Title: Oscillatory correlates of linguistic prediction and modality effects during listening to auditory-only and audiovisual sentences

Authors: Angèle Brunellière[1], Marion Vincent[1] and Laurence Delrue[2]

[1] Univ. Lille, CNRS, UMR 9193 - SCALab - Sciences Cognitives et Sciences Affectives, F-59000 Lille, France

[2] Univ. Lille, CNRS, UMR 8163 - STL - Savoirs, Textes, Langage, F-59000 Lille, France

Please address correspondence to:
Prof. Angèle Brunellière
SCALab, CNRS UMR 9193, Université de Lille, Domaine universitaire du Pont de Bois, BP 60149, 59653 Villeneuve d'Acsq, France
Tel: (+33) 3 20 41 72 04
angele.brunelliere@univ-lille.fr

Abstract:

In natural listening situations, understanding spoken sentences requires interactions between several multisensory to linguistic levels of information. In two electroencephalographical studies, we examined the neuronal oscillations of linguistic prediction produced by unimodal and bimodal sentence listening to observe how these brain correlates were affected by the sensory streams delivering linguistic information. Sentence contexts which were strongly predictive of a particular word were ended by a possessive adjective matching or not the gender of the predicted word. Alpha, beta and gamma oscillations were investigated as they were considered to play a crucial role in the predictive process. During the audiovisual or auditory-only listening to sentences, no evidence of word prediction was observed. In contrast, in a more challenging listening situation during which bimodal audiovisual streams switched to unimodal auditory stream, gamma power was sensitive to word prediction based on prior sentence context. Results suggest that prediction spreading from higher sentence levels to lower word levels is optional during unimodal and bimodal sentence listening and is observed when the listening situation is more challenging. Alpha and beta oscillations were found to decrease when semantically constraining sentences were delivered in the audiovisual modality in comparison with the auditory-only modality. Altogether, our findings bear major implications for our understanding of the neural mechanisms that support predictive processing in multimodal language comprehension.

Count: 217 words (max.: 250 words)
Keywords: oscillations, prediction, audiovisual speech, sentence processing

1. Introduction

Predictive processing is considered to play a key role in successful language comprehension, as it appears to be essential to ensure fast and accurate understanding in sentence processing (see Kuperberg & Jaeger, 2016; Pickering & Gambi, 2018, for reviews). Predictive processing in language comprehension means that words are immediately predicted from higher-level information based on the semantic and syntactic information provided by the prior context. Although evidence of this predictive processing has been widely explored during the visual presentation of written sentences (Delong et al., 2005; Fleur et al., 2020; Foucart et al., 2014, Wicha et al., 2004) and in the auditory stream (Foucart et al., 2015; Otten et al., 2007; Van Berkum et al., 2005; Wicha et al., 2003) while recording electrophysiological brain activity, little is known about oscillatory correlates of linguistic prediction in audiovisual sentences and how they are shaped by the sensory streams delivering linguistic information at sentence level.

Neural oscillations reflect rhythmic fluctuations of excitability in large neuronal ensembles and are known to coordinate neuronal activation across distributed brain regions (König & Schillen, 1991). Predictive processing in language comprehension requires strong efficient inter-network transfer of information across distributed brain regions. To investigate predictive processing in sentences, electrophysiological studies often use the cloze probability test to obtain measures of both the degree of sentence constraint and the expectancy of the target word within a sentence. Participants are asked to complete a sentence frame with the first word that comes to mind: the cloze probability of a word is defined as the proportion of participants who choose that same word to complete the sentence. The expectancy of the target word within a sentence is assessed by the cloze probability of a word and the sentence constraints are determined by the cloze probability of the first word that comes to mind. High-constraint

sentence contexts are characterized by a high cloze probability of the first word that comes to mind, whereas low-constraint sentence contexts are characterized by a low cloze probability of the first word that comes to mind.

Two different time windows placed before the predicted word are often studied as they are crucial when testing the effects of prediction (i.e., the period of sentential context or that related to the processing of adjectives preceding predicted nouns). Evidence of prediction effects in sentence processing requires identifying these effects before the occurrence of the expected words (Pickering & Gambi, 2018), in order to avoid inconclusive interpretations owing to the integration of the incoming word into sentence representation. The first time window corresponds to the processing of sentential context from which linguistic higher-level information based on contextual semantic and syntactic information can elicit strong or weak predictions of upcoming nouns depending on sentence constraints. Significant differences in brain activity between strong and weak sentence constraints during the processing of sentential context suggest predictive processing is involved in language comprehension. However, it does not provide clear evidence of the word prediction hypothesis driven by prior sentence context. Indeed, such differences between strong and weak sentence constraints in brain activity may also be interpreted as reflecting differences in the elaboration of sentence meaning. In contrast, the second time window corresponds to the processing of adjectives preceding predicted nouns and may give a more conclusive interpretation of the word prediction hypothesis. Importantly, these adjectives and predicted nouns share the same linguistic properties, which provides the opportunity to test the brain's reaction to adjectives matching the linguistic properties of predicted words or not. Significant differences in brain activity between mismatching and matching adjectives clearly support the word prediction hypothesis driven by prior sentence context, as only predictive processes can elicit such differences in brain activity.

Two recent review papers (Meyer, 2018; Prystauka & Lewis, 2019) summarized previous electrophysiological studies which measured electroencephalographic (EEG) and magnetoencephalographic (MEG) signals and focused on linguistic predictive processing while reading sentence context. A reduction of alpha/beta power was observed for high-constraint sentence contexts during sentence context processing (Piai et al., 2014; Piai et al., 2015; Rommers et al., 2017; Wang et al., 2018). Li et al. (2017) also found reduced beta power in high-constraint sentence contexts compared to low-constraint sentence contexts. Regarding studies focusing on the processing of adjectives preceding predicted nouns, Molinaro et al. (2017) observed lower beta power for gender-mismatching adjectives than for those matching the predicted nouns. This was observed only for transparent nouns in second language speakers but neither for opaque nouns nor in native speakers. Predictive linguistic processing therefore seems to depend on two factors: linguistic experience and saliency of predicted linguistic properties based on the upcoming words. In line with these findings, Bastiaansen et al. (2012) assumed that beta oscillations (16 – 20 Hz) were associated to maintain a set of information in the working memory and reflect the top–down propagation of predictions to lower processing levels. Other electrophysiological studies showed that prediction shifts the attention to the expected event, leading to the suppression of alpha activity (8 – 12 Hz) when attention is strongly engaged (Foxe et al., 1998; Mayer et al., 2016). In addition to alpha and beta oscillations associated with the engagement of top-down predictions, low gamma activity around 30 Hz (i.e. lower than 45 Hz, see Penolazzi et al., 2009; Weiss & Mueller, 2003) is known to reflect the matching between the incoming input and the top-down predictions (see Lewis & Bastiaansen, 2015; Meyer, 2018, for reviews). Accurate matching between the incoming input and top-down predictions triggers an increase in low gamma activity, so that gamma-band power increases with semantically congruent sentences compared to semantically

incongruent ones (Hald et al., 2006; Penolazzi et al., 2009; Rommers et al., 2013; Schneider et al., 2008).

Contrary to previous electrophysiological studies with written sentences, auditory signals are accompanied by visual signals during natural face-to-face conversational speech. Neural oscillations have already been examined during the audiovisual processing of speech and visually iconic gestures (Drijvers et al., 2017, 2018). Interestingly, alpha/beta power was more reduced in response to semantically mismatching vs. matching gestures when speech was clear. In contrast, beta power was less reduced in response to semantically mismatching vs. matching gestures when speech was degraded with moderate noise vocoding. The question is therefore whether this reduction in alpha/beta power is associated with the matching between the incoming input and semantic top-down predictions in audiovisual processing (i.e., less matching between incoming input and semantic top-down predictions when speech is degraded). In line with these findings, alpha/beta power may be particularly involved in predictive linguistic processing with audiovisual speech stimuli, as visual attention and the motor system are strongly engaged during the processing of interactions between speech signals and visual articulatory movements (Fridriksson, et al., 2008; Hall et al., 2005; Lange et al., 2013; Skipper et al., 2005). Beta-band oscillations have already been reported in the motor domain (see Engel & Fries, 2010, for review) and they may reflect the engagement of language production systems in the retrieval of conceptual representations during language comprehension (for review, see Prystauka & Lewis, 2019). Alpha-band oscillations are known to be associated with attentional engagement and successful listening comprehension (Boudewyn, & Carter, 2018). Until now, neural oscillations focusing on the processing of audiovisual stimuli and semantic content have only been investigated for spoken words accompanied by visually iconic gestures, but not for speech accompanied by face and lip movements at sentence level.

To our knowledge, no studies have yet explored the oscillatory correlates of word prediction from listening to sentence context in interaction with the nature of sensory streams delivering sentential information. This study aims at investigating this issue to obtain a realistic view of the dynamic brain processes involved in spoken language comprehension. Although audiovisual speech may offer substantial benefits of processing at different levels, including sub-lexical, lexical and sentence levels (e.g., Basirat et al., 2018; Brunellière et al., 2013, 2020), neuronal oscillations associated with linguistic prediction when listening to unimodal and bimodal sentences have not yet been investigated. Using the same approach as previous electrophysiological studies focusing on linguistic predictive processing of sentences (Foucart et al., 2015; Otten et al., 2007; Van Berkum et al., 2005; Wicha et al., 2003), we presented semantically constraining spoken sentences followed by a possessive adjective that either matched the gender of the expected (albeit not presented) word or not (see Figure 1). As in natural conversational speech, participants were asked to listen to spoken sentences and understand their meaning (see Figure 1A). As in Foucart et al. (2015), the expected words were never presented after the critical adjectives when participants listened to the sentences in order to avoid any interference effect between the processing of adjectives and that of predicted words, owing to strategies related to linguistic violations. From the beginning of the experiment, participants were informed that after the listening phase, they had to perform a lexical recognition task in which they had to indicate whether they had been previously exposed to these words (see Figure 1C). This task ensured the participants' attention during sentence listening and showed us that the meaning of sentence contexts had been well computed by observing more false alarms for the predicted words than for new words that were not exposed or expected during the listening phase.

Consistent with the goal of our study, we manipulated the presentation modality of sentence context so that the linguistic information was conveyed either in an auditory-only

manner or audiovisually (see Figure 1A). In a first experiment (Experiment 1), we explored the oscillatory correlates of prediction from sentence context listening in unimodal and bimodal situations without any interruption of sensory signals. However, we are sometimes exposed to challenging situations in which one sensory signal can be interrupted and missed because of defective communication tools or/and variability in the transmission of information. In a second experiment (Experiment 2), we thus explored whether oscillatory correlates of prediction from sentence context listening were more easily observable in challenging listening situations by interrupting one of the sensory streams. Sentence context was not always presented in the same modality as the adjective in Experiment 2. It either included the critical adjective and was conveyed in an auditory-only manner, or it was presented audiovisually and was followed by the critical adjective conveyed in an auditory-only manner. The realization of the two experiments allowed us to study a realistic view of the dynamic brain processes involved in spoken language comprehension.

As in the work of Molinaro et al. (2017) on written sentences, we investigated whether beta power is also a brain correlate of linguistic prediction when listening to audiovisual sentences with face and lip movements and sentences presented in an auditory-only manner. Dynamics in the alpha and gamma frequency bands were also examined as they were previously found to be involved in core predictive processes. In Experiment 1, we explored the specificity of oscillatory correlates associated with linguistic prediction depending on the modality of sentence context. As we tested the hypothesis that oscillatory correlates associated with linguistic prediction may differ depending on the modality of sentence context, oscillatory correlates of linguistic prediction were therefore analyzed separately by modality. Since the suppression of alpha/beta power may be involved in predictive linguistic processing of audiovisual stimuli owing to the engagement of visual attention and the motor system, we hypothesized that alpha/beta power suppression between expected and unexpected adjectives

may be observed with audiovisual stimuli but not with auditory-only stimuli. An alternative hypothesis is that the suppression of alpha/beta power may be associated with the processing of audiovisual stimuli independently of predictive linguistic processing. If so, alpha/beta power should be suppressed with the processing of audiovisual stimuli in comparison with unimodal stimuli presented in an auditory-only manner. This finding would reflect modality effects. In line with the assumption that accurate matching between the incoming input and top-down predictions triggers an increase in low gamma power (see Lewis & Bastiaansen, 2015), we expected that an increased gamma activity between expected and unexpected adjectives should be found with audiovisual stimuli and this prediction effect should also be shown with auditory-only stimuli.

In challenging listening situations, gamma activity was found to be higher for high- than for low-predictable words when there was more acoustic degradation, suggesting a greater benefit of top-down predictions from sentential contexts when the acoustic degradation is stronger (Obleser & Kotz, 2011). In addition, Molinaro et al. (2017) showed word prediction associated with lower beta power for gender-mismatching adjectives only in second language speakers, which is another type of challenging situations. In Experiment 2, interrupting the visual stream from audiovisual stimulations also constituted a challenging situation because the sensory input was suddenly degraded. In line with the role of predictive processing in successful language comprehension, we hypothesized that word prediction would be more strongly triggered in challenging listening situations when the incoming sensory information is difficult to analyze; higher-level information based on the semantic and syntactic information provided by the prior context would thus be more used to predict the upcoming word. As in Experiment 1, we used the same approach to investigate the oscillatory correlates of word prediction from listening to sentence context in interaction with the nature of sensory streams delivering sentential information and the oscillatory correlates associated with modality effects.

