).

Perception Similarly to production, there were no main effects of word or phonotactic frequency. There was, however, a significant interaction between word frequency and anteriority ($X^2(2,19) = 11.419$, P = 0.003) (see Figs 3 and 4). Post hoc tests suggest this interaction is due to HWF items (M = 1.14) being more positive than low-frequency items (M = 1.05) across the scalp, with a greater voltage increase at anterior sites (high frequency, P < 0.001, M = 1.24 (SE = 0.05); low frequency, P = 0.045, M = 1.03 (SE = 0.03), diff = 0.21 mV) and central sites (high frequency, P < 0.001, M = 1.19 (SE = 0.03); low frequency, P = 0.001, M = 1.1 (SE = 0.03), diff = 0.08 mV) compared to posterior sites (Mhigh = 1 (SE = 0.03), Mlow = 1.03 (SE = 0.03), diff = 0.03 mV). For phonotactic frequency we found a significant interaction with laterality ($X^2(2,19) = 16.487$, P < 0.001), with a trend towards significance between low- and high-frequency items at mid electrodes (P = 0.058, Mlow = 1.31 (SE = 0.04), Mhigh = 0.93 (SE = 0.04)), a significant voltage increase from left towards mid sites for low-frequency items (P < 0.001, M = 0.98 to 1.31 (SEs = 0.04), diff = 0.33 mV), which was absent for high-frequency items (P = 1,
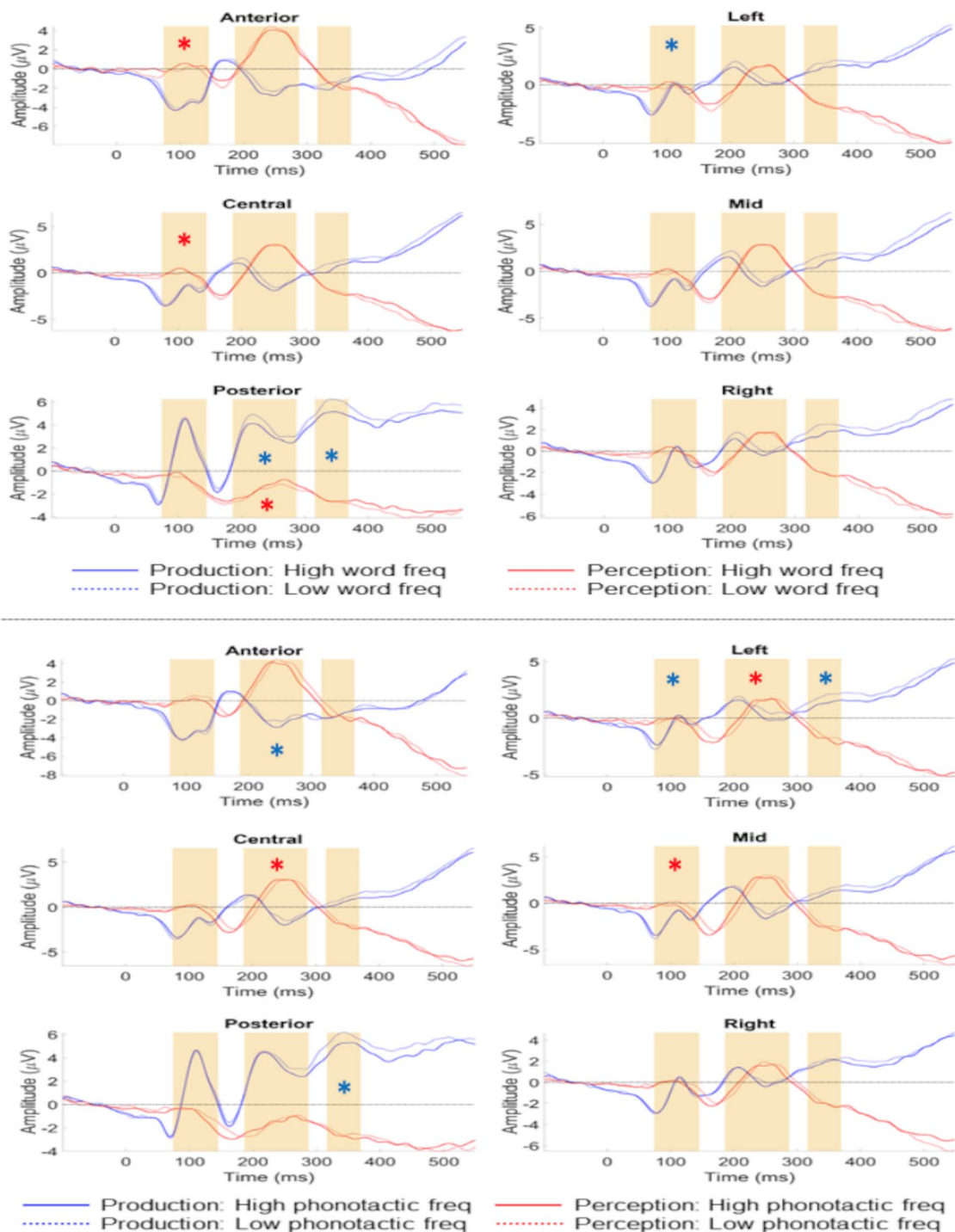
Figure 3. ERP results for the production and perception tasks at each electrode location. The upper panel displays the ERPs for high (full line) versus low (dotted line) word frequency in production (blue) and perception (red). The lower panel displays the ERPs for high (full line) versus low (dotted line) phonotactic frequency in production (blue) and perception (red). TWs for analyses are highlighted with yellow rectangles, and the significant interactions of scalp site with word or phonotactic frequency, respectively, are depicted with a blue asterisk for production and a red one for perception. Note that in production, effects for both variables were significant in all three time windows, while in perception, effects for both variables were significant for the first two TWs, but not for the last TW. Note that for visualization purposes, ERPs are presented here as the averaged means, but the actual data-analyses were performed for single-trial ERP data with mixed linear effect models.
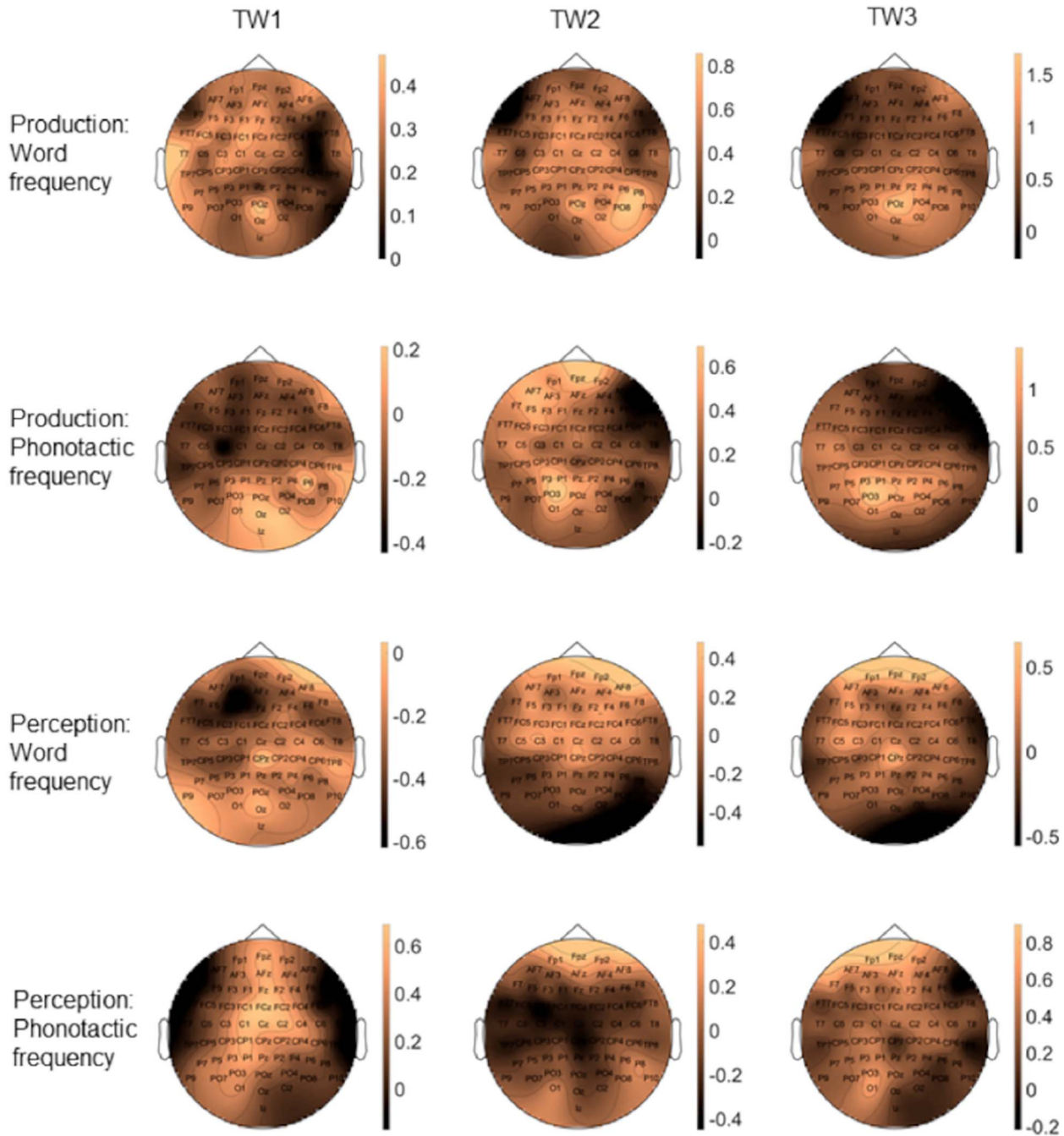
Figure 4. Topographic scalp distributions of the word frequency and phonotactic frequency effects in production and perception at each TW of analyses.

