Intracranial recordings reveal high-frequency activity in the human temporal-parietal cortex supporting non-literal language processing

Shweta Soni[1], Jacqueline Overton[2], Julia W. Y. Kam[3,4], Penny Pexman[3], Akshay Prabhu[5], Nicholas Garza[5], Ignacio Saez[6] and Fady Girgis[1,5]*

[1]Department of Clinical Neurosciences, University of Calgary, Calgary, AB, Canada, [2]Department of Neuroscience and Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, United States, [3]Department of Psychology, University of Calgary, Calgary, AB, Canada, [4]Hotchkiss Brain Institute, University of Calgary, Calgary, AB, Canada, [5]Department of Neurological Surgery, University of California, Davis, Davis, CA, United States, [6]Department of Neuroscience, Neurosurgery and Neurology, Icahn School of Medicine at Mount Sinai, New York, NY, United States

Objective: Non-literal expressions such as sarcasm, metaphor and simile refer to words and sentences that convey meanings or intentions that are different and more abstract than literal expressions. Neuroimaging studies have shown activations in a variety of frontal, parietal and temporal brain regions implicated in non-literal language processing. However, neurophysiological correlates of these brain areas underlying non-literal processing remain underexplored.

Methods: To address this, we investigated patterns of intracranial EEG activity during non-literal processing by leveraging a unique patient population. Seven neurosurgical patients with invasive electrophysiological monitoring of superficial brain activity were recruited. Intracranial neural responses were recorded over the temporal-parietal junction (TPJ) and its surrounding areas while patients performed a language task. Participants listened to vignettes that ended with non-literal or literal statements and were then asked related questions to which they responded verbally.

Results: We found differential neurophysiological activity during the processing of non-literal statements as compared to literal statements, especially in low-Gamma (30–70Hz) and delta (1–4Hz) bands. In addition, we found that neural responses related to non-literal processing in the high-gamma band (>70Hz) were significantly more prominent at TPJ electrodes as compared to non-TPJ (i.e., control) electrodes in most subjects. Moreover, in half of patients, high-gamma activity related to non-literal processing was accompanied by delta-band modulation.

Conclusion: These results suggest that both low- and high-frequency electrophysiological activities in the temporal-parietal junction play a crucial role during non-literal language processing in the human brain. The current investigation, utilizing better spatial and temporal resolution of human intracranial electrocorticography, provides a unique opportunity to gain insights into the localized brain dynamics of the TPJ during the processing of non-literal language expressions.

1 Introduction

Non-literal figures of speech such as sarcasm and metaphor convey meanings or intentions that are different from their literal expression (Gibbs and Colston, 2012). Because these types of statements are ubiquitous in social communication (Billow, 1981; Gibbs, 2000; Hancock, 2004; Whalen et al., 2009; Burgers et al., 2012; Biddle et al., 2020) and disruptions in processing these types of expressions are implicated in neurodevelopmental disorders (Happé, 1993; Dennis et al., 2001; Mo et al., 2008), it is crucial to understand the neural mechanisms that underlie the processing of these figures of speech.

The interpretation of non-literal language requires an understanding of the context of communication (Searle, 1979; Grice, 1989) as well as the general intention of the speaker (Katz and Pexman, 1997). Therefore, understanding non-literal meanings requires the utilization and conceptual blending of different information that may be simultaneously accessible during language comprehension (Fauconnier and Turner, 2008) which makes the whole process cognitively more engaging and demanding as compared to literal interpretation (Levorato and Cacciari, 2002; Citron et al., 2019). Moreover, the listener requires access to the speaker's mental states and intents about the utterance to determine meanings that are not literal (Channon et al., 2005; Spotorno et al., 2012). Therefore, non-literal comprehension involves an element of theory of mind, which is defined as the ability to attribute mental states to oneself and others (Premack and Woodruff, 1978). For example, autistic children with impaired theory of mind processing are unable to comprehend complex non-literal statements as compared to neurotypical children with intact theory of mind processing (Happé, 1993; Sullivan et al., 1995).

From a neuroscientific standpoint, because of the relation between the processing of non-literal statements and those involving theory of mind, there is an overlap in brain regions associated with the two constructs. Both theory of mind and non-literal processing incorporate and share various functional networks such as the frontoparietal control network (Cole et al., 2013), the default mode network (Buckner et al., 2008; Andrews-Hanna et al., 2010), and the ventral attention network (Dosenbach et al., 2007), as well as areas implicated in working memory and inhibition (Bohrn et al., 2012; Rapp et al., 2012; Yang, 2014; Reyes-Aguilar et al., 2018). This idea that brain regions enabling theory of mind processes are engaged in comprehending non-literal meanings is reinforced by imaging studies contrasting non-literal and literal language processing (Eviatar and Just, 2006; Wang et al., 2006; Wakusawa et al., 2007; Rapp et al., 2010; Shibata et al., 2010; Prat et al., 2012; Spotorno et al., 2012; Uchiyama et al., 2012; Schneider et al., 2015; Baptista et al., 2018; Filik et al., 2019) and their meta-analyses (Bohrn et al., 2012; Rapp et al., 2012). Recently, Hauptman et al. (2023) suggested that non-literal comprehension is co-supported by mechanisms that process literal meaning and those that facilitate broad social inference such as theory

of mind (Hauptman et al., 2023). Therefore, it would be reasonable to expect enhanced BOLD activation during fMRI tasks looking at non-literal comprehension in brain regions that are part of the ToM network, namely the temporal-parietal junction (TPJ), the brain area most often associated with theory of mind (Saxe and Kanwisher, 2003; Leslie et al., 2004; Aichhorn et al., 2009; Spreng and Grady, 2010; Schurz et al., 2014). Because non-literal communication calls upon theory of mind skills (Happé, 1993), these imaging studies serve as compelling evidence for the link between theory of mind and non-literal processing.

In this vein, fMRI studies have found a direct relationship between the TPJ and non-literal processing (Prat et al., 2012; Spotorno et al., 2012). For example, reading phrases that were contextually non-literal (e.g., "*This campaign has really been a hit.*" in the context of a disappointing campaign) activated bilateral TPJ when compared to reading phrases that were contextually literal (e.g., "*This campaign has really been a hit.*" in the context of an excellent campaign) (Spotorno et al., 2012). This finding was corroborated by other studies exploring neural substrates involved in non-literal processing that demonstrated BOLD activation in both hemispheres in regions that anatomically are part of the TPJ. These regions included, the middle and superior temporal gyri, and the inferior parietal cortex (Lee and Dapretto, 2006; Stringaris et al., 2006; Rapp et al., 2007; Boulenger et al., 2009; Mashal et al., 2009; Bosco et al., 2017).

In addition to fMRI studies, scalp EEG studies have employed spectral analysis of electrophysiological responses to gain insights into how brain oscillations are modulated at different frequencies during non-literal comprehension (van den Brink et al., 2012; Spotorno et al., 2013; Sun et al., 2022; Champagne-Lavau et al., 2023; Hubbard et al., 2023). For example, understanding sentences conveying nonliteral meaning elicited an increase in gamma, alpha and delta power as compared to sentences that conveyed literal meaning, suggesting the engagement of different mechanisms such as integration of multiple information and executive functioning while understanding nonliteral sentences (Spotorno et al., 2013). Moreover, the semantic decoding of non-literal expressions (i.e., metaphors) requires more attention and efforts as compared to literal expressions, as indexed by enhanced synchronization in delta and theta band during different time windows throughout widespread brain regions (Sun et al., 2022).

Non-literal language interpretation is associated with increased TPJ activity in fMRI studies and general electrical oscillatory perturbations on scalp EEG studies. While these techniques are non-invasive and well-validated, they can be limited by low spatial resolution (scalp EEG), low temporal (fMRI) resolution, and poor signal-to-noise ratio for individual trials. Intracranial EEG can help mitigate these drawbacks due to its excellent temporal and spatial precision as well as high signal-to-noise ratio, which can better elucidate the role of low and high-frequency oscillatory activity in the TPJ during non-literal processing. Therefore, we examined intracranial EEG (iEEG) from neurosurgical patients undergoing awake brain operations in the region of the TPJ during a task

investigating non-literal versus literal language comprehension. Specifically, we examined whether comprehending non-literal statements, such as sarcasm, metaphor, and simile, gives rise to distinct patterns of brain oscillatory activity in the TPJ as compared to comprehending literal statements. This novel methodology offers a more comprehensive depiction of the neurobiology of non-literal language processing. Given the link between the TPJ and non-literal processing, we hypothesized that non-literal processing would elicit a stronger response in low (1–30 Hz) and high (>30 Hz) frequency bands as compared to literal processing in the TPJ and that greater low and high frequency activity would be observed at TPJ electrodes as compared to non-TPJ electrodes.

2 Materials and methods

2.1 Participants

Seven neurosurgical patients (age in years: 58.14 ± 13.71) undergoing awake craniotomies for clinical purposes were recruited to participate in the study. All participants consented to the study according to the guidelines of the Declaration of Helsinki that was approved by the institutional ethics review board of the University of California, Davis (approval number: 1201780-12). Participants reported no history of other illnesses and did not show any signs of impairment in executive functioning and language skills. Patients were chosen if they met the following specific criteria: They must be able to participate in a cognitive task, the required operation must expose the temporal-parietal junction cortex (TPJ), and this exposed cortex must be normal, meaning that it possesses normal signal characteristics on standard MRI sequences and it appears normal on direct visualization during surgery. As per routine clinical practice in awake brain surgery, surface strip electrodes were used to record after-discharge potentials during eloquent cortex mapping to minimize the risk of seizures and enhance the safety of the mapping procedure. Once clinical mapping was complete, and if the patient was comfortable, fully alert, cooperative, and able to participate in a cognitive task, the strip electrodes were repositioned over the TPJ, and the cognitive task was administered for research purposes. The data for two patients were not usable because the patients did not tolerate testing: one patient became nauseated at the beginning of the task and could not continue, and a second opted out of the study after a few trials. Data from another patient were excluded because of noisy electrophysiologic recordings that resulted in most trials being discarded during analysis. Therefore, data from three patients were excluded from the analyses. The remainder of this section pertains to the remaining four patients whose data were included in the analyses (see Table 1 for patients' demographic details).

