


Cortical Processing of Arithmetic and Simple Sentences in an Auditory Attention Task

Joshua P. Kulasingham,¹ Neha H. Joshi,^{1,2*} Mohsen Rezaeizadeh,^{1,2*} and  Jonathan Z. Simon^{1,2,3}

¹Department of Electrical and Computer Engineering, ²Institute for Systems Research, and ³Department of Biology, University of Maryland, College Park, Maryland, 20742

Cortical processing of arithmetic and of language rely on both shared and task-specific neural mechanisms, which should also be dissociable from the particular sensory modality used to probe them. Here, spoken arithmetical and non-mathematical statements were employed to investigate neural processing of arithmetic, compared with general language processing, in an attention-modulated cocktail party paradigm. Magnetoencephalography (MEG) data were recorded from 22 human subjects listening to audio mixtures of spoken sentences and arithmetic equations while selectively attending to one of the two speech streams. Short sentences and simple equations were presented diotically at fixed and distinct word/symbol and sentence/equation rates. Critically, this allowed neural responses to acoustics, words, and symbols to be dissociated from responses to sentences and equations. Indeed, the simultaneous neural processing of the acoustics of words and symbols was observed in auditory cortex for both streams. Neural responses to sentences and equations, however, were predominantly to the attended stream, originating primarily from left temporal, and parietal areas, respectively. Additionally, these neural responses were correlated with behavioral performance in a deviant detection task. Source-localized temporal response functions (TRFs) revealed distinct cortical dynamics of responses to sentences in left temporal areas and equations in bilateral temporal, parietal, and motor areas. Finally, the target of attention could be decoded from MEG responses, especially in left superior parietal areas. In short, the neural responses to arithmetic and language are especially well segregated during the cocktail party paradigm, and the correlation with behavior suggests that they may be linked to successful comprehension or calculation.

Key words: cocktail party; isochronous; MEG; selective attention; TRF

Significance Statement

Neural processing of arithmetic relies on dedicated, modality independent cortical networks that are distinct from those underlying language processing. Using a simultaneous cocktail party listening paradigm, we found that these separate networks segregate naturally when listeners selectively attend to one type over the other. Neural responses in the left temporal lobe were observed for both spoken sentences and equations, but the latter additionally showed bilateral parietal activity consistent with arithmetic processing. Critically, these responses were modulated by selective attention and correlated with task behavior, consistent with reflecting high-level processing for speech comprehension or correct calculations. The response dynamics show task-related differences that were used to reliably decode the attentional target of sentences or equations.

Received Feb. 3, 2021; revised July 6, 2021; accepted July 14, 2021.

Author contributions: J.P.K., N.H.J., M.R., and J.Z.S. designed research; J.P.K., N.H.J., and M.R. performed research; J.P.K., N.H.J., M.R., and J.Z.S. analyzed data; J.P.K., N.H.J., M.R., and J.Z.S. wrote the first draft of the paper; J.P.K., N.H.J., M.R., and J.Z.S. edited the paper; J.P.K. and J.Z.S. wrote the paper.

This work was supported by the Defense Advanced Research Projects Agency Grant N660011824024, National Science Foundation Grants SMA-1734892 and DGE-1449815, and the National Institutes of Health Grant R01-DC014085. We thank Ellen Lau for her guidance in the early stages of this project and Anna Namyst for help with MEG measurements.

*N.H.J. and M.R. contributed equally to this work.

The authors declare no competing financial interests.

Correspondence should be addressed to Joshua P. Kulasingham at joshuapk@umd.edu.

<https://doi.org/10.1523/JNEUROSCI.0269-21.2021>

Copyright © 2021 the authors

Introduction

Comprehension and manipulation of numbers and words are key aspects of human cognition and share many common features. Numerical operations may rely on language for precise calculations (Spelke and Tsivkin, 2001; Pica et al., 2004) or share logical and syntactic rules with language (Houdé and Tzourio-Mazoyer, 2003). During numerical tasks, frontal, parietal, occipital, and temporal areas are activated (Menon et al., 2000; Dehaene et al., 2003, 2004; Arsalidou and Taylor, 2011; Maruyama et al., 2012; Dastjerdi et al., 2013; Harvey et al., 2013; Harvey and Dumoulin, 2017). Bilateral intraparietal sulcus (IPS) is activated by presenting numbers using Arabic or alphabetical notation (Pinel et al., 2001) or speech (Eger et al., 2003).

Posterior parietal and prefrontal areas are activated for both arithmetic and language (Price, 2000; Göbel et al., 2001; Venkatraman et al., 2006; Zarnhofer et al., 2012; Bemis and Pylkkänen, 2013). However, some cortical networks activated by numerical stimuli (e.g., IPS), differ from those underlying language processing, even when the stimuli are presented using words (Monti et al., 2012; Park et al., 2012; Amalric and Dehaene, 2016, 2019). Lesion studies (Dehaene and Cohen, 1997; Varley et al., 2005; Baldo and Dronkers, 2007) further provide evidence that the neural basis of numerical processing is distinct from that of language processing (Gelman and Butterworth, 2005; Amalric and Dehaene, 2017).

The dynamics of these neural processes have also been investigated. Evoked responses to arithmetic have been found in parietal, occipital, temporal and frontal regions (Iguchi and Hashimoto, 2000; Kou and Iwaki, 2007; Ku et al., 2010; Jasinski and Coch, 2012; Maruyama et al., 2012; Iijima and Nishitani, 2017). Arithmetic operations can even be decoded from such responses (Pinheiro-Chagas et al., 2019). Speech studies differentiate early auditory evoked components from later components reflecting linguistic and semantic processing in temporal, parietal, and frontal regions (Obleser et al., 2003, 2004; Koelsch et al., 2004; Lau et al., 2008; Baggio and Hagoort, 2011). Linear models of time-locked responses to continuous speech called temporal response functions (TRFs) have also revealed dynamical processing of linguistic features (Di Liberto et al., 2015; Brodbeck et al., 2018a,b; Broderick et al., 2018).

To investigate cortical processing of spoken language and arithmetic, we use a technique pioneered by Ding et al. (2016) of presenting isochronous (fixed rate) words and sentences. There, the single syllable word rate, also the dominant acoustic rate, is tracked strongly by auditory neural responses, as expected. However, cortical responses also strongly track the sentence rate, completely absent in the acoustics, possibly reflecting hierarchical language processing (Sheng et al., 2019; Jin et al., 2020; Luo and Ding, 2020). When subjects selectively attend to one speech stream among several, in a “cocktail party paradigm,” the sentence rate is tracked only for the attended speaker (Ding et al., 2018). Similarly, cocktail party studies using TRFs show early auditory responses regardless of attention, and later attention-modulated responses to higher order speech features (Ding and Simon, 2012; Brodbeck et al., 2018b). Attention modulates activation related to numerical processing as well (Castaldi et al., 2019).

Here, magnetoencephalography (MEG) is used to study the cortical processing of short spoken sentences and simple arithmetic equations, presented simultaneously at fixed sentence, equation, word and symbol rates, in an isochronous cocktail party paradigm. This study is motivated by several questions, of increasing complexity. The most basic is whether isochronously presented equations allow segregation of equation-level from symbol-level neural processing in the frequency domain. We demonstrate strong evidence for this segregation. The next level is whether equation-level and sentence-level processing show shared or distinct cortical activity areas. We demonstrate evidence for both: shared activity in the left temporal lobe, and distinct equation processing in bilateral IPS and occipital lobe. Finally, we address whether the cocktail party listening paradigm can further differentiate between them, and we find that it does: selective attention allows greater differentiation between the higher-level processing, and, critically, also surfaces neural correlations with behavioral measures.

Materials and Methods

Participants

MEG data were collected from 22 adults (average age 22.6 years, 10 female, 21 right handed) who were native English speakers. The participants gave informed consent and received monetary compensation. All experimental procedures were approved by the Internal Review Board of the University of Maryland, College Park. To ensure that the subjects could satisfactorily perform the arithmetic task, only subjects who self-reported that they had taken at least one college level math course were recruited.

Speech stimuli

Monosyllabic words were synthesized with both male and female speakers using the ReadSpeaker synthesizer (<https://www.readspeaker.com>, “James” and “Kate” voices). The language stimuli consisted of four-word sentences, and the arithmetic stimuli consisted of five-word equations. Hereafter, arithmetic words are referred to as “symbols,” arithmetic sentences as “equations,” non-arithmetic words as “words,” and non-arithmetic sentences as “sentences.” The words and symbols were modified to be of constant durations to allow for separate word, symbol, sentence, and equation rates, so that the neural response to each of these could be separated in the frequency domain. The words and symbols were constructed with fixed durations of 375 and 360 ms, respectively, giving a word rate of 2.67 Hz, a symbol rate of 2.78 Hz, a sentence rate of 0.67 Hz, and an equation rate of 0.55 Hz. All the words and symbols were monosyllabic, and hence the syllabic rate is identical to the word/symbol rate. These rates are quite fast for spoken English, and, although intelligible, can be difficult to follow in the cocktail party conditions. Because neural signals below 0.5 Hz are very noisy, however, it was not deemed appropriate to reduce the rates further; preliminary testing showed that these rates were a suitable compromise between ease of understanding and reasonable neural signal-to-noise ratio. In addition, the rates were selected such that each trial, made of either 10 equations or 12 sentences, would have the same duration (18 s), allowing for precise frequency resolution at both rates.

The individual words and symbols were shortened by removing silent portions before their beginning and after their end, and then manipulated to have fixed durations, using the overlap-add resynthesis method in Praat (Boersma and Weenick, 2018). The words and symbols were, respectively, formed into sentences and equations (described below) and were lowpass filtered below 4 kHz using a third order elliptic filter (the air-tube system used to deliver the stimulus has a lowpass transfer function with cutoff \sim 4 kHz). Finally, each stimulus was normalized to have approximately equal perceptual loudness using the MATLAB *integratedLoudness* function.

The equations were constructed using a limited set of symbols consisting of the word “is” (denoting =), three operators [“plus” (+), “less” (–), and “times” (×)], and the eleven English monosyllabic numbers (“nil,” “one” through “six,” “eight” through “ten,” and “twelve”). The equations themselves consisted of a pair of monosyllabic operands (numbers) joined by an operator, an “is” statement of equivalence, and a monosyllabic result; the result could be either the first or last symbol in the equation (e.g., “three plus two is five” or “five is three plus two”). The equations were randomly generated with repetitions allowed, to roughly balance the occurrences of each number (although smaller numbers are still more frequent since there are more mathematically correct equations using only the smallest numbers). The fact that there were a limited set of symbols and that the same symbol “is” occurs in every sentence, in either the second or fourth position, are additional regularities, which contribute to additional peaks in the acoustic stimulus spectrum at the first and second harmonic of the equation rate (1.11 and 1.66 Hz) as seen in Figure 1 (and borne out by simulations). Although less than ideal, it is difficult to avoid in a paradigm when restricting to mathematically well-formed equations. Hence, we do not analyze the neural responses at those harmonic frequencies, since their relative contributions from auditory versus arithmetic processing are not simple to estimate. The sentences were also constructed with two related syntactic structures to be similar to the two equation formats: verb in second position (e.g., “cats drink warm milk”) and verb in third position (e.g., “head

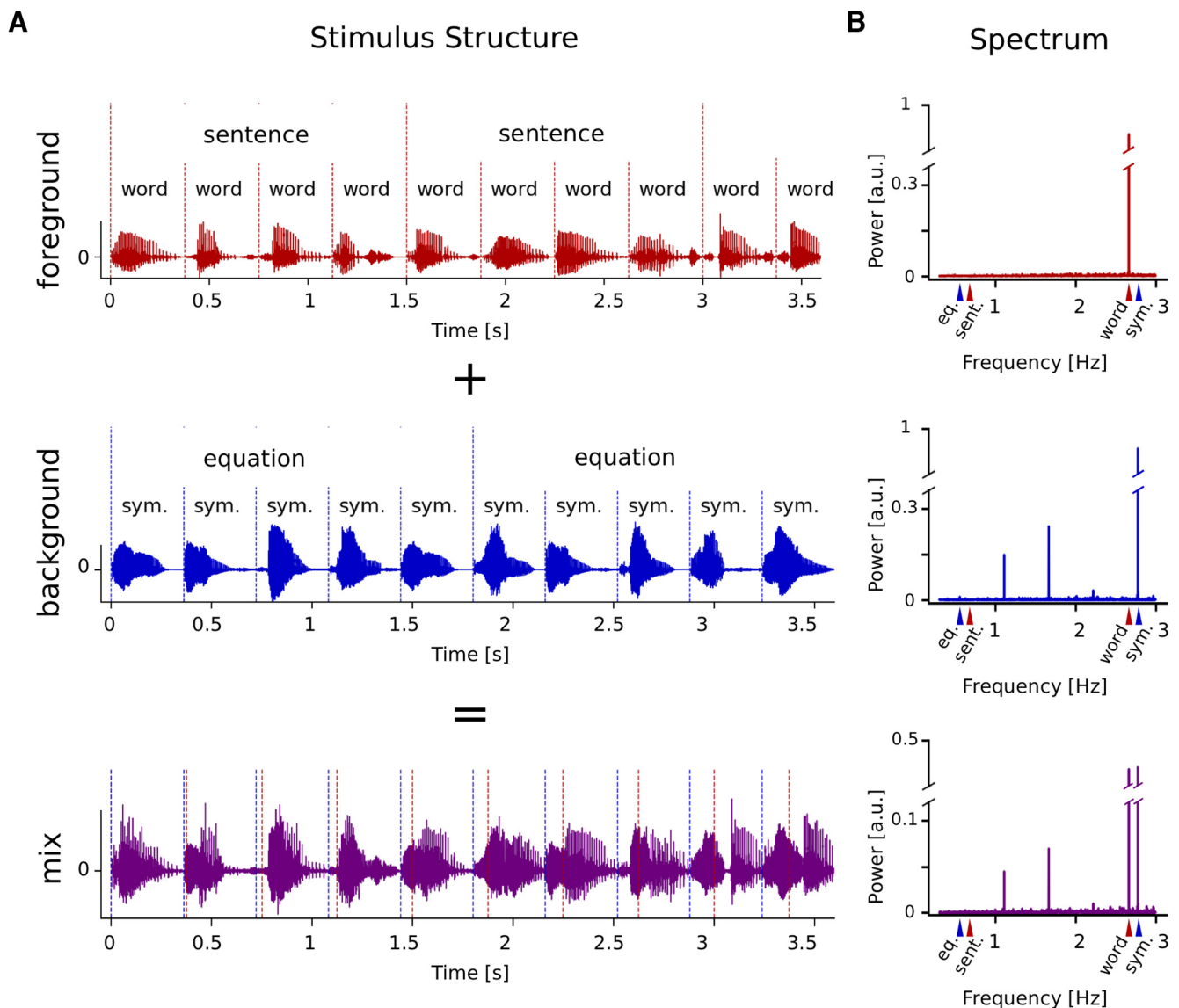


Figure 1. Stimulus structure. **A**, The foreground, background, and mix waveforms for the initial section of the stimulus for a two-speaker attend-language trial. The sentence, equation, word, and symbol structures are shown. The word and symbol rhythms are clearly visible in the waveforms. The mix was presented diotically and is the linear sum of both streams. **B**, The frequency spectrum of the Hilbert envelope of the entire concatenated stimulus for the attend-sentences condition (432-s duration). The sentence (0.67 Hz), equation (0.55 Hz), word (2.67 Hz), and symbol (2.78 Hz) rates are indicated by colored arrows under the x-axis. Clear word and symbol rate peaks are seen in the foreground and background, respectively, while the mix spectrum has both peaks. Note that there are no sentence rate or equation rate peaks in the stimulus spectrum. The appearance of harmonics of the equation rate are consistent with the limited set of math symbols used.