M = 0.86 to 0.92 (SEs = 0.04), diff = 0.06 mV), and a significant increase from mid to right sites for high-frequency items (P < 0.001, M = 0.93 (SE = 0.04) to 1.19 (SE = 0.03), diff = 0.26 mV), which was absent for low-frequency items (P = 0.610, M = 1.31 to 1.34 (SEs = 0.04), diff = 0.03 mV) (see Figs 3 and 4).

Mixed Models Analysis: TW2 186 to 287 ms. Production There were no main effects of word nor phonotactic frequency. Word frequency did interact with anteriority (X^2(2,19) = 7.752, P = 0.021) (see Figs 3 and 4). Post hoc tests showed a stronger increase in positivity from anterior to posterior electrodes for low-frequency items (P < 0.001; M = −0.67 (SE = 0.04) to 0.93 (SE = 0.03), diff = 1.6 mV) compared to high-frequency items (P < 0.001;

M = −0.67 (SE = 0.04) to 0.76 (SE = 0.03), diff = 1.43 mV). Phonotactic frequency interacted with anteriority (X^2(2,19) = 40.678, P < 0.001), with a marginally significant difference between high- and low-frequency items at anterior sites (P = 0.053, Mhigh = −0.82 (SE = 0.04), Mlow = −0.52 (SE = 0.04), diff = 0.3 mV), and a stronger increase in positivity from central to anterior sites for low-frequency items (P < 0.001; M = −0.81 (SE = 0.03) to −0.52 (SE = 0.04), diff = 0.29 mV) compared to high-frequency items (P < 0.001; M = −0.89 (SE = 0.03) to −0.82 (SE = 0.04), diff = 0.07 mV) (see Figs 3 and 4). Phonotactic frequency also interacted with laterality (X^2(2,19) = 6.293, P < 0.043), with stronger positivity in left compared to right sites for low-frequency items (P = 0.002,

M = 0.02 to −0.17 (SEs = 0.03), diff = 0.19 mV), which was absent for high-frequency items (P = 0.947, M = −0.15 to −0.2 (SEs = 0.03), diff = 0.05 mV).

Perception No main effects of word or phonotactic frequency were found. The interaction between lexical frequency and anteriority was significant (X^2(2,19) = 33.636, P < 0.001), where post hoc inspection suggests a stronger increase in negativity from anterior to posterior electrodes for low-frequency items (P < 0.001; M = 3.78 (SE = 0.04) to 0.06 (SE = 0.03), diff = 3.72 mV) compared to high-frequency items (P < 0.001; M = 3.55 (SE = 0.04) to 0.11 (SE = 0.03), diff = 3.44 mV) (see Figs 3 and 4). Also the phonotactic frequency by anteriority interaction was significant (X^2(2,19) = 15.256, P < 0.001), with low-frequency items being more negative than high-frequency items at central electrodes (P = 0.049, Mlow = 2.64 (SE = 0.03), Mhigh = 2.98 (SE = 0.03)), and a stronger increase in positivity from posterior to central sites for high frequency (P < 0.001; M = 0.15 to 2.97 (SEs = 0.03), diff = 2.82 mV) than low-frequency items (P < 0.001; M = 0.03 to 2.64 (SEs = 0.03), diff = 2.61 mV). The phonotactic frequency by laterality interaction was also significant (X^2(2,19) = 33.699, P < 0.001), with trends for low-frequency items being more negative than high-frequency items at left (P = 0.06, Mlow = 1.67 (SE = 0.03), Mhigh = 1.99 (SE = 0.03)) and mid sites (P = 0.067, Mlow = 2.59 (SE = 0.04), Mhigh = 2.9 (SE = 0.04)), and a stronger increase in negativity from right to left electrodes for low-frequency items (P < 0.001; M = 2.01 to 1.68 (SEs = 0.03), diff = 0.33 mV) compared to high-frequency items (P = 0.123; M = 2.07 to 1.99 (SEs = 0.03), diff = 0.08 mV) (see Figs 3 and 4).

Mixed Models Analysis: TW3 316 to 369 ms. Production No main effects of either frequency variable were found. Word frequency interacted with anteriority (X^2(2,19) = 51.868, P < 0.001) (see Figs 3 and 4). Post hoc tests demonstrated a trend towards more positivity for low compared to high-frequency items at posterior sites (P = 0.068, Mlow = 3.24 (SE = 0.04), Mhigh = 2.67 (SE = 0.04)), and a stronger increase from frontal towards posterior electrodes for low frequency (P < 0.001; M = −0.63 (SE = 0.05) to 3.24 (SE = 0.04), diff = 3.87 mV) than high-frequency items (P < 0.001;, M = −0.67 (SE = 0.05) to 2.67 (SE = 0.04), diff = 2 mV). Phonotactic frequency interacted with anteriority (X^2(2,19) = 8.049, P = 0.018)(see Figs 3 and 4). Post hoc tests suggest that the phonotactic frequency by anteriority interaction was mainly driven by a stronger increase in positivity from anterior to posterior sites for low (P < 0.001; M = −0.64 (SE = 0.05) to 3.17 (SE = 0.04), diff = 2.53 mV) compared to high-frequency items (P < 0.001; M = −0.67 (SE = 0.05) to 2.73 (SE = 0.04), diff = 3.4 mV).

Perception No main effects of word frequency or phonotactic frequency were found, nor any interactions with the channel variables (see Figs 3 and 4).