2.2 Surgery and electrode localization

Patients underwent a craniotomy to expose the brain for mapping and subsequent tumor resection. Prior to resection, patients were implanted with 1 to 4 strips (6 to 8 channel) of platinum-iridium electrodes, each with a 3 mm exposed recording area and 10 mm center to center inter-contact spacing (Ad-Tech Medical, Racine, WI). Intra-operative photos were taken once the strips were in position, and these were then used to map exact electrode locations on a 3-dimensional MRI surface reconstruction for each patient. The anatomical location of each electrode was determined by the neurosurgeon (e.g., TPJ, primary sensory cortex, superior temporal gyrus, etc.), and locations projected onto a 3-D brain rendering generated from an anatomical atlas within an online neuroimaging software suite (Freesurfer, Charlestown, MA) using MATLAB (Mathworks, Natick, MA). Importantly, anatomical coverage across patients was based on clinical requirements. All patients except for one underwent left sided operations as per the standard clinical strategy of carrying out speech mapping in the left hemisphere.
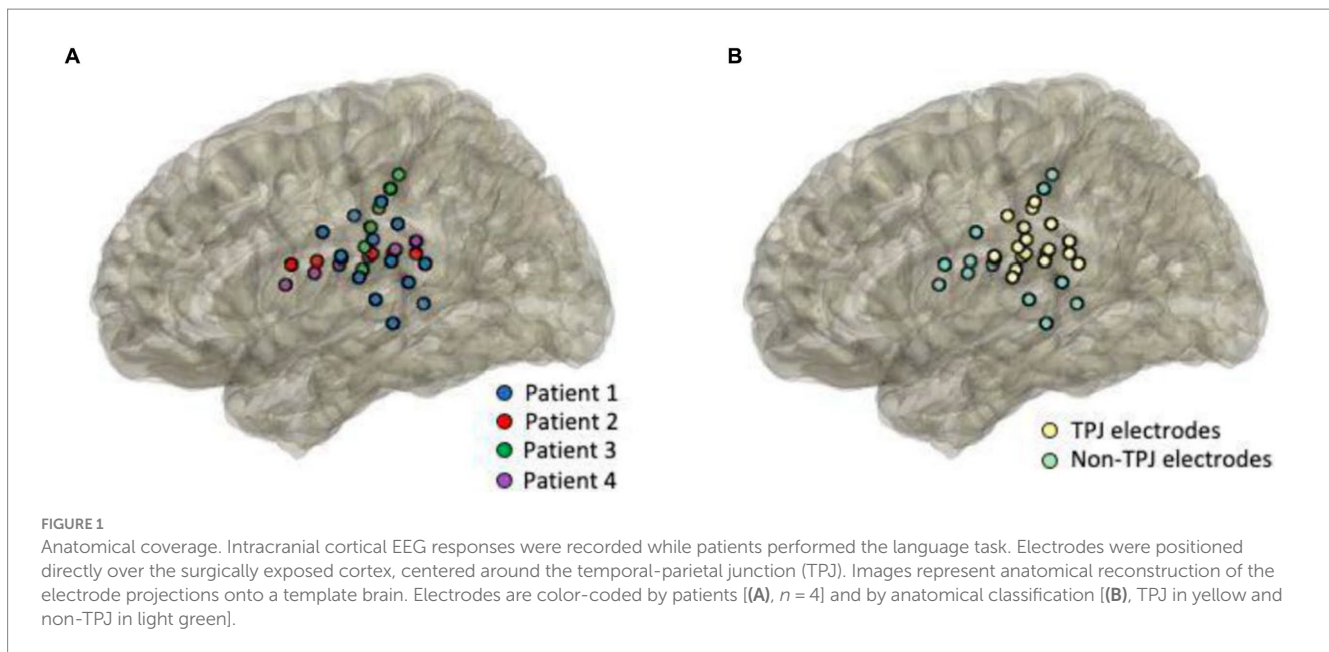
Electrode strips were placed directly onto the cortical surface and positioned over the TPJ area for experimental testing (Figure 1). Electrodes that were placed within the anatomical borders of the TPJ were classified as TPJ electrodes. Electrodes outside of the TPJ were placed over a variety of adjacent areas, including motor and auditory cortices, and were categorized as non-TPJ (control) electrodes (see Figure 1A for individual patient electrode placement and Figure 1B for anatomical classification of electrodes). All patients had at least 3 contacts in the TPJ area (mean number of TPJ contacts = 4.25 ± 2.5, range: 3–8) as well as in non-TPJ areas (mean number of non-TPJ contacts = 3.5 ± 1, range: 3–5), but the number of channels included in the analysis was lower (mean TPJ = 2.75 ± 1.5, mean non-TPJ = 2 ± 0) due to our bipolar derivation analytical strategy (TPJ/non-TPJ: 5/2 for S01, 2/2 for S02, S03, and S04). Bipolar referencing is considered advantageous for analyzing iEEG signals as it emphasizes activity with high local specificity by removing signal artifacts shared by nearby electrodes and widely dispersed neural response components (Mercier et al., 2022). Electrodes were excluded from analyses if there was abnormal tissue at that site.

2.3 Language task

We adopted a simple mentalization task (Happé, 1994; Stone et al., 1998; Jolliffe and Baron-Cohen, 1999; Gregory et al., 2002) in which participants make inferences about the mental states and intentions of others based on statements made in different situations. An audio testing format was chosen over a visual or computer-based format due to the limitations associated with the surgical conditions. Participants

TABLE 1 Participant demographics and testing details.

Sub#	Age	G	Handed-ness	Diagnosis	Surgery side	# of TPJ	# of non-TPJ	# of vignettes completed	Accuracy
S01	65	F	Right	Tumor	Right	8	5	5	85%
S02	61	M	Right	Tumor	Left	3	3	5	63.2%
S03	44	F	Right	Tumor	Left	3	3	6	100%
S04	76	M	Right	Tumor	Left	3	3	5	57.9%

FIGURE 1
Anatomical coverage. Intracranial cortical EEG responses were recorded while patients performed the language task. Electrodes were positioned directly over the surgically exposed cortex, centered around the temporal-parietal junction (TPJ). Images represent anatomical reconstruction of the electrode projections onto a template brain. Electrodes are color-coded by patients [**(A)**, *n* = 4] and by anatomical classification [**(B)**, TPJ in yellow and non-TPJ in light green].

were lying supine on the operating table with their heads fixed in a clamp and turned approximately 70 degrees. In addition, there were sterile drapes impeding parts of their visual field. Therefore, looking at a computer screen or reading a paragraph of text was too onerous for the participants, making a listening task with verbal responses the most feasible option.
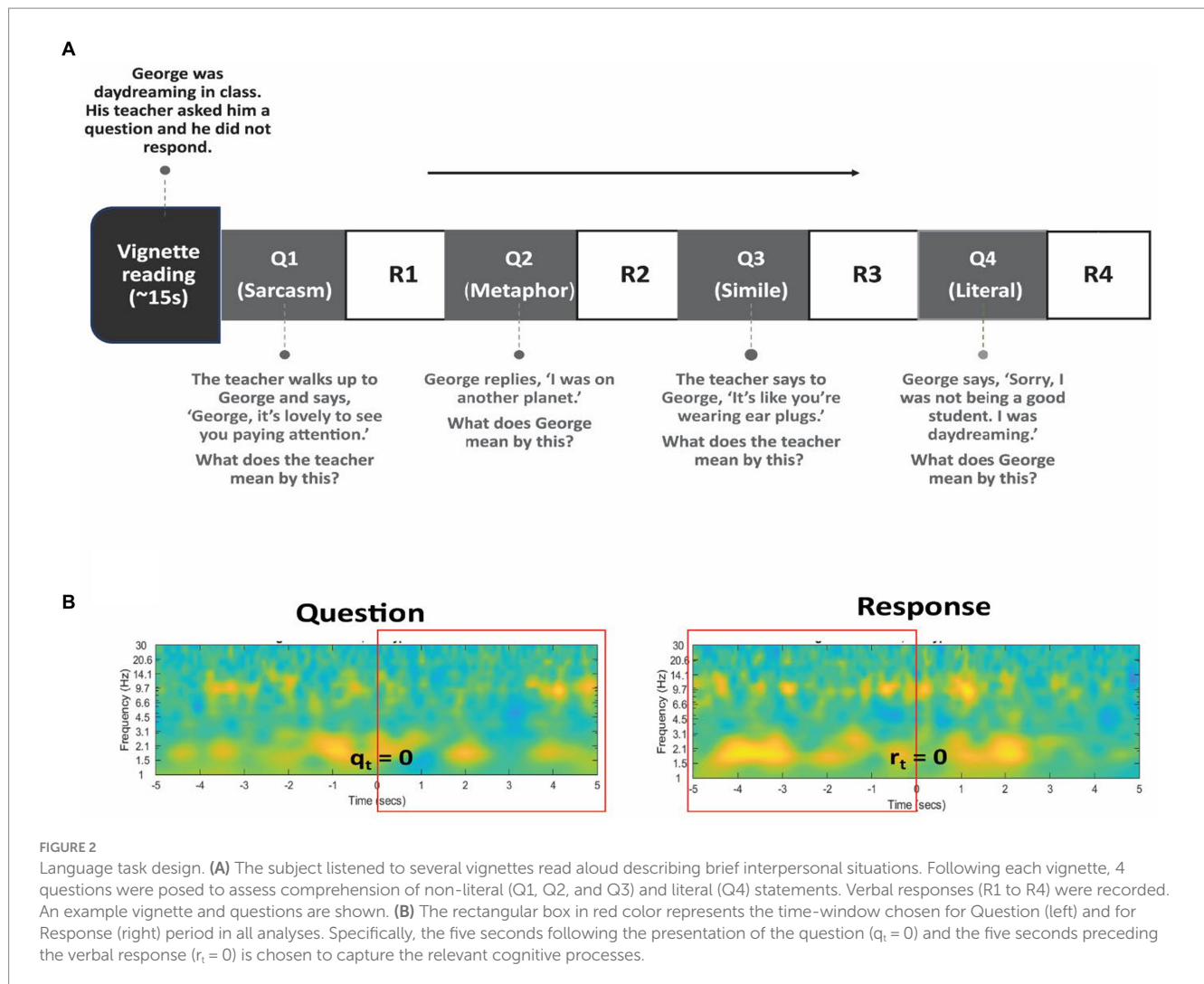
Participants were read a series of vignettes which presented brief interpersonal situations (see Supplementary material). These were stories adapted from initial psychological work on theory of mind (Happé, 1994) and expanded to involve various non-literal figures of speech. Following each vignette, four questions were asked that required different forms of mentalization and assessed the participant's ability to comprehend non-literal and literal expressions in speech (Figure 2A). The first three questions of each vignette required the interpretation and identification of non-literal statements, including sarcasm, metaphor, and simile, while the fourth question involved interpreting a literal statement. Participants responded verbally, and audio was recorded using an external microphone and timestamped to EEG signals. Testing time was limited to approximately 10 min due to surgical constraints, during which time patients were asked to listen to and respond to the four questions to a maximum of six vignettes. The behavioral performance was indexed by the percentage accuracy of the completed questions. In addition to the percentage accuracy, we measured reaction times after each question.

2.4 iEEG recording and analyses

iEEG was recorded during the language task in all available channels and timestamped with events in the task to allow parsing of neuronal activity in relation to task epochs. Number of channels varied across participants between 6 and 13 (see Table 1). Channels were amplified × 10,000, analog filtered (0.01–1,000 Hz) with a 4,096 Hz digitization rate (except in one case where a digitization rate of 2,048 Hz was used). Electrophysiological data were imported to MATLAB for preprocessing and analyses with custom scripts.

