Sensory-modality independent activation of the brain network for language

Sophie Arana[1,2,*], André Marquand[1], Annika Hultén[1,2], Peter Hagoort[1,2], Jan-Mathijs Schoffelen[1]

1 Donders Institute, Radboud University, Nijmegen, The Netherlands
2 Max Planck Institute for Psycholinguistics

*** s.arana@donders.ru.nl**

Abstract

The meaning of a sentence can be understood, whether presented in written or spoken form. Therefore it is highly probable that brain processes supporting language comprehension are at least partly independent of sensory modality. To identify where and when in the brain language processing is independent of sensory modality, we directly compared neuromagnetic brain signals of human subjects either reading or listening to sentences. We used multiset canonical correlation analysis to align individual subject data in a way that boosts those aspects of the signal that are common to all, allowing us to capture word-by-word signal variations, consistent across subjects and at a fine temporal scale. Quantifying this consistency in activation across both reading and listening tasks revealed a mostly left hemispheric cortical network. Areas showing consistent activity patterns include not only areas previously implicated in higher-level language processing, such as left prefrontal, superior & middle temporal areas and anterior temporal lobe, but also parts of the control-network as well as subcentral and more posterior temporal-parietal areas. Activity in this supramodal sentence processing network starts in temporal areas and rapidly spreads to the other regions involved. The findings do not only indicate the involvement of a large network of brain areas in supramodal language processing, but also indicate that the linguistic information contained in the unfolding sentences modulates brain activity in a word-specific manner across subjects.

Introduction

Language can be realized in different modalities: amongst others through writing or in speech. Depending on whether the sensory input modality is visual or auditory, different dedicated networks of brain areas are activated in order to derive meaning from the physical stimulus. Besides the obvious differences in the recruitment of brain circuits that process low-level sensory information, differences in linguistic features across sensory modalities prompt a differential activation of brain areas involved in higher-order processing as well. For instance, speech is enriched with meaningful prosodic cues,

but also requires a coarticulated signal to be parsed into individual words. Written text contains cues in the form of punctuation and spaces, and has the advantage of instantaneous availability of full information compared to the temporally unfolding nature of speech. These, and other, differences are paralleled in the brain's response, and thus the sensory modality in which language stimuli are presented determines the dominant spatiotemporal patterns that will be elicited [14]. Regardless of these differences, the same core message can be conveyed in either modality. Therefore, language processing models of the past and present ([12], [13],[16]) usually include not only early sensory (up to 200ms) processing steps, but also contain late (200 - 500 ms), more abstract, and supposedly supramodal processing steps. While early processing is largely unimodal, and supported by brain regions in the respective primary and associative sensory areas, later processes (for instance lexical retrieval and integration) that activate several areas within the temporo-frontal language network are assumed to do so independent of modality. In order to gain insight into the location and timing of brain processes representing this latter, higher order processing of the linguistic content, researchers so far had to rely on carefully manipulated experimental conditions. As a result, our current understanding of how the brain processes language across different modalities reflects a large variety in tasks (semantic decision task [9], error detection task [7], [10], passive hearing/listening [20], orthogonal tasks like size judgment [26])and stimulus material (ranging from words [9] to sentences [3] and stories [4]). Despite this wealth of experimental findings and resulting insights, an important interpretational limitation stems from the fact that the majority of studies employ modality specific low-level baseline conditions (tone pairs and lines, spectrally rotated speech and false fonts, non-words, white noise [24]) to remove the sensory component of the processing. It can be difficult to assess whether the resulting effects are comparable across auditory and visual experiments. Here, we overcome these limitation by achieving a direct comparison that does not require modality-specific baseline conditions and that leverages word-by-word variation in the brain response. Using magnetoencephalography (MEG) signals from 200 subjects, we performed a quantitative assessment of the sensory modality independent brain activity following word onset during sentence processing. We identified widespread left hemispheric involvement, starting from 325 ms after word onset in the temporal lobe and rapidly spreading to anterior areas. These findings not only provide

a quantitative confirmation of earlier findings in a large study sample, they also indicate that supramodal linguistic information conveyed by the individual words in their sentence context leads to subtle fluctuations in brain activation patterns that are correlated across different subjects.

Materials and Methods

Experimental Procedure and MEG Data Acquisition. The MEG data forms part of a large publicly available dataset [32], and has been used in other publications as well [23], [31], [19], [22]. Native Dutch speakers (n = 204, 102 males), with a mean age of 22 years, participated in the experiment. In the current analysis, data from 200 subjects were included. Exclusion of four subjects was due to technical issues during acquisition, which made their datasets not suitable for MCCA. All subjects were right-handed, had normal or corrected-to-normal vision, and reported no history of neurological, developmental, or language deficits. The study was approved by the local ethics committee (CMO, the local "Committee on Research Involving Human Participants" in the Arnhem – Nijmegen region) and followed the guidelines of the Helsinki declaration. All subjects gave written informed consent before participation. Half of the subjects performed the task in only the visual, the other half in only the auditory modality. In the visual modality, words were presented sequentially on a back-projection screen, placed in front of them (vertical refresh rate of 60 Hz) at the center of the screen. The auditory versions of the stimuli were recorded by a female native Dutch speaker and presented binaurally via plastic tubes. From the total stimulus set of 360 sentences six subsets of 120 sentences were created. This resulted in six different groups of subjects who saw (or heard) the same subset of stimuli although in a different (randomized) order with some overlap of items between groups. MEG data were collected with a sampling frequency of 1,200 Hz. Throughout the measurement the head position was continuously monitored using custom software [35].

Artifact Rejection and Alignment of MEG activity. All analyses were done with custom-written MATLAB scripts and FieldTrip [29] and the corresponding code will be made availabe via the Donders Repository. Data were bandpass filtered between 0.5 and 20 Hz, and epoched according to sentence onset, each epoch varying in length,

depending on the number of words within each sentence. Samples contaminated by artifacts were replaced by NaN before further analysis. Since all sentences had been presented in random order, we reordered sentences for each subject to yield the same order across subjects. Subsequently, the signals of the auditory subjects were temporally aligned to the signals of the visual subjects, ensuring coincidence of the onset of the individual words across modalities (Fig. 1A). This alignment was needed to accommodate for differences in word presentation rate. The alignment was achieved by first epoching the auditory subject's signals into smaller overlapping segments. Each segment's first sample corresponded to one of the word onsets as annotated manually according to the audio file while each segment's length depended on the duration of the visual presentation of the corresponding word. Finally, all segments were concatenated again in the original order. By defining segments that were longer than the corresponding auditory word duration, the neural response to each word is fully taken into account and matched to the visual signal, even in the case of short words where the response partly coincided with the next word presentation. MEG data were then downsampled to 120 Hz.

Source Reconstruction and MCCA. We used linearly constrained minimum variance beamforming (LCMV) [36] to reconstruct activity onto a parcellated cortically constrained source model. The sourcemodel consisted of 8,196 dipole locations and was further reduced to 191 parcels per hemisphere following previous work [31]. For each parcel, we obtained a parcel-specific spatial filter as follows. We concatenated the spatial filters of the dipoles comprising the parcel, and obtained a set of time courses of the event-related field at each parcel, and performed a principal component analysis on the result. We selected for each parcel the first five spatial components explaining most of the variance in the signal. MCCA [30], [11] was applied to find projections of those five spatial components that would transform the subject-specific signals so as to boost similarities between them. Canonical correlation analysis (CCA) is a standard multivariate statistical method often used to investigate underlying relationships between two sets of variables. Classically, canonical variates are estimated by transforming the two sets in a way that optimizes their correlation. We applied a generalized version of the classical approach (MCCA) [21], which extends the method to multiple sets, in our case multiple subjects. In this context the optimal linear combination of the spatial components is the one that maximizes the correlation over samples between two subjects

while generalizing to all subject pairings, i.e. the ratio of between-subject covariance and within-subject covariance. The covariance matrix needed to compute the projection is estimated across time points for each sentence and subsequently averaged across sentences. We used a 5-fold cross-validation scheme, within which we computed the projections on a subset of sentences and applied them to all left out sentences (Fig. 1B), so as to limit overfitting effects (see supplementary methods for details).

