NeuroImage 163 (2017) 206–219

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

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

Kristof Strijkers[a,*], Albert Costa[b,c], Friedemann Pulvermüller[d,e]

[a] Aix-Marseille Université (AMU), CNRS, Laboratoire Parole et Langage (LPL), UMR 7309, 13100, Aix-en-Provence, France
[b] Center for Brain and Cognition, CBC, Universitat Pompeu Fabra, Barcelona, Spain
[c] Institució Catalana de Recerca i Estudis Avançats, ICREA, Barcelona, Spain
[d] Brain Language Laboratory, Department of Philosophy and Humanities, WE4, Freie Universität Berlin, 14195, Berlin, Germany
[e] Berlin School of Mind and Brain, Humboldt Universität zu Berlin, 10099, Berlin, Germany

ARTICLE INFO

Keywords:
Language
Speech production
Spatiotemporal brain activity
MEG
Sensorimotor circuits

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.

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 and 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 and Fadiga, 2010; Kiefer and 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 and Levelt, 2009; Indefrey, 2011; Indefrey and Levelt, 2004; Levelt et al., 1998;

* Corresponding author. Centre National de la Recherche Scientifique (CNRS), Laboratoire Parole et Langage (LPL), Université d'Aix-Marseille, 3, Place Victor Hugo, 13331, Marseille, France.
E-mail address: Kristof.Strijkers@gmail.com (K. Strijkers).

https://doi.org/10.1016/j.neuroimage.2017.09.041
Received 15 May 2017; Received in revised form 20 August 2017; Accepted 20 September 2017
Available online 22 September 2017
1053-8119/

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 and 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 and 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 and 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 et al., 2000; Dell, 1986; Dell et al., 2013; Goldrick et al., 2011; Rapp and 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 and 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., 2014; 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 <200 ms 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. (2014) 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 and Costa, 2016). Similarly, in Miozzo et al. (2014), 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. (2014) 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. (2014), 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