Across-Session Consistency of Context-Dependent Language Processing: Towards a Clinical Tool

Natascha Roos[1]
Supervisors: Vitória Piai[1, 2]

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

Transient aphasias after tumor removal from the left hemisphere are commonly exhibited by patients after the surgery. In most cases, the deficits resolve within weeks. The present study serves to set the parameters for a clinical tool to track this process of language recovery over time. The main aims were to determine a suitable imaging method, the corresponding across-session consistency, and the effect size for a shorter testing duration. For this purpose, 30 native Dutch speakers were tested with magnetoencephalography (MEG) or functional magnetic resonance imaging (fMRI) while performing the same picture naming as sentence completion task (15 participants per method). Sentences were either constrained or unconstrained towards the picture, such that participants could retrieve the target word through sentence context (constrained sentences) or had to wait for the picture to appear (unconstrained sentences) to be able to name it. Behavioral results show a strong reaction time effect for picture naming in the MEG as well as the fMRI experiment, verifying that constrained sentence context primes the target word before the picture is shown. The MEG results reveal alpha-beta power decreases (10 - 20 Hz) in the left temporal and inferior parietal lobe that yield a Dice coefficient quantifying the across-session consistency of activation of 0.49. Analyzing only the first half of all MEG sessions reduces the area of the alpha-beta power decreases and lowers the Dice coefficient to 0.35 but increases the significance of the power decrease cluster. The fMRI results reveal BOLD signal increases for constrained over unconstrained sentences mostly in the left inferior temporal and parietal lobe but also bilaterally in motor areas, resulting in a Dice coefficient of 0.43. Analysis of the first half of the fMRI sessions diminishes the obtained BOLD increase clusters and lowers the Dice coefficient to 0.31. Based on spatiality, consistency, and significance of the obtained effect profiles with each method, the findings of the present study lead to conclude that MEG is a more suitable imaging method for the clinical tool than fMRI.

Keywords: MEG, fMRI, tumor, aphasia

Corresponding author: Natascha Roos; **E-mail:** natascha.roos1693@gmail.com

The present study is the first step of a project aiming to develop a clinical tool to track language-related changes of brain activity in patients. More precisely, the broader aim is to scan patients with tumors in language areas of the brain before as well as after they undergo tumor surgery to investigate the recovery process of language in the brain. Many of these patients show deficits in language use shortly after the surgery. However, these deficits are mostly temporary and can resolve within the following weeks. To capture and better understand this recovery process of the brain, the first step of the project is a study with healthy control participants to assess the optimal parameters for patient testing. This is what constitutes the present study and will be referred to as such in the following report. The longitudinal patient study will be a future step within the project, after the clinical tool is finalized and ready for patient testing.

Patients with tumors in language areas in the left hemisphere of the brain often suffer from aphasias after tumor removal. The term aphasia describes language deficits due to brain damage in one or more language domains such as speaking, comprehending, reading, or writing (see Dronkers & Baldo, 2010, for a review of different types of aphasia). However, these patients' language deficits due to resection surgery are mainly temporary and can resolve within the following weeks. For example, 110 patients performed language tests before undergoing surgery as well as two to three days and one month after (Wilson et al., 2015). Most patients showed normal language scores before surgery and a major decrease in performance two to three days after the surgery. But these decreases mostly resolved within one month such that the language scores of the second postsurgical test did not significantly differ from presurgical test scores. Further, this study also shows that the location where the tumor was resected determines the language domains that are affected, and thus the type of aphasia that is observed.

A meta-analysis of hemodynamic studies conducting language tasks with stroke patients suffering from aphasias concluded that the activated brain areas in patients are coherent across studies, including task-related activity in the right-hemisphere (Turkeltaub, Messing, Norise & Hamilton, 2011). Further, another study with aphasic stroke patients employing functional magnetic resonance imaging (fMRI) has shown that the amount of right-hemisphere activity for language tasks depends on the size of left-hemisphere lesions and influences language recovery (Skipper-Kallal, Lacey, Xing & Turkeltaub,2017). This suggests that the brain can compensate by employing the right hemisphere for language functions if the left hemisphere is lesioned.

To address this neuroplasticity of language with electrophysiological measures, Piai, Meyer, Dronkers and Knight (2017a) conducted a combined electroencephalography (EEG), behavioral, and structural connectivity study with patients that had suffered a stroke in the left hemisphere. In previous electrophysiological studies, employing a sentence completion task with neurotypical control subjects, the authors observed power decreases of brain activity in the alpha-beta frequency range (8-30 Hz) for context-driven word retrieval (Piai, Roelofs & Maris, 2014; Piai, Roelofs, Rommers & Maris, 2015). Interestingly, when conducting the same task with left-hemisphere stroke patients they discovered the same power decrease in the alpha-beta range as in neurotypical control subjects, but lateralized to the right instead of the left hemisphere in patients (Piai et al., 2017a; Piai, Rommers & Knight, 2017b).

These findings indicate that left and right hemispheres can perform similar neuronal computations and that patients with lesions in the left hemisphere can supportively draw on their intact right hemisphere for language use. As such, the brain can uniquely reorganize itself for language use in recovery. This stresses the individuality of the functional organization of the brain, which has been argued to be an important aspect for further investigation to enhance surgery outcomes and patient therapy (Duffau, 2005).

In accordance with this argumentation, a recent review shows that different language tasks have resulted in diverging findings in terms of the functional organization of language in the brain, suggesting that it might be diversely independent of the language domains and processes that are involved (Bradshaw, Thompson, Wilson, Bishop & Woodhead, 2017). This divergence can have either individual or task-dependent impacts, but might also merely derive from differences in methodology across studies. Thus, the authors of the review call for an increase in methodological consistency to increase the comparability of results from different studies. This would also help to gain further insights into the independent organization of language processes in the brain.

Following this line of reasoning, Wilson, Bautista, Yen, Lauderdale and Eriksson (2017) investigated the validity and reliability of different language production and comprehension paradigms to identify language areas with fMRI. 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. They conclude that sentence completion tasks provide the best-balanced combination of validity and reliability. However, they also point out general limitations of language mapping with fMRI in individuals for clinics and research, and prompt for equal assessments and comparisons of different paradigms.

To date, there have been several findings demonstrating transient aphasias after left-hemisphere damage (Skipper-Kallal et al., 2017; Wilson et al., 2015), with beneficial right hemisphere compensation in recovery (Piai et al., 2017a, 2017b; Turkeltaub et al., 2011). Additionally, many authors agree on the fact that more coherent methodology across studies would yield better comparable findings and help to obtain further insights to the functional organization of language in the brain (Bradshaw et al., 2017; Duffau, 2005; Wilson et al., 2017). If one paradigm was revealed to take the lead over others in assessing the processes of language recovery in patients, it could be widely employed as a standard clinical tool. Accumulating all findings obtained with the same tool would commonly contribute to gain further insights into neurorehabilitation. Finally, this would help to improve the predictions of language recovery as well as patient care, enriching it with more individualized therapy.

The Present Study

To develop such a clinical tool requires a spatially defined and reliable effect that can be captured within a short testing duration. As a first step in this direction, the present study was conducted with healthy control participants to determine the optimal parameters for this tool. More precisely, this study approaches three main questions: What is the most suitable imaging method for the clinical tool, how spatially reliable is the brain activity of our paradigm over time, and how does the effect size change for half the testing duration?

In order to reveal the appropriate imaging method for patient testing, participants were scanned with MEG or fMRI. This serves as a more direct comparison of the effect profiles obtained per method, while participants perform the same experimental task. As outlined above, the common standard method to localize brain functions in tumor patients so far has been fMRI, but not necessarily because it is methodologically more suitable.

In hemodynamic methods such as fMRI, the obtained signal is a blood oxygen level dependent (BOLD) response. This is based on the framework that neurons that are active consume oxygen from the blood, causing a subsequent increase in blood flow called a hemodynamic response function (HRF). This HRF leads to a higher blood oxygen level in the local vessels which increases the signal intensity for fMRI and yields the BOLD signal that is measured. As such, fMRI measures neuronal activity only indirectly with a rather slow temporal resolution depending on the HRF, which can have a delay of about two seconds and last between 6 to 12 seconds before the signal decays.

In contrast, electrophysiological methods, such as MEG, yield a direct measure of brain activation based on the magnetic fields of neuronal activity. Here, the requirement is that neurons are activated in synchrony which initiates an electric current in the brain and induces a magnetic field around it. Contrary to fMRI, MEG captures neuronal activity with a high temporal resolution at the level of milliseconds and thus allows to track the time course of neuronal sources. But as MEG is only measured close to the scalp, the spatial resolution of fMRI is more precise, especially for subcortical structures. Nevertheless, MEG is successfully being employed to determine the dominant hemisphere for language, as it is done in patients before undergoing brain surgery (Findlay et al., 2012).

This shows that both methods are certainly possible to employ for a clinical tool, even though they measure different aspects of neuronal activation and have been shown to provide divergent findings (Kujala et al., 2014; Liljestrom, Hulten, Parkkonen & Salmelin, 2009; Vartiainen, Liljeström, Koskinen, Renvall & Salmelin, 2011). Additionally, keeping in mind that the patient study involves surgery in between testing sessions, it is noteworthy that prior brain surgery can affect and impair the BOLD signal of the whole hemisphere for fMRI (Kim et al., 2005). Hence, the present study questions fMRI as the clinical standard to localize brain functions by comparing it to MEG for the same task. Employing a more suitable imaging method for these purposes could possibly improve patient care and rehabilitation therapy in the future.

