Behavioral/Cognitive

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] and Jan-Mathijs Schoffelen[1]
[1]Donders Institute, Radboud University, 6525EN Nijmegen, The Netherlands, and [2]Max Planck Institute for Psycholinguistics, 6525XD Nijmegen, The Netherlands

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 200 human subjects (102 males) 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 included not only areas previously implicated in higher-level language processing, such as left prefrontal, superior and 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 indicate not only the involvement of a large network of brain areas in supramodal language processing but also that the linguistic information contained in the unfolding sentences modulates brain activity in a word-specific manner across subjects.

Key words: amodal; auditory; canonical correlation analysis; Magnetoencephalography; MEG; visual

Significance Statement

The brain can extract meaning from written and spoken messages alike. This requires activity of both brain circuits capable of processing sensory modality-specific aspects of the input signals as well as coordinated brain activity to extract modality-independent meaning from the input. Using traditional methods, it is difficult to disentangle modality-specific activation from modality-independent activation. In this work, we developed and applied a multivariate methodology that allows for a direct quantification of sensory modality-independent brain activity, revealing fast activation of a wide network of brain areas, both including and extending beyond the core network for language.

Introduction

Language can be realized in different modalities: among others, through writing or speech. Depending on whether the sensory input modality is visual or auditory, different networks of brain areas are activated to derive meaning from the stimulus. In addition to different brain circuits being recruited to 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 co-articulated signals to be parsed into individual words. Written text has the advantage of instantaneous availability of full information compared with the temporally unfolding nature of speech. These 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 (Hagoort and Brown, 2000).

Regardless of low-level differences, the same core message can be conveyed in either modality. Therefore, language-processing models of the past and present (Geschwind, 1979; Hickok and Poeppel, 2007; Hagoort, 2017) not only include early sensory (up to 200 ms) processing steps but also contain late (200–500 ms), more abstract, and supposedly supramodal processing steps. Whereas early processing is largely unimodal and supported by

Received Sept. 20, 2019; revised Feb. 6, 2020; accepted Feb. 13, 2020.

Author contributions: A.H., P.H., and J.-M.S. designed research; A.H. and J.-M.S. performed research; J.-M.S. contributed unpublished reagents/analytic tools; S.A., A.M., and J.-M.S. analyzed data; S.A. and J.-M.S. wrote the paper.

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

The authors declare no competing financial interests.

Correspondence should be addressed to Sophie Arana at s.arana@donders.ru.nl.

https://doi.org/10.1523/JNEUROSCI.2271-19.2020

brain regions in the respective primary and associative sensory areas, later processes (for instance, lexical retrieval and integration) that activate several areas within the temporofrontal language network are assumed to do so independent of modality.

To gain insight into the location and timing of brain processes representing this latter, higher-order processing of the linguistic content, researchers so far have relied on carefully manipulated experimental conditions. As a result, our current understanding of how the brain processes language across different modalities reflects a large variety of tasks [semantic decision task (Chee et al., 1999), error detection task (Carpentier et al., 2001; Constable et al., 2004), passive hearing/listening (Jobard et al., 2007), size judgment (Marinkovic et al., 2003)] and stimulus material [words (Chee et al., 1999), sentences (Bemis and Pylkkänen, 2013), and stories (Berl et al., 2010; Regev et al., 2013; Deniz et al., 2019)]. Despite this wealth of experimental findings and resulting insights, an important interpretational limitation stems from the fact that the majority of studies use modality-specific low-level baseline conditions (tone pairs and lines, spectrally rotated speech and false fonts, nonwords, white noise; Lindenberg and Scheef, 2007) to remove the sensory component of the processing. It is difficult to assess how far such baselines are comparable across auditory and visual experiments. Recent fMRI work has demonstrated sensory modality-independent brain activity by directly comparing the BOLD response across visual and auditory presentations (Regev et al., 2013; Deniz et al., 2019). Yet, fMRI signals lack the temporal resolution to allow for a temporally sufficiently fine-grained investigation of the response to individual words.

Few studies have used magnetoencephalography (MEG) to study supramodal brain activity, and all were based on event-related averaging (Marinkovic et al., 2003; Vartiainen et al., 2009; Bemis and Pylkkänen, 2013; Papanicolaou et al., 2017). Averaged measures capture only generic components in the neural response. Although generic components make a large contribution to the neural activity measured during language processing, there also exists meaningful variability in the neural response that is stimulus-specific and robust (Ben-Yakov et al., 2012). A complete analysis of the supramodal language network needs to tap into these subtle variations as well.

Here, we overcome previous limitations by achieving a direct comparison without relying on modality-specific baseline conditions, leveraging word-by-word variation in the brain response. Using MEG signals from 200 subjects, we performed a quantitative assessment of the sensory modality-independent brain activity following word onset during sentence processing. The MEG data form part of a large publicly available dataset (Schoffelen et al., 2019) and have been used in other publications (Lam et al., 2016, 2018; Schoffelen et al., 2017; Hultén et al., 2019). 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 provide a quantitative confirmation of earlier findings in a large study sample. Importantly, they also indicate that supramodal linguistic information conveyed by the individual words in sentence context leads to subtle fluctuations in brain activation patterns that are correlated across different subjects.

Materials and Methods

Subjects. A total of 204 native Dutch speakers (102 males), with an age range of 18–33 years (mean 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 our analysis pipeline. 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 (Central Committee on Research involving Human Subjects 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 and received monetary compensation for their participation.

Experimental design. The subjects were seated comfortably in a magnetically shielded room and presented with Dutch sentences. From the total stimulus set of 360 sentences, six subsets of 120 sentences were created. This resulted in six different groups of subjects who were presented with the same subset of stimuli, although in a different (randomized) order with some overlap of items between groups. Within each group of subjects, half of them performed the task in only the visual modality, the other half in only the auditory modality. In the visual modality, words were presented sequentially on a back-projection screen and placed in front of them (vertical refresh rate of 60 Hz) at the center of the screen within a visual angle of 4°, in a black monospaced font on a gray background. Each word was separated by an empty screen for 300 ms, and the intersentence interval was jittered between 3200 and 4200 ms. Mean duration of words was 351 ms (minimum 300 ms and maximum 1400 ms), depending on word length. The median duration of whole sentences was 8.3 s (range 6.2–12 s). Auditory sentences had a median duration of 4.2 s (range 2.8–6.0 s) and were spoken at a natural pace. The duration of each visual word was determined by the following quantities: (i) the total duration of the audio version of the sentence/word list (audiodur), (ii) the number of words in the sentence (nwords), (iii) the number of letters per word (nletters), and (iv) the total number of letters in the sentence (sumnletters). Specifically, the duration (in ms) of a single word was defined as follows: (nletters/sumnletters) * (audiodur + 2000 − 150 * nwords), where item-independent parameters were chosen for the optimal balance between readability and "natural" reading pace. In the auditory task, the stimuli were presented via plastic tubes and ear pieces to both ears. Before the experiment, the hearing threshold was determined individually, and the stimuli were then presented at an intensity of 50 dB above the hearing threshold. A female native Dutch speaker recorded the auditory versions of the stimuli. The audio files were recorded in stereo at 44,100 Hz. During postprocessing, the audio files were low-pass filtered at 8500 Hz and normalized so that all audio files had the same peak amplitude and same peak intensity. All stimuli were presented using the Presentation software (version 16.0, Neurobehavioral Systems). Sentences were presented in small blocks of five sentences each, along with blocks containing scrambled sentences, which were not used here. See Lam et al. (2016) for more details about the stimulus material used. To check for compliance, 20% of the trials were randomly followed by a yes/no question about the content of the previous sentence/word list. Half of the questions addressed the content of the sentence (e.g., "Did grandma give a cookie to the girl?"), whereas the other half addressed one of the main content words (e.g., "Was the word 'grandma' mentioned?"). Subjects answered the question by pressing a button for "yes"/"no" with their left index and middle fingers, respectively.

