CORTEX 68 (2015) 76–85

Available online at www.sciencedirect.com

ScienceDirect

Journal homepage: www.elsevier.com/locate/cortex

Special issue: Research report

Speaker's hand gestures modulate speech perception through phase resetting of ongoing neural oscillations

Emmanuel Biau [a], *Mireia Torralba* [a], *Lluis Fuentemilla* [c,d], *Ruth de Diego Balaguer* [c,d] *and Salvador Soto-Faraco* [a,b,*]

[a] *Multisensory Research Group, Center for Brain and Cognition, Universitat Pompeu Fabra, Barcelona, Spain*
[b] *Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain*
[c] *Department of Basic Psychology, University of Barcelona, Barcelona, Spain*
[d] *Bellvitge Research Biomedical Institute (IDIBELL), Cognition and Brain Plasticity Unit, Barcelona, Spain*

ARTICLE INFO

Article history:
Received 8 July 2014
Reviewed 27 August 2014
Revised 21 October 2014
Accepted 17 November 2014
Published online 20 December 2014

Keywords:
Beats
Audiovisual speech
Low frequency oscillations
EEG

ABSTRACT

Speakers often accompany speech with spontaneous beat gestures in natural spoken communication. These gestures are usually aligned with lexical stress and can modulate the saliency of their affiliate words. Here we addressed the consequences of beat gestures on the neural correlates of speech perception. Previous studies have highlighted the role played by theta oscillations in temporal prediction of speech. We hypothesized that the sight of beat gestures may influence ongoing low-frequency neural oscillations around the onset of the corresponding words. Electroencephalographic (EEG) recordings were acquired while participants watched a continuous, naturally recorded discourse. The phase-locking value (PLV) at word onset was calculated from the EEG from pairs of identical words that had been pronounced with and without a concurrent beat gesture in the discourse. We observed an increase in PLV in the 5–6 Hz theta range as well as a desynchronization in the 8–10 Hz alpha band around the onset of words preceded by a beat gesture. These findings suggest that beats help tune low-frequency oscillatory activity at relevant moments during natural speech perception, providing a new insight of how speech and paralinguistic information are integrated.

1. Introduction

Speakers naturally accompany their discourse with spontaneous hand gestures that highlight relevant points in speech and help to structure their narrative (Casasanto & Jasmin, 2010; McNeill, 1992). These so-called 'beat' gestures are rapid biphasic movements of the hands devoid of semantic content but exquisitely synchronized with lexical prosody. Indeed, the

* *Corresponding author*. Deptartment de Tecnologies de la Informació i les Comunicacions, Universitat Pompeu Fabra, Roc Boronat, 138, 08018 Barcelona, Spain.
E-mail address: salvador.soto@icrea.cat (S. Soto-Faraco).
http://dx.doi.org/10.1016/j.cortex.2014.11.018
0010-9452/

apex (i.e., the maximum extension phase of the movement) co-occurs rather consistently with the stressed syllable of the affiliate word (Krahmer & Swerts, 2007) conferring a reliable temporal sequence between visual (hand beat) and acoustic (word) information (Biau & Soto-Faraco, 2013; Leonard & Cummins, 2011). Previous ERP studies have investigated the integration between gestures describing an object (i.e., iconic gestures) and concurrent speech (Habets, Kita, Shao, Ozyurek, & Hagoort, 2011; Obermeier & Gunter, 2015; Obermeier, Holle, & Gunter, 2011). For instance, Obermeier and Gunter (2014) showed that the ERP signature of semantic integration between gesture and speech was affected by incongruence when an approximate temporal overlap between the gesture fragment and its affiliate word was maintained (between −200 msec and +120 msec around identification point of word). However, one must note that contrary to the iconic gestures used by Obermeier and Gunter (2015), beat gestures do not convey semantic content, and therefore the temporal relationship between the apex and the target word might be more important. Indeed, Treffner, Peter, and Kleidon (2008) showed that when the apex of a beat was aligned with a word, this word was perceived as more prominent in the utterance suggesting that the correct integration between beats and speech is synchrony-dependent and might occur within a narrower time window. Another argument for the importance of alignment in time regards the functional role of beats: As the production of a beat modulates acoustic features of the stressed word in speech production and its perceived saliency properties for the listener (Krahmer & Swerts, 2007), one could infer that beats have a predictive value, i.e., are susceptible to diminish the uncertainty about when the corresponding acoustic cues (word onsets) occur, to facilitate speech processing (Arnal & Giraud, 2012). In the present study, we advance a possible neural mechanism whereby visual information from the speaker's gestures is integrated with auditory information on the listener's side in natural speech context.

One important feature of beat gestures bears on their natural correspondence with syllabic rhythm of speech (i.e., stressed syllables; Krahmer & Swerts, 2007), which is one of the most relevant organizational units in spoken language production (Peelle & Davis, 2012) and perception (Ghitza & Greenberg, 2009; Mehler, Dommergues, Frauenfelder & Seguí, 1981). Indeed, the speech signal is characterized by envelope modulations organized around a syllable-based temporal pattern (4–8 Hz) within the acoustic flow (Giraud & Poeppel, 2012; Greenberg, 1999; Peelle & Davis, 2012). Luo and Poeppel (2007) showed that during speech perception neural activity in the auditory cortex undergoes phase synchronization within theta range matching the syllable-based spectro-temporal structure of the utterance. Of particular relevance for the present study, phase synchronization of neural oscillations in the range of 4–8 Hz has been proposed as a potential mechanism enabling predictive coding based on the correlation between audio-visual speech cues (Arnal & Giraud, 2012; Lakatos, Karmos, Mehta, Ulbert, & Schroeder, 2008; Schroeder & Lakatos, 2009; Schroeder, Lakatos, Kajikawa, Partan, & Puce, 2008). In particular, Arnal and Giraud (2012) argued that this resonance between theta oscillatory activity and regular relevant acoustic cues reflects temporal anticipation that can be facilitated by preceding visual speech information. It must be noted that the sight of human actions, in general, has been previously associated with a decrease in the alpha and beta rhythmic activity in the EEG (for example Muthukumaraswamy & Johnson, 2004; Pfurtscheller, Neuper, & Krausz, 2000). In the case of gestures, Quandt, Marshall, Shipley, Beilock, and Goldin-Meadow (2012) showed that the power in the alpha/beta frequency bands was sensitive to the characteristic of social gestures, as iconic gestures affected more the alpha/beta activities than simple deictic gestures. These results suggested that the neural system is sensitive not only to pure action's perception but also to the communicative value of the gesture.