As illustrated in Figure 1, the material used was in French, unlike previous electrophysiological studies demonstrating prediction effects on words preceding a predictable noun in semantically constraining sentences (e.g., Foucart et al., 2015; Otten et al., 2007; Van Berkum et al., 2005; Wicha et al., 2003). In both experiments, we explored oscillatory activity in the alpha, beta and gamma frequency bands during the processing of critical adjectives, providing direct evidence of word prediction based on prior sentence context in interaction with the nature of sensory streams delivering sentential information.

< Insert Figure 1 here>

2. Material and Methods: Experiment 1

2.1 Participants

Thirty-two French-speaking students from the University of Lille, aged between 20 and 25 years old (21 females, mean age: 21.9; SD age: 1.48), took part in Experiment 1. We decided on a sample size that was a multiple of 4 (as it facilitated a perfect counterbalancing of participants per experimental list), similar to that used in previous electrophysiological studies focusing on neural oscillations with audiovisual materials (Drijvers et al., 2017, 2018) and larger than Molinaro et al. (2017). Participants had normal or corrected-to-normal vision and none of them self-reported any hearing or language impairment. All were right-handed as assessed by the Edinburgh Handedness Inventory (Oldfield, 1971). They received monetary compensation for their participation (10€) or credits for courses. Before the beginning of the experiment, participants gave their written informed consent. Five participants were excluded during the EEG pre-processing stage owing to excessive blinking and movement artifacts. The study (including Experiments 1 and 2) was approved by the Research Ethics Committee of the University of Lille.

2.2 Stimuli

The experimental stimuli consisted of a set of 120 pairs of strongly semantically constraining sentence frames (mean cloze probability: 0.74, SD: 0.15). Their selection was based on the classical cloze procedure in which 45 French speakers (who did not take part in Experiments 1 and 2) were asked to complete sentence contexts with the first word that came to their mind. The mean length of sentence frames was 17.7 words (9 – 18 words). Each selected sentence frame ended by the possessive adjective referring to the second person /ta/ or /tɔ̃/, which was either in accordance with the gender of the expected final noun or not (see Figure 1B). These possessive adjectives were not included in the sentence frames during the cloze procedure test. When a noun was provided as the first word that came to mind, an adjective also had to be provided with the noun to complete the sentence frame. This helped determine the cloze probability of the first word that came to mind and verify the plausibility of the use of possessive adjectives. The expected final noun was never presented at the end of the sentence and the two possessive adjectives beginning with an initial unvoiced plosive segment (here, /t/) provided a clear physical marker on the spectrogram, which helped detect the onset of target adjectives easily. In this study, we manipulated two different factors: gender agreement of the adjective with the expected final noun (congruency of gender: unexpected gender vs. expected gender) and modality of sentence presentation (audiovisual modality vs. auditory-only modality). This generated the four following experimental conditions: Auditory-only and Expected gender (AO-EG), Auditory-only and Unexpected gender (AO-UG), Audiovisual and Expected gender (AV-EG), Audiovisual and Unexpected gender (AV-UG). To avoid exposing participants to repeated presentations of the same sentence frame, we constructed four equivalent experimental lists of 30 trials per modality. All experimental modalities (AO-EG, AO-UG, AV-EG, AV-UG) were equally represented in each list. In addition to the experimental stimuli, 120 filler sentences were created to avoid the development of focused attention on the two critical adjectives in the experimental stimuli. These fillers were semantically and

grammatically congruent sentences in which one of the two critical adjectives (/ta/ or /tɔ̃/) was introduced either at the beginning or in the middle of a sentence (for example: “Quand je suis passé devant ta vitrine, ça m’a donné envie de manger un éclair”, *As I walked past your window display, it made me want to eat an eclair*).

Regarding the recording of the stimuli, a French-speaking male speaker was asked to pronounce the sentences several times with natural prosody at a normal speaking rate. The video recording featured a full-face frontal view of the speaker recorded simultaneously with the auditory stream during the production of sentences. To make sure that intonation, speaking rate, duration and visual movements were equivalent up to the word before the target adjective, we used a splicing technique (using Adobe Premiere Pro). A recording of each sentence context up to the word just before the target adjectives was selected for each sentence, so that intonation and speaking rate would sound natural (e.g., in the example shown in Figure 1B, “Tu m'as bien aidé quand j'étais indécis. Je l'ai suivi…”, *You helped me when I was undecided. I followed…*). Fragments coming from other recordings of the same sentence frame (e.g., in the example, “ta conseil” *your advice*, “ton conseil” *your advice*) then completed the sentence to create two new versions: one containing the expected adjective and the other containing the unexpected adjective. We ensured that the montage of video stimuli was not perceptible under auditory and visual streams for all stimuli by asking naïve persons to judge whether the stimuli had been pronounced naturally by the speaker. Importantly, after the adjective, auditory and visual distortions replaced the occurrence of the expected final word, so that the expected final word was not heard or perceived visually. By imposing a visual progressive wave using Adobe Premiere Pro, the visual distortion enabled the shape of the speaker to be seen, although the movements of the speaker’s face could not be perceived. For the auditory stream, we used Cool Edit and generated brown noise at the mean intensity of the two adjectives for each sentence frame. The duration of auditory and visual distortions was identical for one sentence frame with

a mean duration of 0.93 s (range: 0.7-1.54 s). Adjective onset was detected and duration values were extracted using the Praat speech editing software (version 5.3; Boersma and Weenink, 2011). The mean duration of the adjective was 0.072 s (range: 0.035-0.136 s) and did not vary significantly between the expected and unexpected gender conditions (p>.2). Filler sentences had the same distortions as critical sentences, although the duration of distortions varied between five values (0.04, 0.1, 0.2, 0.3 or 0.4 s) and they were never placed in the final part of the sentences. All audiovisual sentences were presented with a 0.28-s linear fade-in ramp and a 0.18-s linear fade-out ramp. The mean duration of sentences was 4.69 s.

2.3 Experimental procedure

As illustrated in Figure 1A, each trial began with a red fixation cross presented in the center of the monitor for 0.5 s, followed by the presentation of a sentence. A black screen then appeared for 2 s and was replaced by a grey fixation cross in the center of the monitor for 1 s. The audio part was played binaurally at a comfortable sound level via headphones, and the video part was played on a computer monitor placed 100 cm away from the participant. To minimize artefacts, participants were asked to focus their gaze on the center of the screen and to keep their eyes as still as possible. They were encouraged to avoid moving unless the grey fixation cross was displayed. Participants listened to 16 practice sentences prior to the set of four 10-min blocks of 60 trials, each containing sentences from all experimental conditions and fillers presented in random order. During the experiment, they were instructed to listen to the sentences carefully for comprehension, without any further tasks (for similar approaches, see Van Den Brink & Hagoort, 2004, Brunellière et al., 2020). After the listening task, participants performed a lexical recognition task to examine whether predictive mechanisms induced a memory trace of expected (although not presented) words (see Figure 1C). They were asked to indicate as quickly and accurately as possible whether they had heard the word or not during the listening task by pressing one of the two buttons on a response box. These response buttons

were counterbalanced across participants. Each word among a set of 320 words was presented randomly at the center of the screen. For each participant, half of the words (160) never appeared in the sentences during the listening task, and the remaining half (160) came from each experimental condition and from fillers. Among the words which never appeared during the listening task, half of them were new (80) and the other half were expected from the sentence frame, yet muted (80). Among the latter words, 20 words were expected from each experimental condition (AO-EG, AO-UG, AV-EG, AV-UG) and they were equivalent to new words (p>.2) in terms of psycholinguistic properties (lexical frequency, length and neighborhood). During both listening and lexical recognition tasks, participants sat in a shielded room.

2.4 EEG recording and pre-processing

The electrical signal was recorded from the scalp using a 128-channel Biosemi Active Two AD-box and was digitized at 1024 Hz. Two electrodes measured eye movements from the right eye and two additional electrodes were placed over the right and left mastoids. As recommended with the Biosemi Active Two AD-box, individual electrodes were adjusted to a stable offset lower than 20 mV. Artefact rejection was performed using the Cartool software (https://sites.google.com/site/cartoolcommunity/home) under a rejection criterion of 100 µV for any channel, in a segment starting 2 s before and ending 2 s after the onset of the adjectives and after the onset of sentences. When a difference amplitude from one time frame to the next in the EEG segment was superior to 100 µV over one electrode, the segment was rejected. Blinks and eye movements as well as other muscle artifacts were therefore removed. The number of accepted trials was equal across all four experimental conditions with an average of 29 accepted trials (p>.2; AO-EG: 28.6; AO-UG: 28.7; AV-EG: 28.8; AV-UG: 28.7). Time-frequency and statistical analyses were then conducted using the Fieldtrip Toolbox (Oostenveld et al., 2011). The EEG signal was first re-referenced offline to an average mastoid reference (left and right). Data were then segmented time-locked to the onset of sentences and that of the

adjective; only the segments accepted after the trial rejection step were included. Epochs were defined as 4-s segments (-2 to + 2 s) for both sentence onset and adjective onset. Segments for sentence onset were built because a time window between -0.50 and -0.20 s relatively to sentence onset was used as a baseline during the computing of time-frequency representations (TFRs). It helped rule out any bias due to the processing of sentence context and the nature of sensory streams delivering sentential information (see, supplementary material). The length of the baseline was based on that previously used by Wang et al. (2018). Finally, a bandpass filter (2nd-order Butterworth filter, 0.01 – 100 Hz) was applied.

2.5 EEG analyses: Time-frequency representations of power and statistics

Following the procedure by Wang et al. (2018), time-frequency representations (TFRs) of single trials were computed for each participant, channel and epoch, in two overlapping frequency ranges. For low frequencies ranging from 2 to 30 Hz, a 0.5-s Hanning window was applied in 2-Hz frequency steps and 0.01-s time steps. In the high frequency range (25 – 100 Hz), the Slepian multitaper approach was used. Power estimates were calculated with a 0.2-s window, 10-Hz frequency smoothing, in 5-Hz frequency steps and 0.05-s time steps. As expected, time-frequency representations of power over the baseline time window did not differ across experimental conditions. We divided the TFRs of adjective onset epochs in each experimental condition by the baseline consisting in the average of all sentence onset epochs over a time window between -0.50 and -0.20 s relatively to sentence onset (see supplementary material). A log transformation (10*log10) was then applied to power values to provide them in decibels.

As in Wang et al. (2018), we used a [-1; 1]-s time window time-locked to the target word, so that this window was long enough to observe brain activity but shorter than that used for data segmentation, in order to avoid a ringing artifact on the signal. To achieve the purpose

of the study, cluster-based permutation statistics (Maris, & Oostenveld, 2007) were performed across participants over all 128 electrodes in three different frequency bands (alpha: 8 – 12 Hz, beta: 16 – 20 Hz, gamma: 25 – 40 Hz[1]), for the [-1; 1]-s time window time-locked to adjective onset. This non-parametric statistical procedure optimally solves the multiple comparison issue. To investigate whether oscillatory activity in the alpha, beta and gamma frequency bands reflected word prediction with audiovisual and auditory-only stimuli, we compared TFRs between the EG and UG conditions in the interval after the presentation of the adjective (0 to 1 s) over each modality (AV or AO). If an effect of word prediction was found over any modality, we examined a potential two-way interaction between the two factors (Congruency of gender and Modality) by performing permutation tests on the differences in TFRs for UG minus EG between the AV and AO modalities. To establish oscillatory activity in the alpha, beta and gamma frequency bands associated with a modality effect, we compared TFRs for the AV modality to those for the AO modality in the interval prior to the presentation of the adjective (−1 to 0 s) and in a 1-s post-adjective window (0 to 1 s). All statistical comparisons were quantified using a *t*-test and a threshold of 95th quantile was applied to determine cluster candidates. Cluster-level statistics were computed by adding the *t*-values within each cluster. All adjacent data points according to the adjacent neighbors' design exceeding significance level (0.05 %) were grouped into clusters. Significance probability was calculated using the Monte Carlo method, with 1,000 random permutations. Statistical analyses with Student *t* tests and ANOVAs were also conducted on behavioral data for the lexical recognition task. By using the theory of signal detection, we tested whether participants accurately performed this task

[1] Like Wang, Hagoort, and Jensen (2018), we also quantified high gamma power with a 60–90 Hz range. Over this frequency band, there were neither significant differences between the expected and unexpected gender conditions within one modality nor a modality effect.

above chance level and whether there were more false alarms for predicted words than for new words not presented during the listening phase.

3. Results: Experiment 1

Behavioral Effects in Lexical Recognition Task: investigating the use of semantic constraints from sentence contexts

Participants performed the lexical recognition task accurately, as estimated by signal detection (*d*-prime) using hit responses (words heard during sentence listening and participant pressing the button corresponding to 'heard words') and false alarms (foils never presented during the sentence listening phase, but for which participants pressed the button corresponding to 'heard words'). Individual values of *d*-prime significantly differed from the null hypothesis of chance performance ($t(27)=10.99$, $p=10^{-8}$), showing that participants paid attention to the sentences (mean value of *d*-prime: 0.45). However, this value of *d*-prime was quite low, as we inserted words expected from the sentence frames but never presented in the listening phase. An ANOVA analysis on false alarms revealed a main effect of this manipulation: $F(4,104)=15.07$, $p=10^{-7}$. Participants produced more false alarms for expected yet unheard words than for new words (i.e., unexpected and unheard words, 0.21, $p=1.16\times10^{-4}$), suggesting that participants paid attention to the meaning of sentence contexts. There were no significant effects between experimental conditions (AO-EG: 0.42, AO-UG: 0.34, AV-EG: 0.39 AV-UG: 0.37) on false alarms.