MEG data acquisition and structural imaging. MEG data were collected with a 275 axial gradiometer system (CTF). The signals were analog low-pass filtered at 300 Hz and digitized at a sampling frequency of 1200 Hz. The subject's head was registered to the MEG sensor array using three coils attached to the subject's head (nasion, and left and right ear canals). Throughout the measurement, the head position was continuously monitored using custom software (Stolk et al., 2013). During breaks, the subject was allowed to reposition to the original position if needed. Participants were able to maintain a head position within 5 mm of their original position. Three bipolar Ag/AgCl electrode pairs were used to measure the horizontal and vertical electrooculograms and the electrocardiogram.

A T1-weighted magnetization-prepared rapid acquisition gradient echo pulse sequence was used for the structural images, with the following parameters: volume TR = 2300 ms, TE = 3.03 ms, 8° flip angle, 1 slab, slice matrix size = 256 × 256, slice thickness = 1 mm, field of view = 256

mm, isotropic voxel size = 1.0 × 1.0 × 1.0 mm. A vitamin E capsule was placed as a fiducial marker behind the right ear to allow a visual identification of left–right consistency.

Preprocessing. 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 due to eye movements, muscular activity, and superconducting quantum interference device jumps were replaced by Not a Number before further analysis. Because 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. 1*A*). 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, whereas 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 was 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 down sampled to 120 Hz.

Source reconstruction. We used linearly constrained minimum variance beamforming (Van Veen et al., 1997) to reconstruct activity onto a parcellated cortically constrained source model. For this, we computed the covariance matrix between all MEG sensor pairs as the average covariance matrix across the cleaned single-trial covariance estimates. This covariance matrix was used in combination with the forward model, defined on a set of 8196 locations on the subject-specific reconstruction of the cortical sheet to generate a set of spatial filters, one filter per dipole location. Individual cortical sheets were generated with the FreeSurfer package (version 5.1, http://surfer.nmr.mgh.harvard.edu; Dale et al., 1999), coregistered to a template with a surface-based coregistration approach using Caret software (Van Essen Laboratory at the Washington University School of Medicine) (Van Essen et al., 2001), and subsequently down sampled to 8196 nodes using the MNE software (https://mne.tools/stable/index.html; Gramfort et al., 2014). The forward model was computed using the FieldTrip single-shell method (Nolte, 2003), where the required brain–skull boundary was obtained from the subject-specific T1-weighted anatomical images. We further reduced the dimensionality of the data to 191 parcels per hemisphere (Schoffelen et al., 2017). For each parcel, we obtained a parcel-specific spatial filter as follows. We concatenated the spatial filters of the dipoles comprising the parcel and used the concatenated spatial filter to obtain a set of time courses of the reconstructed signal at each parcel. Next, we performed a principal component analysis and selected for each parcel the first five spatial components explaining most of the variance in the signal.

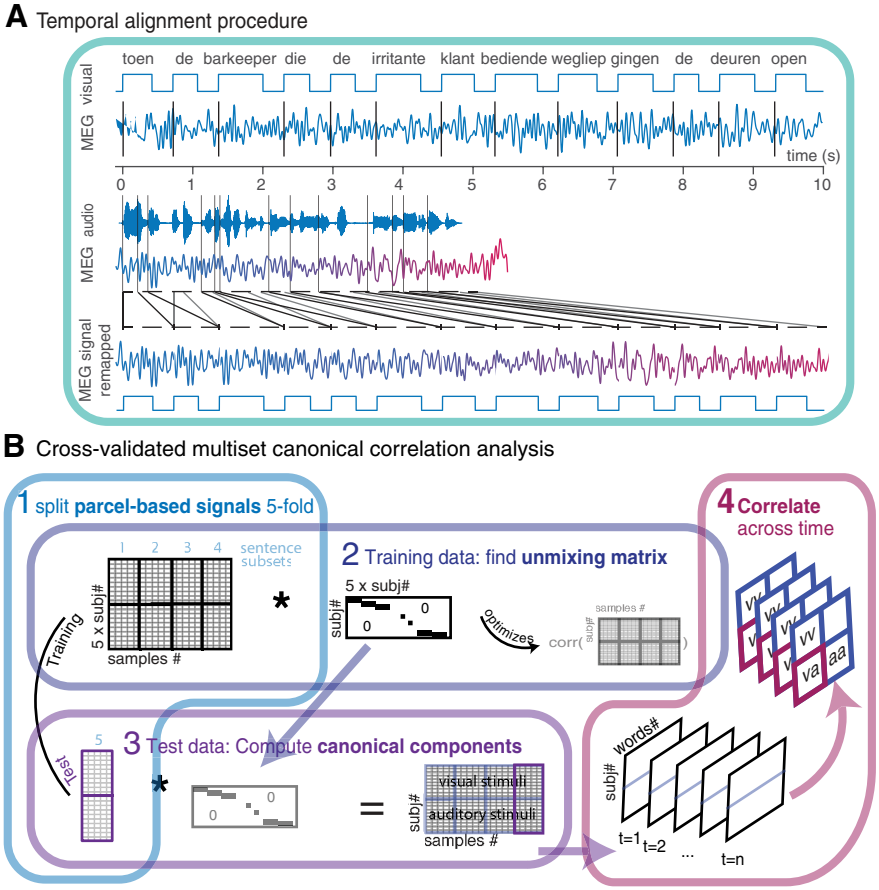

Figure 1. Analysis pipeline. ***A***, Temporal alignment procedure. MEG signals of auditory and visual subjects differed in length due to different presentation rates. To achieve alignment between signals of auditory and visual subjects, auditory signals were epoched into overlapping segments. Each segment's 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 points for the multiset 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, whereas 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 was 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) was interpreted as quantifying supramodal activation.