Additional Analyses: Linear Regression. We also carried out a linear regression analysis on the three TWs, akin to a more typical ANOVA style analysis of the data, given that this still concerns the most performed analyses on neurophysiological data and is therefore interesting as a comparison. For brevity and clarity, we will not discuss the statistical details of these analyses—they can be consulted in full in the online repository at https://osf.io/hp2 me/—but only descriptively report the main findings. In general, the results were quite similar to the mixed models analysis in that they confirmed the presence of word and phonotactic frequency effects for both modalities in all TWs (with the exception of TW3 in perception; just as in the mixed models analyses). The biggest, and quite noteworthy, difference with the mixed models analyses is that the linear regression analyses displayed main effects of both frequency variables in several of the TWs

analyzed (see the osf repository). In other words, the word and phonotactic frequency effects were more marked and broader in the traditional regression analyses compared to the mixed effect models analyses.

A deeper investigation into these discrepancies suggested that there is a substantial amount of variance attributable to the items used. This variance is accounted for as random variation in the mixed effects models, i.e., the variance is treated as "noise" and not as part of the variables of interest. In linear regression, this variance is instead (at least partially) grouped with variation in the variables of interest, leading to a higher chance of a significant effect. In this manner, we believe that the mixed effect models analysis is a better approach to single out the "pure" frequency effects, since item variance between conditions is better controlled for compared to traditional ANOVAs and linear regression analyses. This difference between the types of analyses might have important consequences for all neurophysiological research, which relies on conditions containing physically different stimuli. For present purposes, we will not go any further into this interesting methodological issue (but see Fairs et al. 2020), and instead highlight that both analyses point to the same theoretical insight, namely early parallel access of word and phonotactic frequency in both production and perception. Below we will discuss what this finding signifies for the temporal dynamics of brain language models.

General Discussion

We investigated the temporal dynamics of word activation by comparing the event-related brain potential elicited by lexico-semantic and phonological knowledge in production versus perception, and analyzing the data at the single-trial level with mixed linear effects models. The results showed significant word frequency and phonotactic frequency effects (in interaction with scalp location, respectively) in early TWs (74–145 ms and 186–287 ms) for both word production and perception (see Fig. 3). At a later TW (316–369 ms), word and phonotactic frequency effects only remained significant for word production, but not for word perception (see Fig. 3). This data pattern provides evidence for ultra-rapid parallel activation dynamics underpinning word retrieval in both production and perception. The speed of linguistic activation, it's parallel nature for different word knowledge and the identical temporal profile (early on) when speaking and understanding places important novel constraints on brain language models: 1) The data provide strong evidence for parallel IMs where words are represented in the brain as unified cell assemblies across the language modalities (Pulvermüller 1999, 2018; Strijkers 2016; Strijkers and Costa 2016), 2) The early parallel activation, which is followed at a later stage of processing by modality-specific activation fits the predictions of Hebbian-like neural assembly models where early word ignition is identical between production and perception and later reverberation is task- and context specific. 3) The similarities in cortical dynamics between word production and word perception suggest the same mental representations underpin both modalities and therefore production–perception interactions in conversation may tap into the same neural resources for lexical representations between speakers and listeners.

Parallel Cortical Dynamics of Word Processing

The main objective of this study was to shed novel light on a central and longstanding question in the psycho- and neurolinguistic literature: Are words processed in a hierarchical sequential

manner or in a parallel integrated manner? The data were clear-cut and demonstrated that lexico-semantic and phonological-phonetic word knowledge are rapidly activated in parallel, both when uttering and comprehending words[3]. The observation of the same parallel activation pattern across the language modalities is crucial, since the parallel effects cannot be attributed to physical variance between stimuli nor to a cascade of activation over sequentially organized levels of word processing.

With regard to physical variance between stimuli (e.g., sensorial differences between low- and high-frequency stimuli), in the present study, the input is different between production (visually presented objects) and perception (auditory presented words), but nonetheless the same temporal effects between the modalities were present for the word frequency and phonotactic frequency effects. Therefore, explaining the data as due to sensorial differences between the low- and high-frequency conditions is highly unlikely (if not impossible) since it would mean that by chance the same sensorial differences are present for both the visual and auditory input, for both manipulations and in both early TWs. Instead, parsimony clearly favors a linguistic explanation since across the two modalities, identical word processing components are targeted. Adding further support to this conclusion is the observation that we replicated the typical ERP morphology found for word frequency and phonotactic frequency in prior production and perception studies in isolation. In production, ERPs elicited by low-frequency items (both word and phonotactic) display more positive going waveforms than high-frequency items (Strijkers et al. 2010, 2011; Baus et al. 2014; Bürki et al. 2015), while in perception ERPs elicited by low-frequency items typically display more negative going waveforms than high-frequency items (Hauk et al. 2006; MacGregor et al. 2012; Hunter 2013; Winsler et al. 2019). Here both these ERP patterns were replicated and integrated within the same participants and for the same stimuli, lending strong reliability to the observed results.

The current cross-modal timing cannot be reconciled with those brain language models assuming functional delays between word components in the range of 100 s ms (Friederici 2002, 2011; Indefrey and Levelt 2004; Indefrey 2011). Moreover, the argument that the overlapping time course of lexico-semantic and phonological word knowledge could reflect fast sequentially and interactivity rather than parallel activation finds little support in the current results. If lexico-semantic and phonological-phonetic word processing are functionally (and thus temporally) distinct, we should have seen an interaction between the language modalities with phonotactic frequency manifesting *prior* to word frequency in perception *and* that same phonotactic frequency manifesting *after* word frequency in production. No such interaction was present, and instead the ERP data revealed effects of lexico-semantic and phonological word

processing, which were indistinguishable between language production and perception in our early time windows (74–145 ms; 186–287 ms).

The fact that we can contrast the activation time course of a lexico-semantic and phonologic variable across two modalities offers an important advantage over the traditional within-modality studies, namely that the points of comparison are doubled. For example, taking the first TW where we observed significant effects for both variables (74–145 ms), this means that in order to claim fast sequentiality, the maximal possible onset delay that remains between lexico-semantic and phonological access in production and (vice versa respectively) in perception has to be less than 35 ms. At this point one may question how such small (potential) delays would still support the notion of sequentiality. For one, these short delays perfectly fit with axonal conduction velocities between different brain regions (Miller 1996; Matsumoto et al. 2007), hereby denoting physical distance delays in the brain, not functional ones. In a similar vein, it has been recently shown from recordings with micro-wire electrodes that small spatiotemporal delays around 30 ms offer an ideal temporal window in the human brain to form Hebbian cell assemblies (Roux et al. 2021). Furthermore, if sequential activation would be so fast that it becomes indistinguishable to assess between such early TWs (as observed here), what happens for the remaining 400–500 ms in those models? The elegance of sequential models is that they link different linguistic operations to different points in time that gradually enroll over the entire period of speech planning. Unlike for parallel IMs (see also next section), no theory nor predictions are available for sequential models, which would explain what happens during the bulk of speech planning if the serial activation onsets are as fast as observed here. Also, assuming rapid interactivity does not explain the data pattern we observed in the present study: In interactive models (Dell 1986; Friederici 2011) lexico-semantic and phonological processing can overlap due to the feedback activity from a layer higher in the hierarchy (e.g., phonology) onto a layer lower in the hierarchy (e.g., lexical). However, in this case the overlap produces itself after initial sequential feedforward activation. In other words, an interactive brain language model predicts first sequential activation dynamics followed by parallel activation dynamics, while in the current study we observed the reverse pattern with parallel effects very early on and modality-specific effects later on. In sum, for sequential and interactive models to account for the current data pattern, critical properties of why hierarchical sequentiality or interactivity was proposed as a temporal processing dynamic to begin with would need to be entirely dropped or changed. In contrast, parallel IMs exactly predict the temporal pattern we observed in this study, even the ultra-rapid parallel effects in the earliest time window (74–145 ms) (Pulvermüller 2018).

