Melt into the group

Electrophysiological Evidence of Gestalt Perception of Human Dyad

Karima Mersad[a], Céline Caristan[a]

[a] Institut de Psychologie, Université Paris Descartes, France

Corresponding author: karima.mersad@parisdescartes.fr

Abstract

It has been showed recently that the human brain has dedicated networks for perception of human bodies in synchronous motion or in situation of interaction. However, below motion and interaction, how does the brain process a simple plurality of humans in close positioning? We used EEG frequency tagging technique to investigate integration of human dyad elements in a global percept. We presented to participants images of two silhouettes, a man and a woman flickering at different frequencies (5.88 vs. 7.14 Hz). Clear response at these stimulation frequencies reflected response to dyad parts, both when the dyad was presented upright and inverted. However, an emerging intermodulation component (7.14 + 5.88 = 13.02 Hz), a nonlinear response regarded as an objective signature of holistic representation, was significantly enhanced in upright relatively to inverted position. Inversion effect was significant only for the intermodulation component as opposed to stimulation frequencies revealing that dyad configuration perception overrides structural properties of dyad elements. Inversion effect was not significant for a pair of non-human objects. Our results show that merely facing two humans in close positioning leads to perceptually bind them and suggest that the perception of individuals is of different nature when they form a plurality.

Introduction:

As intuitive social animals, all of us have noticed that an individual alone and the same individual in a group are not the same psychological beings. Yet, research in social psychology has revealed how important the influence of others is on our behavior. In Solomon Asch's seminal study on social conformity, participants often went with the majority (actually partners of the experimenter responding according to a prearranged plan) giving the same wrong answer in a trivial visual-perception task (Asch, 1955). To have influence on our minds and behavior, others do not need to be active. Indeed, in his work on social facilitation (Zajonc, 1965) showed that the mere presence of conspecifics could increase performance in certain tasks and have detrimental effects on other tasks. This effect of the group upon individuals likely reflects sensitivity to complex and subtle dynamics of human aggregates. It must have its roots in specific perceptual mechanisms amongst which vision could be the most important apparatus (Nakayama et al., 2010). Behavioral and brain-imaging studies have started to characterize, in particular, brain mechanisms dedicated to the perception of individuals engaged in social interaction (Abassi & Papeo, 2020; Isik et al., 2017; Papeo et al., 2017; Walbrin et al., 2018). It is remarkable that this set of experimental research presupposes that a plurality of individuals is essentially represented as the result of explicit interactions between those individuals. An alternative to this view is that the human brain tends to process any plurality of individuals who are merely close together in space as a coherent unit based on a global configuration. Little is known about the process of human group perception, independently of interaction phenomenon, when an individual merely faces a plurality of conspecifics.

As physical entities, humans are visually perceived as faces and/or bodies. Faces and bodies are special stimuli of the environment; they are present from the very first moments of life, and all along development. These singular animate objects are crucial to our living, and even to our survival: they provide information, knowledge, provoke emotion, and might have to do with what philosopher Spinoza calls *conatus*, that is " the effort by which each thing, as far as it can by its own power, strives to persevere in its being ". Unsurprisingly, evolution has endowed human with highly specialized perceptual visual systems so as to process human faces and bodies differently than other objects. For example, when present in a visual scene, faces are immediately detected as they automatically elicit rapid saccades towards them (Crouzet et al., 2010). The human brain exhibit a specific electrophysiological signature when processing this special visual object (Bentin et al., 1996; Bötzel et al., 1995; Eimer, 2011) and

selectively activate dedicated brain networks localized in the right hemisphere (Rossion et al., 2012). Similarly, human bodies's posture, silhouette or movement perception elicits specific behavioral and cerebral mechanisms. For example, adults identify conspecific's biological motion like walking in few hundred milliseconds by mere observation of few point lights displayed on the joints of the walker (Johansson, 1973, 1976). This perceptual expertise is attested by neuroimaging studies in adults revealing the existence of dedicated brain regions to body perception. For example, a study by (Jiang et al., 2001) showed that adults selectively activate the right hemisphere's lateral occipital area when perceiving human bodies compared to other types of objects.

What are the characteristics of the mechanism underpinning face and body perception? In a seminal publication, Young et al. (1987) brought insight on this question. They showed that when aligning the upper half part of a face (A) with the lower half part of another (B), the resulting composite face induces a percept of a new face. Most interestingly, this new face is perceived coherent enough to impair the upper half of face A recognition as belonging to the face A. These results have been interpreted by their authors as a demonstration that the human brain do not process face component features locally and as separate parts but rather integrate them in a strongly coherent percept based on their global configuration (distance and relative position between parts). This interpretation is supported by the finding that the perception of the composite face as a whole is impaired when the faces are presented upside-down (Young et al., 1987). Indeed, the inversion effect that is poorer accuracy and longer reaction times for faces recognition when faces are upside down is classically taken as evidence for holistic or configural processing. Importantly for our purpose Reed et al., (2006) using the inversion paradigm, evidenced that human body perception phenomenon involves, as well, configural processing and the computation of relations among body components (arms, feets etc..).

Besides, recent work has revealed distinctive mechanisms and cortical specialization for the perception of a group of two humans (a dyad) in interaction. In their study Papeo et al., (2017) showed that inversion impairs detection of facing dyads (two persons – faces and bodies – positioned face-to-face seemingly interacting) more than non-facing dyad. Moreover, facing dyads has been found to recruit more attention than non-facing dyads as showed by Papeo & Abassi (2019) using a visual search task. Finally, compared with non-facing bodies, facing bodies increase activity in specific areas of the visual cortex (Abassi & Papeo, 2020). These behavioral and brain-imaging methods provide evidence for configural processing of a group of (two, facing) humans, but leaves unanswered an important question: does the

presence of a group configuration modifies the perception of individuals that form the group, merely by the fact of their presence inside a group?