Given that beat gestures are strongly associated with speech prosody and bear a predictive value within the spoken signal, we hypothesize that the visual input from concurrent beat gestures might increase phase coherence of neural oscillations around the affiliate word by phase resetting of ongoing oscillatory activity. According to prior studies, this might very well be in the range of 4–8 Hz. We tested this hypothesis by presenting participants with a natural audiovisual speech discourse while recording their EEG from scalp electrodes. If the phase resetting hypothesis holds, then it should be reflected by an increase of phase coherence at low frequencies at word onsets associated with gestures, compared to equivalent words pronounced in the absence of a concurrent beat gesture.

2. Material and methods

2.1. *Participants*

Twenty native Spanish speakers (12 female; mean age = 23.8 ± 3.4 years) participated in the experiment after giving written informed consent. All participants were right-handed and had normal or corrected-to-normal vision. The experiment was approved by the local ethics committee of the University Pompeu Fabra (CIEC Parc del Mar).

2.2. *Stimuli*

We used an official discourse (17′duration) of the former Spanish President (Mr. J.L. Rodriguez Zapatero) in which only the head and the upper part of the body were visible (see Fig. 1). We selected 77 pairs of words (e.g., "*crisis*") that were uttered once with a beat gesture and once without at separate moments within the same discourse (mean word length 6.4 ± 2.3 letters; mean lexical frequency 632 ± 1717 as per the LEXESP database, Sebastián-Gallés, Martí, Carreiras, & Cuetos, 2000). Words with and without gesture were not different in average duration (respectively .50 sec ± .02 and .46 sec ± .02; $p = .68$) or average intensity (respectively 69.28 dB ± .58 and 69.36 dB ± .56; $p = .92$). Word onsets (with or without beat) were determined by visual and acoustic inspection of the spectrogram using *Praat* v.5. We selected the word synchronized with the apex visually, and analyzed the corresponding spectrogram to calculate the mean F0, F1 and F2 of each word. No difference between conditions was found for the mean F0 (gesture: 123.22Hz ± 23.37; no gesture: 121.36Hz ± 18.54;

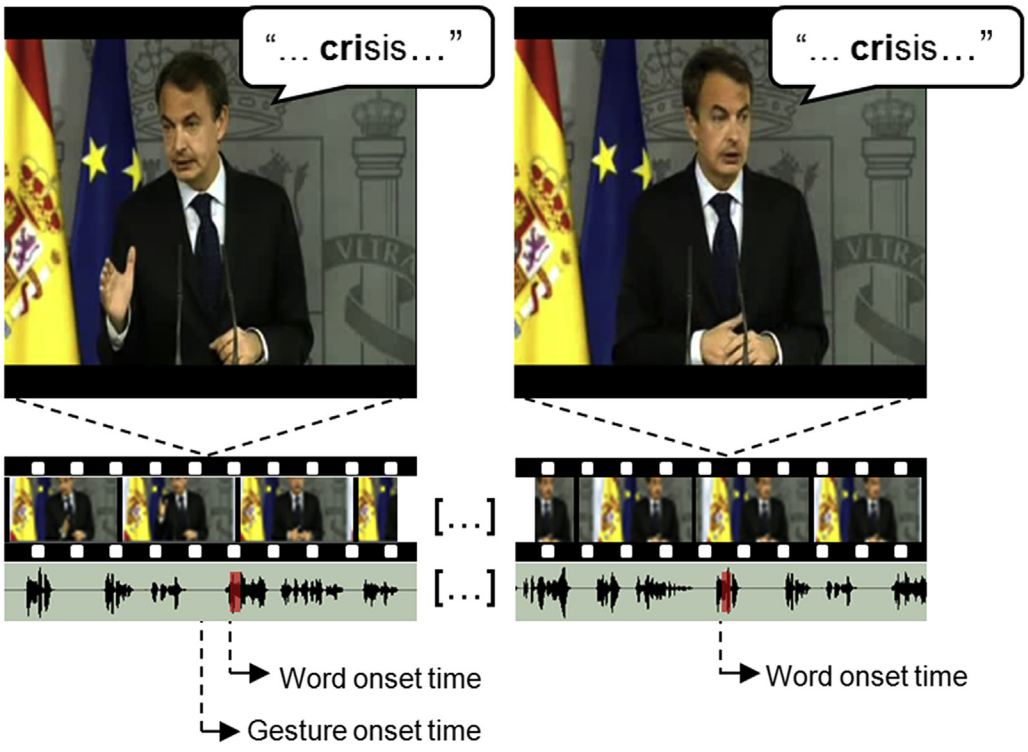

Fig. 1 – Example of video-frames for the gesture (left) and no gesture (right) conditions associated to the same stimulus word "*crisis*" at different moments within the AV discourse. The video frames are shown at the word onset time for each condition. Below, the oscillogram of corresponding audio track fragments (section corresponding to the target word shaded in red). The onset of the gesture corresponding to the word (gesture condition) is also marked on the oscillogram.

$p = .56$), F1 (gesture: 707.89 Hz ± 195.63 no gesture: 705.27 Hz ± 198.27; $p = .93$), F2 (gesture: 1963.28 Hz ± 194.40 no gesture: 1915.60 Hz ± 222.04; $p = .16$). In addition, we also marked the onset time of the gestures synchronized with the selected words in the gesture condition. To do so, we considered as onset the first video frame of the gesture (right when the preparation phase is initiated, see McNeill, 1992). On average, the gesture onset took place 200 ± 102 msec before the acoustic onset of the corresponding word.

We also created an Audio-only version (A) of the discourse by removing the visual information and presenting the sound track with a central fixation cross on a black background. This condition was included to measure possible acoustically-based effects between words pronounced with a beat gesture and equivalent words pronounced without gesture.

2.3. *Experimental procedure*

All the participants attended to the two versions but the order of presentation was counterbalanced across participants (AV first then A, or, A first then AV), through a monitor and headphones. Following a 5s fixation cross, the first version of the discourse (for example AV) began with four one-minute resting breaks inserted every 5 min. Thereafter, participants had 3 min to complete a memory questionnaire (ten two-alternative forced-choice questions with one proposition that appeared during the discourse and one that did not). After a break of 5 min, and 5 sec fixation cross, the second version of the discourse began (for example A if the first one was AV), followed by a second memory questionnaire.

2.4. *Electroencephalogram (EEG) signal recording*

EEG data were recorded (Brain Vision Recorder 1.05; Brain Products) during the discourse at a rate of 500 Hz from 31 thin Ag/AgCl electrodes placed according to the 10–20 convention, with a reference at the participant's tip of the nose. Impedances were kept below 3 kΩ during recording. EEG trials were time-locked to the onset of words pronounced with a gesture and words without gesture the same way in the AV and the A modalities of presentation.

