Received: 1 November 2019 | Revised: 4 May 2020 | Accepted: 13 May 2020

DOI: 10.1111/ejn.14785

RESEARCH REPORT

EJN European Journal of Neuroscience FENS WILEY

Across-session consistency of context-driven language processing: A magnetoencephalography study

Natascha Marie Roos[1] | **Vitória Piai**[1,2]

[1]Donders Center for Cognition, Radboud University, Nijmegen, The Netherlands

[2]Department of Medical Psychology, Radboudumc, Donders Centre for Medical Neuroscience, Nijmegen, The Netherlands

Correspondence
Natascha Marie Roos, Donders Center for Cognition, Radboud University, Montessorilaan 3, 6525 HR, Nijmegen, The Netherlands.
Email: n.roos@donders.ru.nl

Funding information
Nederlandse Organisatie voor Wetenschappelijk Onderzoek, Grant/Award Number: Gravitation Grant 024.001.006

Abstract

Changes in brain organization following damage are commonly observed, but they remain poorly understood. These changes are often studied with imaging techniques that overlook the temporal granularity at which language processes occur. By contrast, electrophysiological measures provide excellent temporal resolution. To test the suitability of magnetoencephalography (MEG) to track language-related neuroplasticity, the present study aimed at establishing the spectro-temporo-spatial across-session consistency of context-driven picture naming in healthy individuals, using MEG in two test–retest sessions. Spectro-temporo-spatial test–retest consistency in a healthy population is a prerequisite for studying neuronal changes in clinical populations over time. For this purpose, 15 healthy speakers were tested with MEG while performing a context-driven picture-naming task at two time points. Participants read a sentence missing the final word and named a picture completing the sentence. Sentences were constrained or unconstrained toward the picture, such that participants could either retrieve the picture name through sentence context (constrained sentences), or could only name it after the picture appeared (unconstrained sentences). The context effect (constrained versus unconstrained) in picture-naming times had a strong effect size and high across-session consistency. The context MEG results revealed alpha–beta power decreases (10–20 Hz) in the left temporal and inferior parietal lobule that were consistent across both sessions. As robust spectro-temporo-spatial findings in a healthy population are required for working toward longitudinal patient studies, we conclude that using context-driven language production and MEG is a suitable way to examine language-related neuroplasticity after brain damage.

KEYWORDS

facilitation, language production, test–retest reliability, temporal lobe | prediction

Abbreviations: BOLD, Blood Oxygen Level Dependent; DICS, Dynamical Imaging of Coherent Sources; fMRI, Functional Magnetic Resonance Imaging; MEG, Magnetoencephalography; MNI, Montreal Neurological Institute.

[Correction added on 25 June 2020, after first online publication: Peer review history statement has been added.]

The peer review history for this article is available at https://publons.com/publon/10.1111/EJN.14785

1 | INTRODUCTION

Damage to the brain very often leads to deficits affecting everyday life. In order to overcome these deficits, the brain can adapt to the damage and reorganize itself, a process also known as neuroplasticity. For instance, there have been several findings demonstrating temporary language deficits after left-hemisphere damage (Skipper-Kallal, Lacey, Xing, & Turkeltaub, 2017; Wilson et al., 2015), with beneficial recruitment of the right hemisphere for recovery (Piai, Meyer, Dronkers, & Knight, 2017; Turkeltaub, Messing, Norise, & Hamilton, 2011). This suggests that, in order to overcome language deficits due to brain lesions, different brain areas can be recruited (Brownsett et al., 2014; Kiran, 2012).

To track language-related neuroplasticity, data are ideally longitudinal in nature. This necessitates knowing how spatially reliable and consistent a given effect under study really is. Previous studies targeted this issue by capturing brain activation of language production at multiple time points, testing healthy controls (Meltzer, Postman-Caucheteux, McArdle, & Braun, 2009) as well as aphasic stroke patient subjects (Eaton et al., 2008). Both studies employed functional magnetic resonance imaging (fMRI) and concluded this method to be a suitable technique to longitudinally examine neuroplasticity of language production.

Following this line of reasoning, another study investigated the validity and reliability of different language production and comprehension paradigms to identify language areas with fMRI (Wilson, Bautista, Yen, Lauderdale, & Eriksson, 2017). Here, validity refers to the property of the paradigm to activate all and only those brain areas that have been shown to be essential for language processing, whereas reliability describes the consistency of measurements across more than one session. The authors concluded that sentence completion tasks provide the best-balanced combination of validity and reliability. However, they also pointed out general limitations of language mapping with fMRI in individuals, for clinical and research purposes. More specifically, the authors argued that a priori selected regions of interest improve results, but become problematic if the focus is to study brain *re*organization.

These as well as other authors prompted the field for similar assessments of validity and reliability measures of different paradigms (Bradshaw, Thompson, Wilson, Bishop, & Woodhead, 2017; Duffau, 2005). This would not only be beneficial for clinical purposes regarding applications in therapy, but also to create a thorough ground for interpreting existing research findings and serve as a base for future studies. In fact, it would be important to establish which distinct findings across studies really reflect differences on the brain level, rather than differences in methodologies across studies such as experimental paradigm, scanner type, or analysis strategies.

To date, the most common method to localize brain functions in neurological patients is fMRI. Here, the obtained signal is the blood oxygen level dependent (BOLD) response. As such, fMRI measures neuronal activity only indirectly with a rather slow temporal resolution dependent on the hemodynamic response (Veldsman, Cumming, & Brodtmann, 2015). Also, the relationship between the measured BOLD signal and the underlying neuronal activity is not always completely evident (Ekstrom, 2010). Importantly, brain damage can affect neurovascular coupling, impacting the BOLD response (Geranmayeh, Wise, Leech, & Murphy, 2015; Kim et al., 2005; Veldsman et al., 2015).

By contrast, direct measures of neuronal activity at the sub-second timescale in which language processes happen, as provided by electrophysiological methods such as magnetoencephalography (MEG), may be highly informative (Traut et al., 2019). Contrary to fMRI, MEG captures neuronal activity with a high temporal resolution and thus allows to track the time course of neuronal sources. MEG has also been successfully employed to determine the language-dominant hemisphere in patients before undergoing brain surgery (Findlay et al., 2012) as well as neuroplastic changes following resective surgery (Traut et al., 2019).

It has been shown that MEG and fMRI may measure different aspects of neuronal activity (Kujala et al., 2014; Liljeström, Hultén, Parkkonen, & Salmelin, 2009; Vartiainen, Liljeström, Koskinen, Renvall, & Salmelin, 2011), both possibly overlooking some aspects of ongoing brain activity. This emphasizes the need for additional techniques that are employed to measure brain activity to investigate neuroplasticity. The approach of investigating neuroplasticity by looking at brain activity and its changes over time bears the same methodological limitations as measuring brain activity (i.e., different methods measure different aspects). Therefore, given that every technique has caveats, a complete understanding of neuroplasticity necessitates finding converging evidence from other techniques, additional to fMRI. Ideally, electrophysiological and hemodynamic techniques would both be employed in a complementary way. In particular, whereas hemodynamic measures might be more suitable for examining the involvement of ("new") brain areas, electrophysiology might be more informative regarding spectro-temporal aspects of neuroplasticity ("what" these areas are doing, e.g., Piai et al., in press).

