Title

Event related spectral perturbations of gesture congruity: Visuospatial resources are recruited for multimodal discourse comprehension.

Authors

Momsen, Jacob

Gordon, Jared

Wu, Ying

et al.

Publication Date

2021-05-01

DOI

10.1016/j.bandl.2021.104916

Peer reviewed

HHS Public Access

Author manuscript

Brain Lang. Author manuscript; available in PMC 2024 August 02.

Published in final edited form as:

Brain Lang. 2021 May ; 216: 104916. doi:10.1016/j.bandl.2021.104916.

Event related spectral perturbations of gesture congruity: Visuospatial resources are recruited for multimodal discourse comprehension

Jacob Momsen[1], **Jared Gordon**[2], **Ying Choon Wu**[3], **Seana Coulson**[1,2]

[(1)]Joint Doctoral Program Language and Communicative Disorders, San Diego State University and UC San Diego

[(2)]Cognitive Science Department, UC San Diego

[(3)]Swartz Center for Computational Neuroscience, UC San Diego

Abstract

Here we examine the role of visuospatial working memory (WM) during the comprehension of multimodal discourse with co-speech iconic gestures. EEG was recorded as healthy adults encoded either a sequence of one (low load) or four (high load) dot locations on a grid and rehearsed them until a free recall response was collected later in the trial. During the rehearsal period of the WM task, participants observed videos of a speaker describing objects in which half of the trials included semantically related co-speech gestures (congruent), and the other half included semantically unrelated gestures (incongruent). Discourse processing was indexed by oscillatory EEG activity in the alpha and beta bands during the videos. Across all participants, effects of speech and gesture incongruity were more evident in low load trials than in high load trials. Effects were also modulated by individual differences in visuospatial WM capacity. These data suggest visuospatial WM resources are recruited in the comprehension of multimodal discourse.

Keywords

alpha suppression; beta suppression; iconic gestures; individual differences; representational gestures; speech-gesture integration; working memory

Corresponding author: Seana Coulson, Cognitive Science, 9500 Gilman Dr., La Jolla, CA 92093, USA.

Author Contributions

Jacob Momsen: Data curation, Investigation, Project administration, Visualization, Writing – Original draft preparation. **Jared Gordon:** Investigation, Software, Visualization, Writing – Editing and Review. **Ying Choon Wu:** Conceptualization, Methodology, Resources, Software, Writing – Editing and Review. **Seana Coulson:** Conceptualization, Funding acquisition, Methodology, Supervision, Writing – Original draft preparation.

Declaration of interests: None.

1 Introduction

Communication occurring in natural environments is often accompanied by nonverbal cues that convey important information beyond language (Hall et al, 2019). One type of cue, co-speech gestures, occurs spontaneously alongside speech and can impact the meaning of a speaker's message (McNeill et al, 1994). During discourse comprehension, this kind of non-linguistic communicative information is rapidly integrated with language (Özyürek et al, 2007; Hagoort, 2004). Thus, it is perhaps unsurprising that gestures in multimodal discourse can exert a real-time influence on neural indices of speech processing (Wu & Coulson, 2010; Holle & Gunter, 2007). In fact, some evidence suggests that this integration happens automatically, and that gesture interpretation is an obligatory component of discourse comprehension (Kelly et al, 2010). However, relatively little is known about the cognitive and neural resources that underlie the comprehension of co-speech gestures.

The current study focuses on iconic gestures, i.e. gestures characterized by having forms that resemble the concepts they represent. Unlike emblems such as the "OK" sign, iconic gestures are rarely produced without speech, and often rely on the accompanying speech to provide a meaningful context for their interpretation (Willems & Hagoort, 2007). Iconic gestures tend to depict visual or spatial properties of objects or actions, and there is evidence for similarities in the conceptual processing of iconic gestures and of pictures (Wu & Coulson, 2011). Previous research has shown that iconic gestures are more likely to be spontaneously produced when people attempt to describe visually primed concepts, and that producing iconic gestures can facilitate spatial working memory (WM) processes (Masson-Carro et al, 2017; Morsella & Krauss, 2004). Because of the relationship between iconic gestures, spatial imagery, and spatial cognitive processing, it might be the case that their comprehension involves brain mechanisms supporting analog visuospatial representations and working memory.

In accordance with the *visuospatial resources hypothesis*, previous behavioral research has suggested that visuospatial WM plays a key role in speech-gesture integration (see Coulson & Wu, 2014 for a review). For example, Wu and Coulson, (2014) had participants watch videos of a speaker describing objects and actions while producing gestures either related or unrelated to the concurrent speech. Following the video, participants judged whether or not a picture probe was related to the preceding discourse; response times on this picture classification task were linearly related to individual measures of visuospatial WM ability (Wu & Coulson, 2014). In a follow-up experiment, the addition of a concurrent visuospatial WM task disproportionately impacted performance on the same picture classification task, suggesting that visuospatial WM was especially important for interpreting discourse with co-speech iconic gestures.

Alternatively, and in line with research illustrating a tight link between gesture and language processing networks in the brain (Willems et al, 2007; Kircher et al, 2009; Özyürek, 2014; Zhao et al, 2018), iconic gestures may instead be directly mapped onto linguistic representations during online processing, which may consequently recruit verbal WM resources during their integration with speech. For example, the posterior middle temporal gyrus (pMTG), a central component of the language network, is thought to play

an important role for processing semantic features of iconic, emblematic, and metaphoric gestures alike (Green et al, 2009; Willems et al, 2009; Andric et al, 2013; Kircher et al, 2009). The current study tests the *visuospatial resources hypothesis* by using cortical measures of discourse comprehension to index online speech-gesture integration while participants' visuospatial WM resources are compromised by a second task. The logic of the dual-task design is that if gesture comprehension requires neural resources that overlap with the activation and maintenance of visuospatial representations, manipulations of visuospatial WM load should modulate electroencephalographic (EEG) indices of speech-gesture comprehension.

1.1 Neural Oscillations and Multimodal Discourse Processing

The transformation of electroencephalogram (EEG) or magnetoencephalogram (MEG) data into event related spectral perturbations (ERSPs) is a valuable method for observing stimulus-related brain activity, including phase locked activity evident in event related potentials (ERPs), as well as non-phase locked activity (see Pfurtscheller & Silva, 1999 for a review). It is well established that both alpha- and beta-band activity are highly associated with human action and gesture observation (Hari et al, 1998; Babiloni, 2002). Related research generally reports power reduction when subjects expend resources analyzing spatial or motoric features of others' actions. These studies suggest that increased perceptual and cognitive efforts to analyze gestural information might modulate low frequency activity related to sensorimotor representations of relevant dynamic body features (Avanzini et al, 2012; Quandt et al, 2012). Suppression of low frequency Rolandic activity generated by sensorimotor cortices, often described as mu rhythms, has also been studied in terms of action preparation, execution, and perception processes (see Hari, 2006 for review). This literature demonstrates that alpha power suppression over central electrode sites occurs during both the execution of motor acts and during the passive observation of another's hand or body movement (Lepage & Theoret, 2006; Iacoboni & Dapretto, 2006; Casile et al, 2011). Despite this documented relationship between action observation and motor network activation, there is no current evidence for this mechanism playing a role in the comprehension of representational body movements.

Previous MEG and EEG studies of representational gestures have identified modulations of alpha and beta band activity as indices of co-speech gesture processing (Drijvers et al, 2018; Drijvers et al, 2019; He, et al, 2015; He, et al, 2018). Using MEG to compare oscillatory activity to videos with gestures that were either semantically related or unrelated to concurrently spoken verbs, Drijvers and colleagues found greater alpha and beta suppression for videos containing incongruent gestures (Drijvers et al., 2018; 2019). Drijvers and colleagues interpreted their data in terms of the *functional inhibition hypothesis*: the proposal that alpha synchronization reflects the inhibition of task-irrelevant cortical activity, and, that alpha suppression reflects cortical disinhibition and increased neural engagement with a current task (Klimesch, 2007; Jensen & Mazaheri, 2010). Consequently, alpha and beta band suppression over the left inferior frontal gyrus (LIFG) was interpreted as increased activity related to cross-modal integration processes, and suppression over occipital cortex as increased visual attention to gesture content (Drijvers et al., 2018).

Using EEG and fMRI to investigate speech-gesture integration, He et al, (2015) similarly found decreases in alpha power when participants watched videos of speakers gesturing and speaking their native language (German) compared to videos where speakers gestured and spoke a language that was foreign (i.e., when speech-gesture integration was not possible). Additionally, both alpha and beta suppression was found when comparing speech-gesture trials to videos that included speech content without any gestures (He et al., 2015; He et al, 2018). These studies suggest that the demands of speech-gesture integration are indexed by low frequency brain activity, particularly in the alpha and beta bands (Drijvers et al, 2018). Accordingly, we used similar oscillatory power changes as a proxy for multimodal discourse comprehension in the current study.

1.2 The Present Study

Here we examined the relationship between gesture comprehension and visuospatial WM by recording EEG as healthy adults interpreted multimodal discourse under varying levels of visuospatial WM load. Gesture comprehension was indexed by comparing ERSP activity during discourse videos with gestures whose relationship to speech content was semantically congruent versus incongruent. In view of the literature reviewed above, we assumed that the demands of discourse comprehension would be indexed by power decreases in low frequency oscillatory activity, namely greater alpha and beta suppression for incongruent than congruent gestures. Moreover, we expected that increased loads on visuospatial WM would occupy cognitive resources that would otherwise be dedicated to interpreting the gestures, serving to reduce the magnitude of gesture congruity effects during these trials (Wu and Coulson, 2014). We also hypothesized that discourse congruity effects would differ as a function of participants' visuospatial WM capacity, such that participants with greater WM abilities would be more sensitive to the differences between congruent and incongruent gestures and thus display larger effects.

2 Methods and materials

2.1 Participants

Participants were 46 fluent English speakers with no reported neurological or learning disorders (30 females; mean age = 20.0, SD = 1.6). All participants gave informed consent and received academic credit for participation in the experimental protocol. Data collected from 8 additional participants were removed from the final analysis due to excessive artifacts or other equipment malfunctions.

2.2 Working Memory tasks

Before the EEG recording session, two computerized screening tests were given to each participant to assess their WM capacities (see Wu & Coulson, 2014 for a full description of these tasks). The Corsi Block task was used to assess visuospatial WM, and required that participants memorize and recall the order of squares that flashed in random sequences on the computer screen (Total score mean = 21.0, SD = 4.5; Span score mean = 7.6, SD = 1.0). The Sentence Span task, commonly used as a measure of verbal WM capacity, required that participants listen to a series of unrelated sentences while memorizing the final word of

each. Scores were based on the number of final words the participant could correctly recall. (Total score mean = 33.4, SD = 4.4; Span score mean = 3.7, SD = 0.72).