Statistical Evaluation of Sensory Modality Independent Activation. After applying the projection to the data we computed a time-resolved Pearson correlation between all possible subject pairings. To this end, we first epoched the resulting canonical components according to individual word onsets and selected only content words (nouns, adjectives & verbs) for subsequent steps. We then subtracted the mean across samples and subsequently computed the correlation across words for each point in time and subject pairing. Statistical significance was assessed with a non-parametric randomization test [27], under the null hypothesis that the original correspondence of lexical items between different-modality subjects was a random draw of 1000 potential word orders (cluster-corrected $P<0.05$). Specifically, we shuffled the word order of the visual subjects with respect to the order in the auditory subjects.

Prevalence inference. We evaluated the overlap in the results across all six subgroups using information prevalence inference [1]. For each of the six sets of data, we obtained spatial maps of time-resolved supramodal correlations, as well as 1000 permutation estimates after word order shuffling (see above). We then tested the majority null hypothesis of the prevalence of the effect being smaller or equal to a threshold prevalence. For this, we computed the largest threshold such that the corresponding null hypothesis could still be rejected at the given significance level α. This was done after concatenating the minimum statistic from all parcels and time points, using the maximum statistic to correct for multiple comparisons. For each parcel we evaluated the highest threshold at which the prevalence null hypothesis could be rejected at a level of $\alpha = 0.05$ (see Fig. S1 and S2 for cortical maps showing thresholds averaged and per time point).

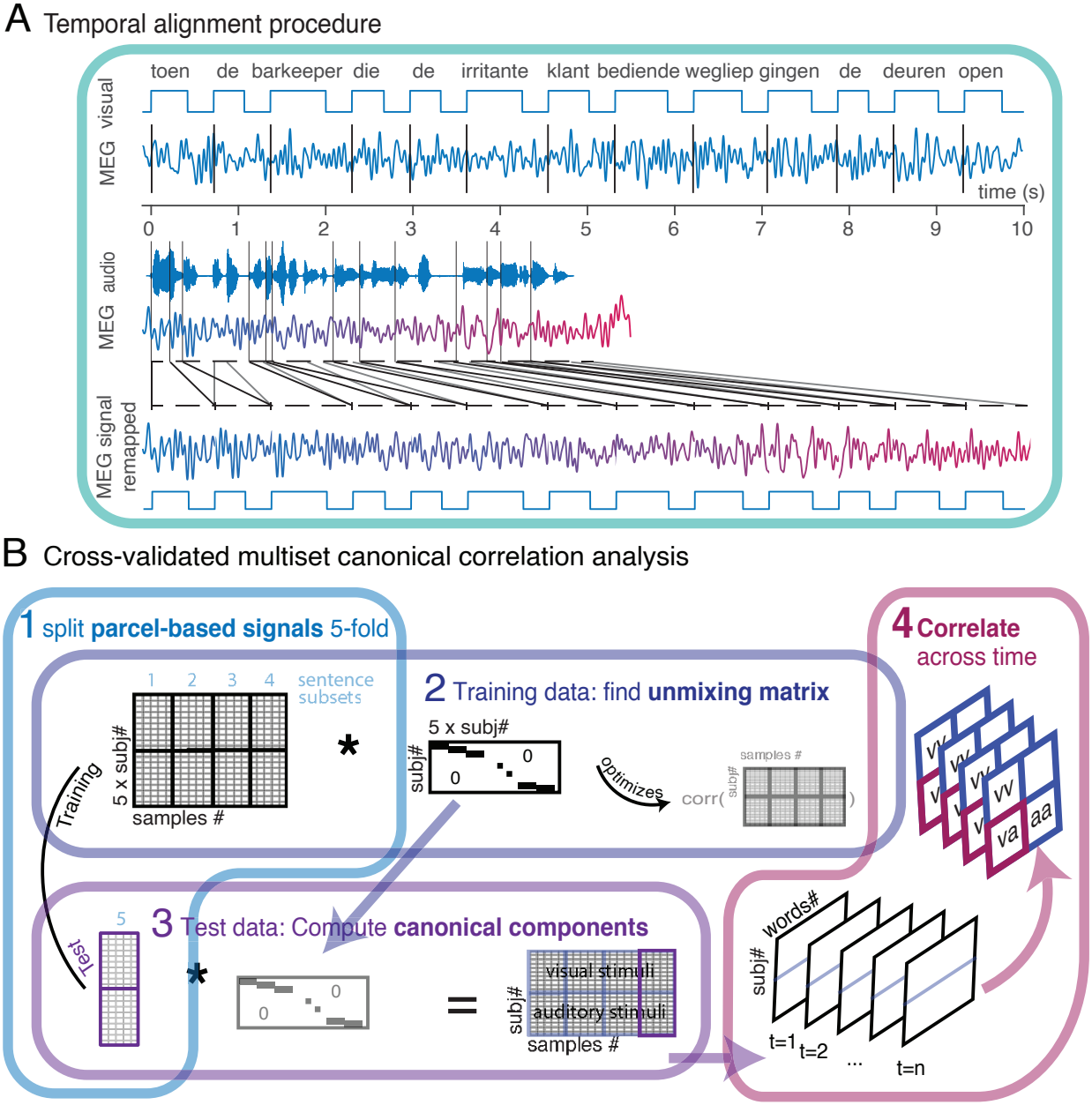


Figure 1. Analysis pipeline. (A) Temporal alignment procedure. MEG signals of auditory and visual subjects differed in length due to different presentation rates. To achieve alignment between signals of auditory and visual subjects, auditory signals were epoched into overlapping segments. Each segment's first sample corresponds to the auditory word onset but each segment's length depends on the duration of the equivalent visual stimulus. Segments were then concatenated in original order to recover signal for the full sentence length. This way, the neural response to each word is fully taken into account in further comparisons, including in the case of short words for which stimulus late processing coincided with the next word presentation. (B) Starting point for the multi-set canonical correlation analysis were parcel-based neural signals for all subjects, consisting of five spatial components each. 1. Signals for all sentence trials were split into five subsets and for cross-validation one subset of sentences was left out as test data, while the remaining four subsets served as training data. 2. Based on the training dataset only

an unmixing matrix was found, per parcel, defining the linear combination of the five spatial components so that the correlation across sets (subjects) and time samples were maximized. The cross-covariance was computed between all subjects' spatial components and across time collapsing over sentence trials. 3. The projection was applied to the test data to compute canonical variables for the left out sentence trials (purple outline) for all subjects. Steps 2 and 3 were repeated for all folds until each sentence subset had been left out once and the resulting canonical variables were concatenated until the entire signal was transformed. 4. Canonical variables were epoched according to word onsets and for each time point a subject-by-subject correlation matrix was computed across words. Correlation between cross-modal subjects (pink outline) were interpreted as quantifying supramodal activation.

