NeuroImage 194 (2019) 55–67

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Native and non-native listeners show similar yet distinct oscillatory dynamics when using gestures to access speech in noise

Linda Drijvers [a,b,*], Mircea van der Plas [c], Asli Özyürek [a,b,d], Ole Jensen [c]

[a] Radboud University, Centre for Language Studies, Erasmusplein 1, 6525 HT, Nijmegen, the Netherlands
[b] Radboud University, Donders Institute for Brain, Cognition, and Behaviour, Montessorilaan 3, 6525 HR, Nijmegen, the Netherlands
[c] School of Psychology, Centre for Human Brain Health, University of Birmingham, Hills Building, Birmingham B15 2TT, United Kingdom
[d] Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525 XD, Nijmegen, the Netherlands

ARTICLE INFO

Keywords:
Degraded speech
Gesture
Non-native language comprehension
Magnetoencephalography
Multimodal integration
Oscillations
Semantics

ABSTRACT

Listeners are often challenged by adverse listening conditions during language comprehension induced by external factors, such as noise, but also internal factors, such as being a non-native listener. Visible cues, such as semantic information conveyed by iconic gestures, can enhance language comprehension in such situations. Using magnetoencephalography (MEG) we investigated whether spatiotemporal oscillatory dynamics can predict a listener's benefit of iconic gestures during language comprehension in both internally (non-native versus native listeners) and externally (clear/degraded speech) induced adverse listening conditions. Proficient non-native speakers of Dutch were presented with videos in which an actress uttered a degraded or clear verb, accompanied by a gesture or not, and completed a cued-recall task after every video. The behavioral and oscillatory results obtained from non-native listeners were compared to an MEG study where we presented the same stimuli to native listeners (Drijvers et al., 2018a). Non-native listeners demonstrated a similar gestural enhancement effect as native listeners, but overall scored significantly slower on the cued-recall task. In both native and non-native listeners, an alpha/beta power suppression revealed engagement of the extended language network, motor and visual regions during gestural enhancement of degraded speech comprehension, suggesting similar core processes that support unification and lexical access processes. An individual's alpha/beta power modulation predicted the gestural benefit a listener experienced during degraded speech comprehension. Importantly, however, non-native listeners showed less engagement of the mouth area of the primary somatosensory cortex, left insula (beta), LIFG and ATL (alpha) than native listeners, which suggests that non-native listeners might be hindered in processing the degraded phonological cues and coupling them to the semantic information conveyed by the gesture. Native and non-native listeners thus demonstrated similar yet distinct spatiotemporal oscillatory dynamics when recruiting visual cues to disambiguate degraded speech.

1. Introduction

Adverse listening conditions during language comprehension can be caused by external factors, such as noise (Peelle, 2018), but also internal factors, such as when understanding language as a non-native listener (Lecumberri et al., 2010). Especially under such adverse listening conditions, listeners can improve comprehension by integrating information from the auditory modality, such as speech, and visual modalities, such as visible speech and co-speech gestures. Brain oscillations are thought to have a mechanistic role in enabling the integration of information from these different auditory and visual modalities (Kayser and Logothetis, 2009; Schroeder et al., 2008; Senkowski et al., 2011; Varela et al., 2001). The engagement of brain areas that are relevant for this integration process is often thought to relate to a suppression of low-frequency oscillatory power in the alpha (8–12 Hz) and beta (13–30 Hz) band (Jensen and Mazaheri, 2010; Klimesch et al., 2007; Payne and Sekuler, 2014). Oscillatory power modulations have shown to be predictive of the degree of non-semantic (e.g., beeps and flashes (Hipp et al., 2011)), and semantic audiovisual integration of an ambiguous stimulus (e.g., speech degradation, (Drijvers et al., 2018a; 2018b)). Here, we investigate how brain oscillations support semantic audiovisual integration when listeners face adverse listening conditions induced by both internal factors

* Corresponding author. Radboud University, Centre for Language Studies, Donders Institute for Brain, Cognition and Behaviour, Wundtlaan 1, 6525 XD, Nijmegen, the Netherlands.
E-mail address: linda.drijvers@mpi.nl (L. Drijvers).

https://doi.org/10.1016/j.neuroimage.2019.03.032
Received 9 September 2018; Received in revised form 12 March 2019; Accepted 15 March 2019
Available online 21 March 2019

(i.e., non-nativeness) and external factors (i.e., speech degradation).

When listeners face adverse listening conditions induced by an external factor, such as speech degradation, studies on unimodal auditory degraded speech comprehension have demonstrated less suppressed alpha power when speech was degraded, possibly reflecting an increased auditory cognitive load when language processing is inhibited (Obleser and Weisz, 2012; Weisz et al., 2011; Wostmann et al., 2015). In multimodal adverse listening conditions, however, semantic information conveyed by iconic gestures has been shown to enhance language comprehension (Drijvers and Ozyürek, 2017; Holle et al., 2010). These iconic gestures (e.g., a 'mixing' gesture when describing a recipe) can convey semantic information that illustrates objects, actions or spatial relationships (McNeill, 1992) and are thought to be automatically integrated with speech (Kelly et al., 2010) on both a neural and behavioral level (see for an overview, Özyürek, 2014). Imaging studies relying on fMRI that investigated the spatial correlates of this process suggested that the semantic integration of speech involves left-inferior frontal gyrus (LIFG), posterior middle temporal gyrus (pMTG), superior temporal sulcus (STS), visual and motor regions (Dick et al., 2014; Green et al., 2009; Straube et al., 2012; Willems et al., 2009, 2007; Zhao et al., 2018). EEG studies on the temporal correlates of speech-gesture integration reported low-frequency oscillatory modulations to gestures that had both a semantic and non-semantic relation to speech (Biau et al., 2015; Biau and Soto-Faraco, 2015; He et al., 2018, 2015, 2011). In line with studies on non-semantic audiovisual integration (Hipp et al., 2011) and studies on the neural correlates of speech-gesture integration, we demonstrated in a previous MEG study that oscillatory power modulations in LIFG, left-temporal, motor and visual regions can predict how much a listener can benefit from gestures during degraded speech comprehension (Drijvers et al., 2018a). However, it is unknown whether similar oscillatory modulations can predict how much a listener can benefit from the semantic information conveyed by gestures when internal factors cause an adverse listening condition during language comprehension, such as when understanding language as a non-native listener.

When an internal factor, such as non-nativeness, impacts language comprehension, previous research demonstrated that semantic information conveyed by gestures can enhance language comprehension (Dahl and Ludvigsen, 2014; Sueyoshi and Hardison, 2005). However, in a recent EEG study that investigated how the N400 component was modulated by the semantic congruency of gestures in clear and degraded speech, an N400 effect for non-native listeners was observed only when speech was clear but not when speech was degraded. Thus, although non-native listeners seem to benefit from gestural enhancement during degraded speech comprehension, speech-gesture integration seems to be hindered for non-native listeners when speech is degraded (Drijvers and Özyürek, 2018). A potential explanation for these findings is that non-native listeners need more phonological cues to benefit from the semantic information that is conveyed by the gesture (Drijvers and Özyürek, 2019). This is in line with previous behavioral work that demonstrated that non-native listeners can only utilize auditory semantic-contextual cues for comprehension when the auditory signal is of sufficient quality to allow access to semantic cues (Bradlow and Alexander, 2007; Golestani et al., 2009; Hazan et al., 2006; Mayo et al., 1997; Oliver et al., 2012; Zhang et al., 2016). However, it is unknown which brain areas engage in this process over time, and how this differs from native listeners, who are not challenged by internally induced adverse listening conditions when understanding language.

The current paper investigates whether spatiotemporal oscillatory dynamics can predict how much a listener can benefit from semantic information conveyed by gestures in internally induced (i.e., non-nativeness) and externally induced adverse listening conditions (i.e., speech degradation). Using the same paradigm as in Drijvers et al. (2018a) where only external factors induced an adverse listening condition, we presented participants with videos of an actress who uttered an action verb in clear or degraded speech, while making a gesture or not. After watching the videos, the participants had to indicate which verb they heard in a cued-recall task. An internally induced adverse listening condition was created by testing highly proficient non-native speakers of Dutch with sufficient vocabulary knowledge of Dutch, as low-proficient participants would not recognize all verbs, or be focused solely on the gesture. An externally induced adverse listening condition was created by manipulating speech quality by noise-vocoding. We used the already acquired MEG data from native listeners (described in Drijvers et al., 2018a) to compare to the oscillatory activity observed in non-native listeners during semantic audiovisual integration.

On a behavioral level, we expected that non-native listeners would show a similar gestural enhancement effect as native listeners on the cued-recall task. However, we predicted that non-native listeners would overall be less accurate and slower than native listeners when answering what verb they heard in the videos. This would, in line with previous literature (Drijvers and Ozyürek, 2017; Drijvers and Özyürek, 2019), indicate that although non-native listeners benefit from gestures during degraded speech comprehension, they might be hindered in resolving the degraded auditory cues and coupling those cues to the semantic information that is conveyed by the gesture.

On a neural level, our central hypothesis was that a suppression of alpha (8–12 Hz) and beta power (15–20 Hz) would reflect engagement of brain regions that are relevant for comprehension during gestural enhancement of degraded speech comprehension. Based on what was observed in previous work on native listeners (Drijvers et al., 2018a), we predicted that for non-native listeners, gestural enhancement would rely on the engagement of the extended language network (LIFG, and left-temporal regions), motor and visual regions to perform this semantic audiovisual integration. This would be similar as what was observed for native listeners. More specifically, this would mean that as for native listeners, we expected that for non-native listeners a larger alpha power suppression in the extended language network would reflect stronger engagement of these regions when unification load is higher (Wang et al., 2012; Drijvers et al., 2018a, 2018b). We expected larger alpha power suppression over visual regions, reflecting a larger allocation of visual attention to gestures when speech is degraded. A larger beta power suppression over motor regions would reflect a larger engagement of these regions during gestural observation when speech is degraded (Caetano et al., 2007; Kilner et al., 2009; Koelewijn et al., 2008). Lastly, we expected that the observed oscillatory power modulations in non-native listeners would correlate with the benefit a non-native listener would experience during degraded speech comprehension, similar as to what was observed for native listeners (Drijvers et al., 2018a).

However, previous work suggested that non-native listeners might only be able to utilize semantic cues when the auditory signal is of sufficient quality to allow access to these semantic cues (Bradlow and Alexander, 2007; Golestani et al., 2009; Hazan et al., 2006; Mayo et al., 1997; Oliver et al., 2012; Zhang et al., 2016). Therefore, we conducted some exploratory analyses to investigate possible differences between the two groups. On the basis of our previous results (for native listeners in Drijvers et al., 2018a), we expected less engagement of the LIFG for non-native listeners compared to native listeners. This would reflect that when speech is degraded, it is more difficult for non-native listeners to unify the degraded auditory cues with the semantic information that is conveyed by the gesture.

