# Alpha and Beta Oscillations Index Semantic Congruency between Speech and Gestures in Clear and Degraded Speech

**Linda Drijvers[1], Asli Özyürek[1,2], and Ole Jensen[3]**

## Abstract

■ Previous work revealed that visual semantic information conveyed by gestures can enhance degraded speech comprehension, but the mechanisms underlying these integration processes under adverse listening conditions remain poorly understood. We used MEG to investigate how oscillatory dynamics support speech–gesture integration when integration load is manipulated by auditory (e.g., speech degradation) and visual semantic (e.g., gesture congruency) factors. Participants were presented with videos of an actress uttering an action verb in clear or degraded speech, accompanied by a matching (mixing gesture + "mixing") or mismatching (drinking gesture + "walking") gesture. In clear speech, alpha/beta power was more suppressed in the left inferior frontal gyrus and motor and visual cortices when integration load increased in response to mismatching versus matching gestures. In degraded speech, beta power was less suppressed over posterior STS and medial temporal lobe for mismatching compared with matching gestures, showing that integration load was lowest when speech was degraded and mismatching gestures could not be integrated and disambiguate the degraded signal. Our results thus provide novel insights on how low-frequency oscillatory modulations in different parts of the cortex support the semantic audiovisual integration of gestures in clear and degraded speech: When speech is clear, the left inferior frontal gyrus and motor and visual cortices engage because higher-level semantic information increases semantic integration load. When speech is degraded, posterior STS/middle temporal gyrus and medial temporal lobe are less engaged because integration load is lowest when visual semantic information does not aid lexical retrieval and speech and gestures cannot be integrated. ■

## INTRODUCTION

Oscillatory dynamics are thought to subserve the integration of complex information from multiple modalities (Varela, Lachaux, Rodriguez, & Martinerie, 2001), such as during multisensory integration (Schepers, Schneider, Hipp, Engel, & Senkowski, 2013; Kayser & Logothetis, 2009; Schroeder, Lakatos, Kajikawa, Partan, & Puce, 2008; Senkowski, Schneider, Foxe, & Engel, 2008). Low-frequency oscillatory power decreases in the alpha and beta bands are often related to the engagement of brain areas, whereas increases are often related to disengagement or functional inhibition of task-irrelevant brain regions (Jensen & Mazaheri, 2010; Klimesch, Sauseng, & Hanslmayr, 2007; Pfurtscheller & Lopes da Silva, 1999). In line with this, previous research revealed that oscillatory power increases can predict the degree of nonsemantic audiovisual integration of an ambiguous stimulus (e.g., beeps and flashes; Hipp, Engel, & Siegel, 2011). However, it is poorly understood how these mechanisms translate to semantic audiovisual integration, such as in multimodal speech processing.

Investigating whether similar oscillatory mechanisms also support more realistic situations is particularly relevant when considering face-to-face communication, which integrates auditory (e.g., speech) and visual (e.g., gestures) modalities. Under adverse listening conditions, speech comprehension can be enhanced by the visual semantic information conveyed by iconic gestures (Drijvers & Özyürek, 2017; Holle, Obleser, Rueschemeyer, & Gunter, 2010). These iconic gestures can illustrate object attributes, actions, and space (McNeill, 1992) and are known to affect clear and degraded speech comprehension (e.g., Drijvers & Özyürek, 2018; Drijvers, Özyürek, & Jensen, 2018; Zhao, Riggs, Schindler, & Holle, 2018; Drijvers & Özyürek, 2017; Dick, Mok, Raja Beharelle, Goldin-Meadow, & Small, 2014; Straube, Green, Weis, & Kircher, 2012; Holle et al., 2010; Green et al., 2009; Willems, Özyürek, & Hagoort, 2007, 2009; see Özyürek, 2014, for a review). For example, when the semantic information that is conveyed by these gestures mismatches clear speech, previous studies have demonstrated that semantic integration load increases and audiovisual integration might be hindered. For example, previous fMRI studies have demonstrated

This paper is part of a Special Focus deriving from a symposium at the 2017 annual meeting of Cognitive Neuroscience Society, entitled "Top–Down Functions of Neural Oscillations for Speech and Language Processing."

[1]Radboud University, [2]Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands, [3]University of Birmingham

more BOLD activation in the left inferior frontal gyrus (LIFG) when semantic integration load increased and gestures mismatched rather than matched clear speech (Willems et al., 2007, 2009). Similar effects have been demonstrated in EEG studies, where the N400, an ERP component that is sensitive to semantic unification operations, was more negative when gestures mismatched than matched clear speech (e.g., Kelly, Kravitz, & Hopkins, 2004). Extending on this, recent work demonstrated that the difference in N400 amplitude (i.e., the N400 effect) in response to mismatching compared with matching gestures is larger in clear than in degraded speech, which indicated that listeners are more hindered when integrating gestures with degraded speech (Drijvers & Özyürek, 2018). These results suggest that, when speech is degraded, the mismatching gesture cannot aid to disambiguate the remaining auditory cues to facilitate speech comprehension and integration load is lowest as it is not possible. However, it is unknown what neural mechanisms underlie speech–gesture integration in clear and adverse listening conditions, and it is unknown how semantic integration occurs when the integration load is manipulated by auditory factors (e.g., speech degradation) and visual semantic factors (e.g., congruency of gestures). Therefore, the current study aims to get insight in what oscillatory mechanisms support the semantic integration of speech and gestures in both clear and degraded speech.

Studies on unimodal degraded speech processing have consistently demonstrated less suppressed alpha power as a function of speech intelligibility (i.e., enhanced alpha power in response to degraded speech), which has been interpreted as possibly reflecting the allocation of resources and the functional inhibition of task-irrelevant neural activity during speech comprehension in challenging listening situations. This might be due to a higher auditory cognitive load when language processing is inhibited because of speech degradation (Drijvers, Mulder, & Ernestus, 2016; Wilsch, Henry, Herrmann, Maess, & Obleser, 2015; Strauß, Wöstmann, & Obleser, 2014; Weisz & Obleser, 2014; Becker, Pefkou, Michel, & Hervais-Adelman, 2013; Obleser & Weisz, 2012; Obleser, Wöstmann, Hellbernd, Wilsch, & Maess, 2012). During audiovisual processing of speech in noise, other work has revealed that beta power localized in the STS was less suppressed in high noise compared with no or low noise, possibly reflecting disturbed or altered audiovisual speech processing (Schepers et al., 2013). The abovementioned studies, however, do not include a visual semantic component, such as iconic co-speech gestures. In the visual domain, previous research on speech–gesture integration has identified that, during gestural enhancement of degraded speech comprehension, low- and high-frequency oscillatory power modulations in the LIFG and left temporal, motor, and visual regions predicted a listener's benefit from gestures during degraded speech comprehension (Drijvers et al., 2018). However, it is unknown how oscillatory activity supports speech–gesture integration when this integration is modulated by auditory (speech degradation) and visual semantic (gesture congruency) factors. The spatiotemporal characteristics of this integration process are needed to reveal which brain areas are engaged and disengaged in this process over time, such as when integration load is increased and a gesture mismatches rather than matches clear speech and also when integration load is lowest as it is not possible to integrate the two inputs, such as when a gesture mismatches rather than matches degraded speech.

