DOCTOR OF PHILOSOPHY

Electrophysiological studies of the time-course of 3D shape perception and object recognition

Oliver, Zoe

Award date:
2017

Awarding institution:
Bangor University

Electrophysiological studies of the time-course of 3D shape perception and object recognition

Zoe J. Oliver

PRIFYSGOL
BANGOR
UNIVERSITY

Submitted for the degree of Doctor of Philosophy

2017

School of Psychology

Bangor University

Acknowledgements

First and foremost I would like to thank Charles Leek for the opportunity to undertake this work but also for his encouragement and advice throughout my PhD. I couldn't have completed this without your enthusiasm, vast knowledge and support.

I would also like to thank the rest of my thesis committee: James Intriligator and Filipe Cristino for their encouragement and insightful comments throughout. Thank you to Mark Roberts for his assistance in the ERP analysis and again to Filipe Cristino for his extensive Matlab knowledge and guidance with programming.

I also wish to thank my family for their belief in me and encouragement to pursue my academic career. And in a similar vein, a special thank you to Roisin McKelvey for her support throughout this journey and to Laura Ingram for the encouragement during the writing stage.

Last but not least, I would like to express my gratitude to my partner, Tom Gardner. Thank you for your constant encouragement, helpful suggestions and unrelenting belief and support over the years.

Contents

List of Tables and Figures

Abstract

This thesis reports the results of three novel studies using event-related potentials (ERPs) to examine (1) how different kinds of shape information across local and global spatial scales are computed, and integrated, during the perception of 3D object shape, (2) the role of stereo information in 3D shape processing and (3) the temporal dynamics of shape information processing.

In experiment 1 we examined the time course of information processing at local and global spatial scales during object recognition and the role of stereo information in this processing. ERPs were recorded whilst participants completed a recognition memory task where they distinguished objects learned in training sessions from distracters that were either locally- or globally-similar. Participants completed the training and recognition task in either mono or stereo viewing conditions. The behavioural data showed a stereo advantage in object recognition and enhanced generalisation between trained and untrained views. The ERP data showed that during mono viewing, perceptual sensitivity was greatest for distracters with different local parts to targets during the N1. For stereo viewing, perceptual sensitivity was greatest for distracters with different 3D spatial configuration during the N2/P3 component. The findings show that there is differential ERP sensitivity to shape processing at local and global spatial scales and that stereo information is important in object recognition. The results, therefore, challenge theoretical models of object recognition that do not attribute functional significance to both 2D and 3D shape information.

In experiment 2 we investigated the perceptual integration of information from local and global spatial scales, to find a temporal-spatial EEG marker for this integration. A Navon-type paradigm was used, whereby participants' attention was directed to either the local or global level of the stimulus. The local and global levels could be either congruent or incongruent and the rationale was that congruency effects can be used as a functional marker for integration,

as congruency effects presumably arise at the point of global/local integration. ERPs were recorded whilst participants made orientation decisions about hierarchical stimuli made up of Gabor patches oriented to either the left or right. We found that there were interference effects evident in the behavioural and ERP data, particularly, global interference was evident at the N2/P3. We suggest that the global interference is evidence of global/local integration.

In experiment 3 we aimed to examine the robustness of the integration of local and global information found in experiment 2 using more complex stimuli and a different task. Our stimuli comprised sets of geometrically coherent (possible) or incoherent (impossible) objects. The impossible objects comprised local parts that were geometrically coherent and global configuration that was possible. The objects' impossibility becomes apparent when integrating the local and global levels of information. The rationale, therefore, was that the first point that the ERPs differed between possible and impossible object conditions would reflect the integration of information at local and global spatial scales. Using ERPs, we compared the processing of possible and impossible objects in a simple classification task. We found that there were no early processing differences for possible and impossible objects. However there were differences at the N2/P3 (from around 300ms post-stimulus onset) and we were able to verify that these differences did reflect the perceptual integration of local and global shape information. The results provided evidence of the generalisability of the integration effect at the N2/P3 with a more complex stimulus set.

The main empirical findings in this thesis show: (1) processing of information from local and global spatial scales first occurs at the N1 and information from these spatial scales are integrated at the N2/P3; (2) local and global processing occur at least partly in parallel; (3) stereo information plays a role in object recognition and ought to be included in models of object recognition; and (4) our findings challenge models of object recognition that do not include independent coding of object parts and their spatial relations.

1 Chapter I

1.1 Aims of the Thesis and Overview of Empirical Studies

The overarching aims of this thesis are to investigate how the human visual system perceives and recognises three-dimensional (3D) object shape. We test two specific theoretical hypotheses: (1) The computation of 3D object shape representations involves distinct processing of shape information across local (fine) and global (coarse) spatial scales; (2) Stereo disparity differentially modulates the perceptual processing of object shape across these spatial scales. We also aimed to investigate the time-course of the integration of information at local (fine) and global (coarse) spatial scales. To test these hypotheses we conducted three novel studies combining psychophysical tasks and high-density electroencephalography (EEG).

The thesis is structured as follows: in Chapter I, we outline of theories of object recognition and review processing at local and global spatial scales, followed by an overview of the temporal dynamics of object recognition. Chapter II comprises a brief overview of EEG methodology, followed by the particular methods used in Chapters III, IV and V.

In Chapter III we examine the role of stereo information in 3D shape processing, particularly any modulation of processing shape information at global and local spatial scales. The rationale was based on a recent ERP study by Leek, Roberts, Oliver, Cristino and Pegna (2016) who found differential sensitivity to local part structure and global shape configuration of complex 3D objects. The role of stereo information in 3D object perception is unclear; stereo information (i.e., local depth disparity) appears to facilitate processing of 3D surfaces properties, however the effects of stereo disparity on the perceptual matching of object shape across changes in viewpoint provide a mixed picture with stereo advantages reported in some studies but not in others. As stereo disparity seems to enhance regions of surface curvature at local part boundaries, we might expect to see an advantage in processing 3D spatial configuration when viewing objects in stereo. However, it may be that stereo information facilitates object recognition only when monocular shape cues are insufficient to identify an object; perhaps stereo advantage is only present when a match cannot be made based on 2D image cues. To investigate the role of stereo disparity in global and local processing we recorded ERPs and used a recognition memory task in which observers had to first memorize a sub-set of complex novel 3D objects (targets) and subsequently discriminate them from visually similar non-target (not previously memorised) objects. We then contrasted effects of target/non-target similarity defined by local part and global 3D shape configuration under conditions of stereo and mono viewing.

In Chapter IV we examine the processing of shape information at different spatial scales in non-object stimuli. We wanted to find ERP signatures for the integration of local and global information. We recorded ERPs whilst participants completed a simple orientation detection task using basic visual stimuli. We used arrays of Gabor patches orientated either to the left or right. The stimuli were designed to be similar to Navon letters, a type of hierarchical stimuli, but much more basic. Differences in local and global processing are reported frequently in ERP studies, with varying findings. As we used a Navon-style task with hierarchical stimuli, we expected to find evidence of a global precedence effect (GPE), including global interference, which presumably arises at the level of local and global integration.

In Chapter V, we examine the time course of integration of information at local and global spatial scales in complex objects. To do this, we utilised impossible objects – the difficulty in perception of impossible objects stems from the inability to form the representation of a coherent 3D structure, as one that exists in the 3D world does not exist. We recorded ERPs and used a simple classification task where participants had to decide whether objects were geometrically possible or impossible. Our impossible stimuli include local and global features that, alone, are possible. However, when the local and global features are integrated to form a 3D representation, the object's impossibility becomes apparent. Therefore, the rationale was that the differences in the ERPs for possible and impossible object trials should reflect the integration of local and global shape information.

In conclusion, the results of the work provide new empirical evidence that the perception of 3D object shape does involve processing of shape information across local (fine) and global (coarse) spatial scales, and that these processes are differentially modulated by stereo visual input. The integration of information across local (fine) and global (coarse) spatial scales occurs at the N2/P3 time frame. The findings challenge current theoretical models that do not attribute functional significance to the distinction between local and global information, or stereo visual input, during the perception and recognition of 3D objects in human vision.

1.2 The problem of object recognition in the human visual system

Human vision depends on a complex biological system: light reflects off surfaces and projects an image onto the retina, the brain then translates the information (colour, edges, vertices etc.) into a representation of shape that we retrieve from a stored representation in the brain. It is considered that from the retina, information flows through the visual cortex from V1 to higher visual areas (V2, V3, V4 etc.), processing information from colours, edges, vertices to more complex analyses of shape such as curvature. More fine-grained analysis then takes place, including top-down activation of temporal regions to identify objects by matching the information given to a representation of an object. Object recognition requires matching perceptual information to a stored representation in long-term memory (LTM).

However, the apparent ease with which we are able to perceive and recognise 3D object shape belies the complexity of the underlying computational problem. For example we are able to recover information (internal representations) about 3D object shape across variations in sensory data brought about by changes in object viewpoint, lighting (e.g., shading, shadow), colour, texture, size (scale) and motion. This poses a question concerning the extent of the information about an object that is encoded and stored in LTM and the bio-computational processes that support the derivation of 3D object shape to support image classification.

1.2.1 Models of object recognition

Different theories of object recognition make different claims about how shape is represented, making claims about viewpoint dependency, which low-level features make up shape representation and reference frames. The low-level features could be as basic as pixels (Liu, Knill & Kersten, 1995); more complex such as volumetric parts (Marr & Nishihara, 1978); categorical properties of object parts (Biederman, 1987; Hummel & Biederman, 1992; Hummel, 2001); surfaces (Leek, Reppa & Arguin, 2005) or something in between such as image fragments (Ullman, 2007); edges and vertices (Lowe, 1987; Poggio & Edelman, 1990); or collections of edges and vertices (Fukushima & Miyake, 1982; Riesenhuber & Poggio, 1999; 2002). Theories should also consider the reference frames that are stored with object representations: they could be viewer-centred (Hummel & Biederman, 1992; Poggio & Edelman, 1990; Riesenhuber & Poggio, 2002; Ullman & Basri, 1991); object-centred (Marr & Nishihara, 1978); or a mixture of viewer-centred and object-centred (Hummel, 2001; Hummel & Stankiewicz, 1996a; Lowe, 1987; Ullman, 1989).

1.2.1.1 Structural descriptions accounts

Structural description accounts of object recognition propose that objects are represented in terms of their parts' spatial relations to each other. Part attributes and their relations are represented explicitly and independently of each other. (Biederman, 1987; Dickson, Pentland & Rosenfeld, 1992; Hummel & Biederman, 1992; Marr & Nishihara, 1978). This account was proposed as a solution to the problem of view-invariant object recognition. The structural descriptions accounts differ in terms of what they propose as the component parts: Biederman (1987) suggests non-accidental properties (NAPs) as building blocks of volumetric parts; Leek et al. (2005) suggest that surfaces and their spatial configurations are the basic parts; whereas for Marr and Nishihara (1978), depth is important – with a progression from lines to contours, contours to surfaces, then surfaces to parts. Next we will consider in detail the structural description approaches of Biederman, Marr, and Leek et al..

1.2.1.1.1 Recognition by components model

Biederman (1987) proposed the 'Recognition by Components' (RBC) account, purporting that recognition of objects is a process whereby the visual input is segmented using concavities, then into an arrangement of simple geometric components such as cylinders, blocks, wedges and cones (termed geons). These components are derived from contrasts of properties of edges in a 2D image, properties including collinearity, curvature, symmetry, parallelism and cotermination.

The presumed sub-processes that lead to object recognition are shown in Figure 1 – the first stage is edge extraction: responding to differences in surface characteristics such as luminance, texture and colour, this provides an edge-based representation of the object. From edge extractions, there is detection of non-accidental properties (NAPs) and parsing at regions of concavity in parallel (NAPs provide constraints on the possibilities of identity of components). The next stage is determining of components, then matching those components to object representations, then finally object identification.

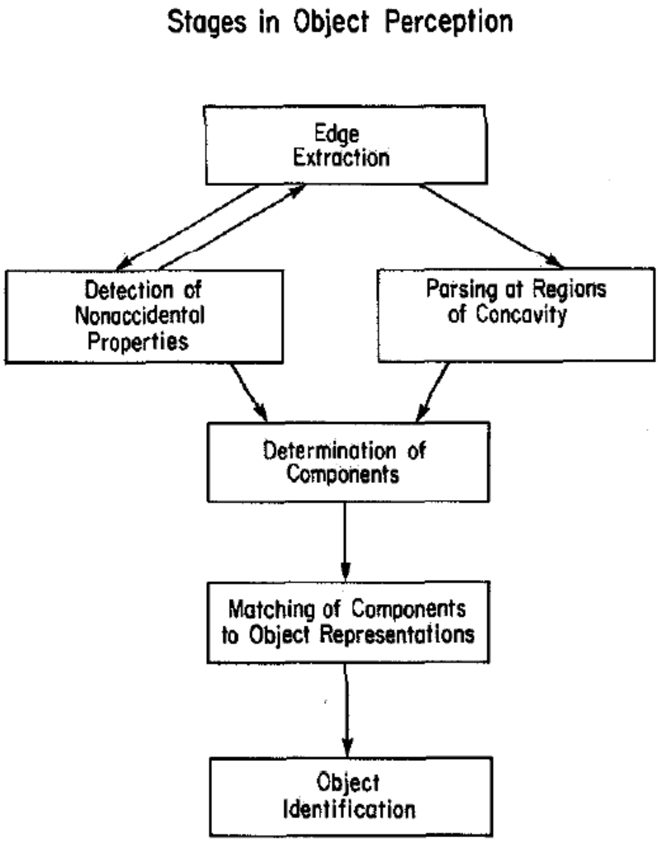


Figure 1. Presumed processing stages in object recognition. From Biederman (1987).

NAPs can be seen in Figure 2, they include collinearity (straight line in the image, presumes that the edge producing that line in the 3D world is also straight); curvilinearity (smoothly curved elements are similarly inferred to be produced by smoothly curved features in the 3D world); symmetry (if the image is symmetrical, we assume that the object producing the image is also symmetrical); parallelism (when edges are parallel, we assume that real-world edges are also parallel). These features are non-accidental as they would rarely be produced by accidental alignments of viewpoint and object features and so are usually unaffected by variations in viewpoint.

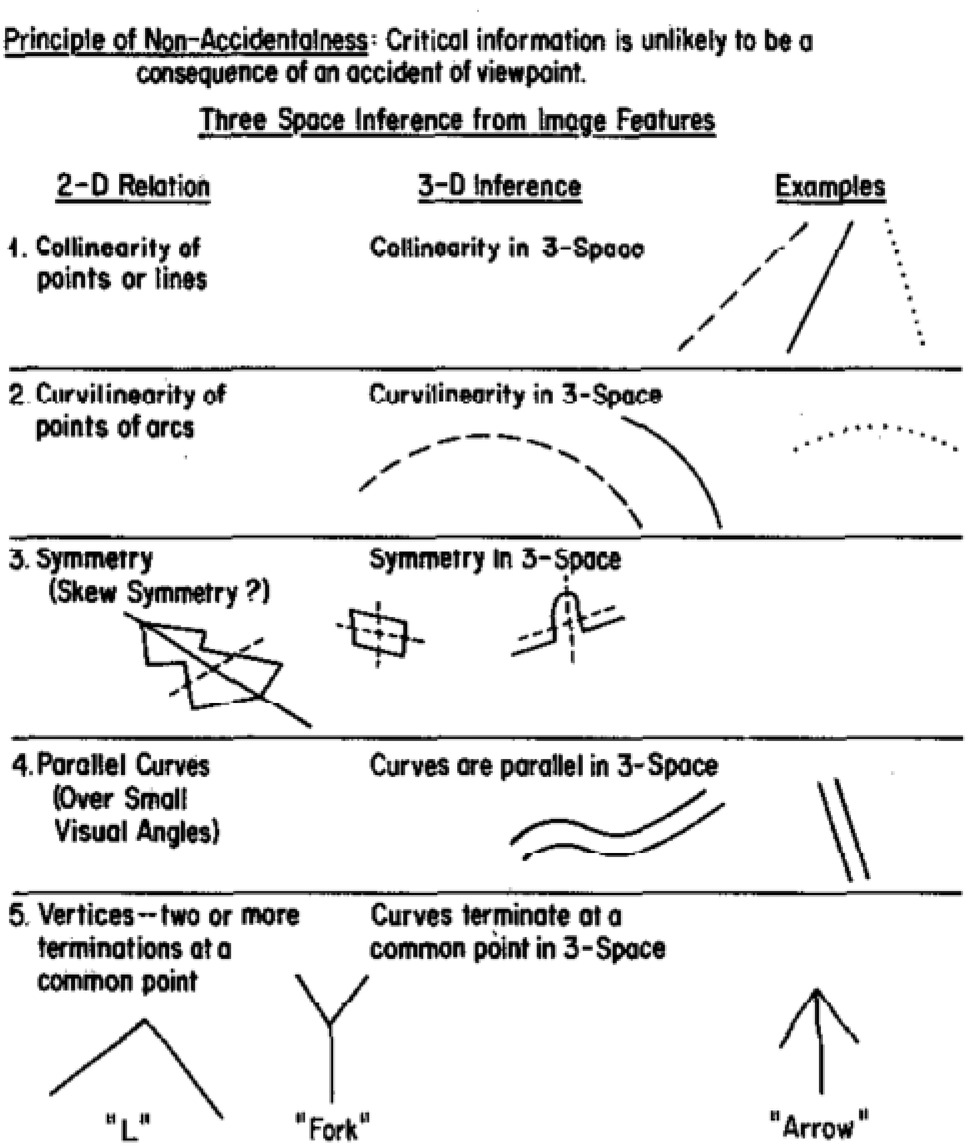


Figure 2. Five non-accidental relations. From Biederman (1987).

The arrangement of the shape primitives is necessary for representations of a particular object. Different arrangements of the same primitives yield different representations (see Figure 3). The representation of a particular object, therefore, needs to be a structural description that expresses the relations among components.

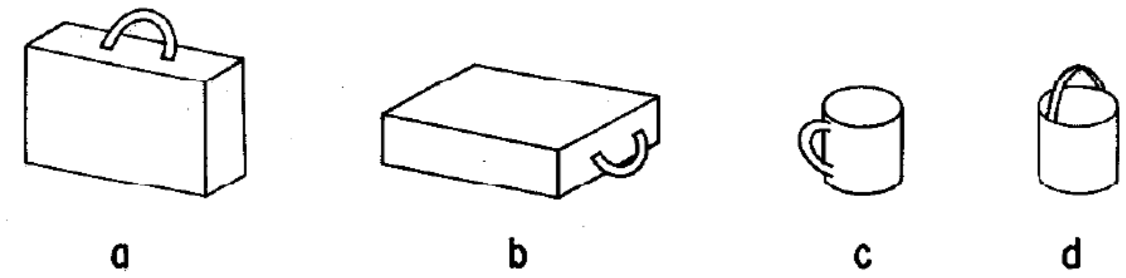


Figure 3. Different arrangements of the same components can produce different objects. From Biederman (1987).

Biederman and Blickle (1985, cited in Biederman 1987) provide some support for the RBC theory. They found that when identifying line drawings of objects with deleted parts objects at vertices of edges (non-recoverable degradation), the recovery of object components was disrupted, making it difficult to recognise, compared to when parts are removed from the midsection of a curve (recoverable degradation). Their results provided evidence that some contours presented in an image are vital for object identification, namely vertices of edges. A further study from Biederman (1987) used a similar task, showing image of line drawings with varying amounts deleted (25%, 45% and 65%) from internal and external contours, with varying exposure times (100, 200 or 750ms). Again, the deleted contours were either at the vertices or mid segments, but differing from the previous experiment, the removed mid segments did not bridge the components of collinearity or curvature. In the 100ms exposure condition, the 65% deleted contour condition resulted in higher error rates for removal of vertices compared to the midsection, but this disparity in error rates was reduced when exposure was longer and when there was a lower percentage of the image missing. They concluded that identification from filling-in of contours at mid segments and vertices can be completed within one second, but when there is a misleading component breaking the curvature, the image produced cannot be recovered, even with longer exposure time. However, the RBC model does not explain how representations are matched to new exemplars of objects.

1.2.1.1.2 Marr's Stage Model

Marr and Nishihara (1978) proposed that sequences of representations are required for the recovery of 3D information from 2D images (see Figure 4). There are three main stages: the primal sketch; 2½ D sketch; and 3D model representation. The primal sketch uses light information from noise in the visual field (VF) and involves three stages: detection of zero-crossings; formation of the raw primal sketch; grouping and formation of higher-level constructs. The 2½D sketch uses information from stereopsis, optical flow, texture, shading, occluding contours, surface contours and motion parallax. The function of the 2½D sketch is to represent orientation and depth of surfaces and discontinuities from a specified view in a viewer-centred system. 3D model representation involves the conversion of the 2½D sketch to a 3D representation, this enables recognition from different viewpoints, as it allows one to mentally rotate from any viewpoint perceived to match a stored representation.

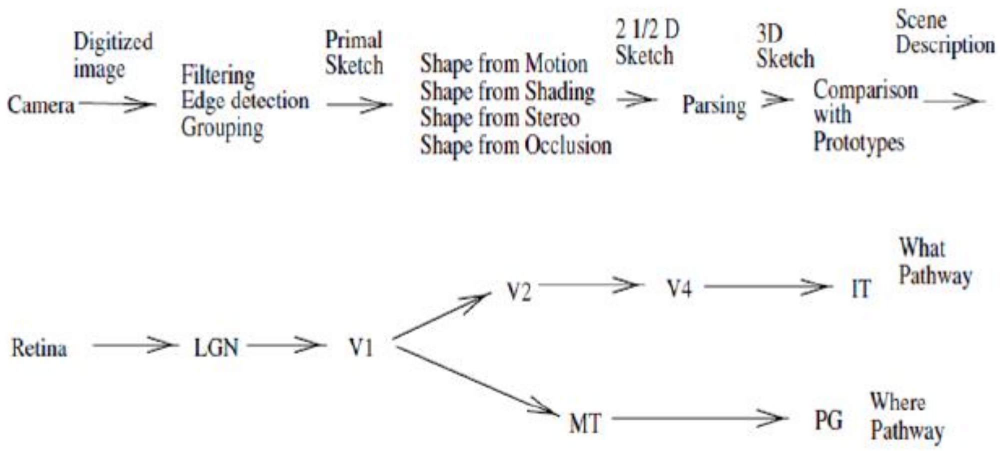
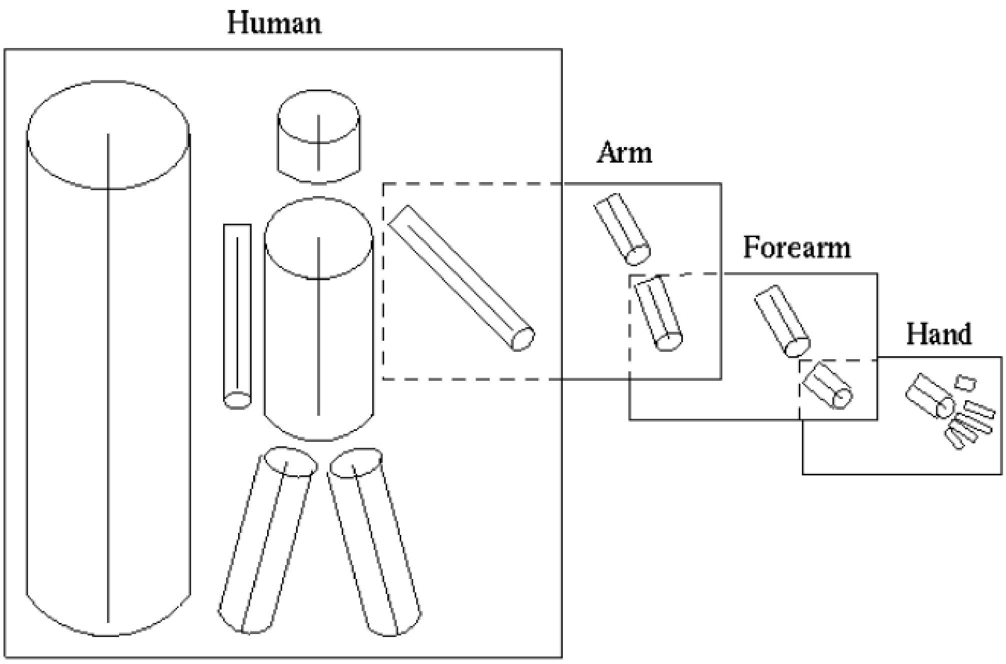


Figure 4. Marr's flow chart of visual processing, explained in hierarchical manner. Local edges and tokens are computed and grouped to infer surface orientations of objects. The 2.5D sketch is then parsed and matched against stored 3D prototypes. From Lee (2003).

Primitives for Marr and Nishihara carry information about local surface orientation and distance (relative to the viewer) at thousands of evenly spaced locations in the visual field. There are two aspects of a representation's primitives: the type of shape information it carries and its size. They define two classes of shape primitive: surface based (2D) and volumetric (3D). The simplest surface primitives provide information about size and location of small pieces of the surface. More complex surface primitives give information about orientation and depth. Volumetric primitives carry information about spatial distribution of a shape; the simplest ones provide information about a location and an extent (providing a corresponding spherical location in space); when a vector is added to the information about a spherical location in space, we have the length of the cylinder and diameter, and a further vector can provide information



about rotational orientation, then another to specify curvature in the axis.

Figure 5. Marr's proposed hierarchical arrangement of the 3D model representation. From Marr (1982).

1.2.1.1.3 Surface based model

The surface based model was outlined in Leek, Reppa and Arguin (2005). They proposed that surfaces are the primitives used for shape representation; this is done using 2D edge-bounded polygons that are used to approximate the shapes of surfaces in 3D objects. Surfaces belonging to objects are approximated by bounded 2D regions defined by discontinuities including luminance, chrominance, texture and retinal display (see Figure 6).

Spatial configuration of object surfaces is described by a surface configuration map. This is for both visual stimulus input and for stored representations of objects. During perception, there is a 2D configuration map computed for surfaces, which is viewpoint specific. The stored surface configuration maps for known objects contain spatial adjacency for all known surfaces of objects, and may be considered as 3D model representations. The surfaces and 2D configuration map from the visual input is then matched to the stored 3D configuration map to determine what the object is.

Leek et al. (2005) used a whole part matching paradigm to investigate if volumetric components have some special status. They found that there was an advantage for volumetric components but also for surfaces – these findings further motivate a surface-based model. Reppa, Greville and Leek (2015) also provided evidence for a surface-based model of high-level shape representation using a whole-part matching task; matching surfaces, volumetric parts or closed-contour fragments to whole objects. There was a performance cost in matching volumetric parts to wholes when the volumes showed surfaces that were occluded in the whole object. This was found for both same and different viewpoints, and regardless of target-distracter similarity.

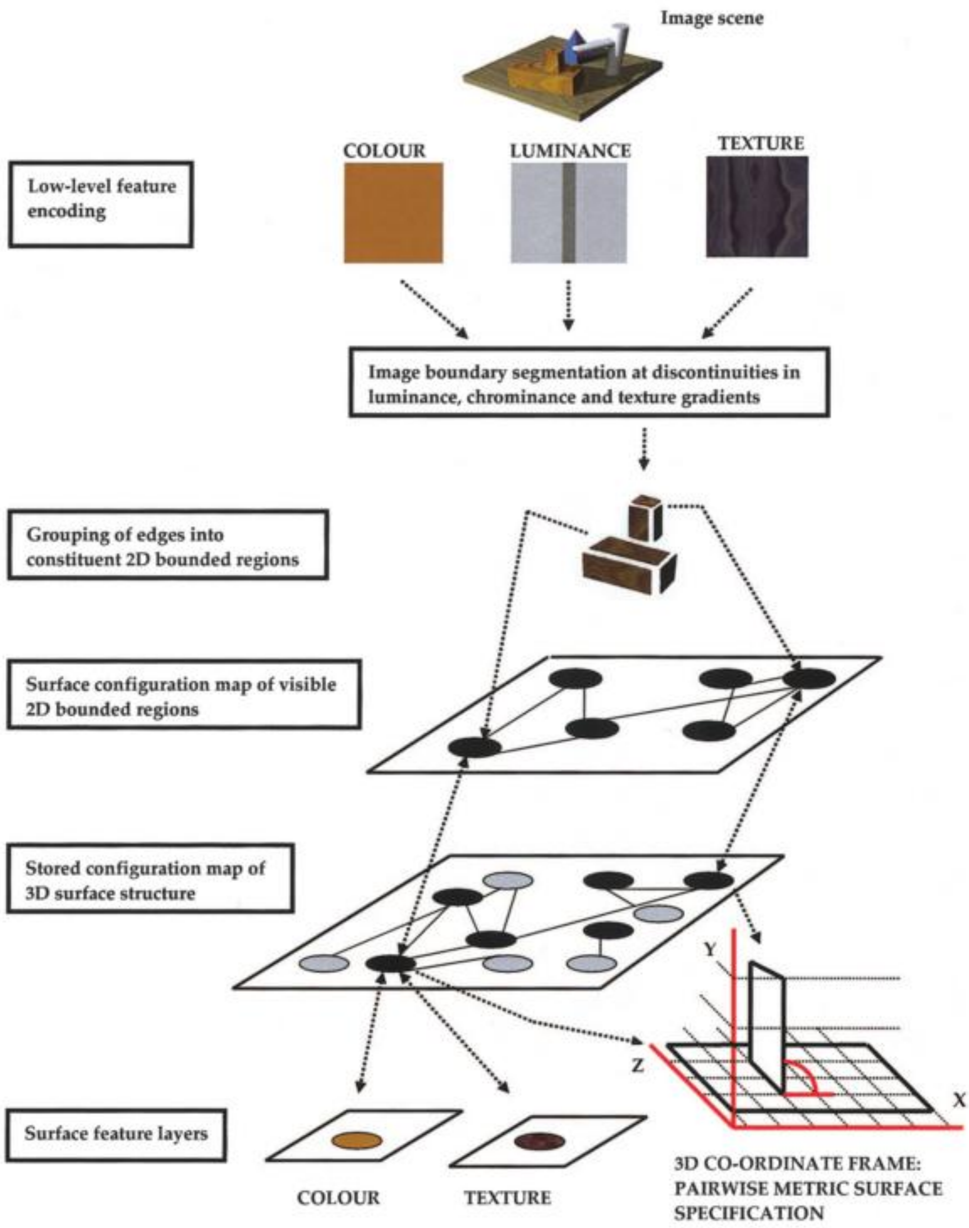


Figure 6. A schematic outline of the surface-based representations hypothesis. Two-dimensional (2D) edge-bound polygons are used to approximate the shapes of object surfaces. The spatial configuration of visible surfaces in the stimulus (black circles) is encoded in a perceptual surface configuration map. Individual surface attributes (e.g. colour and texture) are encoded separately in feature layers linked to each surface (shown only for stored shape representations). A three-dimensional (3D) surface configuration map is used to encode the configuration of all known surfaces (grey circles represent known but currently occluded surfaces). Metric surface attributes of shape and orientation are specified for each pair of surfaces using a surface-centred 3D coordinate reference frame. From Leek, Reppa and Arguin (2005).

Another object recognition model that highlights the importance of surfaces was outlined by Hummel (2001): a revised version of the RBC model; JIM.3. Like Leek at al.'s (2005) surface-based model, this model also contains a level of surface representation. On this account, surface structure derives from grouping of edges bounding each surface, and this level of representation in turn, outputs activation both to a map of surface attributes and a geon-based object shape model.

Evidence from integrative agnosia provides support for structural description models, those with integrative agnosia are unable to integrate objects' parts and spatial relations, therefore providing evidence that parts and their spatial relations are separate. More specifically, individuals with damage to stored shape representation or access to them can be sensitive to surface properties of objects. (eg. Chainey & Humphreys, 2001; Humphrey, Goodale, Jakobson & Servos, 1994; Servos, Goodale & Humphrey, 1993). Patients tended to make fewer errors with real objects and photographs than with line drawings (eg. Chainey & Humphreys, 2001; Davidoff & Wilson, 1985; Farah, 1990), this may be due to the greater amount of information about surface structure in real objects and photographs relative to line drawings, this provides some support for a surface-based model for shape representation.

1.2.1.2 Image-based accounts

Image-based accounts of object recognition claim that objects are represented as vectors or features and/or feature coordinates. All aspects of object shape are perceptually integral to one another and with the viewpoint in which the object is depicted (Cichy, Khosla, Pantazis, Torralba & Oliva, 2016; Edelman & Intrator, 2001; 2003; Khaligh-Razavi & Kriegeskorte, 2014; Krizhevsky, Sutskever & Hinon, 2012; Olshausen, 1993; Poggio & Edelman, 1990; Reisenhuber & Poggio, 1999; 2002; Ullman & Basri, 1991). The goal of image-based accounts was originally to understand how neural networks could accomplish pattern recognition.

1.2.1.2.1 HMAX Model

Reisenhuber and Poggio (1999) proposed the HMAX model, a biologically inspired account of object recognition, emulating the feedforward architecture of stages of object recognition in the cortex. As it is based on the organisation of the visual cortex, it includes parallel and gradual increases of feature complexity and receptive field size. Initially many cells are required for simple features, and then in higher areas, neurons are tuned to a large number of complex features and show invariance to scale and position (see Figure 7).

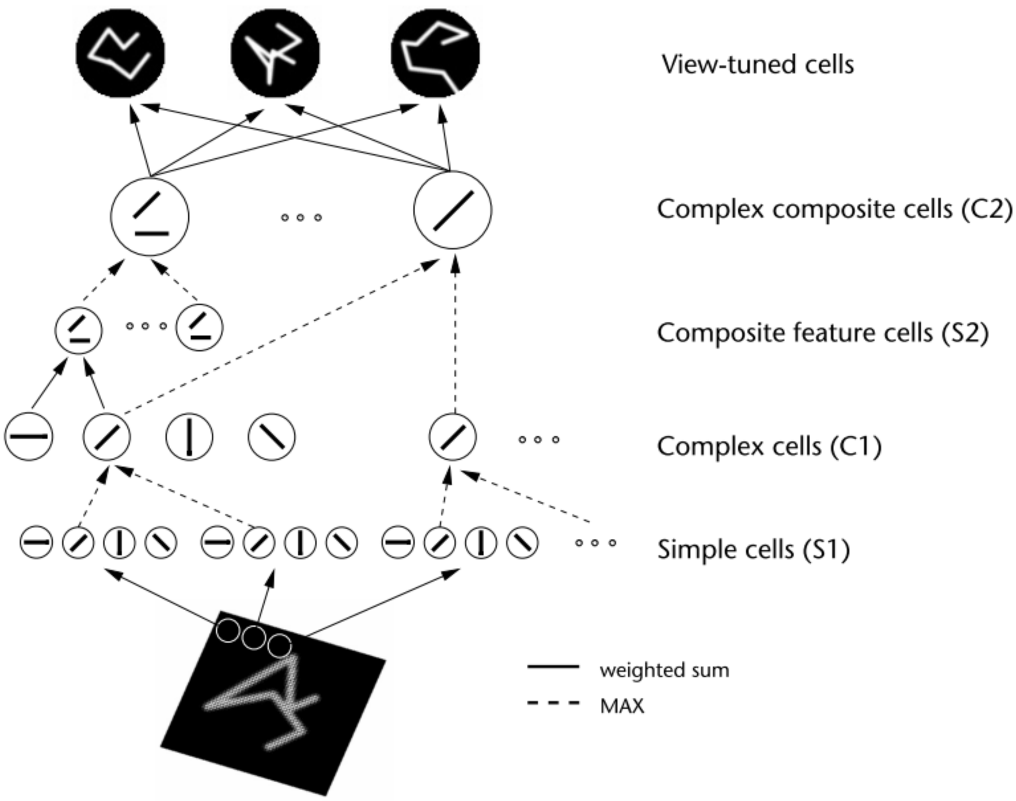


Figure 7. A diagram of the HMAX model, consisting of hierarchical layers. The first layer (S1) shows four different orientations (0°, 45°, 90° and 135°) and consists of simple Gabor filters at several spatial scales. The second layer (C1) pools the filter outputs spatially and across nearby scales. The third layer (S2) is tuned to a combination of orientation, and the fourth later (C2) provides further spatial and scale invariance. The C2 outputs are directly fed to a classifier. From Reisenhuber and Poggio (1999).

The original HMAX model had 4 layers: S1, S2, C1 and C2; where S and C stand for simple and complex units. The first input layer has filters for different orientations and areas of the visual field. C1 units accumulate responses by 'max' pooling operations over the S1s (akin to averaging orientations). The complexity of S input and scale invariance increases progressing through the hierarchical layers. C2 spatially pools output from S2 units and provides spatial invariance. The alternating architecture of S and C combining simpler low-level features into more complex ones gives increased feature detection specificity and enhanced invariance.

The HMAX model was improved by the addition of feature-learning stages (Serre, Wolf, Bileschi, Riesenhuber & Poggio, 2007). A learning module like this assumes that each unit measures similarity between a given stored view and a given input image. The output of all units are added and if they are above a threshold (if they are similar enough to the representation) the output is '1', if they are not the output is '0'. Therefore, over time, weights and threshold adjustments optimise classification of exemplars of objects and on this view, objects are recognised based on interpolation between small numbers of stored views.

An issue with image-based accounts concerns generalisability across objects and the issue of 3D representation. For example, unlike Marr (and arguably Biederman's volumetric geons), image-based accounts attribute no functional significance to 3D object structure, or the recovery of 3D object representations and the potential role of stereo input.

1.2.1.3 Neurological models of object recognition/processing

Object recognition is usually described as a hierarchical, feedforward, and bottom-up process. However, Bar (2003) proposed an interactive model of visual object recognition that highlighted the importance of top-down facilitation during processing as well as feedforward processing. Bar (2003) proposed that low spatial frequency (LSF) information facilitates object recognition by initiating a top-down process projected from the orbitofrontal cortex (OFC) to the visual cortex after a partial analysis of an image, with recurrent processing until a representation is matched. The rationale is that LSF content is coarser than HSF, therefore is used first to minimise the number of potential object matches and feeds information to the fusiform gyrus, meanwhile HSF information is processed along the ventral stream (occipital cortex) to the fusiform gyrus (FG) to perform the finer analyses.

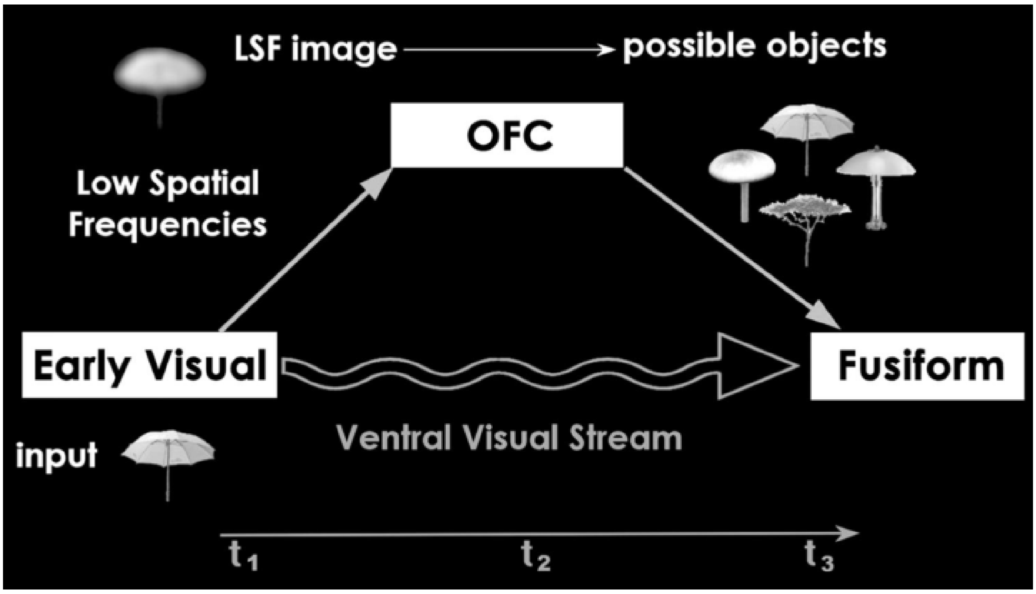


Figure 8. An illustration of the proposed model. A low spatial frequency (LSF) representation of the input image is projected rapidly, possibly via the dorsal magnocellular pathway, from early visual cortex to the OFC, in parallel to the systematic and relatively slower propagation of information along the ventral visual pathway. This coarse representation is sufficient for activating a minimal set of the most probable interpretations of the input, which are then integrated with the bottom-up stream of analysis to facilitate recognition. From Bar et al. (2006).

Providing support for this theory, Bar et al. (2006) used MEG and fMRI to investigate activity in early visual areas, the OFC and the FG during object recognition. The hypothesis was that there would be some differential response to low and high spatial frequencies in the OFC earlier than object recognition. They found that this was the case: there was stronger activation for LSFs in the OFC earlier than in the temporal cortex. They concluded that this activity is evidence for early feed-forward projection of LSFs which precedes the synchronised activity between the OFC and the FG, which is presumed to relay candidate interpretations based on the LSF content.