2. Methods

2.1. Participants

The non-native listener group was formed by thirty-two right-handed German advanced learners of Dutch (mean age = 23.09, 15 males) who reported normal hearing, normal or corrected-to-normal vision, no language, motor or neurological impairments. All participants were students at Radboud University who were paid to participate in the study, and were recruited on the basis of the following criteria: They had lived or

studied in the Netherlands for at least 1 year, had to use Dutch regularly (minimally once per week) for their studies and/or their personal lives, and acquired Dutch after age 12. We excluded two participants due to unreported metal (1) in their bodies and left-handedness (1). The data of the non-native listener group was compared to the data of the native listener group (n = 30) reported in Drijvers et al. (2018a). All participants gave written consent before participation.

2.2. LexTALE assessment

As we aimed to recruit highly-proficient non-native German speakers of Dutch to introduce internal ambiguity, we assessed the proficiency level of our (potential) participants with the Dutch version of the Lexical Test for Advanced Learners of English (LexTALE), a vocabulary test using non-speeded visual lexical decision (Lemhöfer and Broersma, 2012). In this test, participants are presented with 60 words (40 Dutch words, 20 nonwords) of which they have to decide whether is a real word in Dutch or not. Nonwords are constructed of strings created by either changing a few letters in a real Dutch word, or by recombining existing Dutch morphemes. As we were aiming for an intermediate-high proficiency level of our participants, we only included participants who scored at a B2 level or higher (above 67.5%). After the experiment, we used an adapted version of the LexTALE test consisting of 40 verbs that were used in the experiment, and 20 non-words that were constructed on the basis of the stimuli used in the experiment to ensure that the German participants were familiar with the verbs that were used in the MEG experiment (similar to Drijvers and Özyürek, 2018).

2.3. Stimulus materials

We used the same stimuli as in Drijvers et al. (2018a). These stimuli consisted of 160 2-s video clips of a woman uttering Dutch verbs in either clear or degraded speech, while producing a gesture or not. All verbs that were used were highly frequent Dutch action verbs (see for pre-tests and earlier behavioral experiment, Drijvers and Ozyürek, 2017, and for German participants, Drijvers and Özyürek, 2018). All verbs were unique and no verbs were repeated. The actress in the videos was visible from the knees up, and was wearing neutral-colored clothing and stood at a neutrally colored background (see Fig. 1A).

In short, all videos had an average length of 2000 ms. The preparation of the gesture (i.e., the first frame in which the actress moved her hand), was at 120 ms. The stroke of the gesture commenced at approximately 550 ms, followed by speech onset at approximately 680 ms, gesture retraction at 1380 ms and gesture offset at 1780 ms. Note that the stroke of the gesture started on average 130 ms before speech onset, maximizing the overlap between speech and gesture for mutual enhancement and

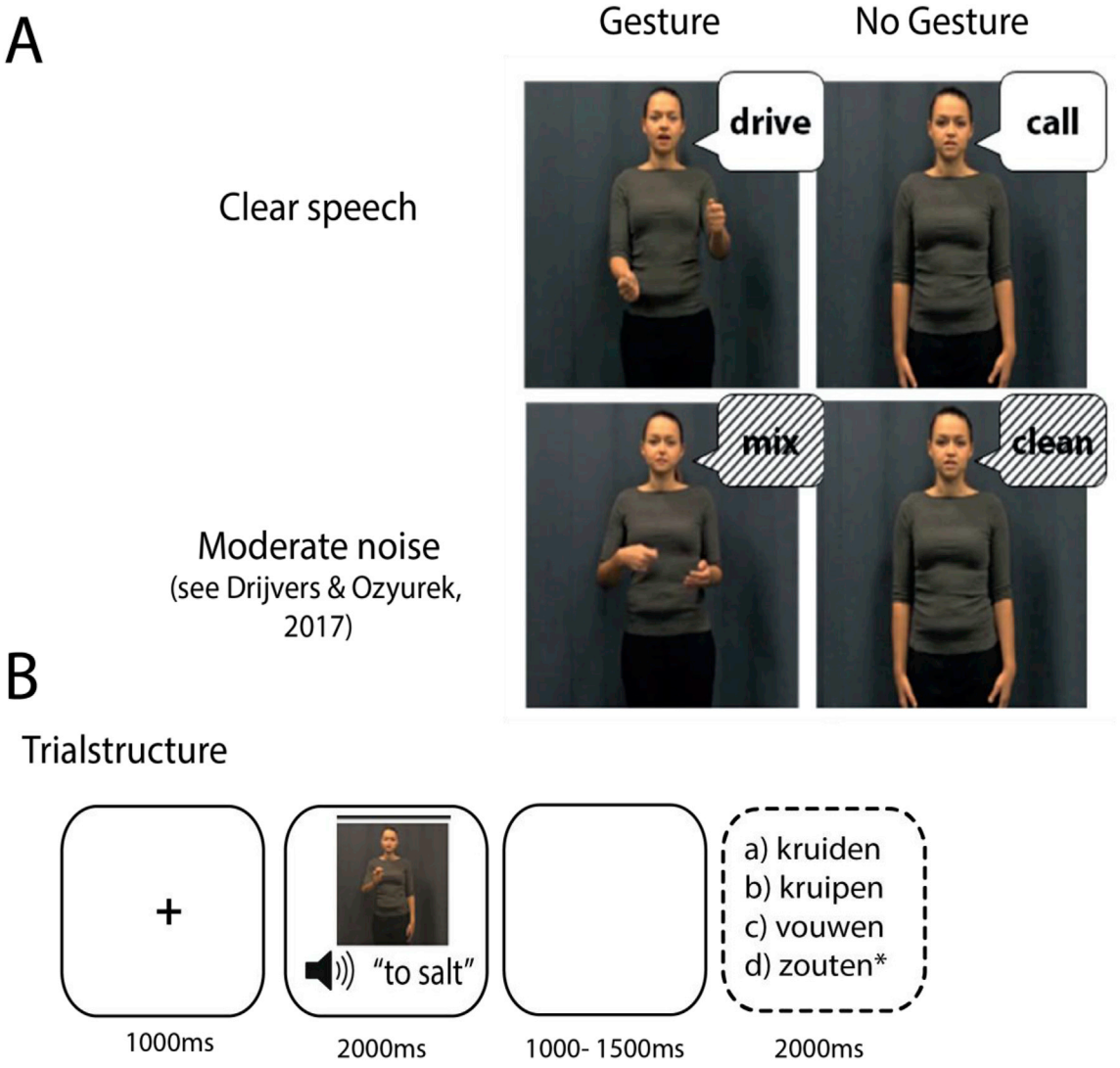

Fig. 1. A: Illustration of the different conditions and stimuli, based on Drijvers et al. (2018a). B: Structure of a trial: Participants were presented with a fixation cross and watched and listened to the video. After a delay period, they had to indicate out of 4 options which verb they heard in the video.

comprehension (Habets et al., 2011).

The sound in the videos was intensity-scaled to 70 dB, de-noised with *Praat* (Boersma and Weenink, 2015) and recombined with their corresponding video files in Adobe Premiere Pro. All the 'cleaned' audio files were noise-vocoded by using 6 noise-vocoding bands (see for pretests; Drijvers and Özyürek (2017) and Drijvers and Özyürek (2019)). We used 6 noise-vocoding bands because pretests had shown that 6-band noise-vocoding allowed for the most gestural enhancement in both non-native and native speakers. Noise-vocoding obtained by band-pass filtering each speech file between 50 and 8000 Hz, and dividing the signal by 6 logarithmically spaced bands between 50 and 8000 Hz. This resulted in cut-off frequencies at 50 Hz, 116.5 Hz, 271.4 Hz, 632.5 Hz, 1473.6 Hz, 3433.5 Hz and 8000 Hz. We used half-wave rectification to extract the amplitude envelope and multiplied the amplitude envelope with the noise-bands, before recombining the bands to create the degraded speech signal (Shannon et al., 1995). The speech sounds from the presented videos were presented to the participant through plastic MEG compatible air tubes.

We presented the stimuli in four conditions to probe gestural enhancement of degraded speech comprehension (similar as in Drijvers et al., 2018a): a clear speech condition with no gesture (CO, clear speech only), a degraded speech condition with no gesture (DO, degraded speech only), a clear speech condition with a matching gesture (CM, clear speech + matching gesture) and a degraded speech condition with a matching gesture (DM, degraded speech + matching gesture). All four conditions contained 40 videos, and none of the verbs overlapped in any condition.

2.4. Procedure

All participants were required to take an online LexTALE test to see whether they met the participation criteria. If a participant scored above 60%, the participant was invited for the MEG experiment. For the MEG experiment, participants were asked to attentively listen to and watch the videos. All participants were instructed that they would encounter a cued-verb recall task after each video where they would be asked to indicate which verb they heard in the videos by means of a right-hand button-press on a 4-button box. We included the cued-verb recall task to ensure that participants were paying attention to the videos, and to calculate whether these behavioral responses could be predicted by their oscillatory modulations (as was found for native listeners in Drijvers et al., 2018a).

Every trial started with a fixation cross (1000 ms), which was followed by the experimental video (2000 ms). After a short interval (1000–1500 ms, jittered), the subject had to indicate which verb they heard. Following the response, there was a 1000 ms pause upon which the next trial would start. The order of the stimuli was pseudo-randomized per subject, with the constraint that the same condition could not occur more than twice in a row along with the constraint that each video would only be presented once. The videos were divided into four mixed blocks of 40 trials each. No verbs were repeated. After each block, the participants could take a self-paced break. If any significant head-movement occurred (>5 mm), the experiment was paused and the subject was brought back to the original starting position.

2.5. MEG data acquisition

We followed all procedures described in Drijvers et al. (2018a). We recorded MEG with a 275-channel axial gradiometer CTF MEG system. All data were filtered online with a 300 Hz low pass filter, digitized at 1.2 kHz and stored for offline data analyses. The head position of the participants with respect to the gradiometers was measured by using three tracking coils (placed at the left and right ear canal and at the nasion to monitor head position in real-time (Stolk et al., 2013). Four channels of the CTF system were malfunctioning throughout all recordings (MLC11, MLC32, MLF62, MRF66). We recorded all participants' eye gaze by using an Eyelink 1000 eyetracker, to monitor eye-blinks during the task. Participants' electrocardiogram (ECG) and horizontal and vertical electrooculogram (EOG) were recorded for artifact rejection purposes. A neck brace was applied to reduce head-movements in the MEG (Lozano-Soldevilla et al., 2014). In the MEG, the subject was positioned in a seated position at 70 cm distance to the screen, similar as in Drijvers et al. (2018a). All stimuli were back-projected onto a semi-translucent screen by using a PROPixx projector with a resolution of 1920x1080 and a refresh rate of 120 Hz. All stimuli were presented at full screen through Presentation software (Neurobehavioral Systems, Inc.).

