Hand gestures as visual prosody: BOLD responses to audio-visual alignment are modulated by the communicative nature of the stimuli

Emmanuel Biau [a]
Luis Moris Fernandez [a]
Henning Holle [c]
César Avila [d]
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 Psychology, University of Hull, UK.

[d] Department of Psychology, Universitat Jaume I, Castelló de la Plana, Spain.

Corresponding author: Emmanuel Biau

Dept. de Tecnologies de la Informació i les Comunicacions
Universitat Pompeu Fabra
Roc Boronat, 138
08018 Barcelona
Spain

+34 691 752 040

emmanuel.biau@free.fr

ABSTRACT

During public addresses, speakers accompany their discourse with spontaneous hand gestures (beats) that are tightly synchronized with the prosodic contour of the discourse. It has been proposed that speech and beat gestures originate from a common underlying linguistic process whereby both speech prosody and beats serve to emphasize relevant information. We hypothesized that breaking the consistency between beats and prosody by temporal desynchronization, would modulate activity of brain areas sensitive to speech-gesture integration. To this aim, we measured BOLD responses as participants watched a natural discourse where the speaker used beat gestures. In order to identify brain areas specifically involved in processing hand gestures with communicative intention, beat synchrony was evaluated against arbitrary visual cues bearing equivalent rhythmic and spatial properties as the gestures. Our results revealed that left MTG and IFG were specifically sensitive to speech synchronized with beats, compared to the arbitrary vision-speech pairing. Our results suggest that listeners confer beats a function of visual prosody, complementary to the prosodic structure of speech. We conclude that the emphasizing function of beat gestures in speech perception is instantiated through a specialized brain network sensitive to the communicative intent conveyed by a speaker with his/her hands.

Speech perception; Gestures; Audiovisual speech; Multisensory Integration; MTG; fMRI.

1. INTRODUCTION

In everyday life, most communicative interactions between humans involve auditory and visual information. Indeed, in addition to auditory speech, listeners often have visual access to the speaker's lips, head, body posture and hand gestures. Here we concentrate on the communicative impact of the cospeech gestures that speakers produce with their hand movements while talking to someone (McNeill, 1992). By combining behavioral and physiological measures like event-related potentials (ERPs), prior studies have demonstrated that, for example, gestures describing an object or an action (i.e. iconic gestures) alter semantic processing of the spoken message (Kelly et al., 2004; Kelly et al., 2009; Wu & Coulson, 2010) or help disambiguate semantically complex sentences (Holle et al., 2007). These studies suggest that gestures provide information not present in the verbal modality alone, and support the idea that both streams of information are in fact components of a common integrated language system (McNeill, 1992; Kelly, Creigh & Bartolotti, 2009).

Many fMRI studies have investigated the degree to which gestures and speech recruit common brain areas. For example, a recent study by Dick et al. (2014) established the implication of a fronto-temporal network of language-related areas when iconic gestures provide complementary information to speech. The Superior Temporal Sulcus (STS) and the Middle and Superior Temporal Gyri (MTG/STG), which are well known to respond to audiovisual (AV) speech (Nath and Beauchamp, 2012; Calvert et al., 2000; Callan et al., 2004; Macaluso et al., 2004; Meyer et al., 2004; Campbell, 2008), have been found to be sensitive to the semantic relationship and congruency between gestures and the spoken message (Marstaller & Burianova, 2014). Greater BOLD responses in the STS, inferior parietal lobule and precentral sulcus were found for the perception of spoken sentences accompanied by semantically corresponding iconic gestures, as compared to meaningless movements or auditory-only versions (Holle et al., 2010; Holle et al., 2008). Willems et al, (2009) also found greater activations in the left STS/MTG when spoken sentences were presented with simultaneous pantomimes (i.e. speech-independent gestures) whose shape matched the verb of the utterance in meaning, as compared to incongruent ones. Additionally, the left Inferior Frontal Gyrus (IFG) has been often found to respond to the manipulation of the semantic relationship between gesture and speech (Marstaller & Burianova, 2014; Willems et al., 2009; Willems et al., 2007), suggesting a role in the integration of both streams of information to support sentence comprehension (Glaser et al., 2013; Uchiyama et al., 2008; Willems et al., 2007; Hagoort, 2005).

Although very relevant, these past studies have focused mostly on the neural correlates of hand gestures conveying semantic content, leaving aside other important functions of gestures, like their role as prosodic markers of speech (Guellaï, Langus & Nespor, 2014). Additionally, in these prior studies, participants were typically presented with single sentences where gesture-speech interactions happen in an impoverished context (i.e., short speech fragments containing an isolated gesture corresponding to a critical word). If one considers gestures and speech as two complementary sides of a common underlying language system, a natural continuous flow of visual (gestural) and audio (speech) streams might be essential for the system to remain fully functional (Hubbard et al., 2009; Biau & Soto-Faraco, 2013; Biau et al., 2015).

In the present study, we address the neural correlates of spontaneous beat gestures. As compared to the more commonly studied iconic gestures, beats are much less sophisticated in semantic content. Generally, beats are rapid biphasic flicks of the hand with no semantic content, serving to highlight relevant information and structure the narrative discourse (McNeill, 1992; So et al., 2012). These kinds of gestures are, by far, the most frequent class of co-speech gesture, and their use is very evident in public addresses, such as political discourses. Based on several evidences, it is now widely hypothesized that beat gestures may also play a role in prosodic processing (Guellaï, Langus & Nespor, 2014). First, beats seem to be very precisely aligned with speech envelope. The functional phase of beats - the moment of maximum extension of the movement, called the “apex” – is temporally aligned with the pitch accent of its affiliate spoken word, increasing its prominence by modulating the acoustic properties of the accentuated syllable (Yasinnik, Renwick & Shattuck-Hufnagel, 2004; Krahmer & Swerts, 2007; Treffner and al., 2008; Leonard & Cummins, 2010). Second, the speakers use the timing of gesture’s apexes to pack related information together, possibly playing a role in the syntactic organization of sentences supported by prosody (Holle et al., 2012; Guellaï, Langus & Nespor, 2014). The few studies that have investigated the neural correlates of beat gestures support the prosodic hypothesis too. For instance, Biau & Soto-Faraco (2013) found that beats modulate early ERPs time-locked to the affiliate words onset, within the latency window corresponding to phonological processing. Holle et al. (2012) also found that beats in complex sentences modulated the P600 ERP component, associated to syntactic analysis. Finally, in an fMRI study, observers watched a speaker producing beats while spontaneously speaking (Hubbard et al., 2009). The authors reported greater activations in the left STG/S in response to speech when it was accompanied by beats as compared to unrelated sign language gestures. They also reported greater BOLD responses in the bilateral posterior STG/S, including the Planum Temporale (PT) for speech accompanied by beats compared to a still body. Using beats from an actual fragment of continuous discourse ensured that gestures were produced in a legitimate context and frequency. In addition, spontaneous speech production ensured that the temporal relationship between the continuous beats stream and the rhythm of speech was maintained as in natural language conversation (Biau et al., 2015).