Results

We analyzed evoked neuromagnetic signals from 200 subjects, who were either reading or listening to sentences. These data are part of the Mother Of Unification Studies (MOUS) dataset [32], which has been used in other publications as well [23],[31],[19], [22]. We reconstructed the brain activity onto a cortically constrained source model, using an anatomical atlas-based parcellation (see supplementary material). For each of these parcels we quantified the degree of supramodal activation by directly correlating signals of subjects being visually stimulated with those of subjects receiving auditory stimulation. We computed those correlations across words and for each time point separately. Comparing between different subjects' brain responses is challenging, however, due to differing position or orientation of neuronal sources relative to the MEG sensors. In order to make the individual subject data more comparable to each other we first boosted signal traits common to all using multi-set canonical correlation analysis (MCCA), a multivariate statistical method classically used to assess common aspects across multiple datasets [21]. Using a cross-validated estimation procedure we estimated canonical variables for all subjects on all concatenated data trials (see supplementary materials for details), that replaced the original source time-courses. From the MCCA-transformed signals we extracted epochs, defined relative to the onset of the individual words (a total of 757 content words). Finally, we computed time-resolved Pearson correlation coefficients across word epochs between all possible subject pairings (Fig 1B illustrates the analysis pipeline). Averaging the correlations across those subject pairings for which subjects were stimulated either in the same sensory modality, or each in a different modality, allowed us to evaluate the modality-specific brain response and the supramodal

response, respectively.

As per the study design, the subjects were assigned to one of six stimulus sets. Within each of these sets, all subjects heard or read the exact same set of sentences. Different subgroups were presented with different sets of sentences, which prevented us from combining across all 200 subjects. Initially, we constrained our analysis to the first set of 33 subjects (henceforth exploratory dataset). We then analyzed the remaining sets of subjects (confirmatory dataset) using the same analysis pipeline and report on how prevalent the effects are across all six datasets.

For the exploratory dataset, Figure 2 shows that early sensory cortical areas only show correlated activity for the group of subjects receiving the stimuli in the corresponding sensory modality, for the visual (red), and auditory (blue) modalities. It also highlights the need for performing the MCCA, in order to identify the intersubject correlations. Only after MCCA does cortical activity in visual and auditory areas become significantly correlated across those subjects performing the task in the visual or auditory domain respectively (Fig 2A). Critically, we next evaluated correlations averaged over all cross-modal subject pairings.

Supramodal activation patterns. We observed significant supramodal correlated activation patterns in mostly left-lateralized cortical areas (Fig 3, see supplementary Fig S4 for results in the right medial hemisphere). The effect has a large spatial and temporal extent, becoming apparent as early as 250 ms and lasting until 700 ms after word onset. Parcels in middle superior temporal gyrus (STG) contribute to the effect at the earliest time points, followed by the posterior and anterior part of the STG and ~50 ms later the anterior temporal pole. Supramodal correlated activation in ventral temporal cortex follows a similar temporal and spatial pattern, with supramodal correlations starting out more posterior around 292 ms and evolution towards the middle anterior temporal lobe at 308 ms. Other areas that express supramodal activity at relatively early time points are medial prefrontal cortex and primary auditory cortex (250 ms), followed by subcentral parietal regions and supramarginal gyrus at around 300 ms, and finally dorsolateral frontal cortex (325 ms). By the time 375 ms have passed, the entire orbito-frontal cortex, anterior and dorsolateral prefrontal cortex as well as inferior frontal gyrus (IFG) show strong supramodal subject correlation. Supramodal activation in the frontal lobe further extends towards posterior regions including pre- and postcentral gyrus. At around 400

ms supramodal subject correlation in the anterior temporal pole reaches its peak. In addition to the lateral cortical areas, correlated activity also extends to left dorsal and ventral anterior cingulate cortex as well as left fusiform gyrus. The spatio-temporal patterns of supramodal activation described so far are robust, also when ordinal word position is controlled for in the statistical evaluation (see supplementary Figure S1).

Prevalence Inference. So far, the presented results reflect an analysis of a set of 33 subjects, based on which we observed a network of areas expressing significant supramodal correlated activity. We then quantified the extent to which the spatiotemporal patterns observed in the first dataset can be replicated in the confirmatory dataset, and generalize to the population. We performed population inference using a permutation-based implementation of information prevalence inference [1]. This type of statistical inference allows to formulate a complex null hypothesis, i.e. that the prevalence of the effect is smaller than or equal to a threshold prevalence, where the threshold can be realized by different values (see supplementary materials). Based on the minimum statistic across datasets the global null hypothesis of no effect in any subgroup of subjects in the population, was tested. For those parcels at which the global null hypothesis could be rejected, we infer that at least in one of the datasets an effect of supramodal processing was present. In addition, we evaluated the majority null hypothesis of whether, in the majority of subgroups in the population (threshold > 0.5), the data contains an effect.

The global null hypothesis (no information in any set of subjects in the population) could be rejected at a level of $\alpha = 0.05$ in on average 40 parcels per time point (between 325 and 617 ms after onset, std. $= 31.48$). For those parcels, the average largest lower bound $\gamma 0$ at which the prevalence null hypothesis can be rejected is shown in supplementary Figures S1 and S2. For those parcels for which the largest bound $\gamma 0$ is larger than or equal to 0.5 we can infer that in the majority of the datasets the activity patterns were similar across subjects independent of modality.

This majority null hypothesis could be rejected (at a level of $\alpha = 0.05$) in 90% of parcels that also showed a global effect (Fig 4). Compared to the temporal pattern of the largest nominal suprathreshold cluster from the cluster-based permutation test conducted on the exploratory dataset, the effect became significant in the majority of datasets later in time and was less long lasting (325 - 608 ms). Given this time span, the majority null hypothesis was rejected in on average 42% of those parcels

contributing most to the largest cluster. The orbitofrontal cortex and IFG showed an involvement in supramodal processing in both analyses, but the effect there was much more temporally sustained in the exploratory dataset. In addition, according to the exploratory dataset, supramodal activation of STG occurred almost 100 ms earlier as compared to IFG. Based on the confirmatory dataset, however, supramodal correlated activation in IFG and STG appeared almost simultaneously. Finally, the exploratory analysis revealed supramodal activation in primary and premotor areas extending over the entire left dorsolateral surface, of which only the most ventral parcels close to the Sylvian fissure were significantly supramodal in the majority of datasets. Thus, the spatial extent of the effect was partly reduced for prevalence inference, compared to the cluster-based permutation approach on the exploratory data. Nevertheless, widely overlapping anatomical regions were indicated by both analyses, encompassing dorsolateral frontal gyrus and the middle & superior parts of the temporal lobe at first and later also inferior frontal and orbito-frontal cortex, as well as anterior temporal lobe.

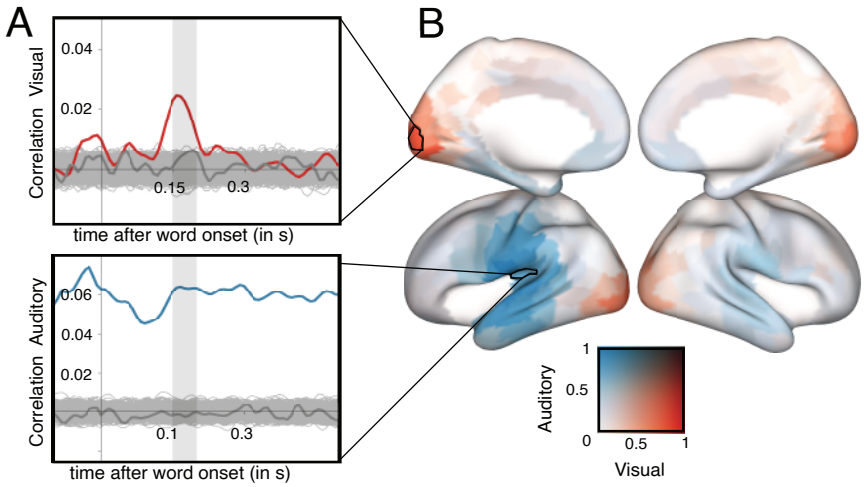


Figure 2. Specificity of the within-modality correlated activity patterns. (A) Time-resolved correlation values averaged across all visual subject pairings for a parcel in left primary visual cortex (upper panel) and all auditory subject pairings for a parcel in left primary auditory cortex (lower panel), before (dark grey line), and after MCCA (blue and red lines). Light grey lines show recomputed correlation values for 1000 random permutations of word order across subjects. Notably, signals of auditory subjects highly correlate even before word onset. This is likely due to a more varied distribution of information in the auditory signal caused by the continuous nature of auditory stimulation and as a result differing time points at which individual words become uniquely recognizable. The MCCA is blind to the stimulus timing and will thus find canonical variables that yield maximal correlations at any timepoint if possible. (B) Cortical map of the spatial distribution of correlations, comparing within modality visual subject pairs (red) with auditory subject pairs (blue). Correlation strength is expressed as the Pearson correlation coefficient averaged over a time window from 150 to 200 ms post word-onset and normalized by the maximum value of that window.

