LANGUAGE, COGNITION AND NEUROSCIENCE, 2017
VOL. 32, NO. 8, 1064–1067
https://doi.org/10.1080/23273798.2016.1274415

COMMENTARY

OPEN ACCESS

Importance of response time in assessing the cerebral dynamics of spoken word production: comment on Munding et al. (2016)

Ardi Roelofs and Natalia Shitova

Centre for Cognition, Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands

ARTICLE HISTORY Received 12 October 2016; Accepted 5 December 2016

Introduction

The cerebral dynamics of spoken word production is a lively debated issue. Recently, Munding, Dubarry, and Alario (2016a) wrote an opinion article on this issue, which was published together with several commentaries by colleague experts, and a response to the commentaries by Munding, Dubarry, and Alario (2016b). Here, we address a critical issue that has been only briefly mentioned in some of the commentaries (i.e. Laganaro, 2016; Piai, 2016), namely the importance of taking response time (RT) into account when assessing the cerebral dynamics of word production.

Munding et al. (2016a) performed a meta-analysis of magnetoencephalography (MEG) studies to evaluate serial, cascaded, and parallel models of spoken word production. In serial models, a particular stage of word planning begins only if the previous stage has been completed; in cascaded models, stages are sequential, but a particular stage may begin before the previous stage is done; and in parallel models, stages are not sequential but concurrent. Based on their analysis, Munding et al. state that “the ensemble of activity is not indicative of serial processing stages but rather suggestive of concurrent or a strong temporally overlapping cascade of processes” (p. 449). Moreover, the meta-analysis was used to evaluate stage duration estimates by Indefrey and Levelt (2004; Indefrey, 2011). According to Munding et al., articulatory activity occurred

> before 400 ms, where Indefrey (2011) proposes an onset of 600 ms post-stimulus. Likewise, phonological code retrieval is frequently reported 75 ms before predicted, and onsets for lemma selection are spread to both sides of the predicted onset of 200 ms. (p. 455)

In estimating the stage onsets for picture naming, Indefrey and Levelt (2004) assumed a mean naming latency of 600 ms, but latencies may, of course, be different. In particular, mean latencies varied between about 500 and 1000 ms across studies in the meta-analysis of Munding et al. (2016a). Clearly, if the mean RT is 1000 ms, speakers do not complete all planning stages within 600 ms and then wait for 400 ms to initiate articulation. Thus, some kind of rescaling is needed for the stage durations if the mean RT is shorter or longer than 600 ms. Consequently, depending on the RT in a study, stage onsets may occur earlier or later than estimated by Indefrey and Levelt. We demonstrate that if RTs, and consequently stage durations, differ among studies, collapsing time windows of MEG activity across studies yields overlap of stage activity even under serial models.

Rescaling of stage durations

Assume that word planning stages are serial. The estimates of the stage durations by Indefrey and Levelt (2004; Indefrey, 2011) held for a mean naming RT of 600 ms. Mean RTs ranged from about 500 to 1000 ms across studies in the meta-analysis of Munding et al. (2016a). Thus, rescaling is needed for the stage durations if the mean RT is shorter or longer than 600 ms, as Laganaro (2016) and Piai (2016) pointed out in their commentary to Munding et al. In their response, Munding et al. (2016b) indicated to be sceptic about rescaling because “the scaling function is unknown” and “fractionating speech response times into meaningful durations has proven difficult even based on recordable responses measured on single trials” (p. 480). However, although it seems difficult to rescale the empirically observed MEG responses, rescaling may be applied in deriving predictions from models.

