



HAL
open science

Blending into the Crowd: Electrophysiological Evidence of Gestalt Perception of a Human Dyad

Karima Mersad, Céline Caristan

► **To cite this version:**

Karima Mersad, Céline Caristan. Blending into the Crowd: Electrophysiological Evidence of Gestalt Perception of a Human Dyad. *Neuropsychologia*, 2021, 160, pp.107967. 10.1016/j.neuropsychologia.2021.107967 . hal-03326597

HAL Id: hal-03326597

<https://hal.science/hal-03326597v1>

Submitted on 22 Aug 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

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 shown 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 suggesting an interplay between the processing of global dyad configuration, on the one hand, and the structural properties of dyad elements, on the other hand. Our results show that merely facing two humans in close positioning leads to perceptually bind them and that the perception of individuals might be of different nature when they form a plurality.

Introduction:

What is the effect on mental activity of the presence of a group of other persons, if studied objectively? This question, asked more than a century ago by W.H. Burnham (Burnham, 1910) has started only recently to receive answers from research on visual group perception. Groups are highly salient entities in our environment. Their complex dynamics lead to emergent collective behavior that is crucial for a human being to perceive and understand, one critical example being panic signals conveyed by a crowd (Helbing et al., 2000). High sensitivity to complex and subtle dynamics of human aggregate is likely to have its roots in specific perceptual mechanisms, where vision could be the most important apparatus (Nakayama et al., 2010). Indeed, research has showed that the human visual system gathers, from a human ensemble, rapid and precise information at the ensemble level. For example, observers extract average emotion and identity from a crowd of faces (Haberman & Whitney, 2009). Observers are also able to extract a summary representation of the direction on which a group is looking (Sweeny & Whitney, 2014).

Beside these *ensemble coding* studies, focusing on statistical summary representation of human ensembles, behavioral and brain-imaging studies have started to characterize 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.

Groups are constituted of physical entities, visually perceived as faces and/or bodies. Faces and bodies are singular stimuli of our environment. Present from the very first moments of life and all along the development of human beings, they prove to be critical for our life and even for our survival: they provide information, knowledge, trigger emotions and might be related to what the 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 in order 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 will respond with a specific electrophysiological signature when processing this specific 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, the posture of human bodies, silhouettes or perception of movement trigger specific behavioral and cerebral mechanisms. For example, adults identify the biological motion of conspecifics like walking in few hundred milliseconds by the mere observation of a few point-lights displayed on the joints of the walker (Johansson, 1973, 1976). This perceptual expertise is detected by neuroimaging studies in adults revealing the existence of brain regions dedicated to body perception. For example, in a study (Jiang et al., 2001), the author showed that adults selectively activate the lateral occipital area of the right hemisphere when perceiving human bodies compared to the perception of 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 to 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 results in the 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 the 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 decreased accuracy and longer reaction times in face recognition when faces are upside down, is typically considered as an evidence for an holistic or a configural processing. What is of major importance for our purpose, Reed et al. (2006) used the inversion paradigm in order to provide the evidence that the human body perception phenomenon also involves a configural processing and the computation of relations among body components (arms, feet etc.).

Besides, recent work has revealed the presence of distinctive mechanisms and a cortical specialization in the perception of a group of two humans (a dyad) in interaction. In their study Papeo et al. (2017) showed that inversion impairs the detection of facing dyads (two persons – faces and bodies – positioned face-to-face seemingly interacting) more than non-facing dyad. Moreover, facing dyads have been found to draw 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).

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 f_1 and f_2 frequencies). In addition, to generate steady state visual evoked potentials (SSVEP) at f_1 , f_2 and at their harmonics (mf_1 and mf_2 , 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 the fact that some neurons integrate information arriving from the two parts of the image and perform non-linear processing. IM responses occur at frequencies that are sums and differences of the different harmonics ($mf_1 \pm mf_2$).

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. Benefitting from 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 positioned close to each other. Our hypothesis is that dyad, just like individual faces and bodies, is processed as a visually structure unit, even when there is no apparent interaction between its parts. To assess the emergence of Gestalt-like properties resulting from the perception of a dyad, we recorded EEG response while simultaneously presenting two human silhouettes tagged at different frequencies. We intended to dissociate objectively the response to elements forming the dyad from the response to the dyad as a whole entity. 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., f_1 , f_2 , $2f_1$, $2f_2$, $3f_1$,...). While perceiving two human

beings next to each other tends to evoke a group configuration with a social meaning, perceiving two chairs one beside the other should not evoke the same social meaning. Consequently, we expected the perception of dyad stimulus to trigger a higher response at IM frequencies (f_1+f_2 , f_1-f_2 , $2f_1+f_2$, etc.) compared to the perception of the chairs. As the mechanism expected goes beyond the simple pairings of objects belonging to the same semantic category, involving a 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.

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 for the following reasons: The results of 2 participants were excluded due to the presence of continuous noise in their recording making the data unexploitable whereas the results of 2 participants were excluded due to an error made by the experimenter during the recording. A written informed consent approved by the Research Ethical Committee of the University of Paris was obtained from all participants prior to their participation in the experiment.