The present study was conducted with healthy control participants as a first step in this direction, as we do not expect any spectro-temporo-spatial changes between sessions for healthy participants. This means that we expect the profiles of brain activity to be consistent across sessions regarding the brain areas they appear in, as well as their spectro-temporal content. Thus, for the purpose of the present study, our definition of an appropriate across-session consistency focuses on aspects of space, frequency, and time. It is, however, not

based on statistical correlations between power values, as we do not expect those to remain fixed across sessions (e.g., they could differ due to different signal-to-noise ratios). In other words, if the profiles of brain activity differ in their spectro-temporo-spatial appearance per session, we would not consider our approach appropriate for the purpose of investigating neuroplasticity.

The goal was to determine how reliable the brain activity obtained with MEG is over time, as well as to evaluate how the effect size changes for shorter testing duration. Especially in studies with clinical populations, the duration of testing sessions should be kept as short as possible. But shorter testing times in that case also equal less acquired data. An ideal paradigm to investigate neuroplasticity would provide good spectro-temporo-spatial test–retest reliability despite a restricted amount of data.

We used the same sentence completion as picture-naming paradigm (i.e., context-driven picture naming) as previously employed in electrophysiological studies by our group (Klaus, Schutter, & Piai, 2020; Piai, Klaus, & Rossetto, 2020; Piai et al., 2017; Piai, Roelofs, & Maris, 2014; Piai, Roelofs, Rommers, & Maris, 2015; Piai, Rommers, & Knight, 2018). This paradigm has repeatedly elicited robust oscillatory power decreases in the alpha–beta range (8–25 Hz) prior to picture onset in healthy participants as well as stroke patients. Herein, sentences are presented without the last word, which is presented as a picture. The task is to pay attention to the sentence and name the picture. Sentences appear in two conditions such that the sentence context is either constrained or unconstrained toward the target word (for an example, see Figure 1). The sentence context in constrained sentences enables participants to retrieve information about the target word (and potentially prepare to name the picture) already before it appears. Unconstrained sentences, however, provide little information about the target word, and participants must wait for the picture to appear until they can retrieve conceptual and lexical information to name it. This means that trials of different conditions only vary in sentence context and this variation also yields the effect of interest. Thereby, the present paradigm offers a precise contrast between conditions and prevents from capturing condition-specific differences that are of no interest. Further, participants only perform one task which eliminates the risk of capturing possible task switching demands.

2 | METHOD

The present study was approved by the accredited ethical reviewing committee, CMO Arnhem-Nijmegen, following the Declaration of Helsinki. It was conducted at the Donders Institute for Cognitive Neuroimaging in Nijmegen in the Netherlands.

2.1 | Participants

A total of 15 native Dutch speakers, consisting of seven females and eight males, aged between 18 and 50 years (Mdn = 25, three participants older than 35) participated in the study for monetary compensation or course credits. All were healthy and right-handed, with normal or corrected-to-normal vision (no glasses).

2.2 | Materials

The stimuli consisted of 224 target words with a corresponding picture. This was a photograph depicting the target word on white background, if possible (except for landscape-like target words such as *forest* or *mountain*). Each target word was the last word of one constrained and one unconstrained sentence. As such, each target word had one corresponding sentence per condition, yielding 448 experimental sentences. All linguistic materials were in Dutch, taken from previous studies (Piai et al., 2014, 2015). Pictures were collected from the BOSS database (Brodeur, Dionne-Dostie, Montreuil, & Lepage, 2010) and via online search. The length of the

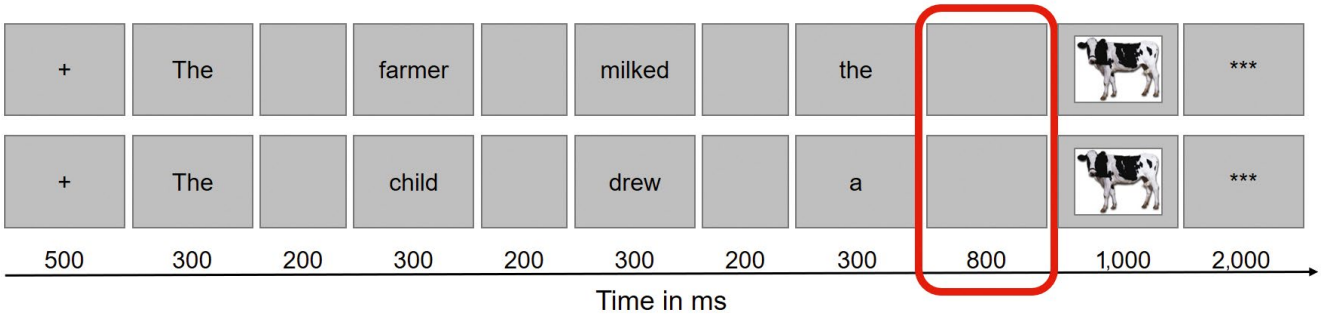


FIGURE 1 Overview of the constrained (top) and unconstrained (bottom) trial for the target picture cow. Boxes represent the screens that participants saw, with presentation durations below and the analyzed time window of interest circled in red [Colour figure can be viewed at wileyonlinelibrary.com]

target picture names varied from 2 to 11 phonemes (mean length = 5). Sentence length varied between 4 and 13 words including the target word (mean length = 7) and was as similar as possible for both sentences associated with the same target word.

2.3 | Design

The stimuli were presented in three main lists uniquely divided in half, controlled for frequency, word length, and initial picture-name phoneme. Each half was pseudorandomized using Mix (van Casteren & Davis, 2006) so that there were at least 20 trials between the first and the second appearance of the same target picture and a maximum of five repetitions of trials with the same condition. Participants were randomly assigned to one of the three main lists. As the study consisted of a test and a retest session scheduled between 14 and 28 days apart (Mdn = 20), each participant was presented with half of the target pictures per session, alternating the order of which half was presented first. Thus, one session consisted of 112 target pictures, once preceded by a constrained sentence and once by an unconstrained sentence, yielding 224 trials per session.

2.4 | Procedure

Each session started with instructing the participants about the task and the scanning session and clarifying possible questions. Then, participants signed the consent forms and were screened for MEG and MRI compatibility. Before entering the magnetically and electrically shielded room, they were familiarized with the pictures of the experiment and the corresponding target names. These were presented in a slide show with four pictures on one slide, and the target names printed below. Each session started with four practice trials in the MEG chair, so that participants knew what to expect and had the chance to clarify remaining doubts before the start of the experiment. Stimuli were presented with Presentation software (*Neurobehavioral Systems*, n.d.) and projected on a screen in front of the participants in the MEG room. Figure 1 shows a trial overview for both sentences for the same target word with the presentation durations in the experiment, highlighting the time window of interest. Each trial started with a 500 ms fixation cross and consisted of a word-by-word presentation of the sentence in the center of the screen. Each word was presented for 300 ms followed by a 200 ms blank screen. Words were presented in black on a gray background. The picture was then presented for 1,000 ms. The task was to silently read the sentences attentively and name the pictures with the words that participants were familiarized with. Also, they were asked to keep fixation to the center of the screen and move their jaw and head as little as possible.