To investigate which anatomical pathways trigger top-down facilitation for object recognition, Kverega, Boshyan and Bar (2007) used dynamic causal modelling (DCM) with fMRI. They proposed that the rapid activation of the OFC is triggered by a magnocellular projection to the OFC which generates “initial guesses” based on magnocellular information. The magnocellular pathway conveys low-resolution, achromatic information, whereas the parvocellular pathway and koniocellular pathways conduct information more slowly and resolve finer details and chromatic contrasts, but require much higher luminance contrasts to detect achromatic stimuli. They used well-known objects that were HSF or LSF. The reasoning was that LSF images should be predominantly carried by the magnocellular pathway, whereas HSF would be predominantly carried by the parvocellular pathway. They found that LSF (magnocellular-biased) stimuli were recognised faster than HSF (parvocellular-biased) stimuli and activated OFC more than HSF stimuli, whereas HSF stimuli activated the ventro-temporal object recognition regions to a greater extent than did LSF stimuli. Also, for LSF stimuli, activity in the OFC was correlated with a recognition speed advantage, whereas larger BOLD response in the fusiform cortex was associated with an increase in recognition RT for HSF stimuli. Finally, LSF stimuli increased conduction of information from the middle occipital gyrus (MOG) to OFC and from OFC to FG, whereas HSF stimuli increased conduction of information from MOG to FG.

There is a wealth of evidence that shape perception and object recognition occur very early (Thorpe, Fize & Marlot, 1996). Presumably, this is via bottom-up processing and is suggested to occur within 130-215ms. Recent findings from Bar et al. (2006) indicate that we need to incorporate top-down feedback into models of object recognition. Bottom-up models cannot fully explain the visual constancy of human recognition (Serre et al., 2007). Instead, it is suggested that the bottom-up pathway provides the initial input and hypothesis to test using top-down processes (Serre et al., 2007b). Top-down processes are even more important when the stimulus is impoverished or ambiguous.

1.2.1.4 Overview of object recognition theories

These different theoretical models make contrasting claims about the functional architecture of object recognition and about the nature of the shape information that mediates object recognition. Two key differences among these hypotheses that are examined in this thesis concern: (1) how the perceptual processing of 3D shape may involve differential processing, and subsequent integration, of shape information across spatial scales and (2) the potential contribution of stereo input to the perceptual derivation of 3D object shape. We consider each of those two issues in turn.

1.2.2 Shape information at different spatial scales

Shape information can be described at coarse to fine spatial scales. Global features are those that can be detected at a coarse spatial scale, such as edge collinearity, elongation, symmetry, aspect ratio and global outline. Whereas, local object features are computed at a finer spatial scale, for example, edge boundaries, corners, surface depth, vertices, curvature, colour and texture. Some theoretical models support the distinction between local and global shape information, suggesting that mental representations of complex 3D objects comprise both local higher-order parts, and the global spatial configuration of these parts. (e.g., Biederman, 1987; Marr & Nishihara, 1978). Image-based theories, however, do not make a distinction between shape information at different spatial scales.

1.2.3 The role of stereo information in object recognition

Stereo information might facilitate perceptual analyses of 3D object shape by providing cues to properties like local surface slant, global depth orientation and 3D shape configuration. Some current theories of shape perception attribute little, if any, significance to stereo information (e.g., Bülthoff & Edelman, 1992; Chan, Stevenson, Li & Pizlo, 2006; Li, Pizlo & Steinman, 2009; Riesenhuber & Poggio, 1999; Serre, Oliva & Poggio, 2007). For example, in the HMAX model (Riesenhuber & Poggio, 1999), recognition is accomplished using a multi-layer system in which hierarchically structured representations of object shape are computed from monocular image contour. In other work, Pizlo (2008; see also Li et al., 2009; Pizlo et al., 2010) has shown that veridical 3D object shape can be computed from 2D shape information when the derivation is constrained by simplicity constraints (symmetry, compactness, planarity and minimum surface area). Elsewhere, the contribution of stereo is neither ruled out nor explicitly incorporated into the proposed theoretical framework (e.g., Biederman, 1987; Leek, Reppa & Arguin, 2005; Ullman, 2006). This contrasts with some earlier theoretical models of object recognition that have proposed an important role for stereo input – such as in the recovery of local surface depth in the $2^1/_2$D sketch of Marr (1982; see also Marr & Nishihara, 1978).

1.2.3.1 Three dimensional (3D) shape recovery model

Pizlo (2007; 2010) outlined a computational account of 3D shape perception, based on recovery of 3D shape from a limited number of a priori constraints in 2D images. These equate to volume (maximal 3D compactness), surface (minimal surface area) and contours (maximal 3D symmetry and maximum planarity of contours); they do not use stereo visual input, depth, surfaces or learning in their model. These components form a "combinatorial map", this describes how the space containing the shape is partitioned into: volumes; surfaces that bound those volumes; and contours that bound those surfaces. A geometric interpretation is created from these partitions by adding information about planarity of surfaces, straightness of contours, and equality of line segments and positions of endpoints of line segments.

Li, Pizlo and Steinman (2009) provided evidence for the success of this computational model using abstract symmetrical polyhedral objects. They made additions to their computational model to allow for recognition of natural objects and natural man-made objects. In this amended computational model, 2D contours, rather than 2D points were used as the input to the model. They found that symmetry could be used to recover both symmetrical and nearly symmetrical objects. This is significant as it allows the recovery of 3D shapes such as the human body. They suggest that a single 2D retinal image is not sufficient to produce a representation of 3D shape, and could have been produced by infinitely many 3D shapes. They suggest that the way the possible 3D shapes is reduced is by using a priori simplicity constraints on the possible 3D interpretations. The visual system can only do this if it knows something about the nature of 3D shapes and Pizlo et al. propose that characteristics such as 3D symmetry, volume and a combinatorial map could have been learned through interaction with the environment, but suggest that a more likely proposal is that these things are innate, acquired during evolution.

1.3 Local and global processing of object shape

The availability of different types of shape information is dependent on spatial scale. Some information can be detected at a relatively coarse spatial scale, such as edge co-linearity, elongation, symmetry, aspect ratio, and global outline, these can be described as global features. Other useful shape information can be computed from finer spatial scale such as edge boundaries, corners, surface depth, vertices and curvature, and colour and texture, these are local features. It has been suggested that local and global features of objects are processed in different ways (Bullier, 2001; Hegde, 2008; Heinz, Johannes, Münte & Magun, 1994; Heinz, Hinrichs, Scholz, Burchert & Mangun, 1998; Lamb & Yund, 1993; Peyrin, Michel, Schwartz, Thut, Seghier et al., 2010; Peyrin, Baciu, Segebarth & Marendaz, 2004; Peyrin, Chauvin, Chokron & Marendaz, 2003; Robertson & Lamb 1991; Schyns & Oliva, 1994). The terms local, global, coarse, fine and spatial scale are all poorly defined and inconsistent in the literature. Shape information appears to be acquired at multiple spatial scales during perception and spatial scale can be thought of in at least two ways: the size of the sampling window, from narrow (focussed) to broad (diffuse); and as high and low spatial frequencies. There could be a broad sampling window, with high or low spatial frequency information and the same for a narrow sampling window. Presumably there is a trade-off: the broader the sampling window, the more information, therefore it is more economical to use low spatial frequency information. And for a narrow sampling window, there is more high spatial frequency information, but not exclusively, so sampling is not binary, but more likely a continuum of sampling at different spatial scales. The terms 'fine' and 'coarse' can be used to describe the granularity of information at high and low spatial frequencies, whereas local and global defines more the size of the sampling window (regions of the visual field or object it is extracted from). Therefore, care should be taken when directly comparing studies defining local and global in terms of spatial frequency and those using definitions involving sampling windows due to the inevitability of both low and high spatial frequencies in broad and narrow sampling windows.

1.3.1 Global Precedence Effect

David Navon was the first to investigate local and global processing with hierarchical stimuli (Navon, 1977), now known as Navon letters (see Figure 9), he proposed that local and global information is processed differently. After a series of experiments, Navon (1977) observed two phenomena related to processing the local and global elements of stimuli: global information is processed faster than local information; and when the local and global elements of a stimulus are congruent, people can voluntarily attend to the global pattern without being affected by the local features, but cannot process the local features without attending to the global pattern (global interference), he called this the global precedence effect (GPE).

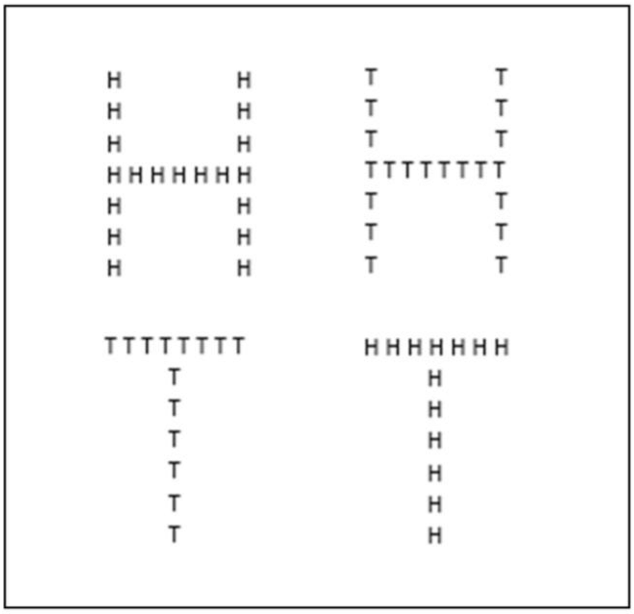


Figure 9. An example set of hierarchical letter stimuli (Navon letters) made up of small letters (local level), either congruent or incongruent to the larger letter (global level).

Navon (1977) proposed that perception proceeds from a global analysis to a more local analysis, based on findings of global precedence. He suggested that while we can attend to just the global level of a scene, we are not able to ignore the global level when processing local elements and as we cannot skip global processing, it seems that this is a necessary stage of perception. Many studies have provided support for GPE (e.g. Beaucousin, Simon, Cassotti, Pineau, Houde & Poirel, 2013; Han, He & Woods, 2000; Proverbio, Minniti & Zani, 1998; Yamaguchi, Yamagata & Kobayashi, 2000). Poirel, Pineau and Mellet (2008) found global precedence regardless of the meaningfulness of the stimuli; the global level was always processed faster than the local level. However, the interference effect occurred only for meaningful stimuli; objects rather than non-objects. They suggest that global precedence is a sensory mechanism and therefore automatic for all stimuli, whereas global interference reflects a cognitive mechanism, which is related to identification. Hence, the global level of the stimuli will always be processed faster, but there may not always be interference from the global level when processing local information.

1.3.2 Hemispheric effects

Hemispheric differences are often found in processing of hierarchical stimuli, with greater activity in the left hemisphere for local processing and greater in the right for global processing. (Delis, Robertson & Efron, 1986; Lamb, Robertson & Knight, 1989; 1990; Robertson & Lamb, 1991; Robertson, Lamb & Knight, 1988; Van Kleek, 1989). There is frequent report of hemispheric asymmetry in EEG studies for local and global processing (Heinz et al., 1998; Leek, Roberts, Oliver, Cristino & Pegna, 2016; Mangun, Heinz, Scholz & Hinrichs, 2000; Volberg & Hübner, 2004; Yamaguchi et al., 2000), and also from imaging studies (Han, Weaver, Murray, Kang, Yund & Woods, 2002). A similar pattern has been reported in patient populations, where, during copy-drawing, patients with left hemisphere damage are impaired in reporting the local level of Navon-type letters, with relatively intact report of the global level, and vice versa for those with right hemisphere damage (Robertson, Lamb & Knight, 1988; Robertson & Lamb, 1991).

These hemispheric asymmetries appear to occur more often with centrally presented stimuli than with laterally presented ones in ERP and behavioural studies, Volberg and Hübner (2004) investigated if presentation or response conflict between levels was more important in finding hemispheric asymmetries in an ERP study. They observed hemispheric differences that were more pronounced for incongruent stimuli (conflicting), suggesting that conflicting stimuli are more important in observing hemispheric differences than central presentation of stimuli. Volberg and Hübner (2008) conducted a further study to disentangle possible effects of task difficulty and hemispheric differences in local and global processing; they found that increased difficulty in the incongruent condition of the task did not account for hemispheric asymmetries.

Studies using bilateral stimulus presentation have found that reaction times are faster for stimuli presented in the left than the right visual field for global targets (Flevaris, Bentin & Robertson, 2010; Schlösser, Hübner & Studer, 2009; Van Kleeck, 1989; Volberg & Hübner, 2006). This suggests that global shape information is processed faster when projected to the right cerebral hemisphere. Furthermore, stimuli presented in the left side of space provide a benefit for global processing and the same for the right side of space and local processing (Christie et al., 2012; Yovel, 2001) and the same was true for the left and right sides of objects (Christie et al., 2012).

Using Navon-type stimuli and an RT paradigm, global precedence was assessed in patients with unilateral focal lesions in posterior regions and compared with neurotypical participants. When patients had damage in the right hemisphere, they had impairments in global processing compared to controls and when damage was in the left hemisphere local processing was impaired (Delis, Robertson & Efron, 1986; Lamb, Robertson & Knight, 1990; Lamb et al., 1989; Robertson et al., 1988). Bultitude and Woods (2010) also provide evidence for dominance of the right hemisphere for global processing and left for local using prism adaptation studies using Navon letters.

One explanation for the apparent hemispheric asymmetry for processing local and global elements in Navon letters is that the letter stimuli provide a left hemisphere advantage due to the linguistic information, however, the hemispheric differences have also been found in non-letter hierarchical stimuli (e.g. Han, He, Yund and Woods, 2001). There are, nonetheless discrepancies in results, which may stem from the type of stimuli used (Fink, Marshall, Halligan, Frith, Frackoiak & Dolan, 1997). It has been reported that there is a reversal of hemispheric asymmetry when the stimuli change from letters to meaningful objects, the authors speculate that this may be due to left hemisphere bias for language, whereas object recognition may be driven by the right hemisphere. Others find no hemispheric asymmetries when displaying images unilaterally (e.g. Han, Yund & Woods, 2003).

It may be the case that hemispheric effects observed for Navon-type tasks reflect an integration process, and not just differences between processing local and global elements of stimuli (Flevaris, Bentin & Robertson, 2010; Hübner & Volberg, 2005; Martens & Hübner, 2013); integration of level and identity. The rationale is that first, letter information is available without level information (global or local), and the hemispheres will not differ in processing here. However, when stimuli are incongruent (at the global and local levels) the whole stimuli have to be taken into account, requiring the integration of letter and level information. Therefore, hemispheric asymmetries will only be observed when there are incongruent hierarchical stimuli; when binding of content and the level is required. Hübner and Volberg (2005) conducted a study to determine whether letters and levels were processed in combination. They suggested that letters and levels are coded separately at early stages of visual processing and are integrated at a later stage as conjunction errors indicated that the letters and levels remained unbound at this early stage (very short masked presentation). In a further experiment, they found that refocusing attention from the local to global levels was more difficult than vice versa. They concluded that there were no hemispheric differences between local and global performance for letter identification (using congruent stimuli) and so suggest that differences stem from integration of levels and letters.

Martens and Hübner (2013) used electrophysiological measures to test the hypothesis that asymmetric differences in local/global studies reflect a binding process, not early perceptual processing, testing the assumptions of Hübner and Volberg (2005). They assumed that there should be no differences in the left and right hemispheres for congruent stimuli (as no binding is required), but should see an effect for incongruent stimuli. They expected decreased cortical activity due to repetition of stimuli processed by particular hemisphere (right for global, left for local). As expected, they found that local targets tended to elicit smaller responses in the left hemisphere than global stimuli, and global targets elicited smaller responses in the right hemisphere than did local. These effects were found only for incongruent stimuli, therefore assumed to be related to the binding of level and letter.

Flevaris, Bentin and Robertson (2010) investigated binding of local and global information in relation to relative spatial frequencies (SFs). The aim was to examine whether the medium of hierarchical binding is attentional selection of task-relevant SFs and if hemispheric asymmetries in hierarchical binding are mediated by attentional selection of relatively high versus low SFs. They found that attentional selection of relatively low SFs reduced hemispheric asymmetry for global conjunction errors and facilitated binding of letters to the global level in the RH and vice versa.

An issue to consider in the processing of global and local elements of an object concerns the reference frame of the hierarchical objects; this may be dependent on the scale of the object and where attention is focussed. In a study displaying hierarchical objects either in the left or right hemifield, local information was more easily detected when on the right side of objects and global information was easier to detect on the left side of objects (Christie et al., 2012). This extends the findings of hemispheric specialisation, but within hemifields, possibly implicating an object-centred reference frame, this is supported by neurologically damaged patients (Kleinman et al., 2007), where neglect following left hemisphere injury tended to be allocentric rather than egocentric – with patients neglecting the right sides of objects rather than the right side of space.

Hemispheric asymmetry may also differ for the preparatory responses to the cues (which level to attend to) and the stimuli themselves. Yamaguchi, Yamagata and Kobayashi (2000), in an ERP study, found asymmetric hemispheric activity to local and global processing at both the target presentation and after the stimulus cue (arrows indicating whether they should attend to the local element of the stimulus or the global level), 240ms after cue stimulus. They suggest that this was evidence of top-down allocation of attention to global and local features and claimed that the early difference they found in local and global cueing indicates the shifting of attention to local and global. However, Volberg and Hübner (2007) found no evidence of asymmetric local and global responses to cues. They used arrows and colours to cue the local or global levels and found hemispheric differences in the arrow cue condition, whereas in the colour cue condition, there were no differences between left and right hemispheres. They suggest that the arrow cues had to be processed in a local/global manner and differences did not just reflect preparatory processes. They, therefore suggest that the findings of Yamaguchi et al. (2000) were due to the nature of the cueing stimulus: the local cue consisted of arrows pointing inwards, and the global cue consisted of arrows pointing outwards, both creating a virtual square, possible at both the local and global levels.

1.3.3 Spatial Frequency and Double Filtering by Frequency theory

The processing of local and global elements is seemingly interrelated with processing of high and low spatial frequencies (SFs), Broadbent (1977) was among the first to suggest that the functional distinction between processing of global and local levels of a stimulus was based on the spatial frequency content. Boeschoten, Kemner, Kenemans and Engeland (2005), in an ERP study, provided direct evidence that global and local levels are processed according to their SF content. They found that performance on local and global tasks was affected by removing high and low SFs from their hierarchical shape images: the removal of low SFs decreased the activity associated with processing of a global target. They concluded that processing of global information depended on the low SF content and processing of local information depended on its high SF content.

Hughes (1986) suggested that there is a relationship between global precedence and spatial frequencies, with a processing advantage for LSF channels. Shulman, Sullivan, Gish and Sakoda (1986) found that after adaptation to LSFs, global processing was enhanced, and after adaptation to HSFs, local processing was enhanced. Dale and Arnell (2014) replicated these findings and Shulman and Wilson (1987) found that attention to global elements of a stimulus facilitates LSF detection, but not detection of HSFs.

Hemispheric asymmetries are seemingly affected by the SF content of images: removing low SFs, comparable to global information, diminishes hemispheric effects (Han et al., 2002). As asymmetry is diminished by removing low SFs from the stimuli, this suggests some role for SF filtering in extrastriate cortical areas during the processing of hierarchical stimuli. The double filtering by frequency hypothesis (DFF), put forward by Ivry and Robertson (1998), aims to explain how global perception is associated with low SFs and processing bias of the right hemisphere and local perception is associated with high SFs and processing bias of the left hemisphere. DFF theory develops this by proposing that visual attention selects and is directed to relatively low SFs by the right hemisphere and relatively high SFs by the left hemisphere, suggesting a direct causal relationship between SF selection and global versus local perception. This theory also aims to explain the large literature on laterality of high and low SFs. Robertson and Ivry (2000) suggested that the hemispheres first perform symmetric (between both hemispheres) filtering of visual information, then the functional hemispheric asymmetry (that is reported in many studies) arises from a second filtering. The second filtering includes filters that are skewed in the two hemispheres: the right hemisphere operates on a low-pass filter and the left hemisphere operates on high-pass filter. This means that the representations each hemisphere receives are slightly different, therefore each hemisphere is more suited to different parts of a local and global task.

Flevaris, Bentin and Robertson (2011b) tested the DFF hypothesis by measuring EEG at posterior left and right sites whilst attention was directed to either local or global levels after selection of relatively high or low SFs in a previous stimulus. They found that attentional selection for SF modulates preparatory activity in global versus local perception. They found greater alpha reduction in the right hemisphere than left in preparation for global targets, they did not find this pattern for local and left. They concluded that hemispheric asymmetry in global versus local tasks occurs in a top-down fashion as well as bottom-up, as they found modulation in the preparatory processes before the hierarchical (Navon) stimuli were shown. The SFs modulate relative neural activity in the left and right hemispheres when involved with preparing for local and global. In a later study, they found that the left hemisphere preferentially bound the local level of stimuli, and the right hemisphere had a preference for binding to the global level. Also, binding is modulated by attentional selection of high and low spatial frequencies. Therefore, attention to high SFs facilitated binding to the local level in the left hemisphere and selection of low SFs facilitated binding to the global level in the right hemisphere (Flevaris et al., 2014).

In line with this, differential hemispheric effects have been observed for processing of high and low SFs using PET (Romei, Driver, Schyns & Thut, 2011; Smith, Gosselin & Schyns, 2006). Fink, Marshall, Halligan and Dolan (1999) conducted a PET study using hierarchical stimuli. They used non-representational figures (lines to give the impression of rectangles) with all orientations incongruent at the local and global levels; the task was to identify the orientation of the stimuli at either the global or local level. They found that there were hemispheric differences in early visual areas for processing of local and global levels of stimuli. However, they did not find that local and global information was processed in the left and right hemispheres only, it was dependent on processing level and stimulus characteristics.

However, others suggest that attention to local and global features is not lateralised, but is mapped in multiple visual cortical areas. Using fMRI, Sasaki, Hadjikhani, Fischl, Liu, Marret, Dale and Tootell (2001) found that attention to local features activated the foveal representation, where sensitivity to higher SFs was highest, whereas when attending to global features, there was increased sensitivity to lower SFs at more peripheral eccentricities. This implicates the position of the stimulus in the visual field interacting with SF in the apparent processing differences. Also, Han et al., (2003) found that early global processing, at the N190, was diminished when low spatial frequencies were removed, but this was not the case for the later N300, therefore low spatial frequencies seem to play an important role in early, but not later processing of global stimuli.

1.3.4 Failure to replicate GPE

There are several studies that fail to fully replicate GPE (e.g. Martens & Hübner, 2013; Roalf, Lowery & Turetsky, 2006). It may be that Navon letters, or similar hierarchical stimuli are problematic in that they are too artificial, and therefore the effects (GPE) may be paradigmatic. This is evidenced by the widely varying findings from studies using slightly different stimuli and tasks. It seems that the emergence of global precedence depends on a number of factors; things that can disrupt or diminish GPE include: visual angle/size – GPE is diminished when visual angle exceeds 7-10 degrees (Kinchla & Wolfe, 1979; Lamb and Robertson, 1990), however this effect is modulated when eccentricity of both levels is equated (Amirkhiabami & Lovegrove, 1999; Navon & Norman, 1983); spatial certainty – GPE is less likely when participants know where the stimuli will appear (Lamb and Robertson, 1988); central rather than peripheral display (Grice, Canham & Boroughs, 1983; Pomerantz, 1983; Yund et al., 2002b), however, this is disputed (Luna, Merino & Marcos-Ruiz, 1990; Navon & Norman, 1983); spacing of elements – GPE is less likely to occur with sparse than dense elements (Kimchi, 1985; Martin, 1979; Yovel, Yovel & Levi, 2001), this can be the number of elements or the distance between them; exposure duration – GPE is more likely to occur with a short exposure duration (Luna, 1993; Paquet & Merikle, 1984); and differences have been reported between males and females, with local advantage for females, but not males (Roalf et al., 2006). Furthermore, GPE is consistently found to be eliminated by contrast balancing; removing most of the low spatial frequency content in images (e.g. Boeschoten, Kemner, Kenemans & Engeland, 2005; Han, Weaver, Murray, Kang, Yund & Woods, 2002; Han et al., 2003; Hughes, Fendrich & Reuter-Lorenz, 1990; Jiang & Han, 2005).

As complex scenes do not contain simply one global and one local element, Krakowski (2015) introduced an intermediate level in their hierarchical stimuli. They found that size of the level was not the issue, as global and intermediate levels were processed in the same way. They controlled for the size of elements, which seems to rule out the possibility that global elements are processed first, simply because they are larger than local, or intermediate.

Differences in findings and conclusions in the literature may stem from confusion in defining terms. One is the difference between coarse-to-fine and global-to-local. A second possible source of differences is the definition of spatial scale. Spatial frequency is an often used expression of spatial scale and is easily quantified, however in studies such as that of Sanocki (1993), spatial scale was defined using terms 'large-scale' and 'small-scale'. Large scale meant large overall shapes of objects and small-scale meant smaller target details. A definition such as this cannot be quantified in the same way that spatial frequency can. This also creates difficulty in comparing results from studies such as this to those using spatial frequency. In the following empirical chapters, we use the terms local and global, or local spatial scales and global spatial scales to mean information at narrow and broad sampling windows, respectively. This is because information at local and global spatial scales comprise both low and high spatial frequencies. In a broader (global) sampling window, there is more information, therefore it is more economical to use low spatial frequency information. And for a narrow (local) sampling window, there is more high spatial frequency information, but not exclusively, so sampling is not binary, but more likely a continuum of sampling at different spatial scales.

1.4 Temporal Dynamics of Object Recognition

Our primary focus in this thesis is on elucidating how shape information may be processed, and integrated, across local and global spatial scales during the perception and recognition of 3D object shape, and about the role of stereo visual input in these processes. These two issues are inherently linked to the temporal dynamics of 3D shape processing.

Most models of object recognition focus on local to global processing, for example, the HMAX model (Riesenhuber & Poggio, 1999) proposes processing of edges first, then a gradual move up the hierarchy to whole object processing. Similarly, structural descriptions accounts tend to include detection of edges or surfaces as the first steps in object recognition (Biederman, 1987; Leek et al., 2005; Marr & Nishihara, 1978). Seemingly, these models are not compatible with the global precedence effect (GPE) literature, which suggests that we process global information first, perhaps when not even attending to it. GPE could, however, be described to be consistent with a coarse-to-fine model of visual recognition, with the global information conveyed by rapid magnocellular visual channels, allowing for rapid initial perceptual analysis of visual inputs. This early analysis allows for guidance of the subsequent analysis of local information conveyed by slow parvocellular visual channels, this could also be explained by early global information sent to the orbitofrontal cortex (OFC) and the high SF information takes a slower route along the ventral stream (Bar, 2003; Bar et al., 2006).

The GPE often observed using hierarchical stimuli could be interpreted as representing global-to-local processing, where global image features take precedence over local ones. However, it could be the case that it takes longer to deploy attention to local features, like a 'zooming of attention'. The latter explanation implies that the GPE may just be an issue of the size of elements; global elements are larger than local ones in Navon letters. It also may be an issue of processing time, Reynolds (1981), using illusory triangles made up of Pacmen, found that when a mask immediately followed Pacmen (with a SOA of 50ms), the majority of participants saw the local Pacmen but not the global triangle, however when the SOA was 100-125ms, all participants saw the triangle. Therefore, it seems that when there is more processing time, more information is available, but not necessarily global first.

1.4.1 Coarse-to-fine processing in scene perception

Scene classification appears to be an extremely rapid process, based on coarse analyses (e.g., Bullier, 2001; Hegdé, 2008; Heinz, Johannes, Münte & Mangun, 1994; Heinz, Hinrichs, Scholz, Burchert & Mangun, 1998; Lamb & Yund, 1993; Peyrin et al., 2010; Peyrin, Chauvin, Chokron & Marendaz, 2003; Schyns & Oliva, 1994). In scene classification, where the task was to identify the presence of an animal in a complex scene, categorisation occurred around 150ms after stimulus onset (Thorpe et al., 1996). This indicates that complex natural scenes can be processed (at least the gist) ultra-rapidly through feed-forward processing, as presumably 150ms would not leave sufficient time for feedback processing. Subsequent studies have provided confirmatory evidence of this (Johnson and Olshausen, 2003; 2005; Fabre-Thorpe, 2011; Rousselet, Thorpe & Fabre-Thorpe, 2004).

Although scene classification can occur rapidly, it may be the case that more detailed parts of a scene or an object require more processing time, including the identity of an object in a scene (Johnson & Olshausen, 2003). One problem with a fast, feed-forward processing hypothesis is that natural scenes are more complex than the photographs shown in experiments: factors such as occlusion, variations in lighting, shadows etc. make it more difficult to segment objects in a scene than is apparent in experiments. Studies of scene perception, using simulated occlusion, have found that top-down factors such as prior knowledge of the objects that are occluded are needed to determine the identity of the occluded object. (Bar, 2004; Johnson and Olshausen, 2005).

The Reverse hierarchy theory (Hochstein & Ahissar, 2002; Ahisssar & Hochstein, 2004) takes increased processing time for fine-grained information into account, suggesting that rapid categorisation, or 'vision at a glance' is mediated by high-level areas, and more detailed analysis, or 'vision with scrutiny' involves progressively lower visual areas, containing more fine-grained information. 'Vision at a glance' involves a feedforward sweep up the visual hierarchy (Felleman & Van Essen, 1991). Then, 'vision with scrutiny' is mediated by feedback processing, in the reverse direction, down the hierarchy. Other coarse-to-fine processing models incorporate spatial frequency, postulating that the visual system first extracts low SF-based gist, then followed by more fine-grained, higher SF information. The coarse information is used to constrain the interpretation of the information at a finer spatial scale (Bar, 2003; 2004; Peyrin et al., 2005).

1.4.2 Models of Local and Global Processing

There is a wealth of research on hemispheric differences in global and local processing (see Chapter 1.3.2) and these differences are related to the question of parallel or serial processing. If the two hemispheres are preferentially processing local or global information, the processes could be occurring in parallel. Results showing that global information interferes with the processing of local information have been interpreted as indicating that hierarchical levels of information are processed in either a top-down or bottom-up fashion, but other authors suggest that these levels of processing may be evaluated independently and in parallel (Lamb and Robertson, 1988, Lamb, Robertson and Knight, 1989 and Robertson, Lamb and Knight, 1988). It may be that information is processed serially, or in parallel, or a combination of both (see Figure 10).

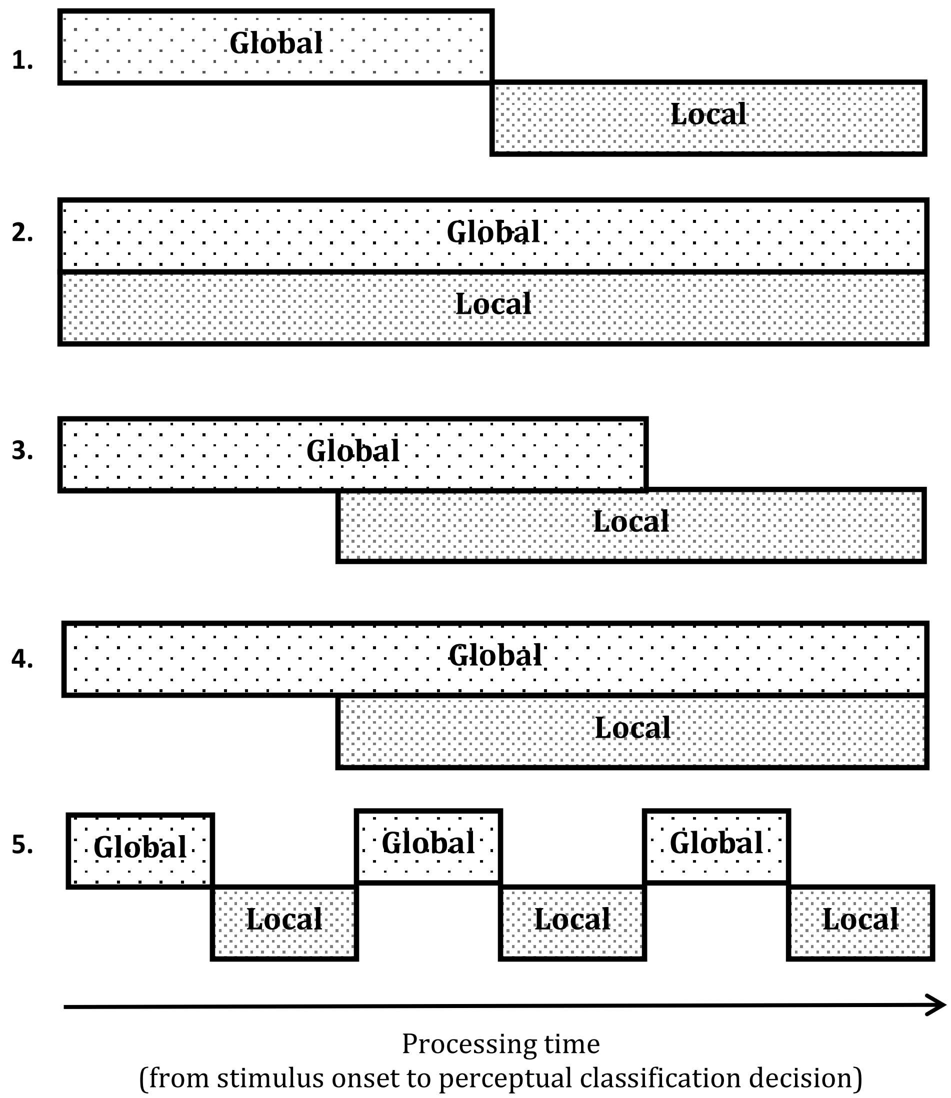


Figure 10. Hypothetical models illustrating processing of local and global shape information from stimulus onset to a perceptual classification decision: the hypothetical point in time where the system has accumulated enough information from global and local channels to make a decision. Model 1 illustrates serial processing at local and global scales, global processing occurs first, then when global processing is compete, local processing occurs. Model 2 illustrates fully parallel processing of global and local information. Model 3 illustrates a temporally overlapping, partially parallel process. Model 4 illustrates another partially parallel process, where global processing starts first, but both local and global processing continue at the same rate. Model 5 illustrates a process whereby global information is processed first, but then local information is sampled; global and local information are not processed in parallel, but are sampled one at a time, intermittently.

GPE is not the only explanation of the pattern of results observed in studies of global and local scene acquisition. Loftus and Harley (2004) divided theories into three categories: independence theories, stating that high and low SFs are acquired at the same time and combine additively (as in Figure 10, model 2); GPE theories, holding that global information is processed first, but global and local combine additively (as in Figure 10, models 1, 3 and 4); and interactive theories, stating that the global information affects local information acquisition rate (as in Figure 10, models 3, 4 and 5). This is another version of a global to local theory, but incorporating interference. Loftus and Harley (2004) tested these theories with a series of experiments. They used number strings of high or low spatial frequencies with masks of the differing spatial frequency to see if the SF information presented before the target influenced processing time. Their findings were inconsistent with independence theories, as LSF and HSF information was not acquired over the same time-course: for short durations, LSF performance exceeded that of HSF, whereas the opposite was true when there was longer processing time. Also, LSF and HSF information was differentially effective at different times after stimulus onset. Loftus and Harley (2004) concluded that global precedence theory explains a number of studies' findings coherently. Schyns and Oliva (1994) also provided data supporting global-to-local theories, using scenes with LSF or HSF information. They found that when presented for short durations, LSF images were perceived, whereas at longer durations, HSF scenes were perceived. They, therefore, concluded that LSF and HSF information was perceived at different time courses: LSF first, then HSF.

One example of a global-to-local theory is the contingency hypothesis, provided by Sanocki (1993). The theory posits that local information acquisition is contingent on global information acquisition – based on the assumption that the visual system cannot process all of the LSF and HSF information in the sensory input. Sanocki (1991; 1993; 2003) provided empirical support for this theory with priming studies, the results showed that a global prime enhanced discrimination of similar images with extra local details. They suggest that they provide evidence for an interactive theory, as the presence of a prime could only have facilitated acquisition of local information from the target.

In support for an (at least partly) parallel processing model of object recognition, Bar et al. (2006) proposed a model of visual object recognition where low spatial frequencies facilitate object recognition by initiating a top-down process projected form orbitofrontal to the visual cortex. They used MEG and fMRI to test this theory and found that there was a differential response to low and high SFs in the OFC earlier than object recognition, with stronger activation for low SFs in the OFC earlier than in the temporal cortex and at around 115ms. They concluded that this activity is evidence for early feed-forward projection of low SF information, which precedes the activity between the OFC and the fusiform gyrus (which they suggest represents feedback based on the low SF content). Low SF content is coarser, used first to minimise possibilities of what an object might be and feeds information to the fusiform gyrus, meanwhile high SF information is processed along the ventral stream (occipital cortex) to the fusiform gyrus to perform the finer analysis. On this view, processing of low SF information starts earlier than high SF information, but high SF information is processed in the same time-window, just through a ventral visual route.

It is possible, however, that visual processing order is flexible. Schyns and colleagues offer an alternative to the coarse-to-fine model (Morrison & Schyns, 2001 (for review); Gosselyn & Schyns, 2001; Oliva & Schyns, 1997; Schyns & Oliva, 1999). The alternative is based on the theory that the visual system does not operate as rigidly as often supposed: different spatial scales may be processed in different orders, dependent on the task and stimuli used: sometimes low spatial frequency is more useful than high spatial frequency and vice versa. If the order in which information from different spatial scales is, in fact, flexible, then experiments looking for the way the visual system uses spatial scale information may be futile. What we can do, is investigate what shape information, at particular spatial scales, is most relevant for a particular task, or classification context. Similarly, the question of the absolute speed of object recognition is also ill-posed, as it depends on the complexity of the stimuli, familiarity with the stimuli, the task and, potentially, a host of other factors. Instead, we can look for relative timing of different perceptual processes such as global and local processing, integration of information at different spatial scales and recognition indexing.

1.4.3 Electrophysiological components relating to global and local processing differences

In this section, we examine evidence from electrophysiology concerning the temporal dynamics of 3D object shape representation and recognition. Many event-related potential (ERP) studies have identified differential processing of local and global information, namely hemispheric differences (e.g. Heinz et al., 1998; Leek, Roberts, Oliver, Cristino & Pegna, 2016; Mangun, Heinz, Scholz & Hinrichs, 2000; Volberg & Hübner, 2004; Yamaguchi et al., 2000). Generally, information at a global spatial scale elicits greater activity in the right hemisphere, whereas information at local spatial scales elicits greater left hemisphere activity.

Findings from ERP studies have also identified early modulations related to local and global processing at the P1 (Jiang & Han, 2005; Han, He & Woods, 2000), but more frequently around the N1 and P2 components, approximately 150-240ms post stimulus onset (e.g., Beaucousin, Simon, Cassotti et al., 2013; Leek, Roberts, Oliver, Cristino & Pegna, 2016; Proverbio, Minniti & Zani, 1998; Yamaguchi, Yamagata & Kobayashi, 2000). Some studies using hierarchical stimuli find that global and local processing differences occur later, at the N2 and P3 (Han et al., 2000; Heinz & Münte, 1993; Heinze et al., 1998; Malinowksi et al., 2002; Volberg & Hübner, 2004; Yamaguchi et al., 2000).

A number of studies using Navon letters provide a range of different findings relating to the neural mechanisms underlying global and local processing. Jiang and Han (2005) found evidence of a GPE at the P1 (around 80-120ms). Similar results were found in previous ERP studies (Han, Fan, Chen & Zhuo, 1997; Han, He & Woods, 2000). Whereas, Beaucousin et al. (2013) found a global interference effect in the N1 time range: there were comparable N1 amplitudes for local and global levels of stimuli when letters were presented at both levels, but when the local and global levels were incongruent N1 amplitudes were reduced for local processing, compared to global. Other studies identify a later N2 (around 250ms post stimulus onset) as the point where global and local processing differ (e.g., Malinowksi et al., 2002; Volberg & Hübner, 2004; Yamaguchi et al., 2000).

Several ERP studies, however, have found evidence of global and local processing differences at two components. For example, Heinze and Münte (1993) used Navon letters to find ERP correlates of local and global processing and found differences at the N2, then the later P3. They suggest that the N2 (between 150 and 400ms) may be an index of early local and global target perception, with a later P3 difference reflecting target classification. Subsequent ERP studies have reported similar results. For example, Han, Yund and Woods (2003) found differences between local and global processing at the N1, then at a later N2. They also found that early global processing at N1 was diminished when low spatial frequencies were removed, but this was not the case for the later component, therefore demonstrating the important role of low spatial frequencies in early, but not later processing of global stimuli. Other ERP studies using hierarchical figures of simple shapes rather than letters have identified differential processing at similar components. For example, Boeschoten, Kemner, Kenemans and Engeland (2005) found global and local processing differences at the P1, then again at N2.

It is clear from varying ERP results that the task and stimuli are important in determining results in terms of global and local processing differences. However, broadly, it seems that there may be an early component relating to perception of local and global shape information, then a later component related to some classification process.

Differential processing of local and global shape information has also been found in complex 3D objects. Leek et al. (2016) found early differential perceptual sensitivity to local part structure and global shape configuration of complex 3D objects. A perceptual matching task was used, and different object pairs could share either local parts but differ in global shape configuration, share global shape configuration but have different local parts, or share neither. The results showed differential N1 sensitivity to local and global shape similarity between stimulus pairs occurring at the N1, around 170ms post-stimulus onset.

Also, early perceptual sensitivity to stereo versus mono input has been reported in a perceptual matching task using ERPs (Pegna, Darque, Roberts & Leek, 2016). Using complex, 3D objects and mono and stereo viewing conditions, they found that that there was early perceptual sensitivity to the mode of viewing (mono or stereo) at the N1 (between 160-220ms post-stimulus onset). Also, the results showed later modulation of ERP amplitude during an N2 component (between 240-370ms) for stereo and mono input that was linked to the perceptual matching of shape.

1.5 Summary

- 3D shape representation and recognition represents a fundamental problem in vision.
- Current theoretical models make different claims about the functional architecture, and shape information, that supports 3D shape processing.
- Current evidence was reviewed from electrophysiological studies using Navon-style stimuli and more complex 3D stimuli.
- We propose to examine (1) how different kinds of shape information across local and global spatial scales are computed, and integrated, during the perception of 3D object shape, (2) the role of stereo information in 3D shape processing and (3) the temporal dynamics of shape information processing.