While in language perception few of such very early effects linked to word processing have been reported before (Pulvermüller et al. 2005; MacGregor et al. 2012; Grisoni et al. 2016; Strijkers et al. 2019), to our best knowledge this is the first demonstration of ultra-rapid word access in language production. Although we do not believe the early effect indexes item-specific retrieval of the target word, for which the TW between 186 and 287 ms seems a better candidate based on the previous literature (Vihla et al. 2006; Costa et al. 2009; Strijkers et al. 2010, 2017; Aristei et al. 2011; Miozzo et al. 2015; Python et al. 2018; Feng et al. 2021), its presence both in the production and perception tasks does suggest a linguistic effect. We argue this early effect reflects a difference of "global" word activation between the conditions, with high word and phonotactic frequency items

---

3 To ensure the parallel activations observed between production and perception were indeed the same between the modalities for these early TWs, we additionally analyzed post hoc a TW between 150 and 180 ms after stimulus onset. Even though this TW falls outside the a priori defined maximal GFP windows, upon visual inspection differences between conditions seems apparent (see Figure 3). The results were identical to those of the analyses for TW1 and TW2, displaying significant word and phonotactic frequency effects both in the production (word frequency: $P < 0.001$; phonotactic frequency: $P < 0.001$) and the perception modalities (word frequency: $P = 0.002$; phonotactic frequency: $P < 0.001$). In sum, these additional analyses confirm the main analyses (for complete details and all model output see the online repository: https://osf.io/hp2me/).

causing enhanced activation of a set of potential words linked to the input compared to the low word and phonotactic frequency items. In other words, the differences between the high- and low-frequency conditions trigger almost immediately upon perceiving the input activation within the linguistic system, which is different in its magnitude between frequent and more rare words (this difference could be due to higher resting level of activation and/or stronger connectivity and myelinization between input features and lexical features for the more frequent words). Note also that such ultra-rapid linguistic activations fit well with recent spatiotemporal evidence suggesting that combinatorial processes emerge already within 200 ms of processing (Pylkkänen 2019). Clearly, if we can start integrating two words at such speeds, it logically follows that some information of each single word has to be retrieved even faster.

In sum, the uncovered parallel cortical dynamics of distinct word components in both production and perception cannot be accounted for by the dominant sequential hierarchical brain language models in the literature (Friederici 2002, 2011; Indefrey and Levelt 2004; Hickok and Poeppel 2007; Indefrey 2011; Hickok 2012). In contrast, the data nicely fit parallel IMs, which state that words rapidly ignite as whole in parallel, and do so identically in language production and perception (Pulvermüller 1999, 2018; Strijkers 2016; Strijkers and Costa 2016). To our best knowledge, this is the strongest demonstration yet for parallel cortical dynamics underpinning word processing and the main contribution of our study.

Ignition and Reverberation as Neurophysiological Principles of Word Processing?

Another intriguing observation in our data was that for the late TW of analysis (316—369 ms) language production and perception displayed different patterns. Both the word and phonotactic frequency effects remained significant for the language production task, but were no longer significant for the language perception task[4]. This pattern, where early on in the course of processing the cortical dynamics are the same between production and perception, but later on they become different, fits with the theoretical assumptions put forward in Hebbian-based neural assembly models of word processing (Pulvermüller 1999, 2018; Strijkers 2016; Strijkers and Costa 2016). Within such models, the idea is that early ignition of a cell assembly and later reverberation within that cell assembly can fulfill distinct functional roles. The notion stems from the neurophysiological properties of cell firing and theoretical systems neuroscience (Hebb 1949; Braitenberg 1978; Fuster 2003; Buzsáki 2010; Singer 2013). The firing of neuronal cells is marked by a rapid, explosion-like onset (ignition), followed by slower, sustained activity (reverberation).

Importantly, while originally, it was assumed this slow reverberatory decay of neural activity after cell ignition was mainly noise in the system, by now it is clear that reverberation is sensitive to stimuli- and task-relevant properties (Fuster 1995; Singer and Gray 1995; Tsodyks and Markram 1997; Azouz and Gray 2003). In this manner, cell ignition has been linked to "target identification," and reverberations upon the ignited cells to second-order processes such as memory consolidation, reprocessing, decision-making and meta-cognition (Mesulam 1990; Fuster 1995, 2003; Engel et al. 2001; Dehaene et al. 2006).

A similar idea of distinct functional states linked to the different neuronal firing properties of ignition and reverberation can be adopted to language processing (Pulvermüller 1999, 2002, 2018; Strijkers 2016; Strijkers and Costa 2016; Strijkers et al. 2017), and its plausibility has been computationally demonstrated (Garagnani et al. 2008; Schomers et al. 2017). Specifically, for word processing, the hypothesis is that rapid ignition would reflect the activation of the memory representation of a word associated with the input, and reverberation with reactivation of the ignited word assembly linked to task- and behavior-specific processing such as verbal working memory, semantic integration, grammatical inflections, and articulatory control (Pulvermüller 2002; Strijkers 2016; Strijkers et al. 2017). For example, if one intends to utter the phrase "(s/he) runs," during ignition the "base" representation "(to) run" would quickly become activated, and subsequently during reverberation this "base" representation is morphologically inflected to "runs" in function of the syntactic context and constraints (s/he). So how could such neurophysiological dynamics of ignition and reverberation account for the fact that in our data word and phonotactic frequency effects remained significant for the language production task, but not for the perception task? A possible answer lies in the nature of the task and behavior that participants had to produce: In production, the goal-directed behavior of the participants is to utter the specific word associated with the input. Hence, the input-specific word representation, with its associated lexico-semantic and phonetic knowledge, has to be kept online in order to successfully execute the task and articulate the unique word response. In contrast, in the perception task, the goal-directed behavior of the participants is geared towards the assessment whether the word associated with the input fits in the general category knowledge of "food items." In other words, in the production task, the focus is on the single word representation that needs to be uttered while inhibiting other potential word candidates that could interfere with the goal-directed behavior. Of course, this does not mean that no related lexical or conceptual representations are active in the production task, but that in order to correctly utter a single word, at some point in the process the target representation has to be boosted while related information needs to be inhibited. In contrast, in the perception task, the activated word representation needs to be integrated with other word and category knowledge, thus promoting the additional activation of related words and concepts. This concurrent activation of related words and concepts may therefore cause less pronounced word-specific effects in the perception task compared the production task. Alternatively, the difference may stem from the need to perform a response in production versus the absence of a response for the no-go trials in the perception task. Regardless of the exact source, both explanations highlight that a difference in goal-directed behavior between the two language behaviors could account for the modality-specific differences later on in processing as predicted by neural assembly models of language.

---

4 Note, when inspecting visually the figures (see Figure 3), it seems that after TW3 word and phonotactic frequency effects become significant again in perception as well. To assess this, post hoc we performed an additional mixed linear effect analysis between 400 and 500 ms after stimulus onset. The results of these post hoc analyses (which can be consulted in detail in the online repository: https://osf.io/hp2me/) indeed confirmed significant interactions of both word and phonotactic frequency with electrode location in perception (and also in production). Nevertheless, this post hoc result does not change the fact that, in contrast to the early TWs, from around 300 ms after stimulus onset the ERP effects are different between production and perception, with less robust and extended word and phonotactic frequency effects in perception compared to production.