chef bakes pie”), but unlike the “is” of the arithmetic case, the verb changed with every sentence and there were no analogous harmonic peaks in the sentence case. Deviants were also constructed: deviant equations were properly structured but mathematically incorrect (e.g., “one plus one is ten”); analogously, deviant sentences were syntactically correct but semantically nonsensical (e.g., “big boats eat cake”). Cocktail party stimuli were constructed by adding the acoustic waveforms of the sentences and equations in a single audio channel (Fig. 1) and presented diotically (identical for both ears). The speakers were different (male and female), to simplify the task of segregating diotically presented speech. The mixed speech was then normalized to have the same loudness as all the single speaker stimuli using the abovementioned algorithm.

Experimental design

The experiment was conducted in blocks: four single speaker blocks (2 × 2: male and female, sentences and equations) were followed by eight cocktail party blocks (see Table 1). The order of the gender of the speaker was counterbalanced across subjects. Each block consisted of

Table 1. Experiment block structure

Foreground	Background	Speaker foreground (background)	Number of trials per block
Equations	—	Male	10
Sentences	—	Male	10
Equations	—	Female	10
Sentences	—	Female	10
Sentences	Equations	Female (male)	6
Equations	Sentences	Female (male)	6
Sentences	Equations	Male (female)	6
Equations	Sentences	Male (female)	6
Sentences	Equations	Female (male)	6
Equations	Sentences	Female (male)	6
Sentences	Equations	Male (female)	6
Equations	Sentences	Male (female)	6

The experiment consisted of four single speaker blocks followed by eight cocktail party blocks. Each trial was 18 s in duration and consisted of 10 equations (1.8 s × 10 = 18 s) or 12 sentences (1.5 s × 12 = 18 s). The speaker gender was counterbalanced across subjects (i.e., the order of column 3 was changed).

multiple trials: 10 for single speaker and six for cocktail party as shown in Table 1; 50% of the blocks had one deviant trial. Each trial consisted of 10 equations or 12 sentences (or both, for cocktail party conditions) and was 18 s in duration for all cases ($0.360 \text{ s/symbol} \times \text{five symbols/equations} \times 10 \text{ equations} = 18 \text{ s}$; $0.375 \text{ s/word} \times \text{four words/sentence} \times 12 \text{ sentences} = 18 \text{ s}$). In total, the single speaker conditions had 240 sentences and 200 equations, and the cocktail party conditions had 288 sentences and 240 equations in the foreground. Deviant trials had four equations or five sentences being deviants. At the start of each block, the subject was instructed which stimulus to attend to, and was asked to press a button at the end of each trial to indicate whether a deviant was detected (right button: yes; left button: no). The subjects kept their eyes open, and a screen indicated which voice they should attend to (“Attend Male” or “Attend Female”) while the stimulus was presented diotically. After each trial, the stimulus was paused, and the screen displayed the text “Outlier?” until the subjects pressed one of the two buttons. There was a 2-s break after the button press, after which the next trial stimulus was presented.

Since the deviant detection task was challenging, especially in the cocktail party case, subjects were asked to practice detecting deviants just before they were placed inside the MEG scanner (two trials of language, two trials of arithmetic, with stimuli not used during the experiment). Most subjects reported that it was easier to follow and detect deviants in the equations compared with the sentences. This might arise for several reasons, e.g., because the equations had a restricted set of simple numbers, or because the repetitive “is” symbol helped keep track of equation structure.

This experiment was not preregistered. The data are available at <https://doi.org/10.13016/xd2i-vyke> and the code is available at <https://github.com/jpkulasingham/cortical-sentence-equation>.

MEG data acquisition and preprocessing

A 157 axial gradiometer whole head MEG system (Kanazawa Institute of Technology) was used to record MEG data while subjects rested in the supine position in a magnetically shielded room (VAC). The data were recorded at a sampling rate of 2 kHz with an online 500-Hz low pass filter, and a 60-Hz notch filter. Saturating channels were excluded (approximately two channels on average) and the data were denoised using time-shift principal component analysis (de Cheveigné and Simon, 2007) to remove external noise, and sensor noise suppression (de Cheveigné and Simon, 2008) to suppress channel artifacts. All subsequent analyses were performed in mne-python 0.19.2 (Gramfort, 2013; Gramfort et al., 2014) and eelbrain 0.33 (Brodbeck et al., 2020b). The MEG data were filtered from 0.3 to 40 Hz using an FIR filter (mne-python 0.19.2 default settings), downsampled to 200 Hz, and independent component analysis was used to remove artifacts such as eye blinks, heartbeats, and muscle movements.

Frequency domain analysis

The complex-valued spectrum of the MEG response for each sensor was computed using the discrete Fourier transform (DFT). The preprocessed MEG responses were separated into four conditions: attending math or language, in single speaker or cocktail party conditions. The male and female speaker blocks were combined for all analysis. Within each condition, the MEG responses for each trial were concatenated to form signals of duration 6 min for each of the single speaker conditions and 7.2 min for each of the cocktail party conditions. The DFT was computed for each sensor in this concatenated response, leading to a frequency resolution of 2.7×10^{-3} Hz for the single speaker conditions and 2.3×10^{-3} Hz for the cocktail party conditions. The amplitudes of the frequency spectra were averaged over all sensors and tested for significant frequency peaks (described below, Statistical analysis).

Frequencies of interest were selected corresponding to the equation rate (0.555 Hz), the sentence rate (0.667 Hz), the symbol rate (2.778 Hz), and the word rate (2.667 Hz). Note that the duration of the signals is an exact multiple of both the symbol and the word durations, ensuring that the frequency spectrum contained an exact DFT value at each of these four rates. In addition, the neighboring 5 frequency values (width of ~ 0.01 Hz) on either side of these key frequencies were also selected to be used in a noise model for statistical tests.

Neural source localization

The head shape of each subject was digitized using a Polhemus 3SPACE FASTRAK system, and head position was measured before and after the experiment using five marker coils. The marker coil locations and the digitized head shape were used to co-register the template FreeSurfer “fsaverage” brain (Fischl, 2012) using rotation, translation and uniform scaling. A volume source space was formed by dividing the brain volume into a grid of 12-mm sized voxels. This source space was used to compute an inverse operator using minimum norm estimation (MNE; Hämäläinen and Ilmoniemi, 1994); with a noise covariance estimated from empty room data. Each sensor’s response was concatenated over all trials in each condition, and the Fourier transform was used to compute the complex-valued frequency spectrum, with both amplitude and phase, at that sensor. The values of these spectra at each of the 44 selected frequencies (four frequencies of interest with ten sidebands each) are complex-valued sensor distributions, which were source-localized independently using MNE onto the volume source space, giving complex-valued source activations (Simon and Wang, 2005). The amplitudes of these complex-valued source activations were used for subsequent analysis. Finally, the sideband source distributions were averaged together to form the noise model.

TRFs

The preprocessed single-trial MEG responses in each of the four conditions (excluding deviant trials) were source-localized in the time domain using MNE, similar to the method described above in the frequency domain. The MEG signals were further lowpassed below 10 Hz using an FIR filter (default settings in mne python) and downsampled to 100 Hz for the TRF analysis. These responses were then used along with representations of the stimulus to estimate TRFs. The linear TRF model for P predictors (stimulus representations) is given by

$$y(t) = \sum_{p=1}^P \sum_d \tau^{(p)}(d) x^{(p)}(t-d) + n(t), \quad (1)$$

where $y(t)$ is the response at a neural source at time t , $x^{(p)}(t-d)$ is the time shifted p^{th} predictor (e.g., speech envelope, word onsets, etc., as explained below) with time lag of d , $\tau^{(p)}(d)$ is the value of the TRF corresponding to the p^{th} predictor at lag d , and $n(t)$ is the residual noise. The TRF estimates the impulse response of the neural system for that predictor, and can be interpreted as the average time-locked response to continuous stimuli (Lalor and Foxe, 2010). For this analysis, several predictors were used to estimate TRFs at each neural source using the boosting algorithm (David et al., 2007), as implemented in eelbrain, thereby separating the neural response to different features. The boosting algorithm may result in overly sparse TRFs, and hence an overlapping basis of 30-ms Hamming windows (with 10-ms spacing) was used to allow smoothly varying responses. For the volume source space, the TRF at each voxel for a particular predictor is a vector that varies over the time lags, representing the amplitude and direction of the current dipole activity.

The stimulus was transformed into two types of representations that were used for TRF analysis: acoustic envelopes and rhythmic word/symbol or sentence/equation onsets. Although we were primarily interested in responses to sentences and equations, a linear model with only sentence/equation onsets would be disadvantaged by the fact that these representations are highly correlated with the acoustics. Hence by jointly estimating the acoustic envelope and word onset TRFs in the model, the lower-level acoustic responses are automatically separated, allowing the dominantly higher-level processing to emerge in the sentence/equation TRFs. The acoustic envelope was constructed using the 1- to 40-Hz bandpassed Hilbert envelope of the audio signal (FIR filter used above). The onset representations were formed by placing impulses at the regular intervals corresponding to the onset of the corresponding linguistic unit. The four onset responses were: impulses at 375-ms spacing for word onsets, 360 ms for symbol onsets, 1500 ms for sentence onsets, and 1800 ms for equation onsets. Values at all other time points in these

onset representations were set to zero. In order to separate out responses to stimulus onset and offset, the first and last sentences were assigned separate onset predictors, which were not analyzed further except to note that their TRFs showed strong and sustained onset and offset responses, respectively. The remaining (middle) sentences' onsets were combined into one predictor that was used for further TRF analysis. The same procedure was followed for the equation onset predictors.

For each of the two single-speaker conditions, five predictors were used in the TRF model: the corresponding three sentence/equation onsets (just described), word/symbol onsets and the acoustic envelope. For each of the two cocktail conditions, ten predictors were used in the TRF model: the abovementioned five predictors, for each of the foreground and the background stimuli. The predictors were fit in the TRF model jointly, without any preference given to one of them over another.

The TRF for the speech envelope and the word/symbol onsets were estimated for time lags of 0–350 ms to limit the TRF duration to before the onset of the next word (at 375 ms) or symbol (at 360 ms). The sentence and equation TRFs were estimated starting from 350 ms to avoid onset responses, as well as lagged responses to the previous sentence. The sentence TRF was estimated until 1850 ms (350 ms past the end of the sentence) and the equation TRF was estimated until 2150 ms (350 ms past the end of the equation), to detect lagged responses. These sentence and equation TRFs were used to further analyze high level arithmetic and language processing.

Decoder analysis

All decoding analyses were performed using scikit-learn (Pedregosa et al., 2011) and mne-python software. To investigate the temporal dynamics of responses, linear classifiers were trained on the MEG sensor space signals bandpassed 0.3–10 Hz at 200-Hz sampling frequency. Decoders were trained directly on the sensor space signals, since the linear transformation to source space cannot increase the information already present in the MEG sensor signals. The matrix of observations $\mathbf{X} \in \mathbb{R}^{N \times M}$, for N samples and M sensors in each sample, was used to predict the vector of labels $\mathbf{y} \in \{0, 1\}^N$ at each time point of sentences or equations. The labels correspond to the two attention conditions: attend-equations or attend-sentences. The decoders were trained in the single speaker conditions on time points from 0 to 1500 ms for both 1500-ms-long sentences and 1800-ms-long equations. Therefore, the decoder at each time point learns to predict the attended stimulus type (equations or sentences) using the MEG sensor topography at that time point.

In a similar manner, the operator type in the arithmetic condition was also decoded from the MEG sensor topographies at each time point, in the 720-ms time window of each equation that contained the operator and its subsequent operand. Three decoders were trained for the three comparisons (“plus” vs “less,” “less” vs “times,” and “plus” vs “times”).