Scope of the present study

We hypothesize that beat gestures are produced as an integral part of the language system, providing the listener with visual prosodic information that is aligned with the prosodic contour of the speech message. For this reason, we advance that precise temporal alignment is essential to engage brain processes related to the integration of beats and speech. If this is true, brain activations in relevant integration areas may be sensitive to a breach in the temporal synchrony of beats with respect to their speech affiliates (Marstaller & Burianova, 2014; Hubbard et al., 2009). To test this hypothesis, we used fMRI while participants were presented with video clips in which the video was either synchronized with the audio track or lagged behind 800 milliseconds. With this manipulation, we assumed that when beat’s apexes fall out of synchrony with their affiliated speech accentuations, their highlighting function would falter. Yet, please note that desynchronization between beats and speech involves temporal misalignment at many levels, from mere spatio-temporal correlations of low level features to the misalignment in linguistic functions. Therefore, an integral question in this framework is whether the putative prosodic function of beats relates to a generic mechanism of visual emphasis or, alternatively, whether beats engage a specialized mechanism. Revealing such specialization is essential to attribute any beat-speech interaction effects to a common underlying language system. For instance, it is relevant that in the study by Holle et al. (2012), mentioned above, the authors did not find the same effects on the P600 ERP component when speaker’s moving hands (producing the beats) were replaced with discs following equivalent spatio-temporal trajectories in the visual display. The authors concluded that beats bear additional communicative intentions above and beyond simple visual emphasis (e.g. intentions and postures that come along with the prosodic variations, which might not be the case for an isolated disc).

Following Holle et al.’s logic, we wanted to single out brain areas that play a relevant and specific role in beat-speech integration by looking at the effect of beats-speech (de)synchronization, compared to the same effect when the speaker’s hands are replaced by arbitrary visual cues (i.e., moving discs). We hypothesized that the visual emphasis from arbitrary stimuli may differ from the linguistic function that gestures have when combined with speech (i.e. when beat emphasis is synchronized with the speech prosody). If beat gestures effectively confer a special communicative value to the spoken message, then one should expect disparate effects of audio-visual synchrony for beat gestures as compare to visual cues. We set up a 2x2 design with the factors AV synchrony (synchronous or asynchronous) and visual information (beats or discs) to test how the temporal alignment affects the integration of speech with either type of visual information. The interaction between synchrony and visual information is of essential interest because it allows isolating brain areas in which the impact of synchrony depends on which kind of visual information (beats or discs) accompanies audio speech prosody. Please note that a simple comparison between synchronous-asynchronous would conflate brain areas that are sensitive to generic, low level features as well as more specific linguistic related attributes of the stimuli. Thus, in this study we will mainly concentrate on brain areas where such an interaction arises. According to prior literature, these areas might (though not exclusively) correspond to the ones previously shown to be sensitive to gesture-speech integration, such as the left STS/G but also the left IFG (Holle et al., 2007; Willems et al., 2007; Hubbard et al., 2009; Holle et al., 2010; Marstaller & Burianova, 2014).

2. MATERIAL AND METHODS

2.1 Participants

Nineteen native speakers of Spanish (12 female, age range 19-29) took part in the current study. All participants were right-handed with normal auditory acuity as well as normal or corrected-to-normal vision. Participants gave informed consent prior to participation in the experiment and the study was approved by the University's ethics committee. Due to a technical problem, two participants could not listen to the speech stream during fMRI data acquisition and were therefore excluded from the statistical analysis. Thus, data from 17 participants (12 females, age range: 22.4 ± 2.4 years old) were included in the imaging analysis.

2.2 Material and stimuli

We extracted 44 video clips (18 s duration each) from a political discourse of the former Spanish President Luis Rodríguez Zapatero, recorded at the palace of La Moncloa and available on the official website (*Balance de la acción de Gobierno en 2010*, 12-30-2010; http://www.lamoncloa.gob.es). During the whole public address, the speaker stood behind a lectern, with the upper part of the body in full sight. The video clips were edited using Adobe Premiere Pro CS3. We visually inspected the entire discourse to select relevant segments of speech, containing only beats and cohesive gestures (series of beats that link successive points to a common concept) according to McNeill's definition. Clear iconic gestures were not found but as gesture categories sit along a continuum with fuzzy boundaries, some gestures may fall into multiple categories. Therefore one cannot be absolutely certain that our stimuli never included a minimum of semantic content in the hand shape. However, hand movements always conformed to McNeill's definition of beat gestures. To avoid abrupt onsets and offsets, we introduced 1 second audio-visual fade-in and –out at the beginning and end of each clip (respectively). In all the AV clips, the head of the speaker was masked with a superimposed ellipse-shaped patch in order to remove any facial information, such as lips or eyebrow movements, as well as head movements. After editing, videos were exported using the following parameters: video resolution 960x720, 25 fps compressor Indeo video 5.10, AVI format; audio sample rate 48 kHz 16 bits Mono. As explained below, we created four different versions for each video, corresponding to the four conditions of our experimental design: Beat Synchronous (Bs), Beat Asynchronous (Ba), Disc Synchronous (Ds) and Disc Asynchronous (Ds) (Fig. 1).

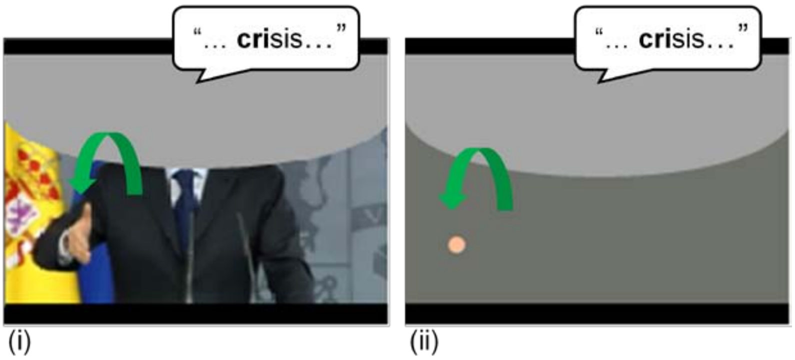

Figure 1. Screenshots from (i) Beat and (ii) Disc conditions. Audio and video streams were either synchronized (Bs and Ds conditions) or desynchronized (audio lagged video by 32 frames, corresponding to 800 ms) with respect to audio in the Ba and Da conditions). Green arrow illustrates the trajectory of a beat gesture and the corresponding disc. The apex of the movement coincided in this case with the Spanish word 'crisis'.