2 Chapter II

ERP methodology

2.1 Brief history of EEG

The discovery of the electroencephalogram (EEG) was made by Hans Berger, he found that it was possible to measure electrical activity of the human brain using electrodes on the scalp, amplifying the signal and plotting changes in signal over time (Berger, 1929). Though EEG is a coarse measure of brain activity, due to the many possible sources of neural activity, event-related potentials (ERPs) make it possible to identify cognitive responses associated with particular cognitive, sensory, and motor events using averaging and other techniques.

2.2 Neurophysiology

Neurons in the brain react to a stimulus, beginning a chain of events that leads to two types of recordable electrical activity: action potentials and post-synaptic potentials. Action potentials are discrete voltage spikes that travel from the beginning of the axon at the cell body to the axon terminals, where neurotransmitters are released. Usually, surface electrodes cannot measure action potentials due to the timing and physical arrangement of axons. As neurons rarely fire at exactly the same time (microseconds apart), action potentials in different axons will typically cancel, and the only way to record the action potentials from a large number of neurons is to place the electrode near the cell bodies and to use a very high impedance electrode that is sensitive only to nearby neurons. Also, the duration of an action potential is only about a millisecond, however, postsynaptic potentials typically last tens or even hundreds of milliseconds. Therefore, ERPs usually reflect postsynaptic potentials rather than action potentials. Post-synaptic potentials are voltages that arise when the neurotransmitters bind to receptors on the postsynaptic cell membrane, leading to a change in potential across the cell membrane.

2.2.1 From cell to scalp

Synaptic inputs to a neuron make it more likely to fire (excitatory post-synaptic potential) or less likely to fire (inhibitory post-synaptic potential). When excitatory neurotransmitters are released at the apical dendrites of a cortical pyramidal cell, current flows from the extracellular space into the cell, creating negativity on the outside of the cell, near the apical dendrites. However, to complete the circuit, current flows out of the cell body and dendrites to create a net positivity, which together create a pair of positive and negative electrical charges a small distance apart – a dipole.

For voltages to be large enough to measure at the scalp, thousands or millions of neurons must produce dipoles at the same time and must be spatially aligned. If dipoles are not spatially aligned, positivity and negativity could be cancelled out by adjacent neurons; whenever individual dipoles are more than 90 degrees from each other, they cancel each other to some extent, with complete cancellation at 180 degrees. Neurons, however, are embedded in densely interconnected columns which tend to fall into synchrony. The current from individual cells is conducted through the brain until it reaches the surface, creating a summated voltage at the scalp. The summation of the individual dipoles, however, is complicated by the many folds in the cortex, as well as the different conductivity levels of the skull and brain tissue and scalp. Electricity travels at nearly the speed of light, therefore the voltages recorded at the scalp reflect what is happening in the brain at the same moment in time.

2.3 Recording principles

As ERPs produce a very small signal compared to the electrical noise in the environment, at least 3 electrodes are needed: recording, reference and ground. The ground electrode accounts for environmental noise and the ERP signal is always the difference between a recording electrode and a reference electrode. A reference electrode can be placed in one of several places on the head, these commonly include: the midline position, which is used to avoid amplifying the signal in one hemisphere more than the other; linked mastoids or earlobes, which includes a mathematical average of electrodes attached at either side of the head. Another option is using an average reference, where the average from all outputs is used as a common reference for each channel.

It is a good idea to always try to reduce noise at the source. A Faraday cage can be used to reduce environmental noise, as well as low electrode impedance and using careful experimental design. When the electrodes pick up the EEG, it must be amplified and converted from a continuous analogue form to a discrete, digital form to view.

2.4 Steps in pre-processing

Several steps are required to transform the raw EEG data for subsequent analyses: re-referencing; filtering; segmentation; baseline correction; artefact treatment; and averaging. Filtering is required to remove high frequency data and reduce slow drift by using high and low-pass filters, however filtering will almost always distort the waveforms to some extent. Low pass filtering (removing high frequency noise) is necessary as the EEG data must be sampled at a rate that is at least twice as high as the highest in the incoming data, also for muscle activity creating high-frequency noise and any noise from external electrical devices. One low-pass filtering option is to use Gaussian impulse response function, for minimal temporal distortion. A Gaussian filter with half amplitude cut-off of 30Hz eliminates most line frequency and muscle noise but with very little disruption to long latency ERP components. For high-pass filtering, the aim is to remove very slow voltage changes of non-neural origin during data acquisition which can be caused by sweat and drifts in electrode impedance which can lead to changes in the baseline voltage of the EEG signal. We can remove these shifts by removing voltages lower than

0.01Hz. The EEG recording must be segmented into trials based on the stimulus markers, and, if necessary, remove incorrect response data. Then a baseline correction is applied, whereby average voltage in a predetermined interval (pre-stimulus) in a segment is subtracted from all data points in that segment. A pre-stimulus interval is used for baseline correction as it is assumed that the voltage is not related to the stimuli.

Good experimental design and careful preparation of the electrode cap can help reduce artefacts in the recording, however artefact treatment is necessary as artefacts are large compared to EEG signal, therefore reduce the signal to noise ratio of averaged ERPs if left in. Also, if they are time-locked to a particular condition (such as blinks in one condition and not another) then averaging would not eliminate them and could create differences between conditions that are not really there. Artefacts can include: alpha waves; muscle activity; cardiac spikes; unstable electrode reference effects; linear drifts; and eye movements. Artefact rejection and correction can be used to eliminate artefacts: rejection is removal of artefacts in single trials by eliminating the entire trial, whereas correction involves an algorithm to try to correct for artefacts. An example of an algorithm to correct for eye blinks is to use electrodes above and below the eye, allowing the experimenter to take advantage of the opposite polarity from the two electrodes from movement of the eyelids during a blink. The reasoning is that the voltage of eye movements should create a proportional voltage increase at frontal scalp sites. Independent components analysis (ICA) can also be used, this allows identification of eye blinks based on inverted polarity of the horizontal and vertical eye movement, then removes effects of the boosted voltage on frontal and nearby electrodes, reducing the voltage by a proportion so there are no effects of eye movement on the other electrode voltages.

Averaging is another necessary step used to minimise EEG noise. Many trials per condition are needed for averaging, as any single trial recording will contain EEG signal plus noise, so the more trials there are, the clearer the EEG signal will be once averaged. This is because the noise is reduced through the averaging process, whereas the waveforms should remain the same (based on the assumption that signal for each trial of the same condition should be the same), therefore the signal to noise ratio increases through the averaging process. There is, however, a problem with latency variability – if different trials have slightly different latencies of peaks, then averaging them can distort the peaks, making it impossible to see the peak. If there is latency variation in the peaks, area measures can be used instead of peak amplitude measures, this looks at area under the curve rather than taking values from a peak.

2.5 Strengths and weaknesses of the ERP method

ERP recordings have very good temporal resolution, this can be of 1ms or better under optimal conditions, whereas hemodynamic measures are limited to a resolution of several seconds due to the sluggish nature of the hemodynamic response. Therefore, ERPs can address some questions that functional magnetic resonance imaging (fMRI) and positron emission tomography (PET) cannot. However, spatial resolution is far superior using hemodynamic measures, with spatial resolution in the millimetre range, whereas with ERP recording, source localisation is difficult due to the inverse problem. The inverse problem stems from the different electrical conductivity of the brain tissue, the skull and the scalp. Because electricity tends to follow the path of least resistance, ERPs tend to spread laterally when they encounter the high resistance of the skull. Therefore, at the skull surface, the distribution of voltage is blurred and an ERP generated in one part of the brain can lead to voltages at different parts of the scalp. This leads to the inverse problem, as we only have scalp projections and a head model, so there are infinite possible inverse solutions that fit the data.

A benefit of using ERPs is that it is non-invasive, compared to single electrode recording and also PET, with which only a small number of trials can be used in case of exposing subjects to excessive radiation. ERP recording is also relatively inexpensive, compared to magnetoencephalography (MEG), fMRI and PET. One disadvantage of the ERP method is that it produces such a small signal that it requires a large number of trials to measure effects, compared to behavioural measures.

An alternative to ERPs, maintaining good temporal resolution, but with superior spatial resolution is MEG. Using MEG resolves the issue of blurring of voltage caused by the high resistance of the skull as it records magnetic fields instead of electrical potentials. An electrical dipole is always surrounded by a magnetic field and these summate in the same way as voltages. The skull is transparent to magnetism, therefore magnetic fields are not blurred by the skull, leading to much greater spatial resolution than is possible with electrical potentials. The magnetic equivalent of an ERP is an event-related magnetic field (ERMF). A dipole that is perpendicular to the surface of the scalp is accompanied by a magnetic field that leaves the head on one side of the dipole and enters back again on the other side. If you place a highly sensitive probe called a SQUID (super-conducting quantum interference device) next to the head, it is possible to measure the magnetic field as it leaves and re-enters the head.

2.6 Component analysis

Components can be identified by finding the maximum amplitude in a particular time window, a peak amplitude measure. Alternatively, mean area measures can be used, where the mean voltage in a particular time window is measured. There are several issues that should be considered when identifying components, including the inflation of Type I error. Luck and Gaspelin (2017) warn against using grand-average data to visually identify components, as if peaks are identified from comparison of conditions where there are differences, this will obviously increase the likelihood that there are differences between the conditions at those selected times. They suggest that better practices to identify components include defining time windows based on the overall grand average, collapsed across conditions, or by identifying a component a priori, perhaps using a localiser contrast. These alternatives ensure that the time windows of interest are identified independently of the factors being manipulated. A further option to identify components in the ERPs is to use mass univariate analyses, which allow identification of the time windows where there is greatest activity.

2.7 Mass Univariate Analysis (MUA)

Mass univariate analysis (e.g., Groppe, Urbach & Kutas, 2011; Guthrie & Buchwald, 1991) involves using pair wise, time-frame by time-frame, permutation tests based on repeated measures t-tests across all electrodes in a given time frame. An a priori criterion for significance testing is adopted in which a threshold of $p<.01$ (two-tailed) must be attained for at least 12 consecutive time frames in at least 5 neighbouring electrodes, based on simulation studies (Guthrie & Buchwald, 1991; Murray, Brunet & Michel, 2008). Type 1 error is controlled by using the above criteria, but the MUA analysis also incorporates corrections for false discovery rate using 'randomisation' tests. Randomisation tests bootstrap the random distribution of differences between conditions against which the statistical significance of the observed difference is calculated. One benefit of this method is that it provides an overview of activity across the entire scalp, rather than looking at waveforms from selected clusters of/or individual electrodes.

2.8 ERP Protocol

In the subsequent empirical chapters, the following protocol was used:

- Eye movements and blinks were corrected using the ICA protocol in Analyser 2 software and segmented data was then visually inspected with trials containing artefacts rejected.
- Epochs that contained muscle or skin potential artefacts were rejected and only trials on which participants gave a correct response were included.
- Activity from all electrodes was sampled at a rate of 1024Hz. Offline 30 Hz (48 db/oct slope) low-pass and 0.1 Hz (48 db/oct slope) high-pass filters were applied to the data.

- All data was re-referenced to an average reference which was then used to generate the grand averages.
- A 100ms pre-stimulus interval for the baseline correction was used.
- Continuous recording took place during the test phase of the experiment and trials were epoched/segmented from 100ms pre-stimulus to 800ms post-stimulus onset.

Early ERP waves were identified based on the topography, global field power (GFP), deflection and latency characteristics of the respective grand average ERPs time-locked to stimulus presentation. Preliminary epochs of interest for each component were defined based on deflection extrema in the mean local field power (e.g., Brunet, Murray & Michel, 2011; Lehmann & Skrandies, 1980; Murray, Brunet & Michel, 2008). Peak detection was time-locked to the electrode of maximal amplitude for each component. Mass univariate analyses were then used to verify the statistical robustness of our earlier analyses. MUA can be used to provide an additional 'bias free' measure of statistical contrasts across all electrodes (unlike using selected clusters in the standard waveform).

In all three experiments, the electroencephalograph (EEG) was recorded continuously through 128 electrodes (see Figure 11) placed on an ECI cap (Electro-Cap International, Ohio, USA) using the Active-Two Biosemi EEG system (Biosemi V.O.F Amsterdam, Netherlands).

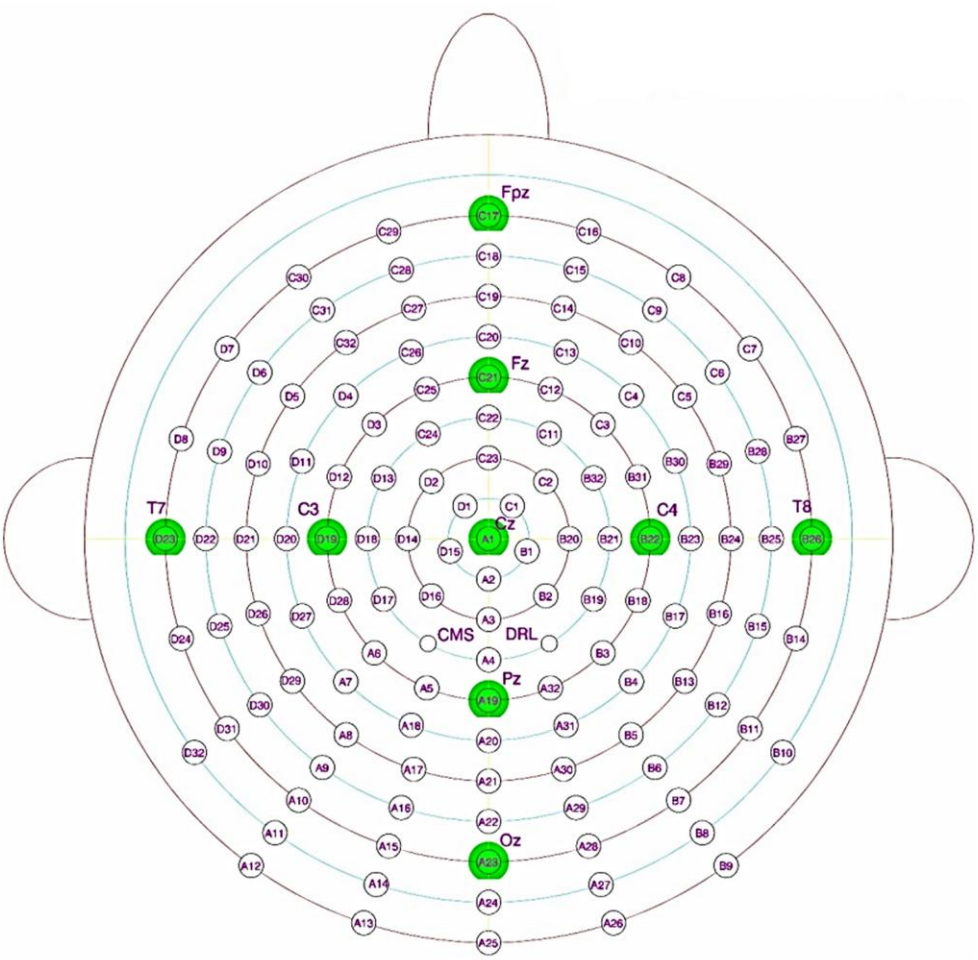


Figure 11. Electrode montage of the 128 channel cap used.

2.9 Summary

- ERP is the best method to investigate the temporal aspects of processing, however not for spatial information.
- It is important to carefully process EEG recordings to avoid inflation of Type I error.
- We used global field power to identify the latency of components, then local peak amplitude measures within those time frames.
- Mass univariate analyses were also conducted in order to identify when the differences between conditions arose using all electrodes, rather than selected clusters.

3 Chapter III

Stereo viewing modulates three-dimensional shape processing during object recognition: A high-density ERP study

The aim of the chapter is to investigate three issues relating to the perceptual processes involved in 3D object recognition. First, the underlying determinants of the differential early perceptual sensitivity to local part similarity and global shape configuration remain to be fully elucidated. This sensitivity could be driven solely by overlap in the 2D projection (i.e. global shape silhouette defined by occluding contour), rather than reflecting similarity in the 3D object part configuration. Second, it remains unclear whether these effects solely reflect processes related to the perceptual derivation of object shape, or whether they would also be observed in a task of recognition requiring the matching of a perceptual representation of 3D object shape to a long-term memory representation. Third, current evidence does not tell us whether stereo disparity modulates differential perceptual sensitivity to local and global shape structure during object recognition. This finding would present a challenge to several current theoretical models that attribute no functional significance to stereo information during 3D object shape perception.

The remainder of this chapter comes from Oliver, Cristino, Roberts, Pegna & Leek, 'Stereo viewing modulates three-dimensional shape processing during object recognition: A high-density ERP study', JEP:HPP (in press).

3.1 Abstract

The role of stereo disparity in the recognition of three-dimensional (3D) object shape remains an unresolved issue for theoretical models of the human visual system. We examined this issue using high-density (128 channel) recordings of event-related potentials (ERPs). A recognition memory task was used in which observers were trained to recognise a sub-set of complex, multi-part, 3D novel objects under conditions of either (bi-) monocular or stereo viewing. In a subsequent test phase they discriminated previously trained targets from untrained distractor objects that shared either local parts, 3D spatial configuration or neither dimension, across both previously seen and novel viewpoints. The behavioural data showed a stereo advantage for target recognition at untrained viewpoints. ERPs showed early differential amplitude modulations to shape similarity defined by local part structure and global 3D spatial configuration. This occurred initially during an N1 component around 145-190ms post-stimulus onset, and then subsequently during an N2/P3 component around 260-385ms post-stimulus onset. For mono viewing, amplitude modulation during the N1 was greatest between targets and distracters with different local parts for trained views only. For stereo viewing, amplitude modulation during the N2/P3 was greatest between targets and distracters with different global 3D spatial configurations and generalised across trained and untrained views. The results show that image classification is modulated by stereo information about the local part, and global 3D spatial configuration of object shape. The findings challenge current theoretical models that do not attribute functional significance to stereo input during the computation of 3D object shape.

Statement of Public Significance: The aim of this research is to elucidate how the human visual system processes sensory information about shapes of three-dimensional (3D) objects so that we can perceive, and recognise, them. We asked whether these processes are sensitive to both monocular and stereo visual input. To answer this question we measured electrophysiological responses generated in the brain while people viewed, and made recognition judgements about, mono or stereo images of 3D objects. The objects could differ from each in terms of their part structure, or overall 3D spatial configuration. The results showed that the visual system processes these sorts of shape properties differently, and that how it does so is influenced differently by mono and stereo visual input. The findings shed new light on the role of stereo information in the visual perception and recognition of 3D object shape.

3.2 Introduction

The human visual system is remarkable for its ability to rapidly and accurately classify three-dimensional (3D) objects despite variability in sensory input (e.g., Arguin & Leek, 2003; Bar, 2003; Bar, Kassam, Ghuman et al., 2006; Cichy, Pantazis & Oliva, 2014; Fabre-Thorpe, 2011; Harris, Dux, Benito & Leek, 2008; Kirchner & Thorpe, 2006; Leek, 1998a; 1998b; Leek, Atherton & Thierry, 2007; Leek, Davitt & Cristino, 2015; Leek & Johnston, 2006; Leek, Roberts, Oliver, Cristino & Pegna, 2016; Tarr & Bulthoff, 1998; Thorpe, Fize & Marlot, 1996; VanRullen & Thorpe, 2001).

One important, and unresolved, issue is whether, and under what conditions, information derived from stereo (binocular) disparity influences the recognition of 3D object shape (e.g., Bennett & Vuong, 2006; Chan, Stevenson, Li & Pizlo, 2006; Cristino, Davitt, Hayward & Leek, 2015; Edelman & Bülthoff, 1990; Koenderink, van Doorn & Kappers, 1992; Li, Pizlo & Steinman, 2009; Norman, Swindle, Jennings et al., 2005; Norman, Todd, & Phillips, 1995; Pegna, Darque, Roberts & Leek, 2016; Pizlo, Sawada, Li, Kropatsch, & Steinman, 2010; Welchman, Deubelius, Conrad, Bülthoff & Kourtzi, 2005). Some current theories attribute little, if any, significance to stereo information (e.g., Bülthoff & Edelman, 1992; Chan et al., 2006; Pizlo, 2008; Riesenhuber & Poggio, 1999; Serre, Oliva & Poggio, 2007). For example, in the HMAX model (Riesenhuber & Poggio, 1999), image classification is accomplished within a multi-layer feedforward architecture in which hierarchically structured edge-based representations of object shape are computed from monocular image contour –see also other recent approaches to image classification based on hierarchical deep networks (e.g., Cichy, Khosla, Pantazis, Torralba & Oliva, 2016; Khaligh-Razavi & Kriegeskorte, 2014; Krizhevsky, Sutskever & Hinon, 2012). Pizlo (2008; see also Li et al., 2009; Pizlo et al., 2010) has proposed that 3D object structure is computed solely from 2D shape information subject to the application of simplicity constraints (symmetry, compactness, planarity and minimum surface area). On other accounts, the contribution of stereo input is not ruled out, but neither explicitly incorporated into the proposed theoretical framework (e.g., Biederman, 1987; Leek, Reppa & Arguin, 2005; Ullman, 2006). This contrasts with theoretical models that have attributed functional significance to certain kinds of stereo-defined shape information in object recognition - such as the computation of local surface depth orientation, and the specification of 3D object structural descriptions (Marr & Nishihara, 1978).

Although binocular disparity has been shown to contribute to the perception of surface properties such as slant, tilt and curvature (e.g., Ban & Welchman, 2015; Norman et al., 1995; Norman et al., 2009; Welchman et al., 2005; Wexler & Ouarti, 2008; Wismeijer, Erkelens, Ee, & Wexler, 2010), its role in the recognition of complex 3D object shape remains unclear. Indeed, it has been argued that although stereo information (i.e., local depth disparity) facilitates processing of 3D surfaces properties this does not, in itself, establish a functional link between stereo vision and the perception (and recognition) of complex (i.e., multi-part) 3D object shape per se (Li et al., 2009; Pizlo, 2008; Pizlo et al., 2010). This issue has been investigated in previous studies by assessing the effects of stereo disparity on the perceptual matching of object shape across changes in viewpoint. The results provide a mixed picture with stereo advantages reported in some studies (e.g., Bennett & Vuong, 2006; Burke, 2005; Burke, Taubert, & Higman, 2007; Chan, et al., 2006; Edelman & Bülthoff, 1990; Hong Liu, Ward, & Young, 2006; Lee & Saunders, 2011; Rock & DiVita, 1987; Simons, Wang & Roddenberry, 2002), but not in others (Humphrey & Khan, 1992; Pasqualotto & Hayward, 2009). Recently, Cristino et al. (2015) have proposed that stereo information is computed during the visual perception of object shape. It is more likely to be used to supplement shape information derived from mono-ocular cues when object recognition (i.e., target/non-target discrimination or view generalization) is facilitated by the derivation of 3D object structure. In support of this hypothesis, they showed that stereo input facilitates the classification of complex multi-part 3D objects across large, but not small, changes in depth rotation. In other recent work, Pegna et al. (2016) have found early perceptual sensitivity to stereo versus mono input in a perceptual matching task using event-related potentials (ERPs). In that study, ERPs were recorded while observers made shape equivalence judgments about pairs of sequentially presented novel 3D objects under conditions of stereo or mono viewing. The results showed an early perceptual sensitivity to the mode of input shown by a negative amplitude modulation between 160-220ms post-stimulus onset. The results also showed later modulation of ERP amplitude during an N2 component between 240-370ms for stereo and mono input that was linked to the perceptual matching of shape[1].

The aim of the current study was to determine whether stereo disparity contributes to object processing during the recognition of 3D object shape. The rationale was based on recent work by Leek et al. (2016) who found evidence for early differential sensitivity of ERP amplitudes to local part structure and global shape configuration of complex 3D objects in mono displays. In that study ERPs were recorded while observers made shape matching judgments to sequentially presented pairs of novel objects under conditions of mono viewing. Different object pairs could either share local parts but differ in global shape configuration, share global shape configuration but have different local parts, or share neither. The results showed differential N1 sensitivity to local and global shape similarity between stimulus pairs occurring around 170ms post-stimulus onset. These findings provide evidence that mental representations of complex 3D object shapes comprise both local higher-order parts, and the global spatial configuration of these parts - consistent with theoretical models, and other empirical evidence, supporting this distinction (e.g., Arguin & Saumier, 2004; Behrmann, Peterson, Moscovitch & Satoru, 2006; Behrmann & Kimchi, 2003; Biederman, 1987; Hummel, 2013; Hummel & Stankiewicz, 1996; Marr & Nishihara, 1978). We hypothesized that one way in which stereo disparity may contribute to recognition is by facilitating the computation of 3D object representations via depth information. These representations could augment a range of shape information including surface depth gradients and curvature, higher-order part boundaries, and the 3D

[1] Throughout the paper we use the term 'mono' to describe non-stereo 'bi-monocular' visual input (that is, where there is no disparity between visual inputs to the left and right eye). Stereo refers to visual input with binocular disparity (i.e., different left and right eye images for a given viewpoint).

spatial configuration of (volumetric) object parts. Of relevance to the current study is whether stereo input might differentially modulate the sensitivity of object recognition processes to local part and global 3D spatial configuration information. For example, under some structural description accounts, object parts are computed directly from 2D image-based input derived from local edge relations (e.g., non-accidental properties or NAPs – Biederman, 1987). This level of representation may be sufficient where object recognition can be based on a parts-based description of object identity, or where the discrimination of target and non-target objects can be achieved based on part composition. In other situations, it may be beneficial to compute a global 3D object model which specifies (amongst other attributes) the spatial configuration of local object parts – for example, where recognition depends on discrimination among objects with similar parts but different spatial configurations.

To test this prediction we used ERPs, which have been previously shown by Leek et al (2016) to show differential amplitude sensitivity to local and global shape structure. Unlike earlier work, we also wanted to examine this issue in the context of an object recognition task rather than the perceptual matching of sequentially presented stimuli. Object recognition differs from perceptual matching in that the former requires indexing a (stored) long-term memory representation of object shape. We used a recognition memory task in which observers had to first memorize a sub-set of complex novel 3D objects (targets) and subsequently discriminate them from visually similar non-target (not previously memorized) objects. We then contrasted effects of target/non-target similarity defined by local part and global 3D shape configuration under conditions of stereo and mono viewing. We predicted that stereo presentation would enhance ERP modulations related to object discrimination weighted towards perceptual analysis of 3D global shape configuration.

3.3 Methods

3.3.1 Participants

Forty Bangor University students (24 female, mean age 21.46, *SD*=3.16, 3 left-handed) participated for course credit. The sample was recruited through an online participation portal. All participants had normal or corrected-to-normal visual acuity. Ethics approval was granted by the Local Ethics Committee and in accordance with British Psychological Society guidelines. Informed consent was obtained and participants were free to withdraw from the study at any time without penalty.

3.3.2 Apparatus & Stimuli

The stimuli comprised a set of 48 novel computer-generated 3D objects. There were 12 target objects and 36 non-targets (distracters) varying in visual similarity to the targets (see Figure 1). Each stimulus comprised a unique spatial configuration of four different volumetric parts. The parts were defined by variation among non-accidental properties (NAPs) comprising: edges (straight vs. curved), symmetry of the cross section, tapering (co-linearity) and aspect ratio (Biederman, 1987). The object models were produced using Strata 3D CX software (Strata, USA), then rendered in Matlab using a stereo camera rig programmed with custom code. To create the stereo images left and right eye images were rendered without 'toeing in' using an Inter Pupillary Distance (IPD) of 62mm. In both mono and stereo viewing conditions, participants wore polarised 3D glasses to view the stimuli presented on a passive inter-leaved 3D stereo monitor (60Hz 27" AOC 3D monitor (D2769VH), resolution = 1920x1080 pixels). In the stereo condition, participants viewed objects rendered from two viewpoints (left eye and right eye). In the (bi-) mono condition, participants viewed the objects with the same (right eye) rendered image presented to both eyes.

The stimuli were then normalised in size across objects to sustain an average on screen size of 17° x 17°). All stimuli were rendered using a mustard yellow colour: R=227, G=190, B=43, and presented on a white background to facilitate figure/ground segmentation. Object models were rendered with shading using a single top-left light source but without (internal or external) cast shadow (Leek et al., 2015).

For each of the 12 target objects, 3 corresponding non-targets were designed: one variation was composed of the same parts arranged in a different spatial configuration (SD - **S**ame Parts/**D**ifferent spatial configuration – 'locally-similar'); a second variation was composed of different parts arranged in the same configuration as the target (DS - **D**ifferent parts/**S**ame spatial configuration – 'globally-similar'); finally, in a third variation comprised different parts and spatial configuration (DD - **D**ifferent parts/**D**ifferent spatial configuration – 'Dissimilar'). Each object was rendered at six different viewpoints varying by 60 degree rotations in depth around a vertical axis perpendicular to the line of sight.

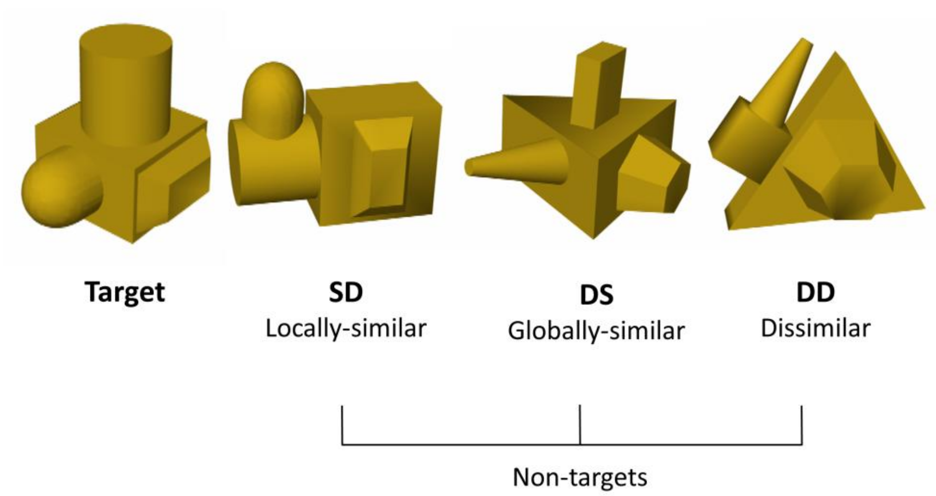


Figure 12. An example of one target object and its three corresponding SD (locally similar), DS (globally similar) and DD (dissimilar) non-targets.

Measures of target/non-target image similarity using three models based on (1) Pixel overlap, (2) Gabor filter bank and (3) HMAX - C1 output layer (Serre, Oliva & Poggio, 2007) were computed on the 2D mono stimulus images using the Matlab Image Similarity Toolbox

(Seibert & Leeds https://github.com/daseibert/image_similarity_toolbox). In the toolbox, the pixel overlap model computes the sums of squared differences in pixel intensity values between images. The Gabor filter bank model projects the image onto a Gabor wavelet pyramid as a model of V1 orientation selectivity (Kay, Naselaris, Prenger & Gallant, 2008), using filters spanning eight orientations, four sizes (image %) and X, Y positions. The Euclidian distance between the resulting vector of filter responses is compared between images. The HMAX model is based on the C1 output layer of the hierarchical feed-forward image classification model of Serre et al (2007). We use this model to provide an estimate of image-based stimulus similarity between target and non-target conditions. Table 1 shows the mean normalised similarity values of the three models for both target vs SD (locally-similar), DS (globally-similar) and DD (dissimilar) distracter image contrasts between trained and untrained viewpoints. A 2 (Viewpoint: Trained, untrained) x 3(Stimulus type: SD; DS; DD) x 3 (Model: pixel overlap; HMAX; Gabor) repeated measures ANOVA, showed no significant main effects. However, there was an interaction between Stimulus type and Model, F (4, 44) =3, p=.029. Post-hoc analyses showed that there were no differences between stimulus types for the pixel overlap or Gabor models. For HMAX there was a significant difference between SD (locally-similar) and DS (globally-similar) stimulus types (p = .02) driven by the lower mean (normalised) similarity values for trained views of target/DS (globally-similar) relative to either target/SD (locally-similar) or target/DD (dissimilar) stimulus contrasts.

Table 1. Table showing mean (SD) normalised (0-1) image similarity between targets and distractors (non-targets) for the Pixel overlap, HMAX and Gabor models. Smaller values indicates lower similarity.

MODEL		**VIEW** TRAINED **M**	**(SD)**	**UNTRAINED M**	**(SD)**
PIXEL OVERLAP	SD (Locally-similar)	0.42	0.19	0.39	0.17
	DS (Globally-similar)	0.34	0.16	0.35	0.18
	DD (Dissimilar)	0.39	0.18	0.39	0.14
HMAX	SD (Locally-similar)	0.31	0.10	0.31	0.09
	DS (Globally-similar)	0.17	0.11	0.31	0.09
	DD (Dissimilar)	0.30	0.14	0.28	0.09
GABOR	SD (Locally-similar)	0.30	0.15	0.31	0.13
	DS (Globally-similar)	0.33	0.11	0.35	0.14
	DD (Dissimilar)	0.26	0.12	0.25	0.07

A 2 (Display: mono/stereo) x 4 (Stimulus type: Target, SD (locally-similar), DS (globally-similar, DD (dissimilar)) mixed factorial design was used, with Display as a between-subjects factor and Stimulus type as a within-subjects factor. Participants were randomly allocated to either the mono or stereo display group. There were 20 participants in each group. The stereo display group completed a verification task to assess their ability to fuse stereo images using interleaved polarised displays. During this task they were seated 60 cm from the screen and shown a random-dot stereogram with an embedded figure eight that was only perceivable with stereo fusion using polarised glasses. Participants were asked to report what they saw. All participants correctly reported the embedded stereo figure. The main study comprised two phases: learning and test. One group completed both the learning and test phases in mono. The other group completed both the learning and test phases in stereo. This aspect of the design ensured that any observed differences between the viewing conditions during the test phase cannot be due to mismatches in stimulus presentation formats between the learning and test phases. During the learning phase for both groups 12 objects were memorised. In the learning phase each target was seen at three viewpoints distinguished by rotations of 120 degrees around a vertical (y) axis defined with reference to the object – see Figure 13. In the test phase, each target and non-target was seen from six different viewpoints distinguished by 60 degree rotations around the y axis. In the learning phase each target was shown at each of three viewpoints three times. In the test phase, the 12 targets were presented at each of six viewpoints three times (216 target trials in total). There were also 36 non-targets (three distracters for each of the 12 targets). Each non-target was presented once at each of the six test viewpoints (six trials per non-target = 216 non-target trials in total, 72 trials per non-target condition). In total there were 432 trials in the test phase comprising equal numbers of target and non-target trials. Trial order was randomised.

3.3.3 Procedure

3.3.3.1 Learning phase

During the learning phase participants in both the stereo and mono groups wore polarised glasses but viewed stereo or mono images depending on the group assignment. The learning phase comprised three identical training sessions conducted over three days in separate training sessions. The purpose of the learning phase was for participants to memorise each of the 12 targets, and an associated unique stimulus number. Only participants who were able to identify targets to a criterion level of 80% after the three training sessions proceeded to the test phase. Each training session comprised a memorisation stage and a verification stage. During the memorisation stage target objects were presented centrally (duration = 3s) on a computer monitor sequentially at three different training viewpoints denoted 0°, 120° and 240° - see Figure 13. Target presentation was preceded by an identification number (1-12). Target identification numbers were randomly assigned across the target set but were the same for all participants. There were 36 trials (12 objects x 3 viewpoints) in each block of memorisation trials. After the memorisation phase, participants completed a verification task in which the 12 targets were shown randomly, one-at-a-time and for unlimited duration (until response), at each of the three viewpoints. After each stimulus, participants provided the associated target number via a key press on a standard PC keyboard. Feedback was given via a 'Correct' or 'Incorrect' message displayed centrally on the monitor. The memorisation and verification tasks were repeated three times per training session (9 times across the three training sessions). All participants completed all three training sessions (regardless of whether they reached criterion accuracy earlier).

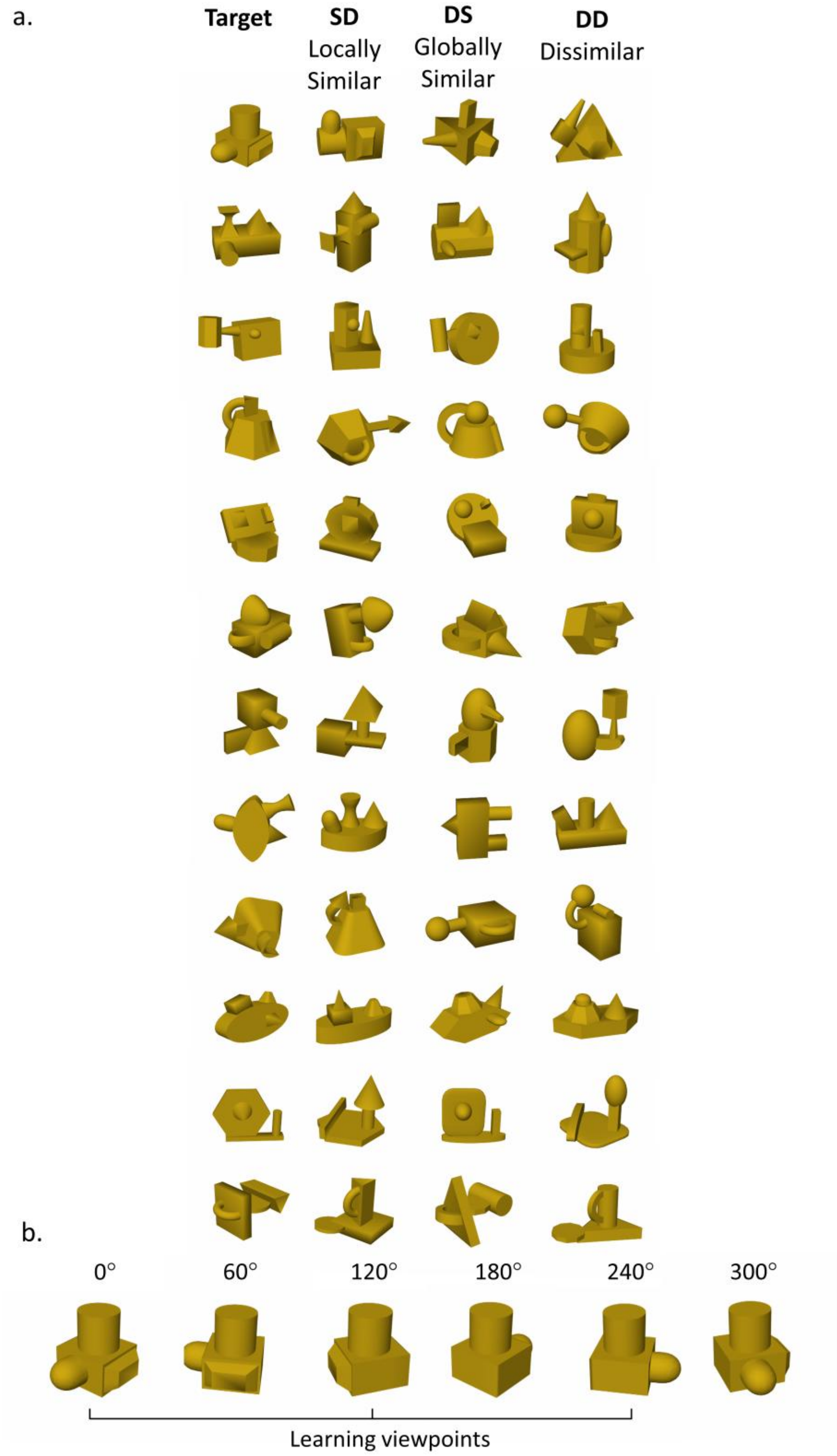


Figure 13. (a) All 12 target objects used in the study, with three distractor objects: SD (locally-similar); DS (globally-similar); DD (dissimilar). (b) One target object at the three learning (0°; 120°; 240°) and additional three test phase viewpoints (60°; 180°; 300°).

3.3.3.2 Test phase

During the test phase, participants in both the stereo and mono groups wore polarised glasses but viewed stereo or mono images depending on the group assignment. After the participants had completed three training sessions and had achieved the criterion level of performance in the learning phase, they completed the test phase involving a recognition memory task. The final training session of the learning phase was completed immediately before the test phase. EEGs were recorded during the test phase (see below). Each trial involved presentation of one stimulus (either a target or non-target) at one of six viewpoints. At the start of each trial a small central fixation cross was presented in the centre of the monitor at 0.7° of visual angle. The duration of the fixation cross was jittered randomly in 50ms increments between 500-800ms. Following onset of the fixation marker the test stimulus was shown for 750ms. This stimulus was replaced by a response screen (centrally presented question mark). All trial events were separated by an inter-stimulus interval of one screen refresh (17ms). Participants were instructed to respond via a button press using a standard PC keyboard ("1" for target and "2" for non-target – with the fore and middle fingers of the right hand respectively for all participants) indicating whether the stimulus shown was one of the 12 objects that they had previously memorised regardless of its orientation. They were alerted to the fact that the stimuli could be presented at previously seen and novel viewpoints. Participants could only respond following onset of the response screen, and not during presentation of the stimulus. This was done to help reduce potential motor response artefacts in the EEG. The response screen remained until a response was given (see Figure 14). The inter-trial interval was a blank screen presented for 1000ms. For the behavioural data the dependent measure was response accuracy. RTs were not collected because keyboard responses were only acquired from the onset of the response screen. This was done to reduce motor artefacts in the ERPs associated with the stimulus event.

