ERP CORRELATES OF THE INTERACTIONS BETWEEN TOP-DOWN AND BOTTOM-UP PROCESSES IN VISUAL OBJECT CATEGORIZATION

by

Szilvia Linnert

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS OF LANCASTER UNIVERSITY FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

LANCASTER UNIVERSITY

September 2017

Declaration

I declare that the thesis is my own work, and has not been submitted in substantially the same form for the award of a higher degree elsewhere.

28th September 2017

Signature

Abstract

Numerous studies have reported category differences between animate and inanimate objects in the early visual ERP components. However, contradictory explanations have been suggested for the underlying processes of such category differences. Both low-level physical differences and higher-level category-specific processes have been shown to modulate the visual ERPs. The present research investigates how physical features (such as amplitude spectrum and spatial frequency) and top-down processes (the categorization task that the participants perform) interact and modulate the visual ERPs. We found ERP correlates of categorical representation for animal and inanimate object categories, as well as early, task-related top-down modulation of the visual ERPs. These results indicate that top-down factors can modulate visual processing both at the level of lower-level physical features and at the level of category representations. The results are discussed in terms of shape- and/or category-selective representations and brain areas in the ventral visual pathway, and they are interpreted within the framework of flexible evidence accumulation processes.

Acknowledgements

The work presented in this thesis was funded by the Faculty of Science and Technology at Lancaster University, and I would like to thank the many participants who took part in my research.

I am very grateful to my supervisor, Prof. Gert Westermann for being supportive, helpful and patient during these years. Thank you to my second supervisor, Prof. Vincent Reid for helping to finalize the study designs and revising the manuscripts. I would also like to thank Dr. Tamás Tompa for first introducing me to this field of research many years ago.

Thank you very much to Eugenio and Sam for all their continuous help and for being there when I needed. Finally, thank you to my family for all their support.

Dedication

For James

Table of Contents

Statement of Authorship

Authorship contributions

PAPER 1: UNDERLYING MECHANISMS OF TASK-RELATED MODULATION OF ANIMAL VS. VEHICLE DIFFERENCES IN THE VISUAL ERPs

Conception and design of study: **Szilvia Linnert**, Gert Westermann, Vincent Reid
Acquisition of data: **Szilvia Linnert**
Analysis and/or interpretation of data: **Szilvia Linnert**
Drafting the manuscript: **Szilvia Linnert**
Revising the manuscript critically for important intellectual content: **Szilvia Linnert**, Gert Westermann
Contribution of principal author: 85%

PAPER 2: VISUAL ERP DIFFERENCES BETWEEN ANIMAL AND VEHICLE CATEGORIES IN A LABEL-IMAGE MATCHING TASK

Conception and design of study: **Szilvia Linnert**, Gert Westermann, Vincent Reid
Acquisition of data: **Szilvia Linnert**
Analysis and/or interpretation of data: **Szilvia Linnert**
Drafting the manuscript: **Szilvia Linnert**
Revising the manuscript critically for important intellectual content: **Szilvia Linnert**, Gert Westermann
Contribution of principal author: 85%

PAPER 3: EARLY ANIMAL-VEHICLE ERP DIFFERENCES REFLECT INTEGRATION OF HIGH AND LOW SPATIAL FREQUENCY INFORMATION

Conception and design of study: **Szilvia Linnert**, Gert Westermann, Vincent Reid
Acquisition of data: **Szilvia Linnert**
Analysis and/or interpretation of data: **Szilvia Linnert**
Drafting the manuscript: **Szilvia Linnert**
Revising the manuscript critically for important intellectual content: **Szilvia Linnert**, Gert Westermann
Contribution of principal author: 85%

PAPER 4: EARLY VISUAL ERPs REFLECT FLEXIBLE MECHANISMS IN INANIMATE OBJECT DISCRIMINATION

Conception and design of study: **Szilvia Linnert**, Gert Westermann, Vincent Reid
Acquisition of data: **Szilvia Linnert**
Analysis and/or interpretation of data: **Szilvia Linnert**
Drafting the manuscript: **Szilvia Linnert**
Revising the manuscript critically for important intellectual content: **Szilvia Linnert**, Gert Westermann, Vincent Reid
Contribution of principal author: 85%

We the undersigned agree with the above stated “proportion of work undertaken” for each of the above manuscripts contributing to this thesis:

Signed: SZILVIA LINNERT Szilvia Linnert

VINCENT REID

GERT WESTERMANN

Epigraph

“Open mind for a different view
And nothing else matters”

James Hetfield & Lars Ulrich

Introduction: Visual category representations and their ERP correlates

Szilvia Linnert

Lancaster University

Introduction: Visual category representations and their ERP correlates

People use categories to process incoming visual information. A category can aggregate very different kinds of objects (e.g., 'Things to take to the beach') but other categories can contain perceptually similar representatives, for instance cats and dogs. Furthermore, the same objects can belong to several categories; for instance, a cat can be categorized as a cat, as an animal, as a mammal, as a four-legged animal etc. The brain somehow has to represent category-specific information, and different processes have to access this knowledge in a flexible manner. Case studies provide evidence that representations can be selectively impaired. For instance, many studies have shown that in visual agnosia the visual recognition of animals, fruits/vegetables, man-made objects or faces can be selectively impaired (Mahon & Caramazza, 2009; Warrington & Shallice, 1984). However, it is important to point out that these patients can still recognize objects by touching them and they can also name these objects and describe them. Patients can also recognize lower-level features, such as lines, edges or simple geometrical shapes. Therefore, this deficit selectively impairs a specific stage of information processing in the visual modality, however, the actual impairment varies between patients, depending on the extent of the injury (e.g. Behrmann & Kimchi, 2003). Research with these patients can provide important information about visual processes and category representations. For instance, several aspects of the two visual systems theory are based on a patient with visual form agnosia (patinet D.F.; see Goodale & Keith Humphrey, 1998; Whitwell, Milner, & Goodale, 2014).

The role of vision is to construct an inner representation of the objects in the outside word. This inner representation is the percept, how we perceive the outside word. Therefore, at the last stage of the visual processing, this percept has to be constructed based on the incoming visual information. According to the two-visual pathway theory, the ventral visual pathway processes information regarding the identity of objects, in other words, it constructs the percept (Goodale & Keith Humphrey, 1998; Goodale & Milner, 1992; Ungerleider & Mishkin, 1982). According to the classical view (Ungerleider, Mishkin, 1982), visual information within this pathway is processed in a feed-forward, hierarchical fashion. The early visual areas process simple features, such as lines and edges. This information is projected to higher-level areas, where more complex features are processed. At the last stage of the visual processing, whole object representations have to be constructed. In addition, category-specific visual impairment in agnosia patients suggests that not only individual objects, but also categories can be represented in the last stages of visual processing.

Object and category representations in the higher-level visual areas

The last stage of the ventral pathway is the inferior temporal cortex (IT) or its human counterpart, the lateral occipital complex (LOC: Grill-Spector, Kourtzi, & Kanwisher, 2001; Malach et al., 1995). Here, complex objects are processed and represented. Neurons in the IT respond to complex stimulus features by pooling information from the lower areas (for review see Tompa & Sáry, 2010). IT neurons in monkeys are shape-selective, they tend to respond to some shapes while activation is decreased for others. Therefore, each neuron has a preferred shape, which can be the basis of object representation. The strongest IT responses have been found for colorful, complex stimuli. In addition, several cells seem to respond to objects that have biological relevance for the monkeys, such as hands and faces. The responses of IT neurons have also been shown to be viewpoint independent for the preferred shape, however only to a limited extent (for review see Tompa & Sáry, 2010).

It seems that beyond shape selectivity, category information is also coded in the IT. Kiani, Esteky, Mirpour, and Tanaka (2007) showed that single neurons and neuron populations responded differently to category exemplars in the monkey IT. In this study, activity was recorded extracellularly from more than 600 neurons while monkeys performed a passive fixation task. Each neuron was tested with around 1000 stimuli. Images contained a big range of colourful animate and inanimate objects. The activities of a single neuron increased or decreased in response to different objects. Activity patterns measured across neurons were used to calculate similarity between the presented objects. Object belong to closer categories in our intuitive hierarchical category structure elicited more similar response patterns. Authors also used multidimensional scaling and cluster analysis in order to examine the distance between neural response patterns for each presented object. They found hierarchical category representations in the activity pattern of neurons. The highest-level categories were animate and inanimate categories. The ‘animate’ object category was divided into subcategories of bodies, hands and faces. The ‘body’ category was divided into two subcategories: the first grouped humans, birds and four-legged animals, while the second contained fishes, insects and lower animals. Faces were split into primate and non-primate categories, and the primate face category was further divided into human and monkey faces.

Brain imaging studies in humans have also shown object specificity in the LOC. Several studies have reported specialized visual areas that show greater activation for one object category compared to another. Specific areas have been identified for faces (face-selective area: Kanwisher, McDermott, & Chun, 1997) and animals (Chao, Haxby, & Martin, 1999) as well as for buildings (place-selective area: Aguirre, Zarahn, &

D’Esposito, 1998), chairs (Ishai, Ungerleider, Martin, Schouten, & Haxby, 1999) and tools (Chao et al., 1999). It has been suggested, that such activity differences reflect distinct categorical representations (Kanwisher et al., 1997). However, most of these categories are basic-level categories. Therefore, exemplars within a category usually share similar shapes, whereas objects belonging to different categories have very distinct shapes. It is possible, therefore, that only individual objects are represented in the LOC instead of object categories, and that these individual object representations are organized by perceptual similarity. With this view, category-specific areas do not reflect real category representations; instead, category-specificity arises as an artifact from experiments that use perceptually very different categories (Kriegeskorte et al., 2008). More recent fMRI studies have found evidence both for distinct, category-based and for continuous, similarity-based representations in the LOC (Kriegeskorte et al., 2008; Mur et al., 2013). Distinct category-based activity patterns were found for animate and inanimate categories. In addition, face and body sensitive activity clusters were found within the animate category, reflecting hierarchical category representations (Kriegeskorte et al., 2008; Mur et al., 2013). Results also indicated that, beside the distinct categorical representation, objects within these categories were represented in a continuous, similarity-based manner (Kriegeskorte et al., 2008).

Therefore, these results suggest that, beside shape-related representations, some categories are represented in the higher-level visual areas, both in monkeys and humans. According to the classical view, these shape and category representations are activated in a feed-forward manner, as the incoming visual input is analyzed (e.g. Ungerleider & Mishkin, 1982). According this view, category specificity arises on top of low-level visual feature processing.

Nevertheless, it is possible that some low-level visual features already contain category-specific information. For instance, the amplitude spectrum (AS) of an image, as a low-level characteristic, has been suggested to provide category specific information (Torralba & Oliva, 2003). The AS of an image refers to its energy at different spatial frequencies and it describes global spatial characteristics. Looking at only the AS content of an image, the image is not recognizable. Nevertheless, it has been shown in computational models that the AS content alone provided 80% accuracy of detecting animals, people and vehicles in natural and man-made scenes (Torralba & Oliva, 2003). In an fMRI study, Andrews, Clarke, Pell, and Hartley (2010) investigated whether category-specific responses in the face- and house-selective visual areas can be explained by sensitivity to the AS information. Intact and phase-scrambled faces and houses were used as stimuli. The phase spectrum of an object provides information about the structure of the objects, such as the shape or configuration of local features. Therefore, phase scrambling preserves global, low-level properties, such as the AS information, but it destructs the recognizable shape information. Therefore, these stimuli contain the AS information of the original images, but they do not contain recognizable objects. Results indicated that AS information did contribute to face- and house-selective activation. However, phase-scrambled images activated the category-selective regions to a much lesser degree than intact ones. Beside the category-specific areas, the results also showed selectivity to AS information in more medial-occipital regions, which process lower-level visual features. This suggests, that category-specific responses in higher-level areas arise from the processing of category-specific low-level features in the lower-level visual areas (Andrews et al., 2010). Therefore, category-specificity does not solely rely on higher-level representations; low-level, category-specific features can also contribute to it.

The nature of visual category representations

The above-mentioned studies investigated objects and category representations in the visual areas. Nevertheless, other studies have suggested the role of prefrontal and temporal cortex in category learning. For instance, in a study by Freedman, Riesenhuber, Poggio, and Miller (2001) monkeys were trained to categorize morphed cat-like and dog-like stimuli. The category boundary between these animals was arbitrary in that the perceptual difference between the two category exemplars closest to the category boundary was the same as the difference between within-category members. Nevertheless, the activity of prefrontal neurons differentiated categories, as activity was similar for in-group exemplars irrespectively of the shape differences (Freedman, Riesenhuber, Poggio & Miller, 2001).

Jiang et al. (2007) tested category learning mechanisms in humans by using fMRI. Participants were trained to categorize randomly generated morphed cars. Training in car categorization led to increased selectivity in the human LOC. However, this selectivity was sensitive to small changes in shape; novel exemplars did not evoke the same activation pattern. This result suggested that LOC activation reflected only shape selectivity and not category membership per se. In contrast, prefrontal areas showed category selectivity that was not driven by perceptual differences as category-selective activation was found also for novel exemplars. Similarly, van der Linden, van Turennout, and Indefrey (2010) found only perceptual learning effects in the visual areas, but no generalization for the novel category exemplars when participants were trained to categorize artificially created birds into arbitrary categories. Interestingly, a category effect, irrespective of perceptual differences, was found in the superior temporal cortex, but not in the prefrontal cortex. A possible explanation for this difference to the results of

Jiang et al. (2007) is the difference in training. While in the study of Jiang et al. (2007) a perceptual training was performed, van der Linden et al. (2010) labeled the categories.

If category learning mechanisms affect prefrontal and temporal areas, but not the visual areas, this suggests that the visual areas contain more hard-wired, less flexible representations. In other words, objects, and some categories are represented in the visual areas. However, novel, arbitrary categories are formed in different parts of the brain. Indeed, in a recent fMRI study, Bugatus, Weiner, & Grill-Spector (2017) reported results confirming this theory. Images of faces, bodies, cars, houses and letter-strings were presented while participants performed different tasks. Participants either had to press a button if a noisy stimulus appeared on the screen, or they performed a 2-back task. In the third, selective attention task, superimposed images of two categories were presented to the participants, and a cue indicated to which category did they have to attend. The distributed neural activity patterns across prefrontal and visual areas for each category and task were examined using multivoxel pattern analysis. Activation in the category-specific visual areas reflected similar patterns for each category, irrespectively of the task, therefore, representations in this visual area seem to be driven by categories. In contrast, in the prefrontal cortex, category representations differed between the tasks therefore, category-selective activation patterns were driven by the task, Therefore, visual areas seem to contain stable visual representation, while flexible representations in the prefrontal cortex enable task-specific, arbitrary categorization (Bugatus et al., 2017). Nevertheless, selective attention, as a top-down factor, affected the activation of category representations also in the visual areas. According to the authors, attentional effects in the visual area were found in studies when the task required restricted processing of the stimulus because of visual competition (see e.g. selective attention task in Bugatus at al., 2017, as described above). Bugatus et al. (2017) therefore suggested that task-dependent category membership and selective attention likely rely on different mechanisms. Selective attention restricts the processing to a specific (attended) feature and modulates the activation of the stable category representations in the visual areas. Different task engagements, on the other hand, modulated the boundaries of category representations in the prefrontal cortex, but not in the visual areas (Bugatus et al., 2017). Overall, attention-related processes in the visual areas indicate that top-down processes can interact with the bottom-up information processes, providing a more flexible information flow. However, top-down processes seem to only activate category-specific areas, but do not change category boundaries. The following section will provide evidence of such top-down activation.

Top-down activation of category specific areas

The classical view of object categorization, that visual information is processed in a serial fashion, is changing (Tompa & Sáry, 2010). Top-down theories emphasize the role of the task as a top-down factor on visual processes. For instance, according to Gilbert and Sigman (2007) the degree of top-down influence is dependent on the task that the visual system has to solve. Gilbert and Sigman (2007) assume that the rapidly changing environment requires flexible information processing. Flexibility can be achieved if the brain areas do not have a fixed role and are able to execute different processes depending on the context. Therefore, the current processing demands always depend on the resonance between the incoming sensory input and task requirements (Gilbert & Sigman, 2007). These theories are in line with the assumption of stable category representations in the higher-level visual areas suggested by Bugatus et al. (2017). Flexible processing of the visual input does not necessarily mean that top-down processes have to modulate the boundaries of category representations. Rather, top-down attention can help to select relevant information more efficiently and quickly by activating specific representations and/or processes according to the task demands.

Top-down theories on visual information processing are supported by several lines of evidence. Task-related activation has been reported at the level of IT neurons in monkeys. For instance, firing rates of IT cells depend on attentional factors and on whether monkeys perform a categorization or discrimination task (for review see Tompa & Sáry, 2010).

In humans, one line of evidence for top-down activation of shape and category selective areas is studies on mental imagery. It has been shown that mental imagery can be intact in patients with visual form agnosia (Rizzi, Piras, & Marangolo, 2010; Servos & Goodale, 1995). For instance, Rizzi et al. (2010) presented a case study showing how mental imagery as a top-down factor can influence bottom-up visual information processing. Their patient had a selective deficit to visually recognize objects, however, her semantic knowledge and mental imagery were intact. She could recognize local parts of an object but could not generate a global percept and therefore could not recognize the whole objects. At the same time, when she was told to point out a specific object (e.g. the bottle) among other objects, she was able to do that. She explained how she did this by saying she imagined the bottle, and she was looking for something similar on the image (Rizzi et al., 2010). This case suggests that visual imagery based on object names (a top-down process) supports bottom-up information processing.

fMRI studies have also shown that mental imagery activates the same category-specific visual areas in the human LOC that are activated by actually looking at the objects. For instance, O'Craven and Kanwisher (2000) found activation in face-selective and place-selective areas of the visual cortex when participants were imagining faces and buildings. The activation, however, was less pronounced for imagery than during actual perception. Reddy, Tsuchiya, and Serre (2010) reported similar results for categories of faces, houses, tools and fruits. In addition, category information could be decoded from the fMRI activity pattern during mental imagery. Using multivariate analysis, classifiers trained on activity during actual image presentation provided similar results when they were tested on activity during mental imagery, and vice versa (Reddy et al., 2010). Therefore, these studies suggest that top-down projections can activate category-specific representations of the visual cortex even in the absence of incoming sensory information.

Shape and category selective areas can be also activated in expectation of a stimulus. Such preparatory activity has been reported in the visual cortex when anticipatory cues predicted the subsequent stimulus (for review see Battistoni, Stein, & Peelen, 2017). In a study by Stokes, Thompson, Nobre, and Duncan (2009), participants had to selectively attend to simple shapes. An auditory cue was given to inform participants of whether they had to attend to the letter ‘O’ or the letter ‘X’. Selective preparatory activity was found in the LOC area for the different shapes. Preparatory activity has also been shown to affect specific, category-selective visual areas. Esterman and Yantis (2010) found such activity in face- and place-selective visual areas following visual images as anticipatory cues. Peelen and Kastner (2011) reported similar result when symbolic visual cues were given to participants in preparation to detect either cars or humans on the presented image. In addition, this activation also facilitated participants’ target detection performance. Preparatory activity was also found in the early visual cortex (V1), suggesting anticipation of low-level features, such as line-orientation. However, preparatory activity in the V1 worsened participants’ performance (Peelen & Kastner, 2011). Finally, preparatory activity was also found when written names (HOUSE or FACE) were presented as anticipatory cues before images of houses and faces (Puri, Wojciulik, & Ranganath, 2009).

Most of the above mentioned studies used symbolic cues in order to elicit preparatory activity. The use of symbolic cues requires participants to actively interpret the cue. Therefore, attentional influence has to be based on top-down processes and not on involuntary bottom-up processes (Battistoni et al., 2017). Moreover, a transcranial magnetic stimulation (TMS) study has provided more direct evidence that preparatory activity in a specific brain area directly affects the processing of the subsequent image (Reeder, Perini, & Peelen, 2015). Participants were instructed to detect car or people in complex scenes, similar to the study by Peelen and Kastner (2011). However, Reeder et al. (2015) applied TMS before image onset over the category specific LOC areas where preparatory activity has been found by Peelen and Kastner (2011). Reeder et al. (2015) found impaired detection of cars or people if TMS was applied over the LOC area compared to detection of the same categories if TMS was applied over the vertex. Moreover, performance was impaired only when participants had to detect general categories (cars, people). When participants had to detect the same person or car on each image – a task that enables detection based on lower level features – TMS over the LOC did not affect the performance.

The results of Puri et al. (2009) also suggest that preparatory activity directly influences the processing of the incoming information. In that study, house and face stimuli were preceded by a written word (HOUSE or FACE). On 70% of the trials, the word and image matched, however, on the remaining 30% they did not. Beside the preparatory activity in the category-specific areas, increased activity was also found during the processing of the visual stimuli if the anticipatory cue precisely predicted the subsequent visual stimulus. This increased activity was not present when the word was incorrect.

Overall, these studies provided evidence that higher-level visual areas contain shape and category specific representations. These representations can be flexibly activated either by top-down or bottom-up processes. Furthermore, top-down processes modulate bottom-up information processing. Again, it is important to point out that top-down factors activate, object or category representations, but do not modulate category boundaries (Bugatus et al., 2017). Therefore, top-down and bottom-up processes interact at the level of visual representations. Category-specific visual areas have been suggested to accumulate and integrate evidence for decision-making (Tremel & Wheeler, 2015). A decision can be made if the accumulated evidence reaches a threshold. Nevertheless, it is possible that the top-down activation of a category-specific area modulates this threshold. In other words, the expectation of an object makes it easier to recognize it. This explanation is in line with the previously presented results of Bugatus et al., (2017), suggesting stable visual category representations, that can be affected by selective attention.

Content-based versus spatial attention

In a recent review, Battistoni et al. (2017) compared the effects of spatial and non-spatial attention on visual processes. They concluded that top-down attention could not only modulate spatial attributes but also the processing of non-spatial attributes, such as low-level features or even object categories. Such non-spatial attention was named content-based attention by the authors (Battistoni et al., 2017). The studies described in the previous section suggest that content based attention selectively pre-activate visual areas before the actual stimulus appears. Thus, it influences bottom-up information processing based on the current goals.

The underlying mechanisms of content-based attention can be investigated at the level of neuronal activity. IT neurons are shape-selective, as they tend to respond to some shapes while activation is decreased for others. Thus, each neuron has a preferred shape (Tompa & Sáry, 2010). Chelazzi, Duncan, Miller, and Desimone (1998) measured activities of single neurons in monkeys in response to a cue that they learned to associate with an object. The monkey's task was to detect the cued object in an array of objects by fixating on it. More than half of the neurons showed sustained activity following the cue if it signalled the preferred shape of the neuron. Cues that were irrelevant to the task (as monkeys only had to maintain their gaze at the central fixation) did not elicit shape-specific activity. Therefore, shape-selective neurons in the IT increase their activity in preparation to detect a behaviourally relevant stimulus.

Content-based attention and spatial attention seem to modulate activity in distinct ways at the level of neuron populations (Ling, Liu, & Carrasco, 2009; Martinez-Trujillo & Treue, 2004). Results suggest that spatial attention only increases the level of activity of a neuron population selective to the attended location. Content-based attention, on the other hand, not only increases neuronal activity, it also fine-tunes the response of the neuron population. This means that content-based attention increases the response of neurons that are sensitive to the attended feature or shape, and decreases the response of neurons that are sensitive to other features or shapes.

Therefore, content-based top-down attention and spatial attention are similar processes, however, they seem to rely on different underlying mechanisms at the level of neuron populations.

Electroencephalogram (EEG) and visual event-related potentials ERPs

Electroencephalogram (EEG) – the recording of the electrical activity of the brain – is an excellent tool to investigate visual processes. Visual information processing is very fast, but the high temporal resolution of the EEG can help us differentiate its components. Furthermore, EEG is non-invasive and is widely applied in human research, in both adult and developmental populations. The EEG activity reflects the average voltage fluctuation of large neural ensembles over time.

One way to analyze EEG is to compute event-related potentials (ERPs). ERPs are averaged signal of EEG segments that are time-locked to a specific event, such as the appearance of a stimulus. By averaging together many segments linked to the appearance of a stimulus type, EEG signal reflecting the processing of that stimulus will be amplified and show up as deflections on the averaged waveform. These positive and negative deflections are called ERP components. In contrast, EEG signals that are not linked to the processing of that stimulus cancel out through averaging. Research using ERP has identified a range of components that are linked to different mental processes.

Visual ERPs are measured over the posterior brain areas and give information about visual processing in the first 300 ms after stimulus onset. The visual ERPs are the P1, N1 (the face specific N1 is called N170), P2, and N2 components (Luck, 2005). The visual P1 is a positive-going component with peak latency between 100 - 130 ms after the stimulus onset. The P1 is sensitive to low-level stimulus parameters (contrast, size, luminance and amplitude spectrum). The visual N1 is a negative-going component with peak latency between 150 - 200 (or sometimes rather between 140 - 180) ms after the onset of the stimulus. This component is sensitive to the task and seems to reflect some sort of discriminative processes (for review see Luck, 2005). Somewhat less is known about the visual P2 (positive component, peaking between 180 - 300 ms after stimulus onset) and N2 (negative component, peaking between 200 - 350 ms after stimulus onset) components. These components are not necessarily distinguishable on the ERP waveforms. They are linked to more extensive visual analysis (e.g. Schendan & Kutas, 2007). It has been suggested that competing or insufficient information in the higher visual areas get reinforcement by re-activating the early visual areas by feedback projections (Di Lollo, Enns, & Rensink, 2000) and that the P2 component reflects this re-activation process (Kotsoni, Csibra, Mareschal, & Johnson, 2007). This explanation is in line with studies finding larger P2 for unrecognizable, noisy stimuli (e.g. Rousselet, Husk, Bennett, & Sekuler, 2008).

Regarding categorization and category differences, several studies have reported N1 amplitude differences between faces and man-made objects (for review see Rossion & Jacques, 2008) or between animate and inanimate categories (e.g. Antal, Kéri, Kovács, Janka, & Benedek, 2000; VanRullen & Thorpe, 2001). Many different interpretations have been suggested in order to explain these differences. Different studies measure these category differences at different electrode locations, using different tasks and different object categories. Still, category differences in the N1 likely reflect some sort of discrimination processes (Luck, 2005). However, the phrase, ‘discrimination process’ can include visual, category-specific processes, as well as decision-making or response-related processes. The next sections of this introduction will summarize studies in more details regarding this question.

Category- and face-selective processes are usually measured over the temporal area, on electrodes T5 and T6, according to the 10 - 20 EEG nomenclature. However, visual ERPs are also analyzed over the occipital areas at electrodes O1, O2 and Oz. The spatial resolution of the EEG/ERP is poor; a signal measured on a single electrode can originate from different parts of the brain. Therefore, the location of an electrode where a specific signal is measured does not necessarily reflect the brain activity of the area under that electrode. Different source localization methods are used in order to find the brain generators for a given EEG/ERP signal. Visual ERPs seem to be generated in the posterior brain areas, where visual information is processed (Luck, 2005). In previous decades, EEG was measured only at a few electrode locations, and results often reported the average ERPs on occipital and temporal electrodes. Nowadays, however, with high-density EEG systems, visual ERPs can be measured at several electrode locations and we can compare the signal between different electrodes. Still, we cannot draw conclusion regarding the exact brain area where these signals originate. Nevertheless, brain imaging studies in humans and single-cell recordings from monkeys provide a substantial amount of information about visual processing in different brain areas. Therefore, interpreting ERP results in a framework provided by different brain imaging methods can help to relate ERP differences to possible underlying mechanisms. The aim of this thesis is to provide this interpretation for category-selective ERP differences.

Category differences in the visual ERPs

Animate-inanimate ERP differences were first reported by Thorpe, Fize, & Marlot (1996). Participants saw images of natural scenes appearing on the screen for 20 ms and they had to respond only if the picture contained an animal. The scenes contained a broad variety of animals (mammals, reptiles, fish etc.), and the non-animal scenes included natural and man-made environments as well as flowers and fruits. Over the frontal brain areas from around 150 ms after stimulus onset, the ERP waves for the no-go (non-animal) trials were more negative than for trials containing animal images (go trials). According to the authors' interpretation, this ERP difference reflected decision-making processes occurring after the completion of visual processing. Thorpe et al. (1996) suggested that the larger negativity for no-go trials might reflect response inhibition. Furthermore, they assumed that such fast decision-making processes should be based solely on feed-forward projections.

Subsequent studies have also found animal vs. non-animal ERP differences at posterior electrode locations, over visual areas (Antal et al., 2001, 2000; Bacon-Macé, Macé, Fabre-Thorpe, & Thorpe, 2005; Codispoti, Ferrari, Junghöfer, & Schupp, 2006; Delorme, Rousselet, Macé, & Fabre-Thorpe, 2004; Johnson & Olshausen, 2003; Kiefer, 2001; Macé, Thorpe, & Fabre-Thorpe, 2005; Proverbio, Del Zotto, & Zani, 2007; Rousselet, Fabre-Thorpe, & Thorpe, 2002; VanRullen & Thorpe, 2001; Zhu, Drewes, Peatfield, & Melcher, 2016). For instance, Antal et al. (2000) investigated the animal-vehicle differences in the ERPs in an animal vs. non-animal two-choice categorization task. Beside the frontal area, they found category differences also over the posterior brain areas with non-animals eliciting more negative ERPs compared to animals between 175 - 205 ms after stimulus onset. Furthermore, they also compared ERP differences for low and high spatial frequency gratings. The spatial frequency gratings elicited only early ERP differences (around 138 - 150 ms after stimulus onset) over the frontal areas. These results indicate two important interpretations regarding the animal vs. non-animal ERP differences over the posterior brain areas. First, these differences are not due to the unbalanced motor responses. In contrast to the go/no-go task, where participants had to press or release a button only for the target (usually animal) category, Antal et al. (2000) used a two-choice categorization task, where participants had to press one of two buttons for each stimulus type. Second, because of the different topography and time window of the category and spatial frequency grating differences, the authors suggested that the posterior category difference reflected higher-level cognitive processes that were e associated with the categorization of complex visual stimuli (Antal et al., 2000). However, this interpretation does not clarify whether categorization reflects visual processes or decision-making processes.

VanRullen and Thorpe (2001) attempted to separate ERP differences reflecting perceptual and decision-making processes. In that study, participants had to perform two go/no-go tasks where either animals or vehicles were the targets. Since vehicles were distractors when animals were targets, and the vice versa, categories could be compared unrelated to the task. Therefore, the authors compared ERPs for animals and vehicles, regardless of whether they were targets or not, and also compared target vs. non-target differences, regardless of their category. Thus, any activity difference between animals and vehicles should reflect perceptual category differences, and differences between targets and non-targets should reflect decision-making processes. The authors found differential activity between the two categories starting from 75 ms after stimulus onset, and they found task-related ERP differences from 150 ms after the stimulus onset. They concluded that the early animal-vehicle activity difference was based on the visual properties of animals and vehicles, while the later activity difference reflected task-related, decision-making processes. Nevertheless, these interpretations were made based on the ERP differences over different areas. The early category difference was most prominent at parietal electrodes, whereas the late target/non-target difference was the largest over the frontal areas. In addition, the animal-vehicle difference at the occipital and frontal electrodes did not disappear in later time-windows with the emergence of task-related differences. Rather, it was maintained in later time windows and it only changed direction at around 160 ms (see Figure 3a in VanRullen & Thorpe, 2001, p. 456). Therefore, the later components do not solely reflect difference between target and non-target trials. Nevertheless, these results do indicate both perceptual and task-specific processes, but their interpretation still leaves open questions. First, in a go/no-go task, decision-making consists of deciding whether the image contains a target object/category or not. In contrast, in a two-choice categorization task, a decision consists in deciding, whether the image (or object) belongs to one category or another. Therefore, a go/no-go task requires the activation of one category representation, whereas in a two-choice task different images/objects can activate different category representations (e.g. animals and vehicles). Second, what kind of visual properties can appear in the visual ERPs? Are these low-level properties or higher-level processes? Do they reflect category-specific processes, or just systematic differences between stimulus sets? The studies presented in the next sections are focused on questions about the role of the categorization method and low-level stimulus properties.

The effect of the categorization task

VanRullen and Thorpe (2001) suggested that the differences in the time-window of the N1 component reflect decision-making processes. In that study, participants had to perform two go/no-go tasks where either animals or vehicles were the targets. In many studies, however, the animal is the only target category in the go/no-go task (e.g. Rousselet et al., 2002; Thorpe et al., 1996). Therefore, differences in the ERPs can reflect either animate-inanimate differences or target vs. non-target differences. As described above, other studies, on the other hand, used a two-choice, animate vs. inanimate task (Antal et al., 2001, 2000). In this task, the decision-making also requires category-specificity. In order to investigate the role of the categorization method, a few studies have compared category differences in different categorization tasks (Delorme et al., 2004; Johnson & Olshausen, 2003; Kincses, Chadaide, Varga, Antal, & Paulus, 2006).

Kincses et al. (2006) compared an animal/non-animal two choice categorization task, an animal/non-animal go/no-go task, and an animal counting task. ERP difference for animal vs. non-animal categories were found earlier in time for the go/no-go task than in the two-choice task. No animal/non-animal difference was found in the animal counting task. This suggests that decision-making processes do play a role in eliciting category differences. Kincses et al. (2006) suggested, that the temporal differences are due to top-down effects. While in the go/no-go task only one category template is needed to perform the task, the two-choice task requires two templates. These templates might have integrated with the bottom-up information processing differently, causing dissimilarities in the ERP waveforms (Kincses et al., 2006).

The role of template as a top-down factor has been also investigated in a single-photograph recognition task in comparison with an animal/non-animal go/no-go task (Delorme et al., 2004). In the single target-photograph recognition task participants had to memorize a target photograph at the beginning of each testing block During the testing phase they had to respond as quickly as possible when they saw the memorized image. This task was assumed to maximize the use of top-down processes, since the recognition can rely on target-specific low-level features, such as edges and contours, that are stored in memory and, as a top-down template, can enable fast recognition. Indeed, the difference between target and non-target waveforms appeared 30 - 40 ms earlier in the photograph recognition task (starting at 135 ms after stimulus onset) compared to the go/no-go task (169 ms after stimulus onset) over the occipital area. Source localization did not show any difference between the two tasks; the differential ERP waveforms were localized in the temporal cortex. Delorme et al. (2004) concluded that the stimuli were processed along the same visual network regardless of the task; however, the speed of the bottom-up processing depended on the top-down influences.

Johnson and Olshausen (2003) compared a two-choice, a go/no-go and a cued target tasks. In the two-choice and go/no-go task participants had to make an animal vs. non-animal decision. In the cued target task, first a category name appeared on the screen and participants had to decide whether the following image belong to that category or not. The category name could be either at superordinate level (‘animal’, ‘furniture’) or at basic level (‘dog’, shoe). In the two-choice and go/no-go task, target vs. non-target amplitude differences were found from around 135 ms after stimulus onset. In the cued target task, target vs. non-target differences appeared later, between around 170 - 210 ms after stimulus onset, depending on the level of the category cue. It has been suggested, that the early differences between 135 -150 ms are due to low-level differences between images, whereas recognition related ERP differences arise later, between 150 - 300 ms (Johnson & Olshausen, 2003). These differences, however, were measured over the frontal areas. Moreover, the authors compared target and non-target ERP waves. In the two-choice and go/no-go task, the target vs. non-target difference was the same as that for the animal and non-animal categories. In the cued target task, however, the cue could belong to many categories. Therefore, in that task, the target vs. non-target differences reflected a congruency decision between the presented word and the following image, instead of animal vs. non-animal differences.

Overall these studies suggest that the categorization task modulates the ERP differences. However, category specific processes and decision-making processes are confounded in these studies. Moreover, these differences are measured at different electrode locations. Delorme et al. (2004) conducted their analysis on both occipital and frontal electrodes, while Johnson & Olshausen (2003) reported results only for the frontal electrodes. Therefore, they do not provide a clear picture about task-related modulations.

The effect of low-level stimulus properties on the visual ERPs

Low-level differences, such as interstimulus perceptual variance (ISVP, Thierry, Martin, Downing, & Pegna, 2007b), and amplitude spectrum information that describes the global spatial characteristics of an image (VanRullen, 2011) have been suggested to play important roles for category differences in the visual ERPs. ISVP refers to the notion that some of the frequently used stimulus categories might be less variable perceptually than other categories. Thierry et al. (2007) pointed out that in studies where faces and cars are used as stimuli, faces are usually presented from full front view, whereas cars are presented from various viewpoints. They suggested that this perceptual variance between the presented faces and cars could contribute to the face-car ERP differences. The suggestion of an ISVP effect on the N1 component is based on the modulation of the face-specific N170 component. Recognition of faces elicits a more negative N170 component than man-made objects or natural scenes (e.g. Rossion and Jacques, 2008). Furthermore, the N170 amplitude is enhanced and delayed for inverted faces related to upright ones (face-inversion effect). To control for this potential confound (Thierry et al., 2007) designed a study in which faces and cars were presented either with high or low ISPV. The high ISVP stimuli differed in size, eccentricity and orientation, whereas low ISPV stimuli were matched for these characteristics. They found face vs. car differences only in the P1 component. In contrast, the N170 amplitudes differed between stimuli with high and low ISPV, irrespectively of the stimulus category. According to the authors, these results question whether the N170 is indeed specific to faces, or the previously reported effects were only due to uncontrolled low-level differences between stimulus sets. However, this interpretation generated a big debate in the literature (Bentin et al., 2007; Rossion & Jacques, 2008; Thierry, Martin, Downing, & Pegna, 2007a). Moreover, Thierry et al. (2007) did find face vs. car P1 amplitude differences and suggested that these differences emerge “from low perceptual invariants” (Thierry et al., 2007, pp. 508), such as outline, contrast or complexity. Such differences do not necessarily reflect differences between stimulus sets; they can as well reflect category-specific low-level features. Furthermore, even if ISVP affects ERP differences between face and non-face objects, it is less likely to play a role in animate-inanimate ERP differences. Animate stimuli usually consist of a large variety of animals, including mammals, bird, reptiles, fish as well as insects, often presented in natural scenes. For inanimate stimuli, some studies use natural and urban scenes without animals, other studies use a variety of man-made objects. It has been argued that such variability rules out the low-level differences that are not category specific (see VanRullen, 2011).

The amplitude spectrum (AS) of an image can also be a possible factor contributing to low-level, category-specific differences in the ERPs, as was also shown in the fMRI study by Andrews et al. (2010) discussed above. The AS contribution to category-specific ERP differences has been investigated in several studies (Bieniek, Pernet, & Rousselet, 2012; Nemeth, Kovacs, Vakli, Kovacs, & Zimmer, 2014; Rousselet, Husk, Bennett, & Sekuler, 2008; Rousselet, Pernet, Bennett, & Sekuler, 2008). However, results have been contradictory. Some studies found AS-related ERP differences (Nemeth et al., 2014; Bruno Rossion & Caharel, 2011), but others did not (Bieniek et al., 2012; Rousselet et al., 2008). For instance, in their study, Bieniek et al., (2012) used faces and houses as stimuli, with the AS either kept intact, swapped with the AS of the other category (e.g. the image contained a visible face but its AS information corresponded to a house), or it contained the averaged AS of faces and houses. In addition, they also used phase-scrambled, unrecognizable images as stimuli. These images were produced by scrambling the phase information of the original face and house stimuli. Phase-randomized stimuli do not contain any recognizable object; however, they contain the AS information of the original image. Phase scrambled stimuli contained either face- or house-specific AS information, or the average of the two. Differences between recognizable faces and houses were present in the N1 time window, peaking at around 165 ms after stimulus onset. These differences were found over the temporal area, where ERP category differences are usually measured. The authors also found differences between face and house AS information, however, these differences occurred earlier, peaking at 142 ms after stimulus onset, and they were present at medial-occipital electrodes (Bieniek et al., 2012). This result is in line with the results of Andrews et al. (2010), who found category-selective activation for Fourier-scrambled images at medial-occipital areas.

Although this study used faces and houses as stimuli, AS has been shown to differ also between animals and man-made objects (Torralba & Oliva, 2003; VanRullen, 2011). Therefore, it is possible that AS also modulates animal vs. vehicle differences; however, these differences might occur in a different time-window and at different electrode locations than higher-level category differences similarly to the ERP differences between recognizable and phase scrambled houses and faces in the study of Bieniek et al., (2012 - see description above).

Role of conscious percept

In contrast to studies emphasizing the role of low-level differences as underlying factors behind ERP category differences, other results suggest that ERP differences reflect some aspects of the percept (the consciously perceived impression of an object) as ERP differences have been found to correlate with behavioral results (Bacon-Macé et al., 2005; Macé et al., 2005). In the study of Bacon-Mace et al. (2005) stimuli were natural images, presented for a duration between 6.25 and 106.25 ms. Each stimulus was followed by a dynamic mask, which is known to interrupt the processing of the previously presented image. Participants performed a go/no-go task, indicating whether an animal was present on the image or not. The accuracy rates were higher when stimuli were presented longer, and they decreased with the shorter presentation times. The ERP differences between animals and non-animals started at 150 ms after stimulus onset. Moreover, the differential ERP activity in the time window of the N1 and P2 components decreased with the shorter stimulus presentation time. The amplitude of the differential brain activity correlated with the ratio of the correct answers. Therefore, the more likely participants recognized the animal in the image, the larger the ERP amplitude differences were between animals and non-animals. It has been suggested that with the longer time before the appearance of the mask, more cues could accumulate to dissociate animal and non-animal images and this led to more correct responses and larger category differences in the ERPs (Bacon-Macé et al., 2005).

Macé et al. (2005) reported similar results when the contrasts of the images were manipulated. They found lower accuracy rates in recognizing animals for images with lower contrast. Furthermore, the animal vs. non-animal differences in the occipital ERPs also decreased with lower contrast. The stimulus presentation time was 28 ms in this study for all stimuli; however, no mask was presented to interrupt the ongoing processing after the stimulus disappeared. Nevertheless, the reaction times increased for images with lower contrast, and the increasing reaction times correlated with the latency of the peak of the differential activity between animal and non-animal images. Again, this was interpreted as a cue accumulation process: the less recognizable the image was, the more time was needed to gather evidence for the recognition of an animal (Macé et al., 2005).

Beside cue accumulation for decision-making, these results also suggest that the differential activity in the N1 and P2 components reflect how participants perceived the images. The bigger the differential activity was, the more likely the images were processed and images containing an animal discriminated from images without an animal. The results of Rokszin et al. (2015) also suggest a similar interpretation. In that study, participants performed a car vs. bird discrimination task on intact and pixel-scrambled stimuli presented either in color and or greyscale versions. Scrambled stimuli were produced by cutting the original images into smaller squares, and these squares were randomly shuffled within an image. The sizes of the shuffled squares were 40x40, 20x20 and 2x2 pixels for the different stimulus types. The accuracy rates were around 95, 80, 70 and 50 % for the intact, 40x40, 20x20, and 2x2 pixels stimuli, respectively. Therefore, the recognizability of the car and bird categories decreased with the smaller square size. The color differences in the visual ERPs, on the other hand, increased from intact to 20x20 pixel images. In this study, the categorization task did not require color discrimination. When the objects were intact the color difference was not salient. With the harder recognizability of the objects, the color difference became more salient. This again suggests that the category differences in the ERPs, at least to some degree, reflect how the image is perceived. If category differences in the visual ERPs reflect at least some aspects of the percept, this means these differences should reflect some kind of visual category representations.

Correspondence between modulation of category-specific representations and modulation of category-specific ERPs

Visual ERPs and category-specific ERP responses are widely used; however, different interpretations have been suggested to explain these differences. The source of contradictory explanations could partly be that different studies measured the category differences at different electrode locations, in different tasks and for different object categories. Furthermore, many studies have looked at the ERP components as distinct entities, and did not consider spatial and temporal factors, or the accumulating evidence about visual information processing, such as the existence of stable category representations in the visual areas and the role of top-down mechanisms. The aim of this section is to point out the similarities between modulations of the visual ERP differences and activation of category representations in the brain.

ERP category differences are mainly reported for animate and inanimate categories, and for faces and man-made objects. According to the studies presented in the previous section, animals, faces and inanimate objects are categorically represented in the LOC (Kiani et al., 2007; Kriegeskorte et al., 2008; Mur et al., 2013). Moreover, source localization methods suggest that the source of the of the animal vs. non-animal ERP difference is in the LOC (Codispoti et al., 2006; Kincses et al., 2006), where animate and inanimate categories are represented. In addition, some studies have provided evidence that the ERP differences, at least partly, reflect some aspects of the percept (Bacon-Macé et al., 2005; Macé et al., 2005; Rokszin et al., 2015). This is in line with the assumption that the visual system reconstructs the outside world as an inner representation and it provides the percept of the outside objects. Furthermore, studies showing that top-down processes such as mental imagery or expectation activate the same category-specific representations as seeing those categories, suggesting that these category representations should reflect some aspects of the percept.

Beside category representations, visual ERPs can also reflect shape-specificity, similarly to LOC representations. ERP differences have been found between inanimate object categories, such as cars vs. non-car vehicles (VanRullen & Thorpe, 2001) and cars vs. shoes and chairs (Rossion et al., 2000). According to Kriegeskorte et al. (2008) these categories are represented by perceptual similarity in the LOC. Furthermore, ERP differences have been reported between computer-generated meaningless objects (Curran, Tanaka, & Weiskopf, 2002; Tokudome & Wang, 2012). In the study of Tokudome and Wang (2012), participants were trained to recognize four different artificially created highly similar 3D objects from four different viewpoints. Before training, the N1 amplitudes were more similar for different objects depicted from the same visual angle and more variable for the different viewpoints of one object. In contrast, after training when participants had learned to differentiate between the objects, the N1 amplitudes were more similar for the different viewpoints of one object, and they varied more between different objects (Tokudome & Wang, 2012). Therefore, before training, N1 amplitudes reflected perceptual similarity, which has been suggested to be the main organizing factor for shape representations in the LOC (Kriegeskorte et al., 2008). In contrast, after training, the N1 amplitudes reflected object-selectivity and viewpoint invariance. This is in line with findings indicating that IT neurons are shape-selective and they represent information in a viewpoint-independent manner (Tompa & Sáry, 2010).

Finally, top-down, attention- and task-related modulation have been shown to affect the firing of IT cells in monkeys (for review see Tompa & Sáry, 2010). Furthermore, expectation, or selective attention, as top-down factors, have been shown to affect bottom-up stimulus processing in humans (Bugatus et al., 2017; Puri et al., 2009). In a previous section, studies suggesting task-related modulation of animate vs. inanimate category differences have been already discussed. However, in those studies, category differences and decision-making processes were somewhat intermixed. In ERP studies, when category differences are measured, participants usually perform some kind of categorization or recognition task. Furthermore, the categories defined by the task are usually the same ones that are measured in the ERPs. Therefore, the content of the categorization task as a top-down factor might contribute to the ERP differences. According to the results of Bugatus et al. (2017), tasks requiring flexible category representations affect more the frontal brain areas. In contrast, visual representations are affected by selective attention. Studies comparing a two-choice animal vs. non-animal task and a go/no-go task, where animals have to be detected, both require the recognition of an animal. Thus, comparing those tasks, it is not clear whether the top-down expectation of the appearance of an animal affects the visual ERPs.

In order to sort out the contribution of top-down and bottom-up processes for ERP differences between object categories, the categorization tasks should manipulate attentional factors on the presented objects. Only a few studies have compared category differences in the visual ERPs between categorization tasks where the content of the task was modulated (e.g. human-dog vs. animate-inanimate task) instead of the method (e.g. two-choice vs. go/no-go task). Paz-Caballero, Cuetos, & Dobarro (2006) compared ERPs elicited by natural and artificial categories between a natural vs. artificial categorization task and a gender decision task, where participants had to decide about the grammatical gender of the presented objects. In the categorization task, artificial objects elicited more negative amplitudes than natural ones between 50 and 200 ms after stimulus onset. In contrast, in the gender decision task, natural images elicited a more negative amplitude compared to the amplitudes for artificial objects (Paz-Caballero et al., 2006). However, natural and artificial categories do not follow the category representations in the LOC. Paz-Caballero et al. (2006) included animals, body parts, fruits and plants in the natural category, whereas according to Kriegeskorte et al. (2008), fruits and plants belong to the inanimate category. In the study of Balas and Koldewyn (2013) stimuli were human and dog faces, presented either as a real version or as a toy. The authors found N1 amplitude differences between humans and dogs only when participants had to perform a human vs.

dog categorization task. When participants had to decide about the animacy of the stimuli (i.e., toy vs. real) the human-dog N1 difference disappeared. Therefore, these two studies indicate that the content of the task, as top-down factor, does modulate the category differences in the visual ERPs.

Content-based task modulation of N1 amplitudes for cars and birds

This section is based on the paper of Linnert, Csifcsák, and Tompa (in prep.), and parts of the paper are reproduced here. The aim of this study was to investigate how the direct attention to the content of the image, as a top-down factor modulates the category differences in the visual ERPs. In order to investigate this, we compared car-bird amplitude differences for the P1 and N1 components in two different categorization tasks. Stimuli were colored and greyscale images of cars and birds in natural scenes. Images were presented in an intact version and in two modified, unrecognizable versions. For pixel-scrambled stimuli (pixel stimuli) the original images were cut into 2x2 pixels of squares and then the small squares were shuffled. Fourier transformed images (Fourier stimuli) were scrambled using a Fourier phase randomization procedure, which preserves global low-level properties, such as the amplitude spectrum information, of the original image (Figure 1). Participants performed a two-choice task to categorize the presented images. One group had to decide whether the presented image depicted a car or a bird (category task, N=17), while the other group had to report whether the image was colored or greyscale (color task, N=18). Previous studies have compared different types of categorization tasks (e.g. go/no-go vs. forced choice task in Johnson & Olshausen, 2003; Kincses et al., 2006). In both of these tasks participant had to perform an animal vs. inanimate categorization. In our tasks, on the other hand, we manipulated the allocated attention to the content of the image. Our goal was to find two tasks, with orthogonal category boundaries, where one task requires the recognition of the object depicted in the image, whereas the other task does not. In this study, the same stimuli could be categorized either as car or bird, or as color or greyscale. Furthermore, both the category and the color were salient features, therefore, they were easy to perceive even if the task did not require it. In this way, we could compare car-bird (and also, color-greyscale) differences when these features were attended and non-attended.

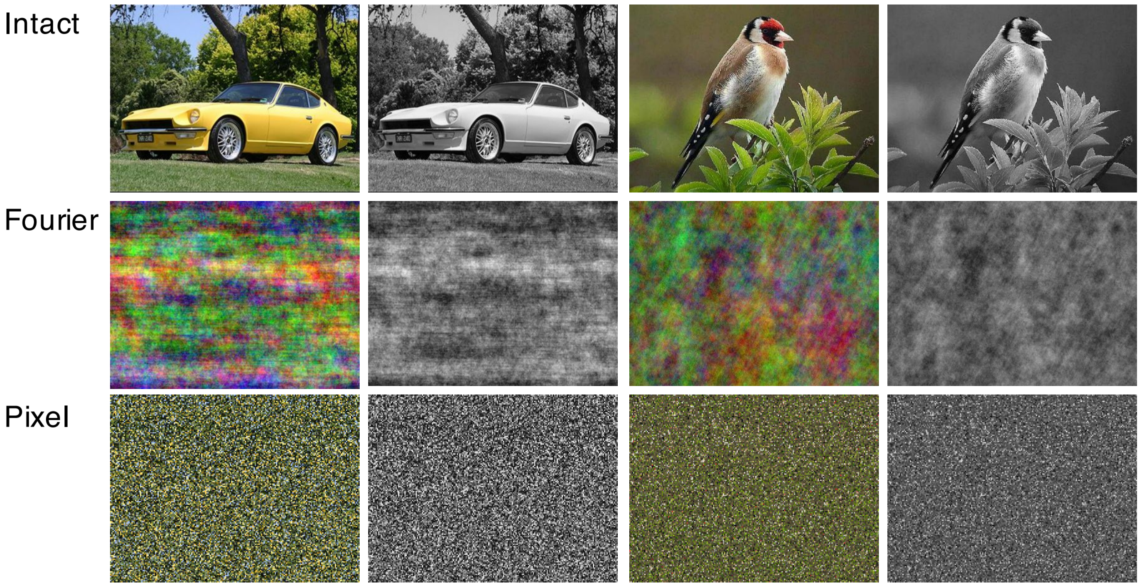


Figure 1. Example stimuli for intact (top row) images and their Fourier-scrambled (middle row) and pixel-scrambled (bottom row) versions. Image adapted from Linnert et al. (in prep).

Visual ERPs were measured over the occipital (electrodes: Oz, O1, O2, 71, 74, 76, 82) as well as the left and right temporal areas (left electrodes: T5, 59, 65; right electrodes: T6, 90, 91,) based on the scalp distribution of the car-bird and color-greyscale differences (see Figure 2). We found category and color differences already at the P1 component; however, the task modulated these differences only at the N1 component.

Only the results concerning task-related category and color differences are summarized here (see ERPs in Figure 3).

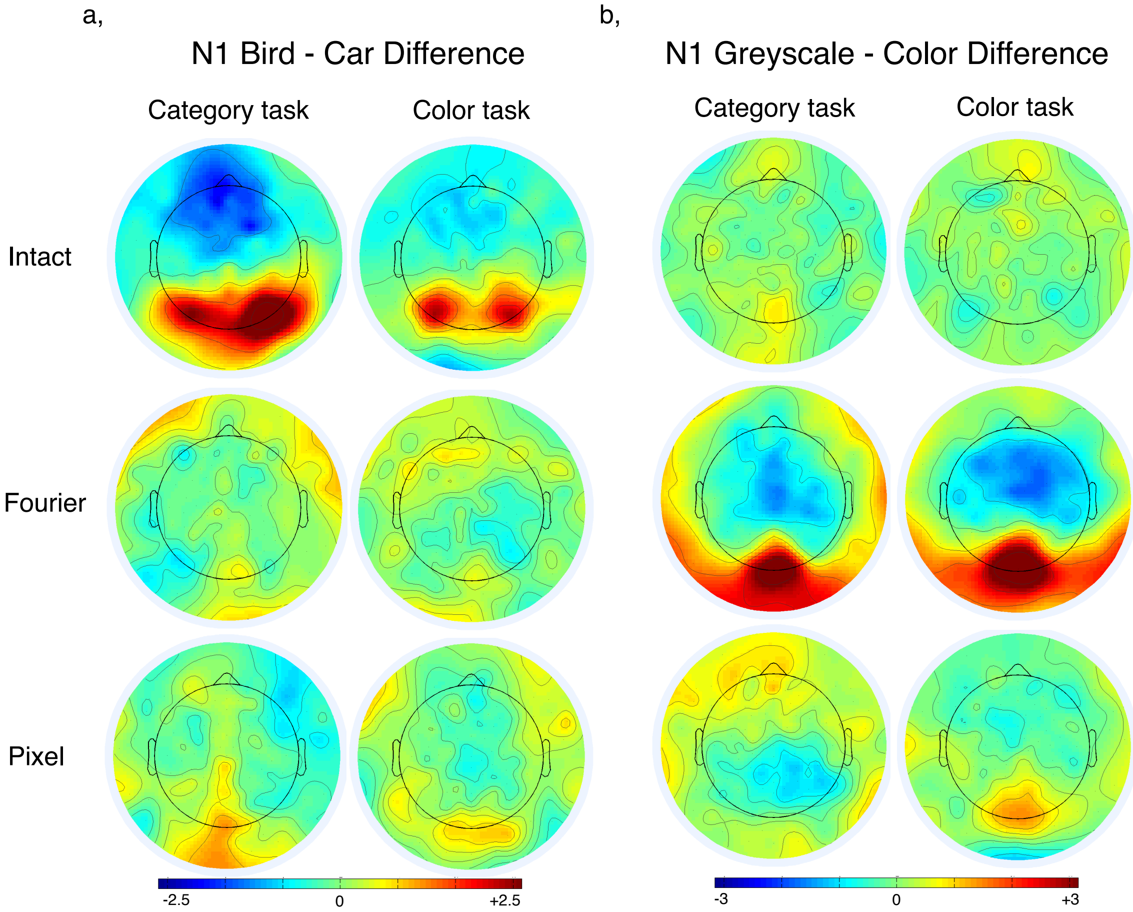


Figure 2. Cortical maps at 170 ms after stimulus onset (at the peak latency of the N1 component) for bird-car ERP differences (a) and greyscale-color differences (b) for the intact (top row), Fourier-scrambled (middle row) and pixel-scrambled images (bottom row), in the category and color discrimination task respectively. Image adapted from Linnert et al. (in prep).