Channels were visually examined, and channels with low signal-to-noise ratio were identified and excluded from subsequent analyses (i.e., electromagnetic noise from hospital equipment, amplifier saturation, poor contact with cortical surface). Signals were down sampled to 1,024 Hz, high-pass filtered above 1.0 Hz and lowpass filtered below 200 Hz with symmetrical (phase true) finite impulse response filters (~35 dB/octave roll-off), and notch filtered to remove line noise (60 Hz and harmonics), using order 8 Butterworth filters. Pairs of adjacent channels were subtracted to produce bipolar re-referenced signals (e.g., given 6 original channels, channels 2 and 5 were selected as reference for channels 1, 3 and 4, 6 respectively, yielding 4 resultant bipolar signals). Periods with excessive noise (e.g., due to patient movement) or epileptiform activity were identified using a custom-written algorithm followed by visual inspection, marked, and excluded from analysis (i.e., timepoints around artifacts were set to NaNs). Because our data consisted of a small number of trials of relatively long duration, we did not reject entire trials due to noise or artifact as is commonly done, but rather excluded noisy periods from statistical analysis and maintained the remainder of the trials.

To obtain an initial comprehensive picture of task-related changes in neural responses, we generated power spectra and event-related spectrograms. For that, we estimated time-frequency representations (TFRs) using complex Morlet wavelet convolution (Cohen, 2019) for each trial excluding artifactual periods as described above. We applied wavelet transform separately in a range of 1 to 30 Hz (in 24 log-spaced steps) and in a range of 30 to 200 Hz (in 30 log-spaced steps) frequencies based on appropriate number of wavelet cycles for both frequency ranges. We examined both low and high frequency bands for several reasons. First, task-related activations during cognitive tasks may occur in several frequency bands (Lakatos et al., 2005). Second, past studies have reported greater synchronization in several frequency bands during non-literal comprehension (Spotorno et al., 2013; Sun et al., 2022; Champagne-Lavau et al., 2023). Finally, the high-frequency broadband activity (>70 Hz) is considered a marker of cortical activation (Parvizi and Kastner, 2018) and related to both

FIGURE 2
Language task design. **(A)** The subject listened to several vignettes read aloud describing brief interpersonal situations. Following each vignette, 4 questions were posed to assess comprehension of non-literal (Q1, Q2, and Q3) and literal (Q4) statements. Verbal responses (R1 to R4) were recorded. An example vignette and questions are shown. **(B)** The rectangular box in red color represents the time–window chosen for Question (left) and for Response (right) period in all analyses. Specifically, the five seconds following the presentation of the question ($q_t = 0$) and the five seconds preceding the verbal response ($r_t = 0$) is chosen to capture the relevant cognitive processes.

neuronal population spiking and BOLD activation (Logothetis et al., 2001; Engell et al., 2012; Jacques et al., 2016; Leszczyński et al., 2023). Given fMRI studies have reported increased activation during non-literal comprehension in the TPJ, this may be manifested in the high-frequency broadband activity in the iEEG signal.

Each trial was separated into a question and response period, with the onset of the audio presentation of the vignettes marking the onset of the question ($q_t = 0$) and the onset of participants' verbal response as the onset of the response period ($r_t = 0$). Spectral power for each period was calculated and baseline corrected in reference to a 4 s time-window prior to the first question. The first three questions comprising the non-literal condition were combined for analysis. So, we averaged spectral power for all non-literal conditions and literal conditions separately across different time-windows for the question (0 to +5 s) and response (−5 to 0 s) period to capture the relevant cognitive processes (Figure 2B). Specifically, the five seconds following the presentation of the question and the five seconds preceding the verbal response presumably engaged the processing of non-literal and literal statements and assessing underlying intentions or meanings. The mean spectral power for non-literal and literal conditions was also separately averaged across TPJ electrodes and non-TPJ electrodes. We carried out two main analyses described below.

To address our first hypothesis, we delineated the neurophysiological response to non-literal and literal conditions by comparing TFRs for non-literal conditions to TFRs for the literal condition (non-literal vs. literal contrast) using the cluster-based permutation method (between-condition comparison). These TFRs have two dimensions: temporal (time-points) and spectral (frequencies). The mean spectral power of each sample— (time, frequency)-pair in a TFR— in the non-literal condition was contrasted to the mean spectral power of the equivalent sample in the literal condition by using the Kolmogorov–Smirnov (KS) statistic and summing the KS-statistic value across connected components or 'cluster' (adjacent (time, frequency)-pairs) that exceeded the KS value at an alpha level of 0.05. This led to a large number of statistical comparisons as the difference between non-literal and literal conditions is evaluated at all samples in the time-frequency plane. To correct for the multiple comparisons, we performed a Monte Carlo simulation by shuffling the trial assignment for each pair of TFRs (i.e., randomly assigning trials to the non-literal and literal conditions), recalculating the KS contrast and sum of KS statistic across suprathreshold (alpha <0.05) clusters, and repeating this process 1,000 times. Significant or suprathreshold clusters were identified by taking the 99th percentile of the resulting distribution

of 1,000 KS-cluster sizes (cluster alpha <0.01) and applying this threshold to the original unshuffled KS-statistic TFR contrasts that resulted into cluster-corrected TFRs. This procedure was repeated using each participant's TFR data and the resulting cluster corrected TFRs were separately obtained for both low (1–30 Hz) and high (30–200 Hz) frequencies in question and response periods at TPJ and non-TPJ electrodes.

To summarize the neural activity from these cluster-corrected TFRs, we calculated the 'proportion of active timepoints' in discrete frequency bands: delta (1 to 4 Hz), theta (4 to 8 Hz), alpha (8 to 13 Hz), beta (13 to 30 Hz), low-gamma (LG: 30 to 70 Hz) and high-gamma (HG: 70 to 200 Hz). To carry out this analysis, we binary coded each time point in the time-frequency plane as active (i.e., if cluster corrected significant activity exists in any frequency at that time point) or inactive (i.e., if no significant clusters emerged at that time point). We then summed the number of active time points (based on suprathreshold clusters) in a particular frequency band within the event of interest (0 to +5 s for question and −5 to 0 s for response) and divided this by the total number of timepoints in that frequency band in the time-frequency plane. Thus, the proportion of active timepoints gives an estimate of the engagement of different frequency bands across the task at both types of electrodes.

To address our second hypothesis, we assessed the differences in non-literal processing-related neural responses between TPJ and non-TPJ electrodes in the above-mentioned frequency bands (between-electrode comparison). Though measuring the proportion of active timepoints shows differences in neural responses between TPJ and non-TPJ electrodes, between-electrode comparison of this metric acquired from cluster-corrected TFRs of only four participants would not provide enough statistical power. Therefore, we took a different approach and first identified timepoints for suprathreshold clusters from the cluster corrected TFR (representing non-literal versus literal contrast) within each frequency band. These clusters indicate significantly more neural activity while processing non-literal statements than literal statements. We then extracted spectral power that corresponds to the timepoints of these suprathreshold clusters from the non-literal TFRs during the question and response periods in each trial and averaged across the period of interest within each frequency band. We referred to it as 'power of active timepoints', representing neural activity in different frequency bands during non-literal processing. This approach not only allowed us to execute trial-balanced comparisons between TPJ and non-TPJ regions but also provided us with sufficient statistical power using more data points (instead of one data point per participant).

2.5 Statistical analysis

All statistical analyses were performed using customized scripts in MATLAB (2022b). Descriptive data are presented as mean ± standard deviation unless otherwise mentioned. We used non-parametric tests because our data is not normally distributed. Comparisons between experimental conditions or electrode types were examined using two-tailed Wilcoxon sign rank test (non-parametric equivalent for matched-samples comparisons) with significance level set at 0.05.

3 Results

Our analyses were based on four patients undergoing awake craniotomy exposing normal temporal-parietal junction (TPJ). They completed at least 4 of 6 language task vignettes (5 ± 0.81 vignettes completed on average). Behavioral performance on the completed questions was high, with mean accuracy of 76.51 (±16.95) across all conditions. Mean accuracy varied slightly across conditions: non-literal questions overall: 75 ± 16.58 (sarcasm: 65 ± 35.7; metaphor: 90 ± 10; simile: 70 ± 17.32) and literal questions: 81.25 ± 20.73. This indicates patients understood and performed the task adequately. Participants took slightly longer while responding to non-literal questions: overall: 7.07 ± 0.93 (sarcasm: 5.16 ± 1.31; metaphor: 8.09 ± 1.40; simile: 7.96 ± 1.19) than literal questions: 6.29 ± 1.38. However, the overall reaction time for non-literal questions was not significantly different from the reaction time for literal questions ($p = 0.25$, Wilcoxon sign-rank test). In addition, we did not find a significant relation between the percentage of accuracy and reaction time for any type of question (Pearson correlation coefficient and value of p for sarcasm: 0.10 and 0.90; metaphor: −0.51 and 0.45, simile: −0.55 and 0.41, literal: 0.50 and 0.46). This signifies that participant's ability to understand and answer correctly did not depend upon the speed they responded with. All the non-literal questions were pooled together in neurophysiological analyses (see Methods).

The TFRs for non-literal and literal conditions from an example patient at TPJ and non-TPJ electrodes during the question and response periods are shown in Figures 3, 4 respectively. In this study, we (1) analyzed non-literal versus literal contrast using a cluster-based permutation approach to examine neural activity specific to non-literal processing at TPJ and non-TPJ electrodes and (2) compared neural activity during non-literal processing between both types of electrodes. The results for both comparisons are described in Sections 3.1 and 3.2, respectively.

3.1 Neural activity differs in non-literal and literal processing mainly in low-gamma and delta bands regardless of electrode type

The cluster-corrected spectral activity from the non-literal versus literal contrast in TPJ and non-TPJ electrodes are shown for an example patient in Figures 5, 6 respectively. Clusters showing significantly increased activity in the non-literal condition as compared to literal condition were not sustained in both question and response periods, but rather appeared in discrete temporal patches. Also, the pattern of increased activity varied across patients.