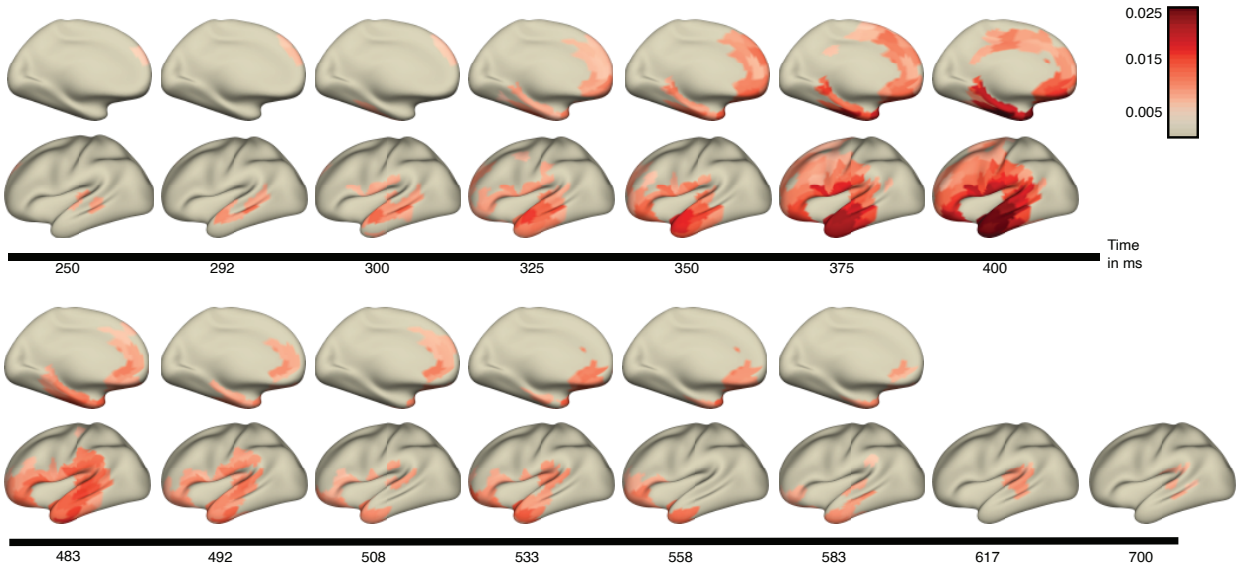


Figure 3. Time-resolved spatial maps of supramodal correlated activity patterns (averaged over all possible cross-modal subject pairings) in the left hemisphere. Medial views of the brain surface are depicted in the first and third row, lateral views in the second and fourth row. Color codes for strength of correlation. Colored parcels were most strongly correlated between cross-modal subject pairs (nonparametric permutation test, corrected for multiple comparisons).

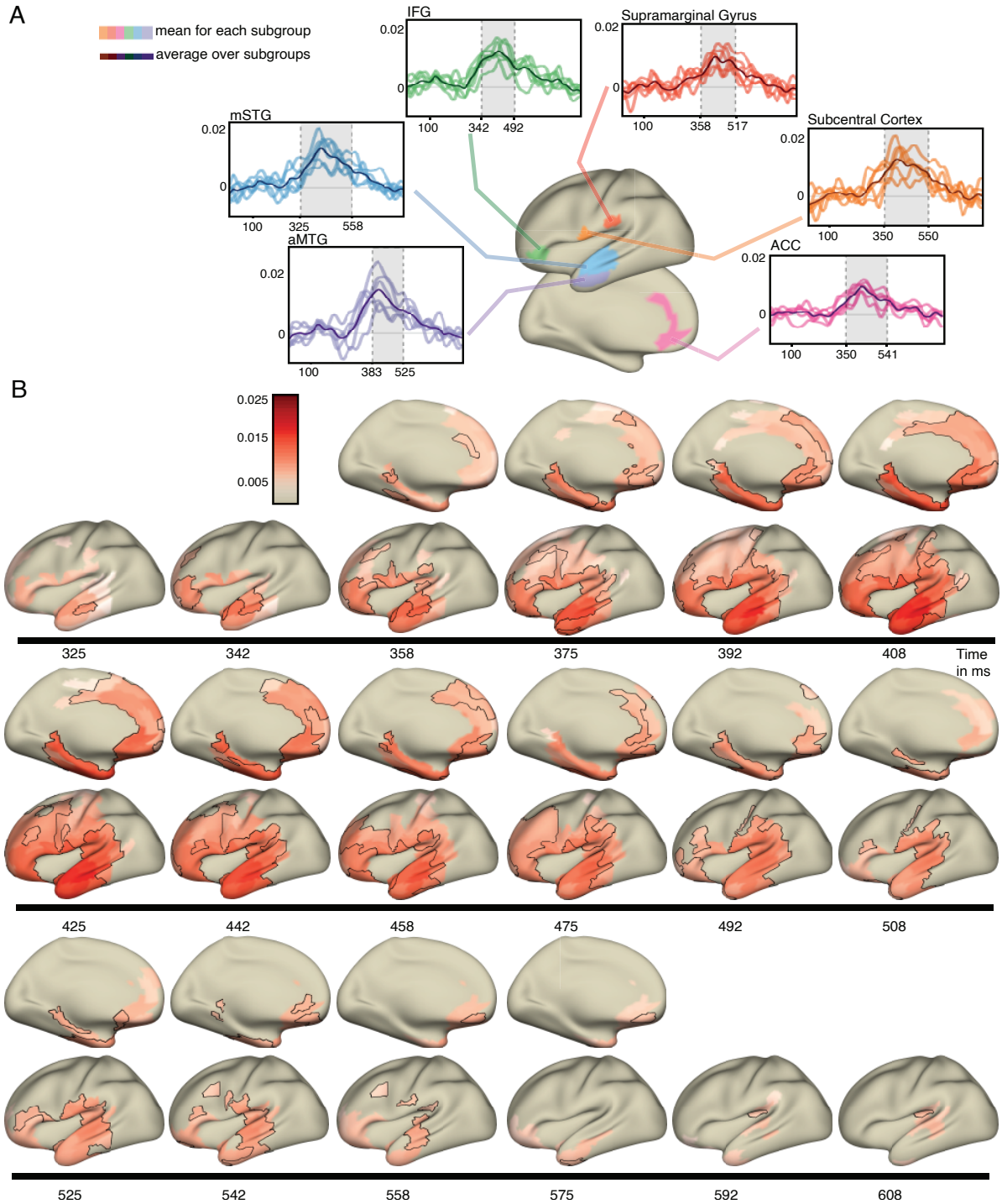


Figure 4. Supramodal correlated activity patterns of word-specific activity consistent across the majority of datasets. (A) Averaged correlation time courses (mean over all possible cross-modal subject pairings) are shown for selected parcels in inferior frontal Gyrus (IFG, green), supramarginal Gyrus (red), Subcentral Cortex (orange), Anterior cingulate cortex (ACC, pink), anterior middle temporal Gyrus (aMTG, purple), and middle superior temporal gyrus (mSTG, blue). Time courses are shown for each dataset individually (light-colored lines) as well as

averaged (dark lines). Grey shaded areas mark statistically significant time points. (B) Time-resolved spatial maps of cross-modal correlations in the left hemisphere. Medial views of the brain surface are depicted in the first, third and fifth row, lateral views in the second, fourth and sixth row. For those parcels that were part of the largest nominal suprathreshold cluster tested on only the exploratory dataset, the mean correlation over all six datasets is shown. Color codes for strength of correlation. In addition, the parcels at which the majority null hypothesis according to prevalence inference could be rejected are outlined in black.

