Validity of chronometric TMS for probing the time-course of word production: a modified replication

Adrian Jodzio[1,2]*, Vitória Piai[2,3], Lennart Verhagen[2], Ian Cameron[2,4] and Peter Indefrey[1,2,5]*

[1]Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

[2] Radboud University, Donders Institute for Brain, Cognition and Behaviour, Nijmegen, The Netherlands

[3]Radboudumc, Donders Institute for Brain, Cognition and Behaviour, Nijmegen, Department of Medical Psychology, Nijmegen, The Netherlands

[4]Faculty of Electrical Engineering, Mathematics and Computer Science (EEMCS), University of Twente, Enschede, The Netherlands

[5]Institut für Sprache und Information at Heinrich Heine University, Düsseldorf, Germany

*Corresponding authors:

adrian.jodzio@donders.ru.nl

Kapittelweg 29, 6525 EN Nijmegen, The Netherlands

+31 (0)24-3610750

indefrey@uni-duesseldorf.de

Universitätsstr. 1, 40225 Düsseldorf, Germany

+49 211 81-15464

Keywords: word production, response-locked analyses, discomfort ratings, peripheral effects, individual variability

Abstract

In the present study, a modified replication of Schuhmann et al. (2012), we used chronometric TMS to probe the time-course of three brain regions during a picture naming task. The left inferior frontal gyrus (IFG), left posterior middle temporal gyrus (pMTG) and left posterior superior temporal gyrus (pSTG) were all separately stimulated in one of five time-windows (225, 300, 375, 450, 525ms) from picture onset. We found posterior temporal areas to be causally involved in picture naming in earlier time-windows, whereas all three regions appear to be involved in the later time-windows. However, chronometric TMS produces non-neuronal effects that may impact behavior, and furthermore, the time-course of any given process is a product of both the involved processing stages along with individual variation in the duration of each stage. We therefore extend previous work in the field by accounting for both individual variations in naming latencies and directly testing for non-neuronal effects of TMS. Our findings reveal that both factors influence behavioral outcomes at the group level. Underlining the importance of accounting for individual variations in naming latencies, especially for late processing stages closer to articulation and recognizing the presence of non-neuronal effects of TMS. The paper advances key considerations and avenues for future work using chronometric TMS to study overt production.

1. Introduction

Typical conversational speech unfolds at a rate of approximately two words per second. It is fast and efficient yet a highly complex and coordinated series of processes is required to produce even a single word. According to a well-supported theory (Levelt et al., 1999), the following processing stages are at the heart of word production: conceptualization, lemma selection, phonological code retrieval, syllabification, phonetic encoding, and articulation. These stages can be operationalized and studied in, for example, the context of picture naming. Here, individuals must retrieve the corresponding concept (conceptualization), select an appropriate lexical label (lemma selection), retrieve the sounds required to form this label (phonological code retrieval), put them together according to phonological rules and intonation patterns (syllabification), program an intricate series of muscles movements (phonetic encoding) and finally articulate a single word (articulation).

Building on the aforementioned theory, Indefrey & Levelt (2004) conducted a comprehensive meta-analysis, henceforth the I&L model, providing both spatial and temporal estimates for each of the proposed processing stages (Indefrey, 2011; Indefrey & Levelt, 2004). Although subsequent proposals for the functional neuroanatomy of language production have been put forth (Hickok, 2012; Price, 2012; Walker & Hickok, 2016), only the I&L model provides temporal estimates for these processes underlying word production. All temporal estimates can be given relative to picture onset and to the naming latency. Assuming a naming latency of 600ms, a common baseline naming latency for picture naming (Indefrey & Levelt, 2004), they are as follows: 1) Conceptualization occurring in widespread brain regions prior to 200ms; 2) Lemma selection in the left mid portion of the middle temporal gyrus (MTG) between 200-275ms; 3) Phonological code retrieval in the left posterior MTG/STG between 275-355ms; 4) Syllabification in the left posterior IFG (pars opercularis) between 355-455ms (duration varies with

number of syllables); 5) Phonetic encoding in bilateral inferior motor cortices between 455- 600ms; 6) Self-monitoring takes place in bilateral STG.

Although much of the literature supports the timing and brain region estimates provided by the I&L model, some contradictory findings have been reported. Strijkers & Costa (2016) have raised issues and pointed out discrepancies that need to be addressed. Among other points, they focus on the findings of a chronometric TMS study conducted by Schuhmann et al. (2012). These researchers stimulated the left mid-portion of MTG (mMTG), left posterior STG (pSTG) and left IFG in five time-windows post-picture onset (150, 225, 300, 400 & 500ms). Compared to a no-stimulation condition, mMTG stimulation was found to significantly increase naming latencies when stimulated at 225-275ms as well as at 400-450ms, pSTG at 400-450ms and IFG at 300-350ms.

Following TMS to mMTG they observed behavioural perturbations, slowing of naming latencies, that match predictions of the I&L model. The earlier perturbation effect (225-275ms) aligns with the lemma selection process thought to take place in mMTG, whereas the later perturbation (400-450ms), also found after pSTG stimulation, aligns a self-monitoring role of these regions at later stages in word production. They also observed effects of TMS interference that on first consideration might seem to contradict the I&L model, such as an earlier effect of IFG stimulation and no evidence of pSTG involvement (Strijkers & Costa, 2016). We note, however, that alternative explanations for these apparent discrepancies are feasible. First and foremost, the earlier IFG involvement might be entirely driven by a shorter overall naming latency of the Schuhmann et al. study (460ms) compared to the original I&L model (assuming a 600ms latency). A concomitant reduction in processing time would explain an earlier recruitment of IFG. Further, the absence of an effect for pSTG is not evidence against the I&L model. Namely, the I&L model assigns both pSTG as well as pMTG to phonological code retrieval. Since the pSTG stimulation site in (Schuhmann et al., 2012) was rather dorsal (Talairach: x = -55; y = -44; z

= 18), it may be the case that TMS interfered with self-monitoring (also assigned to pSTG) but left phonological code retrieval intact.

In the present study, we build on and extend these findings of TMS chronometry of language production, while taking the issue of naming latency variability to heart. As already mentioned for the (Schuhmann et al., 2012) study, the baseline (unperturbed) naming latencies vary greatly across studies (all values are approximated): 450ms (Schuhmann et al., 2009), 460ms (Schuhmann et al., 2012), 525ms (Shinshi et al., 2015), 570ms (Zhang et al., 2018), 600ms (Hämäläinen et al., 2018) and 620ms (Wheat et al., 2013). If naming latencies are off by 50ms between studies, this could mean that the targeted processing stage is also shifted in time, thus the same brain region being stimulated in two different studies may be causally contributing in two distinct time-windows simply as a result of differences in baseline naming latencies. Concurrently, there is also variation among participants within studies. For example, (Shinshi et al., 2015) found that TMS to left IFG compared to sham stimulation significantly increased naming latencies at either 300ms or 375ms for nine participants, at 150ms for one participant, at 225ms for another, and no significant effects were found for a remaining participant. Again, if one participant names pictures 50ms faster than another participant, TMS may perturb the target processing stage in one participant and leave it intact in another. This logic is illustrated in **Fig. 1**.

To foreshadow the methods and results, the current study also observed variability in baseline naming latencies across participants, despite pre-selecting stimuli to evoke naming latencies of around 600ms. In order to address this issue, a baseline-adjusted response-locked analysis was performed. Even if there are differences in baseline naming latencies across participants, it is true that earlier processing stages may still unfold somewhat uniformly across participants. However, variation will inevitably increase as time unfurls, thus affecting later processing stages to a greater extent. By performing a baseline response-locked analysis, we may be better able to elucidate any effects arising in the later time-windows.

Lastly, TMS is also known to elicit non-neuronal effects. When stimulating with TMS, anything under the TMS coil will inevitably be stimulated, this includes any muscles and nerves between the coil and the to-be stimulated cortex. TMS also elicits a “click” noise every time a pulse is triggered. Both the somatosensory and auditory by-products have been shown to evoke a-specific behavioral effects. For example, Holmes & Meteyard (2018) found that subjective discomfort ratings from different stimulation sites were able to predict reaction time differences in previously published studies. Furthermore, studies such as Duecker et al. (2013), have also shown that the timing of stimulation during an ongoing process also matters. For these reasons, discomfort ratings for each stimulation site are acquired in the present study and will be used to test whether subjective discomfort accounts for differences in naming latencies. Moreover, the main effect of time will also be investigated in order to rule out any non-neuronal effects of TMS that may be time-window specific. If there are truly non-neuronal temporal effects of TMS, they should be present across stimulation sites as the effects are not specific to cortical stimulation. If there are site-specific non-neuronal effects of TMS, they should scale with the subjective discomfort.

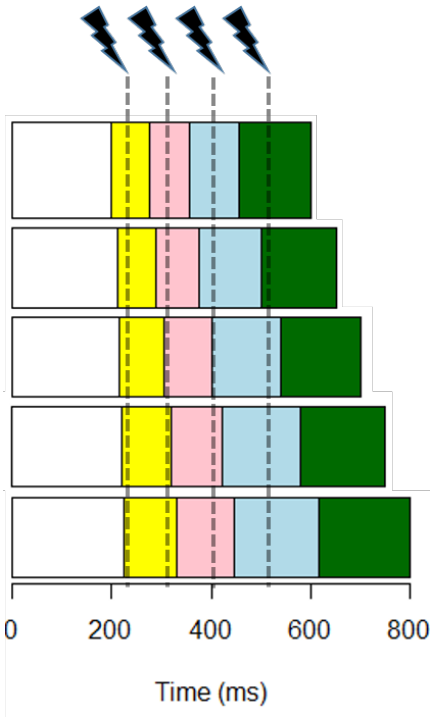

Fig. 1. Schematic of TMS stimulation targeting varying processing stages. The dashed lines indicate the timing of the TMS pulses. Each stage is illustrated by a different color. Note that in the figure the scaling of each stage is proportionate, however this is simply to illustrate the point. It is also conceivable that longer naming latencies may reflect a longer duration in only one processing stage, in which case a different shift of processing stages might occur as a result.

Taken together, the present study aims to conduct a modified replication of the Schuhmann et al. (2012) study, while at the same time exploring alternative factors that may influence the findings such as naming latency variability and non-neuronal effects of TMS. The original study by Schumann and colleagues had a well-designed set-up, which the present study wishes to emulate. At the same time, three modifications are made in order to address the issues raised in the literature (Strijkers & Costa, 2016).