In discussing the issue of rescaling, Indefrey (2011) argued that there are two straightforward options, namely proportional or informed rescaling. In

CONTACT Ardi Roelofs a.roelofs@donders.ru.nl

proportional rescaling (e.g. Schuhmann, Schiller, Goebel, & Sack, 2009), each duration estimate of Indefrey and Levelt (2004) is increased or decreased to the extent that the observed mean RT was longer or shorter than 600 ms. In informed rescaling (cf. Laganaro, Valente, & Perret, 2012), only some duration estimates are rescaled based on knowledge about the factors that caused the mean RT to be longer or shorter than 600 ms. Based on event-related brain potential evidence, Laganaro et al. argued that individual differences in naming RT are for the most part due to differences in lemma retrieval duration (i.e. lexical selection). These individual differences concerned the RTs for a single set of pictures (ranging from 714 ms for fast speakers to 926 ms for slow speakers), whereas participants as well as materials differed among the MEG studies in the meta-analysis of Munding et al. (2016a) with RTs ranging between about 500 and 1000 ms. Thus, it seems unlikely that the difference in RTs among MEG studies is mostly due to individual differences in lemma retrieval duration. Still, it would be important to examine what the predictions of serial models are when the duration of only a single stage (i.e. lemma retrieval) is rescaled.

In examining the consequences of rescaling, we restricted ourselves to the picture naming studies with RTs in the meta-analysis of Munding et al. (2016a), while excluding studies with delayed naming and studies of word reading and phrase production. As pointed out by Laganaro (2016), the estimates of Indefrey (2011) for picture naming cannot straightforwardly be applied to these other tasks. We ended up with seven MEG studies of picture naming with mean RTs ranging from 537 to 998 ms. For each of these studies, we rescaled the stage durations based on the mean RT using both proportional and informed procedures. We assumed a model with strictly discrete stages, following Indefrey (2011): perception and conceptualisation, lemma retrieval, phonological code retrieval, syllabification, and phonetic encoding and articulation. For a mean RT of 600 ms, the estimates of the durations of these stages are 200 ms (perception and conceptualisation), 75 ms (lemma retrieval), 80 ms (phonological code retrieval), 100 ms (syllabification), and 145 ms (phonetic encoding and articulation). The following example numerically illustrates the proportional and informed rescaling procedures. For an empirically observed mean RT of 879 ms and proportional rescaling, the duration of the perception and conceptualisation stage would become $(879/600) \times 200$ ms = 293 ms, the duration of lemma retrieval would become $(879/600) \times 75$ ms = 110 ms, and so forth for the other stages. With informed rescaling, the difference between the observed mean RT and 600 ms, here $879 - 600 = 279$ ms, would be added to the lemma retrieval duration of 75 ms, thus this duration would become $75 + 279 = 354$ ms. Figure 1 displays the predicted stage activities when the stage durations are rescaled for each study based on the empirically observed mean RT.

The figure shows that both under proportional rescaling (panel A) and informed rescaling (panel B), there is overlap of different stages across the seven studies even though the succession of stages is strictly serial within each study. The bottom of the figure shows the corresponding earliest onset and latest offset (i.e. range) of the stages across studies, together with the median offsets. The saturation of the colours indicates the degree of stage overlap across studies, with higher saturation denoting greater overlap (cf. Munding et al., 2016a). Clearly, there is considerable overlap among different stages. Also, there is a successive progression of onsets and offsets. This fits well with the observation of Munding et al. (2016a) that “there is a cascading sequence of overlapping activity moving through the brain” (p. 454). Our analysis shows that serial models are compatible with the MEG data. In the remainder, we make clear that prominent computationally implemented models of word production agree with the outcomes of the meta-analysis.

Relating the MEG findings to computational models

Although Munding et al. (2016a) evaluated serial, cascaded, and parallel models, widely discussed implemented models of spoken word production in the literature, such as the model of Dell (1986), Dell, Schwartz, Martin, Saffran, and Gagnon (1997), Dell, Schwartz, Nozari, Faseyitan, and Coslett (2013), and WEAVER++ (Levelt, Roelofs, & Meyer, 1999; Roelofs, 2008, 2014), have both serial and cascaded aspects. These models make a distinction between spreading activation in a lexical network and selection of nodes from the network. Whereas activation cascades through the network, nodes at different processing levels are selected sequentially.

