Accepted Manuscript

The cortical dynamics of speaking: Lexical and phonological knowledge simultaneously recruit the frontal and temporal cortex within 200 ms

Kristof Strijkers, Albert Costa, Friedemann Pulvermüller

Please cite this article as: Strijkers, K., Costa, A., Pulvermüller, F., The cortical dynamics of speaking: Lexical and phonological knowledge simultaneously recruit the frontal and temporal cortex within 200 ms, *NeuroImage* (2017), doi: 10.1016/j.neuroimage.2017.09.041.

The cortical dynamics of speaking: Lexical and phonological knowledge simultaneously recruit the frontal and temporal cortex within 200 ms.

Kristof Strijkers[1], Albert Costa[2,3] & Friedemann Pulvermüller[4,5].

[1]Aix-Marseille Université (AMU) & CNRS, Laboratoire Parole et Langage (LPL), UMR 7309, 13100,

Aix-en-Provence, France.

[2]Center for Brain and Cognition, CBC, Universitat Pompeu Fabra, Barcelona, Spain.

[3]Institució Catalana de Recerca i Estudis Avançats, ICREA, Barcelona, Spain.

[4]Brain Language Laboratory, Department of Philosophy and Humanities, WE4, Freie Universität Berlin, 14195 Berlin, Germany

[5]Berlin School of Mind and Brain, Humboldt Universität zu Berlin, 10099 Berlin, Germany.

Running title: The cortical dynamics of speaking.

Address for correspondence:

Dr. Kristof Strijkers

Centre National de la Recherche Scientifique (CNRS)

Laboratoire Parole et Langage (LPL) – Université d'Aix-Marseille

3, place Victor Hugo

13331 Marseille, France

Kristof.Strijkers@gmail.com

Abstract

Language production models typically assume that retrieving a word for articulation is a sequential process with substantial functional delays between conceptual, lexical, phonological and motor processing, respectively. Nevertheless, explicit evidence contrasting the spatiotemporal dynamics between different word production components is scarce. Here, using anatomically constrained magnetoencephalography during overt meaningful speech production, we explore the speed with which lexico-semantic versus acoustic-articulatory information of a to-be-uttered word become first neurophysiologically manifest in the cerebral cortex. We demonstrate early modulations of brain activity by the lexical frequency of a word in the temporal cortex and the left inferior frontal gyrus, simultaneously with activity in the motor and the posterior superior temporal cortex reflecting articulatory-acoustic phonological features (+LABIAL vs. +CORONAL) of the word-initial speech sounds (e.g., *M*onkey vs. *D*onkey). The specific nature of the spatiotemporal pattern correlating with a word's frequency and initial phoneme demonstrates that, in the course of speech planning, lexico-semantic and phonological-articulatory processes emerge together rapidly, drawing in parallel on temporal and frontal cortex. This novel finding calls for revisions of current brain language theories of word production.

THE CORTICAL DYNAMICS OF SPEAKING

Efficient and rapid communication is essential for the survival of humans. Indeed, being able to quickly notify our peers of upcoming dangers and problems has high biological relevance and selectivity, and the speed and ease with which we utter complex combinations of words to express our intentions has made speaking our primary communicative tool. Despite this established and necessary speed, the dominant models of brain language mechanisms still suggest that word production is a slow-moving sequential process with substantial functional delays between conceptual, lexical, phonological and articulatory activation, respectively (e.g. Indefrey & Levelt, 2004; Indefrey, 2011). Here we test this assumption and ask whether magnetoencephalography (MEG) and cutting edge source localization production makes it possible to trace early brain correlates of speech production and to localize their origin in the human cortex.

Contrary to the slow activation time courses thought to engender word production, recent neurophysiological research demonstrated very fast brain correlates of speech comprehension. Already within 100-200 ms after the perceptual input, the brain response indexes the processing of its sound structure, syntactic embedding and meaning (e.g., Chanceaux et al., 2012; MacGregor et al., 2012; Näätänen et al., 2007; Pulvermüller et al., 2009). Furthermore, these early neurophysiological differences can be localized to anterior and posterior, and dorsal and ventral brain systems (e.g., Pulvermüller & Fadiga, 2010; Kiefer & Pulvermüller, 2012). This evidence suggests rapid, parallel and distributed mechanisms underpinning the neurobiology of speech comprehension; a view which is notably different from the slow progression through a cascade of processing stages that dominates language production theories (e.g., Caramazza, 1997; Dell, 1986; Levelt et al., 1999). At the cortical level and in line with this sequential, decomposed nature of the word production system, the core components of word production are believed to be subserved by local brain systems, with one-to-one correspondence between linguistic function and cortical areas, that become active in functionally dissociable time intervals (e.g., Hagoort & Levelt, 2009; Indefrey, 2011; Indefrey & Levelt, 2004; Levelt et al., 1998; Salmelin et al., 1994). However, only very few investigations dealing with the cortical dynamics of speech production are available using fine-grained spatiotemporal mapping of brain correlates for distinct word production components (e.g., Strijkers & Costa, 2011; 2016). Here, we aim at filling this gap by tracking the time

course of cortical activations involved in lexico-semantic processes and phonological-articulatory programming. In particular, we trace the spatiotemporal signature of the word frequency effect, a well-established physiological index of lexical processing (e.g., Graves et al., 2007; Sahin et al., 2009; Strijkers et al., 2010; 2011; 2013; Wilson et al., 2009). Crucially, the time-course of the latter is contrasted with the spatiotemporal pattern elicited by articulatory motor programs which differ between speech sounds such as phonemes primarily involving the tip of the tongue (e.g., alveolar [t]) compared with phonemes especially drawing on lip movement (e.g., bilabial [p]).

Most current speech production models imply that conceptual and lexical knowledge is available well before the corresponding phonological and articulatory representations (e.g., Caramazza, 1997; Dell, 1986; Dell & O'Seaghdha, 1992; Levelt et al., 1999). Therefore, these models predict that the earliest brain responses indexing lexico-semantic distinctions related to word frequency precede those of phonological and articulatory processing. In line with this prediction, the dominant spatiotemporal account on word production assumes the following discrete processing sequence housed in specific areas of the left hemisphere (e.g., Indefrey, 2011; Indefrey & Levelt, 2004; Levelt et al., 1998; Salmelin et al., 1994): ~175-250 ms after the onset of the critical stimulus in left anterior middle temporal cortex (MTG): Retrieval of lexico-semantic properties of words; ~250-350 ms in the posterior MTG and superior temporal gyrus (STG): First access to lexical phonological information of the intended words. Further processes necessary for overt speech production are assumed at even later stages and in the left inferior frontal cortex. At ~350-450 ms in the left inferior frontal gyrus (LIFG), a process of syllabification is assumed to be followed by articulation program activation at ~450-600 ms in the inferior premotor and motor areas where the articulators are represented.

Although this specific implementation may be seen to depict a local and strictly serial spatiotemporal progression of temporally non-overlapping processing stages, it is meanwhile well-accepted that the speech production system is not so discrete, but allows for temporal overlap at the different stages (cascading) and functional interactions between the (adjacent) representational levels (e.g., Caramazza, 1997; Dell, 1986; Dell & O'Seaghdha, 1992; Rapp & Goldrick, 2000). This implies that representations downstream in the hierarchy (e.g., phonology) may already become active before critical

processes at upstream levels (lexico-semantics) are completed (e.g., Caramazza, 1997; Costa & Caramazza, 2000; Dell, 1986; Dell et al., 2013; Goldrick et al., 2011; Rapp & Goldrick, 2000). Despite the higher degree of spatiotemporal flexibility compared to Indefrey and Levelt's serial implementation (2004), an interactive activation model still implies that initial spreading of activity across levels is a sequential process where representations lower in the hierarchy become activated significantly before those higher in the hierarchy, separated by functional delays of roughly 100 ms between representational levels (e.g., e.g., Dell & O'Seaghdha, 1992; Laganaro et al., 2012; Llorens et al., 2011; Sahin et al., 2009; Valente et al., 2014). In this manner, the spatiotemporal estimates of word component activation in the Indefrey and Levelt model (2004; Indefrey, 2011) can easily be 'recycled' and representative for sequential interactive activation models as well (e.g., Goldrick et al., 2009; Indefrey, 2011).

That said, there is surprisingly little data available in the language production literature that directly compares the precise time course of cortical area activations elicited by clearly distinct word production components during overt naming (e.g., Strijkers & Costa, 2011; 2016). Such a direct comparison is essential, however, in order to advance in our understanding of the temporal mechanics and neural organization engendering our capacity to speak. This becomes particularly relevant when taking into account that some recent studies of brain indexes during overt speech production, show time courses and cortical sources which are difficult to account for on the basis of the above mentioned slow-sequential activation of local function-specific cortical regions (e.g., Costa et al., 2009; Edwards et al., 2010; Miozzo et al., 2015; Munding et al., 2015; **Ries et al., 2017**; Schuhmann et al., 2012; Strijkers et al., 2010; 2013). For present purposes, two such recent studies are particularly relevant. First, Strijkers and colleagues (2010; see also Strijkers et al., 2013) observed that the lexical frequency effect, the cognate effect (i.e., faster naming latencies for translation words which share phonology) and the language effect (i.e., faster naming latencies in a bilingual's first compared to second language) all produced the same early modulation of electrophysiological responses (P2 latency-range). Especially the latter two effects related to cognates and 1st/2nd language were surprising given that the dimension that traditionally defines these variables is phonological in nature (e.g., Indefrey, 2006; Christoffels et al., 2007). If these effects indeed originate from the phonological processing level, their emergence at <200ms could be seen as a challenge

of classic cascaded models. Rather, the results would potentially suggest parallel retrieval of lexico-semantic (lemma) and lexico-phonological (lexeme) properties during speech preparation. Similar to the work by Strijkers and colleagues, the results of an MEG picture naming study conducted by Miozzo and collaborators (2015) suggested parallelism rather than seriality. The authors performed a multiple-linear regression analysis on the neuromagnetic data of variables related to lexico-semantic (specific semantic features and action features) and lexico-phonological processing (word form features). At around 150 ms after object presentation modulations elicited by lexico-semantic variables became manifest in the left frontotemporal cortex. Importantly, phonological variables elicited modulations in the same latency-range in the left posterior MTG (previously linked to word form processing; e.g., Graves et al., 2007). Although these results suggest near simultaneous lexico-semantic and phonological activations (see also Munding et al., 2015), the data of Strijkers et al. (2010) and Miozzo et al. (2015) only allow for tentative conclusions with respect to the speed with which word production components become active in the brain. One reason why this is so is that neither study could unambiguously separate activation linked to word forms from that linked to lexico-semantic processing. At present it is still uncertain whether variables such as cognate-status solely affect word form processing or may already emerge at the onset of lexical access due to correlations of those variables with earlier lexico-semantic properties (for a detailed account consult, e.g., Strijkers et al., 2010; Strijkers & Costa, 2016). Similarly, in Miozzo et al. (2015), the authors interpret a compound variable as phonological that included word frequency and therefore lexical properties, so that any firm conclusion on early phonology would require further study. Another, related issue, is that according to some authors there is no functional (and thus temporal) division between lexico-semantic (lemma) and word form (lexeme) processing (e.g., Caramazza, 1997).