2.5. *EEG data analysis*

To avoid edge effects in the time frequency domain analysis, we selected long epochs of 4000 msec centered at Word Onset Time (WOT). Please not that the routine uses zero-padding for the non-existing points of the signal at the edges of the time epoch, thus giving inaccurate results at the beginning and end of the period under analysis. Since this problem disappears in the central part of the epoch to analyze, we followed the widely used approach to select time epochs much longer than the window of interest for the time-frequency analysis, obtain the data for the whole time epoch and restrict the analysis to a window which is safely far from contaminated points of the calculation (Aguiar-Conraria, Nuno Azevedo, & Soares, 2008).

We selected words that had no overlap in the window of interest (i.e., another target word −2000 msec to +750 msec from WOT), resulting in 64 valid trials per condition.

2.5.1. *Phase-locking value (PLV)*

Trial to trial phase variability was examined using the PLV (Lachaux, Rodriguez, Martinerie, & Varela, 1999), an amplitude independent measure ranging from 0 (maximal phase variability) to 1 (perfect phase locking). Time-frequency transformation was calculated using complex Morlet wavelet (comprising 6 cycles per wavelet) within a frequency range of 4–20 Hz, in 1 Hz steps over EEG single-trial length. Whenever PLV contrast between conditions at WOT resulted statistically significant (paired t-test gesture versus no gesture conditions, $p < .05$), concurrent amplitude data was analyzed to check whether the PLV differences could be explained by amplitude changes (Fuentemilla, Marco-Pallares, & Grau, 2006; Shah et al., 2004). This procedure allowed selecting the frequency bands and electrodes of interest.

Our main analysis contrasted PLVs from gesture and no-gesture conditions just around WOT in the AV modality of presentation. This way, we could directly test the hypothesis that gesture-induced phase alignment was observable just before word appearance only in those conditions in which a gesture beat signaled the imminent occurrence of a word. We first compared for each electrode and frequency PLVs between conditions from −400 to 400 msec around WOT (repeated measures *t*-test). Next, we used a cluster-based non-parametric permutation test on the resulting *t*-test values (Maris & Oostenveld, 2007). For a given frequency band, significant electrodes were connected in sets on the basis of spatial adjacency (first neighbors). Finally, the nonparametric statistical test was performed by calculating a "Monte Carlo" unbiased *p*-value under the permutation distribution (10^5 samples) and comparing it with an alpha-level of .05.

Finally, we further tested whether the PLV differences were statistically robust over time. To avoid false-positives, we considered significant differences for consecutive segments of at least 18 time points (36 msec) with difference at $p < .05$. This was determined following Guthrie and Buchwald's approach (Guthrie & Buchwald, 1991), after simulation runs of 10^3 pseudo-random series of PLV differences for 20 participants with autocorrelation $r = .93$ of the same duration as the selected window of interest (−400 msec–400 msec with respect to the WOT).

2.5.2. *Phase-preference at WOT*

We hypothesized that if beat gestures impact processing the processing of the upcoming word, this may be reflected by a non-uniform distribution of phases (i.e., neuronal population phases distributed around a preferred value) at WOT. This may be considered as a consequence of a realignment of the ongoing phase state when beat gesture appears, indicating a better brain state to encode the word. We collected the angular data from all participants for each electrode and frequency of interest. Then, we calculated the circular mean of the lumped data and the phase preference across participants by means of a Rayleigh test (Fisher, 1993). Concentration of the angular distribution around the preferred value was assessed by the *kappa* parameter of Von Mises distribution (Aydore, Pantazis, & Leahy, 2013). Circular statistic analyses were performed using CircStat Matlab toolbox (Berens, 2009). Statistical significance (set at $p < .05$) of both Rayleight test and concentration parameter analyses were corrected by Montecarlo sampling on random permutations of both experimental conditions ($N = 10^5$ samples).

2.5.3. *Time course of amplitude*

We computed amplitude time courses for the electrodes and frequency bands of interest between conditions from −400 to 400 msec around WOT. We tested for differences via a stepwise series of repeated measures t-tests (step size of 2 msec) and no correction was applied in the absence of significant difference.

2.5.4. *Audio-only data analysis*

Finally, we applied the exact same whole analysis to the EEG data acquired during the A modality of presentation. First, these analyses allowed us to test that when the visual information is removed, the processing of words pronounced with or without gesture is not different (furthermore, these modality of presentation remains ecologically valid, as it is comparable as listening to the radio). Second, if there is no difference of word processing between conditions in A, the effects potentially observed in AV are only due to the additional visual information.

3. Results

3.1. *Behavioral results*

Participants achieved an overall correct response rate of 80.22 ± 9.82%, which suggests that they paid attention to the discourse. For completeness, we performed a 2-by-2 ANOVA with the factor modality (AV or A) and order of presentation (first or second block). The results showed no effect of modality, as correct response rate in AV (80.5 ± 17%) and A (80.5 ± 15.7%) were not different [$F(1, 19) = 0$; $p = 1$]. In contrast, there was an effect of order, as correct response rates were significantly higher after the second presentation (88.5 ± 11.5%) than the first one (75.5 ± 16.2%); [$F(1, 19) = 4.59$; $p = .046$]. Finally, there was also an interaction between the modality and order of presentation [$F(1, 1) = 12.66$; $p = .0022$]. Correct response rates were greater after the AV modality when it was presented in second, than after the AV modality when it was presented in first ($p = .01$). However, the only interesting comparison was the correct response rate after the first presentation, between AV and A modalities. However, behavioral results will not be discussed any further in the present study.