2.6. MEG data analyses: preprocessing and time-frequency representations of power

We analyzed all MEG data in FieldTrip, an open-source MATLAB toolbox (Oostenveld et al., 2011), and followed the exact same procedure as in Drijvers et al. (2018a). The data were segmented into trials starting 1s before and ending 3 s after the onset of the video. The data were demeaned, detrended and band-stop filtered at 50, 100 and 150 Hz to remove any line noise that could contaminate the data. We then visually inspected the data for overt muscle artifacts, movement artifacts, SQUID jump artifact and other irregular artifacts. All trials with overt artifacts were rejected. We used a semi-automatic rejection routine and removed 4 trials per condition on average which were contaminated by SQUID jump artifacts and muscle artifacts. We then applied independent component analyses to attenuate the signals generated from eye-blinks, eye-movements and cardiac-related activity (Bell and Sejnowski, 1995). As a final step, we went through all single trials again to remove any artifacts that were not removed by ICA or our semi-automatic rejection procedure. We then resampled the data to 300 Hz to speed up the subsequent analyses. For a more intuitive interpretation of the data, we calculated a synthetic planar gradient, as planar gradient maxima are known to be located above the neural sources that may underlie them (Bastiaansen and Knösche, 2000). An approximation of the planar gradient was computed by converting the axial gradiometer data to orthogonal planar gradiometer pairs, and summing the power of the pairs.

2.7. Time-frequency analyses of power

The time-frequency analyses of power were the same as described in Drijvers et al. (2018a). Over a frequency range of 2–30 Hz, we applied a 500-ms Hanning window in frequency steps of 1 Hz and time steps of 50 ms. As we were interested in the gestural enhancement, we compared the difference in power in the Degraded speech + Matching Gesture (DM) and Degraded Speech (DO) conditions to the difference in the Clear Speech + Matching Gesture (CM) and Clear Speech (CO) condition. The power in these conditions was averaged separately for each participant and log10 transformed. We compared the within-group differences between the conditions (DO vs CO, DM vs. CM, DM vs. DO, CM vs. CO), by subtracting the $\log_{10}$ transformed power (i.e., the log ratio, $\log_{10}(A) - \log_{10}(B)$). Similarly, to calculate the gestural enhancement effect, we calculated the difference between DM/DO and CM/CO as $(\log_{10}(DM) - \log_{10}(DO)) - (\log_{10}(CM) - \log_{10}(CO))$. As a time-window of interest, we used the whole window in which speech and gesture were unfolding (0.7–2.0s) for both the within-group and between-group comparisons. To compare the effects of the non-native listeners to the native listeners, we compared this time-window of interest between groups in both the alpha (8–12 Hz) and beta (14–22 Hz) band. Note that in Drijvers et al. (2018a), effects in the gamma band were described. In the non-native group we did not observe any differences in any comparison, nor did we observe reliable differences between the native and non-native group. For comparisons of the single contrasts and interaction effects in the gamma band, please see Supplementary Materials (S1).

2.8. Source analyses

We estimated the sources of observed effects on sensor-level by using dynamic imaging of coherent sources (DICS (Gross et al., 2001),), a beamforming spatial filtering technique. Note that our source analysis served to localize the observed effects on sensor-level, but not to form an additional statistical assessment. Axial gradiometer data were used to perform these analyses. First, a spatial filter was calculated from the cross-spectral density matrix, as well as a lead field matrix. We constructed individual lead fields from our participants by using a realistically shaped single-shell head model based on the participants' own anatomical data from a segmented structural MRI, by dividing the brain volume in a 120 mm spaced grid and warping it to a template brain (MNI).

All within-group source analyses used the time windows in which conditions were found to statistically differ in the sensor analyses. For the alpha band, we thus calculated the CSD at 10 Hz, with 2 Hz frequency smoothing. For the beta band, this effect was centered at 18 Hz, with 4 Hz frequency smoothing. Note that these settings, except for the time windows, are similar to the analyses described in Drijvers et al. (2018a). As the time-windows slightly differed for the non-native and native listeners, we performed a between-group comparison over the whole time window of interest, in both the alpha and beta frequency band to test for between-group differences. We used a common spatial filter over all conditions to project the data through. This common filter was then separately applied to each condition to calculate the power at each gridpoint. This was averaged over trials and $\log_{10}$ transformed. For visualization purposes, we interpolated the grand-average grid of all participants onto the template MNI brain.

2.9. Cluster-based permutation statistics

Non-parametric cluster-based permutation tests were performed across subjects to statistically assess oscillatory power differences between the different conditions and between the non-native and native listener group (Maris and Oostenveld, 2007). The source-level statistics were computed to create thresholded masks to localize any effects that were observed on the sensor-level. We computed the mean difference between two conditions for each x/y/z sample (source) or sample for sensor TFR analysis (sensor), in the frequency ranges of interest (alpha; 8–12 Hz, beta; 14–22 Hz, as determined by a grand-average TFR of all conditions combined) and time window of interest (0.7–2.0 s, from speech onset to video offset). After collecting the difference values of the comparisons, all adjacent values exceeding the threshold of 5% percent were grouped into clusters. This resulted in a distribution of different cluster candidates. The cluster candidates were randomly reassigned 5000 times across all conditions and participants. The cluster with the highest sum of difference values was added to a distribution, resulting in a permutation distribution. The observed cluster values were then compared to this newly created permutation distribution. The clusters that were in the highest or lowest 2.5% were considered significant.

2.10. The relation between alpha and beta oscillations and behavioral cued-recall scores

In Drijvers et al. (2018a) we observed a clear correlation between oscillatory power modulations and the amount of gestural enhancement participants experienced during external ambiguity. As non-native listeners might choose different strategies to process the degraded speech signal or use the gestural information to enhance comprehension, we again correlated an individual's oscillatory power with the behavioral scores that we obtained from the cued-recall task. We calculated this by averaging the power modulation over time points, frequencies and sensors in significant clusters of the interaction effects, which resulted in an individual's power modulation score per frequency band. For the behavioral scores, we calculated an interaction score for the reaction times and amount of correct answers, which was similar to how we calculated the gestural enhancement in oscillatory power. We computed difference scores between the conditions (e.g., DM-DO, CM-CO) and compared these differences to each other, resulting in the amount of behavioral gestural enhancement per participant. Subsequently, we obtained Spearman correlation between these scores an individual's power modulation per frequency band. As our hypotheses were specific on the direction of the power modulation per frequency bands, we used one-tailed tests.

3. Results

Highly-proficient non-native listeners of Dutch watched videos in which an actress would utter an action verb in clear or degraded speech, while making a gesture or not. After every video participants completed a cued-recall task in which they identified what verb they heard in the video. We recorded MEG during the whole experiment, but were interested in oscillatory modulations of power in the alpha and beta frequency band while participants watched the videos, and how the oscillatory dynamics during this time interval related to their behavioral benefit on the cued-recall task.

Our analysis was twofold. First, similar to Drijvers et al. (2018a), we were interested in the behavioral responses as well as oscillatory modulations during gestural enhancement of degraded speech comprehension in non-native listeners (within group). For both the behavioral and neural results, gestural enhancement was calculated as the interaction between the occurrence of a gesture (present/not present) and speech degradation (clear/degraded). Second, we compared the observed behavioral results and oscillatory modulations in non-native listeners to those observed in native listeners, as reported in Drijvers et al. (2018a) (between-group).

3.1. Behavioral results

3.1.1. Non-native listeners (within-group)

3.1.1.1. Gestural enhancement of speech comprehension is largest when speech is degraded. Non-native listeners experienced the most gestural enhancement when speech was degraded, mirroring earlier work on gestural enhancement in native speakers (Drijvers et al., 2018a), and behavioral work on gestural enhancement of degraded speech comprehension in non-native speakers (Drijvers and Özyürek, 2019). A repeated-measures ANOVA with the factors Noise-vocoding (clear speech vs. degraded speech) and Gesture (present vs. not present) on the percentage of correct answers revealed that participants were more able to correctly identify the verb in clear than in degraded speech ($F(1, 29) = 246{,}896$, $p < .0001$, $\eta 2 = 0.895$), and when a gesture was present compared to not present ($F(1,29) = 13{,}88$, $p = .001$, $\eta 2 = 0.324$). A significant interaction between Noise-vocoding and Gesture ($F(1, 29) = 14.238$, $p = .001$, $\eta 2 = 0.329$), indicated that gestural enhancement was largest when speech was degraded. A similar pattern was observed in the reaction times, where listeners were faster when speech was clear than degraded ($F(1,29) = 121.38$, $p < .001$, $\eta 2 = 0.807$) and a gesture was present compared to not present ($F(1,29) = 41.629$, $p < .001$, $\eta 2 = 0.589$). Gestural enhancement was largest when gestures were present and speech was degraded ($F(1.29) = 15.113$, $p = .001$, $\eta 2 = 0.343$), which caused reduced reaction times that was more evident in degraded than in clear speech (see Fig. 2).

