# Functional Identification of Language-Responsive Channels in Individual Participants in MEG Investigations

[1]Mathias Huybrechts*, [1,2,3]Rose Bruffaerts*, [3,4]Alvince Pongos, [3,5]Cory Shain, [3]Benjamin Lipkin, [3,6]Matthew Siegelman, [7,8]Vincent Wens, [9,10]Martin Sjøgård, [3,11]Idan Blank, [7,8]Serge Goldman, [7,8]Xavier De Tiège and [3,12]Evelina Fedorenko

[1]Computational Neurology, Experimental Neurobiology Unit (ENU), Department of Biomedical Sciences, University of Antwerp, Antwerp, Belgium;

[2]Department of Neurosciences, KU Leuven, Leuven, Belgium;

[3]Brain and Cognitive Sciences Department, Massachusetts Institute of Technology, Cambridge, MA, USA;

[4]Department of Bioengineering, UC Berkeley-UCSF, San Francisco, CA, USA;

[5]Laboratory for Computation & Language in Minds & Brains, Stanford University, Stanford, CA, USA;

[6]Department of Psychology, Columbia University, New York, NY, USA;

[7]Laboratoire de Neuroanatomie et Neuroimagerie translationnelles (LN2T), ULB Neuroscience Institute, Université libre de Bruxelles (ULB), Brussels, Belgium;

[8]Service of Translational Neuroimaging, HUB - Erasme hospital, Brussels, Belgium;

[9]Department of Psychiatry, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA

[10]Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital and Harvard Medical School, Charlestown, MA, USA

[11]Department of Psychology, University of California Los Angeles, CA, USA.

[12]McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, MA, USA;

*shared first authorship

Corresponding author: Rose Bruffaerts (rose.bruffaerts@uantwerpen.be), Computational Neurology, Experimental Neurobiology Unit, Department of Biomedical Sciences, University of Antwerp, Antwerp, Belgium, and Evelina Fedorenko (evelina9@mit.edu), Brain and Cognitive Sciences Department, Massachusetts Institute of Technology, Cambridge, MA, USA

CRediT contributions:

Conceptualization: RB and EF

Methodology: MH, RB, AP, MSi, VW, IB and EF

Software: MH, RB, AP, MSi, CS

Investigation: MH, RB, AP, CS, BL, MSi, VW, MSj, IB, SG, XDT, EF

Data curation: MH, RB, MSi, AP, BL

Formal analysis and validation: MH, RB

Visualization: MH, RB

Writing-original draft: MH, RB, EF

Writing-review and editing: CS, SG, MSj, IB, VW

Supervision and project administration: RB and EF

# Abstract

Making meaningful inferences about the functional architecture of the language system requires the ability to refer to the same neural units across individuals and studies. Traditional brain imaging approaches align and average brains together in a common space. However, lateral frontal and temporal cortices, where the language system resides, is characterized by high structural and functional inter-individual variability, which reduces the sensitivity and functional resolution of group-averaging analyses. This issue is compounded by the fact that language areas lay in close proximity to regions of other large-scale networks with different functional profiles. A solution inspired by visual neuroscience is to identify language areas functionally in each individual brain using a 'localizer' task (e.g., a language comprehension task). This approach has proven productive in fMRI, yielding a number of robust and replicable findings about the language system. Here, we extend this approach to MEG. Across two experiments (one in Dutch speakers, n=19; one in English speakers, n=23), we examined neural responses to the processing of sentences and a control condition (nonword sequences). In both the time and frequency domains, we demonstrated that the topography of neural responses to language is spatially stable within individuals but varies across individuals. Consequently, analyses that take this inter-individual variability into account are characterized by greater sensitivity, compared to the group-level analyses. In summary, similar to fMRI, functional identification within individuals yields benefits in MEG, thus opening the door to future investigations of language processing including questions where whole-brain coverage and temporal resolution are both critical.

# 1. Introduction

The functional architecture of the human language network is broadly consistent across individuals (Lipkin et al., 2022). However, the precise topography of this network varies substantially even within homogenous groups of neurotypical adults (Fedorenko et al., 2010). Developing research methods that take these inter-individual differences into account by identifying functional areas in individual participants—an approach known as 'functional localization'—has proven vital in cognitive neuroscience across domains and has been shown to increase sensitivity, functional resolution, accurate effect size estimation, and interpretability (Saxe et al., 2006; Nieto-Castañón & Fedorenko, 2012; Fedorenko, 2021; see Brett et al., 2002 for an early discussion of these issues). So far, functional localizer paradigms have been mostly used in fMRI research (Kanwisher et al., 1997; Saxe & Kanwisher, 2003; Fedorenko et al., 2010; Baldauf & Desimone, 2014), although a handful of studies have applied a similar approach in other recording modalities, including electrocorticography (ECoG) (Cogan et al., 2014; Fedorenko et al., 2016; Regev et al., 2024) and functional near-infrared spectroscopy (fNIRS) (Powell et al., 2018; Y. Liu et al., 2022; Paranawithana et al., 2024). With respect to magnetoencephalography (MEG), functional localization has predominantly been applied in studies of visual processing (J. Liu et al., 2002; De Vries & Baldauf, 2019), but has also shown potential in the language domain with for instance applications in presurgical mapping and inferring the role of different neural frequency bands during language comprehension (Lam et al., 2016; Prystauka & Lewis, 2019; Papanicolaou, 2023).

Numerous studies over the last decade have provided evidence that the language network can be delineated in a robust and replicable way at the individual-subject level using a contrast between language comprehension and a perceptually matched control condition in fMRI (Fedorenko et al., 2010; Mahowald & Fedorenko, 2016; Braga et al., 2020; Fedorenko et al.,

2024; Lee et al., 2024) (see (Lipkin et al., 2022) for data in > 800 individuals). Importantly, the language localizer contrast (language > perceptually matched control condition) has been shown to be robust to input modality (written, spoken, or signed) (Fedorenko et al., 2010; Scott et al., 2017; Richardson et al., 2020; Lee et al., 2024), stimulus content (hand-crafted sentences, sentences extracted from a corpus, or connected passages (Scott et al., 2017)), language (Malik-Moraleda, Ayyash et al., 2022), and the presence or absence of an active task (Fedorenko et al., 2010; Diachek et al., 2020; Ivanova et al., 2020). Furthermore, this network has been shown to be strongly selective for linguistic input (Fedorenko et al., 2011; Monti et al., 2012; Amalric & Dehaene, 2018) (see e.g., (Fedorenko & Blank, 2020) for a review). This broad generalization across paradigm variations and selectivity for language processing jointly suggest that the language brain areas support specifically linguistic computations. Indeed, evidence from dozens of studies has implicated these areas in lexical access, syntactic structure building, and semantic composition during both comprehension and production (e.g. Pallier et al., 2011; Hu, Small et al., 2023; Shain, Kean et al., 2022, 2024; Giglio et al., 2024).

Although fMRI investigations of the language network have yielded important findings, fMRI's poor temporal resolution limits its use for research questions where timing information is critical. Intracranial recordings provide an incredible opportunity to obtain high-spatial and high-temporal resolution data with high signal-to-noise ratio, but the approach is inherently limited with respect to both the population and the sparse brain coverage. In contrast, magnetoencephalography (MEG) enables whole-brain non-invasive measurements of neural activity in typical brains at a millisecond-level resolution. Here, using data from two independent datasets (across two languages: English and Dutch), we establish the feasibility of identifying language-responsive sensors at the individual level in MEG investigations, and provide evidence that this approach is more sensitive than the standard brain-averaging approach.

# 2. Methods

## 2.1 Participants

We recruited 42 healthy young volunteers: 19 native Dutch speakers (18 female, between 19 and 29 years old, mean 23.4 years old) and 23 native English speakers (between 19 and 53 years old, mean 26.7 years old). The study was approved by the local Ethics Committees. All participants provided written informed consent in accordance with the Declaration of Helsinki.

## 2.2 Experimental Design

In identifying the language network, we use a design that has been extensively validated in fMRI (first used in (Fedorenko et al., 2010)), namely, a sentence reading task. Using fMRI, the sentences>nonwords sequences contrast has been shown to robustly and reliably identify the language-selective network (Lipkin et al., 2022; Malik-Moraleda, Ayyash et al., 2022). Participants performed the experiment in their native language. Materials can be downloaded from https://osf.io/vc2bw/. For the sentence condition, 80 12-word-long sentences were constructed in English using a variety of syntactic structures and covering a wide range of topics. The sentences were translated into Dutch, with minimal changes to obtain 12-word-long sentences. For the nonwords condition, care was taken to minimize low-level differences in the phonological make-up of the stimuli compared to the sentence condition. In English, the content words (noun, verb, adjective, adverb) of the sentence condition were syllabified to create a set of syllables that could be re-combined in new ways to create pronounceable nonwords. For syllables that formed real words of English, a single phoneme was replaced (respecting the phonotactic constraints of English) to turn the syllable into a nonword. The syllables were then recombined to create nonwords matched for length (in syllables) with sentence condition. In Dutch, nonwords were derived from the content words in the sentence

condition by means of the pseudoword generator Wuggy (https://github.com/crr-ugent/wuggy). This program generates nonwords that match the original word in subsyllabic structure and respects the transition frequencies specific to Dutch (Keuleers & Brysbaert, 2010).