1) The pSTG & IFG stimulation sites from Schuhmann et al. (2012) are also stimulated in the present study. However, instead of mMTG stimulation, pMTG is chosen as the third stimulation site in order to interfere with phonological code retrieval.
2) Monosyllabic and bisyllabic words are used to investigate how word length may modulate the spatiotemporal aspects of word production from the phonological code retrieval stage onwards.
3) In order to investigate the possibility that faster naming latencies may underlie early contributions of IFG, for the monosyllabic session, stimuli were piloted and the 30 stimuli with naming latencies closest to 600ms were used. Note that the monosyllabic session is directly compared to Schuhmann et al. Stimuli for the bisyllabic session were centered around 650ms based on the 50ms per syllable processing time required during syllabification (Indefrey & Levelt, 2004).

Overall, we expect pMTG stimulation to interfere with picture naming in an earlier time-window as compared to both pSTG and IFG. Similarly, we predict IFG stimulation to have the latest time-window effect among the three stimulation sites. Thus, with respect to the results of Schuhmann et al. (2012), we had the following hypotheses, which were tested in the present study:

1. The stimulation effects in IFG will be replicated, but the time window of these stimulation effects should scale with speech onset latencies.

2. The stimulation effects in pSTG at 400-450ms will be replicated, but there should also be stimulation effects in posterior temporal lobe (pSTG, pMTG or both) in the time window of phonological code retrieval (275-355ms post picture onset).

2. Methods

2.1 Participants

Twenty-four healthy, right-handed, native Dutch speakers (15 females, mean age = 22.5 $\pm$ 3.1 years) were recruited for the present study. All participants gave written informed consent and had no TMS contraindications. One participant was replaced due to issues with the audio recording and two participants were excluded from the analysis due to poor performance as reflected by error rates and delayed responses (>1200ms). All participants were compensated for their time. The study adhered to the Declaration of Helsinki and was approved by the competent reviewing authority ('Medical-Ethical Review Committee', MERC), the 'Commissie Mensgebonden Onderzoek regio Arnhem-Nijmegen' (CMO).

2.2. Materials

177 black and white line drawings were pre-tested, corresponding to 100 monosyllabic and 77 bisyllabic picture names. The drawings were taken from the picture database of the Max Planck Institute for Psycholinguistics in Nijmegen, the Netherlands, to remain consistent with previous studies (Schuhmann et al., 2009, 2012). Twenty participants were recruited and named all the pictures six times each as is the case in the TMS experiment (see below). Half the participants started with the monosyllabic picture names whereas the other half began with the bisyllabic picture names. Both the mono- and bisyllabic stimuli were presented in a pseudorandomized order to properly space out repetitions of the same picture.

A total of sixty pictures were selected to use for the TMS experiment. Because the aim of this experiment is to elucidate the effects of TMS on the time-course of word production, the thirty monosyllabic pictures with naming latencies closest to 600ms and the thirty bisyllabic pictures with naming latencies around 650ms were selected.

2.3 Experimental Lists and Design

For each syllabic type, the thirty pictures were grouped into three sets of 10 pictures each. These picture sets were controlled for onset phonemes and consonant clusters as well as the semantic categories they belonged to. Thus there was a similar proportion of animals, foods and objects in each set. The three picture sets were then assigned to the three stimulation sites, counterbalanced across participants using a lattice square design for each of the syllabic sets. That is, participants saw different pictures for each stimulated site and the site and picture set combination was different across participants.

The study entailed a 3x6x2 within-subject design with three stimulation sites (IFG, pMTG, pSTG), five stimulation time-windows along with a no-stimulation condition (no TMS, 225ms, 300ms, 375ms, 450ms, 525ms), and two word lengths (monosyllabic and bisyllabic). The entire experiment was conducted at the Donders Center for Cognitive Neuroimaging (Radboud University) and comprised three sessions on three separate days. During the first session we acquired a T1-weighted structural scan. If participants already had a T1-weighted scan available, they were exempt from the scanning session. The following two TMS sessions were carried out in the same manner except that monosyllabic stimuli were used in one session and bisyllabic stimuli in the other. The session order was counterbalanced across participants.

Participants named pictures while receiving online TMS to one of three stimulation sites. The order of stimulation sites was counterbalanced across participants and session. In each session, there was a total of six blocks and the TMS coil was moved to the subsequent stimulation site after two blocks. Prior to the first block for a new stimulation site, participants were familiarized with the picture set used for that particular stimulation site via pictures and labels on a sheet of paper. We also included ten practice trials in which participants named each picture in the set once before beginning the first block of every new stimulation site. These ten practice trials were not used in the analysis. Moreover, upon completing

stimulation to a particular site, participants were asked to rate the TMS-specific discomfort (annoyance, pain, muscle twitches) they experienced on a scale of 1 to 10 (Meteyard & Holmes, 2018).

For each stimulation site, each of the ten pictures in the set was presented six times, once for each stimulation time-window (5) and the no TMS condition (1). Thus each participant named ten pictures six times per stimulation site for a total of 60 trials. For all three stimulation sites, participants completed 180 trials in total per session. All trials were recorded with a microphone for later offline analysis. Stimulus presentation, audio recording and TMS stimulation were all controlled with Presentation (Neurobehavioral Systems, Inc., Albany, Calif., USA). The trial timing was kept identical to the Schuhmann et al. (2012) study. A fixation cross was presented for a jittered duration between 5900 and 7900ms, a blank screen for 100ms followed by the target picture for 750ms.

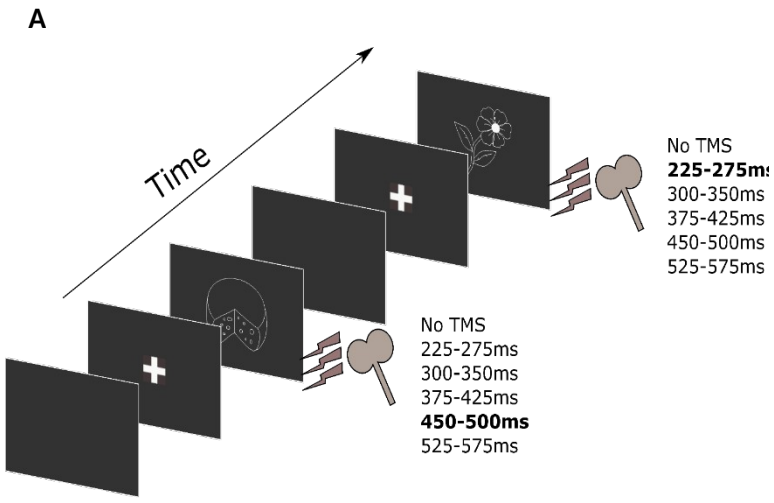

Fig 2. Experimental overview. A. Trial design. Chronometric online TMS was applied in one of five time-windows after picture onset to one of three stimulation sites (blocked design). Bolded times indicate an example time-window for the given trial. B. Stimulation sites (in Talairach coordinates): left IFG (x,y,z = -49.9, 15.1, 20.8), left pSTG (x,y,z = -56.1, -43.4, 14.2) and left pMTG (x,y,z = -56.5, -44.7, 1.7).

2.4 TMS protocol

Frameless infrared-based Neuronavigation (TMS Navigator, Localite, Sankt Augustin, Germany) based on individual anatomical T1-weigthed scans was used to navigate the TMS coil and to maintain its position over the respective target sites throughout the experiment. MR images were acquired using a rapid gradient echo (MPRAGE) sequence (1mm isotropic). The mean Talairach coordinates from Schuhmann et al. were used for localizing the IFG (x,y,z = -49, 13, 26) and pSTG (x,y,z = -55, -44, 18) stimulation sites. On account of whether a more ventral position from the original pSTG site may disrupt phonological code retrieval, the pMTG (x,y,z = -55, -44, 9) stimulation site was localized just inferior to the pSTG stimulation site. Utilizing Localite, Tailarach coordinates were transformed into subject space and adjusted according to each individual's anatomy. If transformed coordinates were positioned in a sulcus then they were adjusted so as to stimulate the apex of the adjacent gyrus. For the temporal-lobe targets, adjustments were only made in the inferior-superior directions (ie. a pMTG transform that was located in the superior temporal sulcus would be shifted inferiorly until it was on the apex of the MTG but no adjustments were made in the anterior-posterior directions). For IFG, the adjustment was made so that it was located superior to the ascending vertical ramus (Schuhmann et al., 2012). Post-adjustment for individual anatomy, the mean Talairach coordinates for our study sample were: IFG (x,y,z = -49.9, 15.1, 20.8), pSTG (x,y,z = -56.1, -43.4, 14.2) and pMTG (x,y,z = -56.5, -44.7, 1.7).

Online triple-pulse TMS (tpTMS) at a frequency of 40Hz and a stimulation intensity of 120% of the resting motor threshold (RMT) was used. Triple-pulse TMS with a frequency of 40 Hz was chosen for replication purpose (Schuhmann et al., 2012), however, it is also the most commonly used in stimulation protocol

chronometric TMS studies investigating picture naming (Hämäläinen et al., 2018; Schuhmann et al., 2009; Shinshi et al., 2015; Zhang et al., 2018). In both sessions, RMT was determined as the minimum intensity evoking a motor potential of at least 0.05mV in the first dorsal interosseous muscle of the right hand in 50% of trials. The mean stimulation intensity used during the experiment was 44.2 ± 7.2% (~67 A/µs) of maximum stimulator output for the first session and 44.7 ± 7.5% (~68 A/µs) for the second session. All TMS pulses were biphasic and applied with a figure eight MagPro MC-B65-HO-8 coil (MagVenture, Farum, Denmark) connected to a Magpro-X-100 magnetic stimulator (MagVenture, Farum, Denmark). The coil was held tangentially to the skull and angled perpendicular to the stimulated gyrus. All participants wore earplugs throughout the duration of the experiment. The current TMS protocol adheres to international safety guidelines (Rossi et al., 2009, 2020).

Participants received triple-pulse TMS in one of five time-windows following picture onset. They were as follows: 1) 225-250-275ms; 2) 300-325-350ms; 3) 375-400-425ms; 4) 450-475-500ms; and 5) 525-550-575ms. Additionally, there was a sixth condition in which no TMS stimulation was applied. The five time-windows consecutively cover time points from 225ms to 575ms post-picture onset. In contrast to previous studies, we chose not to stimulate prior to 225ms since no significant effects were ever reported in this early time-window (Hämäläinen et al., 2018; Schuhmann et al., 2009, 2012; Shinshi et al., 2015; Wheat et al., 2013; Zhang et al., 2018). Furthermore, the phonological processing stages targeted in this study are hypothesized to unfold after 200ms (Indefrey, 2011; Indefrey & Levelt, 2004).