2.5 | Behavioral analysis

In both sessions, vocal responses were audio-recorded to monitor participants' picture-naming performance. Recordings started simultaneously with picture onset and lasted 2,500 ms. Trials in which participants hesitated, stuttered, and responded either with more than one word or later than 2,500 ms after picture onset were considered as errors and not included in the analyses. Trials in which the response was a synonym to the original target word that makes sense in the sentence context of the corresponding trial were marked as correct. If participants' speech onset started prior to picture onset, no reaction time could be measured and the trial was discarded from analyses. Picture-naming times were calculated manually per trial, by means of the spectrogram and oscillogram of the audio-recorded vocal responses using the speech editor Praat (Boersma & Weenink, 2017) and statistically analyzed in R (R Core Team, 2017). The rater was blind for condition and for the purposes of the experiment. The behavioral data were analyzed using a linear mixed-effects regression with fixed effects of condition and session as well as their interaction, and by-participant and by-item random intercepts and random slopes for condition. In addition, we calculated Cohen's *d* (Cohen, 1988) as a measure of effect size for each session separately (i.e., the group mean difference between the two conditions divided by the standard deviation of the difference over participants).

2.6 | MEG acquisition

Participants had to change into the metal-free clothing provided in the MEG laboratory. Then, they were prepared with electrodes attached to their face and body to measure the vertical and horizontal electrooculogram, the mouth electromyogram, and the electrocardiogram. Electrode impedance was kept below 20 kOhm. Before and during the experiment, participants were instructed to restrict blinking to the inserted blinking intervals at the end of each trial indicated by three asterisks (***). MEG data were acquired with a 270 axial gradiometer system (CTF Systems Inc., VSM MedTech Ltd.) and a sampling rate of 1,200 Hz. Participants were positioned in the MEG chair with pillows as they preferred. Localization coils were attached to the left and right ear canal, and the nasion. Head localization was performed in real time (Stolk, Todorovic, Schoffelen, & Oostenveld, 2013), and the head position relative to the sensors at the start of session 1 was stored. This was used at the start of session 2 to place participants in the same position as in session 1. Then, this position

was updated to the real position at the start of session 2, to keep an overview of participants' head movement within the session. The head position was kept as constant as possible across trials and sessions to minimize noise deriving from head position variance. If participants moved more than 8 mm away from their initial position, they were repositioned in the breaks after every block of 28 trials. The scanning for one MEG session lasted approximately 30 min, and participants were in the laboratory for one hour, including preparation time. If not available, structural T1-weighted MRI scans of participants' heads were acquired either after one of the two MEG sessions or on a third occasion.

2.7 | MEG preprocessing

MEG data analyses were performed in Matlab using the FieldTrip toolbox (Oostenveld, Fries, Maris, & Schoffelen, 2011). The data were demeaned to take out drifts, and each trial was segmented to the time window of interest of −800 ms to picture onset. Incorrect trials were not considered, which led to a loss of 0 to 37 trials per session ($M = 6$, $SD = 7$). Subsequently, the data were down-sampled to 600 Hz and trials containing eye movements were discarded by inspection of the electrooculogram channels. This led to a loss of 0 to 26 trials per session ($M = 5$, $SD = 5$). Finally, remaining noisy trials and sensors were marked by means of a trial and sensor overview summary, plotting each trial and sensor by different parameters (variance, minimal value, maximum value, absolute maximum, range, kurtosis, 1/variance, z-value, maximum z-value). Per session, all plotting parameters were used to determine outlying trials and sensors not to be considered for further analyses. With this procedure, 8 to 20 outlying sensors were removed per session ($M = 15$, $SD = 3$; 19 to 20 sensors were removed in only two datasets, 18 sensors in only 4 datasets). Every session consisted of 170 to 213 trials for analysis after preprocessing ($M = 198$, $SD = 10$), including 82 to 110 unconstrained trials ($M = 99$, $SD = 6$) and 81 to 106 constrained trials ($M = 98$, $SD = 6$). To perform the sensor level analysis, we interpolated the missing data points of discarded sensors based on a weighted average of the neighboring sensors located around the missing location, taking into account the distance between sensor locations.

2.8 | MEG analysis

The MEG analyses were based on the differences in brain activity between the constrained and the unconstrained condition in the specified time window of interest (see Figure 1). This is the 800-ms interval between the offset of the last presented word and the onset of the picture in each trial, during which the screen was blank (Piai et al., 2014, 2015, 2017, 2018).

2.8.1 | Sensor level

We analyzed the sensor level time–frequency data during the 800-ms time window. This was computed at frequencies from 2 to 30 Hz with a sliding time window of 200 ms, advanced in steps of 20 ms and 1 Hz. Each time window was multiplied with a Hanning taper with implemented time–frequency transformation based on multiplication in the frequency domain. The difference in power between conditions at the sensor level was assessed by means of a non-parametric cluster-based permutation test (Maris & Oostenveld, 2007) on group level with 1,000 random permutations (alpha = 0.05, two-tailed). To examine the relationship between the neural and behavioral data, a correlation analysis was performed between picture-naming times and the time–frequency representations. Given the hypothesis that the alpha–beta power decreases in the pre-picture interval reflect conceptual preparation and word planning processes (Piai et al., 2015; Piai et al., 2020; Piai & Zheng, 2019), only the constrained condition was used, as conceptual and lexical processes cannot take place in the unconstrained condition before the picture is shown. First, for each participant individually, we correlated picture-naming times with the time–frequency-sensor data points on a trial-by-trial basis using the non-parametric cluster-based permutation framework. The individual correlation matrices then also entered a non-parametric cluster-based dependent-sample *t* test (two-tailed) versus zero, with 1,000 random permutations on the group level. By testing on the group level versus zero, one can reveal whether there is a consistent direction of the correlation (positive or negative) across individuals.

2.9 | Source level

To identify the sources of the obtained brain activity, the MRI scans of participants' heads were realigned with the coordinates of the MEG data by marking the fiducials in both ear canals and the nasion. Then, they were resliced and segmented into brain, scalp, and skull using SPM12, to obtain realistic single-shell models of the inside of the skull with a resolution of 10 mm, serving as individual volume conduction models of participants' heads (Nolte, 2003). Next, the realigned MRI scans were warped to the Montreal Neurological Institute (MNI) space template to obtain subject-specific source model grids in normalized space, so they could be compared across participants. Using the volume conduction models, lead field matrices were computed for each grid point per participant. Then, a frequency domain beamforming technique (DICS)

was applied to estimate the activity at the source level (Gross et al., 2001). The cross-spectral density matrix of the sensor level data for both conditions combined was computed at 15 Hz. Spectral smoothing of 5 Hz yielded a cross-spectral density matrix between 10 and 20 Hz. This frequency range was based on previous findings resulting from the same task and analysis techniques (Piai et al., 2015) as well as the present sensor level findings (see below). As the transition from alpha to beta activity is usually considered around 12–15 Hz, this frequency range is referred to as alpha–beta power in the present report. Together with the lead field matrices, the cross-spectral density matrices were used to calculate a common spatial filter for each grid point. These filters were then applied to the Fourier transformed sensor data per condition to estimate source level power for each grid point. Then, the power estimates for constrained and unconstrained trials were averaged for each participant. The difference in source power between the constrained and unconstrained conditions was evaluated on the group level using the same non-parametric cluster-based permutation test as for the sensor level data mentioned above. A dependent-sample *t* test thresholded at an alpha level of 0.05 served to identify the biggest cluster of neighboring grid points showing a difference between the two conditions. The Monte Carlo p-value of this cluster was calculated as the amount out of 5,000 random permutations that yielded a more extreme effect than the observed one evaluated at an alpha level of 0.05 (two-tailed).