To further investigate the patterns of cortical activity, linear classifiers were trained on the source localized MEG responses at each voxel, with $\mathbf{X} \in \mathbb{R}^{N \times T}$ for N samples and T time points in each sample. The response dynamics of the entire sentence/equation may not be suitable for decoding: since the equations are comprised of five symbols, while the sentences of four words, this might lead to decoding the equations versus sentences based on whether there were five versus four auditory responses to acoustic onsets. To minimize this confound, two types of classifiers were used based on responses to only one word/symbol (and hence with only one acoustic onset): (1) decoding based on first words: the first symbol of each equation and first word of each sentence was used as the sample, with a label denoting attend equations or attend sentences conditions; (2) decoding based on last words: the last symbol or word was used. Words of duration 375 ms were downsampled to match the duration of the symbols (360 ms), to have equal length training samples. This method was used separately for both the single speaker and the cocktail party conditions. The decoder at each voxel learns to predict the attended stimulus type (equations or sentences) using the temporal dynamics of the response at that voxel.

Finally, the effect of attention was investigated using two sets of classifiers for equations and sentences at each voxel. For the attend-equations classifier, the cocktail party trials were separated into samples at

the 12 equation boundaries, and the labels were marked as “1” when math was attended to and “0” when not. The time duration T was 0–1800 ms (entire equation). For the attend-sentences classifier, the cocktail party trials were separated into samples at the ten sentence boundaries and the labels were “1” when attending to sentences and “0” otherwise. The time duration T was 0–1500 ms (entire sentence). Therefore, the attend-equations decoder at each voxel learns to predict whether the equation stimulus was attended to using the temporal dynamics of the response to the equation at that voxel (and similarly for the attend-sentences decoder).

In summary, the decoders at each time point reveal the dynamics of decoding attention to equations versus sentences from MEG sensor topographies, and the decoders at each voxel reveal the ability to decode arithmetic and language processing in specific cortical areas. The trained classifiers were tested on a separate set and the score of the decoder was computed. Logistic regression classifiers were used, with fivefold cross-validation, within-subject for all the trials. The area under the receiver operating characteristic curve (AUC) was used to quantify the performance of the classifiers.

Statistical analysis

Two types of nonparametric permutation tests were performed across subjects to control for multiple comparisons: single threshold max- t tests for the amplitude spectra, and cluster-based permutation tests for the source localized responses. For the former case, the amplitude spectra for each condition were averaged across sensors, and permutation tests were used to detect significant peaks across subjects ($n = 22$). Each frequency value in the spectrum from 0.3 to 3 Hz was tested for a significant increase over the average of the neighboring five values on either side using 10,000 permutations and the single threshold max- t method (Nichols and Holmes, 2002) to adjust for multiple comparisons. In brief, a null distribution of max- t values was calculated using the maximum t values obtained across all frequencies, for each permutation. Any t value in the observed frequency spectra (denoted by t_{obs}) that exceeds the 95th percentile of the max- t null distribution was deemed significant. For these tests, we report the p values and the t_{obs} values, and deliberately omit the degrees of freedom to avoid direct comparison between the two, since the p values are derived entirely from the permutation distribution of max- t values and not from the t distribution. Correlation tests were also performed to investigate associations between different responses (e.g., sentence rate vs equation rate) within each subject. Pearson correlation tests with Holm–Bonferroni correction were used on the responses at the frequencies of interest, after subtracting the average of the five neighboring bins on either side.

Cluster based permutation tests were performed for the source-localized responses. The source distributions for each individual were mapped onto the FreeSurfer fsaverage brain, to facilitate group statistics. To account for individual variability and mislocalization during this mapping, the distributions were spatially smoothed using a Gaussian window with a standard deviation of 12 mm for all statistical tests. The source localized frequency responses were tested for a significant increase over the corresponding noise model formed by averaging the source localized responses of the five neighboring frequencies on either side. Nonparametric permutation tests (Nichols and Holmes, 2002) and threshold free cluster enhancement (TFCE; Smith and Nichols, 2009) were performed to compare the response against the noise and to control for multiple comparisons. A detailed explanation of this method can be found in Brodbeck et al. (2018a,b). Briefly, a test statistic (in this case, paired samples t statistics between true responses and noise models) is computed for the true data and 10,000 random permutations of the data labels. The TFCE algorithm is applied to these statistics, to enhance continuous clusters of large values, and a distribution consisting of the maximum TFCE value for each permutation is formed. Any value in the original TFCE map that exceeds the 95th percentile is considered significant at the 5% significance level. In all subsequent results, the minimum p value and the maximum or minimum t value across voxels is reported as p_{min} , t_{max} , or t_{min} respectively. Note that the p_{min} is derived from the permutation distribution and cannot be derived directly from t_{max} or t_{min} using the t distribution (degrees of freedom are also omitted because of

this reason). Lateralization tests were performed by testing each voxel in the left hemisphere with the corresponding voxel in the right, using permutation tests and TFCE with paired samples t statistics. For the attend math conditions, equations were separated by operator type (+, −, or ×) to test for specific responses to each operator. No significant differences were found between the source localized responses to each operator.

To test for significant effects and interactions, repeated measures ANOVAs were performed on the source localized responses at each frequency of interest after subtracting the corresponding noise model. Nonparametric permutation tests with TFCE were used to correct for multiple comparisons, similar to the method described above. In brief, a repeated-measures ANOVA is performed at each voxel, and then, for each effect or interaction, the voxel-wise F values from this ANOVA are passed into the TFCE algorithm, followed by permutation tests as described earlier. This method detects significant clusters in source space for each significant effect. Note that the maximum F value in the original map within a cluster (F_{\max}) and the p value of the cluster are reported (and degrees of freedom omitted), for the same reasons as those explained in the previous paragraph (i.e., p values are derived from the permutation distribution and not the F distribution).

Several types of repeated measures ANOVAs were performed using the abovementioned method. In the single speaker case, a 2×2 ANOVA with factors stimulus (“language” for words/sentences or “math” for symbols/equations) and frequency (“low” for sentence/equation and “high” for word/symbol) was performed. For the cocktail party case, a $2 \times 2 \times 2$ ANOVA with the added factor of attention (attended or unattended) was performed. In addition, two further ANOVAs were performed to investigate hemispheric effects, using an additional factor of hemisphere for both the single speaker ($2 \times 2 \times 2$ ANOVA) and the cocktail party ($2 \times 2 \times 2 \times 2$ ANOVA) conditions.

To investigate significant ANOVA effects further, *post hoc t* tests across subjects were performed on the responses averaged across voxels within the relevant significant cluster. For this scalar t test, a Holm–Bonferroni correction was applied to correct for multiple comparisons. For these tests, the t values with degrees of freedom, corrected p values and Cohen’s d effect sizes are reported.

Behavioral responses for the deviant detection task were classified as either correct or incorrect, and the number of correct responses for each subject was correlated with the source localized response power of that subject. The noise model for each frequency of interest was subtracted from the response power before correlating with behavior. Nonparametric permutation tests with TFCE were used in a manner similar to that given above. The only difference was that the statistic used for comparison was the Pearson correlation coefficient between the two variables (behavior and response power), and the maximum correlation coefficient across voxels is reported as r_{\max} .

The TRFs were tested for significance using vector tests based on Hotelling’s T^2 statistic (Mardia, 1975). Since the TRFs consist of time-varying vectors, this method tests consistent vector directions across all subjects at each time point and each voxel. The Hotelling’s T^2 statistic was used with non-parametric permutation tests and TFCE as described above, with the added dimension of time, and the maximum T^2 statistic across voxels is reported as T^2_{\max} . This statistic is more suitable than a t statistic based on the amplitude of the TRF vectors, since activation from distinct neural processes may have overlapping localizations (because of the limited resolution of MEG), but different current directions.

Finally, the decoders were tested across subjects for a significant increase in decoding ability above chance ($AUC = 0.5$) at each time point or at each voxel. Multiple comparisons were controlled for using permutation tests and TFCE, similar to the above cases, with AUC as the test statistic.

Results

Behavioral results

After each trial, subjects indicated whether that trial contained a deviant by pressing a button. The single speaker conditions had higher deviant detection accuracies (equations: mean = 89.5%, SD = 10.7%; sentences: mean = 73.4%, SD = 13.8%) than the cocktail party conditions (equations: mean = 79.9%, SD = 13.3%;

sentences: mean = 61%, SD = 19.4%). Subjects reported that the equations were perceptually easier to follow than the sentences, consistent with the fact that the equations were formed using a smaller set of monosyllabic numbers to preserve the symbol rates. The presence of “is” in each equation may have also contributed to subjects tracking equation boundaries.

Frequency domain analysis

The response power spectrum was averaged over all sensors and a permutation test with the max- t method was performed to check whether the power at each frequency of interest was significantly larger than the average of the neighboring five frequency bins on either side (Fig. 2A,B) across subjects ($n = 22$). For the language single speaker condition, the sentence rate (0.67 Hz, $t_{\text{obs}} = 7.25$, $p < 0.001$; note: degrees of freedom not shown since p values are derived from the permutation test, see Materials and Methods, Statistical analysis), its first harmonic (1.33 Hz, $t_{\text{obs}} = 6.11$, $p = 0.0023$), and the word rate (2.67 Hz, $t_{\text{obs}} = 12.98$, $p < 0.001$) were significant (one tailed permutation test of difference of amplitudes with max- t method). Similarly, for the math single speaker condition, the symbol rate (2.78 Hz, $t_{\text{obs}} = 12.39$, $p < 0.001$) and the equation rate (0.55 Hz, $t_{\text{obs}} = 6.29$, $p = 0.0017$) were significant. In this condition, the first and second harmonics of the equation rate were also significant ($t_{\text{obs}} = 7.28$, $p < 0.001$ at 1.11 Hz; $t_{\text{obs}} = 7.77$, $p < 0.001$ at 1.67 Hz). Thus, in both conditions, the responses track the corresponding sentence or equation rhythms that are not explicitly present in the acoustic signal. The harmonic peak (1.33 Hz) in the language condition is consistent with phrase tracking (Ding et al., 2016), and the harmonics in the arithmetic condition (1.11 Hz, 1.66 Hz) are consistent with auditory processing of acoustic properties of the stimulus associated with the limited number of mathematical symbols employed (see Materials and Methods, Speech stimuli), or higher-order processing, or both. Correlation tests within subjects, with Holm–Bonferroni correction, were performed on relevant pairs of responses (after subtracting the neighboring bins). Sentence rate responses were significantly correlated with equation rate responses (Pearson’s $r = 0.576$, $p = 0.015$). Word rate responses were significantly correlated with symbol rate responses ($r = 0.681$, $p = 0.001$). Since such correlations may arise from fluctuating degree of task engagement, or variable neural signal-to-noise ratio across subjects, they were not analyzed further. There were no significant sentence versus word ($r = 0.067$, $p > 0.99$) or equation versus symbol ($r = 0.001$, $p > 0.99$) response correlations.

For the attend-sentences cocktail party condition, only the (attended) word, (unattended) symbol, and (attended) sentence rate responses were significant (one tailed permutation test of difference of amplitudes with max- t method; $t_{\text{obs}} = 9.29$, $p < 0.001$; $t_{\text{obs}} = 10.59$, $p < 0.001$; $t_{\text{obs}} = 5.46$, $p = 0.0176$, respectively) as shown in Figure 2C,D. The (unattended) equation rate response was not significant ($t = 2.99$, $p > 0.99$). On the other hand, for the attend-equations cocktail party condition, the (unattended) word, (attended) symbol, and (attended) equation rate responses were significant ($t_{\text{obs}} = 10.86$, $p < 0.001$; $t_{\text{obs}} = 11.64$, $p < 0.001$; $t_{\text{obs}} = 6.07$, $p = 0.005$, respectively), while the (unattended) sentence rate response was not significant ($t_{\text{obs}} = 2.73$, $p > 0.99$). Responses at the first and second harmonics of the equation rate were also significant in the attend-equations condition (1.11 Hz, $t_{\text{obs}} = 5.31$, $p = 0.027$; 1.67 Hz, $t_{\text{obs}} = 5.09$, $p = 0.04$). Correlation tests within subjects were performed, similar to the single speaker case, on all responses except the non-significant unattended sentence and equation rates. Once again, attended sentence rate responses were significantly correlated with attended equation rate responses ($r = 0.68$, $p = 0.0023$). Word rate