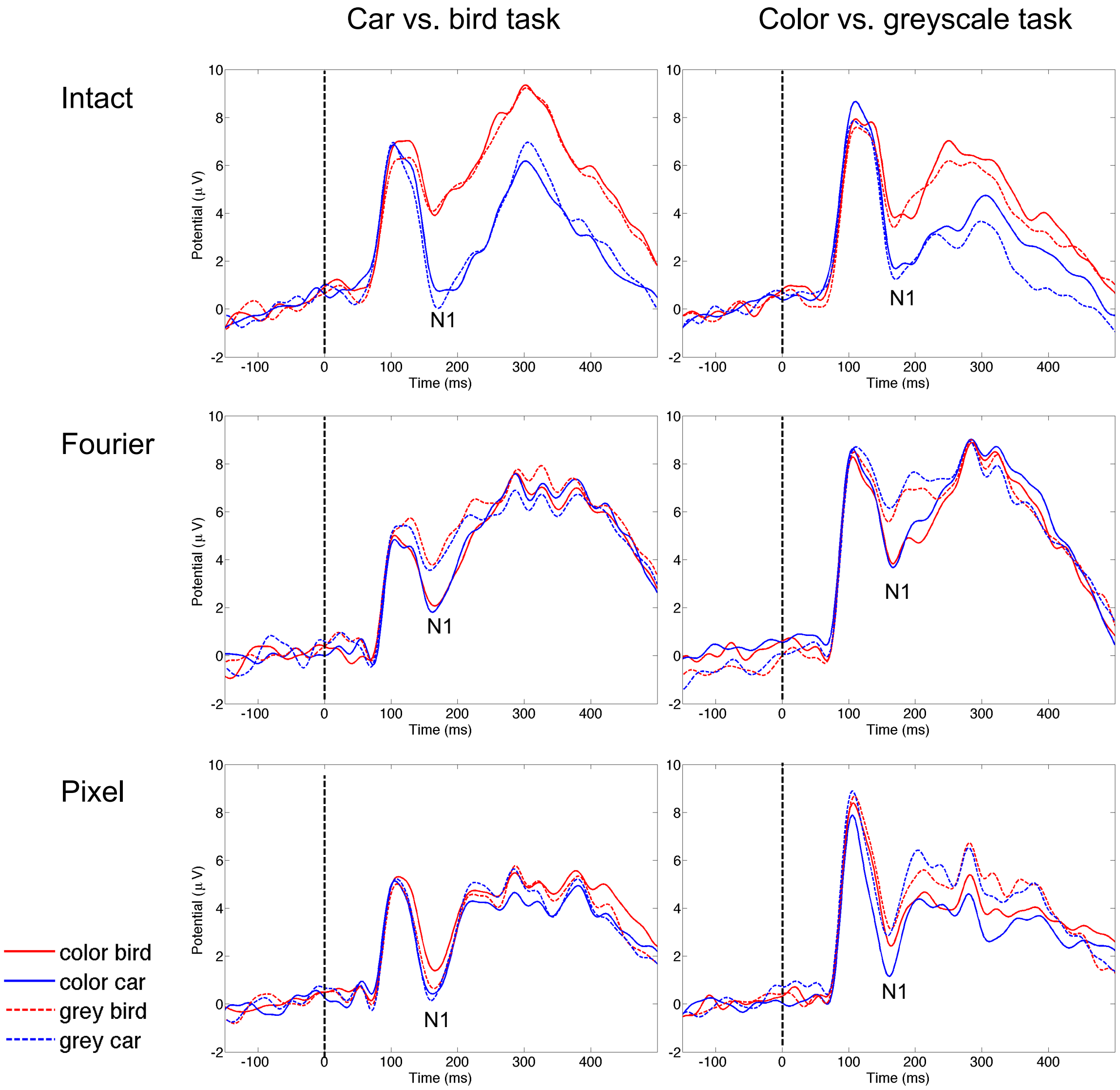


Figure 3. ERPs in the car vs. bird (left) and color vs. greyscale (right) categorization tasks. ERPs for intact images show the average waveforms of the 3 electrodes over the right temporal area (top row). ERPs for Fourier (middle row) and pixel (bottom row) images show the average waveforms of 7 electrodes over the occipital area. Image adapted from Linnert et al. (in prep).

The most pronounced category differences were present for the intact images over the left and right temporal areas for the N1 component; the N1 showed significantly more negative amplitude for cars than birds (see Figure 2a). This car-bird difference was modulated by the categorization task and differed between the two hemispheres (three-way interaction between category task and hemisphere). In the category discrimination task, the car-bird difference was larger over the right hemisphere than over the left. In contrast, in the color discrimination task, category differences were similar over both hemispheres (see Figure 2a, top row). Cars evoked more negative amplitudes than birds also for the pixel images; however, instead of the temporal area, it occurred over the occipital area (see Figure 2a). Category differences for the pixel images were modulated by the color; the car-bird difference was larger for colored images than for greyscale ones. This suggests that the category difference for the pixel images was based on color difference. On the other hand, participants could not categorize the pixel images as animals and vehicles (accuracy rates were at chance level). Therefore, color-based category differences over the occipital area for pixel stimuli did not result in conscious recognition. Moreover, this difference was not modulated by the task. These differences between intact and pixel images indicate that the temporal category differences for the intact images reflected higher-level category-specific processes, possibly activation of category representations. Moreover, task-related top-down projections modulated the category differences for the intact images. For the pixel images category differences only reflected low-level color differences between stimulus sets. The color is unlikely a category-specific feature for birds and cars, as both categories can appear in a broad range of colors. This can be the reason why top-down factors could not modulate color-based category differences for the pixel images.

The color-greyscale difference was most pronounced for Fourier images over the medial occipital area (Figure 2b). However, this difference was not modulated by the task. We also found color difference for the pixel images over the medial occipital area, but not as pronounced as for the Fourier images. Nevertheless, this difference was larger in the color discrimination task than in the category discrimination task (Figure 2b).

Therefore, color selective processes can be accessed by top-down projections, but only when the task specifically requires such discrimination.

Interestingly, we did not find color differences for the intact stimuli, not even in the color discrimination task. On the other hand, animal-vehicle differences occurred also on the color discrimination task for intact images, whereas for the Fourier stimuli, color differences also occurred in the category discrimination task. This indicates that if an image contains an object, then the presence of the object is more salient than the color. However, in the absence of a recognizable shape, the color information becomes more salient. A possible reason for the lack of color difference for the pixel stimuli in the category discrimination task could be that for those stimuli the color was not very salient, especially with the short (200 ms) image presentation.

Overall, these results suggest that selective attention to the presented objects does modulate car-bird category differences in the visual ERPs. Moreover, it shows that both higher-level, category-specific processes and lower-level color-specific processes can be modulated by task-related, top-down projections. Only car-bird differences over the temporal areas resulted in conscious category recognition; therefore, category differences only over this area reflected the conscious percept. This result is in line with top-down activation of category-specific areas in the LOC. Therefore, they suggest that category differences at temporal electrode locations reflect activation of category representations in the LOC.

Aims of the thesis

Taken together, the reviewed evidence suggests that category differences in the visual ERPs are very likely related to shape- and/or category representations in the higher-level visual areas. Furthermore, both our previous study (Linnert et al., in prep) and the study of Balas and Koldewyn (2013) found content-based task modulation in the visual N1 component. Nevertheless, these studies used basic-level categories (car, bird, human face, dog face). These objects have similar shapes, but also belong to animate and inanimate categories. Therefore, category-selectivity can be both based on shape- or category-specific representations. Several studies have reported ERP differences for animate vs. inanimate categories; however, it has not been studied whether the content of the categorization task (e.g. animal-vehicle vs. color-greyscale categorization) can modulate these category differences in the visual ERPs. Furthermore, beside higher-level representations, visual ERPs can also reflect low-level, category-specific differences, such as the amplitude spectrum information. Therefore, it is possible that task-related modulation can operate both at the level of category representations or at the level of lower-level category-specific processes.

The aims of the studies reported in this thesis are to systematically investigate these questions. (1) Can we find content-based task-related ERP modulation for animal and non-animal categories? (see Papers 1, 2 and 3) (2) Do these task-related modulations affect category representations or lower-level category-specific information? (see Papers 1 and 3) (3) Are they based on coarse shape information about the objects, or are they based on fine-grained, detailed representations? (see paper 3) (4) Does the task also modulate ERP differences between inanimate object categories; and what role does the global shape play in this modulation? (see paper 4). In Paper 1 and 2, we investigated the effect of amplitude spectrum information on the animal-vehicle differences in the visual ERPs in different categorization tasks. In Paper 3, we examined whether animal-vehicle ERP differences and their task-related modulation is specific for high or low spatial frequency information. In Paper 4, we investigated whether the task also modulates the

ERP differences for inanimate object categories (tools and vehicles), and whether the presence of global shape plays a role in tool-vehicle discrimination. The next four chapters of the thesis will present these studies. Chapter 6 will provide a general discussion regarding the presented studies.

References

Aguirre, G. K., Zarahn, E., & D'Esposito, M. (1998). An area within human ventral cortex sensitive to "building" stimuli: evidence and implications. *Neuron*, *21*, 373–383. http://doi.org/10.1016/S0896-6273(00)80546-2

Andrews, T. J., Clarke, A., Pell, P., & Hartley, T. (2010). Selectivity for low-level features of objects in the human ventral stream. *NeuroImage*, *49*(1), 703–711. http://doi.org/10.1016/j.neuroimage.2009.08.046

Antal, A., Kéri, S., Kovács, G., Janka, Z., & Benedek, G. (2000). Early and late components of visual categorization: an event-related potential study. *Cognitive Brain Research*, *9*(1), 117–119. http://doi.org/10.1016/S0926-6410(99)00053-1

Antal, A., Kéri, S., Kovács, G., Liszli, P., Janka, Z., & Benedek, G. (2001). Event-related potentials from a visual categorization task. *Brain Research Protocols*, *7*, 131–136. http://doi.org/10.1016/S1385-299X(01)00055-1

Bacon-Macé, N., Macé, M. J. M., Fabre-Thorpe, M., & Thorpe, S. J. (2005). The time course of visual processing: Backward masking and natural scene categorisation. *Vision Research*, *45*, 1459–1469. http://doi.org/10.1016/j.visres.2005.01.004

Balas, B., & Koldewyn, K. (2013). Early visual ERP sensitivity to the species and animacy of faces. *Neuropsychologia*, *51*(13), 2876–2881. http://doi.org/10.1016/j.neuropsychologia.2013.09.014

Battistoni, E., Stein, T., & Peelen, M. V. (2017). Preparatory attention in visual cortex. *Annals of the New York Academy of Sciences*, 1–16. http://doi.org/10.1111/nyas.13320

Behrmann, M., & Kimchi, R. (2003). What does visual agnosia tell us about perceptual organization and its relationship to object perception? *Journal of Experimental Psychology: Human Perception and Performance*, *29*(1), 19–42. http://doi.org/10.1037/0096-1523.29.1.19

Bentin, S., Taylor, M. J., Rousselet, G. A., Itier, R. J., Caldara, R., Schyns, P. G., … Rossion, B. (2007). Controlling interstimulus perceptual variance does not abolish N170 face sensitivity. *Nature Neuroscience*, *10*(7), 801–802. http://doi.org/10.1038/nn0707-801

Bieniek, M. M., Pernet, C. R., & Rousselet, G. A. (2012). Early ERPs to faces and objects are driven by phase, not amplitude spectrum information: evidence from parametric, test-retest, single-subject analyses. *Journal of Vision*, *12*(13), 12. http://doi.org/10.1167/12.13.12

Bugatus, L., Weiner, K. S., & Grill-Spector, K. (2017). Task alters category representations in prefrontal but not high-level visual cortex. *NeuroImage*, *155*, 437–449. http://doi.org/10.1016/j.neuroimage.2017.03.062

Chao, L. L., Haxby, J. V, & Martin, A. (1999). Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nature Neuroscience*, *2*(10), 913–919. http://doi.org/10.1038/13217

Chelazzi, L., Duncan, J., Miller, E. K., & Desimone, R. (1998). Responses of neurons in inferior temporal cortex during memory-guided visual search. *Journal of Neurophysiology*, *80*(6), 2918–2940. http://doi.org/10.1152/jn.1998.80.6.2918

Codispoti, M., Ferrari, V., Junghöfer, M., & Schupp, H. T. (2006). The categorization of natural scenes: Brain attention networks revealed by dense sensor ERPs. *NeuroImage*, *32*, 583–591. http://doi.org/10.1016/j.neuroimage.2006.04.180

Curran, T., Tanaka, J. W., & Weiskopf, D. M. (2002). An electrophysiological comparison of visual categorization and recognition memory. *Cognitive, Affective & Behavioral Neuroscience*, *2*(1), 1–18.

Delorme, A., Rousselet, G. A., Macé, M. J.-M., & Fabre-Thorpe, M. (2004). Interaction of top-down and bottom-up processing in the fast visual analysis of natural scenes. *Cognitive Brain Research*, *19*(2), 103–113. http://doi.org/10.1016/j.cogbrainres.2003.11.010

Di Lollo, V., Enns, J. T., & Rensink, R. a. (2000). Competition for consciousness among visual events: the psychophysics of reentrant visual processes. *Journal of Experimental Psychology. General*, *129*(4), 481–507. http://doi.org/10.1037/0096-3445.129.4.481

Esterman, M., & Yantis, S. (2010). Perceptual expectation evokes category-selective cortical activity. *Cerebral Cortex*, *20*, 1245–1253. http://doi.org/10.1093/cercor/bhp188

Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2001). Categorical representation of visual stimuli in the primate prefrontal cortex. *Science*, *291*(2001), 312–316. http://doi.org/10.1126/science.291.5502.312

Gilbert, C. D., & Sigman, M. (2007). Brain states: Top-down influences in sensory processing. *Neuron*, *54*, 677–696. http://doi.org/10.1016/j.neuron.2007.05.019

Goodale, M. A., & Keith Humphrey, G. (1998). The objects of action and perception. *Cognition*, *67*(1–2), 181–207. http://doi.org/10.1016/S0010-0277(98)00017-1

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, *15*(1), 20–25. http://doi.org/10.1016/0166-2236(92)90344-8

Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision Research*, *41*(10–11), 1409–1422.

http://doi.org/10.1016/S0042-6989(01)00073-6

Ishai, A., Ungerleider, L. G., Martin, A., Schouten, J. L., & Haxby, J. V. (1999). Distributed representation of objects in the human ventral visual pathway. *Proceedings of the National Academy of Sciences*, *96*(16), 9379–9384. http://doi.org/10.1073/pnas.96.16.9379

Jiang, X., Bradley, E., Rini, R., Zeffiro, T., VanMeter, J., & Riesenhuber, M. (2007). Categorization training results in shape- and category-selective human neural plasticity. *Neuron*, *53*, 891–903. http://doi.org/10.1016/j.neuron.2007.02.015

Johnson, J. S., & Olshausen, B. (2003). Timecourse of neural signatures of object recognition. *Journal of Vision*, *3*(7), 499–512. http://doi.org/10.1167/3.7.4

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *17*(11), 4302–4311. http://doi.org/10.1098/Rstb.2006.1934

Kiani, R., Esteky, H., Mirpour, K., & Tanaka, K. (2007). Object category structure in response patterns of neuronal population in monkey inferior temporal cortex. *Journal of Neurophysiology*, *97*(6), 4296–4309. http://doi.org/10.1152/jn.00024.2007

Kiefer, M. (2001). Perceptual and semantic sources of category-specific effects: event-related potentials during picture and word categorization. *Memory & Cognition*, *29*(1), 100–116. http://doi.org/10.3758/BF03195745

Kincses, T. Z., Chadaide, Z., Varga, E. T., Antal, A., & Paulus, W. (2006). Task-related temporal and topographical changes of cortical activity during ultra-rapid visual categorization. *Brain Research*, *1112*, 191–200. http://doi.org/10.1016/j.brainres.2006.07.044

Kotsoni, E., Csibra, G., Mareschal, D., & Johnson, M. H. (2007). Electrophysiological correlates of common-onset visual masking. *Neuropsychologia*, *45*(10), 2285–2293. http://doi.org/10.1016/j.neuropsychologia.2007.02.023

Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., … Bandettini, P. A. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron*, *60*(6), 1126–1141. http://doi.org/10.1016/j.neuron.2008.10.043

Ling, S., Liu, T., & Carrasco, M. (2009). How spatial and feature-based attention affect the gain and tuning of population responses. *Vision Research*, *49*(10), 1194–1204. http://doi.org/10.1016/j.visres.2008.05.025

Linnert, S., Csifcsák, G., & Tompa, T. (in prep). Top-down modulation of animate-inanimate category differences in the visual N1.

Luck, S. J. (2005). *An Introduction to the Event-Related Potential Technique*. Cambridge, Mass: MIT press.

Macé, M. J. M., Thorpe, S. J., & Fabre-Thorpe, M. (2005). Rapid categorization of achromatic natural scenes: How robust at very low contrasts? *European Journal of Neuroscience*, *21*, 2007–2018. http://doi.org/10.1111/j.1460-9568.2005.04029.x

Mahon, B. Z., & Caramazza, A. (2009). Concepts and categories: A cognitive neuropsychological perspective. *Annual Review of Psychology*, *60*(1), 27–51. http://doi.org/10.1146/annurev.psych.60.110707.163532

Malach, R., Reppas, J. B., Benson, R. R., Kwong, K. K., Jiang, H., Kennedy, W. a, … Tootell, R. B. (1995). Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proceedings of the National Academy of Sciences*, *92*(18), 8135–8139. http://doi.org/10.1073/pnas.92.18.8135

Martinez-Trujillo, J. C., & Treue, S. (2004). Feature-based attention increases the selectivity of population responses in primate visual cortex. *Current Biology*, *14*(9), 744–751. http://doi.org/10.1016/j.cub.2004.04.028

Mur, M., Meys, M., Bodurka, J., Goebel, R., Bandettini, P. A., & Kriegeskorte, N. (2013). Human object-similarity judgments reflect and transcend the primate-IT object representation. *Frontiers in Psychology*, *4*, 1–22. http://doi.org/10.3389/fpsyg.2013.00128

Nemeth, K., Kovacs, P., Vakli, P., Kovacs, G., & Zimmer, M. (2014). Phase noise reveals early category-specific modulation of the event-related potentials. *Frontiers in Psychology*, *5*(April). http://doi.org/10.3389/fpsyg.2014.00367

O'Craven, K. M., & Kanwisher, N. (2000). Mental imagery of faces and places activates corresponding stiimulus-specific brain regions. *Journal of Cognitive Neuroscience*, *12*, 1013–1023. http://doi.org/10.1162/08989290051137549

Paz-Caballero, D., Cuetos, F., & Dobarro, A. (2006). Electrophysiological evidence for a natural/artifactual dissociation. *Brain Research*, *1067*, 189–200. http://doi.org/10.1016/j.brainres.2005.10.046

Peelen, M. V, & Kastner, S. (2011). A neural basis for real-world visual search in human occipitotemporal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 12125–12130. http://doi.org/10.1073/pnas.1101042108

Proverbio, A. M., Del Zotto, M., & Zani, A. (2007). The emergence of semantic categorization in early visual processing: ERP indices of animal vs. artifact recognition. *BMC Neuroscience*, *8*, 24. http://doi.org/10.1186/1471-2202-8-24

Puri, A. M., Wojciulik, E., & Ranganath, C. (2009). Category expectation modulates baseline and stimulus-evoked activity in human inferotemporal cortex. *Brain Research*, *1301*, 89–99. http://doi.org/10.1016/j.brainres.2009.08.085

Reddy, L., Tsuchiya, N., & Serre, T. (2010). Reading the mind's eye: Decoding category information during mental imagery. *NeuroImage*, *50*(2), 818–825. http://doi.org/10.1016/j.neuroimage.2009.11.084

Reeder, R. R., Perini, F., & Peelen, M. V. (2015). Preparatory activity in posterior temporal cortex causally contributes to object detection in scenes. *Journal of Cognitive Neuroscience*, *27*(11), 2117–2125. http://doi.org/10.1162/jocn_a_00845

Rizzi, C., Piras, F., & Marangolo, P. (2010). Top-down projections to the primary visual areas necessary for object recognition: A case study. *Vision Research*, *50*(11), 1074–1085. http://doi.org/10.1016/j.visres.2010.03.018

Rokszin, A. A., Győri-Dani, D., Linnert, S., Krajcsi, A., Tompa, T., & Csifcsák, G. (2015). The interplay of holistic shape, local feature and color information in object categorization. *Biological Psychology*, 1–12. http://doi.org/10.1016/j.biopsycho.2015.05.002

Rossion, B., & Caharel, S. (2011). ERP evidence for the speed of face categorization in the human brain: Disentangling the contribution of low-level visual cues from face perception. *Vision Research*, *51*(12), 1297–1311. http://doi.org/10.1016/j.visres.2011.04.003

Rossion, B., Gauthier, I., Tarr, M. J., Despland, P., Bruyer, R., Linotte, S., & Crommelinck, M. (2000). The N170 occipito-temporal component is delayed and enhanced to inverted faces but not to inverted objects: an electrophysiological account of face-specific processes in the human brain. *Neuroreport*, *11*(1), 69–74. http://doi.org/10.1097/00001756-200001170-00014

Rossion, B., & Jacques, C. (2008). Does physical interstimulus variance account for early electrophysiological face sensitive responses in the human brain? Ten lessons on the N170. *NeuroImage*, *39*, 1959–1979.

http://doi.org/10.1016/j.neuroimage.2007.10.011

Rousselet, G. A., Husk, J. S., Bennett, P. J., & Sekuler, A. B. (2008). Time course and robustness of ERP object and face differences. *Journal of Vision*, *8*(12), 1–18. http://doi.org/10.1167/8.12.3

Rousselet, G. a, Fabre-Thorpe, M., & Thorpe, S. J. (2002). Parallel processing in high-level categorization of natural images. *Nature Neuroscience*, *5*(7), 629–630. http://doi.org/10.1038/nn866

Rousselet, G. a, Pernet, C. R., Bennett, P. J., & Sekuler, A. B. (2008). Parametric study of EEG sensitivity to phase noise during face processing. *BMC Neuroscience*, *9*, 98. http://doi.org/10.1186/1471-2202-9-98

Schendan, H. E., & Kutas, M. (2007). Neurophysiological evidence for the time course of activation of global shape, part, and local contour representations during visual object categorization and memory. *Journal of Cognitive Neuroscience*, *19*, 734–749. http://doi.org/10.1162/jocn.2007.19.5.734

Servos, P., & Goodale, M. A. (1995). Preserved visual imagery in visual form agnosia. *Neuropsychologia*, *33*(11), 1383–1394. http://doi.org/10.1016/0028-3932(95)00071-A

Stokes, M., Thompson, R., Nobre, A. C., & Duncan, J. (2009). Shape-specific preparatory activity mediates attention to targets in human visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(46), 19569–19574. http://doi.org/10.1073/pnas.0905306106

Thierry, G., Martin, C. D., Downing, P. E., & Pegna, A. J. (2007a). Is the N170 sensitive to the human face or to several intertwined perceptual and conceptual factors? *Nature Neuroscience*, *10*(7), 802–803. http://doi.org/10.1038/nn0707-802

Thierry, G., Martin, C. D., Downing, P., & Pegna, A. J. (2007b). Controlling for interstimulus perceptual variance abolishes N170 face selectivity. *Nature Neuroscience*, *10*(4), 505–511. http://doi.org/10.1038/nn1864

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, *381*(6582), 520–522. http://doi.org/10.1038/381520a0

Tokudome, W., & Wang, G. (2012). Similarity dependency of the change in ERP component N1 accompanying with the object recognition learning. *International Journal of Psychophysiology*, *83*(1), 102–109. http://doi.org/10.1016/j.ijpsycho.2011.10.012

Tompa, T., & Sáry, G. (2010). A review on the inferior temporal cortex of the macaque. *Brain Research Reviews*, *62*(2), 165–182. http://doi.org/10.1016/j.brainresrev.2009.10.001

Torralba, A., & Oliva, A. (2003). Statistics of natural image categories. *Network*, *14*(3), 391–412. http://doi.org/10.1088/0954-898X/14/3/302

Tremel, J. J., & Wheeler, M. E. (2015). Content-specific evidence accumulation in inferior temporal cortex during perceptual decision-making. *NeuroImage*, *109*, 35–49. http://doi.org/10.1016/j.neuroimage.2014.12.072

Ungerleider, L.G., Mishkin, M. (1982). Two cortical visual systems. In R. J. W. Ingle, David J., Goodale, Melvyn A., Mansfield (Ed.), *Analysis of Visual Behaviour.* (pp. 549–580). Cambridge: The MIT Press.

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In *Analysis of visual behavior* (pp. 549–586). Cambridge, Mass: The MIT Press.

van der Linden, M., van Turennout, M., & Indefrey, P. (2010). Formation of category representations in superior temporal sulcus. *Journal of Cognitive Neuroscience*, *22*, 1270–1282. http://doi.org/10.1162/jocn.2009.21270

VanRullen, R. (2011). Four common conceptual fallacies in mapping the time course of recognition. *Frontiers in Psychology*, *2*, 1–6. http://doi.org/10.3389/fpsyg.2011.00365

VanRullen, R., & Thorpe, S. J. (2001). The time course of visual processing: from early perception to decision-making. *Journal of Cognitive Neuroscience*, *13*, 454–461. http://doi.org/10.1162/08989290152001880

Warrington, E. K., & Shallice, T. (1984). Category specific semantic impairments. *Brain*, *107*(3), 829–853. http://doi.org/10.1093/brain/107.3.829

Whitwell, R. L., Milner, A. D., & Goodale, M. A. (2014). The two visual systems hypothesis: New challenges and insights from visual form agnosic patient DF. *Frontiers in Neurology*, *5*, 1–8. http://doi.org/10.3389/fneur.2014.00255

Zhu, W., Drewes, J., Peatfield, N., & Melcher, D. (2016). Differential visual processing of animal images, with and without conscious awareness. *Frontiers in Human Neuroscience*, *10*(October), 513. http://doi.org/10.3389/FNHUM.2016.00513

PAPER 1

Underlying mechanisms of task-related modulation of animal vs. vehicle differences in the visual ERPs

Szilvia Linnert, Vincent Reid, Gert Westermann

Lancaster University

Abstract

Animal vs. non-animal category differences that occur in visual event-related potentials (ERPs) have been widely studied. It has been shown that such differences are modulated by the task, suggesting that they arise from interactions between bottom-up and top-down processes. However, it is not clear whether the task affects the physical or the shape- or category-specific stages of the processing. The aim of the present study was to investigate whether this task-related ERP modulation depends on amplitude spectrum information (as a physical feature) or occurs irrespectively of low-level physical image characteristics. The stimuli were greyscale images of animals and vehicles with an “X” or an “O” in the background. The amplitude spectrum information was either kept intact in the images or it was equalized across stimuli. Half of the participants had to perform an animal vs. vehicle categorization task, whereas the other half performed an “X” vs. “O” categorization task. We found task-related modulation in the P1 component over the occipital area: a category difference was present only in the animals vs. vehicle categorization task and only when the amplitude spectrum was available. For the P2 and N2, category differences were greater in the animal vs. vehicle categorization task than in the “X” vs. “O” task over the temporal areas. Furthermore, this effect occurred irrespectively of the low-level image modulation. Therefore, we have separated two different top-down mechanisms. The early task-related modulation suggests a top-down process that is selective to physical features. The later task effect occurs irrespectively of the physical differences, suggesting modulation of higher-level visual processes (for instance, activation of visual category representations). Overall, these results are in line with recent theories suggesting flexible interactions between bottom-up visual processes and top-down factors such as task demands.

Underlying mechanisms of task-related modulation of the animal vs. vehicle differences in the visual ERPs

Several studies have reported animate vs. inanimate category differences in the early visual event-related potentials (ERPs) (e.g. Antal et al., 2001; Antal, Kéri, Kovács, Janka, & Benedek, 2000; Codispoti, Ferrari, Junghöfer, & Schupp, 2006; Johnson & Olshausen, 2003; Proverbio, Zotto, & Zani, 2007; Proverbio, Burco, Zotto, & Zani, 2004; VanRullen & Thorpe, 2001). Such differences have been mainly reported for the visual N1, a negative component over the posterior brain areas peaking between 130-200 ms after stimulus onset (e.g. Antal et al., 2001, 2000; Delorme, Rousselet, Macé, & Fabre-Thorpe, 2004; Proverbio et al., 2007). However, some studies have found differences already at the visual P1, a positive component between 70-120 ms after stimulus onset (e.g. VanRullen & Thorpe, 2001). A core question about these observed differences is whether they are due to low-level perceptual processes or to higher-level category-specific information. Both low-level physical differences between stimulus sets (e.g. Thierry, Martin, Downing, & Pegna, 2007; VanRullen, 2011) and higher-level category-related visual processes (Antal et al., 2000; Bacon-Macé et al., 2005) have been suggested to evoke such differences in the visual ERPs. Visual ERPs are widely used in order to investigate visual categorization processes. However, it is hard to withdraw any conclusions if it is not clear what kind of processes can be reflected in these ERP components.

The speed of the appearance of such category specific differences has led to the assumption that they are based on a fast bottom-up sweep of the incoming visual information (S. J. Thorpe & Fabre-Thorpe, 2001). On the other hand, it has also been shown that these early visual components are modulated by top-down processes such as attention (Pollux, Hall, Roebuck, & Guo, 2011; Vogel & Luck, 2000). Recent studies have suggested that category differences in the visual ERPs reflect interaction between task demands and bottom-up processes. Therefore, not only the ERP components, but also the magnitude of the category differences observed in the visual ERPs has been shown to be modulated by the performed categorization task (Balas & Koldewyn, 2013; Linnert, Csifcsak & Tompa, in prep). For instance, in a study by Linnert et al (in prep) participants had to categorize colored and grayscale images of birds and cars either by their category (car vs. bird task) or by their color (color vs. greyscale task). The authors found larger N1 amplitudes for cars compared to birds in both tasks; however, the category difference was larger in the car vs. bird task. They also found task-related N1 color differences for unrecognizable stimuli. The unrecognizable images were produced by randomly scrambling 2x2 pixel squares of the original images. For these stimuli, the N1 amplitudes differed between colored and greyscale images, and this difference only existed when participants performed a color vs. greyscale task but not in the car vs. bird task. In addition, the category difference for the recognizable images showed a temporal scalp distribution, indicating the involvement of higher-level visual areas, whereas the color-based difference for the unrecognizable images showed a medial-occipital scalp distribution, indicating the involvement of lower-level visual areas. These results are in line with the assumption that low-level features are processed in the early stages of the visual pathway, while higher-level features, such as complex object shapes or categories, are processed in the later stages of the visual pathway (Andrews et al., 2010; Lerner, Hendler, Ben-Bashat, Harel, & Malach, 2001). Therefore, the visual ERPs could reflect interaction between top-down and bottom-up processes at the level of physical features (such as color) or at the level of higher-level categorical processes.

Nevertheless, higher-level categorical processes can either reflect activation of category representations, or selectivity for category-specific low-level features. Recent results suggest that some categories, such as animals, faces, bodies, and inanimate objects are indeed represented in the higher-level visual brain areas (Kriegeskorte et al., 2008; Mur et al., 2013). These are also the most extensively studied categories in ERP paradigms; thus, ERP differences for these categories might reflect category representations. On the other hand, category selectivity can also emerge from differential processing of low-level features, such as the amplitude spectrum information (Andrews et al., 2010). Therefore, categorical differences in the ERPs might reflect only category-specific low-level differences.

The role of amplitude and phase spectra as different sources of information for perception have been extensively studied. The amplitude spectrum (AS) of an image refers to its energy at different spatial frequencies. In this way, it describes global, spatial characteristics of an image. However, from looking at an AS, the content of the image is not recognizable. The phase spectrum, on the other hand, provides information about the structure of the objects, such as the shape or configuration of local features. Modifying the AS affects the appearance of an image, but the depicted objects are still visible and recognizable. In contrast, scrambling the phase information (using a Fourier phase randomization procedure - see e.g. Rossion & Caharel, 2011) affects the recognizability of the depicted objects (Gaspar & Rousselet, 2009). Thus, phase scrambled images don’t contain shape information but they contain the AS information of the original image (see Fourier images in the study presented at the end of Intro). Therefore, results regarding the processing of these images indicate the role of AS information in visual processing. It has been suggested that the AS plays a role in visual categorization as a low-level physical factor (Torralba & Oliva, 2003), as it tends to systematically differ between animals and faces, between animals and man-made objects (Bieniek, Pernet, & Rousselet, 2012; VanRullen, 2011) or between natural and artificial scenes (Gaspar & Rousselet, 2009).

The role of the AS has been widely investigated in EEG experiments (e.g., Bieniek, Pernet, & Rousselet, 2012; Nemeth, Kovacs, Vakli, Kovacs, & Zimmer, 2014; Rossion & Caharel, 2011; Rousselet, Husk, Bennett, & Sekuler, 2008; Rousselet, Pernet, Bennett, & Sekuler, 2008); however, contradictory results have been found. Some studies have emphasized the role of the AS, as a low-level factor, to explain category differences in the visual ERPs (e.g. Nemeth et al., 2014; Rossion & Caharel, 2011). These studies have found ERP differences between unrecognizable, phase scrambled cars and faces where the AS information was kept intact. Other results, on the other hand, suggest that the role of phase information is more important (e.g. Bieniek et al., 2012; Rousselet et al., 2008). For instance Bieniek et al. (2012) used both unrecognizable, phase-scrambled and recognizable, AS modulated images of faces and houses as stimuli. They found strong and reliable ERP differences between houses and faces containing phase information, whereas the differences between the phase-scrambled images were weaker and less reliable across the different experimental sessions.

A third possibility is that it is not the amplitude or phase spectrum alone, but instead the interaction between these two kinds of information that is important in visual categorization (Gaspar & Rousselet, 2009). When phase scrambled images were used in different studies, despite the lack of visible shape, participants do perform above chance level in categorizing these images (Linnert et al, in prep; Wichmann, Braun, & Gegenfurtner, 2006). Nevertheless, categorization accuracy for phase scrambled images is much lower compared to images containing phase/shape information. Furthermore, accuracy rates are high in categorizing recognizable, AS equalized images, however, they are decreased compared to categorizing images with intact AS (Drewes, 2006).

Therefore, these studies show that the presence or absence of AS information does influence participants' response accuracy, but not so dramatically than the phase information. Still, it is very likely that both sources of information play a role in visual categorization processes.

The suggestion about the important role of the phase information is in line with the known function of the higher-level areas of the visual system, such as the inferior temporal cortex (IT) (Rousselet et al., 2008). The neurons in IT are shape-selective (for review see Tompa & Sáry, 2010) and it has been suggested that visual forms are represented in this area in a topographical manner (Chao et al., 1999; Ishai et al., 1999; Kriegeskorte et al., 2008). Therefore, it is very likely that the presence or absence of a visual shape in an image would evoke different processing mechanisms compared to the phase-scrambled, unrecognizable images, that contain only the AS information of the original image. Nevertheless, phase-scrambled images do evoke category-selective responses, however, not in IT but in more posterior visual areas (Andrews et al., 2010; Bieniek et al., 2012), suggesting that AS is a category-specific low-level feature.

Taken together, the existing evidence suggests that both AS and phase information play a role in category specific responses, with phase information as a higher-level and AS as a lower-level feature. The ERP is a useful technique to investigate the time-course of categorization. However, it would be important to disentangle the contradicting explanations about the underlying factors of category differences in the visual ERPs. The aims of the present study were to investigate whether AS information contributes to the animate vs. inanimate category differences in the visual ERPs and whether this information is modulated by the task. Stimuli consisted of animals and vehicles appearing on a grey background. The rationale to use single objects instead of natural scenes was to avoid the confound of the effect of the background scene on the AS

(Torralba & Oliva, 2003). The AS information was either kept intact or it was averaged across the stimulus set. If the AS information contributes to the animal-vehicle ERP differences, then the ERPs between these categories should differ between images with intact and averaged AS in terms of magnitude or scalp distribution. In order to investigate top-down task effects, two participant groups performed different categorization tasks. In order to enable multiple categorizations of the images, an “X” or an “O” appeared somewhere next to the object. Half of the participants performed an animal vs. vehicle task, while the other half performed an “X” vs. “O” categorization. If top-down processes modulate category-specific processes at the level where the AS information is processed, then animal-vehicle ERP differences should appear only for images with intact AS and only in the animal-vehicle categorization task. In contrast, if top-down processes modulate visual processes at the level of visual representations (Linnert et al., in prep) then the task should similarly affect the ERP category differences for images with intact and averaged AS. In other worlds, we are interested whether top-down and bottom-up processes interact at lower or higher levels during visual categorization.

Methods

Participants

Thirty-five university students (16 males, 30 right handed), ranging in age from 18 to 33 years, took part in the experiment and received course credits for their participation. Participants had normal or corrected to normal vision. Five additional participants’ data were discarded because of vision problems (N = 2) or because they did not provide enough artifact-free segments (N = 3). Participants were informed about the experiment before they volunteered and signed a consent form. Experiment and procedure were approved by the Lancaster University Ethics Commitee.

Stimuli and materials

The stimuli were greyscale images of easily recognizable animals (n = 150) and vehicles (n = 150) appearing on a grey background. The images represented a diverse range of animals (e.g. fish, insects, mammals, birds etc.), and vehicles (e.g. cars, buses, planes, ships etc.) and they varied in their size and viewpoint angle. Next to the objects an "X" or an "O" appeared somewhere on the grey background (see figure 1). The "X" and "O" was positioned manually on the image close the object. The size of the images was kept constant: W10.71° x H8.38° (W506 x H396 pixels).

For one version of the stimulus set the Fourier amplitude spectrum was kept intact and only the luminance was equalized across the images (intact AS), while for the second version of stimuli both the luminance and Fourier amplitude spectrum was equalized across the images (averaged AS) (see figure 1). The MATLAB based SHINE toolbox (Willenbockel et al., 2010) was used for luminance and spectrum matching. For all images, the luminance histograms of the images were matched by using the average luminance distribution across the whole stimuli set. For the images with averaged AS, we first applied the luminance histogram matching, followed by the amplitude spectrum matching. These two matching processes can distort each other, therefore, we modified the images in 20 iterations (for details see Willenbockel et al., 2010).

Overall, 600 images (150 animals and 150 vehicles both with intact and averaged AS) were presented to the participants. Stimuli were shown on a 15" CRT monitor with 85 Hz refresh rate and 1024 by 768 pixels resolution. Stimuli were presented using E-Prime software.

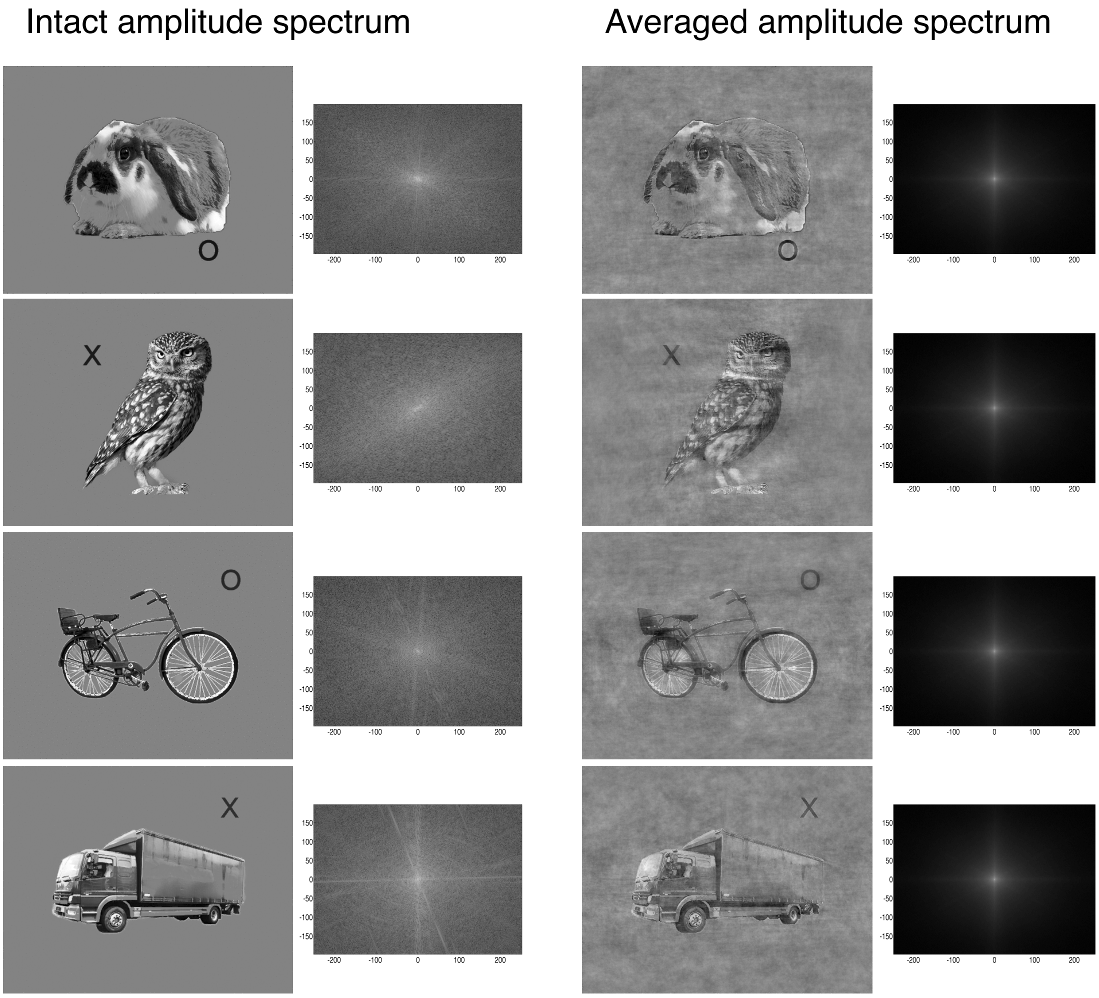


Figure 1. Example stimuli. Images of animals or vehicles containing an X or an O with intact (left) or averaged (right) amplitude spectrum.

Experimental procedure

Participants sat on a chair approximately 80 cm from the screen. The stimulus presentation was around 30 minutes long with a break after every 100 stimuli (around every 5 minutes). The break was as long as the participants wanted and the test proceeded when they pressed a button. The order of the presented stimuli was randomized. The participants performed a two-choice task to categorize the presented images. All participants saw the same images; however, they were randomly assigned to one of two groups. One group had to decide whether the presented image was an animal or a vehicle

(AV task, N = 18), while the other group had to report whether an “X” or an “O” was in the background of the object (XO task, N = 17).

A trial started with a fixation cross displayed in the middle of a grey screen for a random duration between 200 and 350 ms. Then a stimulus was shown in the middle of the grey screen until the participant responded by pressing a button (but for a maximum of 2000 ms). At the end of each trial a blank grey screen was shown for 1500 ms. Participants’ category choices were made by using two buttons on a response box. Within each participant group the left and right buttons were counterbalanced across the participants. Response accuracy was collected together with the recorded EEG.

EEG recording and analysis

We used an EGI (Electrical Geodesics Incorporated, Eugene, OR) GES300 EEG system with a NetAmps USB amplifier and a 128 channel HydroCel Net to record the EEG from the scalp. The EEG was recorded and analyzed with the NetStation software (Electrical Geodesics Incorporated, Eugene, OR). Electrode impedance was kept below 50 kOhm (as is standard for the EGI system). During the recording, the EEG was referenced to Cz, and the sampling rate was 1000 Hz.

During off-line analysis, the raw data were band-pass filtered between 0.1 and 30 Hz. The EEG was segmented from 150 ms before to 800 ms after stimulus onset. Epochs were sorted into the object categories (animals or vehicles) and stimulus types (intact AS or averaged AS). Only epochs containing correct responses were included. Eye movements and bad channels were automatically rejected if the average amplitude in an 80 ms moving window exceeded ±55 μV at EOG channels or ±200 μV at any other channel. In addition, each individual epoch was visually inspected and further epochs or channels were rejected. Epochs in which more than 13 channels contained artifacts were rejected; if fewer than 13 channels contained artifacts, bad channels were substituted by spline interpolation. Participants with fewer than 30 artifact-free epochs per condition were excluded.

Average ERPs were calculated from the artifact-free segments for each stimulus condition. ERPs were baseline corrected using a 150 ms pre-stimulus baseline, and were re-referenced to the average reference. Grand averages were generated using the MATLAB-based EEGLAB (*v.* 13_4_4b) toolbox (Delorme & Makeig, 2004).

Results

Behavioral results

Accuracy and reaction times were recorded during the experiment. Statistical analysis was performed on the ratio of correct responses (defined in %) and on the mean reaction times (defined in ms, excluding responses that were faster than 200 ms and slower than 1000 ms). Within subject factors were the STIMULUS TYPE (intact AS, averaged AS), CATEGORY (animal, vehicle), the between subject factor was the TASK (animal-vehicle categorization, X-O categorization). A Greenhouse-Geisser correction was applied on the p-values for violations of the assumption of sphericity. Bonferroni corrections were applied for multiple comparisons.

The task modulated the ratio of correct responses between images with intact and averaged AS (STIMULUS TYPE * TASK interaction: $F(1, 33) = 5.88$, $p = .02$, $\eta p^2 = .15$). Participants were more accurate for the stimuli with intact AS (M = 97.65%, SE = 0.27%) compared to the images with averaged AS (M = 96.71%, SE = 0.50%) in the animal-vehicle task ($t(16) = 2.63$, $p = .02$). In the X-O task the correct responses did not differ between the stimulus types (intact AS: M = 96.93%, SE = 0.28%; averaged AS: M = 97.13%, SE = 0.34%).

Reaction times were faster for the images with intact AS (*M* = 473.58ms, *SE* = 9.61ms) than for the stimuli with averaged AS (*M* = 489.89ms, *SE* = 9.54ms) (main effect of STIMULUS TYPE: *F*(1, 33) = 79.09, *p*<.01. , $\eta p^2 = .71$). In addition, we also found a CATEGORY * TASK interaction (*F*(1, 33) = 9.06 , *p* < .01, $\eta p^2 = .22$). In the animal-vehicle task participants were faster for the vehicles (*M* = 486.86 ms, *SE* = 11.25 ms) than for the animals (*M* = 499.65 ms, *SE* = 10.10 ms) (*t*(16) = 2.75, *p* = .01). No such difference was found in the X-O task (animal: *M* = 468.68 ms, *SE* = 15.98 ms; vehicle: *M* = 471.75 ms, *SE* = 15.48 ms).

Electrophysiology results

Visual ERPs were analyzed on the occipital and temporal electrodes including electrodes O1, O2, Oz, T5, T6, 59, 65, 66, 67, 71, 72 76, 77, 84, 90, 91. The selection of these electrodes is in line with previous studies (e.g. Balas & Koldewyn, 2013; Bentin, Allison, Puce, Perez, & McCarthy, 1996; Proverbio, Del Zotto, & Zani, 2007; Rossion & Jacques, 2008). Figure 2 shows the scalp distribution of the animal-vehicle ERP differences at the latencies of 50, 75, 100, 125, 150, 175, 200, 225, 250 and 275 ms for images with intact and averaged AS in the animal vs. vehicle (AV task) and in the X vs. O (XO task) tasks. In order to investigate the spatial distribution of the animal-vehicle ERP differences, we performed permutation-based paired tests (p<0.05) with correction for multiple comparisons (false discovery rate) using the MATLAB-based EEGLAB (v. 13_4_4b) toolbox (Delorme & Makeig, 2004). This was done for all selected electrodes at each of the latencies specified above (see Figure 2). The red dots on figure 2 indicate the electrodes with significant animal-vehicle ERP differences at the latencies of the corresponding scalp-maps.

In order to gain additional information about the time-scale of the animal-vehicle ERP differences, we also performed point-by-point tests (permutation with $p<0.05$ and false discovery rate) comparing the ERPs for animals and vehicles at all data points between -150 and 400 ms for images with intact and averaged AS, in the AV task and in the XO task (EEGLAB; Delorme & Makeig, 2004). For this analysis, we averaged ERPs for the occipital and for the temporal electrodes over the left and right hemisphere, respectively (see figure 3 - note that the midline electrodes are not included in this analysis). The distinction between the electrode groups was made based on the ERP waveforms and latencies of the ERP peaks. Results of the point-by-point analysis are shown on Figure 4.

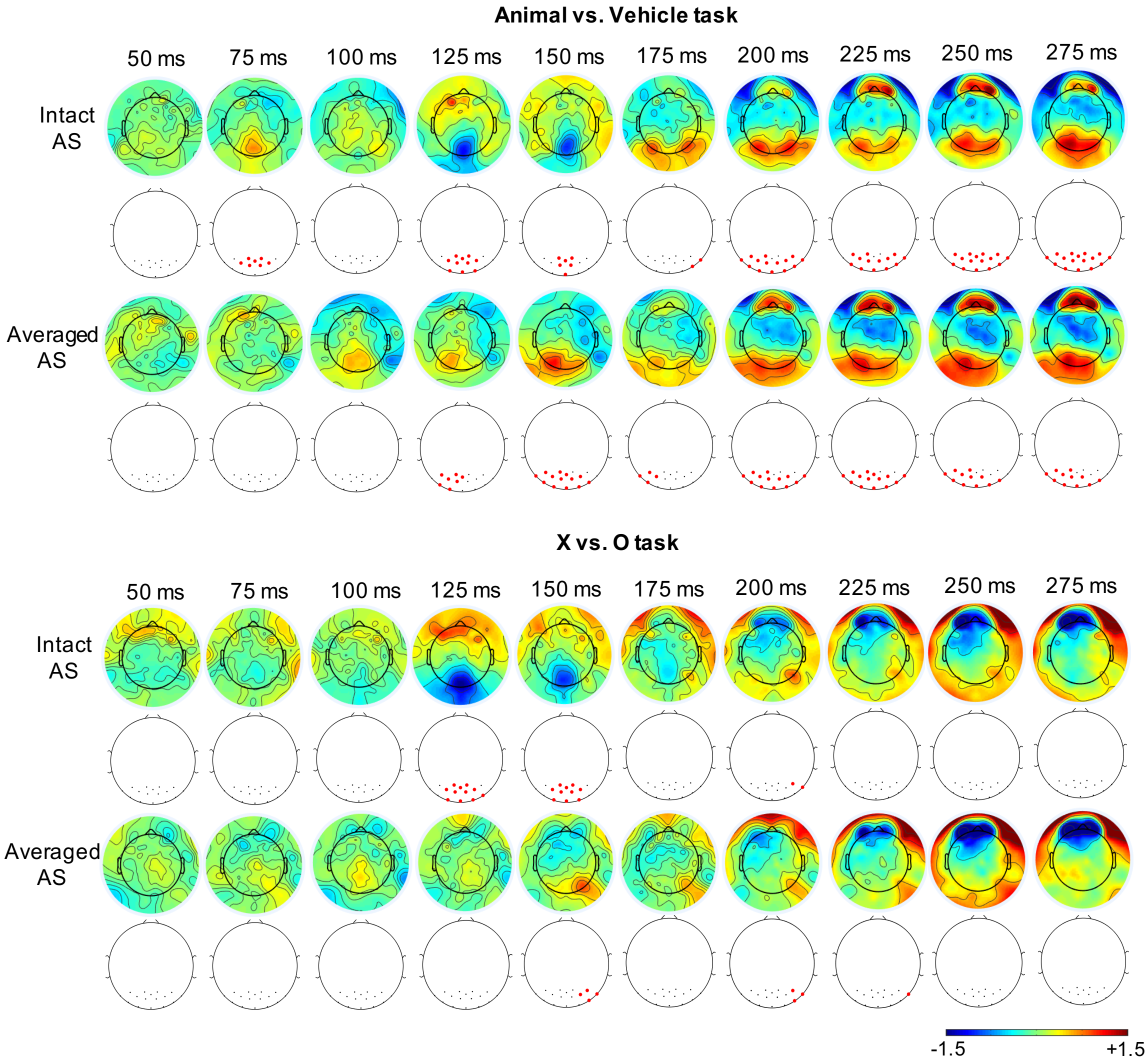


Figure 2. Scalp distribution of the animal-vehicle ERP differences between 50 and 275 ms in the animal vs. vehicle (top rows) and X vs. O task (bottom rows) for images with intact and averaged AS, respectively. The red dots indicate the electrodes with significant animal-vehicle ERP differences at the latency of the corresponding scalp-map. Note that 75 ms, 125-150 ms, 200 ms and 250-275 ms are the peak latencies of the P1, N1, P2 and N2 components, respectively.

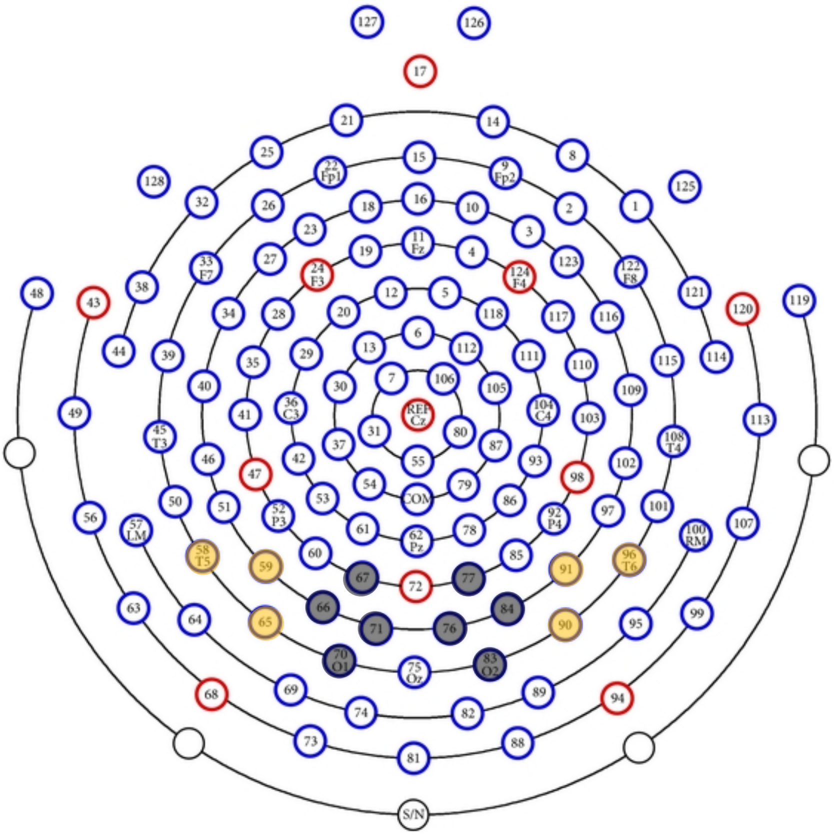


Figure 3. Electrode layout. Occipital electrodes selected for analysis are marked with grey, temporal electrodes are marked with yellow.