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 the 2 chairs appeared different one from the other in order to be associated with a female and a male category respectively. 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 intended to evoke closeness, in upward as well as in inverted position. Silhouettes and chairs pictures subtended approximately the following dimensions of visual angles: Man

silhouette: $7.77 \times 7.67^\circ$, Woman silhouette: $7.20 \times 6.79^\circ$, Female chair: $5.85 \times 7.69^\circ$ Male chair: $6.21 \times 7.69^\circ$. 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 ($0.3 \times 0.1^\circ$) 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 ($5.88 \text{ Hz} + 7.14 \text{ Hz} = 13.02 \text{ Hz}$) 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 100 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 screen. Sitting chair was adjusted so that the eyes of the participant were in the same level as 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 including 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 were 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 randomized 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 Brainvision 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 EEG amplifier.

Data analysis

EEG analysis was conducted 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 to remove slow drifts and very high frequencies. After visual rejection of paroxysmal portions of the continuous recording, data, time-locked to the beginning of the stimulation were segmented into windows of 66.67 s from which were removed the first 2 seconds to avoid artifacts caused by the onset of visual stimulation and to give time to the brain to be entrained by the stimulation. Windows of 64.67 s comprised 380.26 cycles of f_1 (5.88 Hz) and 461.74 cycles of f_2 (7.14). These long duration epoch (64.67 s) allowed us to obtain an amplitude spectrum with a high-frequency resolution ($0.015 \text{ Hz} = 1/64.67$). 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 $250 \mu\text{V}$ or their deviation between fast and slow running averages exceeding $150 \mu\text{V}$. 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 follows: we first computed the difference between the amplitude of the signal at the frequency of interest and 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. The absolute value of this difference was then divided by the standard deviation of the frequency of interest and its neighboring frequencies to obtain SNR value (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: f_1 and f_2 , two harmonics $2f_1$ and $3f_2$ and the second order sum IM component: f_1+f_2 .

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). Our region of interest (ROI) comprised 12 electrodes located over parieto-occipital region: CP1, CPZ, CP2, CP4, P1, PZ, P2, P4, POz, PO4, 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.

The main goal of the present study was to determine whether the whole human dyad stimulus was qualitatively different from the sum of the two silhouettes forming the dyad. Our first analysis was then to examine human trials and to evaluate the response at IM frequency, regarded as a signature of holistic dyad representation. Given behavioral evidence that holistic perception is impaired by inversion, we evaluated inversion effect at IM components. However, IM power variation across Up and Inverted position has to be compared to power variation at the stimulation and harmonic frequencies, to preclude an effect driven by simple dependency of the IMs on the stimulation frequencies (Gordon et al., 2019). We then evaluated the response to dyad elements, identified by the EEG response at stimulation frequencies. The presence of an inversion effect would attest of a configural processing of dyad components. Therefore, focusing on trials presenting human silhouettes, we computed a repeated measure analysis of variance (ANOVA 1) on the selected ROI, with SNR as dependent variable and two factors, Position (Up, Inverted) and Frequency (stimulation, IM).

Lastly, we aimed at testing the social character of the holistic representation by comparing the response to a human dyad to the response to a pair of chairs.

To do so, we performed a first indirect comparison by the mean of an ANOVA 2, similar to ANOVA 1, applied, this time, to Chair trials. We further computed on the same dependent variable, an analysis including Position (Up, Inverted) and Object (Humans, Chairs) factors (ANOVA 3).

Results

As can be observed on Figure 2, amplitude spectrum averaged on all participants, all conditions and all electrodes showed a clear response at the two stimulation frequencies ($f_1 = 5.88$ Hz, and $f_2 = 7.14$ Hz) and their harmonics ($2f_1 = 11.76$ Hz, $2f_2 = 14.28$, $3f_1 = 17.64$ Hz, $3f_2 = 21.42$). 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 $f_1 + f_2$ ($5.88 + 7.14 = 13.02$) corresponding to the second order intermodulation frequency.

Neural representation of whole dyad vs. dyad elements (ANOVA 1)

The analyses were performed on the average SNR at stimulation frequencies f_1 and f_2 . Note that when performing the same analysis on the average of harmonics that prove to be above noise threshold ($2f_1$ and $3f_2$, see Data analysis section), or on the average of all first, second and third harmonics (f_1 , f_2 , $2f_1$, $2f_2$, $3f_1$, $3f_2$), we obtain similar results.

We observed a highly significant main effect of Frequency ($F(1, 29) = 210.6$, $p < 0.001$, $\eta^2 = 0.87$) due to a higher SNR at stimulation frequencies ($M = 4.82$, $SD = 0.89$) than at IM frequency ($M = 2.11$, $SD = 0.87$), effect size $d = 2.41$. We observed no main effect of position ($F(1, 29) < 1$) and a significant interaction between the two factors ($F(1, 29) = 13.73$, $p < 0.001$). Further comparisons showed that inversion effect at stimulation frequency failed to reach significance ($t(29) < 1$) but was significant at IM component ($t(29) = 3.16$, $p = 0.003$, Bonferroni correction: $p_{cor} < 0.025$) with significantly higher response in Up condition ($M = 2.43$, $SD = 1.0$) than in Inverted condition ($M = 1.79$, $SD = 0.69$).

In summary, the analysis of human dyad trials revealed the presence of an IM component that was significantly higher when silhouettes were presented upright than inverted. Stimulation

frequencies, showing higher SNR response than IM component, did not show any pattern of inversion effect.

Neural representation of whole vs. elements of a pair of chairs (ANOVA 2)

The analyses were performed on the average SNR at stimulation frequencies f_1 and f_2 . Similar pattern of results were obtained when performing the same analysis on the average of harmonics that prove to be above noise threshold ($2f_1$ and $3f_2$, see Data analysis section), or on the average of all first, second and third harmonics ($f_1, f_2, 2f_1, 2f_2, 3f_1, 3f_2$).