2.3 Materials

A total of 280 videos were used as discourse primes for the experiment. Also used in Wu and Coulson (2014), discourse primes were constructed from video footage of a speaker describing various objects with both speech and iconic gestures. Videos lasted from 2 to 8 seconds (mean=4.1s, SD=1.3s) and involved descriptions of objects, such as the shape of a rug, and activities, such as swinging a golf club. Congruent and incongruent conditions were created by digitally altering the discourse primes in order to reduce the relatedness of the speaker's co-occurring speech and gesture information. In the congruent condition, videos included the original corresponding audio (speech) and video (gesture) information. In the incongruent condition, the audio and video components of the video clips were switched across other discourse videos. The speaker's face was blurred in all videos so that mismatches between the speaker's speech and his mouth movements were not apparent. This resulted in a collection of 140 incongruent videos that contained all of the same gesture and speech information as the 140 congruent videos. No individual participant viewed or heard more than one version of the same video stimulus during the experiment.

The onset of the videos was the stroke of the first gesture, while the offset was the end of the utterance unit. The variability in the length of the videos was intended to ensure that all clips were coherent. The onset of initial content words in the speech files (i.e., an open-class noun or verb) occurred at various times across the video stimuli (mean = 743ms post video onset; SD = 466ms). This marks the earliest point at which participants could determine the congruency of the videos.

Materials were normed in two studies with participants from the same pool as those in the EEG study. Ten individuals who did not participate in the EEG study rated materials on the congruency between the speech and gestures. On a five-point Likert scale (1 = highly incongruent; 5 = highly congruent), congruent videos were rated on average 3.8 (SD = 0.8), and incongruent videos were rated on average 2.2 (SD = 0.7). Another ten volunteers were shown both types of trials and asked to categorize each one as congruent or incongruent. On average, 81% of congruent trials (SD = .17) and 41% of incongruent trials (SD = .32) were judged correctly (mean $d' = .87$; SD = .7). A matched-pairs t-test revealed that across volunteers, signal detection (d') was significantly greater than zero (where zero indicates no detection of the signal) ($t(18) = 3.1$, $p < .01$). These outcomes suggest that the congruency manipulation was fairly subtle, and that individuals tended to interpret the majority of trials as congruent.

2.4 Procedure

The design of the current study was similar to that in Wu and Coulson (2014), who show that the load manipulation used here affected behavioral indices of gesture comprehension, and that this impact on comprehension was modulated by individual differences in visuospatial WM capacity. The trial structure for the study is illustrated in Figure 1. Participants performed the experimental task in a sound attenuated, dimly lit room. Stimuli

were presented in the middle of a 19” color monitor. Participants were instructed to keep their hand placed on the mouse for the duration of the experiment. Trials began with the initial encoding portion of the visuospatial memory task. Participants viewed a four by four grid on the computer screen while dots flashed in sections of the grid at a rate of 1 dot per second. In the high memory load condition, participants saw a sequence of four dots appear in succession, and in the low load condition, only one dot appeared in the grid. Participants were instructed to memorize the location and order of appearance of the dots in order to perform the free-recall portion of the task later in the trial.

500ms after the appearance of the final dot in the memory task, the discourse prime appeared in the middle of the computer screen. Participants were instructed to watch and attend to the discourse primes, but no explicit task was dedicated to the videos or the pictures that immediately followed. After another 500ms pause, a picture appeared on the screen for 500ms that depicted the object or action described in the discourse prime. Following the picture, a blank four by four grid was presented on the screen for the serial-recall memory task. Participants used their mouse to select areas on the grid where the dots had appeared at the beginning of the trial. Selecting both accurate grid locations and the order of dot presentation was necessary for a correct response in the memory task. Feedback on trial performance was given 500ms after the end of the memory task to complete the trial.

The entire experiment consisted of 10 blocks that contained 14 trials each. Blocks were separated by self-paced breaks

2.5 EEG recording

EEG was recorded in a sound attenuated, electromagnetically shielded chamber with 29 scalp electrodes placed at standard International 10–20 sites. In addition to scalp electrodes, two mastoid electrodes were used to reference the EEG, and three facial electrodes served to detect eye related artifacts during recording. Signals were bandpass filtered (0.01–40Hz) and digitized online at 500 Hz. Scalp electrodes were referenced online to the left mastoid site, and after recording, EEG data were re-referenced to the mean of the left and right mastoid sites. Independent components analysis (Infomax ICA: Bell & Sejnowski, 1995) was applied to the continuous EEG data before subsequent ERSP analysis to remove activities associated with eye movements (Delorme & Makeig, 2004). Following the removal of ICs reflecting the contribution of some eye artifacts to the data, raw EEG epochs containing residual artifacts such as eye movements, drift, or blinks, and were rejected prior to ERSP analysis (mean trial rejection rate = 9.6%; 3.1 correct trials per condition). Trials in which the behavioral response was incorrect were also excluded from all analyses.

2.6 EEG processing and analysis

Time-frequency representations of the EEG data were computed in MATLAB using the Fieldtrip toolbox (Oostenveld et al, 2011). Epochs related to the discourse primes were created by extracting a 500ms baseline before and 2000ms after video onset. Considering the length of the video stimuli, a 2 second epoch size was chosen to ensure that information from every trial only contained the brain response to the discourse (i.e., no video stimulus was shorter than 2000ms). A 400ms Hanning window was applied to each epoch at time

steps of 10ms and frequency steps of 0.5Hz between a range of 3–30Hz. Single trial short Fourier transformed epochs were then averaged within each participant and each of four experimental conditions based on our two-factor design (discourse Video by WM Load). 200ms of baseline data from each condition (300 to 100ms before video onset) was used to log transform ($10*\log_{10}$(epoch power/baseline power)) subjects' time-frequency data before statistical analysis. This baseline correction was used for all visualizations of the data.

The *visuospatial resources hypothesis* contends that the availability of visuospatial WM resources will have measurable impacts on participants' online sensitivity to gesture information. To test this, we investigated the effect of Video congruity separately in each WM Load condition using nonparametric cluster-based permutation tests, where effects of Video congruity were calculated as the power difference between incongruent and congruent videos. Additionally, this hypothesis predicts that individual differences in visuospatial WM capacity will also impact participants' sensitivity to the discourse congruity manipulation. To assess this, we integrated a separate hierarchical regression analysis that allowed main effects of interest (i.e., Video congruity) to covary with participants' offline WM measures. Linear mixed-effects models have developed growing popularity in cognitive neurolinguistic research and are particularly valuable because of their ability to render a more sensitive analysis of experimental factors by accounting for variance associated with subject- and item-level information (Payne et al, 2015; Stites et al, 2017; Alday et al, 2017). This is particularly relevant given the difficulty of exerting robust experimental control over parameters related to complex stimuli such as dynamic speech and gesture videos.

ERSP analysis involves an inherently large hypothesis space of channels, times, and frequencies, which coupled with the lack of strong priors about when and where to expect experimental effects can complicate analysis. Consequently, we used the results of nonparametric tests to constrain measurements of the ERSP data used in the regression analyses. To avoid statistical circularity, our total dataset was divided by separating out odd and even trials within each condition and each subject (as per Kriegeskorte, et al., 2009). This resulted in two independent datasets used separately for the cluster-based permutation and linear mixed effects regression analyses. Cluster-based permutation tests were used to help identify time points, electrode sites, and frequency bands to measure. Independently conducted linear mixed effects regression analyses were then utilized as inferential statistics.

2.6.2 Nonparametric cluster-based analysis—Nonparametric cluster-based permutation tests (Maris & Oostenveld, 2007) help circumvent manual decisions to target specific frequency bands and time windows for expected effects and are consequently well suited for paradigms that lack strong expectations about the timing and location of the effects of interest. For a pairwise conditional comparison (e.g., congruent vs incongruent videos during low WM load trials only), a t-test was performed for each data point across the channel-time-frequency matrix between conditions. Spatiotemporally adjacent test statistics were combined into clusters (alpha =0.05) and subsequently given a cluster-level test statistic based on the sum of contributing t-values. This cluster-level statistic was compared to a null distribution created via a randomized permutation (N=1000) across participants, from which a Monte Carlo p-value estimate was derived (Pernet et al, 2015). Each electrode possessed an average of 5.4 neighbor locations in space.

2.6.3 Linear mixed effects modeling—Analyses of alpha band (8–12 Hz) activity involved measurements of decibel corrected average power 1250–1750ms in each odd numbered trial for each subject at each of the significant electrodes from the cluster (listed below). Analyses of beta band activity involved analogous measurements in the frequency range 13–19 Hz. Single trial measurements were used as the dependent variable in mixed effects regression models. Initial models were used to investigate how the time-frequency data were influenced by factors related to the experimental manipulations, and subsequently used individual differences measures as covariates. To account for distributional biases in the relationship between WM measures and Video- or Load-related oscillations, mixed effects models included interactions with electrode location factors Hemisphere (left/midline/right) and Region (frontal, centroparietal, occipital). Frontal electrodes included 10–20 sites FP1, F3, FC3, F7, FT7 (lh), FPz, Fz, FCz (midline), FP2, F4, FC4, F8, FT8 (rh). Centroparietal electrodes included 10–20 sites C3, CP3, P3, T5, TP7 (lh), Cz, Pz (midline), C4 (rh). Occipital electrodes included sites O1 (lh), Oz (midline), and O2 (rh). All models used random effects structures with random intercepts for subjects and individual discourse videos. Variance associated with electrode location was always modeled using a categorical fixed main effect in each model to remove the influence of signal variability at each electrode from the experimental effects of interest.

To test whether experimental manipulation of VSWM load significantly moderated the effect of speech-gesture congruity, an initial omnibus model was run testing for a significant WM Load by Video congruity interaction. This model contained main effects of Video and Load, a Video x Load interaction, as well as two three-way interactions between Load, Video, and each scalp location factor (Hemisphere, Region).

To test whether experimental effects were moderated by individual differences in WM ability, we used forward model comparison to explore whether model fit was improved by the addition of offline measures of WM capacity. Although our hypotheses were focused on understanding how visuospatial WM influenced co-speech gesture comprehension, other individual differences such as verbal WM ability are known to moderate individuals' relationship to co-speech gestures (e.g., Gillespie et al, 2014). To account for this, we allowed Video congruity to interact with offline measures of verbal WM ability to better specify the hypothesized relationship between visuospatial WM and multimodal discourse processing.

Following the omnibus test, we tested interactions between Video congruity and WM abilities separately in low and high WM load trials. Continuous WM abilities were modeled using z-scored raw performance measures on the Corsi block and Sentence Span tasks. During model comparison, the simplest models included main effects of Video, Region, and Hemisphere, as well as Video x Region and Video x Hemisphere interactions. Next, WM measures were added to subsequent models by introducing four new parameters: a main effect of WM, and Video x WM, Video x Region x WM, and Video x Hemisphere x WM interactions. In all, three hierarchical models were fit and compared: an original model testing the effect of Video without WM interactions, a model fit with additional verbal WM measures, and finally a model fit with both verbal and visuospatial WM measures.