The goal of the present study was to further explore the time course of brain activations associated to clearly temporally and spatially distinct word production components (as hypothesized by sequential hierarchical brain language theories). As in Strijkers et al. (2010) and Miozzo et al. (2015) we explored the spatiotemporal activation of word production without restricting analyses and interpretation by a-priori defined serial or interactive theoretical frameworks, and by utilizing a paradigm that requires overt and immediate speech production. But contrary to Strijkers et al. (2010) and Miozzo et al. (2015),

we aimed at circumventing the confounding factor of correlated activity between the lexico-semantic and lexico-phonological variables by targeting word production processes, which according to traditional speech production theories should be clearly dissociable in time and space.

Here we contrast the spatiotemporal brain activation (by means of MEG recorded during overt object naming) elicited by words that vary in their lexico-semantic versus phonological and articulatory properties. In order to tap into lexico-semantic processes we manipulated the word frequency of the object names, since this psycholinguistic variable is an excellent index of the onset of lexical access (e.g., Graves et al., 2007; Sahin et al., 2009; Strijkers et al., 2010; 2013; Wilson et al., 2009). In object naming the word frequency effect refers to the observation that object names we often produce are processed faster compared to object names we rarely utter (e.g., Oldfield & Wingfield, 1965). The time course of this effect has been traced with electrophysiological techniques to initiate within the first 200 ms of processing (e.g., Baus et al., 2014; Sahin et al., 2009; Strijkers et al., 2010; 2013), and located with fine-grained spatial measures to regions such as the LIFG, the MTG, the inferior temporal cortex (IT) and the STG (e.g., Fiez et al., 1999; Graves et al., 2007; Sahin et al., 2009; Wilson et al., 2009).

To investigate the neurophysiological correlates elicited by phonological and articulatory word properties we relied on the sound-related somatotopic activations observed recently in speech sound perception and production: listening to or uttering syllables starting with labial phonemes produce stronger activations in the motor area controlling lip movements, while listening to or uttering syllables starting with coronal tongue phonemes elicit stronger responses for more inferior parts of the motor cortex controlling tongue movements (e.g., Bouchard et al., 2013; Fadiga et al., 2002; Pulvermüller et al., 2006; Wilson et al., 2004). Moreover, when transcranial magnetic stimulation (TMS) is applied to these motor areas, behavioral dissociations in speech sound and word perception become apparent (e.g., D'Ausilio et al., 2009; Meister et al., 2007; Möttönen et al., 2013; Schomers et al., 2015). These results suggest that differently distributed motor circuits are recruited during language perception and that the distribution of these circuits includes information about phoneme categories. Despite that the functional role of this motor cortex activation in speech perception and understanding remains a debated issue (e.g., Cheung et al., 2016; Mahon & Caramazza, 2008; Pulvermüller, 2013), the established observations of

speech sound somatotopy (for recent reviews see: Schomers & Pulvermüller, 2016; Skipper et al., 2017) provide us with a useful contrast for investigating specific phonological-articulatory processes in speech production. This is because the different motor cortex activation topographies reflecting the processing of different phoneme categories provide a variable, which is fully orthogonal to the brain correlates of other psycholinguistic computations such as conceptual and lexical processing. This orthogonality allows us to single-out the independent spatiotemporal contributions of phonemic processes and lexical frequency in language production. In addition, by contrasting a manipulation sensitive to the onset of lexico-semantic word retrieval (word frequency), argued to be reflected in temporal brain regions, with a manipulation sensitive to articulatory word properties in the motor cortex, we do not merely contrast "adjacent" word components restricted to lexical selection (e.g., Strijkers et al., 2010; Miozzo et al., 2015), but offer a more extreme contrast targeting the 'begin' and 'end' points of the neurolinguistic hierarchy.

To this end, we manipulated the articulator movement (lip vs. tongue) linked to the first phoneme of object names while keeping all other phonological properties between conditions constant (e.g., *M*onkey vs. *D*onkey; see Fig. 1). If an object name starts with labial sounds there should be overall more activation in the motor area controlling lip movement (MLip) compared to the motor area controlling tongue movement (MTongue) and vice versa when an object name initiates with coronal tongue sounds. Note furthermore that the feasibility for separating these close proximate sources (lip and tongue sensitive regions in the motor cortex are only a few cm apart) with MEG has been repeatedly demonstrated in the literature (e.g., Hari et al., 1993; Nakamura et al., 1998; Hari & Salmelin, 2012; Shtyrov et al., 2014). Besides the motor cortex, also superior temporal cortex activity could be modulated by this manipulation because of its crucial role in phonological processing (e.g., DeWitt & Rauschecker, 2012; Hickok & Poeppel, 2007; Indefrey & Levelt, 2004; Leonard et al., 2015; Patterson et al., 2002). Furthermore, superior temporal and inferior frontal brain areas are tightly interlinked with each other both anatomically (e.g., Petrides & Pandya, 2009; Rilling et al., 2008) and functionally (e.g., D'Ausilio et al., 2009; Fadiga et al., 2002; **Guenther, 2016**; Pulvermüller et al., 2006; Meister et al., 2007; Möttönen et al., 2013; Wilson et al., 2004).

THE CORTICAL DYNAMICS OF SPEAKING

Having identified specific brain regions that are responsive to the lexico-semantic and phonemic manipulations respectively, the key question under investigation is at which point in time these predicted cortical activations would manifest. This will allow testing the core assumptions of the Indefrey and Levelt model (2004), specifically, and, more generally, constrain cascaded and interactive models with respect to the speed with which word production components interact. Following these models a progressive brain response both in terms of timing and neuronal sources should be observed: The first modulations elicited by word frequency are expected to initiate around 150 ms after stimulus onset in temporal regions such as the MTG and the IT (Indefrey, 2011; Indefrey & Levelt, 2004; Laganaro et al., 2009; Levelt et al., 1998; Maess et al., 2002; Salmelin et al., 1994), and arguably spread afterwards to inferior frontal and superior temporal cortices (Indefrey, 2011; Indefrey & Levelt, 2004; Sahin et al., 2009); next, the distributional modulations driven by the somatotopic speech sound contrast should occur at the earliest around 300 ms in superior temporal areas from where it spreads towards the frontocentral cortex (between 400-600 ms) (Indefrey, 2011; Indefrey & Levelt, 2004; Koester & Schiller, 2008; Laganaro et al., 2009; Levelt et al., 1998; Salmelin et al., 1994). Note also that those models in which the motor cortex is only involved in articulatory preparation would predict an even later time course for the somatotopic manipulation, starting after premotor processing in the LIFG (around 500-600 ms) and only for frontocentral action-related areas (Indefrey, 2011; Indefrey & Levelt, 2004; Levelt et al., 1998)[1].

To summarize, by exploring the time-course of predicted cortical area activations in response to word frequency and word-initial speech sound, the present study allows for assessing the cortical dynamics underpinning lexico-semantic and phonological-articulatory processes in word production. According to the current dominant model in language production, the prediction is to see an early modulation for the lexico-semantic effect in the temporal cortex and a late modulation for the

---

[1] These temporal predictions are based on the specific estimates provided by Indefrey and Levelt (2004; Indefrey, 2011). It would be naïve, however, to assume that these are fixed entities present for any speech act we perform. Therefore, rather than focusing on the *absolute* numbers provided by Indefrey and Levelt (2004), the *relative distance* in time between different word production components is more relevant. Regardless of the specific time of onset, current models estimate the temporal delay between access to lexical properties and phoneme properties around 100 ms, and between lexical properties and articulatory properties around several 100s of ms.

phonological-articulatory effect in frontal cortex. The orthogonal design of the experiment provides an unambiguous test of this anatomy-function local sequential assumption, and thereby allows us to either confirm or re-assess the cortical dynamics of speaking.

Materials and Methods

Participants. 15 native British English speakers, students at Cambridge University (age: 18-23), were paid to take part in the current study. All were right-handed, had normal or corrected-to-normal vision and did not suffer from neurological disorder. Ethical approval was issued by Cambridge Psychology Research Ethics Committee and informed written consent was obtained from all volunteers.