3.2. *PLV at word onset*

In the AV presentation, the EEG for words pronounced with a beat gesture presented a significantly higher PLV at word onset within the theta (5–6 Hz) frequency range compared to words without gesture (Fig. 2). In contrast, PLV in the alpha (8–10 Hz) frequency range at word onset was significantly lower in the gesture compared to the no gesture condition. The cluster-based permutation test revealed that the increase of theta PLV seen at WOT in the gesture condition was

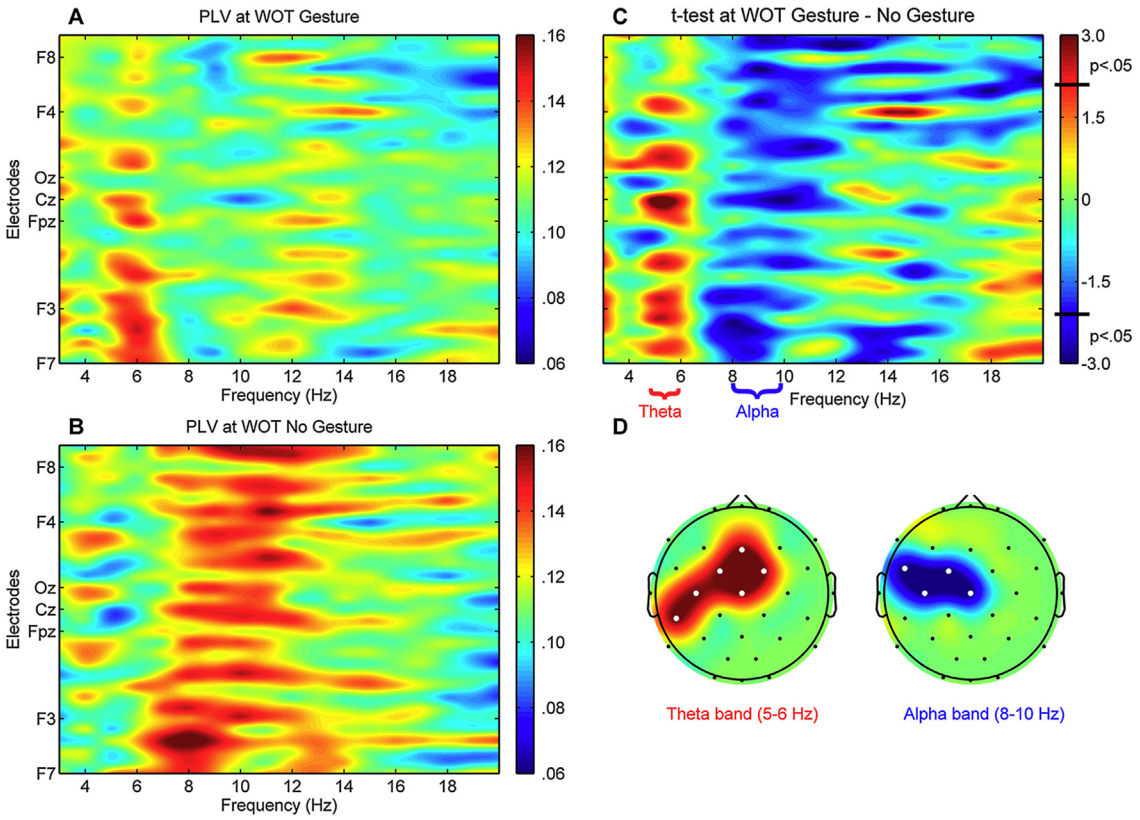

Fig. 2 – The left plots show the color coded PLV at WOT for each frequency and electrode in the (A) gesture and (B) no gesture conditions. (C) Color-coded representation of paired t-test values for the comparison between PLV at WOT in the gesture and no gesture conditions for each electrode and frequency. The frequency bands of interest (shown in the topographic plots D) are labeled in the x axis. (D) Topographic representation of the significant clusters (significant electrodes marked with white dots) for the t-tests shown in C within the theta band (left) and alpha (right) bands (color key as in C).

distributed over left fronto-temporal scalp regions ($p = .0001$, corrected), whereas the PLV difference in the alpha band was distributed over left centro-temporal sites of the scalp ($p = .0002$, corrected) (Fig. 2). Both theta and alpha effects were maximal at Cz [t-value: 3.93, $p = .0009$ (theta) and −3.01, $p = .007$ (alpha)].

We performed parallel analysis between gesture and no-gesture on the EEG recorded in an A condition. Any differences in this condition might reveal acoustically-based effects, unrelated to the sight of the gesture itself. In the A modality, the EEG for words pronounced with a beat gesture compared to words pronounced without gesture revealed no significant difference of PLV at word onset neither within the theta (5–6 Hz) nor within alpha (8–10 Hz) frequency ranges in the electrodes of interest (Fig. 5, Supplementary material).

To further investigate the gesture effect we performed a 2-by-2 ANOVA to test the interaction between the effect of condition (gesture or no gesture) and the effect of modality (AV or A) on the PLV in both theta and alpha bands. We considered the PLV at word onset (which was the initial time point of interest, as the relevant acoustic cue), at Cz electrode (where the observed effect of beat was maximum). Results show marginally significant trends for an interaction between condition and modality both in the theta [$F(1, 38) = 3.98$; $p = .053$] and the alpha bands [$F(1, 38) = 3.02$; $p = .09$], supporting the modulatory effect of gestures in the AV condition but not in the A condition.

3.3. *Time course of PLV and amplitude*

In the AV modality, the onset latency of the PLV difference between conditions at Cz was determined via a stepwise series of repeated measures t-tests (step size of 2 msec). Increments in the gesture condition spanned a window from −76 msec to 104 msec (90 samples) around the WOT in the theta band (Fig. 3). We also observed a significant alpha PLV modulation from −64 msec to 78 msec (71 samples), around the WOT (Fig. 3).

Amplitude analysis in the theta band showed no concurrent difference between gesture conditions in the temporal window of interest for the AV modality of presentation. Outside this window, amplitude differences between gesture and no gesture conditions were significant from −418 to −260 msec and 292 to 338 msec with respect to WOT. In the alpha band, amplitude was not significantly different between gesture and no gesture conditions anywhere in the epoch (Fig. 4). The amplitude analysis results in both theta and alpha bands suggest that PLV effects around WOT fulfill the criteria to be considered oscillatory-based (Fuentemilla et al., 2006; Shah et al., 2004). For the A alone modality of presentation, the