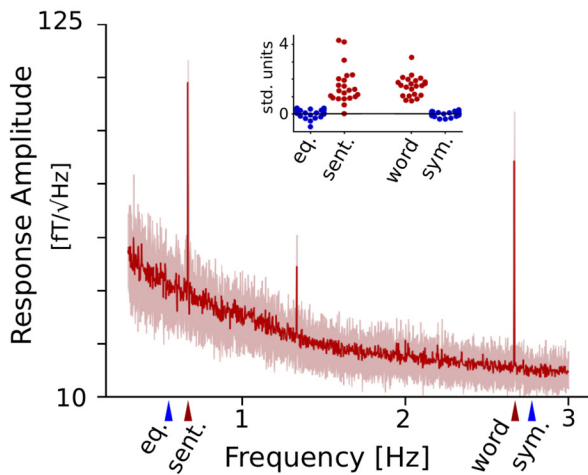
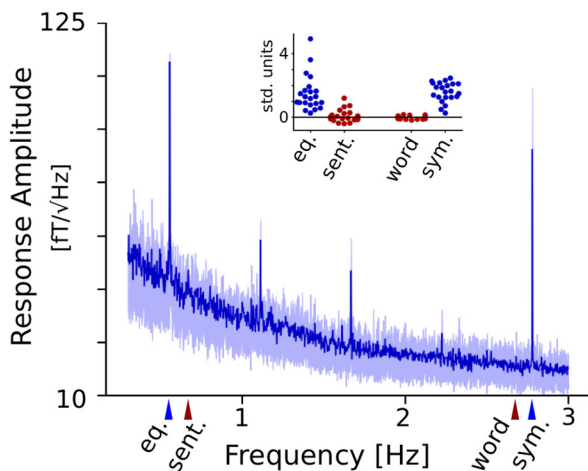
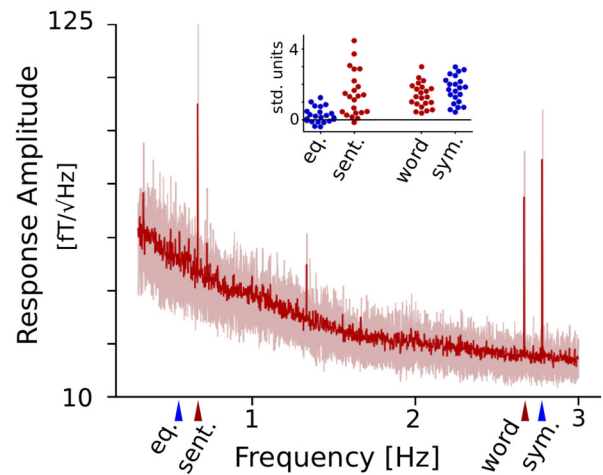
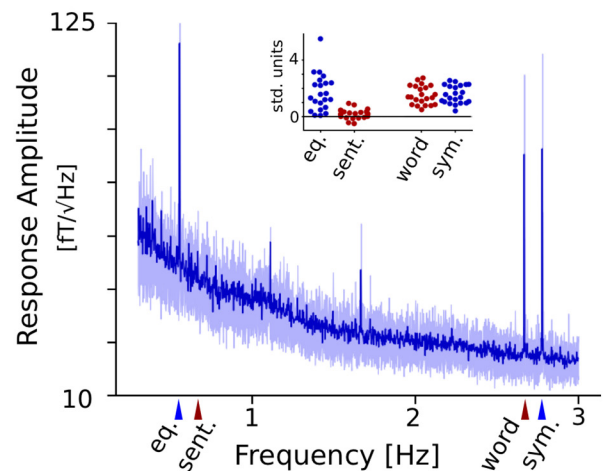
A Single Speaker: Sentences**B** Single Speaker: Equations**C** Cocktail Party: Attend Sentences**D** Cocktail Party: Attend Equations

Figure 2. Neural response spectrum. The MEG response spectrum as a function of frequency for the four conditions. **A:** Single speaker sentences, **B:** Single speaker equations, **C:** Cocktail party attend sentences, **D:** Cocktail party attend equations. The amplitude spectrum, averaged over sensors and subjects, is shown with light shaded regions denoting the first-third quartile range across subjects. Clear peaks are seen at the sentence, equation, word, and symbol rates (indicated by the arrows under the x -axis). These responses were compared against neighboring bins (of width ~ 0.01 Hz, not visible here) for statistical tests. Insets show the average responses at the four frequencies of interest for each subject, after subtracting the neighboring bins. The scale for the insets is standardized within each condition, but with 0 indicating the baseline average activity of the neighboring bins. For the single speaker conditions, peaks appear only at the rates corresponding to the presented stimulus. For the cocktail party conditions, peaks appear at the symbol and word rates regardless of attention, while sentence and equation peaks only appear during the attended condition. There are no analogous sentence or equation peaks during the opposite attention condition.

responses were significantly correlated with symbol rate responses for both attended ($r=0.69$, $p=0.0021$) and unattended cases ($r=0.83$, $p<0.001$). Other correlations were not significant (attended sentence vs attended word: $r=0.12$, $p>0.99$; attended sentence vs unattended word: $r=0.07$, $p=0.74$; attended equation vs attended symbol: $r=0.49$, $p=0.083$; attended equation vs unattended symbol: $r=0.33$, $p=0.4$).

Since the word and symbol rates are present in the acoustics for both conditions, the neural responses at these rates could merely reflect acoustic processing. However, the fact that the sentence and equation rates are significant only in the corresponding attention condition suggests that these

responses may dominantly reflect attention-selective high-level processes. This agrees with prior studies showing a similar effect for language (Ding et al., 2018). Here, we show that this effect occurs even for arithmetic equations. However, arithmetic equations are also sentences, so it is unclear from this result alone if the equation peak reflects merely tracking of sentential structure and not arithmetic processing. To investigate this, we used volume source localization on the responses at the relevant frequencies to determine the cortical distribution of these responses.

The responses at the four frequencies of interest (word, symbol, sentence, and equation rates) were source-localized using

the Fourier transform sensor topographies at these frequencies (see Materials and Methods, Neural source localization). The amplitudes of the resulting complex-valued volume source distributions were used for all subsequent analysis. For each frequency of interest, the source amplitudes in the neighboring five bins on either side were calculated using the same source model, and averaged together as an estimate of the background noise. The response distributions for each of these frequencies were tested for a significant increase over the noise estimate using nonparametric permutation tests with paired sample t statistics and TFCE. For both single speaker conditions, the corresponding word or symbol responses were significant ($t_{max} = 12.85$, $p_{min} < 0.001$, and $t_{max} = 12.77$, $p_{min} < 0.001$, respectively) in the regions shown in Figure 3A,B, with the average response being strongest in bilateral auditory cortex. The word and symbol rate responses were not significantly different ($t_{min} = -2.11$, $t_{max} = 3.63$, $p > 0.08$), consistent with low level auditory processing. The corresponding sentence or equation responses were also significant ($t_{max} = 9.92$, $p_{min} < 0.001$, and $t_{max} = 7.68$, $p_{min} < 0.001$, respectively). The source distribution for sentence responses was predominantly in left auditory cortex and temporal lobe, whereas for equations the response was distributed over areas of bilateral temporal, parietal, and occipital lobes. Despite these visually distinct patterns, the two responses were not significantly different ($t_{min} = -3.00$, $t_{max} = 3.36$, $p > 0.12$), perhaps because large portions of the brain show activity synchronized to the rhythm. Both sentence and equation responses were significantly left lateralized in temporal ($t_{max} = 6.69$, $p_{min} < 0.001$) and parietal ($t_{max} = 3.9$, $p_{min} = 0.009$) areas, respectively. No significant differences were seen in the responses at the equation rate when separated according to operator type (+ vs -: $t_{min} = -2.98$, $t_{max} = 1.64$, $p > 0.34$; - vs \times : $t_{min} = -2.08$, $t_{max} = 3.21$, $p > 0.39$; \times vs +: $t_{min} = -2.26$, $t_{max} = 3.01$, $p > 0.31$).

For the cocktail party conditions, similar results were obtained for both word and symbol rate responses (attend sentences: word rate: $t_{max} = 11.9$, $p_{min} < 0.001$, symbol rate: $t_{max} = 12.8$, $p_{min} < 0.001$; attend equations: word rate: $t_{max} = 11.1$, $p_{min} < 0.001$, symbol rate: $t_{max} = 11.01$, $p_{min} < 0.001$). The response was predominantly in bilateral auditory cortices as shown in Figure 3C,D, and the symbol and word rates were not significantly different ($t_{min} = -4.31$, $t_{max} = 2.33$, $p > 0.16$). The attended sentence or equation rate responses were significant ($t_{max} = 6.78$, $p_{min} < 0.001$, and $t_{max} = 7.87$, $p_{min} < 0.001$, respectively) and the localization was similar to the single speaker case, albeit more bilateral for the equation rate response. Indeed, the sentence rate response was significantly left lateralized ($t_{max} = 5.36$, $p_{min} < 0.001$), similar to the single speaker case, but the equation rate response was not ($t_{max} = 2.97$, $p > 0.067$). However, the spatial distribution of the equation rate response was larger in the left hemisphere (Fig. 3); indeed, the source localization of attended sentence responses and attended equation responses were significantly different ($t_{min} = -4.77$, $t_{max} = 2.39$, $p_{min} = 0.013$), with more equation rate responses in the right hemisphere. This indicates that the equation rate response does not originate from the same cortical regions that give rise to the sentence rate response and that the selective attention task is better able to separate these responses. Perhaps surprisingly, the unattended sentence and equation rates were also significant ($t_{max} = 4.02$, $p_{min} = 0.005$, and $t_{max} = 5.31$, $p_{min} < 0.001$, respectively) in small clusters, although such peaks do not appear in the frequency spectrum averaged across all sensors (Fig. 2). Note however, that some individuals did show small peaks at these

rates even in the average spectrum (see points above zero for unattended rates in the insets of Fig. 2C,D).

A repeated measures ANOVA was performed for the single speaker case on the abovementioned source space distributions (as shown in Fig. 3) for each frequency of interest. The 2×2 ANOVA consisted of factors stimulus (“language” for word/sentence or “math” for symbol/equation) and specific frequency (“high” for word/symbol or “low” for sentence/equation). The ANOVA was performed on the response at each voxel and cluster-based permutation tests with TFCE were used to correct for multiple comparisons (see Materials and Methods, Statistical analysis, for choice of reported statistics). The interaction of stimulus \times frequency was not significant ($F_{max} = 10.38$, $p = 0.149$, but see below for an interaction effect in an ANOVA with a factor of hemisphere). A significant main effect of frequency ($F_{max} = 18.63$, $p = 0.006$) was found in a right auditory cluster and a significant main effect of stimulus type ($F_{max} = 21.67$, $p = 0.003$) was found in the left auditory/temporal area. *Post hoc t* tests across subjects were performed on the responses averaged across voxels within the significant clusters for each effect; p values were obtained from the t distribution and then corrected for multiple comparisons using the Holm–Bonferroni method. These tests revealed that the main effect of stimulus was because of a significant increase in both the sentence over the equation responses ($t_{(21)} = 2.96$, $p = 0.037$, Cohen’s $d = 0.54$) and the word over the symbol responses ($t_{(21)} = 2.85$, $p = 0.038$, Cohen’s $d = 0.52$) in the left auditory/temporal cluster, consistent with increased left temporal activity for language over arithmetic. The main effect of frequency was because of a significant increase in both the word over the sentence responses ($t_{(21)} = 3.67$, $p = 0.01$, Cohen’s $d = 1.16$), and the symbol over the equation responses ($t_{(21)} = 3.15$, $p = 0.028$, Cohen’s $d = 0.97$) in the right auditory cluster.

For the cocktail party case, a similar repeated measures ANOVA was performed, but with an additional factor of attention (attended or unattended) leading to a $2 \times 2 \times 2$ design. A significant three-way interaction of stimulus \times attention \times frequency was found in a right parietal cluster ($F_{max} = 15.18$, $p = 0.024$). *Post hoc t* tests across subjects with Holm–Bonferroni correction were performed on the responses averaged across voxels within this cluster. These revealed a significant increase in the equation responses compared with the sentence responses when attended ($t_{(21)} = 3.71$, $p = 0.0103$, Cohen’s $d = 0.82$), but no significant difference when unattended ($t_{(21)} = 2.27$, $p = 0.09$, Cohen’s $d = 0.65$). There was also no significant difference between word and symbol responses both when attended ($t_{(21)} = -0.32$, $p = 0.75$, Cohen’s $d = -0.06$) and unattended ($t_{(21)} = -0.69$, $p = 0.99$, Cohen’s $d = -0.09$). This is consistent with increased responses to equations in right parietal areas only when attended. In addition to this three-way interaction, several two-way interactions and main effects were also detected but were not analyzed further.

Finally, two further ANOVAs were performed with an additional factor of hemisphere for both the single speaker ($2 \times 2 \times 2$ ANOVA) and cocktail party ($2 \times 2 \times 2 \times 2$ ANOVA). For the single speaker case, the three-way interaction was significant (stimulus \times frequency \times hemisphere: $F_{max} = 18.55$, $p = 0.016$) in superior parietal voxels. For the cocktail party case, the 4-way interaction was not significant (attention \times frequency \times stimulus type \times hemisphere: $F_{max} = 8.31$, $p = 0.115$). However, two three-way interactions involving hemisphere were significant (attention \times frequency \times hemisphere: $F_{max} = 13.71$, $p = 0.031$, frequency \times stimulus \times hemisphere: $F_{max} = 18.12$, $p = 0.017$) in