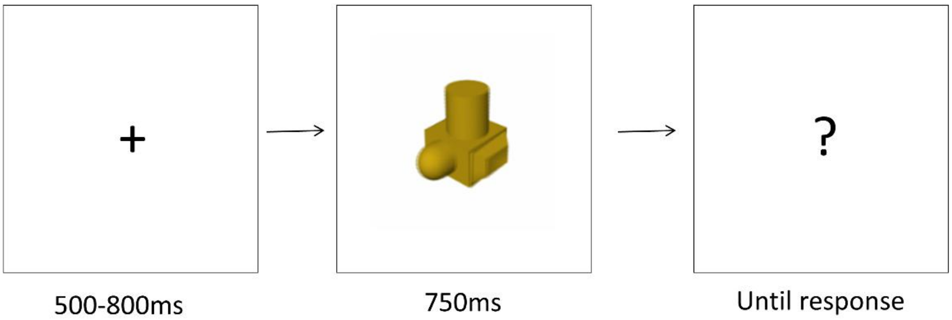


Figure 14. An illustration of the trial sequence comprising: (1) jittered fixation from 500-800ms, (2) stimulus (target or non-target) presentation for 750ms, (3) response prompt.

3.3.4 Electrophysiological recording and processing

The electroencephalograph (EEG) was recorded continuously through 128 electrodes placed on an ECI cap (Electro-Cap International, Ohio, USA) using the Active-Two Biosemi EEG system (Biosemi V.O.F Amsterdam, Netherlands). Eye movements and blinks were corrected using the ICA protocol in Analyser 2 software and segmented data was then visually inspected with trials containing artefacts rejected. Epochs that contained muscle or skin potential artefacts were rejected. Only trials on which participants gave a correct response were included. The mean number of correct trials per subject after artefact rejection was: 189.25 (SS/target), 62.61 (SD/locally-similar) and 67.61 (DS/globally-similar) and 67.82 (DD/dissimilar). Activity from all electrodes was sampled at a rate of 1024Hz. Offline 30 Hz low pass and 0.1 Hz high pass filters were applied to the data. Data were re-referenced to an average reference which was then used to generate the grand averages. We used a 100ms pre-stimulus interval for the baseline correction. Continuous recording took place during the test phase and trials were epoched/segmented from -100ms to stimulus offset (750ms). All ERP data acquired from onset of the response prompt were discarded.

3.3.5 EEG analyses

Four early visual evoked potential components P1, N1, P2 and an N2-P3 complex were identified based on the topography, global field power (GFP) deflection and latency characteristics of the respective grand average ERPs time-locked to stimulus presentation. Preliminary epochs of interest for each component were defined based on deflection extrema in the mean local field power (e.g., Brunet, Murray & Michel, 2011; Lehmann & Skrandies, 1980; Murray, Brunet & Michel, 2008). Peak detection was time-locked to the electrode of maximal amplitude for each component. The latency of peak amplitude was used to define epochs for analyses of four components: Mono - P1 (85-125ms; Peak latency (A10) = 105ms; N1 (145-185ms; Peak latency (B7) = 165ms); P2 (200-240ms; Peak latency (A8) = 220ms); N2-P3 complex (285-385ms; Peak latency (A8) = 335ms); Stereo - P1 (90-130ms; Peak latency (B7) = 110ms); N1 (150–190ms; Peak latency (A11) = 170ms); P2 (195-235ms; Peak latency (A8) = 215ms); N2-P3 (260-360ms; Peak latency (A7) = 310ms).

Symmetrical clusters were extracted over the left (LH) and right (RH) hemispheres comprising nine spatially adjacent posterior electrodes: RH: A32, B3, B4, B5, B6, B7, B8, B10, B11 and LH: A5, A6, A7, A8, A9, A10, A11, D31 and D32, which correspond with electrode locations CP2, P4, P6, P8, PO8 and CP1, P3, P5, P7, PO7 of the extended 10–20 system. These electrode clusters formed the regions-of-interest (ROIs) for the subsequent analyses of contrasts between stimulus conditions. Standard waveform analyses were based on the amplitude data as a measure of differential ERP sensitivity to 3D shape similarity between mono and stereo viewing. Mean amplitudes were analysed using the General Linear Model by way of ANOVA. Greenhouse-Geisser corrections were applied to all analyses of ERP data. Corrected degrees of freedom are reported where applicable. An a priori alpha level of .05 (two-tailed) was adopted. Exact p values are reported (p = x) except where $p < .001$.

3.3.6 Mass Univariate Analyses

Mass Univariate analyses (Groppe, Urbach & Kutas, 2011; Guthrie & Buchwald, 1991; Murray et al., 2008) were used to complement the standard waveform analyses. This involved using pair wise, frame-by-frame, repeated measures t-tests across all 128 electrodes. An a priori criterion for significance was adopted in which a threshold of $p<.01$ (two-tailed) must be attained for at least 12 consecutive time frames in at least 5 neighbouring electrodes over time windows of 150ms (Guthrie & Buchwald, 1991). For this purpose, the mass univariate analyses were conducted on 150ms bins (0-150ms; 151-300ms; 301-450ms) encompassing the P1, N1, P2 and N2/P3 components.

3.4 Results

3.4.1 Behavioural Results

Accuracy data were log transformed prior to statistical analyses.

3.4.1.1 Learning Phase

A 3(Training day) x 2(Display: mono; stereo) mixed ANOVA, with Display as a between subjects factor showed significant main effects of Training day, F (2,60) =58.06, p<.001, with accuracy (% correct) increasing over time, from day one (M=69.48, SD=17.38) to two (M=94.71, SD=8.06), p<.001, and two to three (M=98.09, SD=4), p=.006. There were no differences between mono and stereo display groups and all participants passed criterion by the end of the third training session[2, 3].

3.4.1.2 Test Phase

Figure 15 shows mean percentage correct responses per condition. The data were analysed using a 4 (Stimulus type: Target; SD (locally-similar); DS (globally-similar); DD (dissimilar)) x 2 (Stimulus viewpoint: trained/untrained) x 2 (Display: mono/stereo) mixed ANOVA, with Display as a between subjects factor. There were significant main effects of Stimulus type, F (3, 90) =13.5, p<.001, and Stimulus viewpoint, F (1, 30) =10.41, p =.003, with higher overall accuracy for trained (M=97.05%, SD=2.65) than untrained (M=95.4%, SD=3.42) viewpoints. There was also a significant three-way interaction, F (3, 87) = 3.19, p =.027. To investigate this further we analysed mono and stereo data separately using 4 (Stimulus type) x 2 (Stimulus viewpoint) repeated measures ANOVAs. For the mono viewing group, there was an interaction between Stimulus type and Stimulus Viewpoint, F (3,45)=5.9, p=.002. This derived from significantly higher accuracy for trained than untrained viewpoints for target stimuli,

[2] 9/28 participants reached criterion accuracy after the first training session
[3] Patterns of behavioural and ERP data for the three left-handed participants and the group were the same across conditions.

p=.003 (see Figure 15). In contrast, for the stereo viewing group there were no significant main effects or interactions. Finally, accuracy for targets presented at untrained views was higher for stereo (M=94.68%, SD=5.09) than mono (M=85.19%, SD=14.46) displays (p=.035). This pattern of results is consistent with a stereo advantage in view generalisation for targets between trained and untrained views.

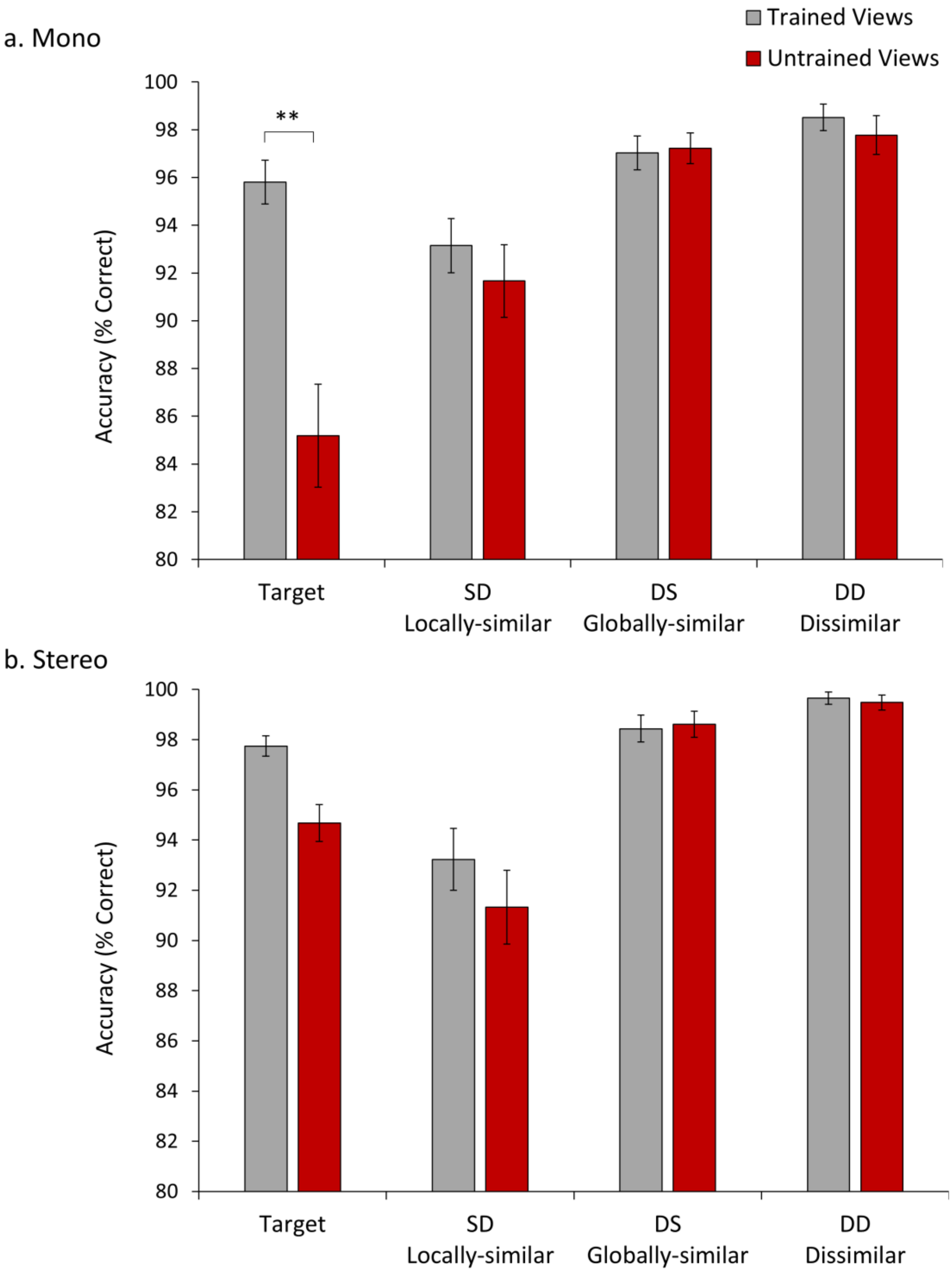


Figure 15. Accuracy for targets in mono and stereo viewing conditions in the test phase. Bars show standard error.

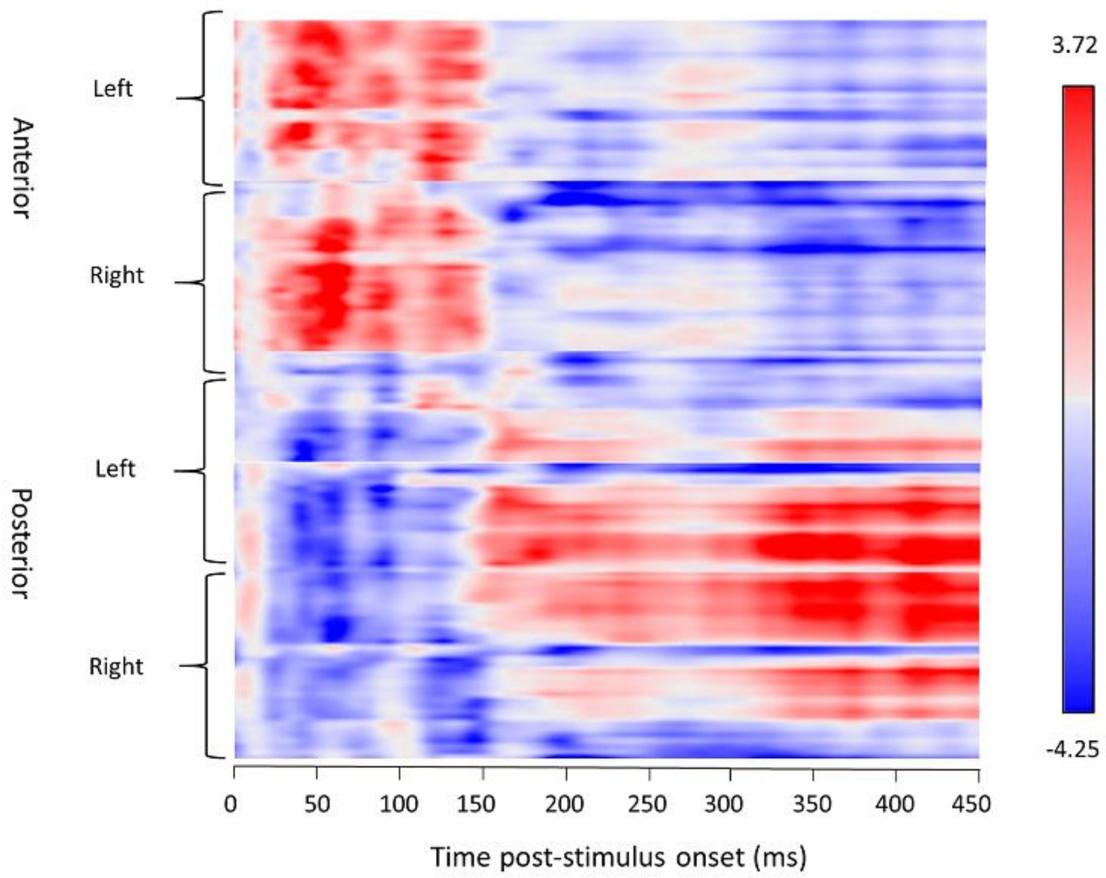
3.4.2 Analyses of ERP data

The aims of these analyses were: (1), to determine whether the ERP showed sensitivity to the manipulation of stereo and mono viewing; (2), to establish whether the ERPs were differentially sensitive to target/non-target shape similarity defined by either shared local parts or global 3D spatial configuration; (3), to determine whether differential perceptual sensitivity to these shape attributes was modulated by mono versus stereo viewing.

3.4.2.1 ERP Analyses I: Perceptual sensitivity to stereo/mono presentation

We first wanted to determine whether our display manipulation of stereo versus mono presentation was sufficient to induce a measurable early perceptual sensitivity in visual evoked potentials. Mass univariate analyses were used to identify a temporal marker defining the earliest time point of differential ERP sensitivity to mono versus stereo viewing. A point-wise mass univariate contrast between the mono and stereo viewing across all conditions revealed differences in the ERP from around 50ms post-stimulus onset over a large group of posterior, temporal-occipital and anterior leads. This difference was sustained during the P1 component over left occipital and some frontal electrodes (see Figure 16). These analyses confirm an early perceptual sensitivity to mono versus stereo viewing.

a.


Anterior
Left
Right
Posterior
Left
Right
3.72
-4.25
0
50
100
150
200
250
300
350
400
450
Time post-stimulus onset (ms)

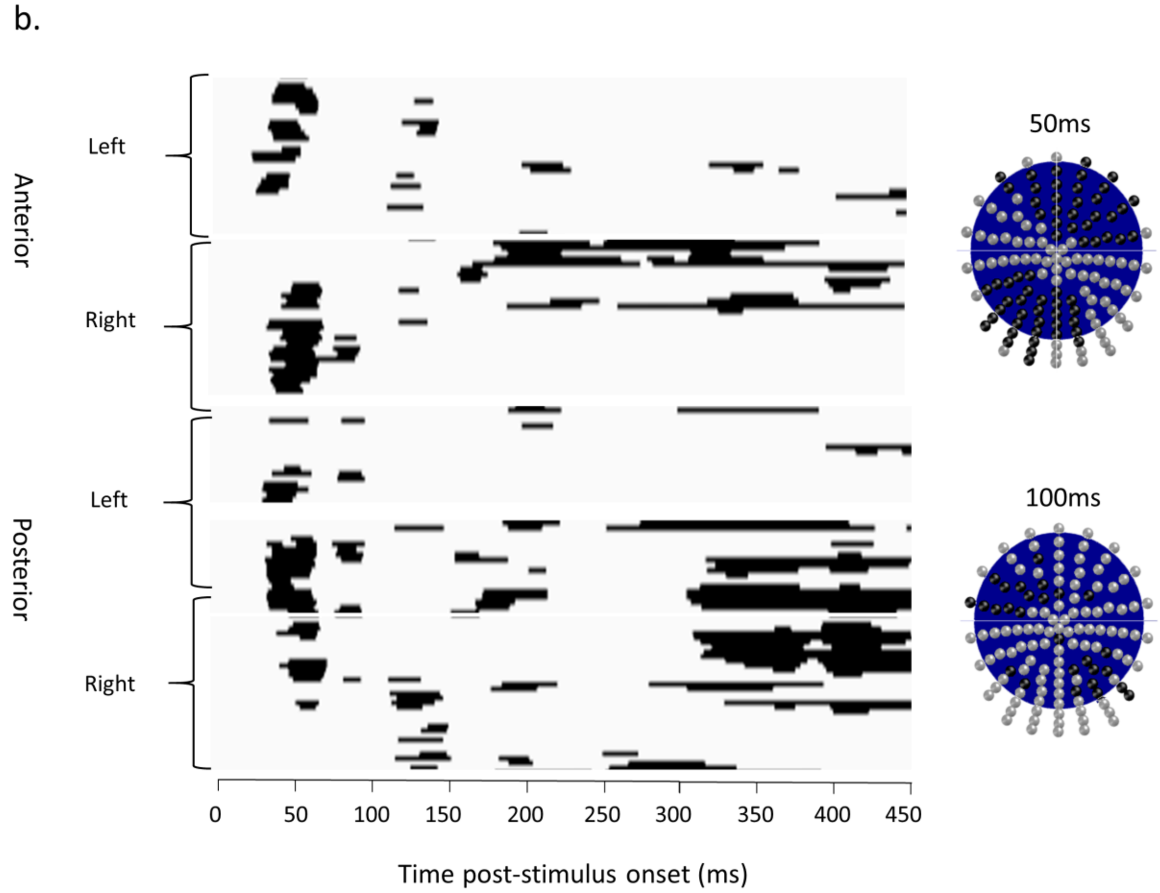


Figure 16. Raster plots of mass univariate contrasts for mono vs. stereo presentation for anterior and posterior left and right hemisphere electrodes (y axis), across time frames from 0-450ms post-stimulus onset (x axis); (a) shows a colour-coded t-map displaying the polarity of contrasts and max/min t values; (b) thresholded plot showing significant pairwise contrasts (p < .01). The electrode montages show the electrodes significant at p<.01 at 50ms (above) and 100ms (below) post-stimulus onset in black.

3.4.2.2 ERP Analyses 2: Perceptual sensitivity to 3D shape similarity as a function of mono/stereo viewing

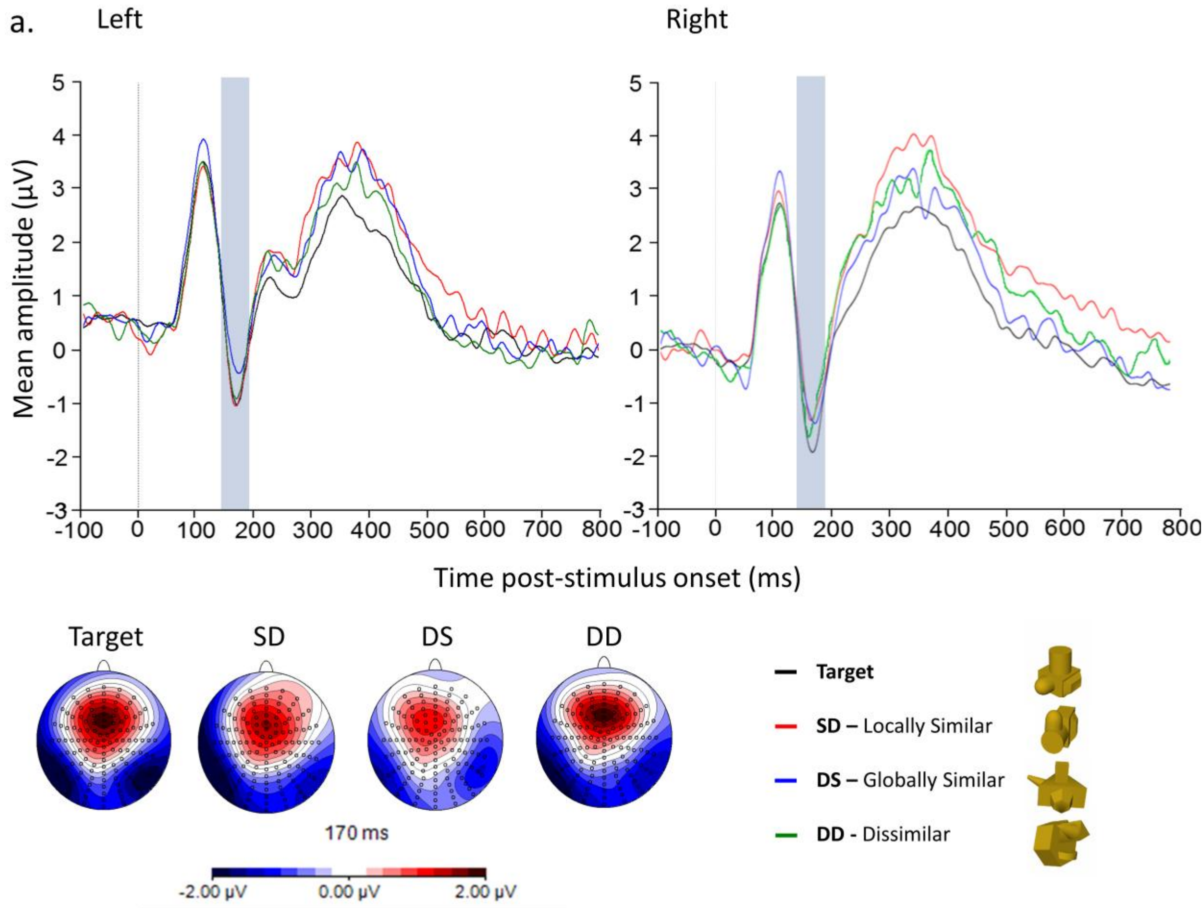
Our next goal was to establish whether perceptual processing of object shape resulted in differential sensitivity to local parts and global 3D shape configuration as a function of mono versus stereo viewing. To do so we conducted both standard waveform analyses and mass univariate contrasts.

3.4.2.2.1 Standard Waveform Analyses

P1. This was defined by a 40ms time window (85-125ms for mono and 90-130ms for stereo). A 4 (Stimulus type: Target; SD (locally-similar); DS (globally-similar); DD (dissimilar)) x 2 (Laterality) x 2 (Display: mono/stereo) mixed ANOVA, with Display as a between subjects factor, showed a main effect of Display, F (1, 30) =5.41, p = .028, with higher amplitudes (μV) for stereo (M=4.03, SD=0.54) than mono viewing (M=2.72, SD=0.15). There was also a main effect of Laterality, F (1, 30) =8.28, p=.007, with greater amplitudes on the right (M=3.63, SD=0.99) than left (M=3.12, SD=0.46) hemisphere electrodes. No other main effects or interactions were significant.

N1. This was defined by a time window of 145-185ms for mono, and 150-190ms for stereo viewing. A 4 (Stimulus type: Target; SD (locally-similar); DS (globally-similar); DD (dissimilar)) x 2 (Laterality) x 2 (Display: mono; stereo) mixed ANOVA, with Display as a between subjects factor, showed a significant three-way interaction, F (2.82, 84.51) =2.98, p=.044. No other main effects or interactions were significant. As can be seen in Figures 17 this interaction derives from the contrasting patterns of amplitude modulation in the SD (locally similar) and DS (globally similar) conditions between mono and stereo viewing. To investigate this further we conducted two separate 4 (Stimulus type: Target; SD (locally-similar); DS (globally-similar); DD (dissimilar)) x 2 (Laterality) repeated measures ANOVAs for the mono and stereo display conditions.

For the mono condition (see Figure 17a), there was a main effect of Stimulus type, F (2.27, 34.05) =3.85, p=.03, driven by a significant difference between the target and DS (globally-similar) non-targets, p=.02, with greater negativity for targets (M=-0.75, SD=0.26) than DS (globally-similar) (M=-0.23, SD=0.25) stimuli. No other main effects or interactions were significant. In contrast, for the stereo condition (see Figure 17b) there was a significant interaction between Stimulus type and Laterality, F (2.76, 41.47) =2.88, p=.046. Post hoc contrasts showed a significant difference between the targets and SD (locally-similar) non-targets in the left hemisphere only, t (15) =2.29, p=.036, with increased negativity for targets (M=-1.39 SD=0.46) compared to SD (locally-similar) (M=-0.97, SD=0.43) stimuli. No other main effects or interactions were significant.



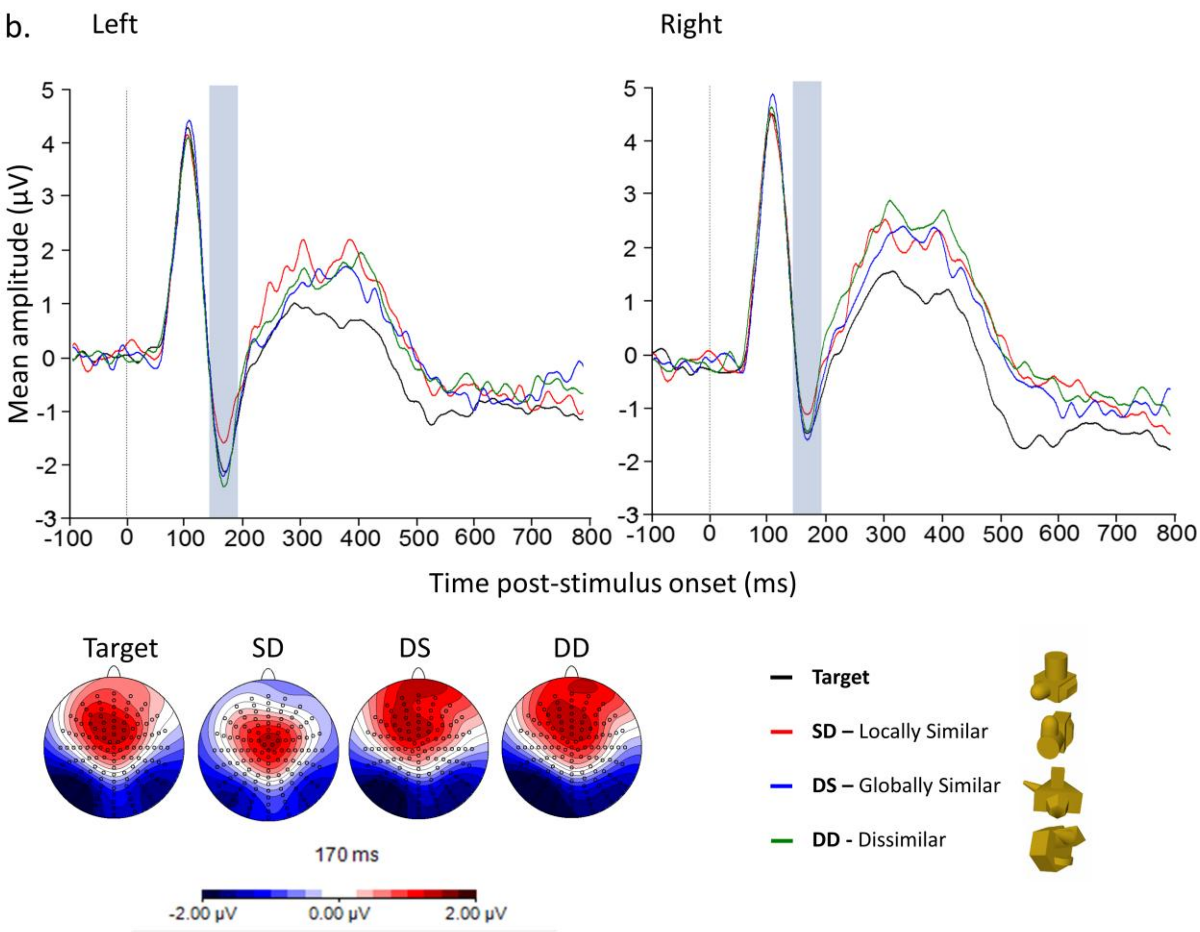
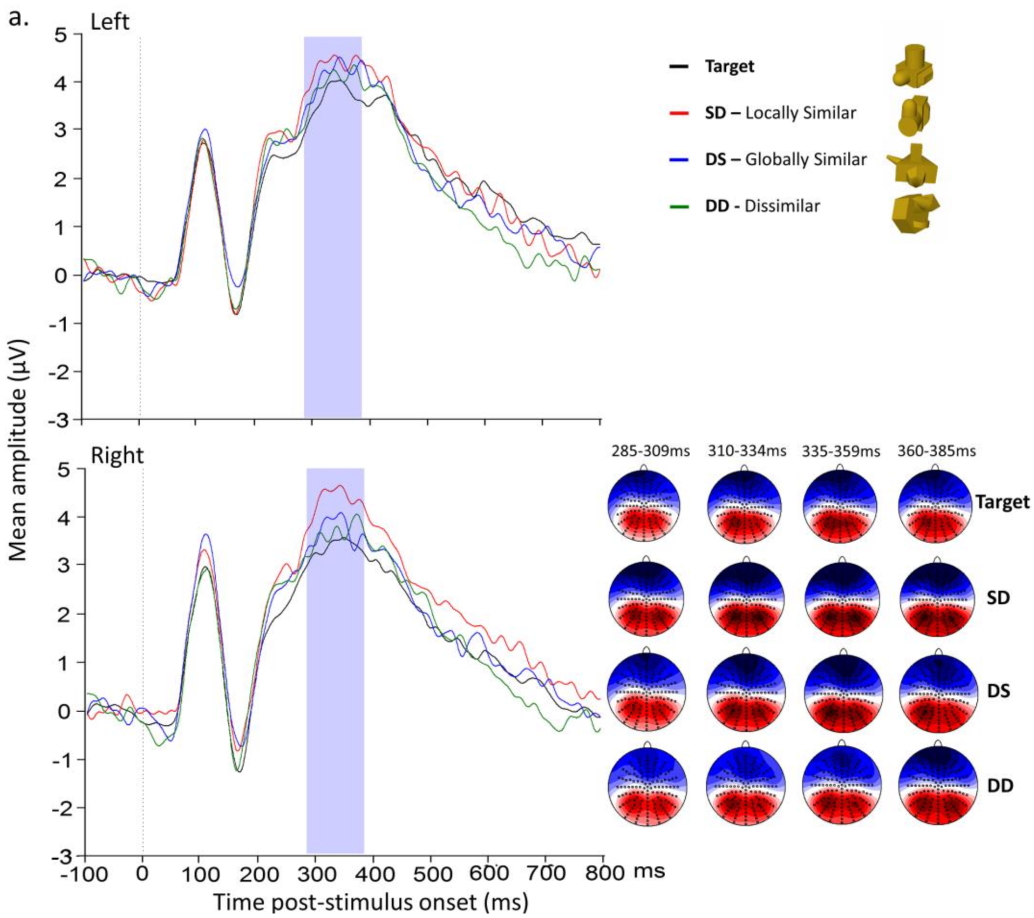


Figure 17. Grand average waveforms for the N1 component (blue highlight) across conditions at the electrode cluster encompassing P7 and PO7 (left hemisphere) and P8 and PO8 (right hemisphere) for (a) Mono and (b) Stereo viewing groups.

P2. This was defined by a time window of 200-240 for mono and 195-235ms for stereo viewing. A 4 (Stimulus type: Target; SD (locally-similar); DS (globally-similar); DD (dissimilar)) x 2 (Laterality) x 2 (Display: mono; stereo) mixed ANOVA, with Display as a between subjects factor showed that no main effects or interactions were significant.

N2-P3 complex The N2-P3 complex was defined by a time window of 285-385ms for mono and 260-360ms for stereo viewing. A 4 (Stimulus type: Target; SD (locally-similar); DS (globally-similar); DD (dissimilar)) x 2 (Laterality) x 2 (Display: mono; stereo) mixed ANOVA, with Display as a between subjects factor showed a significant main effect of Stimulus type, F (2.48, 74.24) =2.97, p=.046. There was also a significant three-way interaction, F (2.82, 84.51) =3.48, p=.022. There were no other significant main effects or interactions. To investigate this further we analysed mono and stereo data separately using 4 (Stimulus type) x 2 (Laterality) repeated measures ANOVAs. For the mono viewing group (Figure 18a) there were no significant main effects or interactions. In contrast, for the stereo viewing group (Figure 18b) there was a significant interaction between Stimulus type and Laterality, F (2.76, 41.47) =4.51, p=.009. Planned comparisons showed that there were no differences between stimulus types in the left hemisphere, but in the right hemisphere mean amplitude for targets was lower than SD (p=.022), DS (p=.024) and DD (p=.002). No other main effects or interactions were significant.


a.
Left
Right
Target
SD – Locally Similar
DS – Globally Similar
DD - Dissimilar
285-309ms
310-334ms
335-359ms
360-385ms
Target
SD
DS
DD
Mean amplitude (μV)
Time post-stimulus onset (ms)
-100
0
100
200
300
400
500
600
700
800 ms

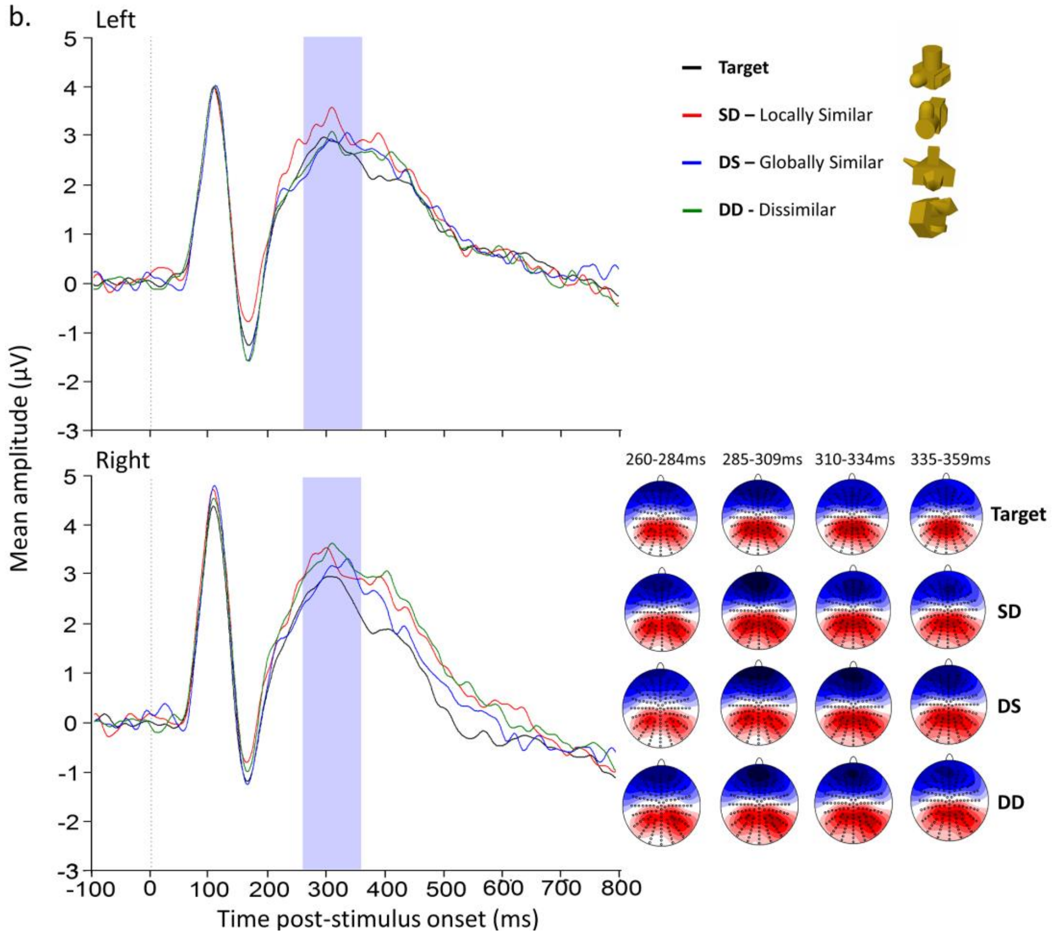
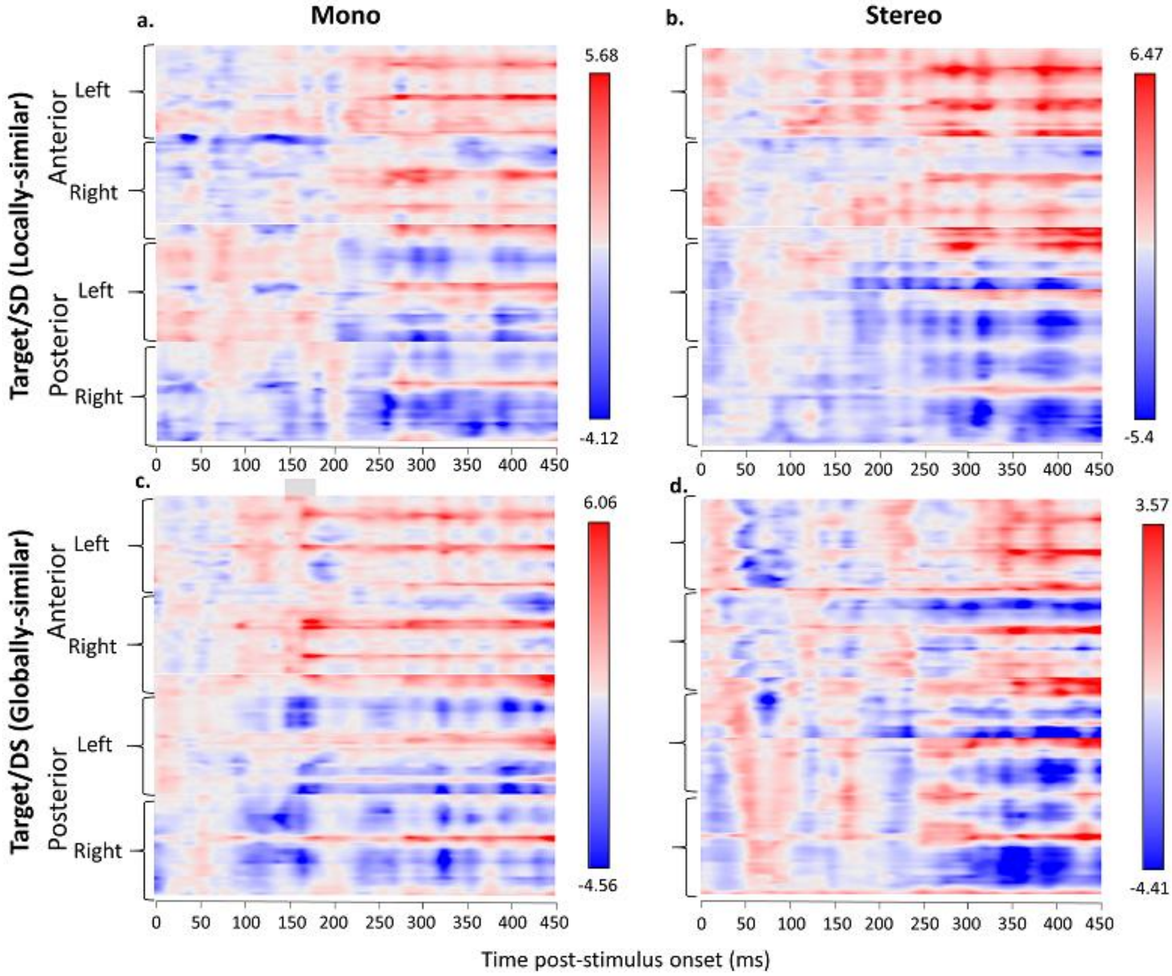


Figure 18. N2-P3 grand average waveforms (highlighted in blue shaded area) for (a) Mono and (b) Stereo viewing groups for all conditions at the electrode clusters encompassing P3 and CP1 (left hemisphere) and P4 and CP2 (right hemisphere).