To gain a similar measure of the effect size at both time points, we also calculated Cohen's *d* (Cohen, 1988) over the whole brain for each session. This was done for each point in the grid by taking the alpha–beta power differences between the two experimental conditions and calculating the standard deviation over all participants of these non-standardized effect sizes. Then, Cohen's *d* maps were computed by dividing the grand average of the non-standardized effect sizes by this standard deviation obtained from the previous step. This yielded an effect size measure for each grid point.

To investigate the robustness of our effect for an MEG experiment of shorter duration, the first 112 trials (out of 224 trials in total) were selected from each session. This served as a representation of half a session and was analyzed in the same way as the full sessions specified above. Importantly, the trial selection for half a session was performed only after the preprocessing step, meaning that incorrect, noisy, and blinking trials were discarded previously. Thus, every representation of half a session consisted of 112 clean and correct trials for all participants, with an average of 55 unconstrained trials ($SD = 4$) and 57 constrained trials ($SD = 4$). To compare the effect size of both sessions, we also calculated Cohen's *d* for the half session, in the same way as for the full session described above.

3 | RESULTS

3.1 | Behavioral results

Table 1 shows an overview of participants' behavioral effect including mean picture-naming times as well as errors per condition and session. Figure 2 shows the mean picture-naming time of each participant per session for unconstrained and constrained contexts. The linear mixed-effects regression yielded a significant effect of condition, showing that participants named the pictures in constrained sentences

TABLE 1 Amount of errors per condition for each session (top) and mean reaction times (and standard deviation over participants' means, *SD*) for picture naming per condition for each session (bottom)

Session	1	2
Errors (%)		
Unconstrained	2.08	3.33
Constrained	2.44	2.14
Reaction Times (ms)		
Unconstrained	738 ($SD = 92$)	759 ($SD = 84$)
Constrained	595 ($SD = 89$)	583 ($SD = 129$)
Effect	143 ($SD = 83$)	177 ($SD = 72$)

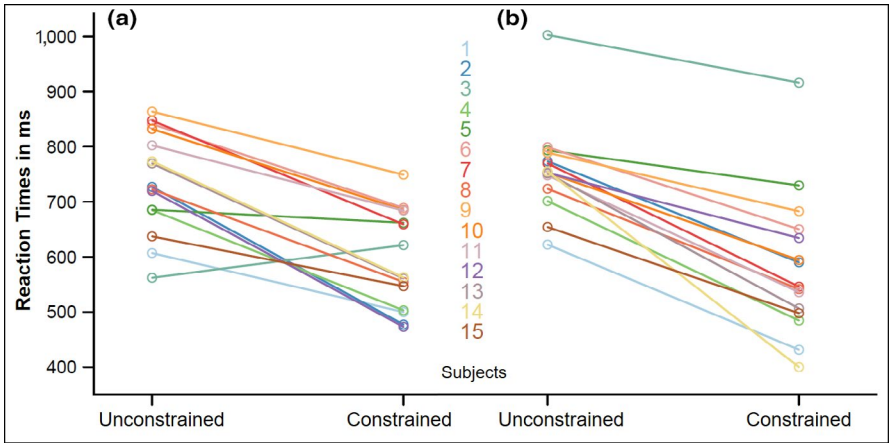


FIGURE 2 Mean reaction times for picture naming per condition in session 1 (a) and session 2 (b). Each line connects both conditions for one participant [Colour figure can be viewed at wileyonlinelibrary.com]

faster than in unconstrained sentences (Estimate = −0.14, *SE* = 0.02, *t* = −7.61, *p* < .001). Further, the model yielded a significant effect of session (Estimate = 0.02, *SE* = 0.01, *t* = 3.39, *p* < .001) and an interaction effect of condition and session (Estimate = −0.03, *SE* = 0.01, *t* = −3.40, *p* < .001). When modeling the linear mixed-effects regression separately per session, we obtained strong effects of condition for both session 1 (Estimate = −0.15, *SE* = 0.02, *t* = −6.60, *p* < .001, Cohen's *d* = 1.58) and session 2 (Estimate = −0.18, *SE* = 0.02, *t* = −9.28, *p* < .001, Cohen's *d* = 1.62). A correlation analysis of the behavioral effect across sessions yielded a Pearson's correlation coefficient of r = 0.58, which is a large effect size. More importantly, all but two individuals in session 1 and all individuals in session 2 showed behavioral facilitation from constrained contexts, indicating that there is high test–retest reliability of the facilitation effect, regardless of its exact magnitude in milliseconds.

3.2 | MEG results

3.2.1 | Sensor level

Figure 3 shows the group-level statistical results of the time–frequency analysis for both sessions. The two left plots of the figure show the time-resolved *t*-values over the 800-ms time window of interest for frequencies from 2 to 30 Hz for selected sensors. The two topography plots show the scalp distribution of the effect in *t*-values, with the selected sensors for the lower plots marked in white. The statistical results of the non-parametric cluster-based permutation test on the differences between conditions yielded significant clusters for session 1 (*p* = .016) and session 2 (*p* = .006).

Figure 4 shows the group-level statistics of the correlation analysis between the neural time–frequency data and the picture-naming times. As in Figure 3, the left plots show the time-resolved *t*-values over time and frequency for selected sensors. The topography plots show the distribution of the correlation values over the scalp, and the selected sensors marked in white. This analysis yielded significant positive correlations for session 1 (*p* = .002) and session 2 (*p* = .01). These results suggest that slower picture naming is associated with a smaller degree of alpha–beta power decreases during the analyzed 800-ms time interval before picture appearance. This finding provides further supports to the claim that the alpha–beta power decreases reflect conceptual and lexical processes during word planning (Piai et al., 2015, 2020; Piai & Zheng, 2019). The more conceptual and lexical planning takes place in this pre-picture interval, indexed by a larger degree of alpha–beta power decreases, the faster people can name the picture once it is presented.

3.2.2 | Source level

The top row of Figure 5 shows the group-level results for the source localization of the power differences for both sessions on the brain surface as well as their overlap, masked by the statistically significant clusters. The top row of Figure 6 depicts the obtained clusters and their spatial distribution in axial slices. However, caution should be taken when interpreting these slices, as the spatial resolution as well as sensitivity of MEG for deeper sources is reduced (Hillebrand et al., 2016). In both figures, color scales show the relative power change (in percentage) in session 1 and session 2 (i.e., the difference between power in the constrained and the unconstrained condition divided by their averaged power). The color scale only applies to the power difference maps for both sessions (Figure 5a,b,d,e and Figure 6a,b,d,e). For the overlap maps (Figure 5c,f and Figure 6c,f), only data points associated with the significant clusters in both sessions are shown.