We observed a highly significant main effect of Frequency ($F(1, 29) = 237.1, p < 0.001, \eta^2 = 0.89$) due to a higher SNR at stimulation frequencies ($M = 4.16, SD = 0.76$) than at IM frequency ($M = 1.79, SD = 0.95$), effect size $d = 2.77$. There was no significant main effect of position ($F(1, 29) < 1$) nor significant interaction between the two factors ($F(1, 29) < 1$).

Neural representation of human dyad vs. object pair (ANOVA 3)

Response at first harmonics (stimulation frequencies)

For the mean f_1 and f_2 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 panel B depicts SNR by condition.

Response at second and third harmonics ($2f_1, 2f_2, 3f_1, 3f_2$)

Responses at second and third harmonics showed more variability when including Chair condition in the analysis, we then present these results separately.

Second harmonics

As SNR was very similar for the two second harmonic $2f_1$ and $2f_2$, we only depict the result for average SNR of these two frequencies. We obtained a main effect of Position ($F(1, 29) = 12.29, p < 0.01, \eta^2 = 0.29$) with a higher SNR for Up ($M = 3.87, SD = 0.80$) than for Inverted trials: ($M = 3.50, SD = 0.85$), effect-size $d = 0.44$. There was no significant main effect of Object ($F(1, 29) < 1$), and a significant interaction between the two factors $F(1, 29) = 17.65, p < 0.01, \eta^2 = 0.38$). Post-hoc analysis revealed that inversion effect reached significance in Chair condition where we found a higher response in Up ($M = 4.11, SD = 0.64$) than Inverted

($M = 3.22$ $SD = 0.82$) trials ($t(29) = 5.79$, $p < 0.01$, Bonferroni correction: $pcor < 0.025$, effect-size $d = 1.21$). This was not the case in Human Condition (Up: $M = 3.64$, $SD = 0.88$; Inverted: $M = 3.79$, $SD = 0.8$; $t < 1$). Figure 4 panel C depicts SNR by condition.

Third harmonics

Unlike the second harmonics, the third harmonics, 3f1 and 3f2 showed different pattern of results that we present separately (Figure 4 panel D).

For 3f1, there was no significant main effect of the factors (Object: $F(1,29) < 1$; Position: $F(1,29) = 2.1$, $p = 0.158$), however, there was a significant interaction ($F(129) = 9.50$, $p < 0.01$) due to a significant inversion effect for Chair condition where we found a higher response for Up ($M = 2.9$, $SD = 0.89$) than for Inverted ($M = 2.23$, $SD = 0.96$) trials, ($t(29) = 3.62$, $p < 0.01$, Bonferroni correction: $pcor < 0.025$, effect size $d = 0.47$). In Human condition, there was no significant difference between Up ($M = 2.61$, $SD = 1.17$) and Inverted ($M = 2.80$, $SD = 1.06$) trials, ($t < 1$).

For the third harmonic 3f2, the only main effect obtained was of the factor Object ($F(1,29) = 5.154$, $p = 0.03$, $\eta^2 = 0.15$) with a higher SNR in Chair ($M = 1.42$, $SD = 0.68$) than in Human ($M = 1.21$, $SD = 0.67$), effect-size $d = 0.31$. There was no significant effect of Object ($F(1,29) < 1$), nor significant interaction between the two factors ($F(1,29) < 1$).

In summary, there was a clear response at stimulation frequencies and their harmonics to human dyad parts and so was the response to the pair of chairs parts. At the third harmonic of the second stimulation frequency f2, there was a significantly higher response to chairs than to human silhouettes. Moreover, at the second and third harmonics of the stimulation frequency f1, there was a significant inversion effect for chair silhouettes that was absent from the response to human silhouettes.

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 a significant main effect of position ($F(1, 29) = 4.37$, $p = 0.04$, $\eta^2 = 0.14$) with higher SNR for Up 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 (see Figure 5 for topographies and electrodes showing significant response at IM component). 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, $p_{cor} < 0.025$), 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 ($t < 1$).

Figure 4 panel A depicts SNR by condition.

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 investigated inversion effect at the level of the dyad and at the level of its components, by measuring, respectively, response at intermodulation frequencies and 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, in response 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 inversion 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.

IM component response to the pair of chairs was significantly reduced with respect to the response to the dyad, even though response to chairs at intermodulation component was not null. The absence of inversion effect weakens the interpretation of a configural processing for chair silhouettes, and confirm previous findings that dyad inversion effect does not generalize to human-object or object-object pairs (Papeo & Abassi, 2019). Note that the difference in response to Chair and Human condition could have been modulated by the difference in the number of trials in the two conditions, thus, the above conclusion is a tentative interpretation of our results, that future experiments should reinforce.

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). In Boremanse et al. (2013), the sum IM, $f1+f2$, had a medial occipital topography and was reduced by physical separation but was still observed when face parts where misaligned but not physically separated. The authors proposed the interpretation that sum IM component likely reflected local spatial interaction in early retinotopic areas.