Likelihood ratio comparisons were used to determine whether the additional WM factors significantly improved model fit.

Linear mixed effects regression results and model comparison information for the analysis of low WM load trials are in Table 1, while the analysis of high WM load trials is in Table 2. All statistical analysis was performed using R using the lme4 package (R Core Team, 2014).

3 Results

To examine the relationship between visuospatial resources and gesture comprehension, a cohort of healthy adults watched videos of a speaker producing speech and gesture while simultaneously performing a visuospatial WM task. Our results include an analysis of the behavioral performance on the WM task as well as the neural response to the discourse videos. Our ERSP analysis featured an initial analysis of even-numbered trials using nonparametric statistical tests to identify regions of the time-frequency-electrode space that were modulated by experimental factors of discourse congruity (congruent vs. incongruent) in each of the WM load conditions (high vs. low). These outcomes were used to define dependent measures for inferential statistics run on the other half of the dataset. As outlined below, analysis involved the construction of linear mixed effects models to explore first, how the overall effect of discourse congruity (congruent vs. incongruent) manifested differently in the two different memory load conditions (high vs. low), and second, how these effects were moderated by our independent measure of visuospatial WM ability.

3.1 Behavioral results

Performance on the visuospatial recall task was very good (total proportion of correct trials = 92.2%). A generalized linear mixed effects logistic regression model was used to estimate the influence of memory Load and Video congruency factors on task performance on each trial. Using random intercepts for each subject and video item, the model revealed that only memory Load significantly predicted task performance ($\beta = -0.58$; SE = 0.13; $P < 0.001$), where the probability of correct recall on a given trial decreased for high WM load trials (see Figure 2A).

A simple linear regression model was fit using participants' verbal and visuospatial WM scores (standardized Sentence span and Corsi span scores) as predictors of the total number of correct trials on the memory task. In this two-parameter model, Sentence span scores were not significant, and so were removed from the analysis. As the sole predictor, Corsi span scores were significant ($\beta = 4.6$; SE = 1.2; $p<0.001$), accounting for approximately 25% of the variance in memory task performance ($R^2 = 0.25$). Figure 2B displays this relationship as a regression of Corsi span scores with the proportion of total correct trials on the recall task. The positive relationship between task performance and our independent measure of visuospatial WM capacity suggests that the dot task worked as intended to tax visuospatial WM resources.

3.2 EEG results

Cluster-based permutation tests returned one cluster statistic below the significance threshold, indicating a rejection of the null hypothesis of exchangeability across distributions

of congruent and incongruent speech and gesture videos under low working memory loads ($p<0.05$). This cluster-statistic was in the negative direction, indicating lower power values during incongruent videos, and was predominately characterized by frequency estimates spanning alpha and beta ranges (~6–19Hz). Electrodes identified by the nonparametric test indicated this difference was distributed across frontal and posterior scalp regions and occurred for approximately 500ms beginning around 1250ms post-video onset. Comparable analysis of ERSPs in high WM load trials revealed no significant results.

Permutation results informed the configuration for linear mixed effect regression models – that is, the choice of electrodes, frequencies, and the time window for measurements – that were used to test the hypotheses that the effect of semantic congruity between speech and gesture is modulated by visuospatial WM load and mediated by differences in participants' visuospatial WM abilities. Since previous work suggests alpha and beta band activity each subserve partially independent processes during the visual perception and processing of body movements—data from 8–12Hz and from 13–19Hz were extracted as separate dependent measures for two distinct sets of regression models.

3.2.1 Alpha and beta band modulations sensitive to speech and gesture congruity—

As noted above, the cluster-based analysis of the discourse congruity effect during low WM load trials revealed that incongruent videos were associated with greater alpha and beta suppression (Figure 4). Omnibus linear mixed effects model testing the WM Load x Video congruity interaction revealed significant Load x Video x Region interactions in both alpha and beta band activity (alpha: $\beta = -0.37$; SE = 0.18; $p<0.05$; beta: $\beta = -0.32$; SE = 0.14; $p<0.025$). These interactions indicate the Video congruity effect was influenced by additional loads on visuospatial resources.

Mixed effects models were also used to investigate the importance of visuospatial WM ability for the Video congruity effect in low WM Load trials using hierarchical model comparison individually for alpha and beta band activity. Forward model comparison suggested that the model of alpha band activity was improved both by the addition of predictors for participants' verbal WM capacity and by their visuospatial WM capacity (see Table 1), as indexed by their performance on separately administered tests. While the addition of verbal WM scores did not improve models of beta band activity, the model with indices of both verbal and visuospatial WM fared considerably better than the simpler Video congruity model (see Table 1).

The same model comparison procedure was used to investigate relationships present in high WM load trials (Table 2). Models of both alpha and beta activity were improved by the addition of verbal WM scores, and further improved by the addition of visuospatial WM scores. Moreover, inspection of the AIC scores in Table 2 indicates that the complex model with both sorts of WM scores performed better than the simple Video congruity model, even when incorporating a penalty for the additional parameters. The inclusion of verbal WM predictors in these models allows us to interpret the importance of interactions with visuospatial WM as continuing to hold, even after controlling for differences due to verbal WM.

For the analysis of both Low and High WM Load trials, both alpha and beta band activity exhibited Video x Region x VSWM interactions, indicating that visuospatial WM abilities moderated the effect of Video congruity differently in each of our designated scalp regions (Low WM Load: alpha activity: $\beta = 0.50$; SE = 0.12; $p<0.001$; beta activity: $\beta = 0.23$; SE = 0.09; $p<0.025$; High WM Load: alpha activity : $\beta = -0.56$; SE = 0.09; $p<0.0001$; beta activity: $\beta = -0.24$; SE = 0.07; $p<0.0001$). To better characterize these interactions, post hoc models were constructed for each Region for low and high WM load trials, respectively, containing fixed effects of Video, Corsi score, and Sentence Span score, as well as the interactions of Video x Corsi and Video x Sentence Span.

Low WM Load trials: A main effect of Video on alpha band activity was revealed in frontal, centroparietal, and occipital channels (frontal: $\beta = -0.43$; SE = 0.07; $p<0.0001$; centroparietal: $\beta = -0.25$; SE = 0.10; $p=0.01$; occipital: $\beta = -0.36$; SE = 0.17; $p<0.05$), indicating more alpha power suppression during incongruent relative to congruent videos (see Figure 5A). Additionally, a marginal Video by VSWM interaction was observed in occipital channels ($\beta = -0.32$; SE = 0.17; $p=0.05$), suggesting participants with larger VSWM capacity exhibited more relative occipital alpha suppression to incongruent than congruent videos.

Analysis of beta band activity indicated a main effect of Video over frontal channels ($\beta = -0.13$; SE = 0.06; $p<0.025$), and significant Video x VSWM interactions in each Region (Frontal: $\beta = -0.20$; SE = 0.06; $p<0.001$; centroparietal: $\beta = -0.18$; SE = 0.07; $p<0.025$; occipital: $\beta = -0.26$; SE = 0.12; $p<0.05$). The negative estimates for these interaction terms indicate that participants with greater visuospatial WM ability exhibited more beta band suppression during incongruent videos relative to congruent videos (Figure 5B).

High Load trials: Analysis of high load trials revealed a main effect of Video in alpha band activity over frontal and occipital channels (frontal: $\beta = -0.24$; SE = 0.08; $p<0.01$; occipital: $\beta = -0.53$; SE = 0.17; $p<0.01$). However, alpha band Video effects were qualified by interaction with visuospatial WM ability (frontal: $\beta = -0.26$; SE = 0.07; $p<0.001$; centroparietal: $\beta = 0.37$; SE = 0.10; $p<0.001$). The negative estimate of the Video x VSWM interaction over frontal channels indicates participants with greater visuospatial WM ability exhibited more alpha suppression during incongruent relative to congruent videos, while the positive estimate of this interaction term for centroparietal sites reflects relative alpha enhancement during incongruent videos (Figure 6A).

Regression analysis of beta band effects in frontal channels revealed a main effect of both Corsi score ($\beta = -0.42$; SE = 0.15; $p<0.01$) and Video ($\beta = -0.31$; SE = 0.06; $p<0.0001$). This suggests that although visuospatial WM ability did not moderate sensitivity to speech-gesture congruity over frontal electrodes, those with high WM abilities displayed more beta suppression during videos overall. Further, Video x VSWM interactions with positive coefficients were present over centroparietal and occipital regions (centroparietal: $\beta = 0.24$; SE = 0.07; $p<0.01$; occipital: $\beta = 0.30$; SE = 0.12; $p<0.025$), suggesting beta enhancement for incongruent relative to congruent videos among those with greater visuospatial WM abilities (Figure 6B).

4 Discussion

The *visuospatial resources hypothesis* is the proposal that visuospatial WM is used to interpret visual and spatial features of co-speech iconic gestures and maintain these representations until they can be integrated with associated speech in order to produce a unified discourse representation. The present study tested this hypothesis by measuring ERSPs as participants viewed video clips with congruent and incongruent gestures under conditions of high and low visuospatial WM load. Increased demands of multimodal integration were indexed by power modulations in the alpha and beta bands of EEG recorded from frontal electrode sites. These effects were larger in trials with low demands on visuospatial WM than in trials where visuospatial WM was more highly taxed, and thus less available. Our results suggest that additional loads on visuospatial resources affect the ability to process gestures, and the magnitude of this impact is related to individual differences in visuospatial WM ability. Taken together, these data suggest the successful integration of speech and iconic gestures in discourse depends on the availability of visuospatial WM.

4.1 VSWM resource availability enhances neural sensitivity to gesture information

The present study shows that the manipulation of visuospatial WM load has measurable impacts on the brain response to multimodal discourse, suggesting a fundamental connection between visuospatial cognitive resources and the processing of iconic gestures. During low WM load trials, videos with incongruent gestures led to more suppression of frontal alpha (Figure 5A) and beta (Figure 5B) activities than those with congruent gestures. The frontal alpha suppression effect was attenuated in high load trials, and linear regression models suggest its attenuation was most evident in participants with lower VSWM capacity (Figure 6A, top). Additionally, levels of beta suppression were much more evident among participants with larger VSWM capacity during high load trials (Figure 6B, top).

Especially in the low load condition where task demands were minimal, gesture congruity effects observed here were reminiscent of alpha and beta suppression effects reported by previous researchers who have used EEG and MEG to study co-speech gestures (Drijvers, et al., 2018a; Drijvers, et al., 2018b; Drijvers, et al., 2019, He, et al., 2015; He, et al., 2018). Comparing the brain response to matching and mismatching speech and gestures, Drijvers and colleagues found that incongruent speech-gesture combinations elicited greater suppression in both the alpha and beta bands than congruent ones, as the alpha activity localized to the LIFG and visual cortex, and beta activity to LIFG, visual, and premotor cortices (Drijvers et al, 2018a; 2018b; 2019). In view of the similar timing and topography between speech-gesture congruity effects in the present study and those reported by Drijvers and colleagues, we adopt a similar interpretation here: reductions in alpha and beta power induced by incongruous gestures reflect heightened engagement with gestural stimuli whose relationship to the concurrent speech is unclear or ambiguous.