Beat conditions: We selected 44 segments (18s each, 450 frames) of the discourse in which the speaker naturally produced spontaneous beats (McNeill, 1992). For each clip, the speaker produced a minimum of 8 beats within the 18 s (mean number of gestures per clip: 12.8 ± 4.2). To create the Beat-Synchronous condition, audio and visual information remained synchronized as in the original discourse, with the speaker's hands fully visible (beat synchrony, Bs). For the beat asynchrony (Ba) condition, audio and visual information were desynchronized by inserting a lag of 800 ms (32 frames), leading to speech preceding beat gestures.

Disc conditions: To create the disc conditions, the video was removed and the hands were replaced by two discs that followed the hand trajectories of the original clips. We defined the junction between the index and the thumb as the reference point of both hands. We used *Skin Color Estimation Application* and *ELAN* software to detect pixel coordinates of hands frame-by-frame in each Beat video (http://tla.mpi.nl/tools/tla-tools/elan; Max Planck Institute for Psycholinguistics, The Language Archive, Nijmegen, The Netherlands; Wittenburg et al., 2006). Reference point coordinates were reviewed and corrected were necessary for both hands using custom-made scripts for Matlab (MATLAB Release 2012b, The MathWorks, Inc., Natick, Massachusetts, United States). The two discs representing the hands had a 40 pixel diameter size and were flesh-colored (Red, Green, Blue color values: 246, 187 and 146) at their corresponding reference point. The background color was set to the average value of a still frame of the speaker (Red Green Blue Value: 110, 114, and 104). We then created a synchronized (Disc Synchrony, Ds) and a desynchronized (Disc Asynchrony, Da) condition following the same process as in the beat condition.

Target videos: To ensure that stimuli were attended, participants performed an auditory detection task. For this, we used two clips from each experimental condition to create 8 targets. For each target video, the fundamental pitch of the original audio tracks was artificially shifted up three semitones (high pitch) for one syllable using Adobe's PitchShift filter while the intensity remained the same. In total, each participant was presented with 36 experimental and 8 target videos.

2.3 Procedure and Instructions

Participants were presented with 44 trials using E-Prime2 software. The order of trials was pseudo-randomized to avoid direct repetition of experimental conditions. Each trial consisted of a fixation cross with variable duration (from 7.5 to 8.5 seconds in steps of 0.25 seconds, uniformly distributed) followed by a video clip. The next trial began automatically after the end of the preceding video. A total of four experimental lists were created, counterbalanced for the four experimental conditions. Each participant saw one of the four lists.

Participants were instructed to perform an auditory detection task and press a button of the fMRI-compatible controller as soon as they detected an artificial pitch change in the voice of the speaker. The hand holding the controller (left or right hand) was counterbalanced across participants (even though target trials were not included in the statistical analysis). Participants were also instructed to always look at the screen during the whole experiment as if they were watching television. Before the fMRI acquisition, participants performed a rapid training with an extra target video presented in both Bs and Ds conditions as an example of artificial pitch change. After the scanning session, participants were given a questionnaire, asking 1) Did you perceive any asynchrony between video and speech during the experiment? 2) What could the moving discs represent? This questionnaire served to ensure that participants correctly attended to all videos. More importantly, it allowed us to evaluate if they could perceive the asynchrony between video and speech.

2.4 fMRI acquisition

Imaging was performed in a single session on a 1.5 T Siemens scanner. We first acquired a high-resolution T1-weigthed structural image (GR\IR TR=2200ms, TE=3.79ms, FA=15º, 256 x 256 x 160, 1mm isotropic voxel size). Functional data was acquired in a single run consisting of 610 Gradient Echo EPI functional volumes (TE = 50 ms, TR = 2000 ms) not specifically co-planar with the Anterior Commisure – Posterior Commisure line, acquired in an interleaved ascending order using a 64× 64 acquisition matrix with a FOV = 224. Voxel size was 3.5 x 3.5 x 3.5 mm with a 0.6 mm gap between slices, covering 94.3 mm in the Z axis.. The functional volumes were placed attempting to cover the whole brain in 23 axial slices. The first four volumes were discarded to allow for stabilization of longitudinal magnetization.

2.5 Imaging data analysing

FMRI data were analyzed using SPM12b (www.fil.ion.ucl.ac.uk/spm) and Matlab R2013b (MathWorks).

2.5.1. Preprocessing

Standard spatial preprocessing was performed for all participants using the following steps: Horizontal AC-PC reorientation; realignment and unwarp using the first functional volume as reference, a least squares cost function, a rigid body transformation (6 degrees of freedom) and a 2nd degree B-spline for interpolation, creating in the process the estimated translations and rotations occurred during the acquisition; slice timing correction using the middle slice as reference using SPM8's Fourier phase shift interpolation; coregistration of the structural image to the mean functional image using a normalized mutual information cost function and a rigid body transformation. The image was then normalized into the Montreal Neurological Institute (MNI) space (Voxel size was changed during normalization to isotropic 3.5 × 3.5 × 3.5 mm and interpolation was done using a 4th B-spline degree). Functional data was smoothed using an 8-mm full width half-maximum Gaussian kernel to increase signal to noise ratio and reduce inter subject localization variability. To add an extra quality control to the movement in participants, we used the Artifact Detection tools (ART) (http://www.nitrc.org/projects/artifact_detect/) with which the composite movement was calculated. This provides a single measure that comprises the movement due to rotation and translation between volumes. All volumes with a composite movement of more than 0.5 mm or more than 9 standard deviations away from the global mean signal of the session were considered as outliers (On average, 1.4% of the volumes per participant were detected as outliers). One regressor per outlier was added at the first level to discard any possible influence of these volumes in the final analysis.

2.5.2. fMRI analysis

The time series for each participant were high-pass filtered at 128 s and pre-whitened by means of an autoregressive model AR(1). At the first level (subject-specific) analysis, box-car regressors modelling the occurrence of the four conditions of interest (Bs, Ba, Ds and Da) and a fifth regressor for trials containing a target, all modelled as 18s blocks, were convolved with the standard SPM12b hemodynamic response function. Additionally, several regressors of no interest were included, including the six movement regressors provided by SPM during the realign process, the extra composite movement regressor calculated with ART and one regressor for each of the volumes considered as outliers. The resulting general linear model produced an image estimating the effect size of the response induced by each of the conditions of interest. The images from the first level were used for the planned critical contrasts in a second level analysis (inter-subject). At the second (inter-subject) level, these images were entered into a random effects factorial design with five levels, corresponding to the four critical conditions, plus an additional subject constant to account for non-condition-specific inter-subject variance. Correction for non-sphericity (Friston et al., 2002) was used to account for possible differences in error variance across conditions and any non-independent error terms for the repeated measures. Statistical images were assessed for cluster-wise significance using a cluster-defining threshold of p<0.001. The 0.05 Family-wise error correction critical cluster size was 31 voxels and was determined using random field theory (Data smoothing FWHM: 11.4mm, 11.2mm, 11.3 mm. Resel Count: 749.2), considering the whole brain as a volume of interest. Contrasts vectors assessing the two main effects and the interaction were used. Although the whole interaction statistical parametric map is presented, the discussion is limited to the clusters that showed an effect of Beat gestures compared to Discs (Bs+Ba > Ds+Da), as our main interest is focused on the parts of the brain that are involved in beat processing (for unmasked results and additional contrasts, please see supplementary online materials). To achieve this, we masked the interaction contrast, corrected as explained above, with the Beat > Discs contrast (p-threshold (unc.) <0.05). MNI coordinates were classified as belonging to a particular anatomical region using the SPM Anatomy Toolbox (Eickhoff et al., 2005).