To track the brain activity of patients before and after the surgery as well as the subsequent language recovery, the patient study will be conducted longitudinally. Accordingly, we would expect to see changes from pre- to post-surgery and subsequent sessions. But to be able to argue that the changes in patients derive from the surgery, we need to know how spatially reliable and consistent the captured

effect really is. Since we do not expect any changes between sessions for healthy control subjects, the present study evaluates the across-session consistency of our paradigm. This is done by visualizing the areas of overlapping brain activity from session 1 and session 2. To quantify this overlap with a measure of overlapping brain activation across sessions, the Dice coefficient is calculated (Wilson et al., 2017).

Further, we evaluate the effect sizes obtained for half the testing duration by only analyzing the first half of each session. Especially for the patient study, the duration of testing sessions should be kept as short as possible. But since half a session equals shorter testing times, it also equals less acquired data. In other words, we want to shorten testing times without risking not capturing enough data for or sacrificing an effect. Therefore, we aim to delineate an amount of data that is necessary to obtain reliable and robust effects to serve the patient study design.

Regarding a spatially defined and reliable effect to employ for the clinical tool, we decided to use the same paradigm as in previous studies by Piai et al (2014, 2015, 2017a, 2017b). As outlined above, this context-dependent sentence completion paradigm has repeatedly elicited robust alpha-beta power decreases in control participants and patients. Herein, sentences are presented word-by-word on a screen in front of the participants. The last word of each sentence is the target word and presented as a picture. The task is to silently read the sentence and name the following picture. The sentences appear in two conditions such that the sentence context is either constrained or unconstrained towards the target word. This means that the sentence context either reveals information about the target word or not. To give an example, the picture for the target word *cow* was a photograph of a cow on white background. The corresponding constrained sentence was "*The farmer milked the [picture]* ", and the unconstrained sentence was "*The child drew a [picture]"*. Thus, the sentence context in constrained sentences enables participants to retrieve information about the target word and accordingly prepare to name the picture, already before it appears. Unconstrained sentences, however, do not give away information about the target word and participants must wait for the picture to appear until they can retrieve the information needed to name it.

This difference between the two conditions determines the time window of interest for the present study, which is the interval between the last word preceding the picture and the onset of the picture. All analyses focus on the differences in brain activity during this time window between the two conditions. More precisely, 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 the capturing of condition-specific differences that are of no interest. Further, participants only perform one task which eliminates the risk of capturing possible task switching demands. As such, the present paradigm proofs to be highly suitable to be employed in a clinical tool.

In line with earlier studies by Piai et al (2014, 2015, 2017a, 2017b) from which the paradigm as well as the stimulus materials were adopted, we hypothesize for the present study faster naming times for the pictures in constrained compared to unconstrained sentences. Further, the MEG brain activity profiles are expected to correspond to the spectral power decrease in the alpha-beta frequency range observed for the same paradigm as outlined above. That is, a decrease of alpha-beta power in the constrained relative to the unconstrained condition. Regarding fMRI, there has been evidence for a correlation between alpha-beta power decreases and BOLD signal increases for covert picture naming (Conner, Ellmore, Pieters, DiSano & Tandon, 2011). Additionally, in another picture naming study comparing MEG to fMRI that correlation between both measures was highest for frequencies in the alpha-beta range (Liljeström, Stevenson, Kujala & Salmelin, 2015). Based on these findings, we expect the BOLD signal to increase for constrained over unconstrained sentences.

Method

The present study falls under the blanket approval for standard studies of the accredited ethical reviewing committee, CMO Arnhem-Nijmegen, following the declaration of Helsinki (2013). It was conducted at the Donders Institute for Cognitive Neuroimaging in Nijmegen in the Netherlands.

Participants

A total of 31 native Dutch speakers aged between 18 and 50 years (*Mdn* = 22) participated in the study for monetary compensation or course credits. All participants were healthy and right-handed, with normal or corrected-to-normal vision (no glasses), and compatible for MEG and MRI (MEG

participants), or MRI only (fMRI participants). The dataset of one female fMRI participant was not considered for analysis due to a large amount of invalid trials and missing field maps for session 1. Thus, an additional participant was recruited so that both groups consisted of 15 participants. The MEG group included 7 females and ranged from 18 to 50 years (*Mdn* = 25), and the fMRI group included 11 females and ranged from 18 to 26 years (*Mdn* = 20).

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. Target words describing landscapes such as *forest*, or *mountain* were shown as full-screen photographs. 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 material was in Dutch and taken from previous studies (Piai et al., 2014, 2015, 2017a, 2017b). Pictures were collected from the BOSS database (Brodeur, Dionne-Dostie, Montreuil & Lepage, 2010) and via online search. The length of the target words 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 kept as similar as possible for both sentences associated with the same target word.

Design

The stimuli were presented in three main lists, uniquely divided in halves controlled for frequency, word length, and initial letter. 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 word, and a maximum of five repetitions of trials with the same condition. Participants were randomly assigned to one of the three main lists. Since the study consisted of a test and a re-test session scheduled between 13 and 28 days apart (*Mdn* = 20), each participant was presented with half of the target words per session, alternating the order of which half was presented first. Thus, one session consisted of 112 target words presented as pictures, once preceded by a constrained sentence and once by an unconstrained sentence, yielding 224 trials per session.

Procedure

Each session started with informing the participants about the task and the scanning session and clarifying possible questions. Then, participants signed the consent forms and were screened according to the employed scanning method. Before entering the scanner room, they were familiarized with the pictures of the experiment and the corresponding target words. These were presented in a slide show with four pictures on one slide and the target words printed below. Each session started with four practice trials in the scanner, 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, Inc.) and projected on a screen in front of the participants in the scanner. Figure 1 shows a trial overview for both sentences for the same target word with the experiment-specific presentation times 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 grey-scaled background. The last word of each sentence was the target word. This was presented as a picture on screen for 1000 ms. The task was to silently read the sentences attentively and name the pictures with the words that participants were familiarized with. Also, participants were asked to keep fixation to the center of the screen and move their jaw and head as little as possible.

In the following sections, the procedure for the behavioral data with participants' reaction times for naming the pictures and error coding is reported first. Then, acquisition, preprocessing, and analysis steps of each method are reported separately, first for the MEG experiment and then for the fMRI experiment. Last, the procedure for calculating the Dice coefficients for both methods is stated.

Behavioral Analysis.

In both scanners, trials were recorded to monitor participants' responses for picture naming. Recordings started simultaneously with picture onset and lasted 2500 ms. Trials in which participants hesitated, stuttered, responded either with more than one word or later than 2500 ms after picture onset were considered as errors and not included in the

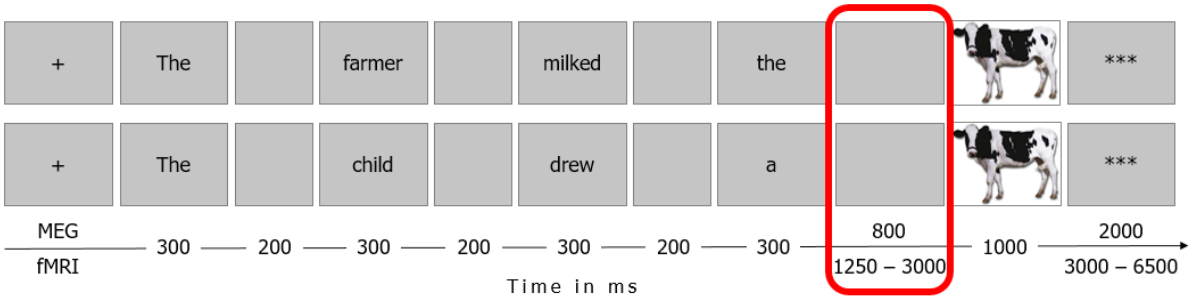

Fig. 1. Overview of the constrained (top) and unconstrained (bottom) trial for the target word cow. Boxes represent the screens that participants saw in the scanner, with experiment-specific presentation times below and the analyzed time window of interest circled in red

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 presentation and recording onset no reaction time could be measured and the trial was discarded from this analysis. Reaction times were calculated using the speech editor Praat (Boersma & Weenink, 2017), blind for condition, and statistically analyzed in R (Team, 2017). The mean reaction time per condition was calculated for each participant and the behavioral effect was evaluated by means of an analysis of variance with condition and session as within-participant variables at an alpha-level of 0.05.

MEG Experiment.

MEG Acquisition.

Participants taking part in the MEG study had to change into the non-magnetic scanner 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 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 showing three asterisks (***). MEG data were acquired with a 270 axial gradiometer system (CTF Systems Inc., VSM MedTech Ltd.) at a sampling rate of 1200 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 relocalized in the breaks after every block of 28 trials. The scanning for one MEG session lasted approximately 30 minutes and participants were in the laboratory for one hour, including preparation time.