For example, Dell (1986) and Dell et al. (1997, 2013) assume a network of conceptual feature, word, and phoneme nodes. Word planning starts by boosting the activation of the relevant conceptual features, activation cascades from conceptual features via word nodes to phoneme nodes, and selection of the highest activated word node occurs after eight time steps. Then, the activation of the selected word is boosted, and the highest activated onset, nucleus, and coda phoneme nodes are selected after eight time steps. Thus, although activation cascades through the network, word and phoneme selection processes are serial in that the phoneme

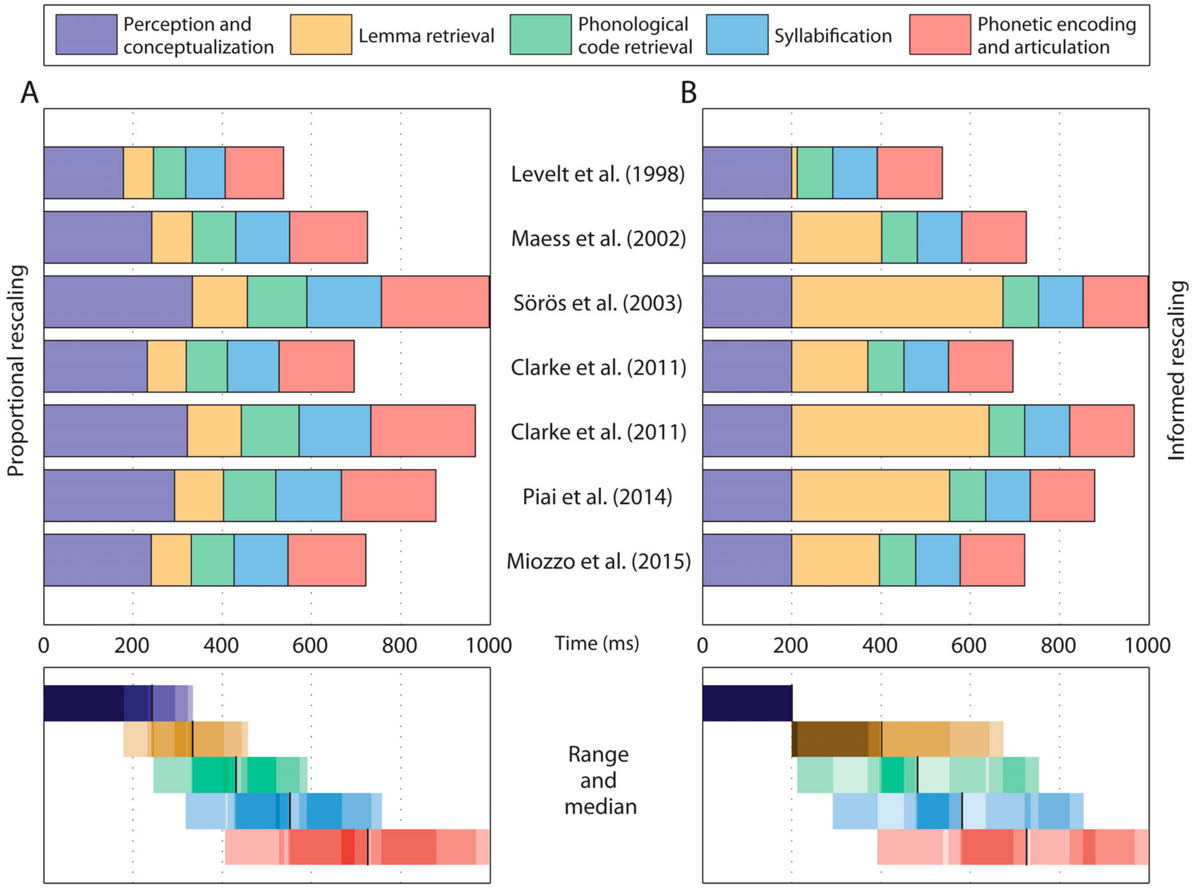

Figure 1. Activity of stages underlying picture naming across studies under a serial model. The stage durations are shown for proportional rescaling (A) and informed rescaling (B) based on the mean picture naming latency. For each type of rescaling, the range of stage durations across studies and the median duration of each stage (vertical line) are shown. The saturation of the colours indicates the degree of stage overlap across studies, with higher saturation indicating more overlap.

selection process starts only after a word node has been selected. The word selection process lasts eight time steps, followed by the phoneme selection process lasting another eight time steps. During the word selection process (i.e. the first eight time steps), phonemes are activated in the network. However, word and phoneme selection processes do not overlap in time (i.e. they occur in the first and second eight time steps). The WEAVER++ model assumes concept, lemma, morpheme, phoneme, and syllable program nodes. Activation spreads from concept to syllable program nodes in a cascading fashion (Roelofs, 2008), but nodes at different processing levels are selected sequentially, as in the model of Dell and colleagues. Thus, extant computational models of word production have both cascaded and serial aspects, in line with the observation of a "cascading sequence of overlapping activity moving through the brain" (p. 454) by Munding et al. (2016a).

