Spatial relations trigger visual binding of people

Parvaneh Adibpour*, Jean-Rémy Hochmann & Liuba Papeo

Institut des Sciences Cognitives—Marc Jeannerod, UMR5229, Centre National de la Recherche Scientifique (CNRS) & Claude Bernard University Lyon 1, Bron, France

Running title: EEG correlates of multiple-body representations
characters in the title: 50
words in the abstract: 262
words in the article (excluding the references) : 4551
figures and tables: 3 figures and 1 table
pages: 22

***Corresponding author:**

Parvaneh Adibpour
CNRS, Institut des Sciences Cognitives - Marc Jeannerod, 67 Boulevard Pinel, 69675, Bron, France
Phone: +33 437911266
E-mail: parvaneh.adibpour@isc.cnrs.fr

Abstract (262 words)

To navigate the social world, humans must represent social entities, and the relationships between those entities, starting with the spatial relationships. Recent research suggests that two bodies are processed with particularly high efficiency in visual perception, when they are in a spatial positioning that cues interaction, i.e. close and face-to-face. Socially relevant spatial relations such as facingness may facilitate visual perception by triggering grouping of bodies into a new integrated percept, which would make the stimuli more visible and easier to process. We used electroencephalography and a frequency-tagging paradigm to measure a neural correlate of grouping (or visual binding), while female and male participants saw images of two bodies face-to-face or back-to-back. The two bodies in a dyad flickered at the frequencies F1 and F2, respectively, and appeared together at a third frequency Fd (dyad frequency). This stimulation should elicit a periodic neural response for each single body at F1 and F2, and a third response at Fd, which would be larger for face-to-face (*vs.* back-to-back) bodies, if those stimuli yield additional integrative processing. Results showed that responses at F1 and F2 were higher for upright than for inverted bodies, demonstrating that our paradigm could capture body-specific activity. Crucially, the response to dyads at Fd was larger for face-to-face (*vs.* back-to-back) dyads, suggesting integration mediated by grouping. Thus, spatial relations that recur in social interaction (i.e., facingness) may promote binding of multiple bodies into a new representation. This mechanism can explain how the visual system contributes to integrating and transforming the representation of disconnected individual body-shapes into structured representations of social events.

Key words: body perception, social interaction, scene perception, frequency-tagging, SSVEP, visual binding

Introduction

The visual world is an array of objects, but our understanding of the world goes beyond single object representation, involving the processing of how individual entities relate to one another. The analysis of relations is essential to represent a scene or an event.

In social scenes (i.e., scenes involving multiple people), certain relational features, such as spatial proximity and positioning angle between bodies, are reliably correlated with the occurrence of social interaction (Zhou et al., 2019). These features are immediately available to our visual system, affecting the way in which individuals represent objects (Glanemann et al., 2016; Hafri et al., 2018; Papeo et al., 2017, 2019). For instance, recent research has shown that, presented at perceptual threshold, two bodies are recognized more easily when they are shown face-to-face than when they are back-to-back (Papeo et al., 2017). Face-to-face bodies are also attended to with higher priority, relative to other multiple-body configurations: in visual search through a crowd, face-to-face bodies recruit attention more strongly, and are searched for more efficiently, than back-to-back bodies (Papeo et al., 2019; Vestner et al., 2019).

Efficient processing of physically disconnected object-shapes in certain spatial configurations could be achieved through perceptual grouping, the binding of parts (here, bodies) into a new unitary representation (Coren & Girgus, 1980). Grouping based on spatial relations can reflect the gestalt laws of low-level vision (Koffka, 1935; Wagemans et al., 2012), but can also be driven by statistically frequent, semantically relevant (Kaiser et al., 2014, 2019) or, as suggested for face-to-face bodies, socially relevant spatial relations (Papeo, 2020). Through grouping, a new composite representation is formed when the spatial arrangement of objects matches an expected, regular, familiar, or meaningful configuration. The processing of multiple objects as a coherent, unitary structure increases efficiency, possibly by reducing the stimulus complexity and/or the competition for selection between objects (Kaiser et al., 2019; McMains & Kastner, 2010; Reddy et al., 2009)

At the neural level, recent functional magnetic resonance imaging (fMRI) studies have reported that visual areas underlying body perception in the lateral occipital cortex, participate in the representation of multiple-body stimuli, showing increased response and more accurate representation of dyads of face-to-face bodies (*vs*. non facing bodies or single bodies) (Abassi & Papeo, 2020; Bellot et al., 2020; Walbrin & Koldewyn, 2019). Likewise, the extensive research on

multiple-object perception has consistently demonstrated an increase in the neural response to pairs of objects seen in familiar (i.e., expected) configurations (e.g., Kaiser et al., 2014; Kim & Biederman, 2011; MacEvoy & Epstein, 2009; Roberts & Humphreys, 2010). Increased activity for regular multiple-object configurations has been specifically associated with markers of integrative processing that could mediate grouping of multiple parts or multiple objects in a composite representation that is more than the sum of the parts (e.g., Baeck et al., 2013; Baldassano et al., 2016; Kaiser & Peelen, 2018; Kubilius et al., 2015). In the above studies, integrative processing has been captured by an increase in the neural response to parts that formed a composite representation (e.g., a screen above a keyboard to form a computer) as compared with the same parts in a spatial relation that does not promote the formation of composite representation (e.g., a screen below a keyboard).

In the case of face-to-face dyads, grouping by virtue of socially relevant spatial relations has been suggested by the participants' efficient performance in attentional and perceptual tasks on the one hand, and increased neural activity in the visual body-selective cortex on the other hand. Here, we sought to test whether grouping could account for the behavioral and neural effects of body-dyad perception, by targeting a neural marker of integration in a frequency-tagging electroencephalography (EEG) study, allowing to separate the response to the whole and parts of a dyad. In frequency-tagging EEG paradigms, stimuli are presented periodically, at a regular frequency. The periodic stimulation entrains a periodic neural activity at the stimulation frequency, which is easily distinguishable in the frequency domain (Norcia et al., 2015; Regan, 1966). By presenting different parts of a multi-part stimulus at different stimulation frequencies, it is possible to dissociate the response to each single part from the response to the integration of those parts into a whole.