From this point of view, the full electrophysiological data pattern we observed in our study can be accounted for by the neural assembly model of words in the brain in the following manner (see Fig. 5): After early initial sensorial activation in response to the input, 1) the first linguistic activation emerges roughly between 75–150 ms after onset denoting the start of "globally" activating potential words fitting the initial sensory analyses, with frequent words triggering a larger set of potential candidates than less frequent words; 2) next, this initial "global word space" activation is further refined and delineated within roughly 150–250 ms leading to the ignition and recognition of the specific lexical item associated with the input; 3) finally, roughly after 300 ms, slower, sequential-like reverberation upon the activated word assembly takes effect in order to embed the recognized target word into the proper linguistic and task context to be able to perform the intended behavior, and modality-specific processing effects become visible. Note that within this framework the sequential effects of word processing reported in the literature (and discussed in the Introduction; Indefrey and Levelt 2004; Friederici 2011; Indefrey 2011) would thus stem from reverbatory processing, but not ignition. Put differently, parallel neural assembly models do not negate the presence and even necessity of serial-like operations in word production and perception, but rather link those observations to a different functional interpretation. That said, the current study does not prove this reverbatory link per se and the proposed framework is tentative. Nevertheless, given the same early parallel effects between production and perception as proposed by parallel IMs, and the prediction of these models that later on in the processing sequence task-specific differences will emerge between production and perception, makes this interpretation a relevant one. Future investigations into this framework of functionally distinct linguistic operations associated with cell ignition and reverberation are important, since it may provide us with linking hypotheses between the algorithms of language processing and their neural implementation.

A Blueprint of Perception–Production Interactions for Conversational Dynamics

To conclude, our data are also relevant to constrain integrated models of production–perception interactions and conversation. While traditionally both language modalities were investigated separately, more recently, there is an increased interest about the nature of language processing under more ecologically valid conditions such as the dyad (Hasson et al. 2012; Pickering and Garrod 2013; Schoot et al. 2016). Importantly, investigations contrasting speaker–listener dynamics have demonstrated neural overlap between the modalities for high-level linguistic operations such as syntactic unification of message-level information (Menenti et al. 2011; Segaert et al. 2012; but see Matchin and Wood 2020), or communicative alignment in conversation Stephens et al. 2010; Silbert et al. 2014; Dikker et al. 2017). These data highlight that the language production and perception systems are more shared than originally thought. Less evident to distil from these data is whether the encoding and decoding of language recruit identical neural populations or rather two interactive, but segregated, neural populations, which are integrated at the highest levels of linguistic processing (e.g., communicative common ground). Indeed, that processes such as communicative alignment or syntactic unification tap into the same neurobiological substrates for production and perception does not necessitate the assumption that lower level building blocks equally rely on
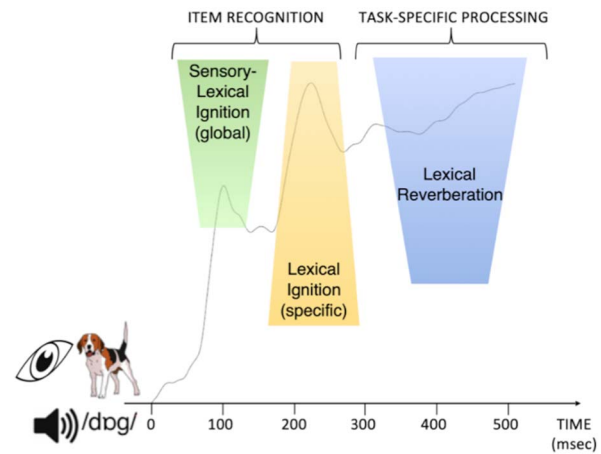

Figure 5. Schematic overview of word processing in a neural assembly framework of ignition and reverberation. Depicted is the global field potential elicited by speech production and speech perception. (1) Upon sensory analyses of a visual stimulus (e.g., picture or spoken word "dog"), the first signs of lexical activation occur very early on (from 75 ms onwards) during "global" sensory-lexical ignition where potential lexical candidates associated with the input become partially activated; (2) from 175 ms onwards, lexical ignition takes effect where the word assembly corresponding to the input becomes active and selected for further processing; (3) from 300 ms onwards, lexical reverberation manifests where the word assembly is re-activated (fully or partially) in function of the specific task and behavior.

shared representations across modalities (Hagoort and Indefrey 2014; Pickering and Garrod 2014; Silbert et al. 2014), and most brain language models still assume at least partial separation between production and perception at the level of words (Scott and Johnsrude 2003; Indefrey and Levelt 2004; Hickok and Poeppel 2007; Friederici 2011; Hickok 2012; Hagoort and Indefrey 2014). Addressing whether this is indeed the case is an important endeavor for future research in order to mechanistically link high-level linguistic processing between interlocutors with the basic linguistic elements on which they rely.

The current data may provide some hints towards this objective, given the overlapping activation time course between production and perception at the level of words. This temporal dynamic fits the notion that words, and their respective linguistic components, share the same cortical representations in production and perception. And though it remains plausible that the time course of processing may be similar between perception and production, but their spatial and functional arrangement is not—an important question to be explored in future research—the similar ERP pattern across the language modalities as observed here neatly follows the predictions of IMs (Pulvermüller 1999, 2018; Strijkers 2016; Strijkers and Costa 2016). Key to neural assembly IMs is that representations are formed through Hebbian-like learning binding together correlated input (perception)—output (production) associations (Pulvermüller and Fadiga 2010; see also Skipper et al. 2017; Shamma et al. 2021); this way, the mental representations of words rely on distributed action-perception circuits, which integrate traditional "production properties" (e.g., motor knowledge) and "perception properties" (e.g., acoustics) into a single coherent unit. From this perspective, what differentiates language production and perception are not the linguistic representations, but the dynamics upon those identical representations. If so, this will have important consequences for interweaving production and perception.

For example, a key element in current conversational models is the assumption that successful alignment (i.e., the coordination and copying of utterances between interlocutors) relies on predictive processing (Federmeier 2007; Pickering and Garrod 2013, 2014; Dell and Chang 2014; Silbert et al. 2014). In short, a prominent idea is that during perception we rely on the production system to make predictions about the upcoming sensorial information and vice versa during production we rely on "efference copies" in the perception system (Pickering and Garrod 2007, 2013; Scott et al. 2009; Tourville and Guenther 2011; Martin et al. 2018). For this mechanism to work, we thus need two separate (though closely interacting) representations, one in production and one in perception. However, if production and perception rely on the same neural representations for words, then this matching mechanism with predictions in one modality and implementations in the other modality would not work. The latter does not mean that prediction would not play a role in conversation; it likely does, and in fact, the very early ERP responses to word knowledge we observed here (75–150 ms) seem a prime candidate to reflect such "partial" predictions of possible words (Grisoni et al. 2016; Strijkers et al. 2019). But the mechanism to implement that prediction would need to be different, where at least the top–down prediction and bottom–up driven analyses of the input converge on the "same" neural representation. Similarly, for speech error monitoring, which necessitates a close interplay between production and perception (Hickok 2012; Pickering and Garrod 2013), the current data fit well with recent findings demonstrating that forward predictions stemming from the cerebellum go beyond mere articulatory control, but include lexical and phonological levels of processing (Runnqvist et al. 2016, 2021). That is, if words are reflected in the brain as holistic action-perception circuits, it follows that sensory motor to cerebellar connectivity will also affect more abstract levels of linguistic processing. Finally, and still largely uncharted territory, the manner to link up low-level language processes with higher level ones in conversational contexts will be markedly different whether the same or different word representations need to be unified onto a common ground between speakers and listeners. Continued research exploring high-level production–perception interactions in ecologically valid conversational contexts, together with more basic, controlled investigations of low-level spatiotemporal similarities and differences between production and perception will be valuable endeavors to address these questions.

Conclusion

Our ERP data revealed surprisingly rapid and parallel activation of lexico-semantic and phonology-phonetic word knowledge, which was indistinguishable between the production and perception of language. These data contrast with the traditional sequential hierarchical models of word production and perception, and instead offer strong support for parallel integration models based on Hebbian-like cell assembly theory where words reflect the Gestalts of language in the brain (Pulvermüller 1999, 2018; Strijkers 2016; Strijkers and Costa 2016). Additional support for this conclusion stems from the observation that the early cross-modal parallel activation of word components was followed later on by distinct ERP effects between production and perception. We interpreted these findings within a neural framework of cell ignition and reverberation being associated with different cognitive states of word processing: rapid, parallel ignition with word activation and retrieval, and later reverberation with behavior-specific processing upon the ignited word assembly in

function of language modality. In this manner, this study may offer a novel and exciting way to map basic neurophysiological brain dynamics onto cognitive operations necessary to express linguistic behavior.