To help participants stay engaged, we used a memory probe version of the language localizer task in the Dutch version, where participants are asked to decide (by pressing one of two buttons) whether a word or nonword, presented at the end of each sentence or nonword list, was in the preceding trial. Probes were restricted to content words in the sentence condition and nonwords in the nonwords condition. Half of the trials required a positive response. In the English version, a button-press icon was presented at the end of each sentence or nonword list where participants are asked to press a button. This difference between the Dutch and English version was introduced to reduce the duration of the English experiment by several minutes, in order to combine the paradigm with other, unrelated studies. As noted above, previous fMRI work demonstrated that the sentences>nonwords localizer contrast is robust to different tasks including passive viewing (Fedorenko et al., 2010; Diachek et al., 2020).

The words/nonwords were presented one at a time in a rapid serial visual presentation paradigm at a fixed rate per word/nonword (385 ms per (non)word in Dutch, 400 ms in English). Each word/nonword was presented in the center of the screen in capital letters without punctuation and appeared immediately after the previous (non)word without a blank transient screen. Then, at the end of the trial, the memory probe or button-press icon was presented, followed by a variable interval, between 0.5s and 2.25s. Participants could respond any time after the probe or button-press icon appeared and until the next trial started. The experiment lasted ~20 minutes. Five participants in the English dataset were scanned using a similar paradigm but with a different set of materials (8-word-long sentences and 8-nonword-long sequences). Previous fMRI work has shown that the sentences>nonwords localizer contrast is robust to such variation (Fedorenko et al., 2010).

## 2.3 MEG Data Acquisition

Continuous MEG data were recorded using a whole-head 306 channel (102 magnetometers, 204 planar gradiometers) TRIUX system (MEGIN, Espoo, Finland) either at the HUB - Hôpital Erasme (Brussels, Belgium) or at the Martinos Imaging Center (McGovern Institute for Brain Research at MIT, Cambridge, MA, USA). Participants were tested in an upright seated position (68° recline). Four head-position indicator (HPI) coils were used to record the head position within the MEG helmet every 200 ms. The participant's head shape was digitally recorded by means of a 3D digitizer (Fastrak Polhemus, Inc., Colchester, VA, USA) along with the position of the HPI coils and fiducial points (nasion, left and right periauricular). MEG signals were recorded at a sampling rate of 1000 Hz with on-line band filter between 0.1 and 330 Hz.

## 2.4 MEG Data Processing

Initial preprocessing of the raw data used MaxFilter version 2.2 (MEGIN, Espoo, Finland): temporal signal space separation (Taulu et al., 2003) was applied to remove noise from external sources and from HPI coils for continuous head-motion correction (correlation threshold: 0.98, 10 s sliding window), and to virtually transform data to a standardized head position. The latter facilitates comparison across experiments. MaxFilter was used to automatically detect and virtually reconstruct noisy channels. Further preprocessing was performed using MNE-Python version 1.8.0 (Gramfort, 2013). Preprocessing consisted of high-pass filtering at 0.1 Hz, low-pass filtering at 300 Hz and notch filtering at 50Hz or 60Hz (with respective higher harmonics) depending on the line frequency. Data were downsampled to 500 Hz and epoched at the level of the sentence or nonword list (500ms before onset of the first (non)word until 400 ms after onset of the last (non)word). Visual inspection of all epochs was performed, and epochs with clear artefacts were marked as bad and excluded from further analysis. Denoising was performed by visually identifying and removing ICA components consistent with eye blinks

and cardiac artefacts from a total of 60 components extracted using fastICA (Hyvarinen, 1999) with default parameters on the concatenated epochs (excluding bad epochs).

A quantitative marker of noise was derived per participant by first averaging the baseline intervals across all sensors for each trial and then calculating the standard deviation of these trial-averaged values across all trials. Here and later in the study, the baseline is defined as the neutral interval of 500 ms prior to trial onset when only a fixation cross is presented. Participants with markers of noise greater than two standard deviations from the average noise marker across participants scanned in the same MEG device were removed from further analysis. This procedure led to the removal of one participant from the English dataset. Two additional participants were removed from the English dataset: one due to excessive motion and one due to the presence of high-amplitude "mu rhythm" apparent upon visual inspection.

Language processing related to the presentation of the word or nonword stimuli is reflected in both the time and frequency domain of the recorded MEG signal (Beres, 2017; Prystauka & Lewis, 2019; Coolen et al., 2024). Analyses that focus on the time domain and rely on event-related potentials have reliably shown effects related to prediction and integration of lexical information during language comprehension (Sun & Luo, 2024). However, such analyses only capture a subset of changes in the signal associated with the presentation of the stimuli. Specifically, oscillatory activity that is not strictly time-or phase-locked to the stimulus will be averaged out in this analytic approach (Maguire & Abel, 2013; Morales & Bowers, 2022). To this end, we analyzed the epochs in the time and frequency domains separately. The spectral analysis offers an alternative, complementary perspective to the time domain analysis for understanding the neural computations associated with language comprehension (Lam et al., 2016).

In the time domain analysis, we grouped each pair of planar gradiometers into a single effective gradiometer derived as their Euclidean norm (hence measuring the amplitude of the field's tangential gradient, independently of orientation) (Chetail et al., 2018). To facilitate between-participant comparisons, we then calculated the percent signal change at every timepoint in the sentence and nonword conditions for every trial compared to the mean amplitude of its baseline. We omitted the first 4 (non)words in each trial (n=80 per condition) from the analysis in order to focus on the part of the trial where the between-condition differences might be most pronounced (given what is known about the processes related to sentence meaning construction (Pallier et al., 2011; Fedorenko et al., 2016)), although this choice proved not to impact the results in the end (see discussion). Finally, we averaged the percent signal change values across all included word/nonword positions (i.e., the 5th (non)word and all subsequent ones) within a condition to derive a single "amplitude-based" value per channel per participant.

In the spectral analysis, we adopted a similar approach: we compared the bandlimited power of the sentence and nonword trials in the 204 planar gradiometers to that of the baseline (as percent power change) for each neural frequency band. We used five bands: $\theta$ (3-8Hz), $\alpha$ (8-13Hz), $\beta$ (13-30Hz), $\gamma_{low}$ (30-60Hz) and $\gamma_{high}$ (60-90Hz). We divided the $\gamma$-band into a lower and higher frequency component given that previous studies have demonstrated distinctions between these sub-bands or linked effects specifically to one of them (Towle et al., 2008; Fedorenko et al., 2016; Lam et al., 2016; Hashimoto et al., 2017). First, we computed the frequency estimates of the sentence trial epochs and nonword trial epochs (from the 5th (non)word onwards, as in the time domain analyses) and their respective baselines using multitapers with 7 Discrete Prolate Spheroidal Sequences (DPSS). Then, we calculated the percent signal change in the power estimates of the sentence and nonword conditions for every trial compared to the baseline interval for that trial. Finally, we averaged all percent signal

change values within each condition, which yielded a single “power-based” percent change value for each channel per participant per frequency band.

## 2.5 Statistical analysis

To test whether MEG allows for a robust identification of language-responsive sensors at the individual participant level, we evaluated the stability of the language responses within individuals across time (i.e., between the odd- vs. even-numbered trials). To do so, we performed three analyses. First, we examined the Spearman correlation in the size of the sentence effect (percent signal change for the sentence condition relative to the baseline) over all channels within each participant across odd- and even-numbered trials and compared this to the correlations between different participants (Wilcoxon rank sum test). Spearman instead of Pearson correlation was used because it is more robust to potential outliers (Rousselet & Pernet, 2012). This analysis asks: if a sensor shows a strong response to sentences in one half of the data, does it also show a strong response to sentences in the other half of the data in the same participant compared to another participant?

Second, we used the data from the odd-numbered trials to define sensors of interest (SOIs) in each participant. SOIs were defined as the 10% of sensors with the highest increase in percentage signal change in the sentence condition relative to the baseline. We then examined the effect size for the sentence condition in the even-numbered trials in these SOIs relative to the baseline and to the effect size for the nonword condition using a signed rank test.

Third, to test whether identification of language-responsive sensors at the individual level is superior to identifying language-responsive sensors at the group level, we compared the results of the second analysis above to a version of the analysis where we performed the same calculation using group-averaged maps (where the responses are averaged across participants from either the English or Dutch dataset in each sensor). In particular, we used the data from

the odd-numbered trials to create a group map and defined SOIs as the 10% of sensors with the highest increases in percent signal change in the sentence condition relative to the baseline. We then examined the effect size for the sentence condition in the even-numbered trials in these SOIs for each participant. Critically, unlike in the individual-level analyses, the SOIs in the group analysis are the same for all participants. To compare the effect sizes in SOIs defined individually vs. based on a group-level map, we used a signed rank test.

In all analyses, we employed an exploratory-confirmatory approach to control for false positives. The English dataset served as the exploratory sample, where we identified key neural effects of interest. Then, we tested whether these effects could be replicated in the Dutch dataset using the same statistical analysis. Only results that met predefined confirmation criteria – statistical significance ($p<0.05$), consistent directionality, and comparable effect size – were considered robust. This approach reduces the likelihood of false positives while enhancing the generalizability of our approach.

# 3. Results

## 3.1 The topography of neural responses to language is stable within but varies between individuals

In the English dataset, the average correlation of the amplitude-based sentence effect sizes across all channels within each participant across odd- and even-numbered trials was 0.61 (s.d. 0.22). The mean correlation between pairs of participants was significantly lower compared to the within-participant correlation (mean rho: 0.16, s.d. 0.24; $P < 0.0001$; Figure 1A). This finding was confirmed in the Dutch dataset as the average within-participant correlation (0.63 s.d. 0.23) was also significantly higher than the average between-participant correlation (0.25 s.d. 0.26; $P<0.0001$; Figure 1B). Furthermore, the effect was comparable in size and

directionality to that of the English dataset, aligning with the predefined confirmation criteria of the exploratory-confirmatory approach.