The above paradigm has proven effective in capturing neural effects of binding visual features based on Gestalt principles (Aissani et al., 2011; Alp et al., 2016), and parts in an object representation (e.g., a face; Boremanse et al., 2013). In particular, using frequency-tagging EEG, Boremanse and colleagues (2013) could distinguish the responses to the two halves of a face, each presented at a different regular frequency, from the response to the whole face, emerging when the brain integrates information from the two halves, at a third regular frequency. Critically, the response to whole-faces was larger when the halves gave rise to a face, relative to when their arrangement violated the canonical face configuration (i.e., the halves were next to each other but

misaligned; Boremanse et al., 2013). Thus, the response to whole-faces was not elicited by the mere occurrence of more facial features, but by the integration of those features in a well-formed face.

In the present study, we used a frequency-tagging paradigm to obtain objective responses to the parts and whole of a visual scene, where the constituting parts were two single bodies and the whole was the dyad. Each of the two bodies flickered at a different frequency to elicit a frequency-tagged response at the corresponding frequency (response to parts: F1 and F2). The two bodies appeared together in full view at a third frequency, that we called Fd (i.e., dyad-related frequency). To isolate body-specific activity, in half of the trials, bodies were presented inverted upside-down, a condition that disrupts body-specific neural activity, while providing matched visual stimulation (Stekelenburg & de Gelder, 2004). In different trials, the two bodies faced toward or away from the center. Therefore, when they appeared together at Fd, they either faced toward each other (facing dyad) or away from each other (non-facing dyad). We reasoned that, if the face-to-face positioning gives rise to a new integrated representation of the two bodies, the response at Fd should be enhanced for facing (vs. non-facing) dyads. By targeting body-specific response (i.e., the channels showing body-selective activity at F1 and F2), we tested whether effects of body positioning arise already in the neural correlates of body perception, as recent neuroimaging studies suggest (Abassi & Papeo, 2020; Bellot et al., 2020; Walbrin & Koldewyn, 2019). Enhanced response to facing dyads will provide initial evidence for multiple-body integration, triggered by specific –socially relevant– spatial relations. Grouping through integration may begin to account for the neural mechanism that transforms a group of bodies in the representation of a social event such as an interaction.

Materials and Methods

Participants

Twenty healthy adults (mean age 23.3, SD = 4.6; 13 female) with normal or corrected-to-normal vision participated in the study. The sample size was decided based on typical samples sizes in related EEG research (Alp et al., 2016; Boremanse et al., 2013; Kaiser et al., 2020) The study was approved by the local ethics committee. All participants gave written informed consent and received monetary compensation for participation.

Stimuli

Gray-scale images of an identical body in 6 different postures, seen in profile view, were created with Daz3D (Daz Productions, Salt Lake City) and the Image Processing Toolbox of MATLAB (MathWorks). Four facing dyads (Figure 1.a) were created by placing two body postures face-to-face. Four non-facing dyads were created by swapping the position of the two bodies in the facing dyads (Figure 1.b). The distance between the center of each body and the center of the screen was fixed (77 pixels); furthermore, the distance between two bodies (i.e., between their inner edges in the horizontal axis) was matched between a facing dyad and its corresponding non-facing dyad (mean = 38.2 ± 5.4 pixels). Thus, the two bodies were equally distant from the central fixation across facing and non-facing dyads.

For every (facing and non-facing) dyad, we created an inverted version by flipping the image by 180°. This yielded four variations of the same stimulus (i.e., upright facing, upright non-facing, inverted facing, and inverted non-facing), corresponding to the four experimental conditions of the study (Figure 1.b).

EEG Recordings

Brain activity was recorded using a 128-electrode EEG net (EGI, Eugene, USA), with a reference on the vertex. Recordings were continuously digitized at a sampling rate of 1000 Hz (net amp 400 system EGI, Eugene, USA). Participants were seated on a chair, one-meter away from a computer screen where the stimuli were presented. Single bodies in the screen had a mean visual angle of approximately 4.6° in height and 2.1° in width, and their inner edge were placed at 0.3° from the center. Dyads subtended a mean visual angle of 4.6° in height and 4.2° in width.

Paradigm

The experiment consisted of four blocks. Each block included four trials of 70.91 seconds (~1.18 min) each, in which the same dyad was presented in one of the four experimental conditions (upright facing, upright non-facing, inverted facing, inverted non-facing).

Each trial consisted of fast periodic visual stimulation, where the two bodies flickered at frequency F1 of 6.6667 Hz (~ 6.66 Hz) and frequency F2 of 5.4545 Hz (~ 5.45 Hz), respectively (Figure 1.c). Those frequencies were chosen so to prevent the overlap with the alpha band frequency (8–12 Hz) and, thus, to maximize the signal-to-noise ratio of the evoked responses

(Boremanse et al., 2013; Regan, 1989). That choice also took into account the refresh rate of the display (60Hz), to allow an integer number of frames for each cycle of stimulus presentation (60 Hz/6.6667 Hz ≈ 9 frames, and 60 Hz /5.4545 Hz ≈ 11 frames), preventing dropped frames or inaccuracies in stimuli frequency. The contrast in the stimuli was sinusoidal-modulated (figure 1.c). Stimuli were presented in cycles of 99 frames (9 x 11 = 99); in most frames, the two bodies were visible at variable contrasts, but appeared together in their full maximal contrast only once during a cycle (at frames 51-52, see Figure 1.c), corresponding to frequency Fd (dyad-related frequency) of 0.6061 Hz (60 Hz/99 ≈ 0.6061 Hz). Note that, two input frequencies, F1 and F2, can give rise to non-linear interactions at frequencies nF1±mF2 (with n and m being integers; Zemon & Ratliff, 1984). Therefore, besides the response at Fd, the response at frequencies nF1±mF2 can reflect the interaction between processing stimulus 1 at F1 and stimulus 2 and F2. Thus, in our analyses, we also targeted the interaction frequencies, which in our study occurred at F1-F2 = 1.2121 Hz, and corresponded to the second harmonic of the response at Fd (2 x 0.6061 Hz = 1.2121 Hz).