Multi-input frequency tagging electrophysiological technique has been recently used to probe configural processing while dissociating the activity underlying the response to whole configuration from the response to local elements forming the image. In this paradigm, different parts of the visual stimuli are presented at different frequencies (for example, two images presented at f1 and f2 frequencies). In addition, to generate steady state visual evoked potentials (SSVEP) at f1, f2 and at their harmonics (nf1 and mf2, m and n being integer numbers), this method allows to capture additional components, known as intermodulation terms (IM). These components are not present in the input and result from non-linear interaction between populations of neurons responding to the image parts (two, in this example). IM responses occur at frequencies that are sums and differences of the different harmonics (nf1 ± mf2).

Frequency Tagging technique is becoming an important tool for studying human cognition, providing objective signatures of neural processes underlying cognitive functions. Using this method Boremanse et al., (2014) brought experimental evidence of an objective neural signature of perceptual binding of face parts. Taking benefit of the same technique Alp et al., (2016) could predict precisely and subsequently demonstrate the presence of neural processes specifically involved in the emergence of illusory surface perception (a Kanisza square). More recently Radtke et al., (2020) used Frequency Tagging method to show that the extraction of the meaning of a scene (Gist perception) is correlated with neural association between simultaneously presented objects sharing semantical relation.

In the present study, we aimed at targeting the neural process that takes place during the perception of a group of humans in its simplest form: two individuals in close positioning. Our hypothesis is that dyad, just like individual faces and bodies, is processed as a visually structure unit, even when there is no clearly depicted interaction between its parts. To assess the emergence of Gestalt-like properties resulting from the perception of a dyad, we recorded EEG response to the simultaneous presentation of two human silhouettes tagged at different frequencies. We intended to dissociate objectively the response to local elements forming the dyad from the response to the global dyad configuration. To better investigate the social character of the holistic representation, we also presented pairs of chairs as non-human control stimuli. We expected the perception of individual human silhouettes (i.e. parts of the dyad) and individual chairs (parts of the pair) to give rise to EEG response at the precise two fundamental frequencies and their harmonics (e.g., f1, f2, 2f1, 2f2, 3f1,..). Because perceiving

two nearby humans can evoke a configuration having social meaning but perceiving two adjacent chairs evokes no distinctive configuration, we expected dyad stimulus to trigger a higher response at IM frequencies (f1+f2, f1-f2, 2f1+f2, etc) than chairs. As the mechanism expected here goes beyond simple pairings of objects belonging to the same semantic category and to involve holistic perception, we also presented inverted silhouettes of dyad and of pair of chairs. We predicted that the response to the global dyad configuration at IM components would be reduced by inversion but not the response to individual silhouettes at stimulation frequencies (Jacques & Rossion, 2009; Suzuki & Cavanagh, 1995).

Method

Participants

30 adults (three left-handed; 17 females; age range: 18–43, mean age 20.09 with normal or corrected-to-normal vision participated in the study. 4 additional participants were tested but their data were not included in the analysis: 2 participants due to the presence of continuous noise in their recording, making the data unexploitable and 2 participants due to an error made by the experimenter during the recording. Written informed consent was obtained from all participants before experiment that was approved by the Research Ethical Committee of the University of Paris.

Stimuli

Two black silhouettes, a man and a woman (head and shoulders) and two black chairs were used as central displays on a white background. Silhouettes, known to induce strong inversion effect (Stein et al., 2012) were used instead of pictures to avoid processing of local face features. The image of chairs were selected to be different from each other and to match a female and male category, for example, the 'female' chair was slightly thinner (see Figure 1). Stimuli were displayed as pairs of humans and chairs presented upward or upside-down. To facilitate perceptual binding of silhouettes constituting the pair, silhouettes identity was held constant and their position, slightly leaning to each other, evoked closeness, in upward as well as in inverted position. Silhouettes and chairs pictures subtended approximately the following dimensions of visual angles: Man silhouette: 7.77x7.67°, Woman silhouette: 7.20x6.79°, Female chair: 5.85x7.69° Male chair: 6.21x7.69°. Distances between stimuli were

controlled between human dyad and chair pairs: mean distance between the two closest points: 1.61° of visual angle for both type of stimuli. At the center of silhouette pairs, a black cross (0.3x0.1°) was displayed at random times at the relative same location (middle of the faces and upper chair parts) in upright and inverted condition. The selection of chairs as control stimuli was inspired by a recent study showing a stronger inversion effect for pairs of bodies than for pairs of chairs (Papeo et al., 2017). Care has been taken for stimuli selection. Firstly, to maximize signal-to-noise ratio, the fundamental frequencies and the sum intermodulation (IM) component (5.88 Hz + 7.14 Hz = 13.02 Hz) were located outside of the alpha (8–12 Hz) range (Regan, 1989). In addition, both 5.88 and 7.14 were integer divisions of the 1000 Hz refresh rate. Finally, stimulation frequencies were selected in the (4-9 Hz) theta-band interval reported to elicit maximal responses to contrast-modulated stimulation with social stimuli like faces (Boremanse et al., 2014).

Procedure

After electrode-cap placement, participants were seated in a dimly lit room at a distance of approximately 70 cm of the sceen. Sitting chair was adjusted so that participant's eyes were level with the center of the screen. Stimuli were displayed using an in-house application written in Python Psychopy software (Peirce et al., 2019). Figure 1 represents an example of trial time course. Each trial lasted 70 s during which participants viewed pairs of stimuli contrast modulated respectively at f1 and f2. Each participant viewed a total of 12 trials comprising 4 trials where human silhouettes were presented in upward position, 4 trials where human silhouettes were presented in Inverted position, 2 trials where chair silhouettes where presented in upward position and 2 trials where chair silhouettes were presented in Inverted position. Each human silhouette was alternatively flickering at f1 or f2 and was alternatively appearing on the left or right side in a random and counterbalanced fashion. Each chair was alternatively appearing on the left or right side in a random and counterbalanced fashion. Note that unlike (target) Human stimuli, (control) Chair stimuli were not counterbalanced for flickering frequency (a fixed frequency was associated to each chair). This resulted in less trials in Chair than in Human condition. Our rationale was to reduce the fatigue and discomfort driven by the cumulative flickering throughout the experiment and maintain the quality of the recorded data. Two trials were separated by a white screen during 8 s and a break was proposed after 6 trials. The order of presentation of trials was randomized across participants. To help maintaining attention throughout the duration of the trial, participants had to press the space bar whenever the cross was appearing or disappearing at

random times at the center of the screen, between the two silhouettes. EEG activity was recorded using a Brainvison amplifier system (Brain Products) with 64 electrodes referenced to the vertex. EEG was digitized at a 1000 Hz sampling rate. To record a highly precise timing of trial onset, triggers were sent from the parallel port of the stimulation computer to the amplifier.