Multiset canonical correlation analysis. Multiset canonical correlation analysis (MCCA; de Cheveigné et al., 2019; Parra, 2018) 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; Kettenring, 1971), which extends the method to multiple sets—here, multiple subjects. In our case, we found linear combinations of the five spatial components for each of two subjects, so that the correlation across time between those subjects was maximized. Because we had more than two subjects, we found each subject's own linear combination, which maxi-

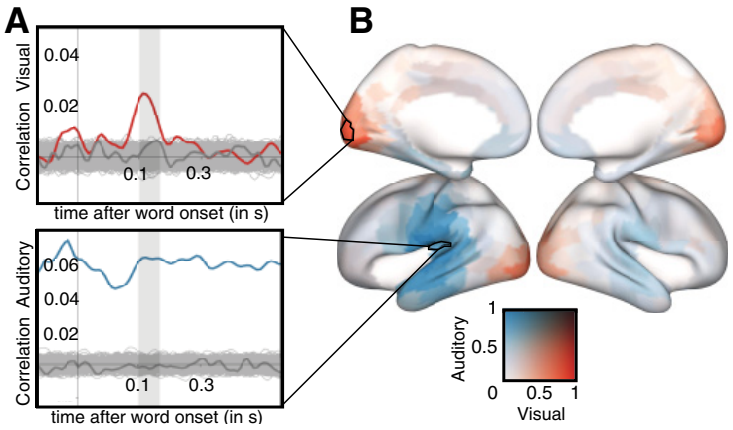

Figure 2. Specificity of the within-modality correlated activity patterns. ***A***, Time-resolved correlation values averaged across all visual subject pairings for a parcel in the left primary visual cortex (top) and all auditory subject pairings for a parcel in the left primary auditory cortex (bottom) before (dark-gray line) and after MCCA (blue and red lines). Light-gray 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 time point if possible. ***B***, Cortical map of the spatial distribution of correlations, comparing visual modality subject pairs (red) with auditory modality subject pairs (blue). Correlation strength is expressed as the Pearson correlation coefficient averaged over a time window from 150 to 200 ms after word onset and normalized by the maximum value of that window.

mized the correlation across time between all subjects from both modality groups (auditory and visual stimulation). Following Parra (2018), we obtained the optimal projection as the eigenvector with the largest eigenvalue of a square matrix $D^{-1}R$, where R and D are square matrices:

$$R = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,N} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ a_{N,1} & a_{N,2} & \cdots & a_{N,N} \end{pmatrix},$$

$$D = \begin{pmatrix} a_{1,1} & 0 & \cdots & 0 \\ 0 & a_{2,2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{N,N} \end{pmatrix}$$

Where a^{lk} are cross-covariance matrices between subject pairs, and D contains only the diagonal blocks of within-subject covariances (Parra, 2018). In our case, the cross-covariance matrices were of size 5×5, containing the cross-covariance between all five spatial components for a given subject (pair). The cross-correlation was computed across time points for each sentence and subsequently averaged across sentences. It is important to note here that the canonical variates resulting from the optimal projection do not reflect sentence averages anymore but have the same temporal resolution as the original source signals. CCA is prone to overfitting and known to be unstable (Dinga et al., 2019). For reliable CCA estimates, the number of samples should be much larger than the number of features (i.e., a sample-to-feature ratio of 20/1 is recommended; Stevens, 2012). We estimated the canonical variables over concatenated data, which included between 756 and 1453 samples per sentence compared with only five features (spatial components), which provides a decent sample-to-feature ratio. Further, we estimated our canonical variables out of sample using fivefold cross-validation to limit overfitting. We randomly split all sentences into five subsets, estimating projections on 96 sentences and applying them to the 24 left out sentences (Fig. 1*B*).

Statistical analysis. As per the study design, the subjects were assigned to one of six stimulus sets. Different groups of subjects were presented with different sets of sentences. Because MCCA relies on commonalities across datasets, we could only combine data from subjects who received the exact same stimulation. We therefore applied MCCA for each subgroup of subjects who listened to or saw the same stimuli separately. Initially, we constrained our analysis to the first set of 33 subjects (henceforth exploratory dataset). 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, and verbs) for subsequent steps. Before computing the correlation, we subtracted the mean across samples. For each pair of subjects, we computed the correlation between two sets of observations [i.e., a pair of vectors with each data point reflecting the subject-specific neural signal for each of the individual words (lexical items), at a given time point relative to word onset, and at a given cortical location]. Correlation coefficients of cross-modality pairings—that is, correlations between subjects reading and subjects listening to the sentences—were interpreted as capturing supramodal processing. We used a permutation test with clustering over time and space (parcels) for familywise error rate correction for statistical inference (Maris and Oostenveld, 2007). We used 1000 randomizations of the epoched words, the default maximum-sum cluster statistic, and a minimum spatial neighbor count of 0 and estimated individual thresholds from the randomization distribution for each parcel time point based on a cluster-forming α of 0.01 and a one-sided test. To this end, we randomized word order for the source-reconstructed parcel time series of the auditory subjects to test for exchangeability of the exact word pairing across sensory modalities. By destroying the one-to-one mapping of individual lexical items, the null distribution allowed for a distinction between individual item-specific shared variance and shared variance due to a more generic response. We also computed modality-specific responses as a quality check of the analysis pipeline given the well known spatiotemporal activity patterns of early sensory brain areas. For this, we averaged correlation across either only pairs of subjects reading or only pairs of subjects listening. These correlations were not constrained to content words but computed across all words. For statistical inference, we again used a permutation test with the same parameters as described earlier. This time, however, we randomized word order for the source-reconstructed parcel time series of both auditory and visual subjects, thereby destroying the one-to-one mapping of individual items within both modalities.

Finally, we analyzed the remaining sets of subjects (confirmatory dataset) using the analysis pipeline described earlier. We evaluated the overlap in the results across all six subgroups using information prevalence inference (Allefeld et al., 2016). Prevalence inference allows formulation of 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). 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 used the smallest observed average correlation across subgroups as the second-level test statistic. We then tested the majority null hypothesis of the prevalence of the effect being smaller than 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 in time and space (parcels). 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 Figs. 7 and 8 for cortical maps showing thresholds averaged and per time point).

To ensure that MCCA as a preprocessing step did not artificially increase correlations between cross-modality subjects, we conducted an additional control analysis on the exploratory dataset. For this, we addi-