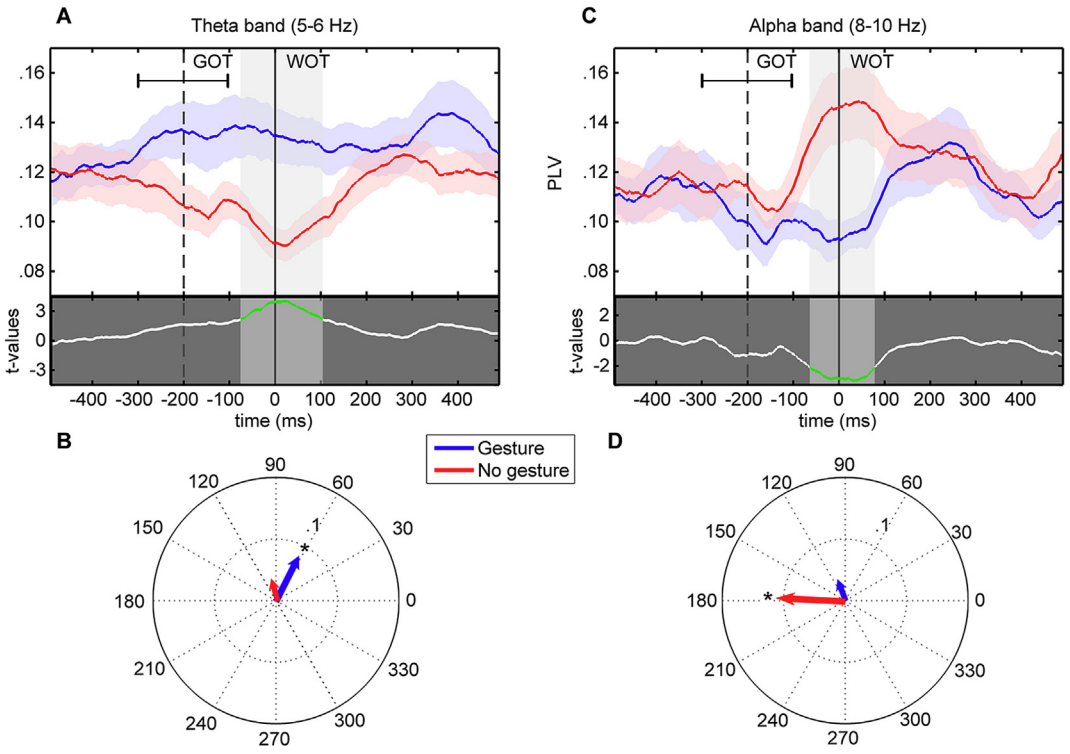

Fig. 3 – PLV time course in 5–6 Hz theta (A) and 8–10 Hz alpha (C) frequency bands at Cz electrode for the gesture (blue line) and no gesture (red line) conditions. For both frequency bands, the mean average ± standard deviation of gesture onset time (GOT) is represented respect to word onset time (WOT). The lower part of each plot displays the paired t-test values between gesture and no gesture conditions. The shaded bands indicate significant time intervals (highlighted in green in the t-test line). (B) and (D) display the mean phase at WOT at Cz expressed in polar coordinates. Arrow length indicates the strength of the phase preference (significant conditions are marked with *) in the gesture (blue arrow) and no gesture (red arrow) condition for theta (B) and alpha (D) frequency bands.

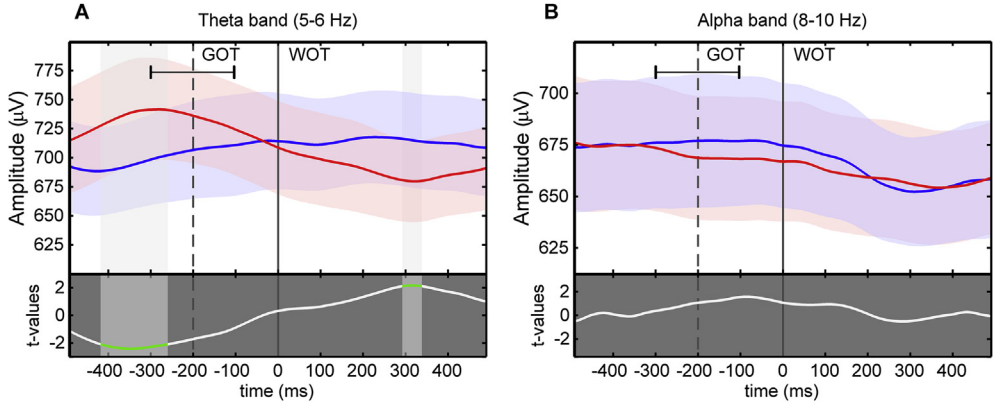

Fig. 4 – Amplitude time course in 5–6 Hz theta (A) and 8–10 Hz alpha (B) frequency bands at Cz electrode for the gesture (blue line) and no gesture (red line) conditions. For both frequency bands, the mean average ± standard deviation of gesture onset time (GOT) is represented respect to word onset time (WOT). The lower part of each plot displays the paired t-test values between gesture and no gesture conditions. The shaded bands indicate significant time intervals (highlighted in green in the t-test line).

time course of PLV and amplitude analyses in the theta and alpha bands revealed no significant differences between gesture and no gesture conditions (Figs. 6 and 7, Supplementary material).

3.4. *Phase-preference for theta and alpha band at WOT*

Rayleigh test in the AV modality of presentation indicated that the ongoing oscillatory phases at WOT within theta were

uniformly distributed in the no gesture condition ($p = .96$) but not in the gesture condition ($p = .03$, corrected) with a significant concentration ($\kappa = .16$, $p = .03$, corrected) around the mean value (1.1 rad) (Fig. 3). In contrast, for the alpha range, the phase distribution of oscillatory activity was uniformly distributed in the gesture condition ($p = .98$) but not in the no gesture condition ($p = .001$, corrected), with a concentration around the mean value (3.1 rad) statistically significant ($\kappa = .21$, $p = .001$, corrected).

In the audio-only modality, the ongoing oscillatory phases at WOT within theta were not uniformly distributed in the gesture and no gesture conditions (respectively $p = .0004$ and $p = .00002$) with kappa concentration parameters $k = .16$ and .18. For the alpha band, the oscillatory phases at WOT were uniformly distributed in both gesture and no gesture conditions (respectively $p = .3$ and $p = .2$) (Fig. 6, Supplementary material).

4. Discussion

During natural discourses speakers often accompany their speech with spontaneous beat gestures. Previous studies have described the role of beats in production as well as their potential benefits on perception (Holle et al., 2012; Krahmer & Swerts, 2007; McNeill, 1992; So, Chen-Hui, & Wei-Shan, 2012; Wang & Chu, 2013). The objective of the present study was to investigate the potential effect of natural beats on the listener's side by looking at the ongoing neural oscillatory dynamics relevant in speech perception. Based on previous studies addressing oscillatory dynamics in audio-visual speech contexts, we hypothesized that if spontaneous beats have an effect on the processing of speech, our results may demonstrate an increase of synchronization in the theta frequency activity at the corresponding word onset. If this modulation of theta activity is effectively due to the presence of a preceding reliable visual cue, the increase of synchronization may be observed before the word onset occurs, in order to facilitate its processing.