Using MEG, we investigated the spatiotemporal oscillatory neuronal dynamics underlying audiovisual integration in a multimodal semantic context. Participants were presented with videos of an actress uttering an action verb in clear or degraded speech, accompanied by a matching or mismatching gesture, following the design of Drijvers and Özyürek (2018). On the basis of the oscillatory modulations that we observed in Drijvers et al. (2018), we expected that the neural integration of speech and gesture relies on an extended network, involving the language network (including LIFG/posterior STS [pSTS]/middle temporal gyrus [MTG]), the motor cortex, and the visual cortex. In line with the functional inhibition notion, our general hypothesis was that a relative decrease of alpha and beta power would reflect engagement of task-relevant brain regions, whereas enhanced alpha and beta power would reflect areas that do not need to be engaged for the task at hand or are less engaged in one condition compared with another condition (Jensen & Mazaheri, 2010; Klimesch et al., 2007). In clear speech, we thus expected that, when visual semantic congruency would increase integration load (i.e., when a gesture would mismatch rather than match the clear speech; see Drijvers & Özyürek, 2018), alpha and beta power would be more suppressed for mismatching compared with matching gestures. We expect that this larger suppression would occur in the language network, as well as the visual and motor cortices, reflecting increased visual attention to mismatching compared with matching gestures (Drijvers et al., 2018; Stothart & Kazanina, 2013 [for nonsemantic input]), a larger engagement of the motor system during observation of mismatching compared with matching gestures (Kilner, Marchant, & Frith, 2009; Koelewijn, van Schie, Bekkering, Oostenveld, & Jensen, 2008; Caetano, Jousmäki, & Hari, 2007), and a higher semantic unification load (Drijvers & Özyürek, 2018; Drijvers et al., 2018). This higher semantic unification load then occurs because the mismatching semantic information of the gesture is harder to integrate with clear speech than matching semantic information (Drijvers & Özyürek, 2018; Wang et al., 2012). Although we thus expect that mismatching gestures increase integration load in clear speech, we expect that, in degraded speech, mismatching gestures result in the lowest integration. In degraded speech, the gestural information cannot be coupled to the remaining auditory cues in the degraded speech signal, which would hinder integration (Drijvers

& Özyürek, 2018). This is opposed to matching gestures in degraded speech, which can enhance recognition of degraded speech (as, e.g., in Drijvers & Özyürek, 2017; Holle et al., 2010). Therefore, in degraded speech, we expect that alpha and beta power will be less suppressed when a gesture mismatches rather than matches degraded speech, reflecting less engagement of task-relevant brain regions during speech–gesture integration.

## METHODS

### Participants

Thirty-two right-handed Dutch native participants, recruited from Radboud University (mean age = 23.2 years, 14 men) participated in this experiment. All participants had normal hearing, normal or corrected-to-normal vision, no language, motor, or neurological disabilities, and gave written consent before participating in this experiment. Three participants (two women) were excluded because of technical failure (two) and excessive (head) motion artifacts (movement > 1 cm or >60% of the trials affected). The final data set included the data of 29 participants.

### Stimulus Materials

Participants were presented with 160 video clips that contained an actress who uttered a highly frequent action verb in clear or degraded speech, accompanied by a matching or mismatching iconic gesture. All of these video clips were pretested as part of the study by Drijvers and Özyürek (2017). To ensure that the verbs would fit with the gestures, we presented participants with the videos without their audio file and asked them to write down the verb they associated with the movement. We then showed participants the verb we originally matched the video with and asked them to indicate on a 7-point scale how much this verb fitted with the movement in the video. The results revealed a mean recognition rate of 59% over all gesture videos, which indicates that the gestures are potentially ambiguous without speech and thus might need speech for successful comprehension. Our rating task resulted in a mean score of "iconicity" of 6.1 (*SD* = 0.64), and all videos that scored under 5 on a 7-point scale were discarded.

In all videos that were used in this experiment (see Figure 1), the actress would always appear in the middle of the screen, where she was visible from her knees upward. She wore neutrally colored clothing and was standing in front of a dark blue neutral background. The gestures that she made were not instructed but made by her on the fly. The actress did not receive any feedback on the gestures she made. For the mismatching gestures, the experimenter would mention two verbs to the actress, of which the first one had to be the spoken verb and the second one had to be the to-be-gestured verb (e.g., "to drive" and "to mix," uttering "drive" while making a mixing gesture). This method was chosen as our stimuli show the face of the actress, and we could therefore not replace the audio track of the video with another verb's audio track, as the visible speech would be different. To determine which verbs were used as mismatching gestures, we divided the list of verbs in the mismatching condition in two separate lists and combined the verbs on the first list with the gesture that matched the verbs on the second list. In all videos, the preparation of this gesture (counted as the first frame where the actress moved her hand) was at 120 msec. At 550 msec, the stroke of the gesture would occur. Speech onset was at 680 msec, and the retraction of the gesture started at 1380 msec. The gesture offset was at 1780 msec

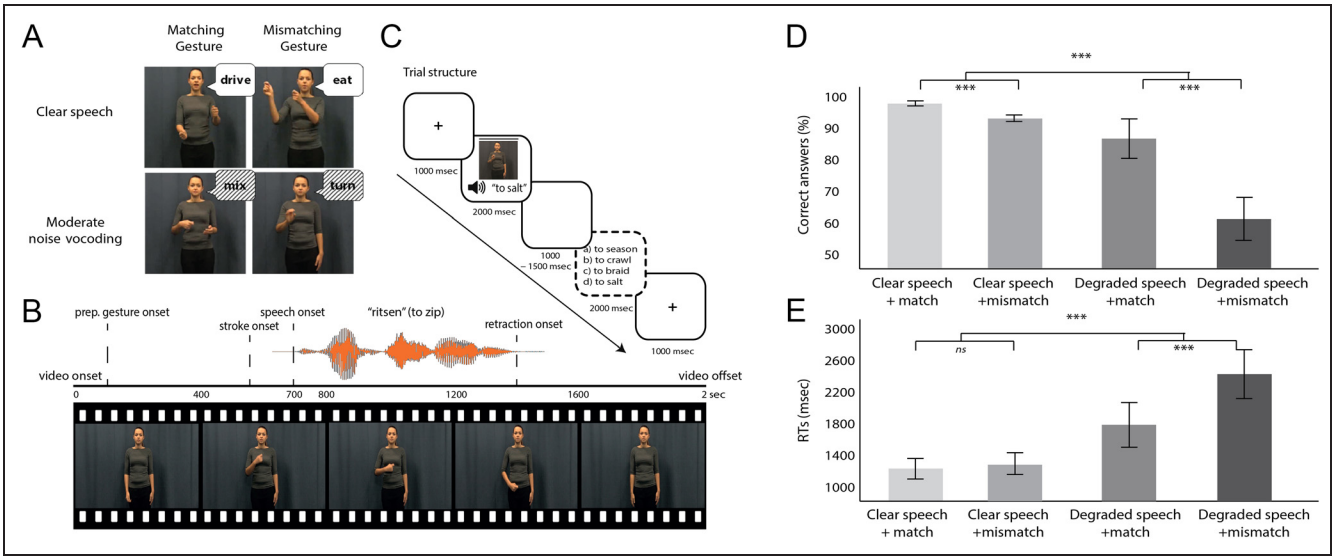


**Figure 1.** (A) Illustration of the different conditions and stimuli. (B) Illustration of the structure of the videos. (C) Structure of the trial. (D) Percentage of correct answers per condition. (E) RTs in milliseconds per condition. Error bars represent *SD*. ***$p < .01$.

(see Figure 1B). As speech onset was at 680 msec and stroke onset was at 550 msec, the overlap between the meaningful part of the gesture and the speech was optimal for mutual enhancement for comprehension (as previously demonstrated in Habets, Kita, Shao, Özyürek, & Hagoort, 2011).