Whether the left or right body oscillated at F1 or F2 was kept constant for facing and non-facing dyads within a block, and was counterbalanced across blocks. Throughout a trial, a black cross was presented at the center of the screen roughly aligned with the shoulders of the bodies (whether upright or inverted). The cross color changed from black to red for 200 msec, eight times at random intervals during the trial duration (i.e., 1.18 min). Participants were instructed to press the spacebar on a keyboard with their right hand when they detected the color change. This task was meant to maintain the participants' vigilance and fixation on the center of the screen.

Trials were separated by a blank screen. After 20 seconds of blank screen, participants could start a new trial when they felt ready. The order of blocks and trials was randomized across participants. The experiment duration did not exceed 30 min. Stimulus presentation, communication of triggers with the recording EEG system, and response collection were controlled with Psychtoolbox (Brainard, 1997) through MATLAB.

Preprocessing

EEG recordings were band-pass filtered between 0.1 and 100 Hz using a zero-phase lag filter, and were further processed using the EEGLAB (Delorme & Makeig, 2004) and Brainstorm (Tadel et al., 2011) toolboxes in MATLAB. Recordings were segmented for the duration of the trials, based on the trigger marking the onset of each trial. In each trial, the first 3.3 seconds of the data

(corresponding to two cycles of stimulus presentation) were removed, which is assumed to be the time needed for the entrainment to become effective on the brain activity. The analysis considered the following 66 seconds of the data for each trial, corresponding to 360 cycles for the stimulation frequency of 6.66 Hz, and 440 cycles for the stimulation frequency of 5.45 Hz, and 40 cycles of the dyad frequency 0.606 Hz. The final 1.6 seconds (corresponding to one cycle of stimulus presentation) were not analyzed, to take into account the potential imprecision of boundary markers.

Frequency domain analysis

In the periodic activity evoked by the periodic presentation of the stimuli, responses to the two individual bodies were frequency-tagged at F1 and F2, while the response to the dyads was frequency-tagged at Fd. Time-series from each trial were transformed into the frequency domain using discrete Fourier transform, with a high-frequency resolution 0.0152 Hz (1/66 sec = 0.0152 Hz), which allows studying the response at the frequencies of interest with high precision. For each resulting frequency spectrum, the Signal-to-Noise Ratio (SNR) was then computed as the amplitude at each frequency divided by the mean amplitude of the 80 neighboring frequencies (40 frequency bins on each side), after excluding the 4 immediately adjacent bins (Boremanse et al., 2013; Leleu et al., 2020). The SNR were then averaged across trials of the same experimental condition, separately for each participant. Mean SNR values were analyzed to determine the response to individual bodies and dyads, at the corresponding tagged frequencies.

Statistical analyses and results

Grand-average identification of frequency-tagged responses

We first evaluated how many participants showed above-noise-level responses at the predefined frequencies, irrespective of experimental conditions and spatial distribution of the responses, to ascertain that our stimulation was effective in evoking frequency-tagged responses. To this aim, we averaged the SNR spectrum over all EEG sensors and trials for each participant (figure 2.a). Then, we compared the SNR at the frequencies F1 and F2 (6.66 Hz and 5.45 Hz), Fd (0.606 Hz) and their harmonics and nF1±mF2, against the noise level (80 neighboring frequency bins excluding the immediately adjacent 4 bins), with a one-tailed *t*-test. This analysis showed significant responses at the main frequencies of interest F1, F2 and Fd, as well as at multiple

harmonics of those frequencies (2 x F1/F2/Fd, 3 x F1/F2/Fd, 4 x F1/F2/Fd). Note that the even harmonics of Fd also correspond to harmonics of F1-F2 (e.g., 2Fd = F1-F2).

For both single-body and dyad-related responses, we constrained the analyses to the first two harmonics, since these responses were the most robust in 18 out of 20 participants and allowed keeping the same number of harmonics for both responses (statistical analysis of all harmonic responses is presented in Table 1). Thus, to study the response to single bodies, we averaged the SNR at 6.66 Hz and 2 x 6.66 Hz (body 1) and at 5.45 Hz and 2 x 5.45 Hz (body 2); to study the response to dyads, we averaged the SNR at 0.606 Hz and 2 x 0.606 Hz (1.21 Hz).

Responses at F1, F2 and their harmonics were predominantly distributed over posterior electrodes with medial distribution. The response to Fd and 2xFd was more broadly distributed over the scalp, encompassing posterior and anterior areas (Figure 2).

Definition of body-selective response

Previous fMRI studies have reported that the encoding of spatial relations between multiple bodies begins in visual areas for body perception (Abassi & Papeo, 2020; Walbrin & Koldewyn, 2019). To follow up on those studies, we isolated body-specific activity by contrasting the response to upright bodies and the response to inverted bodies at F1, F2, 2xF1 and 2xF2. Inversion keeps all the low-level visual information, while disrupting body configuration, which abolishes or significantly reduces body-specific response in the visual cortex, corresponding to posterior (occipital) areas of the brain (Brandman & Yovel, 2010; Stekelenburg & de Gelder, 2004; Taylor et al., 2010)

The comparison between SNR values for upright *vs.* inverted bodies was carried out for each sensor separately (*t*-test, one-tailed). Neighboring sensors with significantly higher amplitude for upright bodies ($t(19)>1.7$) were clustered and tested for significance using nonparametric cluster-mass test (5000 permutations) (Maris & Oostenveld, 2007). This analysis revealed a posterior cluster of electrodes with significantly higher response to upright than inverted bodies ($p = 0.03$; Figure 2.b).

Effect of positioning and orientation of body dyads

In the cluster of electrodes that showed the body-specific response (see Section 3.2), we studied whether positioning (facing or non-facing) and/or orientation (upright or inverted) of the

stimuli affected the response to dyads. To this end, the mean SNR values for Fd and 2Fd over the predefined cluster were entered into a repeated-measures analysis of variance (ANOVA), with factors Positioning (facing, non-facing) and Orientation (upright, inverted). This analysis revealed a main effect of Positioning, $F(1,19) = 5.5$, $p = 0.030$, with higher activity for facing than for non-facing dyads (Figure 2.c). There was no significant effect of Orientation, $F(1,19) = 2.2$, $p = 0.152$, and no interaction, $F(1,19) < 1$, *n.s.*

In a second analysis, we tested the effects of bodies positioning, orientation and their interaction on the SNR values, for each sensor across the scalp. Clusters of neighboring sensors showing main effects or interaction (F(19)>4.3) were tested for significance with nonparametric cluster-mass permutation tests. This analysis highlighted an anterior cluster, showing higher activation for facing than for non-facing dyads (cluster-mass permutation test: $p = 0.03$; Figure 2.c). We found no significant effect of orientation and no interaction (ps > 0.05).