In the frequency domain analysis, the average within-participant correlations of the power-based percent changes were significantly higher than the average between-participant correlations (P<0.0001 for all frequency bands apart from high gamma where P<0.001;Figure 1C,D). For each frequency band, similar effect sizes where obtained in the English and Dutch datasets (Table 1).

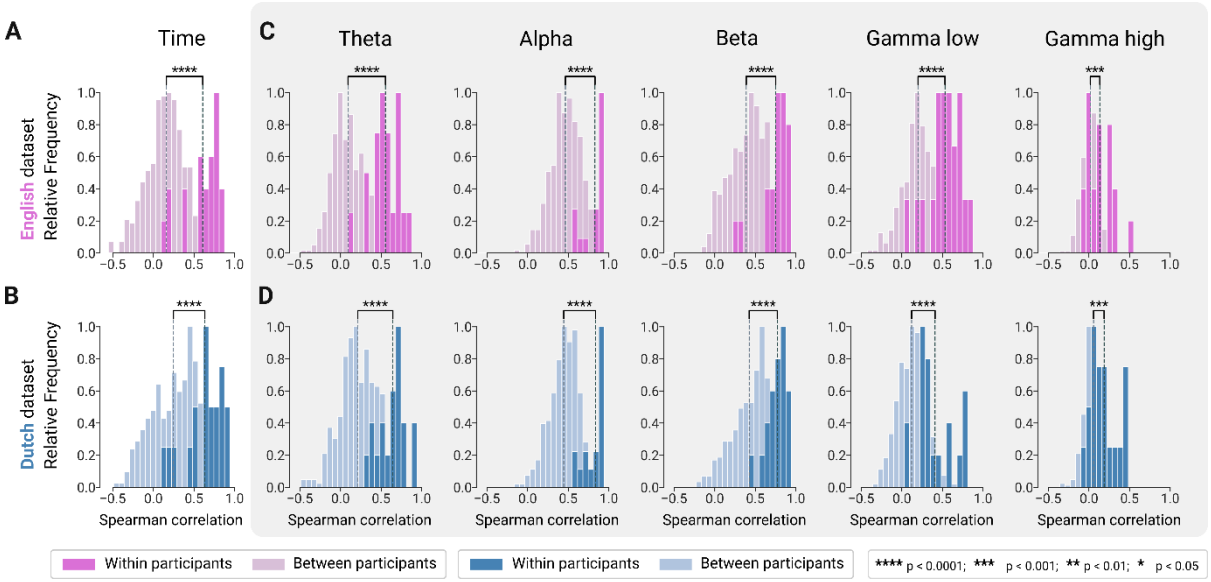



***Figure 1**: Distribution of the correlation (Spearman) values for the size of the sentence>baseline effect across channels within individuals across odd- and even-numbered trials (darker shades) and between data halves taken from different individuals (lighter shades) in the A) English and B) Dutch datasets in the time domain analysis. Similar distributions for the spectral analysis are represented for the C) English and D) Dutch datasets for each of the neural frequency bands. Relative frequency counts were plotted because of the different number of possible combinations between vs within individuals. Vertical striped lines indicate the mean of each distribution.*

***Table 1**: Within and between participant correlation values in the English and Dutch datasets for the time domain and each of the neural frequency bands.*

| ***English*** | Time | Theta | Alpha | Beta | Gamma low | Gamma high |
|---|---|---|---|---|---|---|
| Within correlations | 0.61 (s.d. 0.22) | 0.55 (s.d. 0.17) | 0.83 (s.d. 0.12) | 0.75 (s.d. 0.17) | 0.53 (s.d. 0.20) | 0.13 (s.d. 0.14) |
| Between correlations | 0.16 (s.d. 0.24) | 0.09 (s.d. 0.19) | 0.46 (s.d. 0.18) | 0.39 (s.d. 0.21) | 0.20 (s.d. 0.19) | 0.02 (s.d. 0.10) |
| P value | P < 0.0001 | P < 0.0001 | P < 0.0001 | P < 0.0001 | P < 0.0001 | P < 0.001 |
| ***Dutch*** | Time | Theta | Alpha | Beta | Gamma low | Gamma high |
| Within correlations | 0. 63 (s.d. 0.23) | 0.64 (s.d. 0.16) | 0.84 (s.d. 0.12) | 0.77 (s.d. 0.12) | 0.41 (s.d. 0.22) | 0.19 (s.d. 0.16) |
| Between correlations | 0.25 (s.d. 0.26) | 0.21 (s.d. 0.22) | 0.44 (s.d. 0.17) | 0.42 (s.d. 0.21) | 0.12 (s.d. 0.17) | 0.06 (s.d. 0.11) |
| P value | P < 0.0001 | P < 0.0001 | P < 0.0001 | P < 0.0001 | P < 0.0001 | P < 0.001 |

Inspection confirmed the similarity between the topographies of the signal changes during the sentence condition in the odd- and even-numbered trials at the individual level (Figure 2). When assessing the topography of the language-responsive SOIs that were selected in different participants, inter-individual differences were notable: although some channels were consistently selected in a substantial fraction of participants, other channels were only selected in a small fraction of individuals. In the English dataset, certain sensors were consistently selected in up to 45%, or 9 of the 20 participants in the time domain and up to 75% or 15 of 20 participants in the alpha band. Similarly, in the Dutch dataset, some sensors were consistently selected in up to 42%, or 8 of the 19 participants in the time domain and up to 68% or 13 of 19 participants in the alpha band.

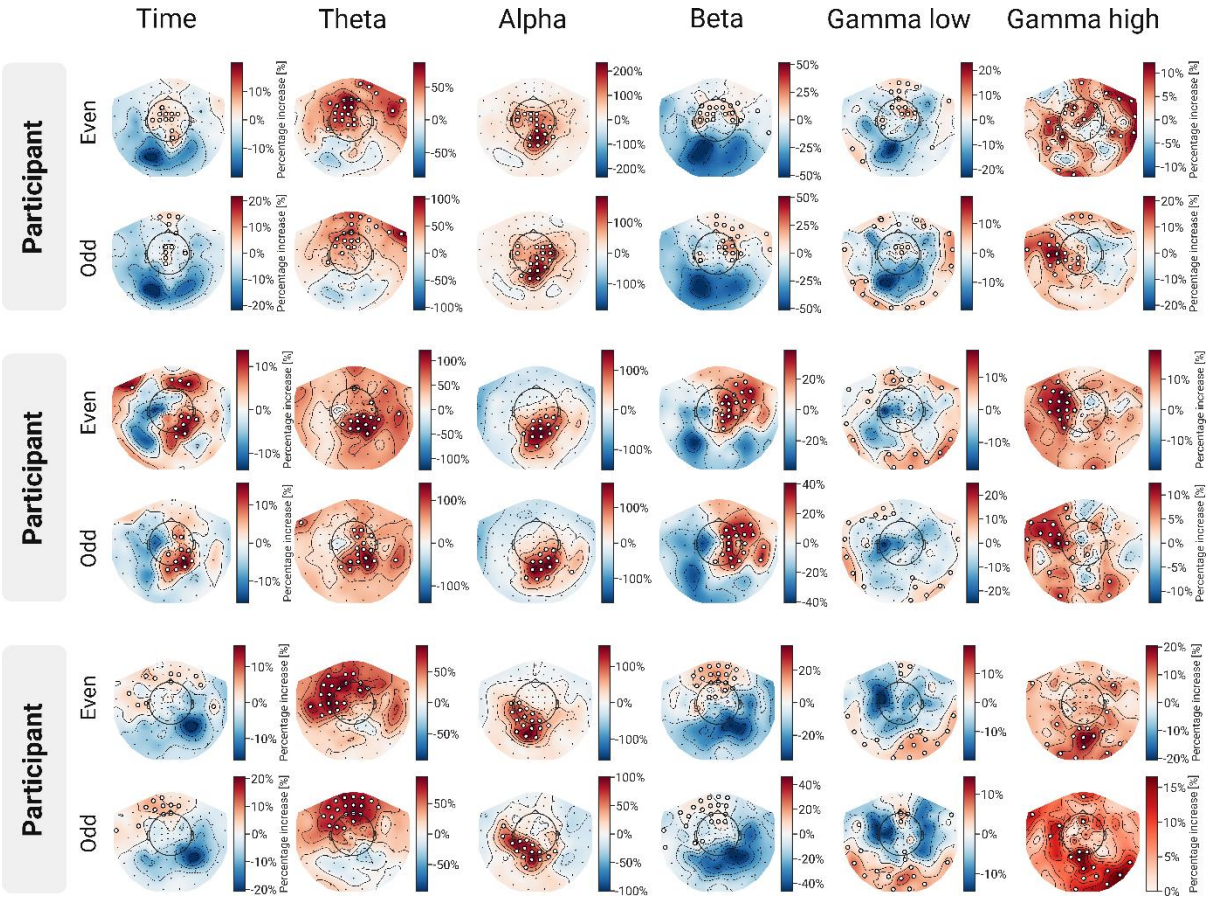



***Figure 2***: *Topographies of the effect sizes (percentage change) during the sentence condition in the odd- and even-numbered trials in time domain and frequency domain in sample participants of the English (top two participants) and Dutch dataset (bottom participant) with selected SOIs indicated by white dots.*

## 3.2 Analyses based on individual-level sensors of interest yield more robust responses compared to the group-level analysis

The amplitude-based effect size for the sentence condition in the even-numbered trials with individually defined SOIs (based on the odd-numbered trials) in the English dataset was 3.17% (s.d. 4.04%) signal increase compared to baseline (Figure 3A) and was reliably greater than the effect size for the nonwords condition (P = 0.024). At the level of individual participants, 16 of the 20 showed a sentence > nonwords effect. This was further confirmed in the Dutch dataset where the effect size for the sentence condition in the individually defined SOIs was 5.54%

(s.d. 6.04%) signal increase compared to baseline in the even-numbered trials (Figure 3B) and was reliably greater than the effect size for the nonwords condition (P=0.026). At the level of individual participants, 14 of the 19 showed a sentence > nonwords effect (Table 2).