The power decreases for constrained relative to unconstrained sentences in the alpha–beta frequency range from 10 to 20 Hz were statistically significant for both session 1 (*p* = .0092) and session 2 (*p* = .0052). The clusters of power changes were exclusively lateralized to the left hemisphere

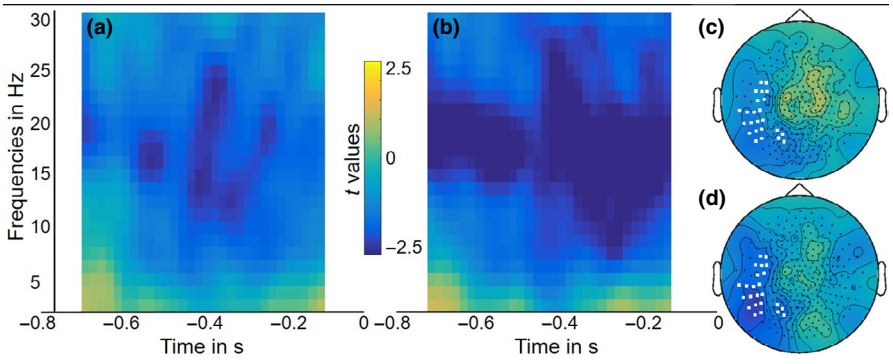


FIGURE 3 Group-level time-resolved frequency representations (standardized as *t*-values) averaged over selected sensors (marked as white in c and d) for (a) session 1 and (b) session 2. The scalp topographies show the averaged *t*-values over the 800-ms time window of interest from 2 to 30 Hz for (c) session 1 and (d) session 2 [Colour figure can be viewed at wileyonlinelibrary.com]

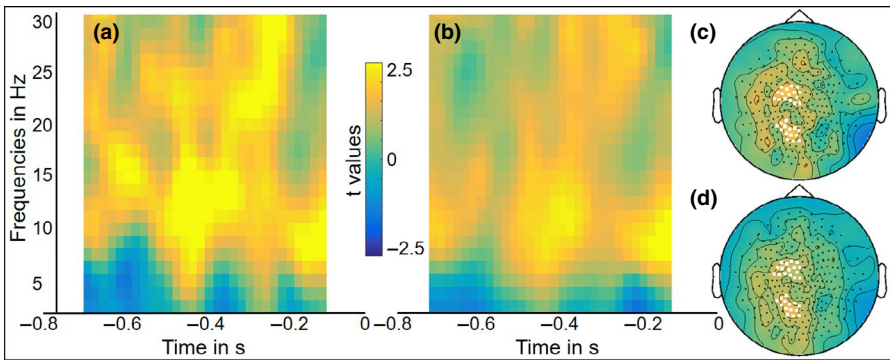


FIGURE 4 Group-level time-resolved frequency representations of the correlation between sensor-level data and behavioral reaction times for picture naming (standardized as *t*-values) averaged over selected sensors (marked as white in c and d) for (a) session 1 and (b) session 2. The scalp topographies show the averaged *t*-values over the 800-ms time window of interest from 2 to 30 Hz for (c) session 1 and (d) session 2 [Colour figure can be viewed at wileyonlinelibrary.com]

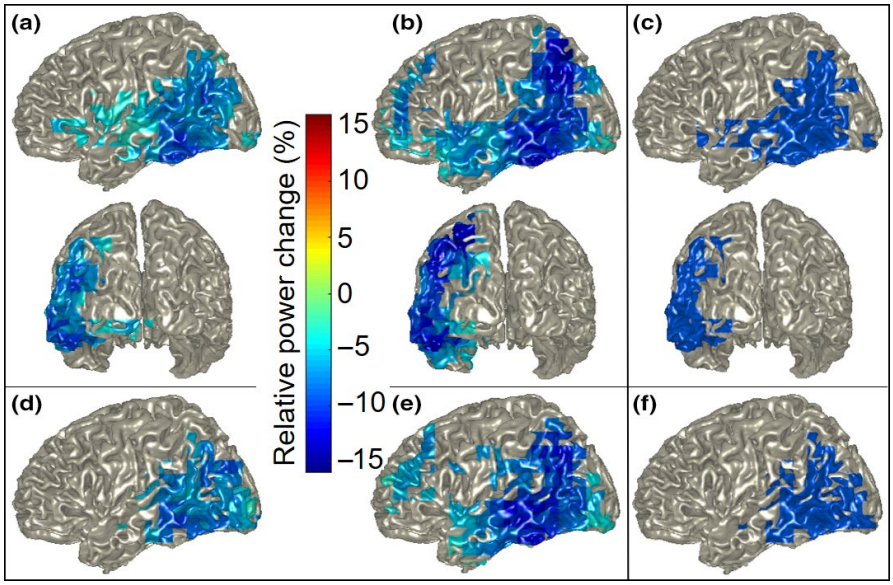


FIGURE 5 Group-level source localization power decreases for constrained relative to unconstrained sentences at 10 – 20 Hz for (a) full session 1 and (b) full session 2 in left and posterior views. Power decreases for (d) first half of session 1 and (e) first half of session 2 in the left hemisphere. Depicted areas are masked by statistically significant clusters. Overlap map of group-level source localization showing areas of power changes common to both sessions for (c) the full sessions and (f) the first half of both sessions. Note: color scales do not apply to overlap maps [Colour figure can be viewed at wileyonlinelibrary.com]

in both sessions. In session 1, the strongest power decreases around 10% were observed in the posterior part of the brain, as shown in Figure 5a. More precisely, they extended from the inferior to the superior temporal gyrus and then posteriorly to the inferior parietal lobule. Anteriorly, a weaker decrease around 5% extended into the inferior frontal gyrus. In session 2, the strongest power decreases were obtained around 15% and extended over the posterior temporal lobe and further over the inferior parietal lobule up to the superior parietal lobule, as shown in Figure 5b. Anteriorly, a weaker decrease of 5% – 10% extended over the anterior temporal lobe and prefrontal cortex. These effects remained unchanged when only participants below the age of 35 were included ($n = 12$), and the power decreases for constrained over unconstrained sentences did not correlate with age.

The top row of Figure 7 depicts the group-level Cohen's *d* effect size maps on the brain surface for the left and right hemisphere per session, resulting from the Cohen's *d* calculation per grid point. Here, color scales show the values of Cohen's *d*. In line with the source localization maps in Figure 5, the Cohen's *d* maps in Figure 7 show a more widespread effect for session 2 (b) than for session 1 (a), mostly lateralized to the left hemisphere in both sessions.

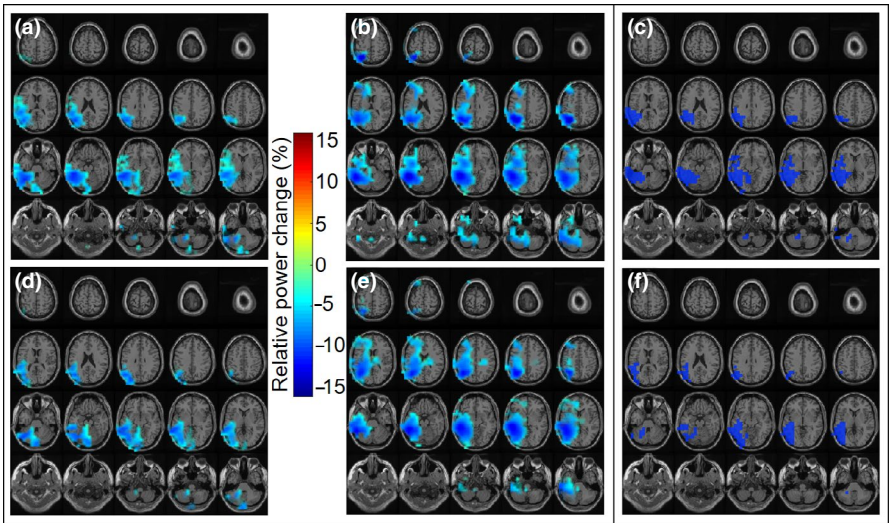
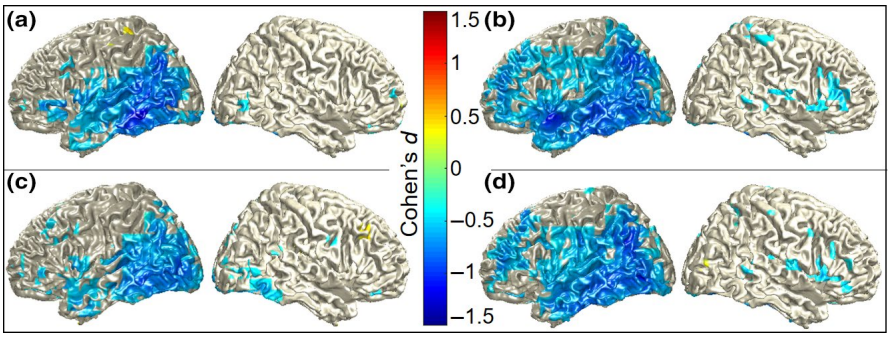