3.4.2.2.2 Mass Univariate Contrasts across all 128 electrodes

Mass univariate analyses were used to complement our standard waveform analyses of the effects of mono and stereo viewing on the discrimination between targets and critical SD (locally-similar) and DS (globally-similar) non-targets. Unlike the standard analysis, the mass univariate approach allows us to examine the patterns of contrasts between conditions across all 128 electrodes (rather than restricting the analysis to the 9 electrode cluster in each hemisphere). The temporal distributions of these contrasts across all 128 electrodes for mono viewing are shown in Figure 19 a-h.


a.
Mono
b.
Stereo
Target/SD (Locally-similar)
Anterior
Posterior
Left
Right
5.68
-4.12
6.47
-5.4
c.
d.
Target/DS (Globally-similar)
6.06
-4.56
3.57
-4.41
0 50 100 150 200 250 300 350 400 450
Time post-stimulus onset (ms)

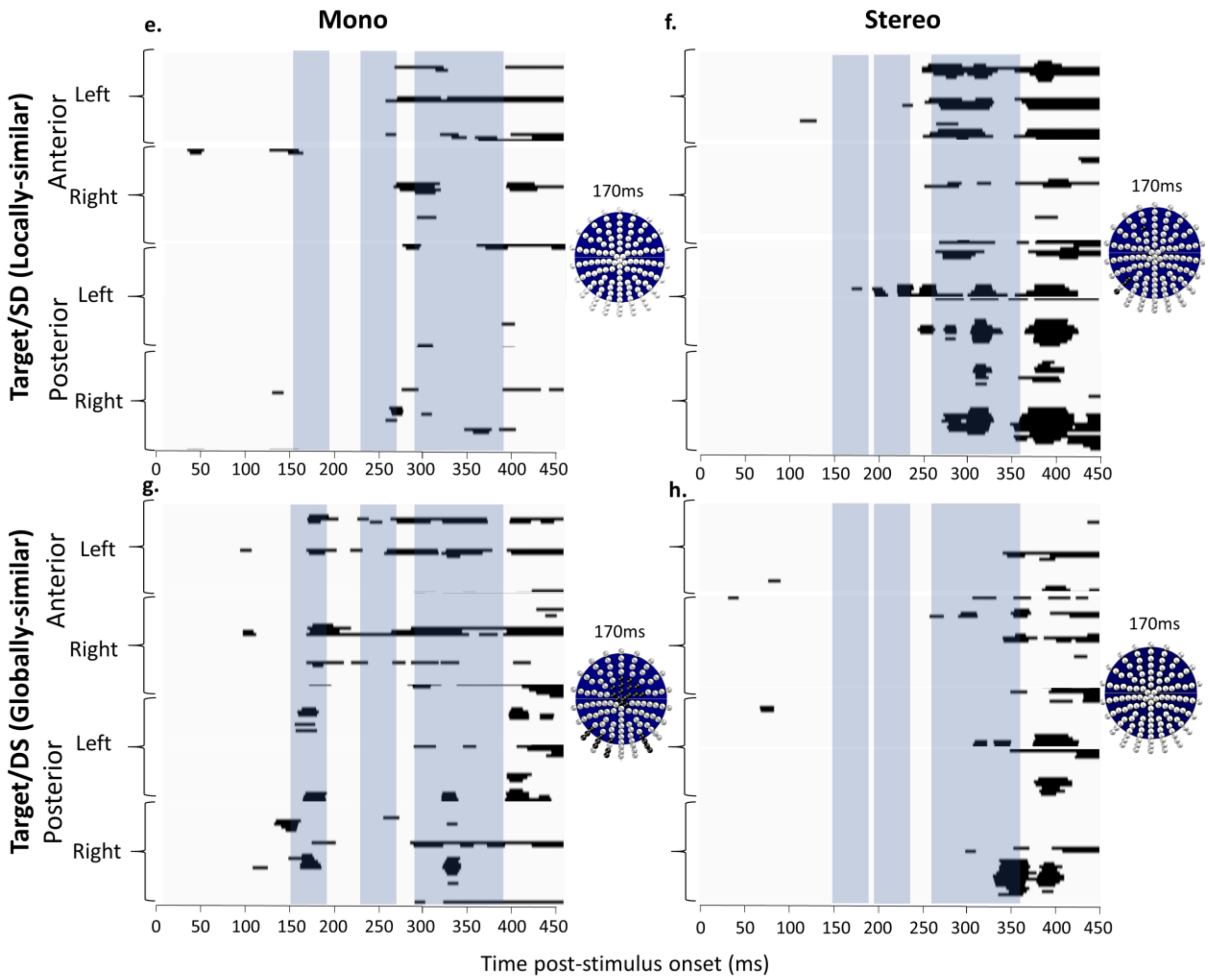


Figure 19. Raster plots of mass univariate contrasts for (a/e) Mono Target-SD (Locally-similar); (b/f) Stereo Target-SD (Locally-similar); (c/g) Mono Target-DS (globally-similar) and (d/h) Stereo Target-DS (globally-similar). Posterior/anterior and right/left electrodes are shown (y axis) across time frames from 0-450ms post-stimulus onset; (a-d) show colour-coded t-maps displaying the polarity of contrasts and max/min t values; (e-h) thresholded plots showing significant pairwise contrasts ($p < .01$). The electrode montages show the electrodes significant at $p<.01$ at 50ms (above) and 100ms (below) post-stimulus onset in black for each contrast. The blue highlighted areas show the N1, P2 and N2/P3 components.

These mass univariate contrasts show the differential sensitivity between targets and SD/DS non-targets for mono and stereo viewing in the N1, P2 and N2/P3 components. A time series plot of the frequency distribution of significant differences is shown in Figure 20. These data were analysed as a non-parametric time-series using the Friedman test. For the N1 during mono viewing there was a higher frequency of significant differences between targets and DS (globally-similar) non-targets in both the left, $\chi^2(1) = 4$, $p = .046$ and right hemispheres, $\chi^2(1) = 5$,

p=.025. For stereo viewing there was a higher frequency of significant differences between targets and SD (locally-similar) non-targets in the left hemisphere only, χ^2 (1) =4, p=.046. The same pattern for stereo viewing was also found during the P2 (χ^2 (1) =4, p=.046), but there was no significant differences for the mono group. The N2/P3 component also showed a striking contrast in perceptual sensitivity to SD (locally-similar) and DS (globally-similar) non-targets between mono and stereo viewing. For mono viewing there was a higher frequency of significant differences between targets and DS (globally-similar) non-targets in the right hemisphere, χ^2 (1) =10, p=.002. The opposite pattern was found for stereo viewing with a higher frequency of significant differences between targets and SD (locally-similar) non-targets in the left hemisphere, χ^2 (1) =6.4 p=.011.

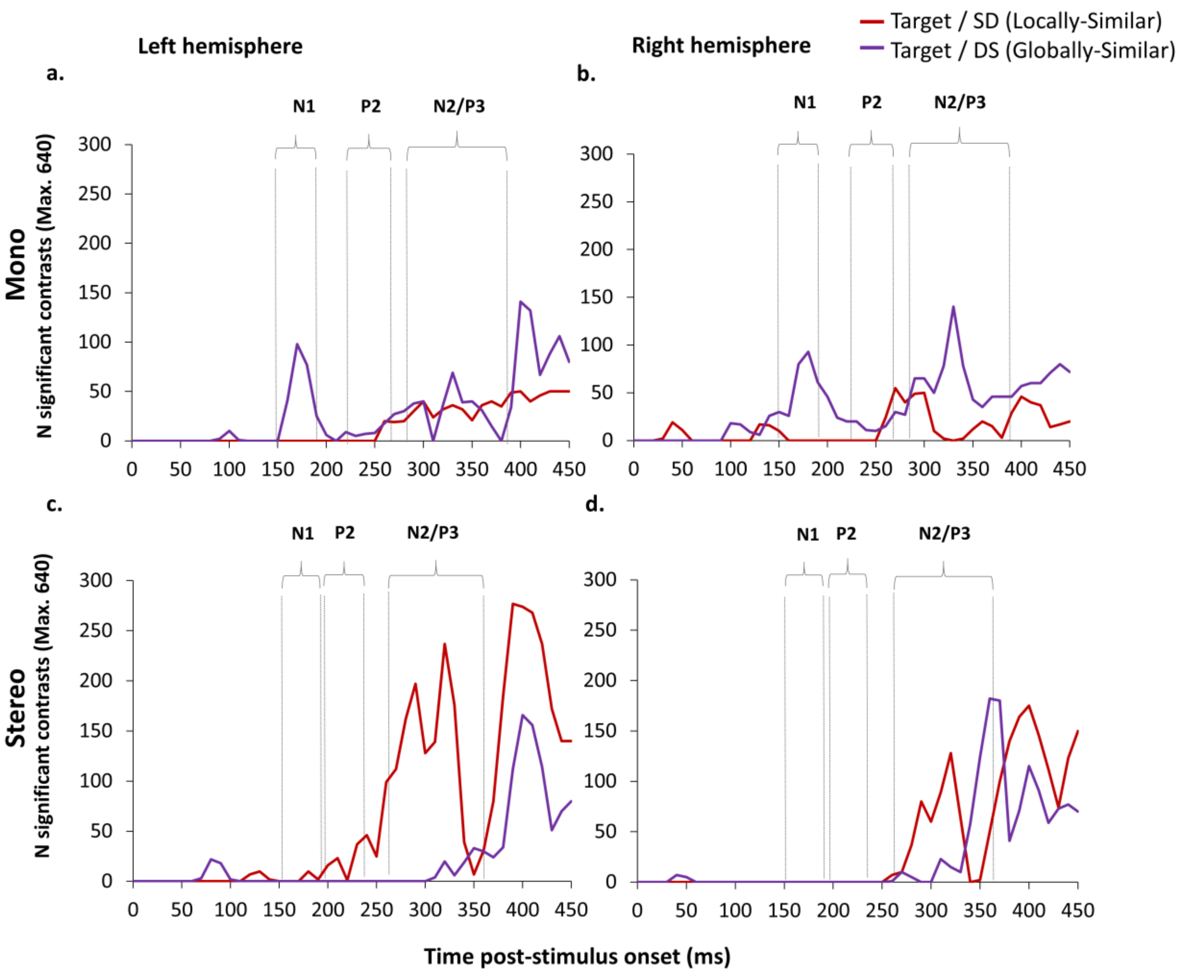


Figure 20. Time series distribution showing the frequency of significant difference contrasts from the mass univariate analysis between 0 and 450ms. Contrasts shown are between Target and SD

(locally-similar) in red and Target and DS (globally-similar) non-targets in purple for both mono in the (a) left and (b) right hemispheres and stereo in the (c) left and (d) right hemispheres.

3.4.2.3 Further analyses II: Effects of training viewpoint

The analyses so far show differential sensitivity to SD (locally-similar) and DS (globally-similar) non-targets between mono and stereo viewing. In brief, during mono viewing there is a greater response modulation to target versus DS (globally-similar) non-targets in both the left and right hemisphere that begins during the N1 and continues into the later N2/P3 component. During stereo viewing, there is a greater response modulation to target versus SD (locally-similar) non-targets that is predominant in the left hemisphere and which begins during the N1 but only peaks during the later N2/P3. In a final analysis, we wanted to examine whether these differential response patterns are modulated by viewpoint familiarity; that is, whether they generalise across image classification at trained and untrained views. Figure 21 shows a time series plot of the frequency distribution of significant differences between target and non-target conditions for trained and untrained viewpoints. The data were analysed as a non-parametric time-series using the Friedman test. For the mono viewing group the higher frequency of significant differences between target and DS (globally-similar) distracters in the left and right hemispheres during the N1 was found for trained viewpoints but did not generalise to untrained viewpoints (LH: $\chi^2(1)=4$, p=.046, RH: $\chi^2(1)=4$, p=.046). In contrast, for the stereo viewing group, there were no differences between trained and untrained viewpoints at the N1. For the mono group at the N2/P3, however, there was a higher frequency of significant differences between target and SD (locally-similar) distracters for trained than untrained viewpoints in the left hemisphere ($\chi^2(1)=6.4$, p=.011). There was also a higher frequency of differences between target and DS distracters in the left and right hemispheres for trained than untrained viewpoints (LH: $\chi^2(1)=10$, p=.002; RH: $\chi^2(1)=10$, p=.002). For the stereo group, there was a higher frequency of significant differences between target and SD (locally-similar) distracters for trained than untrained viewpoints in the left hemisphere ($\chi^2(1)=6.4$, p=.011) and a higher frequency of differences between target and DS (globally-similar) distracters for trained than untrained viewpoints in the right hemisphere (χ^2 (1) =6.4, p=.011).

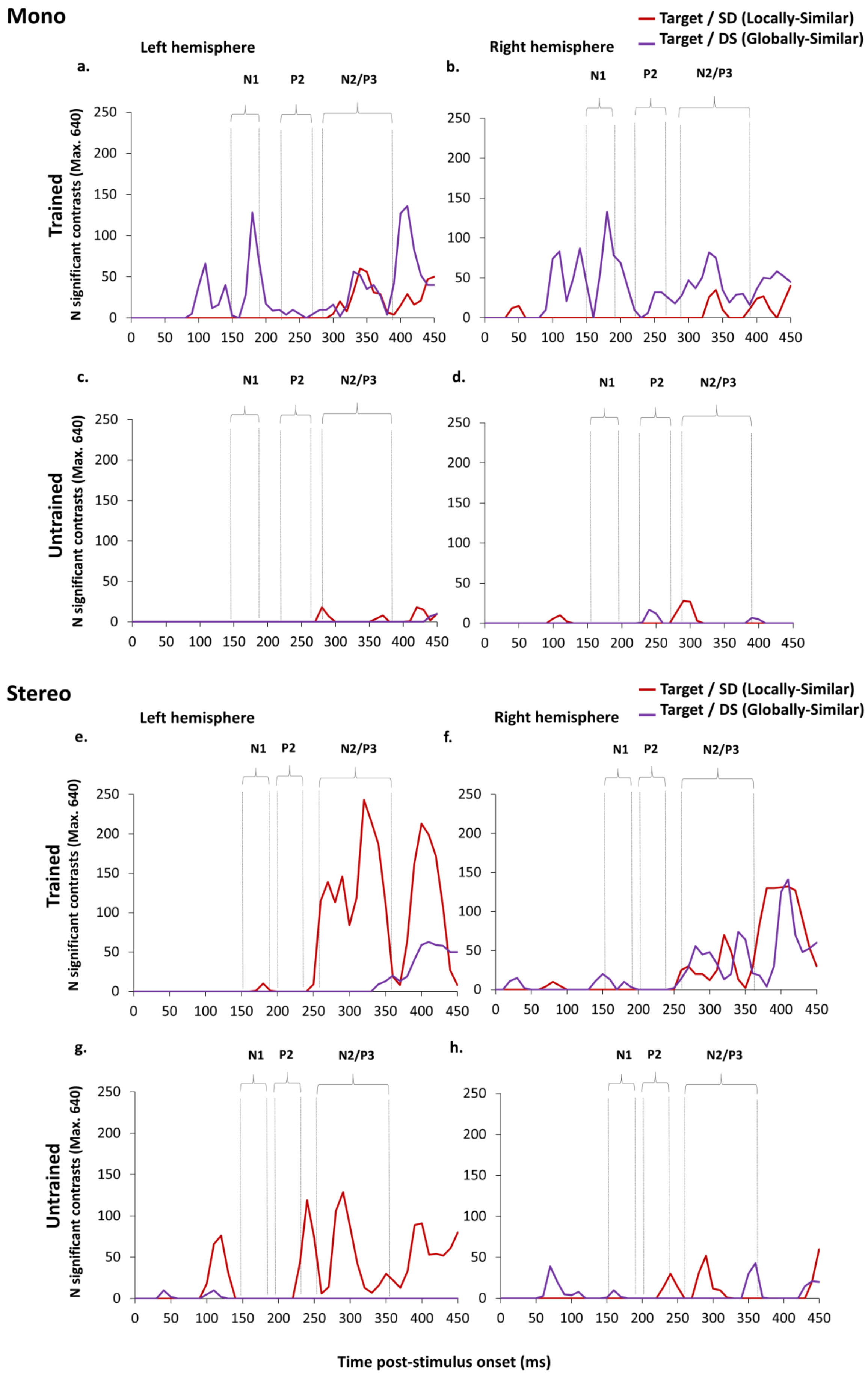

Mono
Target / SD (Locally-Similar)
Target / DS (Globally-Similar)
Left hemisphere
Right hemisphere
a.
b.
c.
d.
N1
P2
N2/P3
Trained
Untrained
N significant contrasts (Max. 640)
0
50
100
150
200
250
300
350
400
450
Stereo
e.
f.
g.
h.
Time post-stimulus onset (ms)

Figure 21. Time series distribution showing the frequency of significant difference contrasts from the mass univariate analysis between 0 and 450ms. Contrasts shown are between target and SD (locally similar) in red and target and DS (globally-similar) non-targets in purple for mono (a-d)/stereo (e-h) viewing, left and right hemispheres and trained versus untrained views.

3.5 Discussion

The main findings can be summarised as follows: First, the behavioural data provided evidence for an advantage in view generalisation for stereo over mono displays. This was shown by higher accuracy in target classification of untrained views for stereo displays. Second, the ERP data showed differential amplitude responses to mono versus stereo viewing as early as 50-100ms post-stimulus onset, with higher amplitudes on the P1 component for stereo displays. Third, we observed differential amplitude modulations of evoked potentials to targets and non-targets defined by shared parts (SD; locally-similar) or shared spatial configuration (DS; globally-similar) starting at the N1 component between 145-200ms post-stimulus onset. N1 amplitudes for mono displays showed greater differential sensitivity to DS (globally-similar) non-targets. For stereo displays, there was a greater differential amplitude modulation for SD (locally-similar) non-targets in left hemisphere electrodes. Fourth, a pattern of differential amplitude modulation was also found at the later N2/P3 component around 260-385ms post-stimulus onset. This was most clearly shown in the mass univariate analysis. For mono viewing, there was a higher frequency of significant differences between targets and DS (globally-similar) non-targets. For stereo viewing, there was a higher frequency of significant differences between targets and SD (locally-similar) non-targets. Fifth, under mono viewing, the differential sensitivity to DS (globally-similar) non-targets was found for trained but not untrained views. In contrast, the amplitude sensitivity in stereo viewing to SD (locally-similar) non-targets was found with both trained and untrained views.

These new empirical findings have several important implications for models of object recognition. First, the results provide new evidence that the representation of complex 3D object shape involves the specification of higher-order part structure and 3D part configuration. This is shown by the differential sensitivity in the ERPs to shape differences between targets and non-targets defined by either shared local parts or 3D shape configuration. These differences emerged during the N1 component between approximately 145-200ms post-stimulus onset, and were also found during the N2/P3 component around 260-385ms post-stimulus onset. This finding is consistent with theoretical models, and other supporting empirical evidence, that the perceptual representation of complex 3D object shape involves the specification of higher-order part structure and global 3D spatial configuration (e.g., Arguin & Saumier, 2004; Behrmann, et al., 2006; Behrmann & Kimchi, 2003; Biederman, 1987; Hummel & Stankiewicz, 1996; Marr & Nishihara, 1978). The results challenge theoretical models which do not attribute functional significance to these properties of object shape representations - including the hierarchical, feed-forward HMAX deep (i.e., multi-layer) network architecture (e.g., Riesenhuber & Poggio, 1999; Serre et al., 2007), and others (e.g., Bulthoff & Edelman, 1992; Chan et al., 2006; Khaligh-Razavi & Kriegeskorte, 2014; Krizhevsky et al., 2012; Li & Pizlo, 2011; Li et al., 2009; Pizlo, 2008).

Second, the results also provide new evidence that the recognition of complex 3D object shape can be modulated by stereo visual input. This was shown in both the behavioural and ERP data patterns. Behaviourally, we found an advantage for object recognition under conditions of stereo viewing in relation to classification accuracy for targets presented at previously untrained views. This observation adds to a growing body of behavioural evidence that stereo input can facilitate 3D object recognition - at least under some conditions (e.g., Bennett & Vuong, 2006; Burke, 2005; Burke et al., 2007; Chan, et al., 2006; Edelman & Bülthoff, 1990; Hong Liu et al., 2006; Lee & Saunders, 2011; Rock & DiVita, 1987; Simons et al., 2002). According to Cristino et al. (2015), stereo input provides additional cues to 3D object shape including, for example, the specification of surface slant, curvature polarity and 3D part configuration. We also found differential modulation of ERP amplitudes during mono and stereo viewing as a function of target/non-target shape similarity. Notably, we found evidence for differential modulation of ERP amplitudes under mono and stereo viewing for DS (globally-similar) and SD (locally-similar) distractors. This shows that stereo viewing can modulate perceptual processing of different attributes of 3D shape - contrary to the predictions of theoretical models that do not attribute functional significance to stereo information in the derivation of 3D object representations (e.g., Bulthoff & Edelman, 1992; Chan et al., 2006; Li & Pizlo, 2011; Li et al., 2009; Pizlo, 2008; Reisenhuber & Poggio, 1999; Serre et al., 2007). One interpretation of the results is that stereo viewing enhances processing of information about the 3D spatial configuration of object parts, and that this information facilitates the classification of SD (locally-similar) distracters as non-targets on the basis of their distinct global 3D spatial configuration. In contrast, under conditions of mono viewing, we found early differential sensitivity to DS (globally-similar) distracters that shared spatial configuration but not local parts (that is, where targets and distractors can be differentiated on the basis of distinct local parts). This raises the possibility that, in the absence of stereo input (as is the case in most previous empirical studies of object processing), the perceptual analysis of 3D object shape is weighted towards differences in 2D local shape attributes. Furthermore, the enhanced processing of local part structure did not generalize to untrained views, suggesting that under monocular viewing conditions object shape processing may be weighted towards an 'image-based' processing strategy. Taken together, these findings suggest that mental representations of 3D object shape in human vision are rich in structure, encoding both 2D image-based local features, and 3D shape properties, broadly consistent with a 'hybrid' approach to object recognition mediated by representations combining both 2D and 3D object structure (Foster & Gilson, 2002; Hummel, 2013; Hummel & Stankiewicz, 1996)[4].

A recent study by Leek et al. (2016), using a sequential novel object matching task under conditions of mono viewing only, also reported early differential perceptual sensitivity to shape differences defined by either shared parts or global spatial configuration. In that work, differential sensitivity in perceptual matching of novel 3D objects was – as in the current study,

[4] In relation to the HMAX hypothesis in particular, it is of interest to note that in terms of image similarity, we also found lower mean (normalised) HMAX target-distractor similarity values for trained views. This could potentially have also contributed to the differential sensitivity of ERP amplitudes to DS and SD non-targets found for mono viewing consistent with an image-based processing strategy. However, this would not account for the why the opposite pattern of amplitude modulation was found with stereo input.

found to emerge earliest on amplitude modulations during the N1 component over posterior electrodes between objects sharing either local parts or global spatial configuration. The current data extend these findings in several important ways. First, we have shown that this differential perceptual sensitivity extends to an object recognition task where observers are required to match a perceptual description of 3D object shape to a (previously learned) long-term memory representation. Second, the results also show that this differential perceptual sensitivity is modulated by mono versus stereo input – in which mono viewing enhances local differences in part structure, while stereo viewing enhances differences in global 3D spatial configuration. Third, we also found that this stereo viewing effect generalises across changes in 3D object viewpoint, whereas perceptual sensitivity to local differences in part structure found under conditions of mono viewing were restricted to trained viewpoints.

An additional important issue arises from our observation of early perceptual sensitivity of ERPs to shape similarity between targets and distracters on the N1 component. This implies that some properties of the shapes of unfamiliar 3D objects can modulate perceptual processing prior to recognition (Bar, 2003; Bar et al., 2006; Leek et al., 2016). One interpretation of this effect is that the early perceptual modulation reflects partial activation of stored (i.e., target) shape representations on the basis (in this case) of parts-based object descriptions. More broadly, this hypothesis is consistent with a conception of object shape processing that is based on parallel analyses of shape across multiple spatial scales (e.g., Bar, 2003; Bar et al., 2006; Hedge, 2008; Heinz, Johannes, Münte & Mangun, 1994; Heinz, Hinrichs, Scholz, Burchert & Mangun, 1998; Navon, 1977; Peyrin, Chauvin, Chokron & Marendaz, 2003; Peyrin, Baciu, Segebarth & Marendaz, 2004; Peyrin et al., 2010).

Finally, one other issue merits brief discussion. Although our primary goal was to examine whether mono versus stereo visual input differentially modulates the perceptual processing of 3D object shape during recognition, we also observed an early perceptual sensitivity, and lateral asymmetry, to stereo disparity. We found the earliest differential responses to mono versus stereo input from around 50ms post-stimulus onset over a large group of posterior, temporal-occipital and anterior leads. This difference was sustained during the P1 component over left occipital and some frontal electrodes. Additionally, we also found greater P1 amplitudes for right over left hemisphere electrode sites. We have taken this to reflect early perceptual sensitivity to mono- versus stereo input in our design. One might argue that these differences do not reflect the resolution of stereo disparity per se, but rather sensitivity to the presentation of different images to the left and right eye in the stereo condition. However, if this were the case, we would expect to find differences between mono- and stereo presentation in all conditions regardless of target-distracter similarity. The observed interactions between stimulus type and viewing condition show that this was not the case.

In summary, we investigated whether stereo viewing modulates perceptual processing of 3D object shape. A recognition memory task was used in which observers were trained to recognise a sub-set of 3D novel objects under conditions of either mono or stereo viewing. In a subsequent test phase, they discriminated trained objects from non-targets that shared either local parts, 3D spatial configuration or neither dimension, across both previously trained and novel viewpoints. The behavioural data showed a stereo advantage for generalisation between trained and untrained views. ERPs amplitudes also showed early differential sensitivity to local part, and 3D spatial configuration, similarity between targets and distracters. This occurred during an N1 component from 145-200ms post-stimulus onset and during an N2/P3 component from 260-385ms post-stimulus onset. For mono viewing, amplitude modulation during the N1 was greatest between targets and distracters with different local parts for trained views only. For stereo viewing, amplitude modulation during the N2/P3 was greatest between targets and distracters with different global 3D spatial configurations and generalised across trained and untrained views. The results show that image classification is modulated by stereo information about the local part, and global 3D spatial configuration of object shape. The findings challenge current theoretical models that do not attribute functional significance to stereo input during the computation of 3D object shape.

3.6 Summary

- This study examined the time course of local and global information processing for object recognition in mono and stereo viewing conditions.
- We used a training paradigm with subsequent object recognition task whilst recording ERPs. Groups learned objects and performed the recognition task in either mono or stereo viewing conditions.
- The main results showed that local and global processing of object shape are processed differentially at the N1 and later N2/P3 and local and global processing are modulated by stereo/mono viewing.
- We conclude that stereo disparity is important for object recognition.

4 Chapter IV

Global interference effects using an orientation detection task: A high-density ERP study

Previously, in Chapter III, we reported differential ERP sensitivity to shape processing at global and local spatial scales. In Chapter IV we wanted to elucidate the processes involved in the perceptual integration of these signals. More specifically, we wanted to find out whether there is an ERP signature for the integration of global and local information. A classic paradigm for investigating integration of information at local and global spatial scales is the use of hierarchical displays in the Navon task. The task typically involves large letters (the global level) composed of smaller letters (the local level), with identification of the letter at either the local or global level. The hierarchical letters can include the same large and small letters (congruent display) or small letters different to the large letter (incongruent). Response to the global level is typically faster than to the local level. Also, reaction times are slowed when responding to the local level in incongruent trials due to the presence of a different letter at the global level; global-to-local interference. Navon described these two phenomena as the global precedence effect (GPE). The GPE has been reported in many experiments including several ERP studies (e.g., Beaucousin et al, 2013; Han et al., 2000; Proverbio et al., 1998; Yamaguchi et al., 2000).

A limitation, however, of most previous studies is the use of mixed object displays, such as Navon letters. These hierarchical stimuli place atypical processing demands on the perceptual system as we rarely encounter these sorts of global/local conflicts in object recognition. Also, the use of letter stimuli will be influenced by prior knowledge. Our aim was to find an ERP signature for local/global integration using simple displays of orientated Gabor elements. The rationale is that we can use congruency effects as a functional marker for integration, as congruency effects presumably arise at the point of global/local integration.

4.1 Abstract

The aim was to find a temporal-spatial EEG marker for local and global integration. As the congruency effect presumably arises at a level of perceptual processing where information from global and local channels are integrated. We used very low-level visual stimuli with manipulation of spacing and contrast and a simple orientation detection task with arrays of Gabor patches orientated either to the left or right whilst ERPs were recorded. We found that reaction times for congruent trials were faster than for incongruent trials and error rate was greater for incongruent trials. However, no difference in interference levels between local report and global report trials was observed. Global interference effects were, however, clear in the ERPs, with greater frequencies of congruency differences in local compared to global report conditions. We found evidence of global interference in the ERP data from around 350ms post-stimulus onset. We suggest that this is evidence of global/local integration.

4.2 Introduction

David Navon was the first to investigate local and global processing with hierarchical stimuli (Navon, 1977), now known as Navon letters. He proposed that local and global information is processed differently. After a series of experiments, Navon (1977) observed two phenomena related to processing the local and global elements of stimuli: global information is processed faster than local information; and global information can be voluntarily attended to without being affected by local features, but local features cannot be attended to without processing the global information (global interference). Together, these phenomena are known as the global precedence effect (GPE).

The GPE led Navon to propose that perception proceeds from a global analysis to a more fine-grained analysis, and that global processing seems to be a necessary stage of perception as we are unable to ignore the global level when processing local elements of a scene. Many studies have provided support for GPE (e.g. Beaucousin, Simon, Cassotti, Pineau, Houde & Poirel, 2013; Han, He & Woods, 2000; Proverbio, Minniti & Zani, 1998; Yamaguchi, Yamagata & Kobayashi, 2000). Poirel, Pineau and Mellet (2008) found partial evidence for GPE regardless of the meaningfulness of the stimuli; the global level was always processed faster than the local level. However, the interference effect occurred only for meaningful stimuli; objects rather than non-objects. They suggested that global precedence is a sensory mechanism and therefore automatic for all stimuli, whereas global interference reflects a cognitive mechanism, which is related to identification.

Hemispheric differences are often found in processing of hierarchical stimuli, whereby the right hemisphere and left hemisphere are biased for global and local processing of images, respectively (Delis, Robertson & Efron, 1986; Lamb, Robertson & Knight, 1989; 1990; Robertson & Lamb, 1991; Robertson, Lamb & Knight, 1988; Van Kleek, 1989). There is frequent report of hemispheric asymmetry for local and global processing in EEG studies (Heinz et al., 1998; Leek et al., 2016; Mangun et al., 2000; Volberg & Hübner, 2004; Yamaguchi, Yamagata & Kobayashi, 2000), and from imaging studies (Han, Weaver, Murray, Kang, Yund & Woods, 2002), with greater left hemisphere activity for local (finer spatial scale) processing, and greater right hemisphere activity for global information (at a coarser spatial scale). A similar pattern has been reported in patient populations, where during copy-drawing, patients with left hemisphere damage are impaired in reporting the local level of Navon-type letters, with relatively intact report of the global level, and vice versa for those with right hemisphere damage (Robertson, Lamb & Knight, 1988; Robertson & Lamb, 1991).

These hemispheric asymmetries occur more often with centrally presented stimuli than with laterally presented ones in ERP studies (e.g. Han, Weaver, Murray, Kang, Yund & Woods, 2002; Volberg & Hübner, 2004). Volberg and Hübner (2004) investigated whether the central presentation or the response conflict (incongruent local and global levels in stimuli) was more important in finding hemispheric asymmetry in an ERP study with bilateral displays, they observed hemispheric differences that were more pronounced for incongruent stimuli, suggesting that conflicting stimuli are more important in observing hemispheric differences than central presentation of stimuli. Volberg and Hübner (2008) conducted a further study to disseminate possible effects of task difficulty and hemispheric differences in local and global processing; they found that increased difficulty in the incongruent condition of the task did not account for hemispheric asymmetries.

Also, studies using bilateral stimulus presentation have found that reaction times are faster for stimuli presented in the left than the right visual field (VF) for global targets (Flevaris, Bentin & Robertson, 2010; Schlösser, Hübner & Studer, 2009; Van Kleeck, 1989; Volberg & Hübner, 2006). This suggests that global shape information is processed faster when projected to the right cerebral hemisphere. Furthermore, stimuli presented in the left side of space provide a benefit for global processing and the same for the right side of space and local processing (Christie et al., 2012; Yovel, 2001) and the same was true for the left and right side of objects (Christie et al., 2012). There is also evidence of specialisation of upper and lower VFs for processing local and global information, respectively (e.g. Christman, 1993; Previc, 1990).

GPE is less likely to occur with sparse than dense elements (Kimchi, 1985; Martin, 1979; Yovel, Yovel & Levi, 2001). Presumably this is because the sparse spacing breaks up the global form, therefore, densely spaced displays should make the global elements more prominent. Several studies also support the idea that local and global processing are carried out based on high and low spatial frequency information, with early global processing effects diminishing when low spatial frequencies are removed (Boeschoten, Kemner, Kenemans & Engeland, 2005; Han et al., 2003; Hughes, Fendrich & Reuter-Lorenz, 1990; Jiang & Han, 2005). This is presumably because global processing preferentially uses low spatial frequency information. Based on this, our stimulus displays included either sparsely-spaced, high-contrast local elements or densely-spaced, low-contrast local elements to maximise local and global processing, respectively.

Differences in local and global processing have been reported to occur as early as the P1 (e.g. Jiang & Han, 2005); and the N1 (e.g. Beaucousin, Simon, Cassotti, Pineau, Houde & Poirel, 2013; Han et al., 2003); however, many studies find that first differences between local and global appear at or after the N2, from about 250ms after stimulus onset (e.g. Han, He & Woods, 2000; Heinze et al., 1998; Heinze & Münte, 1993; Malinowksi et al., 2002; Volberg & Hübner, 2004; Yamagucchi, Yamagara & Kobayashi, 2000). Heinze and Münte (1993) claimed that the N250 is an index of global and local target perception, whereas the subsequent P3 reflects a later stage of target classification. Therefore, the ERP correlates of global and local processing appear to be heavily dependent on the task and stimuli used. The displays used in the current experiment are very low-level, comprising arrays of Gabor patches. An advantage of these low-level stimuli is that they should avoid the possibility of producing highly lateralised waveforms that letter stimuli, often used in Navon-type experiments, might as we were interested in seeing if we could see any hemispheric effects specific to global and local processing. Therefore, they may provide a more bias-free measure of local and global processing differences. They are also free from any semantic ties, so there is no top-down influence as in the letter or object stimuli that are often used.

In Chapter III we found evidence for the distinct perceptual processing of shape information at global and local spatial scales during 3D object recognition. This raises the question of when this information in integrated during online perceptual processing of object shape during the computation of an integrated, geometrically coherent, 3D object shape representation. Our aim was to find an ERP signature for local/global integration using simple displays of orientated Gabor elements. We used low-level visual stimuli, with manipulation of spacing and contrast. A pilot study using the same stimuli, but without ERP recording, showed that the stimuli were sufficient to elicit an interference effect (see Appendix 1). The rationale is that we can use congruency effects as a functional marker for integration, as congruency effects presumably arise at the point of global/local integration.

4.3 Methods

4.3.1 Participants

20 Bangor University students (10 female, mean age 20.9, *SD*=2.61, 1 left-handed) participated for course credit. The sample was recruited through an online participation portal. All participants had normal or corrected-to-normal visual acuity. Ethics approval was granted by Bangor University. Informed consent was obtained and participants were free to withdraw from the study at any time without penalty.

4.3.2 Stimuli

Stimuli used were displays made up of Gabor patches made in Matlab (Psychtoolbox). An array of either 16 (4x4) or 36 (6x6) Gabor patches was displayed in one of 4 quadrants of the screen: upper left, upper right, lower left and lower right. The critical area of the array of

Gabor patches was a diagonal line at either -45 or 45 degrees, which comprised patches with orientations (local orientation) that were congruent with or incongruent with the orientation of the diagonal (global orientation). The Gabor patches surrounding the diagonal line were randomly at 0° or 90°, see Figure 22.

There were 4 different types of displays: Gabor patches orientated at 45 degrees and global orientation of line at 45 degrees (congruent) and the same with -45 degrees; Gabors at 45° but global orientation at -45° (incongruent) and vice versa. Individual Gabor patches were 0.5 degree in visual angle, with the display within 5 degrees of visual angle. We included catch trials (10%), in which the orientation of the Gabor patches in the critical diagonal line were not all the same. We manipulated the spacing and the contrast of the displays, there could be 4x4 arrays (sparse spacing) with high contrast local elements (Michelson contrast = 80%) or 6x6 (dense spacing) with low contrast (Michelson contrast = 40%) local elements, see Figure 23 for all conditions.

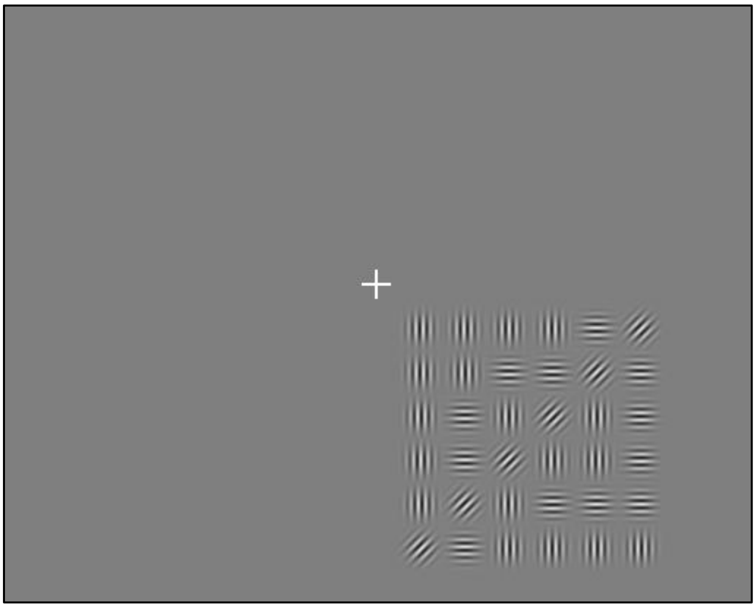


Figure 22. An example of a trial where the array appears in the lower right quadrant, the 'global' orientation is 45° (the orientation of the diagonal line), as is the 'local' orientation (the orientation of the Gabor patches making up the diagonal), therefore this is a congruent trial. The display is densely spaced and high contrast.

The stimuli were designed to maximally elicit GPE, studies have shown that the emergence of global precedence depends on a number of factors, GPE is elicited when: visual angle is less than 7-10 degrees (Kinchla & Wolfe, 1979; Lamb & Robertson, 1990); when participants don't know where the stimuli will appear (Lamb & Robertson, 1988); when peripheral displays are used (e.g. Grice, Canham & Boroughs, 1983; Pomerantz, 1983; Yund et al., 2002b), though this is disputed (Luna, Merino & Marcos-Ruiz, 1990; Navon & Norman, 1983); and when displays include dense rather than sparse elements (Kimchi, 1985; Martin, 1979; Yovel, Yovel & Levi, 2001).

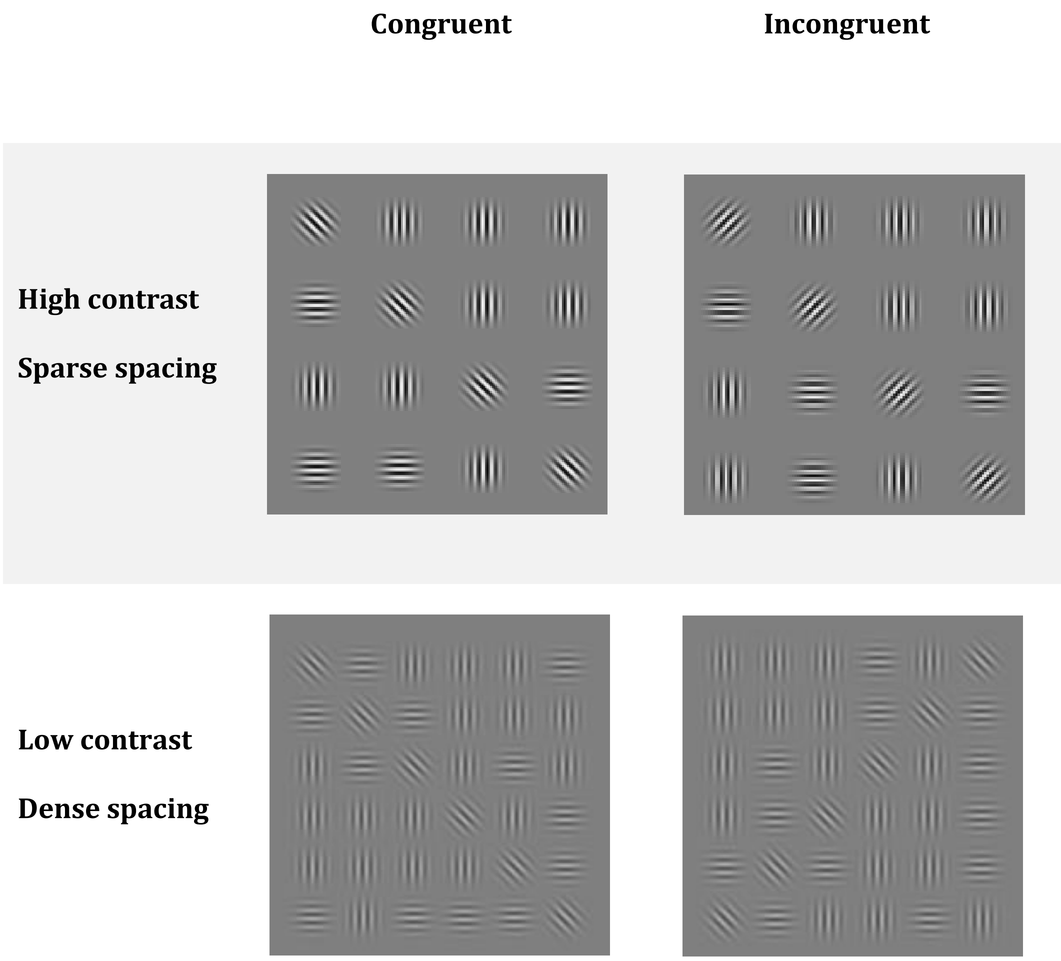


Figure 23. Illustration of Gabor patch arrays for both conditions: high contrast with sparse spacing (locally-weighted) and low contrast with dense spacing (globally-weighted) for both congruent and incongruent trials.