All audio files were presented in clear speech or six-band noise-vocoded speech. This noise-vocoding level was chosen as previous work showed that, at a six-band noise-vocoding level, participants are most able to use gestural information for comprehension (Drijvers & Özyürek, 2017). From the video files, we extracted all audio tracks, denoised the speech, and intensity-scaled the speech to 70 dB by using Praat (Boersma & Weenink, 2006). After degrading 80 of the 160 sound files, all sound files were then recombined again with their corresponding video files, by using a custom-made script in Praat. To degrade the speech signals, we band-pass filtered each audio file between 50 and 8000 Hz and divided the speech signals in logarithmically spaced frequency bands with cutoff frequencies at 50, 116.5, 271.4, 632.5, 1473.6, 3433.5, and 8000 Hz. These frequencies were used to filter white noise to obtain the six bands. Subsequently, the amplitude envelope of these bands was extracted using half-wave rectification. We then multiplied this envelope with the noise bands and recombined the bands, resulting in the degraded signal (Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995). All sound was presented to participants through MEG-compatible air tubes.

The total experiment consisted of four conditions: a clear speech + matching gesture condition (CM), a degraded speech + matching gesture condition (DM), a clear speech + mismatching gesture condition (CMM), and a degraded speech and mismatching gesture condition (DMM). Each condition consisted of 40 items, of which none was repeated in any other condition (see Figure 1A).

### Procedure

Participants were placed in the 275-channel axial gradiometer CTF MEG system, at 70 cm from the projection screen on which the videos were presented. All videos were projected full-screen onto a semitranslucent screen by back projection using an EIKI LC-XL100L projector at a resolution of 1650 × 1080 pixels. The experiment was presented through Presentation software (Neurobehavioral Systems, Inc.). Each trial would start with a fixation cross (1000 msec), followed by a video (2000 msec) and a short delay period (1000–1500 msec, jittered), and ended with a cued-verb recall task in which participants had to identify which verb they just heard in the videos. The answer options in this task would always consist of a semantic competitor, a phonological competitor, an unrelated answer, and the correct answer. Participants had to indicate their choice by pressing a button with their right hand on a four-button box. After the participants had entered their response, a new trial would start after 1500 msec (see Figure 1C). All participants were presented with an individual pseudorandomization of the different videos that ensured none of the conditions would occur more than twice in a row (e.g., two consecutive trials that had degraded speech and a mismatching gesture). Participants were asked to sit as still as possible and not to blink during the videos, but after answering the cued recall task. We measured brain activity with MEG throughout the entire experiment. Participants were able to take a self-paced break per 40 trials.

### MEG Data Acquisition

Whole-head MEG was recorded at a sampling rate of 1200 Hz by using a 275-channel axial gradiometer MEG system. Participants wore recording markers on the nasion and left and right ear canal to monitor their head position in real time, using a MATLAB toolbox (Stolk, Todorovic, Schoffelen, & Oostenveld, 2013). During the breaks, this allowed us to readjust the participants' head position relative to the original position at the start of the experiment if the deviation was larger than 5 mm. We recorded electrocardiogram as well as horizontal and vertical EOGs for artifact rejection purposes. After the experiment, we invited the participants to record a structural MRI of their brain, using a 1.5-T Siemens Magnetom Avanto system with markers attached in the same position as the head coils, to allow us to align the structural anatomy of the participants with the MEG coordinate system. We collected structural MRIs for 22 of 32 participants.

### MEG Data Analysis

All data in this experiment were analyzed by using FieldTrip (Oostenveld, Fries, Maris, & Schoffelen, 2011), an open-source MATLAB toolbox, and custom MATLAB scripts. We preprocessed the data by dividing the data in epochs from −1 sec before video onset until 3 sec after video onset. All data were demeaned and detrended, and line noise was attenuated by using a discrete Fourier transform approach at 50 Hz and its subsequent harmonics. In total, we rejected, on average, ~3 trials per condition, which were contaminated by SQUID jump artifacts and muscle artifacts by using a semiautomatic routine. We then applied independent component analysis to remove all remaining eye movements and cardiac-related activity (Jung et al., 2000; Bell & Sejnowski, 1995). Finally, we went through all single trials and removed any artifacts that were not identified by using independent component analysis or other rejection procedures. We then resampled the data to 300 Hz to speed up analyses.

We computed an approximation of the planar gradient by converting the axial gradiometer data to orthogonal planar gradiometer pairs and computed and summed the power of the pairs. This approach might facilitate the interpretation of the MEG data, as planar gradient maxima are

known to be located above the neural sources that might underlie an effect (Bastiaansen & Knösche, 2000).

### Time–frequency Analyses

Our frequencies of interest ranged from 2 to 30 Hz, in frequency steps of 1 Hz. We applied a 500-msec Hanning window in 50-msec time steps (Mitra & Pesaran, 1999). To calculate the differences between conditions, we compared oscillatory power by averaging the four conditions separately for each participant. Time–frequency representations (TFRs) were log10 transformed, and the difference between the conditions was calculated by subtracting the log10 transformed power [= "log ratio," e.g., log10(A) − log10(B) or log10(CMM) − log10(CM) and log10(DMM) − log10(DM)]. The time window of analysis was always between 0.7 and 2.0 sec, which corresponds to speech onset until video offset.

### Source Analyses

To estimate the sources of our observed effects, we used dynamic imaging of coherent sources (Gross et al., 2001) as a beamforming spatial filtering technique. For this part of the analysis, the axial gradiometer data were used. First, the algorithm computed a common spatial filter from the cross-spectral density (CSD) matrix of the data and a lead field. For all frequency ranges of interest, we used a single Hanning taper. All lead fields of the participants were constructed by using a realistically shaped single-shell head model based on the participants' own individual anatomical data and by identifying the anatomical markers at the nasion and the two ear canals. Each volume was then divided into a 10-mm spaced grid of points and warped to the Montreal Neurological Institute brain template, where the lead field was calculated for each grid point.

The time windows that were used as input for the source analysis were based on the results from the sensor analysis. For the alpha band, we calculated the cross-spectral density between 1.3 and 2.0 sec at 10 Hz, with 2-Hz frequency smoothing. For the beta band, we computed the cross-spectral density between 1.3 and 2.0 sec, centered at 18 Hz with 4-Hz frequency smoothing. We used a common spatial filter containing all of the conditions to project the data through, separately per condition. We then averaged over trials, log10-transformed the data, and calculated the difference between conditions by subtracting the log power for the single contrasts. Finally, the grand-average grid of all participants was interpolated onto the Montreal Neurological Institute template for visualization purposes. Note that we included all trials in our sensor and source level analyses and did not differentiate between correct and incorrect trials, as the cued recall task might have masked the actual comprehension participants might have had when they were watching and listening to the video.

### Cluster-based Permutation Statistics

We performed nonparametric cluster-based permutation tests (Maris & Oostenveld, 2007) across participants to statistically quantify differences between the different conditions in power on source and sensor levels. We used the sensor level statistics to create statistical threshold masks to localize the observed effects on source level. We computed the mean difference between two conditions (e.g., CMM vs. CM, or DMM vs. DM) for each *x/y/z/* sample of our data set in the frequency ranges (alpha: 8–12 Hz, beta: 15–20 Hz) and time window (0.7–2.0 sec, i.e., from speech onset until the end of the video) we defined a priori and on the basis of a grand-average TFR of all conditions combined. After collecting all of the difference values of these comparisons (e.g., CMM vs. CM, or DMM vs. DM), all values where thresholded with the 95th percentile of the entire distribution. The remaining values formed the cluster candidates. All conditions and their corresponding values were randomly reassigned 5,000 times to form the permutation distribution. Out of this distribution, the cluster candidate who had the highest sum of the difference values was added to the permutation distribution. Finally, the actual observed cluster-level summed values were compared against this distribution, and all clusters that fell in the highest or lowest 2.5% were considered significant.