3. RESULTS

3.1 Behavioral results

Participants correctly detected pitch deviation targets on 65.4% ± 31.7 % of the target trials and gave False Alarm (FA) responses only on 7.0% ± 13.6 % of the non-target trials.

3.2 Post-scanning questionnaire

When asked, after the scanning session, whether they perceived any asynchrony between video and speech during the experiment, 12 participants responded “yes”; 3 participants responded “yes, but not in the disc condition” and 2 participants responded “no”. With respect to the second question (“What could the moving discs represent?”), all participants responded “the hand of the speaker. This suggests that the asynchrony between beats and speech was noticeable, even though facial information was removed from videos. Furthermore, this consistent response confirmed that the spatiotemporal characteristics of disc movements successfully mimicked the hand trajectories in the Disc conditions. Both the behavioural and post-scanning questionnaire results suggest that participants were attentive to the AV stimuli.

3.3 fMRI results

3.3.1 Differential effect of AV synchrony depending on visual information

The first contrast of interest concerns the interaction between synchrony and visual information [(Bs-Ba) – (Ds-Da)]. This contrast is of particular interest as it highlights the brain areas where the impact of synchrony depends on which kind of visual information (beats or discs) accompanies speech. We studied this interaction in the areas that showed an effect of Beat > Disc (uncorrected mask p<0.05), as explained in the methods section (see Table 1). This restricts our analysis to areas that are related to beat processing. The results revealed a significant interaction in BOLD responses in two different clusters of the left Middle Temporal Gyrus and Superior Temporal Sulcus (MTG/STS), one more posterior and one more anterior (respectively, pMTG and aMTG/STS).

Additionally, significant interactions in left IFG and left occipital cortex (Brodmann area 18) were observed.

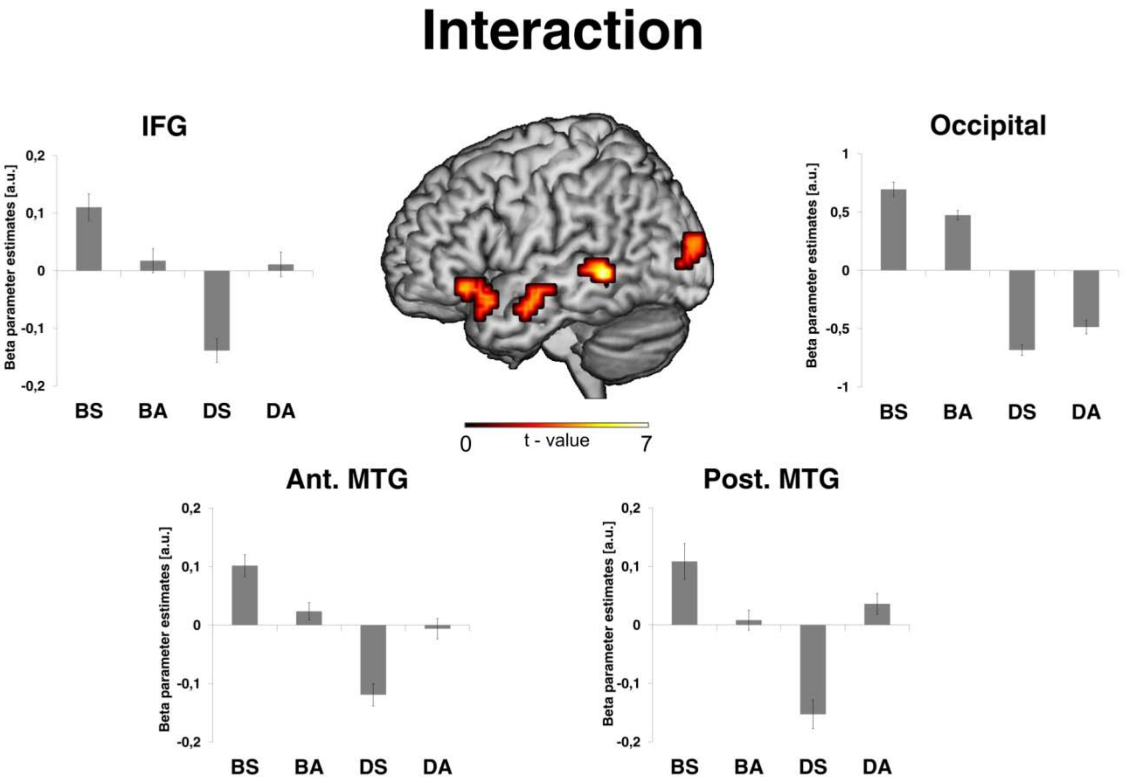

Figure 2. Interaction contrast [(Bs- Ba) – (Ds – Da) inclusively masked with the main effect of Beat (Bs+Da) compared to Disc (Ds+Da) using a p<0.05 cluster-corrected threshold with a minimum cluster size k = 31 and rendered on a 3D brain surface in MNI space (Left hemisphere). Error bars show 1 S.E.M of parameter estimates. IFG: Inferior frontal gyrus (-41 32 -11); Ant.MTG: anterior Middle temporal gyrus (-52 -7 -18); Post. MTG: posterior MTG (-59 - 46 -4); Occipital (-20 -95 14).

These results suggest that synchrony differentially affects speech integration, depending on the content of visual information. In particular, speech-gesture synchrony seems to recruit left-hemisphere brain areas preferentially, as compared to other visual cues which share the same spatio temporal properties but are arbitrary. Post-hoc analysis in the four significant clusters revealed that activations were significantly greater when beats and audio were synchronized (Bs) than asynchronous (Ba). Furthermore, the effect of synchrony on brain's activations was exactly the opposite when beats were replaced by simple discs (see Figure 2; see the significance of post-hoc simple main effects in the

Supplementary Material). It is worth noting that the areas which display this pattern (MTG, IFG and Occipital cortex in the left hemisphere) and the directionality of the numerical effects of beat synchrony are well in line with previous studies investigating gesture perception (Hubbard et al., 2009; Willems et al., 2009; Skipper et al., 2007; Holle et al., 2008, 2010), which further reassures the interpretation of these activations. Yet, despite this is the pattern expected from prior results and support our hypothesis, one should be careful from putting too much weight on it, given the post-hoc nature of the test.