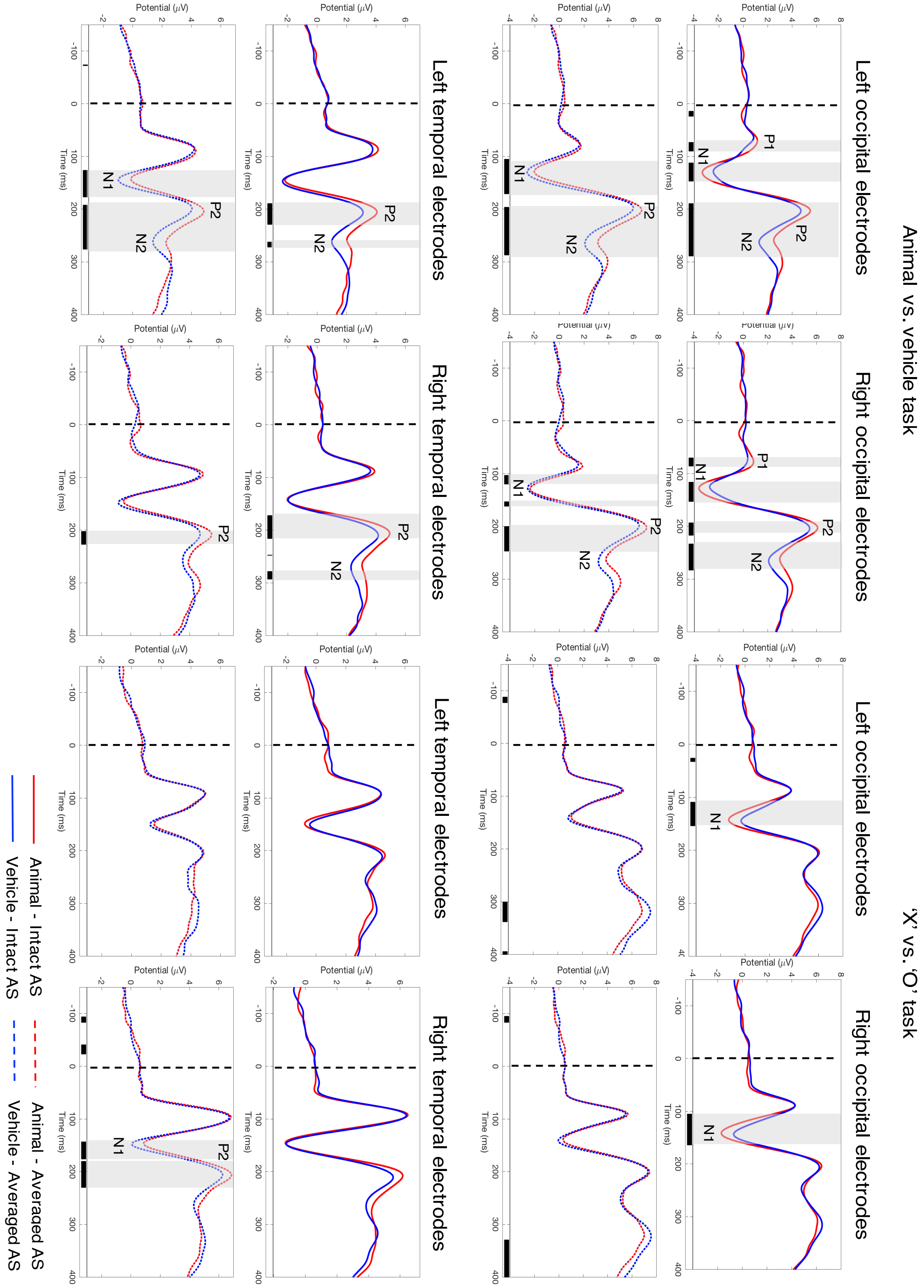


Figure 4. Results of the point-by-point analysis. The continuous lines represent images with intact AS, the dashed lines represent images with averaged AS. The red lines represent animals, the blue lines represent vehicles. Paired analyses were performed on the average waveforms of four occipital electrodes and on the average waveforms of three temporal electrodes over both the left and right hemispheres, separately. Averaged ERPs are presented for the left and right occipital areas (top two rows) and for the left and right temporal areas (bottom two rows) in the AV task (left columns) and in the XO task (right columns). Grey rectangles show the time-windows of the significant differences between the animal and vehicle ERPs.

As it is not sufficient to conclude a difference by only looking at the significant category difference pattern (Nieuwenhuis, Forstmann, & Wagenmakers, 2011) we also examined the interactions between the animal-vehicle ERP differences, the task, and/or the stimulus type. Mean amplitudes were extracted for time windows 70-90 ms (corresponds to P1), for 110-150 ms (corresponds to N1) and for 190-300 ms (corresponds to P2 and N2) over the occipital. Over the temporal area, the mean amplitudes were extracted for time windows 140-180 ms (corresponds to N1), for 190 - 230 ms (corresponds to P2) and for 260 - 290 ms (corresponds to N2). These time windows are also in line with previous literature (Luck, 2005). Similarly to the point-by-point analysis presented above, the occipital area included electrodes O1, 71, 66, 67 over the left hemisphere and O2, 76, 77, 84 over the right hemisphere, and the temporal area included electrodes T5, 59 and 65 over the left hemisphere and electrodes T6, 90, 91 over the right hemisphere. Average ERPs are presented on Figure 5. We conducted 2x2x2x2 mixed ANOVAs with the within subject factors STIMULUS TYPE (intact AS, averaged AS), HEMISPHERE (left, right), and CATEGORY (animal, vehicle), and the between subject factor TASK (AV task, XO task) separately for occipital and temporal locations. A Greenhouse-Geisser correction was applied on the p-values for violations of the assumption of sphericity. Bonferroni corrections were applied for multiple comparisons.

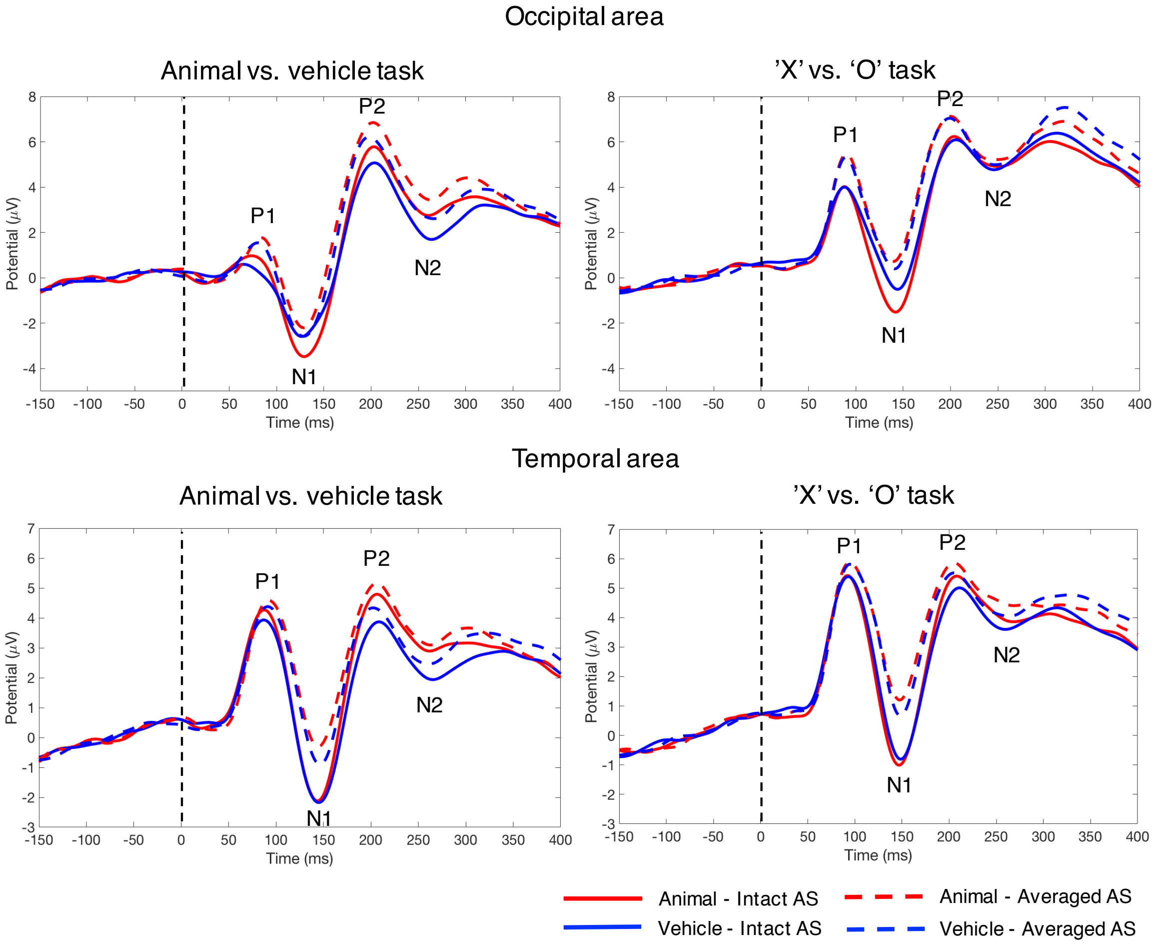


Figure 5. ERP waveforms for the Animal vs. Vehicle (left column) and “X” vs. “O” (right column) categorization task over the occipital (top row) and temporal areas (bottom row) respectively. The continuous lines represent images with intact AS, the dashed lines represent images with averaged AS. The red lines represent animals, the blue lines represent vehicles. ERPs show the averaged waveforms across electrodes and hemispheres.

P1

Figure 2 shows significant animal vehicle differences at 75 ms after stimulus onset on the occipital electrodes, but only for images with intact AS and only in the AV task. Similarly, Figure 4 shows category differences over the left and right occipital area for images with intact AS in the AV task between 70 - 90 ms after stimulus onset. The ANOVA on the mean amplitudes for this time window confirmed the interaction between CATGEORY, STIMULUS TYPE and the TASK ($F(1, 33) = 6.56$, $p = .02$, $\eta p^2 = .17$). The post-hoc t-tests showed category differences only for the images with intact AS in the animal-vehicle task (*t*(16) = 4.37, *p* < .001). The P1 amplitudes were higher for the animals (*M* = .85 µV, *SE* = .79 µV) than for the vehicles (*M* = .28 µV, *SE* = .82 µV) (see Figure 5 and 6a). The interaction between the task and the category indicate that the task modulated the animal-vehicle differences in the time-window of the P1 component. However, this effect was present only for images with intact AS, suggesting that these early ERP differences are due to the amplitude spectrum information. In other words, the animal-vehicle categorization task had a selective effect on the AS information. Therefore, the three-way interaction suggests that the task, as a top-down factor affected the visual processes at an early stage where the AS information was processed.

N1

Figure 2 and 4 show animal vehicle differences for all stimulus types in both tasks both on occipital and temporal electrodes at the time window of the N1 component. Over the occipital areas, we found an interaction between CATEGORY and STIMULUS TYPE (*F(1, 33)* = 44.37*, p* < .001*,* ηp^2 = .57). The N1 amplitudes were higher for animals (*M* = -1.70 µV, *SE* = .74 µV) compared to vehicles (*M* = -.78 µV, *SE* = .69 µV) for images with intact AS (*t*(34) = 6.40, *p* < 001). Conversely, vehicles evoked higher amplitudes (*M* = -.01 µV, *SE* = .77 µV) compared to animals (M = -.36 µV, SE = .81 µV) for images with averaged AS (*t*(34) = -3.03, *p* = .005) (see figure 5 and 6b). These results are in line with the results of the paired analyses. The different direction of the amplitude difference for animals and vehicles between images with intact and average amplitude spectrum suggests that the presence or absence of amplitude spectrum information did affect the visual information processes reflected in the N1.

Over the temporal areas, Figure 2 and 4 show animal-vehicle differences only for images with averaged AS. However, the results of the ANOVA showed only a main effect of CATEGORY ($F(1, 33) = 10.04$, $p = .003$, $\eta p^2 = .23$). The interaction between CATEGORY and STIMULUS TYPE was only marginally significant ($F(1, 33) = 3.97$, $p = .055$, $\eta p^2 = .11$) (Figure 6c).

P2 and N2

Over the occipital area mean amplitudes for the P2 and N2 components were extracted together between 190 ms and 290 ms after stimulus onset. We found an interaction between CATEGORY and TASK ($F(1, 33) = 8.49$, $p = .006$, $\eta p^2 = .21$). Animals elicited more positive amplitudes ($M = 3.16\ \mu V$, $SE = .88\ \mu V$) then vehicles ($M = 2.40\ \mu V$, $SE = .77\ \mu V$) ($t(16) = 3.66$, $p = .002$) in the AV task. In the XO task, amplitudes of animals ($M = 5.69\ \mu V$, $SE = .63\ \mu V$) and vehicles ($M = 5.62\ \mu V$, $SE = .66$) μV did not differ ($t(16) = .64$, $p = .53$) (Figure 2, 4 & 7a).

Over the temporal area, mean amplitudes were extracted separately for the time window of the P2 and N2 components according to the results of the point-by-point analysis. We found significant interaction between CATGEORY AND TASK for both the P2 ($F(1, 33) = 6.64$, $p = .015$, $\eta p^2 = .17$) and N2 ($F(1, 33) = 5.61$, $p = .024$, $\eta p^2 = .15$) components. For the P2 component animal - vehicle amplitude differences were bigger in the AV task ($M = .80\ \mu V$, $SE = .12\ \mu V$) then in the XO task ($M = .36\ \mu V$, $SE = .12\ \mu V$), however, they were significant in both tasks (AV task: $t(16) = 6.70$, $p < .001$; XO task: $t(17) = 3.05$, $p = .007$). For the N2 component, amplitudes of animals and vehicles differed only in the AV task ($M = .77\ \mu V$, $SE = .21\ \mu V$, $t(16) = 3.67$, $p = .002$), but not in the XO task ($M = .18\ \mu V$, $SE = .14\ \mu V$, $t(16) = 1.29$, $p = .21$) (Figure 2, 4, 7b, 7c).

The pairwise comparisons showed animal-vehicle differences also in the XO task for the P2 components over the temporal area (see Figures 2, 4). Nevertheless, the P2 interaction between CATEGORY and TASK and the post-hoc tests showed that the P2 animal-vehicle amplitude differences are smaller in the XO task, compared to the AV task. Unlike for the P1 component, for the P2 and N2 the task-related top-down modulation was irrespective of the presence or absence of the AS information. In the AV task, animal-vehicle differences were present for images both with intact and averaged AS (see pared analyses, figure 2 and 4). The effect of low-level physical features on the P1 and N1 category differences do not seem to affect the category differences in the P2 and N2 components. This suggests that the P2 and N2 components reflect a different stage of the processing. These components might reflect the activation of higher-level category representations, where low-level features no longer modulate visual processes. Moreover, the task-related modulation in the P2 and N2 component suggests that decision-making processes also affect these components.

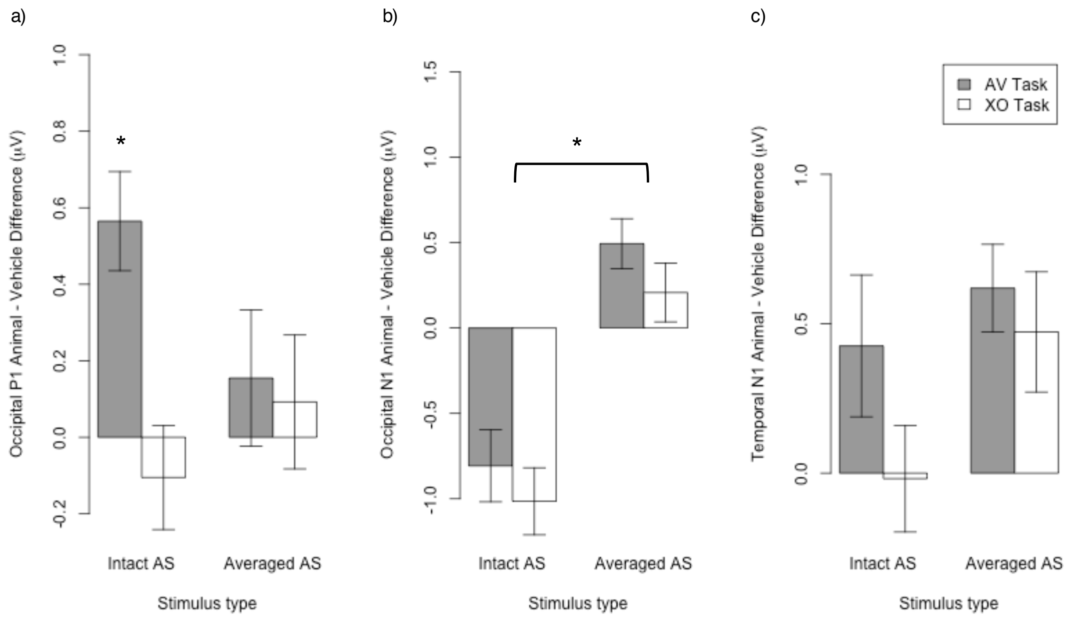


Figure 6. Mean amplitude differences between ERPs for animals and vehicles in the time window of the occipital P1 (a), occipital N1 (b) and temporal N1 (c) components for images with intact and averaged AS in the AV and XO tasks. Error bars represent SE.

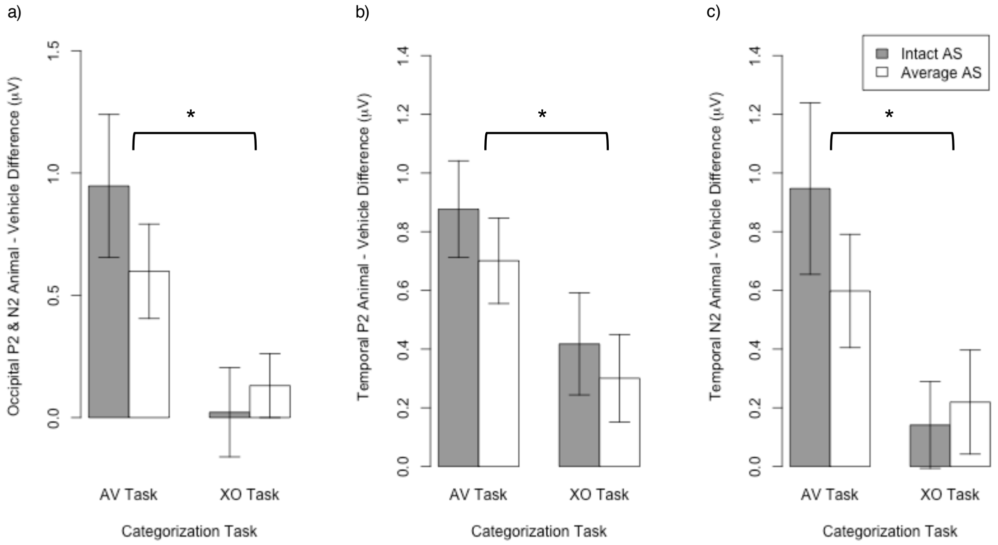


Figure 7. Mean amplitude differences between ERPs for animals and vehicles in the time window of the occipital P2 & N2 (a), the temporal P2 (b) and temporal N2 (c) components for images with intact and averaged AS in the AV and XO tasks. Error bars represent SE.

Comparison across components

In order to compare the animal-vehicle ERP differences across components, we performed an additional ANOVA including COMPONENT (occipital: P1, N1, P2 & N2; temporal: N1, P2, N2) as a within subject factor beside STIMULUS TYPE (intact AS, averaged AS), HEMISPHERE (left, right), and CATEGORY (animal, vehicle), and the between subject factor TASK (AV task, XO task). Planned comparisons were performed for P1-N1 and N1-P2&N2 component pairs over the occipital area, and for N1-P2 and P2-N2 pairs over the temporal area.

Over the occipital area, we found an interaction between COMPONENT, STIMULUS TYPE, CATEGORY and HEMISPHERE (*F(2, 66*) = 3.50, *p* = .036, ηp^2 = .10). For images with intact AS, the animal-vehicle difference differed between components, but not between hemispheres (COMPONENT * CATEGORY: *F(2, 68*) =

35.00, $p < .001$, $\eta p^2 = .51$). The category differences were significant for the N1 (*M* = -.91 µV, *SE* = .14 µV; *t(34)* = -6.40, $p < .001$) and P2 & N2 (*M* = .47 µV, *SE* = .19 µV; *t(34)* = 2.54, *p* = .048) components, and not significant for the P1 (*M* = .22 µV, *SE* = .11 µV; *t(34)* = 2.02 , *p* = .153). The animal-vehicle difference differed between the P1- N1 (*F(1, 34)* = 41.89, $p < .001$, $\eta p^2 = .55$) and between the N1 – P2&N2 (*F(1, 34)* = 53.43, $p < .001$, $\eta p^2 = .61$) components, possible because of change in direction and not due to change in magnitude (figure). For images with averaged AS we found only a main effect of CATEGORY (*F(1, 34)* = 8.38, *p* = .007, $\eta p^2 = .198$). However, the planned comparisons showed that animal-vehicle differences differed between the P1 and N1 (*F(1, 34)* = 4.37, *p* = .044, $\eta p^2 = .114$).

The animal-vehicle difference was modulated by the task over the occipital area (CATEGORY * TASK: *F(1, 33)* = 9.07, *p* = .005, $\eta p^2 = .22$). The category difference was larger and significant only in the AV task (*M* = .33 µV, *SE* = .11 µV; *t(34)* = 2.87 , *p* = .022) compared to the XO task (*M* = -.11 µV, *SE* = .09 µV; *t(34)* = -1.21, *p* = .24).

Over the temporal area, we found a three-way interaction between COMPONENT, STIMULUS TYPE and CATEGORY (*F(2, 66)* = 8.61, $p < .001$, $\eta p^2 = .21$). For images with intact AS, the animal-vehicle difference increased (planned contrast for N1-P2: *F(1, 34)* = 21.08, $p < .001$, $\eta p^2 = .38$) between the N1 (*M* = .20 µV, *SE* = .15 µV; *t(34)* = 1.32, *p* = .59) and P2 (*M* = .64 µV, *SE* = .13 µV; *t(34)* = 5.14, $p < .001$) component, and it did not differ (planned contrast for P2-N2: *F(1, 34)* = .72, *p* = .40, $\eta p^2 = .021$) between the P2 and N2 (*M* = .53 µV, *SE* = .17 µV; *t(34)* = 3.08, *p* = .012) components. For images with averaged AS, animal vehicle differences were significant for all three components (N1: *M* = .54 µV, *SE* = .12 µV; *t(16)* = 4.73, $p < .001$; P2: *M* = .50 µV, *SE* = .11 µV; *t(16)* = 4.57, $p < .001$; N2: *M* = .40 µV, *SE* = .13 µV; *t(16)* = 3.04, *p* = .015), and did not differ between the N1-P2 (*F(1, 34)* = .26, *p* = .61, ηp^2 = .01) and P2-N2 (*F(1, 34)* = .79, *p* = .38, ηp^2 = .02) component pairs.

Over the temporal area, we also found an interaction between CATEGORY and TASK (*F(1, 33)* = 5.80, *p* = .022, ηp^2 = .149). The animal-vehicle difference was larger and significant only in the AV task (*M* = .70 µV, *SE* = .14 µV; *t(16)* = 5.03 , *p* < .001), compared to the XO task (*M* = .26 µV, *SE* = .12 µV; *t(34)* = 2.13 , *p* = .10).

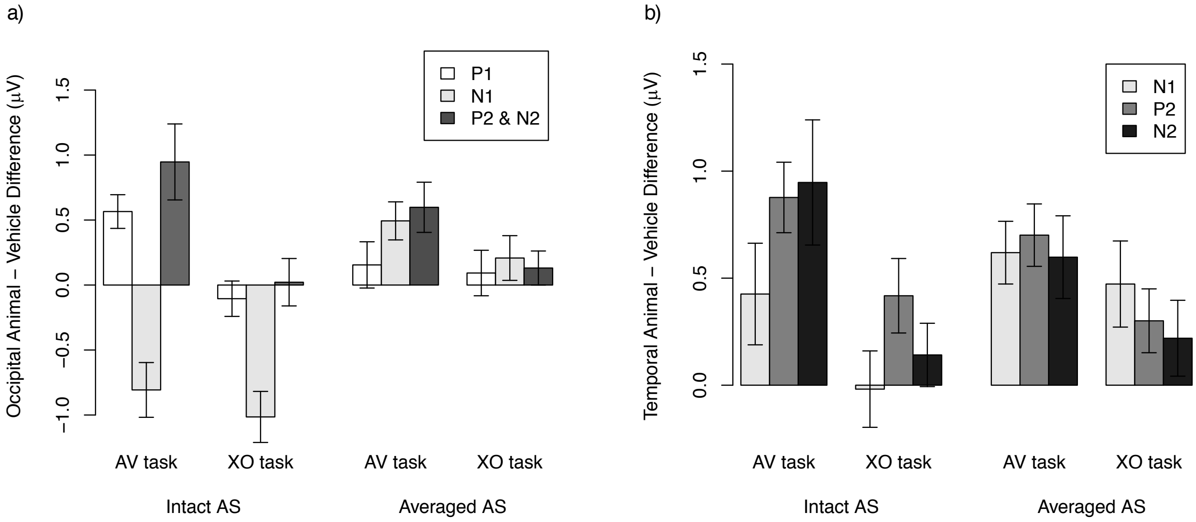


Figure 8. Differences between animal and vehicle amplitudes by components over the occipital (a) and temporal (b) areas. Error bars represent SE.

Discussion

The aim of this study was to investigate the underlying mechanisms of task-related top-down modulation of the animal vs. vehicle category differences in the visual ERPs and the role of amplitude spectrum (AS) information in object categorization. Our findings show that the AS information plays a role in the early stages of categorization and it modulates the category differences over the occipital areas. The later stages of categorization, where category differences in P2 and N2 occurred over the temporal areas, on the other hand, were not affected by the presence or absence of AS information. Furthermore, both stages of categorization were affected by top-down, task-related factors. An amplitude difference in P1 between animals and vehicles was present only when AS information was available, and only when the categorization task was to distinguish animals from vehicles and not to decide whether an X or an O was present in the image. The task effect in the P2 and N2 components, on the other hand, was independent of the physical image characteristics. The animal-vehicle ERP differences were bigger in the animal vs. vehicle task than in the X vs. O task irrespective of whether the AS information was intact or it was averaged across stimuli.

Our results suggest that the early and later animal-vehicle differences in the visual ERPs are elicited by different factors. The P1 category difference in our study is clearly driven by the AS information and it appeared at the occipital electrodes. The P2 and N2 category differences, on the other hand, were not modulated by the presence or absence of the AS information and they appeared at the temporal electrodes. These results are in line with the results of Bieniek, Pernet, & Rousselet (2012). They found ERP difference between phase-scrambled faces and houses, indicating the effect of AS information; however, this difference was more medially distributed, appeared earlier in time, and was weaker than the ERP differences driven by the phase (i.e., object shape) information. The phase information was present in all our stimuli, therefore, in line with other studies (Bieniek et al., 2012; Gaspar & Rousselet, 2009) it is very likely that the latter category differences in the P2 and N2 are driven by the presence of a recognizable shape.

The important role of phase information in visual processing has been linked to the shape selectivity in the IT (Gaspar & Rousselet, 2009). It has been suggested that only individual shapes (objects) are represented in the IT and these shape representations are topographically organized by perceptual similarity (Kriegeskorte et al., 2008). Therefore, category-specific processes might not reflect real category representations but only perceptual differences between object categories used in those experiments (Kriegeskorte et al., 2008). However, usually basic level categories share similar shapes (e.g. cars, birds) but not vehicles and animals; the categories we used in our study. Nevertheless, Wilder, Feldman and Singh (2011) have shown that superordinate level (in their case, animal vs. leaf) categorization can be modeled by using the structure of object shapes. Therefore, the phase information can play a role also in superordinate category discrimination. Furthermore, it has been shown that besides shapes, some categories (such as animate and inanimate objects) are also represented in the IT (Kaiser, Azzalini, & Peelen, 2016; Kriegeskorte et al., 2008; Mur et al., 2013). Therefore, the P2 and N2 amplitude differences for animals and vehicles in our study might reflect the activation of these category representations in the IT.

Regarding the top-down task effect in the P2 and N2 components, it has been shown that shape or category selective areas in the IT can also be activated by top-down processes even in the absence of incoming visual information. For instance, O'Craven & Kanwisher (2000) found activation in face-selective and place-selective areas of the IT when participants were imagining faces and buildings, while Reddy, Tsuchiya, & Serre (2010) reported similar findings for categories of faces, houses, tools and fruits. In addition, anticipatory cues (such as symbols or written category names) before the stimulus presentation also activate category-selective areas, and such preparatory activation in the IT seems to facilitate participants' categorization performance (Peelen & Kastner, 2011; Puri et al., 2009). Furthermore, an anticipatory cue that precisely predicts the subsequent visual stimulus leads to increased activity in category-selective areas during the processing of the visual stimulus (Puri et al., 2009). In our study, participants expected to see either an animal or a vehicle; therefore, such preparatory activity might explain the task-related P2 and N2 modulation.

In line with previous results (Gaspar & Rousselet, 2009; Nemeth et al., 2014; Bruno Rossion & Caharel, 2011), our findings suggest that the AS information affects category-specific ERP differences, at least in the early components. However, the task-dependency of this effect indicates that AS information is not automatically accessed, but is subject to top-down selection. It has been shown that other low-level level features, such as color, can also be accessed by top-down processes (Linnert et al, in prep). Furthermore, Peelen and Kastner (2011) have reported preparatory activity in the early visual areas (V1), and they have suggested that such activity prepares for the detection of low-level category-specific cues. The task-related animal-vehicle difference in the P1 might indicate similar preparatory activity for the detection of the category-specific AS information.

Whether the AS information is part of the category representation of objects has been debated. Andrews, Clarke, Pell and Hartley (2010) have suggested that higher-level category selectivity is, at least partly, based on the processing of selective low-level features (such as the AS). In this case, activity in the lower-level areas should facilitate higher-level activity and eventually behavioral responses. Indeed, in our study participants categorized images as animals or vehicles with higher accuracy when the images contained the original AS information compared to images with averaged AS. Furthermore, the task-related top-down modulation of the AS information in our study, indicates that the visual system does associate AS and category.

Interestingly, task-related modulations in previous studies have been found in the N1 component (Balas & Koldewyn, 2013; Linnert et al., in prep), whereas we found such effects in the P1, P2 and N2 components. Moreover, the P2 and N2 components were not even visible on the ERP waveforms in these two previous studies. The stimuli used in the current study differed from the stimuli used in the studies by Balas and Koldewyn (2013) and Linnert et al. (in prep.). This difference between the stimulus sets could account for the differences between the ERP waveforms reported in these studies. Linnert et al (in prep.) used colorful (and greyscale) images of cars and birds without any low-level images modification. Balas & Koldewyn (2013) used homogenous image categories: participants saw only sixteen images repeatedly, depicting faces of humans and dogs. In our current study, on the other hand, the images were all greyscale, were matched for low-level features, and animals and vehicles belong to heterogeneous categories. The visual system is tuned to process natural stimuli; therefore, it might need different or additional steps to process artificial, greyscale, luminance and AS equalized images. Recently, it has been shown that the shape selective regions of the IT accumulate and integrate evidence for decision-making functions (Tremel & Wheeler, 2015). Natural images contain several cues (e.g. color, luminance, shape), therefore, this evidence gathering process can be quicker and easier for images containing those cues than for images where such features are not present. Possibly, this additional effort or processing mechanism can be shown up as additional components on the visual ERPs. Indeed, it has been suggested that competing or insufficient information in the higher visual areas get reinforcement by re-activating the early visual areas by feedback projections (Di Lollo et al., 2000) and that the P2 component reflects this re-activation process (Kotsoni et al., 2007). This is in line with the view that perception relies on highly dynamic processes, and top-down information also plays a flexible role on these processes (e.g. Ahissar & Hochstein, 2004; Contini et al., 2017; Gilbert & Sigman, 2007). Contini et al. (2017) have also recently pointed out that the time-course of the visual processes, at least partly, depends on the characteristics of the stimulus set used in an experiment.

Overall, the results of this study are in line with previous results reporting higher-level category differences at the temporal electrodes, and differences driven by low-level physical features at the occipital electrodes (Bieniek et al., 2012; for review see Contini et al., 2017). Furthermore, our results indicate that the AS information, as a low-level feature, is processed earlier in time (at the time of the P1 component). In contrast, higher-level category- or phase-specific differences arise later in time (at the time of the P2 and N2 components). This is in line with previous findings showing that sensitivity to low-level features emerges before category-like representations (for review see Contini et al., 2017). Regarding the top-down modulation, we found an early effect of task in the P1, suggesting that the task affected the processing of the AS information. The later task effects in the P2 and N2 were present for images with both intact and averaged AS, suggesting that this later top-down effect modulated higher-level processes, possibly at the level of category representations.

In conclusion, these results suggest that the AS, as a low-level factor affects processes in the early stages of categorization. ERP differences due to the presence of AS information were present only in the earliest ERP component (P1) and only at the occipital electrode locations. The later categorization processes more likely rely on category representations stored in the higher-level visual areas. Furthermore, top-down processes can access and modulate both low-level and higher-level information processing. The differences in the timing of the task-related modulation between the current study and previous studies, and the presence of additional ERP components in this study, suggests that the same ERP component might reflect different underlying processes depending on the stimuli used in the study. A cautious approach should be adopted in defining ERP components in terms of underlying functions.

References

Ahissar, M., & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends in Cognitive Sciences*, *8*(10), 457–464. http://doi.org/10.1016/j.tics.2004.08.011

Andrews, T. J., Clarke, A., Pell, P., & Hartley, T. (2010). Selectivity for low-level features of objects in the human ventral stream. *NeuroImage*, *49*(1), 703–711. http://doi.org/10.1016/j.neuroimage.2009.08.046

Antal, A., Kéri, S., Kovács, G., Janka, Z., & Benedek, G. (2000). Early and late components of visual categorization: an event-related potential study. *Cognitive Brain Research*, *9*(1), 117–119. http://doi.org/10.1016/S0926-6410(99)00053-1

Antal, A., Kéri, S., Kovács, G., Liszli, P., Janka, Z., & Benedek, G. (2001). Event-related potentials from a visual categorization task. *Brain Research Protocols*, *7*, 131–136. http://doi.org/10.1016/S1385-299X(01)00055-1

Bacon-Macé, N., Macé, M. J. M., Fabre-Thorpe, M., & Thorpe, S. J. (2005). The time course of visual processing: Backward masking and natural scene categorisation. *Vision Research*, *45*, 1459–1469. http://doi.org/10.1016/j.visres.2005.01.004

Balas, B., & Koldewyn, K. (2013). Early visual ERP sensitivity to the species and animacy of faces. *Neuropsychologia*, *51*(13), 2876–2881. http://doi.org/10.1016/j.neuropsychologia.2013.09.014

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. http://doi.org/10.1162/jocn.1996.8.6.551

Bieniek, M. M., Pernet, C. R., & Rousselet, G. A. (2012). Early ERPs to faces and objects are driven by phase, not amplitude spectrum information: evidence from parametric, test-retest, single-subject analyses. *Journal of Vision*, *12*(13), 12. http://doi.org/10.1167/12.13.12

Chao, L. L., Haxby, J. V, & Martin, A. (1999). Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nature Neuroscience*, *2*(10), 913–919. http://doi.org/10.1038/13217

Codispoti, M., Ferrari, V., Junghöfer, M., & Schupp, H. T. (2006). The categorization of natural scenes: Brain attention networks revealed by dense sensor ERPs. *NeuroImage*, *32*, 583–591. http://doi.org/10.1016/j.neuroimage.2006.04.180

Contini, E. W., Wardle, S. G., & Carlson, T. A. (2017). Decoding the time-course of object recognition in the human brain: From visual features to categorical decisions. *Neuropsychologia*, *105*, 165–176. http://doi.org/10.1016/j.neuropsychologia.2017.02.013

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. http://doi.org/10.1016/j.jneumeth.2003.10.009

Delorme, A., Rousselet, G. A., Macé, M. J.-M., & Fabre-Thorpe, M. (2004). Interaction of top-down and bottom-up processing in the fast visual analysis of natural scenes. *Cognitive Brain Research*, *19*(2), 103–113. http://doi.org/10.1016/j.cogbrainres.2003.11.010

Di Lollo, V., Enns, J. T., & Rensink, R. a. (2000). Competition for consciousness among visual events: the psychophysics of reentrant visual processes. *Journal of Experimental Psychology. General*, *129*(4), 481–507. http://doi.org/10.1037/0096-3445.129.4.481

Gaspar, C. M., & Rousselet, G. A. (2009). How do amplitude spectra influence rapid animal detection? *Vision Research*, *49*(24), 3001–3012.

http://doi.org/10.1016/j.visres.2009.09.021

Gilbert, C. D., & Sigman, M. (2007). Brain states: Top-down influences in sensory processing. *Neuron*, *54*, 677–696. http://doi.org/10.1016/j.neuron.2007.05.019

Ishai, A., Ungerleider, L. G., Martin, A., Schouten, J. L., & Haxby, J. V. (1999). Distributed representation of objects in the human ventral visual pathway. *Proceedings of the National Academy of Sciences*, *96*(16), 9379–9384. http://doi.org/10.1073/pnas.96.16.9379

Johnson, J. S., & Olshausen, B. a. (2003). Timecourse of neural signatures of object recognition. *Journal of Vision*, *3*(7), 499–512. http://doi.org/10.1167/3.7.4

Kaiser, D., Azzalini, D. C., & Peelen, M. V. (2016). Shape-independent object category responses revealed by MEG and fMRI decoding. *Journal of Neurophysiology*, jn.01074.2015. http://doi.org/10.1152/jn.01074.2015

Kotsoni, E., Csibra, G., Mareschal, D., & Johnson, M. H. (2007). Electrophysiological correlates of common-onset visual masking. *Neuropsychologia*, *45*(10), 2285–2293. http://doi.org/10.1016/j.neuropsychologia.2007.02.023

Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., … Bandettini, P. A. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron*, *60*(6), 1126–1141. http://doi.org/10.1016/j.neuron.2008.10.043

Lerner, Y., Hendler, T., Ben-Bashat, D., Harel, M., & Malach, R. (2001). A hierarchical axis of object processing stages in the human visual cortex. *Cerebral Cortex*, *11*(4), 287–297. http://doi.org/10.1093/cercor/11.4.287

Linnert, S., Csifcsák, G., & Tompa, T. (in prep). Top-down modulation of animate-inanimate category differences in the visual N1.

Luck, S. J. (2005). *An introduction to the event-related potential technique*. Cambridge, Mass: MIT press.

Mur, M., Meys, M., Bodurka, J., Goebel, R., Bandettini, P. A., & Kriegeskorte, N. (2013). Human object-similarity judgments reflect and transcend the primate-IT object representation. *Frontiers in Psychology*, *4*, 1–22. http://doi.org/10.3389/fpsyg.2013.00128

Nemeth, K., Kovacs, P., Vakli, P., Kovacs, G., & Zimmer, M. (2014). Phase noise reveals early category-specific modulation of the event-related potentials. *Frontiers in Psychology*, *5*(April). http://doi.org/10.3389/fpsyg.2014.00367

Nieuwenhuis, S., Forstmann, B. U., & Wagenmakers, E.J. (2011). Erroneous analyses of interactions in neuroscience: a problem of significance. *Nature Neuroscience*, *14*(9), 1105–1107. http://doi.org/10.1038/nn.2886

O'Craven, K. M., & Kanwisher, N. (2000). Mental imagery of faces and places activates corresponding stiimulus-specific brain regions. *Journal of Cognitive Neuroscience*, *12*, 1013–1023. http://doi.org/10.1162/08989290051137549

Peelen, M. V, & Kastner, S. (2011). A neural basis for real-world visual search in human occipitotemporal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 12125–12130. http://doi.org/10.1073/pnas.1101042108

Pollux, P. M. J., Hall, S., Roebuck, H., & Guo, K. (2011). Event-related potential correlates of the interaction between attention and spatiotemporal context regularity in vision. *Neuroscience*, *190*, 258–269. http://doi.org/10.1016/j.neuroscience.2011.05.043

Proverbio, A. M., Burco, F., Del Zotto, M., & Zani, A. (2004). Blue piglets? Electrophysiological evidence for the primacy of shape over color in object recognition. *Cognitive Brain Research*, *18*(3), 288–300.

http://doi.org/10.1016/j.cogbrainres.2003.10.020

Proverbio, A. M., Del Zotto, M., & Zani, A. (2007). The emergence of semantic categorization in early visual processing: ERP indices of animal vs. artifact recognition. *BMC Neuroscience*, *8*, 24. http://doi.org/10.1186/1471-2202-8-24

Puri, A. M., Wojciulik, E., & Ranganath, C. (2009). Category expectation modulates baseline and stimulus-evoked activity in human inferotemporal cortex. *Brain Research*, *1301*, 89–99. http://doi.org/10.1016/j.brainres.2009.08.085

Reddy, L., Tsuchiya, N., & Serre, T. (2010). Reading the mind's eye: Decoding category information during mental imagery. *NeuroImage*, *50*(2), 818–825. http://doi.org/10.1016/j.neuroimage.2009.11.084

Rossion, B., & Caharel, S. (2011). ERP evidence for the speed of face categorization in the human brain: Disentangling the contribution of low-level visual cues from face perception. *Vision Research*, *51*(12), 1297–1311. http://doi.org/10.1016/j.visres.2011.04.003

Rossion, B., & Jacques, C. (2008). Does physical interstimulus variance account for early electrophysiological face sensitive responses in the human brain? Ten lessons on the N170. *NeuroImage*, *39*, 1959–1979. http://doi.org/10.1016/j.neuroimage.2007.10.011

Rousselet, G. A., Husk, J. S., Bennett, P. J., & Sekuler, A. B. (2008). Time course and robustness of ERP object and face differences. *Journal of Vision*, *8*(12), 1–18. http://doi.org/10.1167/8.12.3

Rousselet, G. a, Pernet, C. R., Bennett, P. J., & Sekuler, A. B. (2008). Parametric study of EEG sensitivity to phase noise during face processing. *BMC Neuroscience*, *9*, 98. http://doi.org/10.1186/1471-2202-9-98

Thierry, G., Martin, C. D., Downing, P., & Pegna, A. J. (2007). Controlling for interstimulus perceptual variance abolishes N170 face selectivity. *Nature Neuroscience*, *10*(4), 505–511. http://doi.org/10.1038/nn1864

Thorpe, S. J., & Fabre-Thorpe, M. (2001). Seeking categories in the brain. *Science*, *291*, 260–263. http://doi.org/10.1126/science.1058249

Tompa, T., & Sáry, G. (2010). A review on the inferior temporal cortex of the macaque. *Brain Research Reviews*, *62*(2), 165–182. http://doi.org/10.1016/j.brainresrev.2009.10.001

Torralba, A., & Oliva, A. (2003). Statistics of natural image categories. *Network)*, *14*(3), 391–412. http://doi.org/10.1088/0954-898X/14/3/302

Tremel, J. J., & Wheeler, M. E. (2015). Content-specific evidence accumulation in inferior temporal cortex during perceptual decision-making. *NeuroImage*, *109*, 35–49. http://doi.org/10.1016/j.neuroimage.2014.12.072

VanRullen, R. (2011). Four common conceptual fallacies in mapping the time course of recognition. *Frontiers in Psychology*, *2*, 1–6. http://doi.org/10.3389/fpsyg.2011.00365

VanRullen, R., & Thorpe, S. J. (2001). The time course of visual processing: from early perception to decision-making. *Journal of Cognitive Neuroscience*, *13*, 454–461. http://doi.org/10.1162/08989290152001880

Vogel, E. K., & Luck, S. J. (2000). The visual N1 component as an index of a discrimination process. *Psychophysiology*, *37*(2), 190–203. http://doi.org/10.1111/1469-8986.3720190

Wilder, J., Feldman, J., & Singh, M. (2011). Superordinate shape classification using natural shape statistics. *Cognition*, *119*(3), 325–340. http://doi.org/10.1016/j.cognition.2011.01.009

Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., & Tanaka, J. W. (2010).

Controlling low-level image properties: the SHINE toolbox. *Behavior Research Methods*, *42*(3), 671–684. http://doi.org/10.3758/BRM.42.3.671

Linking statement for Paper 2

In Paper 1 we investigated whether the amplitude spectrum (AS), as a low-level image characteristic contributes to the animate vs. inanimate category differences in the visual ERPs and whether the role of this information is modulated by the task. Stimuli consisted of animal and vehicle images appearing on a grey background. In order to enable multiple categorizations of the images, an “X” or an “O” appeared somewhere next to the object. Half of the participants performed an animal vs. vehicle categorization task, while the other half performed an “X” vs. “O” categorization task. We found task-related modulation in the P1 component over the medial occipital area: a category difference was present only in the animal vs. vehicle categorization task, and only when the amplitude spectrum was available. For the P2 and N2 components, category differences were greater in the animal vs. vehicle categorization task than in the “X” vs. “O” task. Furthermore, this effect occurred irrespectively of the low-level image modulation. Therefore, we have separated two different top-down mechanisms. The early task-related modulation suggests a top-down process that is selective to physical features. The later task effect occurs irrespectively of the physical differences, suggesting modulation of higher-level visual processes (for instance, activation of visual category representations).

It has been shown that the expectation of a specific category, cued by the category label activates higher-level, category specific visual areas. Furthermore, this preparatory activity modulates the processing of the subsequent stimuli. The aim of Paper 2 was to investigate whether a similar mechanism also plays a role in the task-related animal-vehicle ERP difference modulation that we found in Paper 1. In Paper 1, participants expected to see either an animal or a vehicle, therefore, the task could pre-activate the category-specific areas, which could modulate the processing of the presented image.

This process might show up as larger animal-vehicle differences in the visual ERPs. In paper 2, we used the same animal and vehicle images as in Paper 1, but instead of a two-choice categorization task, we used a label-image matching task. Participants first saw the word '*animal*' or '*vehicle*' appearing on the screen, which was followed by an image of an animal or a vehicle – a similar paradigm to those used in fMRI studies reporting preparatory activity in the visual areas. Participants had to decide whether the presented category label and the following image matched or not. All images were presented twice, following the presentation of either the correct, or the incorrect category label. Early visual ERPs reflect visual processes and not decision-making per se. If labels elicit preparatory activity in visual areas and such activity modulate the processing of the subsequent image, then the difference between trials with correct and incorrect labels could evoke differences already during the visual processes and possibly in the visual ERPs before decision-making processes. Therefore, if the task effect observed in Paper 1 was related to a mechanism similar to preparatory activity, then ERPs should differ between trials with correct, and trials with incorrect category labels for animal and vehicle images.

Moreover, we also wanted to investigate whether we find similar animal-vehicle amplitude differences in the visual ERPs as in Paper 1, even in the absence of a two-choice categorization task. The label-image matching task is a categorization task that directs participants' attention to the object on the image, similarly to a two-choice task. However, the decision that participants have to make is not a forced-choice, animal vs. vehicle decision. Using a similar, still different categorization task can show whether ERP differences found in Paper 1 are due to a two-choice task or they are due to the processing of these images, irrespectively of the task.

PAPER 2

Visual ERP differences between animal and vehicle categories in a label-image matching task

Szilvia Linnert, Vincent Reid, Gert Westermann

Lancaster University

Abstract

Recent studies have reported task-related, top-down modulation in the visual ERPs when participants performed two-choice categorization tasks (for instance animals vs. vehicles). Furthermore, the task, as a top-down factor, has been shown to affect both category-specific lower-level physical properties, such as amplitude spectrum (AS) information, and higher-level category representations. Here we investigated the role of AS information and the effects of category label on the visual ERPs in a label-image matching task, where labels correctly predicted the category of the image in 50% of the cases. The word “animal” or “vehicle” was presented on the screen followed by an image of an animal or a vehicle. Participants had to decide whether the label and image matched or not. Stimuli were presented in two versions with AS information either intact or averaged across stimuli. We found animal-vehicle amplitude differences in the N1, P2 and N2 components, but only if the AS information was available in the images. This suggests that the AS information plays an important role in category recognition. The different role of AS information in this study and previous studies, however, suggests different involvement of top-down and bottom-up processes for different task requirements.

Visual ERP differences between animal and vehicle categories in a label-image matching task

Recent studies have shown that category differences in the visual ERPs are modulated by top-down factors, such as the performed categorization task (Balas & Koldewyn, 2013; Linnert, Csifcsák, & Tompa, in prep; Linnert, Reid, & Westermann, Paper 1). For instance, Linnert, Csifcsak and Tompa (in prep) presented color and greyscale images of cars and birds to the participants together with their unrecognizable, 2x2 pixels scrambled versions. Participants either performed a car vs. bird categorization task or a color vs. greyscale task. For the recognizable stimuli, the car-bird difference in the visual N1 component was larger in the car-bird task compared to the color task over temporal areas. For the unrecognizable images the color-greyscale N1 difference was larger in the color task then in the car-bird task, and this effect was measured over the medial occipital area. This result suggests that task as a top-down factor can affect both higher-level category specific processes and lower-level physical processes, such as color processing (Linnert, Csifcsak, & Tompa, in prep). However, it remained open whether ERP differences between object categories such as cars and birds or animate and inanimate objects reflect category representations per se, or only differences between category-specific low-level features.

Torralba and Oliva (2003) have suggested that categorization can be solved in a feed-forward manner relying on low-level information such as the amplitude spectrum (AS). AS reflects the energy distribution of an image in different spatial frequencies. It has been shown that the AS information differs between categories such as faces, animals, man-made objects or between natural and artificial scenes (Bieniek et al., 2012; Gaspar & Rousselet, 2009; Torralba & Oliva, 2003; Rufin VanRullen, 2011). According to Andrews, Clarke, Pell, and Hartley (2010), high-level visual category representations arise from differential processing of the AS information in lower-level visual areas. The role of AS information has also been shown to play a role in eliciting category differences in the visual ERPs (Gaspar & Rousselet, 2009; Linnert et al., Paper 1; Nemeth, Kovacs, Vakli, Kovacs, & Zimmer, 2014; Rossion & Caharel, 2011). A recent study by Linnert et al. (Paper 1) has investigated the effect of AS information on the animal-vehicle differences in the visual ERPs and whether such information can be accessed in a top-down process. Stimuli were images of animals and vehicles on a grey background containing either intact or averaged AS information. Participants either had to discriminate between animals and vehicles or they had to categorize the images based on a letter (X vs. O) appearing somewhere on the image. Linnert et al. (Paper 1) found animal-vehicle differences in the P1 component only when the intact AS information was available in the images and only when participants had to discriminate between animals and vehicles. For the later P2 and N2 components, the category differences were independent of the presence or absence of AS information. There was also a task effect, with the P2 and N2 animal-vehicle amplitude differences larger in the animal-vehicle task than in the X-O task. These results suggest that the animal-vehicle difference in the earlier P1 component reflects low-level category-specific AS differences between stimulus sets, whereas the same differences in the later, P2 and N2 components reflects higher-level category representations. Furthermore, the task as a top-down factor can modulate both lower-level and higher-level representations.

Category selective top-down activation has been reported in several fMRI studies. Visual, face- and place-selective areas have been shown to be activated by mental imagery (O’Craven & Kanwisher, 2000; Reddy et al., 2010). Such category-selective activation suggests that top-down projections can activate representations in the visual cortex even in the absence of incoming sensory information. Similar activation has also been reported in the visual cortex when anticipatory cues, such as auditory cues (Stokes et al., 2009), symbols (Esterman & Yantis, 2010; Peelen & Kastner, 2011) or written labels (Puri et al., 2009) predicted the subsequent stimulus (for review see Battistoni, Stein, & Peelen, 2017). Preparatory activity, similarly to metal imagery, has been found in shape- and category-selective areas for letters (Stokes et al., 2009), for faces and houses (Esterman & Yantis, 2010; Puri et al., 2009) or for cars and humans (Peelen & Kastner, 2011). Furthermore, not only the higher-level visual areas but even V1 can be activated by anticipatory cues; an activation that might reflect the preparation to detect category-specific low-level visual features (Peelen & Kastner, 2011). In addition, preparatory activity in the category-selective areas has been shown to facilitate participants' performance (Peelen & Kastner, 2011) and it also modulates activation in category-selective areas during the processing of the following image (Puri et al., 2009). Together, these results suggest an interaction between top-down and bottom-up processes at the level of visual, category-specific representations.

A possible EEG correlate of preparatory activity that has been identified is a change in the power of alpha-band activity before a stimulus presentation (for review see Battistoni et al., 2017). Similarly to the fMRI activation, such alpha-band power change has also shown area specificity depending on the attended feature (for review see Battistoni et al., 2017). For instance, Snyder and Foxe (2010) reported increased alpha-band activity over the ventral visual regions when anticipatory cues instructed participants to attend to motion, and over the dorsal regions when participants were preparing to attend to color. It is possible that the task-related top-down modulation of category differences in the visual ERPs relies on a mechanism similar to preparatory activity. In most of the studies where category differences are found participants usually perform a categorization task. Therefore, the expectation of the appearing categories could elicit top-down processes. If preparatory activity modulates the processing of the following image (Puri et al., 2009), such a mechanism could explain interaction between top-down and bottom-up processes reflected in the visual ERPs.

The aim of the present study was to investigate whether correct and incorrect category labels, as anticipatory cues, modulate the visual ERPs in response to presentation of category exemplars. In order to investigate this question, instead of a two-choice categorization task, we used a label-image matching task. Participants first saw the word ‘*animal*’ or ‘*vehicle*’ appearing on the screen, which was followed by an image of an animal or a vehicle. Participants had to decide whether the presented category label and the following image matched or not. All images were presented twice, following both the correct and incorrect category label. If the task-effect in previous studies was related to a mechanism similar to preparatory activity, then ERPs should differ between trials with correct and incorrect category label for animal and vehicle images, respectively.

It has been shown that preparatory activity affects both higher-level, category-selective areas and earlier areas processing low-level features (Peelen & Kastner, 2011). Similarly, the task-related top-down processes modulating the visual ERPs have been shown to affect both higher-level category representations and low-level features, such as the AS information (Linnert et al., Paper 1). In order to investigate whether labels modulate higher- or lower-level processes, we used images with intact and averaged AS information, similarly to Linnert et al. (Paper 1). Early visual ERPs reflect visual processes and not decision-making per se. If labels elicit preparatory activity in visual areas and such activity modulate the processing of the subsequent image, then the difference between trials with correct and incorrect labels could evoke differences already during the visual processes and possibly in the visual ERPs before decision-making processes. If presenting a label affects visual processing at early levels where AS information is processed then we would expect effect of label congruency in the early components and only for images with intact AS. In contrast, if label congruency affected both image types, it would show interaction at the level of higher-level category representations.

Finally, we wanted to investigate whether we find similar animal-vehicle amplitude differences in the visual ERPs as Linnert et al. (Paper 1) even in the absence of a two-choice categorization task. Category differences have been found in word-image matching paradigms between natural and artificial categories (Kiefer, 2001) or between dogs and birds (J. W. Tanaka & Curran, 2001). On the other hand, Kincses, Chadaide, Varga, Antal, and Paulus (2006) found different timings and scalp distributions for animal vs. non-animal ERP differences between a go/no-go task and a two-choice categorization task. Therefore, different task requirements can modulate the underlying visual processes, which can be reflected in the visual ERPs. According to Tremel and Wheeler (2015) category-specific areas accumulate and integrate evidence for decision-making and top-down factors can modulate these processes. The label-image matching task is a categorization task that directs participants' attention to the object on the image, similar to a two-choice task. However, the decision that participants have to make is not a forced-choice, animal vs. vehicle decision. It is possible that different information is accumulated in a forced-choice categorization task and in a recognition task, and the AS information might play a different role in different tasks. If so, such processes might also be reflected in the ERPs. If using a similar, still different categorization task results in similar ERP differences to the ones reported in Paper 1, then these differences are due to the processing of these images, irrespectively of the task. In contrast, if we found dissimilar results to Paper 1, then ERP differences are more likely due to the forced-choice categorization task.

Methods

Participants

Fifteen university students (10 females, 11 right handed), ranging in age from 18 to 27 years took part in the experiment and received course credits for their participation. Participants had normal or corrected-to-normal vision. Four additional participants' data were discarded because they did not provide enough artifact-free segments. Participants were informed about the experiment before they volunteered and signed a consent form. Experiment and procedure were approved by the Lancaster University Ethics Committee.

Stimuli and materials

The stimuli used in this study were partly the same as those of Linnert, Reid and Westermann (Paper 1), and are reproduced here. The stimuli were grayscale images of easily recognizable animals (n=65) and vehicles (n=65) appearing on a grey background. The images represented a diverse range of animals (e.g. fish, insects, mammals, birds etc.), and vehicles (e.g. cars, buses, planes, ships etc.) and the animals and vehicles varied in their size and viewpoint angle. The size of the images was kept constant: W10.71° x H8.38° (W506 x H396 pixels).

For one version of the stimulus set the Fourier amplitude spectrum was kept intact and only the luminance was equalized across the images (intact AS), while for the second version of stimuli both the luminance and Fourier amplitude spectrum was equalized across the images (averaged AS) (see Figure 1). The MATLAB based SHINE toolbox (Willenbockel et al., 2010) was used for luminance and spectrum matching. For all images, the luminance histograms of the images were matched by using the average luminance distribution across the whole stimulus set. For the images with averaged AS, we first applied the luminance histogram matching, followed by the amplitude spectrum matching. These two matching processes can distort each other, therefore, we modified the images in 20 iterations (for details see Willenbockel et al., 2010).

Overall, participants saw 520 images: 65 animals and 65 vehicles both with intact and averaged AS and following congruent and incongruent category labels. Stimuli were shown on a 15” CRT monitor with 85 Hz refresh rate and 1024 by 768 pixels resolution. Stimuli were presented using E-Prime software.