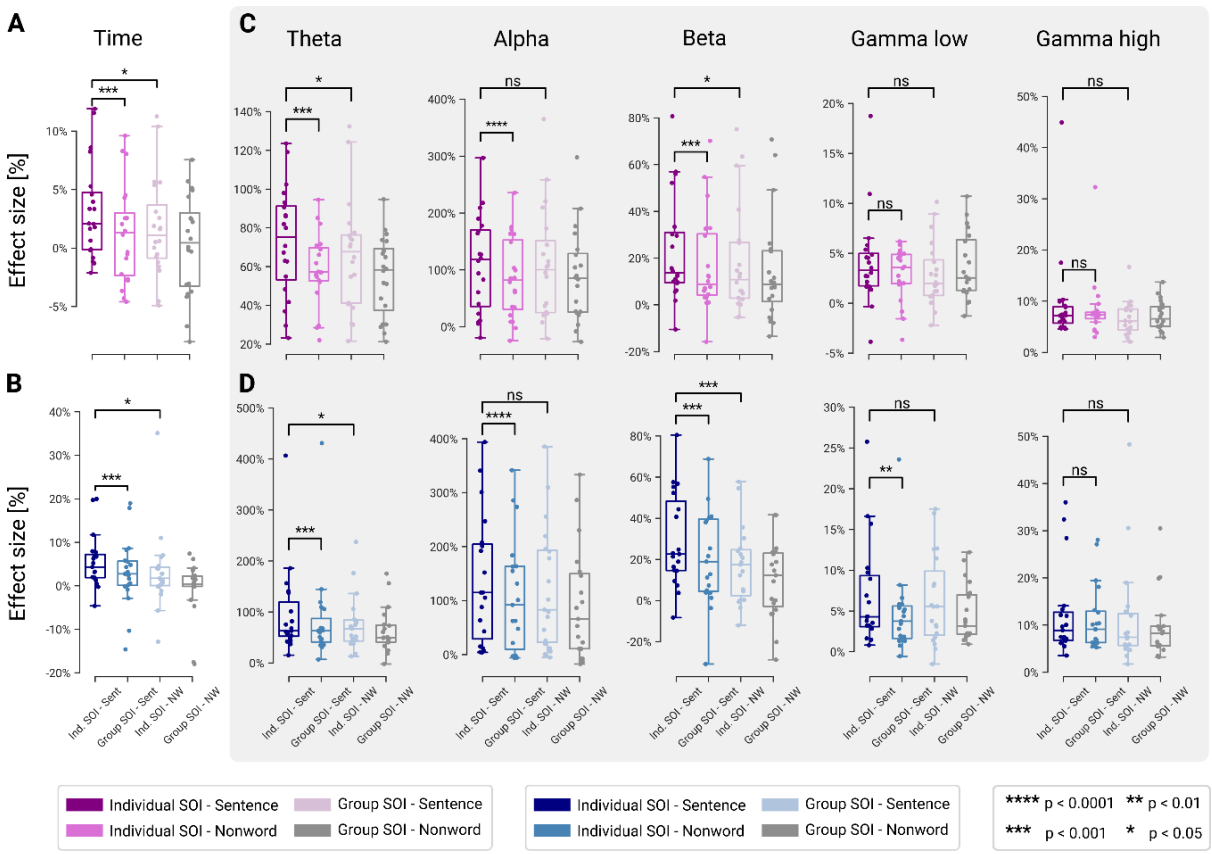



*Figure 3: Mean effect sizes (percent signal change) for the sentence and nonwords conditions when the language-responsive sensors of interest (SOIs) are defined at the individual level vs. at the group level in the A) English and B) Dutch datasets for the time domain analysis. Similar results depicted for each of the neural frequency bands in the frequency domain analysis in the C) English and D) Dutch dataset. In all cases, the SOIs are defined using one half of the data (odd-numbered trials) and the response magnitudes are examined in the other half of the data (even-numbered trials).*

In the frequency domain, in the theta and beta bands, power-based signal increases of 72.95% and 23.12% respectively were observed with the individually defined SOIs in the English dataset (Figure 3C). Both were reliably greater than their nonword conditions (P=0.015 for theta and P=0.014 for beta). At the level of individual participants, 16 out of 20 exhibited a

greater theta power response to sentences compared to nonword lists, and 14 out of 20 showed an increase in beta power for the same contrast. These findings were replicated in the Dutch dataset where for the theta and beta bands, the increases of the sentence condition over baseline were 96.66% and 29.26%. respectively. Both were reliably greater than the nonword condition with P=0.012 for the theta band and P<0.001 for the beta band. Other frequency bands also showed increased levels of power for the sentence condition compared to baseline but did not demonstrate a reliably greater effect for sentences than for nonwords (Table 2).

***Table 2:*** *(Left) Proportion of participants showing an increase in effect size for the sentence condition over the nonword condition with individually defined SOI (Sentence > Nonword). (Right) The proportion of participants showing an increase in effect size in the sentence condition when SOIs are defined at the individual level versus when SOIs are compared at the group level (Individual SOI > Group SOI). Values indicate the number of significant participants (n) out of the total, with corresponding p-values for English and Dutch datasets.*

| | **Individual SOI: Sentence > Nonword** | | | | **Individual SOI > Group SOI** | | | |
|---|---|---|---|---|---|---|---|---|
| | English | | Dutch | | English | | Dutch | |
| | n | p-val | n | p-val | n | p-val | n | p-val |
| Time | 16/20 | 0.024 | 14/19 | 0.026 | 18/20 | < 0.001 | 16/19 | < 0.001 |
| Theta | 16/20 | 0.015 | 16/19 | 0.012 | 16/20 | < 0.001 | 17/19 | < 0.001 |
| Alpha | 12/20 | 0.674 | 14/19 | 0.087 | 19/20 | < 0.0001 | 19/19 | < 0.0001 |
| Beta | 14/20 | 0.014 | 16/19 | < 0.001 | 17/20 | < 0.001 | 15/19 | < 0.001 |
| Gamma Low | 11/20 | 0.294 | 12/19 | 0.441 | 9/20 | 0.596 | 13/19 | 0.006 |
| Gamma High | 14/20 | 0.105 | 11/19 | 0.490 | 10/20 | 0.498 | 12/19 | 0.441 |

To test whether identification of SOIs at the individual level is superior to analysis at the group level, we defined a separate set of SOIs based on the group-level map for the odd-numbered trials. With the group-level SOIs, we found that the amplitude-based effect size for the sentence condition in the English dataset was significantly smaller compared to the analyses that take

inter-individual differences into account ($P < 0.001$, Figure 3A). At the level of individual participants, 18 of the 20 participants showed a larger amplitude-based effect size in the individually defined SOIs compared to the group-level SOI. This was confirmed in the Dutch dataset where the same result was observed ($P<0.001$, Figure 3B). At the level of individual participants, 16 of the 19 showed a larger effect size when SOIs are selected at the individual level effect. Similarly, the power in the theta, alpha, and beta bands was significantly lower when inter-individual differences were not taken into account for both the English and Dutch dataset ($P<0.001$, $P<0.0001$, $P<0.001$, Figure 3C,D; Table 2).

# 4. Discussion

In two independent datasets, MEG recordings allowed the identification of language-responsive sensors at the individual-participant level, both in the time and frequency domains. During sentence reading, neural signal changes could be detected within native individual English speakers and reproduced in native Dutch speakers. We showed that in both datasets, the response to this language comprehension task is spatially consistent (over time) within individuals: similar sensors show strong responses during language processing in two halves of the data. Importantly, the response to language is spatially variable across individuals, presumably due to differences in the functional neuroanatomy. The consequence is that analyses that take inter-individual variability into account have greater sensitivity than the traditional group-level analysis. Overall, the results suggest that the functional identification approach, where sensors of interest are defined in individual participants, can yield advantages in MEG, including greater sensitivity, functional resolution, and interpretability.

## 4.1 Individual-level identification of the language network

In line with prior results using the same language localizer task with fMRI (Lipkin et al., 2022) and intracranial recordings (Fedorenko et al., 2016), we observed that analyses that take into account inter-individual variability in the functional neuroanatomy of language-responsive cortex yield higher sensitivity: language responses are reliably higher when the channels are selected at the individual level, compared to the group level. When using a typical fMRI preprocessing and analysis pipeline, voxel-wise neural responses from each participant are warped from the subject space to a common space based on a brain template and functional correspondence is assumed in each voxel. This assumption has long been shown to be flawed, especially when examining cognitive functions supported by the association cortex (Nieto-Castañón & Fedorenko, 2012; Frost & Goebel, 2012; Fedorenko & Blank, 2020). Analysis of resting-state MEG data demonstrated that functional connectivity patterns enable differentiation between different individuals (Da Silva Castanheira et al., 2021) and we here extend the individual-level neural characterization to the language network.