3.3.2 Effect of type of visual information within temporal synchrony

Looking at the main effect of type of visual cue within the synchronous conditions can reveal differences arising from the type of visual stimulus. The contrast Beat Synchronous > Disc Synchronous revealed a greater BOLD response in various brain areas when speech was accompanied by synchronized beats (Bs), relative to synchronized discs (Ds) (see figure 3 and table 1). Not surprisingly, the greatest difference was observed in the occipital cortex likely due to a pure difference in visual information between conditions. The contrast also revealed differences in beyond visual brain areas, such as a significantly greater BOLD activity in the left MTG/STS, as well as in the left Inferior frontal Gyrus (left IFG) and left hippocampus. The contrast Ds>Bs revealed greater BOLD activity when speech was accompanied by synchronous discs rather than synchronous hand beats in the Superior Parietal areas bilaterally and right Angular Gyrus (see figure 3 and table 1).

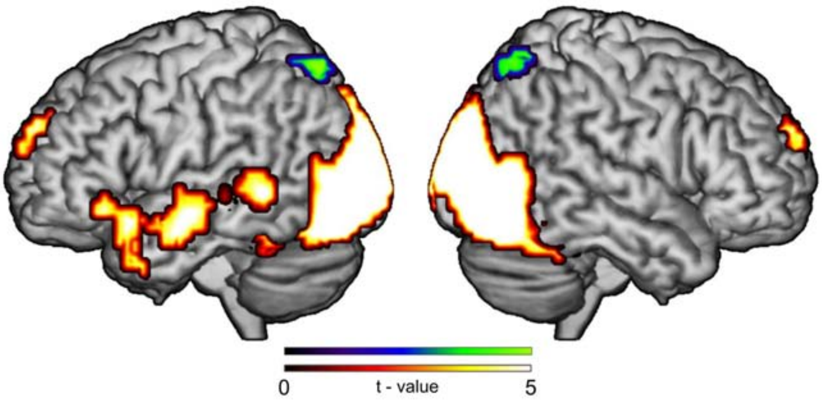

Figure 3. Main effect of Beat Synchronous (Bs) compared to Disc Synchronous (Ds). Statistical maps are thresholded at *P*-uncorrected <0.001 with a minimum cluster size k = 31 and rendered on a 3D brain surface in MNI space. From left to right: left hemisphere, right hemisphere and an axial cut at z=0. Hot colors indicate Bs > Ds. Cold colors indicate Ds> Bs.

3.3.3 Effect of synchrony between beat gestures and speech

The contrasts involving the comparisons Bs>Ba and Ba>Bs, restricted within the beat gesture conditions, revealed no main effect of synchrony, when performed at the whole brain level. Note that this particular result deviates from Hubbard et al. (2009), who reported an effect of synchrony in the left STS/G area. However, it must be mentioned that in Hubbard's study not only the actual synchrony, but also the nature of the gestures themselves was substantially changed between the synchronous and asynchronous condition (beats vs. ASL gestures in the control condition, respectively). In any case, our result implies that despite the BOLD responses for synchronous gestures tend to be larger than the BOLD responses for asynchronous gestures in the areas of significant interaction (as revealed in the interaction analysis). However, as discussed in the introduction, this effect cannot be fully interpreted without factoring in the responses of these areas to the disc synchrony/asynchrony conditions. This is because several low-level generic, as well as language-specific responses to synchrony are conflated in this contrast.

Hemisphere	Region	Corrected Cluster P-Value	Number of Voxels[a]	Z - Score	Coordinates (mm)[b]		
					x	y	z
Interaction [(Bs-Ba) – (Ds-Da)] masked with Beat > Disc (mask p-value <0.05)							
L	Middle Temporal Gyrus	0,043	32	5,93	-59	-46	-4
L	Inferior frontal gyrus	0,048	31	4,36	-41	32	-11
L	Temporal Pole			4,35	-45	14	-18
L	Middle Temporal Gyrus	0,048	31	4,20	-52	-7	-18
L	Middle Temporal Gyrus			4,10	-59	-11	-14
L	Middle Temporal Gyrus			4,09	-59	-4	-21
L	Middle Occipital	0,039	33	4,04	-20	-95	14
L	Inferior Occipital			3,38	-31	-88	4
Beat Synchronous > Disc Synchronous							

R	Lingual Gyrus	0,000	3080	Inf	8	-88	4
L	Cuneus			Inf	-10	-98	18
L	Calcarine			Inf	-3	-88	-4
L	Middle Temporal Gyrus	0,000	151	5,22	-62	-11	-14
L	Temporal Pole			4,75	-48	18	-14
L	Inferior Frontal Gyrus			4,33	-41	28	-11
L	Thalamus	0,006	52	5,20	-24	-28	0
L	Middle Temporal Gyrus	0,001	75	4,90	-55	-46	0
L	Middle Temporal Gyrus			3,93	-48	-32	0
Disc Synchronous > Beat Synchronous							
L	Superior Parietal	0,006	50	4,75	-16	-70	56
R	Superior Parietal	0,009	47	3,73	22	-66	59
	Angular Gyrus			3,49	22	-56	49
	Superior Parietal			3,40	15	-59	63
Beat Synchronous > Beat Asynchronous							
No significantly activate regions							
Beat Asynchronous > Beat Synchronous							
No significantly activate regions							

Table 1.[a] Number of voxels exceeding a voxel-height threshold of $p < 0.001$ using a $p < 0.05$ cluster-extend FWE correction. [b] First three maximum peaks more than 8 mm apart are reported for each cluster.

4. DISCUSSION

In the present study, we investigated the neural correlates of spontaneous beat gestures accompanying continuous, natural spoken discourses. Based on previous reports (McNeill, 1992; Yasinnik et al., 2004; Guellaï et al., 2014; Biau et al., 2015), we hypothesized that beats act as a visual counterpart of prosody. If this is the case, then breaking up the consistency between beat apexes and speech prosody may affect speech processing. In terms of neural expression, we hypothesized that if beats are integrated as linguistically relevant information, brain activity in relevant integration areas may be modulated by an asynchrony between visual and audio streams. As an integral aspect of this question, we addressed whether beats convey additional communicative aspects above and beyond arbitrary visual cues (discs) sharing the same spatiotemporal properties (Holle et al., 2012). Beats are thought to translate speaker intentions, extending body posture accompanying speaker's prosody to emphasize relevant segments of speech, which are available for listeners during speech perception (So et al., 2012; Casasanto & Jasmin, 2009). If this is the case, and beats play a linguistically relevant role above and beyond mere emphasis acting at low-level stages of stimulus processing, then the effect of synchrony for beats should be different as compared to visual discs, in the relevant brain areas. Indeed, this question was answered with the interaction term in our analysis, that indicates that the temporal synchrony of beats with speech prosody has a differential impact on BOLD responses, as compared to other kinds of visual information (here, discs that replaced the speaker's hands). The tendencies in the pattern of the interaction simple contrasts suggest greater activations when beats and speech were presented in synchrony as compared to asynchrony. Instead, the opposite pattern was observed when discs accompanied speech. Based on this significant interaction pattern, we interpret that, in addition to their emphasizing trajectory, beats also convey communicative aspects that simple discs are arguably lacking.