Stimuli. Simple white line drawings presented on a black background were selected (n = 132; 70 from the standardized database of Snodgrass and Vanderwant (1980) and 62 from Google line drawings). Half of the object names started with a labial sound (/b/, /p/, /f/, /v/, and /m/) for which the lip(s) represent the place of articulation and the other half with a coronal tongue sound (/t/, /d/, /s/, /l/ and /k/), for which the tongue is the main articulator. Within each phoneme contrast half of the object names corresponded to a low frequency word and the half to a high frequency word (see Fig. 1). In this manner, an orthogonal stimuli design was created between lexical frequency and word-initial phoneme, where between the different conditions words had identical phonological properties with exception of the first phoneme, which was either lip- or tongue-related, or, in terms of phonological features, [+LABIAL] vs. [+CORONAL] (see Fig. 1). Furthermore, we matched the items per condition on a range of relevant variables: Between the lip and tongue contrasts we matched the items for lexical frequency (mean log frequency lip items = 2.91, **(maximal)** dispersion = 1.38 – 4.51; mean log frequency tongue items = 2.98, dispersion = 1.66 – 4.39), bigram frequency (mean bigram frequency lip items = 3.4, dispersion = 0.9 – 8.1; mean bigram frequency tongue items = 3.6, dispersion = 1.2 – 9.8), phoneme length (mean phoneme length lip items = 4.5, dispersion = 3 - 7; mean phoneme length tongue items = 4.6, dispersion = 3 - 7). For the lexical frequency contrast, there was a clear difference in lexical frequency (mean log frequency low frequency items = 2.41, dispersion = 1.38 – 3.00; mean log frequency high frequency items = 3.50, dispersion = 3.01 – 4.51), but not for bigram frequency (mean bigram frequency low frequency items =

3.5, dispersion = 1.2 – 8.5; mean bigram frequency high frequency items = 3.4, dispersion = 0.9 – 9.8) or phoneme length (mean phoneme length low frequency items = 4.8, dispersion = 3 - 7; mean phoneme length tongue items = 4.3, dispersion = 3 - 7). We also ensured that semantic category membership was kept constant across all conditions (i.e., an equal amount of animals, tools, vehicles, etc.). In addition, 29 students of the University of Cambridge rated on a 7-point Lickert scale all items on the following dimensions: name agreement, familiarity, typicality and action-relatedness. These variables did not show reliable differences between the experimental conditions, except for the expected difference in familiarity between low and high frequency items, with a higher familiarity score for the latter. Finally, to obtain an objective measure of the visual variance between conditions, we calculated the inter-stimulus physical variance (ISPV; Thierry et al., 2007). There was no significant difference in ISPV between the different conditions.

Design. Participants were seated in a magnetically shielded room. Using E-prime software the drawings were presented on a black background in the center of a computer screen. The volunteers were instructed to name aloud the object's name as fast and accurate as possible. Each experimental trial had the following sequence: (1) a fixation cross appeared in the center of the screen during 500 ms; (2) a blank interval replaced the fixation cross for 500 ms; (3) the object was presented on screen for 350 ms; (4) a blank inter-trial interval intervened for 2000 ms. Each participant went through all 132 stimuli twice in completely randomized blocks.

MEG recording and MRI data acquisition. MEG was recorded continuously using a 306-channel whole-head Vectorview system (Elekta Neuromag, Helsinki, Finland) with a bandpass filter from 0.03 to 330 Hz and a sampling rate of 1000 Hz. Vertical and horizontal eye movements were monitored with electrodes placed above and below the left eye and either side of the eyes. EMG activity elicited by mouth movement was monitored with electrodes placed on facial muscles (the obicularis oris and the buccinators). Head position relative to the sensor array was recorded continuously by using five magnetic head-position indicator (HPI) coils. Before the recording, the positions of the HPI coils relative to three anatomical fiducials (nasion, left and right preauricular points) were digitally recorded using a 3-D digitizer (Fastrak Polhemus, Colchester, VA). Approximately 80 additional head points over the scalp

were also digitized to allow the offline reconstruction of the head model and co-registration with individual MRI images. For each participant, high-resolution structural MRI images (T1-weighted) were obtained on a 3 T Tim Trio MR scanner (Siemens, Erlangen, Germany) with $1\times1\times1$ mm isotropic voxels.

MEG data processing. To minimize the contribution of magnetic sources from outside the head and to reduce any within-sensor artifacts, the data from the 306 sensors were preprocessed using the temporal extension of the signal-space separation technique (Taulu & Kajola, 2005), implemented in MaxFilter 2.0.1 software (Elekta Neuromag); correlates of MEG signal originating from external sources were removed and compensation was made for within-block head motion. Using the Minimum Norm Estimates Suite (MNE Suite version 2.6.0, Martinos Center for Biomedical Imaging, Charlestown, MA, USA), average event-related fields (ERFs) starting 100 ms before stimulus onset up to 500 ms after stimulus onset were computed, baseline-corrected (for each individual participant) over the prestimulus period of $-$ 100 to 0 ms and bandpass filtered between 1 and 30 Hz. Epochs were rejected when the magnetic field variation at any gradiometer or magnetometer exceeded 2,000 fT cm^{-1} or 3,500 fT, respectively, or when voltage variation at either bipolar electro-oculograms electrodes or facial muscle electrodes was >150 μV. **As a consequence, a total of 24% (SD: 11%) of the trials were rejected.**

Sensor-level analysis. Overall signal strength of the event-related magnetic fields was quantified as the global signal-to-noise ratio (SNR) across all 306 sensors, including 204 gradiometers and 102 magnetometers. Time-windows for analysis were selected on the basis of prominent peaks identified in the global SNR collapsed over conditions and participants (and smoothened in 5 ms time-steps). The global field power (GFP) showed a triphasic response, with peaks at 115, 180 and 280 ms after stimulus onset displaying a hillock-valley structure which motivated a statistical approach exploring each peak-valley pair separately with a two-level factor Time in order to capture an extensive part of the signal (rather than solely the peak activity). Time-windows of interest were assessed objectively through half-peak measuring (i.e., the median from the peak latency to the immediately preceding through and the proceeding through latencies; e.g., Picton et al., 2000) and kept symmetrical between the 'peak' and 'valley', resulting in the following time-windows of interest for analyses: (100-130 ms, 130-160 ms), (160-200 ms, 200-240 ms) and (260-300 ms, 300-340 ms). Next, the event-related magnetic fields were

THE CORTICAL DYNAMICS OF SPEAKING

quantified as the absolute amplitude of the 102 orthogonal gradiometer pairs by computing the root mean square (RMS) of the amplitudes of the two gradiometers in each pair. The resulting data were used to produce sensor-space grand averages across participants. For each time-window and contrast, a 2 (variable) x 2 (time-course) repeated measures ANOVA assessed significant differences for the frequency contrast and the phoneme contrast, and when relevant a two-tailed paired *t*-test assessed when mean global activation over the entire sensor array differentiated between low and high frequency words on the one hand, and between tongue versus lip first phonemes on the other hand.

Source-level analysis. Cortical sources of the observed neuromagnetic activity were estimated using signals from all 306 sensors and the L2 MNE approach that models the recorded magnetic field distribution with the smallest amount of overall source activity (Hämäläinen & Ilmoniemi, 1994). Individual head models were created for each participant using segmentation algorithms (FreeSurfer 4.3 software, Martinos Center for Biomedical Imaging, Charlestown, MA, USA) to reconstruct the brain's cortical gray matter surface from structural MRI data. Further processing was performed using the MNE Suite 2.6.0 software. The original triangulated cortical surface was down-sampled to a grid by decimating the cortical surface with an average distance between vertices of 5 mm, which resulted in 10,242 vertices in each hemisphere. A single-layer boundary element model containing 5,120 triangles was created from the inner skull surface with a watershed algorithm. Dipole sources were computed with a loose orientation constraint of 0.2 and no depth weighting, and with a regularization of the noise-covariance matrix of 0.1. Current estimates for individual participants were morphed to an average brain using five smoothing steps and, for visualization, grand averaged over all 15 participants.

The Regions of Interest (ROIs) were defined in two steps: First, a data-driven approach was followed by calculating the global source activations elicited to all stimuli and identifying those regions which were significantly different from baseline (thus sensitive to object naming in general) (see Fig 2a). Second, a theory- and prediction-specific approach was followed by selecting from the regions identified in the first step those which are known to be modulated by the word frequency and phoneme manipulations, respectively (see Fig. 2b). In that manner, 5 ROIs were chosen from the global response which have been shown to be sensitive to word frequency in previous fMRI research of object naming

13

THE CORTICAL DYNAMICS OF SPEAKING

(e.g., Graves et al., 2007; Wilson et al., 2009): (1) A region around the mid temporal gyrus and the temporal pole (MTG-TP), (2) a region around the inferior temporal cortex (IT), (3) a region around the left inferior frontal gyrus (LIFG) and two regions surrounding the superior temporal cortex, namely (4) a middle portion of the superior temporal gyrus (mSTG) and (5) a more posterior portion of the STG (pSTG) (see Fig. 2b). These ROIs were created by drawing ellipsoid areas (radii ranging between 0.35-1cm; the MNE suite has a tool to hand draw ellipsoid ROIs which are morphed directly into the cortical surface, for technical details consult: http://www.martinos.org/meg/manuals/MNE-manual-2.7.pdf) around the global activations in response to all stimuli and based (in order to avoid crossing region-boundaries) on the Desikan–Killiany Atlas parcellation of the cortical surface (Desikan et al., 2006), as implemented in the Freesurfer software package (see Fig. 2b). With respect to the somatotopic phoneme manipulation the two ROIs in the motor cortex, an anterior one surrounding the area controlling lip movements (MLip) and a posterior one surrounding the area controlling tongue movements (MTongue), were based on the mean Montreal Neurological Institute (MNI) coordinates of lip and tongue movement and articulation identified in previous research (lips: -56, -8, 46; tongue: -60, -10, 25; Pulvermüller et al., 2006). Furthermore, we included the two STG ROIs, one more anterior (mSTG) and one more posterior (pSTG), given that this part of the brain is generally considered central in the processing of speech sounds (DeWitt & Rauschecker, 2012; Hickok & Poeppel, 2007; Indefrey & Levelt, 2004; Leonard et al., 2015; Patterson et al., 2002).

For statistical analyses we conducted Repeated Measures Analyses of Variance (ANOVAs) of each experimental contrast for those time-windows displaying significant effects in the sensor-level analysis. To explore the word frequency effect Time (2 levels: peak and valley) x ROI (5 levels: MTG-TP, IT, LIFG, mSTG and pSTG) x Frequency (2 levels: high frequency and low frequency) ANOVAs were conducted. To explore the modulations driven by somatotopic phoneme contrast Time (2) x ROI (2 levels: anterior and posterior) x Area (2 levels: Inferior Frontal and Superior Temporal) x Phoneme (2 levels: bilabial and alveolar) ANOVAs were conducted. **Note that this particular statistical specification is chosen because of the strong a-priori prediction that the phoneme contrast should result in a significant and specific Phoneme by ROI interaction in the motor cortex (while for the**

14

superior temporal regions we have no specific predictions beyond that it may be affected by the phoneme manipulation). Where relevant, interactions were broken down by independent ANOVAs and Student t-tests. All reported p-values are Greenhouse-Geisser corrected.