Oscillatory activity in alpha, beta and gamma frequency bands: investigating word prediction effects with audiovisual and auditory-only stimuli

When we compared the TFRs of alpha (8 – 12 Hz), beta (16 – 20 Hz), and gamma (25 – 40 Hz) band activities for the UG and EG conditions in the interval after adjective presentation

(0 to 1 s), no significant clusters were identified for any particular modality in the alpha, beta and gamma frequency bands. Figure 2 illustrates the absence of significant clusters between the UG and EG conditions for the gamma frequency band in the AV modality.

< Insert Figure 2 here >

Oscillatory activity in alpha, beta and gamma frequency bands: investigating modality effects

We compared the TFRs of alpha (8 – 12 Hz), beta (16 – 20 Hz), and gamma (25 – 40 Hz) band activities for the AV and AO modalities prior to (−1 to 0 s) and after (0 to 1 s) adjective presentation. Figure 3 shows a significant cluster for the alpha activity in the observed data between the AV and AO modalities including all electrodes, between −1 to 0 s after adjective onset (p=4×10^{-3}). Another significant cluster including all electrodes was identified between these modalities but in a 0-to-1-s time window after adjective onset (p=9.99×10^{-4}). The alpha power in the AV modality was lower than in the AO modality prior to (−1 to 0 s) and after (0 to 1 s) adjective presentation.

< Insert Figure 3 here >

The cluster-based permutation tests on the beta activity revealed similar results to those found for the alpha activity with significant differences between the AV and AO modalities before and after adjective presentation. Figure 4 shows that these differences corresponded to a first cluster in the observed data including a set of frontocentral, right anterior and right centroparietal electrodes and beginning 1 s to 0.73 s before the onset of adjectives (p=3×10^{-2}), and to a second cluster in the observed data including almost all electrodes and occurring 0.15 s to 0.99 s after adjective onset (p=9.99×10^{-4}). Beta power was lower in the AV than in the AO modality over these two significant clusters.

< Insert Figure 4 here >

Contrary to the alpha and beta frequency bands, there was a significant difference between the AV and AO modalities over the gamma frequency band only after adjective presentation. This corresponded to a cluster in the observed data including frontocentral, central and right centroparietal electrodes and beginning 0.25 s to 0.45 s after adjective onset (p=3.96×10^{-2}). Gamma power was lower in the AV than in the AO modality over this cluster.

< Insert Figure 5 here >

4. Discussion: Experiment 1

Although the lexical recognition task showed that participants paid attention to the meaning of sentence contexts, we found no significant effect of gender congruency after the processing of target adjectives at alpha, beta and gamma frequency bands in Experiment 1. This showed no evidence of word prediction based on sentence context delivered by unimodal or bimodal streams. This may seem surprising based on prior event-related potential studies with written sentences (Delong et al., 2005; Fleur et al., 2020; Foucart et al., 2014, Wicha et al., 2004) or auditory sentences (Foucart et al., 2015; Otten et al., 2007; Van Berkum et al., 2005; Wicha et al., 2003), showing larger amplitude of electrophysiological components for unexpected than for expected gender. Unlike in most previous studies, we never presented the words expected from the semantically constraining sentence contexts. This could explain why evidence of word prediction was difficult to observe in our experiment. Indeed, auditory feedback about the expected information is known to be used for adjusting the internal prediction system until speech errors are eliminated and target speech is achieved thanks to motor control (Guenther & Vladusich, 2012). However, the absence of expected words from the semantically constraining sentence contexts may not be the only factor that reduces word prediction, as Foucart et al. (2015) did not present expected words during the listening phase

either and found an effect of word prediction from pre-nominal adjectives that mismatched the gender of a likely upcoming noun. Other factors may concern the EEG analyses and task instructions during the listening phase. While we performed time-frequency analyses, Foucart et al. (2015) examined ERP responses. However, Molinaro et al. (2017) showed that time-frequency analyses did not prevent finding a prediction effect. Unlike Foucart et al. (2015), we did not ask participants to answer comprehension questions after listening to sentences. In their study, one third of the sentences was followed by a comprehension question. Molinaro et al. (2017) also asked their participants to answer comprehension questions. They observed lower beta power for gender-mismatching adjectives than for those matching the predicted nouns in second-language speakers. Hence, using comprehension questions may have increased the use of high-level semantic constraints from sentence contexts.

The findings of Experiment 1 mainly suggest that prediction spreading from higher sentence levels to lower word levels is optional during unimodal and bimodal sentence listening. In line with the role of predictive processing in successful language comprehension, we hypothesized that word prediction would be more strongly triggered in challenging listening situations when the incoming sensory information is difficult to analyze. The higher-level information based on the semantic and syntactic information provided by prior context would thus be more used to predict the upcoming noun. In Experiment 2, we explored whether oscillatory correlates of prediction from sentence context listening were more easily observable in challenging listening situations by interrupting one of the sensory streams. Materials and methods of Experiment 2 are provided in the following section.

Regarding modality effects, the sensory streams delivering sentence context strongly affected brain dynamics in alpha, beta and gamma frequency ranges before and/or after the processing of target adjectives in Experiment 1. Prior to the presentation of target adjectives, alpha and beta oscillations were reduced by the audiovisual speech delivering sentence context

in comparison with unimodal stimuli presented in an auditory-only manner. This suppression of alpha/beta power was also observed with audiovisual stimuli in comparison to auditory-only stimuli during the processing of target adjectives. This is in line with previous studies showing that visual attention and the motor system are strongly engaged during the processing of interactions between speech signals and visual articulatory movements (Fridriksson et al., 2008; Hall et al., 2005; Lange et al., 2013; Skipper et al., 2005). The audiovisual suppression in alpha activity confirmed that audiovisual events guide attention during the processing of incoming information (e.g., Van der Burg et al., 2008a, 2008b). Beta-band oscillations have already been reported in the motor domain (see Engel & Fries, 2010, for review) and have also been observed during audiovisual integration in the superior temporal cortex (Schepers et al., 2013). Surprisingly, low gamma activity was strongly decreased by the audiovisual modality in comparison with the auditory-only modality during adjective recognition from 0.25 s. This is again consistent with previous studies showing that audiovisual speech contributes to word recognition (e.g., Brunellière et al., 2013, 2020; Buchwald et al., 2009; Fort et al., 2013). The occurrence of such adjectives in speech input can be predicted timely thanks to the parenthetical structure of sentence context in our study (for other EEG experiments with similar structures, see Brunellière et al., 2019; Brunellière et al., 2020). This can be more salient with audiovisual sentences (Brunellière et al., 2020). Visual information thus increases sensitivity to expected sensory information by temporal predictions (Peelle, & Sommers, 2015, for a review), which is in line with the stronger attentional engagement directed towards the adjective observed in alpha activity from listening to audiovisual sentences. Gamma oscillations are also known to be involved in visual processing (e.g., Tallon-Baudry & Bertrand, 1999) and selective visual attention (Fries et al., 2001).

In sum, when listening conditions were optimal, no evidence of word prediction was found. It is clear that oscillatory activity in alpha, beta and gamma frequency ranges shaped by the nature of the sensory streams during the processing of spoken sentences were independent from predictive linguistic processing. In Experiment 2, we investigated whether word prediction emerged when listening to semantically constraining sentences in challenging situations, and whether interrupting and switching sensory streams triggered word prediction in such contexts.

5. Material and Methods: Experiment 2

Thirty-two new participants were selected using the same criteria as those in Experiment 1. Only two participants were excluded during the pre-processing step because of excessive blinking and movement artifacts. The experimental stimuli and design were adapted from Experiment 1. The only difference was that the target adjective and the following noise were always presented in the auditory-only modality, irrespective of the sensory streams delivering sentence context. As a result, the AV modality in Experiment 2 presented a modality switch between sentence context and the rest of the sentence after the adjective, while the AO modality was similar to that used in Experiment 1. Data acquisition, pre-processing and analyses were the same as in Experiment 1[2].

6. Results: Experiment 2

Behavioral Effects in Lexical Recognition Task: investigating the use of semantic constraints from sentence contexts

[2] Over the high gamma activity in a 60–90 Hz range, there were neither significant differences between the expected and unexpected gender conditions within one modality, nor a modality effect.

When the individual values of *d*-prime were compared to the null hypothesis of chance performance, it showed that participants performed the recognition task above chance ($t(30)=13.91$, $p=10^{-7}$). As in Experiment 1, participants paid attention to the stimuli (Experiment 2, mean value of *d*-prime: 0.69). Moreover, this value of *d*-prime was quite low, as we inserted words expected from the sentence frames but never presented in the listening phase. An ANOVA analysis on false alarms revealed a main effect of this manipulation, $F(4,104)=22.27$, $p=10^{-15}$. Participants produced more false alarms for expected yet unheard words than for new words (i.e., unexpected and unheard words, 0.16, $p=1.7\times10^{-5}$), suggesting that they paid attention to the meaning of sentence contexts. There were no significant effects between experimental conditions (AO-EG: 0.36, AO-UG: 0.33, AV-EG: 0.37, AV-UG: 0.32) on false alarms.

Oscillatory activity in alpha, beta and gamma frequency bands: investigating word prediction effects with audiovisual and auditory-only stimuli

As in Experiment 1, we compared the TFRs of alpha (8 – 12 Hz), beta (16 – 20 Hz), and gamma (25 – 40 Hz) band activities for the UG and EG conditions in the 0-to-1-s interval after adjective presentation for each modality. There was a significant difference in gamma activity between the EG and UG conditions for the AV modality (see Figure 2). This corresponded to a cluster in the observed data including over the right anterior and central electrodes between 0.33 s and 0.55 s after the onset of adjectives ($p=1.39\times10^{-2}$). Gamma power was higher in the UG than in the EG condition over this cluster (see Figure 2). No interactive effect between congruency of gender and modality was found by performing permutation tests on the differences in TFRs for UG minus EG between the AV and AO modalities. However, a two-way repeated ANOVA was conducted over the significant cluster on mean gamma power with the two independent variables: congruency of gender (UG and EG conditions) and modality

(AV and AO modalities). It revealed a significant interaction between congruency of gender and modality ($F(1,29)=17.31$, $p=1.36\times10^{-2}$). Post-hoc Tukey *t*-tests confirmed that there was a significant difference in gamma activity between the EG and UG conditions with the AV modality ($p=3.31\times10^{-2}$), whereas this difference was not significant with the AO modality ($p=8.31\times10^{-1}$).

Oscillatory activity in alpha, beta and gamma frequency bands: investigating modality effects

We compared the TFRs of alpha (8 – 12 Hz), beta (16 – 20 Hz), and gamma (25 – 40 Hz) band activities for the AV and AO modalities in the intervals prior to (−1 to 0 s) and after (0 to 1 s) adjective presentation. Similar to Experiment 1 at alpha frequency band, a significant cluster was identified in both these intervals. There was a first cluster in the observed data including all electrodes between 1 s and 0 s before adjective presentation ($p=1.99\times10^{-3}$) and a second cluster in the observed data including all electrodes and occurring around 0 s until 0.68 s after the onset of adjectives ($p=4.99\times10^{-3}$). Over both clusters, alpha power was lower in the AV than in the AO modality (see Figure 3B).

Regarding beta frequency band, a significant cluster was identified between the AV and AO modalities (see Figure 4) in the observed data including right central and right centroparietal electrodes between 1 s to 0.09 s before adjective presentation ($p=4.59\times10^{-2}$). Figure 4B shows that beta power was lower in the AV than in the AO modality. Similar to the beta frequency band, there was a significant cluster identified between the AV and AO modalities in the interval prior to adjective presentation at gamma frequency band (see Figure 5). This corresponded to a cluster in the observed data including almost all electrodes from the scalp occurring between 1 s to 0 s before the onset of adjectives ($p=2.99\times10^{-3}$). Over this cluster, gamma power was lower in the AV than in the AO modality (see Figure 5B).

7. Discussion: Experiment 2

In Experiment 2, we found evidence of word prediction in challenging situations when switching between the two sensory streams delivering the spoken sentences. In line with the functional role of the gamma band described as reacting to the matching between top-down predictions and incoming input (see Lewis and Bastiaansen, 2015; Meyer, 2018, for reviews), gamma power was affected by the congruency of gender of the adjective qualifying the word predicted from sentence context. Although our experimental design provided clear evidence of word prediction before its presentation in challenging listening situations, observing reduced gamma activity after the incoming adjective when its gender matched that of the word predicted from sentence context is a somewhat surprising result. However, increased gamma activity for semantic violations has already been documented in sentence processing (Hagoort et al., 2004) and the matching between top-down predictions and expected word is not always accompanied by an increase in gamma activity (e.g., Hagoort et al., 2004; Li et al., 2017). Some studies have indeed suggested that attention may influence gamma activity, so that gamma fluctuations are dependent on experimental designs and task strategies (Gruber et al., 1999). Importantly, our experiment is the first to provide clear evidence of word prediction associated with gamma fluctuations in sentence processing and to show that word prediction appeared to be an optional process.