In brain recording approaches, like fNIRS and MEG, where a fixed number of sensors are used, a similar assumption is typically made when data are pooled across participants: that the same sensor is functionally equivalent across individuals. However, because of the inter-individual neuroanatomical differences in cortical thickness, gyrification, and total brain volume, and differences in the placement of the recording channels relative to the brain, the same sensor may capture different underlying sources of neural signals depending on the individual's anatomy. This variability in the sources captured by the same channels across participants introduces noise when signals are averaged across participants and complicates interpretation. potentially leading to incorrect conclusions (e.g., see (Powell et al., 2018) for a discussion of this issue in fNIRS).

Source modelling allows to map neural signals from the sensors to an individual's structural brain scan for both fNIRS and MEG. However, group-level analysis after source modelling is associated with the same limitations as group-level analysis performed in a common space using fMRI: specifically, warping the neural signals across participants to a common space wrongly assumes functional equivalence of the same voxel or vertex across participants (Nieto-Castañón & Fedorenko, 2012; Frost & Goebel, 2012; Fedorenko & Blank, 2020). Functional identification in individual participants provides a powerful alternative solution. This approach allows us to circumvent potential differences in the underlying anatomy and focus on the functional responses. Functional identification at the individual level can be performed in sensor space, as we have done here, or in source space. Logistically, a MEG source modelling approach requires additional data acquisition using MRI. Future research will investigate the use of source modelling combined with our functional localizer approach to localize the language-responsive vertices at the individual level.

We established the feasibility of individual-participant functional identification in MEG using an extensively validated language localizer paradigm. Previous work established the replicability of neural responses evoked by language processing using MEG within the same group of individuals (Roos & Piai, 2020). We demonstrated the topographic stability of the language-responsive channels within individuals over time—the critical foundation of individual-level functional localization. The functional identification approach yields greater sensitivity and functional resolution. In particular, with respect to sensitivity: by grouping the selected SOIs, one statistical test can be performed across the ensemble of sensors, avoiding the need for multiple statistical comparisons (Nieto-Castañón & Fedorenko, 2012). Further, effect sizes are more accurately estimated. This greater sensitivity can be helpful in examining subtle effects in some new, critical conditions and has the potential for application in patients with neurological disease, where neural responses may be overall weaker.

These advantages in sensitivity and functional resolution are afforded by the application of the individual-subject analyses in any study regardless of whether a functional localizer task is included. However, if a validated localizer paradigm is used, the study will additionally benefit from greater interpretability. By using the same paradigm to identify the relevant functional subset of the brain across individuals, studies, imaging modalities, and species (in cases of shared cognitive capacities e.g., face processing), we can a) make stronger inferences about the origins of an effect (e.g., the ability to interpret some critical effect as arising within the language-responsive regions), which cannot be done based on anatomy alone because of the inter-individual functional differences, and b) be generally more confident that we are referring to the same system which is critical for knowledge accumulation and for comparing findings across studies and labs. In this way, the functional localization approach aligns with the field's current focus on robust and replicable science (Poldrack et al., 2017). In addition to affording the ability to refer to the same system across studies, studies that rely on functional localization include an internal replication component because half of the data is used to identify the voxels/sensors of interest and the other half is used to quantify the response magnitudes, similar to what is commonly done in other modalities (Peelen & Downing, 2005; Nieto-Castañón & Fedorenko, 2012; Powell et al., 2018). Thus, the use of a functional identification task yields a replication of the effect in every new participant and study, thus reducing type I errors.

We have so far focused on univariate analyses. However, the combined use of functional identification and multivariate analyses, like multivariate pattern analysis (MVPA), could powerfully enable investigations of fine-grained meaning and structure representations within the language network. For example, with respect to semantic knowledge, explicit quantitative models reflecting single-word and contextualized semantic knowledge can be used to test hypotheses about how the brain processes meanings extracted from linguistic input (Bruffaerts

et al., 2019). Focusing on language-responsive sensors can increase both the sensitivity of such an analysis and help interpret the observed effects.

## 4.2 Using MEG to probe linguistic computations

We here advocate individual-level functional identification as an approach to achieve greater sensitivity and interpretability when investigating the language network using MEG. The present work introduces this approach using neural signals in sensor space. When dealing with well-characterized systems, like the language network or the face recognition system, for which certain localizer paradigms have been shown (in fMRI research) to reliably identify the relevant underlying functional neuroanatomy, obtaining the anatomical information from MEG becomes not critical: we already know where these signals are coming from based on dozens or even hundreds of fMRI studies, which are ideally suited for localizing functions. As a result, we argue for leveraging the advantages that MEG has over other recording modalities, the core one being the ability to study the detailed time-course of information processing. Our proposed method provides a means for the selection of language-responsive sensors in which in a second stage different time-resolved processes contributing to sentence comprehension can be studied. Specifically, the fine temporal resolution of MEG offers the potential to study the incremental construction of sentence structure and meaning in real time (e.g., (Heilbron et al., 2022; Ten Oever et al., 2022; Desbordes et al., 2024)). Previous MEG studies have suggested that the temporal dynamics of sentence processing entail both feed-forward as well as recurrent processing (Hultén et al., 2019), that representations extracted from artificial neural network language models capture some aspects of neural signals recorded with MEG (Choi et al., 2021), and that different frequency bands may reflect distinct cognitive processes, such as lexical retrieval, semantic composition, and prediction of upcoming words (Lam et al., 2016; Prystauka & Lewis, 2019).

While the main purpose of this work is not to make inferences about the functional role of each frequency band during language comprehension, we do find that the results align with current research. For example, the observed increase in theta-band power during the sentence condition (Figure 3) supports leading theories proposing that theta synchronization plays a role in retrieving incoming words from memory during sentence processing (Meyer et al., 2015; Meyer, 2018; Prystauka & Lewis, 2019). Similarly, the reported increase in beta-band power aligns with accounts suggesting that beta-band synchronization supports the maintenance of the ongoing cognitive set, facilitating top-down predictions (Lam et al., 2016; Armeni et al., 2019; Prystauka & Lewis, 2019; Lewis et al., 2023). On an additional note, we report differences in the SOIs that are selected at the level of the frequency bands (and the time domain by extension). Sensors that play a key role in one frequency band, might not be equally influential in other frequency bands, highlighting the difference in functional roles that these bands have during language comprehension. This aligns with findings showing that the language network that we can measure with MEG is not driven by a single dominant frequency but rather exhibits a frequency-dependent organization (Coolen et al., 2020).

In summary, the functional identification MEG approach presented here can increase the sensitivity and interpretability of future investigations of the temporal and spectral dynamics of language processing, as needed to decipher the precise computations that enable language comprehension.

## 4.3 Limitations

As is typical with MEG, we used neural changes during sentence processing relative to the baseline to identify the SOIs rather than a contrast between two conditions. In the fMRI implementation of the language localizer, the cortical regions of interest are identified by the sentences > nonwords univariate contrast. It is known that the inverse contrast (nonwords >

sentences) activates the multiple demand network (Duncan, 2010; Fedorenko et al., 2013) which is neuroanatomically in close proximity to the language network (Blank et al., 2014; Braga et al., 2020; Du et al., 2024). The inherently lower spatial resolution of MEG may result in sources from the language and multiple demand networks being captured by the same SOIs. Here, we opted to select SOIs based on the sentence condition and verified that the response to the nonwords condition was significantly lower from the sentence condition. The advantage of using the contrast of the sentence minus the nonwords condition is that non-language processing is subtracted out: in the case of the nonwords condition, this processing includes working memory and visual perceptual processing. As we did not use a contrast and examined temporal aggregates within the post-stimulus time window – in which both visual processing and language processing occurs – some of the selected sensors may reflect sources in the occipital cortex related to lower-level visual processes. However, it is clear that most SOIs are not occipital channels (Figure 2).

Secondly, we opted to include the top 10% of most responsive sensors in the sentence condition. This choice assumes an event related synchronization in each of the neural frequency bands over the course of the sentences. While past research does point in this direction (Prystauka & Lewis, 2019), some have demonstrated opposite effects in which, depending on the anatomical brain region, the alpha and beta band also exhibit an event-related desynchronization (Lam et al., 2016).

Thirdly, we opted to compute the percentage signal change from baseline between the fifth (non)word and the end of the trial to calculate the effect size. This choice did not critically impact the results: similar findings were observed when including signals across the whole trial (Supplementary Tables 1 and 2).

Finally, extracting frequency information from the signal can pose challenges. For one, the detectability of neural activations in higher frequency bands (gamma) is limited in MEG (Jerbi et al., 2009; Muthukumaraswamy, 2013). The lower signal-to-noise ratio for high-frequency activity might explain why the within-participant correlations and the effect sizes were smaller in the gamma bands when compared to the lower-frequency bands (theta, alpha, and beta). Additionally, the paradigm used in this study, due to its time-locked stimulus presentation, is expected to give rise to both evoked responses as well as induced responses. A relevant question related to the earlier notion that the frequency analysis presents a different yet complementary perspective to the time-domain analysis, is that the power-based effect sizes in the frequency domain may be driven by the same evoked responses that shape the results in the time domain. To address this, we performed a control analysis in which we estimate the induced power in each of the frequency bands after subtracting the event-related average from the signal. Our verification analysis (Supplementary Figure 1) confirms that the power-based effects primarily reflect modulations of ongoing oscillations.