Please note that our main analyses focus on the left hemisphere, given that the predictions and models we test are defined mainly or even exclusively for the left hemisphere. Nevertheless, we point out that ANOVAs including the corresponding ROIs in the right hemisphere displayed significant interactions of hemisphere for both the frequency contrast (p = .002) and phoneme contrast (p = .008), respectively, confirming the validity for focusing our prediction-specific analyses on the left hemisphere. A detailed appreciation of the right hemisphere analyses can be found in appendix A. In addition, we also ran an omnibus ANOVA of the left hemisphere including all ROIs and both variables (word frequency and word-initial phoneme) in order to explore potential interactions between word frequency and word-initial phoneme within the defined ROIs. Since the omnibus ANOVA did not reveal interactions that could compromise the results of our main, prediction-specific analyses (as specified above and of which the results are reported below), the results of these analyses are discussed in Appendix B.

Behavioral Results

Incorrect trials (i.e., no-responses, wrong labels, speech errors) were removed prior to analyses of the data (also for the MEG analyses), but were not analyzed themselves given the very low error-rate (<1%). Naming latencies were extracted from the voice recordings files registered during the experiment. Because of technical problems, naming latencies for 3 participants were not recorded due to technical failure, so the statistical analysis for this variable includes 12 participants. We performed a 2x2 repeated-measures ANOVA on the averaged naming latencies of each subject with Word Frequency and word-initial Phoneme as independent variables. Results displayed the expected main effect of Word Frequency ($F(1,11) = 84.03$, MSE = 1161.44, P < .01), with high frequency words being named faster than low frequency words (see Fig. 3). Equally expectedly, there was no main effect of word-initial Phoneme (F < 1) (since it concerns a distributional manipulation and not a load manipulation as in the case of Word

THE CORTICAL DYNAMICS OF SPEAKING

Frequency) nor an interaction between Word Frequency and Phoneme (F(1,11) = 2.71, MSE = 1278.46, P = .13) (see Fig. 3).

MEG Results

Note that for brevity and clarity only significant effects will be reported (with the exception of non-significant effects that are of importance for data interpretation).

Sensor-level Effects.

The GFP of neurophysiological activity calculated across all sensors revealed a reliable difference between low and high frequency words across the to halves of the time-window of 160 – 240 ms after picture onset with a significant interaction between Word Frequency and Time Course (F(1,14) = 7.44, MSE = .01, P = .02). Whereas at 160-200 ms, there was no word frequency effect, (t(14) < 1), the subsequent window (200-240 ms) showed significantly stronger neuromagnetic responses for naming with low frequency words compared with high frequency ones (t(14) = -2.57, P = .02; see Fig. 4a). Differences in global sensor activity between naming conditions involving object names with coronal tongue vs. labial word-initial phonemes were significant between 160-240 ms and 260-340 ms. Tongue-related [+coronal tongue] items were linked with a stronger neuromagnetic response compared to lip-related [+labial] items (160-240 ms: F(1,14) = 7.66, MSE = .04, P = .01; 260-340 ms: F(1,14) = 9.47, MSE = .07, P = .01) (see Fig. 5a)[2]. Based on these results, the subsequent source-level analyses (the main, prediction-specific objective of the study) focus on the time-windows between 160-240 ms and between 260-340 ms.

Source-level Effects.

160-240 ms - the frequency contrast. In this time-window the naming of objects with low frequency names triggered stronger source activations than those with high frequency names when source

---

[2] Note that the difference between 40-80 ms visible in Fig. 5a was not analyzed given the low magnetic field gradient of the peak activity (see Fig. 2a). Nevertheless, when analyzing the labial vs. coronal tongue activity in this time-window we found no significant effects in sensor-space (P = .14) and, as predicted given the low activity level, no significant effects in source-space (all Ps > .50).

activity from all 5 ROIs under consideration was included in the analysis ($F(1,14) = 7.03$, $P = .02$). The interaction of the factors Frequency and ROI just failed the significance criterion ($F(4,56) = 2.80$, $P = .06$), with independent t-test suggesting dominant differences in MTG-TP ($t(14) = -2.67$, $P = .02$) and LIFG ($t(14) = -2.61$, $P = .02$), along with a trend in the pSTG ($t(14) = -1.95$, $P = .07$) and no effect elsewhere (mSTG: $P = .37$; IT: $P = .16$) (see Fig. 4b). None of the other possible interactions with word frequency were significant (all Ps > .31).

160-240 ms – the phoneme contrast. In this time window, source activations between object names starting with a labial phoneme versus those starting with an coronal tongue phoneme were significantly different ($F(1,14) = 3.79$, $P = .02$). Crucially, however, this phoneme effect was further qualified by a significant Time*ROI*Phoneme interaction ($F(1,14) = 8.03$, $P = .01$) and a significant Area*ROI*Phoneme interaction ($F(1,14) = 10.79$, $P < .01$). To explore these predicted interactions, we performed additional Time*ROI*Phoneme ANOVAs for the anterior and posterior brain areas separately. Across the pre- and post-central gyri, we observed a significant effect of the phoneme factor ($F(1,14) = 26.54$, $P < .01$) and, importantly, a significant interaction between ROI*Phoneme ($F(1,14) = 6.71$, $P = .02$). Independent t-tests showed that this interaction was due to a significant increase of source activation in the MTongue when an object name starting with an coronal tongue was under processing as compared with a labial phoneme ($t(14) = -2.59$, $P = .02$) (see Fig. 5b), and the absence of similar significant for the other sensorimotor cortex ROI (MLip). In the STG, we similarly found significant interactions between ROI*Phoneme ($F(1,14) = 9.14$, $P = .01$) and between Time*ROI*Phoneme ($F(1,14) = 5.62$, $P = .03$). ANOVAs performed separately for the two temporal sub-windows (160-200 ms and 200-240 ms) demonstrated that the ROI*Phoneme interaction only reached significance for the earlier one ($F(1,14) = 9.09$, $P = .01$). Between 160-200 ms a cross-over double dissociation between ROI and articulatory type was apparent: Whereas the more posterior/dorsal part of the STG showed significantly stronger source activity for labial compared to coronal tongue phonemes ($t(14) = -3.04$, $P = .01$), the relatively more anterior/ventral portion of the STG showed the reverse pattern with maximal responses for object names beginning with coronal tongue phonemes ($t(14) = -2.66$, $P = .02$) (see Fig. 5b). This pattern in part

paralleled the situation in sensorimotor pre- and postcentral cortex, where the inferior/ventral area showed similar predominance for tongue-related phonemes and articulations.

260-340 ms – the frequency contrast. In this time-window, stronger source activations for low compared with high frequency words was observed ($F(1,14) = 4.62$, $P = .05$) (see Fig 4b). This frequency effect did not interact with the factors ROIs or Time (Time*Frequency: $F < 1$; ROI*Frequency: $F < 1$; Time*ROI*Frequency: $F(1,14) = 2.34$, $P = .09$), indicating the effect was present for the whole time-window and all 5 ROIs under investigation.

260-340 ms – the phoneme contrast. In this time window, object names beginning with coronal tongue sounds produced significantly stronger source activations compared with those beginning with labial sounds ($F(1,14) = 11.33$, $P < .01$). Furthermore, and in contrast to the earlier time-window, the effect was constant over ROIs for both motor and STG areas, as evidenced by the absence of any interactions (all Ps > .13) (see Fig. 5b). The latter pattern is interesting since it indicates that lip-tongue dissociations linked to the word-initial phoneme are not present across the board, but associated with a particular time-window (i.e., 160-240), as one would expect in terms of the time-course of phoneme processing.

Discussion

We investigated the spatiotemporal dynamics of cortical activation underlying the processing of lexico-semantic and phonological-articulatory features in object naming. Lexico-semantic processing was assessed by manipulating word frequency, whereas articulatory-phonological word properties were targeted by manipulating the sound features and motor movements required to utter the initial speech sound of an object name. The results demonstrate an early effect of word frequency and, in parallel, a simultaneous dependency of local brain responses on the place of articulation of word-initial speech sounds. More specifically, between 160-240 ms after picture presentation, activity in left frontal (LIFG) and temporal (MTG) cortical regions, which are known to be sensitive to a word's lexical frequency, was significantly modulated by this variable. In the exact same time-window, phoneme-specific dissociations in cortical activity became apparent in the sensorimotor cortex and the STG. These data replicate well-

THE CORTICAL DYNAMICS OF SPEAKING

established frontotemporal sources found for both manipulations, but demonstrate, contrary to common believe, that they arise with a similar early temporal profile. The surprisingly rapid and near-simultaneous nature of the neural responses indicating differences in word frequency effect and word initial phonemes document that lexico-semantic and phonological-articulatory processes emerge in parallel in speech production, recruiting both anterior and posterior brain systems early on in the course of speech planning. In what follows, we will break down each of these observations in greater detail, and afterwards discuss how the data constrain brain language theories.

The early brain responses sensitive to the frequency of an object's name (160-240 ms) were present across several areas, including left temporal regions (MTG and temporal pole), along with frontal LIFG. In all of these ROIs, low frequency words produced stronger source activations than high frequency words. This effect remained stable and extended towards more inferior and superior regions of the temporal cortex later in time as well (see Fig. 4b). This pattern confirms previously reported heamodynamic studies, which also showed stronger activation in frontotemporal areas during low vs high frequency word processing (e.g., Fiez et al., 1999; Graves et al., 2007; Hauk et al., 2008; Wilson et al., 2009). The early latency of the frequency effect is also consistent with the results from previous electrophysiological recordings in overt naming experiments (e.g., Baus et al., 2014; Strijkers et al., 2010; 2011; 2013), and combines these real-time measurements with cutting-edge source localization, thereby integrating neurophysiological and heamodynamic results into a single spatiotemporal picture. This demonstrates the cross validity of the findings and supports a scientific approach where neurophysiological measurements are recorded congruently with overt speech production (e.g., Strijkers & Costa, 2011).