Notes

Conflict of Interest: None declared.

Author Contributions

K.S. designed to study; K.S., A.M., and S.D. implemented the experiment and collected the data; A.F. analyzed the data; A.F. and K.S. interpreted the data and wrote the manuscript; and all authors reviewed the manuscript.

Fundings

References

Aristei S, Melinger A, Abdel Rahman R. 2011. Electrophysiological chronometry of semantic context effects in language production. *J Cogn Neurosci.* **23**:1567–1586.

Azouz R, Gray CM. 2003. Adaptive coincidence detection and dynamic gain control in visual cortical neurons in vivo. *Neuron.* **37**:513–523.

Bates D, Kliegl R, Vasishth S, Baayen H. 2015. Parsimonious mixed models. arXiv preprint arXiv:1506.04967. 29 May 2018, preprint: not peer reviewed.

Baus C, Sebanz N, de la Fuente V, Branzi FM, Martin C, Costa A. 2014. On predicting others' words: Electrophysiological evidence of prediction in speech production. *Cognition.* **133**:395–407.

Boersma P, Weenink D. 2020. *Praat: Doing phonetics by computer* [computer program]. Version 6.0. 37. http://www.praat.org/.

Braitenberg V. 1978. *Cell assemblies in the cerebral cortex: In theoretical approaches to complex systems.* Berlin, Heidelberg: Springer. pp. 171–188.

Brehm L, Goldrick M. 2016. Empirical and conceptual challenges for neurocognitive theories of language production. *Lang Cogn Neurosci.* **31**:504–507.

Bürki A, Cheneval PP, Laganaro M. 2015. Do speakers have access to a mental syllabary? ERP comparison of high frequency and novel syllable production. *Brain Lang.* **150**:90–102.

Buzsáki G. 2010. Neural syntax: cell assemblies, synapsembles, and readers. *Neuron.* **68**:362–385.

Caramazza A. 1997. How many levels of processing are there in lexical access? *Cogn Neuropsychol.* **14**:177–208.

Caramazza A, Costa A, Miozzo M, Bi Y. 2001. The specific-word frequency effect: Implications for the representation of homophones in speech production. *J Exp Psychol Learn Mem Cogn.* **27**:1430.

Cholin J, Levelt WJ, Schiller NO. 2006. Effects of syllable frequency in speech production. *Cognition.* **99**:205–235.

Costa A, Strijkers K, Martin C, Thierry G. 2009. The time course of word retrieval revealed by event-related brain potentials during overt speech. *Proc Natl Acad Sci.* **106**:21442–21446.

Croot K, Lalas G, Biedermann B, Rastle K, Jones K, Cholin J. 2017. Syllable frequency effects in immediate but not delayed syllable naming in English. *Lang Cogn Neurosci*. **32**:1119–1132.

Dahan D, Magnuson JS, Tanenhaus MK. 2001. Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cogn Psychol*. **42**:317–367.

Dehaene S, Changeux JP, Naccache L, Sackur J, Sergent C. 2006. Conscious, preconscious, and subliminal processing: a testable taxonomy. *Trends Cogn Sci*. **10**:204–211.

Dell GS. 1986. A spreading-activation theory of retrieval in sentence production. *Psychol Rev*. **93**:283.

Dell GS, Chang F. 2014. The P-chain: Relating sentence production and its disorders to comprehension and acquisition. *Philos Trans R Soc B Biol Sci*. **369**:20120394.

Dell GS, O'Seaghdha PG. 1992. Stages of lexical access in language production. *Cognition*. **42**:287–314.

den Hollander J, Jonkers R, Mariën P, Bastiaanse R. 2019. Identifying the speech production stages in early and late adulthood by using electroencephalography. *Front Hum Neurosci*. **13**:298.

Dikker S, Wan L, Davidesco I, Kaggen L, Oostrik M, McClintock J, Rowland J, Michalareas G, van Bavel JJ, Ding M et al. 2017. Brain-to-brain synchrony tracks real-world dynamic group interactions in the classroom. *Curr Biol*. **27**:1375–1380.

Di Liberto GM, Wong D, Melnik GA, de Cheveigné A. 2019. Low-frequency cortical responses to natural speech reflect probabilistic phonotactics. *Neuroimage*. **196**:237–247.

Dubarry AS, Llorens A, Trébuchon A, Carron R, Liégeois-Chauvel C, Bénar CG, Alario FX. 2017. Estimating parallel processing in a language task using single-trial intracerebral electroencephalography. *Psychol Sci*. **28**:414–426.

Dufour S, Brunellière A, Frauenfelder UH. 2013. Tracking the time course of word-frequency effects in auditory word recognition with event-related potentials. *Cognit Sci*. **37**:489–507.

Dufour S, Bolger D, Massol S, Holcomb PJ, Grainger J. 2017. On the locus of talker-specificity effects in spoken word recognition: an ERP study with dichotic priming. *Lang Cogn Neurosci*. **32**:1273–1289.

Duñabeitia JA, Crepaldi D, Meyer AS, New B, Pliatsikas C, Smolka E, Brysbaert M. 2018. MultiPic: A standardized set of 750 drawings with norms for six European languages. *Q J Exp Psychol*. **71**:808–816.

Engel AK, Fries P, Singer W. 2001. Dynamic predictions: oscillations and synchrony in top–down processing. *Nat Rev Neurosci*. **2**:704–716.

Fairs A, Strijkers K. 2021. Can we use the internet to study speech production? Yes we can! Evidence contrasting online versus laboratory naming latencies and errors. *Psych Archive*. 10.31234/osf.io/2bu4c.

Fairs A, Strijkers K, Michelas A, Dufour S. 2020. Evidence for simultaneous lexical and sublexical access in production and perception & roles of item variation. In: LiveMEEG October 2020.

Federmeier KD. 2007. Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*. **44**:491–505.

Feng C, Damian MF, Qu Q. 2021. Parallel Processing of semantics and phonology in spoken production: evidence from blocked cyclic picture naming and EEG. *J Cogn Neurosci*. **33**:725–738.

Fox J, Weisberg S. 2019. *An {R} companion to applied regression*. 3rd ed. Sage Publications: Thousand Oaks.

Friederici AD. 2002. Towards a neural basis of auditory sentence processing. *Trends Cogn Sci*. **6**:78–84.

Friederici AD. 2011. The brain basis of language processing: from structure to function. *Physiol Rev*. **91**:1357–1392.

Friederici AD, Pfeifer E, Hahne A. 1993. Event-related brain potentials during natural speech processing: Effects of semantic, morphological and syntactic violations. *Cogn Brain Res*. **1**:183–192.

Fuster JM. 1995. *Memory in the cerebral cortex: An empirical approach to neural networks in the human and nonhuman primate*. United Kingdom: The MIT Press.

Fuster JM. 2003. *Cortex and mind: Unifying cognition*. Oxford: Oxford University Press.

Ganushchak L, Christoffels I, Schiller NO. 2011. The use of electroencephalography in language production research: a review. *Front Psychol*. **2**:208.

Garagnani M, Wennekers T, Pulvermüller F. 2008. A neuroanatomically grounded Hebbian-learning model of attention–language interactions in the human brain. *Eur J Neurosci*. **27**:492–513.

Grainger J, Holcomb PJ. 2009. Watching the word go by: On the time-course of component processes in visual word recognition. *Lang Ling Compass*. **3**:128–156.