Discussion

The aim of this research was to quantify similarities of the brain response across reading and listening at a fine temporal scale. We identified a widespread left-lateralized brain network, which was activated independent of modality starting from 325 ms after word onset. Importantly, dividing our large study sample into six independent subsets, we could directly quantify the consistency and generalizability of these observed activity patterns. The spatial distribution of the supramodal activation is in line with the known involvement of left hemispheric areas, including parts of left temporal cortex, left inferior parietal lobe, as well as portions of prefrontal cortex [38], [9], [10], [6], [25], [24], [34], [18]. The involvement of both STG and IFG also fits with predictions from the Memory, Unification and Control model (MUC), in which activity reverberating within a posterior-frontal network ([2], [13]) is thought to be crucial for adequate language processing. According to the MUC model temporal and parietal areas support the retrieval of lexical information, while unification processes are supported by inferior frontal cortex. Bidirectional communication [31] between these areas is facilitated by white matter connections. Concerning the temporal dynamics, we observe that temporal areas are supramodally activated at earliest time points and sustain this activation for the longest as compared to other regions. Over time, supramodal activation spreads from middle and posterior left STG to the anterior temporal pole. This rapid progression of activity from posterior to anterior regions mirrors previous observations ([26], [37]), adding to those findings a direct quantitative comparison of the supramodal brain activity.

Beyond the core language network and the single word level. Whereas prior studies on supramodal processes have reported involvement of mainly inferior parts of the frontal cortex [20], [24], [10], [26], [18], [28], we also observed modality-

independent activity in more dorsal regions of the frontal cortex. This dorsal frontal cortical involvement could well be caused by our use of linguistically rich sentence material, of varying syntactic complexity, as opposed to single words or short phrases as stimuli. Indeed, discrepancies with respect to frontal lobe involvement in modality-independent processing have been argued to mainly arise from differences in stimulus material and task demands [6], and a recent meta-analysis has identified that more complex syntax robustly activates more dorsal parts of the left inferior frontal gyrus [15]. Further, a previously published analysis on the visual input modality of this MEG data showed dorsolateral prefrontal cortex to be sensitive to sentence progression effects, when compared against a word list condition [19]. Using syntactically complex sentence material may also have lead to supramodal activation of the anterior cingulate cortex (ACC) bilaterally. The ACC is a midline structure, forming part of a large domain-general executive control network supporting language processing [13], [8]. Specifically, it has been shown to be sensitive to statistical contingencies in the language input supporting a potential role in mediating learning and adaptation in response to predictive regularities in both local experimental as well as global environment [39]. It should be noted that deep sources are normally poorly detectable in MEG [17] and we thus consider any interpretations with respect to the midline structures as tentative. Nonetheless, we generally observe that supramodal processing involves more than the traditional language areas. Importantly, given our analysis approach we can conclude that cortical areas outside the core language network are implicated in supramodal processing in an item-specific way. Our findings therefore go beyond showing a general activation of those areas as compared to baseline but rather reveal consistent word-by-word fluctuations of activation within the recruited areas.

Supramodal orthography-phonology mapping. We observed supramodal activation in post-central and subcentral gyrus, as well as supramarginal gyrus. These areas show significant supramodal activation in the majority of datasets, which coincides temporally with supramodal activation of primary auditory cortex. Increases of neural activity in supramarginal gyrus have been repeatedly elicited by cross-modal tasks [33], such as rhyming judgments to visually presented words [5], for which conversion between orthographic and phonological representations is likely needed. At the same time post- and subcentral areas partly span articulatory motor and somatosensory areas for the

mouth and tongue. Together, the modality-independent activation of these areas suggests that retrieval of phonetic and articulatory mappings is not limited to speech perception only but also occurs during passive reading.

Latency of supramodal processing. The temporal alignment procedure of the subject data, as a necessary preparation step for the MCCA procedure, followed by the estimation of time-resolved intersubject correlations, focused on common signal aspects that are exactly synchronized across subjects. The differences in sensory modality specific characteristics of the input signal require dedicated processing with likely different processing latencies, which may also lead to latency differences in the activation of supramodal areas. For example, in previous work investigating supramodal language processing [26], shorter reaction times during the visual task were observed, yet earlier activity peaks were found for the auditory task in corresponding early sensory cortex and the left anterior temporal lobe. In contrast, other work observed earlier anterior temporal lobe activation for visual, compared to auditory stimulation [3]. Our results indicate that there is a certain degree of overlap across modalities in the temporal window within which supramodal cortical areas are activated. It is possible, that we were able to observe more temporally extensive activation, for instance, activity related to unification processes, given that we used longer sentences as compared to using words [9],[26],[5],[25],[37] or short phrases [3],[7],[6].

In addition, this overlap may have been amplified as a necessary consequence of the MCCA procedure. MCCA allowed us to more directly investigate time-resolved inter-subject correlations and thereby move beyond expressing activation as event-related averages [26]. The use of parcellated MEG source reconstruction in combination with exact temporal alignment of individual sentences across subjects, allowed for the extraction of signal components that are shared across subjects, and provided a way to reduce the spatial variability across subjects, which is commonly observed in more traditional (for instance, dipole fitting) procedures [37]. Evidently, time-resolved correlations between signals from auditory subjects were boosted with less temporal specificity compared to visual subjects (see Fig 2B). This observation was unexpected and may be due to a more continuous stimulation in the auditory experiment. For example, as the sound of a spoken word unfolds, the timing at which it becomes uniquely recognizable will vary across word. Thus, the distribution of information in the auditory

signal is much more varied as compared to the visual. MCCA will pick up on any common relationship across subjects regardless of timing. In our specific application, projections were estimated on concatenated data, effectively making the method blind to word onset boundaries.

In conclusion, this study provides direct neurophysiological evidence for sensory modality independent processes supporting language comprehension in multiple left hemispheric brain areas. We identified a network of areas including domain general control areas as well as phonological mapping circuits over and above traditional higher-level language areas in frontal and temporal-parietal regions, by quantifying between-subject consistency of their respective word-specific activation patterns. These consistent activation patterns were item-specific, and thus likely reflect more than just generic activation during language processing. Based on a prevalence statistical analysis, we conclude that these findings are generalizable to the wider population. Finally, this work indicates that alignment of individual subject data through multi-set canonical correlation analysis might be a promising tool for investigating subtle word-in-context specific modulations of brain activity in the language processing system.

Acknowledgments

We thank Phillip Alday for providing helpful comments. This work was supported by The Netherlands Organisation for Scientific Research (NWO Vidi: 864.14.011, awarded to J.M.S.).

References

1. C. Allefeld, K. Görgen, and J. D. Haynes. Valid population inference for information-based imaging: From the second-level t-test to prevalence inference. *NeuroImage*, 141:378–392, 2016.