In exactly the same early temporal window where the earliest lexical frequency effects were present (160-240 ms), local activation of articulatory motor cortex was differentially modulated by the place of articulation of the initial phoneme of the object name. This differential focal modulation of articulatory activity reflecting phonological information was manifest in a ROI x phoneme type interaction, demonstrating an increase in cortical activity specifically in tongue-related motor and premotor cortex for object names starting with alveolar compared to bilabial sounds (see Fig. 5b). The

phoneme-specific responsiveness of the motor cortex replicates (in part) previous works successfully demonstrating the cortical activation patterns in perisylvian frontal, parietal and temporal areas associated with the production and perception of different phonemes and drew attention to the differences between phoneme specific activation patterns (e.g., Bouchard et al., 2013; Evans & Davis, 2015; Fadiga et al., 2002; Murakami et al., 2015; Pulvermüller et al., 2006). What these previous studies did not clarify, however, is the temporal relationship between any such phoneme specific activation with lexical-semantic processing in language production. Only few previous studies addressed this issue, which is critical to psycholinguistic theories of speech production (e.g., Salmelin et al., 1994; Levelt et al., 1998; Sahin et al., 2009; Strijkers et al., 2010; Miozzo et al., 2015), and none of them used phoneme specific activation, which provides a unique avenue towards specific phoneme-related processes. Our present study is the first observation of such phoneme-specific local modulation of articulatory cortex activity in meaningful language production and links them to the same early time-course as lexico-semantic processing.

While the phoneme-specific differences cannot be explained by co-articulation by context phonemes (e.g., Bouchard & Chang, 2015), as those were exactly balanced between the coronal and labial conditions, it is compelling that these results were obtained in response to many different words and a variety of first phonemes, including voiced and unvoiced stops, fricatives and even nasals (see Methods; Fig. 1). Despite this variance and in agreement with prediction, tongue-related word initiations activated the tongue area more strongly than lip-related ones. The absence of the other arm of the expected double dissociation, i.e. of any relatively stronger brain index of lip-word processing in dorsal lip-related articulatory cortex, is explained by the overall much stronger cortical activity during tongue movements compared with lip movements. In their tongue vs. lip movement localizers as well as in their silent articulatory conditions, Pulvermüller et al. (2006; see also Lotze et al., 2000) found substantially stronger and more widespread activation in sensorimotor cortex when tongue movements and tongue-initial biphone syllables were produced. The stronger and more widespread activations reflecting tongue-related manipulations are due to the fact that the tongue is a larger and biomechanically more complex articulator compared to the lips, therefore occupying a less focal representational space in the motor cortex (e.g., Bartoli et al., 2013; Schomers et al., 2015). Hence, in spite of the lack of a full cross-over pattern, the

phoneme-specific local modulation of neuromagnetic activity in the articulatory cortex revealed by the significant ROI x Phoneme interaction, which furthermore followed the predicted sign (direction) of interaction, demonstrates that the place of articulation of the word initial phoneme is manifest early in the naming process in the motor system's activation topography.

Interestingly **and surprisingly**, this phoneme-specific dissociation in the motor cortex co-occurred with a full double dissociation in the STG between 160-200 ms, with a more superior-posterior portion responding maximally to object names starting with labial phonemes and a more inferior portion responding maximal to object names starting with coronal tongue phonemes (see Fig. 5b). Observing such differential brain response in this particular region fits with a phonological origin of the effect. The STG and adjacent superior temporal sulcus are generally thought to be central for acoustic and phonological processing in both speech perception and production (e.g., DeWitt & Rausschecker, 2012; Hickok & Poeppel, 2007; Indefrey, 2011; Indefrey & Levelt, 2004; Patterson et al., 2002; Shtyrov et al., 2000), and categorical responses to different phonemic input have been demonstrated before with MEG, fMRI and high-density cortical surface recordings (e.g., Chang et al., 2010; Desai et al., 2008; Mesgarani et al., 2014; Obleser et al., 2003). The temporal alignment of this effect with that in the sensorimotor cortex supports models postulating that articulatory-phonological and auditory-phonological representations are merged by action perception circuits spanning frontocentral and superiortemporal parts of the left perisylvian language cortex (e.g., Pulvermüller & Fadiga, 2010).

Moreover, since motor gestures in the inferior frontal cortex are somatotopically organized, thus preserving neighborhood relationships from cortex to motor periphery, and a general neuroscientific principle of cortico-cortical connectivity states such neighborhood preservation through topographical projections (e.g., Braitenberg & Schuz, 1998), indirect somatotopy through frontotemporal connections is predicted in the temporal regions, given the rich dorsal frontotemporal projections between premotor and superiortemporal cortex by way of the arcuate fascicle (e.g., De Schotten et al., 2012; Rilling, 2014). Given this frontotemporal parallelism of region-specific phonemic effects, the dissociations manifest here predict specific connections between anterior-ventral STG and inferior-ventral tongue motor regions, and between posterior-dorsal STG and the more dorsal lip region of the motor cortex; a novel prediction

21

which can be explored in neuroanatomical studies using DTI and tractography. This said, we haste to add that our study remains silent whether or not the parallel activation of the motor and superior temporal cortices also signifies functional causality of activation between those regions. **That is, it offers no proof for this hypothesis**. Nevertheless, the noteworthy spatiotemporal pattern in response to the present phonemic manipulation does emphasize this possibility as an important one to explore in future research.

For now, the significant interactions of the articulator-related activity in both sensorimotor and auditory brain systems show an influence of phonological processes at a much earlier latency than suggested by most current neurocognitive perspectives on language production (e.g., Hagoort & Levelt, 2009; Indefrey, 2011; Indefrey & Levelt, 2004; Koester & Schiller, 2008; Laganaro et al., 2009; Levelt et al., 1998; Sahin et al., 2009; Salmelin et al., 1994). This novel finding coupled with the early frontotemporal manifestations for word frequency provides compelling evidence that lexico-semantic and phonological-articulatory processes are sustained by distributed inter-areal neural circuits that become activated near-simultaneously early on in the speech preparation process[3]. This is the main contribution of this study.

Implications for brain language models

The finding that word frequency modulated activity in the left anterior and posterior-superior temporal cortex, and the left IFG with a similar temporal profile already calls for a revision of serial

---

[3] One potential issue with the lexico-semantic interpretation is that the variable lexical frequency may be sensitive to (correlated with) semantic differences (prior to lexical access) as well (e.g., Graves et al., 2007; Strijkers et al., 2010; see also Introduction). However, this issue is of no relevance for the present objectives and conclusions. First, the available literature investigating the lexical frequency effect in language production with fine-grained temporal measures clearly favors an onset during lexico-semantic processing (e.g., Strijkers & Costa, 2011), and also the neural sources sensitive to lexical frequency in the current study (in particular mid and anterior temporal cortex) correspond with previous work linking those regions to lexico-semantic processing (e.g., Graves et al., 2007; Sahin et al., 2009; Wilson et al., 2009). Second, and even regardless of the above, if our results would indeed index conceptual rather than lexical processing, it would not alter the main conclusion of the current study (i.e., parallel distributed activation of distinct word production components) given that the models we set out to test (e.g., Indefrey & Levelt, 2004; Hagoort & Levelt, 2009; Indefrey, 2011; Hickok, 2012; Dell et al., 2013) all assume that in speech planning conceptual processing occurs prior to lexical processing, and lexical processing prior to phonological and articulatory processing.

models of language production which assume that early lexical processes are confined to the temporal cortex (Indefrey, 2011; Indefrey & Levelt, 2004). It does not, however, speak against the notion of sequential processing of linguistic knowledge per se. Approaches to the brain mechanisms of language, which assign more domain-general functions to the left IFG, such as the unification and control model (e.g., Hagoort, 2005), predict the involvement of this area at each sequentially ordered level of linguistic retrieval (e.g., Hagoort & Levelt, 2009; Sahin et al., 2009). Crucially, however, the results did not only reveal the rapid involvement of frontal and temporal brain areas for the lexico-semantic manipulation, but also for the place of articulation linked to a word's initial phoneme. The combined spatiotemporal patterns found for lexical frequency and the somatotopic phoneme manipulation are at odds with a segregated spatiotemporal organization (e.g., Hagoort & Levelt, 2009; Indefrey, 2011; Indefrey & Levelt, 2004; Levelt et al., 1998; Sahin et al., 2009; Salmelin et al., 1994), and not compatible with the dominant spatiotemporal model of object naming as currently envisioned (Indefrey, 2011; Indefrey & Levelt, 2004). This model predicts a temporal sequence from lexico-semantic to phonological and finally to articulatory information access in object naming, with typical latencies of 150-250ms, 300-400 ms and 500-600 ms after onset of the to-be-named pictures. Our results show simultaneous access to lexical and phonological-articulatory word properties at 160-240ms. In contrast, Indefrey and Levelt's model would have predicted a relative temporal distance of some 100 (between lexical and phonological access) to several 100s of ms (between lexical and articulatory access) both in terms of word component activation as in terms of posterior-anterior neural recruitment. An attempt to reconcile that model with the present data would require to drop the one-to-one mapping of psycholinguistic function to cortical area immanent to it and allow one-to-many mappings instead (i.e., cortical representations of both lexical units and phonemes; see Pulvermüller, 1999; Strijkers & Costa, 2016). A different strategy to reconcile some of Indefrey and Levelt's ideas with the present data, especially since we targeted the first phoneme of a word, might be to maintain that premotor and motor cortex exclusively contributes to articulatory processing but that this process once thought to be substantially delayed in the language production cascade is in fact initiated substantially faster. That is, if the sequential nature of this type of models were to be maintained, the postulates would need to be revised and specified at a much shorter time scale, for example proposing

23

delays of 10-40ms between psycholinguistic process onsets (at the least for word-initial phonemes) and corresponding area activations (for further discussion, see Pulvermüller et al., 2009; Strijkers & Costa, 2016).