To quantify the involvement of different frequency bands, we summarized the spectral activity related to non-literal versus literal processing as the proportion of active timepoints in each frequency band (see Methods). The results revealed relatively consistent activity in the low-gamma band (30–70 Hz) band at TPJ electrodes, which was present in most patients in our dataset during both question (3 out of 4 participants, min = 0% (no activity at all), max = 2.73% active timepoints) and response (4 out of 4 participants, min = 0.27%, max = 6.30%) periods. This pattern was also similar at non-TPJ electrodes as 3 out of 4 participants showed activity in the low-gamma
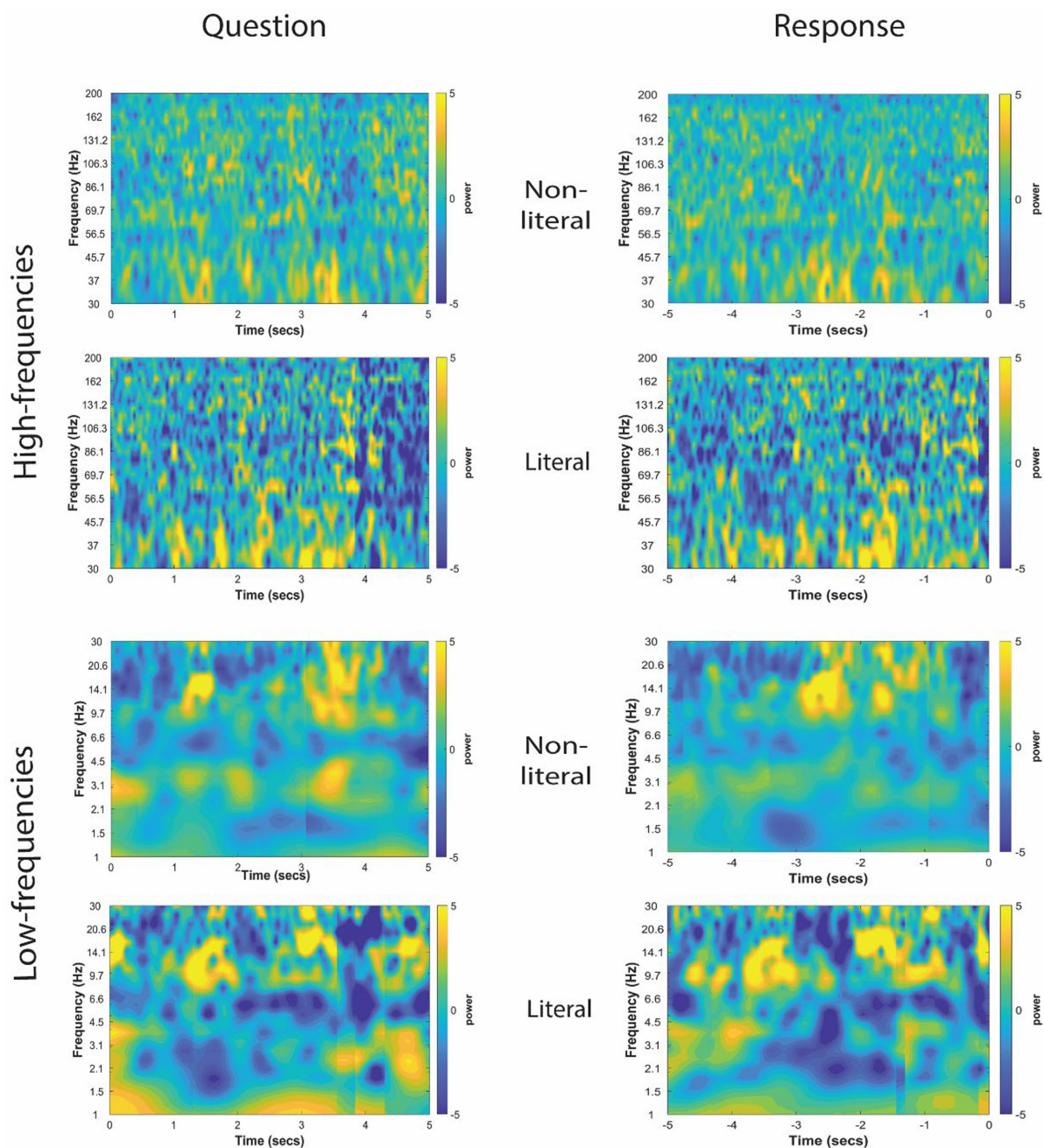
FIGURE 3
Time-frequency representation of intracranial EEG activity at TPJ electrodes during question and response periods in an example participant. Y-axis and X-axis represent frequency in Hertz and time in seconds, respectively. For the Question (left column), $t = 0$ indicates the start time of the question being read by the experimenter, and for the Response (right column) $t = 0$ indicates the start time of the participant's verbal response. Time-frequency representations were averaged for both non-literal and literal conditions showing differential power modulations in both high (30−200 Hz, top panel) and low (1−30 Hz, bottom panel) frequencies.

band in both question (min = 0%, max = 2.15%) and response (min = 0%, max = 1.88%) periods. This proportion of active timepoints was not different between these two periods in any frequency band ($p > 0.25$ at both TPJ non-TPJ electrodes).

Interestingly, there was also activity in the high-gamma band (70–200 Hz) and the delta frequency band (1–4 Hz), but of a less consistent nature, appearing mainly at TPJ electrodes (see Supplementary Table S1). In the high-gamma band, the activity was

found only in 2 of 4 patients (1.20 and 1.12%) in the question period and in 3 of 4 patients (1.18, 0.18 and 0.62%) during the response period, respectively. Noticeably, delta activity was consistently found in all patients during the question period (min = 1.37%, max = 6.35%) but was visible only in 2 out of 4 patients (4.30 and 3.33%) in the response period. The TPJ activity in other frequency bands was negligible (overall mean active points in question and response = 1.56 and 1.41%) for theta [4–8 Hz], 0.07 and 0.68% for alpha [8–12 Hz] and
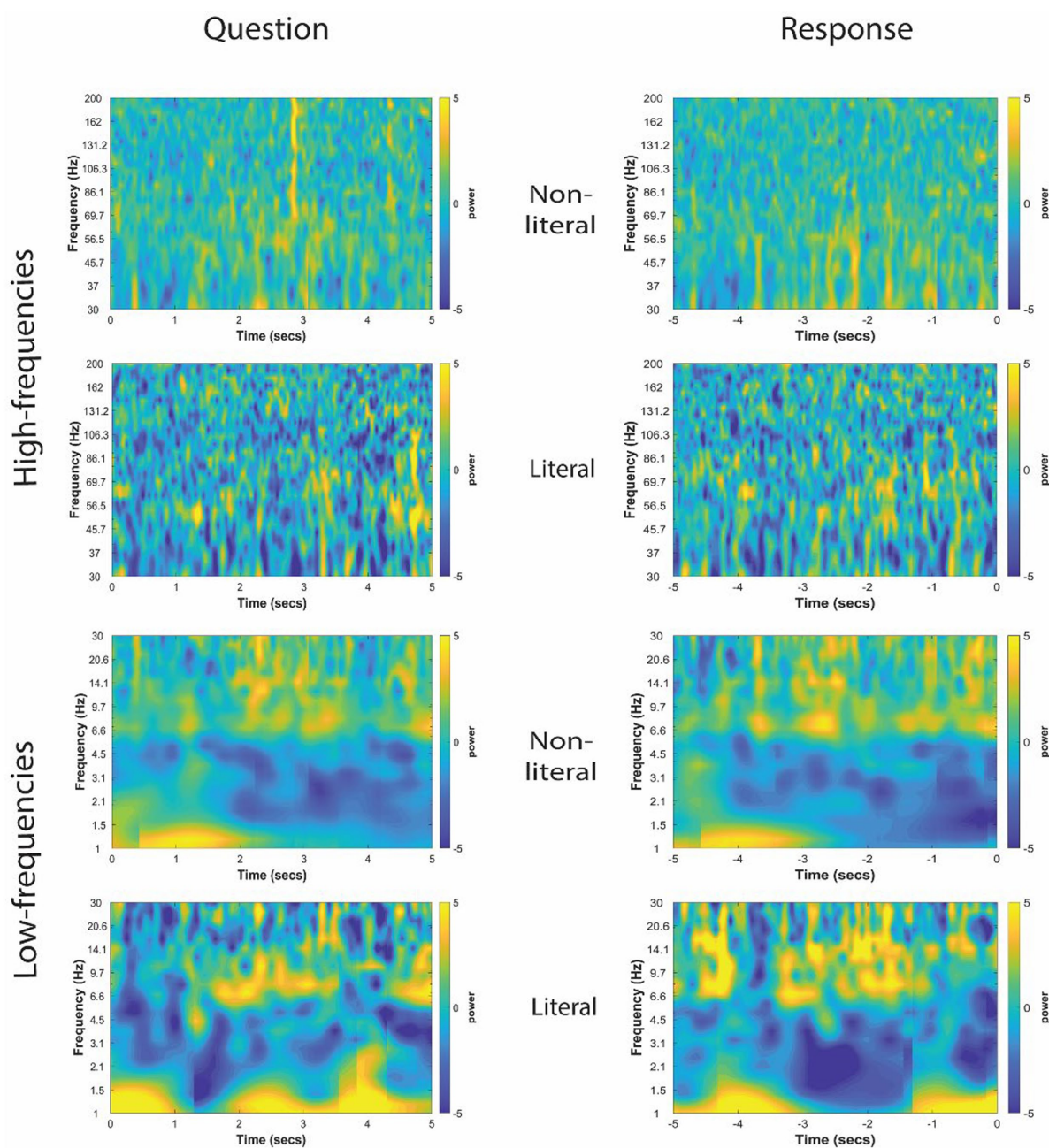
FIGURE 4
Time-frequency representation of intracranial EEG activity at non-TPJ (i.e., control electrodes) during question and response in an example participant. *Y*-axis and *X*-axis represent frequency in Hertz and time in seconds, respectively, where $t = 0$ indicates the beginning of the Question reading by the experimenter (left column) or of the verbal Response made by the patient (right column). Time-frequency representations were averaged for both non-literal and literal conditions showing differential power modulations in both high- (30−200 Hz, top panel) and low- (1−30 Hz, bottom panel) frequencies.

0.96 and 0.02% for beta [12–30 Hz], as well as non-consistent (in question period, 2/4, 1/4, and 1/4 participants for theta, alpha and beta respectively; in response period, 1/4 participants for all three bands). Thus, low-gamma activity might be the marker of neural activity that differs in non-literal and literal processing, which was accompanied by delta-band and high-gamma activity in more than half of the cases at TPJ electrodes. The proportion of active timepoints was not different between question and response periods ($p > 0.05$).