## RESULTS

We presented participants with videos that showed an actress uttering a Dutch action verb, while she simultaneously made a matching or mismatching gesture. Subsequently, participants had to indicate which verb they heard by a button press. Brain activity was recorded by MEG throughout the whole trial, but we focused on the time window from speech onset (0.7 sec) until the end of the video.

### Behavioral Results

A repeated-measures ANOVA with the factors Gesture (matching/mismatching) and Noise (clear/degraded) revealed that, when speech was clear, participants were more able to identify a correct answer on the cued-verb recall task than when speech was degraded, $F(1, 28) = 94.97$, $p < .001$, $\eta^2 = .77$. Similarly, participants found it easier to identify a word when a gesture matched rather than mismatched the speech signal, $F(1, 28) = 72.77, p < .001, \eta^2 = .72$ (see Figure 1D). An interaction effect between Gesture and Noise confirmed that the difference in correct answers when comparing mismatching with matching gestures was larger in degraded speech than in clear speech, $F(1, 28) = 58.45$. $p < .001$, $\eta^2 = .68$ (CM: $M = 97.2\%$, $SD = 1.6\%$; CMM: $M = 92.8\%$, $SD = 2.1\%$; DM: $M = 85.6\%$, $SD = 12.1\%$; DMM: $M = 61.4\%$, $SD = 11.2\%$). Post hoc $t$ tests on the relevant contrasts confirmed that

participants were more able to correctly identify the verb when the verb was accompanied by a matching compared with a mismatching gesture (clear speech: $t(28) = -3.09$, $p < .01$; degraded speech: $t(28) = -8.42$, $p < .001$). We did not observe any reliable differences in the amount of semantic or phonological competitors.

A second repeated-measures ANOVA using the same factors revealed a similar pattern for the RTs as for the correct answers: Participants were quicker to answer when speech was clear compared with degraded, $F(1, 28) = 143.63$, $p < .001$, $\eta^2 = .84$, and when a gesture matched rather than mismatched the speech signal, $F(1, 28) = 59.90$, $p < .001$, $\eta^2 = .68$ (see Figure 1E). The difference in RTs when comparing mismatching with matching gestures was larger in degraded speech than in clear speech, $F(1, 28) = 46.40$, $p < .001$, $\eta^2 = .62$ (CM: $M = 1269.1$, $SD = 360.3$; CMM: $M = 1299.8$, $SD = 378.0$; DM: $M = 1849.9$, $SD = 578.5$; DMM: $M = 2492.4$, $SD = 673.5$). Post hoc $t$ tests on the relevant contrasts confirmed that participants were not quicker to identify the verb when the verb was accompanied by a matching compared with a mismatching gesture in clear speech, $t(28) = 0.82$, $p = .41$, but were quicker to identify the verb when the verb was accompanied by a matching compared with a mismatching gesture in degraded speech, $t(28) = 8.02$, $p < .001$. These behavioral results reveal that gesture facilitates comprehension of degraded speech when the actress made a matching gesture but hindered comprehension when she performed a mismatching gesture.

## Semantic Congruency Effects in Clear Speech

### *Alpha and Beta Power Are More Suppressed When a Gesture Mismatches than Matches Clear Speech*

We first conducted a sensor level analysis over the full time window (0.7–2.0 sec, from speech onset until video offset) to identify differences in oscillatory power between the conditions. We calculated the TFRs of power for the individual trials and averaged them per condition. For TFRs of the single conditions, please see Figure S1A. Figure 2A represents the TFRs of power in response to the contrast CMM versus CM between 2 and 30 Hz, at representative left temporal sensors, based on the topographical plots that visualize this effect in time and space (see Figure 2B). Sensor level analyses confirmed a larger alpha and beta power suppression over left temporal, motor, and occipital areas when speech was clear and a gesture mismatched rather than matched the speech signal (alpha: one negative cluster, $p = .04$, 1.3–2.0 sec; beta: one negative cluster, $p < .01$, 1.3–2.0 sec), suggesting engagement of these areas in response to the mismatching gesture.

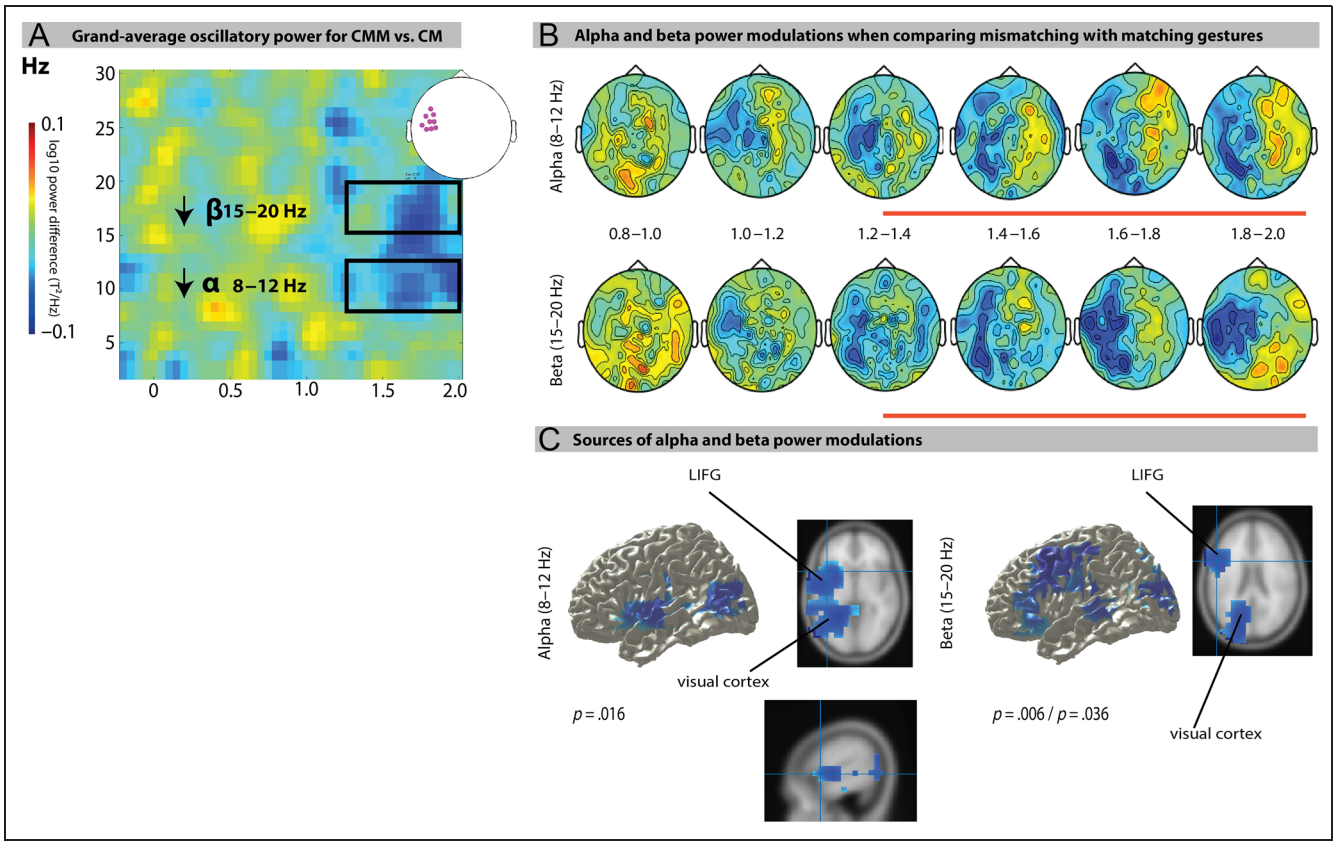


**Figure 2.** (A) TFRs of power of the contrast between CMM versus CM. (B) Topographical distribution of alpha (top) and beta (bottom) power of the contrast CMM versus CM in 200-msec time bins. Orange bars denote significant clusters in the sensor level analyses. (C) Estimated source results of the contrast in the alpha (left) and beta (right) bands, masked by statistically significant clusters.