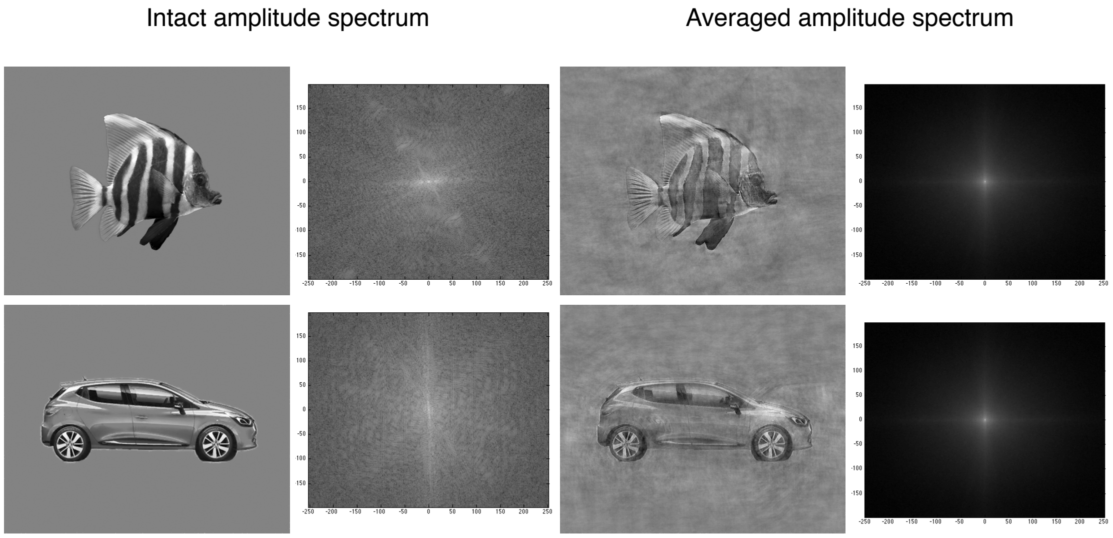


Figure 1. Examples for the luminance and the luminance & amplitude spectrum matched animal and vehicle stimuli.

Experimental procedure

Participants sat on a chair approximately 80 cm from the screen. The stimulus presentation was around 40 minutes long with a break after every 65 stimuli (around every 5 minutes). The break was as long as the participant wanted and the test proceeded when they pressed a button. The order of the presented stimuli was randomized, the distance between the presentations of the two versions of the same stimuli was unpredictable and the presentation order was also random. Participants performed a category label-image matching task.

A trial started with a category label (either "*ANIMAL*" or "*VEHICLE*") displayed in the middle of a grey screen for 1500 ms. Then a fixation cross appeared for a random duration between 500-700 ms which was followed by the image displayed in the center of the screen. The participants had to decide whether the previously presented category label and the image were congruent or not by pressing a button on a response box. The left and right buttons were counterbalanced across the participants: half of the participants had to press the left button if the label and image were congruent and the right button if they were not congruent, and this was reversed for the other half. The image was shown until the participant pressed a button (but for a maximum of 2000 ms). At the end of each trial a blank grey screen was shown for a random duration between 1300 - 1500 ms. Response accuracy and reaction time was collected together with the recorded EEG.

EEG recording and analysis

The details of EEG recording and analysis used in this study were similar as those of Linnert, Reid and Westermann (Paper 1), and are reproduced here. We used an EGI (Electrical Geodesics Incorporated, Eugene, OR) GES300 EEG system with a NetAmps USB amplifier and a 128 channel HydroCel Net to record the EEG from the scalp. The EEG was recorded and analyzed with the NetStation software (Electrical Geodesics Incorporated, Eugene, OR). Electrode impedance was kept below 50 kOhm. During the recording, the EEG was referenced to Cz, and the sampling rate was 1000 Hz.

During off-line analyzis the raw data were band-pass filtered between 0.1-30 Hz. The EEG was segmented from 150 ms before to 800 ms after stimulus onset. Epochs were sorted according to the object categories (animals, vehicles), stimulus types (intact AS, averaged AS) and congruency (congruent, incongruent). Only epochs containing correct responses were included. Eye movements and bad channels were automatically rejected if the average amplitude in an 80 ms moving window exceeded ±55 µV at EOG channels or ±200 µV at any other channel. Epochs in which more than 13 channels contained artifacts were rejected; if fewer than 13 channels contained artifacts, bad channels were substituted by spline interpolation. Participants with fewer than 20 artifact-free epochs per condition were excluded.

Average ERPs were calculated from the artifact-free segments for each stimulus condition. ERPs were baseline corrected using a 150 ms pre-stimulus baseline, and were re-referenced to the average reference. Grand averages were plotted by using the EEGLAB (*v. 13_4_4b*) MATLAB toolbox.

Results

Behavioral results

Repeated measures ANOVAs were performed on the accuracy rates and mean reaction times (RT). Accuracy was defined as the percentage of correct responses. We calculated the mean RT values (correct responses only) for each participant and for each condition. Reaction times values smaller than 200 ms or larger than 1000 ms were excluded. We performed a 2x2x2 ANOVA with the within-subject factors CATGEORY (animal, vehicle), STIMULUS TYPE (intact AS, averaged AS), and CONGRUENCY (label and following image was congruent or incongruent).

We found an interaction between CATEGORY and CONGRUENCY (*F*(1, 14) = 7.44, $p = .016$, $\eta p^2 = .35$) (Figure 2a). Participants were more accurate (*t*(14) = 2.66, $p = .018$) when the animal image was preceded by congruent category label (“animal”; *M* = 96.46%, *SE* = 1.18) then when the preceding category label was incongruent (“vehicle”; *M* = 94.5 %, *SE* = 1.26 %). It was the other way around for vehicle images; participants were more accurate (*t*(14) = -2.17, $p = .048$) when the image was preceded by incongruent category label (“animal”; *M* = 95.72 %, *SE* = 1.6 %) compared to the congruent category label (“vehicle”; *M* = 92.88 %, *SE* = 1.37 %). The accuracy difference between congruent and incongruent trials for images of animals and vehicles indicate that participants did not simply press one of the buttons more likely. However, it is possible that they more likely gave “yes” response to animal images and “no” response to vehicle images.

Regarding the RT, participants were faster in label-image matching when images contained intact AS information (*M* = 604.87 ms; *SE* = 18.63 ms) compared to images with averaged AS (*M* = 619.83 ms; *SE* = 19.03 ms) (main effect of STIMULUS TYPE; *F*(1,14) = 6.98, $p = .019$, $\eta p^2 = .33$). We also found an interaction between CONGRUENCY and CATEGORY (*F*(1, 14) = 7.41, $p = .017$, $\eta p^2 = .35$) (Figure 2b). For animal images, participants were faster (*t*(14) = 5.46, $p < .001$) in the congruent trials (*M* = 586.51 ms, *SE* = 16.11 ms) than in the incongruent trials (*M* = 638.20 ms, *SE* = 21.28 ms). No difference (*t*(14) = 0.91, $p = .376$) was found for vehicle images between congruent (*M* = 607.68 ms, *SE* = 20.65 ms) and incongruent (*M* = 617.03 ms, *SE* = 19.68 ms) trials.

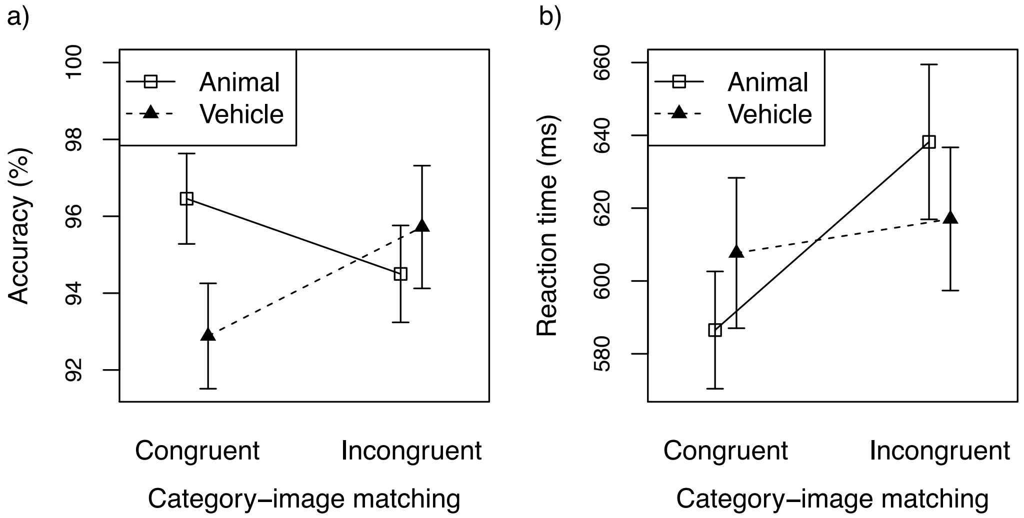


Figure 2. Accuracy rates (a) and reaction times (b) in the congruent and incongruent conditions for animal and vehicle images, respectively.

Electrophysiology results

Visual ERPs were analyzed on the occipital and temporal electrodes including electrodes O1, O2, Oz, T5, T6, 59, 65, 66, 67, 71, 72 76, 77, 84, 90, 91 as in Paper 1. The selection of these electrodes is in line with previous studies (e.g. Balas & Koldewyn, 2013; Bentin, Allison, Puce, Perez, & McCarthy, 1996; Proverbio, Del Zotto, & Zani, 2007; Rossion & Jacques, 2008). Figure 2 shows the scalp distribution of the animal-vehicle ERP differences at the latencies of 75, 100, 125, 150, 175, 200, 225, 250, 275 and 300 ms for congruent and incongruent images with intact and averaged AS information in the animal vs. vehicle (AV task) and in the X vs. O (XO task) tasks. Figure 3 shows the scalp distribution of the congruent-incongruent difference at the latencies of 75, 100, 125, 150, 175, 200, 225, 250, 275 and 300 ms for animal and vehicle images with intact and averaged AS information. In order to investigate the spatial distribution of the animal-vehicle and congruent-incongruent ERP differences, we performed permutation-based paired tests ($p<0.05$) with correction for multiple comparisons (false discovery rate) using the MATLAB-based EEGLAB (v. 13_4_4b) toolbox (Delorme & Makeig, 2004). This was done for all selected electrodes at each of the latencies specified above (see Figure 2). The red dots on figure 2 indicate the electrodes with significant animal-vehicle ERP differences at the latencies of the corresponding scalp-maps.

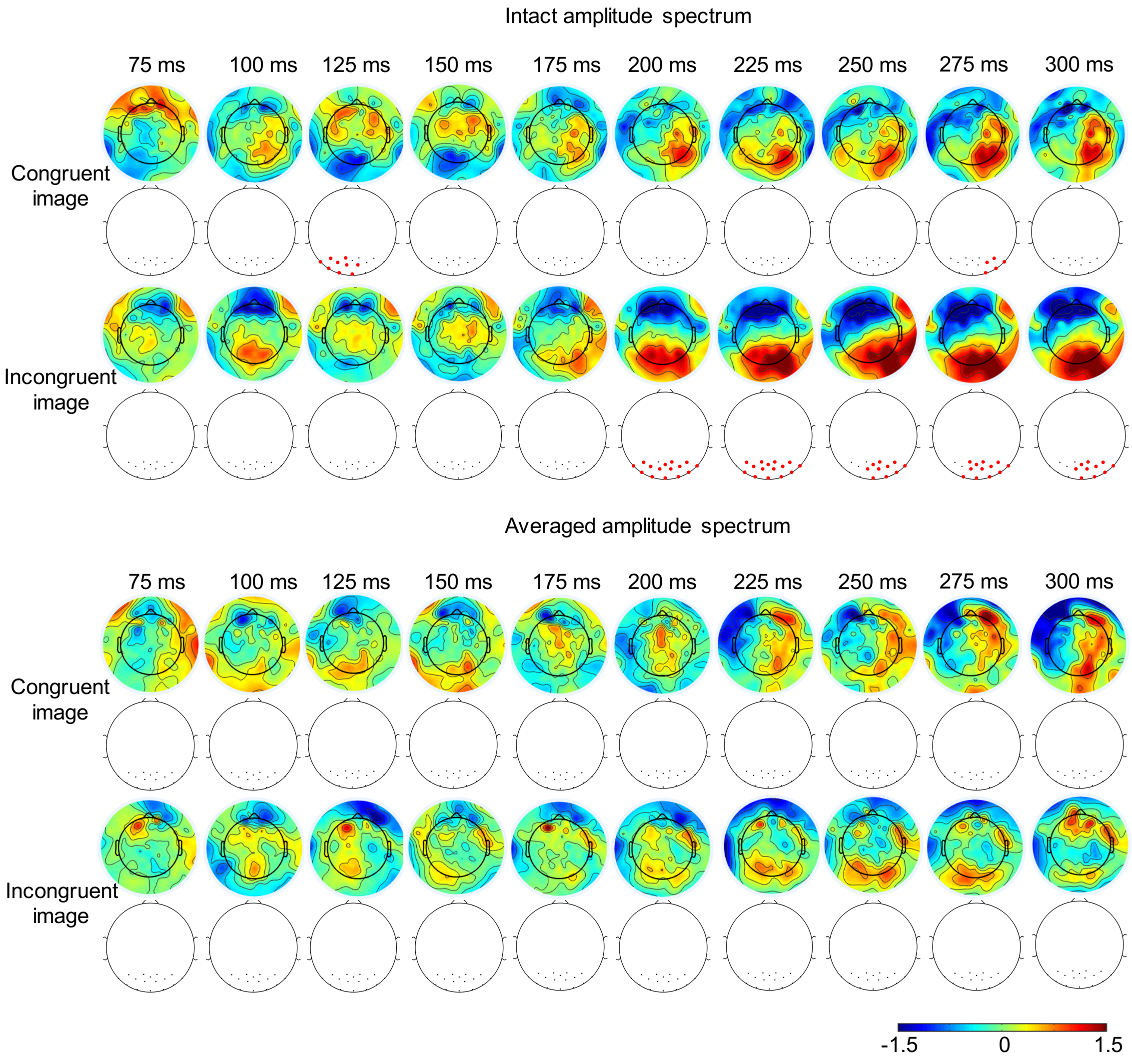


Figure 2. Scalp distribution of the animal-vehicle ERP differences between 50 and 300 ms for images with intact (top rows) and averaged AS (bottom rows) in the congruent and incongruent conditions, respectively. The red dots indicate the electrodes with significant animal-vehicle ERP differences at the latency of the corresponding scalp-map. Note that 75 ms, 125-150 ms, 200 ms and 250-275 ms are the peak latencies of the P1, N1, P2 and N2 components, respectively.

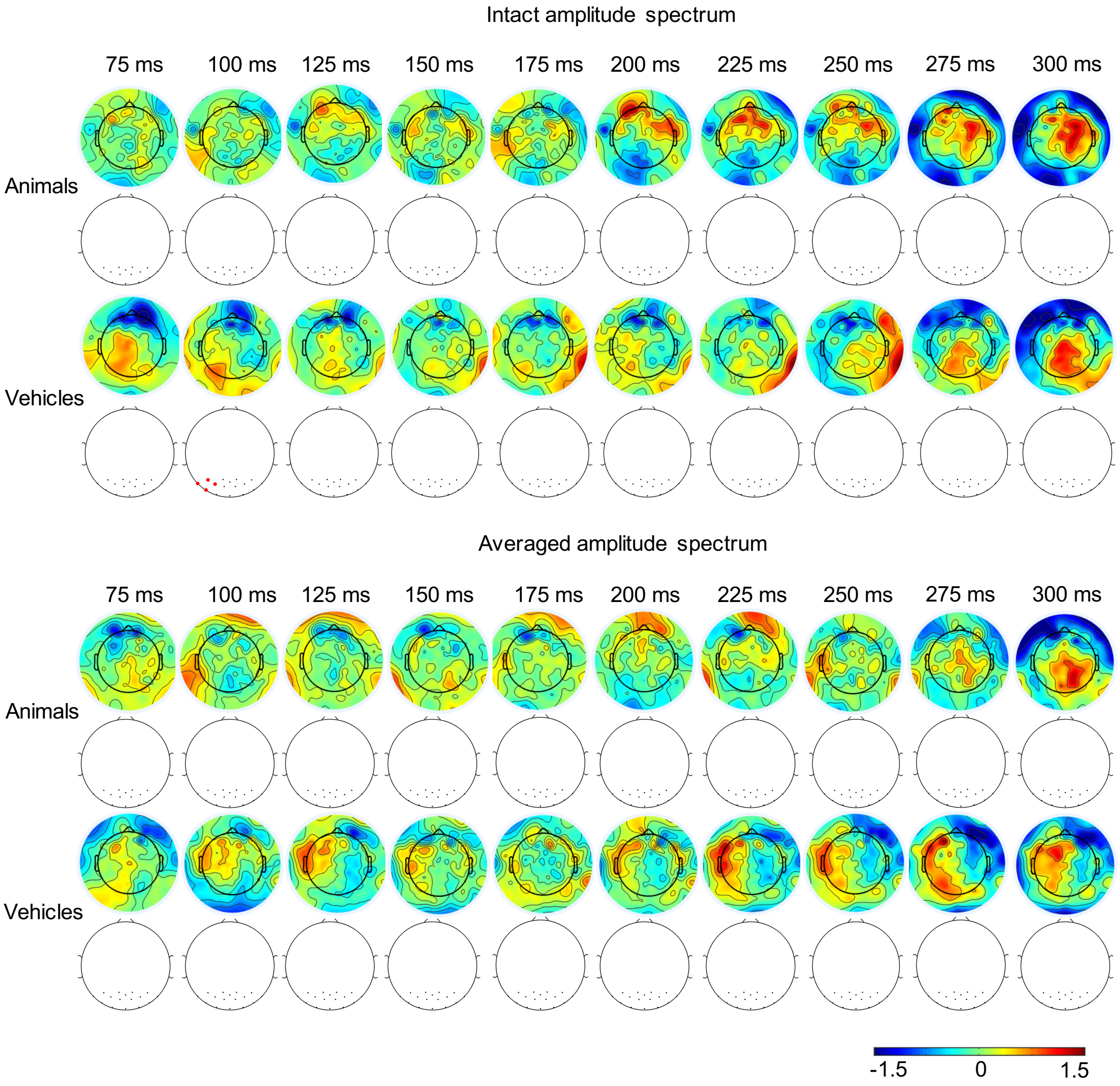


Figure 3. Scalp distribution of the congruent-incongruent ERP differences between 50 and 300 ms for images with intact (top rows) and averaged AS (bottom rows) for animal and vehicle images, respectively. The red dots indicate the electrodes with significant animal-vehicle ERP differences at the latency of the corresponding scalp-map. Note that 75 ms, 125-150 ms, 200 ms and 250-275 ms are the peak latencies of the P1, N1, P2 and N2 components, respectively.

In order to gain additional information about the time-scale of the animal-vehicle ERP differences, we also performed paired tests (permutation with $p<0.05$ and false discovery rate) comparing the ERPs for animals and vehicles for congruent and incongruent images and for congruent and incongruent trials for animals and vehicles at all data points between -150 and 400 ms for images with intact and averaged AS information (EEGLAB; Delorme & Makeig, 2004). For this analysis, similarly to Paper 1, we averaged ERPs for the medial occipital and for the temporal electrodes over the left and right hemisphere, respectively (see figure 4, 5 - note that the midline electrodes are not included in this analysis). The distinction between the electrode groups was made based on the ERP waveforms and latencies of the ERP peaks.

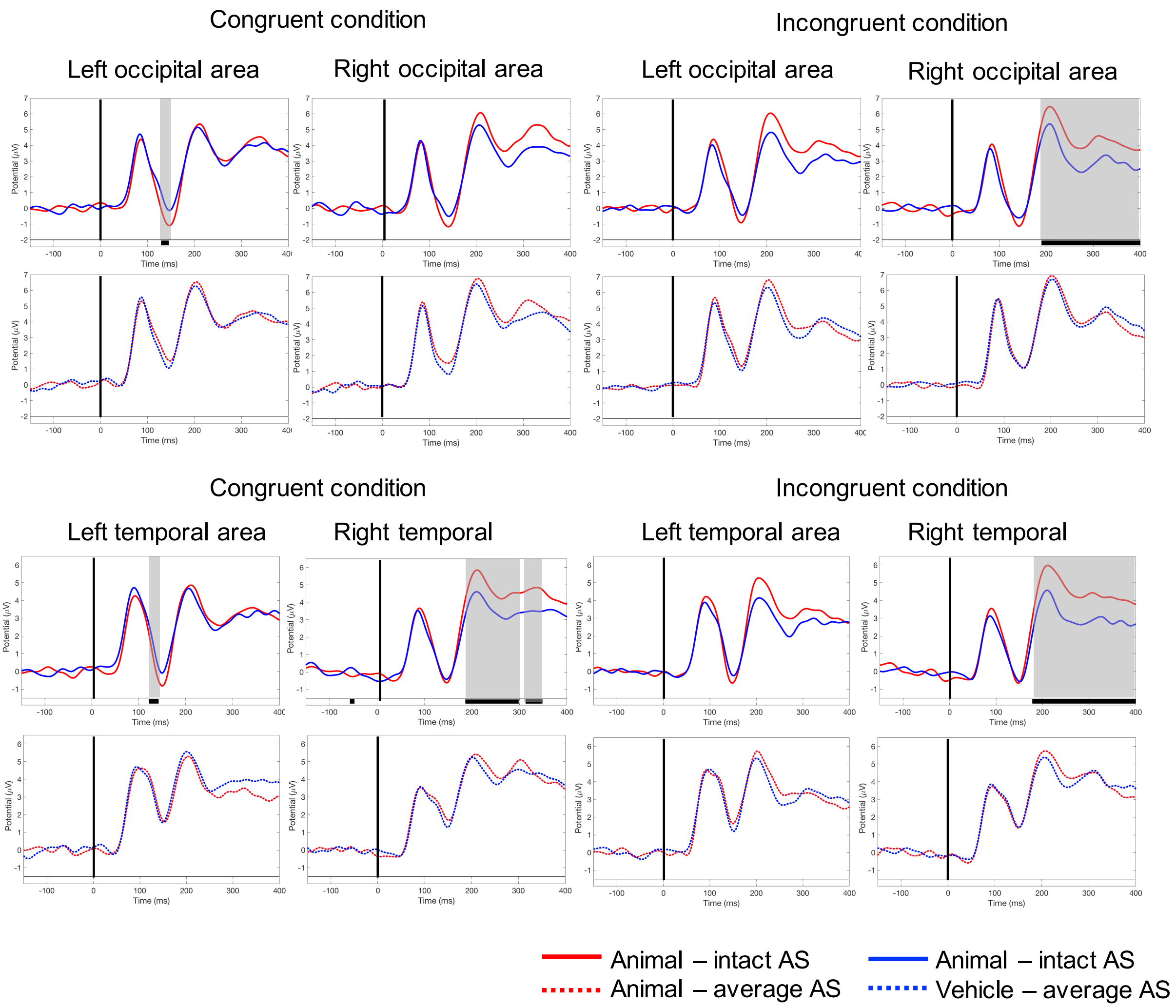


Figure 4. Results of the point-by-point analysis for animal-vehicle ERP differences. The continuous lines represent images with intact AS, the dashed lines represent images with averaged AS. The red lines represent animals and the blue lines represent vehicles. Paired analyses were performed on the average waveforms of four occipital electrodes and on the average waveforms of three temporal electrodes over both the left and right hemispheres, separately. Averaged ERPs are presented for the left and right occipital areas (top two rows) and for the left and right temporal areas (bottom two rows) for the congruent (left columns) and incongruent conditions (right columns). Grey rectangles show the time-windows of the significant differences between the animal and vehicle ERPs.

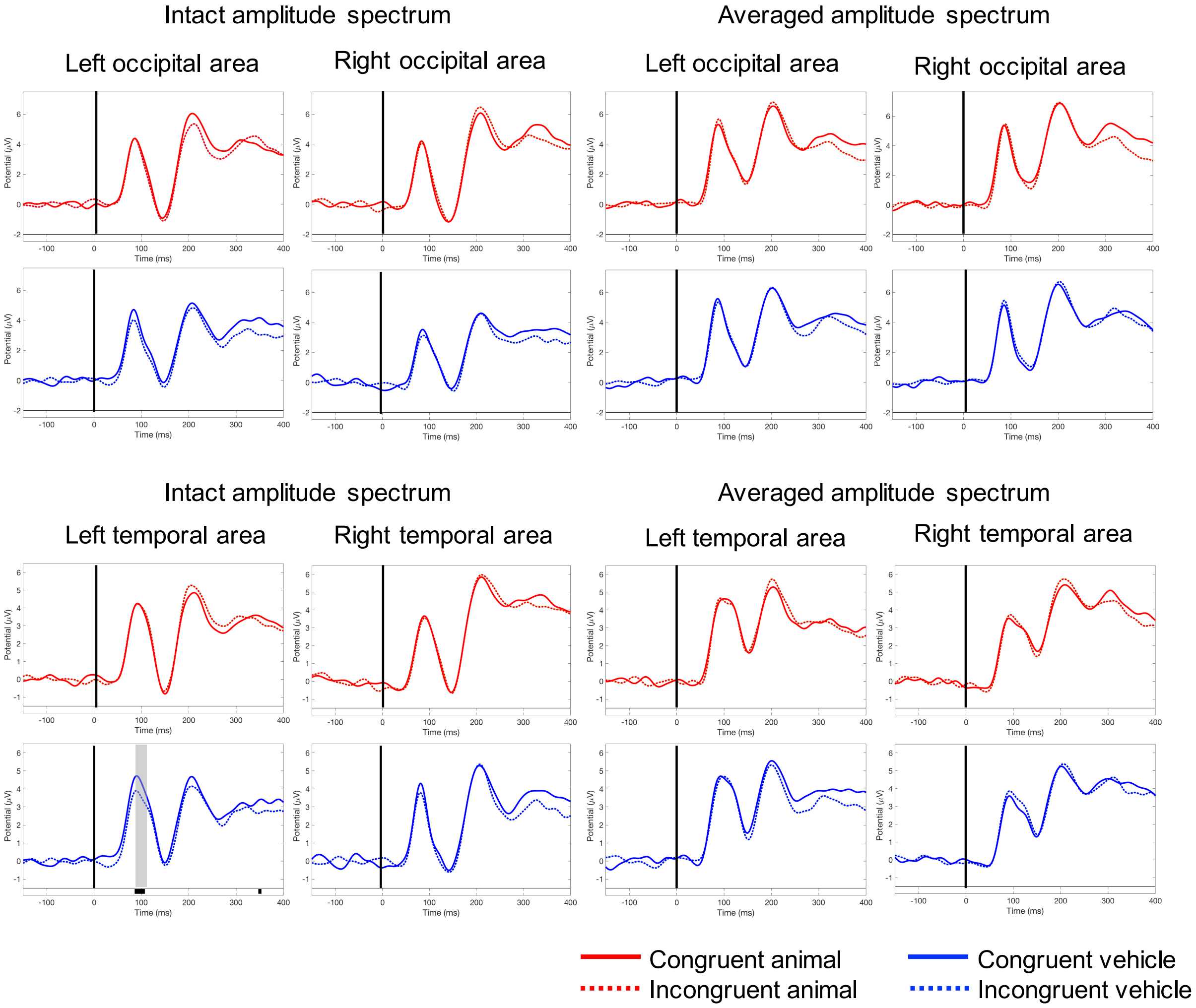


Figure 5. Results of the point-by-point analysis for congruent-incongruent ERP differences. The continuous lines represent congruent trials, the dashed lines represent incongruent trials. The red lines represent animals, the blue lines represent vehicles. Paired analyses were performed on the average waveforms of four occipital electrodes and on the average waveforms of three temporal electrodes over both the left and right hemispheres, separately. Averaged ERPs are presented for the left and right occipital areas (top two rows) and for the left and right temporal areas (bottom two rows) for images with intact (left columns) and averaged AS (right columns). Grey rectangles show the time-windows of the significant differences between the animal and vehicle ERPs.

In order to examine the interactions between animal-vehicle ERP differences, congruency, task, and/or stimulus types, mean amplitudes were extracted for time windows 85 - 105 ms (corresponds to P1), for 130 - 145 ms (corresponds to N1), and for 180 - 300 ms (corresponds to P2 and N2) over the occipital area. Over the temporal area, the mean amplitudes were extracted for time windows 85 - 105 ms (corresponds to P1), for 120 - 140 ms (corresponds to N1) and for 180 - 300 ms (corresponds to P2 and N2).

Similarly to the previous paper (Paper 1) and to the point-by-point analysis presented above, the occipital area included electrodes O1, 71, 66, 67 over the left hemisphere and O2, 76, 77, 84 over the right hemisphere, and the temporal area included electrodes T5, 59 and 65 over the left hemisphere and electrodes T6, 90, 91 over the right hemisphere. Average ERPs are presented on Figure 6. We conducted 2x2x2x2 mixed ANOVAs with the within subject factors STIMULUS TYPE (intact AS, averaged AS), HEMISPHERE (left, right), CATEGORY (animal, vehicle) and CONGRUENCY (label and following image was congruent or incongruent for the occipital and temporal electrodes separately. A Greenhouse-Geisser correction was applied on the p-values for violations of the assumption of sphericity. We report only the highest-level interactions including the category factor.

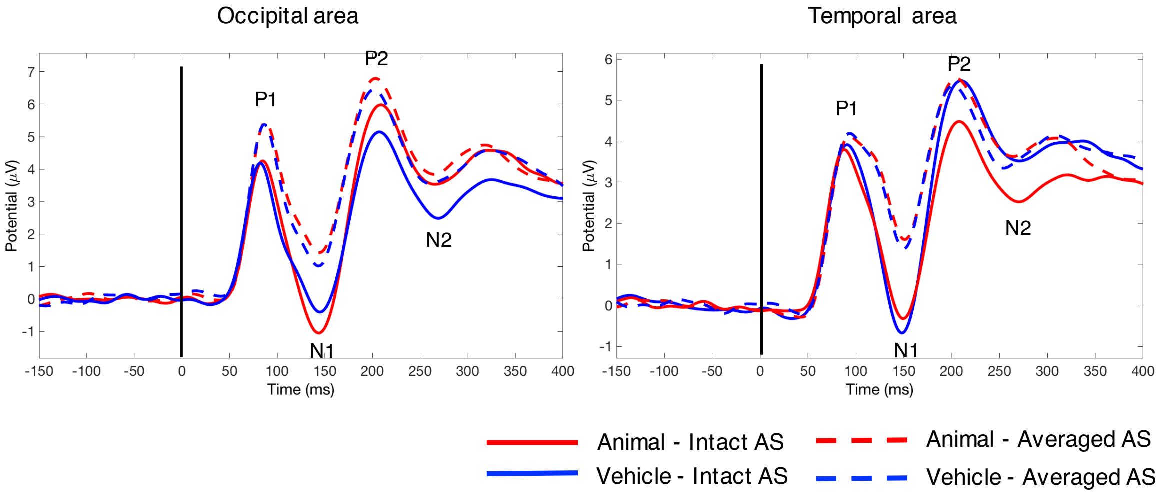


Figure 6. ERP waveforms over the occipital (left) and temporal (right) areas. The continuous lines represent images with intact AS, the dashed lines represent images with averaged AS. The red lines represent animals, the blue lines represent vehicles. ERPs show the averaged waveforms across electrodes and hemispheres.

Occipital area

We did not find any significant differences between animals and vehicles or between congruent and incongruent trials in the P1 component.

The animal-vehicle difference in the N1 differed between the stimulus types (STIMULUS TYPE * CATEGORY interaction: *F(1, 14*) = 15.29, *p* = .002, ηp^2 = .52). For images with intact AS, animals elicited more negative amplitudes (*M* = -.85 µV, *SE* = .74 µV) compared to vehicles (*M* = -.15 µV, *SE* = .87 µV) (*t(14*) = -2.90, *p* = .024). For images with intact AS, amplitudes elicited by animals (*M* = 1.60 µV, *SE* = .75 µV) and vehicles (*M* = 1.18 µV, *SE* = .78 µV) did not differ (*t(14*) = 1.37, *p* = .29).

In the time window of the P2 and N2 components we found a three-way interaction between CATEGORY, STIMULUS TYPE and HEMISPHERE (*F(1, 14*) = 5.32, *p* = .037, ηp^2 = .28). ERPs between animals and vehicles differed only for images with intact AS and only over the right hemisphere (*M* = 1.06 µV, *SE* = .31 µV, *t(14*) = 3.46, *p* = .016) (Figure 7).

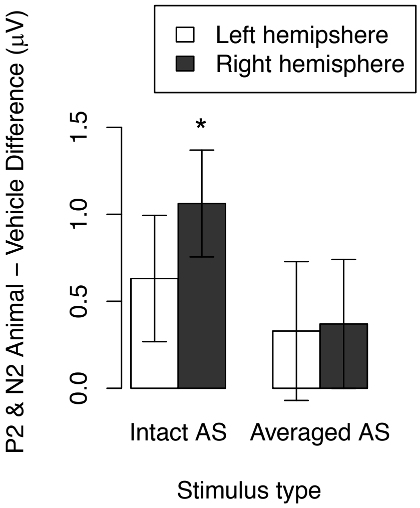


Figure 7. Mean amplitude differences between ERPs for animals and vehicles in the time window of the occipital P2 and N2 components for images with intact and averaged AS over the left and right hemispheres. Error bars represent SE.

Temporal area

We did not find any significant differences between animals and vehicles or between congruent and incongruent trials in the P1 or N1 components over the temporal areas.

In the time-window of the P2 and N2 components we found an interaction between CATGEORY and STIMULUS TYPE ($F(1, 14) = 25.77$, $p < .001$, $\eta p^2 = .65$). The animal vehicle difference was significant only for images with intact AS ($M = .97$ µV, $SE = .22$ µV; $t(14) = 4.49$, $p = .002$), but not for images with averaged AS ($M = .21$ µV, $SE = .31$ µV; $t(14) = .70$, $p = .99$). This confirms the interaction suggested by Figure 4.

Comparison across components

In order to compare the animal-vehicle ERP differences across components, similarly to Paper 1 we performed an additional ANOVA including COMPONENT (occipital: P1, N1, P2 & N2; temporal: N1, P2, N2) as a within subject factor beside STIMULUS TYPE (intact AS, averaged AS), HEMISPHERE (left, right), CATEGORY (animal, vehicle), and CONGRUENCY (congruent, incongruent). Planned comparisons were performed for P1-N1 and N1-P2&N2 component pairs over the occipital area, and for N1-P2 and P2-N2 pairs over the temporal area.

Over the occipital area we found an interaction between CATEGORY, STIMULUS TYPE and COMPONENT ($F(2, 28) = 16.11$, $p < .001$, $\eta p^2 = .54$). For images with intact AS, the animal-vehicle difference changed direction ($F(1, 14) = 13.62$, $p = .002$, $\eta p^2 = .41$) between the P1 ($M = .41$ µV, $SE = .38$ µV) and N1 ($M = -.70$ µV, $SE = .24$ µV) components, and it also differed ($F(1, 14) = 3.63$, $p = .077$, $\eta p^2 = .21$) between N1 and P2 & N2 ($M = .8466$ µV, $SE = .318$ µV). For images with averaged AS, animal-vehicle difference did not differ between the components (P1-N1: *F(1, 14)* = .87, $p = .37$, $\eta p^2 = .06$; N1-P2&N2: *F(1, 14)* = 1.0, $p = .76$, $\eta p^2 = .01$) (Figure 8).

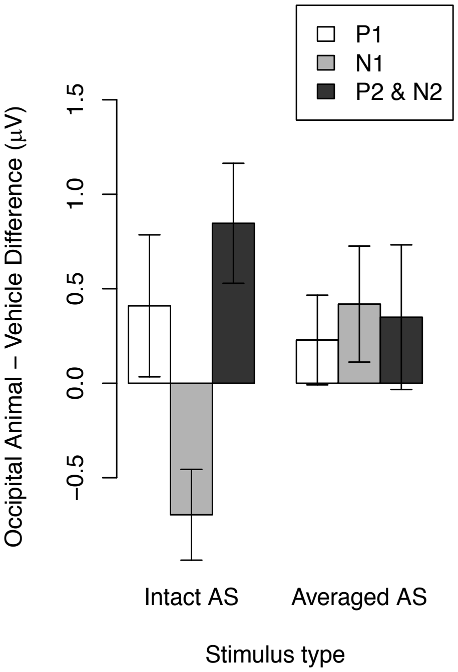


Figure 8. Differences between animal and vehicle amplitudes by components over the occipital area for images with intact and averaged AS. Error bars represent SE.

Over the temporal area we found an interaction between CATEGORY, STIMULUS TYPE and COMPONENT (*F(2, 28)* = 22.51, $p < .001$, $\eta p^2 = .62$) (Figure 9). For images with intact AS, the animal-vehicle difference changed direction (*F(1, 14)* = 8.49, $p = .011$, $\eta p^2 = 3.77$) between the P1 (*M* = .21 µV, *SE* = .18 µV) and N1 (*M* = -.28 µV, *SE* = .17 µV) component, it increased (*F(1, 14)* = 33.59, $p < .001$, $\eta p^2 = .71$) for the time window of the P2 and N2 components (*M* = .97 µV, *SE* = .22 µV). For images with intact AS, the animal-vehicle difference did not differ between the components (P1-N1:

F(1, 14) = 1.98, $p = .18$, $\eta p^2 = .124$; N1-P2 & N2: *F(1, 14)* < .001, $p = .98$, $\eta p^2 < .001$) (Figure 9).

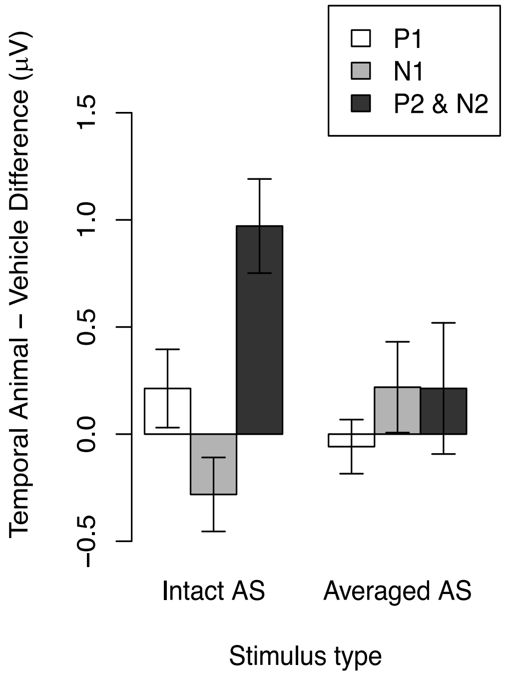


Figure 9. Differences between animal and vehicle amplitudes by components over the temporal area for images with intact and averaged AS. Error bars represent SE.

Discussion

The aim of this study was to investigate effects of AS information on the category differences in the visual ERPs in a label-image matching task. Furthermore, we investigated whether correct and incorrect category labels, as anticipatory cues, modulate the visual ERPs in response to presentation of category exemplars. We found animal-vehicle amplitude differences in the N1, P2, and N2 components, but only when the AS was available in the images. The correct and incorrect category labels did not affect the visual ERPs.

It has been shown that category labels evoke preparatory activity in category-specific areas, and such activity modulates the processing of the following image in the same, category-specific areas (Puri et al., 2009). In Puri et al.'s (2009) fMRI study, images of houses and faces were preceded by their written names (*house* or *face*). Participants had to answer whether the name and the image matched or not. Beside the preparatory activity in the category-specific visual areas, the authors found increased activity during the processing of the visual stimuli when the anticipatory cue correctly predicted the subsequent visual stimulus. Here, we did not find amplitude differences between congruent and incongruent trials for animals and vehicles, A possible reason for the lack of amplitude differences can be that the presentation frequency of correct and incorrect labels affect the underlying processes. In the study of Puri et al. (2009), the percentage of correct anticipatory cues was 70%, whereas in our study it is only 50%. If anticipatory cues are only effective at chance level, they might not evoke preparatory activity. Alternatively, it is possible that the ERP technique is not sensitive enough to measure such differences, especially for highly manipulated images. EEG studies have investigated task-specific preparatory activity by measuring the pre-stimulus alpha power (Battistoni et al., 2017; Snyder & Foxe, 2010). Therefore, pre-stimulus activity can be modulated by task requirements. Furthermore, the results of Van Der Lubbe et al. (2016) suggest that pre-stimulus power in the alpha band predicts the P1 and N1 components to some extent. Therefore, in theory, it would be possible to see the effect of preparatory activity, and in this study specifically, the effect of label congruency on the visual ERPs. However, pre-stimulus activity accounts only for a small part of the ERP signal, beside evoked and induced activity (Van Der Lubbe et al., 2016). Thus, small differences in the pre-stimulus power do not necessarily appear in the subsequent ERPs as amplitude differences, especially because ERPs are baseline corrected using a pre-stimulus interval.

Regarding the category differences in the ERPs, we found amplitude differences between animals and vehicles in the N1, P2 and N2 components, but only when the AS information was available in the stimuli. In the study of Linnert et al. (Paper 1), on the other hand, the animal-vehicle differences in the P2 and N2 components were not modulated by the AS information. The differences between our results and the results of Linnert et al. (Paper 1) could be due to the different task requirements. Task-related, temporal and topographical differences have been previously reported for animal and non-animal ERP differences between go/no-go and two-choice categorization tasks (Kincses et al., 2006). Authors have suggested that less information is necessary for recognizing the presence of an animal in a go/no-go task than for an animal vs. non-animal categorization in a two-choice task. Nevertheless, both the two-choice animal vs. vehicle categorization task (Linnert et al., Paper 1) and the label-image matching task for animal and vehicle categories in this study require the recognition of a category member. It is possible, however, that the involvement of top-down and bottom-up processes differs between the two tasks. Indeed, according to Gilbert and Sigman (2007), the current processing demands always depend on the resonance between the incoming sensory input and task requirements. The P2 and N2 animal-vehicle differences in the previous study (Linnert et al., Paper 1) were larger or present only when the task required participants to discriminate between those categories. This suggests that top-down factors play a crucial role in category differences for the later ERP components in a two-choice task. In the present study, however, labels predicted the subsequent image only at chance level. Therefore, it is possible that bottom-up processes played a more important role in processing the image category than in a two-choice categorization task. The different involvement of top-down and bottom-up processes in the two kinds of task could explain the differences regarding the AS information between the two studies. Low-level, category-specific factors such as the AS information could play a more important role in bottom-up processing. The role of AS information in bottom-up processes is in line with the suggestion that higher-level category specificity arises from the processing of low-level category-specific features, such as the AS information (Andrews et al., 2010).

In sum, our results suggest that category labels affect visual processes, and such effects are reflected in the visual ERPs. The presence of ERP category differences only for images with intact AS information suggests the role of low-level, category-specific features in a label-image matching task. This effect can be due to the fact that labels only predicted the subsequent image category at chance level. Therefore, it is possible that category recognition under these task requirements relied more on bottom-up processes. Overall, these results are in line with previous studies suggesting task-related category differences in the visual ERPs and the flexibility of processes reflected in the ERPs.

References

Andrews, T. J., Clarke, A., Pell, P., & Hartley, T. (2010). Selectivity for low-level features of objects in the human ventral stream. *NeuroImage*, *49*(1), 703–711. http://doi.org/10.1016/j.neuroimage.2009.08.046

Balas, B., & Koldewyn, K. (2013). Early visual ERP sensitivity to the species and animacy of faces. *Neuropsychologia*, *51*(13), 2876–2881. http://doi.org/10.1016/j.neuropsychologia.2013.09.014

Battistoni, E., Stein, T., & Peelen, M. V. (2017). Preparatory attention in visual cortex. *Annals of the New York Academy of Sciences*, *1396*, 92–107. http://doi.org/10.1111/nyas.13320

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. http://doi.org/10.1162/jocn.1996.8.6.551

Bieniek, M. M., Pernet, C. R., & Rousselet, G. A. (2012). Early ERPs to faces and objects are driven by phase, not amplitude spectrum information: evidence from parametric, test-retest, single-subject analyses. *Journal of Vision*, *12*(13), 12. http://doi.org/10.1167/12.13.12

Esterman, M., & Yantis, S. (2010). Perceptual expectation evokes category-selective cortical activity. *Cerebral Cortex*, *20*, 1245–1253. http://doi.org/10.1093/cercor/bhp188

Gaspar, C. M., & Rousselet, G. A. (2009). How do amplitude spectra influence rapid animal detection? *Vision Research*, *49*(24), 3001–3012. http://doi.org/10.1016/j.visres.2009.09.021

Gilbert, C. D., & Sigman, M. (2007). Brain states: Top-down influences in sensory processing. *Neuron*, *54*, 677–696. http://doi.org/10.1016/j.neuron.2007.05.019

Kiefer, M. (2001). Perceptual and semantic sources of category-specific effects: event-related potentials during picture and word categorization. *Memory & Cognition*, *29*(1), 100–116. http://doi.org/10.3758/BF03195745

Kincses, T. Z., Chadaide, Z., Varga, E. T., Antal, A., & Paulus, W. (2006). Task-related temporal and topographical changes of cortical activity during ultra-rapid visual categorization. *Brain Research*, *1112*, 191–200. http://doi.org/10.1016/j.brainres.2006.07.044

Linnert, S., Csifcsák, G., & Tompa, T. (in prep). Top-down modulation of animate-inanimate category differences in the visual N1.

Linnert, S., Reid, V., & Westermann, G. (in prep). Underlying mechanisms of task-related modulation of animal vs. vehicle differences in the visual ERPs.

Nemeth, K., Kovacs, P., Vakli, P., Kovacs, G., & Zimmer, M. (2014). Phase noise reveals early category-specific modulation of the event-related potentials. *Frontiers in Psychology*, *5*, 1-12. http://doi.org/10.3389/fpsyg.2014.00367

O'Craven, K. M., & Kanwisher, N. (2000). Mental imagery of faces and places activates corresponding stiimulus-specific brain regions. *Journal of Cognitive Neuroscience*, *12*, 1013–1023. http://doi.org/10.1162/08989290051137549

Peelen, M. V, & Kastner, S. (2011). A neural basis for real-world visual search in human occipitotemporal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 12125–12130. http://doi.org/10.1073/pnas.1101042108

Proverbio, A. M., Del Zotto, M., & Zani, A. (2007). The emergence of semantic categorization in early visual processing: ERP indices of animal vs. artifact recognition. *BMC Neuroscience*, *8*, 24. http://doi.org/10.1186/1471-2202-8-24

Puri, A. M., Wojciulik, E., & Ranganath, C. (2009). Category expectation modulates baseline and stimulus-evoked activity in human inferotemporal cortex. *Brain Research*, *1301*, 89–99. http://doi.org/10.1016/j.brainres.2009.08.085

Reddy, L., Tsuchiya, N., & Serre, T. (2010). Reading the mind's eye: Decoding category information during mental imagery. *NeuroImage*, *50*(2), 818–825. http://doi.org/10.1016/j.neuroimage.2009.11.084

Rossion, B., & Caharel, S. (2011). ERP evidence for the speed of face categorization in the human brain: Disentangling the contribution of low-level visual cues from face perception. *Vision Research*, *51*(12), 1297–1311. http://doi.org/10.1016/j.visres.2011.04.003

Rossion, B., & Jacques, C. (2008). Does physical interstimulus variance account for early electrophysiological face sensitive responses in the human brain? Ten lessons on the N170. *NeuroImage*, *39*, 1959–1979. http://doi.org/10.1016/j.neuroimage.2007.10.011

Snyder, A. C., & Foxe, J. J. (2010). Anticipatory attentional suppression of visual features indexed by oscillatory alpha-band power increases: A high-density electrical mapping study. *Journal of Neuroscience*, *30*(11), 4024–4032. http://doi.org/10.1523/JNEUROSCI.5684-09.2010

Stokes, M., Thompson, R., Nobre, A. C., & Duncan, J. (2009). Shape-specific preparatory activity mediates attention to targets in human visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(46), 19569–19574. http://doi.org/10.1073/pnas.0905306106

Tanaka, J. W., & Curran, T. (2001). A neural basis for expert object recognition. *Psychological Science*, *12*, 43–47. http://doi.org/10.1111/1467-9280.00308

Torralba, A., & Oliva, A. (2003). Statistics of natural image categories. *Network*, *14*(3), 391–412. http://doi.org/10.1088/0954-898X/14/3/302

Tremel, J. J., & Wheeler, M. E. (2015). Content-specific evidence accumulation in inferior temporal cortex during perceptual decision-making. *NeuroImage*, *109*, 35–49. http://doi.org/10.1016/j.neuroimage.2014.12.072

Van Der Lubbe, R. H. J., Szumska, I., & Fajkowska, M. (2016). Two sides of the same coin: ERP and wavelet analyses of visual potentials evoked and induced by task-relevant faces. *Advances in Cognitive Psychology*, *12*(4), 153–167. http://doi.org/10.5709/acp-0195-3

VanRullen, R. (2011). Four common conceptual fallacies in mapping the time course of recognition. *Frontiers in Psychology*, *2*, 1–6. http://doi.org/10.3389/fpsyg.2011.00365

Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., & Tanaka, J. W. (2010). Controlling low-level image properties: the SHINE toolbox. *Behavior Research Methods*, *42*(3), 671–684. http://doi.org/10.3758/BRM.42.3.671

Linking statement for Paper 3

Spatial frequencies (SFs) in visual stimuli have been suggested to be a flexible source of information according to task demands in object recognition. For instance, a man-made environment can be categorized as a city based on the coarse low spatial frequency (LSF) information, whereas the same scene can be categorized as New York, based on more fine-grained, high spatial frequency (HSF) information. Thus, different spatial frequencies represent different characteristics of objects or scenes. Moreover, HSF and LSF information is projected by different pathways to higher-level areas. Therefore, it is possible that a specific categorization task can selectively rely on SF information as well as modulate the animate-inanimate category differences in the visual ERPs.

In Papers 1 and 2 we have shown that the categorization task modulates animal-vehicle amplitude differences in the visual ERPs. The aim of Paper 3 was to investigate whether the animal-vehicle ERP differences and the task-related category difference modulation relies on specific SF information. In Paper 3, we used the same animal and vehicle images as in the previous studies, however, the images contained either only low or only high SF information, or the SF was kept intact. If category differences in the visual ERPs are based on coarse information, such as global shape, then the temporal course of the animal-vehicle differences for intact and LSF images should be similar. In contrast, if the category differences are based on fine-grained, local features, then the temporal course of such differences should be similar for intact and HSF stimuli. In addition, we wanted to investigate how the categorization task, as a top-down factor, affects the category differences for HSF and LSF images. Again, if the task-related top-down process selectively rely on HSF or LSF information, then the effect of task on the animal-vehicle differences will be similar for intact and HSF, or for intact and LSF stimuli, respectively. On the other hand, if the task-related modulation of animal-vehicle differences is irrespective of the SF content of the stimuli, then top-down and bottom-up processes might interact at the level of category representations.

PAPER 3

Early animal-vehicle ERP differences reflect integration of high and low spatial frequency information

Szilvia Linnert, Vincent Reid, Gert Westermann

Lancaster University

Abstract

Spatial frequencies (SFs) in visual stimuli have been suggested to be a flexible source of information according to task demands in object recognition. It has also been shown that the specific categorization task modulates animal-vehicle differences in the visual ERPs. The aim of this study was to investigate whether this task-dependent animal-vehicle amplitude difference modulation specifically relies on the availability of certain frequency bands in the visual stimuli. Stimuli contained a diverse range of animals and vehicles appearing on a grey background. Images either contained intact SFs, or low or high SF information only. Participants either performed an animal vs. vehicle discrimination task (AV task), or an X vs. O discrimination task (XO task) according to letters appearing on the images. For intact images, we found faster reaction times and more accurate responses. For intact images animal-vehicle amplitude differences occurred early, in the time window of the N1 component. In contrast, for images containing high or low SFs, category differences occurred only in the time window of the P2 component. Animal-vehicle P2 differences were larger in the AV task compared to the XO task. This task related modulation was irrespective of the SF content of the stimuli, suggesting the role of higher-level category representations in this effect. The earlier category differences for images with intact SFs suggest rapid accumulation and integration of category-specific visual cues from different spatial frequencies.

Early animal-vehicle ERP differences reflect integration of high and low-spatial frequency information

Spatial frequency (SF) information plays an important role in visual object recognition. Different spatial frequencies represent different characteristics of objects or scenes. Low spatial frequencies (LSF) convey coarse, global information about the shape of the objects, whereas high-spatial frequencies (HSF) convey fine-grained information about local features and smaller details. Low SFs are processed faster than HSF information, possibly because different SFs are projected by different pathways to higher-level areas (for the role of magno- and parvocellular pathways see Bar, 2003; Bullier, 2001). The faster processing of LSF information suggests that object recognition follows a coarse-to-fine processing course (Bar, 2003; Bullier, 2001; Schyns & Oliva, 1994). For instance, Schyns and Oliva (1994) used hybrid images in which the HSF and LSF content belonged to different scenes. The hybrid images were interpreted based on the LSF information when the images were presented for 30 ms, whereas they were interpreted according to the HSF information, when the presentation time was 150 ms. The coarse-to-fine order of scene processing has been also investigated by neuroimaging and electrophysiology methods (for review see Kauffmann, Ramanoël, & Peyrin, 2014). For instance, Musel et al. (2014) have reported higher activation of the parahippocampal place area for scene sequences when images were presented in a coarse to fine order (first low than high SFs) compared to the activation for fine to coarse sequence order. Beside scene perception, object (Goddard, Carlson, Dermody, & Woolgar, 2016) and face (Goffaux et al., 2011) recognition also seem to follow the coarse-to-fine sequence. For instance, in a recent MEG study it was shown that object identity decoding in temporal areas took longer for HSF images than for LSF images (Goddard et al., 2016). In addition, if LSF information is processed faster, it can also play a role in top-down processes (Bar, 2003; Bullier, 2001). According to Bar (2003) the course, partially processed LSF information activates rough interpretations about object identity in the prefrontal cortex. These possible identity templates facilitate the more detailed processing in the visual areas through feedback projections. In a combined fMRI and MEG study, Bar et al. (2006) found orbitofrontal activity selective to stimuli containing LSF information. Furthermore, they found phase synchrony between frontal and visual areas, which was stronger for LSF or intact images. These results support the roles of LSF information and frontal areas in top-down processes. In further support of Bar's (2003) feedback theory, Bognár et al. (2017) have recently reported that transcranial stimulation over the frontal areas affected the processing of LSF but not HSF objects.

However, the fixed LSF-to-HSF processing course has been challenged. In contrast to the fixed order of coarse-to fine process sequence, Schyns and Oliva (Oliva & Schyns, 1997; Schyns, 1998; Schyns & Oliva, 1997; Schyns & Oliva, 1999) have suggested that SF information can be used flexibly according to task demands (for review on task-dependent face perception see Ruiz-Soler & Beltran, 2006). For instance, a man-made environment can be categorized as a city based on the coarse LSF information, whereas the same scene can be categorized as New York, based on more fine-grained, HSF information. Both behavioral (Oliva & Schyns, 1997; Schyns & Oliva, 1999) and EEG (Valerie Goffaux, Jemel, Jacques, Rossion, & Schyns, 2003) evidence support this theory. For instance, Schyns and Oliva (1999) have shown that different SFs are used in face perception depending on whether the task is to decide about the gender or about facial expression. In this study, hybrid faces were used; in a hybrid image the HSF and LSF information differed in gender (e.g. the HSF face was female, while the LSF was a male), whether the face was expressive or non-expressive and whether the expressive face showed happiness or anger. The authors found that the gender decision was made based on the HSF or LSF information with equal probability. Participants used the HSF information more when they had to decide whether the face was expressive or not, and the type-of-expression decision was made more based on the LSF information. Task-related SF selectivity has also been investigated in the visual ERPs. In the study of Goffaux et al. (2003), the task was to decide about either the gender or the familiarity of intact, HSF and LSF faces. The N170 amplitudes differed between the gender and familiarity task, but only for the LSF images and not for the HSF and intact ones. Craddock et al. (2013), on the other hand, did not find task-related SF modulation for the visual P1 and N1 components. Instead of faces, they used HSF and LSF animal and non-animal images and participants had to perform either a living/non-living categorization, or a grammatical gender decision (in German).