An important difference between the present study and those in the literature, however, is that our paradigm emphasized the working memory task at the expense of discourse comprehension. Indeed, the premise of the study was that increasing the difficulty of the visuospatial memory task would decrease participants' ability to integrate the meaning of the speech and the gestures as they shifted their attentional resources away from the videos.

While the present study did not include a behavioral task to explicitly probe comprehension, the alpha and beta suppression effects we report have previously been associated with behavioral measures of language comprehension at the subject level, which provides strong reason to expect they serve as a neural index of how well participants understood the discourse (Drijvers et al, 2019).

4.2 Gesture processing and sensorimotor recruitment

Fronto-central alpha and beta effects observed in the present study may relate to the sensorimotor mu rhythm induced by the visual presentation of human motor activity (Pellegrino, 1992). Desynchronization of 8–12 Hz oscillatory activity over motor and pre-motor areas during action observation has been taken as evidence that motoric representations are recruited to understand body movements, perhaps via a simulation of the motor percept (Pfurtscheller and Neuper, 1997). If similar activity is partially responsible for the modulation of the alpha band activity observed in the current study, it might indicate that motor representations of the gestures are enhanced during conditions when these representations require a greater allocation of attention or need to be maintained for a longer time in order to integrate their content with unrelated speech.

Regarding speech-gesture congruity effects in the beta band, premotor beta suppression is thought to reflect differential motoric engagement with or “simulation” of speaker movements depending on their semantic relevance to the discourse (Drijvers et al, 2018), aligning with other studies supporting the notion that premotor areas are sensitive to both kinematic and semantic content conveyed by co-speech gestures (Weisburg et al, 2017). By contrast, relative alpha band suppression may indicate increased attentional allocation to hand- or body-based representations to support their integration with unrelated speech (Quandt et al, 2012). Importantly, researchers using combined EEG-BOLD correlational analyses maintain that beta activity plays a less prominent role than alpha in the extraction of semantic information from gestures (He et al, 2018).

The recruitment of sensorimotor cortices in processing iconic gestures may be related to its role in a working memory system hypothesized to support the temporary maintenance of body movements (Sepp, et al., 2019). A representational space for the temporary storage of body-related percepts has long been a component of theories of action imitation (Meltzoff & Moore, 1997), and behavioral research in our lab supports a connection between participants’ ability to reproduce sequences of body configurations from memory (a capacity we refer to as kinesthetic working memory) and the ability to benefit from co-speech gestures (Wu & Coulson, 2015). Neuroimaging research likewise supports a role for somatosensory cortex in the temporary storage of visually presented depictions of body-related stimuli (Galvez-Pol et al., 2018a; 2018b).

A role for sensorimotor cortex in decoding meaningful body movements during communication is in keeping with a growing appreciation of the import of motor related activity in cognitive processes (see Galvez-Pol et al., 2019 and Sepp, et al., 2019 for reviews). Similarly, the recruitment of sensorimotor cortices to maintain body movements aligns with a wider literature demonstrating that WM maintenance is supported by partially non-overlapping networks dedicated to relevant modality specific information (Wager &

Smith, 2003; Lefebvre et al., 2013), and that successful WM performance is achieved by allocating attention to traces in long-term memory stored in diverse areas of the cortex (D'Esposito & Postle, 2015). More research is needed to identify if and precisely how sensorimotor representations in the brain are functionally involved in the comprehension of more abstract visuomotor content such as iconic gestures.

4.3 Visuospatial resource competition and attention to gesturing speakers

Alpha and beta suppression effects in the present study, however, were focused over frontocentral scalp, whereas the mu rhythm has typically been reported over central electrode sites (see Fox, et al., 2016 for a review). In fact, centroparietal alpha band effects observed here differed from those over frontal sites, especially in the high WM load trials (Figure 6A). In participants with superior VSWM, incongruent videos viewed during high load trials elicited greater alpha band activity than congruent ones over centroparietal (and occipital) sites. Posterior alpha enhancement observed in the present study may be related to those in the literature linking alpha enhancement with inhibitory processes used to ignore or suppress distracting, task irrelevant information (Kelly et. Al 2006; Roux and Uhlhaas, 2014; Payne and Sekuler, 2014; Klimesch et al, 2007). For example, alpha power has been shown to increase prior to the onset of distractors in a WM task, and the magnitude of this posterior alpha enhancement during WM maintenance has been related to superior task performance (Bonnefond and Jensen, 2012). Accordingly, the present study suggests that under high VSWM loads, participants with superior VSWM ability were better able to selectively inhibit semantically unrelated gestures.

This explanation of the data is also in keeping with cognitive load theory, a framework that relates WM capacity to inhibitory ability. According to this theory, selective attention mechanisms effectively inhibit distracting information when demands on cognitive functioning are low, but become compromised when cognitive demands are high, resulting in a reduced ability to inhibit task irrelevant information (Lavie et al, 2014). Because cognitive control ability is well measured by span tasks like the Corsi block task used here, cognitive load theory suggests participants with lower WM scores might be unable to optimally suppress discourse information competing with the visuospatial memory task, while those with greater VSWM capacity would be less susceptible to distracting information in the incongruent videos. Consistent with cognitive load theory, our participants' scores on the Corsi block task were associated both with centroparietal alpha enhancement during the videos (Figure 6A, middle) and with better performance on the visuospatial recall task (Figure 2B).

The ability to dynamically control and update attention to speaker movements is important in natural communicative situations, as people change how they treat gestural information depending on factors such as their timing with speech (Obermeier et al, 2011) and their information value (Holle and Gunter, 2007). For example, when speakers intersperse meaningless grooming gestures amongst meaningful ones, listeners respond by reducing their sensitivity to all of the speaker's co-speech gestures (Obermeier et al, 2015). These studies speak to the automaticity of gesture processing—i.e., that co-speech gestural information is always considered when a speaker produces it (Kelly et al, 2010). Evidence

that certain participants excelled at selectively inhibiting misleading gesture information — viz., the systematic variability of incongruity effects in the present study, — undermines the idea that speech-gesture integration is completely obligatory. Indeed, visuospatial WM might help subserve multimodal discourse comprehension by directing selective attention toward or away from gestural cues depending on the semantic or task relevancy, while reduced visuospatial WM abilities may indicate less control over attention to gestures.

Besides the alpha band effects, participants with superior visuospatial WM abilities also showed enhancement effects to incongruent videos in beta band activity recorded over posterior sites (Figure 6B, middle). In a comparison of mismatching speech and gestures when the speech signal was degraded, Drijvers and colleagues found similar beta enhancement effects over left temporoparietal areas which was localized to the left auditory cortex, superior temporal sulcus, MTG, and the medial temporal lobe (Drijvers et al, 2018). The imposition of the visuospatial WM task in the present study may have placed similar attentional demands on our participants that interfered with natural cross-modal integration processes (Oberfield et al, 2016; Talsi et al, 2010). On this interpretation, beta enhancement effects focused over left centroparietal sites may reflect failed cross-modal integration due to the suppression of irrelevant visual information. The fact that beta band enhancement to incongruent speech-gesture videos scaled with visuospatial WM ability lends further support to our suggestion that visuospatial attention influences the degree of multimodal information exchange important for speech and gesture integration.

4.4 Visuospatial Resources Hypothesis

Iconic gestures pose an interesting intersection between visual and semantic processing, as both visual and motoric features of the gestures resemble the information they represent. For example, the trajectory of arm movement during a “swinging” gesture to depict a baseball player’s batting form provides spatial and motoric detail about the event being described, i.e., not only indicating that a swing occurred, but what the swing looked like. These visuospatial details of the gesture need to be interpreted, maintained, and combined with speech information in order to develop a visually refined discourse representation (Wu & Coulson, 2010). By contrast, an emblematic gesture such as a “thumbs-up” provides the same symbolic information regardless of the physical manner in which it is produced. For this reason, we suggest that iconic gestures are apprehended as analog representations that recruit neural networks supporting WM resources for tracking bodily motion through space.

In keeping with this suggestion, neuroimaging data indicates gesture comprehension recruits brain regions that subserve the evaluation of human form and movement, and that networks differ across gesture types, i.e., for emblems, iconic gestures, or grasping actions (Andric & Small, 2012). Similarly, hand movements activate different functional networks depending on whether or not they communicate meaningful information (Corina & Knapp, 2008; Rudner, 2018). Meaningful gestures typically recruit left perisylvian language areas, including the LIFG, as well as the middle and superior temporal gyri (Willems et al, 2007, Andric & Small, 2012; Dick et al, 2014; Demir-Lira et al, 2018), while both meaningful and meaningless gestures activate regions involved in visual and sensorimotor analysis of

body related information, e.g., intraparietal sulcus (IPS), supramarginal gyrus, and ventral and dorsal premotor areas (Andric & Small, 2012; Yang et al, 2015).

The parietal cortex interacts with premotor as well as prefrontal networks responsible for representing visual information for the motor system and spatial working memory operations, respectively (Kravitz et al, 2011). Namely, the intraparietal sulcus has been identified within functional networks mediating visual working memory (Pollman & Cramon, 2000; Corbetta et al, 2002). For example, a study using the Corsi blocks task revealed activation in bilateral occipital and intraparietal cortices, where greater activity in the right IPS was associated with superior Corsi performance (Rotzer et al, 2009). Our demonstration of interference between memory for spatial information and the comprehension of representational gestures is in keeping with an overlap between the brain areas that mediate the processing co-speech gestures and those underlying visuospatial WM. Commonalities between functional networks supporting these processes might underlie the resource competition evidenced in the current study and contributes to a mechanistic motivation for the visuospatial resources hypothesis.

5 Conclusions

In sum, results of the present study suggest an overlap in the neural resources that mediate memory for a sequence of dot locations with those underlying the comprehension of cospeech iconic gestures. This research contributes to a growing literature aimed at scaling neural mechanisms related to human action observation with the comprehension of multimodal discourse. This and other efforts are beginning to shape our understanding of multimodal communication by revealing a functional role for neurocognitive mechanisms related to vision, space, and human motion in our models of language comprehension (Pulvermüller , 2018; Holler & Levinson, 2019).

Acknowledgements

Research was supported by NSF award #BCS-0843946 to S.C. and NIH T32 DC007361 to J.M.