However, one of the studies in the meta-analysis by Munding et al. (2016a) reported deviant findings. In the study of Miozzo, Pulvermüller, and Hauk (2015), conceptual and phonological manipulations influenced the MEG response at peak latencies of 150 and 141 ms, respectively. The mean picture naming RT was 722 ms. It is unlikely that speakers completed phonological code retrieval within 141 ms, and then took the remaining 581 ms to complete syllabification, phonetic encoding, and articulation (which should take some 150–200 ms, according to Indefrey, 2011; Indefrey & Levelt, 2004). In terms of the computational models, perhaps the early MEG response reflected phonological code *activation* rather than phonological code *selection*. Alternatively, perhaps the phonological manipulation was confounded with an uncontrolled but correlated conceptual factor, such as conceptual familiarity (e.g. Shao, Roelofs, & Meyer, 2014). Based on their findings, Miozzo et al. argued for a parallel model (see also Strijkers & Costa, 2016a, 2016b), but see Indefrey (2016) for counterarguments.

Conclusion

A meta-analysis of MEG studies by Munding et al. (2016a) provided evidence for sequential but overlapping stage activity. We showed that if RTs differ among studies, collapsing time windows of MEG activity across studies yields such overlap of stage activity even under serial models. Moreover, sequential but overlapping stage activity is also expected under prominent computationally implemented models of spoken word production, which assume cascading of activation and serial selection. We conclude that serial and serial/cascaded models are compatible with the outcomes of the MEG meta-analysis.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Clarke, A., Taylor, K. I., & Tyler, L. K. (2011). The evolution of meaning: Spatio-temporal dynamics of visual object recognition. *Journal of Cognitive Neuroscience*, *23*, 1887–1899. doi:10.1162/jocn.2010.21544

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*, 283–321. doi:10.1037/0033-295X.93.3.283

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review*, *104*, 801–838. doi:10.1037/0033-295X.104.4.801

Dell, G. S., Schwartz, M. F., Nozari, N., Faseyitan, O., & Coslett, H. B. (2013). Voxel-based lesion-parameter mapping: Identifying the neural correlates of a computational model of word production. *Cognition*, *128*, 380–396. doi:10.1016/j.cognition.2013.05.007

Indefrey, P. (2011). The spatial and temporal signatures of word production components: A critical update. *Frontiers in Psychology*, *2*. doi:10.3389/fpsyg.2011.00255

Indefrey, P. (2016). On putative shortcomings and dangerous avenues: Response to Strijkers & Costa. *Language, Cognition and Neuroscience*, *31*, 517–520. doi:10.1080/23273798.2015.1128554

Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition*, *92*, 101–144. doi:10.1016/j.cognition.2002.06.001

Laganaro, M. (2016). Dynamics of word production and processing speed. *Language, Cognition and Neuroscience*, *31*, 463–464. doi:10.1080/23273798.2015.1096402

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. doi:10.1016/j.neuroimage.2011.10.082

Levelt, W. J. M., Praamstra, P., Meyer, A. S., Helenius, P., & Salmelin, R. (1998). An MEG study of picture naming. *Journal of Cognitive Neuroscience*, *10*, 553–567. doi:10.1162/089892998562960

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1–38. doi:10.1017/s0140525×99001776

Maess, B., Friederici, A. D., Damian, M., Meyer, A. S., & Levelt, W. J. M. (2002). Semantic category interference in overt picture naming: Sharpening current density localization by PCA. *Journal of Cognitive Neuroscience*, *14*, 455–462. doi:10.1162/089892902317361967

Miozzo, M., Pulvermüller, F., & Hauk, O. (2015). Early parallel activation of semantics and phonology in picture naming: Evidence from a multiple linear regression MEG study. *Cerebral Cortex*, *25*, 3343–3355. doi:10.1093/cercor/bhu137

Munding, D., Dubarry, A.-S., & Alario, F.-X. (2016a). On the cortical dynamics of word production: A review of the MEG evidence. *Language, Cognition and Neuroscience*, *31*, 441–462. doi:10.1080/23273798.2015.1071857

Munding, D., Dubarry, A.-S., & Alario, F.-X. (2016b). MEG studies of word production: What next? *Language, Cognition and Neuroscience*, *31*, 480–483. doi:10.1080/23273798.2016.1153117

Piai, V. (2016). The role of electrophysiology in informing theories of word production: A critical standpoint. *Language, Cognition and Neuroscience*, *31*, 471–473. doi:10.1080/23273798.2015.1100749

Piai, V., Roelofs, A., Jensen, O., Schoffelen, J.-M., & Bonnefond, M. (2014). Distinct patterns of brain activity characterise lexical activation and competition in spoken word production. *Plos One*, *9*, e88674. doi:10.1371/journal.pone.0088674

Roelofs, A. (2008). Tracing attention and the activation flow in spoken word planning using eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 353–368. doi:10.1037/0278-7393.34.2.353

Roelofs, A. (2014). A dorsal-pathway account of aphasic language production: The WEAVER++/ARC model. *Cortex*, *59*, 33–48. doi:10.1016/j.cortex.2014.07.001

Schuhmann, T., Schiller, N. O., Goebel, R., & Sack, A. T. (2009). The temporal characteristics of functional activation in Broca's area during overt picture naming. *Cortex*, *45*, 1111–1116. doi:10.1016/j.cortex.2008.10.013

Shao, Z., Roelofs, A., & Meyer, A. S. (2014). Predicting naming latencies for action pictures: Dutch norms. *Behavior Research Methods*, *46*, 274–283. doi:10.3758/s13428-013-0358-6

Sörös, P., Cornelissen, K., Laine, M., & Salmelin, R. (2003). Naming actions and objects: Cortical dynamics in healthy adults and in an anomic patient with a dissociation in action/object naming. *NeuroImage*, *19*, 1787–1801. doi:10.1016/S1053-8119(03)00217-9

Strijkers, K., & Costa, A. (2016a). The cortical dynamics of speaking: Present shortcomings and future avenues. *Language, Cognition and Neuroscience*, *31*, 484–503. doi:10.1080/23273798.2015.1120878

Strijkers, K., & Costa, A. (2016b). On words and brains: Linking psycholinguistics with neural dynamics in speech production. *Language, Cognition and Neuroscience*, *31*, 524–535. doi:10.1080/23273798.2016.1158845