3.2 High-gamma power increases at TPJ to a greater extent than at non-TPJ electrodes

When comparing whether neural activity related to non-literal processing differs in both types of electrodes, our results demonstrate more power of active timepoints at TPJ sites than non-TPJ sites during non-literal-specific processing across participants, mainly in the delta
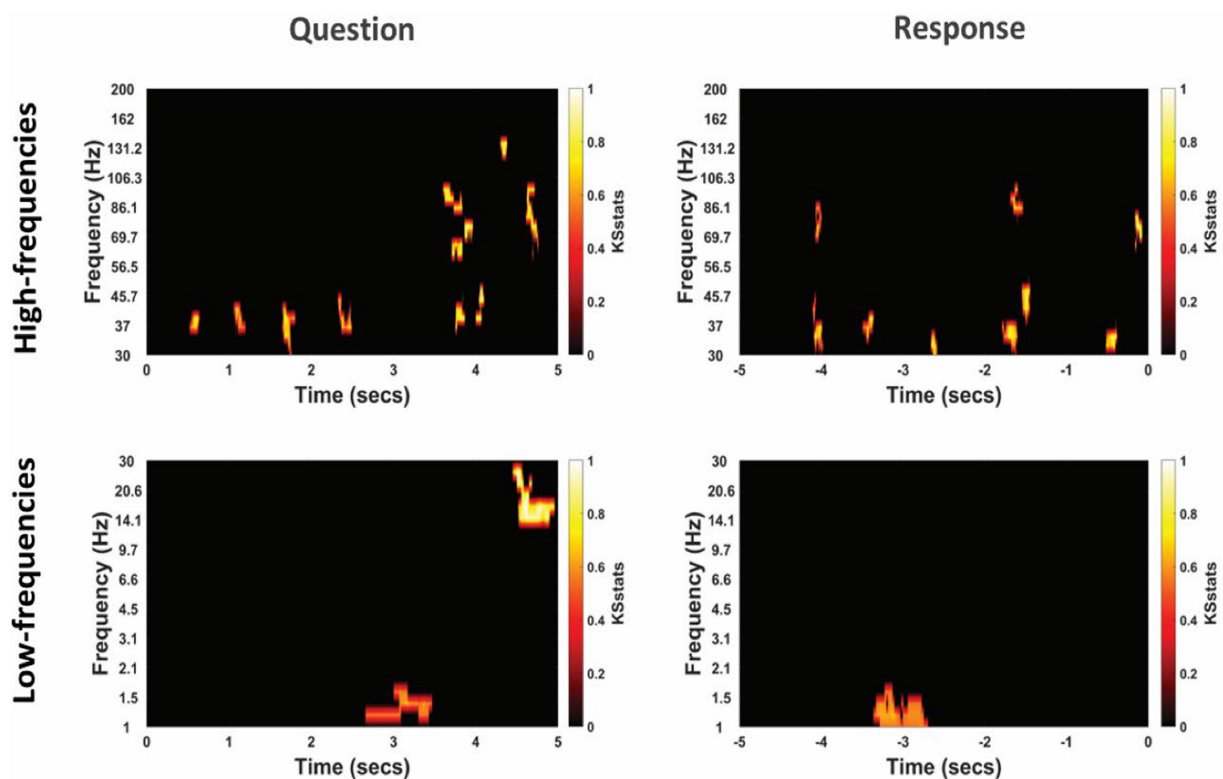
FIGURE 5
Differential power modulation between non-literal and literal conditions at TPJ electrodes for an example participant. Statistical maps show significant differences in power between non-literal and literal conditions after corrected by cluster-based permutation at p = 0.01. Analyses were performed separately for Question (left column) and Response (right column) epochs in both high-frequency (top panel) as well as low-frequency (bottom panel) bands. The t = 0 at the x-axis of these maps indicates the time when the Question was started to be read by the experimenter or when the verbal Response began by the patient. Color intensity of clusters represent KS-statistics resulting from comparing power across all questions/responses and across all TPJ electrodes for each condition (non-literal/literal).

and high-gamma band (see Figure 7). During the question period all participants showed significantly more power at TPJ than non-TPJ in the delta band ($p < 0.001$, Wilcoxon sign-rank test). Moreover, significantly more high-gamma TPJ-power accompanied with delta activity in 2 out of these 4 participants ($p < 0.001$ and $p = 0.0002$ respectively, Wilcoxon sign-rank test). We found a similar pattern of more high-gamma activity in TPJ electrodes during the response period which was present in 3 of 4 participants ($p < 0.001$, $p < 0.001$ and $p = 0.002$ respectively, Wilcoxon sign-rank test). However, we did not find any robust pattern of more power at TPJ than non-TPJ across participants in any other frequency bands such as theta, alpha, beta and low gamma, as only 1 or 2 out of 4 participants showed more TPJ activity in these bands during both question and response periods. Hence, high-gamma and delta activity seem to be indicators of TPJ-specific activity during non-literal processing.

4 Discussion

We investigated patterns of neurophysiological activity involved in interpreting situations that included non-literal expressions in comparison to literal expressions. Importantly, our study is the first, to our knowledge, to directly examine the involvement of the TPJ in processing non-literal stimuli using an iEEG approach. Particularly,

we sought to verify if (1) processing non-literal situations elicit different neural activity compared to literal situations at TPJ and non-TPJ sites, and if (2) the TPJ sites show greater activity during non-literal processing than non-TPJ sites (i.e., electrodes placed over motor or auditory areas). In general, our results demonstrate that iEEG activation for the non-literal condition was significantly different from that for the literal condition in low-gamma (30–70 Hz), high-gamma (70–200 Hz) and delta (1–4 Hz) bands at the TPJ. Specifically, we noticed a robust increase in low-gamma activity across the majority of patients during both question (while the patient was listening to a question about the content of vignettes) and response (before the patient responded to the question) periods. We further show that TPJ electrodes exhibit more non-literal processing-specific activation than non-TPJ electrodes, mainly in the high-gamma (70–200 Hz) band. We found no evidence for an activity increase in other low-frequency bands (theta, alpha, beta).

The key strength of the present work is that it employs intracranial EEG to reveal brain oscillatory dynamics during non-literal processing. The use of intracranial electrodes that capture task-relevant activity from only a few millimeters of the cortex provided direct access to 'local' signals from highly precise anatomical localizations (Holdgraf et al., 2016). In addition to offering good spatial resolution, intracranial EEG enabled the recording of high-frequency activity (broadband gamma, 30–200 Hz). This would
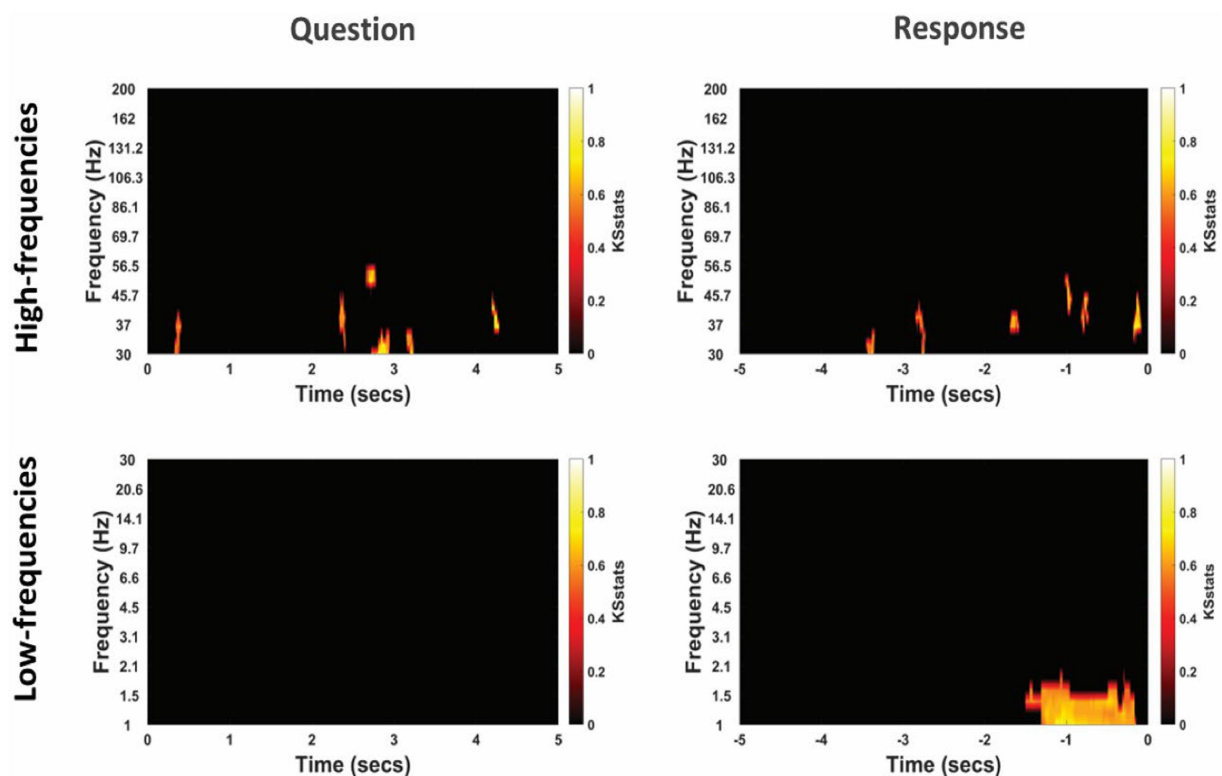
FIGURE 6
Differential power modulation between non-literal and literal conditions at non-TPJ or control electrodes for an example participant. Statistical maps show significant differences in power between non-literal and literal conditions after corrected by cluster-based permutation at $p = 0.01$. Analyses were performed separately for Question (left column) and Response (right column) epochs in both high-frequency (top panel) as well as low-frequency (bottom panel) bands. The $t = 0$ at the x-axis of these maps indicates the time when the Question was started to be read by the experimenter or when the verbal Response began by the participant. Color intensity of clusters represent KS-statistics resulting from comparing power across all questions/responses and across all non-TPJ electrodes for each condition (non-literal/literal).
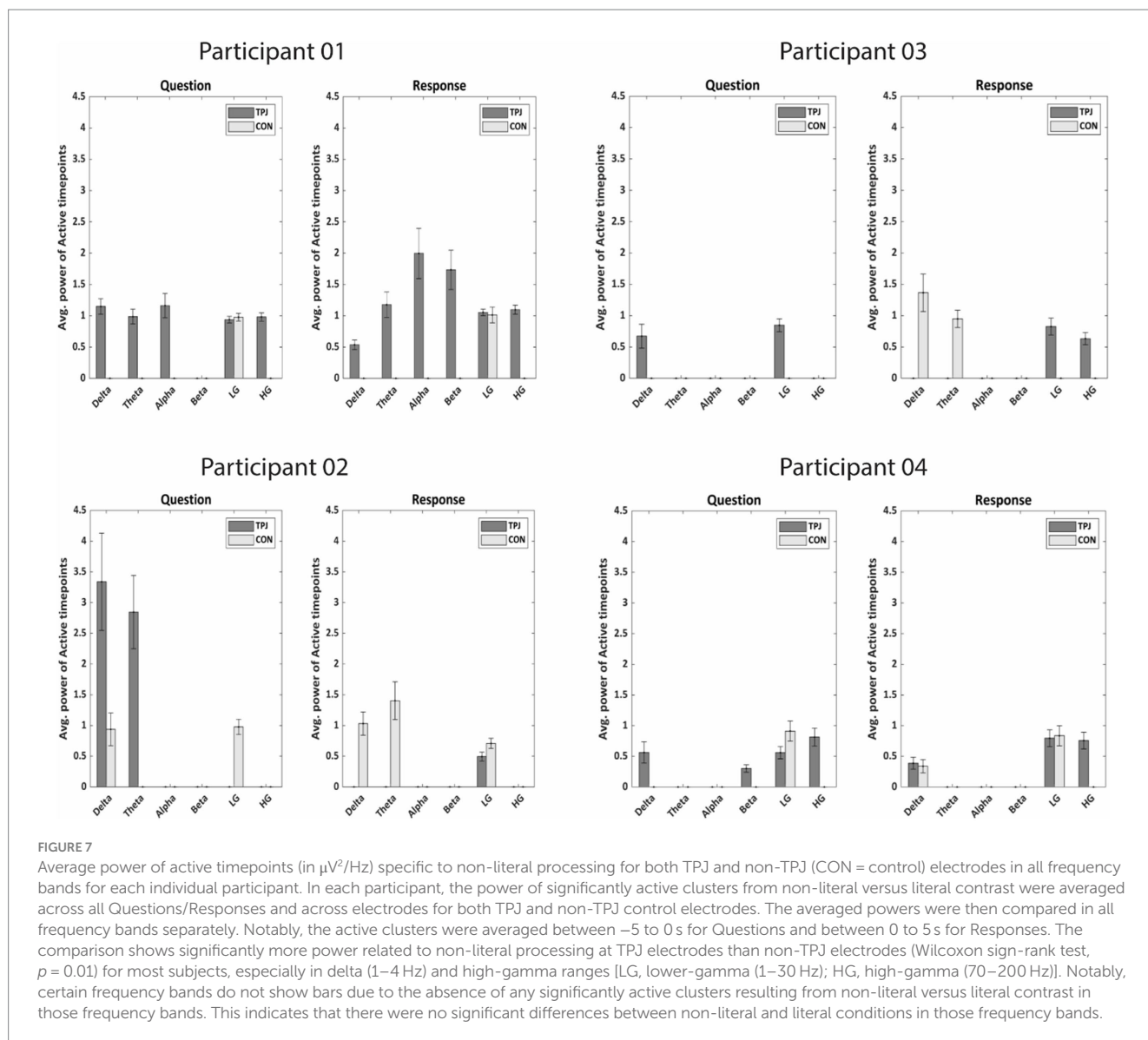
be difficult to achieve using non-invasive approaches where the skull functions as a low-pass filter, attenuating signals above ~40 Hz (Buzsáki et al., 2012). Additionally, investigation of brain activity using time-frequency analysis can reveal subtle differences in brain oscillations that cannot be captured by ERP analyses alone (Cohen, 2014).