# 5. Conclusion

Using an extensively validated fMRI language localizer task, based on sentence reading, we showed that the neural responses recorded with MEG are reproducible at the individual participant level in the time and frequency domain and we generalized these findings across two datasets in two different languages (English and Dutch). We observed that language-responsive sensors are spatially variable across individuals, giving an individual-level approach an advantage over the traditional group-level analysis. This method has a wide range of applications—from the detailed characterization of the time-course of language processing to probing the language network in (small) non-neurotypical populations – and may generally

encourage the use of MEG to study the functional neuroanatomy of human higher-order cognition.

## 6. Data and Code Availability Statements

Initial preprocessing of the data made use of MaxFilter version 2.2, while further processing was performed with MNE-Python version 1.8.0. All scripts used to generate the outputs for this work as well as the task stimuli are available on OSF (https://osf.io/vc2bw/). Raw MEG data can be made available upon reasonable request.

## 7. Acknowledgements

The authors wish to thank dr. Anna Ivanova and dr. Antonin Rovai. They also acknowledge the Athinoula A. Martinos Imaging Center at the McGovern Institute for Brain Research. This work was supported by an FWO research grant to RB (1509318N), an FWO junior and senior postdoctoral fellowship to RB (12I2117N, 12I2121N), an FWO travel grant for a short stay abroad to RB, and a MISTI Belgium seed fund to EF and RB. RB and MH are supported by the Francqui Foundation. AP was supported by a diversity supplement to EF's NIH grant R01-DC016950. EF was additionally supported by NIH awards R01-DC016607, R01-DC016950, and U01-NS121471, and by research funds from the McGovern Institute for Brain Research, the Brain and Cognitive Sciences department, and the Simons Center for the Social Brain.

## 8. References

Amalric, M., & Dehaene, S. (2018). Cortical circuits for mathematical knowledge: Evidence for a major subdivision within the brain's semantic networks. *Philosophical Transactions of*

*the Royal Society B: Biological Sciences*, *373*(1740), 20160515. https://doi.org/10.1098/rstb.2016.0515

Armeni, K., Willems, R. M., Van Den Bosch, A., & Schoffelen, J.-M. (2019). Frequency-specific brain dynamics related to prediction during language comprehension. *NeuroImage*, *198*, 283-295. https://doi.org/10.1016/j.neuroimage.2019.04.083

Baldauf, D., & Desimone, R. (2014). Neural Mechanisms of Object-Based Attention. *Science*, *344*(6182), 424-427. https://doi.org/10.1126/science.1247003

Beres, A. M. (2017). Time is of the Essence: A Review of Electroencephalography (EEG) and Event-Related Brain Potentials (ERPs) in Language Research. *Applied Psychophysiology and Biofeedback*, *42*(4), 247-255. https://doi.org/10.1007/s10484-017-9371-3

Blank, I., Kanwisher, N., & Fedorenko, E. (2014). A functional dissociation between language and multiple-demand systems revealed in patterns of BOLD signal fluctuations. *Journal of Neurophysiology*, *112*(5), 1105-1118. https://doi.org/10.1152/jn.00884.2013

Braga, R. M., DiNicola, L. M., Becker, H. C., & Buckner, R. L. (2020). Situating the left-lateralized language network in the broader organization of multiple specialized large-scale distributed networks. *Journal of Neurophysiology*, *124*(5), 1415-1448. https://doi.org/10.1152/jn.00753.2019

Brett, M., Johnsrude, I. S., & Owen, A. M. (2002). The problem of functional localization in the human brain. *Nature Reviews Neuroscience*, *3*(3), 243-249. https://doi.org/10.1038/nrn756

Bruffaerts, R., De Deyne, S., Meersmans, K., Liuzzi, A. G., Storms, G., & Vandenberghe, R. (2019). Redefining the resolution of semantic knowledge in the brain: Advances made by the

introduction of models of semantics in neuroimaging. *Neuroscience & Biobehavioral Reviews*, *103*, 3-13. https://doi.org/10.1016/j.neubiorev.2019.05.015

Chetail, F., Ranzini, M., De Tiège, X., Wens, V., & Content, A. (2018). The consonant/vowel pattern determines the structure of orthographic representations in the left fusiform gyrus. *Cortex*, *101*, 73-86. https://doi.org/10.1016/j.cortex.2018.01.006

Choi, H. S., Marslen-Wilson, W. D., Lyu, B., Randall, B., & Tyler, L. K. (2021). Decoding the Real-Time Neurobiological Properties of Incremental Semantic Interpretation. *Cerebral Cortex*, *31*(1), 233-247. https://doi.org/10.1093/cercor/bhaa222

Cogan, G. B., Thesen, T., Carlson, C., Doyle, W., Devinsky, O., & Pesaran, B. (2014). Sensory–motor transformations for speech occur bilaterally. *Nature*, *507*(7490), 94-98. https://doi.org/10.1038/nature12935

Coolen, T., Mihai Dumitrescu, A., Wens, V., Bourguignon, M., Rovai, A., Sadeghi, N., Urbain, C., Goldman, S., & De Tiège, X. (2024). Spectrotemporal cortical dynamics and semantic control during sentence completion. *Clinical Neurophysiology*, *163*, 90-101. https://doi.org/10.1016/j.clinph.2024.04.012

Coolen, T., Wens, V., Vander Ghinst, M., Mary, A., Bourguignon, M., Naeije, G., Peigneux, P., Sadeghi, N., Goldman, S., & De Tiège, X. (2020). Frequency-Dependent Intrinsic Electrophysiological Functional Architecture of the Human Verbal Language Network. *Frontiers in Integrative Neuroscience*, *14*, 27. https://doi.org/10.3389/fnint.2020.00027

Da Silva Castanheira, J., Orozco Perez, H. D., Misic, B., & Baillet, S. (2021). Brief segments of neurophysiological activity enable individual differentiation. *Nature Communications*, *12*(1), 5713. https://doi.org/10.1038/s41467-021-25895-8

De Vries, E., & Baldauf, D. (2019). Attentional Weighting in the Face Processing Network: A Magnetic Response Image-guided Magnetoencephalography Study Using Multiple Cyclic Entrainments. *Journal of Cognitive Neuroscience*, *31*(10), 1573-1588. https://doi.org/10.1162/jocn_a_01428

Desbordes, T., King, J.-R., & Dehaene, S. (2024). Tracking the neural codes for words and phrases during semantic composition, working-memory storage, and retrieval. *Cell Reports*, *43*(3), 113847. https://doi.org/10.1016/j.celrep.2024.113847

Diachek, E., Blank, I., Siegelman, M., Affourtit, J., & Fedorenko, E. (2020). The Domain-General Multiple Demand (MD) Network Does Not Support Core Aspects of Language Comprehension: A Large-Scale fMRI Investigation. *The Journal of Neuroscience*, *40*(23), 4536-4550. https://doi.org/10.1523/JNEUROSCI.2036-19.2020

Du, J., DiNicola, L. M., Angeli, P. A., Saadon-Grosman, N., Sun, W., Kaiser, S., Ladopoulou, J., Xue, A., Yeo, B. T. T., Eldaief, M. C., & Buckner, R. L. (2024). Organization of the human cerebral cortex estimated within individuals: Networks, global topography, and function. *Journal of Neurophysiology*, *131*(6), 1014-1082. https://doi.org/10.1152/jn.00308.2023

Duncan, J. (2010). The multiple-demand (MD) system of the primate brain: Mental programs for intelligent behaviour. *Trends in Cognitive Sciences*, *14*(4), 172-179. https://doi.org/10.1016/j.tics.2010.01.004

Fedorenko, E. (2021). The early origins and the growing popularity of the individual-subject analytic approach in human neuroscience. *Current Opinion in Behavioral Sciences*, *40*, 105-112. https://doi.org/10.1016/j.cobeha.2021.02.023

Fedorenko, E., Behr, M. K., & Kanwisher, N. (2011). Functional specificity for high-level linguistic processing in the human brain. *Proceedings of the National Academy of Sciences*, *108*(39), 16428-16433. https://doi.org/10.1073/pnas.1112937108

Fedorenko, E., & Blank, I. A. (2020). Broca's Area Is Not a Natural Kind. *Trends in Cognitive Sciences*, *24*(4), 270-284. https://doi.org/10.1016/j.tics.2020.01.001

Fedorenko, E., Duncan, J., & Kanwisher, N. (2013). Broad domain generality in focal regions of frontal and parietal cortex. *Proceedings of the National Academy of Sciences*, *110*(41), 16616-16621. https://doi.org/10.1073/pnas.1315235110

Fedorenko, E., Hsieh, P.-J., Nieto-Castañón, A., Whitfield-Gabrieli, S., & Kanwisher, N. (2010). New Method for fMRI Investigations of Language: Defining ROIs Functionally in Individual Subjects. *Journal of Neurophysiology*, *104*(2), 1177-1194. https://doi.org/10.1152/jn.00032.2010

Fedorenko, E., Ivanova, A. A., & Regev, T. I. (2024). The language network as a natural kind within the broader landscape of the human brain. *Nature Reviews Neuroscience*, *25*(5), 289-312. https://doi.org/10.1038/s41583-024-00802-4

Fedorenko, E., Scott, T. L., Brunner, P., Coon, W. G., Pritchett, B., Schalk, G., & Kanwisher, N. (2016). Neural correlate of the construction of sentence meaning. *Proceedings of the National Academy of Sciences*, *113*(41). https://doi.org/10.1073/pnas.1612132113