A main difference with our experiment is that the two parts of the dyad were physically separated in the image. Therefore, the sum IM component evidenced in our data when the dyad was in upright position cannot be explained by simple integration of common receptive field signals. Interestingly, unlike Boremanse et al. (2013) with face stimuli, we did not observe any significant response at IM difference ($mf1 - mf2$) components. While IMs can provide insight into the existence of an integration of distinct neural signals, the exact type of non-linear mechanism each IM underly is more challenging to decipher. Why IM components showing significant effects were different among reports of Gestalt-like processing

(Appelbaum et al., 2008; Boremanse et al., 2013)? It has been proposed that the pattern of IM responses depends on where and how the input frequencies are processed with specific neural subtypes, synaptic connections, and neural circuits within and between hierarchical levels of processing (Alp et al., 2016). Here, we used the same input frequencies as in Boremanse et al. (2013). Therefore, the absence of difference IM components in our study is unlikely due to a lower frequency tuning of the neural circuits and might be more related to the higher-level processing of the overall configuration of the image.

Future research should characterize sum and difference IM components involved in holistic perception and disentangle between potential differences in the mechanisms that gave rise to them.

Response to individual silhouettes

The absence of 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). Nevertheless, abundant support in the literature to the face and body holistic perception exists for faces and bodies perceived individually. In our study, not only one, but two busts of silhouettes were presented. Therefore, two levels of wholes-and-parts were available to perceptual processing. In the first level, the wholes are the busts and the parts are the head and the shoulders (we have intentionally removed the internal parts of faces to avoid processing of face components themselves, but clear parts still remain in the busts). In the second level, the whole is the dyad and the parts are the busts forming the dyad. In both levels, in natural configuration, the head (or the heads) is (or are) at the top of the bust (s) which allowed us to test, in both cases, for inversion effect. Our interpretation of the absence of inversion effect for human individual silhouettes is the existence of a potential trade-off between the processing of the configuration of the elements forming the dyad and the processing of the configuration of the dyad as a whole. This trade-off could be particularly present for human stimuli given that only human agents can change their behavior, both at the individual and group level. This pattern of trade-off is coherently reflected in our data: in Chair condition, we have simultaneously observed a significant inversion effect at second (mean 2f1 and 2f2) and third (3f1) harmonic frequencies (pair-element configuration level of processing) and no inversion effect at IM component (whole-pair configuration level of processing). At the contrary, in Human condition, we have simultaneously observed a significant inversion effect at the IM component and no inversion effect at the stimulation frequencies.

The trade-off hypothesis is also supported by research showing that when visually processing a set of multiple objects, the human brain uses ensemble coding mechanisms to extract summary information of all elements of the set, a serial processing of each element being computationally overwhelming and potentially useless. Moreover, in this case, an advantage for whole over individual perceptual processing has been observed. For example, Haberman & Whitney (2009) showed that human use precise summary representation of the mean emotion of an ensemble of faces (4 to 16) while retaining little information about the emotion of the individual members. Similarly, Sweeny & Whitney (2014) showed that observers perceived with high sensitivity the walking behavior of a crowd (2, 6 or 10 visible walkers from a crowd of 12) while they exhibited poorer representation of the individual's behavior. Additionally, 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. Future experiments should assess this hypothesis while controlling for the inversion effect at the individual silhouette when perceived alone.

Grouping non-human vs. human entities

In our data, we observed a significantly higher EEG responses to human dyad than to a pair of chair stimuli. Although this difference could have been modulated by the difference between the number of trials in the two conditions, it likely also reflect differential processing of human and non-human shapes captured in the entrainment measure. Yet, human bodies have a special status in the realm of objects. Behavioral findings suggest that the attentional system is tuned to detecting conspecifics compared to other objects (Bonatti et al., 2002; Stein et al., 2012) and that human body has priority for selective attention (Downing et al., 2004). Regarding pluralities, a behavioral experiment comparing accuracy scores in recognizing human dyads compared to pairs of chairs, showed higher global accuracy scores for humans relatively to chairs (Papeo et al., 2017). At the neural level, there is evidence that neural populations in the occipito-temporal cortex are specialized for processing bodies (Peelen & Downing, 2007) and dyads, at least when dyad elements are facing each other (Abassi & Papeo, 2020). It appears that the human brain devotes to social entities selective attention grounded in a cerebral system allowing priority and efficiency to their visual processing. Importantly, this social perception can automatically guide behavior (Dimberg et al., 2000). Hence, the superiority for humans over objects visual grouping found in our results might

stem from the critical importance for individuals to be able to read group intentions and to tune their behavior consequently.

Facing a dyad vs. observing a dyad facing each other

Did the silhouettes position slightly leaning to each other play a role in the integration process? We cannot exclude that the difference in the angles between Human and Chair conditions could have facilitated visual grouping, a possibility raised, for example, in Zhou et al. (2019). This potential bias should be controlled in future experiments by presenting both silhouettes strictly facing the observer. In Papeo et al. (2017) study, dyads of facing bodies elicited a strong inversion effect on recognition scores 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 group parts is not a necessary condition to perceive a plurality of humans as a structured unit, at least in the case of dyad.

Conclusion

Recent experiments have demonstrated 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 positioned close one from the other involves long-range communication between neural populations responding to each silhouette and triggers configural processing. This suggests that like faces and bodies, any plurality of individuals has a special perceptual status.