FIGURE 6 Group-level source localization power decreases between conditions at 10 – 20 Hz for (a) full session 1, (b) full session 2, (d) first half of session 1, and (e) first half of session 2 in axial brain slices. Depicted areas are masked by statistically significant clusters. Overlap map of group-level source localization showing areas of power changes common to both sessions for (c) the full sessions and (f) the first half of both sessions. Note: color scales do not apply to overlap maps [Colour figure can be viewed at wileyonlinelibrary.com]

FIGURE 7 Group-level Cohen's *d* effect size measures of alpha–beta power differences per grid point for (a) full session 1, (b) full session 2, (c) half session 1, and (d) half session 2 in the left and right hemisphere. Color scales show Cohen's *d*-values, thresholded at 0.5 [Colour figure can be viewed at wileyonlinelibrary.com]



3.3 | MEG across-session consistency

To depict the overlap of brain areas exhibiting power changes in session 1 and 2, a new mask was created based on the significant clusters resulting from the source localization for both sessions. This overlap map thus only shows those areas that were significant in session 1 and session 2 and is shown in Figures 5c and 6c. Note that the color scales do not apply to the overlap maps. In line with the strongest power decreases in session 1 and 2, the area of overlap extends over the posterior temporal lobe to the inferior parietal lobule and anteriorly over the superior temporal gyrus slightly into the inferior frontal gyrus.

3.4 | MEG first half effect size

The bottom rows of Figures 5 and 6 show the results for the source localization of the power changes for the first half of both sessions. Here, the decreases in power between conditions were also significant for both session 1 ($p = .0036$) and session 2 ($p = .0024$). Compared to the full session, the power decreases of the first half of session 1 are restricted to the posterior temporal and inferior parietal lobules, not extending into the inferior frontal gyrus, as shown in Figure 5d. The power decreases in the first half of session 2 are quite consistent with the full session. They extend from the posterior temporal into the inferior parietal lobule and anteriorly over the anterior temporal lobe to the middle frontal gyrus, as shown in Figure 5e.

Likewise, the overlap of the source localization results for the first half of session 1 and the first half of session 2 is depicted by means of the same masking procedure as described above and shown in Figure 5f. This area of overlap extends over the posterior temporal lobe into the inferior parietal lobule, presenting a similar overlap as for the full sessions, but less anterior and continuous.

The bottom row of Figure 7 depicts the group-level Cohen's *d* map of the effect size per grid point for both hemispheres resulting from the half session analysis. When

comparing the Cohen's *d* maps for the first half of session 1 (c) and session 2 (d) to the Cohen's *d* maps of the full sessions 1 (a) and 2 (b), it is clear that the effect sizes and their spatial distributions are similar for the half and full sessions.

4 | DISCUSSION

The present study aimed to determine the across-session consistency of context-driven word production and the spectro-temporo-spatial robustness of this context effect for short testing sessions, with the ultimate goal of informing the development of a suitable approach to track the neuroplasticity of language in neurological patients using electrophysiology. Therefore, healthy participants were tested with MEG while performing a context-driven picture-naming task, similar to sentence completion (Wilson et al., 2017). Replicating previous findings (Klaus et al., 2020; Piai et al., 2014, 2015, 2017, 2018, 2020) in both sessions, the behavioral differences demonstrated faster picture naming for constrained than unconstrained sentences, indicating that the context of constrained sentences provides the necessary information for enabling conceptual and lexical retrieval already prior to the target picture being shown. Also the MEG source level results were in accordance with these findings and revealed the expected power decreases in the alpha–beta frequency range in the left hemisphere. This was even further supported by the trial-by-trial relationship between the MEG sensor level results and the picture-naming times, revealing a positive correlation between the neural and behavioral data for constrained sentences. These results suggest that slower picture naming goes along with a lesser extent of alpha–beta power decreases, which reinforces the link between these alpha–beta power decreases and conceptual and lexical processes of word planning (Piai et al., 2015, 2020; Piai & Zheng, 2019).

The source localization maps of session 1 and session 2 showed a high degree of overlap. The brain areas that exhibited alpha–beta power decreases in both sessions were the posterior temporal lobe and the inferior parietal cortex of the left hemisphere. These were also the areas that showed the strongest power decreases in each session. Even when only analyzing the first half of both MEG sessions, the core areas of strongest power decreases are the same as those for the full sessions, just not as spatially spread out. Also the Cohen's *d* effect size maps are in line with this, showing the effect spreading over these core areas, for the full and the half sessions. This robustness suggests that the left temporal and inferior parietal lobules are the most consistent brain areas for the retrieval of concepts and words guided by contextually constrained sentences. These findings are in line with previous literature showing the involvement of these areas in the lexical-semantic aspects of word production (Binder, Desai, Graves, & Conant, 2009; Griffis, Nenert, Allendorfer, & Szaflarski, 2017; Piai et al., 2018; Price, 2012; Roelofs, 2014).

The left posterior temporal and inferior parietal regions were commonly activated across sessions. The inferior temporal gyrus has previously been shown to be a crucial area for accessing lexical-semantic concepts in object recognition and word production (Price, 2012; Roelofs, 2014). Additionally, the inferior parietal lobule has been argued to be essential in predicting and integrating semantic knowledge (Binder et al., 2009; Price, 2012). This further corroborates the findings of Piai et al. (2018) that established a causal link between the posterior temporal and inferior parietal lobules and context-driven word retrieval. In that study, left-hemisphere stroke patients performed the same task as employed in the present study. Patients whose left middle and superior temporal gyri and inferior parietal lobule were damaged by a stroke did not show a behavioral context effect. In addition, their EEG revealed a reduced alpha–beta context effect. The results of the present study support the claim that these brain regions are critical for context-dependent word retrieval.

Interestingly, the reliability of the clusters at the source level obtained with the Monte Carlo simulations increased from full to half session analysis for both sessions. In other words, the likelihood of finding a more extreme result than the observed one by chance was reduced. This suggests that the quality of the acquired data decreases with the duration of the session, probably partly due to movement and fatigue of the participants. This is further supported by the Cohen's *d* effect size maps for the full and half sessions. These are very similar for the different session lengths (cf., Figure 7a–c,b–d), showing that halving the duration of the session does not entail halving the effect size. Particularly, this shows that it is indeed possible to capture an effect nearly as strong as that of the full session with only half the testing duration. In general, for MEG as a method, this suggests that acquiring more data is not always better and does not necessarily increase the significance or size of the investigated effect.