In sum, our analyses highlighted activations distributed posteriorly and anteriorly, suggesting an integrated representation of the two bodies in a face-to-face (relative to back-to-back) configuration.

Effect of single body direction

In this analysis, we assessed whether the difference between facing and non-facing dyads, observed in the posterior and anterior clusters, was due to the processing of the dyad as a whole (i.e., to the processing of the relation between two bodies), or could be accounted for by a difference in the visual treatment of each single body in the facing *vs.* non-facing condition. If the latter is true, the identified clusters should show different responses not only to dyads (i.e., at dyad-related frequencies Fd and 2Fd), but also to single bodies in the facing vs. non-facing condition (i.e., at the single-body related frequencies F1 and F2).

To this end, we tested the effect of positioning and orientation, separately for each cluster, considering the mean SNR values measured at the single-body related frequencies. In the posterior cluster, this analysis confirmed the effect of orientation, $F(1,19) = 18.5$, $p < 0.001$ (see above Section 3.2), but no effect of positioning, $F(1,19) = 0.2$, $p = 0.643$, and no interaction, $F(1,19) = 0.2$, $p = 0.675$. No effect reached significance in the anterior cluster (orientation: $F(1,19) = 0.5$, $p = 0.498$; positioning, $F(1,19) = 0.7$, $p = 0.417$; interaction: $F(1,19) = 3.2$, $p = 0.089$). This analysis

showed that neither the body direction (inward vs. outward) nor the involvement in a facing or non-facing dyad did modulate the processing of single bodies.

Relationship between posterior and anterior responses to dyads

In a post-hoc analysis, we examined the relationship (i.e., covariation) between the responses to body dyads in the anterior and posterior clusters. Covariation was quantified in terms of correlation between the activity in the anterior and posterior cluster (i.e., mean SNR value across all sensors in a cluster) at the dyad-related frequencies Fd and 2Fd. Correlation between the responses in the two clusters was significant for facing dyads (Pearson's $r = 0.57$, $p = 0.008$), but not for non-facing dyads (Pearson's $r = 0.19$, $p = 0.420$) (Figure 3). As exploratory as it might be, the correlation suggests a functional coupling between the two clusters during perception of facing dyads.

Discussion

Using a frequency-tagging EEG paradigm, we recorded frequency-separable responses to body dyads (at Fd) and to each single body that formed the dyad (at F1 and F2). We first identified body-specific neural response at F1 and F2, using the classic contrast of upright *vs.* inverted bodies (Stekelenburg & de Gelder, 2004). This analysis revealed a posterior cluster, compatible with the occipital activity specifically associated with body perception (Downing, 2001; Peelen & Downing, 2007). Having identified body-specific neural activity, we could test the hypothesis, inspired by previous studies, that the neural correlates involved in body perception also encode for the relative positioning of multiple bodies in a visual scene (Abassi & Papeo, 2020; Bellot et al., 2020; Walbrin & Koldewyn, 2019). In line with those previous studies, we found an effect of positioning in the body-specific cluster, with larger response to facing than to non-facing dyads at Fd. Extended to the whole scalp (i.e., all the sensors), our analysis revealed another cluster showing larger response to facing than to non-facing dyads, with anterior distribution.

Unlike the responses evoked at the dyad-related frequencies, the responses evoked by single bodies did not vary depending on whether a body was part of a facing or a non-facing dyad. Thus, the effect of positioning at Fd captured the response to the overall dyadic stimuli, rather than a change in the response to a single body seen in a facing vs. a non-facing dyad. In other words, the effect reported here appears to reflect the encoding of relative positioning (i.e., the

positioning of a body relative to another), rather than absolute positioning (i.e., body orientation directed toward or away from the center).

The response to dyads was effectively tagged at two frequencies, possibly reflecting different processes: Fd (0.606 Hz), representing the periodic simultaneous presentation of two bodies together, and 1.21 Hz, corresponding to both the second harmonic of Fd (2 x 0.606 Hz = 1.21 Hz), and the nonlinear interaction between the two input frequencies (F1-F2 = 1.21 Hz; Zemon & Ratliff, 1984). From visual inspection of figure 2.a, one could appreciate that the response at 0.606 Hz is comparable in strength to the response at 1.21 Hz. Since harmonics typically have lower amplitudes than the main response, it is unlikely that the latter response only captured the harmonic of the first response. This raises the possibility that the response at 1.21 Hz resulted from both the second harmonic of the response to periodic visual stimulation (i.e. periodicity of dyad presentation), and the nonlinear long-range interaction of the responses to the two stimulation frequencies (F1 and F2) (Alp et al., 2016; Boremanse et al., 2013).

Frequency-tagged responses at dyad-related frequencies mark the neural representation of the two bodies together. Note that, over a trial, the two bodies were almost always present on the screen, though at a varying contrast level. That is, participants experienced the presence of two flickering bodies throughout the trial and were not aware that the bodies were only occasionally (i.e., at Fd) shown together in full view. This characteristics of the paradigm implies that the response to dyads at Fd captured a process that was spontaneous, if not automatic, given certain input characteristics, and independent of an explicit goal or task (Fodor, 1983); most likely a perceptual process.

The larger response to dyads for facing bodies suggests the recruitment of additional processing, beyond the processing of multiple single bodies, which were identical in facing and non-facing dyads. That response involved activations distributed over posterior and anterior areas. The recruitment of posterior areas might capture the activity of the body-selective visual cortex that, in single body perception, encodes the spatial relations between body parts to construct the representation of a body posture (Brandman & Yovel, 2010; Downing et al., 2001) and, in multiple body perception may encode the spatial relations between bodies, when face-to-face, to construct the representation of an interaction (see Papeo, 2020).

