The role of electrophysiology in informing theories of word production: a critical standpoint

Vitória Piai

Department of Psychology and Helen Wills Neuroscience Institute, University of California, Berkeley, USA

Acknowledgments

I am truly indebted to Kristoffer Dahlslätt for the various discussions that gave rise to this commentary and for his patience in commenting on all its previous versions. I also want to thank Chris Holdgraf for feedback on a previous version of this commentary and Roemer van der Meij for helpful input. This work was supported by the Netherlands Organization for Scientific Research under Grant #446-13-009.

Munding, Dubarry, and Alario (this issue) courageously and thoroughly summarise the MEG literature on word production. It is evident that their task was a real undertaking and word production researchers should applaud their efforts. In this commentary, I raise a few issues that are inspired by their report. These comments are not meant as criticism to Munding et al., but mainly reflect what I see as limitations of electrophysiological methods when it comes to making temporally specific claims.

Three language production theoretical models are discussed by Munding et al. The hierarchical state feedback control model (Hickok, 2012) has an elaborated motor control aspect, but little specification regarding earlier stages of word planning. As such, it does not lend itself well as a model against which to evaluate the MEG evidence for serial vs. parallel processing. A second model discussed, Price's model (2012), is solely based on positron emission tomography and functional magnetic resonance imaging (fMRI) data, which have no temporal resolution at the scale of single word production (i.e., often less than 1 second). Moreover, in fMRI the link between the blood-oxygen-level dependent (BOLD) signal and underlying neuronal activity, and hence electrophysiological measures, is still poorly understood (e.g., Ekstrom, 2010; Logothetis, 2008). As such, Price's model may be difficult to assess on the basis of electrophysiological data. By contrast, Indefrey's model comprises early and late stages of word production and is based on electrophysiological evidence along with haemodynamic evidence. Thus, Indefrey’s model possibly has the greatest potential to provide the framework for examining the evidence for either serial or parallel models, so my comments are mostly related to it.

With their review, Munding et al. conclude that no evidence can be found for any of the models under consideration. In my view, it seems important for the field to define *a priori* what kind of electrophysiological evidence is needed to speak to either type of model. For example, if activity in different brain areas overlaps in time, then this would presumably constitute evidence in favour of a parallel model. Conversely, temporally non-overlapping activity would provide evidence for a serial

model. As I will argue below, evidence of this latter type is in practice unlikely to be obtained.

Munding et al. integrated the evidence from all available (and appropriate) MEG studies. However, as acknowledged by the authors themselves, problems emerge from this approach. From my viewpoint, it is problematic to make claims based on high temporal precision when combining information across many studies conducted separately. As I will argue, the issues I present below may have (partly) caused the apparent parallelism and/or the unexpected timing that the authors report.

Firstly, regarding brain activity, the authors note: “the latencies of onset and duration do not match Indefrey's estimates” (p. xxx). Possibly the most important issue in this regard is the fact that speech latencies vary across the studies assessed. Though it is true that word production response times (RTs) are between 600-700 ms on average, RTs in the studies examined were much longer than Indefrey's estimate of 600 ms, on which his model is based. For example, the participants of Piai et al. (2014) had an average RT of 880 ms; the participants of Pylkkänen, Bemis, and Elorietta (2014) had an average RT of 905 ms; the participants of Sörös, Cornelissen, Laine and Salmelin (2003) had an average RT of 1000 ms. On the other end, Klein et al. (2014) used a delayed word reading for the MEG experiment (with no RTs reported), whereas RTs for their behavioural experiment using immediate responses were around 455 ms.

Clearly, it cannot be expected that word production processes will have a fixed timing regardless of how long it takes participants to speak. The challenge of scaling estimates has been raised by Indefrey himself: “the question arises how to rescale the [Indefrey & Levelt] estimates to shorter or longer naming latencies.” (2011, p. 3). While part of the across-study variability in RTs can be explained by differences between participant samples, a large part of the variability likely comes from the experimental manipulations introduced by the researchers. It is conceivable that some of these manipulations will prolong some, but perhaps not all, stages of word production, leading to variability in behavioural measures (e.g., Piai et al., 2014).

The variability in speech latencies due to experimental design may artificially create the impression that there is temporal overlap between activities across studies or that certain processes start earlier or later than hypothesised by the model. Overall, the issue of variable RT and how the onset and the duration of processes should be scaled to account for this variability remain an important issue to work on if our theories are to benefit from the temporal dimension of electrophysiological methods.

Secondly, treating different electrophysiological measures across studies equally may unintentionally create parallelism when combining studies, as measures may differ in their degree of temporal precision. Since temporal smearing is an inherent property of time-resolved power estimation, activity inferred from time-frequency representations potentially shows more temporal overlap and less precise onsets and offsets that are not necessarily reflective of the underlying signal (or process for that matter). Notably, low-pass filters also temporally smear the signal, especially with large cut-off values.