Data analysis

EEG analysis was performed using custom-made MATLAB scripts and EEGLAB toolbox (Delorme & Makeig, 2004). The data were resampled to 250 Hz to increase the speed of data processing by reducing the workload. Data were band-pass filtered between 0.1 and 100 Hz so as to remove slow drifts and very high frequencies. After visual rejection of paroxysmal portions of the continuous recording, data were segmented into windows of 66.67 s comprising exactly 392 cycles of f1 (5.88 Hz) and 476 cycles of f2 (7.14). Our purpose in choosing a long duration epoch (66.67 s) was to obtain an amplitude spectrum with a high-frequency resolution (0.015 Hz = 1/66.67). We then removed the first 2 seconds of the recording to avoid artifacts caused by the onset of visual stimulation and to give time to the brain to be entrained by the stimulation. In a following step, we performed automatic rejection of epochs and channels contaminated by artifacts. More precisely, epochs were considered unsuitable for analysis if their fast average amplitude exceeded 250 μV or their deviation between fast and slow running averages exceeding 150 μV. For each subject, channels that had more than 50% of epochs marked as unsuitable were considered as bad channels. 3.4 % of channels on average were rejected per subject. Trials having more than 80% of bad channels were rejected from the analysis. On average, 6,3 % of epochs by subject were rejected. For each participant, epochs of the same condition were averaged to increase the signal to noise ration of the EEG response. We obtained 4 conditions per participant: HumanUp (Human silhouettes in upside position); HumanInvert (Human silhouettes in inverted position), ChairUp (Chair silhouettes in upside position), ChairInvert (Chair silhouettes in inverted position). The resulting averaged waveforms were then submitted to a Fourier decomposition. For each frequency component, the amplitude was taken as the magnitude of the complex number resulting from the FFT for this frequency. Figure 2 shows the amplitude spectrum describing the contribution of each frequency and their topographical maps. SNR of each frequency was then computed as its z-score. More precisely, for each participant, each condition and each electrode, SNR was computed as the amplitude of the

signal at the frequency of interest divided by the mean of the amplitude of the 30 neighboring frequency bins (15 from both sides), excluding two bins (one from both sides) adjacent to the bin of interest (Rossion & Boremanse, 2011; Srinivasan et al., 1999). To determine which frequency was above noise level, we computed average of frequency SNR of all participants, all electrodes and all conditions. Five frequencies prove to be above noise threshold ($z > 2$, $p < 0.01$): stimulation frequencies: f1 and f2, two harmonics 2f1 and 3f2 and the second order sum IM component: f1+f2. To perform statistical analysis we selected adjacent electrodes with a maximal SNR for the fundamental frequencies, as classically carried out in similar paradigms (e.g., Alp et al., 2016, 2017). 12 electrodes located over parieto-occipital region were selected: CP1, CPZ, CP2, CP4, P1, PZ, P2, P4, POz, PO4, PO8, Oz, O2 (white circles on the maps in Figure 3). Figure 3 depicts SNR frequency spectrum, computed on the region of interest, separately for average Human and Chair conditions in both up and inverted position presentation. A repeated measure analysis of variance (ANOVA) was computed on the same region, with SNR as dependent variable and two factors: Position (Up, Inverted) and Object (Humans, Chairs). We analyzed separately both sets of frequencies selected above noise threshold: on the one hand, stimulation frequencies and their harmonics, on the other hand, IM component.

Results

As can be observed on Figure 2, Amplitude spectrum averaged on all participants, all conditions and all electrodes shows a clear response at the two stimulation frequencies (f1= 5.88 Hz, and f2 = 7.14 Hz) and their harmonics (2f1= 11.76 Hz, 2f2 = 14.28, 3f1= 17.64 Hz, 3f2 = 21.42). As depicted on topographical maps, the response at stimulation frequencies and harmonics was mainly located in occipital region. We observe a typical decrease of spectral amplitude with frequency, nevertheless maintaining second and third order harmonics clearly above noise. Most importantly for our hypothesis, another spectral amplitude peak in the same region is observed at the frequency f1+ f2 (5.88 + 7.14 = 13.02) corresponding to the second order intermodulation frequency. We used SNR values to statistically evaluate and compare stimulation and IM frequencies in the different conditions. To do so, we averaged SNR of both stimulation frequencies f1 and f2, in the one hand, and stimulation frequencies and their second harmonics, on the other hand (the results were similar when including also the third harmonic in the average).

Response at stimulation frequencies and their harmonics

First harmonics (stimulation frequencies)

For the mean f1 and f2 SNR, we observed a significant main effect of Object ($F(1, 29) = 25.64$, $p < 0.001$, $\eta^2 = 0.46$), with a higher SNR for Human stimuli (M = 4.82, SD = 0.89) than for Chair stimuli (M = 4.17, SD =0.76), effect size d = 0.79. We observed no main effect of position ($F(1, 29) < 1$) nor interaction between the two factors $F(1.29) < 1$. Figure 4 depicts SNR by condition.