2. G. Baggio and P. Hagoort. The balance between memory and unification in semantics: A dynamic account of the N400. *Language and Cognitive Processes*, 26(9):1338–1367, 2011.

3. D. K. Bemis and L. Pylkkänen. Basic linguistic composition recruits the left anterior temporal lobe and left angular gyrus during both listening and reading. *Cerebral Cortex*, 23(8):1859–1873, 2013.

4. M. M. Berl, E. S. Duke, J. Mayo, L. R. Rosenberger, E. N. Moore, J. VanMeter, N. B. Ratner, C. J. Vaidya, and W. D. Gaillard. Functional anatomy of listening and reading comprehension during development. *Brain and Language*, 114(2):115–125, 2010.

5. J. R. Booth, D. D. Burman, J. R. Meyer, D. R. Gitelman, T. B. Parrish, and M. M. Mesulam. Modality independence of word comprehension. *Human Brain Mapping*, 16(4):251–261, 2002.

6. D. Braze, W. E. Mencl, W. Tabor, K. R. Pugh, R. Todd Constable, R. K. Fulbright, J. S. Magnuson, J. A. Van Dyke, and D. P. Shankweiler. Unification of sentence processing via ear and eye: An fMRI study. *Cortex*, 47(4):416–431, 2011.

7. A. Carpentier, K. R. Pugh, M. Westerveld, C. Studholme, O. Skrinjar, J. L. Thompson, D. D. Spencer, and R. T. Constable. Functional MRI of language processing: Dependence on input modality and temporal lobe epilepsy. *Epilepsia*, 42(10):1241–1254, 2001.

8. I. Cattinelli, N. A. Borghese, M. Gallucci, and E. Paulesu. Reading the reading brain: A new meta-analysis of functional imaging data on reading. *Journal of Neurolinguistics*, 26(1):214–238, 2013.

9. M. W. L. Chee, K. M. O'Craven, R. Bergida, B. R. Rosen, and R. L. Savoy. Auditory and visual word processing studied with fMRI. *Human Brain Mapping*, 7(1):15–28, 1999.

10. R. T. Constable, K. R. Pugh, E. Berroya, W. E. Mencl, M. Westerveld, W. Ni, and D. Shankweiler. Sentence complexity and input modality effects in sentence comprehension: An fMRI study. *NeuroImage*, 22(1):11–21, 2004.

11. A. de Cheveigné, G. M. Di Liberto, D. Arzounian, D. D. Wong, J. Hjortkjær, S. Fuglsang, and L. C. Parra. Multiway canonical correlation analysis of brain data. *NeuroImage*, 186:728–740, 2018.

12. N. Geschwind. Specializations of the human brain. *Scientific American*, 2(241):180–201, 1979.

13. P. Hagoort. The core and beyond in the language-ready brain. *Neuroscience and Biobehavioral Reviews*, 81:194–204, 2017.

14. P. Hagoort and C. M. Brown. ERP effects of listening to speech compared to reading: the P600 to syntactic visual presentation. *Neuropsychologia*, 38:1531–1549, 2000.

15. P. Hagoort and P. Indefrey. The Neurobiology of Language Beyond Single Words. *Annual Review of Neuroscience*, 37(1):347–362, 2014.

16. G. Hickok and D. Poeppel. The Cortical Organisation fo Speech Processing. *Nature*, 8(5):393–402, 2007.

17. A. Hillebrand and G. R. Barnes. A quantitative assessment of the sensitivity of whole-head MEG to activity in the adult human cortex. *NeuroImage*, 16(3A):638–650, 2002.

18. F. Homae, R. Hashimoto, K. Nakajima, Y. Miyashita, and K. L. Sakai. From perception to sentence comprehension: The convergence of auditory and visual information of language in the left inferior frontal cortex. *NeuroImage*, 16(4):883–900, 2002.

19. A. Hultén, J.-M. Schoffelen, J. Uddén, N. H. Lam, and P. Hagoort. How the brain makes sense beyond the processing of single words – An MEG study. *NeuroImage*, 186:586–594, 2019.

20. G. Jobard, M. Vigneau, B. Mazoyer, and N. Tzourio-Mazoyer. Impact of modality and linguistic complexity during reading and listening tasks. *NeuroImage*, 34(2):784–800, 2007.

21. J. R. Kettenring. Canonical analysis of several sets of variables. *Biometrika*, 58(3):433–451, 1971.

22. N. H. Lam, A. Hultén, P. Hagoort, and J. M. Schoffelen. Robust neuronal oscillatory entrainment to speech displays individual variation in lateralisation. *Language, Cognition and Neuroscience*, 33(8):943–954, 2018.

23. N. H. Lam, J.-M. Schoffelen, J. Uddén, A. Hultén, and P. Hagoort. Neural activity during sentence processing as reflected in theta, alpha, beta, and gamma oscillations. *NeuroImage*, 142:43–54, 2016.

24. R. Lindenberg and L. Scheef. Supramodal language comprehension: Role of the left temporal lobe for listening and reading. *Neuropsychologia*, 45(10):2407–2415, 2007.

25. A. G. Liuzzi, R. Bruffaerts, R. Peeters, K. Adamczuk, E. Keuleers, S. De Deyne, G. Storms, P. Dupont, and R. Vandenberghe. Cross-modal representation of spoken and written word meaning in left pars triangularis. *NeuroImage*, 150:292–307, 2017.

26. K. Marinkovic, R. P. Dhond, A. M. Dale, M. Glessner, V. Carr, and E. Halgren. Spatiotemporal dynamics of modality-specific and supramodal word processing. *Neuron*, 38(3):487–497, 2003.

27. E. Maris and R. Oostenveld. Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, 164(1):177–190, 2007.

28. E. Michael, T. A. Keller, P. Carpenter, and M. A. Just. fMRI investigation of sentence comprehension by eye and by ear: Modality fingerprints on cognitive processes. *Human Brain Mapping*, 13(4):239–252, 2001.

29. R. Oostenveld, P. Fries, E. Maris, and J. M. Schoffelen. FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, 2011:1:9, 2011.

30. L. C. Parra. Multi-set Canonical Correlation Analysis simply explained. *arXiv:1802.03759*, (11 February 2018), 2018.

31. J.-M. Schoffelen, A. Hultén, N. Lam, A. F. Marquand, J. Uddén, and P. Hagoort. Frequency-specific directed interactions in the human brain network for language. *Proceedings of the National Academy of Sciences*, 114(30):8083–8088, 2017.

32. J.-M. Schoffelen, R. Oostenveld, N. H. L. Lam, J. Uddén, A. Hultén, and P. Hagoort. A 204-subject multimodal neuroimaging dataset to study language processing. *Scientific Data*, 6(1):17, 2019.

33. M. W. Sliwinska, M. Khadilkar, J. Campbell-Ratcliffe, F. Quevenco, and J. T. Devlin. Early and sustained supramarginal gyrus contributions to phonological processing. *Frontiers in Psychology*, 3:1–10, 2012.

34. G. Spitsyna, J. E. Warren, S. K. Scott, F. E. Turkheimer, and R. J. S. Wise. Converging Language Streams in the Human Temporal Lobe. *Journal of Neuroscience*, 26(28):7328–7336, 2006.