Continuing on the parallel between single body and multiple-body processing in posterior (occipital) areas, one would have expected a reduction of neural response when dyads were presented inverted (upside-down), relative to the response to upright dyads, similar to what we

observed for single bodies (at F1 and F2). Instead, in the present study, inversion of dyads did not affect the response at Fd (tagged at 0.6 Hz). However, the fact that responses to upright and inverted dyads had similar amplitudes at Fd in the frequency domain, does not rule out the possibility that the same responses had different shapes in the time course, which could not be distinguished in a frequency-tagging design. Future studies should address this question using event-related designs.

Contrary to posterior activation that was present for both facing and non-facing dyads, although magnified in the former condition, anterior activations evoked by facing dyads were virtually absent for non-facing dyads at Fd. One might be tempted to interpret the anterior activations as a selective response to facing dyads, reflecting the binding of two bodies when presented face-to-face. In this perspective, the current results, targeting a neural signature of integrative processes, might provide an account to the frontal activations reported in fMRI studies that have identified the neural network responsive to facing and interacting vs. nonfacing dyads (Abassi & Papeo, 2020; Bellot et al., 2020).

The current interpretation of the facing vs. non-facing dyad effect at Fd is supported by a corpus of studies, where frequency-tagging paradigms were used to mark the integrative processing that mediates grouping of multiple parts into a whole, such as geometrical elements into a new object representation (e.g., the Kanizsa triangle; Aissani et al., 2011; Alp et al., 2016; Gundlach & Müller, 2013), or face-halves in a face (Boremanse et al., 2013). All those studies demonstrated that spatial relations between parts are key to trigger the integration captured by the enhanced response to a canonical vs a noncanonical configuration. Thus, two face halves increased the response at the integration frequency when aligned but not when misaligned (Boremanse et al., 2013). By analogy, in the current study, two bodies increased the response at the alleged integration frequency, when the two bodies appeared simultaneously and faced toward –but not away from– each other. Adding to the extant literature, our results show that grouping can account not only for object formation through binding of disconnected, but spatially organized, parts (Aissani et al., 2011; Alp et al., 2016; Boremanse et al., 2013; Gundlach & Müller, 2013), but also for the formation of scenes (or events) by binding multiple disconnected objects (here, bodies) together (see also Alp et al., 2017; Radtke et al., 2020).

In summary, by contrasting the response to facing vs. non-facing dyads involving the same individual bodies, we were able to disentangle objective neural responses to parts (individual body) and to whole scene (dyad). The response to facing (*vs.* non-facing) dyads described here,

echoes neural effects that have been consistently associated with perceptual grouping (also called visual binding), and suggests that the face-to-face body positioning can trigger a new representation that is more than the sum of the constituent parts. The representational content of the composite representation of the two facing bodies remains a fascinating question for further research. In conclusion, by extending the frequency-tagging paradigm to a new class of stimuli (i.e., multiple-body stimuli), our results contribute to characterizing grouping as a general mechanism that could account not only for the representation of objects from parts, but also for the representation of scenes from objects or events from social entities.

References

Abassi, E., & Papeo, L. (2020). The Representation of Two-Body Shapes in the Human Visual Cortex. *The Journal of Neuroscience*, *40*(4), 852–863.

Aissani, C., Cottereau, B., Dumas, G., Paradis, A.-L., & Lorenceau, J. (2011). Magnetoencephalographic signatures of visual form and motion binding. *Brain Research*, *1408*, 27–40.

Alp, N., Kogo, N., Van Belle, G., Wagemans, J., & Rossion, B. (2016). Frequency tagging yields an objective neural signature of Gestalt formation. *Brain and Cognition*, *104*, 15–24.

Alp, N., Nikolaev, A. R., Wagemans, J., & Kogo, N. (2017). EEG frequency tagging dissociates between neural processing of motion synchrony and human quality of multiple point-light dancers. *Scientific Reports*, *7*(1), 44012.

Baeck, A., Wagemans, J., & Op de Beeck, H. P. (2013). The distributed representation of random and meaningful object pairs in human occipitotemporal cortex: The weighted average as a general rule. *NeuroImage*, *70*, 37–47.

Baldassano, C., Beck, D. M., & Fei-Fei, L. (2016). Human–Object Interactions Are More than the Sum of Their Parts. *Cerebral Cortex*, bhw077. https://doi.org/10.1093/cercor/bhw077

Bellot, E., Abassi, E., & Papeo, L. (2020). *Moving toward versus away from another: How body motion direction changes the representation of bodies and actions in the visual cortex* [Preprint]. Neuroscience.

Boremanse, A., Norcia, A. M., & Rossion, B. (2013). An objective signature for visual binding of face parts in the human brain. *Journal of Vision*, *13*(11), 6–6.

Brainard, D. H. (1997). The Psychophysics Toolbox. Spat Vis. *Spat Vis*, *10*, 433–436.

Brandman, T., & Yovel, G. (2010). The Body Inversion Effect Is Mediated by Face-Selective, Not Body-Selective, Mechanisms. *Journal of Neuroscience*, *30*(31), 10534–10540.

Coren, S., & Girgus, J. S. (1980). Principles of perceptual organization and spatial distortion: The gestalt illusions. *Journal of Experimental Psychology: Human Perception and Performance*, *6*(3), 404–412.

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21.

Downing, P. E., Jiang, Y., Shuman M., Kanwisher, N. (2001). A Cortical Area Selective for Visual Processing of the Human Body. *Science*, *293*(5539), 2470–2473.

Fodor, J. A. (1983). *The modularity of mind.* MIT press.

Glanemann, R., Zwitserlood, P., Bölte, J., & Dobel, C. (2016). Rapid apprehension of the coherence of action scenes. *Psychonomic Bulletin & Review*, *23*(5), 1566–1575.

Gundlach, C., & Müller, M. M. (2013). Perception of illusory contours forms intermodulation responses of steady state visual evoked potentials as a neural signature of spatial integration. *Biological Psychology*, *94*(1), 55–60.