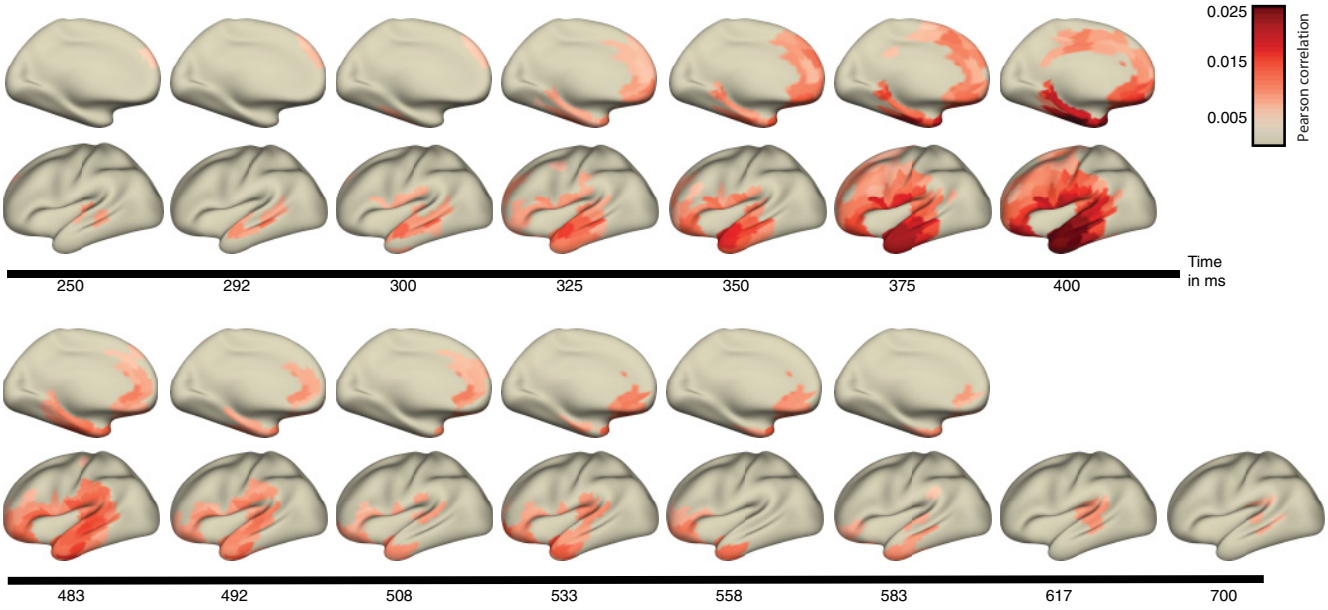

Figure 3. Supramodal correlated activity patterns. Time-resolved spatial maps of supramodal correlated activity patterns (averaged over all possible cross-modal subject pairings) in the left hemisphere. Medial views of the brain surface are depicted in the first and third rows, lateral views in the second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

tionally tested the observed correlation patterns computed on all words (both function and content words) against a null distribution obtained by permuting sentence order 500 times and, importantly, doing this before MCCA. This permutation was not fully unconstrained, because we aimed at aligning sentences across modalities with the same number of words to avoid loss of data and to preserve ordinal word position. Thus, we did a random pairing between sentences with the same number of words after binning the sentences according to their word count. Sentences consisting of 9, 14, or 15 words were infrequent, with fewer than five occurrences each. After each permutation, we performed the temporal alignment between sensory modalities (aligning the word onsets), followed by cross-validated MCCA and computation of the time-resolved correlations of cross-modal subject pairs. Due to the long computation time of the canonical variates, we created this null distribution for the exploratory data only. Results from this additional, conservative permutation test can be found in Figure 4.

Code accessibility. All analyses were performed with custom-written MATLAB scripts (MathWorks) and FieldTrip (Oostenveld et al., 2011), and the corresponding code is available upon request.

Results

Modality-specific activation

We first quantified the similarity between different subjects' brain response within only the exploratory dataset (33 subjects) by correlating word-by-word fluctuations in brain activity between all possible pairs of subjects. 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 displayed in Figure 2, early sensory cortical areas only showed correlated activity for the group of subjects receiving the stimuli in the corresponding sensory modality, for the visual (red) and auditory (blue) modalities. We found that MCCA is a crucial analysis step to reveal meaningful intersubject correlations. Only after MCCA does cortical activity in visual and auditory areas become significantly correlated (cluster-based permutation test, $p = 0.001$ for both) across those subjects performing the task in the visual or auditory domain, respectively (Fig. 2*A*).

Supramodal activation patterns

We averaged between-subject correlations over all cross-modal subject pairings as a metric for supramodal activity. We observed significant supramodal correlated activation patterns in mostly left-lateralized cortical areas (Fig. 3; cluster-based permutation test, $p = 0.001$). The effect had a large spatial and temporal extent, becoming apparent as early as 250 ms and lasting until 700 ms after word onset. Parcels in the middle superior temporal gyrus (STG) contributed 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 the ventral temporal cortex followed a similar temporal and spatial pattern, with supramodal correlations starting out more posterior at ~292 ms and evolution toward the middle anterior temporal lobe at 308 ms. Other areas that expressed supramodal activity at relatively early time points were the medial prefrontal cortex and primary auditory cortex (250 ms), followed by subcentral parietal regions and supramarginal gyrus at ~300 ms, and finally the dorsolateral frontal cortex (DLFC; 325 ms). By the time 375 ms had passed, the entire orbitofrontal cortex, anterior frontal cortex and DLFC, as well as inferior frontal gyrus (IFG) showed strong supramodal subject correlation. Supramodal activation in the frontal lobe further extended toward posterior regions, including the precentral and postcentral gyrus. At ~400 ms, supramodal subject correlation in the anterior temporal pole reached its peak. In addition to the lateral cortical areas, correlated activity also extended to the left dorsal and ventral anterior cingulate cortex (ACC) as well as the left fusiform gyrus. The spatiotemporal patterns of supramodal activation described so far are robust, also when ordinal word position and MCCA overfitting were controlled for in the statistical evaluation (Fig. 4; cluster-based permutation test, $p = 0.002$).

Prevalence inference

Our confirmatory analysis combined all six datasets and tested whether the spatiotemporal patterns observed in the exploratory dataset would generalize to the population. For those parcels at