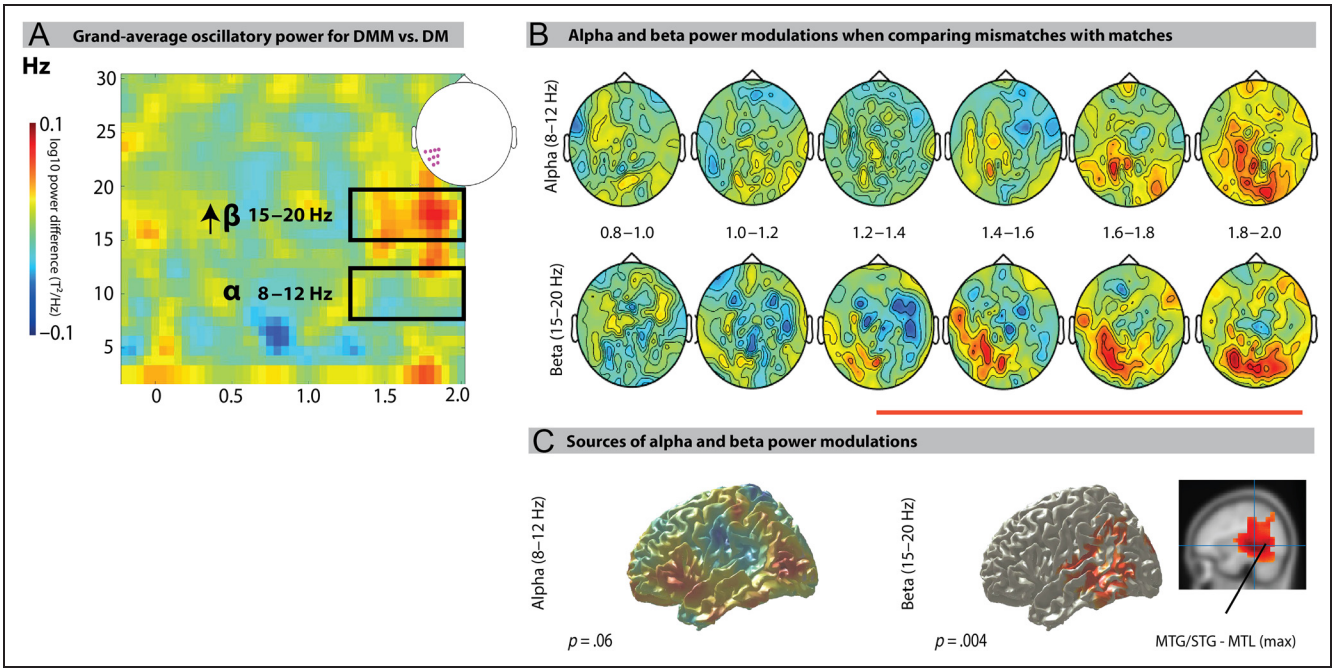


**Figure 3.** (A) TFRs of power of the contrast between CMM versus CM gesture. (B) Topographical distribution of alpha (top) and beta (bottom) power of the contrast DMM versus DM in 200-msec time bins in our time window of interest. Orange bars denote significant clusters in sensor level analyses. (C) Estimated source of the contrast in the alpha (left) and beta (right) bands, masked by statistically significant clusters. Note that, in the beta band, this effect was not statistically significant, but the estimated sources of the difference are included for visualization purposes.

### *Alpha Power Is More Suppressed in LIFG, Left Insula, and Visual Cortex When a Gesture Mismatches than Matches Clear Speech*

We used the time window of the significant clusters from the sensor analyses as input for our source analyses to estimate the sources of the alpha power modulation. Note that the statistical assessment was based on the sensor analysis, not the source level analysis.

Nevertheless, we applied a cluster randomization approach to the source data to find a threshold for when to consider the source estimates reliable. To investigate these underlying sources, we used a frequency-domain spatial beamformer technique (dynamic imaging of coherent sources; Gross et al., 2001). This analysis revealed that the source of the larger alpha power suppression in response to mismatching compared with matching gestures was localized in a widespread cluster including the LIFG, left insula, and visual cortex (one negative cluster, $p = .04$). These results thus suggest engagement of the extended language network when a gesture mismatches clear speech.

### *Beta Power Is More Suppressed in Motor and Visual Regions and LIFG When a Gesture Mismatches Rather than Matches Clear Speech*

We then localized the sources of the sensor-level power difference in the beta band. We localized the beta power difference in the left precentral and postcentral gyrus, the left frontal midline/SMA, LIFG, and the visual cortex (two negative clusters: $p \leq .01$ and $p \leq .04$). In line with our hypotheses and earlier work (Drijvers et al., 2018), this larger beta power suppression over the motor cortex shows that listeners might engage their motor cortex more when a gesture mismatches rather than matches the clear speech signal.

## Semantic Congruency Effects in Degraded Speech

### *Beta Power Is More Enhanced When a Gesture Mismatches than Matches Degraded Speech*

Next, we investigated whether a similar pattern of oscillatory power modulations would emerge when we compared the same conditions in degraded instead of clear speech. For TFRs of the single conditions, please see Figure S1B. The TFR in Figure 3A suggests enhanced beta power but no differences in alpha power. We plotted the topographical distribution of the contrast between mismatching and matching gestures in both frequency bands (see Figure 3B). We found no difference in alpha band power when comparing matching and mismatching gestures in degraded speech (no significant clusters, $p = .06$; see Figure 3) but found a larger beta power over left temporoparietal areas when a gesture mismatched degraded speech (one positive cluster, $p < .001$; Figure 3B and C). Because of the lack of an alpha power difference in DMM versus DM, the difference in CMM versus CM was greater than the difference in alpha power in DMM versus DM (one positive cluster, $p = .012$). The difference in beta power in CMM versus CM was larger than in DMM versus DM (one positive cluster, $p = .004$).

*Enhanced Beta Power Inhibits STS and Medial Temporal Lobe When a Gesture Mismatches Degraded Speech*

We localized the enhanced beta power in response to mismatching compared with matching gestures in degraded speech in the left auditory cortex, STS, MTG, and medial temporal lobe (MTL; one positive cluster, $p < .01$).

## DISCUSSION

We investigated how oscillatory dynamics support the semantic integration of speech and gestures in clear and degraded speech and what the spatiotemporal dynamics are that are associated with speech–gesture integration. We manipulated semantic integration load by presenting participants with videos of an actress who uttered an action verb in clear or degraded speech, accompanied by a matching or mismatching gesture. Our behavioral results demonstrated a semantic congruency effect and a speech degradation effect on performance; participants were slower and less able to correctly identify the verb when gestures mismatched speech and when speech was degraded. These results replicate previous findings and underline the additive effect of speech degradation (e.g., Holle et al., 2010) and semantic congruency between speech and gestures (e.g., Drijvers & Özyürek, 2018; Özyürek, Willems, Kita, & Hagoort, 2007; Willems et al., 2007, 2009; Kelly et al., 2004) on integration load and, subsequently, behavioral performance.

Our neurophysiological results demonstrate that semantic congruency and speech degradation modulated oscillatory activity in the alpha and beta bands. When speech was clear, we observed a larger alpha power suppression over the LIFG and visual cortex and a beta suppression over the LIFG, (pre)motor cortex, and visual cortex when a gesture mismatched rather than matched speech. When speech was degraded, we observed no difference in alpha power when comparing degraded speech and a mismatching gesture with a matching gesture. However, we did observe enhanced beta power over pSTS when a gesture mismatched rather than matched degraded speech. In both the alpha and beta bands, we observed a larger difference between mismatching and matching gestures in clear than degraded speech, suggesting that integration load was lowest in degraded speech (in line with Drijvers & Özyürek, 2018).