Mean first and second harmonics

Results were similar for mean f1, f2, 2f1, 2f2 SNR (Figure 4), we observed a significant main effect of Object ($F(1, 29) = 13.46$, $p < 0.001$, $\eta^2 = 0.32$) with a higher SNR for Humans (M = 3.52, SD = 0.57) than for Chairs: (M = 3.19, SD = 0.51), effect-size d = 0.67. There was no main effect of position ($F(1, 29) < 1$) nor interaction between the two factors ($F(1, 29) = 1.79$, $p = 0.19$).

In summary, there was clear response at stimulation frequencies and their harmonics to human dyad parts and so was the response to the pair of chairs parts. Only one effect was significant, the effect of Object and there was no significant interaction between Object and Position factors.

Response at second order sum intermodulation component

For the single IM frequency component exceeding the noise threshold, f1 + f2 = 13.02, we observed a significant main effect of Object ($F(1,29) = 7.07$, $p = 0.01$, $\eta^2 = 0.19$) with a higher SNR for Human stimuli (M = 2.12, SD = 0.88) than for Chair stimuli (M = 1.79, SD = 0.95), effect size d = 0.35. We observed also a significant main effect of position ($F(1, 29) = 4.37$, $p = 0.04$, $\eta^2 = 0.14$) with higher SNR for upright than inverted position (UP: M = 2.14, SD = 0.98 [Humans: 2.43-1.0 ; Chairs: 1.86-0.89]; inverted M = 1.76, SD = 0.83 [Humans: 1.79-0.59; Chairs: 1.73-1.02]) effect size d = 0.42. There was a significant interaction between the two factors (($F(1, 29) = 4.56$, $p = 0.04$). Figure 3 panel A shows a peak at the second order sum IM component in Human condition that is higher in Up than in Inverted position. Visual inspection of SNR topographical distribution shows that this inversion effect is the largest over right occipital cortex, as reported in previous studies using similar paradigm (e.g., Boremanse et al., 2013) and central parietal areas (Figure 5). Post-hoc comparisons confirmed that the Position effect was driven by a significantly higher response

in HumanUp (M = 2.43 SD = 1.0) condition than in HumanInvert (M = 1.79, SD = 0.69) condition (t (29) = 3.16, p = 0.003, Bonferroni corrected, pcor < 0.025), effect size d = 0.75. The IM response showed similar response magnitude in ChairUp (M = 1.86, SD = 0.89) and ChairInvert (M = 1.73, SD = 1.02) conditions (t < 1).

Discussion:

We hypothesized that the human visual perceptual system involves, for plurality of conspecifics, special processing that obey to gestalt principles. We studied the minimal form of a group of conspecifics, the dyad, aiming to determine whether merely perceiving two individuals in close positioning leads to perceptual grouping. Using EEG Frequency Tagging technique, we presented at two different precise frequencies each element of a dyad of silhouettes. In addition to response to parts occurring at each input frequency, we expected response to the global dyad configuration to be attested by intermodulation components. We expected inversion to impair the (holistic) response at intermodulation frequencies, but not the response to dyad parts, at stimulation frequencies. We compared the response to human dyad silhouettes to the response to silhouettes of a pair of chairs.

Response to global dyad configuration

The critical result from the present experiment was a single IM component exceeding the noise threshold, a second order sum IM response, specific to the dyad presented in upright position. Such IM response that is not present in the input frequencies is characteristic of a particular non-linear system made of population of neurons that receives and jointly processes inputs from both dyad components (Alp et al., 2016; Appelbaum et al., 2008; Regan & Regan, 1988). Thus, our result demonstrates unambiguously that the response to the global dyad configuration is different from the response to its parts. Importantly, the IM component was significantly reduced when the dyad was presented in inverted position, an effect that cannot be accounted by the response to individual silhouettes which probe to be unaffected by inverted presentation. This result supports the hypothesis of a configural processing of the dyad driven by joint, non-linear processing of dyad parts. Hence, the present study brings neural evidence, as previously established for human faces and bodies, that the visual system processes a plurality of bodies in a configural manner, and supports the hypothesis that it is sufficient to a group of individuals to be close in space to be perceived as a coherent entity.

Moreover, IM component response to the pair of chairs was significantly reduced with respect to the response to the dyad. Note that this difference is unlikely the result of the lower number of trials in Chair condition: firstly, Chair and Human distributions are similar (see Figure 4) and show comparable deviation from the mean. Secondly, overall SNR response at IM component to the pair of chairs is clearly above noise ($z > 1.64$, $p < 0.05$, on tailed, see Results section) reflecting efficient processing of pair of chairs stimuli. Importantly, even though response to chairs at intermodulation component is not null, the absence of inversion effect weakens the interpretation of a configural processing for these stimuli, and confirm previous findings that dyad inversion effect does not generalize to object-object pairs (Papeo & Abassi, 2019).

IMs can be generated by local interaction of populations of neurons whose receptive fields span separated regions of visual space tagged at different frequencies such as the border of left and right halves of a face, or figure and ground regions that abut at a boundary. In these cases, IMs has been mainly observed in medial occipital areas and almost disappeared with small physical separation (Appelbaum et al., 2008; Boremanse et al., 2013). However, when tagged stimuli are physically separated in the image, IMs, observed in more high order regions, can only stem from long-range interactions between populations of neurons that represent retinotopically distal elements (Aissani et al., 2011; Boremanse et al., 2013). In our design, stimuli were physically separated in the image. Therefore, the non-linear integration process evidenced in our data when the dyad was in upright position must have been driven by a high-order structure related to a group of (two) humans in natural, common position. It will be of interest in the future to determine if similar mechanisms underlie other human plurality configurations, and to investigate the nature of these spatial high-order social structures.