Moreover, we replicated the oscillatory changes shaped by audiovisual speech in the alpha and beta activities that we had already observed in Experiment 1 during the processing of semantically constraining contexts. This shows that such changes were specifically due to the nature of the sensory streams delivering sentence context. In addition to the changes in the alpha and beta bands, gamma power was decreased by audiovisual speech when a switch

occurred between the two sensory streams delivering spoken sentences. This suggests an attentional preparatory phase before the occurrence of the interrupted visual stream, since previous studies revealed that gamma oscillations play an attentional role (Fries et al., 2008; Gregoriou et al., 2009; Lima et al., 2011; Siegel et al., 2008). As in Experiment 1, alpha activity decreased during the processing of the target adjective. However, since the adjective was always presented in the auditory-only modality in Experiment 2, this decrease can only be interpreted as reflecting the impact of the sensory streams delivering sentence context on the processing of the following adjective. The impact of visual and auditory streams delivering sentence context on alpha activity thus persisted during the processing of incoming information. This reinforces the idea that the dynamics of oscillatory activity persist for several cycles after stimulation (Kösem et al., 2018). A recent magnetoencephalographic study showed that this persistence even affected the processing of incoming linguistic units (Kösem et al., 2018). Taken together, variations in beta and alpha frequency bands associated with modality effects confirmed that visual attention and the motor system are strongly engaged during the processing of interactions between speech signals and visual articulatory movements (Fridriksson et al., 2008; Hall et al., 2005; Lange et al., 2013; Skipper et al, 2005). In contrast, gamma activity appeared to be more sensitive to linguistic and sensory change predictions.

8. General Discussion

In the present study, we investigated the oscillatory correlates of prediction from sentence context listening and how such brain correlates are shaped by the sensory streams delivering linguistic information at sentence level. No evidence of word prediction was found when bimodal audiovisual and auditory streams were presented in a stable manner. However, when a switch from bimodal audiovisual streams to the unimodal auditory stream occurred, evidence of word prediction when listening to spoken sentences appeared in gamma power over the right anterior and central electrodes. Moreover, the processing of sentence context in this

challenging situation was also accompanied by decreased gamma power over most electrodes. In line with previous studies showing an impact of audiovisual speech during the processing of linguistic information (e.g., Brunellière et al., 2013, 2020; Buchwald et al., 2009; Fort et al., 2013), audiovisual speech benefits were observed during listening to sentence contexts and target words embedded in sentential context at various frequency bands. Alpha and beta oscillations were found to decrease when semantically constraining sentences were delivered in the audiovisual modality, compared to the auditory-only modality. The implications of these findings in light of the previous literature are discussed below.

Gamma oscillations carry predictions in challenging listening situations

According to the predictive coding theory (Bar, 2007; Friston, 2005), the brain continuously infers the probabilities of sensory input across the hierarchy of multi-level representations from higher to lower levels to predict upcoming input. During sentence listening, semantic and syntactic information from sentence context is assumed to trigger word prediction and then pre-activation at lower levels (Kuperberg & Jaeger, 2016; Pickering & Gambi, 2018). However, successful language comprehension can be guaranteed by both the word prediction process and the integration of the incoming word into sentence representation or by a larger part of one of these processes depending on contextual situations. Similarly, we did not find any evidence of word prediction in our first experiment, in which the sensory streams were presented in a stable manner from the beginning to the end of the sentence. In contrast, when the bimodal audiovisual streams switched to the unimodal auditory stream, evidence of word prediction with decreased gamma power was observed. Our findings are in line with the hypothesis that gamma power is highly related to predictive processes and reflects the matching between top-down predictions and incoming information (see Lewis & Bastiaansen, 2015; Meyer, 2018, for reviews). Both experiments in sentence processing and in multisensory semantic matching between visual and auditory objects demonstrated increased

gamma activity when the incoming information matched the predicted information (e.g., Hald et al., 2006; Penolazzi et al., 2009; Rommers et al., 2013; Schneider et al., 2008). Although these studies reported this specific pattern of findings in gamma power, other studies did not reveal gamma power changes in matching situations or increased gamma activity in mismatching linguistic situations (e.g., Hagoort et al., 2004; Li et al., 2017). Importantly, attention may influence gamma band activity by affecting the relationship between bottom-up and top-down processes. Experimental settings may act as strategies to modulate these gamma oscillations (Gruber et al., 1999). Previous EEG/MEG studies in non-linguistic domains revealed that attention increased gamma power (Friese, Daume, Göschl, König, Wang & Engel; 2016; Fries, 2009; Jensen, Kaiser & Lachaux, 2007). In line with these findings, we posit that finding mismatching adjectives for the predicted word captured attention, thus reversing the pattern of gamma activity between matching adjectives and mismatching ones.

Pickering and Gambi (2018) posited that generating predictions from higher to lower levels in language comprehension is an optional process, given that predictive processes are very consuming cognitive resources involving a large set of distributed brain areas. Our findings are in line with Pickering and Gambi's study (2018) showing evidence of word prediction during sentence processing only when speech input is degraded. It may be that word predictions from higher levels of sentence representation are engaged owing to the degradation of information from speech input. In contrast, when speech input is transmitted in optimal listening situations, bottom-up activation from speech input can operate without predictions from higher to lower levels for successful understanding of sentences. According to the predictive coding theory (Bar, 2007; Friston, 2005), challenging situations involving a switch from bimodal audiovisual streams to unimodal auditory stream would require updated predictions based on prior knowledge in order to adapt the processing of speech input. We believe that the more predictions are updated when sentence listening is difficult, the more evidence of word

prediction can be found. Future neuroimaging studies should be conducted to examine the properties of linguistic and sensory information from which predictions can be derived at various representation levels in multimodal situations. In our study, it appeared that alpha and beta frequency variations shaped by the nature of the sensory streams were independent from predictive linguistic processing.

Alpha and beta oscillations suppressed from sentence context with audiovisual speech

Power decreases in alpha and beta oscillations are usually associated with high brain activity, while increases are linked to disengagement of brain areas (see Meyer, 2018, for a review). We found decreases in alpha and beta powers when the sentence context was delivered by audiovisual speech in comparison with the auditory-only modality, suggesting a higher engagement of the brain in the audiovisual modality. Such changes in alpha activity confirm that audiovisual events guide attention during the processing of incoming information (Van Der Burg et al., 2008a, 2008b). Some studies have shown that the suppression of alpha activity is connected to behavioral measures of performance (e.g., Boudewyn & Carter, 2018; Haegens et al., 2011). A recent study on the processing of spoken sentences showed that reduced alpha activity was associated with attentional engagement and produced successful comprehension (Boudewyn & Carter, 2018). Beta oscillations (16 Hz – 20 Hz) tend to be related to maintaining information in the working memory and motor engagement (see Bastiaansen et al., 2012; Meyer, 2018, for reviews). During the processing of written sentences, beta power was reduced with high constraining sentences just before the predicted words (e.g., Li et al., 2017; Wang et al., 2018). This is consistent with the findings of previous studies on audiovisual benefits in sentence processing (e.g., Brunellière et al., 2013, 2020). However, the present study essentially sheds new light on the nature of the processes that are impacted. Moreover, a decrease in alpha

activity due to audiovisual speech persisted during the processing of incoming information independently of the later exposure to audiovisual speech. The persistence of alpha activity suppression due to audiovisual speech questions the brain areas that monitor such suppression and its links with behavioral performance in language comprehension. Our study raises issues about the role of alpha and beta oscillations in multimodal language comprehension. These oscillations should therefore be studied more closely if they are linked to the integration of the incoming word into sentence representation.

9. Conclusion

We found that word prediction is optional during unimodal and bimodal sentence listening in optimal situations. However, evidence of word prediction was related to gamma fluctuations in challenging situations, during which bimodal audiovisual streams switched to the unimodal auditory stream. Finally, we showed that audiovisual speech in spoken sentences shapes brain activity at alpha and beta frequency bands, thus demonstrating an impact on attention, memory and motor processes during the processing of linguistic information independently of word prediction processing.

Figure Captions

Figure 1. Overview of experimental procedure and stimuli. (A) Experimental procedure of listening task; (B) Example of experimental stimuli with gender manipulation. The following sentence context: “Tu m'as bien aidé quand j'étais indécis. Je l'ai suivi…” predicted the masculine word “conseil” (advice, in English). The possessive adjective referring to the second person, “ton”, is the masculine form and therefore the expected gender in this example. The possessive adjective referring to the second person, “ta”, is the feminine form and therefore the unexpected gender in this example; (C) Experimental procedure of lexical recognition task AV: Audiovisual modality; AO: Auditory-only modality; EG: Expected gender; UG: Unexpected gender.

Figure 2. Illustration of results based on gender congruency effect on 25 – 40 Hz gamma activity after target adjective for audiovisual modality in Experiments 1 and 2; (A) Spatial distribution of UG-EG difference power over significant time window found in Experiments 1 and 2. Asterisks denote significant clusters of electrodes and circles signal no significant effect of gender congruency at one electrode; (B) Power spectral density at C3 electrode from -1 s to 1 s after onset of target adjective over the two following conditions: Audiovisual modality and Expected gender (AV-EG) and Audiovisual modality and Unexpected gender (AV-UG). C3 electrode belonged to significant clusters of electrodes and was located over right frontocentral sites in Experiment 2. Target adjective started at 0 s. Grey bars denote period of significant clusters.

Figure 3. Illustration of results based on modality effect before and after target adjective in Experiments 1 and 2 at alpha band activity (8 – 12 Hz); (A) Spatial distribution of AV-AO difference power over each significant time window. Asterisks denote significant clusters of electrodes and N.S. signals if an effect of modality was not statistically significant; (B) Power spectral density at B13 electrode over each experimental condition from -1 s to 1 s after the onset of target adjective. B13 electrode belonged to significant clusters of electrodes and was located over right centroparietal sites. Target adjective started at 0 s. Grey bars denote period of significant clusters. AO-EG: Auditory-only modality and Expected gender; AO-UG: Auditory-only modality and Unexpected gender; AV-EG: Audiovisual modality and Expected gender; AV-UG: Audiovisual modality and Unexpected gender.

Figure 4. Illustration of results based on modality effect before and after target adjective in Experiments 1 and 2 at beta band activity (16 – 20 Hz); (A) Spatial distribution of AV-AO difference power over each significant time window. Asterisks denote significant clusters of electrodes and N.S. signals if an effect of modality was not statistically significant; (B) Power spectral density at B13 electrode over each experimental condition from -1 s to 1 s after the onset of target adjective. B13 electrode belonged to significant clusters of electrodes and was located over right centroparietal sites. Target adjective started at 0 s. Grey bars denote period of significant clusters. AO-EG: Auditory-only modality and Expected gender; AO-UG: Auditory-only modality and Unexpected gender; AV-EG: Audiovisual modality and Expected gender; AV-UG: Audiovisual modality and Unexpected gender.

Figure 5. Illustration of results based on modality effect before and after target adjective in Experiments 1 and 2 at gamma band activity (25 – 40 Hz); (A) Spatial distribution of AV-AO difference power over each significant time window. Asterisks denote significant clusters of electrodes and N.S. signals if an effect of modality was not statistically significant; (B) Power spectral density at B13 electrode over each experimental condition from -1 s to 1 s after the onset of target adjective. B13 electrode belonged to significant clusters of electrodes and was located over right centroparietal sites. Target adjective started at 0 s. Grey bars denote period of significant clusters. AO-EG: Auditory-only modality and Expected gender; AO-UG: Auditory-only modality and Unexpected gender; AV-EG: Audiovisual modality and Expected gender; AV-UG: Audiovisual modality and Unexpected gender.

Supplementary material. Diagram illustrating how the power of TFRs related to the adjective onset epochs was computed. A sentence example of 3s duration was shown with the baseline period (labelling baseline TFR) which occurred between -0.50 and -0.20 s relatively to sentence onset.

Acknowledgements

This research was supported by a grant for visual studies (SCV2013-2014) from the French National Research Agency (ANR-11-EQPX-0023) and European funds through the FEDER SCV-IrDIVE program, and another grant from the French National Research Agency (ANR-19-CE28-0006). It was also funded by the University of Lille (AAPEtablissement2015 & AAPEtablissement2014) and the municipal authorities in Lille (AppelLMCU2013). We are very grateful to Jordan Alves, Apolline Delobeau, Chloé Monnier and Laurent Ott for their help in selecting the stimuli and running the experiment. We also thank Benjamin Lob for recording the stimuli. They are also grateful to the anonymous reviewers for their helpful comments. The manuscript was proofread by a native-speaking English copyeditor.

References

Bar, M. (2007). The proactive brain: Using analogies and associations to generate predictions. *Trends in Cognitive Sciences*, *11*(7), 280-289. https://doi.org/ 10.1016/j.tics.2007.05.005

Basirat, A., Brunellière, A., & Hartsuiker, R. (2018). The role of audiovisual speech in the early stages of lexical processing as revealed by ERP word repetition effect. *Language Learning*, *68*, 80-101. https://doi.org/10.1111/lang.12265

Bastiaansen, M. Mazaheri, A., & Jensen, O. (2012). Beyond ERPs: oscillatory neuronal dynamics. In *The Oxford handbook of event-related potential components* (pp. 31–50): Oxford University Press.