References

Andric M, & Small SL (2012). Gesture’s Neural Language. Frontiers in psychology, 3, 99. [PubMed: 22485103]

Avanzini P, Fabbri-Destro M, Volta RD, Daprati E, Rizzolatti G, & Cantalupo G. (2012). The Dynamics of Sensorimotor Cortical Oscillations during the Observation of Hand Movements: An EEG Study. PLoS ONE, 7(5).

Babiloni C. (2002). Human Cortical Electroencephalography (EEG) Rhythms during the Observation of Simple Aimless Movements: A High-Resolution EEG Study. NeuroImage, 17(2), 559–572. [PubMed: 12377134]

Baddeley AD, & Hitch GJ (1974). Working memory. In Bower GA (Ed.), Recent advances in learning and motivation (Vol. 8, pp. 47–90). New York: Academic Press.

Bell AJ, Sejnowski TJ (1995). An information-maximization approach to blind separation and blind deconvolution. Neural Comput. 7 1129–1159. 10.1162/neco.1995.7.6.1129 [PubMed: 7584893]

Casile A, Caggiano V, & Ferrari PF (2011). The Mirror Neuron System. The Neuroscientist, 17(5), 524–538. [PubMed: 21467305]

Conway AR, Cowan N, & Bunting MF (2001). The cocktail party phenomenon revisited: The importance of working memory capacity. Psychonomic Bulletin & Review, 8(2), 331–335. doi:10.3758/bf03196169 [PubMed: 11495122]

Corbetta M, Kincade JM, & Shulman GL (2002). Neural Systems for Visual Orienting and Their Relationships to Spatial Working Memory. Journal of Cognitive Neuroscience, 14(3), 508–523. [PubMed: 11970810]

Corina DP, & Knapp HP (2008). Signed Language and Human Action Processing. Annals of the New York Academy of Sciences, 1145(1), 100–112. doi:10.1196/annals.1416.023 [PubMed: 19076392]

Coulson S. & Wu YC (2014). 187. Multimodal Discourse Comprehension. In: Muller Cornelia, Cienki Alan, Fricke Ellen, Ladewig Silva H., and McNeil David (Eds.), Body-Language-Communication/ Korper-Sprache-Kommunikation, Vol. 2. Handbucher zur Sprache-und Kommunikationswissenschaft/Handbook of Linghuistics and Communication Science. Berlin, New York: Mouton de Gruyter

Cuevas P, Steines M, He Y, Nagels A, Culham J, & Straube B. (2019). The facilitative effect of gestures on the neural processing of semantic complexity in a continuous narrative. NeuroImage,195, 38–47. [PubMed: 30930310]

Delorme A, & Makeig S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. Journal of Neuroscience Methods,134(1), 9–21 [PubMed: 15102499]

Demir-Lira ÖE, Asaridou SS, Beharelle AR, Holt AE, Goldin-Meadow S, & Small SL (2018). Functional neuroanatomy of gesture-speech integration in children varies with individual differences in gesture processing. Developmental Science, 21(5).

D'Esposito M, & Postle BR (2015). The cognitive neuroscience of working memory. Annual review of psychology, 66, 115–142.

Dick AS, Mok EH, Beharelle AR, Goldin-Meadow S, & Small SL (2012). Frontal and temporal contributions to understanding the iconic co-speech gestures that accompany speech. Human Brain Mapping, 35(3), 900–917. doi:10.1002/hbm.22222 [PubMed: 23238964]

Dong S, Reder LM, Yao Y, Liu Y, & Chen F. (2015). Individual differences in working memory capacity are reflected in different ERP and EEG patterns to task difficulty. Brain Research, 1616, 146–156 [PubMed: 25976774]

Drijvers L, Özyürek A, & Jensen O. (2018). Alpha and Beta Oscillations Index Semantic Congruency between Speech and Gestures in Clear and Degraded Speech. Journal of Cognitive Neuroscience, 30(8), 1086–1097. [PubMed: 29916792]

Drijvers L, Özyürek A, & Jensen O. (2018). Hearing and seeing meaning in noise: Alpha, beta, and gamma oscillations predict gestural enhancement of degraded speech comprehension. Human Brain Mapping, 39(5), 2075–2087. [PubMed: 29380945]

Drijvers L, Plas MV, Özyürek A, & Jensen O. (2019). Native and non-native listeners show similar yet distinct oscillatory dynamics when using gestures to access speech in noise. NeuroImage, 194, 55–67. doi:10.1016/j.neuroimage.2019.03.032 [PubMed: 30905837]

Fox NA, Bakermans-Kranenburg MJ, Yoo KH, Bowman LC, Cannon EN, Vanderwert RE, … & Van IJzendoorn MH (2016). Assessing human mirror activity with EEG mu rhythm: A meta-analysis. Psychological Bulletin, 142(3), 291. [PubMed: 26689088]

Galvez-Pol Alejandro & Forster Bettina & Calvo-Merino Beatriz. (2019). Beyond action observation: neurobehavioral mechanisms of memory for visually perceived bodies and actions. 10.31234/osf.io/2vq6p.

Galvez-Pol A, Calvo-Merino B, Capilla A, & Forster B. (2018). Persistent recruitment of somatosensory cortex during active maintenance of hand images in working memory. NeuroImage,174, 153–163. [PubMed: 29548846]

Galvez-Pol A, Forster B, & Calvo-Merino B. (2018). Modulation of motor cortex activity in a visual working memory task of hand images. Neuropsychologia, 117, 75–83. [PubMed: 29738793]

Grabner R, Fink A, Stipacek A, Neuper C, & Neubauer A. (2004). Intelligence and working memory systems: Evidence of neural efficiency in alpha band ERD. Cognitive Brain Research, 20(2), 212–225. [PubMed: 15183393]

Green A, Straube B, Weis S, Jansen A, Willmes K, Konrad K, Kircher T. Neural integration of iconic and unrelated coverbal gestures: A functional MRI study. Human Brain Mapping. 2009; 30:3309–3324. [PubMed: 19350562]

Hagoort P. (2004). Integration of Word Meaning and World Knowledge in Language Comprehension. Science, 304(5669), 438–441. [PubMed: 15031438]

Haier RJ, Siegel BV, Nuechterlein KH, Hazlett E, Wu JC, Paek J, … Buchsbaum MS (1988). Cortical glucose metabolic rate correlates of abstract reasoning and attention studied with positron emission tomography. Intelligence, 12(2), 199–217.

Hall Judith A., Horgan Terrence G., Murphy Nora A. (2019). Nonverbal Communication. Annual Review of Psychology, 70:1

Hari R. (2006). Action–perception connection and the cortical mu rhythm. Progress in Brain Research Event-Related Dynamics of Brain Oscillations, 253–260. doi:10.1016/s0079-6123(06)59017-x

Hari R, Forss N, Avikainen S, Kirveskari E, Salenius S, & Rizzolatti G. (1998). Activation of human primary motor cortex during action observation: A neuromagnetic study. Proceedings of the National Academy of Sciences, 95(25), 15061–15065

He Y, Gebhardt H, Steines M, Sammer G, Kircher T, Nagels A, & Straube B. (2015). The EEG and fMRI signatures of neural integration: An investigation of meaningful gestures and corresponding speech. Neuropsychologia, 72, 27–42. [PubMed: 25900470]

He Y, Steines M, Sommer J, Gebhardt H, Nagels A, Sammer G, … Straube B. (2018). Spatial–temporal dynamics of gesture–speech integration: A simultaneous EEG-fMRI study. Brain Structure and Function, 223(7), 3073–3089. [PubMed: 29737415]

Holle H, & Gunter TC (2007). The Role of Iconic Gestures in Speech Disambiguation: ERP Evidence. Journal of Cognitive Neuroscience, 19(7)

Holler J, & Levinson SC (2019). Multimodal Language Processing in Human Communication. Trends in Cognitive Sciences, 23(8), 639–652. [PubMed: 31235320]

Iacoboni M, & Dapretto M. (2006). The mirror neuron system and the consequences of its dysfunction. Nature Reviews Neuroscience, 7(12), 942–951. [PubMed: 17115076]

Jaeggi SM, Buschkuehl M, Etienne A, Ozdoba C, Perrig WJ, & Nirkko AC (2007). On how high performers keep cool brains in situations of cognitive overload. Cognitive, Affective, & Behavioral Neuroscience, 7(2), 75–89.

Jensen O, Gelfand J, Kounios J, Lisman JE (2002). Oscillations in the alpha band (9–12 Hz) increase with memory load during retention in a short-term memory task. Cereb. Cortex 12, 877–882 10.1093/cercor/12.8.877 [PubMed: 12122036]

Jensen O, & Mazaheri A. (2010). Shaping functional architecture by oscillatory alpha activity: gating by inhibition. Frontiers in human neuroscience, 4, 186. doi:10.3389/fnhum.2010.00186 [PubMed: 21119777]

Kane MJ, Bleckley MK, Conway AR, Engle RW. A controlled-attention view of working-memory capacity. Journal of Experimental Psychology: General. 2001;130(2):169–183. [PubMed: 11409097]

Kane MJ, Hambrick DZ, Tuholski SW, Wilhelm O, Payne TW, & Engle RW (2004). The Generality of Working Memory Capacity: A Latent-Variable Approach to Verbal and Visuospatial Memory Span and Reasoning. Journal of Experimental Psychology: General, 133(2), 189–217. doi:10.1037/0096-3445.133.2.189 [PubMed: 15149250]

Kelly SD, Creigh P, & Bartolotti J. (2010). Integrating Speech and Iconic Gestures in a Stroop-like Task: Evidence for Automatic Processing. Journal of Cognitive Neuroscience, 22(4), 683–694. [PubMed: 19413483]

Kelly SP, Lalor EC, Reilly RB, & Foxe JJ (2006). Increases in Alpha Oscillatory Power Reflect an Active Retinotopic Mechanism for Distracter Suppression During Sustained Visuospatial Attention. Journal of Neurophysiology, 95(6), 3844–3851 [PubMed: 16571739]

Kircher T, Straube B, Leube D, Weis S, Sachs O, Willems K, Green A… et al. (2009). Neural interaction of speech and gesture: Differential activations of metaphoric co-verbal gestures. Neuropsychologia, 47(1), 169–179. [PubMed: 18771673]

Klimesch W, Sauseng P, & Hanslmayr S. (2007). EEG alpha oscillations: the inhibition– timing hypothesis. Brain research reviews, 53(1), 63–88 [PubMed: 16887192]

Lavie N, & Dalton P. (2014). Load Theory of Attention and Cognitive Control. Oxford Handbooks Online.