Frost, M. A., & Goebel, R. (2012). Measuring structural–functional correspondence: Spatial variability of specialised brain regions after macro-anatomical alignment. *NeuroImage*, *59*(2), 1369-1381. https://doi.org/10.1016/j.neuroimage.2011.08.035

Giglio, L., Hagoort, P., & Ostarek, M. (2024). Neural encoding of semantic structures during sentence production. *Cerebral Cortex*, *34*(12), bhae482. https://doi.org/10.1093/cercor/bhae482

Gramfort, A. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, *7*. https://doi.org/10.3389/fnins.2013.00267

Hashimoto, H., Hasegawa, Y., Araki, T., Sugata, H., Yanagisawa, T., Yorifuji, S., & Hirata, M. (2017). Non-invasive detection of language-related prefrontal high gamma band activity with beamforming MEG. *Scientific Reports*, *7*(1), 14262. https://doi.org/10.1038/s41598-017-14452-3

Heilbron, M., Armeni, K., Schoffelen, J.-M., Hagoort, P., & De Lange, F. P. (2022). A hierarchy of linguistic predictions during natural language comprehension. *Proceedings of the National Academy of Sciences*, *119*(32), e2201968119. https://doi.org/10.1073/pnas.2201968119

Hu, J., Small, H., Kean, H., Takahashi, A., Zekelman, L., Kleinman, D., Ryan, E., Nieto-Castañón, A., Ferreira, V., & Fedorenko, E. (2023). Precision fMRI reveals that the language-selective network supports both phrase-structure building and lexical access during language production. *Cerebral Cortex*, *33*(8), 4384-4404. https://doi.org/10.1093/cercor/bhac350

Hultén, A., Schoffelen, J.-M., Uddén, J., Lam, N. H. L., & Hagoort, P. (2019). How the brain makes sense beyond the processing of single words – An MEG study. *NeuroImage*, *186*, 586-594. https://doi.org/10.1016/j.neuroimage.2018.11.035

Hyvarinen, A. (1999). Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*, *10*(3), 626-634. https://doi.org/10.1109/72.761722

Ivanova, A., Siegelman, M., Chandler, C., Pongos, A., Kean, H., & Fedorenko, E. (2020). *Effect of Task on Sentence Processing* [Poster presentation]. Society for the Neurobiology of Language.

Jerbi, K., Ossandón, T., Hamamé, C. M., Senova, S., Dalal, S. S., Jung, J., Minotti, L., Bertrand, O., Berthoz, A., Kahane, P., & Lachaux, J. (2009). Task-related gamma-band

dynamics from an intracerebral perspective: Review and implications for surface EEG and MEG. *Human Brain Mapping*, *30*(6), 1758-1771. https://doi.org/10.1002/hbm.20750

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for Face Perception. *The Journal of Neuroscience*, *17*(11), 4302. https://doi.org/10.1523/JNEUROSCI.17-11-04302.1997

Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods*, *42*(3), 627-633. https://doi.org/10.3758/BRM.42.3.627

Lam, N. H. L., Schoffelen, J.-M., Uddén, J., Hultén, A., & Hagoort, P. (2016). Neural activity during sentence processing as reflected in theta, alpha, beta, and gamma oscillations. *NeuroImage*, *142*, 43-54. https://doi.org/10.1016/j.neuroimage.2016.03.007

Lee, J. J., Scott, T. L., & Perrachione, T. K. (2024). Efficient functional localization of language regions in the brain. *NeuroImage*, *285*, 120489. https://doi.org/10.1016/j.neuroimage.2023.120489

Lewis, A. G., Schoffelen, J., Bastiaansen, M., & Schriefers, H. (2023). Is beta in agreement with the relatives? Using relative clause sentences to investigate MEG beta power dynamics during sentence comprehension. *Psychophysiology*, *60*(10), e14332. https://doi.org/10.1111/psyp.14332

Lipkin, B., Tuckute, G., Affourtit, J., Small, H., Mineroff, Z., Kean, H., Jouravlev, O., Rakocevic, L., Pritchett, B., Siegelman, M., Hoeflin, C., Pongos, A., Blank, I. A., Struhl, M. K., Ivanova, A., Shannon, S., Sathe, A., Hoffmann, M., Nieto-Castañón, A., & Fedorenko, E. (2022). Probabilistic atlas for the language network based on precision fMRI data from >800 individuals. *Scientific Data*, *9*(1), 529. https://doi.org/10.1038/s41597-022-01645-3

Liu, J., Harris, A., & Kanwisher, N. (2002). Stages of processing in face perception: An MEG study. *Nature Neuroscience*, *5*(9), 910-916. https://doi.org/10.1038/nn909

Liu, Y., Sánchez Hernández, F., Ting, F., & Hyde, D. C. (2022). Comparing fixed-array and functionally-defined channel of interest approaches to infant functional near-infrared spectroscopy data. *NeuroImage*, *261*, 119520. https://doi.org/10.1016/j.neuroimage.2022.119520

Maguire, M. J., & Abel, A. D. (2013). What changes in neural oscillations can reveal about developmental cognitive neuroscience: Language development as a case in point. *Developmental Cognitive Neuroscience*, *6*, 125-136. https://doi.org/10.1016/j.dcn.2013.08.002

Mahowald, K., & Fedorenko, E. (2016). Reliable individual-level neural markers of high-level language processing: A necessary precursor for relating neural variability to behavioral and genetic variability. *NeuroImage*, *139*, 74-93. https://doi.org/10.1016/j.neuroimage.2016.05.073

Malik-Moraleda, S., Ayyash, D., Gallée, J., Affourtit, J., Hoffmann, M., Mineroff, Z., Jouravlev, O., & Fedorenko, E. (2022). An investigation across 45 languages and 12 language families reveals a universal language network. *Nature Neuroscience*, *25*(8), 1014-1019. https://doi.org/10.1038/s41593-022-01114-5

Meyer, L. (2018). The neural oscillations of speech processing and language comprehension: State of the art and emerging mechanisms. *European Journal of Neuroscience*, *48*(7), 2609-2621. https://doi.org/10.1111/ejn.13748

Meyer, L., Grigutsch, M., Schmuck, N., Gaston, P., & Friederici, A. D. (2015). Frontal–posterior theta oscillations reflect memory retrieval during sentence comprehension. *Cortex*, *71*, 205-218. https://doi.org/10.1016/j.cortex.2015.06.027

Monti, M. M., Parsons, L. M., & Osherson, D. N. (2012). Thought Beyond Language: Neural Dissociation of Algebra and Natural Language. *Psychological Science*, *23*(8), 914-922. https://doi.org/10.1177/0956797612437427

Morales, S., & Bowers, M. E. (2022). Time-frequency analysis methods and their application in developmental EEG data. *Developmental Cognitive Neuroscience*, *54*, 101067. https://doi.org/10.1016/j.dcn.2022.101067

Muthukumaraswamy, S. D. (2013). High-frequency brain activity and muscle artifacts in MEG/EEG: A review and recommendations. *Frontiers in Human Neuroscience*, *7*. https://doi.org/10.3389/fnhum.2013.00138

Nieto-Castañón, A., & Fedorenko, E. (2012). Subject-specific functional localizers increase sensitivity and functional resolution of multi-subject analyses. *NeuroImage*, *63*(3), 1646-1669. https://doi.org/10.1016/j.neuroimage.2012.06.065

Pallier, C., Devauchelle, A.-D., & Dehaene, S. (2011). Cortical representation of the constituent structure of sentences. *Proceedings of the National Academy of Sciences*, *108*(6), 2522-2527. https://doi.org/10.1073/pnas.1018711108

Papanicolaou, A. C. (2023). Non-Invasive Mapping of the Neuronal Networks of Language. *Brain Sciences*, *13*(10), 1457. https://doi.org/10.3390/brainsci13101457

Paranawithana, I., Mao, D., McKay, C. M., & Wong, Y. T. (2024). Language networks of normal-hearing infants exhibit topological differences between resting and steady states: An fNIRS functional connectivity study. *Human Brain Mapping*, *45*(13), e70021. https://doi.org/10.1002/hbm.70021

Peelen, M. V., & Downing, P. E. (2005). Selectivity for the Human Body in the Fusiform Gyrus. *Journal of Neurophysiology*, *93*(1), 603-608. https://doi.org/10.1152/jn.00513.2004

Poldrack, R. A., Baker, C. I., Durnez, J., Gorgolewski, K. J., Matthews, P. M., Munafò, M. R., Nichols, T. E., Poline, J.-B., Vul, E., & Yarkoni, T. (2017). Scanning the horizon: Towards transparent and reproducible neuroimaging research. *Nature Reviews Neuroscience*, *18*(2), 115-126. https://doi.org/10.1038/nrn.2016.167

Powell, L. J., Deen, B., & Saxe, R. (2018). Using individual functional channels of interest to study cortical development with fNIRS. *Developmental Science*, *21*(4), e12595. https://doi.org/10.1111/desc.12595

Prystauka, Y., & Lewis, A. G. (2019). The *power* of neural oscillations to inform sentence comprehension: A linguistic perspective. *Language and Linguistics Compass*, *13*(9), e12347. https://doi.org/10.1111/lnc3.12347

Regev, T. I., Casto, C., Hosseini, E. A., Adamek, M., Ritaccio, A. L., Willie, J. T., Brunner, P., & Fedorenko, E. (2024). Neural populations in the language network differ in the size of their temporal receptive windows. *Nature Human Behaviour*, *8*(10), 1924-1942. https://doi.org/10.1038/s41562-024-01944-2