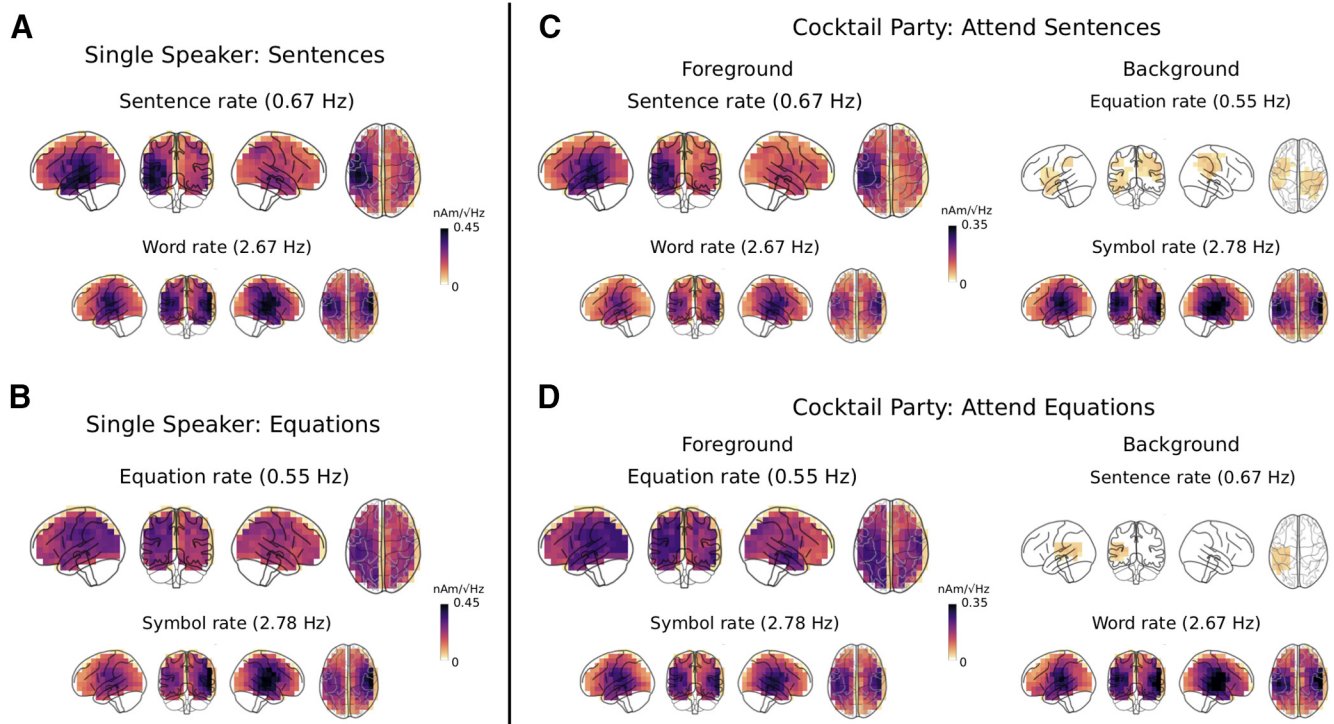


Figure 3. Source localized responses at each frequency of interest. **A:** Single speaker sentences, **B:** Single speaker equations, **C:** Cocktail party attend sentences, **D:** Cocktail party attend equations. The source localized responses at critical frequencies, averaged over subjects and masked by significant increase over the noise model, are shown. Color scales are normalized within each condition to more clearly show the spatial patterns. The word and symbol rate responses are maximal in bilateral auditory cortical areas, while the sentence rate response is maximal in the left temporal lobe. The equation rate responses localize to bilateral parietal, temporal, and occipital areas, albeit with increased left hemispheric activity. Although the background sentence and equation rates also show significant activity, the amplitude of these responses are much smaller than the responses at the corresponding attended rates.

temporal voxels. Other effects involving hemisphere were also found to be significant (frequency \times hemisphere: $F_{\max} = 41.75$, $p < 0.001$, main effect of hemisphere: $F_{\max} = 17.1$, $p = 0.021$), as well as several other effects not involving hemisphere. These effects were not analyzed further, but they indicate that the effects of attention, stimulus and frequency depend significantly on the hemisphere, as already suggested by the lateralized clusters found in the simpler ANOVAs described earlier.

In summary, the ANOVA analysis indicates that, in the single speaker case, low-level responses (word/symbol) are significantly stronger than the higher-level responses (sentence/equation) in right auditory areas and that the language responses (sentence/word) are significantly stronger than the arithmetic responses (equation/symbol) in left auditory/temporal areas. Critically, the ANOVA results for cocktail party indicate that the equation responses are significantly larger than the sentence responses in right parietal areas but only when attended to. ANOVAs also indicate that these effects depend on hemisphere as already suggested by the previous pairwise comparisons.

Behavioral correlations

Behavioral performance was correlated with source localized neural responses using non-parametric permutation tests with TFCE, with Pearson correlation as the test statistic. Deviant detection performance for sentences in the single speaker condition was significantly correlated with the sentence rate neural response ($p_{\min} = 0.02$, maximum correlation in significant regions $r_{\max} = 0.62$) as shown in Figure 4. However, detection of equation deviants in the single speaker condition was not significantly correlated with the equation rate neural response; this may be related to the fact that performance in the single speaker

arithmetic condition was at ceiling for several participants. The performance when detecting sentence deviants in the cocktail party conditions was correlated with the attended sentence rate response ($r_{\max} = 0.62$, $p_{\min} = 0.015$), attended word rate response ($r_{\max} = 0.64$, $p_{\min} = 0.03$) as well as the unattended symbol rate response ($r_{\max} = 0.79$, $p_{\min} = 0.001$). The performance when detecting equation deviants in the cocktail party condition was correlated with the attended equation rate response ($r_{\max} = 0.6$, $p_{\min} = 0.02$) and the unattended word rate response ($r_{\max} = 0.74$, $p_{\min} = 0.04$). It was unexpected that the unattended word and symbol rate responses were significantly correlated with behavior, and possible explanations are discussed in Discussion, Cortical correlates of behavioral performance. Critically, however, sentence and equation rate responses were correlated with behavior only when attended.

TRF analysis

TRF analysis was performed using source localized MEG time signals for each condition after excluding the deviant trials (details in Materials and Methods, TRFs). TRFs were simultaneously obtained for responses to the acoustic envelopes, word/symbol onsets and sentence/equation onsets. Although stimuli with fixed and rhythmic word, symbol, sentence, and equation onsets might lend itself to an evoked response analysis, the fact that the words (or symbols) are only separated by 375 ms (or 360 ms) may lead to high-level late responses overlapping with early auditory responses to the next word (or symbol). In contrast, computing simultaneous TRFs to envelopes and word/symbol onsets in the same model as TRFs to equation/sentence onsets regresses out auditory responses from higher-level responses, providing cleaner TRFs for sentences and equations.

The obtained envelope and word/symbol TRFs were not used for further analysis, since they were dominated by acoustic responses that have been well studied in other investigations (Brodbeck et al., 2018a,b). The volume source localized TRFs are time-varying vectors at each voxel. Activity of nearby neural populations may overlap, even if the activity is driven by different processes, because of the limited spatial resolution of MEG. However, these effects may have different current directions because of the anatomy of the cortical surface. Therefore, a test for consistent vector directions, using Hotelling's T^2 statistic and permutation tests with TFCE, was used to detect group differences in the direction of current flow (see Materials and Methods, Statistical analysis).

The sentence and equation TRFs showed significance over several time intervals and many voxels over the duration of the TRF ($T_{max}^2 = 7.66$, $p_{min} < 0.001$, and $T_{max}^2 = 5.12$, $p_{min} < 0.001$, respectively, see Fig. 5). The TRFs were computed starting from 350 ms after the sentence onset to 350 ms after the end of the sentence, but because of the fixed-period presentation rate without any breaks between sentences in a trial, the TRFs from 0 to 350 ms are identical to the last 350 ms. The large peak at the end (and beginning) of each TRF may either arise from processing of the completion of the sentence/equation, or from preparation (or auditory) processing of the new sentence/sentence/equation, or both. This peak occurs ~60–180 ms after the start of the new sentence/equation, in the typical latency range of early auditory processing. However, spatial distributions of the peak in the equation TRFs seem to indicate patterns that are not consistent with purely auditory processing, especially for the cocktail party condition (described below). Additionally, significant activity is seen throughout the duration of the sentence/equation that is not tied to word/symbol onsets, indicating that lower-level auditory responses have been successfully regressed out. Therefore, the large peaks at the end plausibly reflect processing of the completion of the sentence/equation (with a latency of 420–530 ms after the last word/symbol). The sentence TRF peaks were significant predominantly in the left temporal lobe, while the equation TRF peaks were significant in bilateral temporal, parietal, and motor areas.

Differences in sentence and equation processing were more readily visible in the TRFs for the cocktail party conditions. The test for consistent vector direction revealed similar results to the single speaker conditions (sentence TRF $T_{max}^2 = 5.15$, $p_{min} < 0.001$, equation TRF $T_{max}^2 = 5.24$, $p_{min} < 0.001$) as shown in Figure 6; however, the differences between sentences and equations were more pronounced, especially for the later peaks 410–600 ms after the onset of the last word or symbol. The peaks in the equation TRF were localized to left motor and parietal regions and right inferior frontal areas that are associated with arithmetic processing. This strengthens the hypothesis that these late peaks indicate lagged higher-level processing of the completed equation and not early auditory/preparatory processing of the subsequent equation. Although the cortical localization of sentence TRF peaks remain consistent in left temporal areas

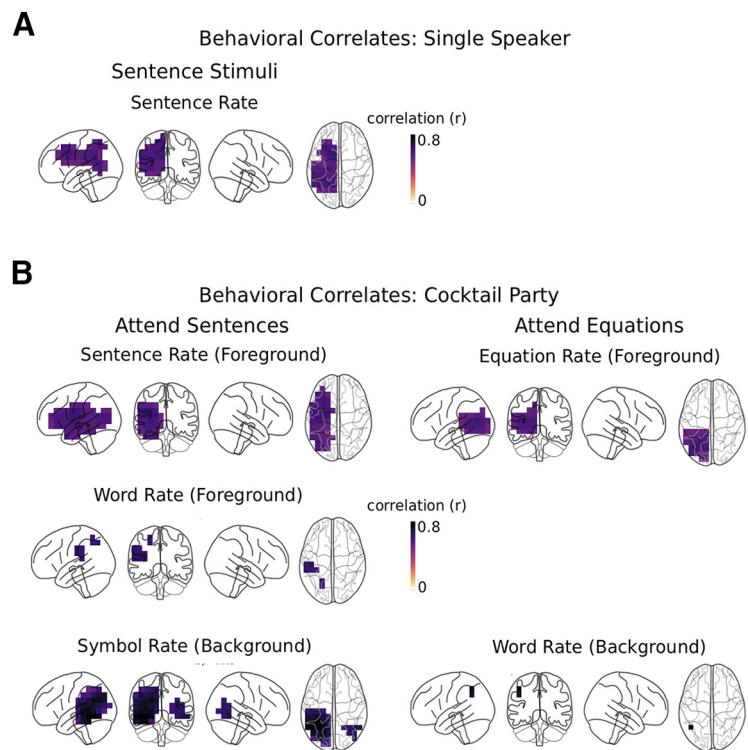


Figure 4. Neural response correlations with behavior. **A:** Single speaker condition, **B:** Cocktail party condition. The source localized responses at the frequencies of interest were correlated with the corresponding deviant detection performance, across subjects. The areas of significant correlation are plotted here (same color scale for all plots). Sentence and equation rate responses are significantly correlated with behavior only if attended, while both attended and unattended word rate responses are significantly correlated with behavior. The sentence rate response is significantly correlated over regions in left temporal, parietal, and frontal areas, while significant correlation for the equation rate response is seen in left parietal and occipital regions.

throughout most of the time course, the equation TRF peaks show several distinct cortical patterns, and may reflect distinct processes. The equation TRF showed strong activity in bilateral IPS, superior parietal and motor areas, while sentence TRFs consistently localized predominantly to regions near left auditory cortex, even more so than in the single speaker case. Therefore, selective attention in the cocktail conditions seems to highlight differences between arithmetic and language processing, and possible explanations are discussed in Discussion. The cocktail party paradigm highlights distinct cortical processes.

Decoder analysis

To further help differentiate between the cortical processing of equations and sentences, two types of linear decoders were trained on neural responses: (1) classifiers at each time point that learned weights based on the MEG sensor topography at that time point; (2) classifiers at each voxel that learned weights based on the temporal dynamics of the response at that voxel. The former was used to contrast the dynamics of equation and sentence processing (Fig. 7A). For the single speaker conditions, all time points showed significant decoding ability across subjects ($t_{max} = 11.3$, $p_{min} < 0.001$), with higher prediction success (as measured by AUC) at longer latencies. For the cocktail party conditions, decoding ability was significantly above chance only at longer latencies ($t_{max} = 6.45$, $p_{min} < 0.001$). While subjects listened to the equations, the identity of the arithmetic operator (e.g., “plus” vs “times” or “less”) was reliably decoded from the MEG sensor topography during the time points when the operator and the subsequent operand were presented (Fig. 7B). Note that decoding accuracy was significantly above chance for time points