Lefebvre C, Vachon F, Grimault S, Thibault J, Guimond S, Peretz I, … Jolicœur P. (2013). Distinct electrophysiological indices of maintenance in auditory and visual short-term memory. Neuropsychologia, 51(13), 2939–2952. doi:10.1016/j.neuropsychologia.2013.08.003 [PubMed: 23938319]

Lepage J-F, & Théoret H. (2006). EEG evidence for the presence of an action observation-execution matching system in children. European Journal of Neuroscience, 23(9), 2505–2510 [PubMed: 16706857]

Malmivuo J. (2011). Comparison of the Properties of EEG and MEG in Detecting the Electric Activity of the Brain. Brain Topography, 25(1), 1–19. [PubMed: 21912974]

Maris E, Oostenveld R, 2007. Nonparametric statistical testing of EEG- and MEG-data. J. Neurosci. Methods 164, 177–190 [PubMed: 17517438]

Masson-Carro I, Goudbeek M, & Krahmer E. (2017). How What We See and What We Know Influence Iconic Gesture Production. Journal of Nonverbal Behavior, 41(4), 367–394. [PubMed: 29104335]

McNeill D, Cassell J & McCullough KE. (1994) Communicative Effects of Speech-Mismatched Gestures, Research on Language and Social Interaction, 27:3,223–237

Meltzoff AN, & Moore MK (1997). Explaining Facial Imitation: A Theoretical Model. Early development & parenting, 6(3–4), 179–192. [PubMed: 24634574]

Meltzoff AN, & Moore MK (1977). Imitation of facial and manual gestures by human neonates. Science, 198, 75–78 [PubMed: 17741897]

Meltzoff AN, & Moore MK (1995). Infants' understanding of people and things: From body imitation to folk psychology. In Bermúdez JL, Marcel A, & Eilan N. (Eds.), The body and the self (pp. 43–69). Cambridge, MA: MIT Press

Meyer L, Obleser J, & Friederici AD (2013). Left parietal alpha enhancement during working memory-intensive sentence processing. Cortex, 49(3), 711–721 [PubMed: 22513340]

Morsella E, & Krauss RM (2004). The Role of Gestures in Spatial Working Memory and Speech. The American Journal of Psychology, 117(3), 411. [PubMed: 15457809]

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

Özyürek A. (2014). Hearing and seeing meaning in speech and gesture: Insights from brain and behaviour. Philosophical Transactions of the Royal Society B: Biological Sciences, 369(1651), 20130296–20130296.

Özyürek A, Willems RM, Kita S, & Hagoort P. (2007). On-line Integration of Semantic Information from Speech and Gesture: Insights from Event-related Brain Potentials. Journal of Cognitive Neuroscience, 19(4), 605–616 [PubMed: 17381252]

Parks RW, Loewenstein DA, Dodrill KL, Barker WW, Yoshii F, Chang JY, … Duara R. (1988). Cerebral metabolic effects of a verbal fluency test: A PET scan study. Journal of Clinical and Experimental Neuropsychology, 10(5), 565–575 [PubMed: 3265709]

Payne L, & Sekuler R. (2014). The Importance of Ignoring. Current Directions in Psychological Science, 23(3), 171–177. [PubMed: 25530685]

Pellegrino GD, Fadiga L, Fogassi L, Gallese V, & Rizzolatti G. (1992). Understanding motor events: A neurophysiological study. Experimental Brain Research, 91(1), 176–180. doi:10.1007/bf00230027 [PubMed: 1301372]

Pernet CR, Latinus M, Nichols TE, & Rousselet GA (2015). Cluster-based computational methods for mass univariate analyses of event-related brain potentials/fields: A simulation study. Journal of neuroscience methods, 250, 85–93. [PubMed: 25128255]

Pesonen M, Hämäläinen H, & Krause CM (2007). Brain oscillatory 4–30 Hz responses during a visual n-back memory task with varying memory load. Brain Research, 1138, 171–177. [PubMed: 17270151]

Pfurtscheller G, & Silva FL (1999). Event-related EEG/MEG synchronization and desynchronization: Basic principles. Clinical Neurophysiology, 110(11), 1842–1857 [PubMed: 10576479]

Pollmann S, & Cramon DY (2000). Object working memory and visuospatial processing: Functional neuroanatomy analyzed by event-related fMRI. Experimental Brain Research, 133(1), 12–22. [PubMed: 10933206]

Proskovec AL, Heinrichs-Graham E, & Wilson TW (2019). Load modulates the alpha and beta oscillatory dynamics serving verbal working memory. NeuroImage, 184, 256–265. [PubMed: 30213775]

Pulvermüller F. (2018). Neural reuse of action perception circuits for language, concepts and communication. Progress in Neurobiology, 160, 1–44. doi: 10.1016/j.pneurobio.2017.07.001 [PubMed: 28734837]

Quandt LC, Marshall PJ, Shipley TF, Beilock SL, & Goldin-Meadow S. (2012). Sensitivity of alpha and beta oscillations to sensorimotor characteristics of action: An EEG study of action production and gesture observation. Neuropsychologia, 50(12), 2745–2751. [PubMed: 22910276]

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rotzer S, Loenneker T, Kucian K, Martin E, Klaver P, & Aster MV (2009). Dysfunctional neural network of spatial working memory contributes to developmental dyscalculia. Neuropsychologia, 47(13), 2859–2865. doi:10.1016/j.neuropsychologia.2009.06.009 [PubMed: 19540861]

Roux F, & Uhlhaas PJ (2014). Working memory and neural oscillations: Alpha–gamma versus theta–gamma codes for distinct WM information? Trends in Cognitive Sciences, 18(1), 16–25 [PubMed: 24268290]

Rudner M. (2018). Working Memory for Linguistic and Non-linguistic Manual Gestures: Evidence, Theory, and Application. Frontiers in psychology, 9, 679. [PubMed: 29867655]

Sepp S, Howard SJ, Tindall-Ford S, Agostinho S, & Paas F. (2019). Cognitive Load Theory and Human Movement: Towards an Integrated Model of Working Memory. Educational Psychology Review, 31(2), 293–317.

The MathWorks MATLAB, signal processing, and statistics toolboxes, Natick, Massachusetts, United States. 2014

Tuladhar AM, Huurne NT, Schoffelen J, Maris E, Oostenveld R, & Jensen O. (2007). Parieto-occipital sources account for the increase in alpha activity with working memory load. Human Brain Mapping, 28(8), 785–792. doi:10.1002/hbm.20306 [PubMed: 17266103]

Wager TD, Smith EE (2003). Neuroimaging studies of working memory: a meta-analysis. Cogn. Affect. Behav. Neurosci. 3, 255–274. [PubMed: 15040547]

Willems RM, & Hagoort P. (2007). Neural evidence for the interplay between language, gesture, and action: A review. Brain and Language,101(3), 278–289 [PubMed: 17416411]

Willems RM, Özyürek A, & Hagoort P. (2007). When Language Meets Action: The Neural Integration of Gesture and Speech. Cerebral Cortex,17(10), 2322–2333. [PubMed: 17159232]

Wu YC, & Coulson S. (2011). Are depictive gestures like pictures? Commonalities and differences in semantic processing. Brain and Language,119(3), 184–195 [PubMed: 21864890]

Wu YC, & Coulson S. (2014). Co-speech iconic gestures and visuo-spatial working memory. Acta Psychologica, 153, 39–50. [PubMed: 25282199]

Wu YC, & Coulson S. (2010). Gestures modulate speech processing early in utterances. NeuroReport, 21(7), 522–526. [PubMed: 20375745]

Wu YC, & Coulson S. (2015). Iconic Gestures Facilitate Discourse Comprehension in Individuals With Superior Immediate Memory for Body Configurations. Psychological Science, 26(11), 1717–1727. [PubMed: 26381507]

Yang J, Andric M, & Mathew MM (2015). The neural basis of hand gesture comprehension: A meta-analysis of functional magnetic resonance imaging studies. Neuroscience & Biobehavioral Reviews, 57, 88–104. [PubMed: 26271719]

Zhao W, Riggs K, Schindler I, & Holle H. (2018). Transcranial Magnetic Stimulation over Left Inferior Frontal and Posterior Temporal Cortex Disrupts Gesture-Speech Integration. The Journal of Neuroscience, 38(8), 1891–1900. [PubMed: 29358361]

Highlights

- EEG was recorded as adults watched discourse videos with gestures that were either congruent or incongruent with speech
- Incongruent gestures led to more neural engagement, indexed by suppression in the alpha and beta bands of the EEG
- When visuospatial working memory (WM) resources were taxed, alpha/beta suppression congruency effects were disrupted
- Oscillatory EEG effects suggest competition between neural resources used for gesture comprehension and for visuospatial WM
- Gesture comprehension recruits visuospatial neurocognitive resources

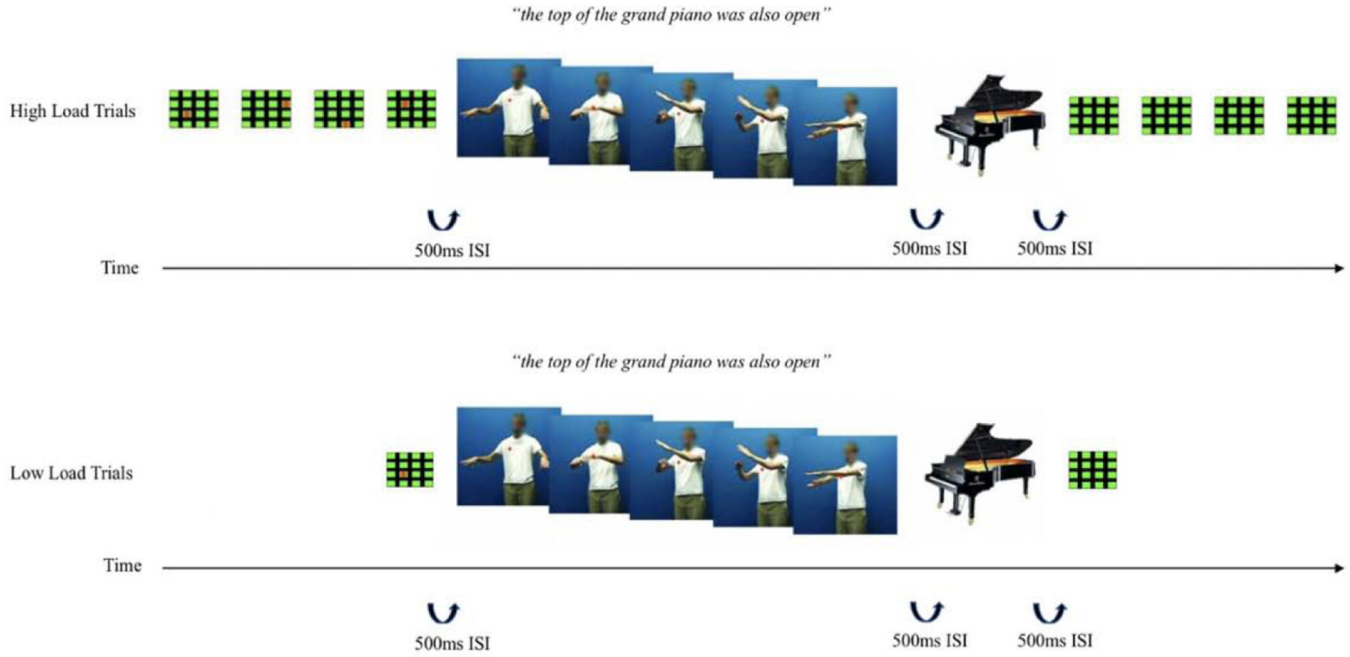

Figure 1:

Summary of experimental paradigm. High load trials involved encoding 4 dots (SOA 1s) followed by a discourse prime and picture probe before a free recall task related to the dot information on the same trial. The low load condition included only a single dot on the memory task but was otherwise identical.