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>
- 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>
- 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.
- Bonatti, L., Frot, E., Zangl, R., & Mehler, J. (2002). The Human First Hypothesis: Identification of Conspecifics and Individuation of Objects in the Young Infant. *Cognitive Psychology*, *44*(4), 388–426. <https://doi.org/10.1006/cogp.2002.0779>
- 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.
- Burnham, W. H. (1910). The group as a stimulus to mental activity. *Science*, 761–767.
- 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.
- Dimberg, U., Thunberg, M., & Elmehed, K. (2000). Unconscious Facial Reactions to Emotional Facial Expressions. *Psychological Science*, *11*(1), 86–89.
<https://doi.org/10.1111/1467-9280.00221>
- Downing, P. E., Bray, D., Rogers, J., & Childs, C. (2004). Bodies capture attention when nothing is expected. *Cognition*, *93*(1), B27–B38.
<https://doi.org/10.1016/j.cognition.2003.10.010>
- Dunbar, R. I. (1998). The social brain hypothesis. *Evolutionary Anthropology: Issues, News, and Reviews: Issues, News, and Reviews*, *6*(5), 178–190.
- Eimer, M. (2011). The face-sensitivity of the n170 component. *Frontiers in Human Neuroscience*, *5*, 119.
- Gordon, N., Hohwy, J., Davidson, M. J., van Boxtel, J., & Tsuchiya, N. (2019). *From Intermodulation Components to Perception and Cognition-a Review*.
- Haberman, J., & Whitney, D. (2009). Seeing the mean: Ensemble coding for sets of faces. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(3), 718–734. <https://doi.org/10.1037/a0013899>

- Helbing, D., Farkas, I., & Vicsek, T. (2000). Simulating dynamical features of escape panic. *Nature*, *407*(6803), 487–490. <https://doi.org/10.1038/35035023>
- 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.
- 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>
- Peelen, M. V., & Downing, P. E. (2007). The neural basis of visual body perception. *Nature Reviews Neuroscience*, *8*(8), 636–648. <https://doi.org/10.1038/nrn2195>
- 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, 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.
- 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>
- Sweeny, T. D., & Whitney, D. (2014). Perceiving Crowd Attention: Ensemble Perception of a Crowd's Gaze. *Psychological Science*, 25(10), 1903–1913. <https://doi.org/10.1177/0956797614544510>
- 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.
- 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. <https://doi.org/10.1038/s41562-019-0618-2>

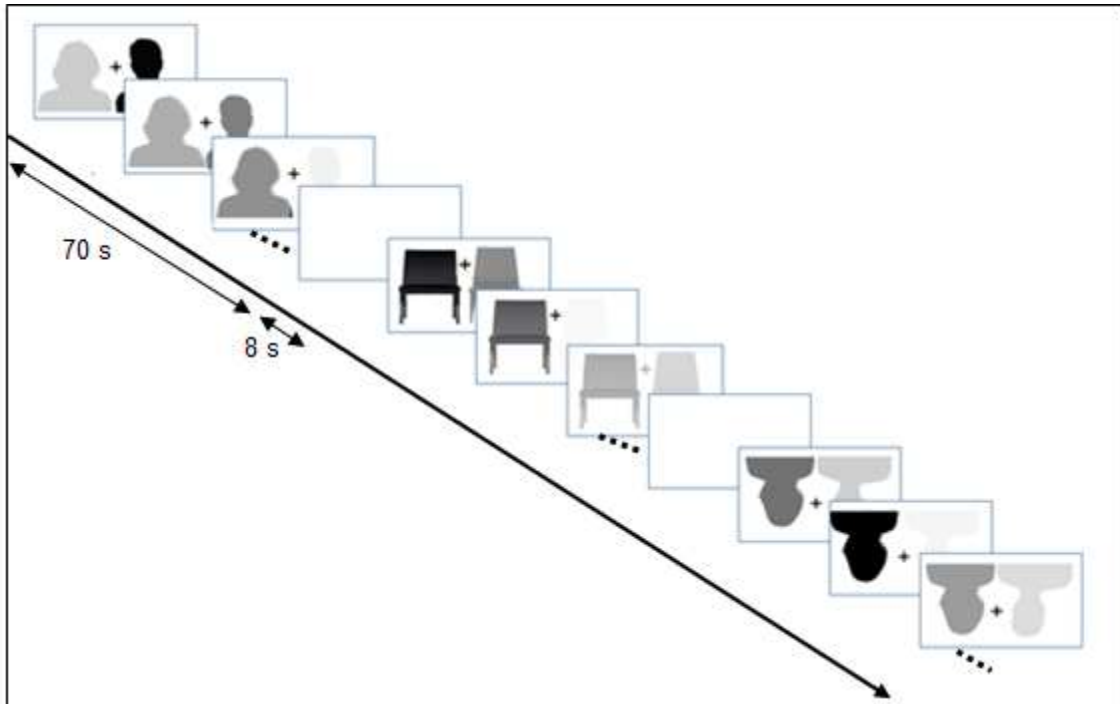


Figure 1. Example of an experiment timecourse. In this example a trial is represented with 3 screen captures. In each trial one pair of stimulus randomly chosen from {Chairs Up, Chairs Inverted, Humans Up, Humans Inverted}, was presented during 70 s while each element of the pair was contrast modulated randomly either at f_1 or f_2 . Two trials were separated by a white screen during 8 s.