The effect of SFs in the visual ERPs has been mainly investigated for face (e.g. Flevaris, Robertson, & Bentin, 2008; Valérie Goffaux, Gauthier, & Rossion, 2003; Halit, de Haan, Schyns, & Johnson, 2006) and scene (e.g. Mu & Li, 2013; Peyrin et al., 2010) processing. Only a few EEG studies have assessed the effects of SFs in animate vs. inanimate categorization (e.g. Craddock et al., 2013; Rokszin, Győri-Dani, Nyúl, & Csifcsák, 2016). However, these studies only compared the ERPs between images with different SF content, but did not analyze the animate-inanimate ERP differences for HSF, LSF and intact images, respectively. Animate-inanimate category differences in the visual ERPs have been shown to be modulated by the categorization task (Linnert, Csifcsák, & Tompa, in prep; Linnert, Reid, & Westermann, Paper 1). Linnert et al. (in prep.) found larger car-bird differences for the visual N1 component in a car vs. bird categorization task compared to the category differences when the same images were categorized as colored or greyscale. In addition, we have shown that such task-related top-down factors can access both lower-level and higher-level category specific processes (Linnert et al., Paper 1). In that study, images of animals and vehicles were categorized either as animals and vehicles, or based on letters (X or O) appearing somewhere on the image. The images contained either intact amplitude-spectrum (AS) information, a low-level, image characteristic that has been shown to systematically differ for different categories (Torralba & Oliva, 2003), or they contained the averaged AS of the stimulus set. We found P1 animal-vehicle differences, but only if the AS information was intact in the images and only in the animal vs. vehicle task. Animal-vehicle differences for the P2 and N2 components were larger in the animal-vehicle task than in the XO task, and this task-related modulation was not affected by the AS information. Therefore, the presence of category-specific, low-level information elicited early animal-vehicle differences, but only when the task demands required such discrimination. The task-effects in the P2 and N2 components, on the other hand, were more likely due to activation of higher-level category representations.

The evidence reviewed above raise the possibility that the specific categorization task can selectively use SF information as well as modulate the animate-inanimate category differences in the visual ERPs. The aim of this study was to investigate whether the task-related category difference modulation in the visual ERPs relies on specific SF information. In this study, we compared the animal-vehicle amplitude differences in the visual ERPs between images containing low or high spatial frequency information or intact, unfiltered images. If category differences in the visual ERPs are based on coarse information, such as global shape, then the temporal course of the animal-vehicle differences for intact and LSF images should be similar. In contrast, if the category differences are based on fine-grained, local features, then the temporal course of such differences should be similar for intact and HSF stimuli. In addition, we wanted to investigate how the categorization task, as a top-down factor, affects the category differences for HSF and LSF images. Again, if the task related top-down process selectively use HSF or LSF information, then the effect of task on the animal-vehicle differences will be similar for intact and HSF or for intact and LSF stimuli, respectively. On the other hand, if the task-related modulation of animal-vehicle differences are irrespective of the FS content of the stimuli, then top-down and bottom-up processes might interact at the level of category representations.

Methods

Participants

The final sample included thirty-six university students (16 males, 30 right handed, mean age: 23 years, age range: 18-39 years), who received either course credits or 8 GBP for their participation. Participants had normal or corrected to normal vision. Three additional participants' data were discarded because they did not provide enough artifact-free segments (N=2) or because of equipment failure (N=1). Participants were informed about the experiment before they volunteered and signed a consent form. Experiment and procedure were approved by the relevant Ethics Commitee.

Stimuli and materials

The stimuli used in this study were partly the same as those of Linnert, Reid and Westermann (Paper 1), and are reproduced here. The stimuli were greyscale images of easily recognizable animals (n=120) and vehicles (n=120) appearing on a grey background. The images represented a diverse range of animals (e.g. fish, insects, mammals, birds etc.), and vehicles (e.g. cars, buses, planes, ships etc.) and they varied in their size and viewpoint angle. Next to the objects an "X" or an "O" appeared somewhere on the grey background (see Figure 1). The size of the images was kept constant: W487 x H377 pixels (W: 10.4 visual degrees, H: 7.9 visual degrees).

The SF information of the stimuli was modulated by multiplying the Fourier transform of the original images by Gaussian filters in MATLAB. Stimuli were manipulated so that they contained either only low or high spatial frequency information (LSF images: <0.9 cycles per degree; HSF images: >4.7 cycles per degree), or the spatial frequency information was kept intact (ISF images). In Paper 1, we have shown that the presence of intact amplitude spectrum (AS) affects the early ERP differences. In order to eliminate this confound, we had to match the AS between stimuli in this study. Nevertheless, if we matched the AS across all stimuli containing different SFs, that would highly modulate the filtered SF content of the images. Therefore, the AS information for each image was replaced with the average AS information of the HSF, LSF and ISF stimulus set respectively. Thus, the averaged AS differed between HSF, LSF and ISF images, but it did not differ between animals and vehicles within the three stimulus types. Therefore, AS could not account for animal-vehicle differences for HSF, LSF and ISF images, respectively. The MATLAB based SHINE toolbox (Willenbockel et al., 2010) was used for luminance and AS spectrum matching. Similarly to Paper 1, we first applied the luminance histogram matching, followed by the amplitude spectrum matching. These two matching processes can distort each other, therefore, we modified the images in 20 iterations (for details see Willenbockel et al., 2010). The images were presented in three versions according to their spatial frequency information.

Overall, 720 images (120 animals and 120 vehicles with high, low and intact SF) were presented to the participants. Stimuli were shown on a 15” CRT monitor with 85 Hz refresh rate and 1024 by 768 pixels resolution. Stimuli were presented using E-Prime software.

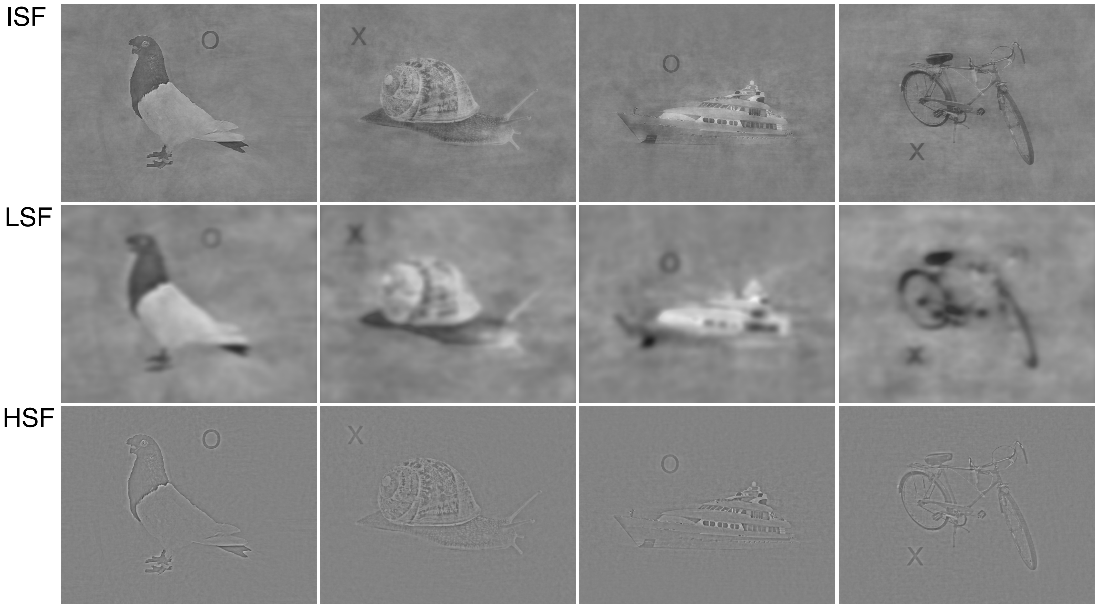


Figure 1. Example stimuli. Stimuli consisted of a diverse range of animals and vehicles on a grey background. An X or O appeared somewhere around the objects. The SF content of the images was intact (top row - ISF), or they contained only either low (middle row - LSF) or high (bottom row - HSF) SF information.

Experimental procedure

The experimental procedure was similar to that of Linnert et al. (Paper 1), and is reproduced here. Participants sat on a chair approximately 80 cm from the screen. The stimulus presentation was around 40 minutes long with a break after every 90 stimuli (around every 5 minutes). The break was as long as the participants wanted and the test proceeded when they pressed a button. The order of the presented stimuli was randomized; the distance between the presentations of the two versions of the same stimulus was unpredictable. The participants performed a two-choice task to categorize the presented images. All participants saw the same images; however, they were randomly assigned to one of two groups. One group had to decide whether the presented image was an animal or a vehicle (AV task, N = 18), while the other group had to report whether there was an “X” or an “O” in the background of the object (XO task, N = 18).

A trial started with a fixation cross displayed in the middle of a grey screen for a random duration between 200 and 350 ms. Then a stimulus was shown in the middle of the grey screen until the participant responded by pressing a button (but for a maximum of 2000 ms). At the end of each trial a blank grey screen was shown for a random duration between 1400 - 1600 ms. Participants' category choices were made by using two buttons on a response box. Within each participant group the left and right buttons were counterbalanced across the participants. Response accuracy and reaction times were collected together with the recorded EEG.

EEG recording and analysis

The details of EEG recording and analysis used in this study were similar as those of Linnert et al. (Paper 1), and are reproduced here. We used an EGI (Electrical Geodesics Incorporated, Eugene, OR) GES300 EEG system with a NetAmps USB amplifier and a 128 channel HydroCel Net to record the EEG from the scalp. The EEG was recorded and analyzed with the NetStation software (Electrical Geodesics Incorporated, Eugene, OR). Electrode impedance was kept below 50 kOhm (as is standard for the EGI system). During the recording, the EEG was referenced to Cz, and the sampling rate was 1000 Hz.

During off-line analysis, the raw data were band-pass filtered between 0.1 and 30 Hz. The EEG was segmented from 150 ms before to 800 ms after stimulus onset. Epochs were sorted into the object categories (animals or vehicles) and stimulus types (HSF, LSF and ISF). Only epochs containing correct responses were included. Eye movements and bad channels were automatically rejected if the average amplitude in an 80 ms moving window exceeded ±55 µV at EOG channels or ±200 µV at any other channel. In addition, each individual epoch was visually inspected and further epochs or channels were rejected. Epochs in which more than 13 channels contained artifacts were rejected; if fewer than 13 channels contained artifacts, bad channels were substituted by spline interpolation. Participants with fewer than 30 artifact-free epochs per condition were excluded.

Average ERPs were calculated from the artifact-free segments for each stimulus condition. ERPs were baseline corrected using a 150 ms pre-stimulus baseline, and were re-referenced to the average reference. Grand averages were generated using the MATLAB-based EEGLAB (*v.* 13_4_4b) toolbox (Delorme & Makeig, 2004).

Results

Behavioral results

Accuracy and reaction times for correct responses were recorded during the experiment. A 3x2x2 mixed ANOVA was performed on the ratio of correct responses (defined in %) and on the mean reaction times (defined in ms, excluding responses that were faster than 200 ms and slower than 1000 ms). Within subject factors were the Stimulus type (HSF, LSF, ISF) and Category (animal, vehicle). The between subject factor was the TASK (AV task, XO task). A Greenhouse-Geisser correction was applied on the p-values for violations of the assumption of sphericity. Bonferroni corrections were applied for multiple comparisons. Planned comparisons were performed for HSF vs. ISF and for LSF vs. ISF stimuli.

Accuracy rates for the three stimulus types were modulated by the task (Stimulus type by Task interaction: $F(2, 68) = 9.578$, $p < .001$, $\eta p^2 = .22$) (Figure 2). In the AV task, accuracy rates were higher for ISF than for LSF images ($t(17) = 6.39$, $p <.001$), but they did not differ between ISF and HSF images ($t(16) = 1.31$, $p = .828$). In the XO task, accuracy rates were higher for ISF than for HSF images ($t(16) = 2.84$, $p = .044$) and showed only marginally significant difference between ISF and LSF images ($t(16) = 2.71$, $p = .06$). For LSF images the accuracy rates were larger ($t(34) = 2.96$, $p = .006$) in the XO task ($M = 95.6$ %, $SE = .851$ %) than in the AV task ($M = 90.74$ %, $SE = 1.44$ %). Accuracy rates did not differ between the two tasks for HSF and ISF images.

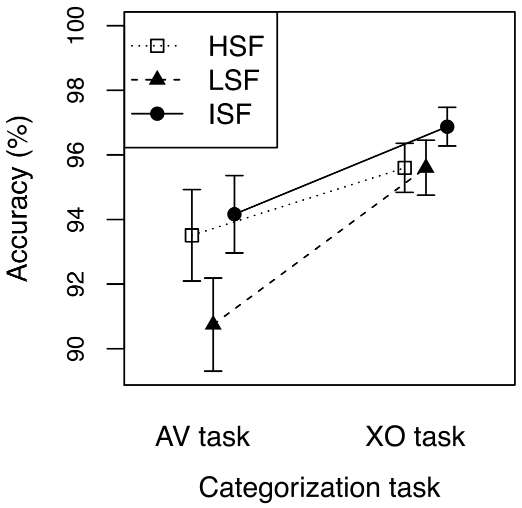


Figure 2. Accuracy rates for ISF, HSF and LSF stimuli in the AV and XO task. In the AV task, accuracy rates differed between ISF and LSF images, in the XO task, accuracy rates differed between ISF and HSF images. Accuracy rates differed between the two tasks only for LSF stimuli. Error bars represent SE.

For the reaction times, we found a three-way interaction between Stimulus type, Category and Task ($F(2, 68) = 14.47$, $p < .001$, $\eta p^2 = .30$). For LSF images, reaction times were faster for vehicles ($M = 566.27$ ms, $SE = 16.88$ ms) compared to animals ($M = 584.46$ ms, $SE = 19.65$ ms) in the AV task, whereas in the XO task, reaction times were faster for animals ($M = 511.10$ ms, $SE = 17.59$ ms), compared to vehicles ($M = 528.38$ ms, $SE = 16.87$ ms) (category by task interaction for LSF images: $F(1, 34) = 26.40$, $p < .001$, $\eta p^2 = .44$). No Task by Category interactions were found for HSF and ISF images. We also compared reaction times for the ISF vs. HSF and ISF vs. LSF image pares separately for animals and vehicles in the AV and XO task. Reaction times differed only between participants were faster responding to both ISF animal and ISF vehicle images compared to LSF animal and vehicle images in both tasks (RT difference between ISF and LSF images in the AV task – animals: *M* = 46.74 ms, *SE* = 5.32 ms; *t(17)* = 8.78 , *p* < .001 ; vehicles: *M* = 26.67 ms, *SE* = 4.85 ms; *t(17)* = 6.11, *p* < .001 ; and in the XO task: animals: *M* = 38.21 ms, *SE* = 3.43 ms; *t(17)* = 11.13, *p* <.001 ; vehicles: *M* = 52.69 ms, *SE* = 4.71 ms; *t(17)* = 11.18, *p* < .001).

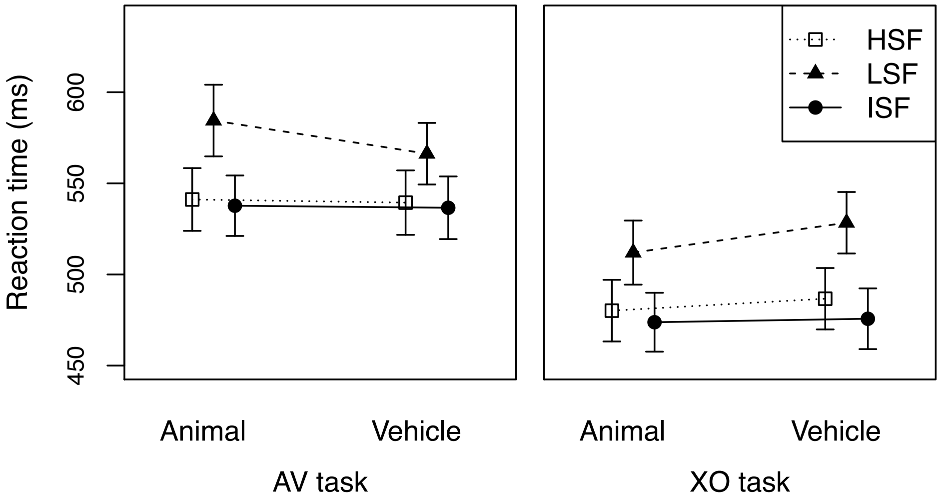


Figure 3. Reaction times for ISF, HSF and LSF stimuli in the AV and XO task for animal and vehicle images. Error bars represent SE.

Electrophysiology results

Visual ERPs were analyzed on the occipital and temporal electrodes including electrodes O1, O2, Oz, T5, T6, 59, 65, 66, 67, 71, 72 76, 77, 84, 90, 91 as in Paper 1. The selection of these electrodes is in line with previous studies (e.g. Balas & Koldewyn, 2013; Bentin, Allison, Puce, Perez, & McCarthy, 1996; Proverbio, Del Zotto, & Zani, 2007; Rossion & Jacques, 2008). Figure 2 shows the scalp distribution of the animal-vehicle ERP differences at the latencies of 50, 75, 100, 125, 150, 175, 200, 225, 250 and

275 ms for images with intact SFs (ISF) and images with low (LSF) and high spatial frequency (HSF) information, in the animal vs. vehicle (AV task) and in the X vs. O (XO task) tasks. In order to investigate the spatial distribution of the animal-vehicle ERP differences, we performed permutation-based paired tests ($p<0.05$) with correction for multiple comparisons (false discovery rate) using the MATLAB-based EEGLAB (v. 13_4_4b) toolbox (Delorme & Makeig, 2004). This was done for all selected electrodes at each of the latencies specified above (see Figure 4). The red dots on Figure 4 indicate the electrodes with significant animal-vehicle ERP differences at the latencies of the corresponding scalp-maps.

In order to gain additional information about the time-scale of the animal-vehicle ERP differences, we also performed paired tests (permutation with $p<0.05$ and false discovery rate) comparing the ERPs for animals and vehicles at all data points between -150 and 400 ms for ISF, LSF and HSF images in the AV task and in the XO task (EEGLAB; Delorme & Makeig, 2004). For this analysis, we averaged ERPs for the occipital and for the temporal electrodes over the left and right hemisphere, respectively (see figure 5 - note that the midline electrodes are not included in this analysis). The distinction between the electrode groups was made based on the ERP waveforms and latencies of the ERP peaks.

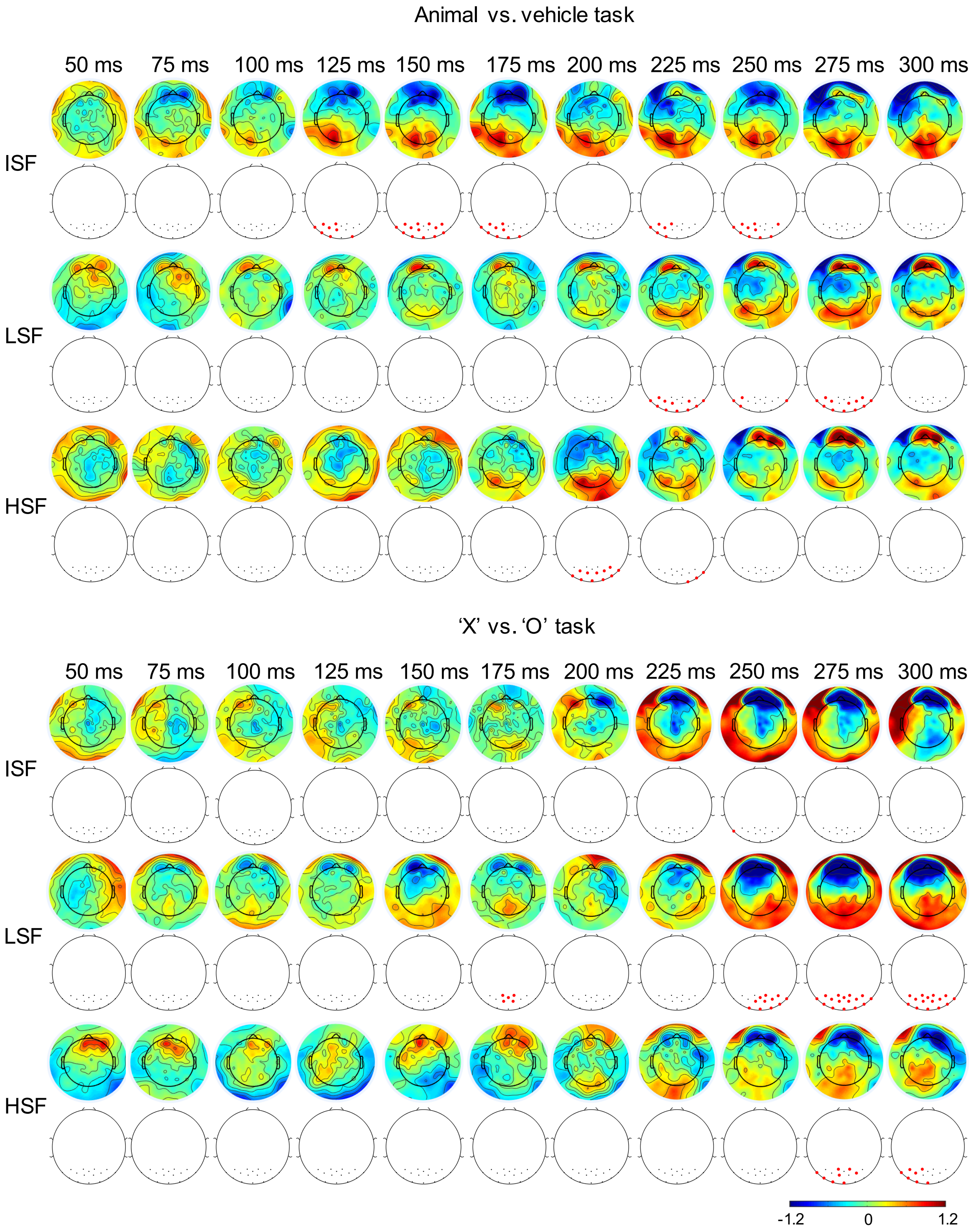


Figure 4. Scalp distribution of the animal-vehicle ERP differences between 50 and 300 ms in the animal vs. vehicle (top rows) and X vs. O task (bottom rows) for ISF, LSF and HSF images. The red dots indicate the electrodes with significant animal-vehicle ERP differences at the latency of the corresponding scalp-map. Note that 75 ms, 125-150 ms, 200 ms and 250-275 ms are the peak latencies of the P1, N1, P2 and N2 components, respectively.

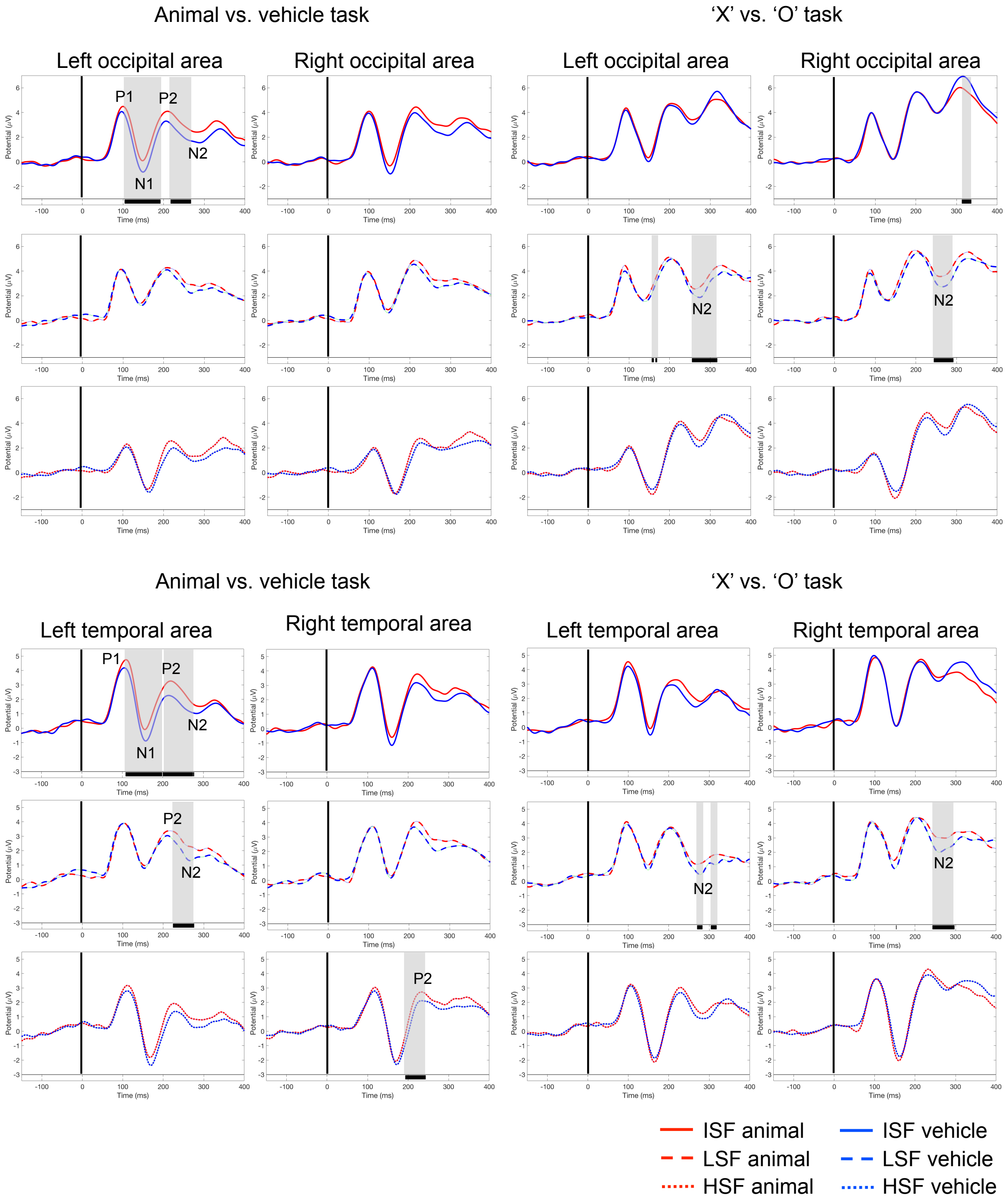


Figure 5. Results of the point-by-point analysis for animal-vehicle ERP differences. The continuous lines represent ISF images, the dashed lines represent LSF images and the dotted lines represent HSF images. The red lines represent animals and the blue lines represent vehicles. Paired analyses were performed on the average waveforms of four occipital electrodes and on the average waveforms of three temporal electrodes over both the left and right hemispheres, separately. Averaged ERPs are presented for the left and right occipital areas (top three rows) and for the left and right temporal areas (bottom three rows) in the AV task (left columns) and in the XO task (right columns). Grey rectangles show the time-windows of the significant differences between the animal and vehicle ERPs.

In order to examine the interactions between animal-vehicle ERP differences, task, and/or stimulus types, mean amplitudes were extracted for time windows 110-190 ms (corresponds to N1), for 210-250 ms (corresponds to P2), and for 250-290 ms (corresponds to N2) over the occipital area. Over the temporal area, the mean amplitudes were extracted for time windows 110-190 ms (corresponds to N1), for 190-240 ms (corresponds to P2) and for 240-280 ms (corresponds to N2). The point-by-point analysis did not reveal any significant animal-vehicle differences in the time window of the P1 component, therefore, we did not perform further analysis for this component. These time windows are also in line with previous literature (Luck, 2005). Similarly to the previous papers (Paper 1 and 2) and to the point-by-point analysis presented above, the occipital area included electrodes O1, 71, 66, 67 over the left hemisphere and O2, 76, 77, 84 over the right hemisphere, and the temporal area included electrodes T5, 59 and 65 over the left hemisphere and electrodes T6, 90, 91 over the right hemisphere. Average ERPs are presented on figure 6. We conducted 3x2x2x2 mixed ANOVAs with the within subject factors STIMULUS TYPE (ISF, LSF, HSF), HEMISPHERE (left, right), and CATEGORY (animal, vehicle), and the between subject factor TASK (AV task, XO task). A Greenhouse-Geisser correction was applied on the p-values for violations of the assumption of sphericity. We report only the highest-level interactions including the category factor.

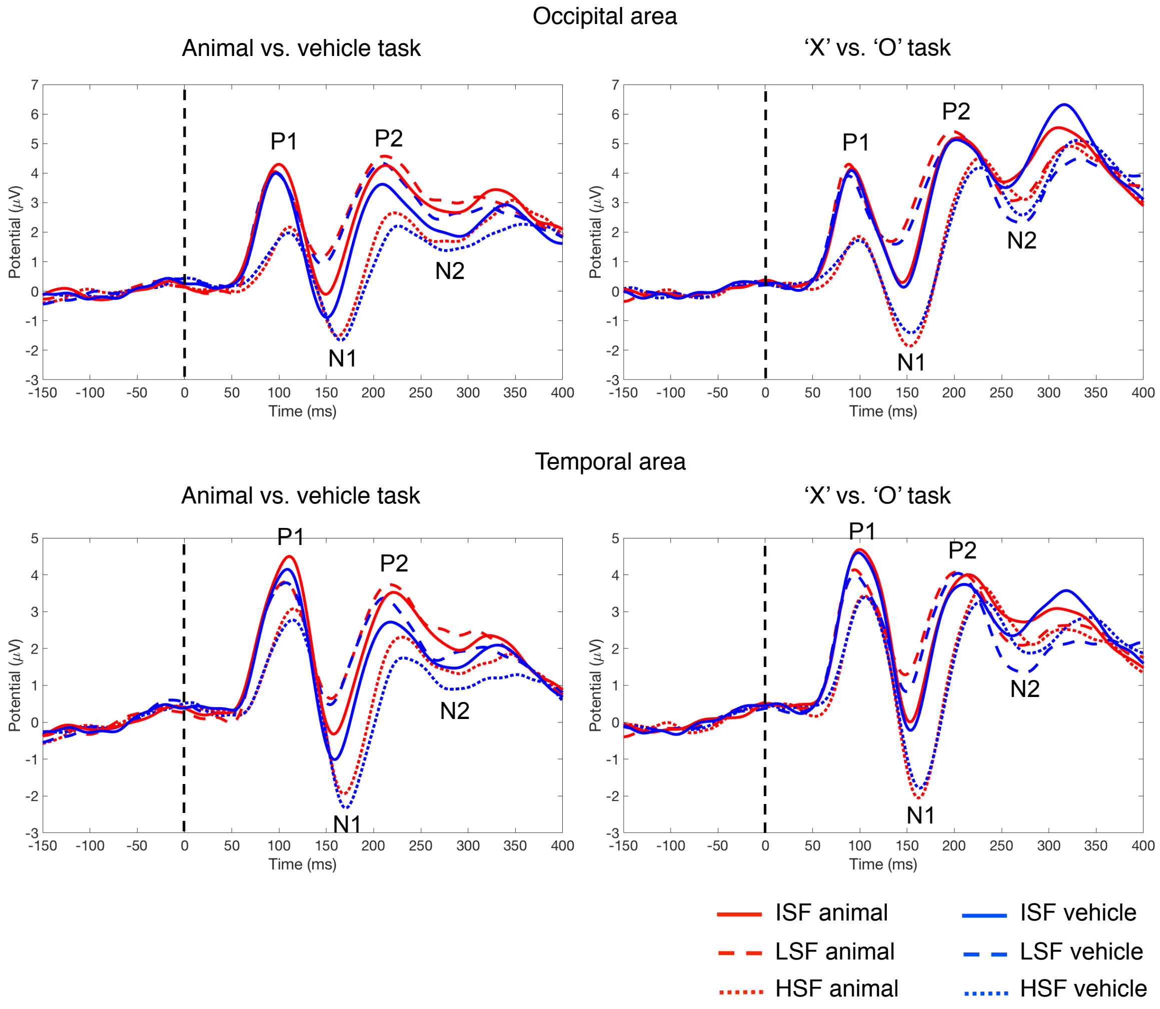


Figure 6. ERP waveforms over the occipital (top) and temporal (bottom) areas in the AV (left) and XO (right) tasks. The continuous lines represent ISF images, the dashed lines represent LSF images and the dotted lines represent HSF images. The red lines represent animals, the blue lines represent vehicles. ERPs show the averaged waveforms across electrodes and hemispheres.

Occipital area

In the time-window of the N1 and P2 components, the animal-vehicle differences were modulated by the Stimulus type and the Hemisphere (CATEGORY*STIMULUS TYPE*HEMISPHERE interaction for N1: $F(2, 66) = 3.44$, $p = .038$, $\eta p^2 = .09$), and for P2: $F(2, 66) = 8.91$, $p < .001$, $\eta p^2 = .21$) (Figure 7). For the N1 component, animal-vehicle differences were significant only for ISF images over the left hemisphere ($M = .61$ µV, $SE = .15$ µV; $t(34) = 4.17$, $p < .001$), but not over the right ($M = .32$ µV, $SE = .13$

µV; $t(34) = 2.40$, $p = .13$). Category differences were not significant for LSF and HSF images over either hemisphere. These results are in line with the results of the point-by-point analysis. Over the left hemisphere, the animal-vehicle difference for the ISF images ($M = .61$ µV, $SE = .15$ µV) was larger compared to category differences for LSF ($M = .16$ µV, $SE = .10$ µV; $F(1, 34) = 6.93$, $p=.013$, $\eta p^2 = .17$) and HSF images ($M = .08$ µV, $SE = .11$ µV; $F(1, 34) = 14.42$, $p = .001$, $\eta p^2 = .30$). For the P2 component, category difference was significant for ISF images over the left hemisphere ($M = .61$ µV, $SE = .18$ µV; $t(34) = 3.43$, $p = .018$). Category differences were not significant for LSF and HSF images over either hemisphere. However, category differences over the left hemisphere did not differ between ISF and HSF (ISF: $M = .61$ µV, $SE = .18$ µV; HSF: $M = .36$ µV, $SE = .15$ µV ;$F(1, 34) = 2.29$, $p = .14$, $\eta p^2 = .06$) and between ISF and LSF (LSF: $M = .25$ µV, $SE = .12$ µV; $F(1, 34) = 2.90$, $p = .097$, $\eta p^2 = .08$) images.

In the time-window of the N2 component, we found an interaction between CATEGORY, STIMULUS TYPE and TASK ($F(2, 66) = 4.3$, $p = .018$, $\eta p^2 = .12$). In the AV task, animal-vehicle difference was significant only for ISF images ($M = .73$ µV, $SE = .24$ µV; $t(34) = 3.06$, $p = .042$) but not for the HSF and LSF ones. Whereas in the XO task, category differences were significant for HSF ($M = .50$ µV, $SE = .16$ µV; $t(34) = 3.23$, $p = .03$) and LSF ($M = .81$ µV, $SE = .15$ µV; $t(34) = 5.51$, $p < .001$) images, but not for the ISF ones. In the AV task, category differences differed between ISF and HSF images ($F(1, 34) = 7.59$, $p = .014$, $\eta p^2 = .31$), but not between ISF and LSF ones. In the XO task, animal-vehicle differences differed between ISF and LSF images only ($F(1, 34) = 4.48$, $p = .05$, $\eta p^2 = .22$) (see figure 7).

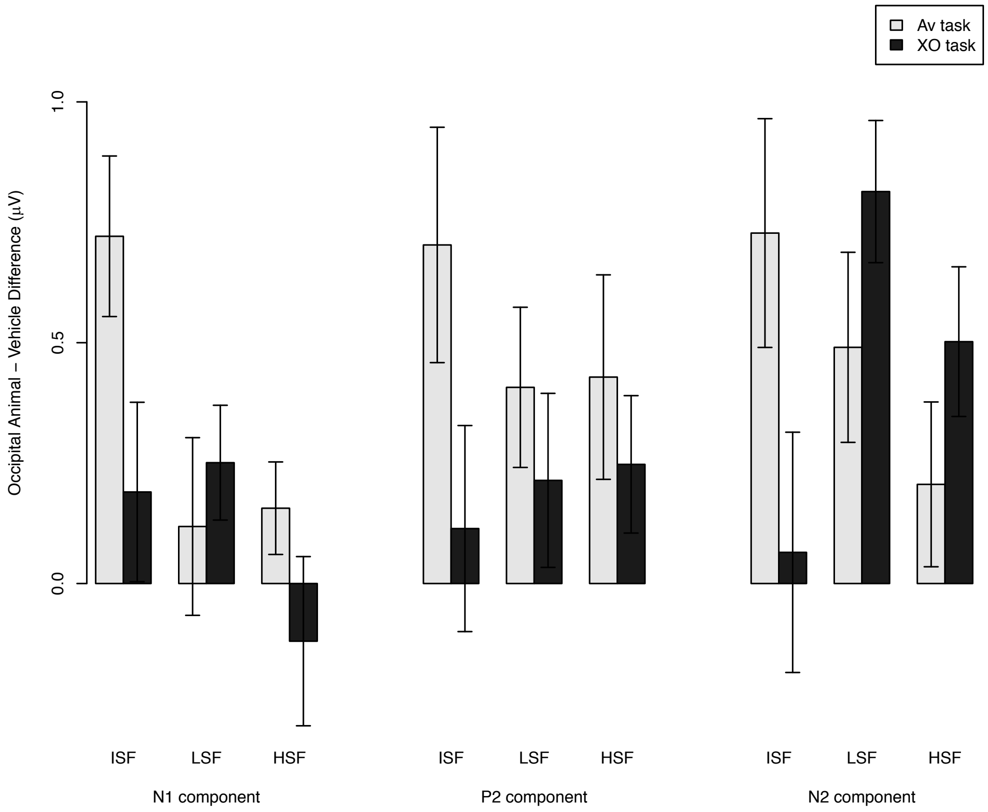


Figure 7. Differences between animal and vehicle amplitudes over the occipital area for ISF, LSF and HSF images in the N1, P2 and N2 components. Error bars represent SE.

Temporal area

In the time window of the N1 component, we found an interaction between the CATGEORY and the STIMULUS TYPE (*F(2, 66*) = 3.25, *p* = .045, ηp^2 = .09). The animal-vehicle difference was significant only for the ISF images (animal: *M* = 1.81 μV, *SE* = .44 μV, vehicle: *M* = 1.4 μV, *SE* = .43 μV; *t(34*) = 4.02, *p* < .001). The ISF category difference was bigger compared to the animal-vehicle differences for the LSF (*M* = .13 μV, *SE* = .10 μV; *t(34*) = 1.21, *p* = .24) and HSF (*M* = .11 μV, *SE* = .10 μV; *t(34*) = 1.15, *p* = .26) images (planned comparisons for ISF vs. LSF: *F(1, 33*) = 5.56, *p* = .024, ηp^2 = .14 and for ISF vs. HSF: *F(1, 33*) =6.17, *p* = .018, ηp^2 = .16).

In the time window of the P2 component, the animal-vehicle differences were modulated by the task only (CATGEORY*TASK interaction: *F(1, 33)* = 4.13, *p* = .05, ηp^2 = .11). The animal vehicle difference was larger in the AV task (*M* = .58 µV, *SE* = .15 µV) than in the XO task (*M* = .24 µV, *SE* = .08 µV), but it was significant in both tasks (AV task: *t(16)* = 3.95, *p* = .001; XO task: *t(16)* = 2.88, *p* = .011). This is in line with the results of the pairwise comparisons, as we found animal-vehicle differences for all stimulus types between 190-240 ms, at least in the AV task (see Figures 4, 5 & 8).

For the N2 component, we only found a main effect of Category (*F(1, 33)* = 43.16, *p* < .001, ηp^2 = .57), suggesting that the category difference was not modulated by the stimulus type or task by this time after stimulus presentation.

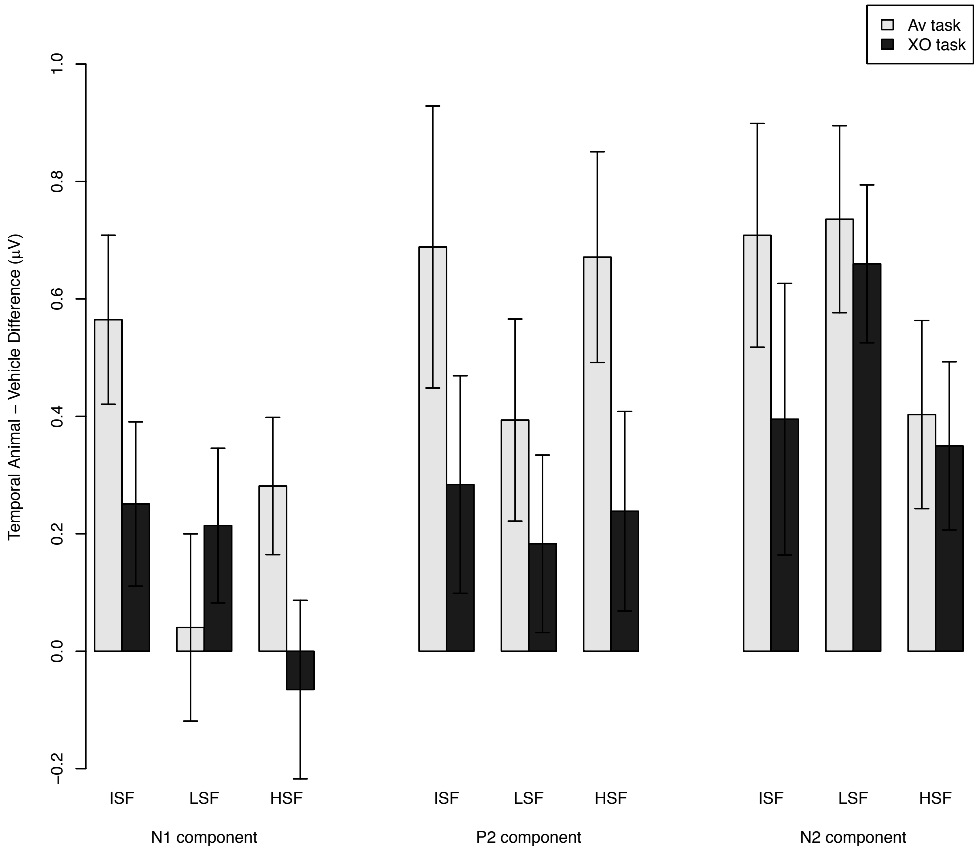


Figure 8. Differences between animal and vehicle amplitudes over the temporal area for ISF, LSF and HSF images in the N1, P2 and N2 components. Error bars represent SE.

Comparison across components

In order to compare the animal-vehicle ERP differences across components, we performed an additional ANOVA including COMPONENT (N1, P2, N2) as a within subject factor beside STIMULUS TYPE (ISF, LSF, HSF), HEMISPHERE (left, right), and CATEGORY (animal, vehicle), and the between subject factor TASK (AV task, XO task) for the occipital and temporal electrodes separately.

Over the occipital area, we found a three-way interaction between COMPONENT, STIMLUS TYPE and CATGEORY (*F(4, 132*) = 4.46, *p* = .02, ηp^2 = .12). For the HSF images, the animal-vehicle difference was not significant for the N1 (*M* = .02 µV, *SE* = .10 µV; *t(34*) = .02, *p* = 2.48), but it was significant both for P2 (*M* = .34 µV, *SE* = .13 µV; *t(34*) = 2.65, *p* = .036) and N2 (*M* = .35 µV, *SE* = .12 µV; *t(34*) = 2.99, *p* = .015). Moreover, the category difference was larger for the P2 than for N1 (*F(2, 68*) = 9.83, *p* = .004, ηp^2 = .22), but it did not differ between P2 and N2 (*F(2, 68*) = .01, *p* = .93, ηp^2 < .001). For the LSF images, the category difference was not significant for the N1 (*M* = .18 µV, *SE* = .11 µV; *t(34*) = 1.66, *p* = .32), and it was significant for the P2 (*M* = .31 µV, *SE* = .12 µV; *t(34*) = 2.57, *p* = .045) and N2 (*M* = .65 µV, *SE* = .13 µV; *t(34*) = 5.16, *p* < .001). The animal-vehicle difference did not differ between N1 and P2, but it did differ between P2 (*F(1, 34*) = 1.79, *p* = .19, ηp^2 = .05) and N2 (*F(1, 34*) = 13.37, *p* = .001, ηp^2 = .28). For the ISF images, animal-vehicle difference was significant for the N1 (*M* = .46 µV, *SE* = .13 µV; *t(34*) = 3.54, *p* = .003), and it was marginally significant for P2 (*M* = .42 µV, *SE* = .17 µV; *t(34*) = 2.48, *p* = .054) and N2 (*M* = .41 µV, *SE* = .18 µV; *t(34*) = 2.27, *p* = .09). The category differences did not differ between components (N1-P2: *F(1, 34*) = .13, *p* = .72, ηp^2 = .004; P2-N2: *F(1, 34*) = .01, *p* = .93, ηp^2 < .001) (see Figure 7).

Over the temporal area, we found a three-way interaction between COMPONENT, STIMULUS TYPE and CATEGORY (*F(4, 132*) = 4.82, *p* = .001, ηp^2 = .13). For the HSF images, the animal-vehicle difference was not significant in the N1 component (*M* = .11 µV, *SE* = .10 µV; *t(34*) = 1.15, *p* = .78), but it was significant in the P2 (*M* = .46 µV, *SE* = .13 µV; *t(34*) = 3.62, *p* = .003) and N2 (*M* = .38 µV, *SE* = .11 µV; *t(34*) = 3.55, *p* = .003) components. Moreover, the animal-vehicle difference was larger in P2 compared to N1 (*F(1, 34*) = 16.94, *p* < .001, ηp^2 = .33) but it did not differ between P2 and N2 (*F(1, 34*) = .67, *p* = .42, ηp^2 = .02). The animal vehicle difference for the LSF images was not significant in the N1 (*M* = .13 µV, *SE* = .10 µV; *t(34*) = 1.21, *p* = .72), and it was significant in the P2 (*M* = .29 µV, *SE* = .12 µV; *t(34*) = 2.54, *p* = .048) and N2 (*M* = .70 µV, *SE* = .10 µV; *t(34*) = 6.76, *p* < .001) components. However, animal-vehicle differences did not differ between the N1 and P2 components (*F(1, 34*) = 3.28, *p* = .08 , ηp^2 = .09), but they differed between the P2 and N2 (*F(1, 34*) = 22.88, *p* < .001, ηp^2 = .40). For the ISF images, animal vehicle differences were significant for all three components (N1: *M* = .41 µV, *SE* = .10 µV; *t(34*) = 4.02, *p* <.001; P2: *M* = .49 µV, *SE* = .16 µV; *t(34*) = 3.18, *p* = .009; N2: *M* = .56 µV, *SE* = .15 µV; *t(34*) = 3.73, *p* = .003), and the category differences did not differ between N1 and P2 (*F(1, 34*) = .46, *p* = .50, ηp^2 = .013), or between P2 and N2 (*F(1, 34*) = .45, *p* = .51, ηp^2 = .013) (see Figure 8).

Discussion

The aim of this study was to investigate animal-vehicle amplitude differences in the visual ERPs in a categorization task containing images with high, low and intact spatial frequencies. In addition, we investigated whether the task-related top-down modulation of category differences in the visual ERPs relies specifically on high or low spatial frequency information, or whether top-down and bottom-up processes interact irrespectively of physical stimulus characteristics. We found animal-vehicle differences already at the N1 component for images with intact SF information. We also found task-related category difference modulation in the time window of the temporal P2. For the P2 component, animal-vehicle differences and their modulation by the categorization task were irrespective of the SF content of the stimuli: for all spatial frequencies, the P2 category difference was larger in the animal vs. vehicle task than in the X vs. O task.

At the time window of the P2 component (from around 200 ms after stimulus onset), animal-vehicle differences occurred independently of the SF content of the images. Moreover, these differences were also modulated by the task. This suggests that by this time category specific processes occurred, and these processes could be affected by top-down modulation. Task related modulation for all stimulus types at the P2 component might mean that animal and vehicle categories are represented separately in different frequency bands. This would mean that different category representations are activated for images containing different SF information. However, it is more likely that both HSF and LSF images activated the same category representations. The results of Goddard et al. (2016) are in line with this explanation. Goddard et al. (2016) presented different everyday objects (faces, body parts, animals, humans, inanimate objects) in HSF and LSF versions in a MEG study. They investigated whether the MEG signal for different objects and conditions could be used to predict the presented stimuli. According to their results, from around 220 ms after stimulus onset, animate-inanimate category differences could be decoded from the MEG signal. Moreover, classifiers trained on HSF trials could predict object identity for LSF trials, and vice versa. This result indicates that object classification for HSF and LSF stimuli rely on similar object representations (Goddard et al., 2016). Results suggesting flexible use of SF information according to task demands also support the idea that category representations are independent of SF information. For further support, it has been shown that category-specific visual areas can flexibly rely on low-level cues. For instance, the activation of the parahippocampal place area (PPA) reflects interaction between SF content and contour of the presented visual scenes. In case of strong contrast, the PPA was more sensitive to LSF information, whereas in case of modified contrast it was more activated by HSF scenes (Kauffmann, Ramanoel, Guyader, Chauvin, & Peyrin, 2015). Therefore, our result indicates that category differences at the time window of the P2 component already reflect high-level visual category representations, and these representations are accessible for top-down processes.

For images containing intact SF information, animal-vehicle differences occurred earlier than for stimuli containing HSF or LSF information only. We found animal-vehicle differences starting at around 110 ms for stimuli with intact SF information in the point-by-point analysis. This suggests that category-specific information is available earlier when stimuli contain broader SF information. In addition, we found more accurate and faster responses for stimuli containing intact SFs compared to high or low SF images. A possible explanation for these results can be that the middle-band SF information is the most informative for discriminating between animals and vehicles. In contrast, it is also possible that animal and object recognition relies on different SFs. Alternatively, this earlier category difference might reflect the benefit of integration of low and high SF information.

A specific role for mid-band frequencies has been suggested for face and object processing both in behavioral (Costen, Parker, & Craw, 1994, 1996) and EEG (Collin, Therrien, Campbell, & Hamm, 2012) studies. Collin et al. (2012) found similar N170 amplitudes for images containing intact and mid-band SFs both for objects and faces. In our study the intact images did contain the middle SF band, whereas HSF and LSF images did not. Therefore, the earlier category differences and task-related modulation for the intact images can be based on the mid-band SFs. Others, on the other hand, have suggested that the processing of different object categories relies on different SF information (Awasthi, Sowman, Friedman, & Williams, 2013; Canário, Jorge, Loureiro Silva, Alberto Soares, & Castelo-Branco, 2016; Harel & Bentin, 2009). In a face vs. building categorization, Awasthi et al. (2013) found faster responses for LSF faces than for HSF ones, whereas for buildings, participants were faster for HSF than LSF images. Furthermore, they found a larger M170 amplitude (the MEG equivalent of the N170) for LSF than HSF faces, whereas amplitudes were larger for HSF compared to LSF buildings. In an fMRI study, Canário et al. (2016) have reported higher activation to LSF compared to HSF stimuli in face and body selective regions, whereas place and object selective areas showed higher activation for HSF than LSF stimuli. Moreover, Harel and Bentin (2009) reported category- and task-specific reliance on different SFs. In that study, participants categorized HSF, LSF and intact airplanes and faces, either at the basic or subordinate level. In basic level categorization, reaction times were similar for HSF, LSF and intact faces, whereas RTs were shorter for intact airplanes compared to HSF and LSF ones. The performance in subordinate categorization was more impaired for LSF airplanes and HSF faces compared to the other conditions. Specifically animal and vehicle categories have not been tested in these studies, still, it is possible that the processing of vehicles overall relies more on HSF information, similarly to objects and airplanes. On the other hand, recognition of animals is very fast; therefore, it can rely on LSF information, similarly to faces. If animal and vehicle recognition relies on different SFs, it could explain the earlier task-related category differences for intact stimuli in our study. Intact stimuli contain all spatial frequencies; therefore, category-specific processes can occur based on different SFs for animals and vehicles, resulting in earlier category differences in the visual N1. However, because of the robust amplitude differences we found between intact vs. LSF and intact vs. HSF images this explanation is unlikely.

Furthermore, the above-mentioned explanations do not take into account the flexibility of visual processes. Participants could categorize all stimulus types quite accurately, and we found category differences in the visual ERPs for all stimulus types by the time-window of the P2 component. This suggests that, even if there is an optimal SF for recognizing animals and vehicles, recognition can occur even if images lack those SFs, and category representations can be activated irrespectively of the physical stimulus characteristics. The results of Halit et al. (2006) suggest that face processing benefits from the presence of broader SF information in the images, instead of relying on a specific frequency band. In that study, beside HSF, LSF and intact stimuli, Halit et al. (2006) used a novel condition which contained both high and low SFs but lacked in the middle band (high/low stimuli). In a face detection task, they found similarly accurate and fast responses for intact and high/low faces but worse performance for LSF and HSF only faces. In addition, the N170 amplitudes for high/low faces were smaller compared to intact faces but larger compared to LSF faces. This finding shows an advantage for faces containing both high and low SF information compared to images containing only high or low SFs. Different SF bands and broader frequencies in an image activate more neurons that are tuned to different SFs in the early visual areas (Collin et al., 2012). Visual areas integrate and accumulate evidence for decision-making processes (Tremel & Wheeler, 2015). The activation of larger numbers of neurons could lead to faster and/or more effective evidence accumulation, related to the recognition task. Therefore, stimuli containing broader scales and more types of low-level information, including, for instance, color, amplitude spectrum, luminance, contrast and spatial frequencies, could be analyzed more effectively. The presence of these different cues in an image can provide more evidence and it can promote faster or stronger activation of category representations in the higher-level areas. Nevertheless, the activation of category representations should be based on the integration of the accumulated evidence. Regarding spatial frequency integration, it has been suggested that such processes start at around 100 ms after the stimulus onset (Kihara & Takeda, 2012; Mu & Li, 2013). Kihara and Takeda (2012) presented intact, high/low (images containing both high and low SFs but lacked in the middle band), HSF and LSF images of scenes, and participants had to decide whether the scenes contained a vehicle or not. Stimuli were presented either for 33 ms, 100 ms or 250 ms. At 100 and 250 ms presentation times, accuracies were higher for intact and high/low images compared to HSF or LSF scenes. Furthermore, for the 100 ms presentation condition, results also indicated that accuracies for high/low images are due to information integration, not only a beneficial effect of wider SF range in the stimuli (probability summation model: Kihara & Takeda, 2012). Mu and Li (2013) used hybrid images of natural and man-made scenes by combing HSF and LSF information from different images belonging to the same or different scene categories. Participants were instructed to categorize the scenes based on the HSF information. Accuracy results suggested that the non-attended LSF information interfered with the processing of HSF information, which indicates attempted integration. Furthermore, the frontal N1 component (peaking at 122 ms) showed LSF interference effect. This suggests the role of frontal areas in SF integration, and that such processes occur as early as 122 ms after stimulus onset (Mu & Li, 2013).

Taken together these possible explanations, our results can be best explained by an evidence integration process, relying on flexible visual features and task demands. Intact images contain more SFs, therefore, evidence regarding object identity can be accumulated from more sources which can result in earlier animal-vehicle ERP differences. For images containing high or low SFs only, on the other hand, evidence accumulation might take longer as there are fewer available sources of information, resulting in ERP category differences only in later time windows. These category differences might reflect visual category representations that are independent of the SF characteristics, and they are accessible for top-down processes.

In summary, we found animal-vehicle amplitude differences for images with intact SFs, as well as for images with high or low SFs only. Furthermore, these category differences were modulated by the categorization task. However, for intact images, category differences occurred earlier, at the time window of the N1 component, whereas for HSF and LSF images such differences appeared only at the time window of the P2. The task-related modulation in the P2 component was independent of the SF content of the stimuli, indicating that the task as a top-down factor affected visual processes at the level of higher-level category representations, where processes are irrespective of the SF content of the image. The presence of broader SFs, and consequently more category specific features in the intact images, could best explain the beneficial effect (earlier category differences and faster and more accurate RTs) of intact SF information.