2.5 Analysis

Naming latencies were determined offline using Praat (Boersma & van Heuven, 2001; Boersma & Weenink, 2019) blinded for stimulation site. Trials with an omitted or erroneous response were considered errors and excluded from all naming latency analyses. Trials in which the TMS pulses made it impossible to correctly determine word onset were also excluded from all naming latency analyses as

well as the error analysis (Monosyllabic: 0.3%; Bisyllabic: 0.33%). In order to remain consistent with Schuhmann et al. (2012), all trials with naming latencies greater than 2SD from the participant's mean (per stimulation site and time-window) were removed from naming latency analyses (Monosyllabic: 4.11%; Bisyllabic: 4.21%). In order to replicate the Schuhmann et al. (2012) study as best as possible, analyses were carried out to test for the effects of TMS pulse time within stimulation site for each word type. However, the present study utilizes mixed effects models to analyze the data as opposed to the ANOVAs used in the Schuhmann study (2012). All analyses were conducted in R (version 4.0.4; www.r-project.org). Furthermore, as a result of indirectly comparing across the three stimulation sites, a Bonferroni correction was applied for all analyses to better control for multiple comparisons. The alpha level for significance is therefore <0.016.

2.5.1 Discomfort Ratings

TMS discomfort ratings were collected for each stimulation site (Meteyard & Holmes, 2018). Since each participant provided one discomfort rating per stimulation site, an ANOVA was used to test for any differences in discomfort among the three stimulation sites as well as across sessions. Tukey's test was then used to compare the three sites to one another while controlling for multiple comparisons.

2.5.2 Errors

Errors were analyzed using a mixed-effects logistic regression using the lmerTest package (Kuznetsova et al., 2017). The errors were analyzed for each word type and stimulation site. The model had errors as the dependent variable, TMS pulse time as the fixed factor, with the no-TMS condition as the reference level, and a by-participant random intercept. Items were originally also entered as random intercepts to remain consistent with the model used for the naming latency analysis (see Section 2.5.3) however; the model did not converge for the error analyses. As mentioned above, the alpha level for significance was set to <0.016.

2.5.3 Standard Analysis of naming latencies

Naming latencies were analyzed using linear mixed-effects models with the lmerTest package in R (Kuznetsova et al., 2017). For the within stimulation site analyses, the models had the following parameters: TMS pulse time as a fixed factor, with the no-TMS condition as the reference level, and included by-item and by-participant random intercepts. As mentioned in the previous section, the alpha level for significance was lowered to <0.016.

2.5.4 Baseline response-locked analysis

The following analysis was performed separately for each word type. In order to have a baseline response-locked stimulation time point for each trial (word), each participant's picture-specific baseline (i.e., no-TMS condition) naming latency was subtracted from the TMS stimulation time point (relative to stimulus onset). The participant and picture-specific baseline naming latency (PP-BNL) was used to account for variability across participants and pictures, and reflects the typical naming speed for a given participant and picture should no TMS interference occur. Since the TMS stimulation comprised three pulses, the time of the last pulse was used (e.g. if stimulation on a given trial occurred at 300, 325, and 350ms, then 350ms was used as the TMS stimulation time point). To illustrate, in trial *t* a participant is presented with picture *C*, for which the PP-BNL is 600ms. In trial *t*, the TMS stimulation occurred in the 300-350 ms time-window relative to stimulus-onset, so the corresponding response-locked time point of trial *t* is -250ms (350-600ms), meaning that the last TMS stimulation in trial *t* was delivered 250ms prior to the participant's PP-BNL for picture *C*.

Next, the trial-specific deviations in naming latencies from the above-mentioned PP-BNL were calculated. Continuing from the above example, if this participant had a trial-specific naming latency (*t*-NL) of 650ms for trial *t*, then the deviation in naming latency would be 50ms (650-600ms). Hence, TMS stimulation

250ms prior to the typical naming time (PP-BNL) resulted in a 50ms increase in naming latency for trial *t*. **Fig. 3** illustrates the aforementioned calculations.

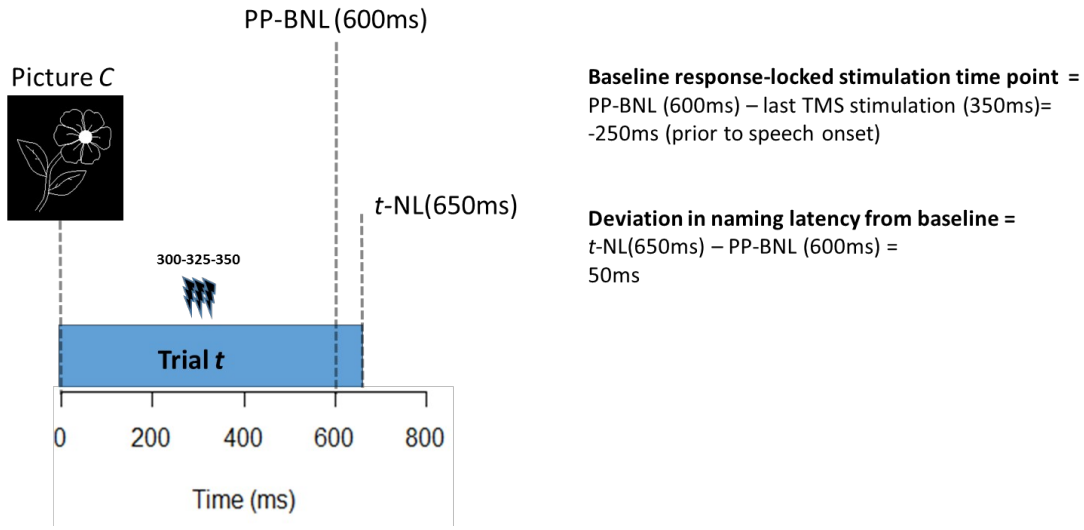

Fig. 3. Visual schematic of baseline response-locked analysis. PP-BNL = Participant and picture-specific baseline naming latency, *t*-NL = trial-specific naming latency.

From the above procedure, the data of each participant results in an irregularly sampled time series for each picture, with baseline response-locked time points on the x-axis and deviation in naming latencies on the y-axis (see Figure 7). To estimate a continuous regular time series for each participant, a moving mean with a Gaussian filter was used to approximate the mean value for each time point. A 100ms time-window was utilized and the Gaussian filter ensured that values in the center of the window had more weight. The window was moved in steps of 25ms. To accept a time-window estimate, we set a minimum of 7 data points contributing to the mean value.

This procedure resulted in a regular time series for each picture for each participant. Note that different groups of pictures were assigned to different stimulation sites. Therefore, the following step necessitated averaging across these groups of pictures in order to obtain time series for each participant

and stimulation site, and subsequently averaging across participants to obtain a site-specific time series. For statistical inference of these timeseries data we used a cluster-based permutation analysis for each site-specific time series whereby the deviations in naming latencies from baseline (in ms) were compared to a theoretical control condition consisting of zeros (i.e., TMS had no effect). The cluster-based permutation analysis was performed using the clusterperm.lmer function in the permutes package (C. C. Voeten, 2019, 2021). The function utilizes an alpha-level significance threshold equivalent to $P < 0.05$.

2.5.5 Non-neuronal TMS effects

Finally, main effects of TMS pulse time, stimulation site and subjective discomfort ratings on naming latencies were analyzed to determine whether non-neuronal effects of TMS might be present. Since the interest is on the non-neuronal effects of TMS, the No TMS condition was not included in this analysis. Naming latencies were analyzed using linear mixed-effects models (lmerTest package, Kuznetsova et al., 2017). Three models were used, all of which included by-item and by-participant random intercepts. First, to test for a main effect of TMS pulse time, the model had naming latencies as the dependent variable and TMS pulse time as the fixed factor, with sequential contrasts. For effects of stimulation site, the model had naming latencies as the dependent variable and stimulation site as the fixed factor, with pMTG acting as the reference level. Lastly, to test for the effect of subjective discomfort, a third model with both stimulation site and subjective discomfort as fixed factors was constructed. Importantly, this third model was compared to the second model using an ANOVA in order to determine whether adding subjective discomfort ratings significantly improves the model fit. Again, the alpha level for significance was lowered to <0.016.

3. Results

3.1 Discomfort Ratings

The ANOVA revealed a main effect for stimulation site ($F(2, 63) = 14.43$, $p < 0.001$). IFG discomfort ratings (mean = 5.06, median = 5, range = 1.66 – 8.33) were found to be significantly higher compared to those of pMTG (mean = 3.65, median = 3.33, range = 1 – 7, $t(21) = 3.517$, $p = 0.002$) and pSTG (mean = 2.95, median = 2.5, range = 1.3 – 6.66, $t(21) = 5.275$, $p < 0.001$). There was no significant difference between pSTG and pMTG discomfort ratings ($t(21) = 1.758$, $p = 0.19$). Moreover, there was also a significant effect in discomfort rating between sessions ($t(21) = -2.69$, $p = 0.008$), where the second session had lower discomfort ratings as compared to the first.

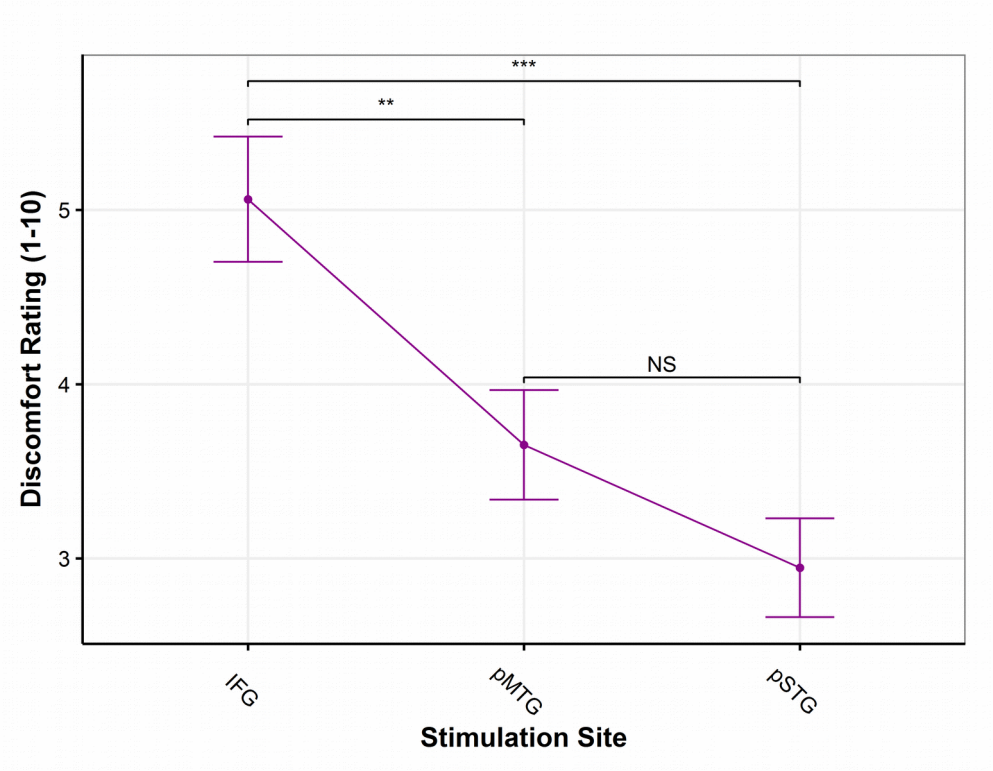

Fig. 4. Mean discomfort ratings collapsed across word type and session. Error bars indicate the standard error of the mean. IFG = inferior frontal gyrus, pMTG = posterior middle temporal gyrus, pSTG = posterior superior temporal gyrus. Discomfort was rated on a scale of 1 to 10. Higher ratings indicate more discomfort.