3.1.2. Non-native listeners vs. native listeners (between-group)

We compared the results of the two groups in a 2 (group; non-native/native) x 2 (gesture; present/not present) x 2 (noise-vocoding; clear/degraded speech) repeated-measures ANOVA for both the correct answers and the reaction times. The analysis of the correct answers revealed no significant differences on any of the interaction terms that contained

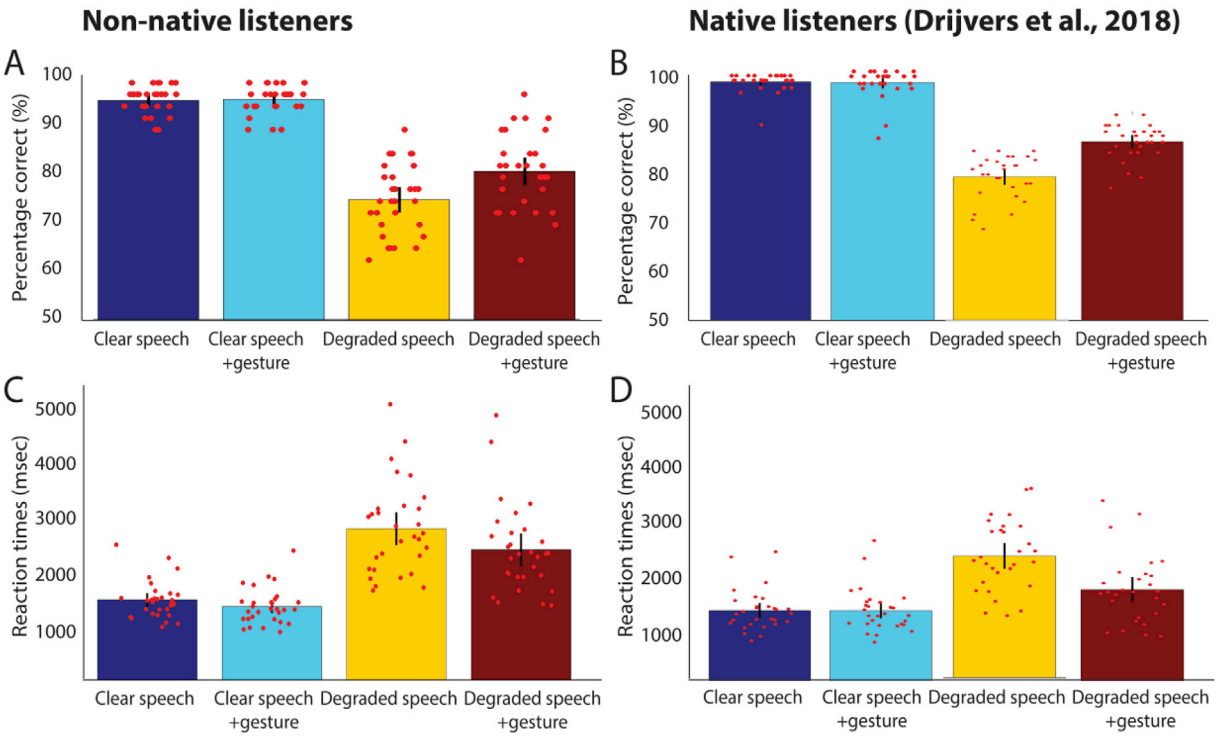

Fig. 2. A/B: Percentage of correct answers per condition for non-native (A) and native listeners (B). Error bars represent SE. Red dots represent an individual participant's data. C/D: Reaction times (in milliseconds) per condition for non-native (C) and native listeners (D). Error bars represent SE. Red dots represent an individual participant's data. Gestural enhancement of degraded speech was similar for non-native and native listeners, but non-native listeners were significantly slower.

the between-group factor, indicating that non-native listeners and native listeners had a similar number of correct answers on clear and degraded speech trials ($F(1,57) = 3.778$, $p = .057$) and trials containing a gesture or no gesture ($F(1,57) = 0.447$, $p = .507$). Gestural enhancement of degraded speech comprehension was not larger for native listeners compared to non-native listeners ($F(1,57) = 3.778$, $p = .306$).

The results of the reaction times revealed different results: native listeners were quicker to answer than non-native listeners on clear and degraded speech trials ($F(1,57) = 15.091$, $p = < .001$), as well as quicker to answer on gesture and no-gesture trials ($F(1,57) = 8.78$, $p = < .001$). Again, there was no three-way interaction of Gesture, Noise-vocoding and Group ($F(1,57) = 0.354$, $p = .554$), indicating that although native and non-native listeners show similar behavioral effects, non-native listeners overall answer more slowly than native listeners. In conclusion, our behavioral results thus revealed that although gestural enhancement of degraded speech comprehension was similar for native and non-native listeners, non-native listeners answered more slowly and were trending towards more incorrect answers.

3.2. MEG results

3.2.1. Non-native listeners (within-group)

3.2.1.1. Alpha power is suppressed in pSTS/MTG, motor and visual regions during gestural enhancement of degraded speech comprehension in non-native listeners. Next we asked how oscillatory activity in the alpha band was modulated during gestural enhancement of degraded speech comprehension in non-native listeners. We first conducted a sensor-level analysis over the full time-window of interest (0.7–2.0, from speech onset until the end of the video) to test for an interaction effect between noise-vocoding and gesture occurrence. This 'gestural enhancement effect' was calculated by comparing the differences between the DMDO and CMCO contrasts (i.e., ($\log_{10}$(DM) - $\log_{10}$(D)) - ($\log_{10}$(CM) - $\log_{10}$(C))). Time-frequency representation (TFRs) of power of individual trials were calculated and averaged per condition. Fig. 3A and B represent the TFRs of power during gestural enhancement of degraded speech comprehension at representative sensors within the non-native listener group. We then visualized the effect in time and space by plotting the topographical distribution of the interaction in the alpha band over time (see Fig. 4A). Sensor-level analyses revealed that alpha power was more suppressed when speech was degraded and a gesture was present (one negative cluster, $p = .006$). This difference between DMDO and CMCO showed a central-parietal onset (0.7–1.0) that progressed over left-temporal and occipital (1.0–1.4) areas to right-temporal areas (1.4–2.0). For comparisons of the single contrasts, please see Supplementary Materials (S2).

To localize the observed effect from the sensor analysis, we conducted source analyses to determine the underlying sources of the negative cluster. We applied a cluster-randomization approach to the source data and used the outcome of this analysis as a threshold for when to consider the source estimates reliable (the statistical assessment of the effect was thus formed by our sensor analyses, not the source analyses). As can be observed by the topographical alpha power distribution plots in Fig. 4, the effect observed in the sensor analysis commences at left-central and parietal regions and progresses over left-temporal and occipital regions to right-temporal regions. We therefore assessed the sources of this cluster in three time windows instead of one to reliably capture the sources of the effect (note that the time-dimension of the data is no longer available using DICS, which would mean that when we would try to visualize the sources over the whole time window, this would result in a source in the middle of the observed loci on sensor level). In the first time-window, from 0.7 to 1.0 s, we observed a larger alpha power suppression when gestures enhanced degraded speech comprehension over STS/MTG, pre/postcentral regions and angular gyrus ($p = .04$, one negative cluster, see Fig. 3C). In the time window from 1.0 to 1.4 s, we observed a larger alpha power suppression over left-temporal (pSTS/STG/MTG) and left- and right-occipital regions ($p = .02$, one negative cluster, see Fig. 3D), and in a final time window from 1.4 to 2.0 s we observed a larger alpha power suppression in right-temporal (pSTS/STG/MTG) regions and, the left temporoparietal junction and left- and right-occipital regions ($p = .004$, one negative cluster, Fig. 3E).

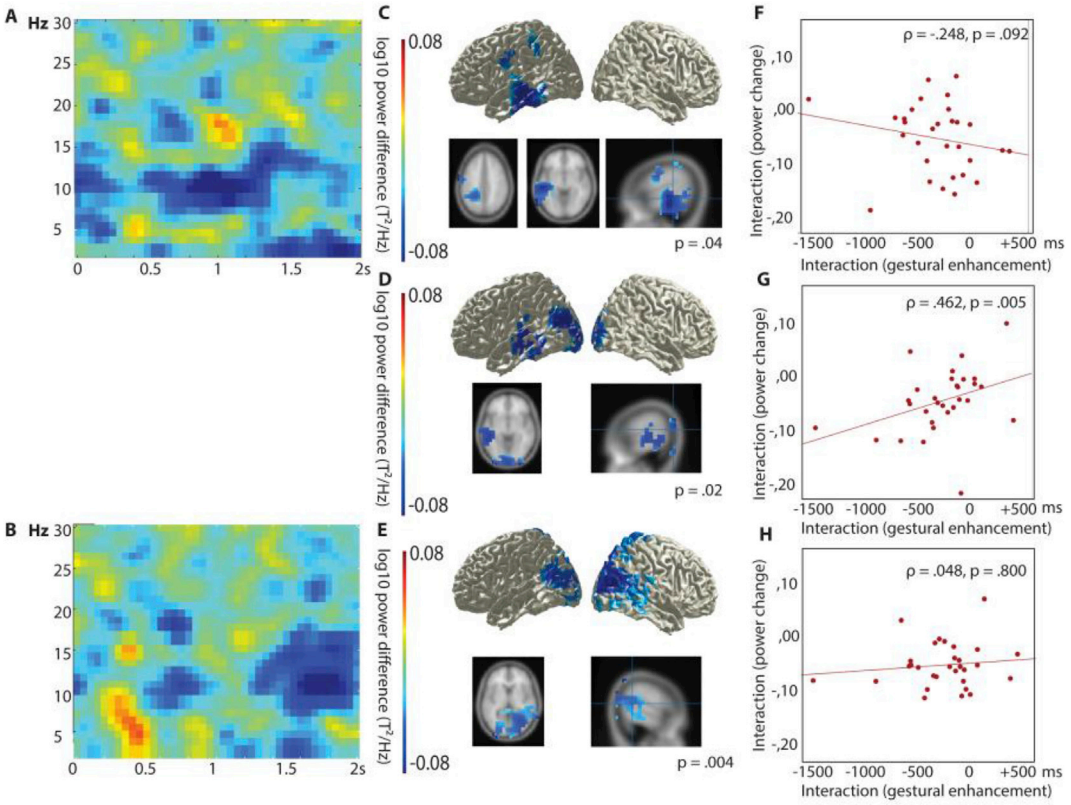

Fig. 3. A: Time-frequency representation of power at representative left-temporal sensors, capturing both the alpha effect from the first time window (0.7–1.0s) and the second time window (1.0–1.4 s). B: Time-frequency representation of power at a representative cluster formed by channels from right-temporo-occipital regions, capturing the late alpha effect (1.4–2.0 s). C: Estimated source results of the first alpha cluster, masked by statistically significant clusters. D: Estimated source results of the second alpha cluster, masked by statistically significant clusters. E: Estimated source results of the third alpha cluster, masked by statistically significant clusters. F: Individual's alpha power modulation in the first time window as a function of individual's gestural enhancement in the cued-recall task. G: Individual's alpha power modulation in the second time window as a function of individual's gestural enhancement in the cued-recall task. H: Individual's alpha power modulation in the third time window as a function of an individual's gestural enhancement in the cued recall task.