The findings reported in the present manuscript suggest that beat gestures effectively modulate brain oscillatory activity that can influence early processing of acoustic features. The PLV analysis brought a clear-cut pattern of strong phase synchronization in the theta 5–6 Hz range with a concomitant desynchronization in the alpha 8–10 Hz range, mainly picked up at left fronto-temporal electrodes. In contrast, when visual information was removed (Audio-only modality), our results showed no difference of PLV nor amplitude between words that had been pronounced with or without a beat gesture in the original discourse. This pattern suggests that the effects observed in the AV condition can be attributed to the sight of gestures, instead of just acoustic differences between gesture and no gesture words in the discourse.

The gesture-induced synchronization in theta tuned-in around 100 msec before the onset of the corresponding affiliate word, and continued for a few tenths of milliseconds after. Given that gestures initiated 200 ± 102 msec before word onsets, we believe that beats lead to an oscillation-based temporal prediction of speech in preparation for the incoming word onset. This interpretation is supported by the lack of amplitude differences within the same time windows of interest in both frequency bands.

Our findings bear on recent proposals that phase locking of ongoing cortical oscillations within theta serves to encode the speech syllabic rhythm and increases sensitivity to segmental cues (Peelle & Davis, 2012). Syllables constitute a fundamental unit in speech perception, with an approximate mean duration of around 200 msec corresponding to the period of a 4–8 Hz theta activity (Ghitza & Greenberg, 2009; Greenberg, 1999; Greenberg, Carvey, Hitchcock, & Chang, 2003; Peelle & Davis, 2012). Some studies have highlighted the role of theta activity during speech segmentation processes. Using MEG recordings, Luo and Poeppel (2007) demonstrated that the ongoing phase pattern of the theta activity in the auditory cortex fits with the spectro-temporal structure of the utterance during auditory speech perception. The relationship between the time-scales present in speech and the theta oscillatory activity may reflect how the brain extracts syllabic periodicity and encodes their timing through theta resynchronization at word/syllable onset (Giraud & Poeppel, 2012). This entrainment of the neural population leads to excitable states alternating predictably in the theta range, thereby improving the sensory processing when the relevant audio input comes at the right moment (Busch, Dubois & VanRullen., 2009; Engel, Fries, & Singer, 2001; Lakatos et al., 2008; Schroeder & Lakatos, 2009; Schroeder et al., 2008). In line with our current results, Arnal and Giraud (2012) argued that low-frequency phase synchronization around periodic acoustic features may be enhanced by a stable preceding visual cue. Arnal, Wyart, and Giraud (2011) focused on the potential of visible lip movements to allow the generation of predictions, reflected by low-frequency activity modulations that are compared with the incoming auditory signal. Arnal and Giraud (2012) defined prediction as the process that decreases the uncertainty about when periodic events are likely to occur in order to facilitate their processing. Here we argue that predictive visual information from the sight of the speaker's beat gestures is integrated with the spoken signal by means of theta synchronization at word onsets during natural speech perception.

As mentioned in the introduction, beats bear an intimate temporal coordination with the spoken signal. At the risk of increasing the noise of our stimuli, we used a real-life natural discourse instead of better controlled artificial recordings in order to preserve a natural temporal structure. In this respect, like it has been reported from the analysis of gestures in other studies, beats initiate shortly before the corresponding targeted word onset (around 200 msec before, Biau & Soto-Faraco, 2013), conferring a systematic order of arrival between visual and acoustic information. In these recordings, the quasi-periodic syllabic tempo of speech is preserved, allowing the oscillatory activity to match with it (Luo & Poeppel, 2007). Beats are entrained with respect to pseudo-periodic acoustic cues, conferring them also a non-random occurrence during speech (Peelle & Davis, 2012). In general, the use of natural materials in our study helps also maintain the integrity of language-related neural processes (Hasson, Malach, & Heeger, 2010) and the appropriate semantic context in which beats are produced and perceived (political addresses, McNeill, 1992).

Previous studies have advanced that beat gestures play the role of highlighters during speech perception, by emphasizing corresponding information (Krahmer & Swerts, 2007; McNeill, 1992). Biau and Soto-Faraco (2013) suggested that beats affect the early processes of speech by modifying the visuospatial context in which the following word occurs. Biau and Soto-Faraco observed that the ERP signal was affected by beats at early latencies and they hypothesized that this modulation of the context may drive the listener's focus of attention to relevant aspects of the spoken signal, at a pre-semantic level. Here, the results suggest that beats are taken into account during natural speech processing, and allow the anticipation of the corresponding word onset by locally tuning theta activity before word onset. This interpretation is corroborated by the absence of significant differences in the Audio-only modality of presentation, suggesting that the modulations of the oscillatory activities observed in the AV modality are not based on acoustic differences between different pronunciations of the same word, but due to the concurrent visual information (furthermore, the audio track in the AV and A was exactly the same). This is a relevant finding, since previous ERP studies (Biau & Soto-Faraco, 2013) could report the consequence of a gesture-driven modulation of auditory evoked responses from word onset.

From a more integrated perspective, seeing the speaker executing a beat may provide with local reliable visual linguistic information for speech parsing. For instance, Holle et al. (2012) showed that ERPs reflecting processing costs for syntactic parsing of ambiguous sentences were significantly reduced when beats emphasized the non-preferred, but correct sentence interpretation. Holle et al. suggested that listeners attribute visual syntactic value to beats when analyzing complex speech structures. Here, the relatively short timing between the beat onset and the consequent modulation of the theta activity puts the focus on a smaller time scale, and suggests that gestures may have a local attentional effect. In particular, beats may engage attention by driving the focus on the affiliate word, and facilitate the anticipation of the acoustic information (target word onset) following the gesture onset. The mechanisms of attention and prediction are related and may work in concert during perception (Summerfield & Egner, 2009).

Albeit speculative, one could think that this quick local attention effect of gestures proposed here might be key for the later-occurring syntactic effects reported in Holle et al. During speech perception, the theta oscillatory activity resets in the auditory cortex with a temporal structure similar to the spectro-temporal structure of the corresponding speech envelope, leading to a temporal alignment of neural excitability at word/syllable onsets (Giraud & Poeppel, 2012). This mechanism has been suggested to be involved in continuous speech parsing (Giraud & Poeppel, 2012; Luo & Poeppel, 2007). The presence of a beat may improve the neural excitability alignment at word onsets because the preparation phase allows anticipating relevant acoustic cues. One can speculate that beats, and their effect on temporal alignment of the neural signal, might have an impact in parsing mechanisms. However, source localization in further investigations may serve to confirm both the auditory cortex origin of the PLV modulation in the theta band and the behavioral expression of gestures on parsing (see Holle et al., 2012).