If not yet available, structural T1-weighted MRI scans of participants' heads were acquired either after one of the two MEG sessions or on a third day.

MEG Preprocessing.

MEG data preprocessing was performed in Matlab using the FieldTrip toolbox (Oostenveld, Fries, Maris & Schoffelen, 2011). The data were demeaned to take out drifts and each trial was cut down to the time window of interest of 800 ms before 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 blinking trials were discarded by means of the vertical electrooculogram channel. 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 and not considered for further analyses. During preprocessing, 8 to 20 sensors ($M = 15$, $SD = 3$) were removed per session and each session consisted of 170 to 213 trials for analysis ($M = 198$, $SD = 10$), including 82 to 110 ($M = 99$, $SD = 6$) unconstrained trials and 81 to 106 ($M = 98$, $SD = 6$) constrained trials.

MEG Analysis.

The MEG analysis is 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 last presented word and the onset of the picture in each trial, during which the screen was blank.

Source Localization.

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 SPM8, to obtain individual volume conduction models of participants' heads. 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 can be compared across participants. Using the volume conduction models, lead field matrices were computed for each grid point per participant. Then, the beamforming technique was applied to estimate the activity at the source-level. 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 is based on previous findings resulting from the same task and analysis technique (Piai et al., 2015). As the transition from alpha to beta activity is usually considered around 12 to 15 Hz, this frequency range is referred to as alpha-beta power for 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 power change was calculated as the difference between power in the constrained and the unconstrained condition divided by the common average. The effect of the power differences from the constrained to the unconstrained condition over all participants was evaluated by means of a non-parametric cluster-based permutation test. A dependent-samples t-test thresholded at an alpha-level of 0.05 served to identify the biggest cluster of neighboring voxels for the effect on the group-level. The p-values of this cluster were calculated as the amount out of 5000 random permutations that yield a larger effect than the observed one, employing a Monte Carlo method with 5000 random permutations.

MEG First Half Effect Size.

To estimate the effect size of our paradigm 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 (*SD* = 4) unconstrained trials and 57 (*SD* = 4) constrained trials.

fMRI Experiment.

fMRI Acquisition.

Participants taking part in the fMRI study had to wear metal-free clothing on their upper body and change into scanner clothing if necessary. Then they were taken into the scanner room and positioned on the scanner bed with cushions underneath their knees and elbows. Their forehead was taped to the lower part of the head coil with crepe tape to minimize their head movement, and an emergency button was placed on their belly. All functional scans were acquired on a 3T Siemens PrismaFit scanner with a 32-channel head coil using echo-planar imaging (EPI) to minimize movement artefacts, employing a multiband sequence (multiband acceleration factor 6, 2 mm isotropic voxels, 66 slices, TR = 1000 ms, TE = 34 ms, FoV = 210 x 210 x 132 mm, flip angle = 60°). Localizer and head scout scans were performed before the start of the experiment to obtain the location of participants' brains. The experiment started with a pulse countdown from nine to zero followed by all 224 trials consecutively. Field maps were acquired at the end of the trials (TR = 620 ms, TE 1 = 4.7 ms, 64 slices, voxel size 2.4 x 2.4 x 2 mm, FoV = 210 x 210 x 132 mm, flip angle = 60°) to calculate voxel displacement maps (VDM) for each session. Structural T1-weighted MPRAGE images (TR = 2300 ms, TE = 3.03 ms, 192 slices, FoV = 256 x 256 x 192 mm, 1 mm isotropic voxels)

for anatomical reference were acquired after session 1. The fMRI experiment had a jittered design to capture the BOLD response at different stages. The interval before picture onset was randomly jittered between 1250 ms and 3000 ms, and the fixation cross between trials was randomly jittered between 3000 ms and 6500 ms. This prolonged the fMRI sessions such that the experimental scanning took approximately 45 minutes.

fMRI Preprocessing.

fMRI preprocessing was performed session-individually using Matlab and SPM12. The first nine volumes of each session were discarded as dummy scans to allow the magnetization to reach a steady state. All other images were realigned with reference to the 10th volume and unwarped by applying the calculated session-specific VDM to reduce movement artefacts of the functional EPI scans. After estimation of coregistration the images underwent segmentation into different tissue types based on probability maps such as grey matter, white matter, cerebrospinal fluid, bone, and soft tissue. Then the images were normalized to MNI space and resliced, and a smoothing Gaussian kernel of twice the voxel size (FWHM = 4 mm) was applied.

fMRI Analysis.

As for the MEG experiment, the fMRI analysis focuses on the difference in brain activity between both conditions in the time window of interest before picture presentation (see Figure 1). This interval from the last word of the sentence to the picture was jittered between 1250 ms and 3000 ms.

General Linear Model.

The fMRI analysis was performed by means of a general linear model (GLM) for each participant. This included five regressors per session containing the onsets of task-specific events as well as the six motion parameters outputted from the session-specific preprocessing to account for further movement artefacts of participants in the scanner. A high-pass filter removed slow signal drifts with periods longer than 128 seconds. The task-specific regressors contained the time-locked onsets of each trial for the first word, the pre-picture interval per condition, and the picture. The fifth regressor modelled all incorrect trials per session (M = 3, SD = 4) by means of the onsets of the pre-picture intervals of incorrect trials. Not considering errors, each session consisted of an average of 110 trials per condition (SD = 2). The onsets were modelled as delta (stick) functions (duration = 0) and convolved with the canonical HRF, i.e., the BOLD response. The design matrix for each participant consisted of the GLM for the first and second session. The contrast of interest was the increase of the BOLD response for constrained over unconstrained trials. Therefore, the BOLD signal of the whole brain at the onset of the pre-picture interval for unconstrained trials was subtracted from that of constrained trials for each participant. These individual t-contrasts then entered the second-level analysis to obtain t-contrasts of the group-level BOLD increases in the whole brain per session.

Table 1.
Amount of total errors and errors per condition for each session. Mean reaction time effect per session and average reaction times per condition.

Experiment	MEG		fMRI	
Session	1	2	1	2
Errors in %				
Total	2.26	2.74	1.49	1.55
unconstrained	2.08	3.33	1.85	1.79
constrained	2.44	2.14	1.13	1.31
Reaction Times in ms				
Effect	143 (SD = 83)	177 (SD = 72)	201 (SD = 59)	203 (SD = 64)
unconstrained	738 (SD = 92)	759 (SD = 84)	896 (SD = 225)	926 (SD = 230)
constrained	595 (SD = 89)	583 (SD = 129)	694 (SD = 269)	723 (SD = 269)

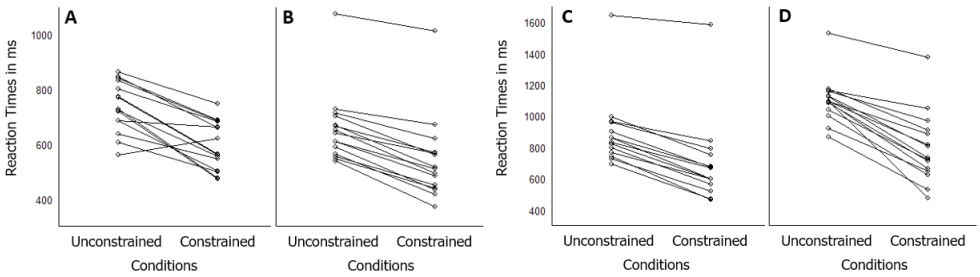

Fig. 2. Mean reaction times per condition in session 1 (A) and session 2 (B) of the MEG experiment, and session 1 (C) and session 2 (D) of the fMRI experiment. Each line connects both conditions for one participant

fMRI First Half Effect Size.

To analyze the first half of each fMRI session, a cutoff point was determined by means of the picture onset of trial 112 (out of 224) of each session. To capture the full BOLD response for this trial, 20 additional volumes were included. Then, the motion parameters from the preprocessing step were modified to match the length of each individual half session. A new GLM was constructed per participant with all onsets for the first 112 trials and the same regressors as specified above. This included an average of 54 unconstrained (*SD* = 4) and 57 constrained (*SD* = 4) trials as well as errors (*M* = 1, *SD* = 2). The individual and group-level t-contrasts were also constructed in the same manner as described above.

Dice Coefficients.

To quantify the extent of overlap of the measured brain activity between session 1 and session 2 and compare it for both experiments, the corresponding Dice coefficients were computed. This measure is calculated by twice the number of overlapping voxels (*overlap*) divided by the sum of activated voxels in session 1 (*ses1*) and session 2 (*ses2*). Here, an outcome of 0 indicates no overlap and an outcome of 1 indicates a perfect overlap of activation between both sessions:

For both methods, this calculation was based on the significant voxels resulting from the analyses. For MEG the statistical voxel threshold was set to 0.05. For fMRI the p-threshold on voxel-level was 0.001 uncorrected, only including Family-wise error (FWE) corrected clusters with $p < 0.05$ (as reported in Table 3).

Results

In this section, the behavioral results for reaction time measurements are reported first, followed by the results for the MEG and the fMRI experiment respectively. Finally, the Dice coefficients as the chosen measure of overlap are reported for both experiments in comparison.