Did the silhouettes position slightly leaning to each other play a role in the integration process? Our data do not allow answering this question. In Papeo et al., (2017) study, dyads of facing bodies elicited a strong inversion effect that disappeared when the exact same two figures faced away from one another. In this study, the behavioral interaction was a critical social feature. The authors concluded: “Just like a pitcher with its spout leaning toward a glass, two facing bodies would be perceptually grouped by virtue of the joint action that they are engaged in”. It is important to note the difference with the present study: the dyad was not facing each other but merely present in front of the participant, either perceived as looking at him or from behind. The conclusion that we can draw from our study is that interaction between the parts is not a necessary condition to perceive a plurality of humans as a structured

unit, at least in the case of dyad. It would be interesting to investigate in more detail the spatial parameters of social configuration and to precise spatial relative positioning that is perceived as social proximity.

Response to individual silhouettes

EEG response at stimulation frequencies f1 and f2 and their harmonics 2f1, 2f2 (response to local dyad parts), showed a highly significant effect of Object and no effect of Position, nor interaction between these factors. This result suggests that the perception of a pair of familiar objects involves the neural processing of the objects category rather than individual objects configuration. The absence of the inversion effect for human individual silhouettes seems surprising, given numerous studies having shown that body pictures and silhouettes are subjected to inversion effect (e.g., Reed et al., 2006). This pattern of perceptual processing has been previously described as the *object–inferiority effect* by (Suzuki & Cavanagh, 1995). In a visual search process they found that the global face-level representation has prior access to local-level face features. Besides, faces holistic representation, at the neural level, is also typically based on the interpretation of how the presence of an intact whole face configuration modifies the response to parts (Jacques & Rossion, 2009; Schiltz & Rossion, 2006). Our interpretation is that from two nearby bodies' perception emerges a unitized representation that is so strong that it renders local dyad bodies constituents inaccessible to configural processing. At the cognitive level, this means that individuals are perceived differently when they are inside a dyad: individual's body structure perception is weakened leaving room to the configural properties of the group.

Conclusion

Recent experiments has revealed that dyad in face-to-face interaction is processed as a unit. These studies postulate that the relevant social feature involving specialized brain processing is the social relation between elements of the dyad, explicitly represented in the visual scene. Here, we show that perceiving two human silhouettes in close positioning involves long-range communication between neural populations responding to each silhouette and triggers configural processing. This suggests that like faces and bodies, pluralities of individuals has a special perceptual status.

Acknowledgments

We are particularly grateful to Marine Vassard for her contribution to stimuli drawing.

References

Abassi, E., & Papeo, L. (2020). The Representation of Two-Body Shapes in the Human Visual Cortex. *Journal of Neuroscience*, *40*(4), 852–863. https://doi.org/10.1523/JNEUROSCI.1378-19.2019

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. https://doi.org/10.1016/j.brainres.2011.05.051

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*, 44012.

Appelbaum, L. G., Wade, A. R., Pettet, M. W., Vildavski, V. Y., & Norcia, A. M. (2008). Figure-ground interaction in the human visual cortex. *Journal of Vision*, *8*(9), 8–8. https://doi.org/10.1167/8.9.8

Asch, S. E. (1955). Opinions and social pressure. *Scientific American*, *193*(5), 31–35.

Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological studies of face perception in humans. *Journal of Cognitive Neuroscience*, *8*(6), 551–565.

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.

Boremanse, A., Norcia, A. M., & Rossion, B. (2014). Dissociation of part-based and integrated neural responses to faces by means of electroencephalographic frequency tagging. *European Journal of Neuroscience*, *40*(6), 2987–2997.

Bötzel, K., Schulze, S., & Stodieck, S. R. (1995). Scalp topography and analysis of intracranial sources of face-evoked potentials. *Experimental Brain Research*, *104*(1), 135–143.

Crouzet, S. M., Kirchner, H., & Thorpe, S. J. (2010). Fast saccades toward faces: Face detection in just 100 ms. *Journal of Vision*, *10*(4), 16–16.

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.

Eimer, M. (2011). The face-sensitivity of the n170 component. *Frontiers in Human Neuroscience*, *5*, 119.

Isik, L., Koldewyn, K., Beeler, D., & Kanwisher, N. (2017). Perceiving social interactions in the posterior superior temporal sulcus. *Proceedings of the National Academy of Sciences*, *114*(43), E9145–E9152.

Jacques, C., & Rossion, B. (2009). The initial representation of individual faces in the right occipito-temporal cortex is holistic: Electrophysiological evidence from the composite face illusion. *Journal of Vision*, *9*(6), 8–8. https://doi.org/10.1167/9.6.8

Jiang, P. E., Shuman, M., & Kanwisher N. (2001). A cortical area selective for visual processing of the human body. *Science*, *293*, 2470–2473.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics*, *14*(2), 201–211.

Johansson, G. (1976). Spatio-temporal differentiation and integration in visual motion perception. *Psychological Research*, *38*(4), 379–393.

Nakayama, K., Reginald Jr, B., Ambady, N., & Shimojo, S. (2010). *The Science of Social Vision*. Oxford University Press.

Papeo, L., & Abassi, E. (2019). Seeing social events: The visual specialization for dyadic human–human interactions. *Journal of Experimental Psychology: Human Perception and Performance*, *45*(7), 877–888. https://doi.org/10.1037/xhp0000646

Papeo, L., Stein, T., & Soto-Faraco, S. (2017). The Two-Body Inversion Effect. *Psychological Science*, *28*(3), 369–379. https://doi.org/10.1177/0956797616685769

Pavlova, M. A. (2012). Biological Motion Processing as a Hallmark of Social Cognition. *Cerebral Cortex*, *22*(5), 981–995. https://doi.org/10.1093/cercor/bhr156

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203.

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. https://doi.org/10.1007/s00221-020-05819-6

Reed, C. L., Stone, V. E., Grubb, J. D., & McGoldrick, J. E. (2006). Turning configural processing upside down: Part and whole body postures. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(1), 73.

Regan, D. (1989). Human brain electrophysiology: Evoked potentials and evoked magnetic fields in science and medicine. *New York, NY: Elsevier, New York*.