3.2 Errors

Error rates were comparable across the two sessions (Monosyllabic: 4.11%; Bisyllabic: 4.21%). **Table 1** provides a statistical summary of the error analysis. During monosyllabic naming there were no significant effects of TMS pulse time within any of the stimulation sites. However, during bisyllabic naming, stimulating IFG at 450ms resulted in significantly more errors as compared to the no-TMS condition. No other effect reached significance during bisyllabic naming.

Table 1. Inferential statistics for errors. IFG = inferior frontal gyrus, pMTG = posterior middle temporal gyrus, pSTG = posterior superior temporal gyrus. Significant effects are p < 0.016.

Monosyllabic session

	B	SE	z	p
IFG				
Intercept	-3.598	0.423	-8.51	**<0.001**
No-TMS vs. 225ms	0.583	0.481	1.211	0.229
No-TMS vs. 300ms	0.266	0.508	0.525	0.6
No-TMS vs. 375ms	1.038	0.453	2.291	0.022
No-TMS vs. 450ms	1.041	0.453	2.297	0.022
pMTG				
Intercept	-3.858	0.481	-8.021	**<0.001**
No-TMS vs. 225ms	<0.001	0.641	0.000	1.000
No-TMS vs. 300ms	-0.522	0.738	-0.707	0.479
No-TMS vs. 375ms	0.641	0.568	1.081	0.280
No-TMS vs. 450ms	0.730	0.558	1.307	0.191
pSTG				
Intercept	-3.676	0.432	-8.502	**<0.001**
No-TMS vs. 225ms	-0.188	0.609	-0.309	0.757
No-TMS vs. 300ms	0.533	0.522	1.022	0.307
No-TMS vs. 375ms	0.968	0.490	1.975	0.048
No-TMS vs. 450ms	1.109	0.482	2.299	0.0215

Bisyllabic session

	B	SE	z	p
IFG				
Intercept	-3.698	0.436	-8.471	**<0.001**
No-TMS vs. 225ms	-0.146	0.541	-0.270	0.787
No-TMS vs. 300ms	-0.146	0.541	-0.270	0.787
No-TMS vs. 375ms	0.250	0.499	0.501	0.616
No-TMS vs. 450ms	1.4408	0.433	3.328	**<0.001**
pMTG				
Intercept	-3.230	0.390	-8.431	**<0.001**
No-TMS vs. 225ms	-0.238	0.490	-0.486	0.627
No-TMS vs. 300ms	-0.238	0.490	-0.486	0.627
No-TMS vs. 375ms	-0.540	0.530	-1.018	0.308
No-TMS vs. 450ms	-0.225	0.490	-0.460	0.645
pSTG				
Intercept	-3.428	0.401	-8.542	**<0.001**
No-TMS vs. 225ms	0.216	0.466	0.464	0.643
No-TMS vs. 300ms	0.311	0.458	0.679	0.497
No-TMS vs. 375ms	0.656	0.435	1.506	0.132
No-TMS vs. 450ms	-0.418	0.542	-0.771	0.441

3.3 Naming Latencies

Although the stimuli had been pre-tested so that monosyllabic stimuli had an average naming latency of 600ms and the bisyllabic stimuli of 650ms, the mean naming latencies in the TMS study turned out to be approximately 550ms for monosyllabic words and 600ms for bisyllabic words. It may be the case that participants were anticipating the TMS stimulation and therefore were more alert and responded faster than during the pre-test where no such anticipation was present. Even though naming was faster than pre-tested, the around 50ms gap between mono- and bisyllabic naming latencies was still present, meaning that our hypothesis concerning word length remained testable. However, as a result of faster naming latencies, it became apparent that many monosyllabic items were named prior to TMS stimulation when stimulation occurred in the 525-575 time-window (41.97%). This issue was also present during the bisyllabic session (22.8%). These trials are problematic since they do not reflect TMS effects, yet removing them skews the data for the 525 time-window and hence affects the model outputs. We therefore removed this time-window from the analysis. This issue did not affect the remaining time-windows.

For the within stimulation site analyses, **Fig. 5** shows the mean naming latencies across time-points for each respective stimulation site. For all three stimulation sites, descriptively, a general trend of facilitation (e.g., compare the 225 ms time window to the no-TMS condition) from TMS is present during the earlier time-windows whereas a general trend towards interference is present in later time-windows. During monosyllabic picture naming, TMS to left IFG significantly increased naming latencies when stimulated in the 450ms window ($p < 0.001$)**.** IFG stimulation did not yield any significant effects for bisyllabic naming. pMTG stimulation also increased naming latencies when stimulation was applied at 450ms ($p < 0.001$) during monosyllabic naming but not during bisyllabic naming. There were however facilitatory effects during the earlier time-windows of 225ms ($p = 0.008$) and 300ms ($p = 0.007$) when

pMTG was stimulated during bisyllabic picture naming. Finally, pSTG stimulation resulted in an interference effect on naming latencies when stimulated at 375ms ($p < 0.001$) for monosyllabic naming. For bisyllabic naming, TMS to pSTG increased naming latencies during the 450ms time-window ($p = 0.003$). A statistical summary of the findings can be found in **Table 2**.

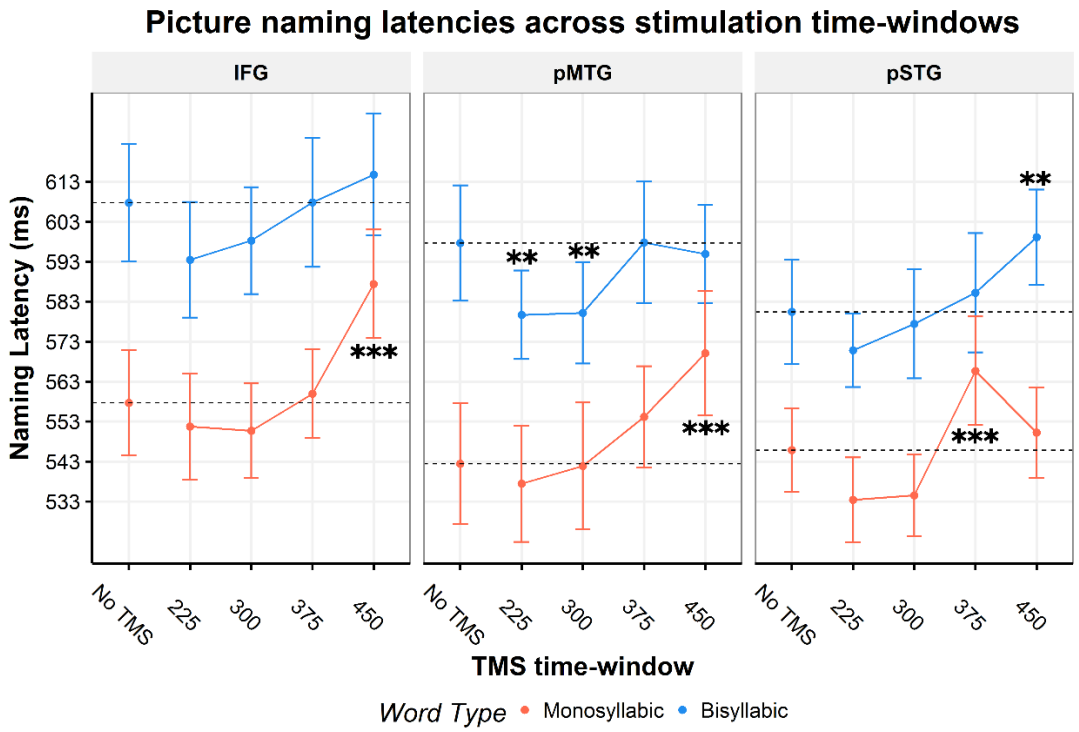

Fig 5. Naming latency results. Mean naming latencies across TMS time-windows for each stimulation site. Error bars indicate the standard error of the mean. Asterisks indicate significant effects relative to the no-TMS condition. ** $p < 0.01$, *** $p < 0.001$. IFG = inferior frontal gyrus, pMTG = posterior middle temporal gyrus, pSTG = posterior superior temporal gyrus.

Table 2. Inferential statistics for naming latencies. IFG = inferior frontal gyrus, pMTG = posterior middle temporal gyrus, pSTG = posterior superior temporal gyrus. Significant effects are p < 0.016.

Monosyllabic Session

IFG	**B**	**SE**	**t**	**p**
Intercept	556.343	12.887	43.172	**<0.001**
No-TMS vs. 225ms	-8.85	6.966	-1.27	0.204
No-TMS vs. 300ms	-5.949	6.966	-0.854	0.393
No-TMS vs. 375ms	2.905	7.066	0.411	0.681
No-TMS vs. 450ms	28.543	7.025	4.063	**<0.001**
pMTG				
Intercept	542.714	14.764	36.758	**<0.001**
No-TMS vs. 225ms	-5.438	5.68	-0.957	0.339
No-TMS vs. 300ms	-0.591	5.667	-0.104	0.917
No-TMS vs. 375ms	11.884	5.73	2.074	0.038
No-TMS vs. 450ms	27.321	5.71	4.785	**<0.001**
pSTG				
Intercept	545.626	10.99	49.65	**<0.001**
No-TMS vs. 225ms	-12.404	5.495	-2.257	0.024
No-TMS vs. 300ms	-10.687	5.546	-1.927	0.054
No-TMS vs. 375ms	20.479	5.561	3.682	**<0.001**
No-TMS vs. 450ms	5.689	5.577	1.02	0.308

Bisyllabic session

IFG	**B**	**SE**	**t**	**p**
Intercept	607.98	14.273	42.597	**<0.001**
No-TMS vs. 225ms	-12.872	6.841	-1.882	0.06
No-TMS vs. 300ms	-9.655	6.836	-1.412	0.158
No-TMS vs. 375ms	0.321	6.87	0.047	0.963
No-TMS vs. 450ms	6.224	7.067	0.881	0.379
pMTG				
Intercept	598.678	12.951	46.226	**<0.001**
No-TMS vs. 225ms	-18.026	6.872	-2.623	**0.009**
No-TMS vs. 300ms	-18.163	6.856	-2.649	**0.008**
No-TMS vs. 375ms	-1.036	6.832	-0.152	0.88
No-TMS vs. 450ms	-3.978	6.936	-0.574	0.566
pSTG				
Intercept	581.518	12.75	45.608	**<0.001**
No-TMS vs. 225ms	-10.291	6.37	-1.616	0.106
No-TMS vs. 300ms	-3.627	6.393	-0.567	0.571
No-TMS vs. 375ms	5.056	6.422	0.787	0.431
No-TMS vs. 450ms	18.619	6.364	2.925	**0.003**

3.4 Baseline Response-Locked Analysis

Fig. 6 shows participants' naming latencies in the baseline (no TMS) condition averaged over pictures and stimulation sites. As mentioned earlier, it is evident that naming latencies across participants varied substantially even after careful selection of materials for which naming latencies were pre-determined to be homogeneously around 600ms. This means that TMS stimulation during, for example, the 300ms time-window might disrupt a given processing stage in one participant and completely miss it in another, simply due to this individual variability in naming speed.