Another analytical challenge is one inherent to many signals in neuroscience, and affects the overall pattern observed by Munding et al. As the authors state themselves, there was variation between the assessed studies in terms of the statistical methods used. The fact that certain studies did not correct for multiple comparisons likely raises the number of false-positives included in Munding et al.'s selection of activities, possibly muddling the overall pattern. More importantly, not all types of statistical tests are equally suitable for inferring *when* events happen. For example, researchers may interpret a specific time point of an effect while the statistical test they employed only controls the false-alarm rate under the null hypothesis of no effect for *any* of the time points tested (see for discussion Piai, Dahlslätt, & Maris, 2015). Accordingly, such tests do not warrant inferences about specific time windows or time points of an effect.

Finally, I want to focus on an even more important question: under the assumption that word production is serial, could seriality, as reflected by temporally non-overlapping activity, ever be detected by electrophysiological measures? I want to make clear that I have no theoretical or other

reason for favouring either serial or parallel models. My scepticism lies in that high temporal resolution does not necessarily mean high temporal precision. This issue is equally applicable to EEG and MEG, and not only when combining studies but also on the single-study level. The MEG (and EEG) signal can be recorded at high sampling rates (1KHz or more), a raw signal that theoretically has the temporal resolution necessary for resolving, for example, a 75-ms activity difference that would be necessary for telling lemma retrieval from phonological code retrieval (e.g., Indefrey, 2011). But it is very difficult to work with the raw signal in our field of research, so processing steps are taken. Here is where part of the problem begins.

A common preprocessing step is low-pass filtering, often used to improve the signal-to-noise ratio of event-related potentials/fields. But depending on the cut-off value used, the temporal smearing can be such that precision is lost, that is, we can no longer resolve small differences. In the spectral domain, temporal smearing also occurs when estimating time-resolved spectral power, making small differences difficult to resolve as well.

Furthermore, we also collect data from more than one person. Speakers vary in their speed of speaking (e.g., Laganaro et al., 2012) – presumably for reasons found in the brain – which introduces even more jittering to onsets/offsets/durations of processes and adds uncertainty to our estimations. Uncertainty is also introduced because even the same person will vary in behaviour – and thus in brain processes – over trials. All this variability contributes to smeared temporal information, and decreased temporal precision. Thus, even if the signal as recorded has high temporal resolution, the precision required for answering the question of serial vs. parallel processing is not easily accomplished in practice.

In conclusion, I concur with the authors that “it is not yet possible to conclusively discard or accept any one current model on the basis of the currently assessed evidence” (p. xxx), but I am also not yet convinced that we ever will conclusively reject or accept serial vs. parallel models on the basis

of electrophysiological evidence alone. This may be seen as an overly negative view of this line of inquiry, but no one would be happier than me if this view were shown to be unfounded. However, even though Munding et al.'s effort did not result in a conclusive answer about serious vs. parallel models of word production, it does nicely illustrate our current understanding of language production processes as investigated through MEG and, by doing so, inspires us towards exciting avenues for future research.

References

Ekstrom, A. (2010). How and when the fMRI BOLD signal relates to underlying neural activity: The danger in dissociation. *Brain Research Reviews*, *62*, 233–244. doi:10.1016/j.brainresrev.2009.12.004

Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience*, *13*, 135–145. doi:10.1038/nrn3158

Indefrey, P. (2011). The spatial and temporal signatures of word production components: a critical update. *Frontiers in Psychology*, *2*, 255. doi:10.3389/fpsyg.2011.00255

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

Klein, M., Grainger, J., Wheat, K. L., Millman, R. E., Simpson, M. I. G., Hansen, P. C., & Cornelissen, P. L. (2014). Early Activity in Broca's Area During Reading Reflects Fast Access to Articulatory Codes From Print. *Cerebral Cortex*. doi:10.1093/cercor/bht350

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

Logothetis, N. K. (2008). What we can do and what we cannot do with fMRI. *Nature*, *453*, 869–878. doi:10.1038/nature06976

Mazaheri, A., & Jensen, O. (2010). Rhythmic pulsing: linking ongoing brain activity with evoked responses. *Frontiers in Human Neuroscience, 4*, 177. doi:10.3389/fnhum.2010.00177

Piai, V., Dahlslätt, K., & Maris, E. (2015). Statistically comparing EEG/MEG waveforms on the basis of successive univariate test statistics: How bad can it be? *Psychophysiology*, *52*, 440-443. doi:10.1111/psyp.12335

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(2),* e88674. doi: 10.1371/journal.pone.0088674

Price, C. J. (2012). A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *NeuroImage*, *62,* 816–847. doi:10.1016/j.neuroimage.2012.04.062

Pylkkänen, L., Bemis, D. K., & Elorrieta, E. B. (2014). Building phrases in language production: An MEG study of simple composition. *Cognition, 133,* 371-384. doi:10.1016/j.cognition.2014.07.001

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