Figure 2 A and B shows the mean reaction time effects for picture naming of each participant in the MEG experiment per session. This effect is the difference between the mean reaction times for unconstrained and constrained sentences. The mean reaction time for unconstrained sentences was 738 ms (*SD* = 92) in session 1 and 759 ms (*SD* = 84) in session 2. For constrained sentences this was 595 ms (*SD* = 89) in session 1 and 583 ms (*SD* = 129) in session 2. Thus, participants named the pictures in constrained sentences faster than in unconstrained sentences. More exact, the mean effect of sentence constraint was 143 ms (*SD* = 83) in session 1, and 177 ms (*SD* = 72) in session 2. Overall, this yields a main effect of condition, $F(1, 14) = 81.15$, $p < 0.001$, no significant effect of session, $F(1, 14) = 0.02$, $p = 0.9$, and no interaction effect, $F(1, 14) = 3.25$, $p = 0.1$.

Figure 2 C and D shows the mean effects for picture naming of all fMRI participants per session. Here, the mean reaction time for unconstrained sentences was 896 ms (*SD* = 225) in session 1 and 926 ms (*SD* = 230) in session 2, and for constrained sentences 694 ms (*SD* = 269) in session 1 and 723 ms (*SD* = 269) in session 2. As in the MEG experiment, participants reacted faster to constrained than unconstrained sentences, with a mean effect of 201 ms (*SD* = 59) in session 1 and 203 ms (*SD* = 64) in session 2. This also yields a main effect of condition, $F(1, 14) = 223.11$, $p < 0.001$, no significant effect of session, $F(1, 14) = 2.32$, $p = 0.15$, and no interaction effect $F(1, 14) = 0.01$, $p = 0.93$.

MEG Results

The top row of Figure 3 shows the group-level results for the source localization of the power

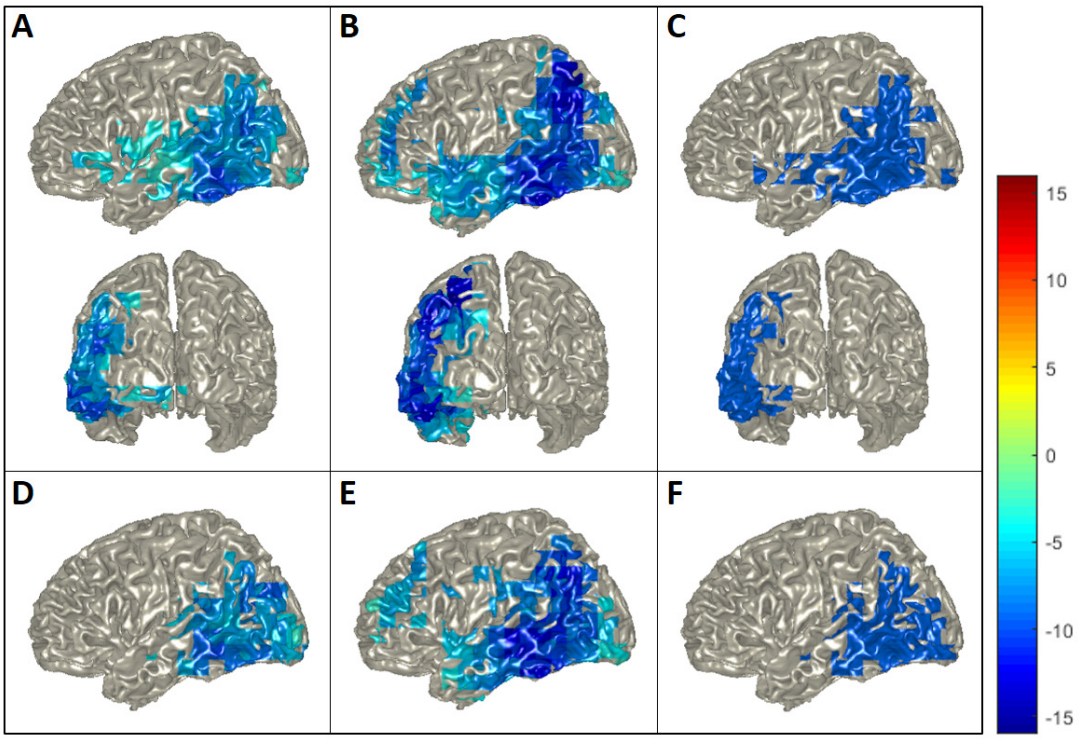

Fig. 3. Group-level source localization power decreases for constrained relative to unconstrained condition at 10 - 20 Hz for full session 1 (A) and full session 2 (B) in left and posterior view. Power decreases for first half of session 1 (D), and first half of session 2 (E) in the left hemisphere. Depicted areas are masked by statistically significant clusters, color bars show relative power change in percentage. Overlap map of group-level source localization showing areas of power changes common to both sessions for the full sessions (C) and the first half of both sessions (F)

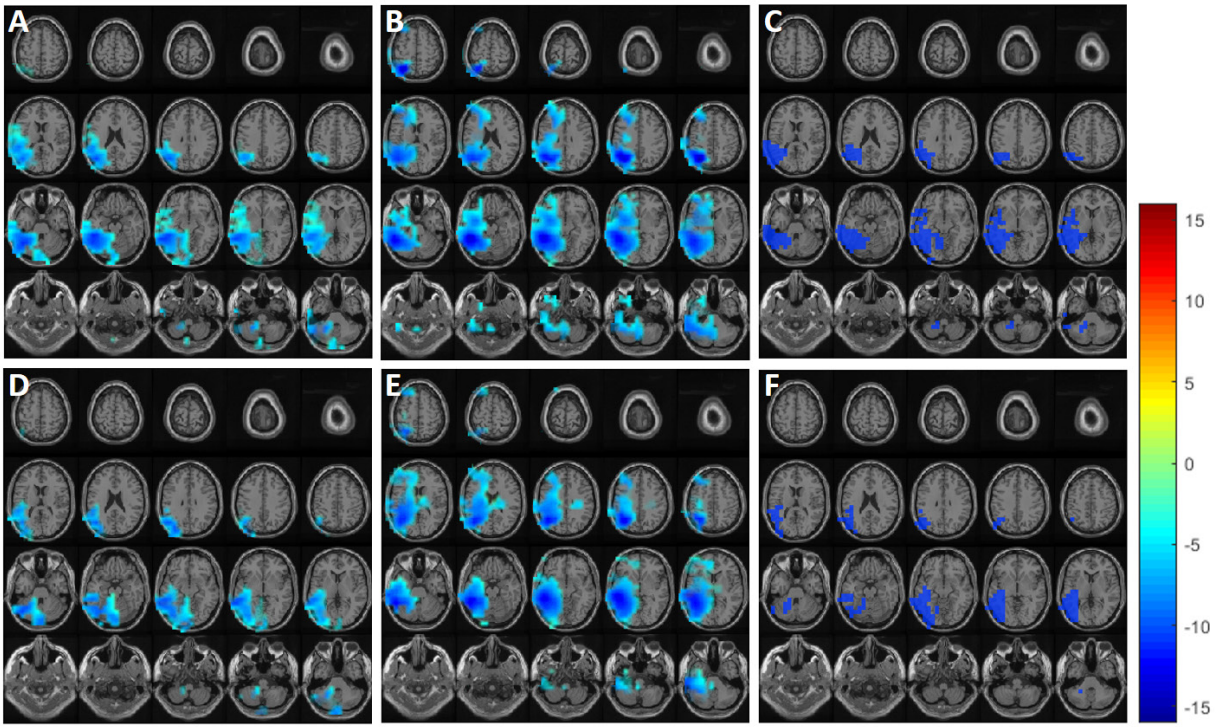

Fig. 4. Group-level source localization power decreases between conditions at 10 - 20 Hz for full session 1 (A), full session 2 (B), first half of session 1 (D), and first half of session 2 (E) in axial brain slices. Depicted areas are masked by statistically significant clusters, color bars show relative power change in percentage. Overlap map of group-level source localization showing areas of power changes common to both sessions for the full sessions (C) and the first half of both sessions (F)