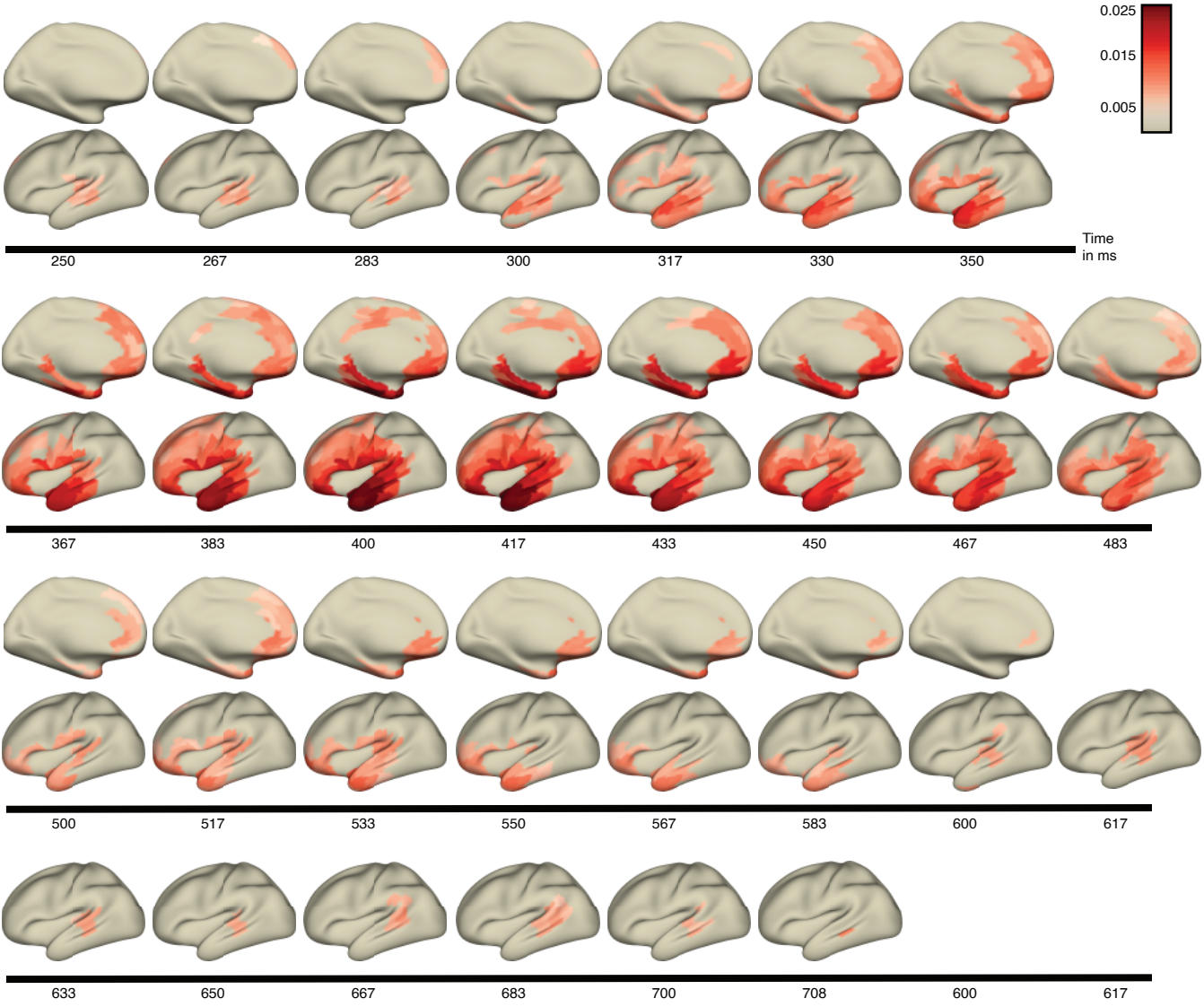
35. A. Stolk, A. Todorovic, J. M. Schoffelen, and R. Oostenveld. Online and offline tools for head movement compensation in MEG. *NeuroImage*, 68:39–48, 2013.

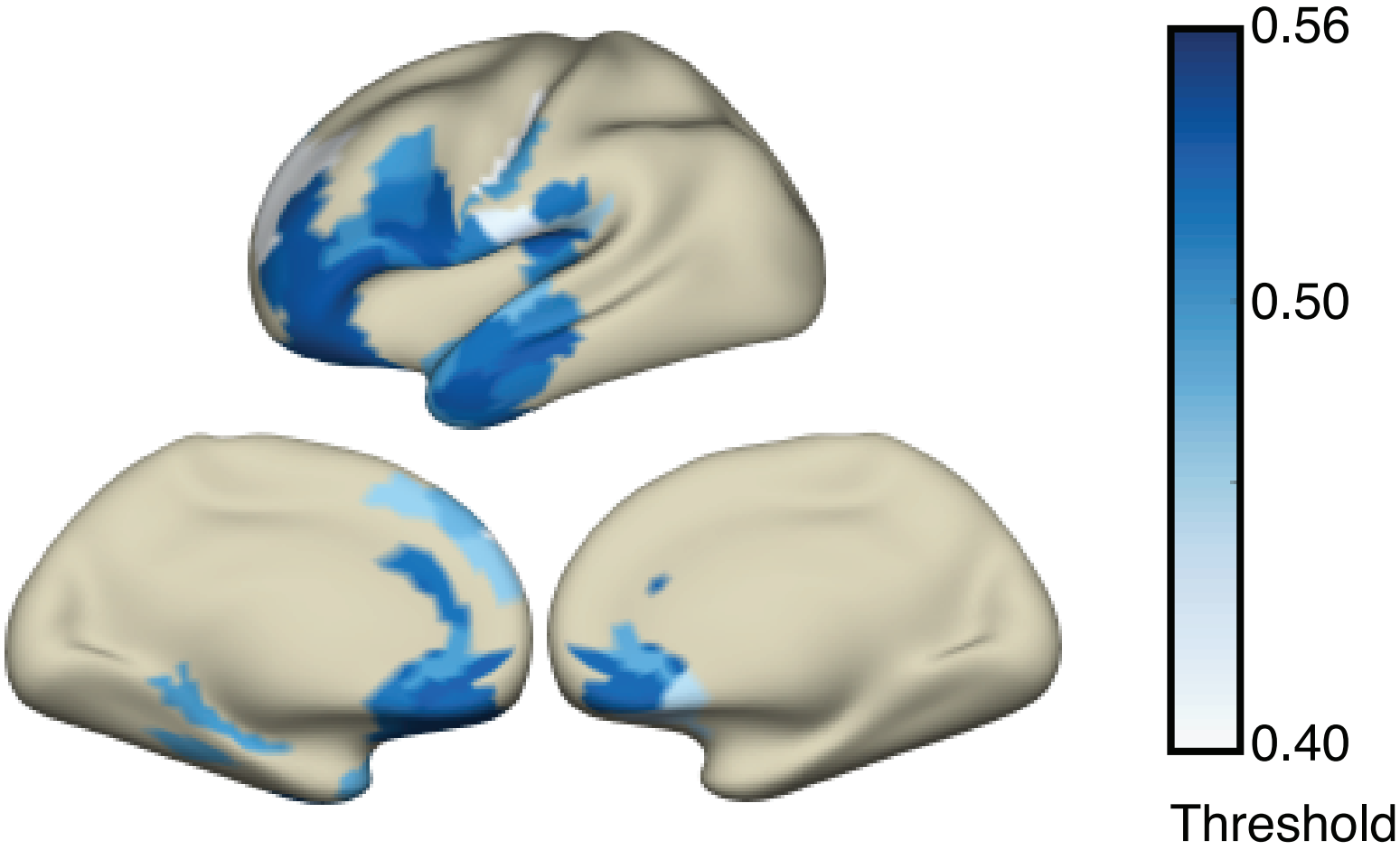
36. B. D. Van Veen, W. van Drongelen, M. Yuchtman, and A. Suzuki. Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. *IEEE Transactions on Biomedical Engineering*, 44(9):867–880, 1997.