References

Awasthi, B., Sowman, P. F., Friedman, J., & Williams, M. A. (2013). Distinct spatial scale sensitivities for early categorization of faces and places: neuromagnetic and behavioral findings. *Frontiers in Human Neuroscience*, *7*, 91. http://doi.org/10.3389/fnhum.2013.00091

Balas, B., & Koldewyn, K. (2013). Early visual ERP sensitivity to the species and animacy of faces. *Neuropsychologia*, *51*(13), 2876–2881. http://doi.org/10.1016/j.neuropsychologia.2013.09.014

Bar, M. (2003). A cortical mechanism for triggering top-down facilitation in visual object recognition. *Journal of Cognitive Neuroscience*, *15*(4), 600–609. http://doi.org/10.1162/089892903321662976

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, a M., Schmidt, A. M., … Halgren, E. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(2), 449–454. http://doi.org/10.1073/pnas.0507062103

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. http://doi.org/10.1162/jocn.1996.8.6.551

Bognár, A., Csete, G., Németh, M., Csibri, P., Kincses, T. Z., & Sáry, G. (2017). Transcranial stimulation of the orbitofrontal cortex affects decisions about magnocellular optimized stimuli. *Frontiers in Neuroscience*, *11*, 1–11. http://doi.org/10.3389/fnins.2017.00234

Bullier, J. (2001). Integrated model of visual processing. *Brain Research Reviews*, *36*(2–3), 96–107. http://doi.org/10.1016/S0165-0173(01)00085-6

Canário, N., Jorge, L., Loureiro Silva, M. F., Alberto Soares, M., & Castelo-Branco, M. (2016). Distinct preference for spatial frequency content in ventral stream regions underlying the recognition of scenes, faces, bodies and other objects. *Neuropsychologia*, *87*, 110–119. http://doi.org/10.1016/j.neuropsychologia.2016.05.010

Collin, C. A., Therrien, M. E., Campbell, K. B., & Hamm, J. P. (2012). Effects of band-pass spatial frequency filtering of face and object images on the amplitude of N170. *Perception*, *41*(6), 717–732. http://doi.org/10.1068/p7056

Costen, N. P., Parker, D. M., & Craw, I. (1994). Spatial content and spatial quantisation effects in face recognition. *Perception*, *23*(2), 129–146. http://doi.org/10.1068/p230129

Costen, N. P., Parker, D. M., & Craw, I. (1996). Effects of high-pass and low-pass spatial filtering on face identification. *Perception & Psychophysics*, *58*(4), 602–612. http://doi.org/10.3758/BF03213093

Craddock, M., Martinovic, J., & Müller, M. M. (2013). Task and spatial frequency modulations of object processing: An EEG Study. *PLoS ONE*, *8*(7), 1–12. http://doi.org/10.1371/journal.pone.0070293

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. http://doi.org/10.1016/j.jneumeth.2003.10.009

Flevaris, A. V., Robertson, L. C., & Bentin, S. (2008). Using spatial frequency scales for processing face features and face configuration: An ERP analysis. *Brain Research*, *1194*, 100–109. http://doi.org/10.1016/j.brainres.2007.11.071

Goddard, E., Carlson, T. A., Dermody, N., & Woolgar, A. (2016). Representational dynamics of object recognition: Feedforward and feedback information flows.

NeuroImage, *128*, 385–397. http://doi.org/10.1016/j.neuroimage.2016.01.006

Goffaux, V., Gauthier, I., & Rossion, B. (2003). Spatial scale contribution to early visual differences between face and object processing. *Cognitive Brain Research*, *16*(3), 416–424. http://doi.org/10.1016/S0926-6410(03)00056-9

Goffaux, V., Jemel, B., Jacques, C., Rossion, B., & Schyns, P. G. (2003). ERP evidence for task modulations on face perceptual processing at different spatial scales. *Cognitive Science*, *27*(2), 313–325. http://doi.org/10.1016/S0364-0213(03)00002-8

Goffaux, V., Peters, J., Haubrechts, J., Schiltz, C., Jansma, B., & Goebel, R. (2011). From coarse to fine? Spatial and temporal dynamics of cortical face processing. *Cerebral Cortex*, *21*(2), 467–476. http://doi.org/10.1093/cercor/bhq112

Halit, H., de Haan, M., Schyns, P. G., & Johnson, M. H. (2006). Is high-spatial frequency information used in the early stages of face detection? *Brain Research*, *1117*(1), 154–161. http://doi.org/10.1016/j.brainres.2006.07.059

Harel, A., & Bentin, S. (2009). Stimulus type, level of categorization, and spatial-frequencies utilization: Implications for perceptual categorization hierarchies. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(4), 1264–1273. http://doi.org/10.1037/a0013621

Kauffmann, L., Ramanoel, S., Guyader, N., Chauvin, A., & Peyrin, C. (2015). Spatial frequency processing in scene-selective cortical regions. *NeuroImage*, *112*, 86–95. http://doi.org/10.1016/j.neuroimage.2015.02.058

Kauffmann, L., Ramanoël, S., & Peyrin, C. (2014). The neural bases of spatial frequency processing during scene perception. *Frontiers in Integrative Neuroscience*, *8*, 37. http://doi.org/10.3389/fnint.2014.00037

Kihara, K., & Takeda, Y. (2012). Attention-free integration of spatial frequency-based information in natural scenes. *Vision Research*, *65*(21), 38–44.

http://doi.org/10.1016/j.visres.2012.06.008

Linnert, S., Csifcsák, G., & Tompa, T. (in prep). Top-down modulation of animate-inanimate category differences in the visual N1.

Linnert, S., Reid, V., & Westermann, G. (in prep). Underlying mechanisms of task-related modulation of animal vs. vehicle differences in the visual ERPs.

Luck, S. J. (2005). *An introduction to the event-related potential technique*. Cambridge, Mass: MIT press.

Mu, T., & Li, S. (2013). The neural signature of spatial frequency-based information integration in scene perception. *Experimental Brain Research.*, *227*(3), 367–77. http://doi.org/10.1007/s00221-013-3517-1

Musel, B., Kauffmann, L., Ramanoël, S., Giavarini, C., Guyader, N., Chauvin, A., & Peyrin, C. (2014). Coarse-to-fine categorization of visual scenes in scene-selective cortex. *Journal of Cognitive Neuroscience*, *26*(10), 2287–2297. http://doi.org/10.1162/jocn_a_00643

Oliva, A., & Schyns, P. G. (1997). Coarse blobs or fine edges? Evidence that information diagnosticity changes the perception of complex visual stimuli. *Cognitive Psychology*, *34*(1), 72–107. http://doi.org/10.1006/cogp.1997.0667

Peyrin, C., Michel, C. M., Schwartz, S., Thut, G., Seghier, M., Landis, T., … Vuilleumier, P. (2010). The neural substrates and timing of top-down processes during coarse-to-fine categorization of visual scenes: a combined fMRI and ERP study. *Journal of Cognitive Neuroscience*, *22*(1994), 2768–2780. http://doi.org/10.1162/jocn.2010.21424

Proverbio, A. M., Del Zotto, M., & Zani, A. (2007). The emergence of semantic categorization in early visual processing: ERP indices of animal vs. artifact recognition. *BMC Neuroscience*, *8*, 24. http://doi.org/10.1186/1471-2202-8-24

Rokszin, A. A., Győri-Dani, D., Nyúl, L. G., & Csifcsák, G. (2016). Electrophysiological correlates of top-down effects facilitating natural image categorization are disrupted by the attenuation of low spatial frequency information. *International Journal of Psychophysiology*, *100*, 19–27. http://doi.org/10.1016/j.ijpsycho.2015.12.006

Rossion, B., & Jacques, C. (2008). Does physical interstimulus variance account for early electrophysiological face sensitive responses in the human brain? Ten lessons on the N170. *NeuroImage*, *39*, 1959–1979. http://doi.org/10.1016/j.neuroimage.2007.10.011

Ruiz-Soler, M., & Beltran, F. S. (2006). Face perception: An integrative review of the role of spatial frequencies. *Psychological Research*, *70*(4), 273–292. http://doi.org/10.1007/s00426-005-0215-z

Schyns, P. G. (1998). Diagnostic recognition: task constraints, object information, and their interactions. *Cognition*, *67*, 147–179. http://doi.org/10.1016/S0010-0277(98)00016-X

Schyns, P. G., & Oliva, A. (1994). From blobs to boundary edges: Evidence for time- and spatial-scale-dependent scene recognition. *Psychological Science*, *5*(4), 195–201.

Schyns, P. G., & Oliva, A. (1997). Flexible, diagnosticity-driven, rather than fixed, perceptually determined scale selection in scene and face recognition. *Perception*, *26*(8), 1027–1038. http://doi.org/10.1068/p261027

Schyns, P. G., & Oliva, A. (1999). Dr. Angry and Mr. Smile: when categorization flexibly modifies the perception of faces in rapid visual presentations. *Cognition*, *69*(3), 243–265. http://doi.org/10.1016/S0010-0277(98)00069-9

Torralba, A., & Oliva, A. (2003). Statistics of natural image categories. *Network*, *14*(3), 391–412. http://doi.org/10.1088/0954-898X/14/3/302

Tremel, J. J., & Wheeler, M. E. (2015). Content-specific evidence accumulation in inferior temporal cortex during perceptual decision-making. *NeuroImage*, *109*, 35–49. http://doi.org/10.1016/j.neuroimage.2014.12.072

Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., & Tanaka, J. W. (2010). Controlling low-level image properties: the SHINE toolbox. *Behavior Research Methods*, *42*(3), 671–684. http://doi.org/10.3758/BRM.42.3.671

Linking statement for Paper 4

In the previous papers, we investigated the animal-vehicle ERP differences and their task-related modulations. In both Papers 1 and 3, we found interactions between top-down and bottom-up processes at the level of category representations. fMRI studies suggest that animate and inanimate categories are represented in the higher-level visual area. Nevertheless, brain imaging studies have also found category-specific areas for inanimate object categories, such as tools, chairs, or buildings. It has been suggested that the higher-level visual areas do not categorically represent these categories. Rather, category specificity can be based on perceptual similarity. Indeed, fMRI studies have found evidence for continuous similarity-based object representations in the higher-level visual areas. With this view, category-specific areas would reflect perceptual dissimilarity between inanimate object categories.

Category differences in the visual ERPs are typically investigated using animate and inanimate categories, but ERP differences have also been found between non-animal categories such as cars vs. non-car vehicles or cars vs. shoes and chairs. In line with the suggestion that inanimate object representations are organized by perceptual similarity, a similarity-based mechanism can also explain the category differences reported for the visual ERPs such as between inanimate object categories. However, it remains, an open question what kind of perceptual, similarity-based processes are reflected in the P1 and N1 differences between different inanimate object categories.

The aim of Paper 4 was to investigate ERP differences between tool and vehicle inanimate object categories. Perceptual similarity can be based on the global shape of the objects, or on category-specific object features. Therefore, we presented images of tools and vehicles either in intact version, or in scrambled, still easily recognizable version. In the scrambled version, the global shape was distorted. If category differences for intact and scrambled images are modulated by similar factors in the same ERP components than inanimate objects are likely represented by specific object features. In contrast, if we find dissimilar ERP differences for intact and scrambled objects, that would suggest inanimate object representation based on global shape. Furthermore, tools and vehicles were presented either as toys or as real objects. This enabled us to investigate the task effect on the ERP differences, as participants either performed a tool vs. vehicle categorization task or a toy vs. real object categorization task. Therefore, we could investigate whether top-down and bottom-up processes interact at the level of global shape or at the level of object fragments.

PAPER 4

Early visual ERPs reflect flexible mechanisms in inanimate object discrimination

Szilvia Linnert, Vincent Reid & Gert Westermann,

Lancaster University

Author Note

A version of this paper is currently under review:

Linnert, Reid and Westermann (submitted). The role of configural information in inanimate object categorization. *Neuropsychologia.*

Abstract

Category differences observed in visual event-related potentials could be based on shape- or category-selective processes. Here we investigated the underlying mechanisms of discriminating inanimate object categories in the presence and absence of the global shape. Stimuli contained intact and scrambled images of tools and vehicles that were either real objects or toys. We found evidence for task related (tool-vehicle or real-toy categorization) modulation in the P1 component, which was lateralized in the presence of the global shape. In addition, intact tools elicited left hemispheric N1 differences, suggesting the activation of action-related knowledge. This indicates the role of configural information for the conceptual representation of the tool category. Differential processing of the scrambled real vehicle category, on the other hand, suggests that fragment-based representations may play an important role for the vehicles. Overall, these results indicate both top-down influences and category specific processing mechanisms in the visual ERP components.

Keywords

Categorization, inanimate objects, visual processing, global shape, visual ERPs

Category selective processing mechanisms for inanimate objects reflected in the early visual ERPs

Numerous studies investigating the neuroscience of object categorization have found differences in early visual event-related potentials (ERPs) over posterior regions when participants categorize pictures from different categories (e.g. Antal et al., 2001; Delorme, Rousselet, Macé, & Fabre-Thorpe, 2004; Kiefer, 2001; Proverbio, Del Zotto, & Zani, 2007; VanRullen & Thorpe, 2001). The most consistent category differences have been found in the N1 component (a negative deflection peaking between 100-200 ms poststimulus over posterior sites). However, a few studies have reported category differences even in the preceding P1 component (a positive deflection peaking between 80-120 ms poststimulus over posterior sites) (e.g. VanRullen & Thorpe, 2001). Typically, experiments investigating these effects have used animate and inanimate categories. The animate category usually contains a variety of animal species from insects to mammals, while inanimate stimuli consist of natural and industrial scenes or man-made objects. However, category differences have been also reported for basic categories, such as between faces and cars (for a review see Rossion & Jacques, 2008). Furthermore, beside the animate vs. inanimate categories, N1 differences have been found between non-animal categories such as cars vs. non-car vehicles (VanRullen & Thorpe, 2001), cars vs. shoes and chairs (Rossion et al., 2000), natural and artificial scenes (Paz-Caballero, Cuetos, & Dobarro, 2006), or computer-generated meaningless stimuli (Curran, Tanaka, & Weiskopf, 2002).

Different explanations have been proposed for these category differences. It has been suggested that the N1 differences between animate and inanimate categories reflect either low-level physical differences between stimulus sets such as interstimulus variability (Thierry, Martin, Downing, & Pegna, 2007) or Fourier spectrum (VanRullen, 2011); higher-level category discrimination processes (e.g. Antal, Kéri, Kovács, Janka, & Benedek, 2000) or decision-making processes (Johnson & Olshausen, 2005; VanRullen & Thorpe, 2001). Therefore, despite several possible explanations, the underlying mechanisms for such category differences are not clear. However, it has been suggested that the processing of low-level features (such as the Fourier spectrum) by the early visual areas contributes to category selectivity in the higher-level visual areas (Andrews, Clarke, Pell, & Hartley, 2010). Therefore, differential processing for category specific low-level features can be part of high-level category discrimination processes. This suggests that such category differences in visual ERPs reflect the activation of category representations of some sort.

In order to understand the category differences reflected in the visual ERPs it is important to take into account theories related to visual category representations in the brain. In the following paragraph, we will present neuroimaging results suggesting shape- and/or category-selectivity in the visual brain areas and link those results with EEG studies.

Several studies have reported specialized visual areas in the human lateral occipital complex (LOC) that show greater activation for one object category compared to another. Specific areas have been identified for faces (Kanwisher et al., 1997) and animals (Chao, Haxby, & Martin, 1999) as well as for buildings (Aguirre et al., 1998), chairs (Ishai et al., 1999) and tools (Chao et al., 1999). Such activity differences might reflect distinct categorical representations (Kanwisher et al., 1997). However, it is also possible that only individual objects are represented in the LOC instead of object categories, and that these individual object representations are organized by perceptual similarity (Kriegeskorte et al., 2008). With this view, category-specific areas do not reflect real category representations but arise as an artifact from experiments that use perceptually very different categories (Kriegeskorte et al., 2008). More recent fMRI studies have found evidence both for distinct category-based and for continuous similarity-based representations in the LOC (Kriegeskorte et al., 2008; Mur et al., 2013). Distinct category-based activity patterns were found for animate and inanimate categories. In addition, face and body sensitive activity clusters were found within the animate category, reflecting hierarchical category representations (Kriegeskorte et al., 2008; Mur et al., 2013). Results also indicated that, beside the distinct categorical representation, objects within these categories were represented in a continuous, similarity-based manner (Kriegeskorte et al., 2008). It is important to note that according to the above mentioned results, animals, faces and inanimate objects are categorically represented in the LOC (Kriegeskorte et al., 2008). These are also the most frequently used categories in ERP experiments. Visual ERP differences between these categories might therefore reflect genuine category representations.

Categories, however, are not only visually represented. It is possible that visual category representations are part of a representational network for conceptual categories. The tool category has been widely studied from this point of view. Passive viewing of tools activates not only the tool-selective LOC regions but also left hemispheric motor regions of the brain. Such left lateralized motor activation was found only for tools but not for other objects (Chao & Martin, 2000). This motor activation is due to the activation of action-related knowledge, stored at least partly in the left dorsal visual pathway (Mahon et al., 2007; Martin, 2007). On the basis of these results, it has been suggested that different properties of objects are represented in different brain areas. For instance, sensory information about the object, such as shape, is stored in sensory areas, whereas motor-related information, such as how to use it, is stored in motor areas. In other words, concepts of categories are represented in distinct, but functionally connected networks (for review see Martin, 2007). Therefore, even if the tool-selective LOC area reflects only a perceptual similarity within the tool category (Kriegeskorte et al., 2008), this area could be part of a network that represents the conceptual tool category (Martin, 2007). Therefore, a tool-selective ERP response could also reflect the conceptual representation of the tool category. Indeed, EEG experiments have also shown specific, left-lateralized processing for tools in the somatosensory areas of the parietal lobe (Proverbio, Azzari, & Adorni, 2013; Proverbio, Adorni, & Aniello, 2011). Furthermore, left hemispheric animal-tool differences have been found over the posterior visual areas, however, not for the P1 or N1 but for later components (Sitnikova, West, Kuperberg, & Holcomb, 2006). Therefore, a visual ERP difference for the tool category vs. another inanimate object category could represent conceptual category differences, even in the lack of categorical tool representation in the LOC.

On the other hand, according to the results of Kriegeskorte et al. (2008) inanimate objects in the LOC are represented and organized by perceptual similarity and not in a categorical manner. What object features can play a role in perceived similarity? The global shape of the objects has been suggested to play an important role in the perceptual description of categories (Dilkina & Lambon Ralph, 2012). It has been proposed that topographically organized form representations in the ventral visual cortex are the underlying mechanism of object recognition (Chao et al., 1999; Ishai et al., 1999). Nevertheless, characteristic object parts also seem to play an important role in category descriptions (Dilkina & Lambon Ralph, 2012). It has been suggested that categories are represented as a hierarchy of category-specific features, so called fragments. On this view, features at the level of intermediate complexity (such as a wheel of a car) are optimal for categorization (Ullman, 2007; Ullman, Vidal-Naquet, & Sali, 2002). ERP differences have also been found between different inanimate object categories (Curran, Tanaka, & Weiskopf, 2002; Paz-Caballero, Cuetos, & Dobarro, 2006; Rossion et al., 2000; VanRullen & Thorpe, 2001) for which distinct category representations in LOC have not been found. Even so, ERP differences could also reflect perceptual similarity differences between inanimate object categories, if inanimate objects are represented in a similarity-based manner in the LOC. A recent ERP result suggests that the N1 component can also reflect similarity-based modulation (Tokudome & Wang, 2012). In this study participants were trained to discriminate between four different artificially created highly similar objects from four different viewpoints. Before training, the amplitudes of the N1 component showed viewpoint selectivity. The N1 amplitudes varied between the viewpoints but did not vary between the objects. After training, however, when participants could more easily differentiate between the objects, the amplitude of the N1 component showed shape selectivity, while the viewpoint selectivity disappeared. Therefore, a similarity-based mechanism can also explain the category differences reported for the P1 and N1 visual ERPs.

In sum, different theories have argued that categories can be represented in the visual areas or as part of a representational network, as well as based on perceptual similarity. Furthermore, ERPs can also reflect these processes. Given the described research, it remains an open question whether P1 and N1 differences between different inanimate object categories reflect conceptual category representations or a perceptual, similarity-based visual distinction. Furthermore, if a similarity-based mechanism plays a role in inanimate category discrimination, it remains open whether this is based on global shape or specific object parts. The aim of the present study was to address these questions by investigating the category differences in the visual ERPs between inanimate object categories and the possible underlying mechanisms by comparing intact and scrambled images. We were interested in the role of the global shape and the role of object fragments in category discrimination processes for tool and vehicle categories by analyzing the P1 and N1 visual components. We used images of four object categories: real tools, real vehicles, toy tools and toy vehicles. Tools and vehicles can be differentiated by visual similarity – either by their shape or by specific object parts (e.g. most vehicles have wheels). Toy and real categories, on the other hand, cannot be easily differentiated by visual similarity because they share visually similar exemplars (e.g. a real and a toy car have similar shapes and share specific features such as wheels). Therefore, discriminating toy and real categories minimizes perceptual similarity-related factors.

An object contains both the overall shape and specific object features. In order to investigate the role of object parts, the global shape should be disrupted. Therefore, beside the intact object, we used two types of scrambled images. In scrambled-recognizable images, the overall shape was disrupted yet the objects were still recognizable as the scrambled squares contained identifiable fragments of the objects. Therefore, in the absence of global shape, category discrimination for the scrambled-recognizable images should reflect the role of category-specific objects parts. In scrambled-unrecognizable images, the objects were not recognizable but the low-level image characteristics (such as luminance and color) were retained.

As discussed previously, tools seem to evoke specific, left-hemispheric processing mechanisms that also appear in the visual ERPs (Matheson, Newman, Satel, & McMullen, 2014; Sitnikova et al., 2006). Therefore, we expected the tool-vehicle ERP differences to be modulated between the hemispheres. However, it is possible that in the absence of global shape, tools will not activate action representations, and therefore the use of scrambled-recognizable images might decrease hemispheric differences. On the other hand it has been shown that isolated fragments of faces elicit similar N1 amplitudes to whole faces (Bentin et al., 1996). Therefore, if fragments of tools and vehicles are useful for recognition, the ERPs should show category differences also for the recognizable-scrambled images.

Finally, it is important to note that in a typical ERP experiment participants have to perform a categorization task in which they decide whether a presented stimulus belongs to one category or another. Therefore, the knowledge about the conceptual categories as a top-down factor might also modulate early visual ERPs. Indeed, a recent result showed task-dependent N1 category effects (Balas & Koldewyn, 2013). In this study stimuli were real or doll faces of humans and dogs. Participants performed a real vs. toy and a human vs. dog categorization task on the same stimuli in different blocks of trials. The N1 amplitudes differed for humans and dogs, but only in the human vs. dog categorization task. When participants had to decide about the animacy of the stimuli the human-dog N1 difference disappeared, suggesting a top-down task effect on the N1. In order to further investigate such task effects, in our study the categorization task differed between participants: half of them had to categorize the stimuli as tools and vehicles, while the other half had to categorize them as real objects and toys. These different categorization tasks enabled us to investigate whether the conceptual category, as top-down factor, modulates the category differences in the visual ERPs. Similarly to the results of Balas & Koldewyn (2013), we expected increased tool-vehicle amplitude differences in the tool vs. vehicle categorization task.

Methods

Participants

Thirty-four university students (13 males, 28 right handed), ranging in age from 18 to 26 years (mean=19.77 years, SD=1.81 years) took part in the experiment and received either course credits or were paid 7.5 GBP for their participation. Participants had normal or corrected to normal vision, and normal color vision (tested with Ishiara figures). Four additional participants did not complete the task according to the instructions, as they did not press any button for unrecognizable stimuli. Pressing buttons in one condition and not pressing in others can contribute to the differences between stimulus types, therefore, their data were discarded. Participants were informed about the experiment before they volunteered and signed a consent form. The experiment and procedure were approved by the Lancaster University Ethics Committee.

Stimuli and materials

The stimuli were color images (*N*=238) of easily recognizable objects that belonged to one of the following four categories: real tool (*N*=58), real vehicle (*N*=63), toy tool (*N*=58), or toy vehicle (*N*=59) (Figure 1). Images were presented on a grey background. The images represented a diverse range of tools (e.g. manual and electric screwdrivers, hammers, scissors etc.), and vehicles (e.g. cars, buses, planes, ships etc.). The images were collected from the internet and the objects varied in their size and viewpoint angle. The height and width of the pictures varied according to the contained object. However, the area of the images (number of pixels) was held constant (*M*=66551.38 pixels, *SD*=240.05 pixels), and it did not differ between the stimulus types. The width of the images varied between 4.87 – 10.71 visual degrees (229 – 506 pixels), while the height varied between 2.65 – 6.15 degrees (131 – 291 pixels).

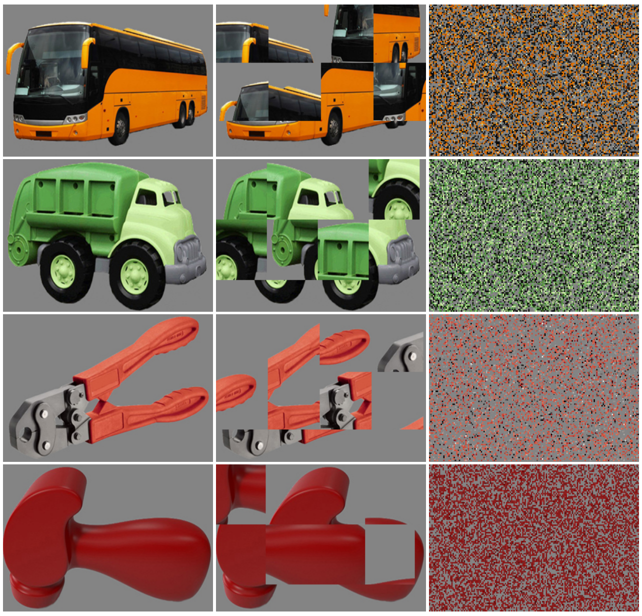


Figure 1. Example stimuli. Intact, scrambled-recognizable (80x80) and scrambled-unrecognizable (2x2) stimulus types: real vehicles, toy vehicles, real tools and toy tools (from top to bottom).

In order to manipulate the level of informative features (especially the shape) three kinds of stimuli were used: easily recognizable, intact images and two scrambled versions of the same stimuli. Scrambled images were constructed by randomly shuffling smaller areas of the image (Photoshop). The sizes of the randomized areas were 80x80 and 2x2 pixels, respectively (see Figure 1). The 80x80 scrambled stimuli were still recognizable, however the global shape of the objects was disrupted. The 2x2 scrambled images were used as control stimuli to see whether low-level features, such as color or relative object size, could help in the categorization. The stimuli (N=238) were presented in intact, 80x80 and 2x2 scrambled versions, thus overall 714 stimuli were presented to the participants.

Stimuli were shown on a 15” CRT monitor with 85 Hz refresh rate and 1024 by 768 pixels resolution. Stimuli were presented using E-Prime software.

Experimental procedure

Participants sat on a chair approximately 80 cm from the screen. They performed a two-choice task to categorize the presented images. All participants saw all four image categories (real tool, real vehicle, toy tool, and toy vehicle), however, they were randomly assigned to one of two groups. One group had to decide whether the presented image was a toy or a real object (N=17), while the other group had to report whether a tool or a vehicle was shown on the screen (N=17). Participants had to categorize also the scrambled stimuli and they were asked to guess if they were not sure. Response accuracy and reaction time data was collected together with the recorded EEG.

The stimulus presentation was around 30 minutes long with a break after every 102 stimuli (around every 4 minutes). The break was as long as the participant wanted and the test proceeded when they pressed a button.

A trial started with a fixation cross displayed in the middle of a grey screen for a random duration between 200 and 350 ms. Then a stimulus appeared in the middle of the grey screen for 1000 ms. The order of stimuli was randomized. The presentation order of different versions of the same image (intact, 80x80, 2x2) and the intervals between their appearances were unpredictable. At the end of each trial a blank grey screen was shown for 500 ms. Participants' category choices were made by using two buttons on a response box. Within each participant group the left and right buttons were counterbalanced across the participants: e.g. half of the participants had to press the left button for the tools and the right button for the vehicles, while the other half pressed the button in the opposite distribution.

EEG recording and analysis

We used an EGI (Electrical Geodesics Incorporated, Eugene, OR) GES300 EEG system with a NetAmps USB amplifier and a 128 channel HydroCel Net to record the EEG from the scalp. The EEG was recorded and analyzed with NetStation software (Electrical Geodesics Incorporated, Eugene, OR). Electrode impedance was kept below 50 kOhm. During the recording, the EEG was referenced to Cz, and the sampling rate was 1000 Hz.

During off-line analyzis the raw data were band-pass filtered between 0.1 - 30 Hz. The EEG was segmented from 300 ms before to 1500 ms after stimulus onset. Epochs were sorted into the four object categories and three stimulus types (intact, 80x80 scrambled, 2x2 scrambled). In case of the intact and the 80x80 scrambled images only epochs containing correct responses were included. For the 2x2 scrambled stimuli, epochs were included only if one of the buttons was pressed, irrespectively of whether the response was correct or incorrect. Eye movements and bad channels were automatically rejected if the average amplitude in an 80 ms moving window exceeded ±55 μV at EOG channels or ±200 μV at any other channel. In addition, each individual epoch was visually inspected and further epochs or channels were rejected. Epochs in which more than 13 channels contained artifacts were rejected; if fewer than 13 channels contained artifacts, bad channels were substituted by spline interpolation. Participants with fewer than 22 artifact-free epochs per condition were excluded.

Average ERPs were calculated from the artifact-free segments for each condition. ERPs were baseline corrected using a 150 ms pre-stimulus baseline, and were re-referenced to an average reference. Grand averages were plotted using the MATLAB-based EEGLAB (*v.* 13_4_4b) toolbox (Delorme & Makeig, 2004).

Results

Behavioral results

For each participant, the proportion of correct responses (%) and the mean reaction times (excluding reaction times lower than 200 ms and higher than 1000 ms) were calculated for the stimulus types. 3x2x2x2 ANOVAs were performed on the correct response proportions and the reaction times (SPSS Statistics software - version 22) including STIMULUS TYPE (intact, 80x80 scrambled and 2x2 scrambled), FUNCTION (toy, real) and OBJECT (tool vehicle) as within-subject factors and TASK (tool vs. vehicle and real object vs. toy categorization) as between-subject factor. A Greenhouse-Geisser correction was applied on the p-values for violations of the assumption of sphericity. Bonferroni corrections were applied for multiple comparisons.

The accuracy results are presented in Figure 2. We found an interaction including all factors (STIMULUS TYPE * FUNCTION * OBJECT * TASK: $F(2, 62) = 3.84$, $p = .027$, $\eta p^2 = .11$). Separate ANOVAs were performed on the three stimulus types. For both the intact and 80x80 images the accuracy for real objects were higher in the tool-vehicle task than in the toy-real task, while the accuracy for toys did not differ between the categorization tasks (FUNCTION * TASK interaction: intact: $F(1, 31) = 9.76$, $p = .004$, $\eta p^2 = .24$; 80x80: $F(1, 31) = 16.30$, $p < .001$, $\eta p^2 = .35$). For the intact images the accuracy was higher for the toy tools compared to the real ones, but the accuracy for the vehicles did not differed in terms of function (OBJECT * FUNCTION interaction: $F(1, 31) = 17.77$, $p < .001$, $\eta p^2 = .36$). The accuracy for the 80x80 images showed different pattern – the participants were more accurate for the real vehicles than for toys, but no such difference were found for the tools (OBJECT * FUNCTION interaction: $F(1, 31) = 15.10$, $p = .001$, $\eta p^2 = .33$). For the 2x2 stimuli we found only a main effect of TASK ($F(1, 31) = 5.96$, $p = .021$, $\eta p^2 = .16$); accuracy in the toy vs. real categorization task ($M = 58.78\%$, *SE* = 2.58%) was higher than in the tool vs. vehicle categorization task (*M* = 51.62%, *SE* = 1.49%). The accuracy in the toy vs. real task was above chance (one sample *t*-test against 50% chance level: *t*(15) = 3.412, *p* = .004) but it did not did not differ from chance in the tool vs. vehicle task (*t*(16) = 1.09, *p* = .29).

The average correct response proportions differed between the three stimulus types (main effect of STIMULUS TYPE: *F*(2, 62) = 528.80, *p* < .001, ηp^2 = .95). The ratio of correct responses for the intact images (*M* = 93.38%, *SE* = 1.02%) was higher than for the 80x80 stimuli (*M* = 89.52%, *SE* = 1.04%) (within subject contrast: *F*(1, 31) = 64.30, *p* < .001, ηp^2 = .68). The accuracy for the 2x2 images (*M* = 55.20, *SE* = 1.47) was significantly lower compared to the 80x80 images (within subject contrast: *F*(1, 31) = 495.76, *p* < .001, ηp^2 = .94).

No significant differences were found for the reaction times (Figure 3).

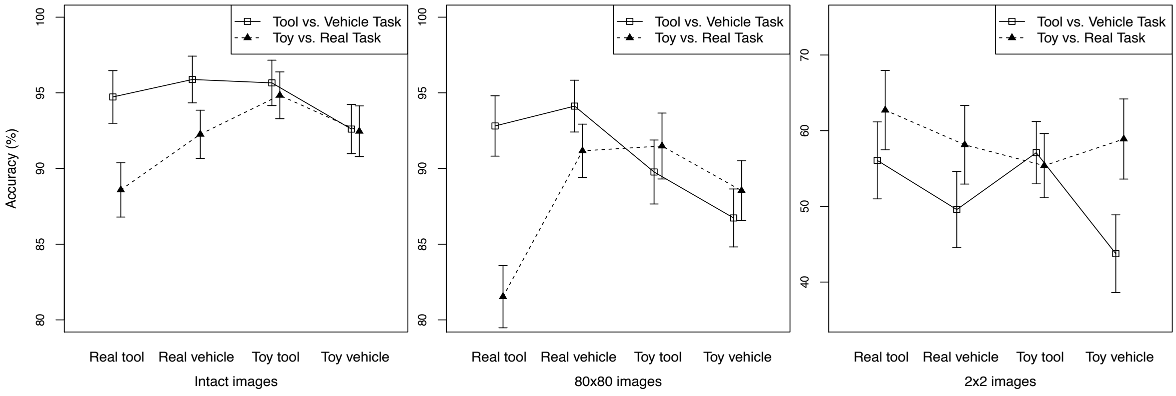


Figure 2. Accuracy rates for the intact (left), 80x80 (middle) and 2x2 (right) stimuli in the tool vs. vehicle task (continuous line) and in the toy vs. real task (dotted line) for the four stimulus types. The error bars represent the standard error.

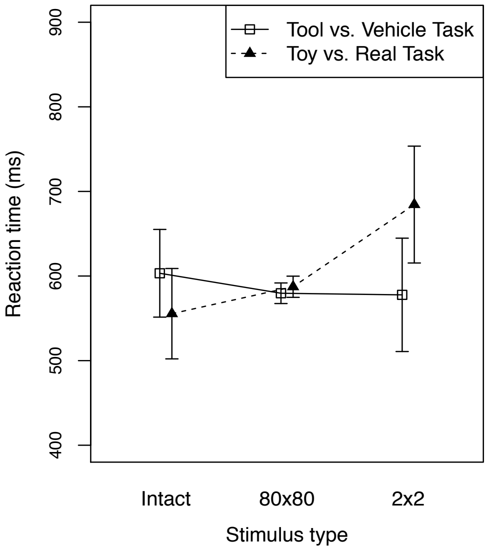


Figure 3. Reaction times for the intact, 80x80 and 2x2 stimuli in the tool vs. vehicle task (continuous line) and in the toy vs. real task (dotted line). The error bars represent the standard error.

EEG results

Category differences in the visual ERPs are usually measured on electrodes P7 and P8 (e.g. Balas & Koldewyn, 2013; Bentin et al., 1996; Proverbio et al., 2007). We analyzed the ERPs on the EGI equivalent of P7, P8 and surrounding electrodes over the left and right temporal areas: 58 (equivalent to P7), 65 (equivalent to PO7) and 59 over the left hemisphere; 96 (equivalent to P8), 90 (equivalent to PO8), and 91 over the right hemisphere. The scalp maps of the vehicle-tool differences also justify the selection of these particular electrodes (see Figure 4). The mean amplitude of the P1 and N1 visual ERP components were analyzed. Time windows were defined between 80-120 ms for the P1 component and between 140-180 ms for the N1 component. The time windows were based on visual inspection of the grand average ERPs, and these time windows fitted well with the literature. Repeated measures ANOVAs were performed on the mean amplitudes for the P1 and N1 components, respectively, by using SPSS Statistics software (version

22). The within subject factors were the STIMULUS TYPE (intact, 80x80 scrambled and 2x2 scrambled), OBJECT (tool and vehicle), and FUNCTION (real object and toy). As the functions are not equally distributed between the hemispheres, an additional within-subject factor was the HEMISPHERE (left and right). The between subject factor was the TASK (tool vs. vehicle and real object vs. toy categorization). A Greenhouse-Geisser correction was applied on the p-values for violations of the assumption of sphericity. Bonferroni corrections were applied for multiple comparisons.

The ERP waveforms for the object categories and stimulus types are presented in Figure 5. The aim of the study was to investigate the category differences and the possible underlying mechanisms by comparing intact and scrambled stimulus types. Therefore, we analyzed only the highest-level significant interactions containing the STIMULUS TYPE, OBJECT and/or FUNCTION factors. In case of significant interactions, separate ANOVAs were performed as post-hoc tests on the stimulus types using the factors involved in the interaction.

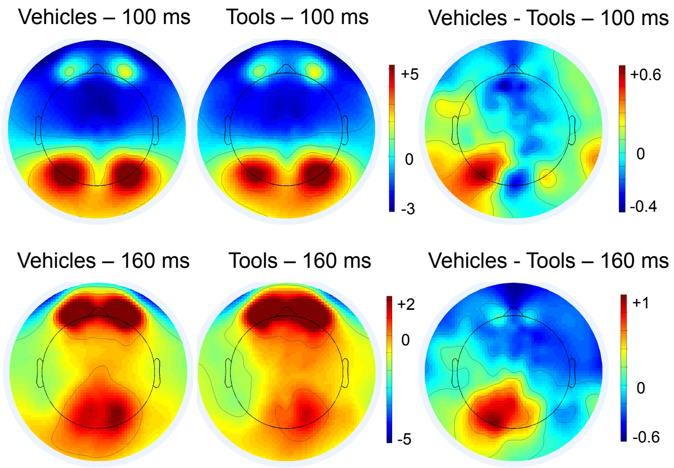
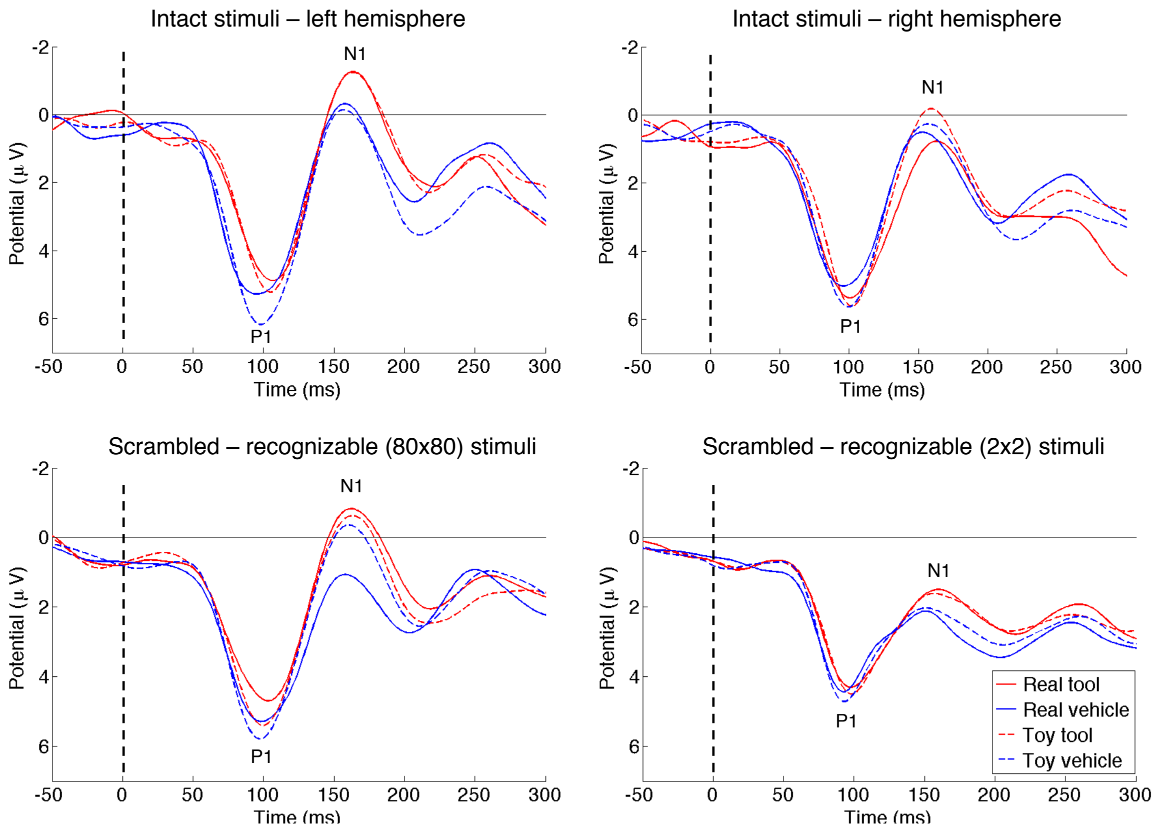


Figure 4. Scalp maps for the intact vehicles (left column), tools (middle column) and for their difference (right column) at the times of the P1 (top row) and N1 (bottom row) peaks. These maps show the average activation in the two categorization tasks.



Intact stimuli – left hemisphere

Intact stimuli – right hemisphere

Scrambled – recognizable (80x80) stimuli

Scrambled – recognizable (2x2) stimuli

Figure 5. ERP waveforms. The red lines represent the tools, the blue lines represent the vehicles, the continuous lines represent the real objects, and the dashed lines represent the toys. The first row shows the ERPs for the intact stimuli in the tool vs. vehicle categorization task over the left and right hemispheres. The second row shows the ERPs for the scrambled – recognizable (left) and scrambled – unrecognizable (right) images. The waveforms for the scrambled stimuli are the average of the two hemispheres and the two categorization tasks.

Vehicle – tool contrast

For the P1 component the highest level significant interaction was found involving the STIMULUS TYPE, OBJECTS, HEMISPHERE and TASK factors ($F(2, 64) = 3.57$, $p = .034$, $\eta p^2 = .10$) (Figure 6). The post hoc ANOVAs were conducted separately on the stimulus types. For the intact images, we found an interaction between OBJECT, HEMISPHERE and TASK ($F(1, 32) = 8.29$, $p = .007$, $\eta p^2 = .21$). The further tests revealed that for the intact images the vehicle-tool contrast was bigger for the left hemisphere than for the right, but only in the tool vs. vehicle categorization task ($t(16) = 3.80$, $p = .002$) and not in the toy vs. real task ($t(16) = -.10$, $p = .93$) (Figure 7).

No such effects were found for the 80x80 and 2x2 scrambled images. Thus, hemispheric lateralization of category differences is task-dependent and therefore, likely modulated by a top-down process. As this task-related lateralization only exists for intact objects it could be caused either by global shape information or by abstract category representations.

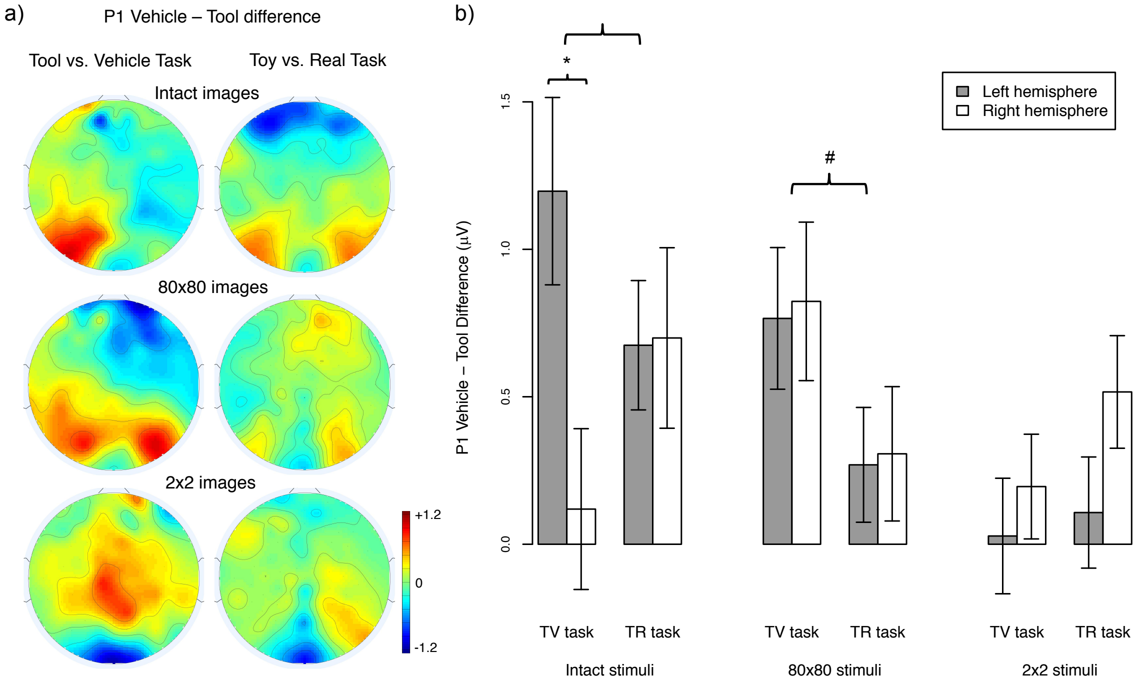


Figure 6. a) Scalp distribution for the vehicle-tool amplitude differences at 80-120 ms for the intact, 80x80 and 2x2 images (from top to bottom) in the tool vs. vehicle task (left column) and in the toy vs. real task (right column). b) P1 vehicle-tool amplitude differences for the intact, 80x80 and 2x2 stimuli (from left to right) over the left (dark columns) and right (white columns) hemispheres in the tool vs. vehicle (TV task) and in the toy vs. real (TR task) categorization task. The error bars represent the standard error. We found a hemispheric difference for the intact stimuli, but only in the tool vs. vehicle task. The 80x80 images showed a marginally significant task effect (#: p=0.051). For the 2x2 images we found a significant tool-vehicle amplitude difference, however, it was not modulated by the task or the hemisphere.

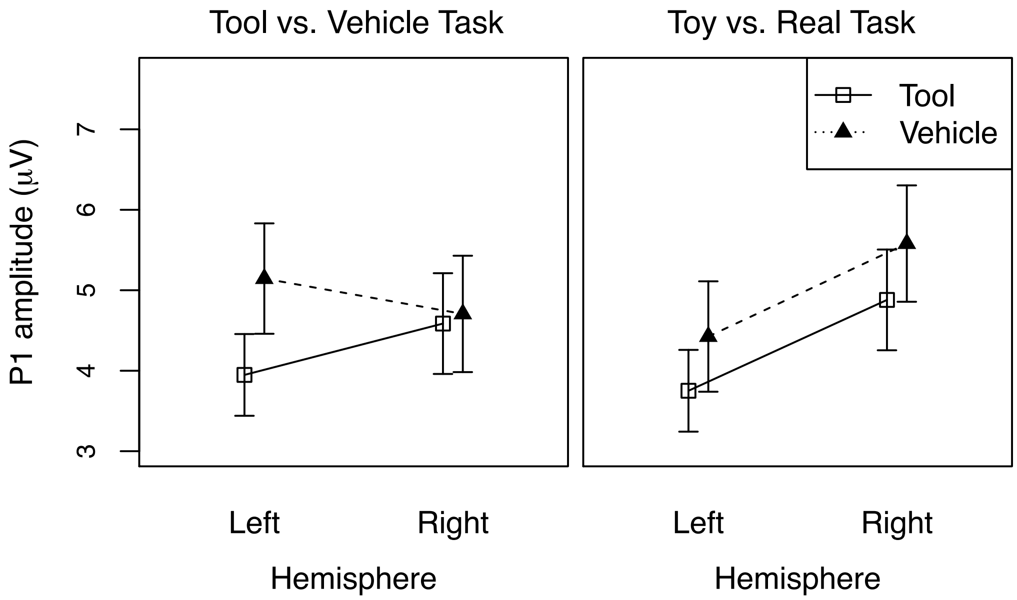


Figure 7. P1 amplitudes for intact tools and vehicles in the tool vs. vehicle task (left panel) and in the toy vs. real task (right panel) over the left and right hemispheres. The error bars represent the standard error. The amplitudes for tools and vehicles differed between the hemispheres but only in the tool vs. vehicle categorization task (left panel). No hemispheric differences were found for the toy vs. real task (right panel).

Regarding the question of whether object features, without the presence of the global shape, lead to category differences or task-related modulation, we found only a main effect of OBJECT for the 80x80 images ($F(1, 32) = 18.70$, $p < .001$, $\eta p^2 = .37$). Figure 4 also indicates that for the 80x80 images the vehicle-tool contrast was bigger in the tool vs. vehicle categorization task than in the toy vs. real categorization task. However, statistically this effect was only marginally significant (OBJECT*TASK interaction for the 80x80 images: $F(1, 32) = 4.10$, $p = .051$, $\eta p^2 = .11$). Therefore, recognizable parts of vehicles and tools do elicit category differences, and it is possible that top-down processes also modulate the processing of these object parts.

For the 2x2 images no significant interactions were found. The ANOVA revealed only a main effect the OBJECT ($F(1, 32) = 4.52$, $p = .041$, $\eta p^2 = .12$). The P1 amplitudes were bigger for vehicles ($M = 3.96$, $SE = 0.40$) than for tools ($M = 3.78$, $SE = 0.51$). The

2x2 images were not recognizable, therefore, the vehicle – tool differences for this stimulus type likely reflect low-level differences between the stimulus sets. However, for the intact images the vehicle-tool contrast was modulated by the hemisphere and task. Furthermore, the cortical maps for the vehicle – tool differences for the 2x2 images show a more medial distribution (Figure 6a). These results suggest that despite the low-level stimulus differences between the stimuli, the presence of the global shape influenced the early visual processing of the stimuli.

In case of the N1 component a STIMULUS TYPE*OBJECT*HEMISPHERE interaction was found ($F(2, 64) = 4.04$, $p = .022$, $\eta p^2 = .112$) (Figure 8). The post-hoc test indicated that this effect was driven by the intact stimuli. The category difference was bigger and only present over the left hemisphere for the intact images (OBJECT*HEMISPHERE interaction for the intact images: $F(1, 33) = 7.12$, $p = .012$, $\eta p^2 = .18$). No hemispheric difference was found for the 80x80 and 2x2 scrambled stimuli. Only a main effect of OBJECT was found for both the 80x80 ($F(1, 33) = 25.07$, $p < .001$, $\eta p^2 = .43$) and 2x2 ($F(1, 33) = 10.77$, $p = .002$, $\eta p^2 = .25$) images, where the N1 amplitudes for vehicles were bigger than the amplitudes of tools. Furthermore, similarly to the P1 category difference, the vehicle-tool difference is distributed more medially for the 2x2 images (Figure 7a).

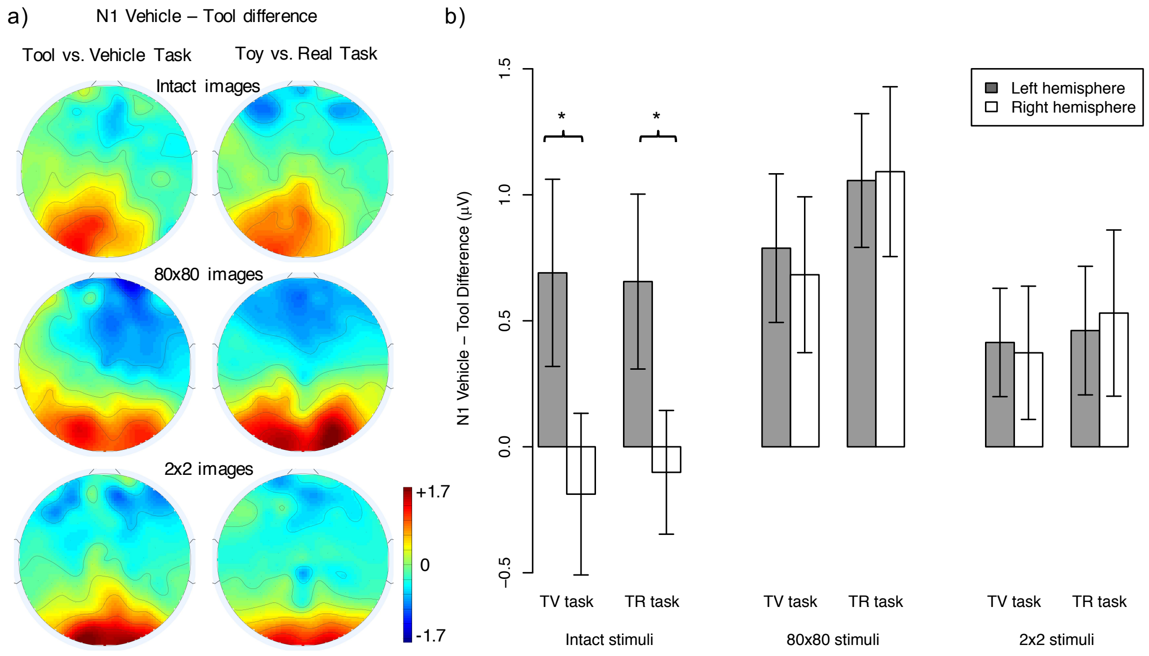


Figure 8. a) Scalp distribution for the vehicle-tool amplitude differences between 140-180 ms for the intact, 80x80 and 2x2 images (from top to bottom) in the tool vs. vehicle task (left column) and in the toy vs. real task (right column). b) N1 vehicle-tool amplitude differences for the intact, 80x80 and 2x2 stimuli (from left to right) over the left (dark columns) and right (white columns) hemispheres in the tool vs. vehicle (TV task) and in the toy vs. real (TR task) categorization task. The error bars represent the standard error. We found a hemispheric difference for the intact images, irrespectively of the task. The vehicle-tool amplitude differences for the 80x80 and 2x2 images were not modulated by the hemisphere or the task.

Similarly to the P1 component, the N1 category difference for the intact images revealed a left hemispheric discrimination processes in the presence of the global shape. Left lateralized representations have been associated with the tool category (Mahon et al., 2007; Martin, 2007). Therefore, we performed further post-hoc tests to see whether the left hemispheric vehicle-tool contrasts were due to the tool category. For the P1 component, the left lateralized difference was not due to a single category but was driven by the task (Figure 7). The P1 amplitudes did not differ between the hemispheres either for the tools or for the vehicles, irrespective of the categorization task. In contrast, the left hemispheric category difference in the N1 was clearly driven by the tools (Figure 9). The

N1 amplitudes for tools differed between the left and right hemisphere (*t*(33) = 2.91, *p* = .012), whereas no such difference was found for the vehicles.

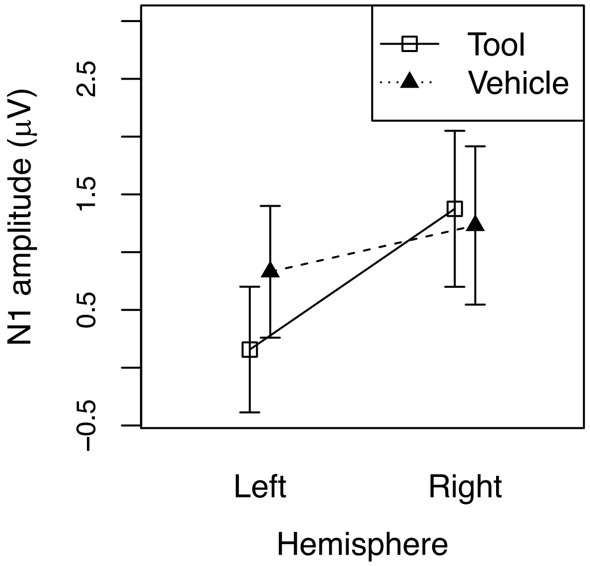


Figure 9. N1 amplitudes for intact tools and vehicles over the left and right hemispheres. The error bars represent the standard error. A category difference was found only over the left hemisphere. Furthermore, the N1 amplitudes differed between the two hemispheres only for the tools.

Real - toy contrast

Regarding the real-toy contrast the P1 component showed only a significant main effect of FUNCTION (*F*(2, 64) = 12.59, *p* = .001, $\eta p^2 = .28$) (Figure 10). However, this P1 effect did not differ between stimulus types, suggesting low-level differences between the toy and real stimuli.