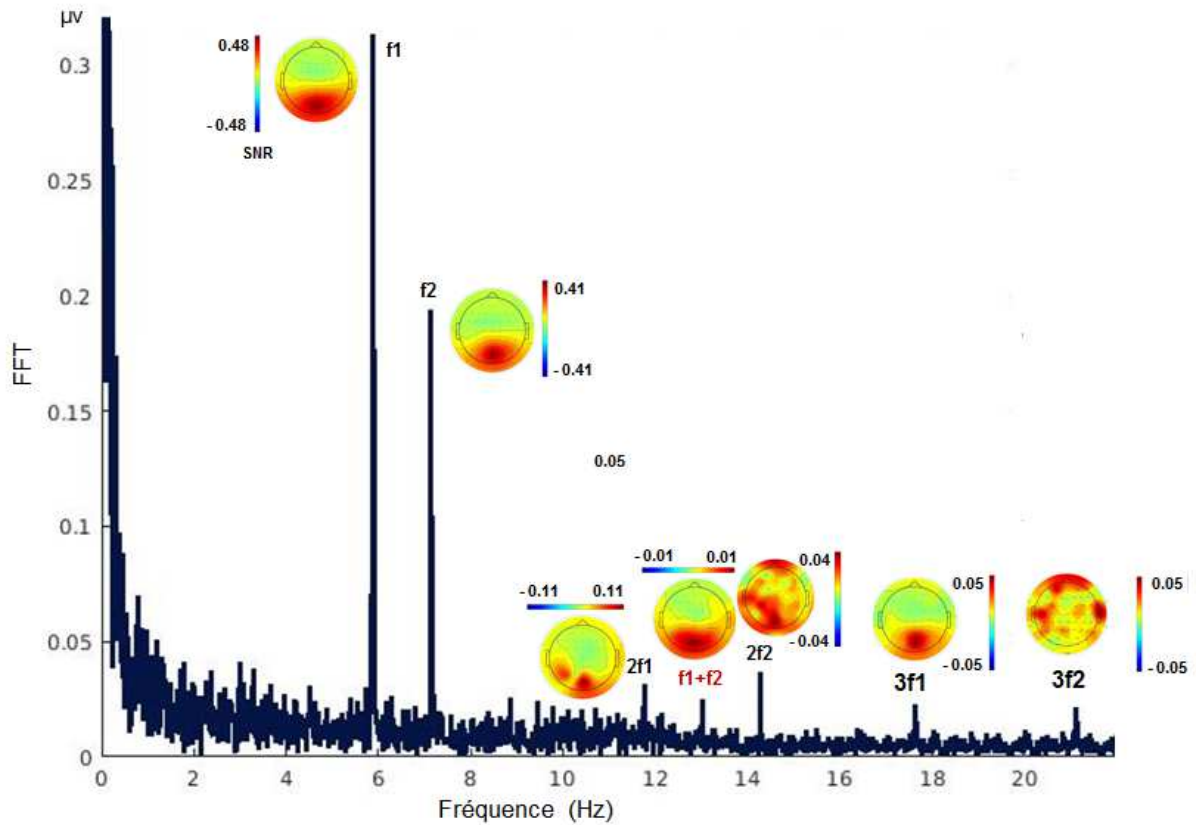
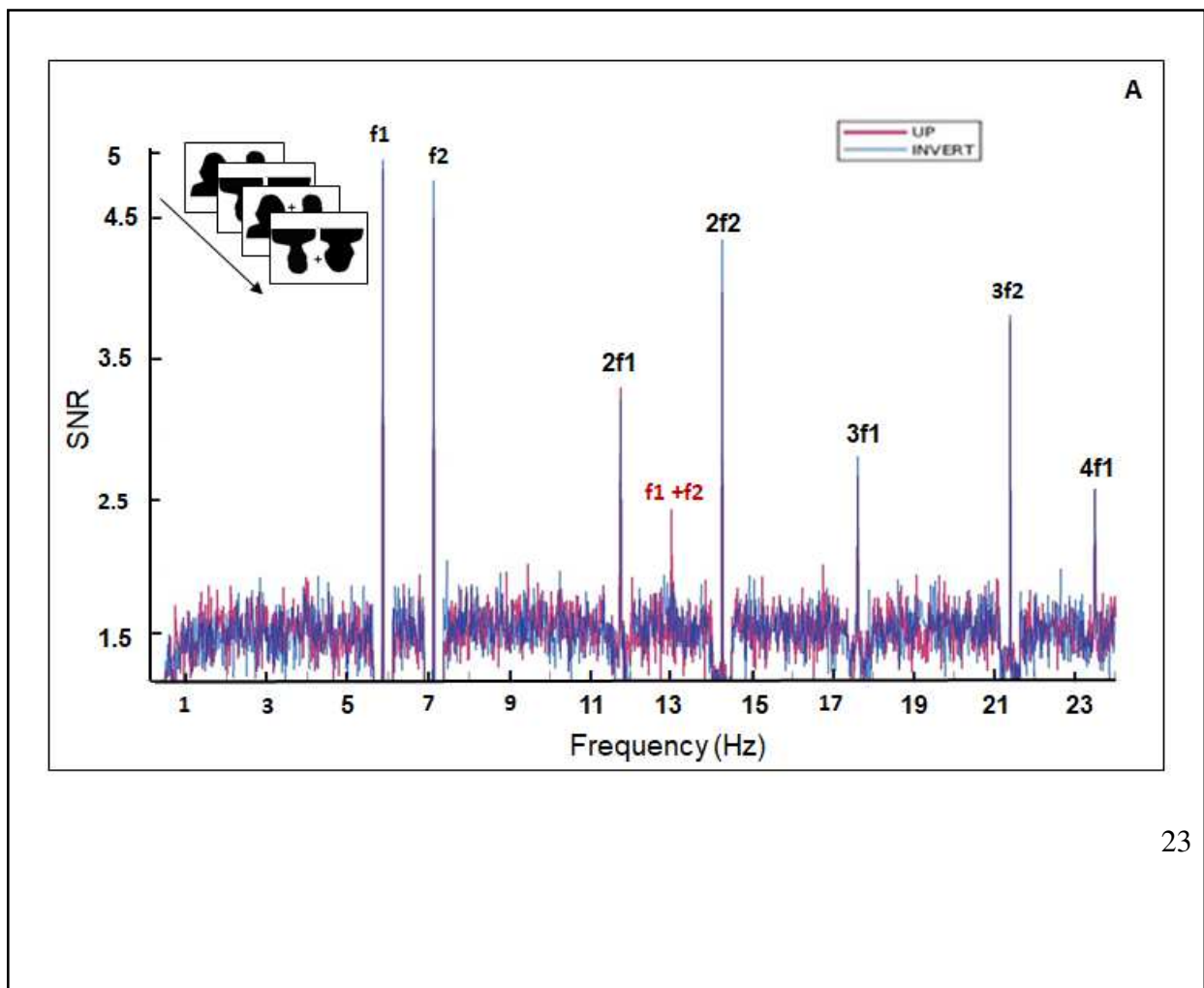