Boersma, P. & Weenink, D. (2011). Praat: doing phonetics by computer [Computer program]. Version 3.4, retrieved 2 Jan 2011 from http://www.praat.org/

Boudewyn, M.A., & Carter, C.S. (2018). I must have missed that: Alpha-band oscillations track attention to spoken language. *Neuropsychologia*, *117*, 148-155. https://doi.org/10.1016/j.neuropsychologia.2018.05.024

Brunellière, A., Auran, C., & Delrue, L. (2019). Does the prosodic emphasis of sentential context cause deeper lexical-semantic processing? *Language, Cognition & Neuroscience*, *34*, 29-42. https://doi.org/10.1080/23273798.2018.1499945

Brunellière, A., Delrue, L., & Auran, C. (2020). The contribution of audiovisual speech to lexical-semantic processing in natural spoken sentences. *Language, Cognition & Neuroscience*, *35*, 694-711. https://doi.org/10.1080/23273798.2019.1641612

Brunellière, A., Sánchez-García, C., Ikumi, N., & Soto-Faraco, S. (2013). Visual information constrains early and late stages of spoken-word recognition in sentence context.

International Journal of Psychophysiology, *89*, 136-147. https://doi.org/10.1016/j.ijpsycho.2013.06.016

Buchwald, A.B., Winters, S.J., & Pisoni, D.B. (2009). Visual speech primes open-set recognition of spoken words. *Language & Cognitive Processes*, *24*, 580-610. https://doi.org/10.1080/01690960802536357

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, *8*, 1117-1121. https://doi.org/10.1038/nn1504

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, https://doi.org/10.1044/2016_JSLHR-H-16-0101

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. https://doi.org/10.1002/hbm.23987

Engel, A., & Fries, P. (2010). Beta-band oscillations--signalling the status quo? *Current Opinion in Neurobiology*, *20*, 156-165. https://doi.org/10.1016/j.conb.2010.02.015

Fleur, D.S., Flecken, M., Rommers, J., & Nieuwland, M.S. (2020). Definitely saw it coming? The dual nature of the pre-nominal prediction effect. *Cognition*, *204*, e104335. https://doi.org/10.1016/j.cognition.2020.104335.

Fort, M., Kandel, S., Chipot, J., Savariaux, C., Granjon, L., & Spinelli, E. (2013). Seeing the initial articulatory gestures of a word triggers lexical access. *Language and Cognitive Processes*, *28*, 1207-1223. https://doi.org/10.1080/01690965.2012.701758

Foucart, A., Martin, C.D., Moreno, E.M., & Costa, A. (2014). Can bilinguals see it coming? Word anticipation in L2 sentence reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(5), 1461-1469. https://doi.org/10.1037/a0036756

Foucart, A., Ruiz-Tada, E., & Costa, A. (2015). How do you know I was about to say "book"? Anticipation processes affect speech processing and lexical recognition. *Language, Cognition and Neuroscience*, *30*, 768-780. https://doi.org/10.1080/23273798.2015.1016047

Foxe, J.J., Simpson, G.V., & Ahlfors, S.P. (1998). Parieto-occipital–10 Hz activity reflects anticipatory state of visual attention mechanisms. *Neuroreport, 9*(17), 3929-3933. https://doi.org/10.1097/00001756-199812010-00030

Fridriksson, J., Moss, J., Davis, B., Baylis, G.C., Bonilha, L., & Rorden, C. (2008). Motor speech perception modulates the cortical language areas. *NeuroImage*, *41*, 605-613. https://doi.org/10.1016/j.neuroimage.2008.02.046

Fries, P., Reynolds, J.H., Rorie, A.E., & Desimone, R. (2001). Modulation of oscillatory neuronal synchronization by selective visual attention. *Science*, *291*(5508), 1560-1563. https://doi.org/10.1126/science.1055465

Fries, P., Scheeringa, R., & Oostenveld, R. (2008). Finding gamma. *Neuron*, *58*(3), 303-305. https://doi.org/10.1016/j.neuron.2008.04.020

Friston, K. (2005). A theory of cortical responses. *Philosophical transactions of the royal society B*, *360*, 815-836. https://doi.org/10.1098/rstb.2005.1622

Gregoriou, G.G., Gotts, S.J., Zhou, H., & Desimone, R. (2009). High-frequency, long-range coupling between pre-frontal and visual cortex during attention. *Science*, *324*, 1207-1210. https://doi.org/10.1126/science.1171402

Gruber, T., Müller, M.M., Keil, A., & Elbert, T. (1999). Selective visual-spatial attention alters induced gamma band responses in the human EEG. *Clinical Neurophysiology*, *110*, 2074-2085. https://doi.org/10.1016/s1388-2457(99)00176-5

Guenther, F.H., & Vladusich, T. (2012). A neural theory of speech acquisition and production. *Journal of neurolinguistics, 25(5)*, 408-422. https://doi.org/10.1016/j.jneuroling.2009.08.006

Haegens, S., Händel, B.F., & Jensen, O. (2011). Top-Down controlled alpha band activity in somatosensory areas determines behavioral performance in a discrimination task. *Journal of Neuroscience, 31*, 5197-5204, https://doi.org/10.1523/JNEUROSCI.5199-10.2011

Hagoort, P., Hald, L.A., Bastiaansen, M., & Petersson, K.M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, *304*(5669), 438-441. https://doi.org/10.1126/science.1095455

Hald, L. A., Bastiaansen, M.C.M., & Hagoort, P. (2006). EEG theta and gamma responses to semantic violations in online sentence processing. *Brain and Language*, *96*(1), 90-105. https://doi.org/10.1016/j.bandl.2005.06.007

Hall, D.A., Fussell, C., & Summerfield, A.Q. (2005). Reading fluent speech from talking faces: Typical brain networks and individual differences. *Journal of Cognitive Neuroscience*, *17*, 939-953. https://doi.org/10.1162/0898929054021175

König, P., & Schillen, T. (1991). Stimulus-dependent assembly formation of oscillatory responses: I. Synchronization. *Neural Computation*, *3*, 155-166. https://doi.org/10.1162/neco.1991.3.2.155.

Kösem, A., Bosker, H.R., Takashima, A., Meyer, A., Jensen, O. & Hagoort, P. (2018). Neural Entrainment Determines the Words We Hear. *Current Biology, 28,* 2867-2875, https://doi.org/10.1016/j.cub.2018.07.023

Kuperberg, G.R., & Jaeger, T.F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, *31,* 32-59. https://doi.org/10.1080/23273798.2015.1102299

Lange, J., Christian, N., Schnitzler, A. (2013). Audio–visual congruency alters power and coherence of oscillatory activity within and between cortical areas. *NeuroImage*, *79,* 111-120. http://dx.doi.org/10.1016/j.neuroimage.2013.04.064

Lewis, A.G., & Bastiaansen, M. (2015). A predictive coding framework for rapid neural dynamics during sentence-level language comprehension. *Cortex, 68,* 155-168. https://doi.org/10.1016/j.cortex.2015.02.014

Li, X., Zhang, Y., Xia, J., & Swaab, T.Y. (2017). Internal mechanisms underlying anticipatory language processing: evidence from event-related-potentials and neural oscillations. *Neuropsychologia, 102*, 70-81. http://dx.doi.org/10.1016/j.neuropsychologia.2017.05.017

Lima, B., Singer, W., Neuenschwander, S. (2011). Gamma responses correlate with temporal expectation in monkey primary visual cortex. *Journal of Neuroscience, 31*, 15919-15931 10.1523/JNEUROSCI.0957-11.2011

Maris, E. & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods, 164,* 177-190. https://doi.org/10.1016/j.jneumeth.2007.03.024.

Mayer, A., Schwiedrzik, C.M., Wibral, M., Singer, W., & Melloni, L. (2016). Expecting to see a letter: Alpha oscillations as carriers of top–down sensory predictions. *Cerebral Cortex, 26,* 3146-3160. https://doi.org/10.1093/cercor/bhv146

Meyer, L. (2018). The neural oscillations of speech processing and language comprehension: state of the art and emerging mechanisms. *European Journal of Neuroscience*, *48*, 2609-2621. https://doi.org/10.1111/ejn.13748

Molinaro, N., Giannelli, F., Caffarra, S., & Martin, C. (2017). Hierarchical levels of representation in language prediction: The influence of first language acquisition in highly proficient bilinguals. *Cognition*, *164*, 61-73. https://doi.org/10.1016/j.cognition.2017.03.012

Obleser, J., & Kotz, S.A (2011). Multiple brain signatures of integration in the comprehension of degraded speech. *NeuroImage, 55,* 713-723. https://doi.org/10.1016/j.neuroimage.2010.12.020

Oldfield, R.C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*, 97-113. https://doi.org/10.1016/0028-3932(71)90067-4

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. https://doi.org/10.1155/2011/156869

Otten, M., Nieuwland, M.S., & Van Berkum, J.A. (2007). Great expectations: specific lexical anticipation influences the processing of spoken language. *BMC Neuroscience, 8,* 89. https://doi.org/10.1186/1471-2202-8-89

Peelle, J.E, & Sommers, M.S. (2015). Prediction and constraint in audiovisual speech perception. *Cortex*, *68*, 169-181. https://doi.org/10.1016/j.cortex.2015.03.006

Penolazzi, B., Angrilli, A., & Job, R. (2009). Gamma EEG activity induced by semantic violation during sentence reading. *Neuroscience Letters, 465*(1), 74-78. https://doi.org/10.1016/j.neulet.2009.08.065

Piai, V., Roelofs, A., & Maris, E. (2014). Oscillatory brain responses in spoken word production reflect lexical frequency and sentential constraint. *Neuropsychologia, 53*, 146-156. https://doi.org/10.1016/j.neuropsychologia.2013.11.014

Piai, V., Roelofs, A., Rommers, J., & Maris, E. (2015). Beta oscillations reflect memory and motor aspects of spoken word production. *Human Brain Mapping, 36*(7), 2767-2780. https://doi.org/10.1002/hbm.22806

Pickering, M.J., & Gambi, C. (2018). Predicting while comprehending language: A theory and review. *Psychological Bulletin, 144*, 1002-1044. https://doi.org/10.1037/bul0000158

Prystauka, Y., & Lewis, A.G. (2019). The power of neural oscillations to inform sentence comprehension: A linguistic perspective. *Language and Linguistics Compass, 13*, e12347. https://doi.org/10.1111/lnc3.12347

Rommers, J., Dijkstra, T., & Bastiaansen, M.C.M. (2013). Context-dependent semantic processing in the human brain: evidence from idiom comprehension. *Journal of Cognitive Neuroscience, 25*, 762-776. https://doi.org/10.1162/jocn_a_00337

Rommers, J., Dickson, D.S., Norton, J.J., Wlotko, E.W., & Federmeier, K.D. (2017). Alpha and theta band dynamics related to sentential constraint and word expectancy. *Language, Cognition and Neuroscience, 32*(5), 576-589. https://doi.org/10.1080/23273798.2016.1183799

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, http://dx.doi.org/10.1016/j.neuroimage.2012.11.066

Schneider, T.R., Debener, S., Oostenveld, R., & Engel, A.K. (2008). Enhanced EEG gamma-band activity reflects multisensory semantic matching in visual-to-auditory object priming. *NeuroImage, 42*, 1244-1254. https://doi.org/10.1016/j.neuroimage.2008.05.033

Skipper, J.I., Nusbaum, H., & Small, S.L. (2005). Listening to talking faces: Motor cortical activation during speech perception. *NeuroImage, 25*, 76-89. https://doi.org/10.1016/j.neuroimage.2004.11.006

Siegel, M., Donner, T.H., Oostenveld, R., Fries, P., Engel, A.K. (2008). Neuronal synchronization along the dorsal visual pathway reflects the focus of spatial attention. *Neuron, 60*, 709-719. https://doi.org/10.1016/j.neuron.2008.09.010

Tallon-Baudry, C., & Bertrand, O. (1999). Oscillatory gamma activity in humans and its role in object representation. *Trends in Cognitive Sciences, 3*(4), 151-162. https://doi.org/10.1016/S1364-6613(99)01299-1

Van Berkum, J.J., Brown, C.M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 443-467. https://doi.org/10.1037/0278-7393.31.3.443

Van den Brink, D., & Hagoort, P. (2004). The influence of semantic and syntactic context constraints on lexical selection and integration in spoken-word comprehension as revealed by ERPs. *Journal of Cognitive Neuroscience, 16*, 1068-1084. https://doi.org/10.1162/0898929041502670

Van der Burg, E., Olivers, C.N.L., Bronkhorst, A.W., & Theeuwes, J. (2008a). Pip and pop: nonspatial auditory signals improve spatial visual search. *Journal of Experimental Psychology: Human Perception and Performance, 34*, 1053-1065. https://doi.org/10.1037/0096-1523.34.5.1053

Van der Burg. E., Olivers, C.N.L., Bronkhorst, A.W., & Theeuwes, J. (2008b). Audiovisual events capture attention: Evidence from temporal order judgments. *Journal of Vision, 8*, 1-10, https://doi.org/10.1167/8.5.2

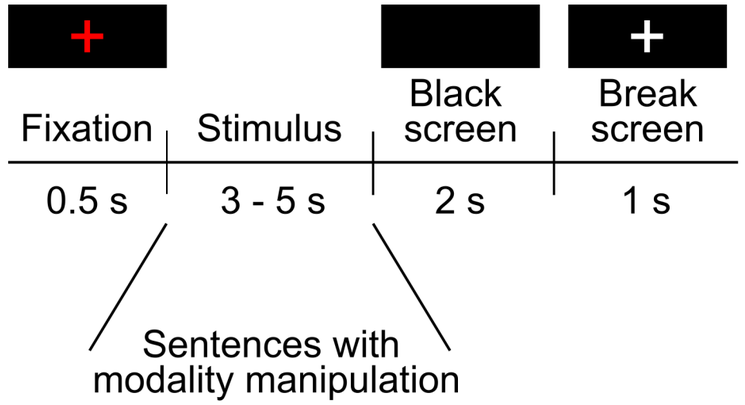
Wang, L., Hagoort, P., & Jensen, O. (2018). Language prediction is reflected by coupling between frontal gamma and posterior alpha oscillations. *Journal of Cognitive Neuroscience, 30*, 432-447. https://doi.org/10.1162/jocn_a_01190