The alpha–beta power decreases we found have previously been argued to reflect memory-related processes as well as motor preparation for word production, especially in the beta frequency range (Piai et al., 2015). However, the analysis of the present data only yielded left-lateralized significant clusters of alpha–beta decreases, whereas motor preparation for speaking activates speech motor areas in both hemispheres (Saarinen, Laaksonen, Parviainen, & Salmelin, 2006; Salmelin & Sams, 2002). This suggests that the motor activity in the right hemisphere was not strong enough to be captured in the present study. Further, none of the obtained clusters in the left hemisphere clearly involved left speech motor areas, suggesting that the captured activity most likely did not reflect any motor preparation. Thus, the present findings indicate that the power decreases between 10 and 20 Hz in the pre-picture interval mostly reflect the context-driven

retrieval of concept and lexical information. Additionally, when focusing on the neuroplasticity of language, motor activity is not a major aspect of interest. In that case, a paradigm as the present one, which only elicits left-lateralized effects in neurotypical control subjects, seems to be a suitable option to investigate the neuroplasticity of language production with electrophysiology.

When comparing the results from both sessions, session 2 showed a stronger and spatially more distributed alpha–beta power decrease than session 1. In line with that are also the Cohen's *d* effect size maps that showed a more spatially distributed effect size for session 2 than for session 1. Especially the involvement of the middle frontal gyrus in session 2 is surprising, as there was no such frontal activity obtained in session 1. As previously found, frontal MEG power for picture naming can vary between participants (Liljeström et al., 2009), such that it is not a consistent feature across the group. Alternatively, this could represent a familiarization effect with the whole experimental procedure on the side of the participants, as session 2 was an analog replication of session 1 and only differed in the stimulus materials. This could be examined in future studies by including a third session of the same paradigm, to determine the activity captured at another time point following session 2 and help to further investigate this potential familiarization effect.

As depicted in models of word retrieval, the process of word production undergoes several stages before the word can be produced (Roelofs, 1992, 2014). In the present study, the target concept is either primed by the sentence context (constrained trials) or represented in the picture (unconstrained trials). When the concept is accessed, further information about the word and its phonology are retrieved. Only then, at the later stages of word retrieval, the preparation for articulation takes place. For the findings of the present study, this indicates that the large alpha–beta power decreases detected with MEG are mostly reflecting computations underlying the early stages of conceptual and lexical retrieval, which happen at a millisecond time scale. These findings are of theoretical importance as they inform neuroanatomical models of spoken-word production.

5 | CONCLUSION

The present study aimed at determining whether context-driven word production, measured in combination with MEG, shows enough spectro-temporal and spatial across-session consistency for being an appropriate method to track neuroplasticity after brain damage. The findings of this study convincingly show that this is a suitable approach for this purpose. The left-lateralized power decreases in the temporal and inferior parietal lobule were observed similarly in both sessions, as well as in the half session analyses. This further provides the opportunity to keep the duration of such testing sessions for neurological populations short. With these characteristics, MEG proofs to be a suitable technique to longitudinally track the reorganization of language functions after brain damage.

ACKNOWLEDGEMENTS

This research was supported by Gravitation Grant 024.001.006 of the Language in Interaction Consortium from Netherlands Organization for Scientific Research.

CONFLICT OF INTEREST

The authors declare that there are no conflicts of interest.

AUTHOR CONTRIBUTIONS

Both authors conceived and planned the study. NR performed the experiment and analyzed the data. NR and VP interpreted the data and revised the manuscript. NR wrote the original manuscript, and VP provided feedback.

DATA AVAILABILITY STATEMENT

The data and code are made available via the Donders Repository (https://data.donders.ru.nl/).

ORCID

Natascha Marie Roos https://orcid.org/0000-0002-1645-9110
Vitória Piai https://orcid.org/0000-0002-4860-5952

REFERENCES

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex*, *19*(12), 2767–2796. https://doi.org/10.1093/cercor/bhp055

Boersma, P., & Weenink, D. (2017). *Praat, software for speech analysis and synthesis*. Amsterdam, The Netherlands: University of Amsterdam.

Bradshaw, A. R., Thompson, P. A., Wilson, A. C., Bishop, D. V. M., & Woodhead, Z. V. J. (2017). Measuring language lateralisation with different language tasks: A systematic review. *PeerJ*, *5*, e3929. https://doi.org/10.7717/peerj.3929

Brodeur, M. B., Dionne-Dostie, E., Montreuil, T., & Lepage, M. (2010). The bank of standardized stimuli (BOSS), a new set of 480 normative photos of objects to be used as visual stimuli in cognitive research. *PLoS One*, *5*(5), e10773. https://doi.org/10.1371/journal.pone.0010773

Brownsett, S. L. E., Warren, J. E., Geranmayeh, F., Woodhead, Z., Leech, R., & Wise, R. J. S. (2014). Cognitive control and its impact on recovery from aphasic stroke. *Brain*, *137*(1), 242–254. https://doi.org/10.1093/brain/awt289

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (pp. 2). Hillsdale, NJ: L. Lawrence Earlbaum Associates.

Duffau, H. (2005). Lessons from brain mapping in surgery for low-grade glioma: Insights into associations between tumour and brain plasticity. *The Lancet Neurology*, *4*(8), 476–486. https://doi.org/10.1016/S1474-4422(05)70140-X

Eaton, K. P., Szaflarski, J. P., Altaye, M., Ball, A. L., Kissela, B. M., Banks, C., & Holland, S. K. (2008). Reliability of fMRI for studies of language in post-stroke aphasia subjects. *NeuroImage*, *41*(2), 311–322. https://doi.org/10.1016/j.neuroimage.2008.02.033

Ekstrom, A. (2010). How and when the fMRI BOLD signal relates to underlying neural activity: The danger in dissociation. *Brain Research Reviews*, *62*(2), 233–244. https://doi.org/10.1016/j.brainresrev.2009.12.004

Findlay, A. M., Ambrose, J. B., Cahn-Weiner, D. A., Houde, J. F., Honma, S., Hinkley, L. B. N., … Kirsch, H. E. (2012). Dynamics of hemispheric dominance for language assessed by magnetoencephalographic imaging. *Annals of Neurology*, *71*(5), 668–686. https://doi.org/10.1002/ana.23530

Geranmayeh, F., Wise, R. J. S., Leech, R., & Murphy, K. (2015). Measuring vascular reactivity with breath-holds after stroke: A method to aid interpretation of group-level BOLD signal changes in longitudinal fMRI studies. *Human Brain Mapping*, *36*(5), 1755–1771. https://doi.org/10.1002/hbm.22735

Griffis, J. C., Nenert, R., Allendorfer, J. B., & Szaflarski, J. P. (2017). Linking left hemispheric tissue preservation to fMRI language task activation in chronic stroke patients. *Cortex*, *96*, 1–18. https://doi.org/10.1016/j.cortex.2017.08.031

Gross, J., Kujala, J., Hämäläinen, M., Timmermann, L., Schnitzler, A., & Salmelin, R. (2001). Dynamic imaging of coherent sources: Studying neural interactions in the human brain. *Proceedings of the National Academy of Sciences*, *98*(2), 694–699. https://doi.org/10.1073/pnas.98.2.694

Hillebrand, A., Nissen, I. A., Ris-Hilgersom, I., Sijsma, N. C. G., Ronner, H. E., van Dijk, B. W., & Stam, C. J. (2016). Detecting epileptiform activity from deeper brain regions in spatially filtered MEG data. *Clinical Neurophysiology*, *127*(8), 2766–2769. https://doi.org/10.1016/j.clinph.2016.05.272

Kim, M. J. J., Holodny, A. I., Hou, B. L., Peck, K. K., Moskowitz, C. S., Bogomolny, D. L., & Gutin, P. H. (2005). The effect of prior surgery on blood oxygen level-dependent functional MR imaging in the preoperative assessment of brain tumors. *American Journal of Neuroradiology*, *26*(8), 1980–1985. http://www.ajnr.org/content/26/8/1980

Kiran, S. (2012). What is the nature of poststroke language recovery and reorganization? ISRN Neurology, 2012.