In parallel to theta synchronization, we found that gestures led to a desynchronization of the 8–10 Hz alpha activity within the same temporal window around word onset. Previous studies have shown that alpha desynchronization is observed at the onset of predicted stimuli (Rohenkohl & Nobre, 2011; Thut, Nietzel, Brandt, & Pascual-Leone, 2006). The desynchronization in the lower 8–10 Hz alpha band has been hypothesized to reflect attentional effects, independently from intelligibility during speech perception (Krause, Porn, Lang, & Laine, 1997). Krause et al. (1997) showed that listening to auditory speech backward compared to speech forward increased the alpha desynchronization only in the upper band (10–12 Hz) but did not affect it in the 8–10 Hz band. In the present study, the greater alpha modulation at the word onset is also found in the lower band, which is compatible with the possible attentional effect proposed here.

Altogether, an interesting aspect of our results is that the effect of gestures on speech is of a modulatory nature, which reflects on the phase but not on the strength of the signal (as it was shown in the power analysis, see Results section). In this respect the present pattern of results departs considerably from multisensory interactions based on a super or supra-additively profile, more classically found when comparing the strength of neural responses of bimodal stimulation with the sum of single modality presentations (e.g., Van Wassenhove, Grant, & Poeppel, 2005). Here, the responses we measured were based on the onset of the auditory words (accompanied with visual information both in the gesture and in the no gesture conditions), and the outcome was an alignment in phase for the gesture-words, rather than signal increase. This is why we interpret this pattern as a cross-modal modulation of one modality (visual) onto another (audition), without necessarily implying integration in the sense of signal summation (Lakatos, Chen, O'Connell, Mills, & Schroeder, 2007).

An alternative to the account above is that the modulation of oscillatory activity in the AV condition reported here reflects some visual-only contribution independent of the processing of speech. For instance, the alpha modulation has been reported to reflect body gesture perception (Streltsova, Berchio, Gallese, & Umilta', 2010). The EEG recorded from a silent version of the video clip, which could be the natural control condition for these purely visual effects, would involve differences in semantic, syntactic and prosodic content with the other conditions that would make interpretation of differences non-trivial. Even if the present data cannot completely rule out a pure effect of simple biological motion perception, some facts suggest an interpretation baed on the communicative value of gestures in speech perception. For instance, our data analyses were time locked to the onset of spoken words (not to the gestures), and thus any speech related effects should have been easier to pick up than visual-only modulations. Note for instance, that the PLV time profiles peak at word onset, showing a precise timing of the gesture effect on auditory speech. This temporal pattern coincides well with studies that investigated the time window of gesture-speech integration from an ERP perspective (Habets et al., 2011; Obermeier & Gunter, 2015;

Obermeier et al., 2011). Moreover, please note that the two relevant conditions in our study (words with gesture and words without gesture, in the AV modality) contained body motion visual information (the upper part of the speaker's body was visible in both conditions), with the only difference being the presence/absence of a hand beat gesture. Finally, the spectral specificity of our effects coincides with other studies addressing the perception of communicative gestures that have shown a decrease of the alpha power during the silent observation of an actor (Quandt, Marshall, Shipley, Beilock, & Goldin-Meadow, 2012). Interestingly, in Quandts' study alpha activity was sensitive to the type of gesture, with a greater effect with iconic gestures than simple pointing (i.e., deictic), suggesting not only an effect of action's perception.

Beats are likely to be perceived as communicative gestures rather than simple random movements. Through daily social experience, listeners learn to attribute linguistic relevance to beat gestures, also because they gesture when they speak. Consequently, listeners seem to have an understanding (if only implicit) of the sense of a beat at a precise moment. For example, Leonard and Cummins (2011) showed that when facial information is absent, listeners can detect an asynchrony from around 200 msec when the apex of a beat lags after the corresponding segment of speech. This suggests that listeners associate the apex with the pitch accent of the affiliate word. So et al., (2012) showed that adults remembered more words when they had been heard accompanied with a beat gesture than words pronounced alone. As this memory improvement was not found with children, the authors suggested that beat gestures serve meta-cognitive functions acquired during early lifetime, experiencing communicative interactions (McNeill, 1992). Thus, beat gestures seem to be implicit and silent visual information shared between speakers and listeners.

To conclude, our results suggest that gestures might be anchors modulating early stages of speech processing, potentially through attentional mechanisms. The present results reveal, for the first time, a principled mechanism that sustains the possible integration of visual information of beat gestures with auditory speech perception, reflected by low-frequency oscillatory activity. Furthermore, this mechanism suggests the generality of low-frequency oscillations as the basis for integration of very different kinds of audio-visual communicative information.

Acknowledgements

This research was supported by the Spanish Ministry of Science and Innovation (PSI2013-42626-P), AGAUR *Generalitat de Catalunya* (2014SGR856) and, the European Research Council (StG-2010 263145).

Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.cortex.2014.11.018.

REFERENCES

Aguiar-Conraria, L., Nuno Azevedo, N., & Soares, M. A. (2008). Using wavelets to decompose the time–frequency effects of monetary policy. *Physica A, 387*, 2863–2878.

Arnal, L. H., & Giraud, A. L. (2012). Cortical oscillations and sensory predictions. *Trends in Cognitive Sciences, 16*(7), 390–398.

Arnal, L. H., Wyart, V., & Giraud, A. L. (2011). Transitions in neural oscillations reflect prediction errors generated in audiovisual speech. *Nature Neuroscience, 14*(6), 797–801.

Aydore, S., Pantazis, D., & Leahy, R. M. (2013). A note on the phase locking value and its properties. *NeuroImage, 74*, 231–244.

Berens, P. (2009). CircStat: a MATLAB Toolbox for circular statistics. *Journal of Statistical Software, 31*(10).

Biau, E., & Soto-Faraco, S. (2013). Beat gestures modulate auditory integration in speech perception. *Brain and Language, 124*(2), 143–152.

Busch, N. A., Dubois, J., & VanRullen, R. (2009). The phase of ongoing EEG oscillations predicts visual perception. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience, 29*(24), 7869–7876.

Casasanto, D., & Jasmin, K. (2010). Good and bad in the hands of politicians: spontaneous gestures during positive and negative speech. *PloS One, 5*(7), e11805.

Engel, A. K., Fries, P., & Singer, W. (2001). Dynamic predictions: oscillations and synchrony in top-down processing. *Nature Reviews. Neuroscience, 2*(10), 704–716.

Fisher, N. I. (1993). *Statistical analysis of circular data*. Cambridge University Press.

Fuentemilla, L., Marco-Pallares, J., & Grau, C. (2006). Modulation of spectral power and of phase resetting of EEG contributes differentially to the generation of auditory event-related potentials. *NeuroImage, 30*(3), 909–916.