Our data thus constrain brain language models of word production in (at least) two important ways: one, the temporal progression of activity cycles through word production components (and their underlying cortical sources) at a much faster rate than traditionally assumed; two, the neural organization of word components relies on distributed frontotemporal circuits, calling for one-to-many mappings between linguistic function and brain function. These spatiotemporal constraints depict a much more flexible form of brain-language integration in word production compared to Indefrey and Levelt's implementation (2004). Such higher flexibility may be provided by the more dynamical interactive and cascaded models of language production (e.g., Caramazza, 1997; Dell, 1986). Implementing such interactive architecture in terms of cortical dynamics is harder than a serial and discrete structure, since the boundaries of linguistic functionality in terms of time-course and space are fuzzier. The present results can offer some hints in this regard. First, the data highlight that interactivity between different linguistic components progresses notably faster than typically assumed, in the range of 10s of ms instead of 100s of ms (e.g., Dell & O'Seaghdha, 1992; Goldrick et al., 2009; Sahin et al., 2009; Llorens et al., 2011). The latter may suggest that the degree of interactivity likely goes beyond "adjacent" processing layers (e.g., Dell & O'Seaghdha, 1992), but rapidly cascades through the whole system allowing for the representations at the lowest levels and those at the highest levels of the linguistic hierarchy to dynamically interact as well. Second, our results indicate that the different representational layers involved in word production, such as lexico-semantics and phonology, are not reflected in the brain by a single specific region or local proximate areas, but rather draw on distributed frontotemporal circuits. This means that the different layers (linked to different word components) in interactive activation models of language production will not follow a strict spatial hierarchy from posterior to anterior functionally-segregated brain areas (i.e., parallel-distributed representations within an hierarchical processing layers, but not across the hierarchy), but need to be distributed in space as well (i.e., parallel-distributed representations crossing the cortical hierarchy; see also Dell et al., 2013).

THE CORTICAL DYNAMICS OF SPEAKING

Current neurocomputational models, which hypothesize a dual-route architecture underpinning language production (e.g., Ueno et al., 2011; Hickok, 2012; 2014; see also e.g., Dell et al., 2013; **Guenther et al., 2006**; **Guenther, 2016**; Pickering & Garrod, 2013), may explain some of the current observations. For example, in Hickok's hierarchical state feedback control (HSFC) model of speech production (2012; 2014; see also **Guenther, 2016**), the activation of lexical representations (stored in mid and anterior temporal cortex) triggers two distinct processing pathways responsible for phonological and articulatory encoding: One to the LIFG and the motor cortex denoting 'production syllables/phonemes' and one to the superior temporal cortex and the supramarginal gyrus denoting 'perception syllables/phonemes'. Selection of the correct phonemes and eventual articulation is achieved when the phonological information in frontal and temporal brain systems matches. As a consequence of this particular neuroanatomy, perception phonemes in auditory cortex may become activated with temporal overlap as the production phonemes stored in the motor cortex; a structure which seems to capture the source activations in response to our phonemic manipulation. That said, and in spite of the refreshing perspective offered by this approach, since the model does not make explicit temporal predictions, it remains an open question whether a hierarchical dual-stream model of language production can indeed account for the early time-course of the parallel effects encountered (e.g., Strijkers, 2015). For instance, the model suggests a temporal sequence whereby Broca's area and superior temporal cortex are first activated, followed by motor cortex and inferior parietal cortex. This claim does not seem to be supported by the current data where sensorimotor and superior temporal areas became simultaneously involved in phonological processing. Similarly, it is not entirely clear how the model would explain the observation that also lexico-semantic processes in frontotemporal networks emerged around the same time as the phonological effects, given that the dual phonological stream should be engaged after lexico-semantic retrieval (e.g., Hickok, 2012; 2014). Of course, as mentioned in the previous paragraph, assuming that cascading over hierarchical layers can operate at a very short time-scale could be sufficient to capture the findings made in the current study.

Another popular brain language model, which makes both spatial and temporal predictions and which can account for all our observations, is built upon Hebb's concept of cell assemblies (1949) and the

influential binding-by-synchrony framework of neural coding (e.g., Singer & Gray, 1995). According to this framework distributed input bind to a coherent mental representation through the synchronization of the firing rates between the 'input neurons' and 'output neurons' across the cortical hierarchy. Neural populations that consistently and coherently are active together in time (temporal correlation) will form an assembly which represents a percept, action or mental event as a functional whole ('gestalt'), while neural populations whose firing rates are out of synchrony will decouple (e.g., Braitenberg, 1978; Singer & Gray, 1995; Gray, 1999; Fries, 2005; Fuster, 2003; Singer, 2013). In other words, neural communication will not only be defined by the hierarchical anatomy of the brain, but also by the gestalt-principles between the input and output of a mental operation. Given that words, in minimalistic terms, concern the consistent binding of meaning with sounds and given that aspects of this link can be mapped by correlation, it has been proposed that meaningful units are cortically implemented as distributed neuronal circuits that bind phonological and semantic information into higher order lexicosemantic units **(for a recent in-depth review see Pulvermuller, 2017)**. Phonological information is embedded into these circuits not as a separate processing level but in the form of phonological subunits of the lexicosemantic circuit, and bound with semantic features that near-simultaneously ignite when a meaningful unit is being activated in the language production process (e.g., Pulvermüller, 1999; 2017; Pulvermüller & Fadiga, 2010; Strijkers, 2016; Strijkers & Costa, 2016).

The results of the present word production study are remarkably consistent with this notion, because a neural assembly model cannot only explain the full spectrum of the present data, but in fact a-priori predicted the several outcomes:

1. The early simultaneous activation of lexical and phonological word properties in picture naming. This is because distributed cell assemblies represent words where semantic, lexical, sound and articulatory word properties grounded in perception, action and domain-general (integrating or 'switching/relaying') brain systems group together and rapidly ignite in parallel.

2. The phoneme-specific modulation of local articulatory motor cortex activation. This is because tongue and lip have different motor representations and these contribute differently to the phonological sub-circuits of word-related cell assemblies.

3. The near-simultaneous frontrocentral articulatory and superiortemporal auditory activation reflecting phonological processing. This is because correlational learning and connectivity structure between articulatory and auditory processing yield distributed perisylvian assemblies that act as integrated action-perception circuits to represent phoneme categories.

Let us stress that the notion of rapid, near-simultaneous activation of different linguistic knowledge organized in distributed frontotemporal brain circuits does not mean that no sequential and local brain responses associated to a particular linguistic operation can be identified (Friederici, 2002; Hagoort, 2005; Holcomb & Grainger, 2006; Indefrey & Levelt, 2004; Levelt et al., 1998; Pulvermüller et al., 2009; Sahin et al., 2009; Salmelin et al., 1994; Van Turennout et al., 1998). Neural assembly models do not deny the presence and even need of local and sequential brain processes to a cognitive event, but rather refute this idea as a sufficient mechanism. While a purely sequential hierarchical model is bound to the cortex based on the principle of forward connectedness between adjacent processing layers, neural assembly theory integrates parallel processing in distributed cell assemblies with hierarchical processing in local cell structures through flexible effective cortical communication (e.g., Braitenberg, 1978; Buszaki, 2010; Engel et al., 2001; Fries, 2005; Fuster, 2003; Hebb, 1949; Singer & Gray, 1995). With respect to language, a neural assembly view predicts that word recognition is biologically underpinned by the rapid ignition of word-related cell assemblies, where small temporal delays in the order of 10s of ms are explained by cortical conduction times (Pulvermüller et al., 1999). Ignition is followed by slower sustained activity due to reverberation of activity in the circuit, which produces well-timed spatiotemporal activation patterns underlying subsequent language operations, including for instance verbal working memory, grammar, semantic integration and articulation (e.g., Pulvermüller, 1999; 2005; Pulvermüller & Fadiga, 2010; Strijkers, 2016; Strijkers & Costa, 2016; Schomers et al., 2017). In other words, a Hebbian-like cell assembly view on words in the brain does not question the local and sequential observations reported for word production (e.g., Indefrey & Levelt, 2004; Indefrey, 2011), but rather the interpretation of some of the previous observations. While initial recognition of all components (e.g., lexico-semantics and phonology) a to-be-uttered word would emerge cortically near-simultaneously and functionally in parallel, linguistic operations upon that core representation, such as grammatical combination, semantic

context integration or motor control may draw on sequential mechanisms that is, by both the later mechanism of reverberation of the activated cell assembly (e.g., motor control: i.e., timed articulation over the motor cortex to be able to sequentially utter the parallel ignited phonemes of a word; e.g., Strijkers, 2016; Strijkers & Costa, 2016), and the serial activation of different cell assembly circuits (e.g., semantic integration of the ignited word assembly into a larger context/discourse)[4]. Exploring in detail such precise activation patterns in future research will be of importance to better understand how the basic neurophysiological elements of the brain are linked to the complex abstract elements of language.

Conclusion.

By isolating the temporal contribution of predicted brain activity in response to lexico-semantic processing versus that of predicted brain activity in response to phonological and articulatory properties, we were able to demonstrate that brain indexes of word production components, which traditional psycholinguistic theories assume to be segregated in time, in fact display simultaneous dynamics. The results question the classic assumptions of discrete one-to-one mappings of psycholinguistic levels of processing onto separate brain regions along with the idea that serial/cascaded processing time steps at a grain size of 100s of milliseconds provide a realistic perspective on the neurobiological basis of language. Rather than linking the components of word production to the brain based on hierarchical discreteness (e.g., Indefrey & 2004; Indefrey, 2011) or through limited interactivity ('globally modular, but locally interactive'; e.g., Dell & O'Seaghdha, 1992; Rapp & Goldrick, 2000; Brehm & Goldrick, 2016), our results call for functional-anatomical interactivity and temporal near-simultaneity or even synchrony. We

---

[4] Besides functional activation differences between word stem 'recognition' (ignition) and linguistic operations upon that word stem (reverberations) such as grammatical inflection or articulatory control to reconcile sequential word component activations with parallel word component activations, another important factor that may affect the sequentiality vs. simultaneity of spatiotemporal word dynamics concerns top-down related task-dependent modulations (e.g., Fargier & Laganaro, 2016; Strijkers & Costa, 2016). That is, the nature and speed of word component activation may vary in function of the type of speech act engaged or the goal-directed behaviour of the speaker. Regardless this interesting and important issue (which is beyond the current scope), at least for overt object naming, the task on which most of the models discussed here are mainly (or even exclusively) based, the data highlight a notably faster and more widespread cortical activation dynamic of word production than traditionally assumed.

have suggested that this may be achieved, in part, by dynamical hierarchical models allowing for fast interactivity subserved by two or more cortical processing streams (e.g., Ueno et al., 2011; **Guenther, 2016**; Hickok; 2012; 2014; Dell et al., 2013; Pickering & Garrod, 2013) (although many traditional serial aspects seem immanent to these proposals and will need to be reassessed). An alternative and new perspective is offered by non-hierarchical neural assembly models where symbol representations in the brain act as 'gestalts' both in space and time (e.g., Pulvermüller, 1999; Pulvermüller & Fadiga, 2010; Strijkers, 2016; Strijkers & Costa, 2016). Regardless whether our data will call for dropping the traditional sequential hierarchical view that dominates word production theories, the main contribution here is the demonstration that slow-scale sequentiality by itself cannot be the whole story, as during simple overt language production lexico-semantic and phonological-articulatory word properties draw in parallel on temporal and frontal cortices within the first 200 ms of speech planning.