Graves WW, Grabowski TJ, Mehta S, Gordon JK. 2007. A neural signature of phonological access: distinguishing the effects of word frequency from familiarity and length in overt picture naming. *J Cogn Neurosci*. **19**:617–631.

Grisoni L, Dreyer FR, Pulvermüller F. 2016. Somatotopic semantic priming and prediction in the motor system. *Cereb Cortex*. **26**:2353–2366.

Hagoort P. 2008. The fractionation of spoken language understanding by measuring electrical and magnetic brain signals. *Philos Trans R Soc B Biol Sci*. **363**:1055–1069.

Hagoort P, Brown CM. 2000. ERP effects of listening to speech: semantic ERP effects. *Neuropsychologia*. **38**:1518–1530.

Hagoort P, Indefrey P. 2014. The neurobiology of language beyond single words. *Ann Rev Neurosci*. **37**:347–362.

Halgren E, Dhond RP, Christensen N, Van Petten C, Marinkovic K, Lewine JD, Dale AM. 2002. N400-like magnetoencephalography responses modulated by semantic context, word frequency, and lexical class in sentences. *Neuroimage*. **17**:1101–1116.

Hasson U, Ghazanfar AA, Galantucci B, Garrod S, Keysers C. 2012. Brain-to-brain coupling: a mechanism for creating and sharing a social world. *Trends Cogn Sci*. **16**:114–121.

Hauk O. 2016. Only time will tell–why temporal information is essential for our neuroscientific understanding of semantics. *Psychon Bull Rev*. **23**:1072–1079.

Hauk O, Davis MH, Ford M, Pulvermüller F, Marslen-Wilson WD. 2006. The time course of visual word recognition as revealed by linear regression analysis of ERP data. *Neuroimage*. **30**:1383–1400.

Hebb DO. 1949. *The organization of behavior: A neuropsychological theory*. United Kingdom: Psychology Press.

Hickok G. 2012. Computational neuroanatomy of speech production. *Nat Rev Neurosci*. **13**:135–145.

Hickok G, Poeppel D. 2007. The cortical organization of speech processing. *Nat Rev Neurosci*. **8**:393–402.

Holcomb PJ, Neville HJ. 1990. Auditory and visual semantic priming in lexical decision: A comparison using event-related brain potentials. *Lang Cognit Process*. **5**:281–312.

Hunter CR. 2013. Early effects of neighborhood density and phonotactic probability of spoken words on event-related potentials. *Brain Lang*. **127**:463–474.

Indefrey P. 2011. The spatial and temporal signatures of word production components: a critical update. *Front Psychol*. **2**:255.

Indefrey P. 2016. On putative shortcomings and dangerous future avenues: response to Strijkers & Costa. *Lang Cogn Neurosci*. **31**:517–520.

Indefrey P, Levelt WJ. 2004. The spatial and temporal signatures of word production components. *Cognition* **92**:101–144.

Janssen N, van der Meij M, López-Pérez PJ, Barber HA. 2020. Exploring the temporal dynamics of speech production with EEG and group ICA. *Sci Rep*. **10**:1–14.

Laganaro M, Alario FX. 2006. On the locus of the syllable frequency effect in speech production. *J Mem Lang*. **55**:178–196.

Laganaro M, Morand S, Schnider A. 2009. Time course of evoked-potential changes in different forms of anomia in aphasia. *J Cogn Neurosci* **21**:1499–1510.

Lenth R. 2020. *Emmeans: Estimated marginal means, aka least-squares means*. R package version 1.5.1.

Levelt WJ, Praamstra P, Meyer AS, Helenius P, Salmelin R. 1998. An MEG study of picture naming. *J Cogn Neurosci*. **10**:553–567.

Levelt WJ, Roelofs A, Meyer AS. 1999. A theory of lexical access in speech production. *Behav Brain Sic*. **22**:1–38.

Levelt WJ, Wheeldon L. 1994. Do speakers have access to a mental syllabary? *Cognition*. **50**:239–269.

MacGregor LJ, Pulvermüller F, Van Casteren M, Shtyrov Y. 2012. Ultra-rapid access to words in the brain. *Nat Commun*. **3**:1–7.

Mahon BZ, Caramazza A. 2008. A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *J Physiol Paris*. **102**:59–70.

Mahon BZ, Navarrete E. 2016. Modelling lexical access in speech production as a ballistic process. *Lang Cogn Neurosci*. **31**:521–523.

Martin CD, Branzi FM, Bar M. 2018. Prediction is Production: The missing link between language production and comprehension. *Sci Rep*. **8**:1–9.

Matchin W, Wood E. 2020. Syntax-sensitive regions of the posterior inferior frontal gyrus and the posterior temporal lobe are differentially recruited by production and perception. *Cereb Cortex Commun*. **1**:tgaa029.

Matsumoto R, Nair DR, LaPresto E, Bingaman W, Shibasaki H, Lüders HO. 2007. Functional connectivity in human cortical motor system: a cortico-cortical evoked potential study. *Brain*. **130**:181–197.

Menenti L, Gierhan SM, Segaert K, Hagoort P. 2011. Shared language: overlap and segregation of the neuronal infrastructure for speaking and listening revealed by functional MRI. *Psychol Sci*. **22**:1173–1182.

Mesulam MM. 1990. Large-scale neurocognitive networks and distributed processing for attention, language, and memory. *Ann Neurol*. **28**:597–613.

Miller R. 1996. *Axonal conduction time and human cerebral laterality: A psychological theory*. United Kingdom: CRC Press.

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. *Cereb Cortex*. **25**:3343–3355.

Näätänen R, Paavilainen P, Rinne T, Alho K. 2007. The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clin Neurophysiol*. **118**:2544–2590.

New B, Pallier C, Brysbaert M, Ferrand L. 2004. Lexique 2: A new French lexical database. *Behav Res Methods*. **36**:516–524.

O'Haver T. 2020. iPeak (https://www.mathworks.com/matlabcentral/fileexchange/23850-ipeak), MATLAB Central File Exchange.

O'Rourke TB, Holcomb PJ. 2002. Electrophysiological evidence for the efficiency of spoken word processing. *Biol Psychol*. **60**:121–150.

Osterhout L, Holcomb PJ. 1995. Event-related potentials and language comprehension. *Electrophysiol Mind*. 171–215.

Oostenveld R, Fries P, Maris E, Schoffelen JM. 2011. FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput Intell Neurosci*. **2011**:1–9.

Pickering MJ, Garrod S. 2007. Do people use language production to make predictions during comprehension? *Trends Cogn Sci* **11**:105–110.

Pickering MJ, Garrod S. 2013. An integrated theory of language production and comprehension. *Behav Brain Sci*. **36**:329–347.

Pickering MJ, Garrod S. 2014. Neural integration of language production and comprehension. *Proc Natl Acad Sci*. **111**:15291–15292.

Price CJ. 2012. A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *Neuroimage*. **62**:816–847.

Pulvermüller F. 1999. Words in the brain's language. *Behav Brain Sci*. **22**:253–279.

Pulvermüller F. 2002. *The neuroscience of language: On brain circuits of words and serial order*. United Kingdom: Cambridge University Press.

Pulvermüller F. 2018. Neural reuse of action perception circuits for language, concepts and communication. *Prog Neurobiol*. **160**:1–44.

Pulvermüller F, Fadiga L. 2010. Active perception: sensorimotor circuits as a cortical basis for language. *Nat Rev Neurosci*. **11**:351–360.

Pulvermüller F, Shtyrov Y, Hauk O. 2009. Understanding in an instant: neurophysiological evidence for mechanistic language circuits in the brain. *Brain Lang*. **110**:81–94.

Pulvermüller F, Shtyrov Y, Ilmoniemi R. 2005. Brain signatures of meaning access in action word recognition. *J Cogn Neurosci*. **17**:884–892.

Pylkkänen L. 2019. The neural basis of combinatory syntax and semantics. *Science*. **366**:62–66.