### Alpha/Beta Power Is More Suppressed over Visual Cortex to Allow for Increased Visual Attention to Mismatching Compared with Matching Gestures during Clear Speech

Both alpha and beta power were more suppressed over visual regions when a gesture mismatched rather than matched clear speech. This effect occurred from when the meaningful part of the gesture and speech were unfolding until the end of the video (1.3–2.0 sec). The larger alpha/beta suppression over visual regions suggests that the visual system is more engaged when a listener observes a mismatching gesture than a matching gesture and that more visual attention is allocated to a mismatching gesture compared with a matching gesture. We suggest that, when all auditory cues are still intact, a mismatching gesture will generate a larger mismatch response, causing increased visual attention to these mismatching gestures compared with matching gestures as a result of the detection of mismatching semantic information. Similar results have been found by Stothart and Kazanina (2013), who reported a poststimulus alpha suppression for deviant visual stimuli, potentially reflecting a shift in attentional resources after the detection of change. In line with this, we suggest that this sustained poststimulus alpha power suppression might reflect a shift in visual attentional resources to the gesture after the detection of mismatching information. Note that the loci of the clusters in the alpha and beta bands slightly seem to differ: The maximum of the cluster in the beta band can be localized to BA 18, whereas the maximum of the alpha cluster is estimated in BA 19. This suggests that the observed beta effect is not simply a harmonic of the observed alpha activity.

### Alpha/Beta Power Is More Suppressed over LIFG Due to Increased Semantic Unification Load in Clear Speech

We observed a larger suppression of alpha and beta power when a gesture mismatched rather than matched clear speech. This larger suppression in response to a mismatching gesture was localized in the LIFG. Previous studies have proposed that the LIFG is sensitive to unification operations from units that are retrieved from memory, the unification of information from different modalities, and lexical access operations (Hagoort, 2013). For example, in a study on unimodal speech comprehension, sentences with incongruent sentence endings yielded larger beta power over the LIFG. This was interpreted as reflective of a higher semantic unification load that was evoked by the incongruent sentence endings, which required a stronger engagement of the task-relevant brain network (Wang et al., 2012). Similarly, we demonstrated in a previous study that alpha/beta power is more suppressed in the LIFG when integration load increases (Drijvers et al., 2018). In line with this work, we suggest that the larger alpha and beta power suppression over the LIFG is reflective of the larger engagement that is required from the LIFG when a mismatching gesture increases semantic unification load to resolve the mismatch between the auditory information and the visual semantic information.

Note that previous studies (e.g., He et al., 2015; Dick et al., 2014; Straube, Green, Bromberger, & Kircher, 2011; Holle et al., 2010; Green et al., 2009; Willems

et al., 2009) have discussed the role of the LIFG in speech–gesture integration and that earlier work has argued that the LIFG is sensitive to the semantic relationship of speech–gesture pairs when a new and unified representation of the gestural input and speech needs to be constructed (Willems et al., 2009), which is the case for incongruent gestures. This interpretation was later extended by Holle et al. (2010) who argued that LIFG activity reflects a modulation or revision of the integrated speech–gesture information. This interpretation partially fits our findings. When post hoc visualizing the oscillatory modulations in the single conditions, we observed that the LIFG revealed suppressed activity compared with baseline in all conditions. The contrast between the mismatching and matching gestures thus shows that this suppression is larger when a mismatching gesture is paired with clear speech than when a matching gesture is paired with clear speech. However, the unification of these movements with the speech signal involved engagement of the LIFG in all single conditions (i.e., degraded speech + matching gesture, degraded speech + mismatching gesture, clear speech + matching gesture, and clear speech + mismatching gesture). This suggests that the observed larger alpha/beta suppression over the LIFG might reflect the increased semantic processing load that is imposed by the mismatching gesture but that the LIFG is engaged in all single conditions to unify the gesture with the speech signal, in line with the results we observed in Drijvers et al. (2018), where we studied gestural enhancement of degraded speech comprehension and where semantic congruency was not manipulated. This suggests that the LIFG possibly has a unifying function of the different inputs irrespective of congruency but that an increased integration load also increases engagement of the LIFG to unify the inputs.

### Motor Beta Suppression Reveals Stronger Simulation of Mismatching Gesture in Clear Speech

Beta power was more suppressed over the precentral cortex and SMA when a gesture mismatched rather than matched clear speech. This effect occurred in a similar time window as the alpha modulation (1.3–2.0 sec) and lasted from when the speech and gesture were unfolding until the end of the video. The larger suppression for mismatching compared with matching gestures suggests that engagement of the motor system is modulated by the semantic fit of the information that is conveyed by the gestures, which is in line with previous studies on action observation (e.g., Schaller, Weiss, & Müller, 2017; Klepp, Niccolai, Buccino, Schnitzler, & Biermann-Ruben, 2015; Weiss & Mueller, 2012; van Elk, van Schie, Zwaan, & Bekkering, 2010). We interpret this effect as showing that listeners more strongly engage their motor system to "simulate" the mismatching gesture to reevaluate whether it fits with the processed speech signal. Note that we did not observe a similar effect when speech was degraded. This suggests that, when speech is degraded, matching and mismatching gestures are simulated equally when auditory cues are not reliable and a reevaluation of the fit of the gesture is hindered. This would be in line with current and previous works that suggest that integration load is lowest when speech is degraded (Drijvers & Özyürek, 2017).

### Enhanced Beta Power over STS/MTG and MTL Reveals Hindered Semantic Integration and Lexical Retrieval When Gestures Mismatch Degraded Speech

When speech was degraded, we did not observe reliable differences in alpha power when comparing mismatching and matching gestures. However, beta power was less suppressed in response to mismatching compared with matching gestures (1.3–2.0 sec) when speech was degraded. This is in line with previous work on nonsemantic audiovisual speech processing, which demonstrated a similar smaller beta suppression in noisy speech when comparing audiovisual with audio-only conditions. This effect was localized to the STS (Schepers et al., 2013). This underlines that modulations of oscillatory activity in the STS play a role in audiovisual speech processing under clear and adverse listening conditions. Previous studies have shown that suppressed beta band activity plays a role in tasks where information from different modalities needs to be integrated (Kopell, Kramer, Malerba, & Whittington, 2010) and in naturalistic audiovisual processing (Kayser & Logothetis, 2009). When speech is degraded and the semantic information that is conveyed by the gesture cannot be matched to the degraded auditory cues, pSTS/MTG might be less engaged because of the hindered audiovisual integration. Similarly, as the meaningful information from a mismatching gesture will not aid in resolving the degraded speech signal, lexical retrieval might be hindered (Hannemann, Obleser, & Eulitz, 2007), which is demonstrated by less involvement of the MTL when listeners process mismatching as compared with matching gestures.