Single Speaker

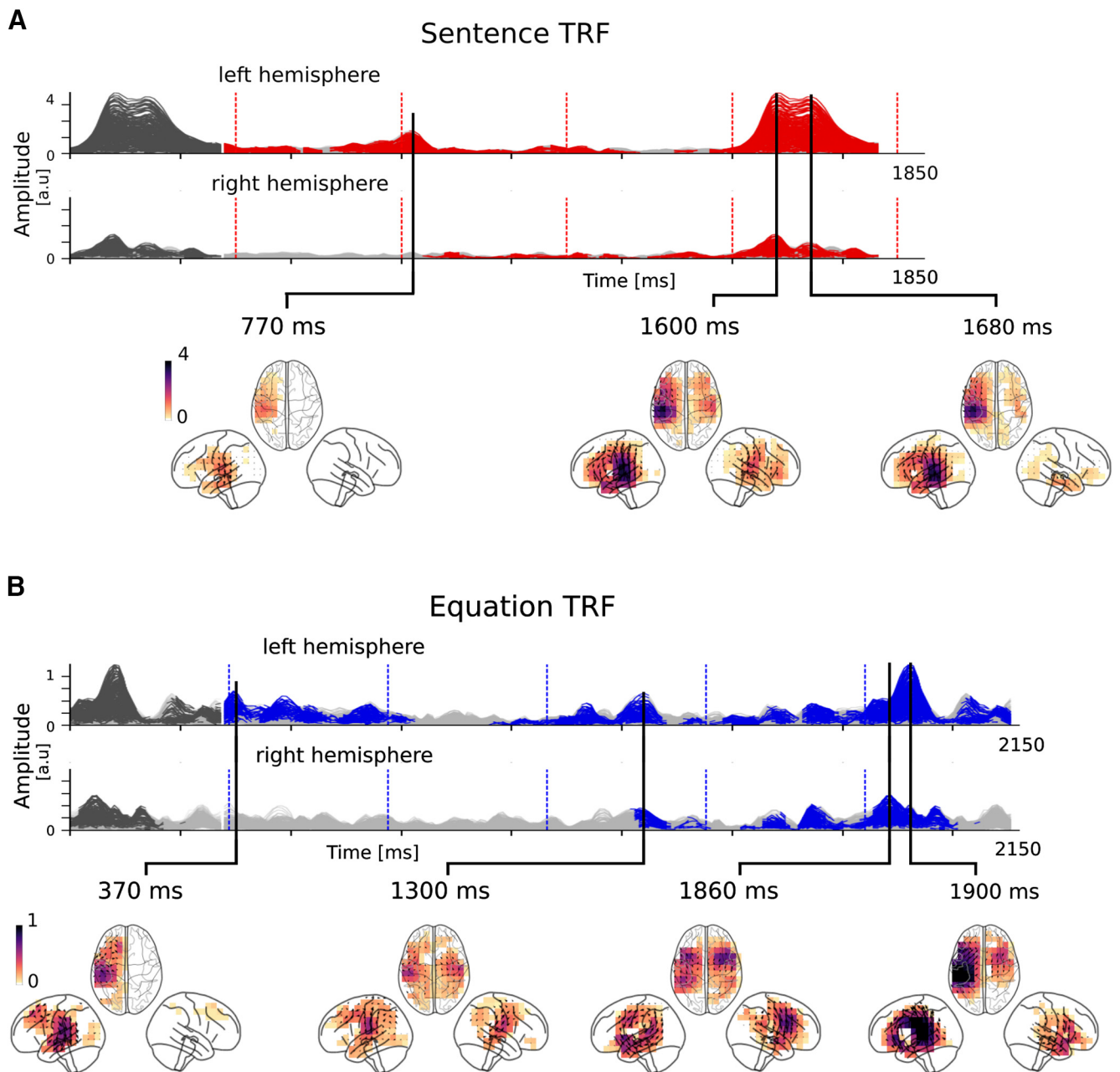


Figure 5. TRFs in the single speaker conditions. **A:** Sentence TRF, **B:** Equation TRF. Overlay plots of the amplitude of the TRF vectors for each voxel, averaged over subjects. For each TRF subfigure, the top axis shows vector amplitudes of voxels in the left hemisphere and the bottom axis correspondingly in the right hemisphere. Each trace is from the TRF of a single voxel; non-significant time points are shown in gray, while significant time points are shown in red (sentence TRF) or blue (equation TRF). The duration plotted corresponds to that of a sentence or equation, plus 350 ms; because of the fixed presentation rate, the first 350 ms (shown in gray) are identical to the last 350 ms. The large peak at the end (and beginning) of each TRF may either be ascribed to processing of the completion of the sentence/equation, or to the onset of the new sentence/equation, or both. Word and symbol onset times are shown in red and blue dashed lines, respectively; it can be seen that response contributions associated with them have been successfully regressed out. Volume source space distributions for several peaks in the TRF amplitudes are shown in the inlay plots, with black arrows denoting current directions (peaks automatically selected as local maxima of the TRFs). Although most of the TRF activity is dominated by neural currents in the left temporal lobe, the equation TRFs show more bilateral activation.

~250–300 ms after the offset of the operator. This is considerably late for decoding based on mere auditory responses to acoustic features of the operator.

Decoders at each voxel were also trained to differentiate attention to equations versus sentences, based on the dynamics of the response at that voxel during the first or last words (Fig. 7C). The prediction success (as measured by AUC) was significant for large areas in the single speaker conditions both for first

words ($t_{max} = 5.1$, $p_{min} < 0.001$) and last words ($t_{max} = 5.4$, $p_{min} < 0.001$) decoders. The AUC for first words decoders was significant for all regions in the left hemisphere except for areas in the inferior and middle temporal gyrus and all regions on the right hemisphere except the occipital lobe. The AUC for last words was significant for all regions in the left hemisphere and parts of frontal temporal lobes in the right hemisphere. For the cocktail party conditions the source-localized regions of significant

Cocktail Party

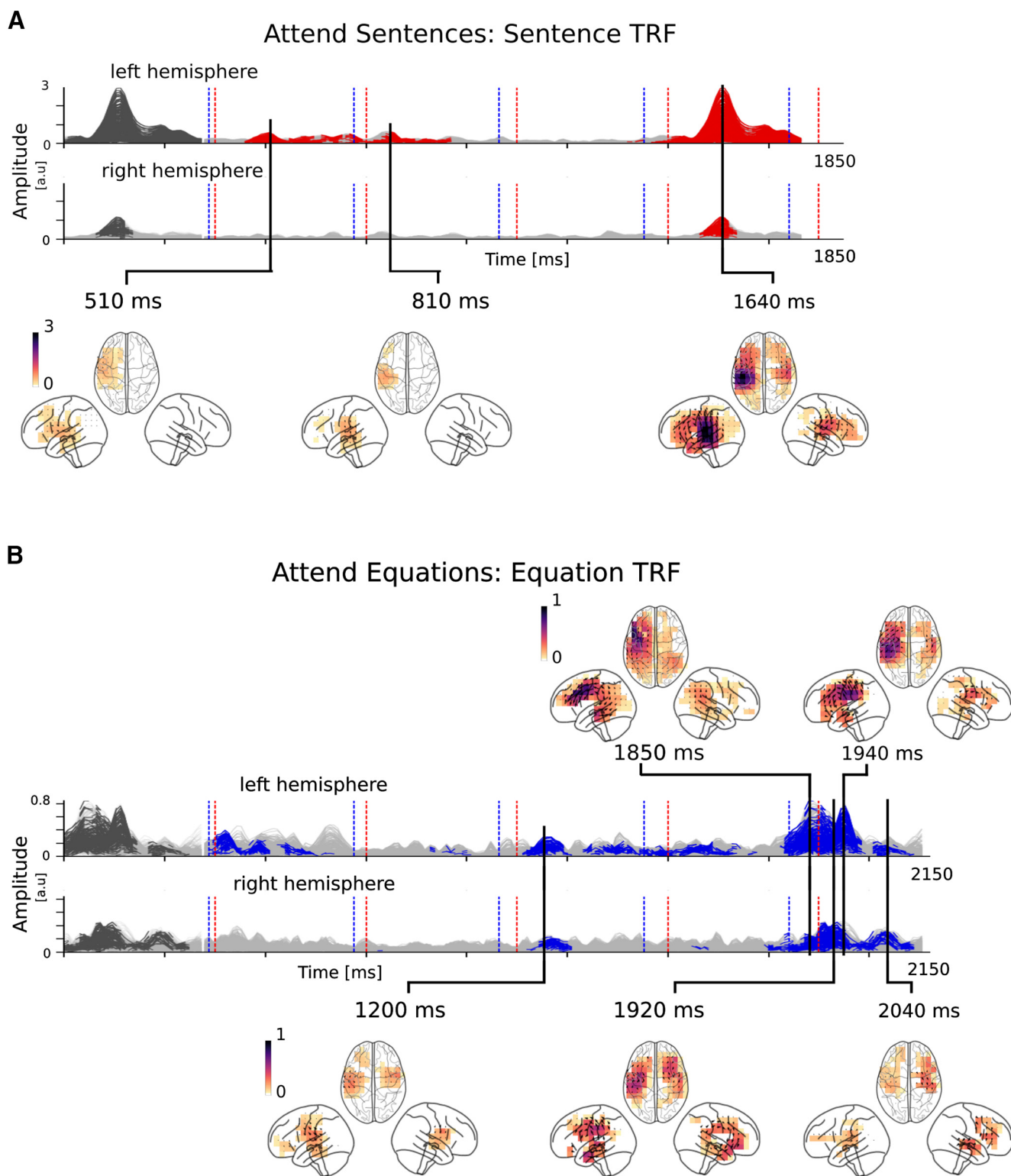


Figure 6. TRFs in the cocktail party conditions. **A:** Sentence TRF while attending to sentences, **B:** Equation TRF while attending to equations. Overlay plots of the TRF for each voxel averaged over subjects are shown as those in Figure 5. Word and symbol onset times are shown in red and blue dashed lines, respectively, and are marked in both sentence and equation TRFs since both stimuli were present in the cocktail party conditions; again, it can be seen that responses contributions associated with them have been successfully regressed out. Differences between sentence and equation TRFs arise at later time points, with sentence TRFs being predominantly near left temporal areas, while equation TRFs are in bilateral temporal, motor, and parietal regions.

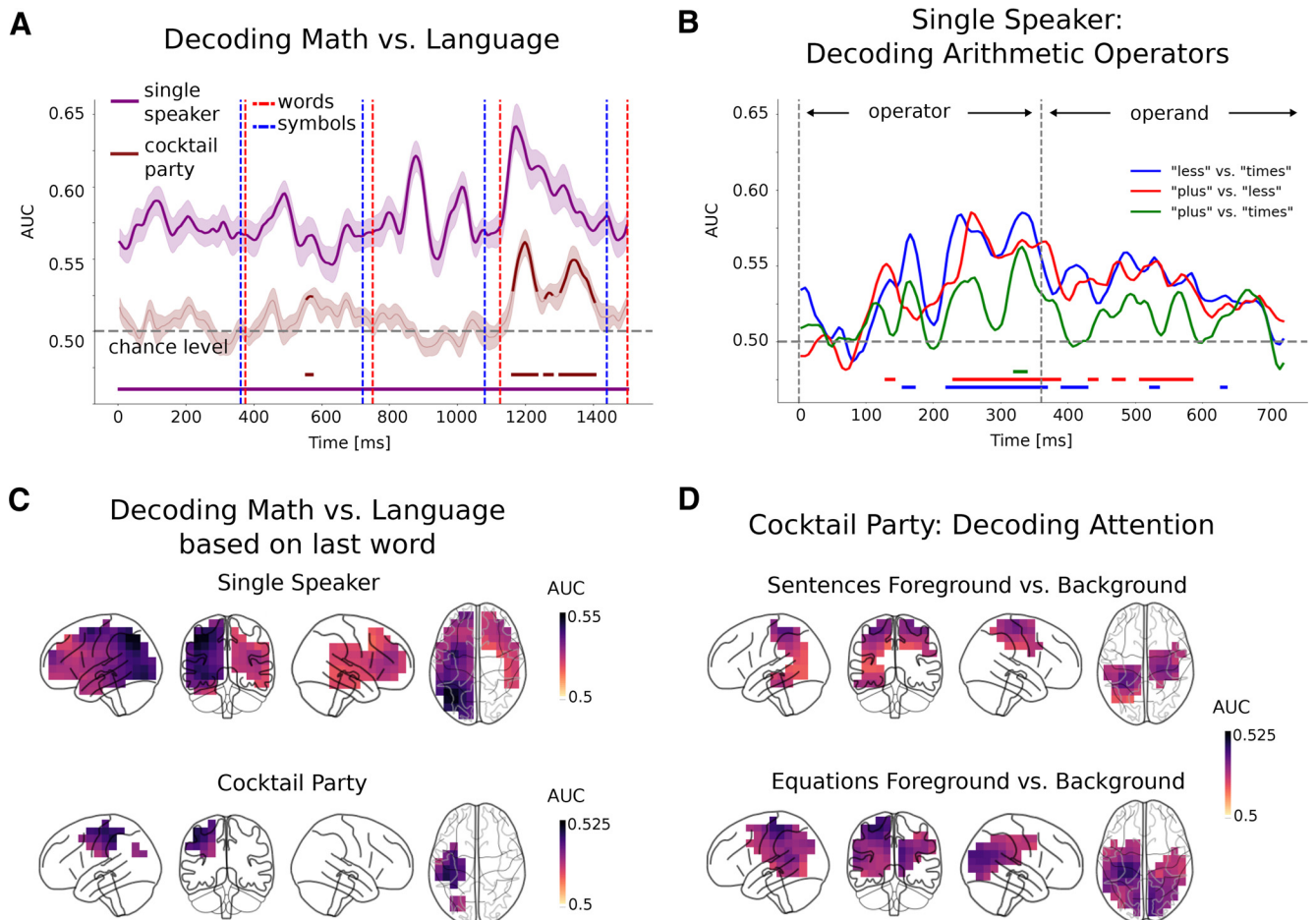


Figure 7. Decoding arithmetic and language processing. **A**, Performance of decoding attention condition (math vs language) at each time point using MEG sensors for single speaker (purple) and cocktail party (brown). Prediction success is measured by AUC, which is plotted (mean and SEM across subjects); time points where predictions are significantly above chance are marked by the horizontal bars at the bottom (every time point is significantly above chance for the single speaker case). The word and symbol onsets are also shown, and the decoding performance increases toward the end of the time window. **B**, Decoding arithmetic operators from sensor topographies. The time window of the operator and the subsequent operand was used for the three types of decoders. Time intervals where predictions are significantly above chance are marked by the colored horizontal bars at the bottom: all three operator comparisons could be significantly decoded. **C**, Decoding math versus language based on the last word. During the single speaker conditions, most of the brain is significant. However, for the cocktail party conditions, more focal significant decoding is seen in IPS and superior parietal areas. Decoding based on the first word resulted in similar results (data not shown). **D**, Decoding attention in the cocktail party conditions (AUC masked by significance across subjects). The sentence responses in foreground and background were decoded in left middle temporal and bilateral superior parietal areas. The equation responses in foreground and background were decoded in bilateral parietal areas.

prediction success were much more focal: the AUC was significant only in the IPS and superior parietal areas for both first words ($t_{max} = 5.3$, $p_{min} < 0.001$) and last words ($t_{max} = 4.3$, $p_{min} = 0.014$) decoders. These results suggest that the activity of voxels in left IPS and superior parietal areas is most useful for discriminating between attending to equations versus sentences. Finally, decoders at each voxel were also trained to decode the attention condition (foreground vs background) from the response to the entire sentence or equation (Fig. 7D). The AUC was significant in bilateral parietal areas for decoding whether arithmetic was in foreground versus background ($t_{max} = 5.2$, $p_{min} < 0.001$), consistent with areas involved in arithmetic processing. For decoding whether language was in foreground versus background, the AUC was significant ($t_{max} = 5.1$, $p_{min} = 0.002$) in left middle temporal areas, consistent with higher level language processing, and bilateral superior parietal areas, consistent with attention networks that are involved in discriminating auditory stimuli. Therefore, the decoding analysis is able to detect different cortical areas that may be involved in attention to language and arithmetic.