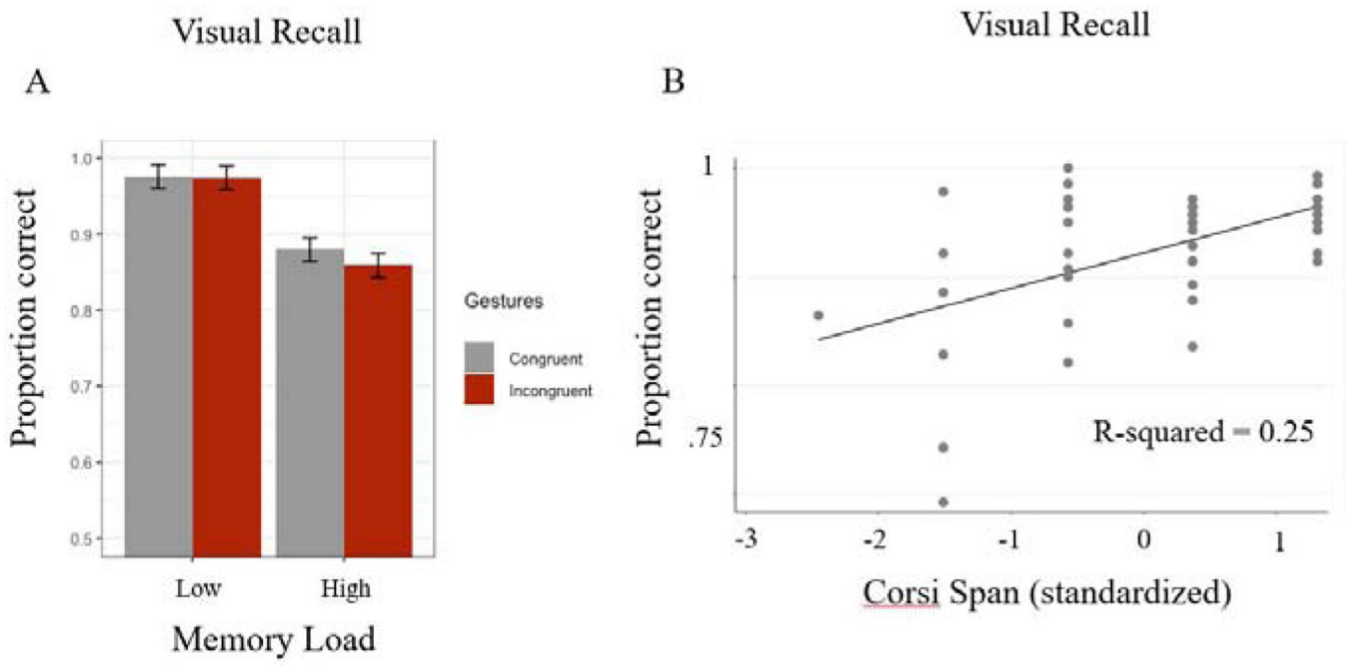

Figure 2.

A) Proportion of correct trials (with 95% confidence intervals) in each task condition across all participants. B) Relationship between Corsi Span scores and visual recall performance during the EEG task.

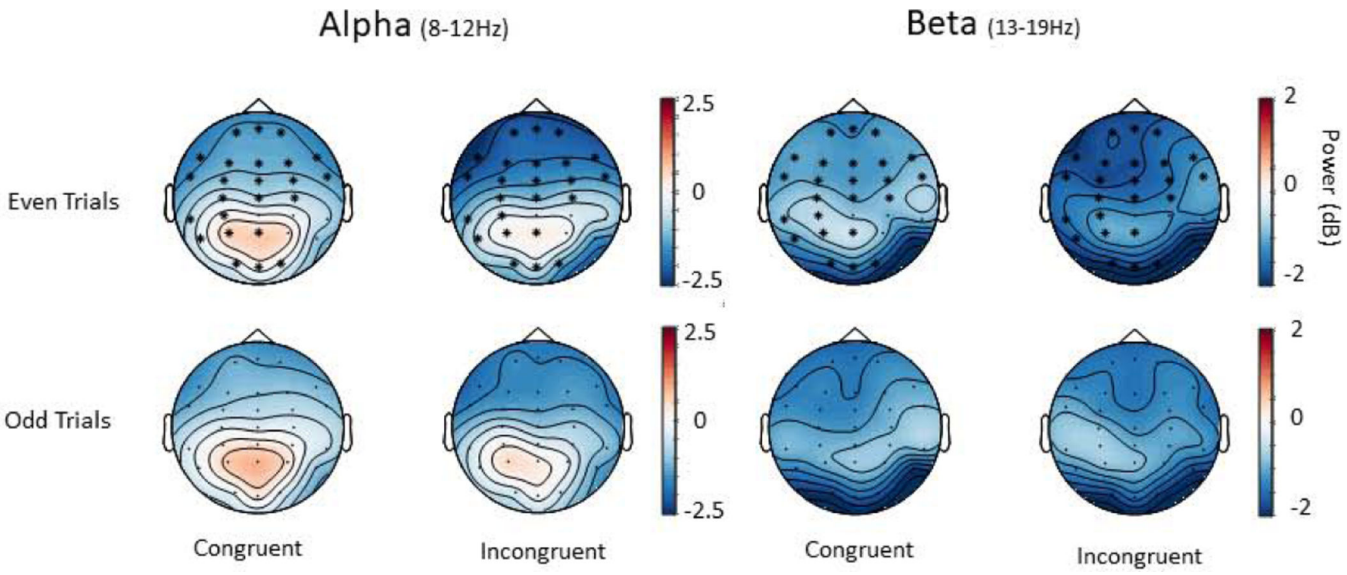

Figure 3.

Topographical depiction of the significant effect of Video congruency in Low load trials revealed by cluster-based permutation analysis of even trials (top row). Electrodes marked with asterisks signify channels from the significant cluster permutation analysis that revealed the effect of Video congruity in low WM load trials. This analysis revealed a cluster spanning alpha and beta band activity, which was analyzed separately via regression models using the other data partition, i.e. odd trials. Odd trials (bottom row) show a similar distribution and magnitude of the low WM load Video effect compared to data from even trials but were not included in the cluster analysis and thus do not display asterisks over electrodes.

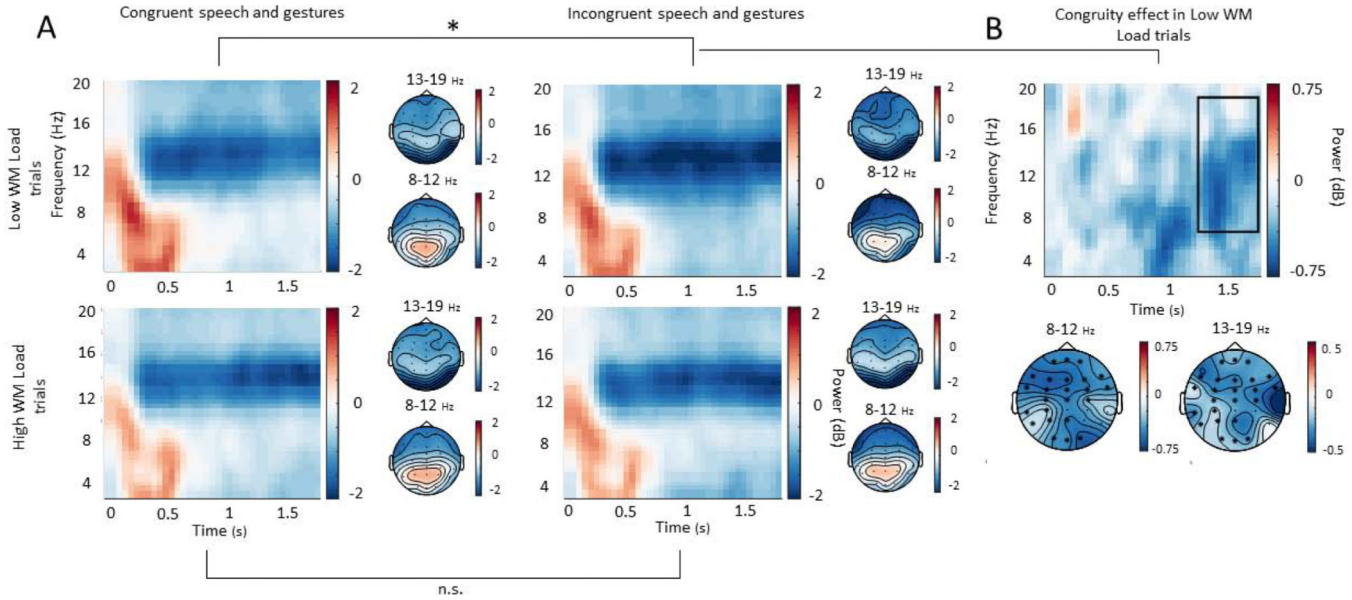

Figure 4.

Spectrogram and topographical representations of time-frequency data corresponding to the 2×2 Video by WM Load factor design across all trials. Cluster permutation results revealed relative power suppression across alpha and beta band frequencies from approximately 1250–1750ms post-video onset. Topographical maps depict averaged alpha (top) and beta (bottom) band activity across this time range. 0ms corresponds to video onset. B. Representations of the effect of Video in low WM load trials.