References

Bartoli E et al. 2013. Listener–speaker perceived distance predicts the degree of motor contribution to speech perception. Cereb. Cortex bht257.

Baus C. et al. 2014. On predicting others' words: Electrophysiological evidence of prediction in speech production. Cognition 133:395-407.

Bouchard KE, Mesgarani N, Johnson K, Chang EF. 2013. Functional organization of human sensorimotor cortex for speech articulation. Nature 495:327-332.

Bouchard, KE, Chang, EF. 2014. Control of spoken vowel acoustics and the influence of phonetic context in human speech sensorimotor cortex. J. Neurosci. 34:12662-12677.

Braitenberg V. 1978. Cell assemblies in the cerebral cortex. In Theoretical approaches to complex systems (pp. 171-188). Springer Berlin Heidelberg.

Braitenberg V, Schüz A. 1998. Statistics and Geometry of Neuronal Connectivity (Berlin: Springer).

Brehm L, Goldrick M. 2016. Empirical and conceptual challenges for neurocognitive theories of language production. Lang. Cogn. Neurosci. 1-4.

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

THE CORTICAL DYNAMICS OF SPEAKING

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

Chanceaux M, Vitu F, Bendahman L, Thorpe S, Grainger J. 2012. Word processing speed in peripheral vision measured with a saccadic choice task. Vision Res. 56:10–19.

Chang EF, et al. 2010. Categorical speech representation in human superior temporal gyrus. Nat. Neurosci. 13:1428-1432.

Cheung, C, Hamilton, LS, Johnson, K, Chang, EF. 2016. The auditory representation of speech sounds in human motor cortex. Elife, 5, e12577.

Christoffels IK, Firk C, Schiller NO. 2007. Bilingual language control: An event-related brain potential study. Brain Res. 1147:192-208.

Costa A, Caramazza A, Sebastian-Galles N. 2000. The cognate facilitation effect: implications for models of lexical access. J. Exp. Psychol.: Learn. Mem. Cogn. 26(5): 1283-1296.

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. USA. 106:21442-21446.

D'Ausilio A, et al. 2009. The motor somatotopy of speech perception. Curr. Biol. 19:381–385.

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

Dell GS. 1986. A spreading-activation theory of retrieval in sentence production. Psychol. Rev. 93:283–321.

Dell GS, Schwartz MF, Nozari N, Faseyitan O, Coslett HB. 2013. Voxel-based lesion-parameter mapping: Identifying the neural correlates of a computational model of word production. Cognition, 128(3): 380-396.

Desai R, Liebenthal E, Waldron E, Binder JR. 2008. Left posterior temporal regions are sensitive to auditory categorization. J. Cogn. Neurosci. 20:1174-1188.

Desikan RS, et al. 2006. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. NeuroImage. 31:968-980.

De Schotten MT, Dell'Acqua F, Valabregue R, Catani M. 2012. Monkey to human comparative anatomy of the frontal lobe association tracts. Cortex 48:82-96.

DeWitt I, Rauschecker J. 2012. Phoneme and word recognition in the auditory ventral stream. Proc. Natl. Acad. Sci. USA. 109:E505-E514.

Edwards E, et al. 2010. Spatiotemporal imaging of cortical activation during verb generation and picture naming. Neuroimage 50:291–301.

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

Evans S, Davis MH. 2015. Hierarchical organization of auditory and motor representations in speech perception: evidence from searchlight similarity analysis. Cereb. Cortex 25:4772-4788.

Fadiga L, Craighero L, Buccino G, Rizzolatti G. 2002. Speech listening specifically modulates the excitability of tongue muscles: a TMS study. Eur. J. Neurosci. 15:399-402.

Fargier R, Laganaro M. 2016. Spatio-temporal Dynamics of Referential and Inferential Naming: Different Brain and Cognitive Operations to Lexical Selection. Brain Topogr. 1-16.

Fiez JA, Balota DA, Raichle ME, Petersen SE. 1999. Effects of lexicality, frequency, and spelling-to-sound consistency on the functional anatomy of reading. Neuron. 24:205–218.

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

Fries P. 2005. A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. Trends Cogn. Sci. 9:474-480.

Fuster JM. 2003. Cortex and Mind: Unifying Cognition (Oxford, UK: Oxford Univ. Press).

Goldrick M, Dell GS, Kroll J, Rapp B. 2009. Sequential information processing and limited interaction in language production. *Science letters online*.

Goldrick M, Baker HR, Murphy A, Baese-Berk M. 2011. Interaction and representational integration: Evidence from speech errors. Cognition, 121(1), 58-72.

Gray CM. 1999. The temporal correlation hypothesis of visual feature integration: still alive and well. Neuron. 24: 31-47.

THE CORTICAL DYNAMICS OF SPEAKING

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.

Guenther FH, Ghosh SS, Tourville JA. 2006. Neural modeling and imaging of the cortical interactions underlying syllable production. Brain Lang. 96:280-301.

Guenther FH. 2016. *Neural control of speech*. Mit Press.

Hagoort P, Levelt WJM. 2009. The speaking brain. Science. 326:372-373.

Hagoort P. 2005. On Broca, brain, and binding: a new framework. Trends Cogn. Sci. 9:416-423.

Hämäläinen MS, Ilmoniemi RJ. 1994. Interpreting magnetic fields of the brain: minimum norm estimates. Med. Biol. Eng. Comput. 32:35-42.

Hari R. et al. 1993. Functional organization of the human first and second somatosensory cortices: a neuromagnetic study. Eur. J. Neurosci. 5:724-734.

Hari R, Salmelin R. 2012. Magnetoencephalography: from SQUIDs to neuroscience: Neuroimage 20th anniversary special edition. Neuroimage 61:386-396.

Hauk O, Davis MH., Pulvermüller F. 2008. Modulation of brain activity by multiple lexical and word form variables in visual word recognition: a parametric fMRI study. Neuroimage. 42:1185-1195.

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

Hickok G. 2014. The architecture of speech production and the role of the phoneme in speech processing. Lang. Cogn. Neurosci. 29:2-20.

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

Holcomb PJ, Grainger J. 2006. On the time course of visual word recognition: an event-related potential investigation using masked repetition priming. J. Cogn. Neurosci. 18:1631-1643.

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

Indefrey P. 2006. A Meta   analysis of Hemodynamic Studies on First and Second Language Processing: Which Suggested Differences Can We Trust and What Do They Mean?. Lang. Learn. 56: 279-304.

Indefrey P. 2011. The spatial and temporal signatures of word production components: a critical update. Front. Psychol. 2:255:1-16.

Kiefer M, Pulvermüller F. 2012. Conceptual representations in mind and brain: theoretical developments, current evidence and future directions. Cortex. 48: 805-825.

Koester D, Schiller N. 2008. Morphological priming in overt language production: Electrophysiological evidence from Dutch. NeuroImage. 42:1622-1630.

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.

Laganaro M, Valente A, Perret C. 2012. Time course of word production in fast and slow speakers: A high density ERP topographic study. NeuroImage. 59: 3881-3888.

Leonard MK, Bouchard KE, Tang C, Chang EF. 2015. Dynamic encoding of speech sequence probability in human temporal cortex. J. Neurosci. 35:7203-7214.

Levelt WJM, Praamstra P, Meyer AS, Helenius P, Salmelin R. 1998. A MEG study of picture naming. J. Cogn. Neurosci. 10:553-567.

Levelt WJM, Roelofs A, Meyer AS. 1999. A theory of lexical access in speech production. Behav. Brain Sci. 22:1-38.

Llorens A, Trébuchon A, Liégeois-Chauvel C, Alario, FX. 2011. Intra-cranial recordings of brain activity during language production. Front. Psychol. 2: 375.

Lotze M, Seggewies G, Erb M, Grodd W, Birbaumer N. 2000. The representation of articulation in the primary sensorimotor cortex. Neuroreport 11:2985-2989.

MacGregor LJ, Pulvermüller F, van Casteren M, Shtyrov Y. 2012. Ultra-rapid access to words in the brain. Nat. Commun. 3:7-11.

THE CORTICAL DYNAMICS OF SPEAKING

Maess B, Friederici AD, Damian M, Meyer AS, Levelt WJM. 2002. Semantic category interference in overt picture naming: an MEG study. J. Cogn. Neurosci. 14:455-462.

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.

Mesgarani N, Cheung C, Johnson K, Chang EF. 2014. Phonetic feature encoding in human superior temporal gyrus. Science 343:1006-1010.

Meister IG, Wilson SM, Deblieck C, Wu AD, Iacoboni M. (2007). The essential role of premotor cortex in speech perception. Curr. Biol. 17:1692–1696.

Miozzo M, Pulvermüller F, Hauk O. 2014. Early parallel activation of semantics and phonology in picture naming: Evidence from a multiple linear regression MEG study. Cereb. Cortex bhu137.

Möttönen R, Dutton R, Watkins KE. 2013. Auditory-motor processing of speech sounds. Cereb. Cortex 23:1190–1197.

Munding D, Dubarry AS, Alario FX. 2015. On the cortical dynamics of word production: a review of the MEG evidence. Lang. Cogn. Neurosci. 1-22.