Regan, M. P., & Regan, D. (1988). A frequency domain technique for characterizing nonlinearities in biological systems. *Journal of Theoretical Biology*, *133*(3), 293–317.

Rossion, B., & Boremanse, A. (2011). Robust sensitivity to facial identity in the right human occipito-temporal cortex as revealed by steady-state visual-evoked potentials. *Journal of Vision*, *11*(2), 16–16. https://doi.org/10.1167/11.2.16

Rossion, B., Hanseeuw, B., & Dricot, L. (2012). Defining face perception areas in the human brain: A large-scale factorial fMRI face localizer analysis. *Brain and Cognition*, *79*(2), 138–157.

Schiltz, C., & Rossion, B. (2006). Faces are represented holistically in the human occipito-temporal cortex. *NeuroImage*, *32*(3), 1385–1394. https://doi.org/10.1016/j.neuroimage.2006.05.037

Srinivasan, R., Russell, D. P., Edelman, G. M., & Tononi, G. (1999). Increased Synchronization of Neuromagnetic Responses during Conscious Perception. *The Journal of Neuroscience*, *19*(13), 5435–5448. https://doi.org/10.1523/JNEUROSCI.19-13-05435.1999

Stein, T., Sterzer, P., & Peelen, M. V. (2012). Privileged detection of conspecifics: Evidence from inversion effects during continuous flash suppression. *Cognition*, *125*(1), 64–79. https://doi.org/10.1016/j.cognition.2012.06.005

Suzuki, S., & Cavanagh, P. (1995). Facial organization blocks access to low-level features: An object inferiority effect. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(4), 901–913. https://doi.org/10.1037/0096-1523.21.4.901

Walbrin, J., Downing, P., & Koldewyn, K. (2018). Neural responses to visually observed social interactions. *Neuropsychologia*, *112*, 31–39.

Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configural information in face perception. *Perception*, *16*, 747–759.

Zajonc, R. B. (1965). Social facilitation. *Science*, *149*, 269–27.

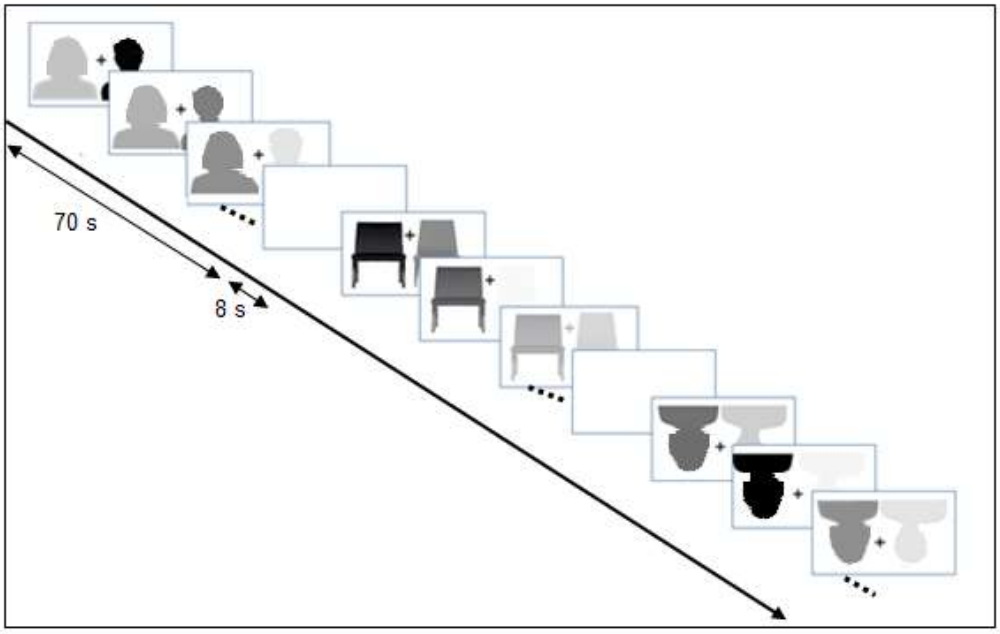

Figure 1. *Example of an experiment timecourse. In this example a trial is represented with 3 screen captures. In each trial one pair of stimulus randomly chosen from {Chairs Up, Chairs Inverted, Humans Up, Humans inverted}, was presented during 70 s while each element of the pair was contrast modulated randomly either at f1 or f2. Two trials were separated by a white screen during 8 s.*