Discussion

We investigated the cortical locations and temporal dynamics of neural responses to spoken equations and sentences. Sentence responses consistently localized to left temporal areas. In contrast, equation responses consistently showed bilateral parietal activity, with other variations depending on analysis method (e.g., motor activity in TRFs). This may be because of different mechanisms involved in equation processing, although further investigation would be needed to support this claim. Cortical patterns consistent across different analysis methods (frequency domain, TRFs, decoders) are illustrated in schematic Figure 8.

Sentence and equation rate responses

As expected, MEG responses to acoustic features source-localized to the bilateral auditory cortex, and sentence rate responses source-localized to the left temporal cortex, consistent with speech and language areas (Friederici, 2002, 2011; Vandenberghe et al., 2002; Hickok and Poeppel, 2007; Binder et al., 2009), similar to prior isochronous speech studies (Sheng et al., 2019). In contrast, equation rate responses localized to left parietal,

temporal, and occipital areas. Arithmetic processing can activate IPS and parietal (Dehaene et al., 2003), angular gyrus (Göbel et al., 2001), temporal (Tang et al., 2006), and even occipital areas (Maruyama et al., 2012; Harvey and Dumoulin, 2017), perhaps because of internal visualization (Zago et al., 2001). Equation responses also localized to the right temporal and parietal areas in cocktail party conditions, confirming that arithmetic processing is more bilateral than language processing (Dehaene and Cohen, 1997; Amalric and Dehaene, 2017, 2019). Critically, ANOVA analysis indicated that attended equation and sentence responses are significantly different. Unexpectedly, significant neural responses at the unattended sentence and equation rates were found in smaller temporal (consistent with language processing) and parietal (consistent with arithmetic processing) areas, respectively. Some subjects may have been unable to sustain attention to the instructed stream for the entirety of this diotic stimulus and so briefly switched their attentional focus.

Left hemispheric dominance of equation responses

Equation responses were left dominant in both single speaker and cocktail party conditions. This could reflect left-lateralized language processing since equations were presented using speech. However, arithmetic processing may also show left dominance (Pinel and Dehaene, 2010), perhaps because of precise calculations (Dehaene, 1999; Pica et al., 2004) or arithmetic fact retrieval (Dehaene et al., 2003; Grabner et al., 2009). These fast-paced stimuli required rapid calculations, and may have resulted in increased reliance on rote memory, which activates left hemispheric areas (Campbell and Austin, 2002). Specific strategies employed for calculation may also result in left lateralization; multiplication of small numbers is often performed using rote memory (Delazer et al., 1999; Ischebeck et al., 2006; Fehr et al., 2007), while subtraction is less commonly performed using memory and shows more bilateral activation (Schmithorst and Brown, 2004; Prado et al., 2011). Addition may recruit both these networks, depending on specific strategies used by individuals (Arsalidou and Taylor, 2011). We found no significant differences in equation responses when separated by operation type, perhaps because of individual variation in procedural calculation or retrieval strategies within the same operation (Tschentscher and Hauk, 2014). However, operation types were successfully decoded from the overall MEG signals (Fig. 8B), consistent with prior work (Pinheiro-Chagas et al., 2019), although not as robustly as decoding stimulus type or attention. Overall, left-hemispheric dominance of equation responses is supported by a combination of speech processing, precise calculations, and arithmetic fact retrieval.

Cortical correlates of behavioral performance

Neural responses to sentence, equation, word, and symbol rates were correlated with performance in detecting deviants, consistent with language-only isochronous studies (Ding et al., 2017).

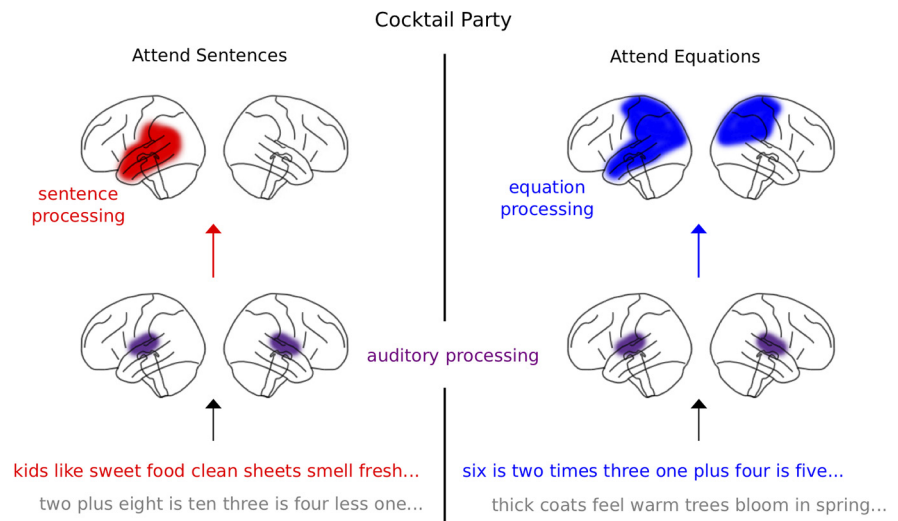


Figure 8. Schematic of cortical processing of sentences and equations. A schematic representation of sentence and equation processing is shown. Exemplars of both foreground and background of stimuli are shown at the bottom. The areas that were most consistent across all analysis methods (frequency domain, TRFs and decoders) are shown.

Sentence responses correlated with behavior in language areas, such as left auditory cortex, superior and middle temporal lobe and angular gyrus (Price, 2000; Binder et al., 2009; Karuza et al., 2013) in both single speaker and cocktail party conditions. In contrast, equation responses correlated with behavior in cocktail party conditions in posterior parietal areas, which are known to predict competence and performance in numerical tasks (Grabner et al., 2007; Lin et al., 2012, 2019; Lasne et al., 2019). The lack of significant behavioral correlations for equation responses in single speaker conditions may be because of several subjects performing at ceiling; equations had a restricted set of only 14 unique symbols, and the presence of the “is” symbol in every equation might be structurally useful in tracking equation boundaries. Unexpectedly, behavioral correlations were also found for background symbol rate responses in parietal and occipital areas (and for background word rate responses in a small parietal region). Some studies show that acoustic features of background speech may also be tracked (Fiedler et al., 2019; Brodbeck et al., 2020a). Since background word and symbol rates were present in the stimulus acoustics, increased effort or attention could enhance both behavioral performance and auditory responses at these rates. Representations of the background could enhance attentional selectivity in challenging cocktail party environments as suggested by Fiedler et al. (2019). However, note that sentence and equation responses were only correlated with behavior when attended. Overall, behavioral correlations in temporal and parietal regions suggest that these responses may reflect improved comprehension because of neural chunking of speech structures or successful calculations (Blanco-Elorrieta et al., 2020; Chen et al., 2020; Jin et al., 2020; Kaufeld et al., 2020; Teng et al., 2020).

Dynamics of arithmetic and language processing

TRFs for equation and sentence onsets were jointly estimated along with speech envelopes and word/equation onsets to regress out auditory responses, analogous to prior work with linguistic and auditory TRFs (Brodbeck et al., 2018a,b; Broderick et al., 2018). Isochronous speech studies have found slow rhythmic activity (Zhang and Ding, 2017), which did not appear in our

TRFs, perhaps because of implicit high-pass filtering (boosting favors sparse TRFs). Instead, we found large TRF peaks at sentence/equation boundaries. Prior studies have found late evoked responses specific to numbers and equations (Avancini et al., 2015). Large peaks appeared in both sentence and equation TRFs 410–600 ms after the onset of the last word/symbol and may reflect processing of the completion of the sentence/equation. Sentence TRF peaks localized to left temporal areas, while equation TRF peaks showed activity in bilateral parietal and temporal areas involved in numerical processing (Abd Hamid et al., 2011; Amalric and Dehaene, 2017), and motor areas, perhaps reflecting procedural calculation strategies (Tschemtscher and Hauk, 2014). The peak latencies were similar to prior arithmetic ERP studies (Iguchi and Hashimoto, 2000; Iijima and Nishitani, 2017). These sentence and equation TRF peaks may reflect several mechanisms; both shared (language processing, decision-making), and separate (semantic vs arithmetic processing), and further work is needed to disentangle these mechanisms. Finally, the cortical patterns of TRF peaks showed more differences in the cocktail party than the single speaker conditions, suggesting that selective attention focuses the underlying cortical networks.

Decoding equation and sentence processing

Numbers and arithmetic operations have been previously decoded from cortical responses (Eger et al., 2009; Pinheiro-Chagas et al., 2019). In this study, the attended stimulus type (sentences or equations) was reliably decoded in single speaker conditions in several cortical regions, perhaps because of highly correlated responses across cortex for this task. In contrast, decoding accuracy during cocktail party conditions was significant in left IPS and superior parietal areas, suggesting that these regions are most important for discriminating between arithmetic and language processing. Both the attend-equations and the attend-sentences states could be decoded from bilateral superior parietal areas, perhaps because of general attentional networks in fronto-parietal areas, or attentional segregation of foreground and background speech based on pitch or gender (Hill and Miller, 2010; Kristensen et al., 2013). Additionally, decoding the attend-equations state was significant in bilateral parietal areas, consistent with arithmetic processing, while decoding the attend-sentences state was significant in the left middle temporal lobe, consistent with language processing (Hickok and Poeppel, 2007). Overall, MEG responses contain enough information to decode arithmetic versus language processing, selective attention, and arithmetic operations.

The cocktail party paradigm highlights distinct cortical processes

Differences in source-localized sentence and equation responses were more prominent in the cocktail party than in the single speaker conditions for all analyses (frequency domain, TRFs and decoders). Responses to both stimuli presented simultaneously may have helped control for common auditory and preattentive responses. Abd Hamid et al. (2011) found fMRI activation in broader areas for spoken arithmetic with a noisy background than in quiet, perhaps because of increased effort. However, in our case, the background stimulus was not white noise, but rather meaningful non-mathematical speech. Our TRF analysis, which regresses out responses to background speech, as well as the selective attention task itself, may highlight specific cortical processes that best separate the arithmetic and language stimuli.

In summary, neural processing of spoken equations and sentences involves both overlapping and non-overlapping

cortical networks. Behavioral correlations suggest that these neural responses may reflect improved comprehension and/or correct arithmetic calculations. Selective attention for equations focuses activity in temporal, parietal occipital and motor areas, and for sentences in temporal and superior parietal areas. This cocktail party paradigm is well suited to highlight the cortical networks underlying the processing of spoken arithmetic and language.

References

- Abd Hamid AI, Yusoff AN, Mukari SMS, Mohamad M (2011) Brain activation during addition and subtraction tasks in-noise and in-quiet. *Malays J Med Sci* 18:3–15.
- Amalric M, Dehaene S (2016) Origins of the brain networks for advanced mathematics in expert mathematicians. *Proc Natl Acad Sci USA* 113:4909–4917.
- Amalric M, Dehaene S (2017) Cortical circuits for mathematical knowledge: evidence for a major subdivision within the brain's semantic networks. *Philos Trans R Soc B Biol Sci* 373:20160515.
- Amalric M, Dehaene S (2019) A distinct cortical network for mathematical knowledge in the human brain. *Neuroimage* 189:19–31.
- Arsalidou M, Taylor MJ (2011) Is $2+2=4$? Meta-analyses of brain areas needed for numbers and calculations. *Neuroimage* 54:2382–2393.
- Avancini C, Soltész F, Szűcs D (2015) Separating stages of arithmetic verification: an ERP study with a novel paradigm. *Neuropsychologia* 75:322–329.
- 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.
- Baldo JV, Dronkers NF (2007) Neural correlates of arithmetic and language comprehension: a common substrate? *Neuropsychologia* 45:229–235.
- Bemis DK, Pyllkänen L (2013) Basic linguistic composition recruits the left anterior temporal lobe and left angular gyrus during both listening and reading. *Cereb Cortex N Cortex* 23:1859–1873.
- Binder JR, Desai RH, Graves WW, Conant LL (2009) Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cereb Cortex* 19:2767–2796.
- Blanco-Elorrieta E, Ding N, Pyllkänen L, Poeppel D (2020) Understanding requires tracking: noise and knowledge interact in bilingual comprehension. *J Cogn Neurosci* 32:1975–1983.
- Boersma P, Weenick D (2018) Praat: doing phonetics by computer [computer program]. Version 6.0.43. Available at <http://www.praat.org>. Accessed September 8, 2018.
- Brodbeck C, Hong LE, Simon JZ (2018a) Rapid transformation from auditory to linguistic representations of continuous speech. *Curr Biol* 28:3976–3983.e5.
- Brodbeck C, Presacco A, Simon JZ (2018b) Neural source dynamics of brain responses to continuous stimuli: speech processing from acoustics to comprehension. *Neuroimage* 172:162–174.
- Brodbeck C, Jiao A, Hong LE, Simon JZ (2020a) Neural speech restoration at the cocktail party: auditory cortex recovers masked speech of both attended and ignored speakers. *PLoS Biol* 18:e3000883.
- Brodbeck C, Brooks TL, Das P, Reddigari S, Kulasingham JP (2020b) christi-anbrodbeck/Eelbrain: 0.33. Zenodo. Available at <https://zenodo.org/record/4060224>. Accessed October 9, 2020.
- Broderick MP, Anderson AJ, Di Liberto GM, Crosse MJ, Lalor EC (2018) Electrophysiological correlates of semantic dissimilarity reflect the comprehension of natural, narrative speech. *Curr Biol* 28:803–809.e3.
- Campbell JID, Austin S (2002) Effects of response time deadlines on adults' strategy choices for simple addition. *Mem Cognit* 30:988–994.
- Castaldi E, Piazza M, Dehaene S, Vignaud A, Eger E (2019) Attentional amplification of neural codes for number independent of other quantities along the dorsal visual stream. *Elife* 8:e45160.
- Chen Y, Jin P, Ding N (2020) The influence of linguistic information on cortical tracking of words. *Neuropsychologia* 148:107640.
- Dastjerdi M, Ozker M, Foster BL, Rangarajan V, Parvizi J (2013) Numerical processing in the human parietal cortex during experimental and natural conditions. *Nat Commun* 4:2528.
- David SV, Mesgarani N, Shamma SA (2007) Estimating sparse spectro-temporal receptive fields with natural stimuli. *Network* 18:191–212.