Hafri, A., Trueswell, J. C., & Strickland, B. (2018). Encoding of event roles from visual scenes is rapid, spontaneous, and interacts with higher-level visual processing. *Cognition*, *175*, 36–52.

Kaiser, D., Häberle, G., & Cichy, R. M. (2020). Cortical sensitivity to natural scene structure. *Human Brain Mapping*, *41*(5), 1286–1295.

Kaiser, D., & Peelen, M. V. (2018). Transformation from independent to integrative coding of multi-object arrangements in human visual cortex. *NeuroImage*, *169*, 334–341.

Kaiser, D., Quek, G. L., Cichy, R. M., & Peelen, M. V. (2019). Object Vision in a Structured World. *Trends in Cognitive Sciences*, *23*(8), 672–685.

Kaiser, D., Stein, T., & Peelen, M. V. (2014). Object grouping based on real-world regularities facilitates perception by reducing competitive interactions in visual cortex. *Proceedings of the National Academy of Sciences*, *111*(30), 11217–11222.

Kim, J. G., & Biederman, I. (2011). Where Do Objects Become Scenes? *Cerebral Cortex*, *21*(8), 1738–1746.

Koffka, K. (1935). Principles of Gestalt Psychology Harcourt, Brace & World. *Inc. New*.

Kubilius, J., Baeck, A., Wagemans, J., & Op de Beeck, H. P. (2015). Brain-decoding fMRI reveals how wholes relate to the sum of parts. *Cortex*, *72*, 5–14.

Leleu, A., Rekow, D., Poncet, F., Schaal, B., Durand, K., Rossion, B., & Baudouin, J. (2020). Maternal odor shapes rapid face categorization in the infant brain. *Developmental Science*, *23*(2).

MacEvoy, S. P., & Epstein, R. A. (2009). Decoding the Representation of Multiple Simultaneous Objects in Human Occipitotemporal Cortex. *Current Biology*, *19*(11), 943–947.

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

McMains, S. A., & Kastner, S. (2010). Defining the Units of Competition: Influences of Perceptual Organization on Competitive Interactions in Human Visual Cortex. *Journal of Cognitive Neuroscience*, *22*(11), 2417–2426.

Norcia, A. M., Appelbaum, L. G., Ales, J. M., Cottereau, B. R., & Rossion, B. (2015). The steady-state visual evoked potential in vision research: A review. *Journal of Vision*, *15*(6), 4.

Papeo, L. (2020). Twos in human visual perception. *Cortex*.

Papeo, L., Goupil, N., & Soto-Faraco, S. (2019). Visual Search for People Among People. *Psychological Science*, *30*(10), 14.

Papeo, L., Stein, T., & Soto-Faraco, S. (2017). The Two-Body Inversion Effect. *Psychological Science*, *28*(3), 369–379.

Peelen, M. V., & Downing, P. E. (2007). The neural basis of visual body perception. *Nature Reviews Neuroscience*, *8*(8), 636–648.

Radtke, E. L., Schöne, B., Martens, U., & Gruber, T. (2020). Electrophysiological correlates of gist perception: A steady-state visually evoked potentials study. *Experimental Brain Research*, *238*(6), 1399–1410.

Reddy, L., Kanwisher, N. G., & VanRullen, R. (2009). Attention and biased competition in multi-voxel object representations. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(50), 6.

Regan, D. (1966). Some characteristics of average steady-state and transient responses evoked by modulated light. *Electroencephalography and Clinical Neurophysiology*, *20*(3), 238–248.

Regan, D. (1989). Methods for recording steady-state evoked potentials. Human brain electrophysiology. Evoked potentials and evoked magnetic fields in science and medicine. New York: Elsevier. *New York: Elsevier*.

Roberts, K. L., & Humphreys, G. W. (2010). Action relationships concatenate representations of separate objects in the ventral visual system. *NeuroImage*, *52*(4), 1541–1548.

Stekelenburg, J. J., & de Gelder, B. (2004). The neural correlates of perceiving human bodies: An ERP study on the body-inversion eɟect. *Neuroreport*, *15*(5), 777–780.

Tadel, F., Baillet, S., Mosher, J. C., Pantazis, D., & Leahy, R. M. (2011). Brainstorm: A User-friendly Application for MEG/EEG Analysis. *Intell. Neuroscience*, *2011*, 8:1–8:13.

Taylor, J. C., Roberts, M. V., Downing, P. E., & Thierry, G. (2010). Functional characterisation of the extrastriate body area based on the N1 ERP component. *Brain and Cognition*, *73*(3), 153–159.

Vestner, T., Tipper, S. P., Hartley, T., Over, H., & Rueschemeyer, S.-A. (2019). Bound together: Social binding leads to faster processing, spatial distortion, and enhanced memory of interacting partners. *Journal of Experimental Psychology: General*, *148*(7), 1251–1268.

Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., & von der Heydt, R. (2012). A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure–ground organization. *Psychological Bulletin*, *138*(6), 1172–1217.

Walbrin, J., & Koldewyn, K. (2019). Dyadic interaction processing in the posterior temporal cortex. *NeuroImage*, *198*, 296–302.

Zemon, V., & Ratliff, F. (1984). Intermodulation components of the visual evoked potential: Responses to lateral and superimposed stimuli. *Biological Cybernetics*, *50*(6), 401–408.

Zhou, C., Han, M., Liang, Q., Hu, Y.-F., & Kuai, S.-G. (2019). A social interaction field model accurately identifies static and dynamic social groupings. *Nature Human Behaviour*, *3*(8), 847–855.

Acknowledgments: This work was supported by ANR grant awarded to J-R. H. (Project: TACTIC, Grant Agreement ANR-16-CE28-0006) and European Research Council Starting Grant awarded to L.P. (Project: THEMPO, Grant Agreement 758473).

Table and Figures captions

Table 1. Significance of the responses at stimulation frequencies of interest and their multiple harmonics as assessed on the grand-average SNR of all EEG sensors.

Significant responses at the frequencies of interest correspond to SNR values higher that noise-level SNR. For responses to individual bodies, significance was tested for frequencies <50 Hz; responses to dyads, significance was tested for frequencies <5 Hz. P-values are Bonferroni corrected for the number of harmonics tested a frequency.