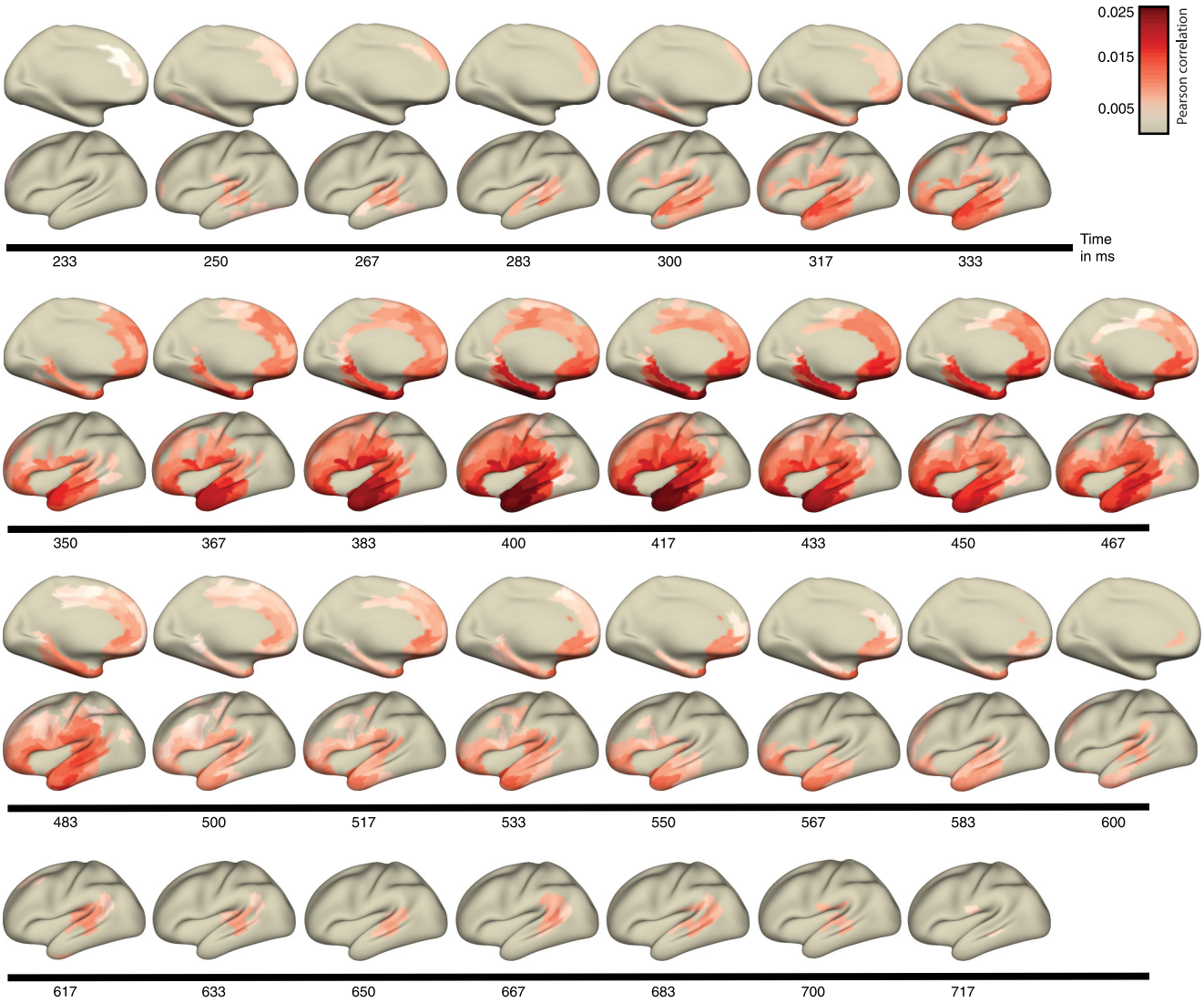

Figure 4. Significant supramodal correlated activity patterns as assessed by an additional permutation test. Time-resolved spatial maps of supramodal correlated activity patterns were estimated across all words (both function and content words) and masked by significance as evaluated by a more conservative shuffling procedure. We permuted sentence order 500 times before MCCA to control for artificially increased correlations due to overfitting. Shown here are correlations at several time points for the left hemisphere. Medial views of the brain surface are depicted in the first, third, and fifth rows, lateral views in the second, fourth, sixth, and seventh rows. Color codes are for strength of correlation. Colored parcels were most strongly correlated between cross-modal subject pairs.

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 (Fig. 5). In addition, we evaluated the majority null hypothesis of whether, in the majority of subgroups in the population, the data contain an effect (threshold > 0.5, significant parcels under the majority null hypothesis outlined in black in Fig. 5*B*).

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 40 parcels per time point on average (between 325 and 617 ms after onset, SD = 31.48). 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 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. 5, black outline; see also Fig. 6 for results in the right medial hemisphere). For parcels at which the global null hypothesis could be rejected, the average largest lower bound $\gamma 0$ at which the prevalence null hypothesis can be rejected is shown in Figures 7 and 8. Compared with 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 42%, on average, 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 the STG occurred almost 100 ms earlier as compared with the IFG. Based on the confirmatory dataset, however, supramodal correlated activation in the IFG and STG appeared almost simultaneously. Finally, the exploratory analysis revealed supramodal activation in primary

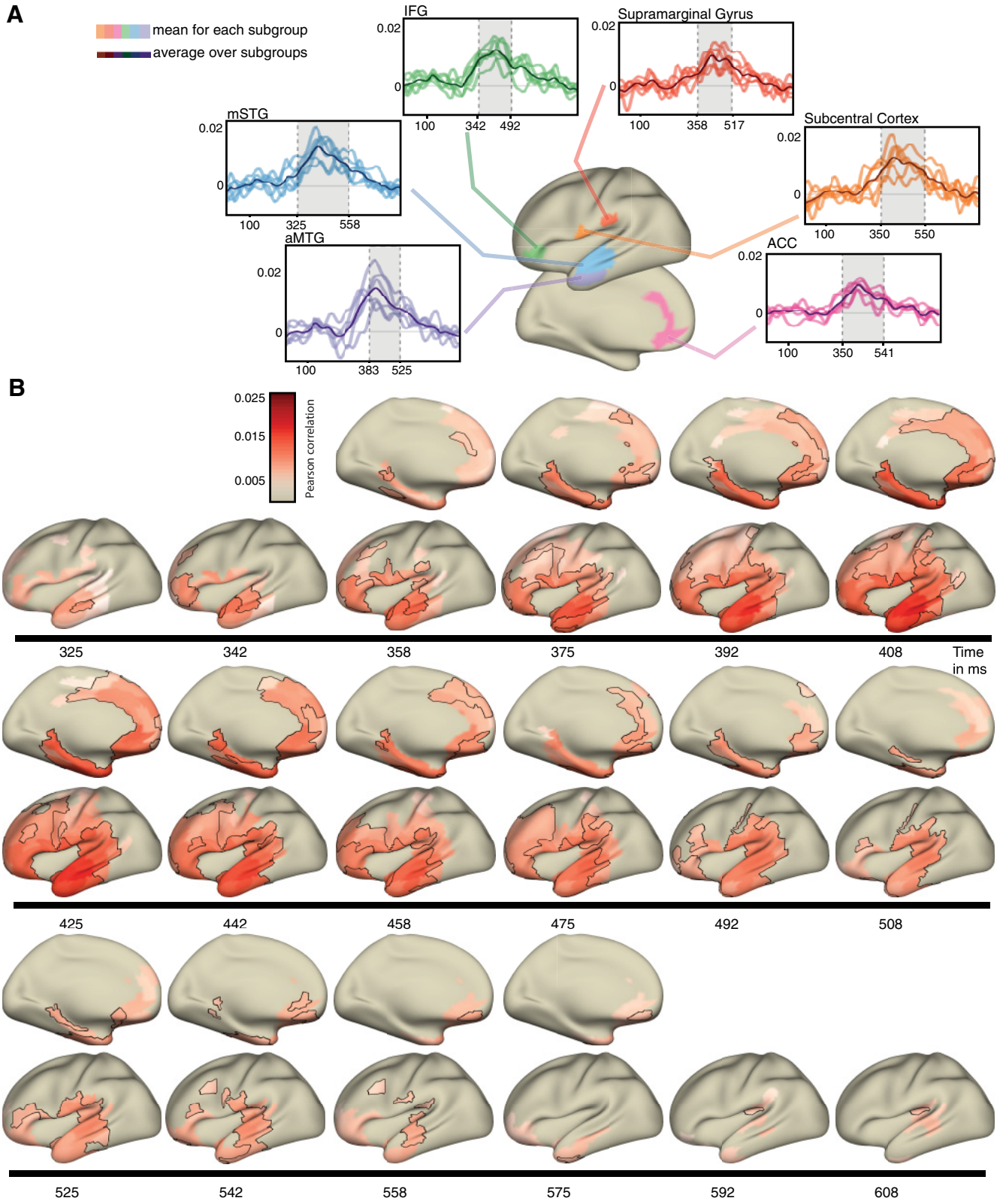

Figure 5. Supramodal correlated activity patterns consistent across the majority of datasets. Supramodal correlated activity patterns of word-specific activity were consistent across the majority of datasets. ***A***, Averaged correlation time courses (mean over all possible cross-modal subject pairings) are shown for selected parcels in the IFG (green), supramarginal gyrus (red), subcentral cortex (orange), 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). Gray 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 rows, lateral views in the second, fourth, and sixth rows. 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 are 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.