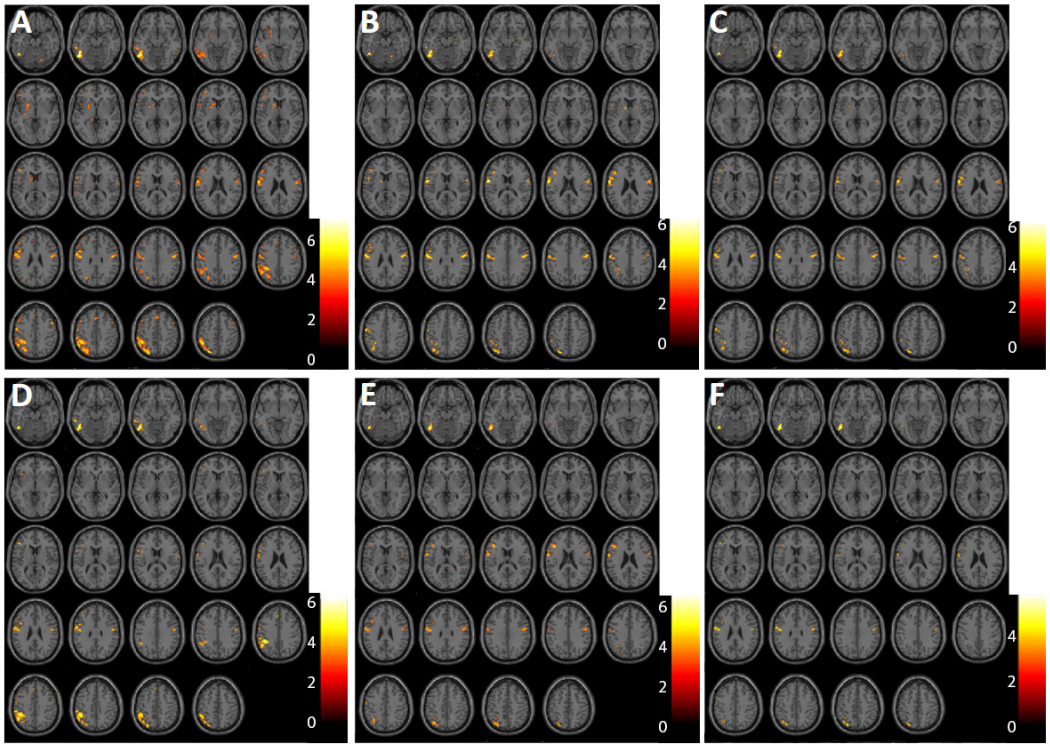

Fig. 5. Clusters of group-level BOLD increases for constrained over unconstrained trials for full session 1 (A), full session 2 (B), first half of session 1 (D), and first half of session 2 (E) in axial slices (MNI space -18 to 54 mm, steps of 3 mm). Conjunction showing areas of group-level BOLD increases common to both sessions for the full sessions (C) and the first half of both sessions (F). Color bars show t-values, $p < 0.001$ on voxel-level

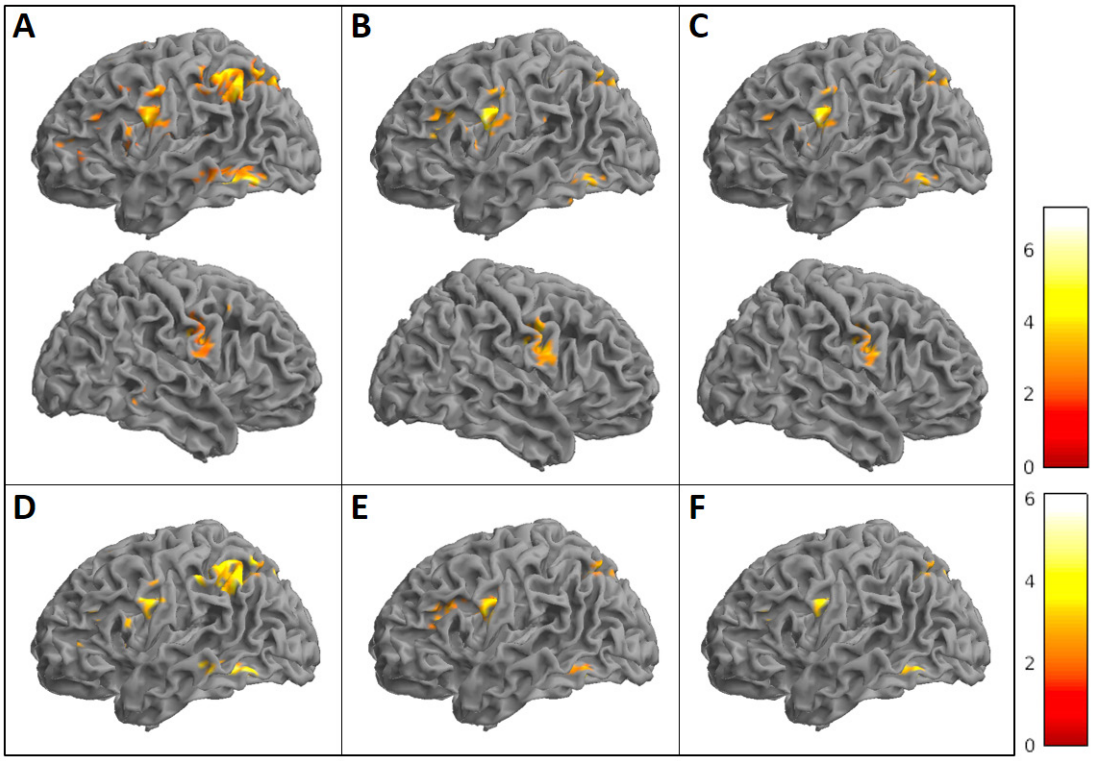

Fig. 6. Group-level BOLD increases for full session 1 (A) and full session 2 (B) projected to the surface of the left (top) and right (bottom) hemisphere. BOLD increases in the first half of session 1 (D), and first half of session 2 (E) projected to the surface of the left hemisphere. Conjunction showing areas of group-level BOLD increases common to both sessions for the full sessions (C) and the first half of both sessions (F). Color bars show t-values, $p < 0.001$ on voxel-level

differences for both sessions on the brain surface as well as their overlap, masked by the statistically significant clusters.

The top row of Figure 4 depicts the obtained clusters and their spatial distribution inside the brain in axial slices. However, caution should be taken when interpreting these slices, as the spatial resolution as well as sensitivity of MEG for subcortical sources is reduced (Hillebrand et al., 2016). In both figures, color scales show the percentage of power change in session 1 and session 2.

The power decreases of constrained relative to unconstrained sentences in the alpha-beta frequency range from 10 to 20 Hz were statistically significant for both session 1 (p = 0.0092) and session 2 (p = 0.0052). The significant clusters of power changes were exclusively lateralized to the left hemisphere in both sessions. In session 1, the strongest power decreases around 10% were observed in the posterior temporal lobe, as shown in Figure 3 A. More precisely, they extended from the inferior to the superior temporal gyrus, and then posteriorly to the inferior parietal lobe. Anteriorly a weaker decrease around 5% extended until 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 angular and supramarginal gyrus up to the superior parietal gyrus, as shown in Figure 3 B. Anteriorly a weaker decrease of 5 - 10% extended over the anterior temporal lobe and the orbital inferior frontal gyrus to the middle frontal gyrus.

MEG Across-Session Consistency.

To depict the overlap of brain areas exhibiting power changes in both sessions, a new mask was created. This was based on the significant clusters resulting from the source localization for both sessions and only included voxels that were active in both session 1 and session 2. This overlap is shown in Figure 3 C. 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 gyrus, and anteriorly over the superior temporal gyrus slightly into the inferior frontal gyrus.

MEG First Half Effect Size.

The bottom rows of Figure 3 and Figure 4 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 = 0.0036) and session 2 (p = 0.0024). Compared to the full session, the power decreases of the first half of session 1 are restricted to the posterior temporal lobe and inferior parietal gyrus, not extending to the inferior frontal gyrus, as shown in Figure 3 D. The power decreases in the first half of session 2 are quite consistent with the full session. They extend from the posterior temporal lobe to the inferior parietal gyrus, and anteriorly over the anterior temporal lobe to the middle frontal gyrus, as shown in Figure 3 E.

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 3 F. This area of overlap extends over the posterior temporal lobe to the inferior parietal gyrus, presenting a similar overlap as for the full sessions, but less anterior and continuous.

fMRI Results

The upper part of Table 3 lists significant clusters of neighboring voxels showing a BOLD increase for the contrast of interest per session, meaning areas that show a stronger BOLD response for constrained than unconstrained trials. The corresponding brain regions are based on the MNI space voxel coordinates obtained with SPM12 and determined by using the Automated Anatomical Labeling (AAL) atlas (Tzourio-Mazoyer et al., 2002). The top row of Figure 5 shows the whole-brain BOLD increases in session 1 and session 2 as well as their overlap by means of the same selected axial slices. The top row of Figure 6 depicts these clusters of BOLD increases in session 1 and session 2 and their overlap projected to the brain surface for both hemispheres.

The BOLD increases for session 1 are shown in Figure 5 A and Figure 6 A. Here, left-hemisphere BOLD increases were detected in a cluster in the inferior occipital lobe and inferior temporal gyrus, a large cluster in the superior and inferior parietal gyrus, a cluster in the post- and precentral gyrus, followed by clusters in the superior frontal gyrus, the middle and triangular inferior frontal gyrus, the pallidum, the caudate, and the supplementary motor area and medial superior frontal gyrus. Clusters of BOLD increase in the right hemisphere were detected in the precentral and middle frontal gyrus, the cerebellum (crus 2), and the postcentral gyrus.

BOLD increases for session 2 are shown in Figure 5 B and in Figure 6 B. Left-hemisphere

clusters showing BOLD increases were located in the fusiform and inferior temporal gyrus, the Rolandic operculum and postcentral gyrus, the superior and inferior parietal and angular gyrus, the inferior parietal gyrus, and the triangular inferior frontal gyrus. For session 2 only one cluster in the postcentral gyrus was significant in the right hemisphere.

fMRI Across-Session Consistency.