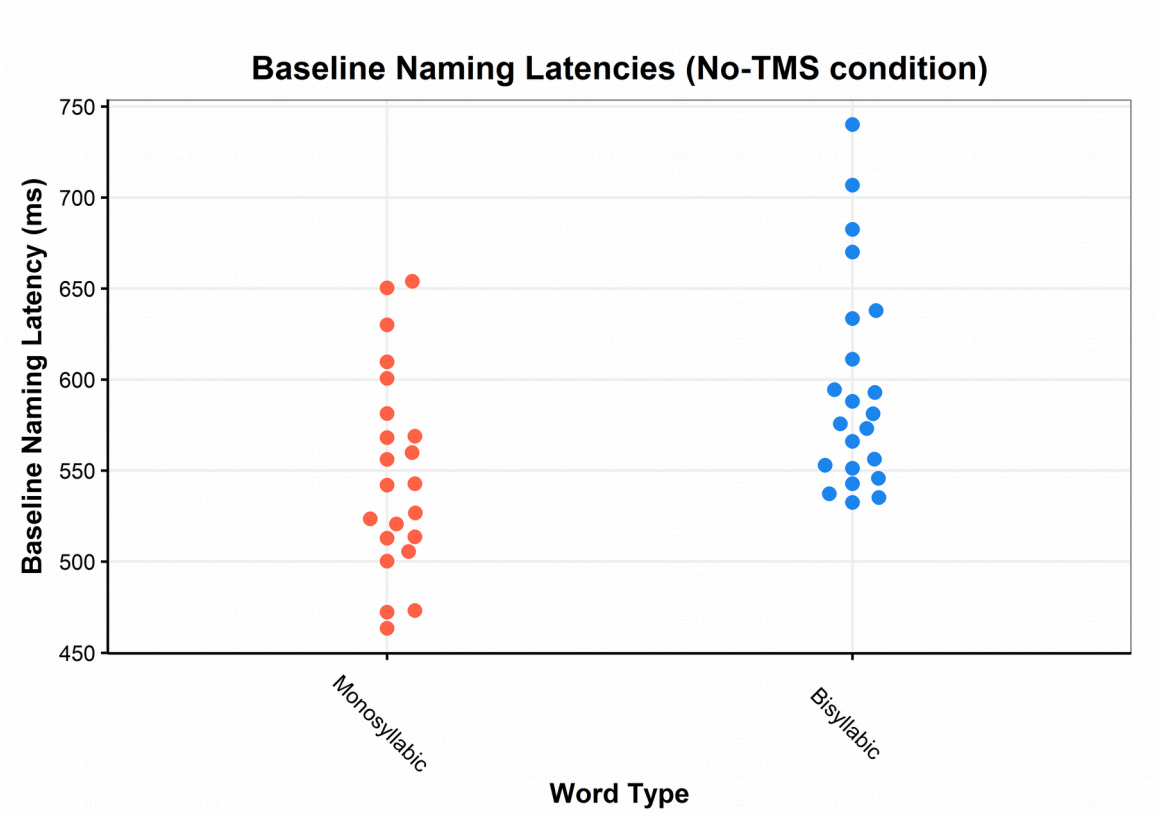

Fig. 6. Baseline Naming Latencies. Participants' baseline naming latencies during the no-TMS condition. Each dot represents one participant.

With varying naming latencies, the temporal estimates of the initial stages of word-production might be more consistent when measured from stimulus-onset, whereas the temporal estimates of later stages

may be more consistent when measured from the response. The more processing moves forward in the chain of events, the greater the chance that different participants will be at different stages of the word production process. For this reason, the baseline response-locked analysis was performed in order to better understand effects at late stages of word production.

A cluster-based permutation analysis was performed to examine whether certain time points had naming latencies that significantly deviated from baseline naming latencies. Generally speaking, the same qualitative trend is observed across all stimulation sites and word types: facilitation effects at earlier time points and interference effects closer to speech onset. IFG stimulation led to interference in naming latencies when stimulation occurred between 150-50ms prior to speech onset for monosyllabic naming whereas this was limited to 125-100ms prior to speech for bisyllabic naming. Stimulating pMTG resulted in significant interference in naming latencies between 125-25ms prior to speech onset during monosyllabic naming and 125-75ms during bisyllabic naming. Furthermore, pMTG stimulation also resulted in significant facilitation between 300-225ms prior to speech onset, roughly corresponding to the facilitation effects found in the standard analysis. Finally, stimulating pSTG resulted in significant facilitation in naming latencies between 275-225ms and significant interference between 150-50ms prior to speech onset for monosyllabic naming. During bisyllabic naming, pSTG stimulation interfered with naming between 100-50ms prior to speech onset.

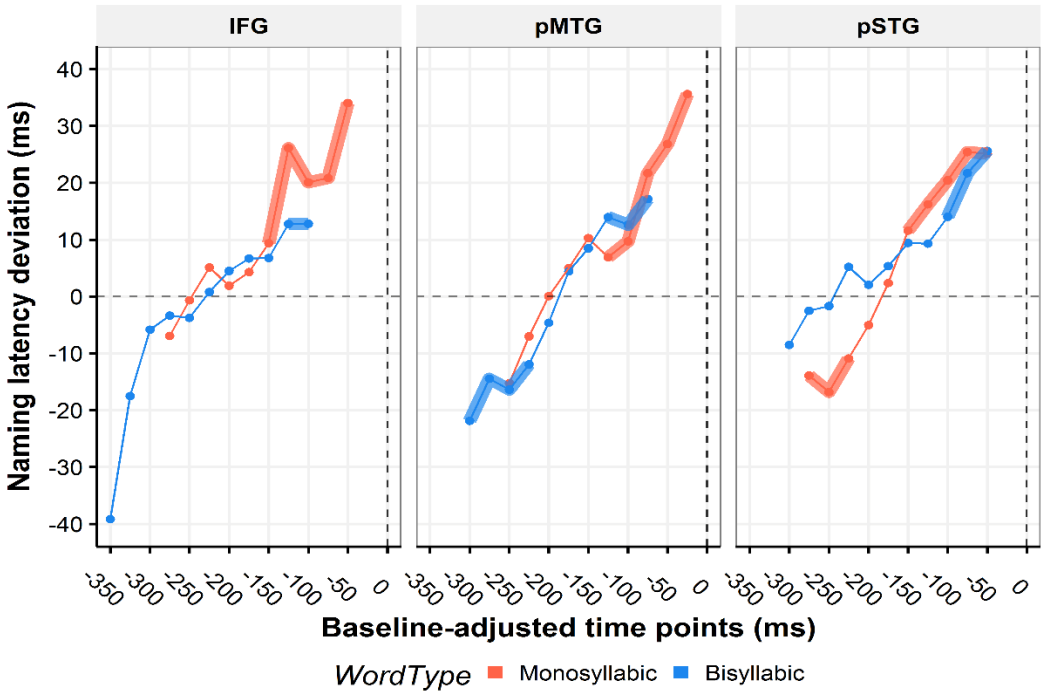

Fig 7. Baseline Adjusted Response-Locked analysis. Response-locked time series for each stimulation site. Bolded lines represent significant clusters ($p<0.05$) yielded from the cluster-based permutation analysis. IFG = inferior frontal gyrus, pMTG = posterior middle temporal gyrus, pSTG = posterior superior temporal gyrus.

3.5 Non-neuronal effects of TMS

In line with this descriptive trend of facilitation-like effects in earlier time-windows and interference-like effects in later ones, the model investigating main effects of TMS pulse time did yield a significant effect but only for the 300 vs. 375ms time-window ($p < 0.001$). Specifically, as compared to the 300ms time-window, stimulating any of the three regions at 375ms post-stimulus onset resulted in significantly longer naming latencies. Furthermore, the model for stimulation site also yielded an effect. Stimulating IFG ($p < 0.001$) as compared to pMTG resulted in longer naming latencies. This effect was not present for

pSTG stimulation as compared to pMTG. The last model with stimulation site and subjective discomfort as fixed factors, IFG stimulation was still significant, although to a lesser degree ($p = 0.01$) and subjective discomfort was also found to be significant ($p < 0.001$). Upon preforming an ANOVA, it became clear that adding subjective discomfort as a fixed factor significantly improved the model fit ($p < 0.001$). A statistical summary can be found in **table 3**.

Table 3. Main and non-neuronal effects for naming latencies. IFG = inferior frontal gyrus, pMTG = posterior middle temporal gyrus, pSTG = posterior superior temporal gyrus. Significant effects are $p < 0.016$.

Main effect of TMS pulse

Condition/parameter	B	SE	t	p
Intercept	572.034	10.677	53.576	**<0.001**
300ms vs. 225ms	1.78	2.892	0.616	0.538
375ms vs. 300ms	15.581	2.912	5.351	**<0.001**
450ms vs. 375ms	6.859	2.941	2.332	0.0197

Main effect of Stimulation site

Condition/parameter	B	SE	T	p
Intercept	568.844	10.766	52.837	**<0.001**
IFG vs. pMTG	13.31	2.542	5.237	**<0.001**
pSTG vs. pMTG	-4.679	2.543	-1.840	0.066

Main effect of Stimulation site and Subjective Discomfort

Condition/parameter	B	SE	t	p
Intercept	554.022	11.527	48.065	**<0.001**
IFG vs. pMTG	7.584	2.944	2.576	**0.01**
pSTG vs. pMTG	-1.843	2.644	-0.697	0.486
Subjective Discomfort	4.054	1.056	3.838	**<0.001**

ANOVA

Model	# Parameters	AIC	p
Stimulation site	6	55008	
Stimulation site + Subjective discomfort	7	54995	**<0.001**

4. Discussion

The present study investigated the temporal contribution of the left IFG, pMTG & pSTG to picture naming. We sought to replicate previous findings of (Schuhmann et al., 2012), while addressing issues regarding inconsistencies with the literature. Moreover, we also conducted a baseline response-locked analysis in order to control for the variation in naming latencies among participants under the rationale that the timing of processing stages is likely to increasingly diverge across participants, as well as from established temporal estimates, as time unfolds. Therefore, the standard, stimulus-locked analysis should be more informative of earlier time-window effects, whereas response-locked analysis should be more informative for the late stages. Finally, non-neuronal effects of TMS were also investigated.