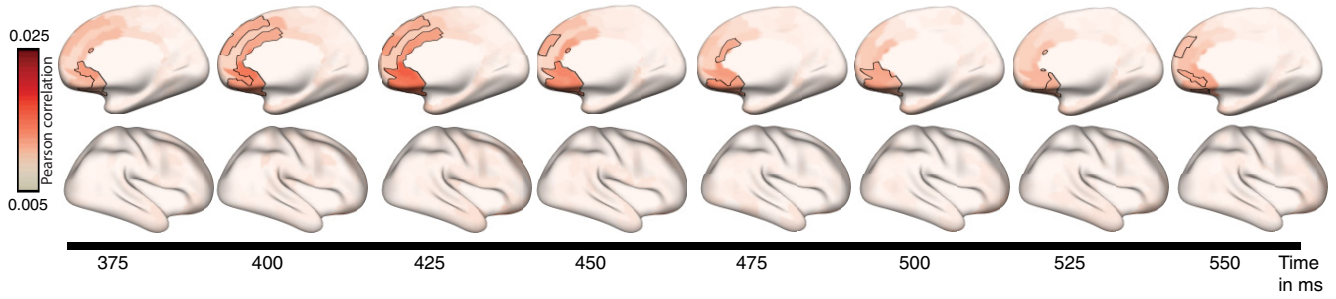

Figure 6. Time-resolved spatial maps of cross-modal correlations for the right hemisphere. The average correlation over all six datasets is shown. Color codes are for strength of correlation. In addition, the parcels at which the majority null hypothesis according to prevalence inference could be rejected are outlined in black.

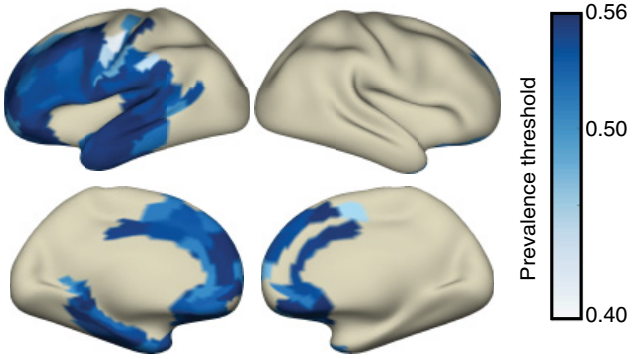

Figure 7. Cortical map of maximum prevalence threshold $\gamma 0$ per parcel. For those parcels at which the global null hypothesis could be rejected, the mean (over time) maximum threshold is plotted, for which the null hypothesis can be rejected ($\alpha = 0.05$). Given the sample size of six datasets, the number of second-level permutations, and a significance level of $\alpha = 0.05$, the maximal possible threshold that can be reached is 0.5633.

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 with the cluster-based permutation approach on the exploratory data. Nevertheless, widely overlapping anatomical regions were indicated by both analyses, encompassing the dorsolateral frontal gyrus and the middle and superior parts of the temporal lobe at first, and the inferior frontal and orbitofrontal cortex as well as the anterior temporal lobe later.

Discussion

Our aim was to quantify similarities of the brain response across reading and listening at a fine temporal scale. To this end, we correlated word-by-word fluctuations in the neural activity across subjects receiving either auditory or visual stimulation. We identified a widespread left-lateralized brain network, activated independently of modality starting 325 ms after word onset. Importantly, dividing our large study sample into six subsets, we could directly quantify the consistency and generalizability of these activity patterns. The spatial distribution of the supramodal activation is in line with the known involvement of left-hemispheric areas, including parts of the left temporal cortex, left inferior parietal lobe, as well as the prefrontal cortex (Chee et al., 1999; Homae et al., 2002; Constable et al., 2004; Spitsyna et al., 2006; Lindenberg and Scheef, 2007; Vigneau et al., 2011; Braze et al., 2011; Liuzzi et al., 2017). The involvement of both the STG and IFG fits predictions from the memory, unification, and control model (MUC), in which activity reverberating within a posterior-frontal network (Baggio and Hagoort, 2011; Hagoort, 2017) is thought to be crucial for language processing. According to the MUC model, temporal and parietal areas support the retrieval of lexical information, whereas unification processes are supported by the inferior frontal cortex. Bidirectional communication (Schoffelen et al., 2017) between these areas is facilitated by white matter connections. We observe that temporal areas are supramodally activated at the earliest time points and sustain activation the longest compared with other regions. Over time, supramodal activation spreads from the middle and posterior left STG to the anterior temporal pole. This rapid progression of activity from posterior to anterior regions mirrors previous observations (Marinkovic et al., 2003; Vartiainen et al., 2009), adding to those findings a direct quantitative comparison of the supramodal brain activity.

Beyond the core language network and the single word level

We observed modality-independent activity in the dorsal frontal cortex in addition to more widely reported inferior parts of the frontal cortex (Michael et al., 2001; Homae et al., 2002; Marinkovic et al., 2003; Constable et al., 2004; Jobard et al., 2007; Lindenberg and Scheef, 2007). This could be because we used linguistically rich sentence material of varying syntactic complexity as opposed to single words (Chee et al., 1999; Booth et al., 2002; Marinkovic et al., 2003; Vartiainen et al., 2009; Liuzzi et al., 2017) or short phrases (Carpentier et al., 2001; Braze et al., 2011; Bemis and Pylkkänen, 2013). Indeed, discrepancies with respect to frontal lobe involvement in modality-independent processing seem to mainly arise from differences in stimulus material and task demands (Braze et al., 2011). A recent meta-analysis has identified that more complex syntax robustly activates dorsal parts of the left IFG (Hagoort and Indefrey, 2014). Further, a previously published analysis of these MEG data showed the DLFC to be sensitive to sentence progression effects (Hultén et al., 2019). Two previous fMRI studies using narratives (Regev et al., 2013; Deniz et al., 2019) added to the debate. Regev et al. (2013) correlated BOLD responses evoked by different modalities. They reported supramodal activation in the left frontal lobe, extending beyond inferior frontal regions. Deniz et al. (2019) studied modality-independent brain areas by modeling semantic features of the stimulus in one modality and used the model to predict the BOLD signal in the other modality. They reported BOLD signals in the prefrontal cortex to be well predicted across modalities. In summary, although complex stimuli consistently activate prefrontal areas beyond the inferior frontal cortex, the exact stimulus features which cause this supramodal activation are still debated.