Our current results also contribute to recent discussions over the role and involvement of pSTS/MTG and LIFG in speech–gesture integration. Although the role of pSTS and LIFG has been discussed, MTG has often been found to be involved in speech–gesture integration. Some studies have shown that MTG is modulated by semantic congruency (Dick, Goldin-Meadow, Solodkin, & Small, 2012; Green et al., 2009, see Özyürek, 2014) and activity in the MTG has been linked to coupling sound and meaning. However, the role of (p)STS has been debated. Some studies have argued that STS is sensitive to semantic aspects of speech–gesture integration. For example, in an fMRI study, stronger activation for ambiguous words that were paired with iconic (i.e., semantic) compared with grooming (nonsemantic) gestures was observed (Holle, Gunter, Rüschemeyer, Hennenlotter, & Iacoboni, 2008). Moreover, a larger involvement of

the (p)STS has been reported in response to congruent iconic gestures coupled with degraded speech compared with clear speech (Holle et al., 2010), but not when comparing complementary versus redundant gestures (Dick et al., 2012). Other studies have argued that pSTS is mostly involved in the mapping and coupling of lower-level audiovisual information, which might already have a stable common object representation, but not to semantic congruency in speech–gesture integration (e.g., Dick et al., 2012; Willems et al., 2007, 2009). Similarly, studies on audiovisual integration (e.g., lips and speech) have suggested that the STS might be related to associating the auditory and visual modalities at a lower-level stage of multimodal matching (e.g., Beauchamp, 2005; Beauchamp, Argall, Bodurka, Duyn, & Martin, 2004; Beauchamp, Lee, Argall, & Martin, 2004; Callan et al., 2004; Calvert, 2001). Our current results suggest that indeed pSTS is sensitive to hindered audiovisual integration but that this is not solely caused by semantic congruency. Note that, although we observed a difference in oscillatory power in pSTS/MTG in degraded speech when comparing mismatching with matching gestures, we did not observe a modulation of oscillatory activity in pSTS/MTG when speech was clear. This suggests that pSTS/MTG is less engaged when speech is degraded. This might occur because integration processes are hindered when the visual semantic information cannot help to retrieve or disambiguate the degraded lexical item, which increases integration load. We thus tentatively propose that the LIFG and pSTS indeed work together to integrate speech and gestures (cf. Willems et al., 2009) but that the role of LIFG is not solely modulatory or revising in nature (see, e.g., Holle et al., 2010; Willems et al., 2009). Instead, the LIFG unifies higher-level semantic information from multiple inputs, irrespective of whether a stable common representation exists on which the input streams can be mapped (see Holle et al., 2010; Willems et al., 2009). However, when speech is degraded and a gesture mismatched speech, integration load was lowest when the gesture could not be integrated and disambiguate the degraded signal, resulting in less engagement from MTL and lower-level areas such as the pSTS/MTG.

## Conclusion

The present work is the first study that investigated how oscillatory modulations can inform us about the processes underlying speech–gesture integration in clear and degraded speech as well as what the spatiotemporal dynamics are that are associated with this process. We set out to investigate how the semantic integration of speech and gestures is supported when integration load is manipulated by auditory (e.g., degraded speech) and visual (e.g., gesture congruency) factors. Our results provide novel insight by revealing how low-frequency oscillations support semantic audiovisual integration in clear and degraded speech: When gestures mismatch clear speech, listeners engage the LIFG and motor and visual regions when semantic unification load increases because of the gesture. When speech is degraded, pSTS/MTG and MTL are less engaged, possibly reflecting the hindered integration of gestures and the degraded signal when the gesture does not disambiguate the degraded speech or aid lexical retrieval. Our results thus reveal that low-frequency oscillatory modulations can index congruency between speech and gestures in a semantic context and demonstrate that low-frequency power modulations do support not only nonsemantic audiovisual integration but also semantic integration. This suggests a domain-general mechanistic role of brain oscillations in enabling integration of different modalities and engaging/inhibiting brain areas that do not contribute to this integration process.

## Acknowledgments

This work was supported by Gravitation grant 024.001.006 of the Language in Interaction Consortium from Netherlands Organization for Scientific Research. O. J. was supported by James S. McDonnell Foundation Understanding Human Cognition Collaborative Award (220020448) and the Royal Society Wolfson Research Merit Award. We thank Nick Wood (†), for helping us in editing the video stimuli, and Gina Ginos, for being the actress in the videos.

Reprint requests should be sent to Linda Drijvers, Centre for Language Studies, Donders Institute for Brain, Cognition and Behaviour, Radboud University, Wundtlaan 1, 6525 XD Nijmegen, The Netherlands, or via e-mail: linda.drijvers@mpi.nl.

## REFERENCES

Bastiaansen, M. C. M., & Knösche, T. R. (2000). Tangential derivative mapping of axial MEG applied to event-related desynchronization research. *Clinical Neurophysiology, 111,* 1300–1305.

Beauchamp, M. S. (2005). See me, hear me, touch me: Multisensory integration in lateral occipital–temporal cortex. *Current Opinion in Neurobiology, 15,* 145–153.

Beauchamp, M. S., Argall, B. D., Bodurka, J., Duyn, J. H., & Martin, A. (2004). Unraveling multisensory integration: Patchy organization within human STS multisensory cortex. *Nature Neuroscience, 7,* 1190–1192.

Beauchamp, M. S., Lee, K. E., Argall, B. D., & Martin, A. (2004). Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron, 41,* 809–823.

Becker, R., Pefkou, M., Michel, C. M., & Hervais-Adelman, A. G. (2013). Left temporal alpha-band activity reflects single word intelligibility. *Frontiers in Systems Neuroscience, 7,* 121.

Bell, A. J., & Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Computation, 7,* 1129–1159.

Boersma, P., & Weenink, D. (2006). Praat (Version 4.5) [Computer software]. Amsterdam: Institute of Phonetic Sciences.

Caetano, G., Jousmäki, V., & Hari, R. (2007). Actor's and observer's primary motor cortices stabilize similarly after seen or heard motor actions. *Proceedings of the National Academy of Sciences, U.S.A., 104,* 9058–9062.

Callan, D. E., Jones, J. A., Munhall, K., Kroos, C., Callan, A. M., & Vatikiotis-Bateson, E. (2004). Multisensory integration sites identified by perception of spatial wavelet filtered visual speech gesture information. *Journal of Cognitive Neuroscience, 16,* 805–816.

Calvert, G. A. (2001). Crossmodal processing in the human brain: Insights from functional neuroimaging studies. *Cerebral Cortex, 11,* 1110–1123.

Dick, A. S., Goldin-Meadow, S., Solodkin, A., & Small, S. L. (2012). Gestures in the developing brain. *Developmental Science, 15,* 165–180.

Dick, A. S., Mok, E. H., Raja Beharelle, A., Goldin-Meadow, S., & Small, S. L. (2014). Frontal and temporal contributions to understanding the iconic co-speech gestures that accompany speech. *Human Brain Mapping, 35,* 900–917.

Drijvers, L., Mulder, K., & Ernestus, M. (2016). Alpha and gamma band oscillations index differential processing of acoustically reduced and full forms. *Brain and Language, 153–154,* 27–37.

Drijvers, L., & Özyürek, A. (2017). Visual context enhanced: The joint contribution of iconic gestures and visible speech to degraded speech comprehension. *Journal of Speech, Language, and Hearing Research, 60,* 212–222.

Drijvers, L., & Özyürek, A. (2018). Native language status of the listener modulates the neural integration of speech and iconic gestures in clear and adverse listening conditions. *Brain and Language, 177–178,* 7–17.

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,* 2075–2087.

Green, A., Straube, B., Weis, S., Jansen, A., Willmes, K., Konrad, K., et al. (2009). Neural integration of iconic and unrelated coverbal gestures: A functional MRI study. *Human Brain Mapping, 30,* 3309–3324.

Gross, J., Kujala, J., Hämäläinen, M., Timmermann, L., Schnitzler, A., & Salmelin, R. (2001). Dynamic imaging of coherent sources: Studying neural interactions in the human brain. *Proceedings of the National Academy of Sciences, U.S.A., 98,* 694–699.

Habets, B., Kita, S., Shao, Z., Özyürek, A., & Hagoort, P. (2011). The role of synchrony and ambiguity in speech–gesture integration during comprehension. *Journal of Cognitive Neuroscience, 23,* 1845–1854.