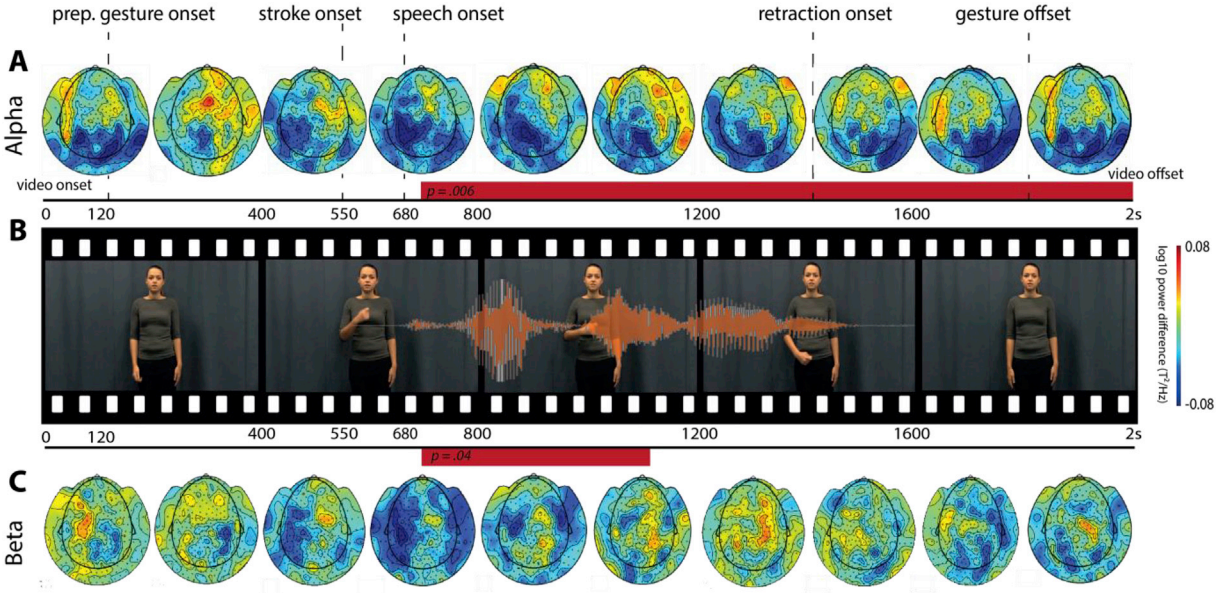

Fig. 4. A: Topographical distribution of alpha power over the whole video interval for the gestural enhancement effect, binned per 200 ms. Red bar on timeline represents significant cluster in sensor-level analysis. B: Structure of the video. Orange waveform represents speech. C: Topographical distribution of beta power over the whole video interval for the gestural enhancement effect, binned per 200 ms. Red bar under timeline represents the significant cluster of the sensor analysis.

3.2.1.2. Left-temporal individual alpha power modulations predict gestural benefit during degraded speech comprehension in non-native listeners. We correlated a participant's individual power modulations in the alpha band with the benefit from gestures participants experienced during the behavioral task. Here, we reasoned that the interaction effect of the accuracy scores might not reliably capture the gestural enhancement a participant experiences during the video. For example, a participant may have not understood the verb while watching the video, but might have only recognized the verb when they were presented with the different answering options. This would result in a correct score, but a slower reaction time than when the participant did already recognize the verb while watching the video. As stated in our hypotheses, we however also expected that gestural enhancement should speed up reaction times. If indeed participants solely selected the right answer when they recognized it in the cued answer options but did not understood it during video presentation, this speeding up of reaction times as a result of gestural enhancement would not occur. Therefore, we calculated an individual's speeding or slowing caused by gestural enhancement by calculating the difference in reaction times between DMDO and CMCO and correlating it with an individual's power modulation. To investigate whether the power modulations that were estimated over specific regions were predictive of gestural enhancement of degraded speech comprehension on the behavioral task, we correlated an individual's behavioral scores with power modulations in the three different time windows. These analyses revealed that the more a listener's alpha power was suppressed in the second time window, the more a listener benefitted from gestural enhancement of degraded speech comprehension (1.0–1.4 s, Spearman's rho = ,462, $p = .005$, one-tailed, see Fig. 3F). Note that this is the time window where most of the meaningful part of the speech and most of the meaningful part of the gesture have unfolded. This correlation was not found in the early time window (0.7–1.0 s, Spearman's rho = −.248 p <. 0.092 one-tailed, Fig. 3G), nor in the late time window (1.4–2.0 s, Spearman's rho = 0.048, $p = .80$, Fig. 3H). Note that similar results were obtained when correlating the individual's behavioral scores on accuracy with the power modulations in the three different time windows. Here, we observed a correlation between a listener's alpha power and the gestural enhancement effect in accuracy in the second time window (1.0–1.4 s, Spearman's rho = −.353 p <. 0.032 one-tailed), but we observed no correlation in the early alpha time window (0.7–1.0 s, Spearman's rho = −.103 p <. 0.301 one-tailed) or the late alpha time window (1.4–2.0 s, Spearman's rho = .241 p <. 0.108 one-tailed, see Figs. S1A–C for visualizations of these correlations). These results indicate that the individual power modulations over left-temporal regions and occipital regions predict the behavioral benefit of a gesture a non-native listener experiences during degraded speech comprehension.

3.2.1.3. Beta power is more suppressed over LIFG and motor regions when gestures enhance degraded speech comprehension. Next we followed a similar procedure when we analyzed sensor-level differences in the beta band (14–22 Hz, range determined on grand-average TFR of all conditions combined, see Fig. 5A). We studied the spatiotemporal course of the effect by plotting the topographical distribution of the gestural enhancement effect (Fig. 3B). We there observed a left-lateralized effect in an early time window. Sensor-level analyses of the interaction effect indeed confirmed that beta power was more suppressed when speech was degraded and a gesture was present (one negative cluster, $p = .04$). This effect occurred when the stroke of the gesture and speech were unfolding (0.7–1.1 s). For comparisons of the single contrasts, please see Supplementary Materials (S3).

We then used source-analysis to estimate the source of the gestural enhancement effect. These analyses demonstrated that the larger beta suppression could be localized to the LIFG, and left pre- and post-central gyrus (one negative cluster, $p = .03$, Fig. 5B).

3.2.1.4. Non-native listener's individual beta power in motor cortex and LIFG predicts gestural benefit during degraded speech comprehension. We correlated a listener's individual beta power with the amount of speeding/slowing a listener experienced during gestural enhancement of degraded speech comprehension in the cued-recall task, and observed a correlation between the amount of beta suppression in the motor cortex/LIFG and the behavioral scores: the more a listener's beta power was suppressed over motor regions and LIFG, the more a listener could benefit from gestural enhancement of degraded speech comprehension (Spearman's rho = 0.438, $p = .008$, one-tailed, Fig. 5C). A similar effect was observed when correlating an individual's beta power with the gestural enhancement effect in their accuracy scores (Spearman's rho = −0.494, $p = .004$, one-tailed, Fig. S1D).

3.2.2. Non-native listeners vs. native listeners (between-group)

3.2.2.1. Native listeners' alpha power is more suppressed in LIFG and ATL than in non-native listeners. We then compared the results of the non-native listeners to the results of the native listeners reported in Drijvers et al. (2018a) to test for between-group differences in the gestural enhancement effect. To this end, we first calculated sensor-level differences in the alpha band (8–12 Hz) between native and non-native listeners by comparing the gestural enhancement effect in the time window of interest (0.7–2.0 s), using between-group cluster-based permutation tests. Here we observed a larger alpha power suppression reflecting the difference in the gestural enhancement effect over left-frontal regions for native compared to non-native listeners over the entire time window (0.7–2.0 s, one negative cluster, $p = .02$, Fig. 6A).

We then estimated the source of this difference in alpha power between the groups and observed a larger alpha power suppression for native than non-native listeners in LIFG and ATL (one negative cluster, $p = .04$, Fig. 6B).

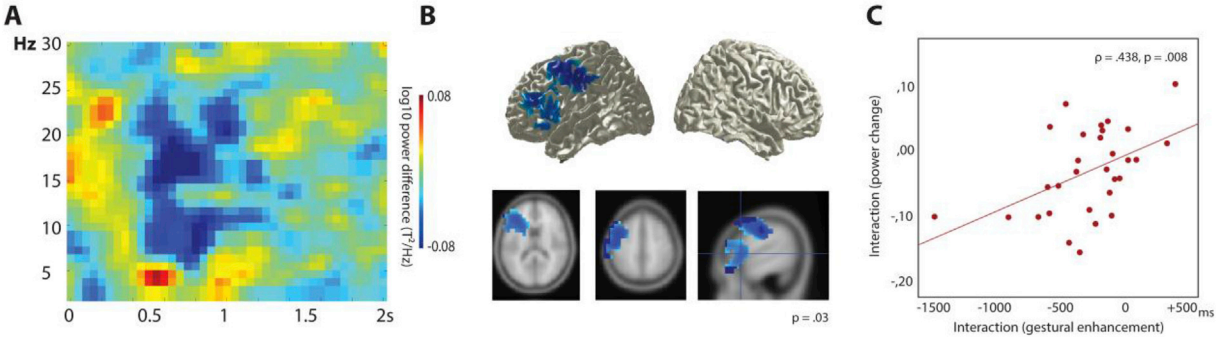

Fig. 5. A: Time-frequency representation of power at representative left-frontal/left-motor sensors. B: Estimated source results of the beta cluster, masked by statistically significant clusters. C: Individual's beta power modulation as a function of individual's gestural enhancement in the cued-recall task.

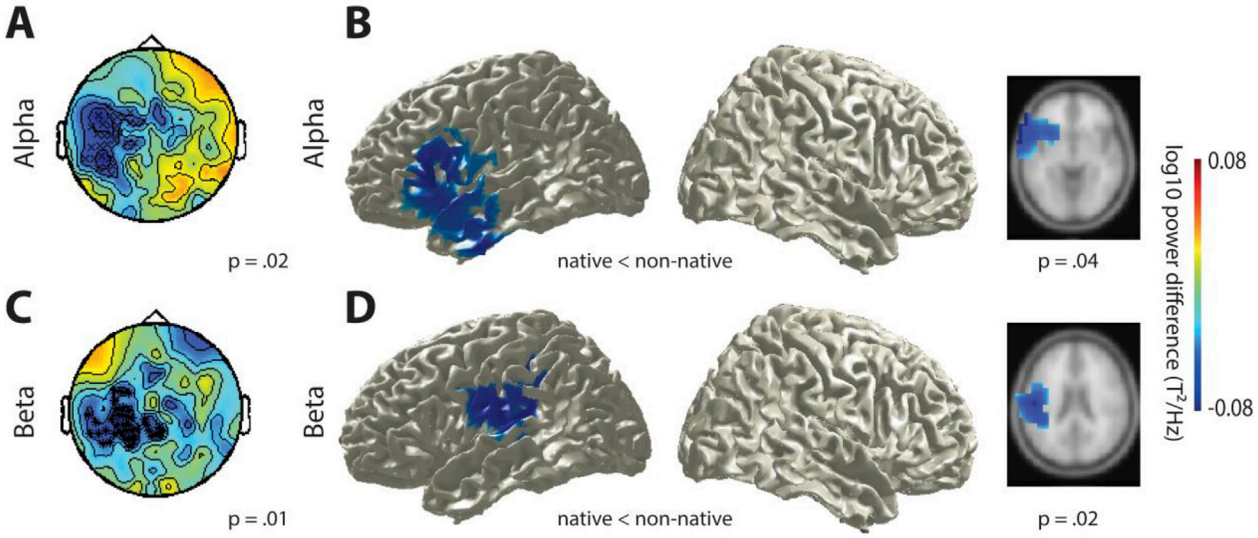

Fig. 6. A: Topographical plot of difference in alpha power between non-native and native listeners on sensor level. B: Estimated source results of the alpha cluster, masked by statistically significant clusters. C: Topographical plot of difference in beta power between non-native and native listeners on sensor level. D: Estimated source results of the beta cluster, masked by statistically significant clusters.