Figure 2. Amplitude Spectrum describing each component frequency contribution (0 – 22 Hz) on average on all electrodes, all participants and all conditions, as well as topographical maps of stimulation frequencies, three harmonics and the second order sum IM component $f1+f2 = 13.02$.



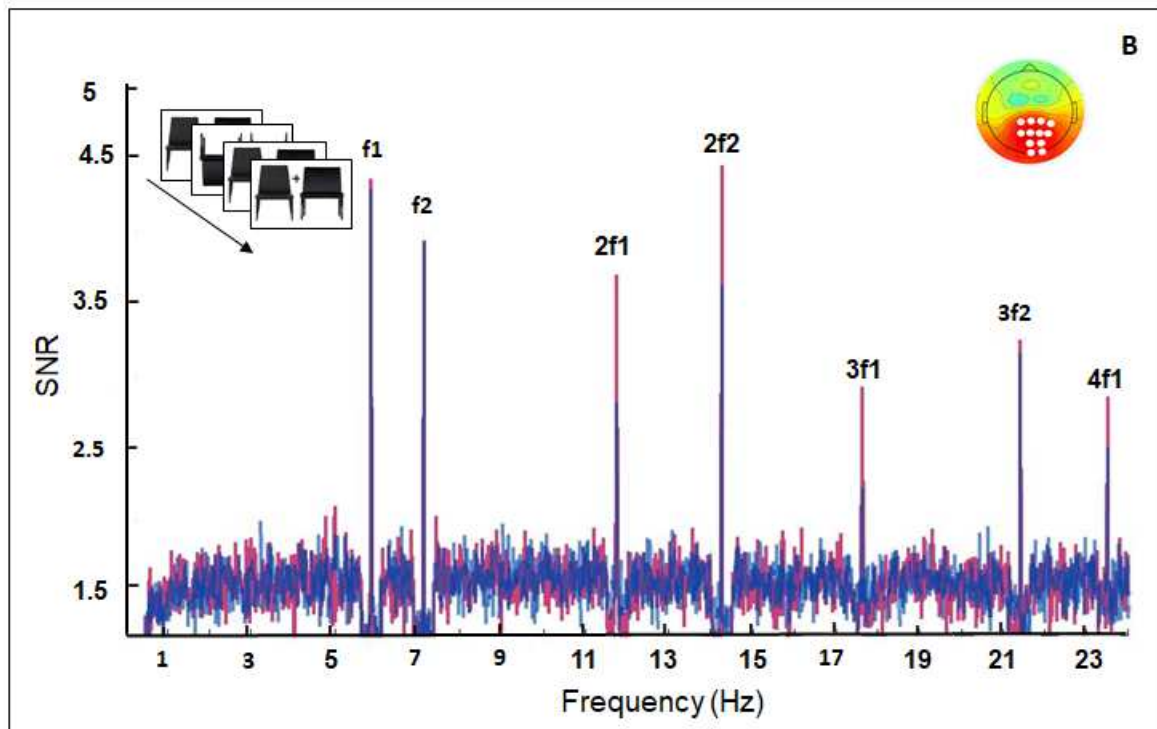


Figure 3. Signal to Noise Ratio (0-23 Hz) averaged across participants and cortical region of analysis (white dots on topographical map) for Humans (A) and Chairs (B) conditions. Up and Inverted trials are represented on both figures, with different colors, in transparency. Clear response appears at 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.