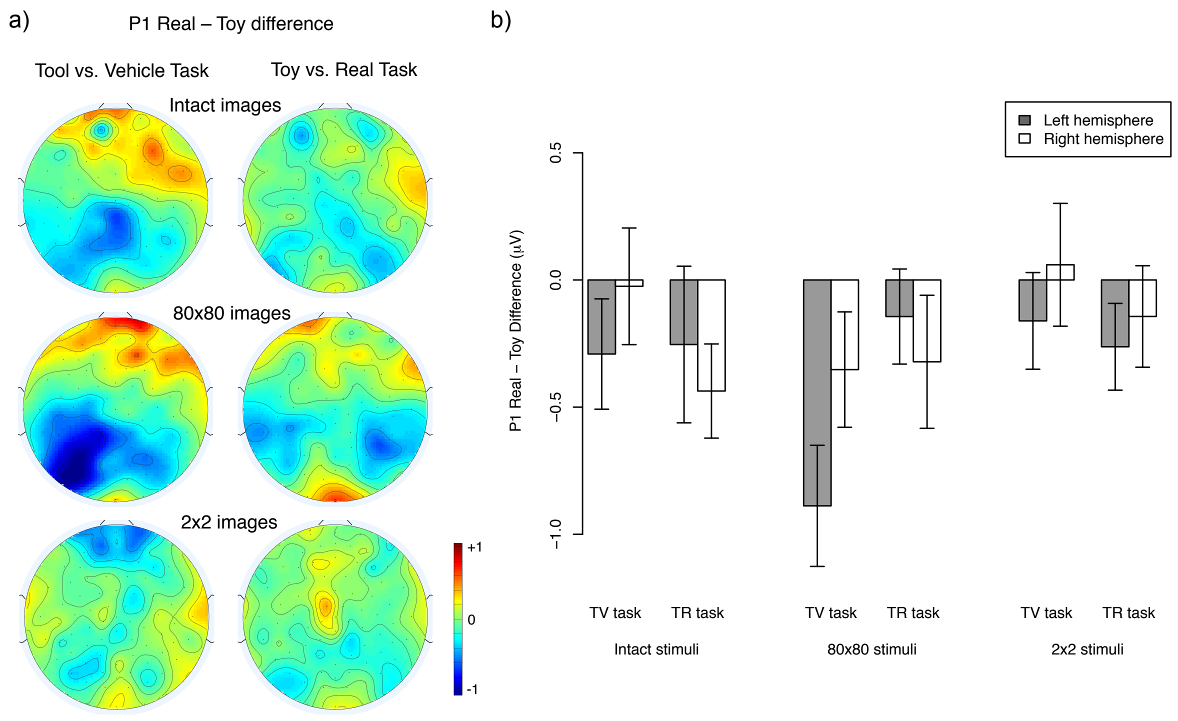


Figure 10. a) Scalp distribution for the real-toy amplitude differences between 80-120 ms for the intact, 80x80 and 2x2 images (from left to right). b) N1 real-toy amplitude differences for the intact, 80x80 and 2x2 stimuli (from left to right) over the left (dark columns) and right (white columns) hemispheres. The error bars represent the standard error.

For the N1 component a STIMULUS TYPE*FUNCTION*HEMISPHERE interaction was found ($F(2, 64) = 3.42$, $p = .039$, $\eta p^2 = .10$) (Figure 9). For the intact images, the real-toy contrast was bigger (and significant only $t(33) = 4.22$, $p < .001$) over the right hemisphere than over the left (FUNCTION*HEMISPHERE interaction for the intact images: $F(1, 33) = 14.80$, $p = .001$, $\eta p^2 = .31$) (Figure 11). Such a hemispheric effect was not found for either of the scrambled stimuli. On the other hand, Figure 11 shows some hemispheric effect also for the 80x80 images. It is not sufficient to conclude a difference by only looking at the significant category difference pattern (Nieuwenhuis et al., 2011). Comparing the hemispheric effect for the intact and 80x80 images indicated no significant difference (within subject contrast: $F(1,32) = 2.89$, $p = .099$, $\eta p^2 = .10$). Therefore, based on these results it is not clear whether the right hemispheric advantage in discriminating toys and real objects is driven by the global shape or can also be elicited by object parts.

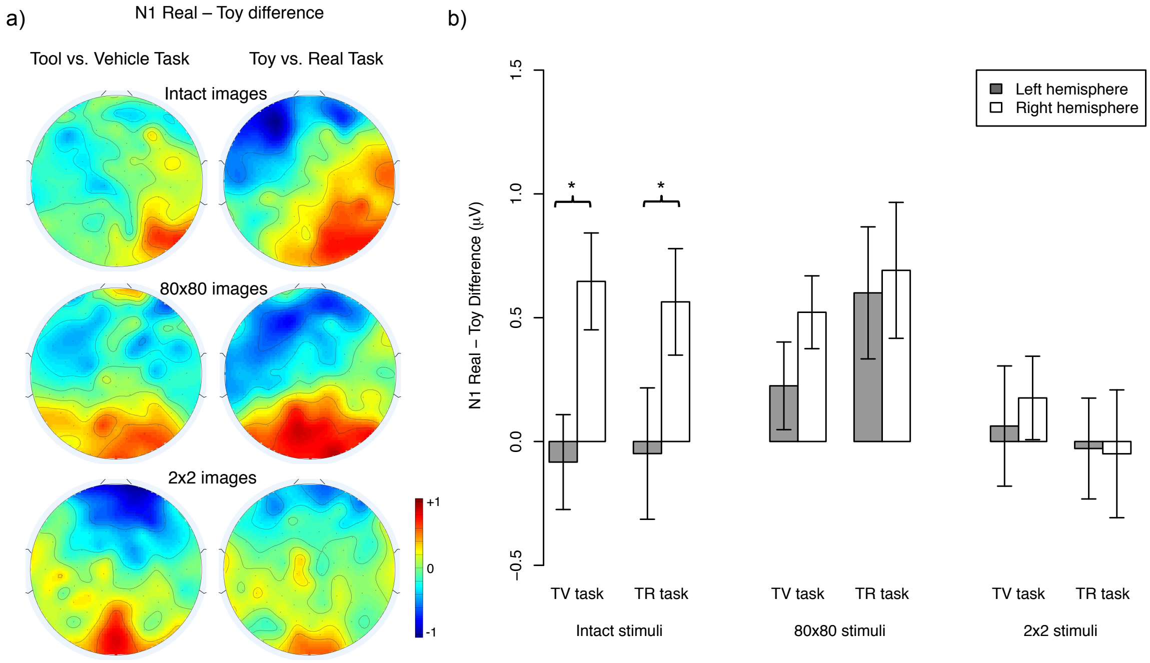


Figure 11. a) Scalp distribution for the real-toy amplitude differences between 140-180 ms for the intact, 80x80 and 2x2 images (from left to right). b) N1 real-toy amplitude differences for the intact, 80x80 and 2x2 stimuli (from left to right) over the left (dark columns) and right (white columns) hemispheres. The error bars represent the standard error. We found a hemispheric effect only for the intact images. However, statistically the hemispheric effect did not differ between the intact and 80x80 images.

The main effect of FUNCTION was found only for the 80x80 images ($F(1, 33) = 18.64$, $p < .001$, $\eta p^2 = .36$) but not for the 2x2 stimuli. This indicates that low-level differences did not contribute towards the toy-real N1 differences.

Interaction between objects and function

For the N1 component the significant STIMULUS TYPE * OBJECT * FUNCTION * TASK interaction ($F(2, 64)=3.45$, $p = .038$, $\eta p^{2} = .10$) was also found. However, the post-hoc tests did not indicate any task-related modulation for either of the stimulus types. On the other hand, we found a significant interaction between OBJECT and FUNCTION for the 80x80 images ($F(1, 32) = 24.52$, $p < .001$, $\eta p^2 = .43$). Figure 12 shows the N1 amplitudes for the 80x80 images. The N1 amplitude for the real vehicles was significantly smaller than for the other three object types (real vehicle vs. real tool: $t(33) = 7.04$, $p < .001$; real vehicle vs. toy vehicle: $t(33) = 6.48$, $p < .001$; real vehicle vs. toy tool: $t(33) = 6.57$, $p < .001$). No OBJECT*FUNCTION interaction was found for the intact and 2x2 stimuli. This interaction reveals that both the vehicle – tool and the real – toy contrast for the 80x80 images were based on the large N1 difference for the real vehicles. Therefore, they might not reveal real discrimination processes between categories.

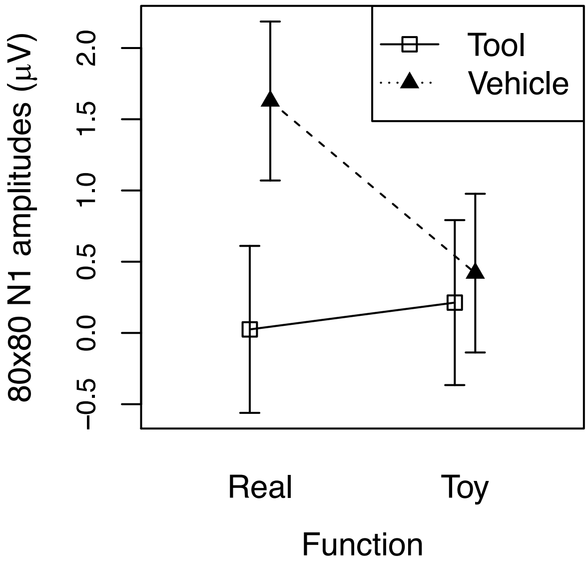


Figure 12. N1 amplitudes for the 80x80 images. The N1 amplitudes for the real vehicles differed from the amplitudes for the other three object categories.

Discussion

The aim of the study was to investigate underlying mechanisms of perceptual similarity-based discrimination processes for inanimate object categories. The main findings are the following. (1) In the presence of the global shape, the vehicle-tool differences in the N1 component were left-lateralized and this effect was driven by the tool category. (2) When the global shape was distorted, real vehicles elicited different N1 amplitudes compared to the other three stimulus types. (3) We found task related top-down modulation in the P1 for both types of recognizable stimuli. However, the pattern of such top-down effects differed for the images containing global and distorted shape. (4) Low-level differences were present between the tool and vehicle images. However, neither the left lateralized discrimination processes for the intact images or the specific N1 difference for scrambled recognizable real vehicles were based on low-level differences.

The category differences showed dissimilar patterns depending on whether the global shape was present or distorted, which suggests different processing mechanisms for whole objects and object parts. In the presence of global shape, the vehicle and tool discrimination process was left lateralized, and this effect was driven by the tool category. Left lateralized visual ERP differences have been reported for tool vs. animal categories, however, only for later ERP components than those investigated here (Sitnikova et al., 2006). In that experiment, however, participants had to perform a naming task instead of direct category discrimination. Therefore, the different task might explain the early ERP differences in our study. Furthermore, the hemispheric difference for the N1 component in our study was clearly due to the tool category. The left lateralized processing for the tools is not surprising. It has been shown that tools evoke specific action-related processes in the left hemisphere (Chao & Martin, 2000; Mahon et al., 2007; Martin, 2007). If the attributes of categories are represented in functionally connected networks (Martin, 2007), such left-lateralized motor activation might also influence visual processing. Indeed, it has been shown that information processing in the ventral visual areas is modulated by the activation of additional brain areas. The dorsal areas, where action-related attributes are processed, can influence the processing by object recognition areas (for review see Cloutman, 2013). Therefore, it is possible the hemispheric differences for tool processing are based on the activation of left-lateralized action knowledge. This interpretation is in line with the interpretation of left-hemispheric animal-tool ERP differences reported by Sitnikova et al., (2006). This would suggest that the N1 amplitude reflects not only visual processing, but that it is modulated by the processing of other, object-related attributes.

However, it is important to note that in our study left hemispheric processes were present only when the global stimulus shape was available. The vehicle-tool difference for the scrambled-recognizable images showed specific processing mechanisms for the real vehicles, and this process were not lateralized. Nevertheless, participants were able to recognize tools with high accuracy even if the global shape was distorted. Therefore, object fragments are informative enough for categorization; however, the different involvement of the left hemisphere for intact and scrambled images suggests different processing mechanisms for whole objects and their fragments. One possibility is that N1 differences for scrambled objects do not reflect real category discrimination processes and consequently, conceptual category representations. Alternatively, it is possible that tool fragments alone are not enough to activate action-related knowledge. Finally, in theory, ERP responses could be affected by the presences of horizontal and vertical lines in the 80x80 images. However, the presence of horizontal and vertical lines unlikely affects the real vehicle category selectively as we found in the N1.

If left-hemispheric processes for the intact images reflect category representations, then the lack of lateralization for the scrambled images might indicate that fragments of objects do not activate conceptual representations at the time of the N1 component. Indeed, we found smaller N1 amplitudes for real vehicles compared to the other object categories when the global shapes were distorted. This shows that the category differences for both the vehicle-tool and toy-real categories were due to the real vehicles. This indicates that category N1 differences for the scrambled images do not reflect category discrimination processes. It has been suggested that the larger N1 amplitude can be interpreted as enhanced visual analysis (J. Tanaka, Luu, Weisbrod, & Kiefer, 1999) or more difficult perceptual encoding (Jacques & Rossion, 2007). It is possible that fragments of real vehicles have higher diagnostic value than fragments of tools or even toy vehicles. A similar mechanism has been suggested for categorizing images of natural scenes (J. Vogel & Wallraven, 2006). In their study, the role of image fragments varied for different scene categories. Participants were similarly accurate in categorizing intact and scrambled versions of forest and plain scenes, whereas accuracy rates decreased for scrambled images of river and lake scenes compared to their intact versions. This result suggests that the vehicle-tool and toy-real amplitude differences in our study might not reflect real discrimination processes for the scrambled images. Rather, the smaller N1 amplitude for real vehicle fragments reflects less effort in processing compared to the other object categories.

It is also possible that the presence of a whole object is necessary to activate the action-knowledge network. This could explain the lack of left lateralized processes for the scrambled images. Indeed, it has been shown that the posterior LOC regions are more sensitive to scrambled images containing local features, while the more anterior and lateral areas in the LOC contain global representations (Grill-Spector et al., 2001; Lerner et al., 2001). If whole objects and object parts are represented in different LOC areas, then they might also elicit different processing mechanisms.

However, both the lack of action-knowledge activation for the scrambled tools and the lack of real category discrimination for scrambled objects suggest that the N1 category differences do not reflect real conceptual category discrimination when the global shape is not available. However, participants were quite accurate in recognizing objects by their fragments (accuracy: 89.52%). If the N1 component does not reflect discrimination processes, category representations should be activated later in time. Indeed, whereas several studies have reported robust N1 category differences for intact faces and cars (for review see Rossion & Jacques, 2008), Harel, Ullman, Epshtein and Bentin (2007) have found category differences for face and car fragments only in a later ERP component (N2). Harel et al. (2007) suggested that fragments of faces might elicit different processing mechanisms than whole faces. This could explain the delayed category difference in their experiment. However, unlike Harel et al. (2007) we did find some kind of category differences for the N1 component. This difference can be explained by the stimuli used in the two studies. Harel et al. (2007) used isolated, greyscale fragments of cars and faces, whereas we used color images, and although the object parts were scrambled we presented all parts of the objects. Indeed, a recent study reported the advantage of colors for categorizing scrambled images of animals and vehicles (Rokszin et al., 2015). Regarding the underlying mechanism of category discrimination, overall our results indicate that the processing of different object categories relies on different mechanisms. The categorization of tools primarily relies on shape-based processing. The presence of global shape enables the activation of conceptual tool representations in the early stages of processing. Unlike for the tools, the activation of conceptual representation of vehicles cannot be identified with a specific mechanism. However, the small N1 amplitude and high accuracy rates for the real vehicles suggest that fragment-based representations play an important role for the vehicle category. Therefore, these results suggest different processing mechanisms that are selective to object categories.

One possible explanation for these results relies on the specific structure of tools. They usually have a handle and a functional part. It has been shown that the handles of tools attract early attentional processes (Matheson et al., 2014). In their study, animals and tools were presented on the screen. These were followed by a dot that appeared on either side of the objects for 100 ms. The task of the participants was to indicate whether the dot appeared on the left or the right side of the screen by pressing a button. Responses were more accurate when the dot was shown next to the object handle than when it was shown next to the functional part. Furthermore, attention related ERP modulation was also found. The P1 component was bigger when the dot was presented next to the handle than when it appeared on the side of the functional part. No such differences were found for the animals (Matheson et al., 2014). Therefore, it is possible that the handles of tools have an important role in the activation of action-knowledge. However, a handle can be defined in relation to the whole object. When the global shape is disrupted the intrinsic structure of the tool disappears; therefore, an early attentional bias towards the handle cannot be present. Furthermore, when only fragments of tools are available the functional tool part might be a better candidate to help in accurate recognition than fragments of handles. This could explain why only the intact tools elicited left-lateralized processing mechanisms. Vehicles, on the other hand, are often partly covered by other objects in everyday scenes, but people can still easily identify them. The small N1 amplitude and high accuracy rates for the real vehicle fragments suggest that fragment-based representations play an important role in recognizing real vehicles in everyday life.

We found task related top-down modulation for the P1 component both for the intact and scrambled recognizable images. When the global shape was present, the tool-vehicle amplitude difference was left lateralized when participants had to discriminate between tools and vehicles, but not when they discriminated between toys and real objects. When the shape was distorted, the vehicle-tool difference was larger in the tool vs. vehicle task than in the toy vs. real task, however, this effect was only marginally significant and it was not lateralized.

Similar task related modulation of the visual ERPs has been recently reported. In one study (Balas and Koldewyn, 2013) participants were shown humans and dogs either as real faces or faces of toys. The authors found N1 amplitude differences only when participants had to perform a human vs. dog categorization task. When participants had to decide about the animacy of the stimuli (i.e., toy vs. real) the human-dog N1 difference disappeared. If the ERP category differences flexibly follow the categories inflicted by the categorization task, than ERP category differences might reflect decision-making processes (Johnson & Olshausen, 2003). However, we found task related modulation in the P1 component instead of the N1. Furthermore, in our study the task effect showed a different pattern when the global shape was present. The task did not enhance the P1 category differences; instead it shifted the category difference to the left hemisphere. Therefore, decision-making processes as an underlying mechanism for the observed task effect cannot explain our results.

It is possible that in our study the task modulation for the P1 component reflects an early interaction between top-down expectations and bottom-up information processes. Indeed, flexible top-down and bottom-up interactive processes have been suggested in vision (e.g. Gilbert & Sigman, 2007). It has been shown that the expectation of simple geometric shapes tune neurons in the early visual cortex (McManus, Li, & Gilbert, 2011). Furthermore, at the level of brain areas, it has been shown that top-down factors can activate category-specific areas in the ventral visual stream. In fMRI experiments, participants learned that specific cues predicted the category of the subsequent stimulus. These anticipatory cues activated category-specific visual areas. Such anticipatory activities have been reported following visual cues such as images (Esterman & Yantis, 2010), symbols (Peelen & Kastner, 2011) or the written name of the categories (Puri et al., 2009). Tool-specific areas can be activated not only by looking at images of tools, but also by answering questions about tools (Chao et al., 1999). Our participants knew in advance that their task was to categorize tools and vehicles. Therefore, the expectation of a tool category could prepare the visual system by activating the network representing tools. This network also involves the left lateralized action-knowledge network (Martin, 2007), which could explain the left lateralized tool-vehicle difference for the P1 component in the tool vs. vehicle task.

The lack of lateralization for the scrambled images, similarly to the N1 effect, could suggest that the global object has to be present for the activation of the action-knowledge network. Still, the presence of a task effect indicates an interaction between top-down and bottom-up processes at the level of object part representations. Therefore, this result suggests that the posterior LOC areas containing the representations of object parts (Grill-Spector et al., 2001) can be accessed both by top-down and bottom-up processes.

Regarding the possible low-level differences between vehicle and tool images, our results suggest that such differences are not the underlying factors of the P1 and N1 differences for the intact and scrambled-recognizable images. It has been suggested that early ERP differences are driven by low-level physical differences between stimuli that are independent from the perceptual or conceptual object categories (e.g. Thierry, Martin, Downing, & Pegna, 2007). In our study the stimuli were not matched for low-level differences such as the luminance or color. Indeed, we found significant vehicle-tool differences even for the unrecognizable-scrambled stimuli, which indicates the role of low-level physical factors in these early ERP components. However, our results suggest that low-level differences alone cannot explain the P1 and N1 category differences found for the intact and scrambled-recognizable stimuli. The ERP category differences were modulated by additional factors in the presence of a recognizable object, or object parts, in the stimuli. In the presence of global shape, we found task-related modulation and a left hemispheric advantage in category discrimination. In the presence of object parts the N1 component reflected a specific mechanism for the real vehicle parts. In addition, the cortical map of the vehicle-tool difference for the unrecognizable images showed more medial distribution compared to the recognizable stimuli. This is in line with the suggestion that low-level features activate the earlier visual areas (Contini et al., 2017). Therefore, the presence of a recognizable object or object parts in the stimuli elicited additional processing mechanisms other than simply low-level discrimination processes. Furthermore, despite the significant category differences in the ERP, the accuracy of category judgments was only slightly above chance level for the unrecognizable images. Therefore, category differences in the P1 and N1 components do not necessarily lead to conscious recognition.

Regarding the toy and real categories, toy-real ERP differences were present for the intact and scrambled-recognizable images. Furthermore, we found right-hemispheric processes for the toy-real distinction in the N1 component. This hemispheric effect was present both for the intact and, at least partly, for the scrambled recognizable images.

Therefore, these results do not clearly suggest different processing mechanisms for these two image types. The members of the toy and real categories share similar shapes and specific object parts (for instance a toy and real car have a similar global shape and they both have wheels). Therefore, the toy-real distinction cannot be based on perceptual similarity. Presumably, a toy vs. real discrimination is based on the processing of detailed information such as material, texture or the proportion of the parts. Indeed, right hemispheric processing for surface properties have been reported (Cant & Goodale, 2007). This observation is in line with our result of a right-hemispheric advantage for toy-real discrimination both for intact and scrambled images, as surface properties can be processes even in the absence of shape.

In regard of the visual ERP components, our results show that similar amplitude differences for the visual P1 and N1 ERP components can reflect different underlying processes. The results suggest that N1 category differences can reflect simply low-level differences (unrecognizable stimuli), but also conceptual category representations (intact tools). Flexible visual processing mechanisms and consequently flexible modulation of the visual ERP components can explain why so many explanations have been given to the category differences reflected in the ERPs. Furthermore, the results also show that category representations can be reflected in the N1 or in subsequent components. Similarly, it has been suggested that decision-making processes can be reflected either in the N1 or subsequent ERP components depending on the task (Johnson & Olshausen, 2005, 2003).

There is an emerging view in the recent EEG literature about the flexibility of the ERP components. For instance Van Der Lubbe, Szumska, and Fajkowska (2016) have suggested that the interpretation of visual P1 and N1 components should be reconsidered in the light of their findings. They showed that the P1 amplitude was related to stimulus-locked activity in the alpha frequency band, while the N1 amplitude was related to stimulus-locked activity in the theta band when emotional faces were used as stimuli in a go/no-go task. It is a question whether P1 and N1 amplitudes are related to similar frequency bands for different stimuli and in different categorization task. Furthermore, in a recent review, Contini et al. (2017) have suggested that novel EEG analyzing methods indicate dynamic time-courses for object processing, which includes both bottom-up and top-down processes. The authors have also pointed out that results related to the time-course of categorization are always specific to the stimulus set used in those experiments. This is in line with our assumption that category differences in the visual ERPs might reflect different processes depending on the stimuli and task used in the experiment.

Overall these results indicate flexible processing mechanisms for different object categories and stimulus types. Consequently, the visual ERPs do not reflect fixed processing mechanisms and the category differences appearing in the N1 component can reflect very different underlying processes. Rather, the visual ERPs might reflect different processes depending on the stimuli and task used in the experiment. Finally, our results show that even the earliest ERP components, around 100 ms after presentation of a stimulus, reflect interactions between perceptually driven bottom-up and conceptually driven top-down processes.

References

Aguirre, G. K., Zarahn, E., & D'Esposito, M. (1998). An area within human ventral cortex sensitive to "building" stimuli: evidence and implications. *Neuron*, *21*, 373–383. https://doi.org/10.1016/S0896-6273(00)80546-2

Antal, A., Kéri, S., Kovács, G., Janka, Z., & Benedek, G. (2000). Early and late components of visual categorization: an event-related potential study. *Cognitive Brain Research*, *9*(1), 117–119. https://doi.org/10.1016/S0926-6410(99)00053-1

Antal, A., Kéri, S., Kovács, G., Liszli, P., Janka, Z., & Benedek, G. (2001). Event-related potentials from a visual categorization task. *Brain Research Protocols*, *7*, 131–136. https://doi.org/10.1016/S1385-299X(01)00055-1

Balas, B., & Koldewyn, K. (2013). Early visual ERP sensitivity to the species and animacy of faces. *Neuropsychologia*, *51*(13), 2876–2881. https://doi.org/10.1016/j.neuropsychologia.2013.09.014

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. https://doi.org/10.1162/jocn.1996.8.6.551

Cant, J. S., & Goodale, M. A. (2007). Attention to form or surface properties modulates different regions of human occipitotemporal cortex. *Cerebral Cortex*, *17*(3), 713–731. https://doi.org/10.1093/cercor/bhk022

Chao, L. L., Haxby, J. V, & Martin, A. (1999). Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nature Neuroscience*, *2*(10), 913–919. https://doi.org/10.1038/13217

Chao, L. L., & Martin, a. (2000). Representation of manipulable man-made objects in the dorsal stream. *NeuroImage*, *12*(4), 478–484. https://doi.org/10.1006/nimg.2000.0635

Cloutman, L. L. (2013). Interaction between dorsal and ventral processing streams: Where, when and how? *Brain and Language*, *127*(2), 251–263. https://doi.org/10.1016/j.bandl.2012.08.003

Contini, E. W., Wardle, S. G., & Carlson, T. A. (2017). Decoding the time-course of object recognition in the human brain: From visual features to categorical decisions. *Neuropsychologia*, *105*, 165–176. https://doi.org/10.1016/j.neuropsychologia.2017.02.013

Curran, T., Tanaka, J. W., & Weiskopf, D. M. (2002). An electrophysiological comparison of visual categorization and recognition memory. *Cognitive, Affective & Behavioral Neuroscience*, *2*(1), 1–18.

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. https://doi.org/10.1016/j.jneumeth.2003.10.009

Delorme, A., Rousselet, G. A., Macé, M. J.-M., & Fabre-Thorpe, M. (2004). Interaction of top-down and bottom-up processing in the fast visual analysis of natural scenes. *Cognitive Brain Research*, *19*(2), 103–113. https://doi.org/10.1016/j.cogbrainres.2003.11.010

Dilkina, K., & Lambon Ralph, M. A. (2012). Conceptual structure within and between modalities. *Frontiers in Human Neuroscience*, *6*, 333. https://doi.org/10.3389/fnhum.2012.00333

Esterman, M., & Yantis, S. (2010). Perceptual expectation evokes category-selective cortical activity. *Cerebral Cortex*, *20*, 1245–1253. https://doi.org/10.1093/cercor/bhp188

Gilbert, C. D., & Li, W. (2013). Top-down influences on visual processing. *Nature*

Reviews Neuroscience, *14*(5), 350–363. https://doi.org/10.1038/nrn3476

Gilbert, C. D., & Sigman, M. (2007). Brain states: top-down influences in sensory processing. *Neuron*, *54*, 677–696. https://doi.org/10.1016/j.neuron.2007.05.019

Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision Research*, *41*(10–11), 1409–1422. https://doi.org/10.1016/S0042-6989(01)00073-6

Harel, A., Ullman, S., Epshtein, B., & Bentin, S. (2007). Mutual information of image fragments predicts categorization in humans: Electrophysiological and behavioral evidence. *Vision Research*, *47*, 2010–2020. https://doi.org/10.1016/j.visres.2007.04.004

Ishai, A., Ungerleider, L. G., Martin, A., Schouten, J. L., & Haxby, J. V. (1999). Distributed representation of objects in the human ventral visual pathway. *Proceedings of the National Academy of Sciences*, *96*(16), 9379–9384. https://doi.org/10.1073/pnas.96.16.9379

Jacques, C., & Rossion, B. (2007). Early electrophysiological responses to multiple face orientations correlate with individual discrimination performance in humans. *NeuroImage*, *36*(3), 863–876. https://doi.org/10.1016/j.neuroimage.2007.04.016

Johnson, J. S., & Olshausen, B. A. (2005). The earliest EEG signatures of object recognition in a cued-target task are postsensory. *Journal of Vision*, *5*(4), 2. https://doi.org/10.1167/5.4.2

Johnson, J. S., & Olshausen, B. a. (2003). Timecourse of neural signatures of object recognition. *Journal of Vision*, *3*(7), 499–512. https://doi.org/10.1167/3.7.4

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *17*(11), 4302–

4311. https://doi.org/10.1098/Rstb.2006.1934

Kiefer, M. (2001). Perceptual and semantic sources of category-specific effects: event-related potentials during picture and word categorization. *Memory & Cognition*, *29*(1), 100–116. https://doi.org/10.3758/BF03195745

Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., … Bandettini, P. A. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron*, *60*(6), 1126–1141. https://doi.org/10.1016/j.neuron.2008.10.043

Lerner, Y., Hendler, T., Ben-Bashat, D., Harel, M., & Malach, R. (2001). A hierarchical axis of object processing stages in the human visual cortex. *Cerebral Cortex*, *11*(4), 287–297. https://doi.org/10.1093/cercor/11.4.287

Mahon, B. Z., Milleville, S. C., Negri, G. A. L., Rumiati, R. I., Caramazza, A., & Martin, A. (2007). Action-related properties shape object representations in the ventral stream. *Neuron*, *55*(3), 507–520. https://doi.org/10.1016/j.neuron.2007.07.011

Martin, A. (2007). The representation of object concepts in the brain. *Annual Review of Psychology*, *58*, 25–45. https://doi.org/10.1146/annurev.psych.57.102904.190143

Matheson, H., Newman, A. J., Satel, J., & McMullen, P. (2014). Handles of manipulable objects attract covert visual attention: ERP evidence. *Brain and Cognition*, *86*(1), 17–23. https://doi.org/10.1016/j.bandc.2014.01.013

McManus, J. N. J., Li, W., & Gilbert, C. D. (2011). Adaptive shape processing in primary visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 9739–9746. https://doi.org/10.1073/pnas.1105855108

Mur, M., Meys, M., Bodurka, J., Goebel, R., Bandettini, P. A., & Kriegeskorte, N. (2013). Human object-similarity judgments reflect and transcend the primate-IT object representation. *Frontiers in Psychology*, *4*, 1–22.

https://doi.org/10.3389/fpsyg.2013.00128

Nieuwenhuis, S., Forstmann, B. U., & Wagenmakers, E.-J. (2011). Erroneous analyses of interactions in neuroscience: a problem of significance. *Nature Neuroscience*, *14*(9), 1105–1107. https://doi.org/10.1038/nn.2886

Paz-Caballero, D., Cuetos, F., & Dobarro, A. (2006). Electrophysiological evidence for a natural/artifactual dissociation. *Brain Research*, *1067*, 189–200. https://doi.org/10.1016/j.brainres.2005.10.046

Peelen, M. V, & Kastner, S. (2011). A neural basis for real-world visual search in human occipitotemporal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 12125–12130. https://doi.org/10.1073/pnas.1101042108

Proverbio, A. M., Adorni, R., & Aniello, G. E. (2011). 250 ms to code for action affordance during observation of manipulable objects. *Neuropsychologia*, *49*(9), 2711–2717. https://doi.org/10.1016/j.neuropsychologia.2011.05.019

Proverbio, A. M., Azzari, R., & Adorni, R. (2013). Is there a left hemispheric asymmetry for tool affordance processing? *Neuropsychologia*, *51*(13), 2690–2701. https://doi.org/10.1016/j.neuropsychologia.2013.09.023

Proverbio, A. M., Del Zotto, M., & Zani, A. (2007). The emergence of semantic categorization in early visual processing: ERP indices of animal vs. artifact recognition. *BMC Neuroscience*, *8*, 24. https://doi.org/10.1186/1471-2202-8-24

Puri, A. M., Wojciulik, E., & Ranganath, C. (2009). Category expectation modulates baseline and stimulus-evoked activity in human inferotemporal cortex. *Brain Research*, *1301*, 89–99. https://doi.org/10.1016/j.brainres.2009.08.085

Rokszin, A. A., Győri-Dani, D., Linnert, S., Krajcsi, A., Tompa, T., & Csifcsák, G. (2015). The interplay of holistic shape, local feature and color information in object categorization. *Biological Psychology*, 1–12. https://doi.org/10.1016/j.biopsycho.2015.05.002

Rossion, B., Gauthier, I., Tarr, M. J., Despland, P., Bruyer, R., Linotte, S., & Crommelinck, M. (2000). The N170 occipito-temporal component is delayed and enhanced to inverted faces but not to inverted objects: an electrophysiological account of face-specific processes in the human brain. *Neuroreport*, *11*(1), 69–74. https://doi.org/10.1097/00001756-200001170-00014

Rossion, B., & Jacques, C. (2008). Does physical interstimulus variance account for early electrophysiological face sensitive responses in the human brain? Ten lessons on the N170. *NeuroImage*, *39*, 1959–1979. https://doi.org/10.1016/j.neuroimage.2007.10.011

Sitnikova, T., West, W. C., Kuperberg, G. R., & Holcomb, P. J. (2006). The neural organization of semantic memory: Electrophysiological activity suggests feature-based segregation. *Biological Psychology*, *71*, 326–340. https://doi.org/10.1016/j.biopsycho.2005.07.003

Tanaka, J., Luu, P., Weisbrod, M., & Kiefer, M. (1999). Tracking the time course of object categorization using event-related potentials. *NeuroReport*, *10*(4), 829–835. https://doi.org/10.1097/00001756-199903170-00030

Thierry, G., Martin, C. D., Downing, P., & Pegna, A. J. (2007). Controlling for interstimulus perceptual variance abolishes N170 face selectivity. *Nature Neuroscience*, *10*(4), 505–511. https://doi.org/10.1038/nn1864

Tokudome, W., & Wang, G. (2012). Similarity dependency of the change in ERP component N1 accompanying with the object recognition learning. *International Journal of Psychophysiology*, *83*(1), 102–109. https://doi.org/10.1016/j.ijpsycho.2011.10.012

Ullman, S. (2007). Object recognition and segmentation by a fragment-based hierarchy. *Trends in Cognitive Sciences*, *11*(2), 58–64. https://doi.org/10.1016/j.tics.2006.11.009

Ullman, S., Vidal-Naquet, M., & Sali, E. (2002). Visual features of intermediate complexity and their use in classification. *Nature Neuroscience*, *5*(7), 682–687. https://doi.org/10.1038/nn870

Van Der Lubbe, R. H. J., Szumska, I., & Fajkowska, M. (2016). Two sides of the same coin: ERP and wavelet analyses of visual potentials evoked and induced by task-relevant faces. *Advances in Cognitive Psychology*, *12*(4), 153–167. https://doi.org/10.5709/acp-0195-3

VanRullen, R. (2011). Four common conceptual fallacies in mapping the time course of recognition. *Frontiers in Psychology*, *2*, 1–6. https://doi.org/10.3389/fpsyg.2011.00365

VanRullen, R., & Thorpe, S. J. (2001). The time course of visual processing: from early perception to decision-making. *Journal of Cognitive Neuroscience*, *13*, 454–461. https://doi.org/10.1162/08989290152001880

Vogel, J., & Wallraven, C. (2006). Categorization of natural scenes : local vs . global information. *International Journal of Computer Vision*, *4*(3), 33–40. https://doi.org/10.1145/1278387.1278393

Discussion: Flexible mechanisms in information accumulation

Szilvia Linnert

Lancaster University

Summary

The aim of the thesis was to investigate the underlying factors of the category differences in the visual ERPs. Additionally, we investigated the interactions between bottom-up and top-down visual processes, reflected as task-related modulation of the ERP category differences. Specifically, we were interested whether category differences are based on lower-level physical differences or higher-level category representations, and as well as, whether top-down factors access lower- or higher-level processes. In Papers 1 and 2, we investigated the effect of amplitude spectrum (AS) information on the animal-vehicle differences in the visual ERPs in different categorization tasks. In Paper 3, we examined whether animal-vehicle ERP differences and their task-related modulation are specific to high or low spatial frequency information. In Paper 4, we investigated whether the task also modulates the ERP differences for inanimate object categories (tools and vehicles), and whether the presence of global shape plays a role in tool-vehicle discrimination.

In Paper 1, we investigated whether animal-vehicle ERP differences are based on the AS, as a low-level, category-specific characteristic and whether the task, as a top-down factor, accesses bottom-up processes at the level of the AS information. For the P1 component, animal-vehicle differences occurred only for images with intact AS and only in the animal-vehicle task. In contrast, category differences in the later P2 and N2 components were irrespective of the AS content of the images, and they were larger when participants had to differentiate animals and vehicles compared to when the task did not require the processing of these categories. Therefore, the early P1 reflected interaction between the AS and the task demand, whereas later ERP components indicated higher-level, category-specificity and its modulation by the task.

We also examined the role of AS on the animal-vehicle ERP differences without a forced choice categorization task. In Paper 2, participants had to match category labels and the subsequent images, therefore, instead of two-choice categorization, this task required a decision whether the presented category label and the following image matched or not. The label-image matching task is a categorization task that directs participants' attention to the object on the image, similar to a two-choice task. However, the decision that participants have to make is not a forced-choice, animal vs. vehicle decision. Using a similar, still different categorization task can show whether ERP differences found in Paper 1 are due to a two-choice task or they are due to the processing of these images, irrespectively of the task. The category differences in the N1, P2 and N2 components were similar to those in Paper 1; however, they were only present when the AS information was available in the images. This similarity suggests that he category differences in the ERPs are mainly due to the processing of the images. This similarity is discussed in detail later on (see 'Animal-vehicle differences in the visual ERPs' section).

In Paper 3, we investigated whether category differences are based on low spatial frequency (SF) information, which reflects coarse, global shape information about objects. In that study, we found P2 animal-vehicle differences irrespective of the spatial frequency content of the images, but these differences were present only in the animal-vehicle task, but not when participants had to identify a letter displayed on the image. This suggests that category differences are not specifically based on high or low SFs. However, when images contained intact SFs, the task modulated the category differences earlier, already in the N1 component. Therefore, the more information is present in the images the earlier the category differences can emerge.

In Paper 4, we investigated how the presence of global shape and recognizable object fragments affect ERP differences for tool and vehicle categories, and how the task affects these differences. The presence of global shapes elicited different ERP category differences than object fragments. The hemispheric differences indicated tool-specific processes, but only for whole objects. Additionally, the task also modulated these tool-specific processes in the presence of the global shape.

In the following sections, these results will be discussed from the point of view of (1) category representations, (2) top-down processes, (3) information accumulation and integration in the visual areas for decision-making, and (4) the function of ERP components in terms of underlying processes.

Category representations in the visual ERPs

The last stage of the ventral visual pathway is the inferior temporal (IT) cortex in monkeys, and its human counterpart, the lateral occipital complex (LOC). Neurons in the IT are shape-selective and their firing patterns seem to represent object categories (Kiani et al., 2007; Tompa & Sáry, 2010). In an single cell recording study, Kiani et al. (2007) found category-selective activity patterns in the monkey IT for animate and inanimate categories. Within the inanimate category, they found category-selectivity for human, monkey and non-primate faces, as well as for bodies for higher-level (humans, birds, for legged animals) and lower-level animals (e.g. fish, insects). Category-specific areas have also been reported in the human LOC using brain imaging methods. Specific areas have been found for faces (Kanwisher et al., 1997) and body parts (Downing, 2001). Furthermore, Chao, Haxby, and Martin (1999) reported animal-specific activity pattern across different LOC areas. Moreover, in humans category-selectivity has also been found for inanimate object categories, such as for buildings (Aguirre, Zarahn, & D'Esposito, 1998), chairs (Ishai et al., 1999), tools (Chao et al., 1999) and visual scenes

(Epstein & Kanwisher, 1998). However, recent fMRI results have suggested that only animate and inanimate categories, as well as faces and bodies are categorically represented in the human LOC (Kriegeskorte et al., 2008). Nevertheless, higher-level visual areas also contains continuous, similarity-based object representations (Kriegeskorte et al., 2008; Mur et al., 2013). According to these results, similar objects are represented closer to each other in a continuous space. On this view, category-specific areas for inanimate objects do not reflect real category representations; instead, category-specificity arises as an artifact from experiments that use perceptually very different categories (Kriegeskorte et al., 2008). Additionally, these category-specific responses are stable, as their boundaries do not change with task requirements (Bugatus et al., 2017; Linda L Chao et al., 1999).

Beside brain imaging methods, the EEG signal also seems to contain categorical code. By using multivariate pattern analysis, both object- and category-selective patterns can be decoded from EEG or MEG signals (for review see Contini, Wardle, & Carlson, 2017). In addition, given the good temporal resolution of EEG, this analysis can also reveal the time-course of category-specific activity. Results have indicated that object-specific information can be decoded earlier, while category specificity emerges later in time. Animate and inanimate category distinctions could be decoded as early as 160 ms after stimulus onset in one study, or later, at around 240 ms after onset, depending on the stimulus set used in the study (Contini et al., 2017).

Animal-vehicle differences in the visual ERPs

Category differences in the visual ERPs have been found for categories that have also shown differential activity in brain imaging studies. Therefore, one aim of this thesis was to discover whether animal/non-animal ERP differences reflect category representations. We investigated this question in three studies (Paper 1, 2, 3), using the same stimuli and similar categorization tasks. In Paper 1, animal and vehicle images were presented either with intact or averaged AS. In Paper 3, we used the same images with different SF content. In both studies, stimuli had to be categorized either as animals and vehicles or according to the letter appearing on the images. In Paper 2, similarly to Paper 1, animal and vehicle images contained either the original or the averaged AS information. However, in this study, instead of a categorization task, participants had to match category labels with the following images. If the category differences in the ERPs reflect higher-level category representations and consequently the percept, then animal-vehicle differences should not be modulated by low-level image properties and results should show similar differences across these studies. Moreover, even if the modulation of different low-level image properties can alter ERP differences, the direction of the animal-vehicle differences (e.g. more negative amplitudes for vehicles compared to animals) should not change between studies. If similar animal and vehicle images elicit the same category differences despite the different modulation of low-level properties (AS, SFs) and task requirements across studies, it would suggest that he category differences in the ERPs are mainly due to the processing of these categories.

It has been shown that ERP category differences reflect some aspect of the percept (Bacon-Macé et al., 2005; Macé et al., 2005; Rokszin et al., 2015). Nevertheless, reflecting the percept requires that representations should not be sensitive to low-level characteristics. This is because the percept reflects the end-result of visual processing, which is an integrated image of different characteristics even if these were processed separately. Indeed, this is what we found in Papers 1 and 3 for the P2 and/or N2 components. In Paper 1, category differences in the early P1 and N1 components were modulated by the AS content of the images. In contrast, animal-vehicle differences in the

P2 and N2 components were irrespective of the AS information. In addition, these late animal-vehicle differences were distributed over more lateral occipital areas, which contain category representations, in contrast to the medial occipital distribution of the differences related to AS information (see Figures 3, 4, 5 & 6 in Paper 1). Similarly, in Paper 3, animal-vehicle differences in P2 were irrespective of the SFs; despite the different low-level characteristics of the images, these characteristics did not modulate the animal-vehicle differences. By the time of the P2 component, amplitudes for vehicles were more negative than for animals for all stimulus types despite the differential processing of AS and SF information. Therefore, animal-vehicle differences by the time of the P2 component are irrespective of AS and SF content of the images, suggesting that they reflect higher-level category representations.

It has also been shown that representations in the higher-level areas are stable (Bugatus et al., 2017), therefore, if ERP differences reflect these representations, these differences should be stable across studies. We used the same images in Paper 1 and 2, therefore we can compare the results. In both studies, P1 and N1 animal-vehicle differences were distributed over medial occipital areas, and P2 and N2 differences were more laterally distributed. Additionally, the direction of the difference was similar in both studies, at least for images with intact AS. That is, the N1 amplitudes were lager for animals than for vehicles over the medial occipital areas. For the P2 and N2 components, the amplitudes were more negative for vehicles compared to animals over the lateral-occipital areas. For images with averaged AS, we found differences only in Paper 1, but not in Paper 2. This difference is therefore likely due to the task (animal-vehicle categorization in Paper 1 vs. word-image matching in Paper 2). This assumption is also supported by the results of Paper 3. In that paper, stimuli with intact SFs were matched for AS. Therefore, these stimuli were similar to animal and vehicle images in Paper 1 with intact AS. In addition, participants performed a two-choice categorization task in both studies. Animal-vehicle differences showed a slightly different scalp distribution in the two studies. In Paper 1, P1 and N1 differences were distributed over the medial occipital areas, while P2 and N2 differences were distributed more laterally (Figures 3, 4, 5, 6 in Paper 1). In contrast, category differences in Paper 3 were distributed similarly for all ERP components, including both temporal and occipital electrodes (Figure 7 in Paper 3). Nevertheless, animal-vehicle ERP differences were similar in both studies. That is, the P1 did not differ between the two categories, while amplitudes for vehicles were more negative than for animals for the N1, P2 and N2 components.

Overall, our results show that animal-vehicle differences in the P2 and N2 are irrespective of low-level stimulus characteristics, and these differences were stable across studies. Furthermore, these results are in line with previous studies reporting more negative ERPs for inanimate objects compared to animals in a two-choice categorization task (Antal et al., 2001; Antal, Kéri, Kovács, Janka, & Benedek, 2000; Linnert, Csifcsák, & Tompa, in prep). However, in studies where a go/no-go task was performed, amplitudes were usually more negative for animals. Nevertheless, the results of VanRullen and Thorpe (2001) suggest that processes reflecting visual categories and decision-making processes overlap in the ERPs in a go/no-go paradigm. When these two processes are intermixed, it can alter the ERP differences. In a two-choice categorization task, on the other hand, both visual processing and decision-making rely on category specific information; therefore, decision-making does not modulate category specificity. Additionally, studies differ in the types of images used as non-animal category. In many papers, images of natural and artificial scenes mixed with man-made objects and are used as inanimate category (e.g. Antal et al., 2001, 2000; Thorpe, Fize, & Marlot, 1996), while other studies used only man-made objects, such as vehicles (R VanRullen & Thorpe, 2001). Additionally, a few studies include plants and fruits as stimuli (Antal et al., 2000; Kiefer, 2001), but while Antal et al. (2000) included these images to the non-animal category, Kiefer (2001) included them in the natural category together with animals. According to Kriegeskorte et al., (2008), plants and fruits belong to the inanimate category, together with man-made objects. Furthermore, a specific areas have been discovered that process scenes and buildings (Aguirre et al., 1998; Epstein & Kanwisher, 1998); thus, scenes and buildings are likely processed differently, than man-made objects. Therefore, if category differences in the visual ERPs reflect visual category representations, the choice of inanimate category could also modulate ERP differences. The categorization task and the choice of stimulus categories can explain the differences between studies. Nevertheless, we found similar animal-vehicle differences in three separate studies, using different categorization tasks and different low-level image modulation, which suggests the role of category representations as underlying factors of the ERP differences.

Differences between inanimate object categories

In line with the suggestion that inanimate object representations are organized by perceptual similarity (Kriegeskorte et al., 2008), a similarity-based mechanism can also explain the category differences reported for the visual ERPs such as between inanimate object categories. In paper 4, we investigated whether perceptual similarity is based on the global shape of the objects, or on category-specific object features for inanimate, vehicle and tool object categories. Moreover, the special nature of the tool category enabled us to investigate whether visual ERPs are modulated by the processing of other, object-related attributes. Vehicle and tool categories are not represented per se in the LOC (Kriegeskorte et al., 2008). Nevertheless, tool-specific activity patterns have been found in the higher-level visual areas (Linda L Chao et al., 1999). Such tool specific patterns could be based on visual similarity, such as the shape information (Kriegeskorte et al., 2008). Alternatively, tool specific activity in the visual areas could be part of a network that represents different attributes of an objects and related knowledge in distributed areas (Haxby et al., 2001). The results of Paper 4 suggest that the activation of such a tool-specific network also modulates the ERP differences between tools and vehicles. On the other hand, results also indicate the role of shape information in the visual ERPs. We found left hemispheric tool-vehicle ERP differences reflecting tool-specific processes only for intact images, but not for images with distorted shape information. This is in line with findings suggesting shape-selectivity in the higher-level visual areas (for review see Tompa & Sáry, 2010). Also, shape can be an important organizing factor for the continuous object representational space in these visual areas (Dilkina & Lambon Ralph, 2012). Furthermore, the importance of the global shape is in line with the assumption that high-level visual representations do reflect the percept, and consequently, higher-level category specificity.

Task-related modulation

In a previous study we have found task-related modulation for car-bird N1 differences (Linnert et al., in prep). One aim of this thesis was to investigate the underlying mechanisms of such modulation. We examined the animal-vehicle ERP differences in different tasks and by modulating low-level image characteristics, such as the AS and SF contents. In Paper 1, we found interactions between top-down and bottom-up processes at the level of both low-level category-specific processes in the P1, and at the level of category representations in the P2 and N2. Results of Paper 3 suggested that top-down processes do not separately access high or low SF information, and we found interactions at the level of category representation also in this study. Furthermore, task-related modulation in Paper 4 also suggests interactions at the level of higher-level category representations for inanimate tool and vehicle categories.

Top-down activation of shape- or category-selective areas has been previously reported in mental imagery (O'Craven & Kanwisher, 2000; Reddy et al., 2010) and in expectation of a subsequent stimulus (Esterman & Yantis, 2010; Peelen & Kastner, 2011; Puri et al., 2009; Stokes et al., 2009). Bugatus et al. (2017) have reported results indicating that stable higher-level category representations in the visual areas are affected by selective attention, but category boundaries do not change in different categorization tasks. The authors suggested that this is because selective attention only restricts the processing to the attended object or category, but it does not modulate category boundaries. In other words, selective attention might focus visual processing in order to "see" or perceive the attended object as quickly as possible. Preparatory activity is also explicable in this framework, as the expectation of an object selectively activates a visual area that is responsible for the processing of that object. Furthermore, such selective activity modulates the processing of the subsequent stimuli and increases performance.

Selective attentional processes can also explain our findings about task-related modulation of ERP category differences. If the ERP category differences flexibly follow the categories inflicted by the categorization task, then ERP category differences might reflect decision-making processes. Despite the fact that these modulations are task-related, results suggest that they do not reflect task-related decision-making per se. First, in our previous study (Linnert et al., in prep) we did not find color-greyscale difference for intact cars and birds, even if the categorization task required color distinction. In contrast, we did find car-bird (Linnert et al., in prep) and animal-vehicle differences

(Paper 1, 3), even if the task did not required the differentiation of those categories. Second, in our previous study (Linnert et al., in prep) and in Paper 4, the categorization task did not modulate the magnitude of the category difference, rather, it modulated the hemispheric distribution of those differences. Third, decision-making could not rely solely on lower-level physical stimulus characteristic as we still found task related modulation at the level of AS. Therefore, task-related modulation of the ERP category differences might rather reflect participants' expectation or selective attention to the content of the images.

Attentional processes, however, might differ in different types of categorization tasks. We did not find direct evidence indicating that correctly or incorrectly expecting an animal or a vehicle would modulate the ERPs (Paper 2). Nevertheless, in Paper 2, participants performed a word-image matching task, whereas in the other studies, the task-related modulation was found in categorization tasks. Both tasks require the recognition of the object; however, they might distribute attention differently. In a word-image matching task the main focus can be to identify the object. In contrast, for categorization, the main focus can be to differentiate the two categories from each other. It is possible that this differentiation shows up in the ERP as larger category difference. The results of Koida and Komatsu (2007) are in line with this explanation. They investigated activity differences at the level of color-selective neurons in the monkey IT in a color categorization and a color discrimination task. They found increased activity in the categorization task, and decreased activity in the discrimination task for the same neurons.

Therefore, selective attention facilitates fast and precise processing of the target object or category. Nevertheless, it is important to point out that higher-level visual representations are also activated automatically, without any attention. For instance, Peelen, Fei-Fei, and Kastner (2009) have found category-specific activity for humans and cars even if the task did not require the processing of those images and if they were presented outside of the focus of attention. Similar results have been reported for ERP animate-inanimate differences (Zhu et al., 2016). In that study, images were presented in a continuous flash suppression paradigm, meaning that noisy images are continuously presented in one eye, which suppresses the processing of the animate or inanimate images presented in the other eye. After stimulus presentation, participants were asked whether they saw an image, and if they did, what was the category of that image. ERPs were compared between animal and non-animal trials in the seen and unseen condition, respectively. ERPs differed between the two categories in both conditions, however, the direction of the difference changed between conditions. In the seen condition animals elicited larger amplitudes compared to vehicles. In contrast, in the unseen condition non-animals elicited larger amplitudes than animals between 150 - 300 ms. This indicates that differential processing of animals and non-animals takes place even in the absence of conscious recognition (Zhu et al., 2016).

We also found category differences even if the task did not require the processing of those categories. The magnitude of the category difference was strong between cars and birds in the color vs. greyscale task (Linnert et al., in prep), as well as between tools and vehicles in the toy vs. real task (Paper 4). We found animal vs. vehicle differences in paper 1 and 3 also in the X vs. O categorization task; however, these differences were not so strong compared to the colored images in Linnert et al. (in prep) and in Paper 4. Nevertheless, the objects on these images are easily recognizable and salient, and participants can easily perceive these categories even if they did not have to categorize them. Therefore, these results show category-selectivity in the absence of task requirement but not in the absence of conscious perception, as in the study by (Zhu et al., 2016).

Overall, our results indicate that category representations are activated automatically, even if those categories are not relevant to the task. Therefore, the task does not elicit these differences; rather, it increases the magnitude of the differences, or modulates hemispheric processing, possibly through a mechanism that focuses the attention to category-specific features both at the level of low-level physical features and higher-level category representations.

Information accumulation

The evidence discussed above suggests that higher-level visual areas contain stable category representations and these representations are affected by selective attention. Moreover, the category differences in the visual ERPs also reflect these attributes of the visual processes. These phenomena can be examined in the framework of information accumulation and integration for decision-making.

Decision-making processes can be understood as evidence accumulation towards that decision. According to the accumulation-to-threshold mechanism, when the accumulated evidence reaches a threshold, the decision can be made. Tremel and Wheeler (2015) have reported evidence showing that face- and house-selective visual areas accumulate information related to the processing of those categories for decision-making. In that study, participants performed a two-choice face vs. house categorization task on videos presenting either a noise-degraded face or house. The authors investigated the temporal dynamics of the activity in the category selective areas, and compared this with the reaction times. When participants' decision was fast, activity in the category selective regions started to rise later, but it showed a steeper increase and quickly reached the peak activity. In contrast, longer reaction times were associated with an earlier rise in activity, but the increase was shallow and the peak activity was lower and it was reached later in time. Images of houses elicited accumulative activity also in the face-region; however, this activity was smaller compared to activity for faces, and it did not correlate with the reaction times. In contrast, no such activity was found for faces in the house-specific area. Thus, these results suggest category-specific evidence accumulation in category-selective areas and they support an integration-to-threshold mechanism in decision-making processes. Interestingly, such an accumulation-like activity change was also found for noise-only trials, when participants still reported seeing a face or house. Therefore, the accumulative activity is percept-dependent; however, it does not necessarily rely only on actual visual features (Tremel & Wheeler, 2015). Nevertheless, the authors did not discuss what kind of evidence could elicit such activity if visual features are not present.

Top-down or feedback processes have been suggested to modulate evidence accumulation (Simen, 2012). Thus, it is possible, that accumulative activity for noisy stimuli (Tremel & Wheeler, 2015) is based on top-down processes, such as expectation of the appearance of a face or house. The results of Puri et al. (2009) are in line with this explanation. They found that top-down processes, such as the expectation of a house or a face, elicit activation in category-specific regions. Following such preparatory activity, the overall activity was increased in those areas for the subsequent image. Moreover, category-specific areas showed greater category selectivity; the difference between the activation for faces and houses was larger following preparatory activity (Puri et al., 2009). Therefore, if category-specific areas accumulate and integrate evidence to a threshold, and this can be measured in the activity pattern in those areas, then top-down activation in expectation to a certain category should modulate information accumulation processes. According to this view, information processing, and also evidence accumulation rely on a flexible interaction between top-down and bottom-up processes depending on the task demands (Gilbert & Sigman, 2007).