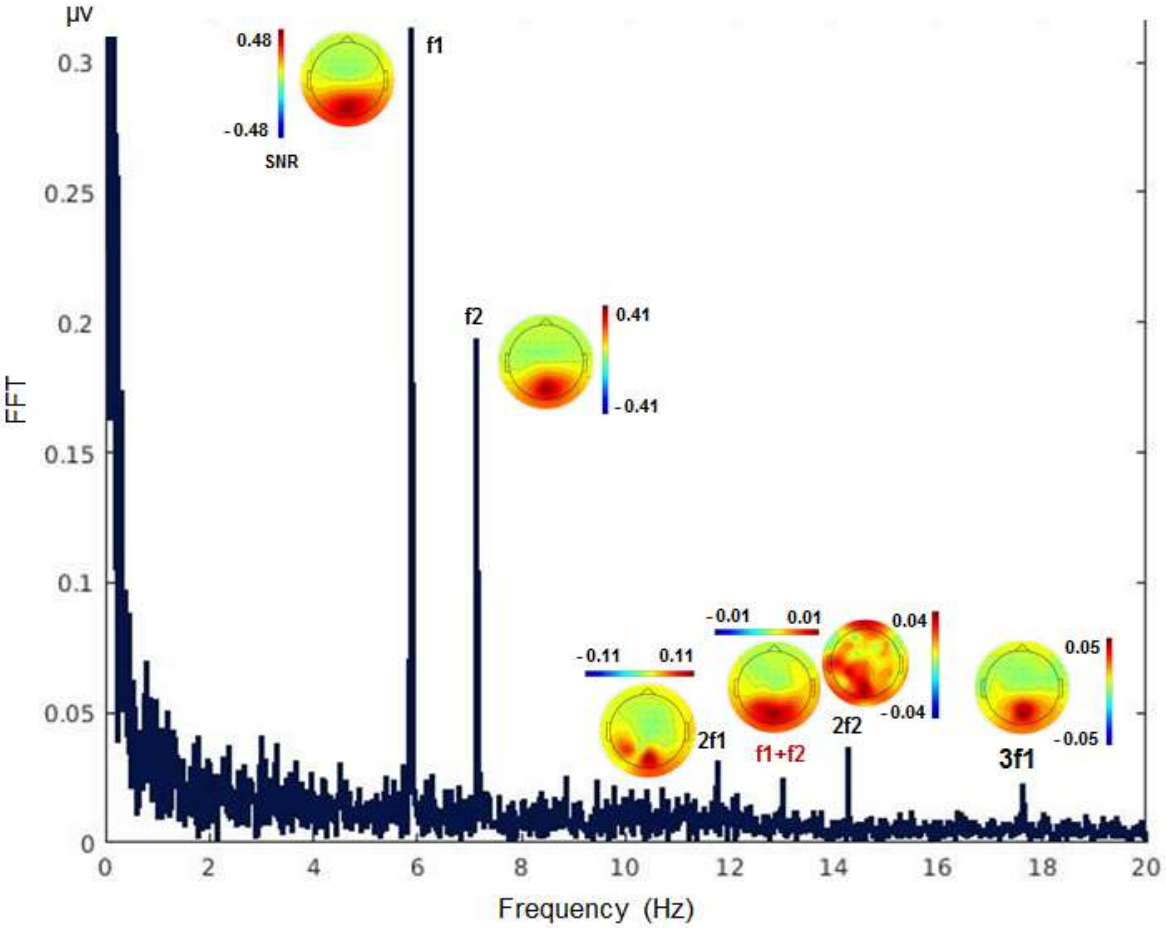

Figure 2. *Amplitude Spectrum describing each component frequency contribution (0 – 20 Hz) on average on all participants and all conditions, as well as topographical maps of stimulation frequencies, three harmonics and the second order sum IM component f1+f2 =13.02.*

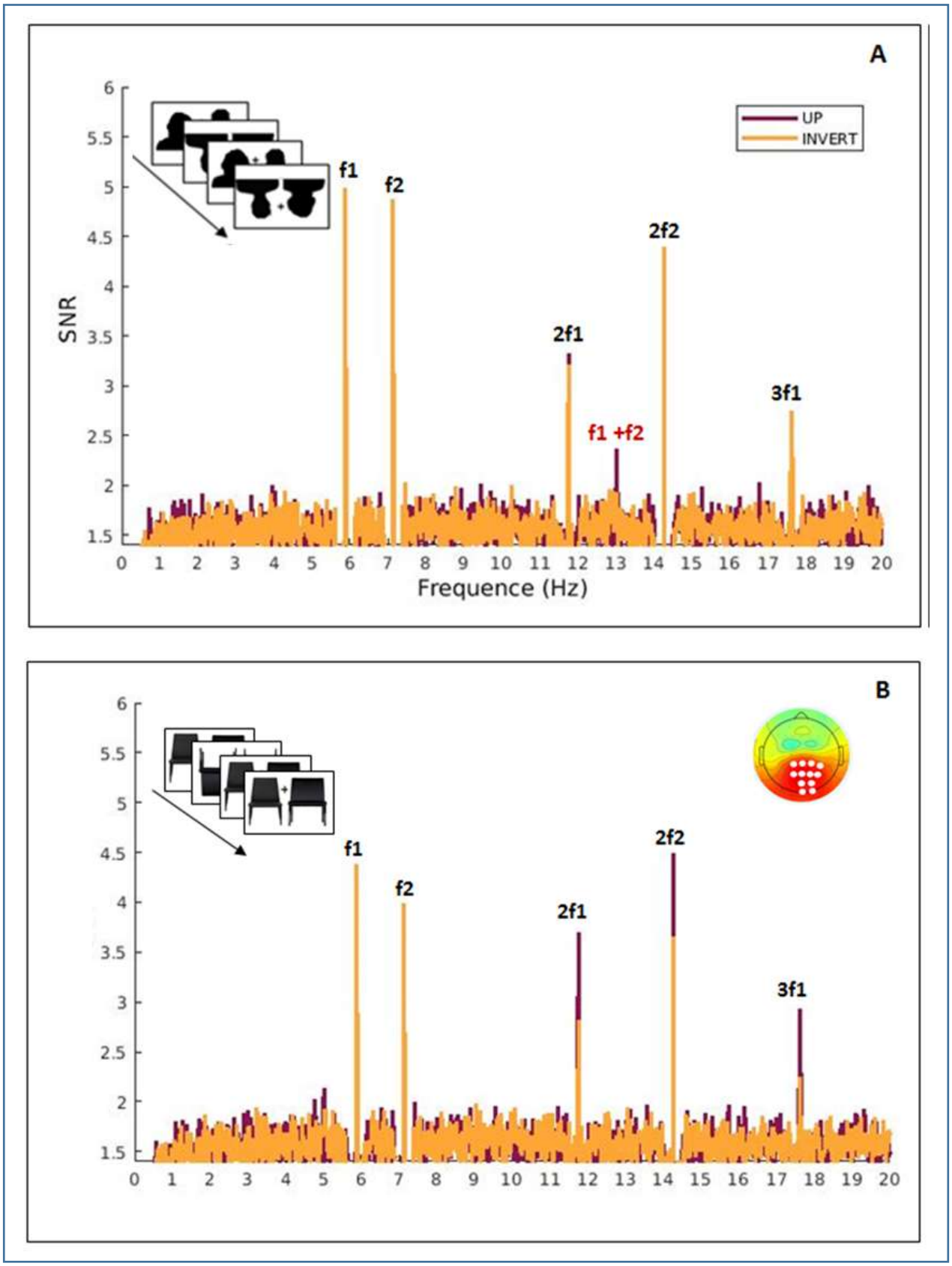

Figure 3. *Signal to Noise Ration (0-20 Hz) averaged across participants and cortical region of analysis (white dots on topographical map) for Humans (A) and Chairs (B) conditions. Up and Inverted trials are represented on both figures, with different colors, in transparency. Clear response appears at stimulation frequencies and harmonics in both conditions. The*

peak at the second order sum IM component (f1+f2 =13.02), present in Humans condition, is higher for Up than Inverted trials and is absent from Chairs condition.