3.2.2.2. Native listeners' beta power is more suppressed over primary somatosensory cortex than in non-native listeners. On sensor-level, we observed a larger beta power suppression (14–22 Hz) for native compared to non-native listeners over the whole time window (0.7–2.0 s, one negative cluster, $p = .01$, 6C). The source of this effect was estimated over primary sensory cortex and left insula (one negative cluster, $p = .02$, Fig. 6D).

4. Discussion

We set out to investigate the spatiotemporal oscillatory dynamics that support gestural enhancement of degraded speech comprehension in non-native listeners, and how these oscillatory modulations compare to earlier results observed in native listeners (Drijvers et al., 2018a). Using this manipulation, we investigated how much benefit from gesture a listener has when resolving language in both externally induced (speech degradation) and internally induced (non-nativeness) adverse listening conditions.

On a behavioral level, we observed a similar gestural enhancement effect for non-native listeners as for native listeners. Although the gestural enhancement effect was similar for both groups, non-native listeners were significantly slower in providing their answers to the cued-recall task. This was partly in line with our hypothesis, as we also expected to observe a significantly lower accuracy score for non-native compared to native listeners.

On a neural level, our central hypothesis was that a suppression of alpha and beta power would reflect engagement of task-relevant brain regions during gestural enhancement of degraded speech comprehension, similar as in Drijvers et al. (2018a, b). This hypothesis was confirmed. More specifically, we demonstrated that when gestures enhanced degraded speech comprehension for non-native listeners, we observed a larger alpha power suppression that commenced at central-parietal regions, and which over time was observable in left-temporal, occipital and right-temporal regions. We observed an early beta power suppression in LIFG and motor regions during gestural enhancement of degraded speech comprehension. We found distinct correlations for two successive time windows in the beta (0.7–1.1) and alpha band (1.0–1.4) (i.e., the time window in which the meaningful part of the gesture and speech were unfolding), that revealed that an individual's power modulations could predict an individual's gestural benefit when resolving language in adverse listening conditions.

Importantly, when comparing the gestural enhancement effect of non-native listeners to the gestural enhancement effect observed in native listeners, native listeners demonstrated more alpha suppression in LIFG and ATL, as well as a larger beta power suppression in primary somatosensory cortex and left insula. Below we discuss the putative role of these spatiotemporal effects during gestural enhancement of speech comprehension in internally and externally induced adverse listening conditions. We will first summarize the results for the non-native listeners (within-group), followed by a comparison between the non-native and native listener groups (between-group).

4.1. Non-native listeners (within-group)

4.1.1. Non-native listeners who more strongly engage motor regions and LIFG benefit more from gestures during degraded speech comprehension

In line with our hypotheses, we observed a stronger engagement of motor regions and LIFG during gestural enhancement of degraded speech comprehension. This effect was observed in an early time window (0.7–1.1 s). In line with our previous work on native listeners (Drijvers et al., 2018a), this might suggest that the motor system is engaged to simulate the observed gesture more strongly when speech is degraded (Klepp et al., 2015; van Elk et al., 2010; Weiss and Mueller, 2012), possibly to extract meaningful information to aid ongoing degraded speech comprehension. In line with previous work, we also suggest that the larger beta power suppression over LIFG reflects a larger engagement of LIFG in this time window to unify the gestural information with the degraded speech signal. Similar results have been observed when unification load was higher due to semantic congruency (Wang et al., 2012; of gestures, see Drijvers et al., 2018b) or speech degradation (Hervais-Adelman et al., 2012; Obleser et al., 2007; Wild et al., 2012).

Importantly, a listener's individual oscillatory power modulation correlated with a listener's individual behavioral benefit of gestural information during degraded speech comprehension: the more an individual listener's beta power was suppressed in this time window, the more gestural benefit a listener experienced in the cued-recall task. This suggests that listeners might optimize their processing strategy by immediately engaging motor regions and LIFG to extract semantic information from the gesture and continuously attempt to unify incoming information with speech to aid retrieval of the speech input. We will interpret this effect in more detail below, when we compare this to the patterns we observed in native listeners in Drijvers et al. (2018a).

4.1.2. A left-lateralized network of motor regions, AG, pSTS/MTG and STG is engaged during gestural enhancement of degraded speech comprehension

In the same time window as the beta effect discussed in the previous paragraph, we observed a stronger alpha power suppression (0.7–1.0 s) in pSTS/STG/MTG, left motor regions and left angular gyrus. Activation of the pSTS/MTG has been repeatedly found in studies on speech-gesture integration, and is thought to reflect an initial matching of the audiovisual stimuli (Dick et al., 2014, 2012; Holle et al., 2010; Willems et al., 2009, 2007). In line with these studies, we propose that the stronger alpha power suppression might reflect early engagement of the language system to perform an initial integration of lower-level characteristics of the audiovisual input.

Note that the abovementioned effects in the beta band (0.7–1.1 s) occur in a similar time window as the current effect in the alpha band (0.7–1.0 s), but that the beta, and not the alpha (0.7–1.0 s) effects correlate with gestural benefit during degraded speech comprehension. This confirms that the alpha band effect indeed might reflect an initial matching of, possibly lower-level, audiovisual information, that is similar for all non-native listeners and does not relate to the gestural enhancement that a listener experiences per se. The engagement of the AG, which is often seen as an association and supramodal integration hub (Binder et al., 2009), and the motor system, which engages more strongly during gestural enhancement of degraded speech, might aid in this integration process.

4.1.3. Non-native listeners who more strongly engage visual regions and left-temporal regions, experience more gestural benefit during degraded speech comprehension

In contrast to the alpha effect in the first time window (0.7–1.0 s), an alpha effect in the subsequent time window (1.0–1.4s) over pSTS/STG/MTG *did* predict how much gestural benefit a non-native listener experiences during gestural enhancement of degraded speech comprehension. We suggest that a listener's power modulation in this time window is predictive of an individual's gestural benefit during degraded speech comprehension because the semantic information from the gestures that is being unified with the unfolding degraded auditory cues in an earlier time window, as was demonstrated by our beta effects (0.7–1.0 s), aids subsequent lexical access of the degraded input (Hagoort, 2013; Lau et al., 2008). Post-hoc power-power correlations between an individual's beta power in the early time window (0.7–1.0 s) and an individual's alpha power in the second time window (1.0–1.4 s) concur with this proposed interpretation: listeners who more strongly show beta suppression in the first time window, also show a larger alpha suppression in the second time window (Spearman's rho = 0.408, $p = .013$, one-tailed). Similarly, listeners who demonstrated a larger alpha suppression over visual regions might have allocated more visual attention to the gestures when speech is degraded to aid comprehension.

4.1.4. Right-temporal regions engage when more neural resources are recruited for comprehension

We then observed a larger alpha power suppression (1.4–2.0s) over right-temporal and right-occipital regions. An individual's power modulation in this time window did not correlate with subsequent comprehension on the cued-recall task, suggesting that this effect might be general for all non-native listeners, and not per se related to the gestural enhancement effect. fMRI studies have suggested that right-lateralized regions are often recruited during non-native language processing (Higby et al., 2013; Leonard et al., 2011). This might suggest that non-native listeners try to recruit more top-down information to facilitate comprehension and unification of the two input streams (Skipper et al., 2007, 2006). We thus suggest that these effects show that right-lateralized regions might be more engaged when non-native listeners require more neural resources for comprehension, especially when auditory cues are not reliable enough to map the semantic information from the gesture to. This is also in line with previous results where we observed right-lateralized effects when comparing matching and mismatching gestures (Drijvers and Özyürek, 2018), where non-natives seemed to recruit additional resources to process mismatching semantic information.

4.1.5. Non-native listeners vs. native listeners (between-group)

As mentioned in the Introduction, we also set out to conduct exploratory analyses to compare differences between non-native listeners and native listeners. In line with the behavioral results observed in native listeners (Drijvers et al., 2018a), we observed that for non-native listeners gestural enhancement was largest when speech was degraded. This gestural enhancement effect was similar for non-native and native listeners. However, we observed that non-native listeners were significantly slower in answering on the cued-recall task. As participants were cued with four answering options, it might have been easier to recognize the degraded verb during this answering period than when the participants watched the video. This might have masked the actual comprehension difficulties that the listeners experienced in the video. However, this does not affect the reaction times. When a non-native listener for example might have experienced more difficulty in understanding the speech during the video, it might take longer to find the correct answer in the cued-recall task. These results thus indicate that although the gestural enhancement effect seemed similar for native and non-native listeners, non-native listeners possibly were sometimes hindered in processing and coupling the degraded auditory information to the semantic information conveyed by the gesture, as demonstrated by the slower reaction times.

4.1.6. Non-native listeners might face more difficulty when retrieving gestural semantic information and unifying it with degraded auditory cues than native listeners

Finally, we compared the oscillatory modulations observed in non-native listeners to the modulations observed in native listeners during gestural enhancement of degraded speech comprehension. Here, we expected to observe less engagement of LIFG for non-native compared to native listeners. We confirmed this hypothesis and observed a larger alpha power suppression for native listeners in LIFG, but also anterior temporal lobe (ATL) than for non-native listeners. As the ATL has been implicated as a domain-general semantic hub (Wong and Gallate, 2012) and we have found converging evidence for the engagement of the LIFG during the unification of degraded speech and gestures (Drijvers et al., 2018a,b), this might suggest that when speech is degraded, it might be more difficult for non-native listeners to access the semantic information from the gesture and unify it to the degraded auditory cues. This difficulty in semantic access might be due to the fact that non-native listeners need more available auditory cues to facilitate access to the semantic cues conveyed by the gesture. In turn, this might cause these areas to be less engaged in non-native listeners than in native listeners, and might explain the slower reaction times in the cued-recall task, despite the similar gestural enhancement effect that was observed. This might also explain the lack of an effect in the gamma band within the non-native listener group, as well of the lack of an effect between the two groups: gamma power modulations might be similar for both groups when gestures disambiguate degraded speech, but might be less pronounced in the non-native listener group.