- de Cheveigné A, Simon JZ (2007) Denoising based on time-shift PCA. *J Neurosci Methods* 165:297–305.
- de Cheveigné A, Simon JZ (2008) Sensor noise suppression. *J Neurosci Methods* 168:195–202.
- Dehaene S, Cohen L (1997) Cerebral pathways for calculation: double dissociation between rote verbal and quantitative knowledge of arithmetic. *Cortex* 33:219–250.
- Dehaene S, Spelke E, Pinel P, Stanescu R, Tsivkin S (1999) Sources of mathematical thinking: behavioral and brain-imaging evidence. *Science* 284:970–974.
- Dehaene S, Piazza M, Pinel P, Cohen L (2003) Three parietal circuits for number processing. *Cogn Neuropsychol* 20:487–506.
- Dehaene S, Molko N, Cohen L, Wilson AJ (2004) Arithmetic and the brain. *Curr Opin Neurobiol* 14:218–224.
- Delazer M, Girelli L, Semenza C, Denes G (1999) Numerical skills and aphasia. *J Int Neuropsychol Soc* 5:213–221.
- Di Liberto GM, O'Sullivan JA, Lalor EC (2015) Low-frequency cortical entrainment to speech reflects phoneme-level processing. *Curr Biol* 25:2457–2465.
- Ding N, Simon JZ (2012) Emergence of neural encoding of auditory objects while listening to competing speakers. *Proc Natl Acad Sci USA* 109:11854–11859.
- Ding N, Melloni L, Zhang H, Tian X, Poeppel D (2016) Cortical tracking of hierarchical linguistic structures in connected speech. *Nat Neurosci* 19:158–164.
- Ding N, Melloni L, Yang A, Wang Y, Zhang W, Poeppel D (2017) Characterizing neural entrainment to hierarchical linguistic units using electroencephalography (EEG). *Front Hum Neurosci* 11:481.
- Ding N, Pan X, Luo C, Su N, Zhang W, Zhang J (2018) Attention is required for knowledge-based sequential grouping: insights from the integration of syllables into words. *J Neurosci* 38:1178–1188.
- Eger E, Sterzer P, Russ MO, Giraud AL, Kleinschmidt A (2003) A supramodal number representation in human intraparietal cortex. *Neuron* 37:719–725.
- Eger E, Michel V, Thirion B, Amadon A, Dehaene S, Kleinschmidt A (2009) Deciphering cortical number coding from human brain activity patterns. *Curr Biol* 19:1608–1615.
- Fehr T, Code C, Herrmann M (2007) Common brain regions underlying different arithmetic operations as revealed by conjunct fMRI–BOLD activation. *Brain Res* 1172:93–102.
- Fiedler L, Wöstmann M, Herbst SK, Obleser J (2019) Late cortical tracking of ignored speech facilitates neural selectivity in acoustically challenging conditions. *Neuroimage* 186:33–42.
- Fischl B (2012) FreeSurfer. *Neuroimage* 62:774–781.
- Friederici AD (2002) Towards a neural basis of auditory sentence processing. *Trends Cogn Sci* 6:78–84.
- Friederici AD (2011) The brain basis of language processing: from structure to function. *Physiol Rev* 91:1357–1392.
- Gelman R, Butterworth B (2005) Number and language: how are they related? *Trends Cogn Sci* 9:6–10.
- Göbel S, Walsh V, Rushworth MFS (2001) The mental number line and the human angular gyrus. *Neuroimage* 14:1278–1289.
- Grabner RH, Ansari D, Reishofer G, Stern E, Ebner F, Neuper C (2007) Individual differences in mathematical competence predict parietal brain activation during mental calculation. *Neuroimage* 38:346–356.
- Grabner RH, Ansari D, Koschutnig K, Reishofer G, Ebner F, Neuper C (2009) To retrieve or to calculate? Left angular gyrus mediates the retrieval of arithmetic facts during problem solving. *Neuropsychologia* 47:604–608.
- Gramfort A (2013) MEG and EEG data analysis with MNE-Python. *Front Neurosci* 7:267.
- 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.
- Hämäläinen MS, Ilmoniemi RJ (1994) Interpreting magnetic fields of the brain: minimum norm estimates. *Med Biol Eng Comput* 32:35–42.
- Harvey BM, Dumoulin SO (2017) A network of topographic numerosity maps in human association cortex. *Nat Hum Behav* 1:0036.
- Harvey BM, Klein BP, Petridou N, Dumoulin SO (2013) Topographic representation of numerosity in the human parietal cortex. *Science* 341:1123–1126.
- Hickok G, Poeppel D (2007) The cortical organization of speech processing. *Nat Rev Neurosci* 8:393–402.
- Hill KT, Miller LM (2010) Auditory attentional control and selection during cocktail party listening. *Cereb Cortex* 20:583–590.
- Houdé O, Tzourio-Mazoyer N (2003) Neural foundations of logical and mathematical cognition. *Nat Rev Neurosci* 4:507–514.
- Iguchi Y, Hashimoto I (2000) Sequential information processing during a mental arithmetic is reflected in the time course of event-related brain potentials. *Clin Neurophysiol* 111:204–213.
- Iijima M, Nishitani N (2017) Cortical dynamics during simple calculation processes: a magnetoencephalography study. *Clin Neurophysiol Pract* 2:54–61.
- Ischebeck A, Zamarian L, Siedentopf C, Koppelstätter F, Benke T, Felber S, Delazer M (2006) How specifically do we learn? Imaging the learning of multiplication and subtraction. *Neuroimage* 30:1365–1375.
- Jasinski EC, Coch D (2012) ERPs across arithmetic operations in a delayed answer verification task. *Psychophysiology* 49:943–958.
- Jin P, Lu Y, Ding N (2020) Low-frequency neural activity reflects rule-based chunking during speech listening. *Elife* 9:e55613.
- Karuza EA, Newport EL, Aslin RN, Starling SJ, Tivarus ME, Bavelier D (2013) The neural correlates of statistical learning in a word segmentation task: an fMRI study. *Brain Lang* 127:46–54.
- Kaufeld G, Bosker HR, ten OS, Alday PM, Meyer AS, Martin AE (2020) Linguistic structure and meaning organize neural oscillations into a content-specific hierarchy. *J Neurosci* 40:9467–9475.
- Koelsch S, Kasper E, Sammler D, Schulze K, Gunter T, Friederici AD (2004) Music, language and meaning: brain signatures of semantic processing. *Nat Neurosci* 7:302–307.
- Kou H, Iwaki S (2007) Modulation of neural activities by the complexity of mental arithmetic: an MEG study. *Int Congr Ser* 1300:539–542.
- Kristensen LB, Wang L, Petersson KM, Hagoort P (2013) The interface between language and attention: prosodic focus marking recruits a general attention network in spoken language comprehension. *Cereb Cortex* 23:1836–1848.
- Ku Y, Hong B, Gao X, Gao S (2010) Spectra-temporal patterns underlying mental addition: an ERP and ERD/ERS study. *Neurosci Lett* 472:5–10.
- Lalor EC, Foxe JJ (2010) Neural responses to uninterrupted natural speech can be extracted with precise temporal resolution. *Eur J Neurosci* 31:189–193.
- Lasne G, Piazza M, Dehaene S, Kleinschmidt A, Eger E (2019) Discriminability of numerosity-evoked fMRI activity patterns in human intra-parietal cortex reflects behavioral numerical acuity. *Cortex* 114:90–101.
- Lau EF, Phillips C, Poeppel D (2008) A cortical network for semantics: (de)constructing the N400. *Nat Rev Neurosci* 9:920–933.
- Lin JFL, Imada T, Kuhl PK (2012) Mental addition in bilinguals: an fMRI study of task-related and performance-related activation. *Cereb Cortex* 22:1851–1861.
- Lin JFL, Imada T, Kuhl PK (2019) Neuroplasticity, bilingualism, and mental mathematics: a behavior-MEG study. *Brain Cogn* 134:122–134.
- Luo C, Ding N (2020) Cortical encoding of acoustic and linguistic rhythms in spoken narratives. *Elife* 9:e60433.
- Mardia KV (1975) Assessment of multinormality and the robustness of Hotelling's T2 test. *J R Stat Soc Ser C Appl Stat* 24:163–171.
- Maruyama M, Pallier C, Jobert A, Sigman M, Dehaene S (2012) The cortical representation of simple mathematical expressions. *Neuroimage* 61:1444–1460.
- Menon V, Rivera SM, White CD, Glover GH, Reiss AL (2000) Dissociating prefrontal and parietal cortex activation during arithmetic processing. *Neuroimage* 12:357–365.
- Monti MM, Parsons LM, Osherson DN (2012) Thought beyond language: neural dissociation of algebra and natural language. *Psychol Sci* 23:914–922.
- Nichols TE, Holmes AP (2002) Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Hum Brain Mapp* 15:1–25.
- Obleser J, Lahiri A, Eulitz C (2003) Auditory-evoked magnetic field codes place of articulation in timing and topography around 100 milliseconds post syllable onset. *Neuroimage* 20:1839–1847.
- Obleser J, Lahiri A, Eulitz C (2004) Magnetic brain response mirrors extraction of phonological features from spoken vowels. *J Cogn Neurosci* 16:31–39.

- Park J, Hebrank A, Polk TA, Park DC (2012) Neural dissociation of number from letter recognition and its relationship to parietal numerical processing. *J Cogn Neurosci* 24:39–50.
- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D (2011) Scikit-learn: machine learning in Python. *J Mach Learn Res* 12:2825–2830.
- Pica P, Lemer C, Izard V, Dehaene S (2004) Exact and approximate arithmetic in an Amazonian indigene group. *Science* 306:499–503.
- Pinel P, Dehaene S (2010) Beyond hemispheric dominance: brain regions underlying the joint lateralization of language and arithmetic to the left hemisphere. *J Cogn Neurosci* 22:48–66.
- Pinel P, Dehaene S, Rivière D, LeBihan D (2001) Modulation of parietal activation by semantic distance in a number comparison task. *Neuroimage* 14:1013–1026.
- Pinheiro-Chagas P, Piazza M, Dehaene S (2019) Decoding the processing stages of mental arithmetic with magnetoencephalography. *Cortex* 114:124–139.
- Prado J, Mutreja R, Zhang H, Mehta R, Desroches AS, Minas JE, Booth JR (2011) Distinct representations of subtraction and multiplication in the neural systems for numerosity and language. *Hum Brain Mapp* 32:1932–1947.
- Price CJ (2000) The anatomy of language: contributions from functional neuroimaging. *J Anat* 197 Pt 3:335–359.
- Schmithorst VJ, Brown RD (2004) Empirical validation of the triple-code model of numerical processing for complex math operations using functional MRI and group independent component analysis of the mental addition and subtraction of fractions. *Neuroimage* 22:1414–1420.
- Sheng J, Zheng L, Lyu B, Cen Z, Qin L, Tan LH, Huang MX, Ding N, Gao JH (2019) The cortical maps of hierarchical linguistic structures during speech perception. *Cereb Cortex* 29:3232–3240.
- Simon JZ, Wang Y (2005) Fully complex magnetoencephalography. *J Neurosci Methods* 149:64–73.
- Smith SM, Nichols TE (2009) Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. *Neuroimage* 44:83–98.
- Spelke ES, Tsivkin S (2001) Language and number: a bilingual training study. *Cognition* 78:45–88.
- Tang Y, Zhang W, Chen K, Feng S, Ji Y, Shen J, Reiman EM, Liu Y (2006) Arithmetic processing in the brain shaped by cultures. *Proc Natl Acad Sci USA* 103:10775–10780.
- Teng X, Ma M, Yang J, Blohm S, Cai Q, Tian X (2020) Constrained structure of ancient Chinese poetry facilitates speech content grouping. *Curr Biol* 30:1299–1305.e7.
- Tschemtscher N, Hauk O (2014) How are things adding up? Neural differences between arithmetic operations are due to general problem solving strategies. *Neuroimage* 92:369–380.
- Vandenberghe R, Nobre AC, Price CJ (2002) The response of left temporal cortex to sentences. *J Cogn Neurosci* 14:550–560.
- Varley RA, Klessinger NJC, Romanowski CAJ, Siegal M (2005) Agrammatic but numerate. *Proc Natl Acad Sci USA* 102:3519–3524.
- Venkatraman V, Siong SC, Chee MWL, Ansari D (2006) Effect of language switching on arithmetic: a bilingual fMRI study. *J Cogn Neurosci* 18:64–74.
- Zago L, Pesenti M, Mellet E, Crivello F, Mazoyer B, Tzourio-Mazoyer N (2001) Neural correlates of simple and complex mental calculation. *Neuroimage* 13:314–327.
- Zarnhofer S, Braunstein V, Ebner F, Koschutnig K, Neuper C, Reishofer G, Ischebeck A (2012) The Influence of verbalization on the pattern of cortical activation during mental arithmetic. *Behav Brain Funct* 8:13.
- Zhang W, Ding N (2017) Time-domain analysis of neural tracking of hierarchical linguistic structures. *Neuroimage* 146:333–340.