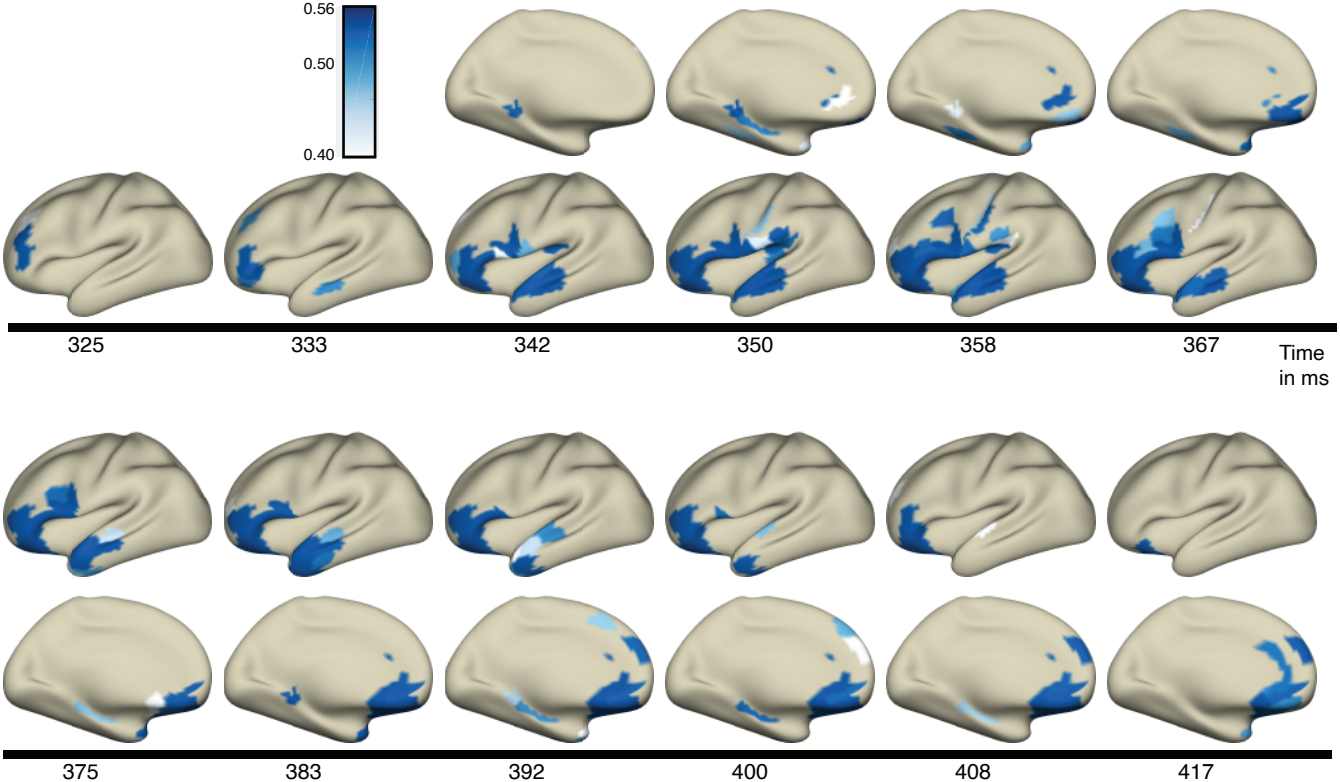
37. J. Vartiainen, T. Parviainen, and R. Salmelin. Spatiotemporal Convergence of Semantic Processing in Reading and Speech Perception. *Journal of Neuroscience*, 29(29):9271–9280, 2009.

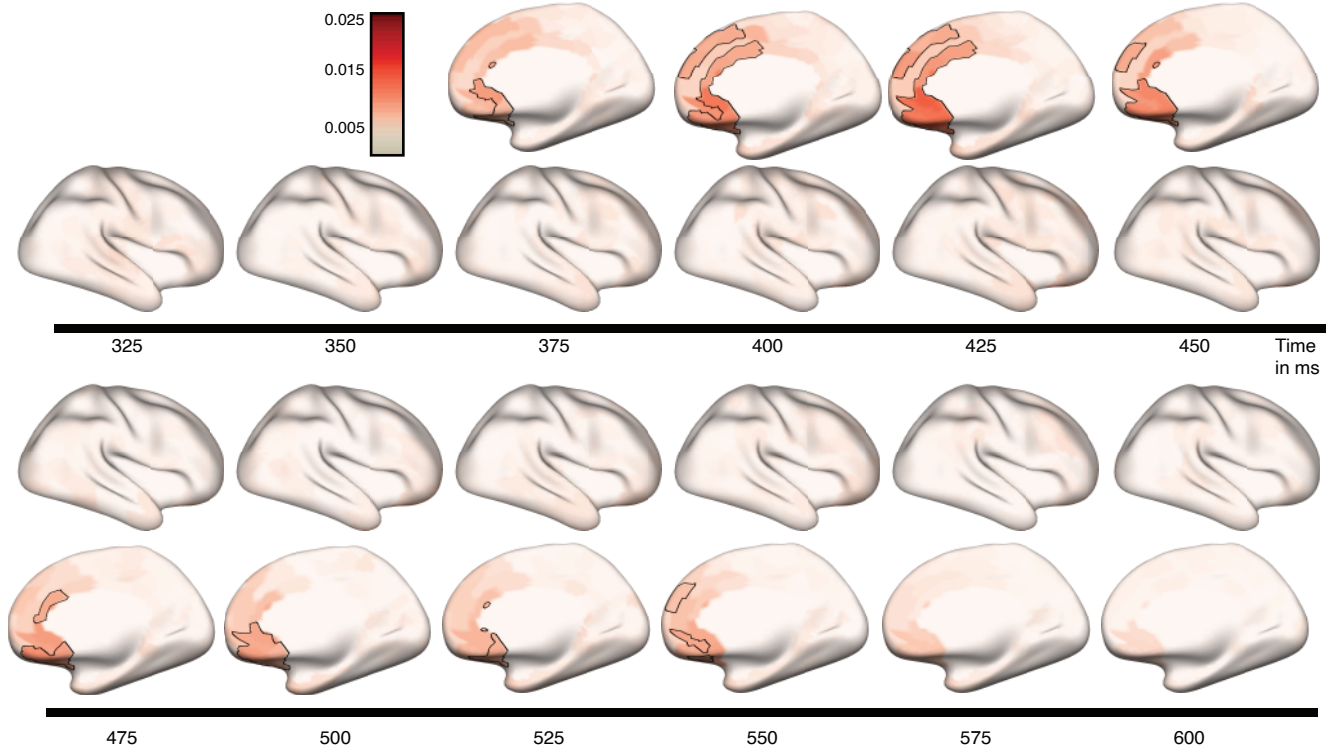
38. M. Vigneau, V. Beaucousin, B. Mazoyer, M. Vigneau, N. Tzourio-Mazoyer, L. Petit, G. Jobard, P.-Y. Hervé, F. Crivello, L. Zago, and E. Mellet. What is right-hemisphere contribution to phonological, lexico-semantic, and sentence processing? *NeuroImage*, 54(1):577–593, 2010.

39. K. Weber, E. F. Lau, B. Stillerman, and G. R. Kuperberg. The Yin and the Yang of prediction: An fMRI study of semantic predictive processing. *PLoS ONE*, 11(3):1–25, 2016.


0.025
0.015
0.005
Time
in ms
250
267
283
300
317
330
350
367
383
400
417
433
450
467
483
500
517
533
550
567
583
600
617
633
650
667
683
700
708
600
617


0.56
0.50
0.40
Threshold


0.56
0.50
0.40
325
333
342
350
358
367
Time
in ms
375
383
392
400
408
417


0.025
0.015
0.005
325
350
375
400
425
450
Time in ms
475
500
525
550
575
600