Richardson, H., Koster-Hale, J., Caselli, N., Magid, R., Benedict, R., Olson, H., Pyers, J., & Saxe, R. (2020). Reduced neural selectivity for mental states in deaf children with delayed exposure to sign language. *Nature Communications*, *11*(1), 3246. https://doi.org/10.1038/s41467-020-17004-y

Roos, N. M., & Piai, V. (2020). Across-session consistency of context-driven language processing: A magnetoencephalography study. *European Journal of Neuroscience*, *52*(5), 3457-3469. https://doi.org/10.1111/ejn.14785

Rousselet, G. A., & Pernet, C. R. (2012). Improving standards in brain-behavior correlation analyses. *Frontiers in Human Neuroscience*, *6*. https://doi.org/10.3389/fnhum.2012.00119

Saxe, R., Brett, M., & Kanwisher, N. (2006). Divide and conquer: A defense of functional localizers. *NeuroImage*, *30*(4), 1088-1096. https://doi.org/10.1016/j.neuroimage.2005.12.062

Saxe, R., & Kanwisher, N. (2003). People thinking about thinking peopleThe role of the temporo-parietal junction in “theory of mind”. *NeuroImage*, *19*(4), 1835-1842. https://doi.org/10.1016/S1053-8119(03)00230-1

Scott, T. L., Gallée, J., & Fedorenko, E. (2017). A new fun and robust version of an fMRI localizer for the frontotemporal language system. *Cognitive Neuroscience*, *8*(3), 167-176. https://doi.org/10.1080/17588928.2016.1201466

Shain, C., Blank, I. A., Fedorenko, E., Gibson, E., & Schuler, W. (2022). Robust Effects of Working Memory Demand during Naturalistic Language Comprehension in Language-Selective Cortex. *The Journal of Neuroscience*, *42*(39), 7412-7430. https://doi.org/10.1523/JNEUROSCI.1894-21.2022

Shain, C., Kean, H., Casto, C., Lipkin, B., Affourtit, J., Siegelman, M., Mollica, F., & Fedorenko, E. (2024). Distributed Sensitivity to Syntax and Semantics throughout the Language Network. *Journal of Cognitive Neuroscience*, *36*(7), 1427-1471. https://doi.org/10.1162/jocn_a_02164

Sun, Y., & Luo, X. (2024). A mapping-knowledge-domain analysis of ERP research on language processing. *Frontiers in Human Neuroscience*, *18*, 1352753. https://doi.org/10.3389/fnhum.2024.1352753

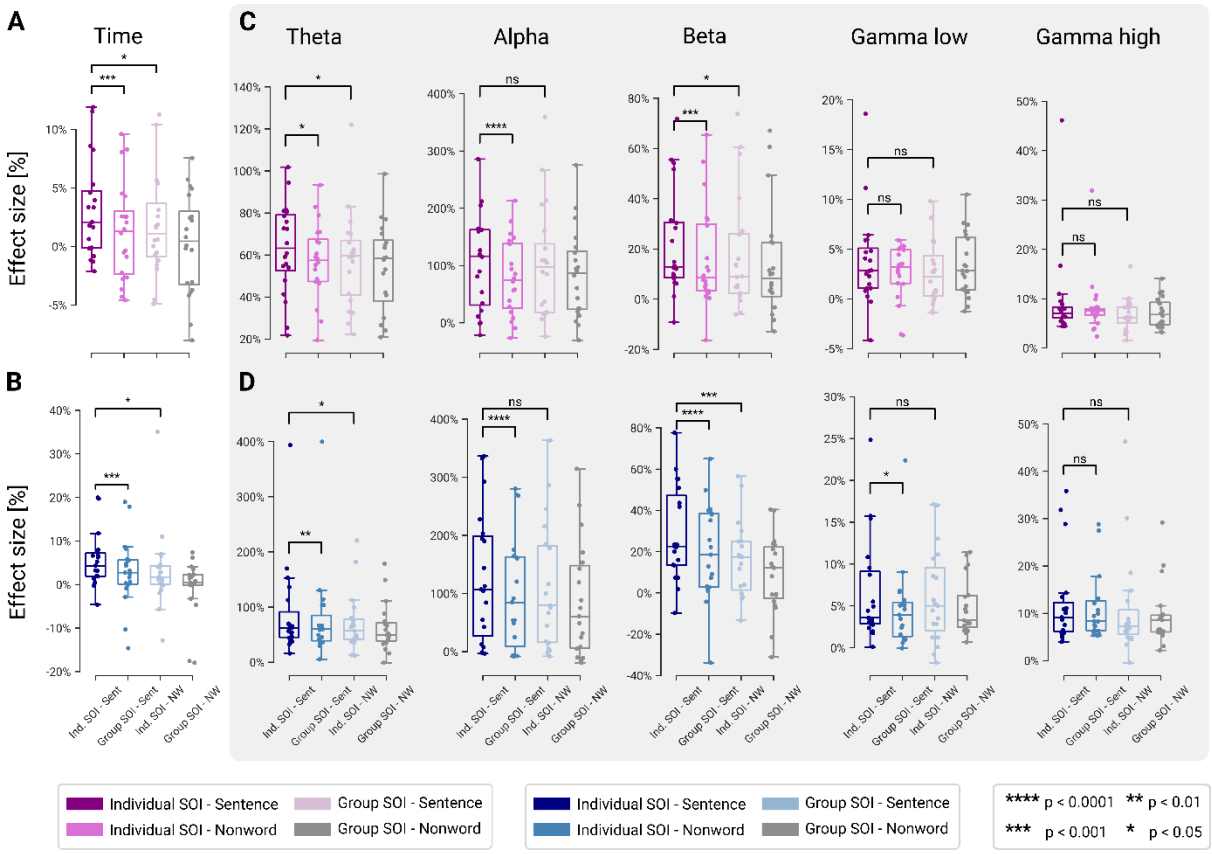
Taulu, S., Kajola, M., & Simola, J. (2003). Suppression of Interference and Artifacts by the Signal Space Separation Method. *Brain Topography*, *16*(4), 269-275. https://doi.org/10.1023/B:BRAT.0000032864.93890.f9

Ten Oever, S., Carta, S., Kaufeld, G., & Martin, A. E. (2022). Neural tracking of phrases in spoken language comprehension is automatic and task-dependent. *eLife*, *11*, e77468. https://doi.org/10.7554/eLife.77468

Towle, V. L., Yoon, H.-A., Castelle, M., Edgar, J. C., Biassou, N. M., Frim, D. M., Spire, J.-P., & Kohrman, M. H. (2008). ECoG gamma activity during a language task: Differentiating expressive and receptive speech areas. *Brain*, *131*(8), 2013-2027. https://doi.org/10.1093/brain/awn147

# Supplementary materials

## Supplementary Figure 1: Induced power effect sizes




***Supplementary Figure 1****: Mean effect sizes (percent signal change) for the sentence and nonwords conditions when the language-responsive sensors of interest (SOIs) are defined at the individual level vs. at the group level in the A) English and B) Dutch datasets for the time domain analysis. Similar results depicted for each of the neural frequency bands in the frequency domain analysis after subtraction of the event-related average in the C) English and D) Dutch dataset. In all cases, the SOIs are defined using one half of the data (odd-numbered trials) and the response magnitudes are examined in the other half of the data (even-numbered trials).*

## Supplementary Table 1: Within vs between participants correlations in English dataset from start word 1

*Average within participants correlation values and between participants correlation values across odd and even-numbered trials when trials are analyzed from the first word onwards instead of from the fifth word and onwards.*

| ***English*** | Time | Theta | Alpha | Beta | Gamma low | Gamma high |
|---|---|---|---|---|---|---|
| Within correlations | 0.58 (s.d. 0.23) | 0.59 (s.d. 0.16) | 0.86 (s.d. 0.10) | 0.77 (s.d. 0.13) | 0.55 (s.d. 0.20) | 0.13 (s.d. 0.13) |
| Between correlations | 0.12 (s.d. 0.23) | 0.12 (s.d. 0.21) | 0.46 (s.d. 0.19) | 0.38 (s.d. 0.24) | 0.20 (s.d. 0.20) | 0.02 (s.d. 0.10) |
| P value | $P < 0.0001$ | $P < 0.0001$ | $P < 0.0001$ | $P < 0.0001$ | $P < 0.0001$ | $P < 0.001$ |

## Supplementary Table 2: Within vs between participants correlations in Dutch dataset from start word 1

*Average within participants correlation values and between participants correlation values across odd and even-numbered trials when trials are analyzed from the first word onwards instead of from the fifth word and onwards.*

| ***Dutch*** | Time | Theta | Alpha | Beta | Gamma low | Gamma high |
|---|---|---|---|---|---|---|
| Within correlations | 0.64 (s.d. 0.20) | 0.67 (s.d. 0.15) | 0.84 (s.d. 0.13) | 0.78 (s.d. 0.13) | 0.41 (s.d. 0.21) | 0.20 (s.d. 0.17) |
| Between correlations | 0.25 (s.d. 0.25) | 0.23 (s.d. 0.21) | 0.47 (s.d. 0.17) | 0.48 (s.d. 0.19) | 0.13 (s.d. 0.17) | 0.06 (s.d. 0.11) |
| P value | $P < 0.0001$ | $P < 0.0001$ | $P < 0.0001$ | $P < 0.0001$ | $P < 0.0001$ | $P < 0.001$ |