One surprising finding of our study is that the effect of synchrony for beats (i.e., greater activity for synchronous as compared to asynchronous beats in left IFG and MTG) was not simply absent for the moving discs, but actually tended to be reversed. When interpreting this cross-over interaction, it is also useful to take into account whether the neural response in these areas represents an activation or deactivation, relative to the implicit fixation cross baseline (see parameter estimates in Fig. 2). Relative to this fixation cross baseline, only speech accompanied by synchronous beats elicited activation in IFG, aMTG and pMTG. This is consistent with the idea that IFG and posterior temporal lobe are crucially involved in comprehending co-speech gestures (Holle et al., 2008, 2010, Willems et al., 2007, 2009). In contrast, a visual emphasis cue presented in asynchrony with speech (regardless of whether emphasis consisted of beats or moving discs) did not activate these areas, which may reflect that temporally incongruent AV stimuli are less likely to be integrated and may even cause suppression in multisensory areas (Noesselt et al., 2007). Interestingly, processing speech accompanied by temporally congruent discs elicited a reduction of activity in IFG, aMTG and pMTG, relative to fixation baseline. Such a deactivation could possibly reflect a phasic inhibitory influence onto IFG, aMTG and pMTG whenever speech is accompanied by temporally congruous but unfamiliar visual emphasis cues, such as moving discs. An influence of stimulus familiarity on AV integration in the temporal lobe has been demonstrated before (Hein et al., 2007) and may extend to unfamiliar speech-accompanying visual emphasis cues, such as moving discs.

Our results are in line with previous fMRI studies that investigated neural correlates of iconic gestures (Holle et al., 2010; Holle et al., 2008; Willems et al., 2009; Willems et al., 2007). Particularly, one previous fMRI addressed natural hand beats co-occurring with continuous speech (Hubbard et al., 2009) and reported a greater engagement of the STS compared to speech alone, an area comparable to the one found in the present study. The authors also reported greater BOLD activation in the left STS/G when speech was presented with the corresponding beat as compared to when presented with unrelated hand movements. Please note that this comparison does not allow one to infer whether the difference in left STS activation was produced by the lack of synchrony between control gestures and speech, the lack of communicative value of control gestures, or an unknown combination of the two. When Hubbard et al. compared speech accompanying beats to beats presented without speech, no difference was observed, suggesting that the modulations in the left STS/G reflect not only processing of biological movement but also integration of speech with the synchronized beat gestures. Indeed, the STS is sensitive to various types of cross-modal correspondence including AV speech (sound-lip correspondence) in various previous studies (Nath and Beauchamp, 2012; Calvert et al., 2000; Callan et al., 2004; Macaluso et al., 2004; Meyer et al., 2004).

In the present study, the interaction contrast suggests that BOLD response in the left MTG was greater when speech was accompanied by beats as compared to discs (regardless of whether they were synchronized or not with speech). At first glance, the greater response to stimuli containing beats in occipital areas compared to those with discs may reflect a pure bottom-up effect of richness of visual information (Figure 3). However, the interaction (Figure 2) revealed also that the significant difference of BOLD activity in the visual areas between beat and disc were dramatically reduced under asynchronous presentations. This suggests that mere physical differences between beats and discs conditions were not sufficient to explain their respective impact of synchrony in the indentified areas. The difference between beats and discs might bring about more profound consequences. For example, in a previous ERP study, Holle et al. (2012) showed that a beat modulated the P600 component reflecting syntactic parsing, whereas a disc following the equivalent trajectory did not. The authors suggested that the lack of communicative intention may explain the failure of simple discs to affect the neural correlates of syntactic parsing. Here, the significant simple contrast Bs>Ds supports this claim as it revealed greater activations not only in the occipital areas (although certainly due to differences of visual information, the results are only orientative), but also in the left MTG and left IFG areas. Indirectly, this result also converges toward the idea a differential response to synchrony for using discs that are not functionally associated with speech as part of a common language system.

According to the effect of interaction on the neural activations, it seems that the MTG responded to some additional language-related aspects associated with beat gestures during speech perception. Previous behavioral studies suggested that some implicit pragmatic and intentional information from the speaker could be extracted from beats, and influence speech encoding. For example, So et al, (2012) showed that adult observers managed to remember more words from a spoken list when the words had previously been accompanied by a beat gesture. As this memory improvement was not found in children, the authors concluded that beat gestures conveyed communicative information but the effect was functionally dependent on experiencing social interactions during development (McNeill, 1992). For example, listeners learn to interpret the speaker's intention to underline relevant information with a beat through social experience. This association of communicative aspects between beats and pitch accentuations was highlighted by Krahmer and Swerts (2007) who showed that listeners perceived words as more salient when accompanied with a beat gesture compared to same words presented in isolation. What is often missing in these studies is whether the value of gestures and their integration of speech simply depended on the general salience of the stimulus, or whether co-speech gestures engaged a more specialized system. Although the listeners in the present study could associate moving discs with movements of the hands and participants were able to detect an asynchrony between discs and speech, synchronized gestures and synchronized discs elicited qualitatively distinct patterns of brain activation (see contrast Bs>Ds). This suggests that during perception listeners distinguished visual information functional related to some aspect of speech (beats) from arbitrary visual cues (discs). Here, this information may require additional processes reflected by the differences of activations in the MTG between beats and discs conditions.

In addition to the above explanation, the possible linguistic aspects engaged when beats are present may be directly related to human movement understanding and body postures, over and above to their interaction with speech. The STS was found to respond to point-light representations of biological movements (Grossman et al., 2004; Pelphrey et al., 2004), actions executed by humans (Thioux et al., 2008) and social visual cues (for reviews, see Nummenmaa & Calder, 2009; Allison, Puce & McCarthy, 2000). Herrington et al, (2009) showed that the posterior STS was significantly more activated for trials in which participants perceived human point-light representations of actions compared to non-human movements. In the present study, the discs did not clearly represent a human form but clearly mimicked the trajectories described by hands during speech. In reference to the present study, listeners could have associated discs trajectories with hands (as they identified in the post-task questionnaire). Yet, whatever aspect of biological motion engaged by left MTG activations in the disc conditions, it was more strongly expressed during beat conditions. Please note, however, that this possible perceptual difference between beat gestures and discs in biological motion cannot explain the whole pattern of results we found in the left MTG, because the interaction term [(Bs – Ba) – (Ds – Da)] effectively controls for the different amounts of biological movement in the beat and disc conditions.