Weiss, S., & Mueller, H. M. (2003). The contribution of EEG coherence to the investigation of language. *Brain and Language, 85*(2), 325–343. https://doi.org/10.1016/S0093-934X(03)00067-1

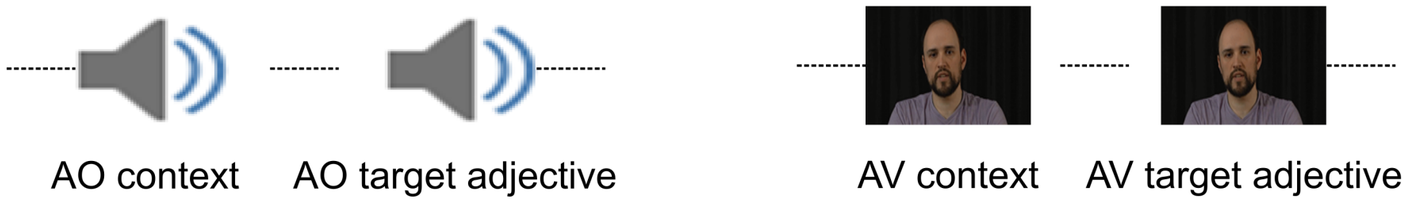
Wicha, N.Y.Y, Bates, E.A, Moreno, E.M., & Kutas, M. (2003). Potato not pope: Human brain potentials to gender expectation and agreement in Spanish spoken sentences. *Neuroscience Letters, 346*, 165-168. https://doi.org/10.1016/S0304-3940(03)00599-8

Wicha, N.Y., Moreno, E.M., & Kutas, M. (2004). Anticipating words and their gender: An event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. *Journal of Cognitive Neuroscience, 16*(7), 1272-1288. https://doi.org/10.1162/0898929041920487

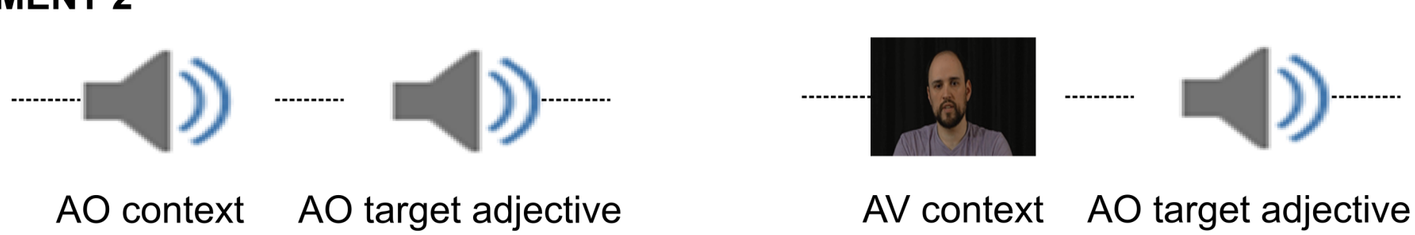
A **LISTENING AND COMPREHENSION TASK**



EXPERIMENT 1



EXPERIMENT 2

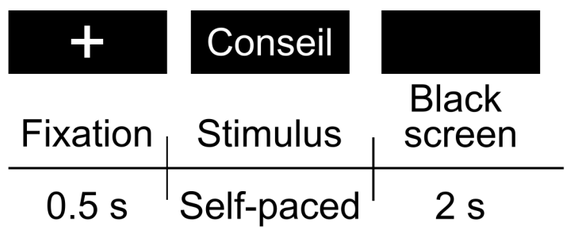


B **EXAMPLE OF SENTENCES WITH GENDER MANIPULATION**

Conditions	Sentence Context	Target Adjective
EG	Tu m'as bien aidé quand j'étais indécis. Je l'ai suivi *You helped me when I was undecided. I followed*	ton *your*
UG	Tu m'as bien aidé quand j'étais indécis. Je l'ai suivi *You helped me when I was undecided. I followed*	ta *your*

C **LEXICAL RECOGNITION TASK**

EXPERIMENTS 1 AND 2

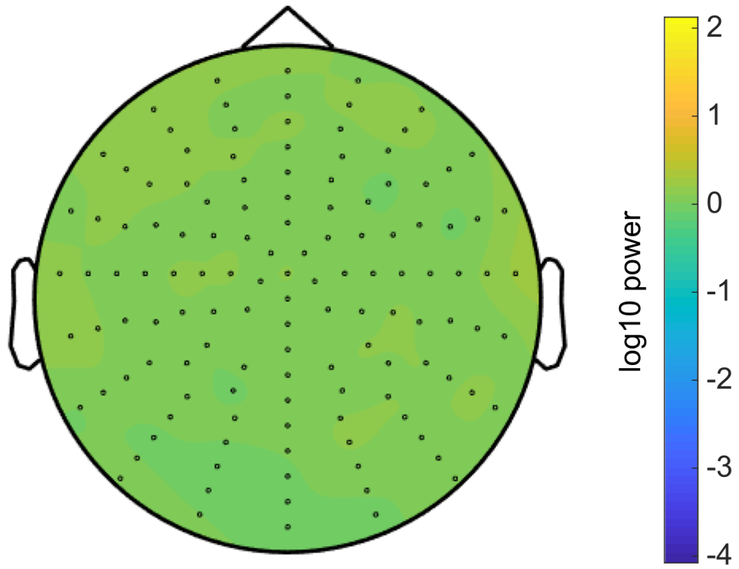


GENDER CONGRUENCY EFFECT ON GAMMA (25 - 40 Hz)

EXPERIMENT 1

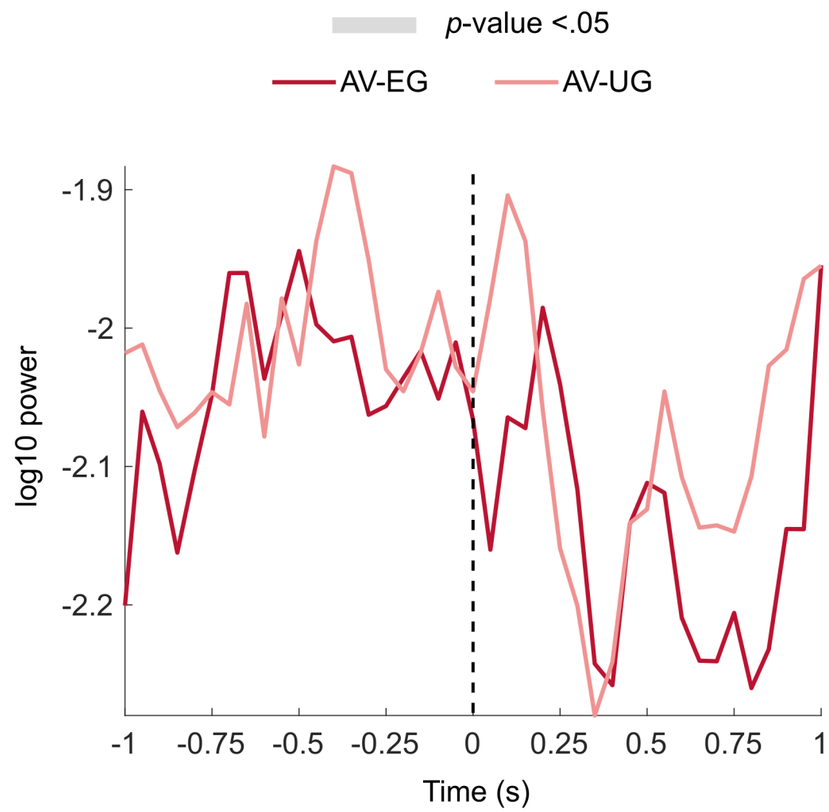
A SPATIAL DISTRIBUTION OF UG-EG DIFFERENCE POWER IN AV MODALITY

* *p*-value <.05

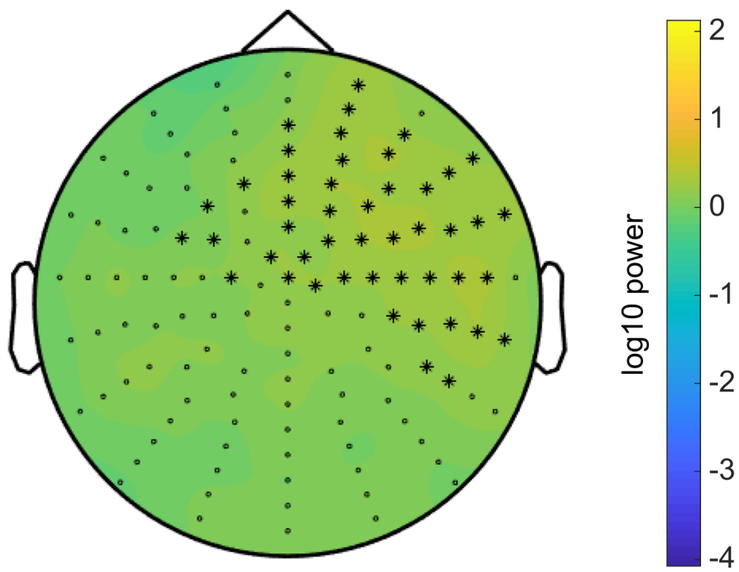


t = [0.33 ; 0.55] s

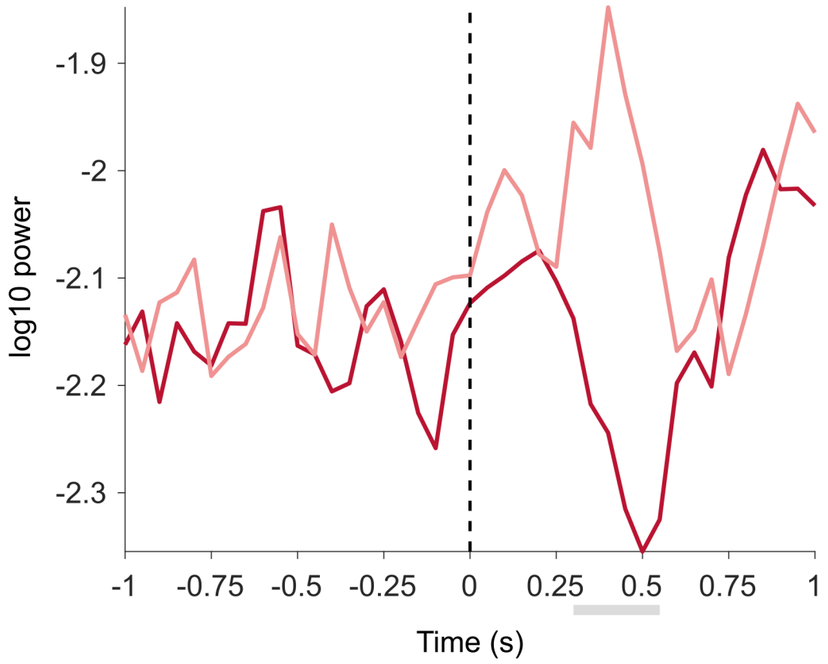
B POWER SPECTRAL DENSITY



EXPERIMENT 2



t = [0.33 ; 0.55] s



EFFECT OF MODALITY - ALPHA (8 - 12 Hz)