The most consistent observation is that TMS has the most perturbing effects, as indexed by a slowing of naming latency, when delivered late in the planning phase, close to articulatory onset. This effect was observed for all three stimulation targets. This could suggest an overall higher susceptibility for perturbation, perhaps with less opportunity for compensation, late in the naming processing, or it could be indicative of a non-neuronal effect of TMS. Nonetheless, irrespective of the general slowing for late TMS, we observed specific effects dependent on stimulation site. Below, we first discuss our findings per region separately, before providing an integrative view over our findings.

4.1 Temporal Dynamics of Word Production

4.1.1 IFG

During monosyllabic naming, IFG seems to contribute to monosyllabic picture naming around 450-500ms. Previous studies have found IFG to be causally involved in earlier time-windows of 200-250ms (Hämäläinen et al., 2018), 225-275ms (Wheat et al., 2013; Zhang et al., 2018), 300-350ms (Hämäläinen et al., 2018; Schuhmann et al., 2009, 2012; Shinshi et al., 2015; Wheat et al., 2013; Zhang et al., 2018),

and 400-450ms (Zhang et al., 2018). IFG involvement at 450-500ms in the present study, therefore, seems rather late. The time window also appears to be late compared to the I&L model estimates (~355-455ms). The standard analysis showed no IFG effects during the bisyllabic naming session.

Analyzing the data in a response-locked fashion showed IFG stimulation to significantly increase naming latency from baseline at 150-50ms prior to speech during monosyllabic naming which corresponds to the 450-500ms time-window in the standard analysis. During the bisyllabic session, IFG stimulation significantly increased naming latency deviations between 125-100ms prior to speech. Importantly, as suggested by the mean baseline naming latencies shown in Figure 6, stimulation at 450ms (latest time-window) fell within 100ms before baseline response onset for about two thirds of the trials during monosyllabic naming. However, during bisyllabic naming the same stimulation only fell within 100ms before response onset in around 10% of the trials. Hence possible interference due to stimulation shortly before articulation could not have been detected for bisyllabic words in the standard analysis. By contrast, the 125-100ms stimulation time point (prior to speech-onset) in the baseline response-locked analysis was able to account for longer latency bisyllabic words, so that significant interference by stimulation shortly before articulation onset could be detected. In other words, what is a non-significant trend in Figure 5 become significant in Figure 7. This pattern of results confirms our assumption that a baseline response-locked analysis should be superior to the standard analysis for the detection of stimulation effects at late processing stages.

In sum, IFG seems to significantly contribute to naming at 150-50ms (monosyllabic words) and 125-100ms (bisyllabic words) prior to speech onset. Based on the 600ms speech onset latency assumed by the I&L model, these intervals would correspond to an interval of 450-550ms post picture onset that is rather late compared to the estimate of the model (355-455ms). However, when comparing our results from the monosyllabic naming session to those of Schuhmann et al. (2012), although the timing relative to stimulus-onset does not correspond, the timing relative to speech-onset does seem to fall in-line. In

our study, IFG stimulation showed effects 150-50ms prior to speech onset corresponding to 400-500ms from stimulus onset (taking the mean naming latency of approximately 550ms in the no-TMS condition to be the baseline). In the Schuhmann et al. (2012) study, the baseline naming latency was approximately 450ms, meaning that their IFG effect at 300-350ms also coincides with the interval of 150-100ms prior to speech onset.

It thus seems that our first hypothesis was confirmed: The time window of stimulation effects in IFG indeed depends on speech onset latencies. The time window of stimulation inducing an IFG effect was shifted by about the increase in onset latencies compared to Schuhmann et al. (2012). We can, therefore, interim conclude that IFG stimulation affects a late process. However, it may well be that this late process is articulatory planning rather than syllabification or phonetic encoding and that the effect is caused by non-neuronal effects of stimulation rather than cortical stimulation, as discussed in more detail below.

4.1.2 pMTG

pMTG stimulation resulted in interference effects in the 450-500ms time-window during monosyllabic naming and facilitatory effects in early time-windows of 225-275ms and 300-350ms during bisyllabic naming. Studies using chronometric TMS to investigate the role of non-IFG regions in picture naming are much fewer. Only one study used online TMS to probe pMTG during picture naming and only in a time-window of -100ms (prior to picture-onset) to 200ms post-picture onset (Acheson et al., 2011). The authors did find this stimulation to significantly delay picture naming, however it is important to note that their stimulation site was placed at the border of pMTG and mMTG, hence stimulation possibly affected the earlier lemma selection stage (Indefrey & Levelt, 2004). There are no other studies with which to compare our results directly. In the I&L model, pMTG is thought to be involved in phonological

code retrieval ~275-355ms. Our monosyllabic effect at 450-500ms is late compared to the model but our effects on naming bisyllabic words at 225-275ms and 300-375ms overlap with the model's estimates.

In the baseline response-locked analysis, pMTG stimulation during monosyllabic naming showed similar results to those of IFG but with a slight shift, namely significant increases in naming latencies when stimulated 125-25ms prior to speech onset. This time-window corresponds to the 450-500ms time-window from the standard analysis. Similar to our findings on IFG, there was also a late effect for pMTG stimulation during bisyllabic naming (125-75ms prior to speech onset) that was not present in the standard analysis. An earlier facilitation effect was also observed when pMTG was stimulated at 300-225ms before speech onset. This roughly corresponds to the facilitation effects found in the 225-350ms time-window from the standard analysis.

Thus, pMTG tends to present a similar issue as IFG, namely overlap in later time-windows between the two baseline response-locked analyses (125-25 and 125-75ms prior to speech onset) and the standard analysis during monosyllabic naming (450-500ms), but no such effect during bisyllabic naming in the standard analysis.

The pMTG site was originally chosen to attempt to interfere with phonological code retrieval thought to take place around 275-355ms post-stimulus onset. Although not present in the monosyllabic session, effects between 225-350ms were observed in the bisyllabic session, which match the theorized time-window. Furthermore, based on the mean naming latency of approximately 600ms in the no-TMS condition, the early significant effects from the baseline response-locked analysis (300-225ms prior to speech onset) correspond to 300-355ms post-stimulus onset, which is also in line with the temporal estimate for phonological code retrieval. However, it is unclear why this effect is only present during bisyllabic naming. One potential reason could lie in the lack of sensitivity of our exploratory baseline response-locked analysis for earlier processing stages. As can be seen in Fig. 7, pMTG also shows a

facilitation effect around 250ms prior to speech onset, but there is insufficient data to look at effects further back in time.

4.1.3 pSTG

pSTG stimulation replicated the interference effect at 400-500ms reported by Schuhmann et al. (2012). We found pSTG stimulation to interfere with the production of monosyllabic words when delivered at 375-425ms (150-50 prior to baseline naming onset) and to interfere with the production of bisyllabic words when delivered at 450-500ms (100-50 prior to baseline naming onset).

The later time-window in the bisyllabic session may reflect the later word onsets, shifting in time appropriately. These findings also fall in line with pSTG's role in self-monitoring as assumed by Indefrey & Levelt (2004). Importantly, the baseline response-locked analysis showed that TMS delivered to pSTG also had an earlier facilitation effect at 275-225ms prior to baseline speech onset which was not observed in the standard analysis. It is worth noting that pSTG stimulation in the 225-275 and 300-350ms time-windows did show an effect in the standard analysis but it did not surpass the alpha level we adopted to control for multiple comparisons.

Overall, the late effects in pSTG are consistent with a role of this region in self-monitoring as assigned in the I&L model. However, similar to our cautionary remark in the IFG effect above, we think that alternative explanations for this late stimulation effect cannot be ruled out (see below).

The facilitation effect at 275-225ms prior to speech onset during monosyllabic naming overlaps in time with the pMTG effect at 300-225ms prior to speech onset during bisyllabic naming. These effects in part confirm our second hypothesis that TMS stimulation in the posterior temporal lobe should not only have a late effect but also affect naming responses when delivered in the earlier time window of phonological code retrieval assumed by the I&L model (275-355ms post picture onset, respectively 325-245ms prior to speech onset). Nonetheless, there are open issues. Effects of stimulation in the two posterior temporal

regions differed between shorter and longer picture names and, most importantly, we found facilitation effects rather than the expected interference.

4.1.4 Network effects

The three target sites of stimulation, IFG, pMTG, and pSTG, together form a closely connected network. The late time-window effects observed across all three sites may also reflect ongoing interactions between these three regions. For example, syllabification processes in IFG may require the phonological code retrieved from posterior temporal areas to remain online. In this case, perturbing posterior temporal areas may interfere with ongoing interactions that cause indirect delays in syllabification. Alternatively, TMS to one brain region has also been shown to affect distal areas in a network. In the motor-domain, stimulating the hand area of the motor cortex in one hemisphere causes (partial) inhibition of the contralateral hand area 8-12ms and 20-50ms after the stimulation (Ni et al., 2009). Although this example is specific to the motor cortex, such effects cannot be excluded for word production considering the white matter connections between IFG, pMTG and pSTG (Catani & Bambini, 2014; Dick et al., 2014; Ortiz et al., 2021).

4.1.5 Facilitation vs. interference

Typically, online TMS protocols tend to interfere with ongoing activity. Unexpected improvements in performance from such protocols have therefore been appropriately termed “paradoxical facilitation” (Bergmann & Hartwigsen, 2021). As the name might suggest, the underlying mechanisms as to how TMS causes such facilitation effects remains largely unclear although some suggestions have been put forward (see Bergmann & Hartwigsen (2021) for an overview). Firstly, TMS is best thought of as adding neural noise to ongoing neural activity. How this noise interacts with ongoing processing will determine whether there is interference or facilitation. An optimal amount of noise may boost the signal and

facilitate behavior, whereas non-optimal amounts may add to the neural noise, decreasing the signal and causing interference (Abrahamyan et al., 2011; Miniussi et al., 2010).

Alternatively, facilitation may also arise from interference in a brain area that is not relevant to the current task. In other words, if a brain area is processing task-irrelevant stimuli or is competing for resources, then interfering with this brain region would facilitate behavior (Luber & Lisanby, 2014; Walsh et al., 1998). However, given the simplicity of our picture naming task, it is unlikely that interfering with resource competition or distractor processing (Piai et al., 2020) underlay the facilitation in the current study. Lastly, referring back to the previous section (4.1.4) about network effects, the final manner in which facilitation may arise is via disinhibition. Stimulating a node in the network may cause disinhibition of a connected node and therefore facilitate behavior (Sandrini, 2011). Given the connectedness of posterior temporal regions with other language network nodes, this final possibility may be likely.