The present results also revealed an interaction between synchrony and visual information effects in the left IFG. Several fMRI studies have showed that the left IFG is sensitive to the semantic relationship between gesture and corresponding speech (Skipper et al., 2007; Willems et al., 2007; Willems et al., 2009; Dick et al., 2009) and may be engaged in the unification of visual (gestures) and audio (speech) complementary streams to facilitate comprehension (Willems et al., 2007; Hagoort, 2005). Recently, a meta-analysis investigating the neural correlates shared between different types of gestures reported a common engagement of the left IFG during the perception of speech accompanied with gestures as compared to a still body (Marstaller & Burianova, 2014). However, beat gestures do not convey semantic content, therefore the IFG responses observed in the present study cannot be explained in terms of semantic integration. Beyond meaning integration, the left IFG was also shown to be involved in the process of syntactic analysis during sentence comprehension (Glaser et al., 2013; Meyer et al., 2012; Obleser et al., 2011; Uchiyama et al., 2008). As beats play a role in syntactic parsing (Holle et al., 2012), our results might correspond to an engagement of this area in the integration of beat information toward the parsing of the spoken stream, as compared to moving discs. When beats were delayed (Ba condition), their apexes felt out from synchrony with pitch accents and likely out of the time window of gesture-speech integration, potentially affecting the AV speech processing load (Habets et al., 2011; Obermeier et al., 2011; Obermeier & Gunter, 2014).

It is worth noting that the simple main effect of synchrony for beat stimuli (contrast Bs vs Ba) in left MTG, IFG and occipital cortex did not reach significance in the whole brain analysis, but it is only revealed by the patterns of activations in the interaction contrasts following up on the interaction. Yet, the post-hoc results obtained for the simple main effects restricted to the interaction areas have to be often interpreted with caution (see Supplementary Materials). In consequence, the interpretation of synchrony effects for beat gestures must be linked to its effects relative to the disc condition. In other words, the disc synchrony manipulation can be seen as a baseline for the beat-synchrony manipulation. However, this is indeed a theoretically relevant type of comparison as discussed Holle et al. (2012). In addition, if we go by the results of previous studies, and extant knowledge the neural correlates of speech, we feel safe in interpreting this pattern in line with the results of the interaction that suggested a difference between synchronous and asynchronous beat conditions (see Figure 2). Note, for example that a similar effect of AV synchrony involving gestures in the left STG/S was reported in Hubbard et al. (2009). In their study, however, as mentioned earlier, Hubbard et al. used unrelated sign language movements as a control condition, which not only constitute a more dramatic asynchrony manipulation altogether (as speech and gestures had completely different rhythms), but also changed the very nature of the visual stimuli from the synchronous to the asynchronous condition. Here, we have looked at these two effects (confounded in Hubbard) separately, and therefore it is not surprising that their individual neural correlates are more subtle. That is, in the present study, although delayed with respect to speech, the rhythm of beats was maintained and might still be associable with the global speech envelope. This may have diminished the detrimental impact of desynchronized gestures on a listener's perception. This may also explain why we did not observe any effect of synchrony in the right auditory cortex related to auditory processing and prosody, as it was reported in Hubbard et al.'s results. A further relevant aspect in our study is that participants were asked to simply focus on an auditory detection task. This is interesting because our results cannot be attributed to an explicit monitoring of speech-gesture synchrony. On the contrary, our auditory detection task may have decreased attention on visual information and effectively weakened the expression of beat synchrony on speech processing networks.

Taken together, the present results provide new insights about the specificity of left MTG and IFG in the processing of multimodal language (for a review, see Campbell, 2008; Özürek, 2014). As participants were not explicitly asked to pay attention to the speaker's hands, this suggests that the temporal correspondence between beats and speech prosody may be picked up automatically. This is in line with previous proposals considering speech and gestures as two side of a same underlying language system (McNeill, 1992; Kelly, Creigh and Bartolotti, 2009). Beats appear to convey additional communicative value such as speakers' intentions, which are not available (or at least, not extracted) from simple visual stimuli (Holle et al., 2012; So et al., 2012; Casasanto & Jasmin, 2009; McNeill, 1992). The access to concurrent gestures during speech perception may engage the listeners and provide a better alignment between listener and speaker, improving speech processing and information encoding. Finally, the fact that the speaker was a well-known former Spanish president may have engaged some political sensitivity from listeners. However, such a possible bias is unlikely to influence our results, since participants viewed the same speaker across all four experimental conditions.

5. CONCLUSION

We investigated the neural correlates of spontaneous beat gestures produced in continuous speech. Our results revealed that the synchrony affected brain's activations differently according to the visual information accompanying speech during perception. We concluded that beats are linguistic information by their trajectories aligned with speech prosody, but also communicative intentions of the speaker.

AKNOWLEDGMENTS

This research was supported by the Ministerio de Economia y Competitividad (PSI2013-42626-P), AGAUR Generalitat de Catalunya (2014SGR856), and the European Research Council (StG-2010 263145).

REFERENCES

Allison, T., Puce, A., & McCarthy, G. (2000). Social perception from visual cues: role of the STS region. *Trends in Cognitive Sciences*, 4(7), 267–278.

Biau, E., & Soto-Faraco, S. (2013). Beat gestures modulate auditory integration in speech perception. *Brain and Language*, 124(2), 143–52.

Biau, E., Torralba , M., Fuentemilla, L., de Diego Balaguer, R., & Soto-Faraco, S. (2015). Speaker's hand gestures modulate speech perception through phase resetting of ongoing neural oscillations. *Cortex*, 68, 76-85.

Brett, M., Anton, J-L., Valabregue, R., & Poline, J-B. Region of interest analysis using an SPM toolbox [abstract] Presented at the 8th International Conference on Functional Mapping of the Human Brain, June 2-6, 2002, Sendai, Japan. Available on CD-ROM in NeuroImage, Vol 16, No 2.

Callan, D. E., Jones, J. A., Callan, A. M., & Akahane-Yamada, R. (2004). Phonetic perceptual identification by native- and second-language speakers differentially activates brain regions involved with acoustic phonetic processing and those involved with articulatory-auditory/orosensory internal models. *NeuroImage*, 22(3), 1182–94.

Calvert, G. A., Campbell, R., & Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Current Biology : CB*, 10(11), 649–57.

Campbell, R. (2008). The processing of audio-visual speech: empirical and neural bases. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 363(1493), 1001–10.

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.