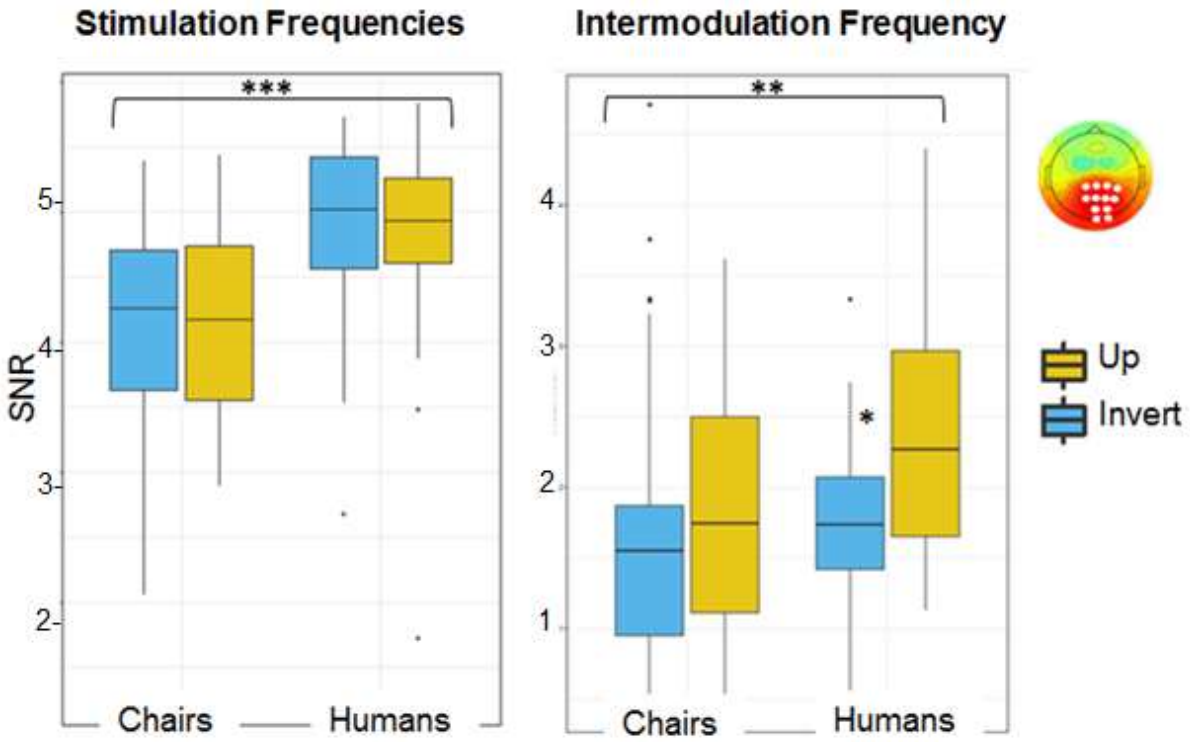

Figure 4. *Signal to Noise Ratio averaged on participants and cortical region of analysis (white dots on topographical map) broken down by level of processing (dyad parts: stimulation frequencies, global dyad configuration: IM frequency) and type of object (Humans, Chairs). The Figure depicts inversion effect in each condition. Boxplots show median, upper and lower quartiles. Error bars represents SE. Outlying points are represented individually. Response to Humans was significantly higher than response to Chairs whatever the level of processing (dyad parts or whole). Inversion effect, however, was significant only for the whole dyad level configuration in intact upright presentation.*

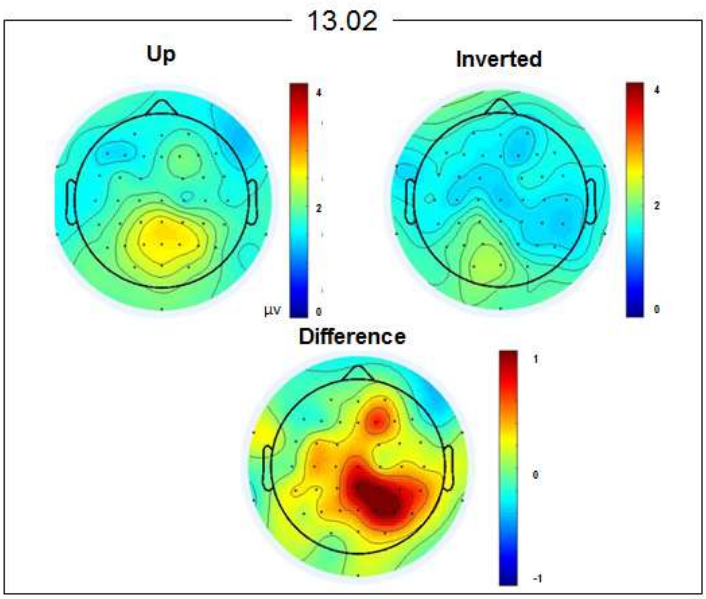

Figure 5. *Topographical maps of SNR at intermodulation frequency f1 + f2 = 13.02 Hz (response to whole dyad configuration (average participants in Humans condition only). The figure depicts the topography of Up and Inverted conditions and their difference (inversion effect). The highest inversion effect appears over right occipito-parietal regions.*

.