EXPERIMENT 1

EXPERIMENT 2

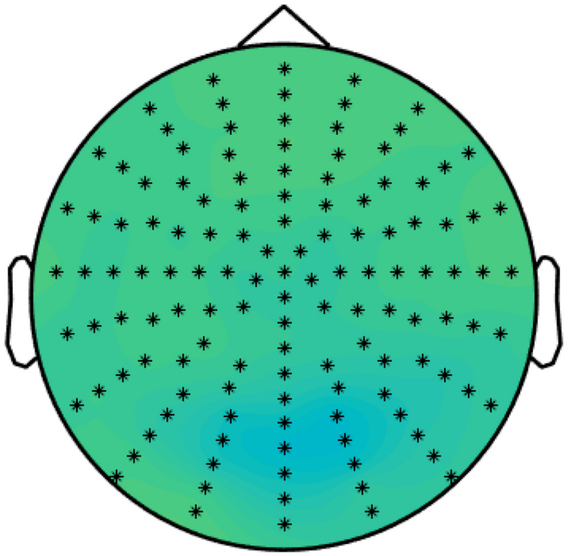
A

SPATIAL DISTRIBUTION OF AV-AO DIFFERENCE POWER

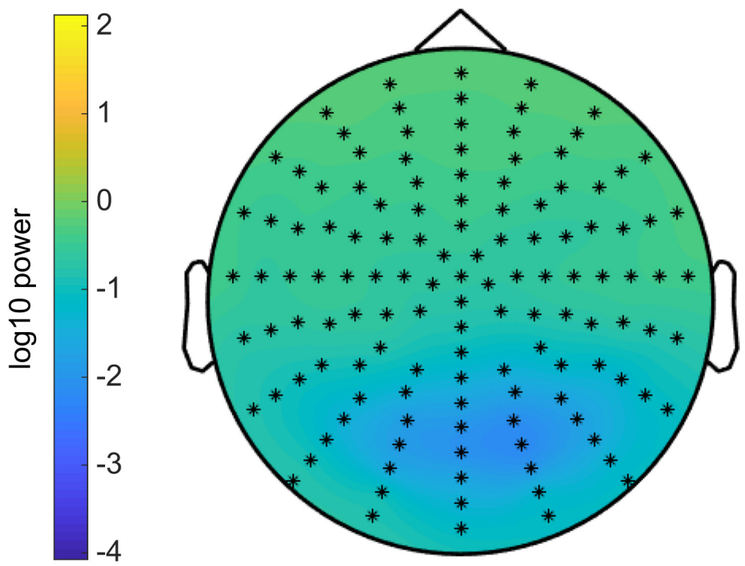
* *p*-value <.05

PRE-ADJECTIVE PERIOD

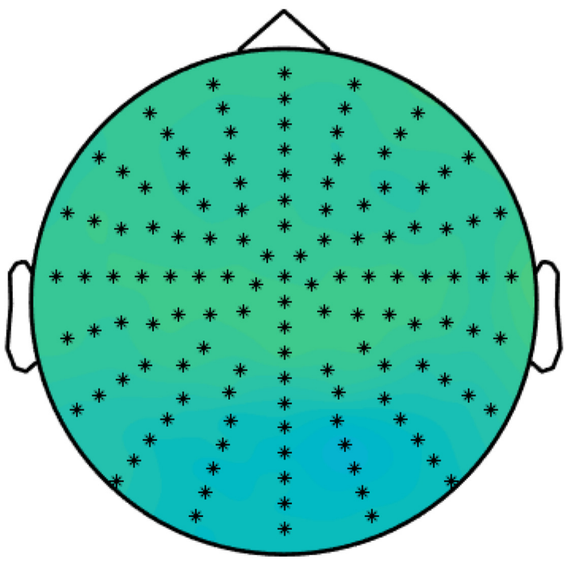
POST-ADJECTIVE PERIOD



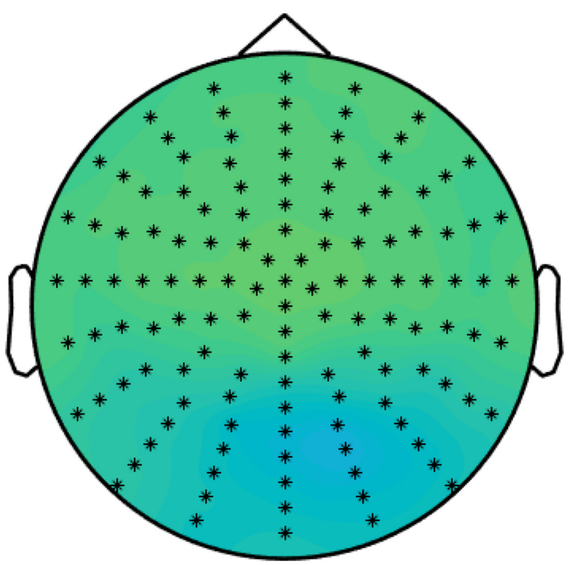
t = [-1 ; 0] s



t = [0 ; 1] s



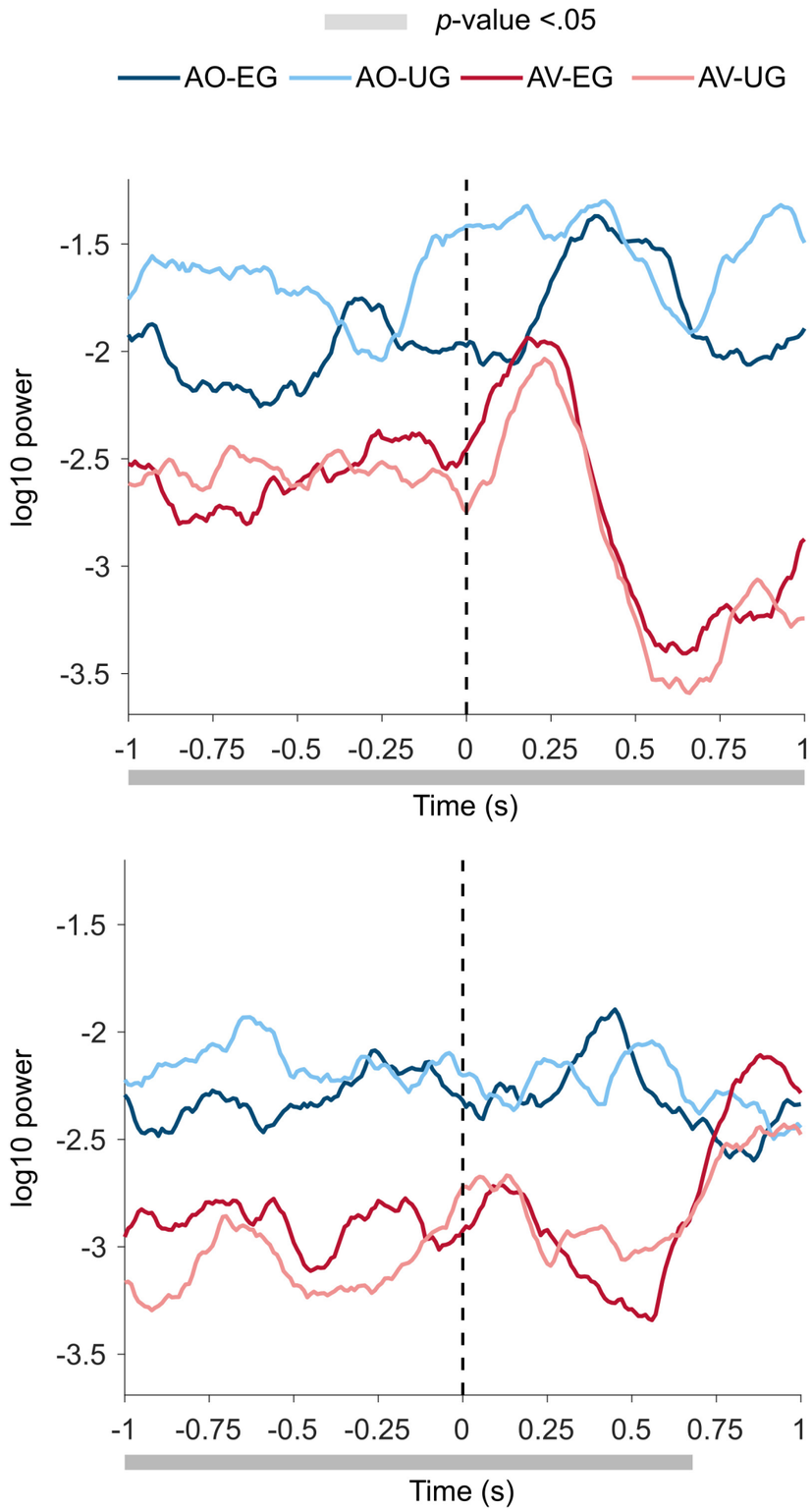
t = [-1 ; 0] s



t = [0 ; 0.68] s

B

POWER SPECTRAL DENSITY



EFFECT OF MODALITY - BETA (16 - 20 Hz)

EXPERIMENT 1

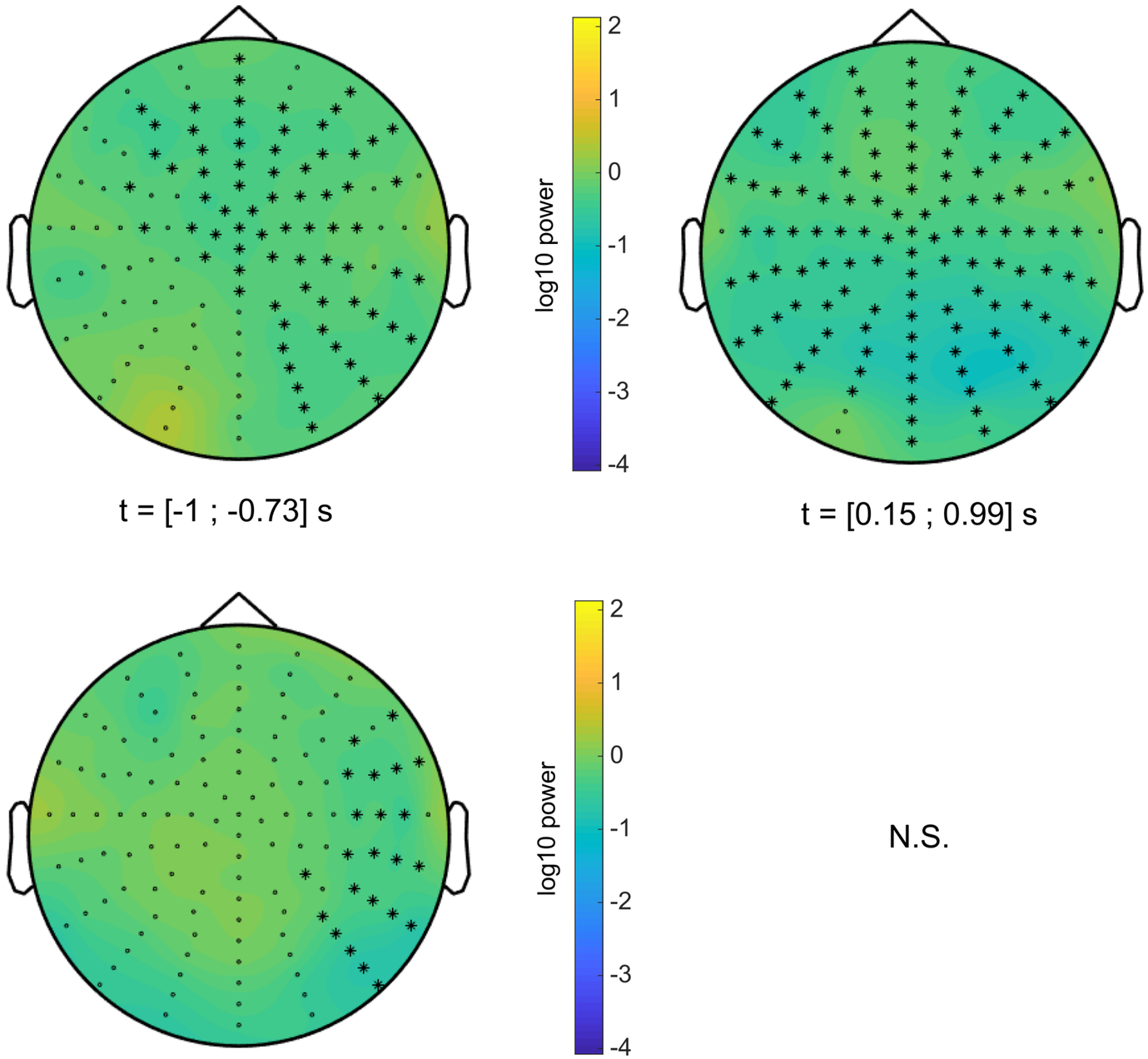
EXPERIMENT 2

A SPATIAL DISTRIBUTION OF AV-AO DIFFERENCE POWER

∗ *p*-value <.05

PRE-ADJECTIVE PERIOD

POST-ADJECTIVE PERIOD



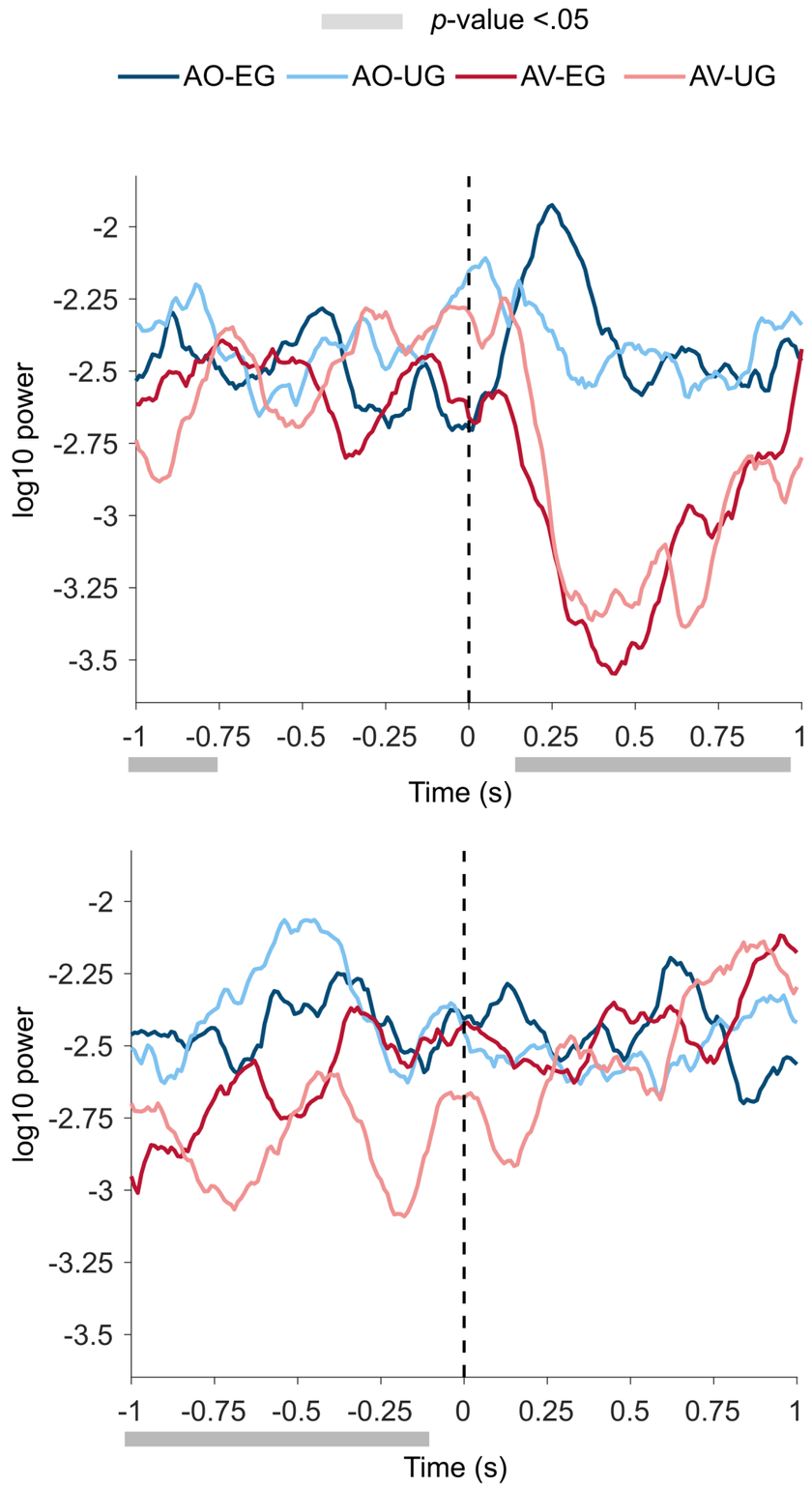
t = [-1 ; -0.73] s

t = [0.15 ; 0.99] s

t = [-1 ; -0.09] s

N.S.