Finally, facilitation effects may also arise from non-neuronal effects of TMS (which will be discussed in the subsequent section). Overall, facilitation effects in online TMS protocols are still quite perplexing and the exact cause remains unknown. Of the three possibilities mentioned above, assuming cortical stimulation, optimal noise or disinhibition of a connected brain region are likely to underlie the facilitation effects observed in the present study.

4.2 Neuromodulation or Non-Neuronal effects of TMS

4.2.1 Stimulation site and subjective discomfort

One important question to ask with any online TMS experiment is whether the findings reflect true neural effects or whether they are a consequence of non-neuronal effects induced by the TMS stimulation. As mentioned earlier, Meteyard & Holmes (2018), found that discomfort varies across the scalp depending on the underlying anatomy, and that subjective discomfort ratings were able to predict reaction time differences in previously published studies (Holmes & Meteyard, 2018). Generally, the

more uncomfortable the stimulation is perceived to be, the slower reaction times. In the present study, discomfort was significantly higher for IFG stimulation as compared to pMTG and pSTG, however there was no difference between pMTG and pSTG. This falls in line with frontal areas being more uncomfortable, as more muscles are present in this region. As for pMTG and pSTG, their scalp locations are in the same general vicinity and therefore the lack of a significant difference in perceived discomfort also makes sense. Interestingly, IFG stimulation also resulted in significantly longer naming latencies as compared to pMTG, whereas no difference in naming latencies was present when comparing pSTG to pMTG stimulation. Therefore, some correspondence between site-specific subjective discomfort and naming latencies seems to exist, with the site with significantly higher discomfort ratings also resulted in significantly longer naming latencies. Furthermore, when comparing the mixed effects models, the model that included discomfort ratings as a fixed factor yielded a significant improvement in model fit. Meaning that subjective discomfort did account for our findings on top of just stimulation site.

4.2.2 Time-windows

Non-neuronal TMS effects can also differ depending on the timing of TMS stimulation during an ongoing process. (Duecker & Sack, 2013) thoroughly investigated the effects that sham stimulation (i.e., stimulation that mimics real TMS but does not stimulate the cortex) has on performance and found that sham stimulation alone can speed up or slow down response times (Duecker et al., 2013; Duecker & Sack, 2015). In their study, the researchers performed a single pulse chronometric TMS experiment where they stimulated the vertex (often used as a control region in TMS studies) with real and sham TMS. Their results showed that both sham and real TMS had effects on the subsequent reaction times as compared to a no TMS condition despite the fact that both sham stimulation and real TMS stimulation over the vertex should not have affected cognitive processes. Our current results show a similar pattern of facilitation effects in earlier time-windows and interference effects in later time-windows to their sham condition during an angle judgement task (Duecker et al., 2013). Interestingly, in the present study

a significant effect of TMS pulse time was present when the 375ms time-windows was compared to the 300ms time-window, irrespective of stimulation site or word type. Although the other contrasts were not significant, a clear linear trend remarkably similar to that found in Duecker et al. (2013) is present. Therefore, it is very likely that non-neuronal effects of TMS are also present in our study.

Although non-neuronal effects of TMS are present, it does not exclude the presence of genuine *cortical effects* of TMS. However, it does mean that the results of the present study should be interpreted in the context of these non-neuronal effects, as it remains unclear to what degree non-neuronal or cortical stimulation are responsible for the behavioral outcomes. As mentioned before, all three of our stimulation sites are interconnected and ongoing interactions among them may also contribute to the similarity on effects. The addition of a non-language stimulation site may help untangle neuromodulatory vs. non-neuronal contributions of TMS.

4.2.3 Non-neuronal effects in Schuhmann et al. (2012)

An immediate question regarding our apparent non-neuronal effects of TMS, is why they were not present in the original Schuhmann et al. (2012) study. Two factors are likely at play. First, they had their participants practice naming the stimuli prior to the experiment until naming latencies were stable. This practice likely explains the overall fast naming latencies (460ms) and could explain the lack of temporal non-neuronal effects of TMS. In the present study, participants were also familiarized with pictures and named each picture once prior to the start of the experiment, however we decided against repeated practice prior to the experiment. The rational for doing so was that repeated practice could result in participants adopting different strategies, for example a stimulus-response type process as opposed to naturalistic picture naming, or that repeated practice may change the temporal recruitment of processing stages involved in word production.

Secondly, the differences also depend on TMS exposure and the perceived discomfort. As mentioned before subjective discomfort varies between stimulation sites but it will also differ among participants. Additionally, participants also seem to habituate to TMS discomfort. As shown in the present study , the same participants stimulated in the exact same sites reported less discomfort for the second session. To our knowledge, this is the only TMS language study that stimulated the same brain regions within the same participants across two different sessions using online chronometric TMS. Although, it is not clear how participants habituate to the discomfort or how that in turn affects behavioural outcomes, it is clear that exposure to TMS does play a role. Thus, if participants from the Schuhmann study had enough TMS exposure through previous studies (habituation) or did not experience significant discomfort, that may explain the lack of non-neuronal TMS effects in their study. In contrast, only 18% (4/22) of participants recruited for the present study had undergone TMS before. This may explain why non-neuronal effects are more pronounced in our study. Future studies might benefit from a habituation block, allowing participants to get used to the specific TMS protocol used in the experiment before beginning data collection, especially for TMS-naive participants.

4.3 Limitations of Chronometric TMS in Overt Production Tasks

4.3.1 TMS muscle twitches and articulatory motor planning

When stimulating the cortex, TMS will inevitably stimulate muscles below or in the vicinity of the coil. Muscle twitches tend to become more problematic as one stimulates regions located lower on the scalp such as temporal regions where more muscle fibers are present. For IFG stimulation, it is very likely that stimulation, regardless of whether it is comfortable or not, will elicit some degree of muscle twitches due to the overlaying temporalis muscle in this location. The temporalis muscle is rather large and responsible for the elevation and retraction of the lower jaw. It becomes clear that stimulating this

muscle can directly interfere with overt articulation and can be likened to having one's arm bumped into as they are reaching to grab an object.

TMS can interfere with articulation by two possible mechanisms. First, the stimulation of the muscle directly, especially with high frequencies, can lead to tetanus-like effects which may delay articulation. This is a muscle specific effect. Second, stimulating the muscle will cause afferent signals to be relayed to the brain. How exactly these afferent signals may affect articulatory motor planning just prior to articulatory onset is unclear.

According to an influential model of speech motor control, the DIVA model (Guenther & Vladusich, 2012), the somatosensory state map is a crucial part of motor planning. In order to properly coordinate the movement of articulators, their initial positions must be known. The DIVA model begins at the level of phonetic encoding, thus somatosensory information becomes crucial for motor planning around 450ms according to the I&L model (Indefrey & Levelt, 2004). Prior to this, it may be possible for the articulators to stabilize after stimulation and therefore not have any effect once motor planning begins.

Although sparse, there are a few studies that directly investigated brain activity elicited by neuromuscular stimulation. Wegrzyk and colleagues (2017), for example, electrically stimulated the right triceps surae (calf muscle) and recorded the subsequent brain activity using fMRI. Electric stimulation of the muscle resulted in activation of M1, S1, S2, premotor cortex, putamen, thalamus, caudate nucleus, cerebellum and many other brain regions. These regions comprise the same motor network recruited during articulatory motor planning (Guenther & Vladusich, 2012). Regardless of how exactly muscular stimulation-induced brain activity may interact with ongoing articulatory motor planning, the vast amount of cortical activation elicited by muscular stimulation cannot be ignored.

No study to date has disentangled the effects of TMS muscular stimulation on articulatory motor planning. This greatly restricts one's ability to infer the true cause of the observed effects, especially in

late time-windows when articulatory planning is underway. Furthermore, recent work demonstrates that different onset phonemes lead to different 'articulatory onset to acoustic onset intervals' (AAI) (Jouen et al., 2021). This further complicates the issue, as some words might be more vulnerable to somatosensory effects than others.

4.3.2 Individual Variability and Confounds

As mentioned before, one of the main problems of chronometric TMS with word production paradigms is the individual variability in naming latencies. If naming speeds are off by 50ms relative to the temporal estimates used to motivate the time windows for stimulation, this could mean appropriately disrupting a processing stage in some participants and completely missing it in others. In Fig. 5, the variation in baseline naming latencies across participants and stimulation sites were shown. It is clear that there is great heterogeneity among participants as well as across stimulation sites even if one carefully selects the materials.

Reaction times can vary even when no stimulation is given. For example, Wheat et al. (2013) found a 34ms time difference for a "no-stimulation" condition between their TMS and sham-TMS sessions. Therefore, even in a condition where no stimulation was applied, there was a difference between the real TMS and sham TMS sessions of a magnitude similar to that of many psycholinguistic manipulations.

On top of accounting for variability in performance that is unrelated to stimulation (baseline naming latencies), it is also important to note that different comparisons yield different results. For example, Zhang et al. (2018) found significant effects of TMS at 225, 300 and 375ms as compared to sham but only at 225ms compared to vertex stimulation.

Therefore, similar to the present study, true TMS effects are likely to present themselves across analyses. Returning to Zhang et al. (2018), the significant effect of TMS at 225ms is present when compared to sham as well as to vertex stimulation, whereas 300 and 375ms fail to reach significance when compared

to vertex. The robustness of the 225ms effect across the two analyses validates this result. Similarly, in the present study, significant time-windows that arise across analyses are also more likely to reflect a true effect as compared to time-windows that only reach significance in one analysis.

5. Conclusion

In summary, the present study utilized online chronometric TMS to probe three different brain regions at one of five time-windows during picture naming. The study sought to replicate and extend the findings of Schuhmann et al. (2012), as well as elucidate empirical inconsistencies of such paradigms in the literature. Overall, IFG, pMTG and pSTG stimulation all significantly affected naming around 150-50ms prior to baseline speech onset. Posterior temporal lobe stimulation in an earlier time-window of 225-350ms facilitated naming. These findings are in line with assumed functional roles of IFG and pSTG in later processing stages (phonetic encoding, articulatory planning, self-monitoring) as well as a functional role of the posterior temporal lobe in phonological code retrieval.