Evidence accumulation processes have been suggested as underlying factors for ERP modulations. For instance, Philiastides and Sajda (2006) have shown that the latency of an ERP component peaking at around 300 ms after the stimulus onset correlated with the reaction time. Instead of the latency, Bacon-Macé et al. (2005) examined the magnitude of the animal vs. non-animal differences. The larger difference in the visual N1 and P2 correlated with higher categorization accuracy. Furthermore, Macé et al. (2005) found a correlation between the reaction time and the peak of the differential activity between animals and non-animals. In all those studies, such a correlation was interpreted as a signal of evidence accumulation.

We cannot correlate the ERP category differences with behavioral measures because of the different task requirements in our studies. The above-mentioned studies compared ERP differences for images with different recognizability. In contrast, we compared category differences between two different tasks. Nevertheless, if category differences in the visual ERPs reflect the percept, this means that by that time the evidence regarding the category decision had to be integrated. Therefore, we can also explain our results in terms of evidence accumulation processes. In Papers 1 and 3 we found larger animal-vehicle differences when participants had to differentiate those categories, compared to the category-independent task (letter discrimination). In line with the results of Puri et al. (2009), this result can reflect greater selectivity, and possibly, more effective evidence accumulation regarding category membership.

In paper 3, we argued that the earlier animal-vehicle difference and its task-related modulation for stimuli containing intact SFs were due to more effective evidence accumulation. If a stimulus contains more SF bands, then category-specific information can be collected parallel from these different SFs. In contrast, if a stimulus contains only low or high SFs, then evidence accumulation should rely only on coarse or fine-grained information alone. In general, stimuli containing broader scales and more types of low-level information, including, for instance, color, amplitude spectrum, luminance, contrast and spatial frequencies, could be analyzed more effectively. The presence of these different cues in an image can provide more evidence and it can promote faster or stronger activation of category representations in the higher-level areas. This can be reflected in the time window or ERP component where the category-specificity appears.

We found interactions between top-down and bottom-up processes at the level of category representations in different ERP components for different stimulus types. For colorful, non-modified stimuli, task-related modulation related to category representation was found in earlier ERP components than for highly modified, luminance and AS-matched stimuli. Task-related category difference modulation was found in the N1 component in (Linnert et al., in prep) for natural scenes containing a car or a bird, and in the P1component in Paper 4 for colorful tools and vehicles. In both studies, stimuli were not matched for any low-level characteristics. In Paper 1 and partly in Paper 3, on the other hand, task-effects at the level of category representations were found only later, for the P2 component. The stimulus categories were different in these papers. Linnert et al. (in prep) used basic level categories, whereas in papers 1 and 3 we used superordinate, animal-vehicle categories. However, in Linnert et al. (in prep), cars and birds were embedded in natural scenes, whereas in Papers 1 and 3 objects appeared on a grey background. Furthermore, in Paper 3, for the stimuli containing intact SFs, we did find task-related modulation in the time window of the N1; however, the magnitude of this modulation was not as strong as in Linnert et al. (in prep). Therefore, it is not likely that the later task-effects in paper 1 and 3 are due to the different category levels. Evidence accumulation processes can explain these differences. In Linnert et al. (in prep), images contained low-level differences, category-specific AS, intact SFs, and even colors. All these factors can contribute to fast evidence accumulation and recognition. In contrast, in Papers 1 and 3, stimuli were highly modified, and for most of the stimuli even the AS information, which has been shown to contribute to category-specific processes, was matched. Therefore, evidence accumulation could take a longer time. This explanation is further supported by the fact that the ERP waveforms look different in these studies. In Linnert et al. (in prep), only the P1 and N1 components were distinguishable on the ERPs, whereas in Papers 1 and 3, the ERP waveforms contained P1, N1, P2, N2 components. The additional components in Papers 1 and 3 can be explained as additional processing to accumulate category-specific evidence. Indeed, it has been suggested that P2 component reflects re-activation processes in case of competing or insufficient information for decision-making (Kotsoni et al., 2007).

Therefore, evidence accumulation processes can explain both the task-related modulation of the ERP differences, and the different temporal course of the interaction between top-down and bottom-up processes in terms of low-level information content of the stimuli. In contrast, selective attention alone, can explain only the task-effect, but not the difference in the temporal course.

Flexibility of visual processing

Our results imply flexible visual processing mechanisms. We showed that top-down and bottom-up processes can interact at different levels, and this interaction depends on the properties of the visual input and on the task requirements. Fitting these results into the framework of evidence accumulation, the results suggest a flexible reliance on low-level visual features and top-down factors in evidence gathering. Category-specific visual areas accumulate evidence from the presented stimuli even if the task does not require categorization or recognition (Peelen et al., 2009; Zhu et al., 2016). Therefore, top-down processes are not necessary, but they can modulate evidence accumulation. According to computational models, evidence accumulation can be modulated in two ways by (1) changing the course of the accumulation process, or (2) changing the threshold (e.g. Domenech & Dreher, 2010). We found top-down, task related modulation selectively at the level of category representations. In contrast, we did not find task-related modulation in all ERP components or for all low-level stimulus properties. In other words, the task did not modulate all stages of early visual processing. These results suggest that the task modulates the threshold rather than the evidence accumulation process. For instance, if ‘animal’ is one of two categories that are expected in a categorization task, then a lower threshold, and consequently less evidence, could be enough in order to make the decision. This is in line with an fMRI study combined with modeling by Domenech and Dreher (2010), showing that the predictability of a subsequent event decreases the decision threshold.

Beside the relative contribution of top-down processes, evidence accumulation also seems to flexibly rely on the bottom-up information sources. For instance, Kauffmann et al. (2015) showed that category-specific visual areas can flexibly rely on low-level cues. Their results indicate that the activation of the scene-specific area (PPA) reflects interactions between spatial frequency content and contour of the presented visual scenes. When images showed strong contrast, the PPA was more sensitive to LSF information, whereas when contrast was modified it was more activated by HSF scenes (Kauffmann et al., 2015). Therefore, visual processes do not need specific low-level cues in order to recognize an object. Rather, the cues contained by a given stimulus can be flexibly used.

Therefore, visual processes can flexibly rely on bottom-up and top-down processes and on the interaction between these processes. This flexibility also shows up in the visual ERPs. ERPs are usually characterized in terms of function; for instance, it has been suggested that the visual P1 reflects low-level image characteristics, and the visual N1 reflects discrimination processes (Luck, 2005). However, flexible processes cannot be characterized by ERP components assigned to fixed functions. Our results indicate that the underlying processes reflected in the ERP, depend on the specific stimuli. We found category discrimination based on low-level features already in the P1, whereas the P2 and N2 components were still sensitive to low-level characteristics such as amplitude spectrum or spatial frequencies. Moreover, we found category-selectivity in different ERP components for different stimuli. Therefore, analyzing only one component according to a function defined in a study that used different stimuli does not necessarily give reliable results. Our findings suggest that ERPs do not reflect a specific function; rather, they can reflect flexible underlying mechanisms depending on the stimulus characteristics and the task. This evidence is in line with an emerging, more flexible view of the ERP components (e.g. Gruber et al., 2014). The original ‘evoked’ view of the ERPs assumed that stimuli elicit fixed latency and polarity components on a background oscillation. In contrast, according to the more recent phase reset model, the appearing stimulus resets the ongoing oscillation and therefore ERPs are not independent from the background oscillation (for review see Sauseng et al., 2007). Recent results of Gruber et al. (2014) and Van Der Lubbe, Szumska, & Fajkowska (2016) support the phase reset model, which indicates that the ongoing oscillation and ERPs are related. Moreover, Gruber et al. (2014) have shown that a late P1 component, peaking around 140 ms after stimulus onset, was preceded by another positive component at around 50 ms after stimulus onset. This contradicts the hypothesis about the fixed ERP components.

Concluding remarks

According to the traditional view of the ventral visual pathway (Ungerleider & Mishkin, 1982), visual processing progresses in a feed-forward fashion, and higher-level cognitive processes cannot access at least the early stages of visual processing (Pylyshyn, 1999). In contrast, more recent results and theories emphasize the flexibility of visual perception. Recent animal research and connectivity analysis of human fMRI data provided evidence of very complex networks of feedback, feed-forward and horizontal connections in visual areas and their connection with other brain regions (e.g. Angelucci et al., 2017; Thomas Yeo et al., 2011). Moreover, top-down projections can access any visual area, and even subcortical regions (for review see Gilbert & Sigman, 2007). It has been suggested that cortical areas do not reflect a given function. Rather, they function as "adaptive processors", where functions can be switched on a moment-by-moment basis according to the incoming sensory information, and top-down factors such as attention, task and expectations (Gilbert & Sigman, 2007). The results presented in this thesis about the flexible emergence of category representations are at least partly compatible with this view.

More recent analysis techniques, such as multivariate pattern analysis or connectivity analysis can reveal complex network patterns (Lang, Tomé, Keck, Górriz-Sáez, & Puntonet, 2012), decode the time-course of object recognition (e.g. Contini, Wardle, & Carlson, 2017) or representational structure of the visual cortex (Kaiser et al., 2016; Kriegeskorte et al., 2008). Nevertheless, as we have shown in this thesis, even the “simple” visual ERP analysis can also provide useful and novel information about flexible visual processes, if they are not considered as fixed components linked to fixed functions.

In summary, the current thesis demonstrates that visual ERPs reflect flexible interactions between top-down and bottom-up processes between task requirements and category-specific low- and higher-level processes. Moreover, it shows that the time-course of these interactions depends of the stimulus characteristics. Our results and their interpretations are in line with the emerging evidence showing that vision and cognition are not separate processes. Furthermore, visual processes are penetrable for top-down modulation even in the early stages of visual processing in order to flexibly and effectively process the incoming visual information according to the task requirements. As we have argued, the nature of this evidence also calls for a more flexible view of the ERPs of early visual processing.

References

Aguirre, G. K., Zarahn, E., & D'Esposito, M. (1998). An area within human ventral cortex sensitive to "building" stimuli: evidence and implications. *Neuron*, *21*, 373–383. https://doi.org/10.1016/S0896-6273(00)80546-2

Angelucci, A., Bijanzadeh, M., Nurminen, L., Federer, F., Merlin, S., & Bressloff, P. C. (2017). Circuits and mechanisms for surround modulation in visual cortex. *Annual Review of Neuroscience*, *40*(1), 425–451. https://doi.org/10.1146/annurev-neuro-072116-031418

Antal, A., Kéri, S., Kovács, G., Janka, Z., & Benedek, G. (2000). Early and late components of visual categorization: an event-related potential study. *Cognitive Brain Research*, *9*(1), 117–119. https://doi.org/10.1016/S0926-6410(99)00053-1

Antal, A., Kéri, S., Kovács, G., Liszli, P., Janka, Z., & Benedek, G. (2001). Event-related potentials from a visual categorization task. *Brain Research Protocols*, *7*, 131–136. https://doi.org/10.1016/S1385-299X(01)00055-1

Bacon-Macé, N., Macé, M. J. M., Fabre-Thorpe, M., & Thorpe, S. J. (2005). The time course of visual processing: Backward masking and natural scene categorisation. *Vision Research*, *45*, 1459–1469. https://doi.org/10.1016/j.visres.2005.01.004

Bugatus, L., Weiner, K. S., & Grill-Spector, K. (2017). Task alters category representations in prefrontal but not high-level visual cortex. *NeuroImage*, *155*, 437–449. https://doi.org/10.1016/j.neuroimage.2017.03.062

Chao, L. L., Haxby, J. V, & Martin, A. (1999). Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nature Neuroscience*, *2*(10), 913–919. https://doi.org/10.1038/13217

Contini, E. W., Wardle, S. G., & Carlson, T. A. (2017). Decoding the time-course of object recognition in the human brain: From visual features to categorical decisions. *Neuropsychologia*, *105*, 165–176. https://doi.org/10.1016/j.neuropsychologia.2017.02.013

Dilkina, K., & Lambon Ralph, M. A. (2012). Conceptual structure within and between modalities. *Frontiers in Human Neuroscience*, *6*, 333. https://doi.org/10.3389/fnhum.2012.00333

Domenech, P., & Dreher, J.-C. (2010). Decision threshold modulation in the human brain. *Journal of Neuroscience*, *30*(43), 14305–14317. https://doi.org/10.1523/JNEUROSCI.2371-10.2010

Downing, P. E. (2001). A cortical area selective for visual processing of the human body. *Science*, *293*(5539), 2470–2473. https://doi.org/10.1126/science.1063414

Epstein, R., & Kanwisher, N. (1998). A cortical representationof the local visual environment. *Nature*, *392*(6676), 598–601. https://doi.org/10.1038/33402

Esterman, M., & Yantis, S. (2010). Perceptual expectation evokes category-selective cortical activity. *Cerebral Cortex*, *20*, 1245–1253. https://doi.org/10.1093/cercor/bhp188

Gilbert, C. D., & Sigman, M. (2007). Brain states: Top-down influences in sensory processing. *Neuron*, *54*, 677–696. https://doi.org/10.1016/j.neuron.2007.05.019

Gruber, W. R., Zauner, A., Lechinger, J., Schabus, M., Kutil, R., & Klimesch, W. (2014). Alpha phase, temporal attention, and the generation of early event related potentials. *NeuroImage*, *103*, 119–129.

https://doi.org/10.1016/j.neuroimage.2014.08.055

Haxby, J. V, Gobbini, M. I., Furey, M. L., Ishai, a, Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science (New York, N.Y.)*, *293*(5539), 2425–2430. https://doi.org/10.1126/science.1063736

Ishai, A., Ungerleider, L. G., Martin, A., Schouten, J. L., & Haxby, J. V. (1999). Distributed representation of objects in the human ventral visual pathway. *Proceedings of the National Academy of Sciences*, *96*(16), 9379–9384. https://doi.org/10.1073/pnas.96.16.9379

Kaiser, D., Azzalini, D. C., & Peelen, M. V. (2016). Shape-independent object category responses revealed by MEG and fMRI decoding. *Journal of Neurophysiology*, *115*, 2246–2250. https://doi.org/10.1152/jn.01074.2015

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *17*(11), 4302–4311. https://doi.org/10.1098/Rstb.2006.1934

Kauffmann, L., Ramanoel, S., Guyader, N., Chauvin, A., & Peyrin, C. (2015). Spatial frequency processing in scene-selective cortical regions. *NeuroImage*, *112*, 86–95. https://doi.org/10.1016/j.neuroimage.2015.02.058

Kiani, R., Esteky, H., Mirpour, K., & Tanaka, K. (2007). Object category structure in response patterns of neuronal population in monkey inferior temporal cortex. *Journal of Neurophysiology*, *97*(6), 4296–4309. https://doi.org/10.1152/jn.00024.2007

Kiefer, M. (2001). Perceptual and semantic sources of category-specific effects: event-related potentials during picture and word categorization. *Memory &*

Cognition, *29*(1), 100–116. https://doi.org/10.3758/BF03195745

Koida, K., & Komatsu, H. (2007). Effects of task demands on the responses of color-selective neurons in the inferior temporal cortex. *Nature Neuroscience*, *10*(1), 108–116. https://doi.org/10.1038/nn1823

Kotsoni, E., Csibra, G., Mareschal, D., & Johnson, M. H. (2007). Electrophysiological correlates of common-onset visual masking. *Neuropsychologia*, *45*(10), 2285–2293. https://doi.org/10.1016/j.neuropsychologia.2007.02.023

Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., … Bandettini, P. A. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron*, *60*(6), 1126–1141. https://doi.org/10.1016/j.neuron.2008.10.043

Lang, E. W., Tomé, A. M., Keck, I. R., Górriz-Sáez, J. M., & Puntonet, C. G. (2012). Brain connectivity analysis: A short survey. *Computational Intelligence and Neuroscience*, *2012*(iii), 1–21. https://doi.org/10.1155/2012/412512

Linnert, S., Csifcsák, G., & Tompa, T. (in prep). Top-down modulation of animate-inanimate category differences in the visual N1.

Luck, S. J. (2005). *An Introduction to the event-related potential technique*. Cambridge, Mass: MIT press.

Macé, M. J. M., Thorpe, S. J., & Fabre-Thorpe, M. (2005). Rapid categorization of achromatic natural scenes: How robust at very low contrasts? *European Journal of Neuroscience*, *21*, 2007–2018. https://doi.org/10.1111/j.1460-9568.2005.04029.x

Mur, M., Meys, M., Bodurka, J., Goebel, R., Bandettini, P. A., & Kriegeskorte, N. (2013). Human object-similarity judgments reflect and transcend the primate-IT object representation. *Frontiers in Psychology*, *4*, 1–22.

https://doi.org/10.3389/fpsyg.2013.00128

O'Craven, K. M., & Kanwisher, N. (2000). Mental imagery of faces and places activates corresponding stiimulus-specific brain regions. *Journal of Cognitive Neuroscience*, *12*, 1013–1023. https://doi.org/10.1162/08989290051137549

Peelen, M. V., Fei-Fei, L., & Kastner, S. (2009). Neural mechanisms of rapid natural scene categorization in human visual cortex. *Nature*, *460*(7251), 94–97. https://doi.org/10.1038/nature08103

Peelen, M. V, & Kastner, S. (2011). A neural basis for real-world visual search in human occipitotemporal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 12125–12130. https://doi.org/10.1073/pnas.1101042108

Philiastides, M. G., & Sajda, P. (2006). Temporal characterization of the neural correlates of perceptual decision making in the human brain. *Cerebral Cortex*, *16*(April), 509–518. https://doi.org/10.1093/cercor/bhi130

Puri, A. M., Wojciulik, E., & Ranganath, C. (2009). Category expectation modulates baseline and stimulus-evoked activity in human inferotemporal cortex. *Brain Research*, *1301*, 89–99. https://doi.org/10.1016/j.brainres.2009.08.085

Pylyshyn, Z. (1999). Is vision continuous with cognition?: The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences*, *22*(03), 341-365; discussion 366-423. https://doi.org/10.1017/S0140525X99002022

Reddy, L., Tsuchiya, N., & Serre, T. (2010). Reading the mind's eye: Decoding category information during mental imagery. *NeuroImage*, *50*(2), 818–825. https://doi.org/10.1016/j.neuroimage.2009.11.084

Rokszin, A. A., Győri-Dani, D., Linnert, S., Krajcsi, A., Tompa, T., & Csifcsák, G. (2015). The interplay of holistic shape, local feature and color information in object categorization. *Biological Psychology*, 1–12. https://doi.org/10.1016/j.biopsycho.2015.05.002

Sauseng, P., Klimesch, W., Gruber, W. R., Hanslmayr, S., Freunberger, R., & Doppelmayr, M. (2007). Are event-related potential components generated by phase resetting of brain oscillations? A critical discussion. *Neuroscience*, *146*(4), 1435–1444. https://doi.org/10.1016/j.neuroscience.2007.03.014

Simen, P. (2012). Evidence accumulator or decision threshold – which cortical mechanism are we observing? *Frontiers in Psychology*, *3*, 1–14. https://doi.org/10.3389/fpsyg.2012.00183

Stokes, M., Thompson, R., Nobre, A. C., & Duncan, J. (2009). Shape-specific preparatory activity mediates attention to targets in human visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(46), 19569–19574. https://doi.org/10.1073/pnas.0905306106

Thomas Yeo, B. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., … Buckner, R. L. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology*, *106*(3), 1125–1165. https://doi.org/10.1152/jn.00338.2011

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, *381*(6582), 520–522. https://doi.org/10.1038/381520a0

Tompa, T., & Sáry, G. (2010). A review on the inferior temporal cortex of the macaque. *Brain Research Reviews*, *62*(2), 165–182. https://doi.org/10.1016/j.brainresrev.2009.10.001

Tremel, J. J., & Wheeler, M. E. (2015). Content-specific evidence accumulation in inferior temporal cortex during perceptual decision-making. *NeuroImage*, *109*, 35–49. https://doi.org/10.1016/j.neuroimage.2014.12.072

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In *Analysis of visual behavior* (pp. 549–586). Cambridge, Mass: The MIT Press.

Van Der Lubbe, R. H. J., Szumska, I., & Fajkowska, M. (2016). Two sides of the same coin: ERP and wavelet analyses of visual potentials evoked and induced by task-relevant faces. *Advances in Cognitive Psychology*, *12*(4), 153–167. https://doi.org/10.5709/acp-0195-3

VanRullen, R., & Thorpe, S. J. (2001). The time course of visual processing: from early perception to decision-making. *Journal of Cognitive Neuroscience*, *13*, 454–461. https://doi.org/10.1162/08989290152001880

Zhu, W., Drewes, J., Peatfield, N., & Melcher, D. (2016). Differential visual processing of animal images, with and without conscious awareness. *Frontiers in Human Neuroscience*, *10*(October), 513. https://doi.org/10.3389/FNHUM.2016.00513

Consolidated references

Aguirre, G. K., Zarahn, E., & D'Esposito, M. (1998). An area within human ventral cortex sensitive to "building" stimuli: evidence and implications. *Neuron*, *21*, 373–383. https://doi.org/10.1016/S0896-6273(00)80546-2

Ahissar, M., & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends in Cognitive Sciences*, *8*(10), 457–464. https://doi.org/10.1016/j.tics.2004.08.011

Andrews, T. J., Clarke, A., Pell, P., & Hartley, T. (2010). Selectivity for low-level features of objects in the human ventral stream. *NeuroImage*, *49*(1), 703–711. https://doi.org/10.1016/j.neuroimage.2009.08.046

Angelucci, A., Bijanzadeh, M., Nurminen, L., Federer, F., Merlin, S., & Bressloff, P. C. (2017). Circuits and mechanisms for surround modulation in visual cortex. *Annual Review of Neuroscience*, *40*(1), 425–451. https://doi.org/10.1146/annurev-neuro-072116-031418

Antal, A., Kéri, S., Kovács, G., Janka, Z., & Benedek, G. (2000). Early and late components of visual categorization: an event-related potential study. *Cognitive Brain Research*, *9*(1), 117–119. https://doi.org/10.1016/S0926-6410(99)00053-1

Antal, A., Kéri, S., Kovács, G., Liszli, P., Janka, Z., & Benedek, G. (2001). Event-related potentials from a visual categorization task. *Brain Research Protocols*, *7*, 131–136. https://doi.org/10.1016/S1385-299X(01)00055-1

Awasthi, B., Sowman, P. F., Friedman, J., & Williams, M. A. (2013). Distinct spatial scale sensitivities for early categorization of faces and places: neuromagnetic and behavioral findings. *Frontiers in Human Neuroscience*, *7*(March), 91. https://doi.org/10.3389/fnhum.2013.00091

Bacon-Macé, N., Macé, M. J. M., Fabre-Thorpe, M., & Thorpe, S. J. (2005). The time course of visual processing: Backward masking and natural scene categorisation. *Vision Research*, *45*, 1459–1469. https://doi.org/10.1016/j.visres.2005.01.004

Balas, B., & Koldewyn, K. (2013). Early visual ERP sensitivity to the species and animacy of faces. *Neuropsychologia*, *51*(13), 2876–2881. https://doi.org/10.1016/j.neuropsychologia.2013.09.014

Bar, M. (2003). A cortical mechanism for triggering top-down facilitation in visual object recognition. *Journal of Cognitive Neuroscience*, *15*(4), 600–609. https://doi.org/10.1162/089892903321662976

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, a M., Schmidt, A. M., … Halgren, E. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(2), 449–454. https://doi.org/10.1073/pnas.0507062103

Battistoni, E., Stein, T., & Peelen, M. V. (2017). Preparatory attention in visual cortex. *Annals of the New York Academy of Sciences*, *1396*, 92–107. https://doi.org/10.1111/nyas.13320

Behrmann, M., & Kimchi, R. (2003). What does visual agnosia tell us about perceptual organization and its relationship to object perception? *Journal of Experimental Psychology: Human Perception and Performance*, *29*(1), 19–42. https://doi.org/10.1037/0096-1523.29.1.19

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. https://doi.org/10.1162/jocn.1996.8.6.551

Bentin, S., Taylor, M. J., Rousselet, G. A., Itier, R. J., Caldara, R., Schyns, P. G., … Rossion, B. (2007). Controlling interstimulus perceptual variance does not abolish

N170 face sensitivity. *Nature Neuroscience*, *10*(7), 801–802. https://doi.org/10.1038/nn0707-801

Bieniek, M. M., Pernet, C. R., & Rousselet, G. A. (2012). Early ERPs to faces and objects are driven by phase, not amplitude spectrum information: evidence from parametric, test-retest, single-subject analyses. *Journal of Vision*, *12*(13), 12. https://doi.org/10.1167/12.13.12

Bognár, A., Csete, G., Németh, M., Csibri, P., Kincses, T. Z., & Sáry, G. (2017). Transcranial stimulation of the orbitofrontal cortex affects decisions about magnocellular optimized stimuli. *Frontiers in Neuroscience*, *11*(APR), 1–11. https://doi.org/10.3389/fnins.2017.00234

Bugatus, L., Weiner, K. S., & Grill-Spector, K. (2017). Task alters category representations in prefrontal but not high-level visual cortex. *NeuroImage*, *155*, 437–449. https://doi.org/10.1016/j.neuroimage.2017.03.062

Bullier, J. (2001). Integrated model of visual processing. *Brain Research Reviews*, *36*(2–3), 96–107. https://doi.org/10.1016/S0165-0173(01)00085-6

Canário, N., Jorge, L., Loureiro Silva, M. F., Alberto Soares, M., & Castelo-Branco, M. (2016). Distinct preference for spatial frequency content in ventral stream regions underlying the recognition of scenes, faces, bodies and other objects. *Neuropsychologia*, *87*, 110–119. https://doi.org/10.1016/j.neuropsychologia.2016.05.010

Cant, J. S., & Goodale, M. A. (2007). Attention to form or surface properties modulates different regions of human occipitotemporal cortex. *Cerebral Cortex*, *17*(3), 713–731. https://doi.org/10.1093/cercor/bhk022

Chao, L. L., Haxby, J. V, & Martin, A. (1999). Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nature Neuroscience*, *2*(10), 913–919. https://doi.org/10.1038/13217

Chao, L. L., & Martin, a. (2000). Representation of manipulable man-made objects in the dorsal stream. *NeuroImage*, *12*(4), 478–484. https://doi.org/10.1006/nimg.2000.0635

Cloutman, L. L. (2013). Interaction between dorsal and ventral processing streams: Where, when and how? *Brain and Language*, *127*(2), 251–263. https://doi.org/10.1016/j.bandl.2012.08.003

Codispoti, M., Ferrari, V., Junghöfer, M., & Schupp, H. T. (2006). The categorization of natural scenes: Brain attention networks revealed by dense sensor ERPs. *NeuroImage*, *32*, 583–591. https://doi.org/10.1016/j.neuroimage.2006.04.180

Collin, C. A., Therrien, M. E., Campbell, K. B., & Hamm, J. P. (2012). Effects of band-pass spatial frequency filtering of face and object images on the amplitude of N170. *Perception*, *41*(6), 717–732. https://doi.org/10.1068/p7056

Contini, E. W., Wardle, S. G., & Carlson, T. A. (2017). Decoding the time-course of object recognition in the human brain: From visual features to categorical decisions. *Neuropsychologia*, *105*, 165–176. https://doi.org/10.1016/j.neuropsychologia.2017.02.013

Costen, N. P., Parker, D. M., & Craw, I. (1994). Spatial content and spatial quantisation effects in face recognition. *Perception*, *23*(2), 129–146. https://doi.org/10.1068/p230129

Costen, N. P., Parker, D. M., & Craw, I. (1996). Effects of high-pass and low-pass spatial filtering on face identification. *Perception & Psychophysics*, *58*(4), 602–612. https://doi.org/10.3758/BF03213093

Craddock, M., Martinovic, J., & Müller, M. M. (2013). Task and spatial frequency modulations of object processing: An EEG study. *PLoS ONE*, *8*(7), 1–12. https://doi.org/10.1371/journal.pone.0070293

Curran, T., Tanaka, J. W., & Weiskopf, D. M. (2002). An electrophysiological comparison of visual categorization and recognition memory. *Cognitive, Affective & Behavioral Neuroscience*, *2*(1), 1–18.

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. https://doi.org/10.1016/j.jneumeth.2003.10.009

Delorme, A., Rousselet, G. A., Macé, M. J.-M., & Fabre-Thorpe, M. (2004). Interaction of top-down and bottom-up processing in the fast visual analysis of natural scenes. *Cognitive Brain Research*, *19*(2), 103–113. https://doi.org/10.1016/j.cogbrainres.2003.11.010

Di Lollo, V., Enns, J. T., & Rensink, R. (2000). Competition for consciousness among visual events: the psychophysics of reentrant visual processes. *Journal of Experimental Psychology*, *129*(4), 481–507. https://doi.org/10.1037/0096-3445.129.4.481

Dilkina, K., & Lambon Ralph, M. A. (2012). Conceptual structure within and between modalities. *Frontiers in Human Neuroscience*, *6*, 333. https://doi.org/10.3389/fnhum.2012.00333

Domenech, P., & Dreher, J.-C. (2010). Decision threshold modulation in the human brain. *Journal of Neuroscience*, *30*(43), 14305–14317. https://doi.org/10.1523/JNEUROSCI.2371-10.2010

Downing, P. E. (2001). A cortical area selective for visual processing of the human body. *Science*, *293*(5539), 2470–2473. https://doi.org/10.1126/science.1063414

Epstein, R., & Kanwisher, N. (1998). A cortical representationof the local visual environment. *Nature*, *392*(6676), 598–601. https://doi.org/10.1038/33402

Esterman, M., & Yantis, S. (2010). Perceptual expectation evokes category-selective cortical activity. *Cerebral Cortex*, *20*, 1245–1253. https://doi.org/10.1093/cercor/bhp188

Flevaris, A. V., Robertson, L. C., & Bentin, S. (2008). Using spatial frequency scales for processing face features and face configuration: An ERP analysis. *Brain Research*, *1194*, 100–109. https://doi.org/10.1016/j.brainres.2007.11.071

Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2001). Categorical representation of visual stimuli in the primate prefrontal cortex. *Science*, *291*(2001), 312–316. https://doi.org/10.1126/science.291.5502.312

Gaspar, C. M., & Rousselet, G. A. (2009). How do amplitude spectra influence rapid animal detection? *Vision Research*, *49*(24), 3001–3012. https://doi.org/10.1016/j.visres.2009.09.021

Gilbert, C. D., & Li, W. (2013). Top-down influences on visual processing. *Nature Reviews Neuroscience*, *14*(5), 350–363. https://doi.org/10.1038/nrn3476

Gilbert, C. D., & Sigman, M. (2007). Brain states: Top-down influences in sensory processing. *Neuron*, *54*, 677–696. https://doi.org/10.1016/j.neuron.2007.05.019

Goddard, E., Carlson, T. A., Dermody, N., & Woolgar, A. (2016). Representational dynamics of object recognition: Feedforward and feedback information flows. *NeuroImage*, *128*, 385–397. https://doi.org/10.1016/j.neuroimage.2016.01.006

Goffaux, V., Gauthier, I., & Rossion, B. (2003). Spatial scale contribution to early visual differences between face and object processing. *Cognitive Brain Research*, *16*(3), 416–424. https://doi.org/10.1016/S0926-6410(03)00056-9

Goffaux, V., Jemel, B., Jacques, C., Rossion, B., & Schyns, P. G. (2003). ERP evidence for task modulations on face perceptual processing at different spatial scales. *Cognitive Science*, *27*(2), 313–325. https://doi.org/10.1016/S0364-0213(03)00002-8

Goffaux, V., Peters, J., Haubrechts, J., Schiltz, C., Jansma, B., & Goebel, R. (2011). From coarse to fine? Spatial and temporal dynamics of cortical face processing. *Cerebral Cortex*, *21*(2), 467–476. https://doi.org/10.1093/cercor/bhq112

Goodale, M. A., & Keith Humphrey, G. (1998). The objects of action and perception. *Cognition*, *67*(1–2), 181–207. https://doi.org/10.1016/S0010-0277(98)00017-1

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, *15*(1), 20–25. https://doi.org/10.1016/0166-2236(92)90344-8

Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision Research*, *41*(10–11), 1409–1422. https://doi.org/10.1016/S0042-6989(01)00073-6

Gruber, W. R., Zauner, A., Lechinger, J., Schabus, M., Kutil, R., & Klimesch, W. (2014). Alpha phase, temporal attention, and the generation of early event related potentials. *NeuroImage*, *103*, 119–129. https://doi.org/10.1016/j.neuroimage.2014.08.055

Halit, H., de Haan, M., Schyns, P. G., & Johnson, M. H. (2006). Is high-spatial frequency information used in the early stages of face detection? *Brain Research*, *1117*(1), 154–161. https://doi.org/10.1016/j.brainres.2006.07.059

Harel, A., & Bentin, S. (2009). Stimulus type, level of categorization, and spatial-frequencies utilization: Implications for perceptual categorization hierarchies. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(4), 1264–1273. https://doi.org/10.1037/a0013621

Harel, A., Ullman, S., Epshtein, B., & Bentin, S. (2007). Mutual information of image fragments predicts categorization in humans: Electrophysiological and behavioral evidence. *Vision Research*, *47*(2007), 2010–2020. https://doi.org/10.1016/j.visres.2007.04.004

Haxby, J. V, Gobbini, M. I., Furey, M. L., Ishai, a, Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science (New York, N.Y.)*, *293*(5539), 2425–2430. https://doi.org/10.1126/science.1063736

Ishai, A., Ungerleider, L. G., Martin, A., Schouten, J. L., & Haxby, J. V. (1999). Distributed representation of objects in the human ventral visual pathway. *Proceedings of the National Academy of Sciences*, *96*(16), 9379–9384. https://doi.org/10.1073/pnas.96.16.9379

Jacques, C., & Rossion, B. (2007). Early electrophysiological responses to multiple face orientations correlate with individual discrimination performance in humans. *NeuroImage*, *36*(3), 863–876. https://doi.org/10.1016/j.neuroimage.2007.04.016

Jiang, X., Bradley, E., Rini, R. a., Zeffiro, T., VanMeter, J., & Riesenhuber, M. (2007). Categorization training results in shape- and category-selective human neural plasticity. *Neuron*, *53*, 891–903. https://doi.org/10.1016/j.neuron.2007.02.015

Johnson, J. S., & Olshausen, B. A. (2005). The earliest EEG signatures of object recognition in a cued-target task are postsensory. *Journal of Vision*, *5*(4), 2. https://doi.org/10.1167/5.4.2

Johnson, J. S., & Olshausen, B. (2003). Timecourse of neural signatures of object recognition. *Journal of Vision*, *3*(7), 499–512. https://doi.org/10.1167/3.7.4

Kaiser, D., Azzalini, D. C., & Peelen, M. V. (2016). Shape-independent object category responses revealed by MEG and fMRI decoding. *Journal of Neurophysiology*, *115*, 2246–2250. https://doi.org/10.1152/jn.01074.2015

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *17*(11), 4302–

4311. https://doi.org/10.1098/Rstb.2006.1934

Kauffmann, L., Ramanoel, S., Guyader, N., Chauvin, A., & Peyrin, C. (2015). Spatial frequency processing in scene-selective cortical regions. *NeuroImage*, *112*, 86–95. https://doi.org/10.1016/j.neuroimage.2015.02.058

Kauffmann, L., Ramanoël, S., & Peyrin, C. (2014). The neural bases of spatial frequency processing during scene perception. *Frontiers in Integrative Neuroscience*, *8*, 37. https://doi.org/10.3389/fnint.2014.00037

Kiani, R., Esteky, H., Mirpour, K., & Tanaka, K. (2007). Object category structure in response patterns of neuronal population in monkey inferior temporal cortex. *Journal of Neurophysiology*, *97*(6), 4296–4309. https://doi.org/10.1152/jn.00024.2007

Kiefer, M. (2001). Perceptual and semantic sources of category-specific effects: event-related potentials during picture and word categorization. *Memory & Cognition*, *29*(1), 100–116. https://doi.org/10.3758/BF03195745

Kihara, K., & Takeda, Y. (2012). Attention-free integration of spatial frequency-based information in natural scenes. *Vision Research*, *65*(21), 38–44. https://doi.org/10.1016/j.visres.2012.06.008

Kincses, T. Z., Chadaide, Z., Varga, E. T., Antal, A., & Paulus, W. (2006). Task-related temporal and topographical changes of cortical activity during ultra-rapid visual categorization. *Brain Research*, *1112*, 191–200. https://doi.org/10.1016/j.brainres.2006.07.044

Koida, K., & Komatsu, H. (2007). Effects of task demands on the responses of color-selective neurons in the inferior temporal cortex. *Nature Neuroscience*, *10*(1), 108–116. https://doi.org/10.1038/nn1823

Kotsoni, E., Csibra, G., Mareschal, D., & Johnson, M. H. (2007). Electrophysiological correlates of common-onset visual masking. *Neuropsychologia*, *45*(10), 2285–2293. https://doi.org/10.1016/j.neuropsychologia.2007.02.023

Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., … Bandettini, P. A. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron*, *60*(6), 1126–1141. https://doi.org/10.1016/j.neuron.2008.10.043

Lang, E. W., Tomé, A. M., Keck, I. R., Górriz-Sáez, J. M., & Puntonet, C. G. (2012). Brain connectivity analysis: A short survey. *Computational Intelligence and Neuroscience*, *2012*, 1–21. https://doi.org/10.1155/2012/412512

Lerner, Y., Hendler, T., Ben-Bashat, D., Harel, M., & Malach, R. (2001). A hierarchical axis of object processing stages in the human visual cortex. *Cerebral Cortex*, *11*(4), 287–297. https://doi.org/10.1093/cercor/11.4.287

Ling, S., Liu, T., & Carrasco, M. (2009). How spatial and feature-based attention affect the gain and tuning of population responses. *Vision Research*, *49*(10), 1194–1204. https://doi.org/10.1016/j.visres.2008.05.025

Linnert, S., Csifcsák, G., & Tompa, T. (in prep). Top-down modulation of animate-inanimate category differences in the visual N1.

Linnert, S., Reid, V., & Westermann, G. (in prep). The role of configural information in inanimate object categorzation. *Neuropsychologia*.

Linnert, S., Reid, V., & Westermann, G. (in prep). Underlying mechanisms of task-related modulation of animal vs. vehicle differences in the visual ERPs.

Luck, S. J. (2005). *An introduction to the event-related potential technique*. Cambridge, Mass: MIT press.

Macé, M. J. M., Thorpe, S. J., & Fabre-Thorpe, M. (2005). Rapid categorization of achromatic natural scenes: How robust at very low contrasts? *European Journal of*

Neuroscience, *21*, 2007–2018. https://doi.org/10.1111/j.1460-9568.2005.04029.x

Mahon, B. Z., & Caramazza, A. (2009). Concepts and categories: A cognitive neuropsychological perspective. *Annual Review of Psychology*, *60*(1), 27–51. https://doi.org/10.1146/annurev.psych.60.110707.163532

Mahon, B. Z., Milleville, S. C., Negri, G. A. L., Rumiati, R. I., Caramazza, A., & Martin, A. (2007). Action-related properties shape object representations in the ventral stream. *Neuron*, *55*(3), 507–520. https://doi.org/10.1016/j.neuron.2007.07.011

Malach, R., Reppas, J. B., Benson, R. R., Kwong, K. K., Jiang, H., Kennedy, W. a, … Tootell, R. B. (1995). Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proceedings of the National Academy of Sciences*, *92*(18), 8135–8139. https://doi.org/10.1073/pnas.92.18.8135

Martin, A. (2007). The representation of object concepts in the brain. *Annual Review of Psychology*, *58*, 25–45. https://doi.org/10.1146/annurev.psych.57.102904.190143

Martinez-Trujillo, J. C., & Treue, S. (2004). Feature-based attention increases the selectivity of population responses in primate visual cortex. *Current Biology*, *14*(9), 744–751. https://doi.org/10.1016/j.cub.2004.04.028

Matheson, H., Newman, A. J., Satel, J., & McMullen, P. (2014). Handles of manipulable objects attract covert visual attention: ERP evidence. *Brain and Cognition*, *86*(1), 17–23. https://doi.org/10.1016/j.bandc.2014.01.013

McManus, J. N. J., Li, W., & Gilbert, C. D. (2011). Adaptive shape processing in primary visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 9739–9746. https://doi.org/10.1073/pnas.1105855108

Mu, T., & Li, S. (2013). The neural signature of spatial frequency-based information integration in scene perception. *Experimental Brain Research.*, *227*(3), 367–77. https://doi.org/10.1007/s00221-013-3517-1

Mur, M., Meys, M., Bodurka, J., Goebel, R., Bandettini, P. A., & Kriegeskorte, N. (2013). Human object-similarity judgments reflect and transcend the primate-IT object representation. *Frontiers in Psychology*, *4*, 1–22. https://doi.org/10.3389/fpsyg.2013.00128

Musel, B., Kauffmann, L., Ramanoël, S., Giavarini, C., Guyader, N., Chauvin, A., & Peyrin, C. (2014). Coarse-to-fine categorization of visual scenes in scene-selective cortex. *Journal of Cognitive Neuroscience*, *26*(10), 2287–2297. https://doi.org/10.1162/jocn_a_00643

Nemeth, K., Kovacs, P., Vakli, P., Kovacs, G., & Zimmer, M. (2014). Phase noise reveals early category-specific modulation of the event-related potentials. *Frontiers in Psychology*, *5*(April). https://doi.org/10.3389/fpsyg.2014.00367

Nieuwenhuis, S., Forstmann, B. U., & Wagenmakers, E.-J. (2011). Erroneous analyses of interactions in neuroscience: a problem of significance. *Nature Neuroscience*, *14*(9), 1105–1107. https://doi.org/10.1038/nn.2886

O'Craven, K. M., & Kanwisher, N. (2000). Mental imagery of faces and places activates corresponding stiimulus-specific brain regions. *Journal of Cognitive Neuroscience*, *12*, 1013–1023. https://doi.org/10.1162/08989290051137549

Oliva, A., & Schyns, P. G. (1997). Coarse blobs or fine edges? Evidence that information diagnosticity changes the perception of complex visual stimuli. *Cognitive Psychology*, *34*(1), 72–107. https://doi.org/10.1006/cogp.1997.0667

Paz-Caballero, D., Cuetos, F., & Dobarro, A. (2006). Electrophysiological evidence for a natural/artifactual dissociation. *Brain Research*, *1067*, 189–200. https://doi.org/10.1016/j.brainres.2005.10.046

Peelen, M. V., Fei-Fei, L., & Kastner, S. (2009). Neural mechanisms of rapid natural scene categorization in human visual cortex. *Nature*, *460*(7251), 94–97.

https://doi.org/10.1038/nature08103

Peelen, M. V, & Kastner, S. (2011). A neural basis for real-world visual search in human occipitotemporal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 12125–12130. https://doi.org/10.1073/pnas.1101042108

Peyrin, C., Michel, C. M., Schwartz, S., Thut, G., Seghier, M., Landis, T., … Vuilleumier, P. (2010). The neural substrates and timing of top-down processes during coarse-to-fine categorization of visual scenes: a combined fMRI and ERP study. *Journal of Cognitive Neuroscience*, *22*(1994), 2768–2780. https://doi.org/10.1162/jocn.2010.21424

Philiastides, M. G., & Sajda, P. (2006). Temporal characterization of the neural correlates of perceptual decision making in the human brain. *Cerebral Cortex*, *16*(April), 509–518. https://doi.org/10.1093/cercor/bhi130

Pollux, P. M. J., Hall, S., Roebuck, H., & Guo, K. (2011). Event-related potential correlates of the interaction between attention and spatiotemporal context regularity in vision. *Neuroscience*, *190*, 258–269. https://doi.org/10.1016/j.neuroscience.2011.05.043

Proverbio, A. M., Adorni, R., & Aniello, G. E. (2011). 250 ms to code for action affordance during observation of manipulable objects. *Neuropsychologia*, *49*(9), 2711–2717. https://doi.org/10.1016/j.neuropsychologia.2011.05.019

Proverbio, A. M., Azzari, R., & Adorni, R. (2013). Is there a left hemispheric asymmetry for tool affordance processing? *Neuropsychologia*, *51*(13), 2690–2701. https://doi.org/10.1016/j.neuropsychologia.2013.09.023

Proverbio, A. M., Burco, F., Del Zotto, M., & Zani, A. (2004). Blue piglets? Electrophysiological evidence for the primacy of shape over color in object recognition. *Cognitive Brain Research*, *18*(3), 288–300. https://doi.org/10.1016/j.cogbrainres.2003.10.020

Proverbio, A. M., Del Zotto, M., & Zani, A. (2007). The emergence of semantic categorization in early visual processing: ERP indices of animal vs. artifact recognition. *BMC Neuroscience*, *8*, 24. https://doi.org/10.1186/1471-2202-8-24

Puri, A. M., Wojciulik, E., & Ranganath, C. (2009). Category expectation modulates baseline and stimulus-evoked activity in human inferotemporal cortex. *Brain Research*, *1301*, 89–99. https://doi.org/10.1016/j.brainres.2009.08.085

Pylyshyn, Z. (1999). Is vision continuous with cognition?: The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences*, *22*(03), 341-365; discussion 366-423. https://doi.org/10.1017/S0140525X99002022

Reddy, L., Tsuchiya, N., & Serre, T. (2010). Reading the mind's eye: Decoding category information during mental imagery. *NeuroImage*, *50*(2), 818–825. https://doi.org/10.1016/j.neuroimage.2009.11.084

Reeder, R. R., Perini, F., & Peelen, M. V. (2015). Preparatory activity in posterior temporal cortex causally contributes to object detection in scenes. *Journal of Cognitive Neuroscience*, *27*(11), 2117–2125. https://doi.org/10.1162/jocn_a_00845

Rizzi, C., Piras, F., & Marangolo, P. (2010). Top-down projections to the primary visual areas necessary for object recognition: A case study. *Vision Research*, *50*(11), 1074–1085. https://doi.org/10.1016/j.visres.2010.03.018

Rokszin, A. A., Győri-Dani, D., Linnert, S., Krajcsi, A., Tompa, T., & Csifcsák, G. (2015). The interplay of holistic shape, local feature and color information in object categorization. *Biological Psychology*, 1–12. https://doi.org/10.1016/j.biopsycho.2015.05.002

Rokszin, A. A., Győri-Dani, D., Nyúl, L. G., & Csifcsák, G. (2016). Electrophysiological correlates of top-down effects facilitating natural image categorization are disrupted by the attenuation of low spatial frequency information. *International Journal of Psychophysiology*, *100*, 19–27. https://doi.org/10.1016/j.ijpsycho.2015.12.006

Rossion, B., & Caharel, S. (2011). ERP evidence for the speed of face categorization in the human brain: Disentangling the contribution of low-level visual cues from face perception. *Vision Research*, *51*(12), 1297–1311. https://doi.org/10.1016/j.visres.2011.04.003

Rossion, B., Gauthier, I., Tarr, M. J., Despland, P., Bruyer, R., Linotte, S., & Crommelinck, M. (2000). The N170 occipito-temporal component is delayed and enhanced to inverted faces but not to inverted objects: an electrophysiological account of face-specific processes in the human brain. *Neuroreport*, *11*(1), 69–74. https://doi.org/10.1097/00001756-200001170-00014

Rossion, B., & Jacques, C. (2008). Does physical interstimulus variance account for early electrophysiological face sensitive responses in the human brain? Ten lessons on the N170. *NeuroImage*, *39*, 1959–1979. https://doi.org/10.1016/j.neuroimage.2007.10.011

Rousselet, G. A., Husk, J. S., Bennett, P. J., & Sekuler, A. B. (2008). Time course and robustness of ERP object and face differences. *Journal of Vision*, *8*(12), 1–18. https://doi.org/10.1167/8.12.3

Rousselet, G. a, Fabre-Thorpe, M., & Thorpe, S. J. (2002). Parallel processing in high-level categorization of natural images. *Nature Neuroscience*, *5*(7), 629–630. https://doi.org/10.1038/nn866

Rousselet, G. a, Pernet, C. R., Bennett, P. J., & Sekuler, A. B. (2008). Parametric study of EEG sensitivity to phase noise during face processing. *BMC Neuroscience*, *9*, 98. https://doi.org/10.1186/1471-2202-9-98

Ruiz-Soler, M., & Beltran, F. S. (2006). Face perception: An integrative review of the role of spatial frequencies. *Psychological Research*, *70*(4), 273–292. https://doi.org/10.1007/s00426-005-0215-z

Sauseng, P., Klimesch, W., Gruber, W. R., Hanslmayr, S., Freunberger, R., & Doppelmayr, M. (2007). Are event-related potential components generated by phase resetting of brain oscillations? A critical discussion. *Neuroscience*, *146*(4), 1435–1444. https://doi.org/10.1016/j.neuroscience.2007.03.014

Schendan, H. E., & Kutas, M. (2007). Neurophysiological evidence for the time course of activation of global shape, part, and local contour representations during visual object categorization and memory. *Journal of Cognitive Neuroscience*, *19*, 734–749. https://doi.org/10.1162/jocn.2007.19.5.734

Schyns, P. G. (1998). Diagnostic recognition: task constraints, object information, and their interactions. *Cognition*, *67*, 147–179. https://doi.org/10.1016/S0010-0277(98)00016-X

Schyns, P. G., & Oliva, A. (1994). From blobs to boundary edges: Evidence for time- and spatial-scale-dependent scene recognition. *Psychological Science*, *5*(4), 195–201.

Schyns, P. G., & Oliva, A. (1997). Flexible, diagnosticity-driven, rather than fixed, perceptually determined scale selection in scene and face recognition. *Perception*, *26*(8), 1027–1038. https://doi.org/10.1068/p261027

Schyns, P. G., & Oliva, A. (1999). Dr. Angry and Mr. Smile: when categorization flexibly modifies the perception of faces in rapid visual presentations. *Cognition*, *69*(3), 243–265. https://doi.org/10.1016/S0010-0277(98)00069-9

Servos, P., & Goodale, M. A. (1995). Preserved visual imagery in visual form agnosia. *Neuropsychologia*, *33*(11), 1383–1394. https://doi.org/10.1016/0028-3932(95)00071-A

Simen, P. (2012). Evidence accumulator or decision threshold – Which cortical mechanism are we observing? *Frontiers in Psychology*, *3*(JUN), 1–14. https://doi.org/10.3389/fpsyg.2012.00183

Sitnikova, T., West, W. C., Kuperberg, G. R., & Holcomb, P. J. (2006). The neural organization of semantic memory: Electrophysiological activity suggests feature-based segregation. *Biological Psychology*, *71*, 326–340. https://doi.org/10.1016/j.biopsycho.2005.07.003

Snyder, A. C., & Foxe, J. J. (2010). Anticipatory attentional suppression of visual features indexed by oscillatory alpha-band power increases: A high-density electrical mapping study. *Journal of Neuroscience*, *30*(11), 4024–4032. https://doi.org/10.1523/JNEUROSCI.5684-09.2010

Stokes, M., Thompson, R., Nobre, A. C., & Duncan, J. (2009). Shape-specific preparatory activity mediates attention to targets in human visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(46), 19569–19574. https://doi.org/10.1073/pnas.0905306106

Tanaka, J., Luu, P., Weisbrod, M., & Kiefer, M. (1999). Tracking the time course of object categorization using event-related potentials. *NeuroReport*, *10*(4), 829–835. https://doi.org/10.1097/00001756-199903170-00030

Tanaka, J. W., & Curran, T. (2001). A neural basis for expert object recognition. *Psychological Science*, *12*, 43–47. https://doi.org/10.1111/1467-9280.00308

Thierry, G., Martin, C. D., Downing, P. E., & Pegna, A. J. (2007a). Is the N170 sensitive to the human face or to several intertwined perceptual and conceptual factors? *Nature Neuroscience*, *10*(7), 802–803. https://doi.org/10.1038/nn0707-802

Thierry, G., Martin, C. D., Downing, P., & Pegna, A. J. (2007b). Controlling for interstimulus perceptual variance abolishes N170 face selectivity. *Nature*

Neuroscience, *10*(4), 505–511. https://doi.org/10.1038/nn1864

Thomas Yeo, B. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., … Buckner, R. L. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology*, *106*(3), 1125–1165. https://doi.org/10.1152/jn.00338.2011

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, *381*(6582), 520–522. https://doi.org/10.1038/381520a0

Thorpe, S. J., & Fabre-Thorpe, M. (2001). Seeking categories in the brain. *Science*, *291*, 260–263. https://doi.org/10.1126/science.1058249

Tokudome, W., & Wang, G. (2012). Similarity dependency of the change in ERP component N1 accompanying with the object recognition learning. *International Journal of Psychophysiology*, *83*(1), 102–109. https://doi.org/10.1016/j.ijpsycho.2011.10.012

Tompa, T., & Sáry, G. (2010). A review on the inferior temporal cortex of the macaque. *Brain Research Reviews*, *62*(2), 165–182. https://doi.org/10.1016/j.brainresrev.2009.10.001

Torralba, A., & Oliva, A. (2003). Statistics of natural image categories. *Network (Bristol, England)*, *14*(3), 391–412. https://doi.org/10.1088/0954-898X/14/3/302

Tremel, J. J., & Wheeler, M. E. (2015). Content-specific evidence accumulation in inferior temporal cortex during perceptual decision-making. *NeuroImage*, *109*, 35–49. https://doi.org/10.1016/j.neuroimage.2014.12.072

Ullman, S. (2007). Object recognition and segmentation by a fragment-based hierarchy. *Trends in Cognitive Sciences*, *11*(2), 58–64. https://doi.org/10.1016/j.tics.2006.11.009

Ullman, S., Vidal-Naquet, M., & Sali, E. (2002). Visual features of intermediate complexity and their use in classification. *Nature Neuroscience*, *5*(7), 682–687. https://doi.org/10.1038/nn870

Ungerleider, L.G., Mishkin, M. (1982). Two cortical visual systems. In R. J. W. Ingle, David J., Goodale, Melvyn A., Mansfield (Ed.), *Analysis of Visual Behaviour.* (pp. 549–580). Cambridge: The MIT Press.

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In *Analysis of visual behavior* (pp. 549–586). Cambridge, Mass: The MIT Press.

van der Linden, M., van Turennout, M., & Indefrey, P. (2010). Formation of category representations in superior temporal sulcus. *Journal of Cognitive Neuroscience*, *22*, 1270–1282. https://doi.org/10.1162/jocn.2009.21270

Van Der Lubbe, R. H. J., Szumska, I., & Fajkowska, M. (2016). Two sides of the same coin: ERP and wavelet analyses of visual potentials evoked and induced by task-relevant faces. *Advances in Cognitive Psychology*, *12*(4), 153–167. https://doi.org/10.5709/acp-0195-3

VanRullen, R. (2011). Four common conceptual fallacies in mapping the time course of recognition. *Frontiers in Psychology*, *2*, 1–6. https://doi.org/10.3389/fpsyg.2011.00365

VanRullen, R., & Thorpe, S. J. (2001). The time course of visual processing: from early perception to decision-making. *Journal of Cognitive Neuroscience*, *13*, 454–461. https://doi.org/10.1162/08989290152001880

Vogel, E. K., & Luck, S. J. (2000). The visual N1 component as an index of a discrimination process. *Psychophysiology*, *37*(2), 190–203. https://doi.org/10.1111/1469-8986.3720190

Vogel, J., & Wallraven, C. (2006). Categorization of natural scenes : local vs . global information. *International Journal of Computer Vision*, *4*(3), 33–40.

https://doi.org/10.1145/1278387.1278393

Warrington, E. K., & Shallice, T. (1984). Category specific semantic impairments. *Brain*, *107*(3), 829–853. https://doi.org/10.1093/brain/107.3.829

Whitwell, R. L., Milner, A. D., & Goodale, M. A. (2014). The two visual systems hypothesis: New challenges and insights from visual form agnosic patient DF. *Frontiers in Neurology*, *5*, 1–8. https://doi.org/10.3389/fneur.2014.00255

Wilder, J., Feldman, J., & Singh, M. (2011). Superordinate shape classification using natural shape statistics. *Cognition*, *119*(3), 325–340. https://doi.org/10.1016/j.cognition.2011.01.009

Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., & Tanaka, J. W. (2010). Controlling low-level image properties: the SHINE toolbox. *Behavior Research Methods*, *42*(3), 671–684. https://doi.org/10.3758/BRM.42.3.671

Zhu, W., Drewes, J., Peatfield, N., & Melcher, D. (2016). Differential visual processing of animal images, with and without conscious awareness. *Frontiers in Human Neuroscience*, *10*(October), 513. https://doi.org/10.3389/FNHUM.2016.00513