Klaus, J., Schutter, D. J. L. G., & Piai, V. (2020). Transient perturbation of the left temporal cortex evokes plasticity-related reconfiguration of the lexical network. *Human Brain Mapping*, *41*(4), 1061–1071. https://doi.org/10.1002/hbm.24860

Kujala, J., Sudre, G., Vartiainen, J., Liljeström, M., Mitchell, T., & Salmelin, R. (2014). Multivariate analysis of correlation between electrophysiological and hemodynamic responses during cognitive processing. *NeuroImage*, *92*, 207–216. https://doi.org/10.1016/j.neuroimage.2014.01.057

Liljeström, M., Hultén, A., Parkkonen, L., & Salmelin, R. (2009). Comparing MEG and fMRI views to naming actions and objects. *Human Brain Mapping*, *30*(6), 1845–1856. https://doi.org/10.1002/hbm.20785

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

Meltzer, J. A., Postman-Caucheteux, W. A., McArdle, J. J., & Braun, A. R. (2009). Strategies for longitudinal neuroimaging studies of overt language production. *NeuroImage*, *47*(2), 745–755. https://doi.org/10.1016/j.neuroimage.2009.04.089

Nolte, G. (2003). The magnetic lead field theorem in the quasi-static approximation and its use for magnetoencephalography forward calculation in realistic volume conductors. *Physics in Medicine and Biology*, *48*(22), 3637–3652. https://doi.org/10.1088/0031-9155/48/22/002

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, *2011*, 1–9. https://doi.org/10.1155/2011/156869

Piai, V., De Witte, E., Sierpowska, J., Zheng, X., Hinkley, L. B., Mizuiri, D., … Nagarajan, S. S. (n.d.). *Language neuroplasticity in brain tumour patients revealed bymagnetoencephalography*. In Press. https://doi.org/10.1162/jocn_a_01561

Piai, V., Klaus, J., & Rossetto, E. (2020). The lexical nature of alpha-beta oscillations in context-driven word production. *Journal of Neurolinguistics*, *55*, 100905. https://doi.org/10.1016/j.jneuroling.2020.100905

Piai, V., Meyer, L., Dronkers, N. F., & Knight, R. T. (2017). Neuroplasticity of language in left-hemisphere stroke: Evidence linking subsecond electrophysiology and structural connections. *Human Brain Mapping*, *38*(6), 3151–3162. https://doi.org/10.1002/hbm.23581

Piai, V., Roelofs, A., & Maris, E. (2014). Oscillatory brain responses in spoken word production reflect lexical frequency and sentential constraint. *Neuropsychologia*, *53*, 146–156. https://doi.org/10.1016/j.neuropsychologia.2013.11.014

Piai, V., Roelofs, A., Rommers, J., & Maris, E. (2015). Beta oscillations reflect memory and motor aspects of spoken word production. *Human Brain Mapping*, *36*(7), 2767–2780. https://doi.org/10.1002/hbm.22806

Piai, V., Rommers, J., & Knight, R. T. (2018). Lesion evidence for a critical role of left posterior but not frontal areas in alpha–beta power decreases during context-driven word production. *European Journal of Neuroscience*, *48*(7), 2622–2629. https://doi.org/10.1111/ejn.13695

Piai, V., & Zheng, X. (2019). Chapter Eight - Speaking waves: Neuronal oscillations in language production. In K. D. Federmeier (Ed.), *Psychology of learning and motivation* (Vol. 71, pp. 265–302). Academic Press. https://doi.org/10.1016/bs.plm.2019.07.002

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

R Core Team (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Core Team.

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, *42*(1), 107–142. https://doi.org/10.1016/0010-0277(92)90041-F

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

Saarinen, T., Laaksonen, H., Parviainen, T., & Salmelin, R. (2006). Motor cortex dynamics in visuomotor production of speech and non-speech mouth movements. *Cerebral Cortex*, *16*(2), 212–222. https://doi.org/10.1093/cercor/bhi099

Salmelin, R., & Sams, M. (2002). Motor cortex involvement during verbal versus non-verbal lip and tongue movements. *Human Brain Mapping*, *16*(2), 81–91. https://doi.org/10.1002/hbm.10031

Skipper-Kallal, L. M., Lacey, E. H., Xing, S., & Turkeltaub, P. E. (2017). Right hemisphere remapping of naming functions depends

on lesion size and location in poststroke aphasia. *Neural Plasticity*, *2017*, 1–17. https://doi.org/10.1155/2017/8740353.

Stolk, A., Todorovic, A., Schoffelen, J.-M., & Oostenveld, R. (2013). Online and offline tools for head movement compensation in MEG. *NeuroImage*, *68*, 39–48. https://doi.org/10.1016/j.neuroimage.2012.11.047

Traut, T., Sardesh, N., Bulubas, L., Findlay, A., Honma, S. M., Mizuiri, D., … Tarapore, P. E. (2019). MEG imaging of recurrent gliomas reveals functional plasticity of hemispheric language specialization. *Human Brain Mapping*, *40*(4), 1082–1092. https://doi.org/10.1002/hbm.24430

Turkeltaub, P. E., Messing, S., Norise, C., & Hamilton, R. H. (2011). Are networks for residual language function and recovery consistent across aphasic patients? *Neurology*, *76*(20), 1726–1734. https://doi.org/10.1212/WNL.0b013e31821a44c1

van Casteren, M., & Davis, M. H. (2006). Mix, a program for pseudorandomization. *Behavior Research Methods*, *38*(4), 584–589. https://doi.org/10.3758/BF03193889

Vartiainen, J., Liljeström, M., Koskinen, M., Renvall, H., & Salmelin, R. (2011). Functional magnetic resonance imaging blood oxygenation level-dependent signal and magnetoencephalography evoked responses yield different neural functionality in reading. *Journal of Neuroscience*, *31*(3), 1048–1058. https://doi.org/10.1523/JNEUROSCI.3113-10.2011

Veldsman, M., Cumming, T., & Brodtmann, A. (2015). Beyond BOLD: Optimizing functional imaging in stroke populations. *Human Brain Mapping*, *36*(4), 1620–1636. https://doi.org/10.1002/hbm.22711

Wilson, S. M., Bautista, A., Yen, M., Lauderdale, S., & Eriksson, D. K. (2017). Validity and reliability of four language mapping paradigms. *NeuroImage: Clinical*, *16*, 399–408. https://doi.org/10.1016/j.nicl.2016.03.015

Wilson, S. M., Lam, D., Babiak, M. C., Perry, D. W., Shih, T., Hess, C. P., … Chang, E. F. (2015). Transient aphasias after left hemisphere resective surgery. *Journal of Neurosurgery*, *123*(3), 581–593. https://doi.org/10.3171/2015.4.JNS141962

How to cite this article: Roos N. M., Piai V. Across-session consistency of context-driven language processing: A magnetoencephalography study. *Eur J Neurosci*. 2020;52:3457–3469. https://doi.org/10.1111/ejn.14785