We quantified the brain oscillatory activity specific to non-literal processing by conducting a between-condition comparison of time-frequency representations (non-literal and literal contrast) and summarized in canonical frequency bands as proportion of active timepoints. One significant finding of this analysis was that patients robustly showed a higher proportion of active timepoints in the low-gamma band both while assessing the situation (question period) and formulating an answer (response period). This finding suggests that low-gamma activity may support the processing of non-literal situations, which is consistent with earlier research that reported increased power in the low-gamma band associated with non-literal processing that is different from literal processing (Spotorno et al., 2013). Unlike that study which observed increased power in the range of 30–35 Hz, low-gamma activation in our study was broader, between 30–70 Hz. We attribute this difference to task design, stimuli and methodologies used in the two investigations. While Spotorno et al. (2013) employed sentences where the target sentence was presented with a context conveying negative or opposite meaning (ironic condition), our stimuli involved a range of

contexts to cue non-literal meanings. Also, the intracranial EEG seems to capture task-related changes in higher frequencies better than the scalp EEG. Broadly, our results are consistent with the reported role of low-gamma in the integration of various information streams in the semantic domain, such as bringing together real-world knowledge and literal meanings (Hagoort et al., 2004; van den Brink et al., 2012).

It is noteworthy that the low-gamma activity specific to non-literal processing was observed in both TPJ and non-TPJ areas. We speculate that increased low-gamma activity during nonliteral versus literal comprehension only reflects mechanisms that enable the unification of semantic and social information. Other mechanisms such as mental inferencing, which is another important aspect of nonliteral understanding and involves inferring others' mental states and intents, are not fully captured by the low-gamma activity in our investigation. If this were true, only TPJ electrodes—which are critical in facilitating theory of Mind processes (Saxe and Kanwisher, 2003; Young et al., 2010)—would exhibit an increase in low-gamma activity. Future research, however, will be required to disentangle linguistic from mentalizing factors that support non-literal interpretation.

The between-electrode analysis performed on the power of significantly active timepoints revealed stronger high-gamma activation at TPJ electrodes than non-TPJ electrodes in 2 and 3 of 4 patients during question and responses periods, respectively, in our

FIGURE 7
Average power of active timepoints (in µV²/Hz) specific to non-literal processing for both TPJ and non-TPJ (CON = control) electrodes in all frequency bands for each individual participant. In each participant, the power of significantly active clusters from non-literal versus literal contrast were averaged across all Questions/Responses and across electrodes for both TPJ and non-TPJ control electrodes. The averaged powers were then compared in all frequency bands separately. Notably, the active clusters were averaged between −5 to 0 s for Questions and between 0 to 5 s for Responses. The comparison shows significantly more power related to non-literal processing at TPJ electrodes than non-TPJ electrodes (Wilcoxon sign-rank test, $p = 0.01$) for most subjects, especially in delta (1–4 Hz) and high-gamma ranges [LG, lower-gamma (1–30 Hz); HG, high-gamma (70–200 Hz)]. Notably, certain frequency bands do not show bars due to the absence of any significantly active clusters resulting from non-literal versus literal contrast in those frequency bands. This indicates that there were no significant differences between non-literal and literal conditions in those frequency bands.

dataset. This suggests that it may be a marker of non-literal processing despite differences in anatomical localization across TPJ sub-regions, which varied across patients. Given that the TPJ has traditionally been implicated in ToM processes (Channon et al., 2005), one likely explanation of this finding is that the high-gamma activity is related to ToM-related mechanisms involved in non-literal comprehension. In non-literal situations, our participants must be assessing characters' mental states and intentions in addition to literal meanings, based on contextual information given in vignettes (Hauptman et al., 2023). This is reflected in high-gamma oscillations during both question and response periods. The increased TPJ activity in our study is corroborated by previous investigations that found enhanced BOLD activation during non-literal comprehension in brain regions (such as medial prefrontal cortex, TPJ) that support ToM processes (Rapp et al., 2010; Shibata et al., 2010; Prat et al., 2012; Spotorno et al., 2012; Uchiyama et al., 2012; Baptista et al., 2018). Thus, our study provides a bridge between non-invasive imaging studies and localized electrophysiological activity normally not observable in human brains.

We also found non-literal processing-related activity in lower frequency bands, particularly in the delta band (1–4 Hz). For both between-condition as well as between-electrode comparisons, delta band activity was highly increased at the TPJ across all participants in question periods but only in 1 and 2 out of 4 participants in response periods. Overall, this makes delta activation comparatively a less robust but an important marker of non-literal comprehension. Functionally, delta oscillations are related to attentional processes (Knyazev, 2012; Ko et al., 2017), and found to reflect executive functioning while understanding non-literal sentences (Spotorno et al., 2013; Sun et al., 2022). Thus, we cannot rule out differences in the attentional state of patients and its effects on neural activity. However, the lack of a relationship between behavioral performance and proportion of active timepoints in the delta band (Pearson correlation, $p > 0.05$) in our data suggests these are not related (see Supplementary Table S2). Despite this caveat, slow ongoing oscillations may be a good candidate to coordinate neuronal activity in engaged brain areas during continuous (i.e., extended over a long period of

time and not stimulus-locked) processes, such as those necessary to understand and assign agency during the question period, which takes at least five seconds in our experimental design. Compared to the temporally precise, local high frequency activity, oscillations in low frequencies are thought to integrate over large spatial regions and long temporal scales (Canolty and Knight, 2010). Given they were observed mainly at TPJ electrodes, these slow oscillations may serve to broadcast relevant ToM-related information needed for interpreting non-literal meaning to other associated brain areas and support interaction between other nodes engaged in non-literal processing.

Interestingly, we did not see consistent activity in the intermediate frequency bands (theta, alpha, beta). We were particularly expecting theta activity (4–8 Hz) based on previous research on non-literal processing (Sun et al., 2022), and due to its association with higher-order cognitive processing in temporal-parietal cortex (Brunetti et al., 2014; Meyer et al., 2015; Wang et al., 2016; Cooper et al., 2017; Bramson et al., 2018). For example, magnetoencephalography recordings during a visual perspective task found a spike in theta power in right posterior temporal-parietal cortex during perspective taking ('embodying' another's viewpoint) but not perspective tracking (merely tracking what another person can see), and this embodied processing could be significantly reduced with transcranial magnetic stimulation to the same area (Wang et al., 2016). We did see greater theta activity at TPJ than non-TPJ sites in question periods in both analyses but only in half of our participants. The lack of consistency in theta activation may be related to the inclusion of non-literal expressions other than metaphors (Sun et al., 2022), or perhaps because our task did not elicit a strong degree of perspective-taking.

One limitation in the present study is that we combined three different types of expressions (sarcasm, metaphor, and simile) under the 'non-literal' condition for our analysis. The primary reason to do so was to take our experimental complexity into account. The experimental conditions were difficult given that subjects were awake on the operating room table, and that resulted in limited number of participants and trials per condition. In addition to increasing the number of trials, combining our non-literal expressions allowed us to improve the statistical power of the TPJ versus non-TPJ contrast during non-literal sentence processing. That being said, there are specific differences in terms of the inferencing involved in these expressions, as well as similarities; for instance, they all need extralinguistic information to be interpreted (see Gibbs and Colston (2012) for an overview of standard pragmatic models). The processing of metaphors mainly relies upon analogical reasoning (Willinger et al., 2017) and structural mappings (Gentner and Bowdle, 2008). Like metaphors, similes are comprehended through controlled, comparison processes (Gentner and Bowdle, 2008; Pambuccian and Raney, 2021). Contrarily, the processing of sarcasm uses frame-shifting and conceptual blending operations for semantic reorganization when inconsistencies between incoming information and an initial interpretation are found (Coulson, 2001). Moreover, sarcasm processing also requires social conceptual knowledge (Akimoto et al., 2014) as well as the understanding of the higher-order nature of the speaker's belief (Pexman and Glenwright, 2007; Gibbs and Colston, 2012). We do not completely disregard the possibility that the interpretation of our current findings may be confounded by the processing variations among different non-literal expressions and hence suggest an important direction for further exploration will be to investigate processing of each nonliteral type separately.