Figure 1. Illustration of stimuli, design and procedures for stimulus delivery. a) The four body dyads used in the study. b) Four variations of a dyad presented in a block. c) Illustration of the flickering presentation of the stimuli: the two bodies flickered at the frequencies F1 and F2, respectively. The contrast-modulated appearance of body images is shown over a cycle of 99 frames, corresponding to frequency Fd, at which both bodies appeared in full contrast. The blue brackets represent a zoomed-in illustration of the frames where the simultaneous appearance of the two bodies at their full contrast occurs (frames 51 and 52, as marked with the blue arrow).

Figure 2. Frequency-tagged responses and their modulation depending on the experimental conditions. a) Grand average Signal-to-Noise-Ratio (SNR) over the group of participants, averaged over all EEG sensors. Peaks are visible at the stimulation frequencies corresponding to single bodies and dyads. Adjacent to the peak, the spatial distribution of responses is shown on a topographical map. b) Response to upright and inverted single bodies in a posterior cluster (body-specific cluster) as highlighted on the topographical maps. Average SNR value within this cluster at frequencies F1 and F2 are highlighted in red for upright bodies and in blue for inverted bodies. c) Response to facing and non-facing dyads in the posterior body-specific cluster of sensors (left), and in the anterior cluster pointed out by the analysis across all EEG sensors (right). These clusters are highlighted on the topographical maps and the average SNR values within each of them is plotted below the corresponding topographies, showing higher activations for facing dyads (in magenta) than for non-facing dyads (in black).

Figure3. Covariation of activations across the separate sensor clusters. Pearson correlation between SNR values of the anterior and posterior clusters showing the effect of body positioning (facing *vs.* non-facing).

Tables and Figures

Table 1.

Frequency	F1 6.66 Hz	2 x F1 13.33 Hz	3 x F1 20.00 Hz	4 x F1 26.66 Hz	5 x F1 33.33 Hz	6 x F1 40.00 Hz	7 x F1 46.66 Hz	8 x F1 above 50 Hz	9 x F1 above 50 Hz
Subjects (out of 20) with significant responses	20	20	19	20	13	18	14	--	--
Statistics	t = 8.7, p <0.001	t = 8.1, p <0.001	t = 5.3, p <0.001	t = 7.3, p <0.001	t = 2.3, p = 0.110	t = 5.6, p <0.001	t = 2.5, p = 0.071	--	--
Frequency	**F2** **5.45 Hz**	**2 x F2** **10.90 Hz**	**3 x F2** **16.36 Hz**	**4 x F2** **21.81 Hz**	**5 x F2** **27.27 Hz**	**6 x F2** **32.72 Hz**	**7 x F2** **38.18 Hz**	**8 x F2** **43.63 Hz**	**9 x F2** **49.09 Hz**
Subjects (out of 20) with significant responses	18	19	20	18	19	17	15	15	13
Statistics	t = 6.2, p <0.001	t = 6.9, p <0.001	t = 6.3, p <0.001	t = 4.0, p = 0.003	t = 4.5, p = 0.001	t = 4.5, p = 0.001	t = 3.7, p = 0.007	t = 2.9, p= 0.04	t = 2.6, p = 0.07
Frequency	**Fd** **0.606 Hz**	**2 x Fd** **1.21 Hz**	**3 x Fd** **1.81 Hz**	**4 x Fd** **2.42 Hz**	**5 x Fd** **3.03 Hz**	**6 x Fdx** **3.63 Hz**	**7 x Fd** **4.24 Hz**	**8 x Fd** **4.84 Hz**	**9 x Fd** **above 5Hz**
Subjects (out of 20) with significant responses	20	19	14	15	16	14	13	13	--
Statistics	t = 7.5, p <0.001	t = 6.4, p <0.001	t = 3.6, p = 0.007	t = 3.2, p = 0.018	t = 2.5, p= 0.083	t = 2.0, p =0.229	t = 2.2, p = 0.163	t = 1.2, p = 0.934	--

Figure 1.

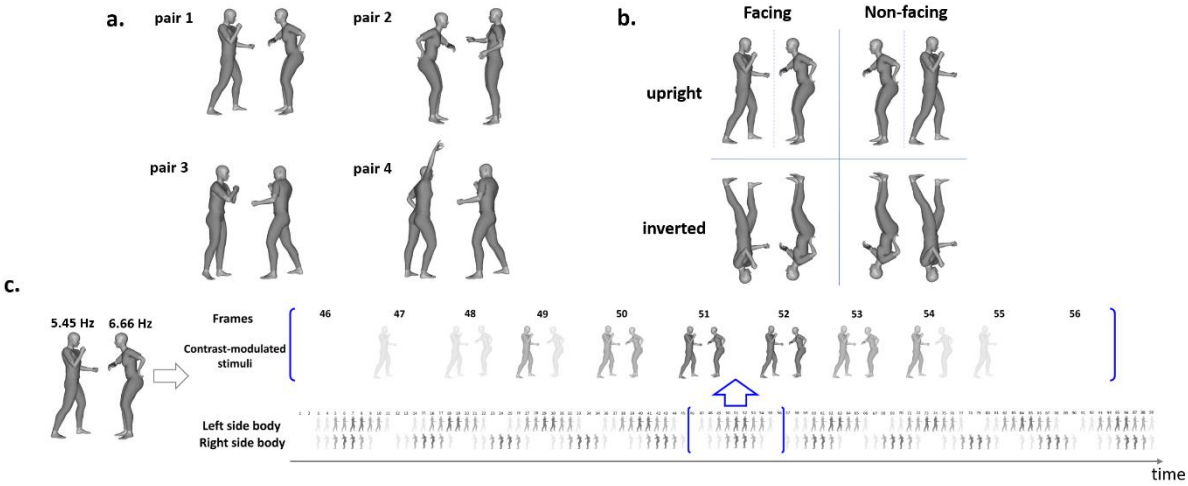

a.
pair 1
pair 2
pair 3
pair 4
b.
Facing
Non-facing
upright
inverted
c.
5.45 Hz
6.66 Hz
Frames
Contrast-modulated stimuli
46
47
48
49
50
51
52
53
54
55
56
Left side body
Right side body
time

Figure2.