To look at the overlap of voxel clusters exhibiting an increase in the BOLD response for constrained over unconstrained trials for both sessions, the contrasts of interest from session 1 and 2 were entered in a conjunction analysis in SPM12. The significant clusters of overlap from both sessions are listed under *Full Conjunction* in Table 3 and shown for the same slices as depicted for the individual sessions in Figure 5 C. Clusters of overlapping BOLD increase in the left hemisphere were located in the fusiform and inferior temporal gyrus, the postcentral gyrus, the superior parietal and angular gyrus, and the inferior parietal gyrus. The overlap of significant BOLD increases in the right hemisphere was limited to the postcentral gyrus.

fMRI First Half Effect Size.

The lower part of Table 3 lists all brain areas with significant clusters of BOLD increases for the contrast of interest per session and their overlap, when analyzing only the first half of each session. These are shown in the bottom row of Figure 5 and Figure 6 for session 1, session 2, and their overlap.

The first half of session 1 only yielded significant clusters in the left hemisphere, shown in Figure 5 D and Figure 6 D. A large cluster was detected in the inferior parietal gyrus, followed by clusters in the fusiform gyrus and inferior occipital lobe, the inferior and middle temporal gyrus, and the precentral gyrus.

As shown in Figure 5 E and Figure 6 E, in the first half of session 2 the significant left-hemisphere clusters were located in the fusiform and inferior temporal gyrus, the post- and precentral gyrus, the inferior and superior parietal gyrus, and the triangular inferior frontal gyrus. In the right hemisphere there was one significant cluster in the postcentral gyrus.

As for the full session data, a conjunction analysis of the contrasts of interest was conducted with SPM12 for the first half of both sessions to look at the overlapping voxel clusters that exhibit a BOLD increase for constrained over unconstrained trials.

Table 2.

Dice coefficient values for full and half session calculation per experiment and average amount of input trials per participant.

Experiment	MEG	fMRI
FULL SESSION		
Dice Coefficient	0.49	0.43
Trials	198 (*SD* = 10)	220 (*SD* = 4)
HALF SESSION		
Dice Coefficient	0.35	0.31
Trials	112 (*SD* = 0)	111 (*SD* = 2)

The locations of these overlaps are listed under *Half Conjunction* in the lower part of Table 3 and are shown in Figure 5 F and Figure 6 F. Overlapping clusters were located in the fusiform and inferior temporal gyrus, the superior and inferior parietal gyrus, and the pre- and postcentral gyrus, all in the left hemisphere.

Table 2 shows the Dice coefficients per experiment and the average amount of trials per participant that the respective calculation is based on. Dice coefficients were calculated as described above for the full and half sessions of both experiments, by means of the respective group-level results. Based on the partition thresholds of Cohen's *d* (Cohen, 1988) as a similar effect size measure, Dice coefficients are classified as *small* (0.20-0.49), *medium* (0.50-0.79), and *large* (0.80-1). The highest Dice coefficient was obtained for the full MEG experiment (0.49), followed by the full fMRI experiment (0.43). The first half analysis resulted in lower Dice coefficients for MEG (0.35) as well as fMRI (0.31).

Discussion

The present study aimed to determine the most suitable imaging method, the across-session consistency, and the effect size for short testing sessions for the present paradigm to develop a clinical tool to track the recovery of language functions in patients. Therefore, participants were tested with MEG and fMRI while performing the selected sentence completion task to be employed in the clinical tool. In both experiments, the behavioral differences in terms of reaction times per condition demonstrate faster word retrieval for constrained than unconstrained sentences, approving that the context of constrained sentences provides the necessary information to retrieve the target word.

Table 3.
Significant clusters of whole-brain BOLD increases for constrained over unconstrained trials. Cluster size given in number of voxels (2 mm isotropic) in the cluster. Coordinates given for maximally activated voxel and up to two local maxima more than 8 mm apart. Voxel-threshold at p = 0.001; cor: Family-wise error (FWE) corrected; unc: unocorrected; l: left; r: right, tr.: triangular; *cases where obtained coordinates deviate 1 mm from AAL region (-47).

Cluster		Voxel			MNI space	Brain structure (AAL)
p(cor)	size	t value	z value	p(unc)	x, y, z (mm)	
FULL SESSION 1						
0.000	531	7.12	5.73	0.000	-48, -58, -14	l inferior occipital lobe
		6.48	5.37	0.000	-46, -50, -16	l inferior temporal gyrus
		5.92	5.02	0.000	-58, -62, -12	l inferior temporal gyrus
0.000	1728	5.87	4.99	0.000	-28, -74, 50	l superior parietal gyrus
		5.71	4.89	0.000	-46, -48, 52	l inferior parietal gyrus
		5.57	4.79	0.000	-50, -54, 42	l inferior parietal gyrus
0.000	717	5.50	4.75	0.000	-58, -2, 22	l postcentral gyrus
		5.22	4.55	0.000	-56, -6, 42	l postcentral gyrus
		5.10	4.48	0.000	-46*, -10, 30	l precentral gyrus
0.015	79	5.47	4.72	0.000	52, 12, 44	r precentral gyrus
		4.46	4.01	0.000	40, 6, 60	r middle frontal gyrus
		3.81	3.51	0.000	46, 12, 52	r middle frontal gyrus
0.042	64	5.28	4.59	0.000	8, -78, -30	r cerebellum (crus2)
0.025	72	5.06	4.45	0.000	-16, 18, 62	l superior frontal gyrus
		3.73	3.44	0.000	-18, 8, 66	l superior frontal gyrus
0.000	321	5.02	4.42	0.000	50, -8, 32	r postcentral gyrus
		4.14	3.77	0.000	58, 2, 20	r postcentral gyrus
		4.01	3.67	0.000	64, -6, 18	r postcentral gyrus
0.000	154	5.01	4.41	0.000	-50, 14, 44	l middle frontal gyrus
		4.67	4.17	0.000	-48, 30, 26	l tr. inferior frontal gyrus
		3.69	3.42	0.000	-50, 22, 28	l tr. inferior frontal gyrus
0.001	134	4.95	4.37	0.000	-24, -4, -4	l pallidum
		4.84	4.29	0.000	-22, 10, -4	l pallidum
		4.81	4.27	0.000	-18, 2, 8	l pallidum
0.021	74	4.82	4.28	0.000	-14, 12, 10	l caudate nucleus
0.034	67	4.56	4.09	0.000	0, 24, 46	l supplementary motor area
		3.61	3.35	0.000	0, 36, 48	l medial superior frontal gyrus
FULL SESSION 2						
0.000	216	6.19	5.19	0.000	-48, -56, -18	l fusiform gyrus
		5.58	4.80	0.000	-46, -48, -16	l inferior temporal gyrus
0.000	611	5.97	5.05	0.000	-54, 0, 16	l Rolandic operculum
		4.93	4.35	0.000	-56, -6, 44	l postcentral gyrus
		4.81	4.27	0.000	-54, -8, 30	l postcentral gyrus
0.000	240	4.85	4.30	0.000	-28, -74, 52	l superior parietal gyrus
		4.68	4.18	0.000	-30, -48, 44	l inferior parietal gyrus

		4.52	4.06	0.000	-36, -64, 44	l angular gyrus
0.000	344	4.69	4.18	0.000	50, -6, 32	r postcentral gyrus
		4.36	3.94	0.000	58, 2, 22	r postcentral gyrus
		4.21	3.82	0.000	64, -2, 18	r postcentral gyrus
0.020	75	4.29	3.89	0.000	-50, -44, 56	l inferior parietal gyrus
		3.88	3.57	0.000	-48, -38, 50	l inferior parietal gyrus
		3.87	3.56	0.000	-38, -40, 42	l inferior parietal gyrus
0.001	135	4.23	3.84	0.000	-42, 22, 24	l tr. inferior frontal gyrus
		4.03	3.68	0.000	-40, 28, 18	l tr. inferior frontal gyrus
		3.84	3.54	0.000	-44, 34, 12	l tr. inferior frontal gyrus
FULL CONJUNCTION						
0.000	209	6.19	5.19	0.000	-48, -56, -18	l fusiform gyrus
		5.58	4.80	0.000	-46, -48, -16	l inferior temporal gyrus
0.000	475	5.50	4.75	0.000	-58, -2, 22	l postcentral gyrus
		4.86	4.30	0.000	-56, -6, 42	l postcentral gyrus
		4.60	4.12	0.000	-48, -10, 30	l postcentral gyrus
0.000	187	4.85	4.30	0.000	-28, -74, 52	l superior parietal gyrus
		4.44	4.00	0.000	-36, -64, 44	l angular gyrus
		3.81	3.51	0.000	-32, -68, 56	l superior parietal gyrus
0.000	244	4.69	4.18	0.000	50, -6, 32	r postcentral gyrus
		4.14	3.77	0.000	58, 2, 20	r postcentral gyrus
		3.83	3.53	0.000	62, -4, 26	r postcentral gyrus
0.020	75	4.29	3.89	0.000	-50, -44, 56	l inferior parietal gyrus
		3.88	3.57	0.000	-48, -38, 50	l inferior parietal gyrus
		3.87	3.56	0.000	-38, -40, 42	l inferior parietal gyrus
HALF SESSION 1						
0.000	1043	6.08	5.12	0.000	-42, -44, 40	l inferior parietal gyrus
		5.08	4.46	0.000	-50, -54, 42	l inferior parietal gyrus
		4.91	4.34	0.000	-54, -32, 44	l inferior parietal gyrus
0.000	238	5.70	4.88	0.000	-46, -56, -14	l fusiform gyrus
		5.29	4.61	0.000	-52, -62, -12	l inferior occipital lobe
		4.81	4.27	0.000	-46, -48, -16	l inferior temporal gyrus
0.049	57	4.60	4.11	0.000	-60, -38, -14	l inferior temporal gyrus
		3.44	3.21	0.000	-52, -46, -6	l inferior temporal gyrus
		3.31	3.10	0.000	-62, -36, -4	l middle temporal gyrus
0.000	160	4.55	4.08	0.000	-46*, -10, 30	l precentral gyrus
		4.46	4.01	0.000	-56, 0, 30	l precentral gyrus
		4.17	3.79	0.000	-60, -6, 26	l precentral gyrus
HALF SESSION 2						
0.000	203	6.79	5.55	0.000	-48, -56, -18	l fusiform gyrus
		4.52	4.06	0.000	-46, -48, -16	l inferior temporal gyrus
0.000	321	5.14	4.50	0.000	-58, -2, 22	l postcentral gyrus
		4.33	3.92	0.000	-46*, -10, 30	l precentral gyrus
		4.23	3.84	0.000	-58, -6, 42	l postcentral gyrus