Hagoort, P. (2013). MUC (Memory, Unification, Control) and beyond. *Frontiers in Psychology, 4,* 416.

Hannemann, R., Obleser, J., & Eulitz, C. (2007). Top–down knowledge supports the retrieval of lexical information from degraded speech. *Brain Research, 1153,* 134–143.

He, Y., Gebhardt, H., Steines, M., Sammer, G., Kircher, T., Nagels, A., et al. (2015). The EEG and fMRI signatures of neural integration: An investigation of meaningful gestures and corresponding speech. *Neuropsychologia, 72,* 27–42.

Hipp, J. F., Engel, A. K., & Siegel, M. (2011). Oscillatory synchronization in large-scale cortical networks predicts perception. *Neuron, 69,* 387–396.

Holle, H., Gunter, T. C., Rüschemeyer, S. A., Hennenlotter, A., & Iacoboni, M. (2008). Neural correlates of the processing of co-speech gestures. *Neuroimage, 39,* 2010–2024.

Holle, H., Obleser, J., Rueschemeyer, S.-A., & Gunter, T. C. (2010). Integration of iconic gestures and speech in left superior temporal areas boosts speech comprehension under adverse listening conditions. *Neuroimage, 49,* 875–884.

Jensen, O., & Mazaheri, A. (2010). Shaping functional architecture by oscillatory alpha activity: Gating by inhibition. *Frontiers in Human Neuroscience, 4,* 186.

Jung, T.-P. P., Makeig, S., Humphries, C., Lee, T.-W. W., McKeown, M. J., Iragui, V., et al. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology, 37,* 163–178.

Kayser, C., & Logothetis, N. K. (2009). Directed interaction between auditory and superior temporal cortices and their role in sensory integration. *Frontiers in Integrative Neuroscience, 3,* 7.

Kelly, S. D., Kravitz, C., & Hopkins, M. (2004). Neural correlates of bimodal speech and gesture comprehension. *Brain and Language, 89,* 253–260.

Kilner, J. M., Marchant, J. L., & Frith, C. D. (2009). Relationship between activity in human primary motor cortex during action observation and the mirror neuron system. *PLoS One, 4,* e4925.

Klepp, A., Niccolai, V., Buccino, G., Schnitzler, A., & Biermann-Ruben, K. (2015). Language-motor interference reflected in MEG beta oscillations. *Neuroimage, 109,* 438–448.

Klimesch, W., Sauseng, P., & Hanslmayr, S. (2007). EEG alpha oscillations: The inhibition-timing hypothesis. *Brain Research Reviews, 53,* 63–88.

Koelewijn, T., van Schie, H. T., Bekkering, H., Oostenveld, R., & Jensen, O. (2008). Motor-cortical beta oscillations are modulated by correctness of observed action. *Neuroimage, 40,* 767–775.

Kopell, N., Kramer, M. A., Malerba, P., & Whittington, M. A. (2010). Are different rhythms good for different functions? *Frontiers in Human Neuroscience, 4,* 187.

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

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago: Chicago University Press.

Mitra, P. P., & Pesaran, B. (1999). Analysis of dynamic brain imaging data. *Biophysical Journal, 76,* 691–708.

Obleser, J., & Weisz, N. (2012). Suppressed alpha oscillations predict intelligibility of speech and its acoustic details. *Cerebral Cortex, 22,* 2466–2477.

Obleser, J., Wöstmann, M., Hellbernd, N., Wilsch, A., & Maess, B. (2012). Adverse listening conditions and memory load drive a common alpha oscillatory network. *Journal of Neuroscience, 32,* 12376–12383.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience, 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, 369,* 20130296.

Özyürek, A., Willems, R. M., 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,* 605–616.

Pfurtscheller, G., & Lopes da Silva, F. H. (1999). Event-related EEG/MEG synchronization and desynchronization: Basic principles. *Clinical Neurophysiology, 110,* 1842–1857.

Schaller, F., Weiss, S., & Müller, H. M. (2017). EEG beta-power changes reflect motor involvement in abstract action language processing. *Brain and Language, 168,* 95–105.

Schepers, I. M., Schneider, T. R., Hipp, J. F., Engel, A. K., & Senkowski, D. (2013). Noise alters beta-band activity in superior temporal cortex during audiovisual speech processing. *Neuroimage, 70,* 101–112.

Schroeder, C. E., Lakatos, P., Kajikawa, Y., Partan, S., & Puce, A. (2008). Neuronal oscillations and visual amplification of speech. *Trends in Cognitive Sciences, 12,* 106–113.

Senkowski, D., Schneider, T. R., Foxe, J. J., & Engel, A. K. (2008). Crossmodal binding through neural coherence:

Implications for multisensory processing. *Trends in Neurosciences, 31,* 401–409.
Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science, 270,* 303–304.
Stolk, A., Todorovic, A., Schoffelen, J.-M., & Oostenveld, R. (2013). Online and offline tools for head movement compensation in MEG. *Neuroimage, 68,* 39–48.
Stothart, G., & Kazanina, N. (2013). Oscillatory characteristics of the visual mismatch negativity: What evoked potentials aren't telling us. *Frontiers in Human Neuroscience, 7,* 426.
Straube, B., Green, A., Bromberger, B., & Kircher, T. (2011). The differentiation of iconic and metaphoric gestures: Common and unique integration processes. *Human Brain Mapping, 32,* 520–533.
Straube, B., Green, A., Weis, S., & Kircher, T. (2012). A supramodal neural network for speech and gesture semantics: An fMRI study. *PLoS One, 7,* e51207.
Strauß, A., Wöstmann, M., & Obleser, J. (2014). Cortical alpha oscillations as a tool for auditory selective inhibition. *Frontiers in Human Neuroscience, 8,* 350.
van Elk, M., van Schie, H. T., Zwaan, R. A., & Bekkering, H. (2010). The functional role of motor activation in language processing: Motor cortical oscillations support lexical-semantic retrieval. *Neuroimage, 50,* 665–677.
Varela, F., Lachaux, J.-P., Rodriguez, E., & Martinerie, J. (2001). The brainweb: Phase synchronization and large-scale integration. *Nature Reviews Neuroscience, 2,* 229–239.
Wang, L., Jensen, O., van den Brink, D., Weder, N., Schoffelen, J. M., Magyari, L., et al. (2012). Beta oscillations relate to the N400m during language comprehension. *Human Brain Mapping, 33,* 2898–2912.
Weiss, S., & Mueller, H. M. (2012). "Too many betas do not spoil the broth": The role of beta brain oscillations in language processing. *Frontiers in Psychology, 3,* 201.
Weisz, N., & Obleser, J. (2014). Synchronisation signatures in the listening brain: A perspective from non-invasive neuroelectrophysiology. *Hearing Research, 307,* 16–28.
Willems, R. M., Özyürek, A., & Hagoort, P. (2007). When language meets action: The neural integration of gesture and speech. *Cerebral Cortex, 17,* 2322–2333.
Willems, R. M., Özyürek, A., & Hagoort, P. (2009). Differential roles for left inferior frontal and superior temporal cortex in multimodal integration of action and language. *Neuroimage, 47,* 1992–2004.
Wilsch, A., Henry, M. J., Herrmann, B., Maess, B., & Obleser, J. (2015). Alpha oscillatory dynamics index temporal expectation benefits in working memory. *Cerebral Cortex, 25,* 1938–1946.
Zhao, W., Riggs, K., Schindler, I., & Holle, H. (2018). Transcranial magnetic stimulation over left inferior frontal and posterior temporal cortex disrupts gesture-speech integration. *Journal of Neuroscience, 38,* 1891–1900.