4.3.3 Apparatus and materials

Stimulus presentation and data collection were performed using Matlab. Stimuli were presented on a 27" AOC 3D monitor (D2769VH), at a resolution of 1920x1080. A chin rest was used to stabilise the participant's head at 60cm viewing distance, and a standard keyboard was used for response collection. An SR Research Eyelink 1000 desktop corneal-reflection eye tracker, sampling at 1000Hz, recorded the right eye during the experiment.

4.3.4 Design

A 2(report level: global, local) x 2(condition: LC Dense/HC Sparse) x 4(visual field: upper left, upper right, lower left, lower right) within subject design was used. There was a total of 32 conditions, each was repeated 20 times, for a total of 640 trials. 64 catch trials were also included (2 for each of the conditions in 'report Global' only), for a total of 704 trials.

4.3.5 Procedure

ERPs were recorded while participants performed a decision task. Also, eye tracking was used, so before the experiment began, calibration was performed. At the start of each trial, participants were told which level of the stimulus they were to attend to (local or global) and were required to press a button indicating the orientation of the relevant level. A small central fixation cross appeared on the screen, the trial did not start until participants maintained central fixation for 500ms, then the test image stayed on the screen until the participant made a response, see Figure 24. Participants were asked to press a button on a keyboard indicating whether the orientation was left or right, "z" for left, "m" for right, or if there was no congruent diagonal line ("spacebar") and reaction time (RT) was recorded from stimulus onset. Central fixation was monitored throughout, if the participant's gaze moved more than 30 pixels away from the centre of the fixation cross, the stimulus disappeared and the fixation cross was replaced by a red cross. If participants lost fixation, the trial started again. Participants had the opportunity to take a break roughly every 15 minutes (after 176 trials). 20 practice trials were performed before the experiment began to ensure that participants understood the task.

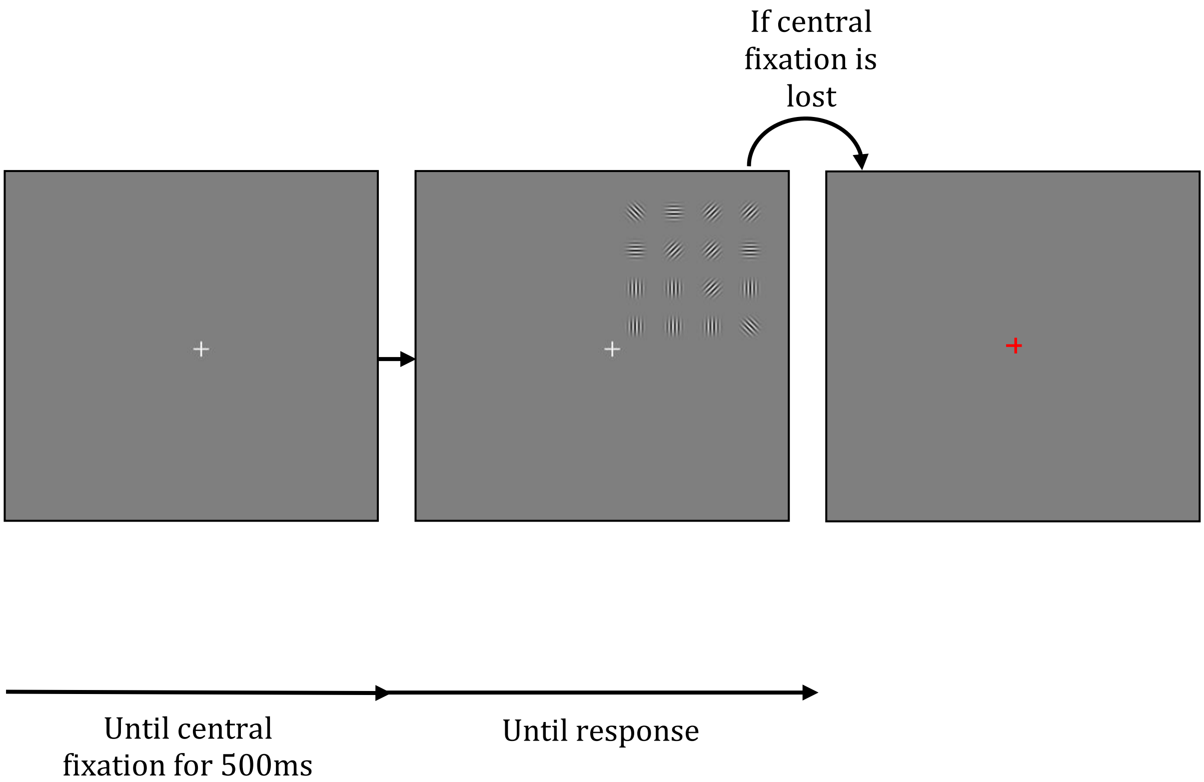


Figure 24. Trial procedure, with central fixation until participant fixates for 500ms, then Gabor array until response.

4.3.6 Electrophysiological recording and processing

The electroencephalograph (EEG) was recorded continuously through 128 electrodes placed on an ECI cap (Electro-Cap International, Ohio, USA) using the Active-Two Biosemi EEG system (Biosemi V.O.F Amsterdam, Netherlands). Eye movements and blinks were corrected using the ICA protocol in Analyser 2 software and segmented data was then visually inspected with trials containing artefacts rejected. Epochs that contained muscle or skin potential artefacts were rejected. Activity from all electrodes was sampled at a rate of 1024Hz. Offline 30 Hz (48 db/oct slope) lowpass and 0.1 Hz (48 db/oct slope) highpass filters were applied to the data. All data was re-referenced to an average reference which was then used to generate the grand averages. We used a 100ms pre-stimulus interval for the baseline correction. Continuous recording took place during the test phase of the experiment and trials were epoched/segmented from 100ms pre-stimulus to 800ms post-stimulus onset.

4.3.7 EEG analyses

Three early ERP components: P1, N1 and N2/P3, were identified based on the topography, global field power (GFP), deflection and latency characteristics of the respective grand average ERPs time-locked to stimulus presentation. Epochs of interest for each component were defined based on deflection extrema in the mean local field power (e.g., Brunet, Murray & Michel, 2011; Lehmann & Skrandies, 1980; Murray, Brunet & Michel, 2008).

Due to the nature of the stimuli, the left/right visual field presentation – with central fixation, the P1 and N1 were identified for left and right hemispheres for the stimuli in the contralateral visual field. As the positivity moved to the opposite visual field another P1 was identified. The same process was used for the N1 – first this negativity appeared in the contralateral visual field to the stimulus presentation, then the negativity occurred in the opposite hemisphere. This was the case for all conditions.

The latency of peak amplitude was used to define time epochs for analyses of the waves for stimuli presented in the upper and lower visual fields, and left and right visual fields separately and for left hemisphere (LH) and right hemisphere (RH) electrodes. LEFT VISUAL FIELD – RH P1 (90-130ms; Peak latency (B12) = 110ms); LH P1 (140-180ms; Peak latency (D29) = 160ms); RH N1 (165-205ms; Peak latency (B6) = 185ms); LH N1 (210-250ms; Peak latency (A9) = 230ms); P3 (350-450ms). RIGHT VISUAL FIELD – LH P1 (90-130ms; Peak latency (D30) = 110ms); RH P1 (145-185ms; Peak latency (B5) = 165ms); LH N1 (160-200ms; Peak latency (A9) = 180ms); RH N1 (200-240ms; Peak latency (B6) = 220ms); N2/P3 (350-450ms).

UPPER AND LOWER VISUAL FIELDS – P1 (100-150ms; Peak latency (B11) = 125ms); N1 (180-230ms; Peak latency (B7) = 205ms); N2/P3 (350-350ms).

Two symmetrical clusters over the left (LH) and right (RH) hemispheres were extracted each consisting of 10 spatially adjacent posterior electrodes: LH: D28, D29, D30, D31, A6, A7, A8, A9, A10, A11 and RH: B3, B4, B5, B6, B7, B8, B11, B12, B13 and B19, which correspond/overlap with electrode locations: LH: CP1, CP3, P3, P5, P7, PO7 and RH: CP2, CP4, P4, P6, P8, PO8 of the extended 10–20 system. These electrode clusters formed the region-of-interest for the subsequent analyses. Mean amplitudes were analysed using the General Linear Model by way of repeated measures ANOVA. Greenhouse-Geisser corrections were applied to all analyses of ERP data. Unless otherwise stated only significant main effects and interactions are reported where corrected $\alpha < .05$. Exact values of p are reported, except where $p < .001$ (two tailed).

4.3.8 Mass Univariate Analyses

Mass Univariate analyses (e.g., Groppe, Urbach & Kutas, 2011; Guthrie & Buchwald, 1991) were used to elucidate the time course of congruency effects, namely global interference. This involved using pair wise, time-frame by time-frame, permutation tests based on repeated measures t-tests across all 128 electrodes from 0-800ms. An a priori criterion for significance testing was adopted in which a threshold of $p<.01$ (two-tailed) must be attained for at least 10 consecutive time frames in at least 5 neighbouring electrodes (Guthrie & Buchwald, 1991; Murray, Brunet & Michel, 2008).

4.4 Results

4.4.1 Behavioural Analyses

4.4.1.1 Error Rate

Mean error rates and reaction times per condition are reported in Table 2. A 4 (Quadrant: LL; LR; UL; UR) x 2 (Congruency: congruent; incongruent) x 2 (Level: Local; Global) x 2 (Condition: HC Sparse; LC Dense) repeated-measures ANOVA revealed that there was a main effect of quadrant, $F(3,48)=3.83$ $p=.015$. Significant differences in error rate (%) were found between LL ($M=14.76$, $SD=16.05$) and LR ($M=12$, $SD=12.14$), $p=.05$, UL ($M=15.83$, $SD=14.92$) and LR ($M=12$, $SD=12.14$), $p=.022$ and UR ($M=17.14$, $SD=16.56$) and LR ($M=12$, $SD=12.14$), $p=.037$. There was also a main effect of congruency, $F(1,16)=82.83$, $p<.001$, with lower error rate (%) for congruent ($M=3.72$, $SD=3.76$) than incongruent stimuli ($M=26.14$, $SD=13.49$).

We included upper/lower and left/right in an ANOVA and found that there was a main effect of upper/lower, $F(1,16)=5.21$, $p=.036$, and also an interaction between upper/lower and left/right, $F(1,16)=4.95$, $p=.041$. In the upper VF, error rate was greater in right than left, whereas in the lower VF, error rate was greater in the left than right, see Table 2 means for effects.

4.4.1.2 RT data

Table 2. Mean error rates and reaction times (RTs) for congruent (C) and incongruent (I) stimuli displayed in four visual field quadrants.

		Error Rate (%)		RT (ms)	
		Mean	SD	Mean	SD
Upper Left	C	3.54	3.15	740.42	186.95
	I	28.12	11.33	916.63	308.43
	Mean	15.83	14.92	828.52	230.44
Upper Right	C	4.46	4.19	759.43	241.37
	I	29.82	14.36	897.89	271.61
	Mean	17.14	16.56	828.66	219.13
Lower Left	C	4.02	4.72	739.72	180.15
	I	25.49	16.25	861.49	259.91
	Mean	14.76	16.05	800.61	193.21
Lower Right	C	2.86	2.84	722.76	195.12
	I	21.14	10.88	860.62	243.27
	Mean	12	12.14	791.69	199.84
Mean Upper	C	4	3.68	749.92	215.29
	I	28.97	12.77	907.26	271.62
	Mean	16.49	15.66	828.59	223.17
Mean Lower	C	3.44	3.88	731.24	187.28
	I	23.32	13.79	861.06	250.79
	Mean	13.38	14.19	796.15	195.13
Mean Left	C	3.78	3.95	740.07	155.99
	I	26.8	13.86	889.06	234.66
	Mean	15.29	15.39	814.57	211.51
Mean Right	C	3.66	3.62	741.09	192.47
	I	25.48	13.29	879.25	204.41
	Mean	14.57	14.64	810.17	208.97

A 4 (Quadrant: LL; LR; UL; UR) x 2 (Congruency: congruent; incongruent) x 2 (Level: Local; Global) x 2 (Condition: HC Sparse; LC Dense) repeated-measures ANOVA revealed that there was a main effect of congruency $F(1,18)=44.91$, $p<.001$, with faster RTs (ms) for congruent (*M*=74.06, *SD*=20.16) than incongruent stimuli (*M*=87.81, *SD*=27.14). No other main effects or interactions were significant.

4.4.2 ERP analyses

The aims of the ERP analyses were: (1) to identify an ERP signature of global/local integration; (2) to establish whether there is a global advantage evident in the ERPs. The rationale was that the congruency manipulation in the local report trials should provide the clearest indicator of a global/local integration effect, global report trials are presented subsequently. ERP analyses were conducted on left/right VF data and upper/lower VF data separately. The behavioural data suggest that there may be processing differences in upper and lower VFs, and left and right VF analyses are included based on hemispheric and visual field differences in local and global processing in the literature.

4.4.2.1 ERP Analyses I: Congruency effects in local report

4.4.2.1.1 Left and Right Visual field presentation

Using only 'report local' trials, looking at congruent vs. incongruent stimuli using only the maximally local condition (HCS).

P1. Using peak amplitude measures, a 2(Congruency: congruent, incongruent) x 2(hemisphere: left, right) x 2(visual field presentation: LVF, RVF) repeated measures ANOVA revealed an interaction of congruency and hemisphere, $F(1,18)=11.72$, $p=.003$. There were greater amplitudes for congruent (*M*=1.79, *SD*=1.49) than incongruent (*M*=1.23, *SD*=1.85) trials in the right hemisphere, $p=.007$, but no significant differences in the left hemisphere.

Latency. A 2(Congruency: congruent, incongruent) x 2(hemisphere: left, right) x 2(visual field presentation: LVF, RVF) repeated measures ANOVA revealed a main effect of congruency, $F(1,18)=5.03$, $p=.038$, with an earlier P1 for congruent (*M*=131.96, *SD*=24.01) than incongruent (*M*=133.45, *SD*=25.84) stimuli.

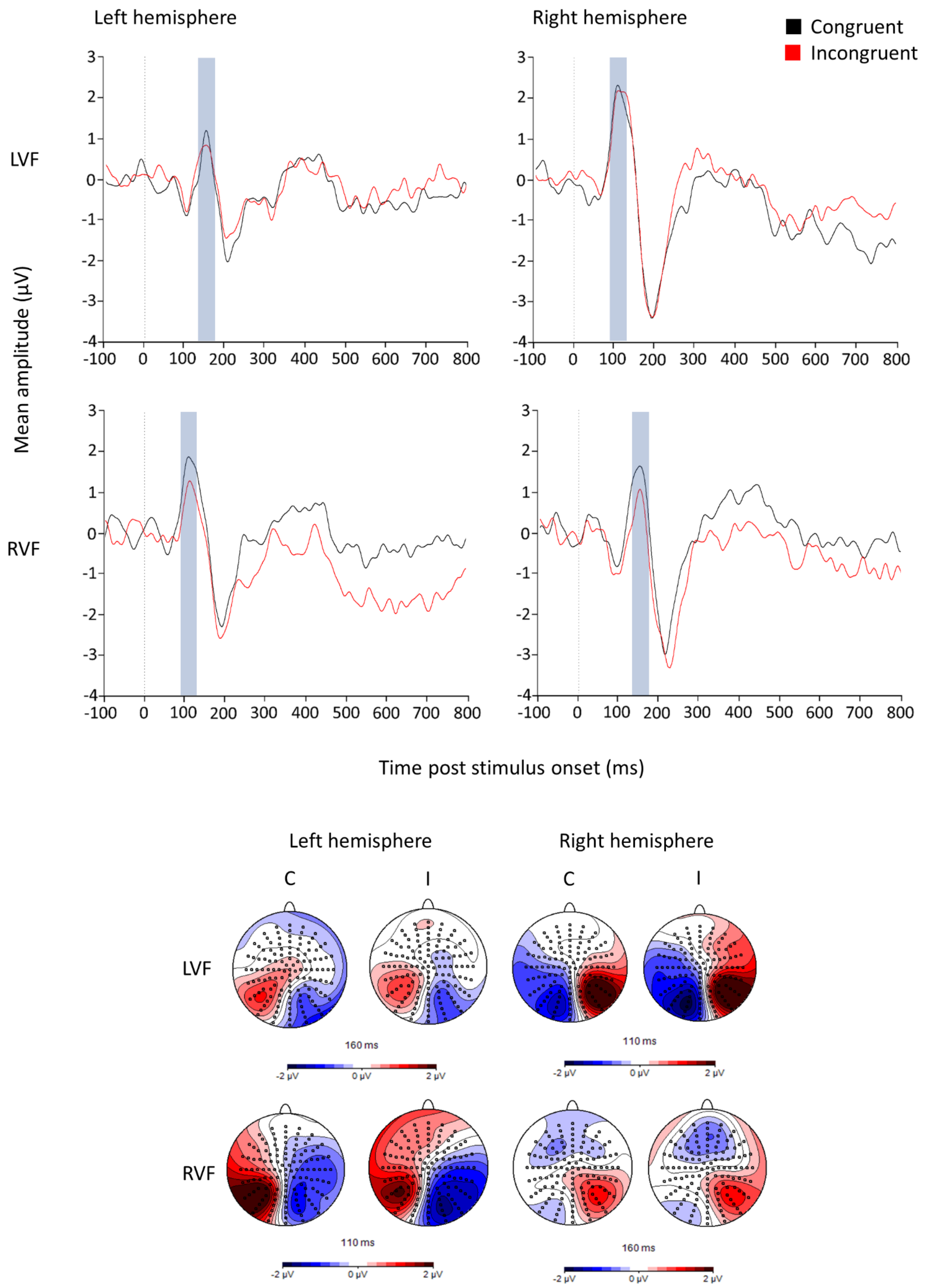


Figure 25. Waveforms for the report local condition from an occipital electrode cluster corresponding to P7, P8, PO7 and PO8 in the 10-20 system. Showing the P1, between 90 and 180ms post-stimulus onset for congruent and incongruent stimuli in the left and right hemispheres.

N1. Using peak amplitude measures, a 2(Congruency: congruent, incongruent) x 2(hemisphere: left, right) x 2(visual field presentation: LVF, RVF) repeated measures ANOVA revealed an interaction of congruency and VF, $F(1,18)=9.22$, $p=.007$. There was greater negativity for incongruent ($M=-3.91$, $SD=2.95$) than congruent ($M=-3.39$, $SD=2.87$) for RVF stimuli, $p=.043$, but no significant differences for LFV stimuli.

Latency. There were no significant main effects of interactions involving congruency.

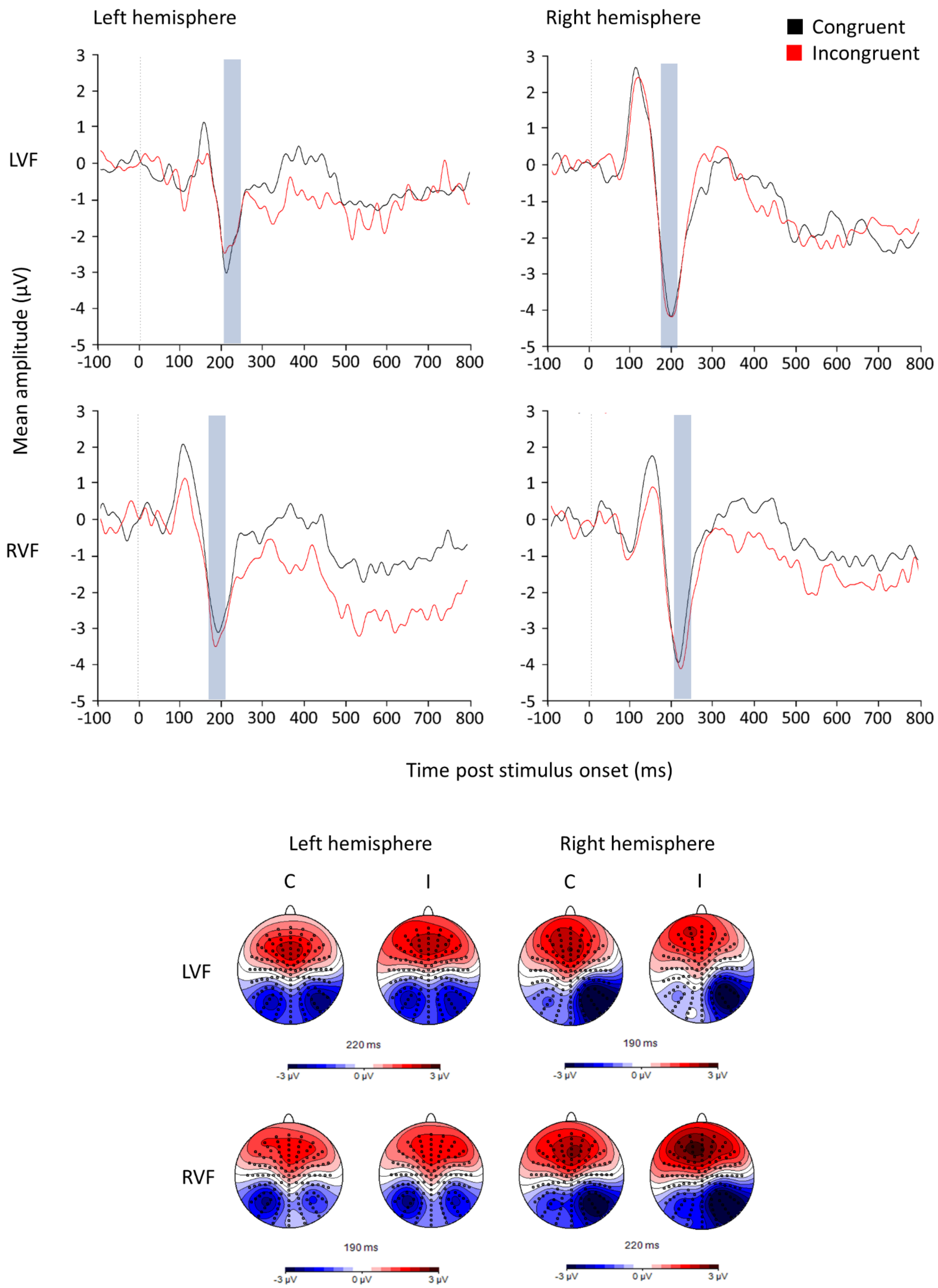


Figure 26. Waveforms for the report local condition from an occipital electrode cluster corresponding to P7, P8, PO7 and PO8 in the 10-20 system. Showing the N1, between 160 and 250ms post-stimulus onset for congruent and incongruent stimuli in the left and right hemispheres.

N2/P3. Using mean amplitude measures, a 2(Congruency: congruent, incongruent) x 2(hemisphere: left, right) x 2(visual field presentation: LVF, RVF) repeated measures ANOVA revealed a main effect of congruency, $F(1,18)=18.41$, $p=.001$, with greater amplitudes for congruent ($M=90.03$, $SD=110.47$) than incongruent ($M=26.03$, $SD=117.23$) stimuli. There was also an interaction between hemisphere and VF, $F(1,18)=4.74$, $p=.045$. In the left hemisphere, there is an amplitude difference between stimuli presented in the LVF ($M=64.58$, $SD=166.51$) and the RVF ($M=11.78$, $SD=181.11$).

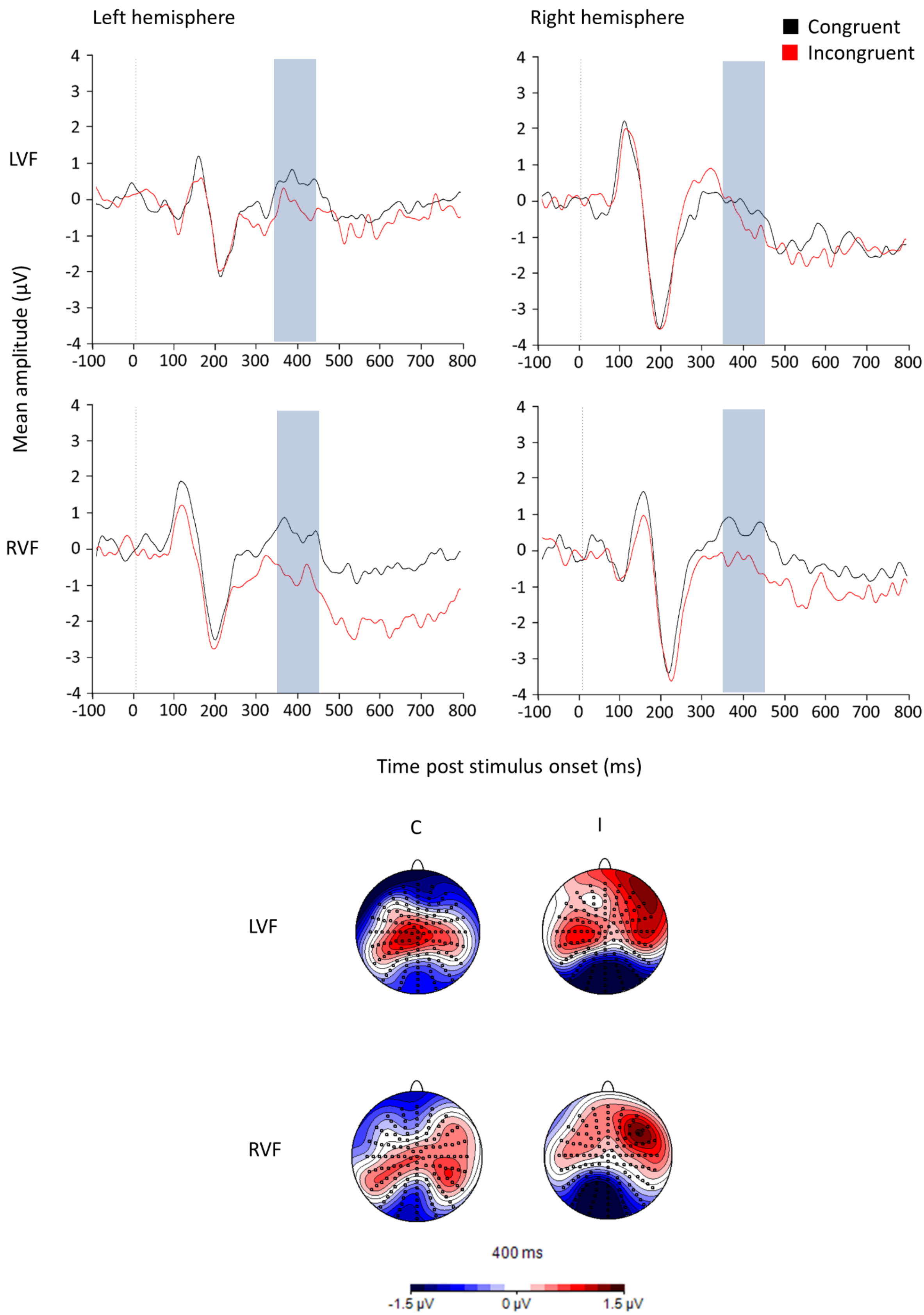


Figure 27. Waveforms for the report local condition from an electrode cluster corresponding to P3, P4, CP3 and CP4 in the 10-20 system. Showing the P3, between 350 and 450ms post-stimulus onset for congruent and incongruent stimuli in the left and right hemispheres.

4.4.2.1.2 Upper and lower visual field presentation

Using only 'report local' trials, looking at congruent vs. incongruent stimuli using only the maximally local condition (HCS).

P1. Using peak amplitude measures, a 2(Congruency: congruent, incongruent) x 2(hemisphere: left, right) x 2(visual field presentation: Lower VF, Upper VF) repeated measures ANOVA revealed a main effect of hemisphere, $F(1,18)=51.86$, $p<.001$. Amplitudes were greater in the right ($M=2.03$, $SD=1.5$) than the left hemisphere ($M=0.39$ $SD=2.27$).

Latency. No significant main effects or interactions were found.

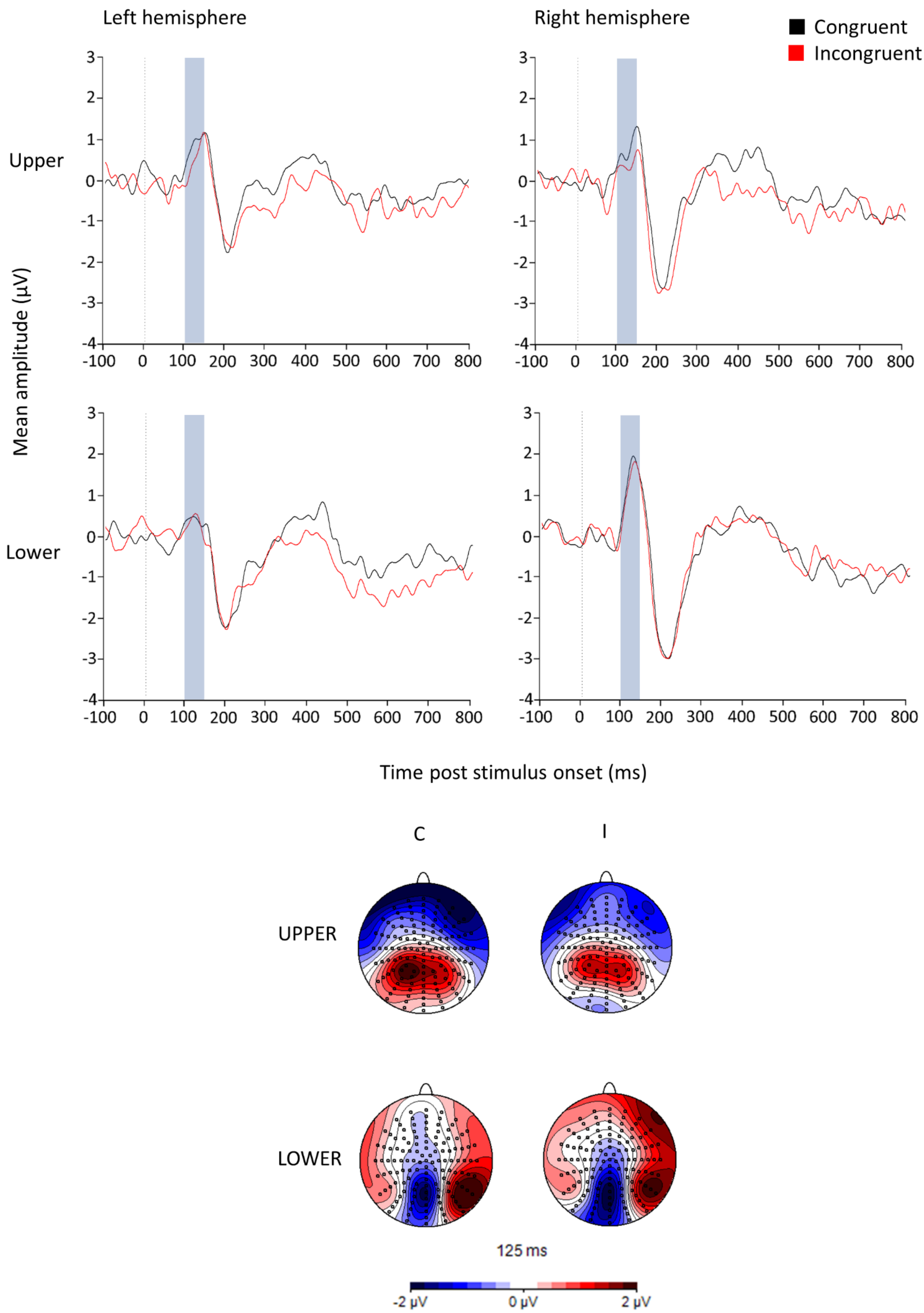
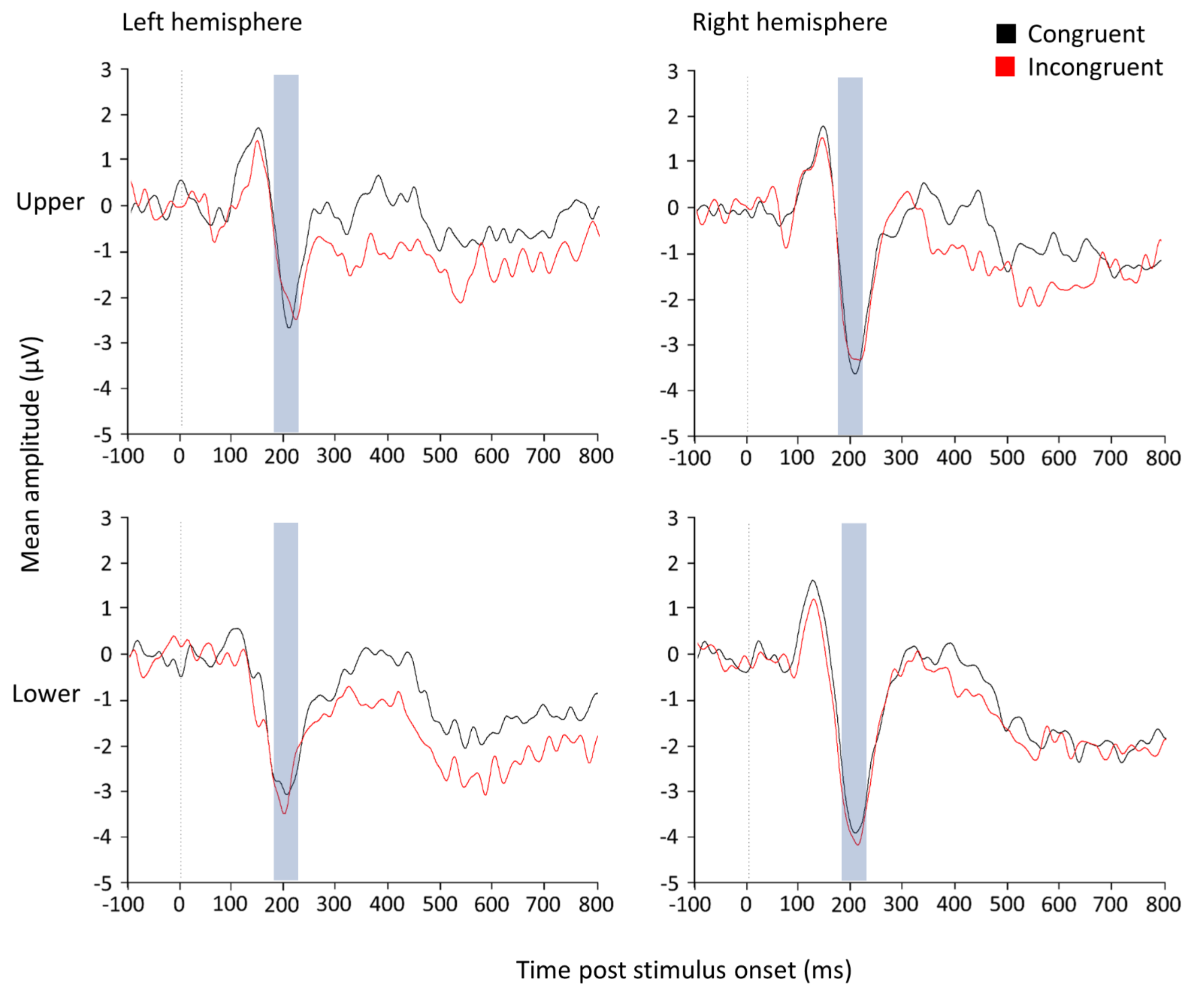


Figure 28. Waveforms for the report local condition from an occipital electrode cluster corresponding to P7, P8, PO7 and PO8 in the 10-20 system. Showing the P1, between 100 and 150ms post-stimulus onset for congruent and incongruent stimuli in the left and right hemispheres.

N1. Using peak amplitude measures, a 2(Congruency: congruent, incongruent) x 2(hemisphere: left, right) x 2(visual field presentation: Lower VF, Upper VF) repeated measures ANOVA revealed a main effect of hemisphere, $F(1,18)=33.26$, $p<.001$, with more negative amplitudes in the right hemisphere ($M=-3.84$, $SD=2.46$) than the left ($M=-1.87$, $SD=3.12$). There was also a main effect of VF, $F(1,18)=20.67$, $p<.001$, with more negative amplitudes for stimuli presented in the lower VF ($M=-3.36$, $SD=2.97$) than the upper VF ($M=-2.16$, $SD=2.97$).

Latency. No significant main effects or interactions were found.



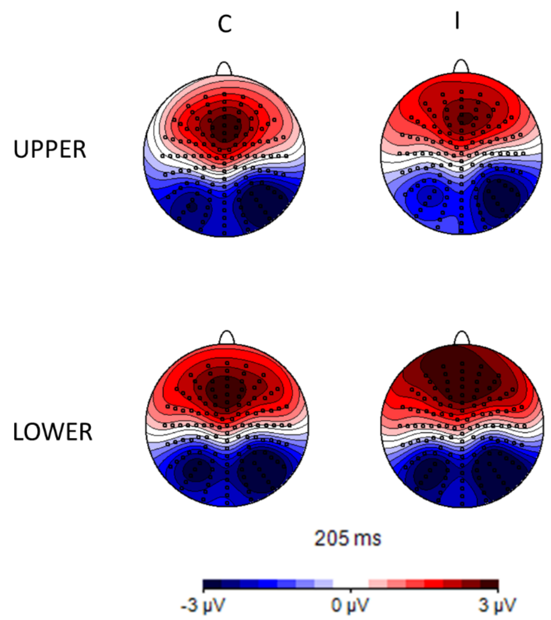


Figure 29. Waveforms for the report local condition from an occipital electrode cluster corresponding to P7, P8, PO7 and PO8 in the 10-20 system. Showing the N1, between 180 and 230ms post-stimulus onset for congruent and incongruent stimuli in the left and right hemispheres.

N2/P3. Using mean amplitude measures, a 2(Congruency: congruent, incongruent) x 2(hemisphere: left, right) x 2(visual field presentation: Lower VF, Upper VF) repeated measures ANOVA revealed a main effect of congruency, $F(1,18)=16.37$, $p=.001$, with greater amplitudes for congruent ($M=85.88$, $SD=153.15$) than incongruent ($M=-11.86$, $SD=146.5$) stimuli. There was also a main effect of VF, $F(1,18)=15.32$, $p=.001$, with greater amplitudes for stimuli presented in the upper VF ($M=68.04$, $SD=150.2$) than the lower VF ($M=29.2$, $SD=155.75$).

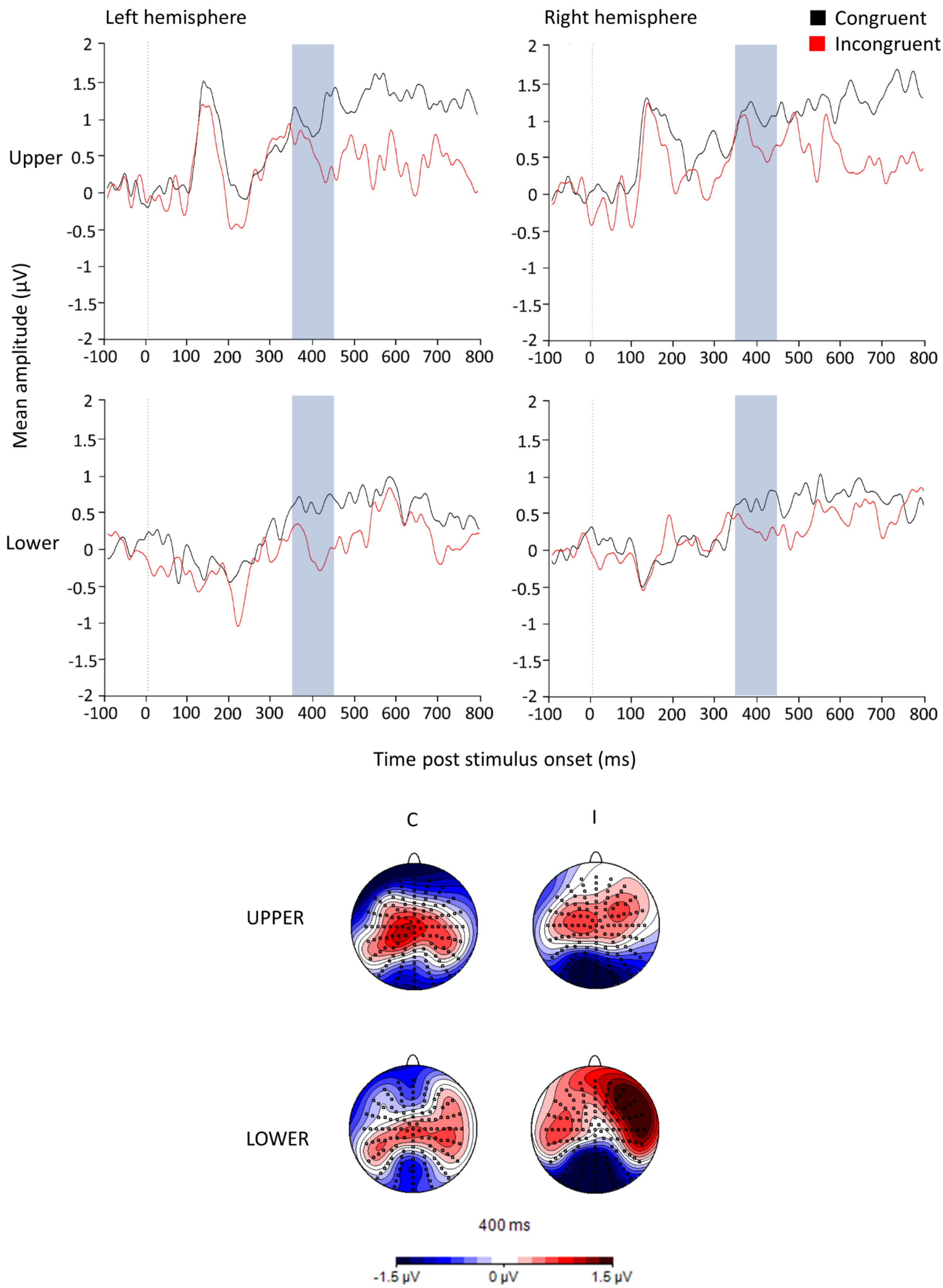


Figure 30. Waveforms for the report local condition from an electrode cluster corresponding to P3, P4, CP3 and CP4 in the 10-20 system. Showing the P3, between 350 and 450ms post-stimulus onset for congruent and incongruent stimuli in the left and right hemispheres.