Ghitza, O., & Greenberg, S. (2009). On the possible role of brain rhythms in speech perception: intelligibility of time-compressed speech with periodic and aperiodic insertions of silence. *Phonetica, 66*(1–2), 113–126.

Giraud, A. L., & Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nature Neuroscience, 15*(4), 511–517.

Greenberg, S. (1999). Speaking in shorthand, A syllable-centric perspective for understanding pronunciation variation. *Speech Communication, 29*, 159–176.

Greenberg, S., Carvey, H., Hitchcock, L., & Chang, S. (2003). Temporal properties of spontaneous speech—a syllable-centric perspective. *Journal of Phonetics, 31*, 465–485.

Guthrie, D., & Buchwald, J. S. (1991). Significance testing of difference potentials. *Psychophysiology, 28*(2), 240–244.

Habets, B., Kita, S., Shao, Z., Ozyurek, A., & Hagoort, P. (2011). The role of synchrony and ambiguity in speech-gesture integration during comprehension. *Journal of Cognitive Neuroscience, 23*(8), 1845–1854.

Hasson, U., Malach, R., & Heeger, D. J. (2010). Reliability of cortical activity during natural stimulation. *Trends in Cognitive Sciences, 14*(1), 40–48.

Holle, H., Obermeier, C., Schmidt-Kassow, M., Friederici, A. D., Ward, J., & Gunter, T. C. (2012). Gesture facilitates the syntactic analysis of speech. *Frontiers in Psychology, 3*, 74.

Krahmer, E., & Swerts, M. (2007). The effects of visual beats on prosodic prominence: acoustic analyses, auditory perception and visual perception. *Journal of Memory and Language, 57*(3), 396–414.

Krause, C. M., Porn, B., Lang, A. H., & Laine, M. (1997). Relative alpha desynchronization and synchronization during speech perception. *Brain Research. Cognitive Brain Research, 5*(4), 295–299.

Lachaux, J. P., Rodriguez, E., Martinerie, J., & Varela, F. J. (1999). Measuring phase synchrony in brain signals. *Human Brain Mapping, 8*(4), 194–208.

Lakatos, P., Chen, C.-M., O'Connell, M. N., Mills, A., & Schroeder, C. E. (2007). Neuronal oscillations and multisensory interaction in primary auditory cortex. *Neuron, 53*(2), 279–292.

Lakatos, P., Karmos, G., Mehta, A. D., Ulbert, I., & Schroeder, C. E. (2008). Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science (New York, N.Y.), 320*(5872), 110–113.

Leonard, T., & Cummins, F. (2011). The temporal relation between beat gestures and speech. *Language and Cognitive Processes, 26*, 10.

Luo, H., & Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron, 54*(6), 1001–1010.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods, 164*(1), 177–190.

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

Mehler, J., Dommergues, J. Y., Frauenfelder, U., & Seguí, J. (1981). The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior, 20*, 298–305.

Muthukumaraswamy, S. D., & Johnson, B. W. (2004). Primary motor cortex activation during action observation revealed by wavelet analysis of the EEG. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology, 115*(8), 1760–1766.

Obermeier, C., & Gunter, T. C. (2015). Multisensory integration: the case of a time window of gesture-speech integration. *Journal of Cognitive Neuroscience, 27*(2), 292–307.

Obermeier, C., Holle, H., & Gunter, T. C. (2011). What iconic gesture fragments reveal about gesture-speech integration: when synchrony is lost, memory can help. *Journal of Cognitive Neuroscience, 23*(7), 1648–1663.

Peelle, J. E., & Davis, M. H. (2012). Neural oscillations carry speech rhythm through to comprehension. *Frontiers in Psychology, 3*, 320.

Pfurtscheller, G., Neuper, C., & Krausz, G. (2000). Functional dissociation of lower and upper frequency mu rhythms in relation to voluntary limb movement. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology, 111*(10), 1873–1879.

Quandt, L. C., Marshall, P. J., Shipley, T. F., Beilock, S. L., & Goldin-Meadow, S. (2012). Sensitivity of alpha and beta oscillations to sensorimotor characteristics of action: an EEG study of action production and gesture observation. *Neuropsychologia, 50*(12), 2745–2751.

Rohenkohl, G., & Nobre, A. C. (2011). Alpha oscillations related to anticipatory attention follow temporal expectations. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 31*(40), 14076–14084.

Schroeder, C. E., & Lakatos, P. (2009). Low-frequency neuronal oscillations as instruments of sensory selection. *Trends in Neurosciences, 32*(1), 9–18.

Schroeder, C. E., Lakatos, P., Kajikawa, Y., Partan, S., & Puce, A. (2008). Neuronal oscillations and visual amplification of speech. *Trends in Cognitive Sciences, 12*(3), 106–113.

Sebastián-Gallés, N., Martí, M. A., Carreiras, M., & Cuetos, F. (2000). *LEXESP: Léxico informatizado del Español*. Barcelona: Edicions Universitat de Barcelona.

Shah, A. S., Bressler, S. L., Knuth, K. H., Ding, M., Mehta, A. D., Ulbert, I., et al. (2004). Neural dynamics and the fundamental mechanisms of event-related brain potentials. *Cerebral Cortex (New York, N.Y.: 1991), 14*(5), 476–483.

So, W. C., Chen-Hui, C. S., & Wei-Shan, J. L. (2012). Mnemonic effect of iconic gesture and beat gesture in adults and children: is meaning in gesture important for memory recall? *Language and Cognitive Processes, 27*(5), 665–681.

Streltsova, A., Berchio, C., Gallese, V., & Umilta', M. A. (2010). Time course and specificity of sensory-motor alpha modulation during the observation of hand motor acts and gestures: a high density EEG study. *Experimental Brain Research, 205*(3), 363–373.

Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences, 13*(9), 403–409.

Thut, G., Nietzel, A., Brandt, S. A., & Pascual-Leone, A. (2006). Alpha-band electroencephalographic activity over occipital cortex indexes visuospatial attention bias and predicts visual target detection. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience, 26*(37), 9494–9502.

Treffner, P., Peter, M., & Kleidon, M. (2008). Gestures and phases: the dynamics of speech-hand communication. *Ecological Psychology, 20*(1), 32–64.

Van Wassenhove, V., Grant, K. W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences of the United States of America, 102*(4), 1181–1186.

Wang, L., & Chu, M. (2013). The role of beat gesture and pitch accent in semantic processing: an ERP study. *Neuropsychologia, 51*(13), 2847–2855.