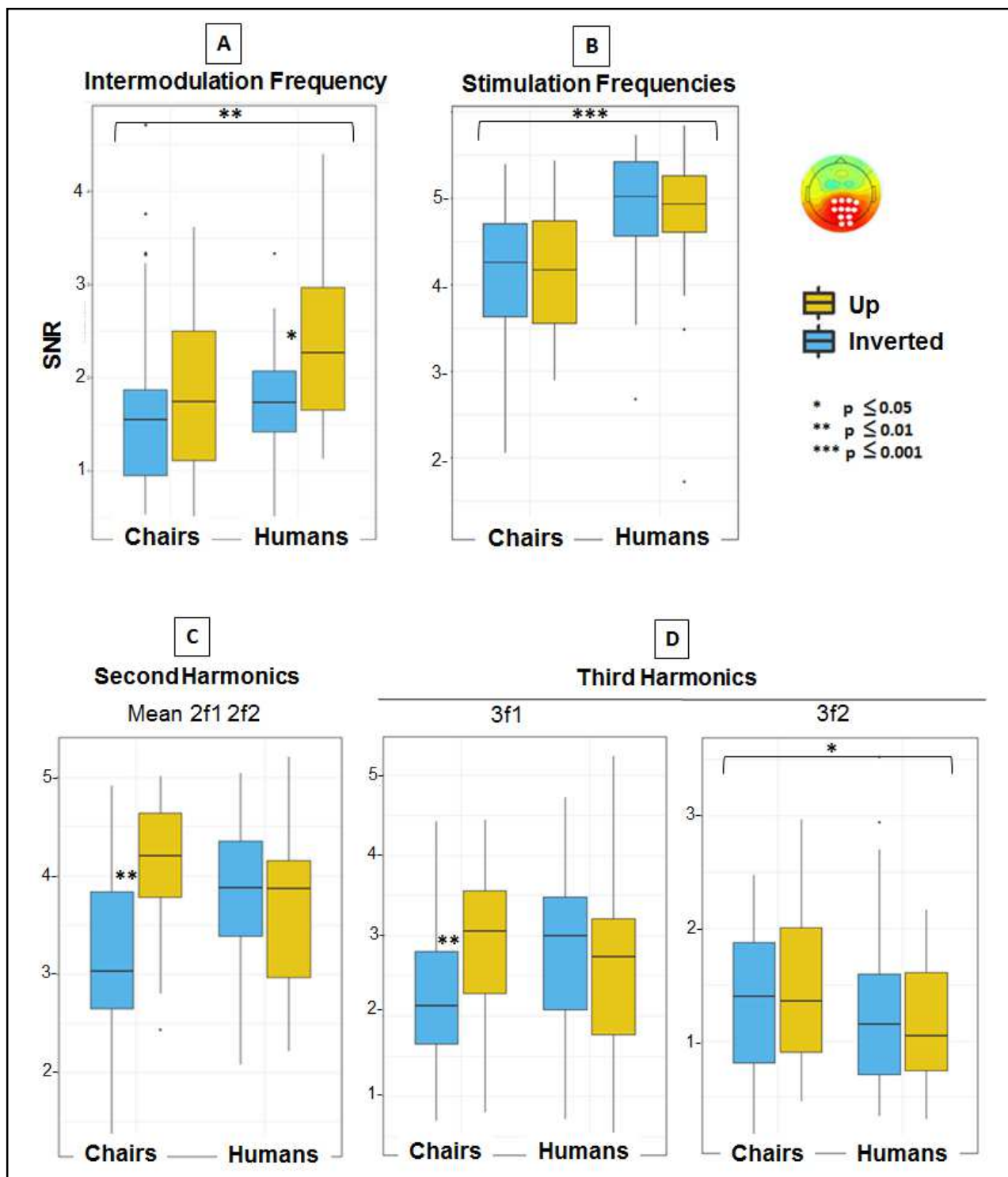


Figure 4. Signal to Noise Ratio averaged on participants and cortical region of analysis (white dots on topographical map) broken down by level of processing and type of object (Humans, Chairs). Dyad parts level of processing: stimulation frequencies and harmonics (panel B: first harmonic, C: second harmonics* and D: third harmonics). Global dyad configuration level: IM frequency (panel A). The figure depicts inversion effect in each condition. Boxplots show median, upper and lower quartiles. Error bars represents SE. Outlying points are represented individually. In Chair condition, we observe simultaneously a significant inversion effect at second (mean 2f1 and 2f2) and third (3f1) harmonic frequencies (pair-element level of processing) and no inversion effect at IM component (whole-pair configuration level of processing). At the contrary, in Human condition, we simultaneously observe a significant inversion effect at the IM component and no inversion effect at the stimulation frequencies.

**Second harmonics, showing the same pattern of result, are depicted on average.*

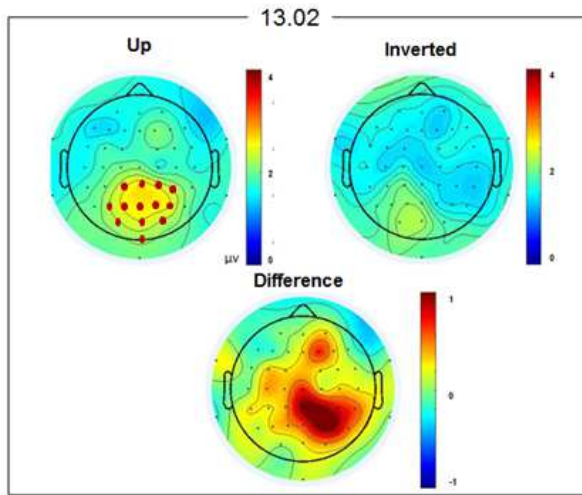


Figure 5. Topographical maps of SNR at intermodulation frequency $f_1 + f_2 = 13.02$ Hz (response to whole dyad configuration (average participants in Humans condition only). The figure depicts the topography of Up (red dots represent electrodes showing significant response at IM frequency) and Inverted conditions and their difference (inversion effect). The highest inversion effect appears over right occipito-parietal regions.