B POWER SPECTRAL DENSITY



EFFECT OF MODALITY - GAMMA (25 - 40 Hz)

EXPERIMENT 1

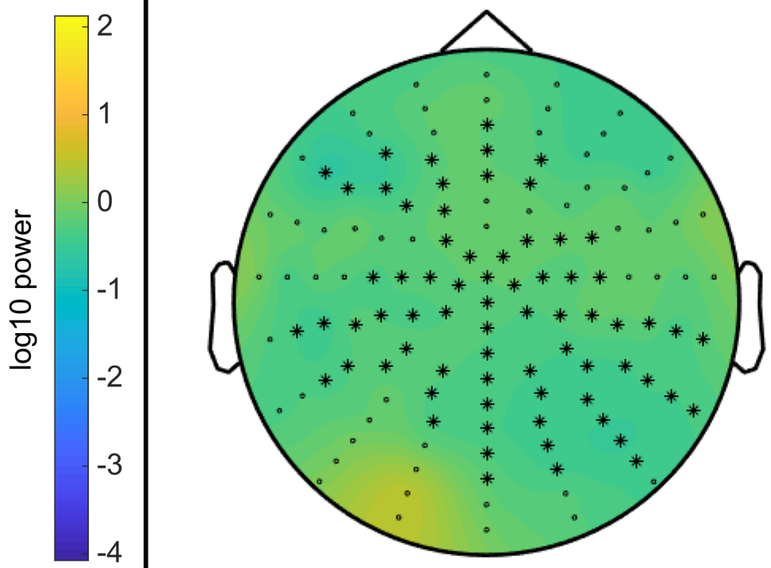
A SPATIAL DISTRIBUTION OF AV-AO DIFFERENCE POWER

∗ *p*-value <.05

PRE-ADJECTIVE PERIOD

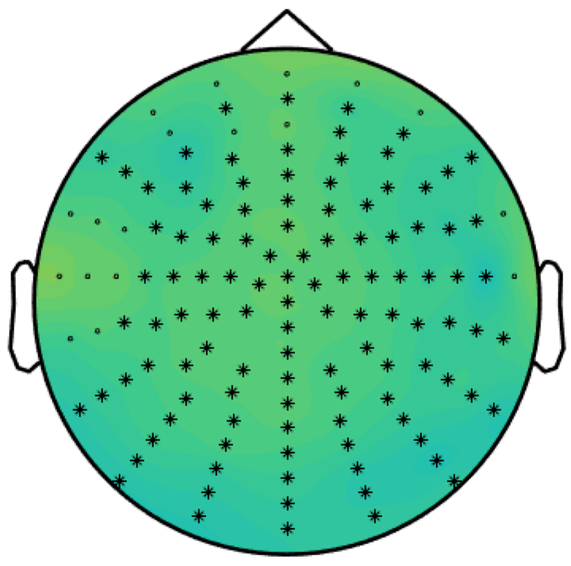
POST-ADJECTIVE PERIOD

N.S.

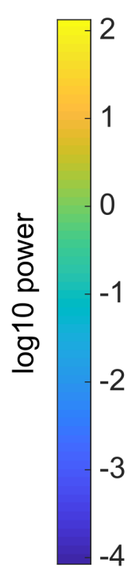


t = [0.25 ; 0.45] s

EXPERIMENT 2

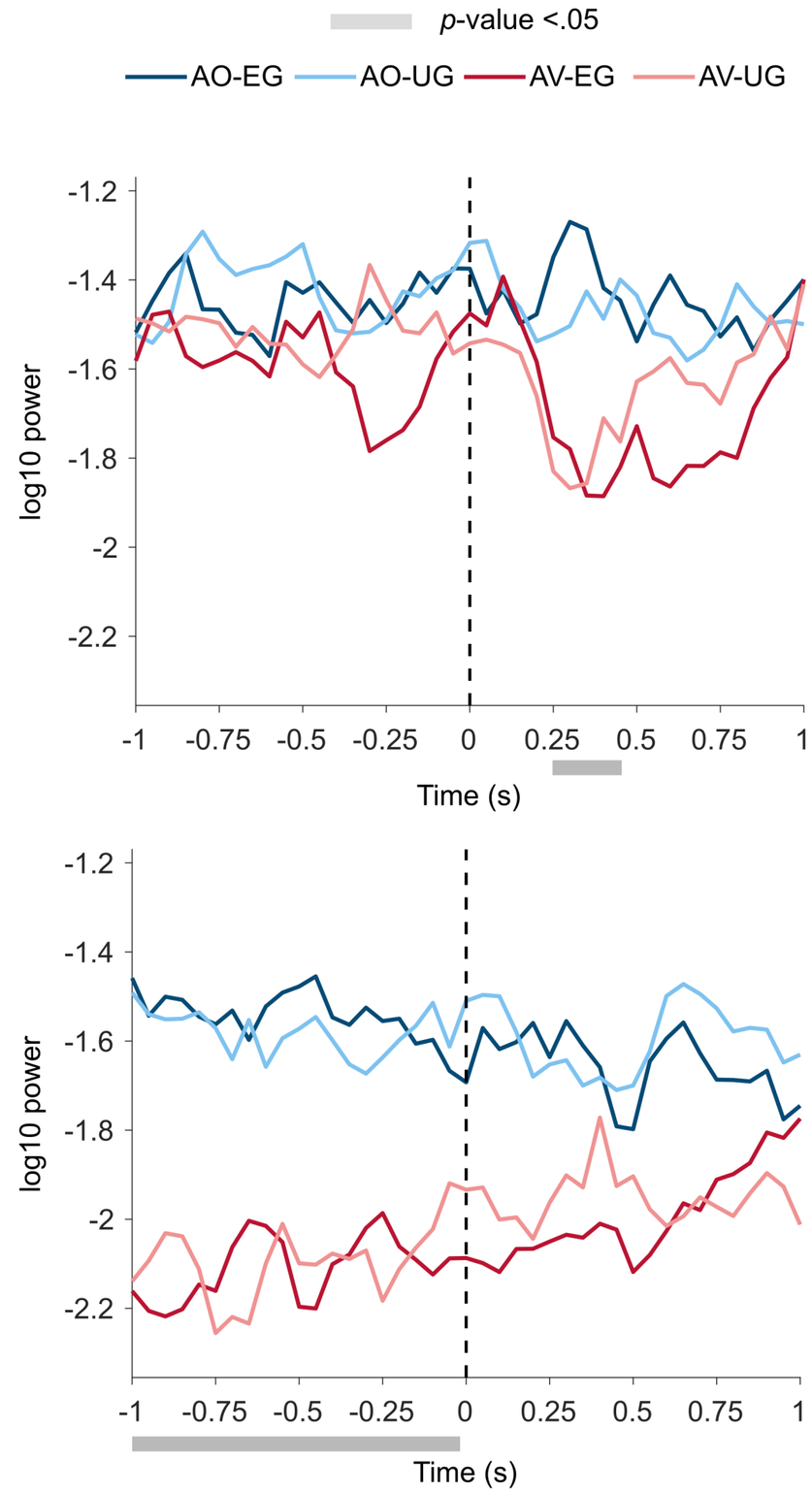


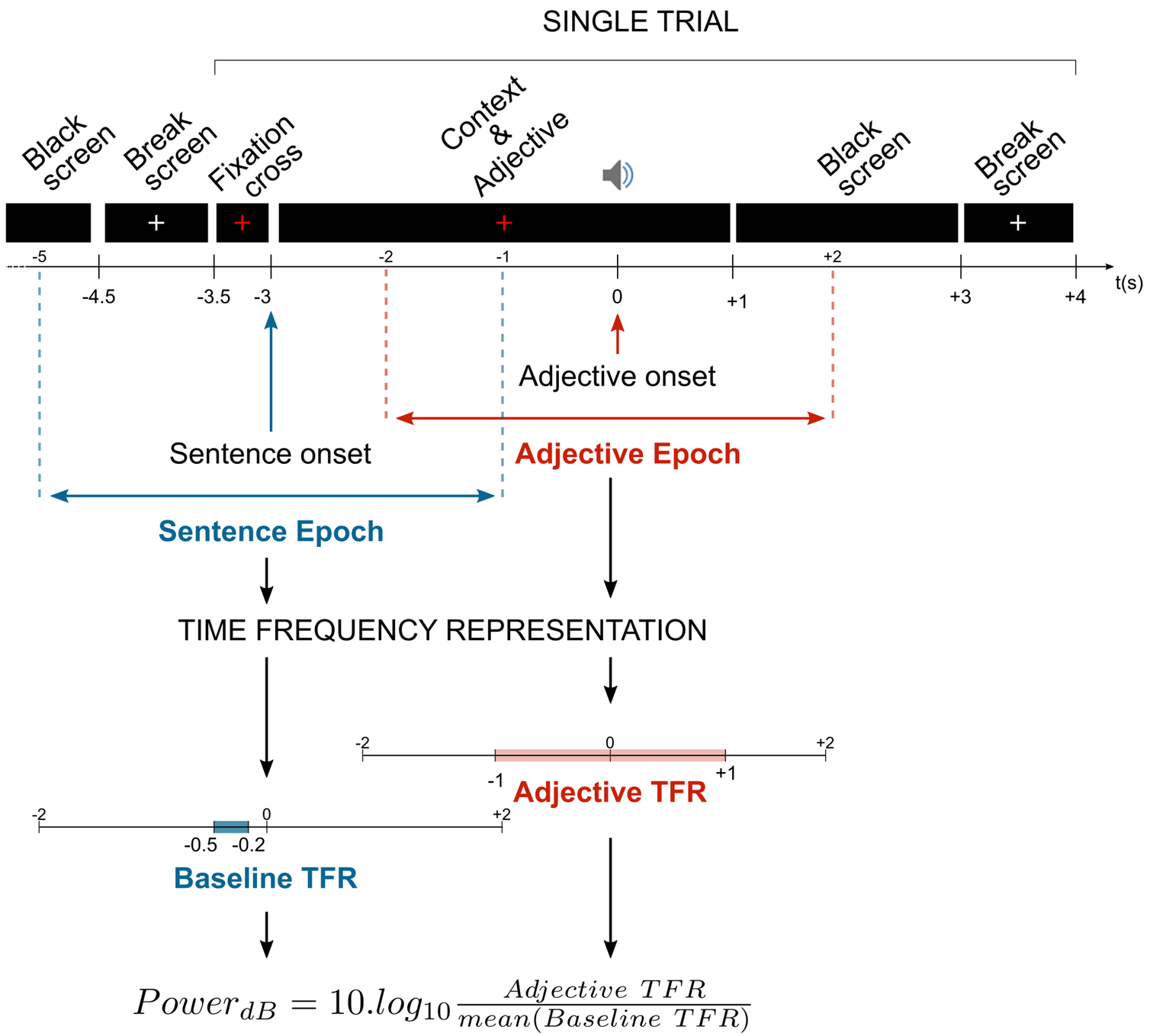
t = [-1 ; 0] s



N.S.

B POWER SPECTRAL DENSITY




SINGLE TRIAL
Black screen
Break screen
Fixation cross
Context & Adjective
Black screen
Break screen
-5
-4.5
-3.5
-3
-2
-1
0
+1
+2
+3
+4
t(s)
Adjective onset
Sentence onset
Adjective Epoch
Sentence Epoch
TIME FREQUENCY REPRESENTATION
-2
0
+2
-1
+1
Adjective TFR
-2
0
+2
-0.5
-0.2
Baseline TFR
$Power_{dB} = 10.log_{10} \frac{Adjective\ TFR}{mean(Baseline\ TFR)}$