4.1.7. Non-native listeners might face more difficulty utilizing phonological information that is conveyed by visible speech to aid degraded speech comprehension

Unexpectedly, our exploratory analyses also revealed a larger beta power suppression over primary somatosensory cortex and left insula for native compared to non-native listeners during gestural enhancement of degraded speech comprehension. Specifically, the lower part of the somatosensory cortex that is sensitive to information from visible speech (i.e., information conveyed by teeth, tongue and lip movements) was less engaged in non-native than native listeners, possibly because non-native listeners are less able than native listeners to simulate the information conveyed by visible speech when speech is degraded. The cluster

overlapped with the left insula, which has been shown to be sensitive to the strength of cross-modal binding (Bushara et al., 2002) as well as being involved in phonological processing (Abdullaev and Melnichuk, 1997; Bamiou et al., 2003; Booth et al., 2007; Tettamanti et al., 2005; Wild et al., 2012), suggesting that the observed effects are consistent with the idea that non-native listeners might face more difficulty using the phonological information that is conveyed by visible speech to aid degraded speech comprehension.

4.1.8. How does gestural enhancement of degraded speech comprehension differ for native and non-native listeners?

Our current results revealed a different spatiotemporal oscillatory profile during gestural enhancement of degraded speech comprehension for non-native listeners compared to native listeners (see for native listeners: Drijvers et al. (2018a)). In the native listeners described in Drijvers et al. (2018a), we observed an alpha power suppression over right STS (0.7–1.0 s), followed by an alpha/beta power suppression in left-motor regions (1.0–1.6), left-temporal and occipital regions (1.6–2.0s). We observed a larger beta power suppression over the motor cortex and extended language network (1.3–2.0 s) and a larger gamma power increase over MTL (1.0–1.5s). These power modulations correlated with an individual's behavioral benefit in the cued-recall task and were suggested to support general unification, integration and lexical access processes during language comprehension, as well as simulation of and increased visual attention to iconic gestures over time.

The observed oscillatory modulations in non-natives suggest similar core processes that support gestural enhancement of degraded speech comprehension as were observed in native listeners in Drijvers et al. (2018a). However, the different spatiotemporal time course of the effects observed in non-native compared to native listeners might suggest that the two listener groups employ different processing strategies. For example, non-native listeners seem to immediately engage motor cortex and LIFG to extract semantic information from the gesture, and might attempt to immediately unify this information with the signal to aid retrieval of the degraded input. Subsequently, when this integration is hindered, non-native listeners engage additional resources by engaging right-temporal regions to aid in comprehension of ambiguous information. Alternatively, native listeners seem to have more access to the degraded phonological information than non-native listeners and might therefore be less hindered in using the semantic information from the gestures to resolve the degraded input. They can therefore already optimize their processing strategy in an early time window, whereas non-native listeners are not able to do this as it is more difficult for them to access the degraded input to map the semantic information from the gesture to. This is in line with unimodal, behavioral studies that investigated the effects of auditory semantic context on non-native degraded speech comprehension (Bradlow and Alexander, 2007; Golestani et al., 2009; Hazan et al., 2006; Mayo et al., 1997; Oliver et al., 2012; Zhang et al., 2016), and fits with our previous behavioral study (Drijvers and Özyürek, 2019) and EEG results (Drijvers and Özyürek, 2018).

5. Conclusion

Our data revealed that spatiotemporal oscillatory dynamics can predict how much a listener benefits from semantic information conveyed by gestures when speech comprehension is challenged by internally (e.g. non-nativeness) and externally (e.g., speech degradation) induced adverse listening conditions. Our behavioral results suggested that although native and non-native listeners revealed a similar gestural enhancement effect in the cued-recall task, non-native listeners were significantly slower than native listeners when indicating which verb they heard in the video. This suggests that non-native listeners possibly faced more difficulty unifying the degraded auditory cues with the semantic information conveyed by gestures. In line with this interpretation, the observed oscillatory modulations in both non-native and native listeners suggest similar core processes that support unification and lexical access processes, as well as simulation of the gesture and increased visual attention to gestures to aid degraded speech comprehension. However, compared to native listeners, non-native listeners might have less access to the phonological information in the degraded signal, as demonstrated by less engagement of the mouth area of the primary somatosensory cortex and left insula. Moreover, non-native listeners might experience more difficulty unifying the semantic information conveyed by the gesture with the speech signal, causing areas that are involved in unification and retrieval (i.e., LIFG and ATL) to be less engaged.

Acknowledgements

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

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.neuroimage.2019.03.032.

References

Abdullaev, Y.G., Melnichuk, K.V., 1997. Cognitive operations in the human caudate nucleus. Neurosci. Lett. 234, 151–155. https://doi.org/10.1016/S0304-3940(97)00680-0.

Bamiou, D.E., Musiek, F.E., Luxon, L.M., 2003. The insula (Island of Reil) and its role in auditory processing: literature review. Brain Res. Rev. 42, 143–154. https://doi.org/10.1016/S0165-0173(03)00172-3.

Bastiaansen, M.C.M., Knösche, T.R., 2000. Tangential derivative mapping of axial MEG applied to event-related desynchronization research. Clin. Neurophysiol. 111, 1300–1305. https://doi.org/10.1016/S1388-2457(00)00272-8.

Bell, A.J., Sejnowski, T.J., 1995. An information-maximization approach to blind separation and blind deconvolution. Neural Comput. 7, 1129–1159. https://doi.org/10.1162/neco.1995.7.6.1129.

Biau, E., Soto-Faraco, S., 2015. Synchronization by the hand: the sight of gestures modulates low-frequency activity in brain responses to continuous speech. Front. Hum. Neurosci. 9, 527. https://doi.org/10.3389/fnhum.2015.00527.

Biau, E., Torralba, M., Fuentemilla, L., de Diego Balaguer, R., Soto-Faraco, S., 2015. Speaker's hand gestures modulate speech perception through phase resetting of ongoing neural oscillations. Cortex 68, 76–85. https://doi.org/10.1016/j.cortex.2014.11.018.

Binder, J.R., Desai, R.H., Graves, W.W., Conant, L.L., 2009. Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. Cerebr. Cortex 19, 2767–2796. https://doi.org/10.1093/cercor/bhp055.

Boersma, P., Weenink, D., 2015. Praat: Doing Phonetics by Computer [WWW Document]. Praat doing phonetics by Comput. [Computer program]).

Booth, J.R., Wood, L., Lu, D., Houk, J.C., Bitan, T., 2007. The role of the basal ganglia and cerebellum in language processing. Brain Res. 1133, 136–144. https://doi.org/10.1016/j.brainres.2006.11.074.

Bradlow, A.R., Alexander, J.A., 2007. Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. J. Acoust. Soc. Am. 121, 2339–2349. https://doi.org/10.1121/1.2642103.

Bushara, K.O., Hanakawa, T., Immisch, I., Toma, K., Kansaku, K., Hallett, M., 2002. Neural correlates of cross-modal binding. Nat. Neurosci. 6, 190.

Caetano, G., Jousmaki, V., Hari, R., 2007. Actor's and observer's primary motor cortices stabilize similarly after seen or heard motor actions. PNAS 104, 9058–9062.

Dahl, T.I., Ludvigsen, S., 2014. How I see what you're saying: the role of gestures in native and foreign language listening comprehension. Mod. Lang. J. 98, 813–833. https://doi.org/10.1111/j.1540-4781.2014.12124.x.

Dick, A.S., Goldin-Meadow, S., Solodkin, A., Small, S.L., 2012. Gestures in the developing brain. Dev. Sci. 15, 165–180. https://doi.org/10.1111/j.0013-9580.2004.458002.x.

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. Hum. Brain Mapp. 35, 900–917. https://doi.org/10.1002/hbm.22222.

Drijvers, L., Ozyürek, A., 2017. Visual context enhanced: the joint contribution of iconic gestures and visible speech to degraded speech comprehension. J. Speech Lang. Hear. Res. 60, 212–222. https://doi.org/10.1044/2016_JSLHR-H-16-0101.

Drijvers, L., Özyürek, A., 2019. Non-native listeners benefit less from gestures and visible speech than native listeners during degraded speech comprehension. Langages and Speech. Advance online publication. https://doi.org/10.1177/0023830919831311.