In conclusion, the present work enhances our understanding of the neurophysiological basis of TPJ activation while processing non-literal expressions. Leveraging a unique dataset, invasive intracranial recordings from neurosurgical patients, we demonstrate that low-gamma activity is robustly enhanced while understanding non-literal statements as compared to literal statements, both during listening to a question and contemplating a response. In addition to low-gamma activation, delta activity, which reflects slower oscillatory dynamics, increases strongly but inconsistently. We further show that high-gamma activation related to non-literal processing is enhanced to a greater extent in the TPJ area than non-TPJ areas. Future studies are recommended to investigate the causal role of TPJ in the comprehension of non-literal expressions and how multi-frequency oscillations operate to coordinate neuronal activity across implicated brain regions.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by the institutional ethics review board and was in accordance with the guidelines of the Declaration of Helsinki. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

Funding

Acknowledgments

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2023.1304031/full#supplementary-material

References

Aichhorn, M., Perner, J., Weiss, B., Kronbichler, M., Staffen, W., and Ladurner, G. (2009). Temporo-parietal junction activity in theory-of-mind tasks: falseness, beliefs, or attention. *J. Cogn. Neurosci.* 21, 1179–1192. doi: 10.1162/jocn.2009.21082

Akimoto, Y., Sugiura, M., Yomogida, Y., Miyauchi, C. M., Miyazawa, S., and Kawashima, R. (2014). Irony comprehension: social conceptual knowledge and emotional response. *Hum. Brain Mapp.* 35, 1167–1178. doi: 10.1002/hbm.22242

Andrews-Hanna, J. R., Reidler, J. S., Sepulcre, J., Poulin, R., and Buckner, R. L. (2010). Functional-anatomic fractionation of the Brain's default network. *Neuron* 65, 550–562. doi: 10.1016/j.neuron.2010.02.005

Baptista, N. I., Manfredi, M., and Boggio, P. S. (2018). Medial prefrontal cortex stimulation modulates irony processing as indexed by the N400. *Soc. Neurosci.* 13, 495–510. doi: 10.1080/17470919.2017.1356744

Biddle, R., Joshi, A., Liu, S., Paris, C., and Xu, G. (2020). Leveraging sentiment distributions to distinguish figurative from literal health reports on twitter. 1217–1227. WWW '20: Proceedings of The Web Conference 2020, Taipei, Taiwan.

Billow, R. M. (1981). Observing spontaneous metaphor in children. *J. Exp. Child Psychol.* 31, 430–445. doi: 10.1016/0022-0965(81)90028-X

Bohrn, I. C., Altmann, U., and Jacobs, A. M. (2012). Looking at the brains behind figurative language-a quantitative meta-analysis of neuroimaging studies on metaphor, idiom, and irony processing. *Neuropsychologia* 50, 2669–2683. doi: 10.1016/j.neuropsychologia.2012.07.021

Bosco, F. M., Parola, A., Valentini, M. C., and Morese, R. (2017). Neural correlates underlying the comprehension of deceitful and ironic communicative intentions. *Cortex* 94, 73–86. doi: 10.1016/j.cortex.2017.06.010

Boulenger, V., Hauk, O., and Pulvermüller, F. (2009). Grasping ideas with the motor system: semantic somatotopy in idiom comprehension. *Cereb. Cortex* 19, 1905–1914. doi: 10.1093/cercor/bhn217

Bramson, B., Jensen, O., Toni, I., and Roelofs, K. (2018). Cortical oscillatory mechanisms supporting the control of human social-emotional actions. *J. Neurosci.* 38, 5739–5749. doi: 10.1523/JNEUROSCI.3382-17.2018

Brunetti, M., Zappasodi, F., Marzetti, L., Perrucci, M., Cirillo, S., Romani, G., et al. (2014). Do you know what I mean? Brain oscillations and the understanding of communicative intentions. *Front. Hum. Neurosci.* 8:36. doi: 10.3389/fnhum.2014.00036

Buckner, R. L., Andrews-Hanna, J. R., and Schacter, D. L. (2008). The Brain's default network. *Ann. N. Y. Acad. Sci.* 1124, 1–38. doi: 10.1196/annals.1440.011

Burgers, C., van Mulken, M., and Schellens, P. J. (2012). Verbal irony: differences in usage across written genres. *J. Lang. Soc. Psychol.* 31, 290–310. doi: 10.1177/0261927X12446596

Buzsáki, G., Anastassiou, C. A., and Koch, C. (2012). The origin of extracellular fields and currents — EEG, ECoG, LFP and spikes. *Nat. Rev. Neurosci.* 13, 407–420. doi: 10.1038/nrn3241

Canolty, R. T., and Knight, R. T. (2010). The functional role of cross-frequency coupling. *Trends Cogn. Sci.* 14, 506–515. doi: 10.1016/j.tics.2010.09.001

Champagne-Lavau, M., Bolger, D., and Klein, M. (2023). Impact of social knowledge about the speaker on irony understanding: evidence from neural oscillations. *Soc. Neurosci.* 18, 28–45. doi: 10.1080/17470919.2023.2203948

Channon, S., Pellijeff, A., and Rule, A. (2005). Social cognition after head injury: sarcasm and theory of mind. *Brain Lang.* 93, 123–134. doi: 10.1016/J.BANDL.2004.09.002

Citron, F. M. M., Cacciari, C., Funcke, J. M., Hsu, C.-T., and Jacobs, A. M. (2019). Idiomatic expressions evoke stronger emotional responses in the brain than literal sentences. *Neuropsychologia* 131, 233–248. doi: 10.1016/j.neuropsychologia.2019.05.020

Cohen, M. X. (2014). *Analyzing neural time series data: Theory and practice.* Cambridge, MA: MIT Press.

Cohen, M. X. (2019). A better way to define and describe Morlet wavelets for time-frequency analysis. *NeuroImage* 199, 81–86. doi: 10.1016/j.neuroimage.2019.05.048

Cole, M. W., Reynolds, J. R., Power, J. D., Repovs, G., Anticevic, A., and Braver, T. S. (2013). Multi-task connectivity reveals flexible hubs for adaptive task control. *Nat. Neurosci.* 16, 1348–1355. doi: 10.1038/nn.3470

Cooper, P. S., Wong, A. S. W., McKewen, M., Michie, P. T., and Karayanidis, F. (2017). Frontoparietal theta oscillations during proactive control are associated with goal-updating and reduced behavioral variability. *Biol. Psychol.* 129, 253–264. doi: 10.1016/j.biopsycho.2017.09.008

Coulson, S. (2001). *Semantic leaps: Frame-shifting and conceptual blending in meaning construction.* Cambridge: Cambridge University Press.

Dennis, M., Lazenby, A. L., and Lockyer, L. (2001). Inferential language in high-function children with autism. *J. Autism Dev. Disord.* 31, 47–54. doi: 10.1023/A:1005661613288

Dosenbach, N. U. F., Fair, D. A., Miezin, F. M., Cohen, A. L., Wenger, K. K., Dosenbach, R. A. T., et al. (2007). Distinct brain networks for adaptive and stable task control in humans. *Proc. Natl. Acad. Sci.* 104, 11073–11078. doi: 10.1073/pnas.0704320104

Engell, A. D., Huettel, S., and McCarthy, G. (2012). The fMRI BOLD signal tracks electrophysiological spectral perturbations, not event-related potentials. *NeuroImage* 59, 2600–2606. doi: 10.1016/j.neuroimage.2011.08.079

Eviatar, Z., and Just, M. A. (2006). Brain correlates of discourse processing: an fMRI investigation of irony and conventional metaphor comprehension. *Neuropsychologia* 44, 2348–2359. doi: 10.1016/j.neuropsychologia.2006.05.007

Fauconnier, G., and Turner, M. (2008). *The way we think: conceptual blending and the mind's hidden complexities.* New York: Basic books.

Filik, R., Țurcan, A., Ralph-Nearman, C., and Pitiot, A. (2019). What is the difference between irony and sarcasm? An fMRI study. *Cortex* 115, 112–122. doi: 10.1016/j.cortex.2019.01.025

Gentner, D., and Bowdle, B. (2008). "Metaphor as structure-mapping" in *The Cambridge Handbook of Metaphor and Thought.* ed. R. W. Gibbs (Cambridge: Cambridge University Press), 109–128.

Gibbs, R. W. (2000). Irony in talk among friends. *Metaphor. Symb.* 15, 5–27. doi: 10.1080/10926488.2000.9678862

Gibbs, R. W. Jr., and Colston, H. L. (2012). *Interpreting figurative meaning.* Cambridge: Cambridge University Press.

Gregory, C., Lough, S., Stone, V., Erzinclioglu, S., Martin, L., Baron-Cohen, S., et al. (2002). Theory of mind in patients with frontal variant frontotemporal dementia and Alzheimer's disease: theoretical and practical implications. *Brain* 125, 752–764. doi: 10.1093/brain/awf079

Grice, P. (1989). *Studies in the way of words.* Cambridge, MA: Harvard University Press.

Hagoort, P., Hald, L., Bastiaansen, M., and Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science* 304, 438–441. doi: 10.1126/science.1095455

Hancock, J. T. (2004). Verbal irony use in face-to-face and computer-mediated conversations. *J. Lang. Soc. Psychol.* 23, 447–463. doi: 10.1177/0261927X04269587

Happé, F. G. E. (1993). Communicative competence and theory of mind in autism: a test of relevance theory. *Cognition* 48, 101–119. doi: 10.1016/0010-0277(93)90026-R

Happé, F. G. E. (1994). An advanced test of theory of mind: understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *J. Autism Dev. Disord.* 24, 129–154. doi: 10.1007/BF02172093

Hauptman, M., Blank, I., and Fedorenko, E. (2023). Non-literal language processing is jointly supported by the language and theory of mind networks: evidence from a novel meta-analytic fMRI approach. *Cortex* 162, 96–114. doi: 10.1016/j.cortex.2023.01.013

Holdgraf, C. R., de Heer, W., Pasley, B., Rieger, J., Crone, N., Lin, J. J., et al. (2016). Rapid tuning shifts in human auditory cortex enhance speech intelligibility. *Nat. Commun.* 7:13654. doi: 10.1038/ncomms13654

Hubbard, R., Bulkes, N., and Lai, V. T. (2023). Predictability and decomposability separately contribute to compositional processing of idiomatic language. *Psychophysiology* 60:e14269. doi: 10.1111/psyp.14269

Jacques, C., Witthoft, N., Weiner, K. S., Foster, B. L., Rangarajan, V., Hermes, D., et al. (2016). Corresponding ECoG and fMRI category-selective signals in human ventral temporal cortex. *Neuropsychologia* 83, 14–28. doi: 10.1016/J.NEUROPSYCHOLOGIA.2015.07.024

Jolliffe, T., and Baron-Cohen, S. (1999). The strange stories test: a replication with high-functioning adults with autism or Asperger syndrome. *J. Autism Dev. Disord.* 29, 395–406. doi: 10.1023/A:1023082928366