Murakami T, Kell CA, Restle J, Ugawa Y, Ziemann U. 2015. Left dorsal speech stream components and their contribution to phonological processing. J. Neurosci. 35:1411-1422.

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.

Nakamura A et al. 1998. Somatosensory homunculus as drawn by MEG. Neuroimage 7:377-386.

Obleser J, Elbert T, Lahiri A, Eulitz C. 2003. Cortical representation of vowels reflects acoustic dissimilarity determined by formant frequencies. Cogn. Brain Res. 15: 207-213.

Oldfield RC, Wingfield A. 1965. Response latencies in naming objects. Q. J. Exp. Psychol. 17:273-281.

Patterson R, Uppenkamp S, Johnsrude I, Griffiths T. 2002. The processing of temporal pitch and melody information in auditory cortex. Neuron. 36:767-776.

Petrides M, Pandya DN. 2009. Distinct parietal and temporal pathways to the homologues of Broca's area in the monkey. PLoS Biol. 7:e1000170.

THE CORTICAL DYNAMICS OF SPEAKING

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

Picton, TW. et al. 2000. Guidelines for using human event‐related potentials to study cognition: Recording standards and publication criteria. Psychophysiology 37:127-152.

Pulvermüller F. 1999. Words in the brain's language. Behav. Brain Sci. 22:253–336.

Pulvermüller F. 2002. The neuroscience of language: on brain circuits of words and serial order. Cambridge University Press.

Pulvermüller F. 2005. Brain mechanisms linking language and action. Nat. Rev. Neurosci. 6:576-582.

Pulvermüller F. 2017. Neural Reuse Of Action Perception Circuits For Language, Concepts And Communication. Progr. Neurobiol. In press.

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

Pulvermüller F, Garagnani M, Wennekers T. 2014. Thinking in circuits: toward neurobiological explanation in cognitive neuroscience. Biol. Cybern. 108:573-593.

Pulvermüller F., et al. 2006. Motor cortex maps articulatory features of speech sounds. Proc. Natl. Acad. Sci. USA. 103:7865-7870.

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.

Rapp B, Goldrick M. 2000. Discreteness and interactivity in spoken word production. Psychol. Rev. 107: 460-499.

Riès SK, Dhillon RK, Clarke A, King-Stephens D, Laxer KD, Weber PB, ... & Lin JJ. 2017. Spatiotemporal dynamics of word retrieval in speech production revealed by cortical high-frequency band activity. Proc. Nat. Ac. Sci. 114:E4530-E4538.

Rilling JK. 2014. Comparative primate neurobiology and the evolution of brain language systems. Curr. Opin. Neurobiol. 28:10-14.

Rilling JK et al. 2008. The evolution of the arcuate fasciculus revealed with comparative DTI. Nat. Neurosci. 11:426-428.

Sahin N, Pinker S, Cash S, 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.

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

Schomers MR, Kirilina E, Weigand A, Bajbouj M, Pulvermüller F. 2015. Causal Influence of Articulatory Motor Cortex on Comprehending Single Spoken Words: TMS Evidence. Cereb. Cortex 25: 3894-3902.

Schomers MR, Pulvermüller F. 2016. Is the sensorimotor cortex relevant for speech perception and understanding? An integrative review. Front. Hum. Neurosci. 10.

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.

Shtyrov Y, Kujala T, Palva S, Ilmoniemi RJ, Näätänen R. 2000. Discrimination of speech and of complex nonspeech sounds of different temporal structure in the left and right cerebral hemispheres. Neuroimage. 12: 657-663.

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. USA. 111:E1918-E1923.

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. Ann. Rev. Neurosci. 18: 555-586.

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.

Snodgrass JG, Vanderwart M. 1980. A standardized set of 260 pictures: norm for name agreement, image

agreement, familiarity, and visual complexity. J. Exp. Psychol. Hum. Learn. Mem. 6:174-215.

Strijkers K. 2015. Can hierarchical models display parallel cortical dynamics? A non-hierarchical

alternative of brain language theory. Lang. Cogn. Neurosci. 1-5.

Strijkers K. 2016. A Neural Assembly–Based View on Word Production: The Bilingual Test Case. Lang.

Learn. 66:92-131.

Strijkers K, Baus C, Runnqvist E, FitzPatrick I, Costa A. 2013. The temporal dynamics of first versus

second language production. Brain Lang. 127:6-11.

Strijkers K, Costa A. 2011. Riding the lexical speedway: A critical review on the time course of lexical

access in speech production. Front. Psychol. 2:356:1-16.

Strijkers K, Costa A. 2016. The cortical dynamics of speaking: Present shortcomings and future avenues.

Lang. Cogn. Neurosci. 1-20.

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 P, Costa A. 2011. Conscious intention to speak facilitates lexical access during

overt object naming. J. Mem. Lang. 65:345–362.

Taulu S, Kajola M. 2005. Presentation of electromagnetic multichannel data: The signal space separation

method. J. App. Physics. 97:124905-124905.

Thierry G, Martin CD, Downing P, Pegna AJ. 2007. Controlling for interstimulus perceptual variance

abolishes N170 face selectivity. Nat. Neurosci. 10:505-511.

Ueno T., Saito S., Rogers TT., Ralph MAL. 2011. Lichtheim 2: synthesizing aphasia and the neural basis

of language in a neurocomputational model of the dual dorsal-ventral language pathways.

Neuron. 72: 385-396.

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.

THE CORTICAL DYNAMICS OF SPEAKING

Van Turennout M, Hagoort P, Brown CM. 1998. Brain activity during speaking: From syntax to

phonology in 40 milliseconds. Science. 280:572–574.

Wilson SM, Saygin AP, Sereno MI, Iacoboni M. 2004. Listening to speech activates motor areas involved

in speech production. Nat. Neurosci. 7:701-702.

Wilson SM, Isenberg AL, Hickok G. 2009. Neural correlates of word production stages delineated by

parametric modulation of psycholinguistic variables. Hum. Brain Mapp. 30:3596-3608.

Figure Captions

<u>Figure 1.</u> Examples of object stimuli used in the experiment divided by condition. Lip and tongue

items were matched pair-wise and equally divided into low and high frequency items.

<u>Figure 2.</u> Global MEG sensor and source activations. (**a**) Global event-related magnetic field gradients

observed in response to all conditions, averaged over all gradiometer pairs and across all participants

(n=15). For the three peaks of activation (120 ms; 180 ms; 280 ms) underlying cortical activity is

represented as mean source strength onto an inflated morphed cortical reconstruction of the left

hemisphere averaged across participants. (**b**) ROIs included in the statistical analyses. Blue patches

correspond to the ROIs included in the analyses of Word Frequency effect and green patches correspond

to the ROIs included in the analyses of Phoneme Somatotopy contrast. The ROIs are represented onto an

inflated morphed cortical reconstruction (MRI-constrained) of the left hemisphere averaged across all

participants (taken at 180 ms for illustration purposes). The ROIs for Word frequency are: LIFG = Region

THE CORTICAL DYNAMICS OF SPEAKING

surrounding left inferior frontal gyrus; MTG-TP = Region surrounding the mid temporal gyrus and temporal pole; IT = Region surrounding the inferior temporal cortex; mSTG = Region surrounding the middle portion of the superior temporal gyrus; and pSTG = Region surrounding the posterior portion of the superior temporal gyrus. The ROIs for Phoneme Somatotopy are: MLip = Region surrounding motor cortex for lip movement; MTongue = Region surrounding motor cortex for tongue movement; mSTG; and pSTG.

Figure 3. **Naming latencies for the word frequency effect and the word-initial phoneme effect.** On the X-axis the naming latencies (in ms) are displayed. On the Y-axis the word-initial phoneme contrast is displayed with Lip indicating those object names initiating with labial and labiodental sounds and Tongue indicating those object names initiating with alveolar and velar sounds. The black bars refer to the low frequency (Low-F) words and the white bars refer to the high frequency (High-F) words. Error bars denote std. error.

Figure 4. **MEG sensor-level and source-level effects for the word frequency contrast**. **(a)** Global event-related magnetic field gradients observed in response to low (blue; transparent light blue denotes std. error) and high frequency (red, transparent light red denotes std. error) object names, averaged over all gradiometer pairs and across all participants (n=15). **(b)** Differences in minimum norm source estimation of the brain responses elicited by low and high frequency object names averaged across all participants (n=15) are plotted onto an inflated morphed reconstruction of the left hemisphere. The source strength differences are represented for the two time-windows analyzed. ROIs showing significantly greater activation in response to low (blue) versus high frequency (red/yellow) object names as revealed by the repeated measures ANOVA are plotted as white circles onto the inflated brain reconstruction. Mean area activations of these ROIs are represented in bar graphs below the inflated brain reconstructions. A white asterisk indicates a significant difference in a particular ROI as evidenced by independent Student t-tests (error bars denote std. error).

Figure 5. **MEG sensor-level and source-level effects for the word-initial phoneme contrast.** **(a)** Global event-related magnetic field gradients observed in response to object names initiating with labial

(green; transparent light green denotes std. error) and coronal (yellow; transparent light yellow denotes std. error) phonemes, averaged over all gradiometer pairs and across all participants (n=15). **(b)** Differences in minimum norm source estimation of the brain responses elicited by object names initiating with labial and coronal phonemes averaged across all participants (n=15) are plotted onto an inflated morphed reconstruction of the left hemisphere. The source strength differences are represented for the two time-windows analyzed (and split up between peak and valley for those ROIs showing an interaction). ROIs showing significantly greater activation in response to object names with labial (blue) versus coronal (red/yellow) phonemes are highlighted in white circles. Mean area activations of these ROIs are plotted in bar graphs where significant interactions of the articulator type with a particular ROI, as evidenced by the repeated measures ANOVA and Student t-tests, are indicated with a white asterisk above the specific type of interaction found (error bars denote std. error).

Lexical frequency

phoneme articulator

LIP
(/p/, /b/, /f/, /v/, /m/)

TONGUE
(/t/, /d/, /s/, /l/, /k/)

HIGH   LOW

[Monkey]   [Bee]

[Sea]   [Donkey]

a

b