Pylkkänen L, Stringfellow A, Marantz A. 2002. Neuromagnetic evidence for the timing of lexical activation: An MEG component sensitive to phonotactic probability but not to neighborhood density. *Brain Lang*. **81**:666–678.

Python G, Fargier R, Laganaro M. 2018. ERP evidence of distinct processes underlying semantic facilitation and interference in word production. *Cortex*. **99**:1–12.

R Core Team 2020. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Riès SK, Dhillon RK, Clarke A, King-Stephens D, Laxer KD, Weber PB, Kuperman RA, Auguste KI, Brunner P, Schalk G et al. 2017. Spatiotemporal dynamics of word retrieval in speech production revealed by cortical high-frequency band activity. *Proc Natl Acad Sci*. **114**:E4530–E4538.

Roux F, Parish GM, Chelvarajah R, Rollings DT, Sawlani V, Hamer H et al. 2021. Oscillations support co-firing of neurons in the service of human memory formation. *bioRxiv* 28 January 2021. 10.1101/2021.01.28.428480 preprint: not peer reviewed.

Runnqvist E, Bonnard M, Gauvin HS, Attarian S, Trébuchon A, Hartsuiker RJ, Alario FX. 2016. Internal modeling of upcoming speech: A causal role of the right posterior cerebellum in non-motor aspects of language production. *Cortex*. **81**:203–214.

Runnqvist E, Chanoine V, Strijkers K, Patamadilok C, Bonnard M, Nazarian B et al. 2021. Cerebellar and Cortical Correlates of Internal and External Speech Error Monitoring. *Cereb Cortex Commun*. https://doi.org/10.1093/texcom/tgab038.

Sahin NT, Pinker S, Cash SS, Schomer D, Halgren E. 2009. Sequential processing of lexical, grammatical, and phonological information within Broca's area. *Science*. **326**:445–449.

Salmelin R, Hari R, Lounasmaa OV, Sams M. 1994. Dynamics of brain activation during picture naming. *Nature*. **368**:463–465.

Scott SK, Johnsrude IS. 2003. The neuroanatomical and functional organization of speech perception. *Trends Neurosci*. **26**:100–107.

Scott SK, McGettigan C, Eisner F. 2009. A little more conversation, a little less action—candidate roles for the motor cortex in speech perception. *Nat Rev Neurosci*. **10**:295–302.

Schomers MR, Garagnani M, Pulvermüller F. 2017. Neurocomputational consequences of evolutionary connectivity changes in perisylvian language cortex. *J Neurosci*. **37**:3045–3055.

Schoot L, Hagoort P, Segaert K. 2016. What can we learn from a two-brain approach to verbal interaction? *Neurosci Biobehav Rev*. **68**:454–459.

Schuhmann T, Schiller NO, Goebel R, Sack AT. 2012. Speaking of which: dissecting the neurocognitive network of language production in picture naming. *Cereb Cortex*. **22**:701–709.

Segaert K, Menenti L, Weber K, Petersson KM, Hagoort P. 2012. Shared syntax in language production and language comprehension—an fMRI study. *Cereb Cortex*. **22**:1662–1670.

Sereno SC, Rayner K, Posner MI. 1998. Establishing a time-line of word recognition: evidence from eye movements and event-related potentials. *Neuroreport*. **9**:2195–2200.

Shamma S, Patel P, Mukherjee S, Marion G, Khalighinejad B, Han C, Herrero J, Bickel S, Mehta A, Mesgarani N. 2021. Learning Speech Production and Perception through Sensorimotor Interactions. *Cereb Cortex Commun*. **2**:tgaa091.

Silbert LJ, Honey CJ, Simony E, Poeppel D, Hasson U. 2014. Coupled neural systems underlie the production and comprehension of naturalistic narrative speech. *Proc Natl Acad Sci*. **111**:E4687–E4696.

Singer W. 2013. Cortical dynamics revisited. *Trends Cogn Sci*. **17**:616–626.

Singer W, Gray CM. 1995. Visual feature integration and the temporal correlation hypothesis. *Annu Rev Neurosci*. **18**:555–586.

Shtyrov Y, Butorina A, Nikolaeva A, Stroganova T. 2014. Automatic ultrarapid activation and inhibition of cortical motor systems in spoken word comprehension. *Proc Natl Acad Sci*. **111**:E1918–E1923.

Skipper JI, Devlin JT, Lametti DR. 2017. The hearing ear is always found close to the speaking tongue: Review of the role of the motor system in speech perception. *Brain Lang*. **164**:77–105.

Stephens GJ, Silber LJ, Hasson U. 2010. Speaker–listener neural coupling underlies successful communication. *Proc Natl Acad Sci*. **107**:14425–14430.

Strijkers K. 2016. A neural assembly–based view on word production: The bilingual test case. *Lang Learn*. **66**:92–131.

Strijkers K, Bertrand D, Grainger J. 2015. Seeing the same words differently: The time course of automaticity and top–down intention in reading. *J Cogn Neurosci*. **27**:1542–1551.

Strijkers K, Chanoine V, Munding D, Dubarry AS, Trébuchon A, Badier JM, Alario FX. 2019. Grammatical class modulates the (left) inferior frontal gyrus within 100 milliseconds when syntactic context is predictive. *Sci Rep*. **9**:1–13.

Strijkers K, Costa A. 2011. Riding the lexical speedway: A critical review on the time course of lexical selection in speech production. *Front Psychol*. **2**:356.

Strijkers K, Costa A. 2016. The cortical dynamics of speaking: Present shortcomings and future avenues. *Lang Cogn Neurosci*. **31**:484–503.

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.

Strijkers K, Costa A, Thierry G. 2010. Tracking lexical access in speech production: electrophysiological correlates of word frequency and cognate effects. *Cereb Cortex*. **20**:912–928.

Strijkers K, Holcomb PJ, Costa A. 2011. Conscious intention to speak proactively facilitates lexical access during overt object naming. *J Mem Lang*. **65**:345–362.

Tourville JA, Guenther FH. 2011. The DIVA model: A neural theory of speech acquisition and production. *Lang Cognit Process*. **26**:952–981.

Tsodyks MV, Markram H. 1997. The neural code between neocortical pyramidal neurons depends on neurotransmitter release probability. *Proc Natl Acad Sci*. **94**:719–723.

Valente A, Bürki A, Laganaro M. 2014. ERP correlates of word production predictors in picture naming: a trial by trial multiple regression analysis from stimulus onset to response. *Front Neurosci*. **8**:390.

Van Petten C, Coulson S, Rubin S, Plante E, Parks M. 1999. Time course of word identification and semantic integration in spoken language. *J Exp Psychol Learn Mem Cogn*. **25**:394.

Van Turennout M, Hagoort P, Brown CM. 1998. Brain activity during speaking: From syntax to phonology in 40 milliseconds. *Science*. **280**:572–574.

Vihla M, Laine M, Salmelin R. 2006. Cortical dynamics of visual/semantic vs. phonological analysis in picture confrontation. *Neuroimage*. **33**:732–738.

Vitevitch MS, Armbrüster J, Chu S. 2004. Sublexical and lexical representations in speech production: effects of phonotactic probability and onset density. *J Exp Psychol Learn Mem Cogn*. **30**:514.

Vitevitch MS, Luce PA. 1998. When words compete: Levels of processing in perception of spoken words. *Psychol Sci*. **9**:325–329.

Vitevitch MS, Luce PA, Pisoni DB, Auer ET. 1999. Phonotactics, neighborhood activation, and lexical access for spoken words. *Brain Lang*. **68**:306–311.

Winsler K, Midgley KJ, Grainger J, Holcomb PJ. 2018. An electrophysiological megastudy of spoken word recognition. *Lang Cogn Neurosci*. **33**:1063–1082.