0.000	220	4.95	4.37	0.000	-34, -68, 46	l inferior parietal gyrus
		4.80	4.26	0.000	-28, -72, 50	l superior parietal gyrus
		4.27	3.87	0.000	-34, -56, 42	l inferior parietal gyrus
0.000	205	4.93	4.35	0.000	-40, 24, 20	l tr. inferior frontal gyrus
		4.54	4.07	0.000	-46, 30, 20	l tr. inferior frontal gyrus
		3.88	3.57	0.000	-42, 32, 12	l tr. inferior frontal gyrus
0.000	179	4.49	4.03	0.000	54, -4, 34	r postcentral gyrus
		4.06	3.70	0.000	60, 2, 24	r postcentral gyrus
HALF CONJUNCTION						
0.000	171	5.70	4.88	0.000	-46, -56, -14	l fusiform gyrus
		4.52	4.06	0.000	-46, -48, -16	l inferior temporal gyrus
		3.60	3.34	0.000	-42, -52, -8	l inferior temporal gyrus
0.001	120	4.45	4.01	0.000	-28, -74, 48	l superior parietal gyrus
		4.14	3.77	0.000	-32, -68, 56	l superior parietal gyrus
		4.00	3.66	0.000	-32, -66, 44	l inferior parietal gyrus
0.001	117	4.33	3.92	0.000	-46*, -10, 30	l precentral gyrus
		4.13	3.76	0.000	-56, -4, 28	l postcentral gyrus

MEG Discussion

The MEG results show the expected alpha-beta power decreases in accordance with previous findings by Piai et al. (2014, 2015, 2017a, 2017b). These decreases have previously been argued to reflect context-dependent aspects as well as motor preparation of word production, especially in the beta frequency range. 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. This suggests that the motor activity in the right-hemisphere was not significant enough to be captured in the present study. Further, none of the obtained clusters in the left hemisphere clearly involved left speech motor areas, meaning that also the captured activity most likely did not reflect any motor preparation. Thus, the present findings indicate that motor preparation is not such a robust aspect of alpha-beta power decreases as it has previously been argued, but that the alpha-beta effect mostly reflects the context-driven retrieval of concept and word information.

Comparing the results from both sessions, session 2 shows a stronger and spatially more distributed alpha-beta power decrease that is more significant than that of 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. This could possibly represent a familiarization effect with the whole experimental procedure on the side of the participants, as session 2 was an analogue replication of session 1 and only differed in the stimulus materials. This could be further investigated by looking at the individual power decrease maps for each participant to see whether this is a common effect across the group. Alternatively, frontal MEG activation for picture naming has also been found to vary between participants (Liljeström et al., 2009).

In terms of consistency, the activation maps of session 1 and session 2 share a high amount of overlap, yielding a marginally small Dice coefficient of 0.49.

When analyzing only the first half of the MEG sessions, the core areas of the strongest power decreases are the same as for the full sessions, but not as spatially spread out. Accordingly, the area of overlapping activity from the half sessions is smaller than the overlap of the full sessions. Also, the Dice coefficient decreased from 0.49 to 0.35. Interestingly, the robustness of the source localization clusters obtained with the non-parametric Monte Carlo permutation test increased from full to half analysis for both sessions. In other words, the likelihood of accidentally finding a stronger result than the observed one out of 1000 permutations reduced by more than one half. This shows that the quality of the acquired data decreases with the duration of the session, probably partly due to movement and fatigue of the participants. In general, for MEG this suggests that acquiring more data is not always better and does not necessarily increase the significance of the investigated effect.

fMRI Discussion

The brain activity captured with BOLD-fMRI revealed significant clusters, mostly but not exclusively in the left hemisphere. Left-hemisphere BOLD increases were spread over the frontal, parietal, and temporal lobe. A strong aspect of both fMRI sessions as well as their overlap is the motor activity in the right and left postcentral gyrus. Obtaining this in the contrast of constrained over unconstrained trials but not in the reverse direction suggests that motor preparation activity is closely linked to concept and word retrieval and depends on participants' knowledge of which word will be articulated. Possibly, participants already imagine to pronounce the word while waiting for the picture to appear. Thus, the motor preparation most likely includes speech planning in terms of adjusting the speech organs for articulation and pronunciation. This could be further investigated with the existing data of the present study by comparing the motor preparation activity of both conditions at the time point when participants could retrieve the concept of the target word. For constrained trials this would be at the same time point as analyzed here, the pre-picture interval. For unconstrained trials, however, the concept is only presented by the appearance of the picture. Thus, an adequate timepoint for this would be after picture onset, but before participants' speech onset to prevent increased scanning artefacts through speech-related motion.

The results of session 1 reveal some clearly visible left-lateralized clusters of BOLD activity. However, this activity majorly diminished from session 1 to session 2, meaning that participants showed less BOLD increases when performing the experiment a second time. This diminishing activity from session 1 to session 2 also largely impairs across-session consistency. The conjunction analysis therefore only captures few spatially overlapping areas of BOLD increases and the corresponding Dice coefficient quantifying the overlap of the full fMRI sessions results in 0.43.

An alternative approach to improve the power of the fMRI experiment and possibly capture stronger BOLD signals in session 2 would be a block design. As the BOLD response also reflects long-lasting processes it is quite sensitive to variations from trial to trial (Liljeström et al., 2009). Therefore, trials could be presented in blocks of sentences in the same condition, instead of in a randomized trial order as it was in the present experiment. That way, participants would constantly activate the same brain regions for the duration of one block. Especially with the sensitive contrast between the two conditions in our paradigm, that only differed in the context of the sentence preceding the picture, an event-related design might be too subtle and noisy to capture strong BOLD increases between conditions. Another possibility to increase power for fMRI is to restrict the analysis to a region of interest. But for the purpose of the present project, to investigate the reorganization of the brain, it is important that all analyses are conducted on a whole-brain level, without any prior spatial restrictions.

The analysis of the first half of fMRI session 1 only yielded about one third of the BOLD increase clusters that resulted from the full session analysis. Also, the number of significant voxels composing the clusters, meaning the cluster size, reduced vastly. This also holds for session 2, but here the half session analysis still revealed all brain areas that were detected to show BOLD increases in the full session analysis. Whereas for session 1 many detected areas of BOLD increases are not significant anymore for the first half of session 1. In line with this, the Dice coefficient of overlap for the first half of both sessions also decreased to 0.31. Especially considering the loss of clusters when only analyzing the first half of session 1 suggests that the power of fMRI data generally increases with the amount of data collected. However, the session duration and the fact that fatigue and motion of the participants over time increasingly introduce noise also need to be considered. These aspects constitute limitations regarding fMRI session duration.

MEG vs. fMRI

The obtained activation maps of both methods do not completely converge across experiments, showing that both methods capture different aspects of neuronal activity. MEG is sensitive to the magnetic fields induced by electrical currents in the brain on a sub-second time scale, whereas fMRI depends on the blood oxygen level that changes in form of a 6 to 12 second response curve. This time constraint of the fMRI signal constitutes a plausible explanation of the divergence of activity detected per method. Liljeström et al. (2009) for example obtained visual activation for 300 ms stimuli with MEG, but not with fMRI. They argue that very short stimuli might not be detected with fMRI, as the BOLD response is rather slow and may also reflect a summation of long-term processes. Other studies agree with this and further discuss that also MEG

might fail to capture activity that shows significant BOLD responses in fMRI, in cases where neurons are not activated in synchrony (Kujala et al., 2014; Vartiainen et al., 2011). Also, MEG is less sensitive to tangential than radial components of neuronal currents (Kujala et al., 2014), meaning that the magnetic fields of neuronal currents that flow radial to the scalp are not detected well with MEG.