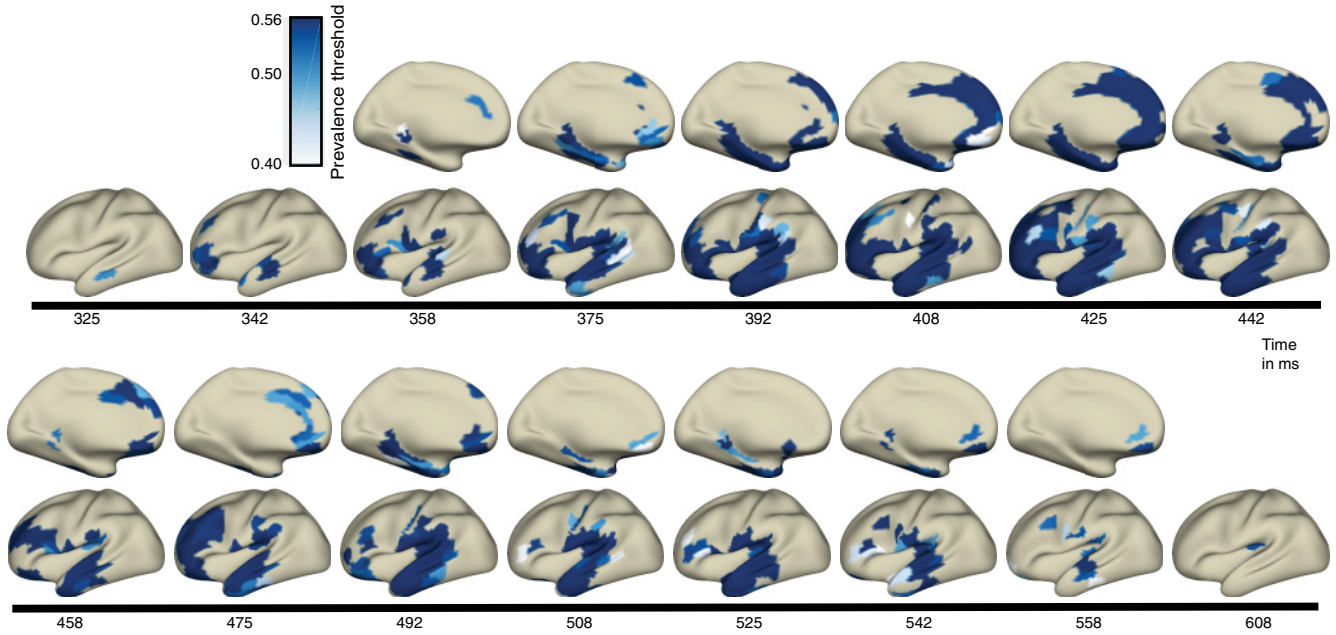

Figure 8. Cortical map of prevalence threshold $\gamma 0$. For those parcels at which the global null hypothesis could be rejected, the maximum threshold is plotted, for which the null hypothesis can be rejected at a level of $\alpha = 0.05$.

Some previous studies using narratives and fMRI reported that supramodal activation was not restricted to the left hemisphere (Jobard et al., 2007; Regev et al., 2013; Deniz et al., 2019). It could be that the previously observed bilateral involvement is due to differences in context-based semantic processing during narratives, compared with the processing of isolated sentences in our experiment. Menenti et al. (2009) specifically contrasted BOLD activity in response to sentences presented within a neutral or a local context. The authors indeed reported the right frontal cortex to be more sensitive to local discourse context compared with its left-hemispheric homolog. Further research is needed to determine whether this effect of presence and absence of narrative thread similarly affects lateralization of brain activity in MEG.

Even though mostly restricted to the left hemisphere, our results also implicate extralinguistic areas in supramodal processing. Specifically, we find bilateral supramodal activation within the ACC. The ACC is a midline structure, forming part of a domain-general executive control network supporting language processing (Cattinelli et al., 2013; Hagoort, 2017). It is sensitive to statistical contingencies in the language input and thus might play a role in mediating learning and adaptation in response to predictive regularities in both local experimental as well as global environment (Weber et al., 2016). It should be noted that deep sources are normally poorly detectable in MEG (Hillebrand and Barnes, 2002), and we thus consider any interpretations with respect to the midline structures as tentative.

Supramodal orthography–phonology mapping

We observed supramodal activation in the postcentral and subcentral gyrus as well as the supramarginal gyrus, which coincides temporally with supramodal activation of the primary auditory cortex. Activity in the supramarginal gyrus has been repeatedly elicited by cross-modal tasks (Sliwinska et al., 2012), such as rhyming judgments to visually presented words (Booth et al., 2002), for which conversion between orthographic and phonological representations is likely needed. At the same time, postcentral and subcentral areas partly span articulatory motor and somatosensory areas for the mouth and tongue. Together, the supramodal 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.

Leveraging word-by-word variability of the neural response

Neuroelectric brain signals exhibit strong moment-to-moment variability. Whereas some of this variability is related to the experimental stimulation, and is therefore associated with specific cognitive activity, some of it is unrelated, ongoing neural activity. By applying MCCA across subjects, we reduced this type of noise and made subtle word-by-word fluctuations in the MEG signal interpretable. Comparing neural activity across subjects is challenging due to differing position or orientation of neuronal sources relative to the MEG sensors. We used parcellated MEG source reconstruction in combination with exact temporal alignment of individual sentences across subjects. This allowed for the extraction of signal components that are shared across subjects, thus reducing the intersubject spatial variability, which is commonly observed in more traditional (for instance, dipole fitting) procedures (Vartiainen et al., 2009). MCCA thus allowed us to more directly investigate time-resolved intersubject correlations and move beyond event-related averages (Marinkovic et al., 2003). Importantly, our analysis approach allows us to conclude that the identified supramodal activity is word-specific. Our findings therefore go beyond showing a general activation of those areas compared with baseline and rather reveal consistent word-by-word fluctuations of activation within the recruited areas.

Latency of supramodal processing

The temporal alignment procedure, 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 pro-

cessing latencies, which may also lead to latency differences in the activation of supramodal areas. For example, Marinkovic et al. (2003) reported shorter reaction times during the visual task, yet found earlier activity peaks for the auditory task in corresponding early sensory cortex and left anterior temporal lobe. In contrast, other work observed earlier anterior temporal lobe activation for visual compared with auditory stimulation (Bemis and Pylkkänen, 2013). Our results indicate a certain degree of overlap across modalities in the temporal window, within which supramodal cortical areas are activated. It is possible that we observed more temporally extensive activation—for instance, related to unification processes—because we used longer sentences. In addition, any overlap may have been amplified as a necessary consequence of the MCCA procedure. Evidently, correlations between signals from auditory subjects were boosted with less temporal specificity compared with visual subjects (Fig. 2*B*). This observation was unexpected and may be due to more continuous stimulation in the auditory experiment. As the sound of a spoken word unfolds, the timing at which it becomes uniquely recognizable will vary across the word. Thus, the distribution of information in the auditory signal is much more varied compared with the visual signal. 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 word-specific, and thus likely reflect more than just generic activation during language processing. Finally, we show that alignment of individual subject data through MCCA is a promising tool for investigating subtle word-in-context-specific modulations of brain activity in the language system.

References

Allefeld C, Görgen K, Haynes JD (2016) Valid population inference for information-based imaging: from the second-level *t*-test to prevalence inference. Neuroimage 141:378–392.

Baggio G, Hagoort P (2011) The balance between memory and unification in semantics: a dynamic account of the N400. Lang Cogn Process 26: 1338–1367.

Bemis DK, Pylkkänen L (2013) Basic linguistic composition recruits the left anterior temporal lobe and left angular gyrus during both listening and reading. Cereb Cortex 23:1859–1873.

Ben-Yakov A, Honey CJ, Lerner Y, Hasson U (2012) Loss of reliable temporal structure in event-related averaging of naturalistic stimuli. Neuroimage 63:501–506.

Berl MM, Duke ES, Mayo J, Rosenberger LR, Moore EN, VanMeter J, Ratner NB, Vaidya CJ, Gaillard WD (2010) Functional anatomy of listening and reading comprehension during development. Brain Lang 114:115–125.