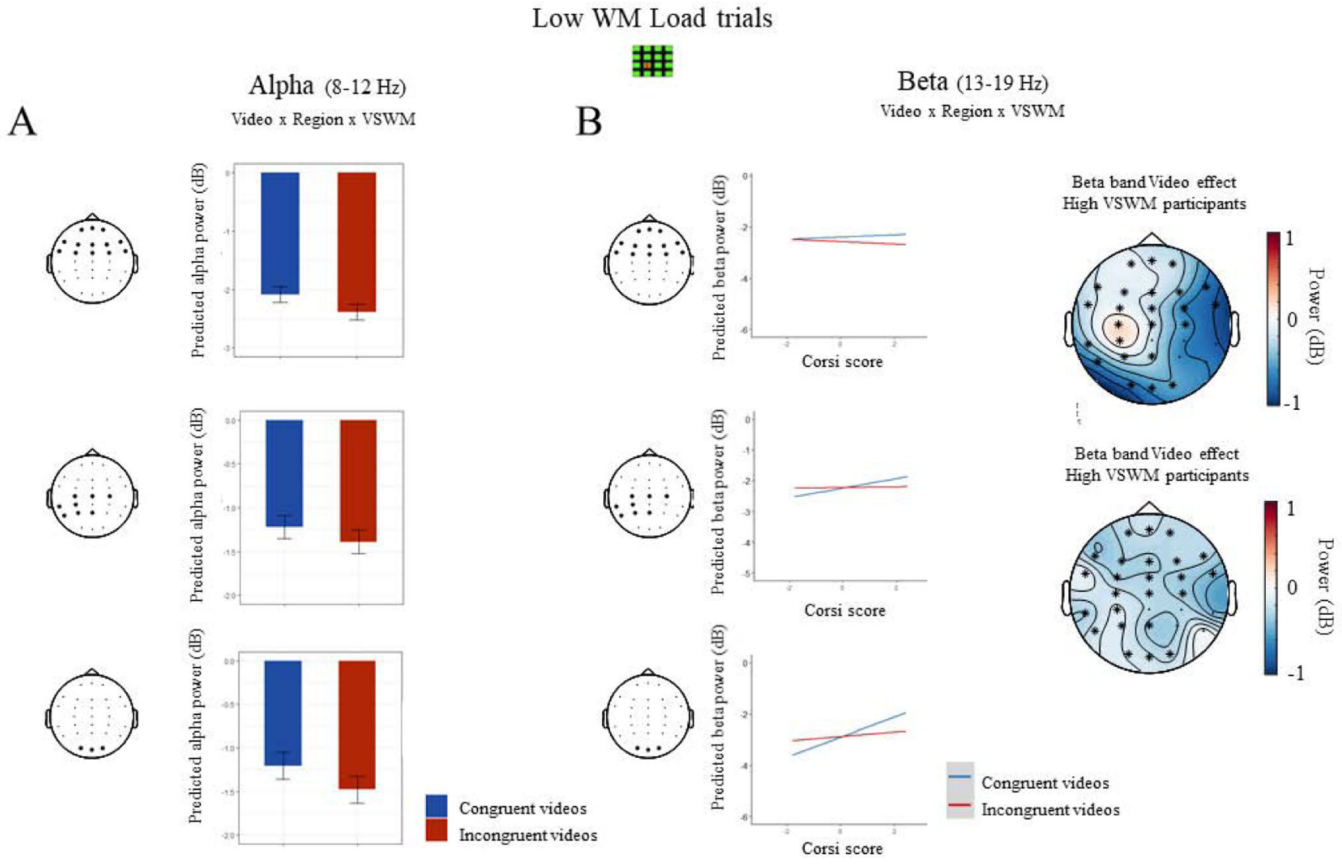

Figure 5.

A) Linear mixed effects models of alpha band activity revealed a main effect of speechgesture congruity during Low WM Load trials. The bar graph displays alpha power measurements averaged across electrodes in each of the scalp regions. B) Analyses of beta band activity indicated a Video by Corsi score interaction in Frontal, Centroparietal, and Occipital scalp regions, indicating larger Video congruity effects in the form of beta band suppression in participants with greater visuospatial WM abilities. Scalp plots show the Video effect in beta band activity (13–19Hz) averaged across all participants (bottom) and in a select group of eleven participants with Corsi scores more than 1 standard deviation above the group average (top). Topographical maps include data corresponding to the Video congruity effect (incongruent minus congruent trials) averaged over 1250–1750ms post-video onset. Topographical data shown here includes the entire dataset.

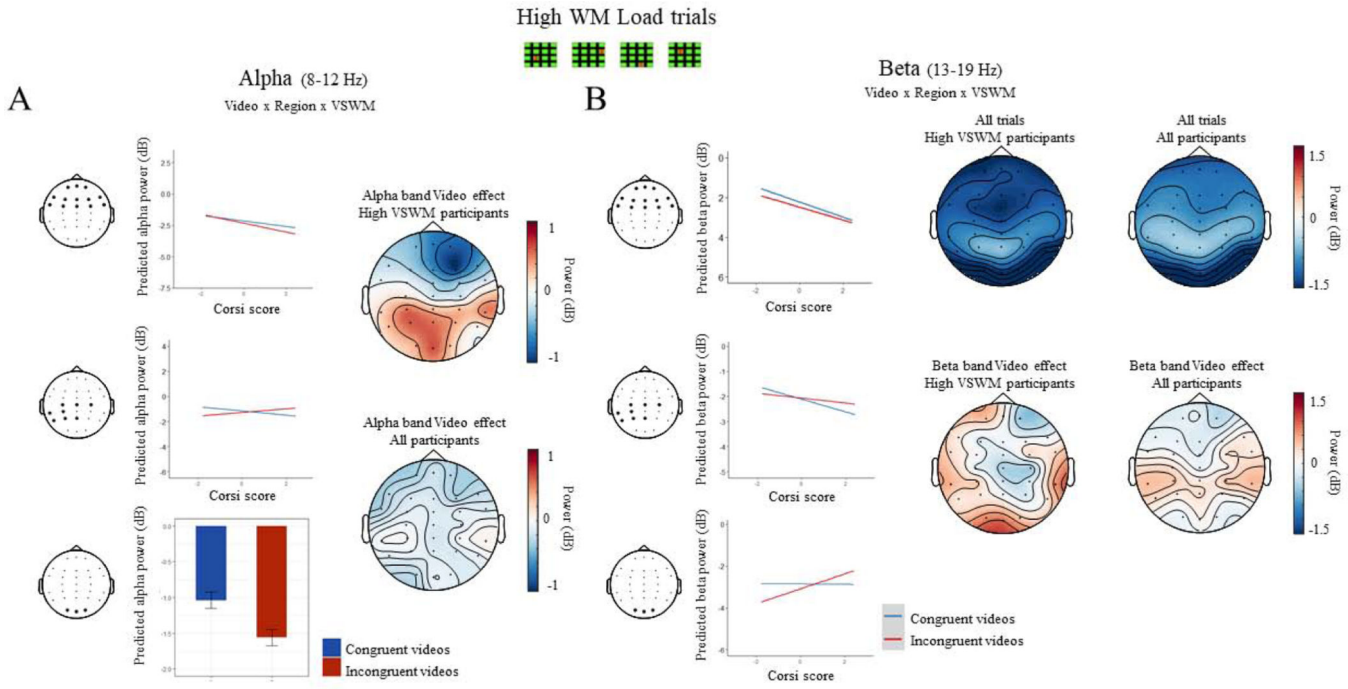

Figure 6.

A) Linear mixed effects models of alpha band activity during high load trials revealed interactions between Video congruity and visuospatial WM ability that differed over regions of the scalp. Over frontal electrodes, participants with higher visuospatial WM abilities displayed more relative alpha suppression during incongruent videos, while over centroparietal sites these participants displayed more relative alpha band enhancement. Scalp plots show the Video effect in alpha band activity (8–12Hz) averaged across all participants (bottom) and in a select group of eleven participants with Corsi scores more than 1 standard deviation above the group average (top). B) (Top) A main effect of Corsi score over frontal channels indicates greater overall beta suppression as a function of increasing visuospatial WM abilities. Topographical maps display data averaged across congruent and incongruent videos in all participants (right) and in participants with superior visuospatial WM abilities (left). B) (Bottom) Regression analyses indicated a Video by Corsi score interaction in left centroparietal regions, indicating beta band enhancement during incongruent videos in participants with greater visuospatial WM abilities. Scalp plots show the Video effect in beta band activity (13–19Hz) across all participants (right) and in those with superior visuospatial WM abilities (left). All depicted topographical maps show ERSP activity averaged from 1250–1750ms post-video onset and includes the total dataset.

Table 1

Model comparisons and output summaries for mixed effects regressions modeling alpha band (top) and beta band (bottom) Video congruity effects in low WM load trials. Model comparisons suggest the addition of visuospatial WM measures as a moderator of speech-gesture congruity significantly improved model fit for both alpha and beta activity.

Alpha (8–12 Hz)			
Model	AIC	logLik	Chi^2
Video	208344	−104140	
with Verbal WM	208332	−104124	31.399***
with VSWM and Verbal WM	**208307**	−104101	**45.293*** **
Parameters	Coefficient	Std Error	t-value
Intercept	−0.13***	0.25	−5.08
Video	−0.32*	0.13	−2.48
Region	−0.72***	0.19	−3.82
Video x Region x Verbal WM	−0.18*	0.08	−2.23
Video x Hemisphere x VSWM	−0.23*	0.09	−2.36
Video x Region x VSWM	0.50***	0.12	4.00
Beta (13–19 Hz)			
Model	AIC	IogLik	Chi^2
Video	191749	−95843	
with Verbal WM	191752	−95834	16.9 (n.s.)
with VSWM and Verbal WM	**191730**	**−95813**	**41.8*** **
Parameters	Coefficient	Std Error	t-value
Intercept	−2.54**	0.18	−13.83
Video x Region x VSWM	0.23*	0.09	2.34

Significance codes:

*** p< 0.001;

** p<0.01;

* p<0.05

Table 2

Model comparisons and output summaries for mixed effects regressions modeling alpha band (top) and beta band (bottom) Video congruity effects in high WM load trials. Model comparisons suggest the addition of visuospatial WM measures as a moderator of speech-gesture congruity significantly improved model fit for both alpha and beta activity.

Alpha (8–12 Hz)			
Model	AIC	IogLik	Chi^2
Video	190260	−95098	
with Verbal WM	190241	−95098	39.58 *
with VSWM and Verbal WM	**190157**	**−95026**	**103.83** ***
Parameters	Coefficient	Std Error	t-value
Intercept	−1.12 **	0.33	−3.38
Region	−0.98 ***	0.19	−5.03
Video x Region	−0.39 *	0.18	−2.15
Video x Verbal WM	−0.36 **	0.14	−2,61
Video x VSWM	0.30 *	0.14	2.12
Video x Region x Verbal WM	0.49 ***	0.08	5.53
Video x Region x VSWM	−0.56 ***	0.09	−6.22
Beta (13–19 Hz)			
Model	AIC	IogLik	Chi^2
Video	173167	−86552	
with Verbal WM	173159	−86537	28.56 **
with VSWM and Verbal WM	**173077**	**−86487**	**101.79** ***
Parameters	Coefficient	Std Error	t-value
Intercept	−2.42 ***	0.19	−12.13
Hemisphere	0.34 *	0.15	2.28
VSWM	−0.35 *	0.16	−2.16
Video x Region	−0.28 **	0.09	−3.05
Video x VSWM	0.22 *	0.11	2.10
Video x Hemisphere x Verbal WM	−0.18 *	0.07	−2.41
Video x Region x Verbal WM	0.19 **	0.07	2.88
Video x Region x VSWM	−0.24 ***	0.7	−3.51

Significance codes:

*** p< 0.001;

** p<0.01;

* $p<0.05$