4.4.2.2 ERP Analyses II: Congruency effects in global report

4.4.2.2.1 Left and Right Visual field presentation

Using only ‘report global’ trials, looking at congruent vs. incongruent stimuli using only the maximally global condition (LCD).

P1. Using peak amplitude measures, a 2(Congruency: congruent, incongruent) x 2(hemisphere: left, right) x 2(visual field presentation: LVF, RVF) repeated measures ANOVA revealed an interaction of congruency and hemisphere, $F(1,18)=7.63$, $p<.013$. Amplitudes were greater for incongruent (M=1.35, SD=1.82) than congruent (M=1.29, SD=1.82) trials in the left hemisphere, p=.031 but not significantly different in the right hemisphere.

Latency. No significant main effects or interactions were found.

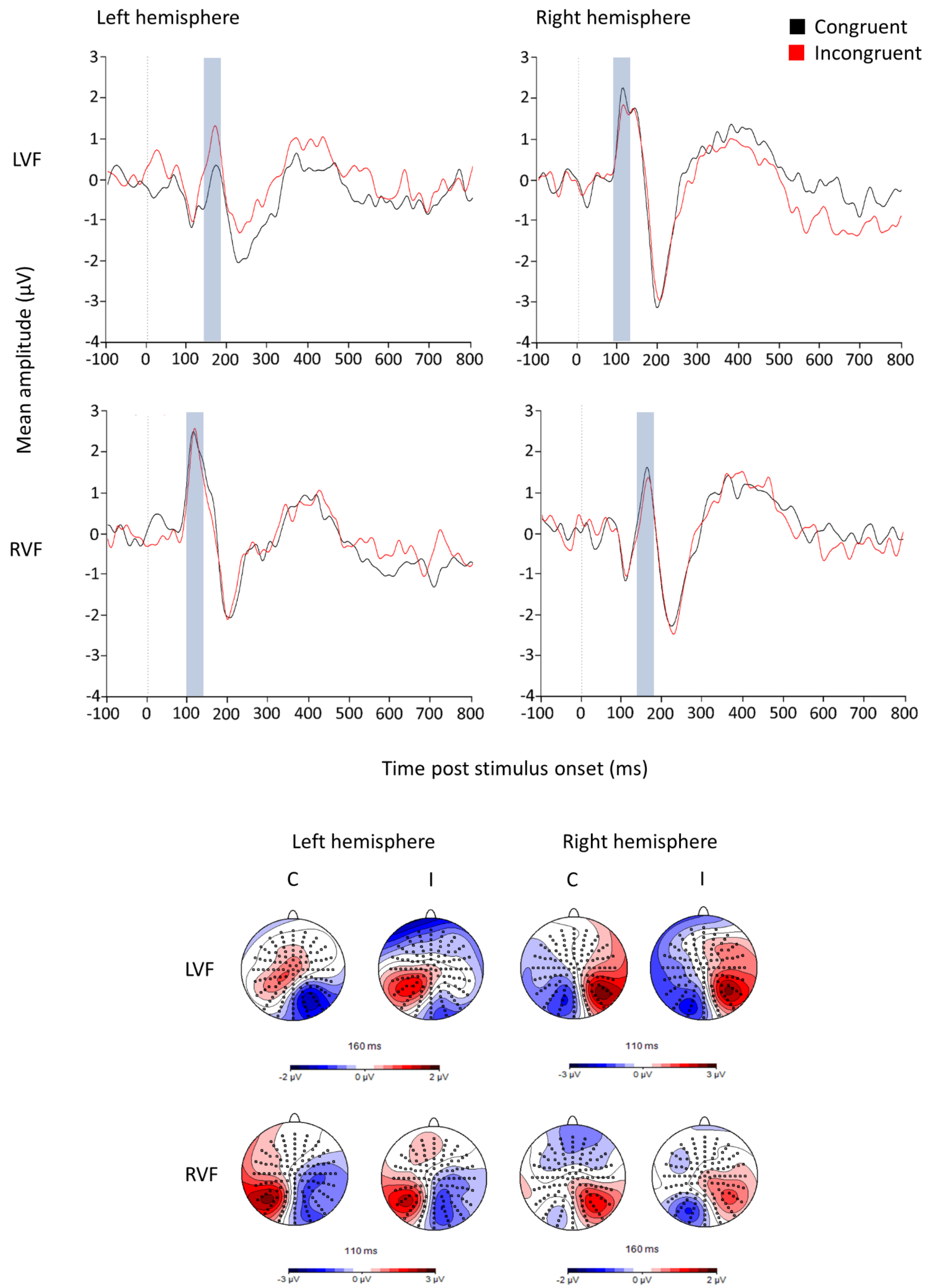


Figure 31. Waveforms for the report global condition from an occipital electrode cluster corresponding to P7, P8, PO7 and PO8 in the 10-20 system. Showing the P1, between 90 and 180ms post-stimulus onset for congruent and incongruent stimuli in the left and right hemispheres.

N1. Using peak amplitude measures, a 2(Congruency: congruent, incongruent) x 2(hemisphere: left, right) x 2(visual field presentation: LVF, RVF) repeated measures ANOVA revealed a main effect of congruency, $F(1,18)=4.78$, $p=.042$, with more negative amplitudes for congruent ($M=-3.89$, $SD=2.9$) than incongruent ($M=-3.01$, $SD=3.3$) stimuli.

Latency. A 2(Congruency: congruent, incongruent) x 2(hemisphere: left, right) x 2(visual field presentation: LVF, RVF) repeated measures ANOVA revealed a main effect of congruency, $F(1,18)=9.44$, $p=.007$. The N1 for congruent stimuli ($M=203.19$, $SD=23.28$) was earlier than for incongruent stimuli ($M=207.4$, $SD=23.22$).

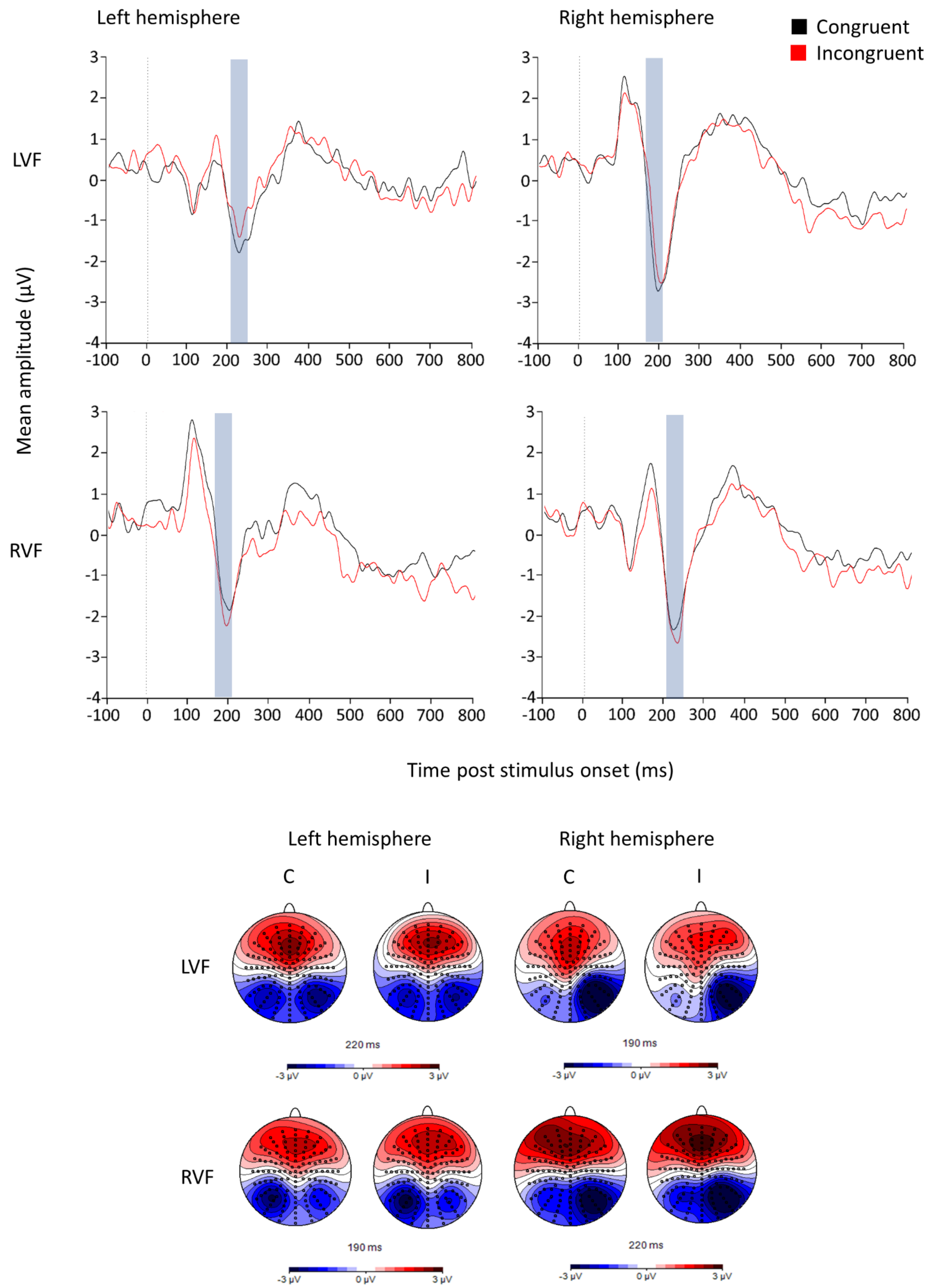


Figure 32. Waveforms for the report global condition from an occipital electrode cluster corresponding to P7, P8, PO7 and PO8 in the 10-20 system. Showing the N1, between 160 and 250ms post-stimulus onset for congruent and incongruent stimuli in the left and right hemispheres.

N2/P3. Using mean amplitude measures, a 2(Congruency: congruent, incongruent) x 2(hemisphere: left, right) x 2(visual field presentation: LVF, RVF) repeated measures ANOVA revealed no significant main effects or interactions.

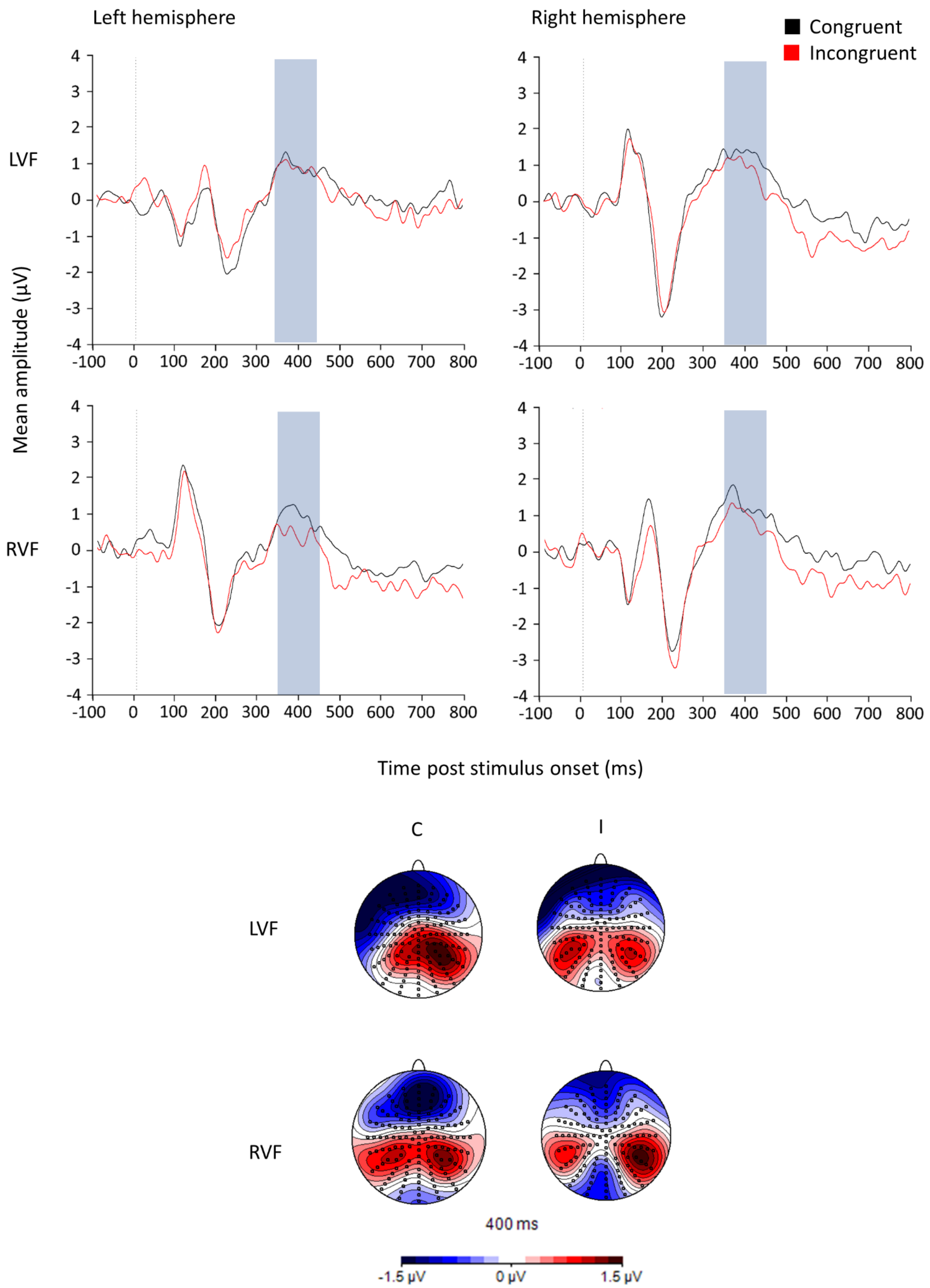


Figure 33. Waveforms for the report global condition from an electrode cluster corresponding to P3, P4, CP3 and CP4 in the 10-20 system. Showing the P3, between 350 and 450ms post-stimulus onset for congruent and incongruent stimuli in the left and right hemispheres.

4.4.2.2.2 Upper and lower visual field presentation

Using only 'report global' trials, looking at congruent vs. incongruent stimuli using only the maximally global condition (LCD).

P1. Using peak amplitude measures, a 2(Congruency: congruent, incongruent) x 2(hemisphere: left, right) x 2(visual field presentation: Lower VF, Upper VF) repeated measures ANOVA revealed a main effect of hemisphere, $F(1,18)=44.66$, $p<.001$. Amplitudes were greater in the right ($M=1.84$, $SD=2.36$) than the left hemisphere ($M=0.39$, $SD=2.38$).

Latency. No significant main effects or interactions were found.

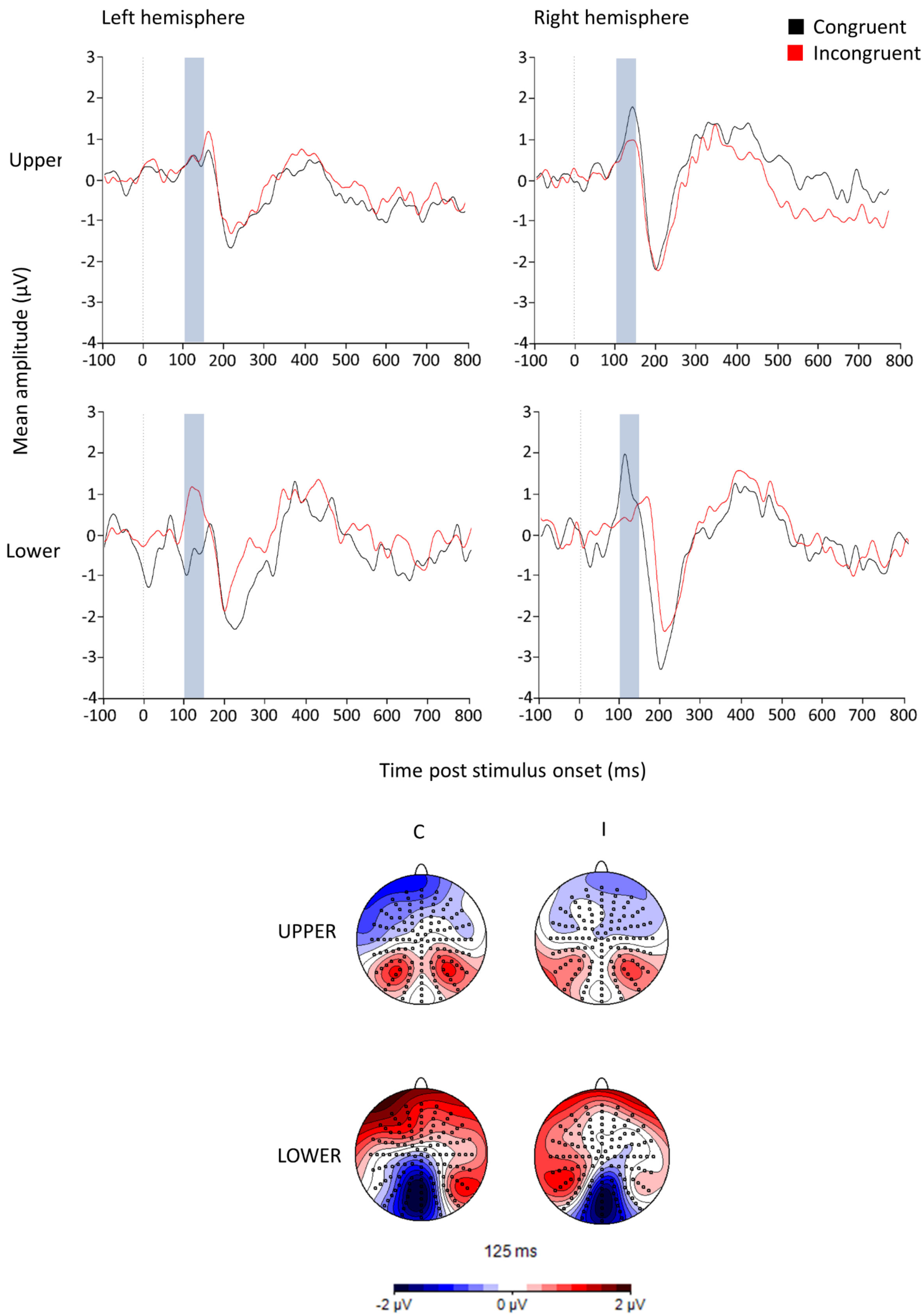


Figure 34. Waveforms for the report global condition from an occipital electrode cluster corresponding to P7, P8, PO7 and PO8 in the 10-20 system. Showing the P1, between 100 and 150ms post-stimulus onset for congruent and incongruent stimuli in the left and right hemispheres.

N1. Using peak amplitude measures, a 2(Congruency: congruent, incongruent) x 2(hemisphere: left, right) x 2(visual field presentation: Lower VF, Upper VF) repeated measures ANOVA revealed a main effect of hemisphere, $F(1,18)=18.99$, $p<.001$, with greater negativity in the right hemisphere ($M=-3.44$, $SD=2.83$) than left ($M=-1.84$, $SD=2.2$). There was also a main effect of VF, $F(1,18)=11.35$, $p=.004$, with more negative N1 for stimuli presented in the lower VF ($M=-3.18$, $SD=3.24$) than the upper VF ($M=-2.16$, $SD=2.78$).

Latency. No significant main effects or interactions were found.

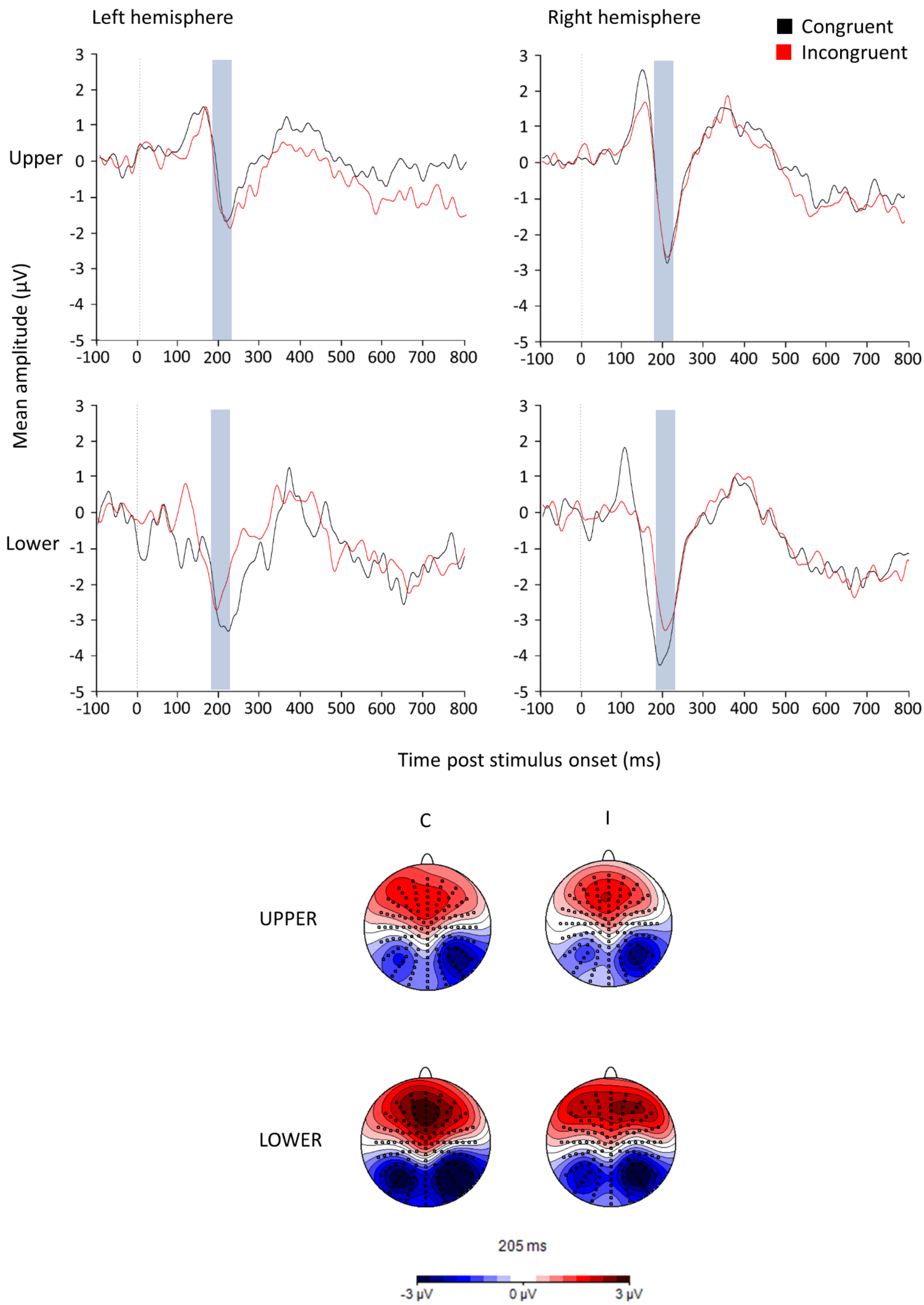


Figure 35. Waveforms for the report global condition from an occipital electrode cluster corresponding to P7, P8, PO7 and PO8 in the 10-20 system. Showing the N1, between 180 and 230ms post-stimulus onset for congruent and incongruent stimuli in the left and right hemispheres.

N2/P3. Using mean amplitude measures, a 2(Congruency: congruent, incongruent) x 2(hemisphere: left, right) x 2(visual field presentation: Lower VF, Upper VF) repeated measures ANOVA revealed an interaction of congruency and hemisphere, $F(1,18)=4.82$, $p=.043$. There was a difference between congruent ($M=91.52$, $SD=159.35$) and incongruent ($M=46.31$, $SD=161.63$) stimuli in the right hemisphere, though this was not statistically significant, $p=.052$

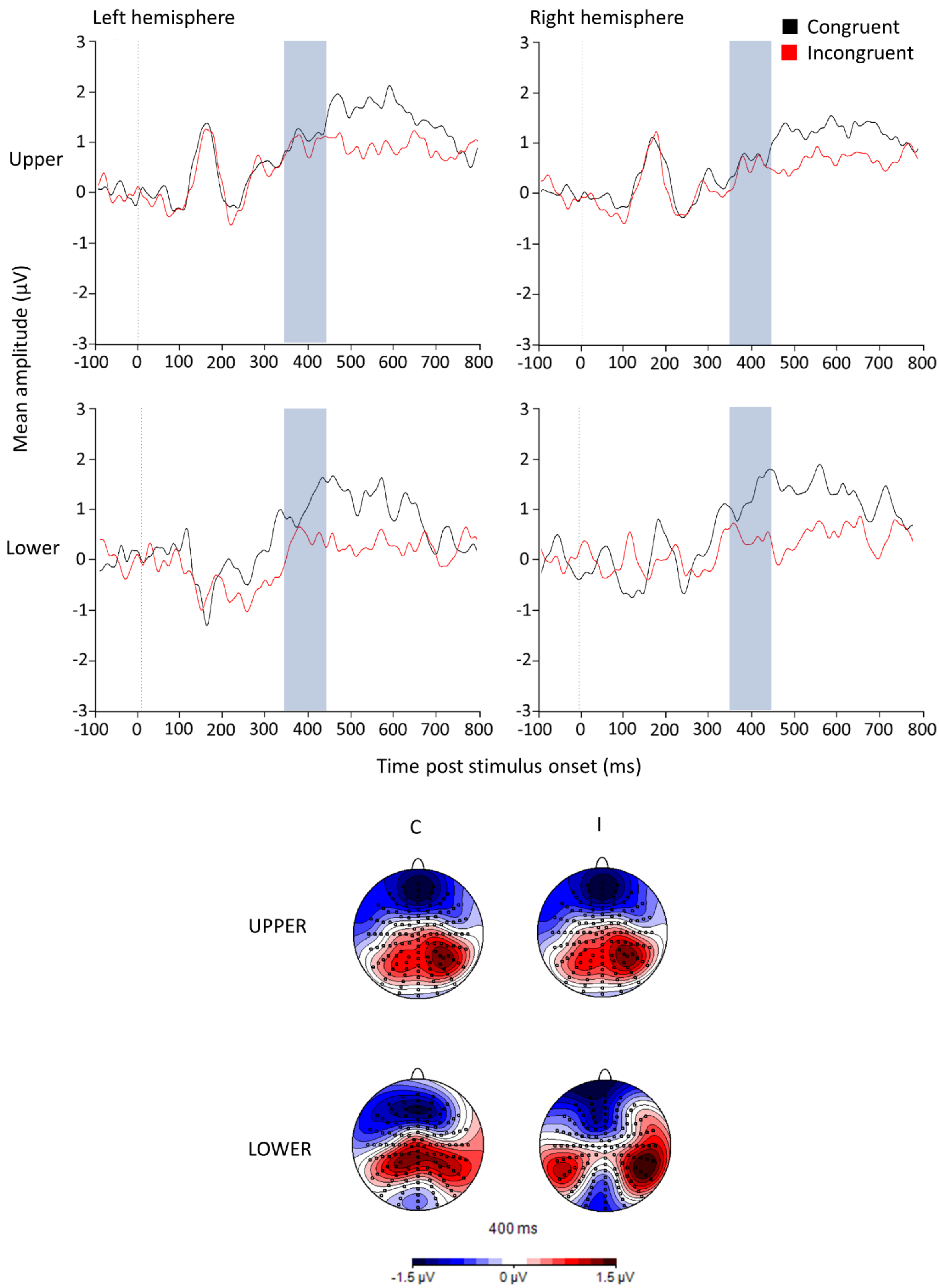


Figure 36. Waveforms from for the report global condition an electrode cluster corresponding to P3, P4, CP3 and CP4 in the 10-20 system. Showing the P3, between 350 and 450ms post-stimulus onset for congruent and incongruent stimuli in the left and right hemispheres.

4.4.2.3 Further Analyses I: Mass Univariate Contrasts across all 128 electrodes

4.4.2.3.1 Left and right visual field presentation

Mass univariate analyses were used to complement our standard waveform analyses of the effects of congruency on discrimination of global or local elements of the stimuli. Unlike the standard analysis, the mass univariate approach allows us to examine the patterns of contrasts between conditions across all 128 electrodes (rather than restricting the analysis to the 10 electrode cluster in each hemisphere). The temporal distributions of these contrasts across all 128 electrodes are shown in Figure 37.

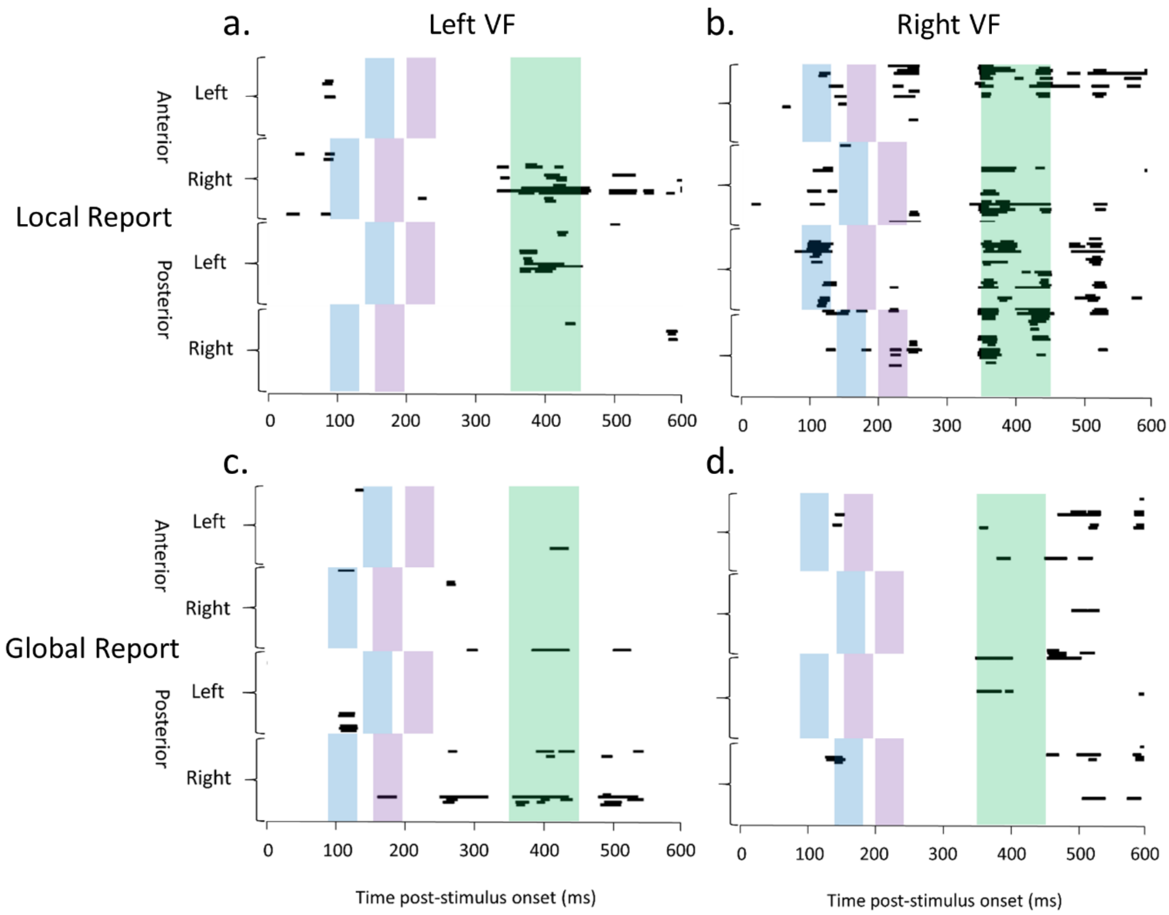


Figure 37. Mass univariate contrasts showing time (x axis) and electrodes (y axis) for the congruent/incongruent stimuli contrast for local and global report for: (a) Local report in the LVF; (b) Local report in the RVF; (c) Global report in the LVF; (d) Global report in the RVF. The highlighted areas show P1 (blue), N1 (purple) and N2/P3 (green). All 128 electrodes are shown, dark areas indicate periods where electrodes are significant at $p<.01$.

Figure 38 shows a time series plot of the frequency distribution of significant differences between congruent and incongruent stimuli, sub-sampled into 10ms bins. These data were analysed as a non-parametric time-series using the Friedman test, which showed that at the P1, frequency distributions were significantly different between local and global report in the left VF, $\chi^2(1)=4$, $p=.046$ and the right VF, $\chi^2(1)=5$, $p=.025$ in the left hemisphere. However, there were no significant differences between local and global report in the left VF or right VF in the right hemisphere. At the N1, frequency distributions were different for local and global report in the right VF, $\chi^2(1)=4$, $p=.046$ in the right hemisphere only. At the N2/P3, frequency distributions were significantly different for local and global report in the left VF, $\chi^2(1)=8$, $p=.005$ and the right VF, $\chi^2(1)=10$, $p=.002$ in the left hemisphere. There were also significant differences between local and global report in the left VF, $\chi^2(1)=10$, $p=.002$ and the right VF, $\chi^2(1)=9$, $p=.003$ in the right hemisphere.

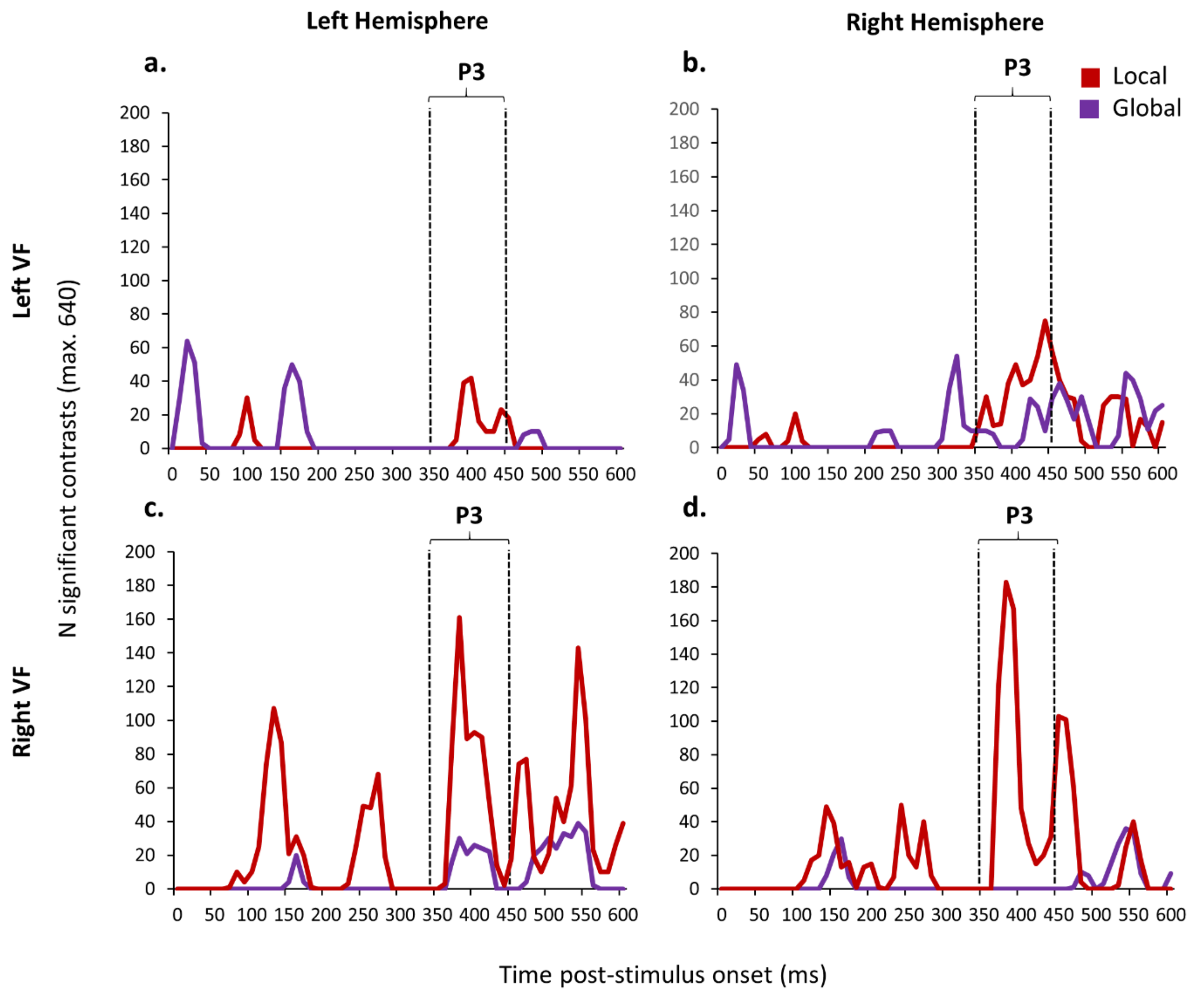


Figure 38. Time series distribution showing the frequency of significant difference contrasts from the mass univariate analysis between 0 and 600ms. Contrasts shown are between congruent and incongruent stimuli in: (a) left hemisphere for LVF stimuli; (b) right hemisphere for LVF stimuli; (c) left hemisphere for RVF stimuli; (d) right hemisphere for RVF stimuli. The dotted lines show the N2/P3 (350-450ms).

These findings complement those of the waveform analyses: the time series distribution from the mass univariate analysis revealed that differences between congruent and incongruent stimuli arise at the N2/P3 in the local report condition, but not in the global report condition. There are also some differences at the N1, but to a lesser extent. There were differences at the P1 in the left hemisphere, with more differences between congruent and incongruent stimuli for local report trials.

4.4.2.3.2 Upper and lower visual field presentation

Mass univariate analyses were used to complement our standard waveform analyses of the effects of congruency on discrimination of global or local elements of the stimuli. Unlike the standard analysis, the mass univariate approach allows us to examine the patterns of contrasts between conditions across all 128 electrodes (rather than restricting the analysis to the 10 electrode cluster in each hemisphere). The temporal distributions of these contrasts across all 128 electrodes are shown in Figure 39.

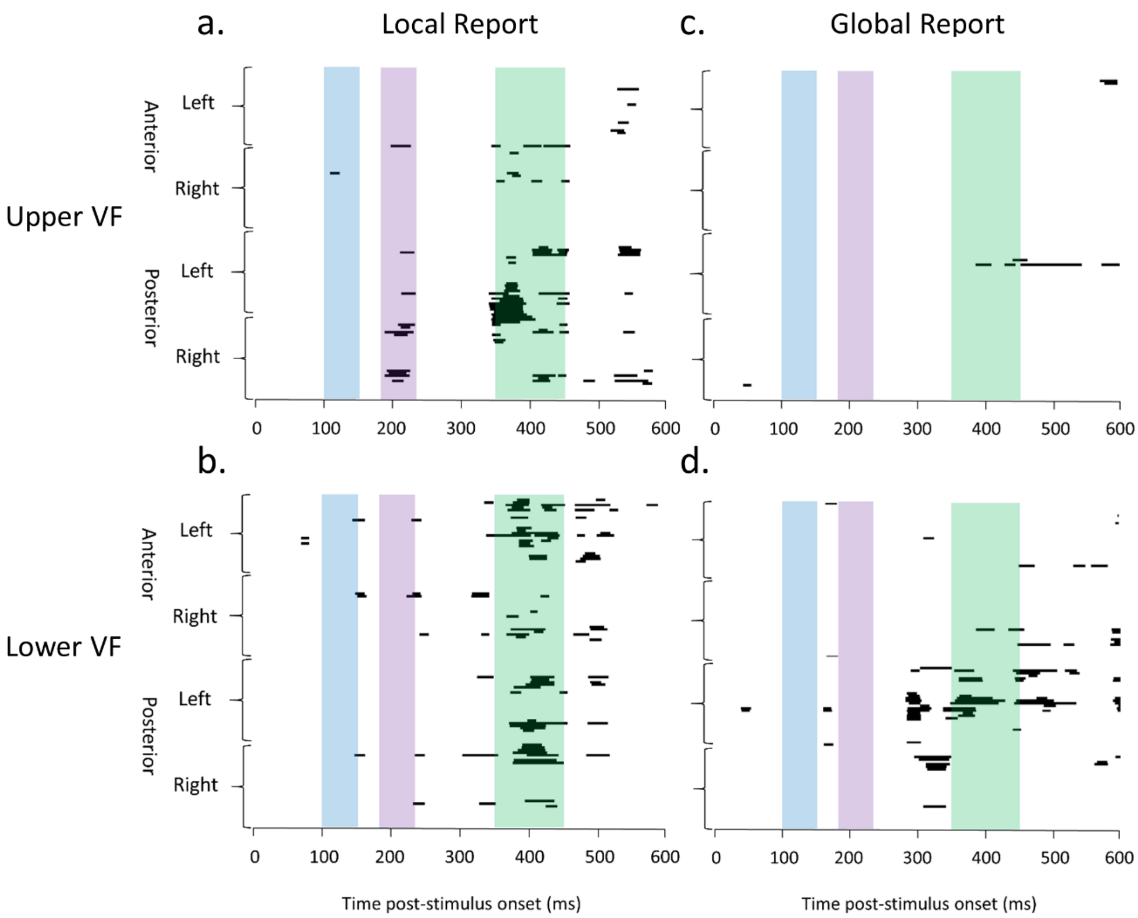


Figure 39. Mass univariate contrasts showing time (x axis) and electrodes (y axis) for the congruent/incongruent stimuli contrast for local and global report in (a) Local report in the upper VF; (b) Local report in the lower VF; (c) Global report in the upper VF; (d) Global report in the lower VF. The highlighted areas show P1 (blue), N1 (purple) and N2/P3 (green). All 128 electrodes are shown, dark areas indicate periods where electrodes are significant at p<.01.