Dick, A. S., Mok, E. H., Raja Beharelle, A., Goldin-Meadow, S., & Small, S. L. (2014). Frontal and temporal contributions to understanding the iconic co-speech gestures that accompany speech. *Human Brain Mapping*, *35*(3), 900–17.

Dick, A. S., Goldin-Meadow, S., Hasson, U., Skipper, J. I., & Small, S. L. (2009). Co-speech gestures influence neural activity in brain regions associated with processing semantic information. *Human Brain Mapping*, 30(11), 3509–26.

Friston, K. J., Glaser, D. E., Henson, R. N. A., Kiebel, S., Phillips, C., & Ashburner, J. (2002). Classical and Bayesian inference in neuroimaging: applications. *NeuroImage*, *16*(2), 484–512.

Glaser, Y. G., Martin, R. C., Van Dyke, J. A., Hamilton, A. C., & Tan, Y. (2013). Neural basis of semantic and syntactic interference in sentence comprehension. *Brain and Language*, *126*(3), 314–26.

Grossman, E. D., Blake, R., & Kim, C.-Y. (2004). Learning to see biological motion: brain activity parallels behavior. *Journal of Cognitive Neuroscience*, 16(9), 1669–79.

Guellaï, B., Langus, A., & Nespor, M. (2014). Prosody in the hands of the speaker. *Frontiers in Psychology*, *5*, 700.

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), 1845–54.

Hagoort, P. (2005). On Broca, brain, and binding: a new framework. *Trends in Cognitive Sciences*, 9(9), 416–23.

Hein, G., Doehrmann, O., Müller, N. G., Kaiser, J., Muckli, L., & Naumer, M. J. (2007). Object familiarity and semantic congruency modulate responses in cortical audiovisual integration areas. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *27*(30), 7881–7.

Herrington, J. D., Nymberg, C., & Schultz, R. T. (2011). Biological motion task performance predicts superior temporal sulcus activity. *Brain and Cognition*, 77(3), 372–81.

Holle, H., & Gunter, T. C. (2007). The role of iconic gestures in speech disambiguation: ERP evidence. *Journal of Cognitive Neuroscience*, 19(7), 1175–92.

Holle, H**.**, Gunter, T. C., Ruschemeyer, S. A., Hennenlotter, A., & Iacoboni, M. (2008). Neural correlates of the processing of co-speech gestures. *Neuroimage, 39*(4), 2010-2024.

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.

Holle, H., Obleser, J., Rueschemeyer, S.-A., & Gunter, T. C. (2010). Integration of iconic gestures and speech in left superior temporal areas boosts speech comprehension under adverse listening conditions. *NeuroImage*, 49(1), 875–84.

Hubbard, A. L., Wilson, S. M., Callan, D. E., & Dapretto, M. (2009). Giving speech a hand: gesture modulates activity in auditory cortex during speech perception. *Human Brain Mapping*, 30(3), 1028–37.

Kelly, S. D., Kravitz, C., & Hopkins, M. (2004). Neural correlates of bimodal speech and gesture comprehension. *Brain and Language*, 89(1), 253–60.

Kelly, S. D., Ozyürek, A., & Maris, E. (2010). Two sides of the same coin: speech and gesture mutually interact to enhance comprehension. *Psychological Science*, 21(2), 260–7.

Kelly, S. D., Ward, S., Creigh, P., & Bartolotti, J. (2007). An intentional stance modulates the integration of gesture and speech during comprehension. *Brain and Language*, 101(3), 222–33.

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.

Leonard, T., & Cummins, F. (2011). The temporal relation between beat gestures and speech. *Language and Cognitive Processes*, 26(10), 1457–1471.

Macaluso, E., George, N., Dolan, R., Spence, C., & Driver, J. (2004). Spatial and temporal factors during processing of audiovisual speech: a PET study. *NeuroImage*, 21(2), 725–32.

Marstaller, L., & Burianová, H. (2014). The multisensory perception of co-speech gestures – A review and meta-analysis of neuroimaging studies. *Journal of Neurolinguistics*, 30, 69–77.

Meyer, M., Steinhauer, K., Alter, K., Friederici, A. D., & von Cramon, D. Y. (2004). Brain activity varies with modulation of dynamic pitch variance in sentence melody. *Brain and Language*, 89(2), 277–89.

Noesselt, T., Rieger, J. W., Schoenfeld, M. A., Kanowski, M., Hinrichs, H., Heinze, H.-J., & Driver, J. (2007). Audiovisual temporal correspondence modulates human multisensory superior temporal sulcus plus primary sensory cortices. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *27*(42), 11431–41.

Nummenmaa, L., & Calder, A. J. (2009). Neural mechanisms of social attention. *Trends in Cognitive Sciences*, 13(3), 135–43.

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), 1648–63.

Obermeier, C., & Gunter, T. C. (2014). Multisensory Integration: The Case of a Time Window of Gesture-Speech Integration. *Journal of Cognitive Neuroscience*, 1–16.

Obleser, J., Meyer, L., & Friederici, A. D. (2011). Dynamic assignment of neural resources in auditory comprehension of complex sentences. *NeuroImage*, *56*(4), 2310–20.

Pelphrey, K. A., Morris, J. P., & McCarthy, G. (2004). Grasping the intentions of others: the perceived intentionality of an action influences activity in the superior temporal sulcus during social perception. *Journal of Cognitive Neuroscience*, 16(10), 1706–16.

Skipper, J. I., Goldin-Meadow, S., Nusbaum, H. C., & Small, S. L. (2007). Speech-associated gestures, Broca's area, and the human mirror system. *Brain and Language*, 101(3), 260–77.

So, W. C., Sim Chen-Hui, C., & Low Wei-Shan, J. (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.

Thioux, M., Gazzola, V., & Keysers, C. (2008). Action understanding: how, what and why. *Current Biology : CB*, 18(10), R431–4.

Treffner, P., Peter, M., & Kleidon, M. (2008). Gestures and Phases: The Dynamics of Speech-Hand Communication. *Ecological Psychology*, *20*(1), 32–64.

Uchiyama, Y., Toyoda, H., Honda, M., Yoshida, H., Kochiyama, T., Ebe, K., & Sadato, N. (2008). Functional segregation of the inferior frontal gyrus for syntactic processes: a functional magnetic-resonance imaging study. *Neuroscience Research*, *61*(3), 309–18.

Willems, R. M., Ozyürek, A., & Hagoort, P. (2007). When language meets action: the neural integration of gesture and speech. *Cerebral Cortex (New York, N.Y. : 1991)*, 17(10), 2322–33.

Willems, R. M., Ozyürek, A., & Hagoort, P. (2009). Differential roles for left inferior frontal and superior temporal cortex in multimodal integration of action and language. *NeuroImage*, 47(4), 1992–2004.

Wu, Y. C., & Coulson, S. (2010). Gestures modulate speech processing early in utterances. *Neuroreport*, 21(7), 522–6.