However, in particular with respect to the effect of stimulation in later time windows, we are also critical in interpreting the results and raise issues regarding the non-neuronal effects of TMS, the effects of individual variability in naming latencies, and muscle stimulation confounds. More than providing theoretical results, the present study's main contribution thus lies in offering new ways of approaching the data and raising key considerations for interpreting results from chronometric TMS studies investigating overt production. Specifically, response-locked analyses are recommended for chronometric TMS studies that have time-windows close to speech onset or where variation in individual naming latencies is present. Lastly, readers and researchers interpreting results of chronometric TMS studies should take care to consider alternative methodological explanations and not assume that behavioral findings are solely driven by cortical stimulation.

Funding

This work was supported by a Research Fellowship of the Max Planck Gesellschaft to Peter Indefrey. Lennart Verhagen was funded by a VIDI grant (18919) by the Dutch Research Council (NWO). Ian Cameron was funded by the European Regional Development Fund (PROJ-00872). Vitória Piai was funded by a VENI grant (451-17-003) by the Dutch Research Council (NWO) and by the Gravitation Grant (024.001.006) of the Language in Interaction Consortium from the Netherlands Organization for Scientific Research (NWO). Ian Cameron was funded by the European Regional Development Fund (PROJ-00872).

References

Abrahamyan, A., Clifford, C. W. G., Arabzadeh, E., & Harris, J. A. (2011). Improving Visual Sensitivity with Subthreshold Transcranial Magnetic Stimulation. *Journal of Neuroscience*, *31*(9), 3290–3294. https://doi.org/10.1523/JNEUROSCI.6256-10.2011

Acheson, D. J., Hamidi, M., Binder, J. R., & Postle, B. R. (2011). A Common Neural Substrate for Language Production and Verbal Working Memory. *Journal of Cognitive Neuroscience*, *23*(6), 1358–1367. https://doi.org/10.1162/jocn.2010.21519

Bergmann, T. O., & Hartwigsen, G. (2021). Inferring Causality from Noninvasive Brain Stimulation in Cognitive Neuroscience. *Journal of Cognitive Neuroscience*, *33*(2), 195–225. https://doi.org/10.1162/jocn_a_01591

Boersma, P., & van Heuven, V. (2001). *Speak and unSpeak with PRAAT*. *5*(9), 7.

Boersma, P., & Weenink, D. (2019). *Praat: Doing Phonetics by Computer [Windows]*. Version 6.1.01, retrieved 12 February, 2019 from http://www.praat.org/

Catani, M., & Bambini, V. (2014). A model for Social Communication And Language Evolution and Development (SCALED). *Current Opinion in Neurobiology*, *28*, 165–171. https://doi.org/10.1016/j.conb.2014.07.018

Dick, A. S., Bernal, B., & Tremblay, P. (2014). The Language Connectome: New Pathways, New Concepts. *The Neuroscientist*, *20*(5), 453–467. https://doi.org/10.1177/1073858413513502

Duecker, F., de Graaf, T. A., Jacobs, C., & Sack, A. T. (2013). Time- and Task-Dependent Non-Neural Effects of Real and Sham TMS. *PLoS ONE*, *8*(9), e73813. https://doi.org/10.1371/journal.pone.0073813

Duecker, F., & Sack, A. T. (2013). Pre-Stimulus Sham TMS Facilitates Target Detection. *PLoS ONE*, *8*(3), e57765. https://doi.org/10.1371/journal.pone.0057765

Duecker, F., & Sack, A. T. (2015). Rethinking the role of sham TMS. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00210

Guenther, F. H., & Vladusich, T. (2012). A neural theory of speech acquisition and production. *Journal of Neurolinguistics*, *25*(5), 408–422. https://doi.org/10.1016/j.jneuroling.2009.08.006

Hämäläinen, S., Mäkelä, N., Sairanen, V., Lehtonen, M., Kujala, T., & Leminen, A. (2018). TMS uncovers details about sub-regional language-specific processing networks in early bilinguals. *NeuroImage*, *171*, 209–221. https://doi.org/10.1016/j.neuroimage.2017.12.086

Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews. Neuroscience*, *13*(2), 135–145. https://doi.org/10.1038/nrn3158

Holmes, N. P., & Meteyard, L. (2018). Subjective Discomfort of TMS Predicts Reaction Times Differences in Published Studies. *Frontiers in Psychology*, *9*, 1989. https://doi.org/10.3389/fpsyg.2018.01989

Indefrey, P. (2011). The Spatial and Temporal Signatures of Word Production Components: A Critical Update. *Frontiers in Psychology*, *2*. https://doi.org/10.3389/fpsyg.2011.00255

Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition*, *92*(1–2), 101–144. https://doi.org/10.1016/j.cognition.2002.06.001

Jouen, A.-L., Lancheros, M., & Laganaro, M. (2021). Microstate ERP Analyses to Pinpoint the Articulatory Onset in Speech Production. *Brain Topography*, *34*(1), 29–40. https://doi.org/10.1007/s10548-020-00803-3

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). **lmerTest** Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13). https://doi.org/10.18637/jss.v082.i13

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *The Behavioral and Brain Sciences*, *22*(1), 1–38; discussion 38-75. https://doi.org/10.1017/s0140525x99001776

Luber, B., & Lisanby, S. H. (2014). Enhancement of human cognitive performance using transcranial magnetic stimulation (TMS). *NeuroImage*, *85*, 961–970. https://doi.org/10.1016/j.neuroimage.2013.06.007

Meteyard, L., & Holmes, N. P. (2018). TMS SMART – Scalp mapping of annoyance ratings and twitches caused by Transcranial Magnetic Stimulation. *Journal of Neuroscience Methods*, *299*, 34–44. https://doi.org/10.1016/j.jneumeth.2018.02.008

Miniussi, C., Ruzzoli, M., & Walsh, V. (2010). The mechanism of transcranial magnetic stimulation in cognition. *Cortex*, *46*(1), 128–130. https://doi.org/10.1016/j.cortex.2009.03.004

Ni, Z., Gunraj, C., Nelson, A. J., Yeh, I.-J., Castillo, G., Hoque, T., & Chen, R. (2009). Two Phases of Interhemispheric Inhibition between Motor Related Cortical Areas and the Primary Motor Cortex in Human. *Cerebral Cortex*, *19*(7), 1654–1665. https://doi.org/10.1093/cercor/bhn201

Ortiz, K. J., Hawayek, M. I., Middlebrooks, E. H., Sabsevitz, D. S., Garcia, D. P., Quinones-Hinojosa, A., & Chaichana, K. L. (2021). Intraoperative Direct Stimulation Identification and Preservation of Critical White Matter Tracts During Brain Surgery. *World Neurosurgery*, *146*, 64–74. https://doi.org/10.1016/j.wneu.2020.10.100

Piai, V., Nieberlein, L., & Hartwigsen, G. (2020). Effects of transcranial magnetic stimulation over the left posterior superior temporal gyrus on picture-word interference. *PLOS ONE*, 18.

Price, C. J. (2012). A review and synthesis of the first 20years of PET and fMRI studies of heard speech, spoken language and reading. *NeuroImage*, *62*(2), 816–847. https://doi.org/10.1016/j.neuroimage.2012.04.062

Rossi, S., Antal, A., Bestmann, S., Bikson, M., Brewer, C., Brockmöller, J., Carpenter, L. L., Cincotta, M., Chen, R., Daskalakis, J. D., Di Lazzaro, V., Fox, M. D., George, M. S., Gilbert, D., Kimiskidis, V. K., Koch, G., Ilmoniemi, R. J., Pascal Lefaucheur, J., Leocani, L., … Hallett, M. (2020). Safety and recommendations for TMS use in healthy subjects and patient populations, with updates on training, ethical and regulatory issues: Expert Guidelines. *Clinical Neurophysiology*. https://doi.org/10.1016/j.clinph.2020.10.003

Rossi, S., Hallett, M., Rossini, P. M., & Pascual-Leone, A. (2009). Safety, ethical considerations, and application guidelines for the use of transcranial magnetic stimulation in clinical practice and research. *Clinical Neurophysiology*, *120*(12), 2008–2039. https://doi.org/10.1016/j.clinph.2009.08.016

Sandrini, M. (2011). The use of transcranial magnetic stimulation in cognitive neuroscience: A new synthesis of methodological issues. *Neuroscience and Biobehavioral Reviews*, 21.

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*(9), 1111–1116. https://doi.org/10.1016/j.cortex.2008.10.013

Schuhmann, T., Schiller, N. O., Goebel, R., & Sack, A. T. (2012). Speaking of Which: Dissecting the Neurocognitive Network of Language Production in Picture Naming. *Cerebral Cortex*, *22*(3), 701–709. https://doi.org/10.1093/cercor/bhr155

Shinshi, M., Yanagisawa, T., Hirata, M., Goto, T., Sugata, H., Araki, T., Okamura, Y., Hasegawa, Y., Ihara, A. S., & Yorifuji, S. (2015). Temporospatial identification of language-related cortical function by a combination of transcranial magnetic stimulation and magnetoencephalography. *Brain and Behavior*, *5*(3), n/a-n/a. https://doi.org/10.1002/brb3.317

Strijkers, K., & Costa, A. (2016). The cortical dynamics of speaking: Present shortcomings and future avenues. *Language, Cognition and Neuroscience*, *31*(4), 484–503. https://doi.org/10.1080/23273798.2015.1120878

Voeten, C. C. (2019). *Analyzing time series data using the permutes package*. 15.

Voeten, C. C. (2021). *Analyzing time series data using clusterperm.lmer*. 7.

Walker, G. M., & Hickok, G. (2016). Bridging computational approaches to speech production: The semantic–lexical–auditory–motor model (SLAM). *Psychonomic Bulletin & Review*, *23*(2), 339–352. https://doi.org/10.3758/s13423-015-0903-7

Walsh, V., Ellison, A., Battelli, L., & Cowey, A. (1998). *Task-speci® c impairments and enhancements induced by magnetic stimulation of human visual area V5*. 8.

Wegrzyk, J., Ranjeva, J.-P., Fouré, A., Kavounoudias, A., Vilmen, C., Mattei, J.-P., Guye, M., Maffiuletti, N. A., Place, N., Bendahan, D., & Gondin, J. (2017). Specific brain activation patterns associated with two

neuromuscular electrical stimulation protocols. *Scientific Reports*, *7*(1), 2742. https://doi.org/10.1038/s41598-017-03188-9

Wheat, K. L., Cornelissen, P. L., Sack, A. T., Schuhmann, T., Goebel, R., & Blomert, L. (2013). Charting the functional relevance of Broca's area for visual word recognition and picture naming in Dutch using fMRI-guided TMS. *Brain and Language*, *125*(2), 223–230. https://doi.org/10.1016/j.bandl.2012.04.016

Zhang, Q., Yu, B., Zhang, J., Jin, Z., & Li, L. (2018). Probing the Timing Recruitment of Broca's Area in Speech Production for Mandarin Chinese: A TMS Study. *Frontiers in Human Neuroscience*, *12*, 133. https://doi.org/10.3389/fnhum.2018.00133