Katz, A. N., and Pexman, P. M. (1997). Interpreting figurative statements: speaker occupation can change metaphor to irony. *Metaphor. Symb.* 12, 19–41. doi: 10.1207/s15327868ms1201_3

Knyazev, G. G. (2012). EEG delta oscillations as a correlate of basic homeostatic and motivational processes. *Neurosci. Biobehav. Rev.* 36, 677–695. doi: 10.1016/j.neubiorev.2011.10.002

Ko, L.-W., Komarov, O., Hairston, W. D., Jung, T.-P., and Lin, C.-T. (2017). Sustained attention in real classroom settings: an EEG study. *Front. Hum. Neurosci.* 11:388. doi: 10.3389/fnhum.2017.00388

Lakatos, P., Shah, A. S., Knuth, K. H., Ulbert, I., Karmos, G., and Schroeder, C. E. (2005). An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. *J. Neurophysiol.* 94, 1904–1911. doi: 10.1152/jn.00263.2005

Lee, S. S., and Dapretto, M. (2006). Metaphorical vs. literal word meanings: fMRI evidence against a selective role of the right hemisphere. *NeuroImage* 29, 536–544. doi: 10.1016/j.neuroimage.2005.08.003

Leslie, A. M., Friedman, O., and German, T. P. (2004). Core mechanisms in 'theory of mind. *Trends Cogn. Sci.* 8, 528–533. doi: 10.1016/j.tics.2004.10.001

Leszczyński, M., Barczak, A., Kajikawa, Y., Ulbert, I., Falchier, A. Y., Tal, I., et al. (2023). Dissociation of broadband high-frequency activity and neuronal firing in the neocortex. *Sci. Adv.* 6:eabb0977. doi: 10.1126/sciadv.abb0977

Levorato, M. C., and Cacciari, C. (2002). The creation of new figurative expressions: psycholinguistic evidence in Italian children, adolescents and adults. *J. Child Lang.* 29, 127–150. doi: 10.1017/S0305000901004950

Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., and Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature* 412, 150–157. doi: 10.1038/35084005

Mashal, N., Faust, M., Hendler, T., and Jung-Beeman, M. (2009). An fMRI study of processing novel metaphoric sentences. *Laterality* 14, 30–54. doi: 10.1080/13576500802049433

Mercier, M. R., Dubarry, A.-S., Tadel, F., Avanzini, P., Axmacher, N., Cellier, D., et al. (2022). Advances in human intracranial electroencephalography research, guidelines and good practices. *NeuroImage* 260:119438. doi: 10.1016/j.neuroimage.2022.119438

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

Mo, S., Su, Y., Chan, R. C. K., and Liu, J. (2008). Comprehension of metaphor and irony in schizophrenia during remission: the role of theory of mind and IQ. *Psychiatry Res.* 157, 21–29. doi: 10.1016/j.psychres.2006.04.002

Pambuccian, F., and Raney, G. (2021). A simile is (like) a metaphor: comparing metaphor and simile processing across the familiarity spectrum. *Canadian J. Exp. Psychol.* 75, 182–188. doi: 10.1037/cep0000242

Parvizi, J., and Kastner, S. (2018). Promises and limitations of human intracranial electroencephalography. *Nat. Neurosci.* 21, 474–483. doi: 10.1038/s41593-018-0108-2

Pexman, P. M., and Glenwright, M. (2007). How do typically developing children grasp the meaning of verbal irony? *J. Neurolinguistics* 20, 178–196. doi: 10.1016/j.jneuroling.2006.06.001

Prat, C. S., Mason, R. A., and Just, M. A. (2012). An fMRI investigation of analogical mapping in metaphor comprehension: the influence of context and individual cognitive capacities on processing demands. *J. Exp. Psychol. Learn. Mem. Cogn.* 38, 282–294. doi: 10.1037/a0026037

Premack, D., and Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behav. Brain Sci.* 1, 515–526. doi: 10.1017/S0140525X00076512

Rapp, A. M., Leube, D. T., Erb, M., Grodd, W., and Kircher, T. T. J. (2007). Laterality in metaphor processing: lack of evidence from functional magnetic resonance imaging for the right hemisphere theory. *Brain Lang.* 100, 142–149. doi: 10.1016/j.bandl.2006.04.004

Rapp, A. M., Mutschler, D. E., and Erb, M. (2012). Where in the brain is nonliteral language? A coordinate-based meta-analysis of functional magnetic resonance imaging studies. *NeuroImage* 63, 600–610. doi: 10.1016/J.NEUROIMAGE.2012.06.022

Rapp, A. M., Mutschler, D. E., Wild, B., Erb, M., Lengsfeld, I., Saur, R., et al. (2010). Neural correlates of irony comprehension: the role of schizotypal personality traits. *Brain Lang.* 113, 1–12. doi: 10.1016/j.bandl.2009.11.007

Reyes-Aguilar, A., Valles-Capetillo, E., and Giordano, M. (2018). A quantitative Meta-analysis of neuroimaging studies of pragmatic language comprehension: in search of a universal neural substrate. *Neuroscience* 395, 60–88. doi: 10.1016/j.neuroscience.2018.10.043

Saxe, R., and Kanwisher, N. (2003). People thinking about thinking people: the role of the temporo-parietal junction in "theory of mind". *NeuroImage* 19, 1835–1842. doi: 10.1016/S1053-8119(03)00230-1

Schneider, S., Wagels, L., Haeussinger, F. B., Fallgatter, A. J., Ehlis, A. C., and Rapp, A. M. (2015). Haemodynamic and electrophysiological markers of pragmatic language comprehension in schizophrenia. *World J. Biol. Psychiatry* 16, 398–410. doi: 10.3109/15622975.2015.1019359

Schurz, M., Radua, J., Aichhorn, M., Richlan, F., and Perner, J. (2014). Fractionating theory of mind: a meta-analysis of functional brain imaging studies. *Neurosci. Biobehav. Rev.* 42, 9–34. doi: 10.1016/j.neubiorev.2014.01.009

Searle, J. R. (1979). *Expression and meaning: Studies in the theory of speech acts.* Expression & Meaning. Cambridge: Cambridge University Press.

Shibata, M., Toyomura, A., Itoh, H., and Abe, J. (2010). Neural substrates of irony comprehension: a functional MRI study. *Brain Res.* 1308, 114–123. doi: 10.1016/j.brainres.2009.10.030

Spotorno, N., Cheylus, A., Van Der Henst, J. B., and Noveck, I. A. (2013). What's behind a P600? Integration operations during irony processing. *PLoS One* 8:6839. doi: 10.1371/journal.pone.0066839

Spotorno, N., Koun, E., Prado, J., Van Der Henst, J. B., and Noveck, I. A. (2012). Neural evidence that utterance-processing entails mentalizing: the case of irony. *NeuroImage* 63, 25–39. doi: 10.1016/j.neuroimage.2012.06.046

Spreng, R. N., and Grady, C. L. (2010). Patterns of brain activity supporting autobiographical memory, prospection, and theory of mind, and their relationship to the default mode network. *J. Cogn. Neurosci.* 22, 1112–1123. doi: 10.1162/jocn.2009.21282

Stone, V. E., Baron-Cohen, S., and Knight, R. T. (1998). Frontal lobe contributions to theory of mind. *J. Cogn. Neurosci.* 10, 640–656. doi: 10.1162/089892998562942

Stringaris, A. K., Medford, N., Giora, R., Giampietro, V. C., Brammer, M. J., and David, A. S. (2006). How metaphors influence semantic relatedness judgments: the role of the right frontal cortex. *NeuroImage* 33, 784–793. doi: 10.1016/j.neuroimage.2006.06.057

Sullivan, K., Winner, E., and Hopfield, N. (1995). How children tell a lie from a joke: the role of second-order mental state attributions. *Br. J. Dev. Psychol.* 13, 191–204. doi: 10.1111/j.2044-835X.1995.tb00673.x

Sun, L., Chen, H., Zhang, C., Cong, F., Li, X., and Hämäläinen, T. (2022). Decoding brain activities of literary metaphor comprehension: an event-related potential and EEG spectral analysis. *Front. Psychol.* 13:913521. doi: 10.3389/fpsyg.2022.913521

Uchiyama, H. T., Saito, D. N., Tanabe, H. C., Harada, T., Seki, A., Ohno, K., et al. (2012). Distinction between the literal and intended meanings of sentences: a functional magnetic resonance imaging study of metaphor and sarcasm. *Cortex* 48, 563–583. doi: 10.1016/j.cortex.2011.01.004

van den Brink, D., Van Berkum, J. J. A., Bastiaansen, M. C. M., Tesink, C. M. J. Y., Kos, M., Buitelaar, J. K., et al. (2012). Empathy matters: ERP evidence for inter-individual differences in social language processing. *Soc. Cogn. Affect. Neurosci.* 7, 173–183. doi: 10.1093/scan/nsq094

Wakusawa, K., Sugiura, M., Sassa, Y., Jeong, H., Horie, K., Sato, S., et al. (2007). Comprehension of implicit meanings in social situations involving irony: a functional MRI study. *NeuroImage* 37, 1417–1426. doi: 10.1016/j.neuroimage.2007.06.013

Wang, H., Callaghan, E., Gooding-Williams, G., McAllister, C., and Kessler, K. (2016). Rhythm makes the world go round: an MEG-TMS study on the role of right TPJ theta oscillations in embodied perspective taking. *Cortex* 75, 68–81. doi: 10.1016/j.cortex.2015.11.011

Wang, A. T., Lee, S. S., Sigman, M., and Dapretto, M. (2006). Neural basis of irony comprehension in children with autism: the role of prosody and context. *Brain* 129, 932–943. doi: 10.1093/brain/awl032

Whalen, J. M., Pexman, P. M., and Gill, A. J. (2009). "Should be fun—not!": incidence and marking of nonliteral language in E-mail. *J. Lang. Soc. Psychol.* 28, 263–280. doi: 10.1177/0261927X09335253

Willinger, U., Deckert, M., Schmöger, M., Schaunig-Busch, I., Formann, A. K., and Auff, E. (2017). Developmental steps in metaphorical language abilities: the influence of age, gender, cognitive flexibility, information processing speed, and analogical reasoning. *Lang. Speech* 62, 207–228. doi: 10.1177/0023830917746552

Yang, J. (2014). The role of the right hemisphere in metaphor comprehension: a meta-analysis of functional magnetic resonance imaging studies. *Hum. Brain Mapp.* 35, 107–122. doi: 10.1002/hbm.22160

Young, L., Dodell-Feder, D., and Saxe, R. (2010). What gets the attention of the temporo-parietal junction? An fMRI investigation of attention and theory of mind. *Neuropsychologia* 48, 2658–2664. doi: 10.1016/j.neuropsychologia.2010.05.012