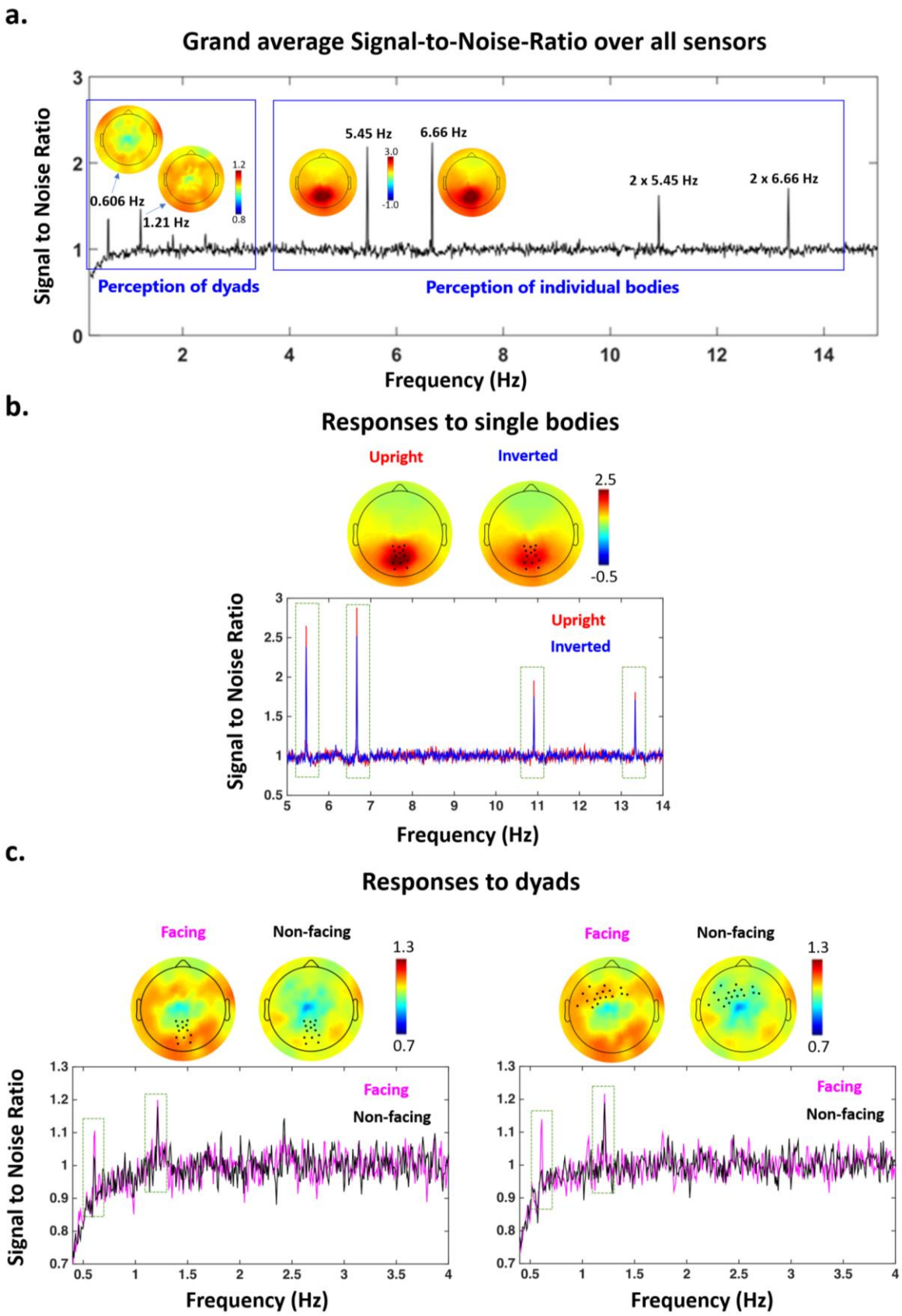
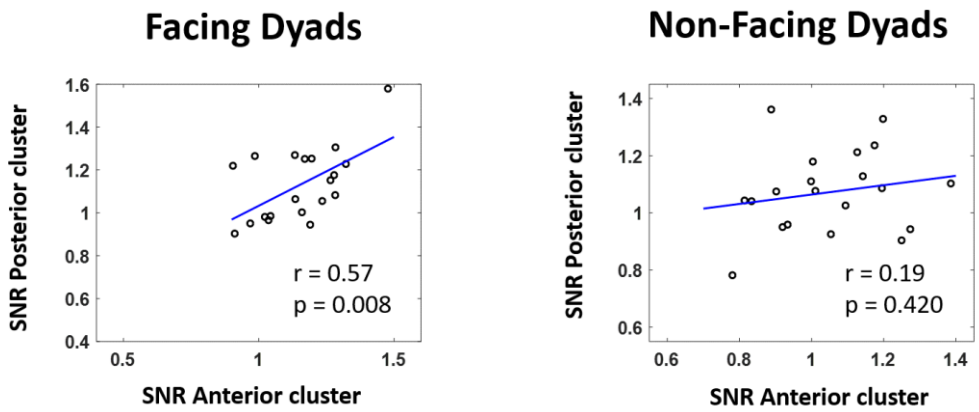

a.
Grand average Signal-to-Noise-Ratio over all sensors
Signal to Noise Ratio
0.606 Hz
1.21 Hz
1.2
0.8
5.45 Hz
6.66 Hz
3.0
-1.0
2 x 5.45 Hz
2 x 6.66 Hz
Perception of dyads
Perception of individual bodies
Frequency (Hz)
b.
Responses to single bodies
Upright
Inverted
2.5
-0.5
Signal to Noise Ratio
Upright
Inverted
Frequency (Hz)
c.
Responses to dyads
Facing
Non-facing
1.3
0.7
Signal to Noise Ratio
Facing
Non-facing
Frequency (Hz)
Facing
Non-facing
1.3
0.7
Facing
Non-facing
Frequency (Hz)

Figure 3.


Facing Dyads
SNR Posterior cluster
SNR Anterior cluster
r = 0.57
p = 0.008
Non-Facing Dyads
SNR Posterior cluster
SNR Anterior cluster
r = 0.19
p = 0.420