Booth JR, Burman DD, Meyer JR, Gitelman DR, Parrish TB, Mesulam MM (2002) Modality independence of word comprehension. Hum Brain Mapp 16:251–261.

Braze D, Mencl WE, Tabor W, Pugh KR, Todd Constable R,T Fulbright RK, Magnuson JS, Van Dyke JA, Shankweiler DP (2011) Unification of sentence processing via ear and eye: an fMRI study. Cortex 47:416–431.

Carpentier A, Pugh KR, Westerveld M, Studholme C, Skrinjar O, Thompson JL, Spencer DD, Constable RT (2001) Functional MRI of language processing: dependence on input modality and temporal lobe epilepsy. Epilepsia 42:1241–1254.

Cattinelli I, Borghese NA, Gallucci M, Paulesu E (2013) Reading the reading brain: a new meta-analysis of functional imaging data on reading. J Neurolinguist 26:214–238.

Chee MW, O'Craven KM, Bergida R, Rosen BR, Savoy RL (1999) Auditory and visual word processing studied with fMRI. Hum Brain Mapp 7:15–28.

Constable RT, Pugh KR, Berroya E, Mencl WE, Westerveld M, Ni W, Shankweiler D (2004) Sentence complexity and input modality effects in sentence comprehension: an fMRI study. Neuroimage 22:11–21.

Dale AM, Fischl B, Sereno MI (1999) Cortical surface-based analysis: I. Segmentation and surface reconstruction. Neuroimage 9:179–194.

de Cheveigné A, Di Liberto GM, Arzounian D, Wong DDE, Hjortkjær J, Fuglsang S, Parra LC (2019) Multiway canonical correlation analysis of brain data. Neuroimage 186:728–740.

Deniz F, Nunez-Elizalde AO, Huth AG, Gallant JL (2019) The representation of semantic information across human cereb cortex during listening versus reading is invariant to stimulus modality. J Neurosci 39:7722–7736.

Dinga R, Schmaal L, Penninx BW, van Tol MJ, Veltman DJ, van Velzen L, Mennes M, van der Wee NJ, Marquand AF (2019) Evaluating the evidence for biotypes of depression: methodological replication and extension of Drysdale et al. (2017). Neuroimage: Clinical 22:101796.

Geschwind N (1979) Specializations of the human brain. Sci Am 2:180–201.

Gramfort A, Luessi M, Larson E, Engemann DA, Strohmeier D, Brodbeck C, Parkkonen L, Hämäläinen MS (2014) MNE software for processing MEG and EEG data. Neuroimage 86:446–460.

Hagoort P (2017) The core and beyond in the language-ready brain. Neurosci Biobehav Rev 81:194–204.

Hagoort P, Brown CM (2000) ERP effects of listening to speech compared to reading: the P600 to syntactic visual presentation. Neuropsychologia 38:1531–1549.

Hagoort P, Indefrey P (2014) The neurobiology of language beyond single words. Annu Rev Neurosci 37:347–362.

Hickok G, Poeppel D (2007) The cortical organization of speech processing. Nature 8:393–402.

Hillebrand A, Barnes GR (2002) A quantitative assessment of the sensitivity of whole-head MEG to activity in the adult human cortex. Neuroimage 16:638–650.

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

Hultén A, Schoffelen JM, Uddén J, Lam NHL, Hagoort P (2019) How the brain makes sense beyond the processing of single words - an MEG study. Neuroimage 186:586–594.

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

Kettenring JR (1971) Canonical analysis of several sets of variables. Biometrika 58:433–451.

Lam NHL, Schoffelen JM, 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.

Lam NH, Hultén A, Hagoort P, Schoffelen JM (2018) Robust neuronal oscillatory entrainment to speech displays individual variation in lateralisation. Lang Cogn Neurosci 33:943–954.

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

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

Marinkovic K, Dhond RP, Dale AM, Glessner M, Carr V, Halgren E (2003) Spatiotemporal dynamics of modality-specific and supramodal word processing. Neuron 38:487–497.

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

Menenti L, Petersson KM, Scheeringa R, Hagoort P (2009) When elephants fly: differential sensitivity of right and left inferior frontal gyri to discourse and world knowledge. J Cogn Neurosci 21:2358–2368.

Michael EB, Keller TA, Carpenter PA, Just MA (2001) fMRI investigation of sentence comprehension by eye and by ear: modality fingerprints on cognitive processes. Hum Brain Mapp 13:239–252.

Nolte G (2003) The magnetic lead field theorem in the quasi-static approximation and its use for magnetoenchephalography forward calculation in realistic volume conductors. Phys Med Biol 48:3637–3652.

Oostenveld R, Fries P, Maris E, Schoffelen JM (2011) FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput Intell Neurosci 2011:156869.

Papanicolaou AC, Kilintari M, Rezaie R, Narayana S, Babajani-Feremi A (2017) The role of the primary sensory cortices in early language processing. J Cogn Neurosci 29:1755–1765.

Parra LC (2018) Multi-set Canonical Correlation Analysis simply explained. arXiv:1802.03759.

Regev M, Honey CJ, Simony E, Hasson U (2013) Selective and invariant neural responses to spoken and written narratives. J Neurosci 33:15978–15988.

Schoffelen JM, Hultén A, Lam N, Marquand AF, Uddén J, Hagoort P (2017) Frequency-specific directed interactions in the human brain network for language. Proc Natl Acad Sci U S A 114:8083–8088.

Schoffelen JM, Oostenveld R, Lam NHL, Uddén J, Hultén A, Hagoort P (2019) A 204-subject multimodal neuroimaging dataset to study language processing. Sci Data 6:1–17.

Sliwinska MW, Khadilkar M, Campbell-Ratcliffe J, Quevenco F, Devlin JT (2012) Early and sustained supramarginal gyrus contributions to phonological processing. Front Psychol 3:1–10.

Spitsyna G, Warren JE, Scott SK, Turkheimer FE, Wise RJ (2006) Converging language streams in the human temporal lobe. J Neurosci 26:7328–7336.

Stevens JP (2012) Canonical correlation. In: Applied multivariate statistics for the social sciences, pp 395–412. New York: Routledge.

Stolk A, Todorovic A, Schoffelen JM, Oostenveld R (2013) Online and offline tools for head movement compensation in MEG. Neuroimage 68:39–48.

Van Essen DC, Drury HA, Dickson J, Harwell J, Hanlon D, Anderson CH (2001) An integrated software suite for surface-based analyses of cerebral cortex. J Am Med Infor Assoc 8:443–459.

Van Veen BD, van Drongelen W, Yuchtman M, Suzuki A (1997) Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. IEEE Trans Biomed Eng 44:867–880.

Vartiainen J, Parviainen T, Salmelin R (2009) Spatiotemporal convergence of semantic processing in reading and speech perception. J Neurosci 29:9271–9280.

Vigneau M, Beaucousin V, Hervé PY, Jobard G, Petit L, Crivello F, Mellet E, Zago L, Mazoyer B, Tzourio-Mazoyer N (2011) What is right-hemisphere contribution to phonological, lexico-semantic, and sentence processing? Neuroimage 54:577–593.

Weber K, Lau EF, Stillerman B, Kuperberg GR (2016) The yin and the yang of prediction: an fMRI study of semantic predictive processing. PLoS One 11:1–25.