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 Lang. 177–178, 7–17. https://doi.org/10.1016/j.bandl.2018.01.003.
Drijvers, L., Özyürek, A., Jensen, O., 2018a. Hearing and seeing meaning in noise: alpha, beta, and gamma oscillations predict gestural enhancement of degraded speech comprehension. Hum. Brain Mapp. 39 (5), 2075–2087. https://doi.org/10.1002/hbm.23987.
Drijvers, L., Özyürek, A., Jensen, O., 2018b. Alpha and beta oscillations index semantic congruency between speech and gestures in clear and degraded speech. J. Cogn. Neurosci. 30, 1086–1097. https://doi.org/10.1162/jocn_a_01301.
Golestani, N., Rosen, S., Scott, S.K., 2009. Native-language benefit for understanding speech-in-noise: the contribution of semantics. Biling. (Camb. Engl). 12, 385–392. https://doi.org/10.1017/S1366728909990150.
Green, A., Straube, B., Weis, S., Jansen, A., Willmes, K., Konrad, K., Kircher, T., 2009. Neural integration of iconic and unrelated coverbal gestures: a functional MRI study. Hum. Brain Mapp. 30, 3309–3324. https://doi.org/10.1002/hbm.20753.
Gross, J., Kujala, J., Hamalainen, M., Timmermann, L., Schnitzler, A., Salmelin, R., 2001. Dynamic imaging of coherent sources: studying neural interactions in the human brain. Proc. Natl. Acad. Sci. U. S. A 98, 694–699. https://doi.org/10.1073/pnas.98.2.694.
Habets, B., Kita, S., Shao, Z., Ozyurek, A., Hagoort, P., 2011. The role of synchrony and ambiguity in speech-gesture integration during comprehension. J. Cogn. Neurosci. 23, 1845–1854. https://doi.org/10.1162/jocn.2010.21462.
Hagoort, P., 2013. MUC (memory, unification, control) and beyond. Front. Psychol. 4, 1–13. https://doi.org/10.3389/fpsyg.2013.00416.
Hazan, V., Sennema, A., Faulkner, A., Ortega-Llebaria, M., Iba, M., Chung, H., 2006. The use of visual cues in the perception of non- native consonant contrasts. J. Acoust. Soc. Am. 119, 1740–1751. https://doi.org/10.1121/1.2166611.
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. https://doi.org/10.1016/j.neuropsychologia.2015.04.018.
He, Y., Gebhardt, H., Str, L., Rondinone, I., Straube, B., 2011. The Missing Power: Language Mediates Sensorimotor-Related Beta Oscillations during On-Line Comprehension of Different Types of Co-speech Gesture (unpublished).
He, Y., Steines, M., Sommer, J., Gebhardt, H., Nagels, A., Sammer, G., Kircher, T., Straube, B., 2018. Spatial–temporal dynamics of gesture–speech integration: a simultaneous EEG-fMRI study. Brain Struct. Funct. 0, 1–17. https://doi.org/10.1007/s00429-018-1674-5.
Hervais-Adelman, A.G., Carlyon, R.P., Johnsrude, I.S., Davis, M.H., 2012. Brain regions recruited for the effortful comprehension of noise-vocoded words. Lang. Cognit. Process. 27, 1145–1166. https://doi.org/10.1080/01690965.2012.662280.
Higby, E., Kim, J., Obler, L.K., 2013. Multilingualism and the brain. Annu. Rev. Appl. Ling. 33, 68–101. https://doi.org/10.1017/S0267190513000081.
Hipp, J.F., Engel, A.K., Siegel, M., 2011. Oscillatory synchronization in large-scale cortical networks predicts perception. Neuron 69, 387–396. https://doi.org/10.1016/j.neuron.2010.12.027.
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. https://doi.org/10.1016/j.neuroimage.2009.08.058.
Jensen, O., Mazaheri, A., 2010. Shaping functional architecture by oscillatory alpha activity: gating by inhibition. Front. Hum. Neurosci. 4, 186. https://doi.org/10.3389/fnhum.2010.00186.
Kayser, C., Logothetis, N.K., 2009. Directed interaction between auditory and superior temporal cortices and their role in sensory integration. Front. Integr. Neurosci. 3, 1–11. https://doi.org/10.3389/neuro.07.
Kelly, S.D., Creigh, P., Bartolotti, J., 2010. Integrating speech and iconic gestures in a Stroop-like task: evidence for automatic processing. J. Cogn. Neurosci. 22, 683–694. https://doi.org/10.1162/jocn.2009.21254.
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. https://doi.org/10.1371/journal.pone.0004925.
Klepp, A., Niccolai, V., Buccino, G., Schnitzler, A., Biermann-Ruben, K., 2015. Language-motor interference reflected in MEG beta oscillations. Neuroimage 109, 438–448. https://doi.org/10.1016/j.neuroimage.2014.12.077.
Klimesch, W., Sauseng, P., Hanslmayr, S., 2007. EEG alpha oscillations: the inhibition-timing hypothesis. Brain Res. Rev. 53, 63–88. https://doi.org/10.1016/j.brainresrev.2006.06.003.
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. https://doi.org/10.1016/j.neuroimage.2007.12.018.
Lau, E.F., Phillips, C., Poeppel, D., 2008. A cortical network for semantics: (de) constructing the N400. Nat. Rev. Neurosci. 9, 920–933. https://doi.org/10.1038/Nrn2532.
Lecumberri, M.L.G., Cooke, M., Cutler, A., 2010. Non-native speech perception in adverse conditions: a review. Speech Commun. 52, 864–886. https://doi.org/10.1016/j.specom.2010.08.014.
Lemhöfer, K., Broersma, M., 2012. Introducing LexTALE: a quick and valid lexical test for advanced learners of English. Behav. Res. Methods 44, 325–343. https://doi.org/10.3758/s13428-011-0146-0.
Leonard, M.K., Torres, C., Travis, K.E., Brown, T.T., Hagler, D.J., Dale, A.M., Elman, J.L., Halgren, E., 2011. Language proficiency modulates the recruitment of non-classical language areas in bilinguals. PLoS One 6. https://doi.org/10.1371/journal.pone.0018240.
Lozano-Soldevilla, D., Ter Huurne, N., Cools, R., Jensen, O., 2014. GABAergic modulation of visual gamma and alpha oscillations and its consequences for working memory performance. Curr. Biol. 24, 2878–2887. https://doi.org/10.1016/j.cub.2014.10.017.
Maris, E., Oostenveld, R., 2007. Nonparametric statistical testing of EEG- and MEG-data. J. Neurosci. Methods 164, 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024.
Mayo, L.H., Florentine, M., Buus, S., 1997. Age of second-language acquisition and perception of speech in noise. J. Speech Lang. Hear. Res. 40, 686–693. https://doi.org/10.1044/jslhr.4003.686.
McNeill, D., 1992. Hand and Mind: what Gestures Reveal about Thought. Chicago University Press, Chicago.
Obleser, J., Weisz, N., 2012. Suppressed alpha oscillations predict intelligibility of speech and its acoustic details. Cerebr. Cortex 22, 2466–2477. https://doi.org/10.1093/cercor/bhr325.
Obleser, J., Wise, R.J.S., Alex Dresner, M., Scott, S.K., 2007. Functional integration across brain regions improves speech perception under adverse listening conditions. J. Neurosci. 27, 2283–2289. https://doi.org/10.1523/JNEUROSCI.4663-06.2007.
Oliver, G., Gullberg, M., Hellwig, F., Mitterer, H., Indefrey, P., 2012. Acquiring L2 sentence comprehension: a longitudinal study of word monitoring in noise. Biling. Lang. Cognit. 15, 841–857. https://doi.org/10.1017/S1366728912000089.
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. https://doi.org/10.1155/2011/156869.
Özyürek, A., 2014. Hearing and seeing meaning in speech and gesture: insights from brain and behaviour. Philos. Trans. R. Soc. B 369, 1–10. https://doi.org/10.1098/rstb.2013.0296.
Payne, L., Sekuler, R., 2014. The importance of ignoring: alpha oscillations protect selectivity. Curr. Dir. Psychol. Sci. 23, 171–177 (The). https://doi.org/10.1177/0963721414529145.
Peelle, J.E., 2018. Listening effort: how the cognitive consequences of acoustic challenge are reflected in brain and behavior. Ear Hear. 39, 204–214. https://doi.org/10.1097/AUD.0000000000000494.
Schroeder, C.E., Lakatos, P., Kajikawa, Y., Partan, S., Puce, A., Program, S., Hall, A.S., 2008. Neuronal oscillations and visual amplification of speech. Trends Cognit. Sci. 12, 106–113. https://doi.org/10.1016/j.tics.2008.01.002.Neuronal.
Senkowski, D., Saint-Amour, D., Höfle, M., Foxe, J.J., 2011. Multisensory interactions in early evoked brain activity follow the principle of inverse effectiveness. Neuroimage 56, 2200–2208. https://doi.org/10.1016/j.neuroimage.2011.03.075.
Shannon, R., Zeng, F.-G., Kamath, V., Wygonski, J., Ekelid, M., 1995. Speech recognition with primarily temporal cues. Science (80-.) 270, 303–304.
Skipper, J.I., Nusbaum, H.C., Small, S.L., 2006. Lending a helping hand to hearing: another motor theory of speech perception. Action to Lang. via mirror neuron Syst 250–286. https://doi.org/10.1017/CBO9780511541599.009.
Skipper, J.I., Wassenhove, V. Van, Nusbaum, H.C., Steven, L., 2007. Hearing lips and seeing voices: how cortical areas supporting speech production mediate audiovisual speech perception. Cerebr. Cortex 17, 2387–2399. https://doi.org/10.1093/cercor/bhl147 (Hearing).
Stolk, A., Todorovic, A., Schoffelen, J.M., Oostenveld, R., 2013. Online and offline tools for head movement compensation in MEG. Neuroimage 68, 39–48. https://doi.org/10.1016/j.neuroimage.2012.11.047.
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. https://doi.org/10.1371/journal.pone.0051207.
Sueyoshi, A., Hardison, D.M., 2005. The role of gestures and facial cues in second language listening comprehension. Lang. Learn. 55, 661–699. https://doi.org/10.1111/j.0023-8333.2005.00320.x.
Tettamanti, M., Moro, A., Messa, C., Moresco, R.M., Rizzo, G., Carpinelli, A., Matarrese, M., Fazio, F., Perani, D., 2005. Basal ganglia and language: phonology modulates dopaminergic release. Neuroreport 16, 397–401. https://doi.org/10.1097/00001756-200503150-00018.
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. https://doi.org/10.1016/j.neuroimage.2009.12.123.
Varela, F., Lachaux, J., Rodriguez, E., Martinerie, J., 2001. The brainweb: phase synchronization and large-scale integration. Nat. Rev. Neurosci. 2, 229.
Wang, L., Jensen, O., van den Brink, D., Weder, N., Schoffelen, J.-M., Magyari, L., Hagoort, P., Bastiaansen, M., 2012a. Beta oscillations relate to the N400m during language comprehension. Hum. Brain Mapp. 33, 2898–2912. https://doi.org/10.1002/hbm.21410.
Weiss, S., Mueller, H.M., 2012. "Too many betas do not spoil the broth": the role of beta brain oscillations in language processing. Front. Psychol. 3, 201. https://doi.org/10.3389/fpsyg.2012.00201.
Weisz, N., Hartmann, T., Müller, N., Lorenz, I., Obleser, J., 2011. Alpha rhythms in audition: cognitive and clinical perspectives. Front. Psychol. 2, 73. https://doi.org/10.3389/fpsyg.2011.00073.
Wild, C.J., Yusuf, A., Wilson, D.E., Peelle, J.E., Davis, M.H., Johnsrude, I.S., 2012. Effortful listening: the processing of degraded speech depends critically on attention. J. Neurosci. 32, 14010–14021. https://doi.org/10.1523/JNEUROSCI.1528-12.2012.
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. https://doi.org/10.1016/j.neuroimage.2009.05.066.
Willems, R.M., Özyürek, A., Hagoort, P., 2007. When language meets action: the neural integration of gesture and speech. Cerebr. Cortex 17, 2322–2333. https://doi.org/10.1093/cercor/bhl141.

Wong, C., Gallate, J., 2012. The function of the anterior temporal lobe: a review of the empirical evidence. Brain Res. 1449, 94–116. https://doi.org/10.1016/j.brainres.2012.02.017.

Wostmann, M., Herrmann, B., Wilsch, A., Obleser, J., 2015. Neural alpha dynamics in younger and older listeners reflect acoustic challenges and predictive benefits. J. Neurosci. 35, 1458–1467. https://doi.org/10.1523/JNEUROSCI.3250-14.2015.

Zhang, L., Li, Y., Wu, H., Li, X., Shu, H., Zhang, Y., Li, P., 2016. Effects of semantic context and fundamental frequency contours on Mandarin speech recognition by second language learners. Front. Psychol. 7, 1–8. https://doi.org/10.3389/fpsyg.2016.00908.

Zhao, W., Riggs, K., Schindler, I., Holle, H., 2018. Transcranial magnetic stimulation over left inferior frontal and posterior temporal cortex disrupts gesture-speech integration Transcranial magnetic stimulation over left inferior frontal and posterior temporal cortex disrupts gesture-speech integration 2. J. Neurosci. 10, 1748–17. https://doi.org/10.1523/JNEUROSCI.1748-17.2017.