For the findings of the present study, this indicates that the large activation clusters detected with MEG are mostly based on computations with millisecond time scales and thus do not evoke a wide-spread corresponding BOLD response. Further, the fMRI results of the present study reveal consistent motor activity that was not reflected in the MEG results, but has previously been detected with the same paradigm in MEG (Piai et al., 2015). This suggests that the lack of captured motor activity with MEG in the present study is due to analysis thresholds, rather than unsynchronized or radially oriented neurons. Although, for this direction of comparing the results across methods, it is important to keep in mind that the MEG analysis in the present study was based on prior knowledge of the frequency of interest (Piai et al., 2014, 2015, 2017a, 2017b). In this sense, the obtained MEG results already diverge from the fMRI results because they are priorly restricted to a frequency range.

As depicted in models of word retrieval, the process of word production undergoes several stages before the word can be produced (Roelofs, 1992). In the present study, the concept of the target word is either presented by sentence context or picture presentation. In constrained trials, participants retrieve the concept based on the information that is given in the sentence context. In unconstrained trials, the concept retrieval depends on identifying the object that is presented in the picture. 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. Relating this to the results of the present study suggests that fMRI better reflects the motor preparation stages of word planning, but not necessarily the early stages of conceptual retrieval. In line with the reasons for diverging findings discussed above, this indicates that the early stages of concept retrieval might occur too fast for a measurable BOLD response to establish. MEG, in contrast, captures the computations underling the early stages of concept retrieval.

The brain regions that seem to be most crucial for concept and word retrieval in contextually constrained sentences in both experiments are the left temporal and inferior parietal lobe. These regions were commonly activated across sessions and captured by both methods in the experiments. The inferior temporal gyrus has previously been shown to be a crucial area to access lexical semantic concepts in object recognition and word production (Price, 2012; Roelofs, 2014). Additionally, the inferior parietal lobe has been argued to be essential in predicting and integrating semantic knowledge (Binder, Desai, Graves & Conant, 2009; Price, 2012). This further corresponds to the findings of Piai et al. (2017b) that established a causal link between these brain regions and context-driven word retrieval. In this study, left-hemisphere stroke patients performed the same task that was employed in the present study. Patients whose left middle and superior temporal gyri and inferior parietal lobe were damaged by the lesion did not show a behavioral effect of sentence context. In addition, their EEG revealed a reduced alpha-beta power decrease. Thus, the results of the preset study and the fact that the left temporal and inferior parietal lobe were captured with MEG as well as fMRI support the claim that these brain regions are essential for context-dependent word retrieval.

To further investigate the divergence between the fMRI and the MEG results, the MEG data of the present study could be analyzed without a priorly specified frequency range and different threshold. This would reveal whether the activity captured with fMRI but not with MEG was not robust enough for the analysis strategies of the present study, or whether this absence is due to undetectable neuronal activity for MEG, as discussed above. However, as the patient study serves to investigate language processing rather than speech preparation, the motor activity is not a major aspect of interest. Especially since the aim is to assess whether brain activity lateralizes from the left to the right hemisphere, it is more suitable to employ a paradigm that only elicits left-lateralized effects in neurotypical control subjects.

When comparing the results for both experiments with one another, the across-session familiarization seems to have a reversed effect across experiments. While MEG results show more and stronger power decreases in session 2 than in session 1, fMRI results show less BOLD increases in session 2 than in session 1. As a next step to further investigate this finding, the data of the present study should be analyzed on the subject-level. This will reveal whether this familiarization effect is consistent across participants, or else only over-represented in some of the participants. Alternatively, this could also be

examined by including a potential third session of the same paradigm for both experiments. This would allow to investigate the activity captured at another time point following session 2 and help to determine the direction of this potential familiarization effect.

Finally, the Dice coefficients of both experiments are higher for the full sessions and decrease when the calculation only includes the first half of the sessions. Notably, the full MEG session overlap yields the highest Dice coefficient, even though the calculation is based on an average of 22 trials less than the full fMRI session overlap. For the half session Dice coefficient the number of included trials is equal, but MEG still yields a higher coefficient than fMRI. Overall, this indicates that MEG trials result in a higher Dice coefficient than fMRI trials, which is further supported by the fact that the half MEG session overlap reveals the highest ratio of Dice coefficient and average number of trials per participant that the calculation is based on. As a further analysis step with the existing data of the present study, the individual Dice coefficients for the activation overlap from session 1 to session 2 for each participant should be calculated. As the patient study will be investigating individual cases with different lesion size and location per patient, the process of recovery will also be evaluated on an individual level. Thus, Dice coefficients on subject-level are ultimately more informative for the purpose of establishing the clinical tool than group-level Dice coefficients.

Conclusion

Recalling the main goal of the project to develop a clinical tool to investigate language recovery in patients, the results of the present study serve to determine the optimal parameters for this purpose. The aim was to find a suitable imaging method that provides spatially reliable effect profiles over time that can be measured within short testing sessions. The obtained results reveal a different focus of the reflected aspect of neuronal activity per method. Particularly, they vary in terms of spatiality and significance for short and long sessions. Also, the Dice coefficients reflect different relationships of consistency per included trials for both methods. By means of the performed analyses of the present study, the results convincingly determine MEG to be a- more suitable method for the clinical tool than fMRI. This does not derive from a direct statistical comparison between the two methods, but rather from a more practical origin. Aspects such as the laterality of the effect or the duration of the testing session are crucial points that should not be neglected when it comes to the development of a clinical tool for patient testing.

Acknowledgements

I want to thank everyone who supported me and my project in the past year in any way, including all members of the *Language Function and Dysfunction* group and the students in the trainee room. A special thanks to Nathalie Meyer for helpful discussions and reviewing my thesis, as well as Mauricio Saldivar for helping me with programming and the layout. Further, thanks to Atsuko Takashima for all her support and guidance for the fMRI part of the project. Finally, I want to thank my supervisor Vitória Piai for giving me the opportunity to work on such a great project during my master internship, as well as for everything I have learned from her in the past year and all the support, suggestions, and guidance I received.

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.

Boersma, P., & Weenink, D. (2017). Praat, software for speech analysis and synthesis.

Bradshaw, A. R., Thompson, P. A., Wilson, A. C., Bishop, D. V., & Woodhead, Z. V. (2017). Measuring
language lateralisation with different language tasks: a systematic review. PeerJ, 5, e3929.

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.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences. 1988, hillsdale, nj: L. Lawrence Earlbaum Associates, 2.

Conner, C. R., Ellmore, T. M., Pieters, T. A., DiSano, M. A., & Tandon, N. (2011). Variability of the relationship between electrophysiology and bold-fmri across cortical regions in humans. Journal of Neuroscience, 31(36), 12855–12865.

Dronkers, N. F., & Baldo, J. (2010). Language: aphasia. In

Encyclopedia of neuroscience. Elsevier Ltd.

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.

Findlay, A. M., Ambrose, J. B., Cahn-Weiner, D. A., Houde, J. F., Honma, S., Hinkley, L. B., … Kirsch, H. E. (2012). Dynamics of hemispheric dominance for language assessed by magnetoencephalographic imaging. Annals of neurology, 71(5), 668–686.

Hillebrand, A., Nissen, I., Ris-Hilgersom, I., Sijsma, N., Ronner, H., Van Dijk, B., & Stam, C. (2016). Detecting epileptiform activity from deeper brain regions in spatially filtered meg data. Clinical Neurophysiology, 127(8), 2766–2769.

Kim, M. 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.

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.

Liljeström, M., Hulten, A., Parkkonen, L., & Salmelin, R. (2009). Comparing meg and fmri views to naming actions and objects. Human brain mapping, 30(6), 1845–1856.

Liljeström, M., Stevenson, C., Kujala, J., & Salmelin, R. (2015). Task-and stimulus-related cortical networks in language production: Exploring similarity of meg-and fmri-derived functional connectivity. Neuroimage, 120, 75–87.

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.

Piai, V., Meyer, L., Dronkers, N. F., & Knight, R. T. (2017a). Neuroplasticity of language in left-hemisphere stroke: Evidence linking subsecond electrophysiology and structural connections. Human brain mapping, 38(6), 3151–3162.

Piai, V., Roelofs, A., & Maris, E. (2014). Oscillatory brain responses in spoken word production reflect lexical frequency and sentential constraint. Neuropsychologia, 53, 146–156.

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.

Piai, V., Rommers, J., & Knight, R. T. (2017b). 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.

Price, C. J. (2012). A review and synthesis of the first 20 years of pet and fmri studies of heard speech, spoken language and reading. Neuroimage, 62(2), 816–847.

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. Cognition, 42(1-3), 107–142.

Roelofs, A. (2014). A dorsal-pathway account of aphasic language production: The weaver++/arc model. Cortex, 59, 33–48.

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.

Stolk, A., Todorovic, A., Schoffelen, J.-M., & Oostenveld, R. (2013). Online and offline tools for head movement compensation in meg. Neuroimage, 68, 39–48.

Team, R. C. (2017). R: A language and environment for statistical computing. vienna, austria: R foundation for statistical computing; 2017. ISBN3-900051-07-0 https://www. Rproject.org.

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.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., … Joliot, M. (2002). Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri singlesubject brain. Neuroimage, 15(1), 273–289.

van Casteren, M., & Davis, M. H. (2006). Mix, a program for pseudorandomization. Behavior research methods, 38(4), 584–589.

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.

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.

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.