Figure 40 shows a time series plot of the frequency distribution of significant differences between congruent and incongruent stimuli, sub-sampled into 10ms bins. These data were analysed as a non-parametric time-series using the Friedman test, which showed that at the P1, frequency distributions were not significantly different for the local and global report conditions. At the N1, frequency distributions were different for local and global report in the upper VF, $\chi^2(1)=4$, $p=.046$ in the right hemisphere only. At the N2/P3, frequency distributions were significantly different for local and global report in the upper VF, $\chi^2(1)=10$, $p=.002$, but not the lower VF, $p=.058$ in the left hemisphere. Also, significant differences between local and global report in the upper VF, $\chi^2(1)=10$, $p=.002$ and the lower VF, $\chi^2(1)=5.44$, $p=.02$ in the right hemisphere.

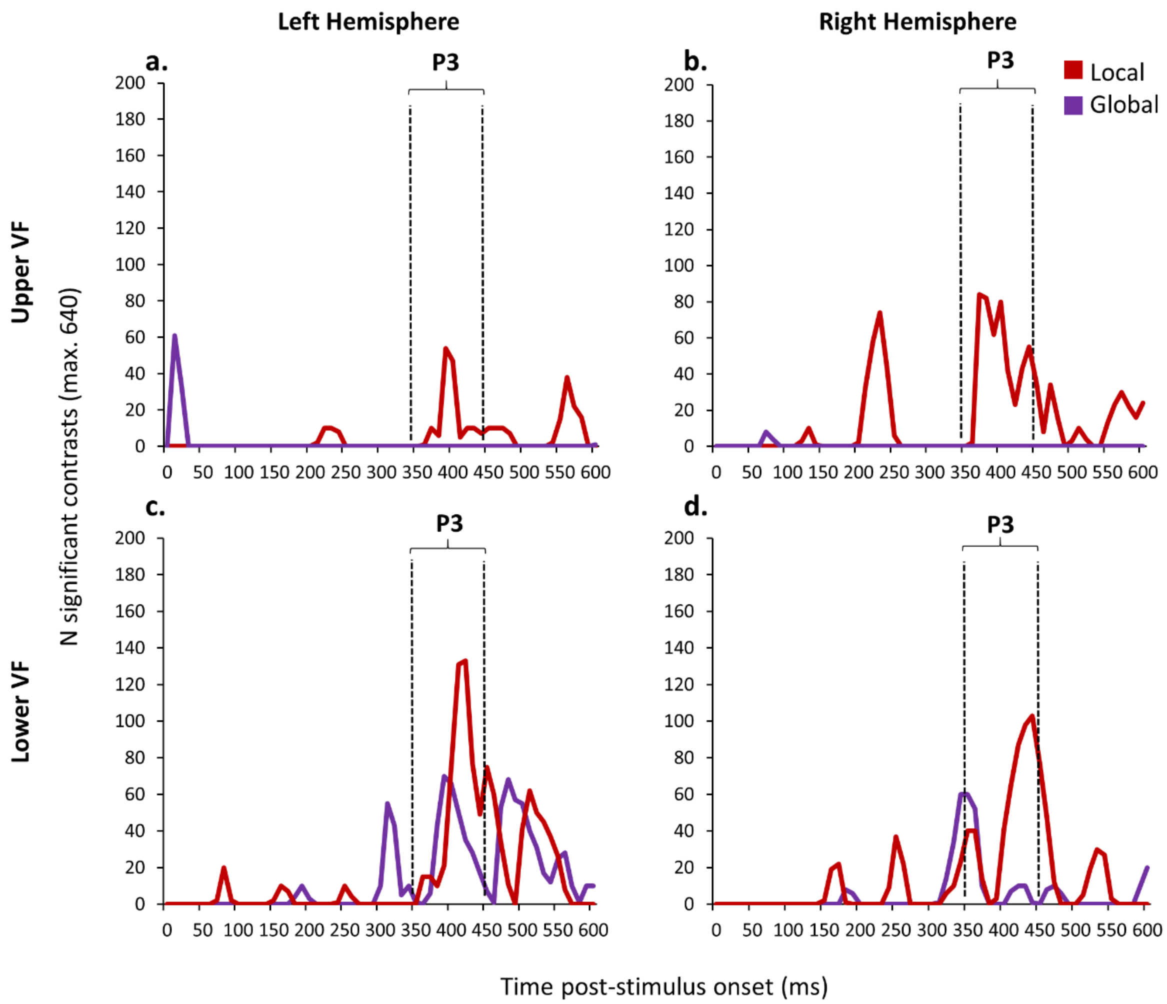


Figure 40. Time series distribution showing the frequency of significant difference contrasts from the mass univariate analysis between 0 and 600ms. Contrasts shown are between

congruent and incongruent stimuli in for (a) left hemisphere for upper VF stimuli; (b) right hemisphere for upper VF stimuli; (c) left hemisphere for lower VF stimuli; (d) right hemisphere for lower VF stimuli. The dotted lines show the N2/P3 (350-450ms).

These findings complement those of the waveform analyses: the time series distribution from the mass univariate analysis revealed that differences between congruent and incongruent stimuli arise at the N2/P3, with more differences for local report than global report. There are also some differences at the N1, but to a lesser extent.

4.4.2.4 Further Analyses II: Global and Local effects

As we found evidence of congruency effects and global interference, we were interested to see if a global advantage was evident in the ERPs. Global advantage is another facet of the GPE, in the ERPs there may be differential processing for global and local report. In the following analyses, we collapsed across congruency.

4.4.2.4.1 Left and right VF presentation

P1. Using peak amplitude measures, a 2(Report: global, local) x 2(hemisphere: left, right) x 2(visual field presentation: left visual field (LVF), right visual field (RVF)) repeated measures ANOVA revealed an interaction of hemisphere and VF presentation, $F(1,18)=6.59$, $p=.019$. Amplitudes in the left hemisphere were greater for stimuli presented in the RVF ($M=1.53$, $SD=1.2$) than LVF ($M=0.6$, $SD=1.3$), $p=.004$. There was no difference between LVF and RVF presentation in the right hemisphere. There were no other main effects of interactions.

Latency. A 2(Report: global, local) x 2(hemisphere: left, right) x 2(visual field presentation: LVF, RVF) repeated measures ANOVA revealed an interaction of hemisphere and VF, $F(1,18)=755.28$, $p<.001$. There was an earlier P1 in the left hemisphere for stimuli presented in the RVF ($M=109.77$, $SD=7.71$) than LVF ($M=156.52$, $SD=9.65$), $p<.001$. In the right hemisphere, P1 for stimuli presented in the LVF ($M=111.62$, $SD=7.74$) was earlier than for stimuli presented in the RVF ($M=157.23$, $SD=9.23$), $p<.001$.

N1. Using peak amplitude measures, a 2(Report: global, local) x 2(hemisphere: left, right) x 2(visual field presentation: LVF, RVF) repeated measures ANOVA revealed a main effect of hemisphere, $F(1,18)=7.12$, $p<.016$. There was a more negative N1 in the right ($M=-3.83$, $SD=2.41$) than the left hemisphere ($M=-2.75$, $SD=2.27$).

Latency. A 2(Report: global, local) x 2(hemisphere: left, right) x 2(visual field presentation: LVF, RVF) repeated measures ANOVA revealed a main effect of VF, $F(1,18)=49.04$, $p<.001$. The N1 for stimuli presented in the RVF ($M=197.36$, $SD=21.27$) was earlier than for those presented in the LVF ($M=207.2$, $SD=20.11$). There was also an interaction of hemisphere and VF, $F(1,18)=167.72$, $p<.001$. The left hemisphere N1 was earlier for RVF ($M=179.14$, $SD=10.71$) than LVF ($M=224.87$, $SD=10$) stimuli, $p<.001$. In the right hemisphere, the N1 for stimuli presented in the LVF ($M=189.52$, $SD=8.58$) was earlier than for those presented in the RVF ($M=215.58$, $SD=10.7$), $p<.001$.

N2/P3. Using mean amplitude measures, a 2(Report: global, local) x 2(hemisphere: left, right) x 2(visual field presentation: LVF, RVF) repeated measures ANOVA revealed an interaction of hemisphere and VF, $F(1,18)=5.66$, $p=.03$. In the left hemisphere there was a difference between stimuli presented in the LVF (M=79.59, $SD=154.74$) and RVF (M=40.93, $SD=150.47$). There was no difference in amplitudes in the right hemisphere.

4.4.2.4.2 Upper and lower VF presentation

P1. Using peak amplitude measures, a 2(Report: global, local) x 2(hemisphere: left, right) x 2(visual field presentation: upper visual field (Upper VF), lower visual field (Lower VF)) repeated measures ANOVA revealed a main effect of hemisphere, $F(1,18)=48.29$, $p<.001$. Amplitudes in the right hemisphere ($M=2.08$, $SD=1.1$) were greater than those in the left ($M=-0.29$, $SD=0.77$). There were no other main effects of interactions.

Latency. No significant main effects or interactions were found

N1. Using peak amplitude measures, a 2(Report: global, local) x 2(hemisphere: left, right) x 2(visual field presentation: Lower VF, Upper VF) repeated measures ANOVA revealed a main effect of hemisphere, $F(1,18)=52.03$, $p<.001$. There was a more negative N1 in the right ($M=-3.86$, $SD=2.03$) than the left hemisphere ($M=-1.43$, $SD=1.65$). There was also a main effect of VF, $F(1,18)=22.47$, $p<.001$. There was a more negative N1 for stimuli presented in the lower VF ($M=3.29$, $SD=1.99$) than the upper VF ($M=-2.01$, $SD=1.6$).

Latency. No significant main effects or interactions were found

N2/P3. Using a time window identified in MUA (350-450ms), then mean amplitude, a 2(Report: global, local) x 2(hemisphere: left, right) x 2(visual field presentation: Lower VF, Upper VF) repeated measures ANOVA revealed a main effect of VF, $F(1,18)=7.37$, $p=.015$, with greater amplitude for stimuli presented in the upper VF ($M=74.97$, $SD=96.18$) than the lower VF ($M=38.52$, $SD=103.98$).

4.4.2.5 Additional analyses: spacing and contrast differences

We were interested to see if the spacing and contrast manipulations in the displays affected global/local processing.

4.4.2.5.1 Maximal global vs minimal global for left and right VF presentation

Using only 'report global' trials, looking at maximally global condition (Low contrast, densely spaced (LCD)) versus the minimally global condition (High contrast, sparsely spaced (HCS)).

P1. Using peak amplitude measures, a 2(Condition: LCD, HCS) x 2(hemisphere: left, right) x 2(visual field presentation: LVF, RVF) repeated measures ANOVA revealed an interaction of hemisphere and VF – $F(1,18)=6.12$, $p=.024$. Amplitudes in the left hemisphere were greater for stimuli presented in the RVF ($M=1.62$, $SD=1.46$) than LVF ($M=0.68$, $SD=1.47$), *p*=.013. There was no difference between stimuli presented in the LVF and RVF in the right hemisphere.

Latency. A 2(Condition: LCD, HCS) x 2(hemisphere: left, right) x 2(visual field presentation: LVF, RVF) repeated measures ANOVA revealed a main effect of condition, $F(1,18)=7.18$, $p=.015$. The P1 for HCS ($M=131.23$, $SD=24.13$) was earlier than for LCD ($M=135.15$, $SD=25.97$) global condition. There was also an interaction of hemisphere and VF (for both global and local report, see above).

N1. Using peak amplitude measures, a 2(Condition: LCD, HCS) x 2(hemisphere: left, right) x 2(visual field presentation: LVF, RVF) repeated measures ANOVA revealed a main effect of hemisphere, $F(1,18)=7.66$, $p<.013$. There was a more negative N1 in the right hemisphere ($M=-3.93$, $SD=2.6$) than the left ($M=-2.77$, $SD=2.66$).

Latency. A 2(Condition: LCD, HCS) x 2(hemisphere: left, right) x 2(visual field presentation: LVF, RVF) repeated measures ANOVA revealed a main effect of condition, $F(1,18)=15.51$, $p=.001$. The N1 for HCS ($M=200.04$, $SD=21.04$) was earlier than for LCD ($M=205.3$, $SD=22.28$). There was a main effect of VF, $F(1,18)=44.29$, $p<.001$. The N1 for stimuli presented in the RVF ($M=197.46$, $SD=20.9$) was earlier than for those presented in the LVF ($M=207.56$, $SD=21.63$). There was also an interaction of hemisphere and VF (for both global and local report, see above).

N2/P3. Using mean amplitude measures, a 2(Condition: LCD, HCS) x 2(hemisphere: left, right) x 2(visual field presentation: LVF, RVF) repeated measures ANOVA revealed a main effect of condition, $F(1,18)=8.94$, $p=.009$, with greater amplitudes for LCD stimuli (M=94.74, $SD=147.86$) than HCS stimuli (M=35.14, $SD=144.34$).

4.4.2.5.2 Maximal local vs minimal local for left and right VF presentation

Using only 'report local' trials, looking at maximally local condition (High contrast, sparsely spaced (HCS)) versus the minimally local condition (Low contrast, densely spaced (LCD)).

P1. Using peak amplitude measures, a 2(Condition: LCD, HCS) x 2(hemisphere: left, right) x 2(visual field presentation: LVF, RVF) repeated measures ANOVA revealed an interaction of hemisphere and VF, $F(1,18)=5.94$, $p=.025$. Amplitudes in the left hemisphere were greater for stimuli presented in the RVF ($M=1.44$, $SD=1.35$) than LVF ($M=0.52$, $SD=1.63$), $p=.009$. There was no difference between stimuli presented in the LVF and RVF in the right hemisphere.

Latency. A 2(Condition: LCD, HCS) x 2(hemisphere: left, right) x 2(visual field presentation: LVF, RVF) repeated measures ANOVA revealed a main effect of condition, $F(1,18)=7.4$, $p=.014$. The P1 was earlier for HCS ($M=132.71$, $SD=23.9$) than LCD ($M=136.06$, $SD=27.98$). There was also an interaction of condition, hemisphere and VF, $F(1,18)=11.68$, $p=.003$. In the left hemisphere, stimuli presented in the LVF showed an earlier P1 for HCS ($M=153.22$, $SD=11.35$) than LCD stimuli ($M=160.72$, $SD=12.81$), $p=.008$. In the right hemisphere, stimuli presented in the RVF showed earlier P1 for HCS ($M=154.71$, $SD=11.37$) than LCD stimuli ($M=162.37$, $SD=12.23$), $p=.018$.

N1. Using peak amplitude measures, a 2(Condition: LCD, HCS) x 2(hemisphere: left, right) x 2(visual field presentation: LVF, RVF) repeated measures ANOVA revealed a main effect of condition, $F(1,18)=6.61$, $p=.019$. There was a more negative N1 for HCS ($M=-3.64$, $SD=2.62$) than LCD stimuli ($M=-2.81$, $SD=2.45$). There was a main effect of hemisphere, $F(1,18)=5.81$, $p=.027$. There was a more negative N1 in the right hemisphere ($M=-3.73$, $SD=2.67$) than the left ($M=-2.72$, $SD=2.36$). There was also an interaction of condition and hemisphere, $F(1,18)=6.06$, $p=.024$. There was greater negativity for HCS ($M=-4.3$, $SD=2.66$) than LCD ($M=-3.16$, $SD=2.6$) stimuli in the right hemisphere, $p=.005$, but not in the left hemisphere.

Latency. A 2(Condition: LCD, HCS) x 2(hemisphere: left, right) x 2(visual field presentation: LVF, RVF) repeated measures ANOVA revealed a main effect of VF, $F(1,18)=40.29$, $p<.001$. The N1 for stimuli presented in the RVF ($M=196.93$, $SD=23.14$) was earlier than for those presented in the LVF ($M=206.83$, $SD=220.86$). There was an interaction of condition, hemisphere and VF, $F(1,18)=7.55$, $p=.013$. In the left hemisphere, stimuli presented in the RVF showed an earlier N1 for LCD ($M=175.06$, $SD=13.94$) than HCS stimuli ($M=180.72$, $SD=11.72$), $p=.028$, but no differences for LVF stimuli. Also, no differences in the right hemisphere. There was also an interaction of hemisphere and VF (for both global and local report, see above).

N2/P3. Using mean amplitude measures, a 2(Condition: LCD, HCS) x 2(hemisphere: left, right) x 2(visual field presentation: LVF, RVF) repeated measures ANOVA revealed a main effect of condition, $F(1,18)=9.44$, $p=.007$, with greater amplitudes for LCD stimuli (M=72.82, $SD=118.71$) than HCS stimuli (M=34.54, $SD=137.21$). There was also an interaction between hemisphere and VF, $F(1,18)=6.07$, $p=.025$. In the left hemisphere there is a difference between stimuli presented in the LVF (M=81.22, $SD=145.01$) and RVF (M=43.78, $SD=137.17$). There is no difference between stimuli presented in the left and right VFs in the right hemisphere.

4.4.2.5.3 Maximal vs. minimal global for upper and lower VF presentation

Using only 'report global' trials, looking at maximally global condition (Low contrast, densely spaced (LCD)) versus the minimally global condition (High contrast, sparsely spaced (HCS)).

P1. Using peak amplitude measures, a 2(Condition: LCD, HCS) x 2(hemisphere: left, right) x 2(visual field presentation: Lower VF, Upper VF) repeated measures ANOVA revealed an main effect of hemisphere, $F(1,18)=36.17$, $p<.001$. Amplitudes in the right hemisphere ($M=2.03$, $SD=1.34$) were greater than those in the left ($M=-0.33$, $SD=0.83$).

Latency. A 2(Condition: LCD, HCS) x 2(hemisphere: left, right) x 2(visual field presentation: Lower VF, Upper VF) repeated measures ANOVA revealed a main effect of condition, $F(1,18)=8.27$, $p=.011$. The P1 for HCS ($M=114.21$, $SD=10.2$) was earlier than for LCD ($M=124.42$, $SD=7.91$) global condition. There was also an interaction of condition and VF, $F(1,18)=7.49$, $p=.015$. For stimuli presented in the upper VF, there was an earlier P1 for HCS ($M=111.33$, $SD=12.2$) than LCD stimuli ($M=126.44$, $SD=11.73$), $p<.001$. There was no difference between stimuli presented in the lower VF.

N1. Using peak amplitude measures, a 2(Condition: LCD, HCS) x 2(hemisphere: left, right) x 2(visual field presentation: Lower VF, Upper VF) repeated measures ANOVA revealed a main effect of condition, $F(1,18)=7.23$, $p=.016$, with more negative N1 for HCS stimuli ($M=-3.09$, $SD=1.71$) than LCD ($M=-2.29$, $SD=2.04$). There was a main effect of hemisphere, $F(1,18)=34.2$, $p<.001$. There was a more negative N1 in the right hemisphere ($M=-3.89$, $SD=2.15$) than the left ($M=-1.49$, $SD=1.77$). There was also a main effect of VF, $F(1,18)=9.34$, $p=.008$, with greater negativity for stimuli presented in the lower VF ($M=-3.22$, $SD=2.25$) than the upper VF ($M=-2.17$, $SD=1.52$). There was an interaction between condition and VF, $F(1,18)=5.13$, $p=.038$. For stimuli presented in the upper VF, there was greater negativity for HCS stimuli ($M=-2.93$, $SD=1.64$) than LCD stimuli ($M=-1.4$, $SD=1.75$), $p=.001$. There were no differences between stimuli presented in the lower VF.

Latency. No significant main effects or interactions were found

N2/P3. Using mean amplitude measures, a 2(Condition: LCD, HCS) x 2(hemisphere: left, right) x 2(visual field presentation: Lower VF, Upper VF) repeated measures ANOVA revealed a main effect of condition, $F(1,18)=8.98$, $p=.009$, with greater amps for LCD stimuli ($M=94.45$, $SD=120.22$) than HCS stimuli ($M=34.79$, $SD=108.91$). There was also a main effect of VF, $F(1,18)=4.61$, $p=.047$, with greater amplitudes for stimuli presented in the upper VF ($M=81.4$, $SD=108.79$) than the lower VF ($M=47.84$, $SD=114.83$).

4.4.2.5.4 Maximal vs. minimal local for upper and lower VF presentation

Using only 'report local' trials, looking at maximally local condition (High contrast, sparsely spaced (HCS)) versus the minimally local condition (Low contrast, densely spaced (LCD)).

P1. Using peak amplitude measures, a 2(Condition: LCD, HCS) x 2(hemisphere: left, right) x 2(visual field presentation: Lower VF, Upper VF) repeated measures ANOVA revealed a main effect of hemisphere, $F(1,18)=40.51$, $p<.001$. Amplitudes in the right hemisphere ($M=2.12$, $SD=0.93$) were greater than those in the left ($M=-0.25$, $SD=1.24$).

Latency. There were no main effects of interactions

N1. Using peak amplitude measures, a 2(Condition: LCD, HCS) x 2(hemisphere: left, right) x 2(visual field presentation: Lower VF, Upper VF) repeated measures ANOVA revealed a main effect of condition, $F(1,18)=5.34$, $p=.035$, with more negative amplitudes for HCS ($M=-2.98$, $SD=1.96$) than LCD stimuli ($M=-2.23$, $SD=1.85$). There was a main effect of hemisphere, $F(1,18)=56.97$, $p<.001$, with more negative amplitudes in the right hemisphere ($M=-3.84$, $SD=2.04$) than the left ($M=-1.37$, $SD=1.76$). There was also a main effect of VF, $F(1,18)=27.64$, $p<.001$, with greater negativity for stimuli presented in the lower VF ($M=-3.36$, $SD=1.9$) than upper VF ($M=-1.85$, $SD=1.86$).

Latency. No significant main effects or interactions were found

N2/P3. Using mean amplitude measures, a 2(Condition: LCD, HCS) x 2(hemisphere: left, right) x 2(visual field presentation: Lower VF, Upper VF) repeated measures ANOVA revealed a main effect of condition, $F(1,18)=8.04$, $p=.012$, with greater amplitudes for LCD ($M=68.92$, $SD=79.8$) than HCS stimuli ($M=-28.82$, $SD=108.18$). There was also a main effect of VF, $F(1,18)=7.84$, $p=.013$, with greater amplitudes for stimuli presented in the upper VF ($M=68.54$, $SD=90.67$) than the lower VF ($M=29.2$, $SD=99.14$).

To see if any differences found between the conditions found in the ERPs were due to the contrast and spacing, we collapsed across local and global report and congruency to compare contrast and spacing conditions.

4.4.2.5.5 Contrast and spacing differences in left and right visual field presentation

P1. Using peak amplitude measures, a 2(condition: LCD, HCS) x 2(hemisphere: left, right) x 2(visual field presentation: LVF, RVF) repeated measures ANOVA revealed an interaction of condition and hemisphere, $F(1,18)=4.6$, $p=.046$. There were greater amplitudes for LCD ($M=1.25$, $SD=1.17$) than HCS stimuli ($M=0.88$, $SD=1.52$) in the left hemisphere, $p=.009$, but no differences in the right hemisphere.

Latency. A 2(condition: LCD, HCS) x 2(hemisphere: left, right) x 2(visual field presentation: LVF, RVF) repeated measures ANOVA revealed a main effect of condition, $F(1,18)=10.25$, $p=.005$, with an earlier P1 for HCS ($M=131.61$, $SD=23.54$) than LCD stimuli ($M=135.97$, $SD=26.54$). There was also an interaction between condition, hemisphere and VF, $F(1,18)=18.58$, $p<.001$. In the left hemisphere, for stimuli presented in the LVF, there was an earlier P1 for HCS ($M=153.37$, $SD=8.77$) than LCD stimuli ($M=159.67$, $SD=11.63$), $p=.004$. There were no differences for RVF stimuli. In the right hemisphere, for stimuli presented in the RVF, there was an earlier P1 for HCS ($M=153.71$, $SD=10.7$) than LCD stimuli ($M=160.67$, $SD=10.15$), $p=.01$.

N1. Using peak amplitude measures, a 2(condition: LCD, HCS) x 2(hemisphere: left, right) x 2(visual field presentation: LVF, RVF) repeated measures ANOVA revealed a main effect of condition, $F(1,18)=17.22$, $p=.001$, with more negative amplitudes for HCS ($M=-3.58$, $SD=2.46$) than LCD stimuli ($M=-2.99$, $SD=2.34$). There was a main effect of hemisphere, $F(1,18)=7.12$, $p=.016$, with more negative amplitudes in the right ($M=-3.83$, $SD=2.46$) than left hemisphere ($M=-2.75$, $SD=2.26$). There was also an interaction of condition and hemisphere, $F(1,18)=4.86$,

p=.0141. In the right hemisphere, there was a more negative N1 for HCS (M=-4.27, SD=2.49) than LCD stimuli (M=-3.38, SD=2.37), p=.001. There was no difference in the left hemisphere.

Latency. A 2(condition: LCD, HCS) x 2(hemisphere: left, right) x 2(visual field presentation: LVF, RVF) repeated measures ANOVA revealed a main effect of condition, F(1,18)=7.21, p=.015, with an earlier N1 for HCS (M=200.93, SD=19.45) than LCD stimuli (M=203.62, SD=23.22).

N2/P3. Using mean amplitude measures, a 2(condition: LCD, HCS) x 2(hemisphere: left, right) x 2(visual field presentation: LVF, RVF) repeated measures ANOVA revealed a main effect of condition, F(1,18)=12.5, p=.003, with greater amps for LCD stimuli (M=83.78, SD=133.54) than HCS stimuli (M=34.84, SD=139.78). There was also an interaction between hemisphere and VF, F(1,18)=5.66, p=.03. In the left hemisphere there is a difference between stimuli presented in the LVF (M=79.59, SD=153.59) and RVF (M=40.93, SD=149.59). There is no difference between stimuli presented in the left and right VFs in the right hemisphere.

4.4.2.5.6 Contrast and spacing differences in upper and lower VF presentation

P1. Using peak amplitude measures, a 2(condition: LCD, HCS) x 2(hemisphere: left, right) x 2(visual field presentation: Lower VF, Upper VF) repeated measures ANOVA revealed a main effect of hemisphere, F(1,18)=48.29, p<.001, with greater amplitudes for the right hemisphere (M=2.08, SD=1.1) than left (M=-0.29, SD=0.77).

Latency. A 2(condition: LCD, HCS) x 2(hemisphere: left, right) x 2(visual field presentation: Lower VF, Upper VF) repeated measures ANOVA revealed a main effect of condition, F(1,18)=13, p=.002, with an earlier P1 for HCS (M=117.03, SD=7.35) than LCD stimuli (M=122.09, SD=6.95).

N1. Using peak amplitude measures, a 2(condition: LCD, HCS) x 2(hemisphere: left, right) x 2(visual field presentation: Lower VF, Upper VF) repeated measures ANOVA revealed a main effect of condition, $F(1,18)=20.52$, $p<.001$, with more negative amplitudes for HCS (M=-3.04, SD=1.75) than LCD stimuli (M=-2.26, SD=1.75). There was a main effect of hemisphere, $F(1,18)=52.03$, $p<.001$, with more negative amplitudes in the right hemisphere (M=-3.86, SD=2.03) than the left (M=-1.43, SD=1.65). There was also a main effect of VF, $F(1,18)=22.47$, $p<.001$, with more negative amplitudes for stimuli presented in the lower VF (M=-3.29, SD=1.99) than upper VF (M=-2.01, SD=1.6).

Latency. There were no main effects or interactions.

N2/P3. Using mean amplitude measures, a 2(condition: LCD, HCS) x 2(hemisphere: left, right) x 2(visual field presentation: Lower VF, Upper VF) repeated measures ANOVA revealed a main effect of condition, $F(1,18)=12$, $p=.003$, with greater amplitudes for LCD (M=81.68, SD=96.73) than HCS stimuli (M=31.8, SD=104.58). There was also a main effect of VF, $F(1,18)=7.37$, $p=.015$, with greater amplitudes for stimuli presented in the upper VF (M=74.97, SD=96.18) than lower VF (M=38.52, SD=103.98).

4.5 Discussion

The aim of the study was to find an ERP signature for global/local integration using the congruency effect (global interference) as a functional marker. The main findings can be summarised as follows: firstly, we found that reaction times for congruent trials were faster than for incongruent trials and error rate was higher for incongruent trials. There was no difference in interference levels between local report and global report trials. Secondly, we found that HCS stimuli were processed differently than LCD stimuli, evidenced by amplitude differences at the N1 and N2/P3, but this did not interact with local and global report. Thirdly, we found congruency effects at the P1 and N1 for both local and global report, with differences between congruent and incongruent stimuli. Fourth, we found evidence of global interference in the ERP data, at the N2/P3; congruency effects were greater for local than global report. Global interference effects were also clear in the mass univariate analyses, with greater frequencies of congruency differences in local compared to global report in this time frame.

The behavioural data did not provide support for the global precedence effect. There were higher error rates for incongruent trials and faster RTs for congruent stimuli. However, we observed this effect for both local report and global report trials. This is contrary to many studies that find a classic behavioural GPE (e.g. Beaucousin et al., 2013; Navon, 1977). The reason for the discrepancy in results (the local interference) may be due to the nature of the stimuli and task we used. The task required attention to both local and global levels for all trials: as when the task was to identify the global orientation, identification of the local orientation was still necessary to identify a coherent global line. Similarly, when reporting the local orientation; the participant had to, first, ensure that all local elements were the same orientation in the global line. In a classic Navon task there would not usually be an interference effect for global report trials, this is because global processing is proposed to be automatic, and therefore is not affected by the incongruent local elements. Due to the automaticity of global processing, when global and local elements are incongruent and the task is to report the local level, the involuntary attention to the global level affects reaction time and accuracy for local report. Therefore, due to the forced attention to both levels of stimulus in our task, effects of global precedence may have been disguised. Several other studies have failed to replicate GPE (e.g. Martens & Hübner, 2013; Roalf, Lowery & Turetsky, 2006). It may be that Navon letters, or similar hierarchical stimuli are problematic in that they are too artificial, and therefore the effects may be paradigmatic. This is evidenced by the widely varying findings from studies using slightly different stimuli and tasks. There is a discrepancy between the behavioural and ERP data, as we observed no global interference in the behavioural data. This can be explained by the stimuli and task used, as the task forced attention to both global and local levels of the stimuli, therefore a global processing advantage may have been masked in the behavioural data.

Several studies using bilateral stimulus presentation have found that RTs are faster for stimuli presented in the left than the right visual field for global targets (Flevaris, Bentin & Robertson, 2010; Schlösser, Hübner & Studer, 2009; Van Kleeck, 1989; Volberg & Hübner, 2006). We did not find evidence of hemispheric asymmetry for local and global processing, again, this may be due to the stimuli. We did, however, an RT and accuracy effect for the lower visual field, driven by the lower right visual field, but this did not interact with target level. Studies have provided evidence of latency advantage in the lower VF (e.g. Gawryszweski, Riggio, Rizzolatti & Umilta, 1987; Levine & McAnany, 2005; Rizzolatti, Riggio, Dascola & Umilta, 1987). This is proposed to be because of an advantage in luminance and contrast threshold sensitivity in the lower visual field (Lundh, Lennerstrand & Derefeldt, 1983; Murray, MacCana & Kulikowski, 1983; Rijsdijk, Kroon & van der Wildt, 1980).

Hemispheric asymmetries are frequently reported in ERP studies using Navon tasks (Heinz et al., 1998; Leek et al., 2016; Mangun et al., 2000; Volberg & Hübner, 2004; Yamaguchi, Yamagata & Kobayashi, 2000). Although we did not find evidence of hemispheric asymmetry for local and global report per se, we did observe hemispheric differences for local and global interference at the P1. In local report trials, there was a congruency effect, but only in the right hemisphere, whereas in global report trials, the congruency effect was only observed in the left hemisphere. These differences may reflect interference from information at the unattended target level, therefore right hemisphere effects reflect global interference in report local trials, and vice versa.

Though we did not find evidence of early differences between local and global report conditions, there were early differences between HCS and LCD stimuli, with earlier P1 for HCS and an amplitude difference at the N1 and the P3. These findings are similar to those of Craddock, Martinovic and Müller (2013), using individual objects or scenes, found P1 and N1 sensitivity to high and low spatial frequencies, with faster responses for high spatial frequency than low spatial frequency stimuli.

We found congruency effects in local and global report conditions at the P1 (at around 120ms) and the N1 (at around 200ms). There were also congruency effects at the N2/P3 (from around 350ms), but only for local report. This is evidence of global interference, as there were considerably more differences in activity for incongruent trials in the local report condition than the global report condition, indicating that the local information in the incongruent global report trials did not interfere significantly with global processing. However, due to the differences in local report trials, it seems that global information was interfering with the local processing. The effects replicate findings from previous studies with global interference at the P1 (Han et al., 1997; 2000; Jiang and Han, 2005) and N1 (e.g. Beaucousin, Simon, Cassotti, Pineau, Houde & Poirel, 2013; Han et al., 2003). A later component around the N2 and P3 has frequently been reported to reflect differences in local and global processing with hierarchical stimuli (e.g. Han, He & Woods, 2000; Han, Yund & Woods, 2003; Heinze et al., 1998; Heinze & Münte, 1993; Malinowksi et al., 2002; Volberg & Hübner, 2004; Yamagucchi, Yamagara & Kobayashi, 2000).

In Chapter III we found evidence of processing of both local and global shape information at the N1. It is difficult to compare findings regarding local and global processing as Experiment 1 (Chapter III) used 3D objects and a recognition memory task, whereas the present experiment uses very basic visual stimuli and a Navon task. We did not find evidence of differential processing of local and global information in the present experiment, however, we did observe global and local differences with regard to congruency at the N2/P3. We suggest that this congruency effect reflects global-to-local interference, which can also be described as the integration of local and global information.

The observed data seem to show one of the two elements of the global precedence effect, as global information appeared to interfere with processing of local information. This has implications for theories of object recognition, as it implies that global information is processed preferentially to local information. Several models of object recognition focus on local to global processing, for example, the HMAX model (Reisenhuber & Poggio, 1999) and structural descriptions accounts tend to include the detection of local elements such as edges or surfaces as the first steps in object recognition (Biederman, 1987; Leek et al., 2005; Marr & Nishihara, 1978). However, our data fit with Bar's (2003) model of object recognition, whereby global (or low spatial frequency) information is processed first along a dorsal route, whereas local information is processed along a slower ventral route, with feedback from the global information to update information about the image. Due to the necessity to process both levels of the stimulus in all trials, we suggest that the interference observed for the incongruent trials at the N2/P3 reflects the integration of global and local levels of the stimulus, as this is required to make the orientation decision.

To summarise, we found that global interference occurred at the N2/P3, around 350ms post-stimulus onset. We propose that this global interference effect reflects the integration of local and global information as it is only present in the condition where the presence of global information interferes with local report. These findings fit with models of object processing that suggest early processing of local and global information occurs in parallel. As global/local integration appears to occur at the N2/P3 in a Navon task using basic visual stimuli, we were interested to see if this effect would generalise to a different paradigm, an image classification task.

4.6 Summary

- This study examined the integration of information at local and global spatial scales.
- We used a Navon-style task with very basic visual stimuli made up of Gabor patches whilst recording ERPs.
- The main results showed that basic visual stimuli elicited both local and global interference effect and that local/global integration occurs from around 350ms post-stimulus onset (N2/P3).

- The results provide partial support for the global precedence effect using very basic visual stimuli.

5 Chapter V

Integrating local and global shape information in impossible objects: A high-density ERP study

Previously, in Chapter IV, we reported a congruency effect at the N2/P3, presumably reflecting local/global integration. The aim of the following experiment was to see if a similar ERP signature, reflecting the integration of local and global information, was present in complex objects without Navon hierarchical displays. The remainder of this chapter is in manuscript form.

5.1 Abstract

The aim of the study was to find an ERP signature for the integration of local and global shape information using complex objects. One way to investigate the integration of local and global information is with the use of impossible objects. Impossible objects defy the laws of geometry; they are 2D drawings that represent objects that cannot exist in 3D space. The difficulty in perception of impossible objects stems from the inability to form the representation of a coherent 3D structure, as one that exists in the 3D world, does not exist. Our impossible stimuli included local and global features that, alone, were possible, however, when combined, or integrated, to form a 3D representation, revealed the impossibility. Using event-related potentials (ERPs), we compared the processing of possible and impossible objects in a simple classification task. We found that there were no early processing differences for possible and impossible objects. However, there was a latency difference at the N2 (around 250ms post-stimulus onset) and another difference in amplitude at the P3 (from around 300ms post-stimulus onset). We propose that the differences that arise at the N2 and P3 encompass an ERP signature of global/local integration.

5.2 Introduction

The visual system is able to perform object classification extremely rapidly (Fabre-Thorpe, 2011; Johnson & Olshausen, 2005; 2003; Rousselet et al., 2004; Thorpe & Fabre-Thorpe, 2001) and our ability to recognise three-dimensional (3D) objects despite changes in sensory input is remarkable (e.g., Arguin & Leek, 2003; Bar, 2003; Bar, Kassam, Ghuman et al., 2006; Cichy, Pantazis & Oliva, 2014; Fabre-Thorpe, 2011; Harris, Dux, Benito & Leek, 2008; Kirchner & Thorpe, 2006; Leek, 1998a; 1998b; Leek, Atherton & Thierry, 2007; Leek, Davitt & Cristino, 2015; Leek & Johnston, 2006; Leek, Roberts, Oliver, Cristino & Pegna, 2016; Tarr & Bulthoff, 1998; Thorpe, Fize & Marlot, 1996; VanRullen & Thorpe, 2001). But there are still some mysteries regarding the time course of perceptual processes in 3D objects**.**

An interesting aspect of object perception concerns the way in which we sample information from different spatial scales. Global features are those that can be detected at a coarse spatial scale, such as edge collinearity, elongation, symmetry, aspect ratio and global outline. Whereas, local object features are computed at a finer spatial scale, for example, edge boundaries, corners, surface depth, vertices, curvature, colour and texture. Findings from ERP studies have identified early modulations related to local and global processing at the P1 (Han, He & Woods, 2000), but more frequently around the N1 and P2 components, approximately 150-240ms post stimulus onset (e.g., Beaucousin, Simon, Cassotti et al., 2013; Proverbio, Minniti & Zani, 1998; Yamaguchi, Yamagata & Kobayashi, 2000). More recent research indicates that local and global processing of complex 3D objects occurs at the N1 (145-215ms), with differential activity when processing local part information and global spatial configuration of objects (Leek, Roberts, Oliver, Cristino & Pegna, 2016; Oliver, Cristino, Roberts, Pegna & Leek, 2017 (Chapter III). These results suggest that local and global object features are processed at this early stage, therefore, presumably there is a time-point where information from these different spatial scales are integrated for the formation of a 3D representation of an object and for recognition to occur.

Findings from Chapter IV indicate that the integration of local and global information occurs around the N2/P3, from around 350ms post-stimulus onset. This experiment used basic visual stimuli and a Navon task, whereby participants attended to either the local or global level of a stimulus with either congruent or incongruent information at the other level. We were interested to see if ERP correlates of global/local integration generalised to a different task and stimuli. An interesting way to investigate the integration of local and global information is with the use of impossible objects. Impossible objects violate the laws of geometry; they are 2D drawings that represent objects that cannot exist in 3D space. The difficulty in perception of impossible objects stems from the inability to form the representation of a coherent 3D structure, as one that exists in the 3D world, does not exist. Object impossibility occurs due to regions of objects where there are structural violations (Schacter, Cooper & Delaney, 1990; Soldan, Hilton & Stern, 2009; Williams & Tarr, 1997). Schacter et al. (1990; 1991) were the first to use impossible objects to investigate LTM representations. They suggested that a structural description system needed to be formed in order to understand objects and found that this could not be formed for impossible objects.

Freud et al. (Freud, Avidan & Ganel, 2013; Freud, Avidan & Ganel, 2015; Freud, Ganel & Avidan, 2013; Freud, Hadad, Avidan & Ganel, 2015) have investigated the underlying perceptual processes that mediate the representation of impossible objects. Freud et al. (2013a) used a Garner classification task with possible and impossible cubes and found Garner interference for both possible and impossible stimuli. They suggested that is evidence for holistic processing of both possible and impossible objects. Freud et al. (2015a), again, used a Garner speeded classification task with possible and impossible cubes. The stimulus images were filtered to contain mostly high or low spatial frequency information. They found that holistic representations can be created even without low SF information for possible, but not impossible objects. They, therefore, suggested that areas of impossibility tend to be high spatial frequency in nature; based on more detail and requiring more fine-grained analysis to interpret. Freud et al. (2013b) suggest that the global 3D structure is incoherent in impossible objects. On the one hand, we agree that the 3D structure is incoherent in impossible objects. However, our 'global' element is not the 3D structure, we define properties at a coarse spatial scale as global. We agree that the local elements in the objects are valid and possible, but we argue that our global element, the spatial configuration, is also possible and it is the combination of the local and global elements that reveals the inconsistencies and impossibility.

Previous work has provided evidence of the visual system's sensitivity to object impossibility, for example, 4-month old infants were found to look longer at impossible than possible objects (Shuwairi, Albert & Johnson, 2007) and cortical regions along the dorsal and ventral visual streams show sensitivity to object impossibility (Freud, Rosenthal, Ganel & Avidan, 2015). These regions include the inferior temporal gyrus, the right superior parietal gyrus and the fusiform gyrus (Wu, 2012). Shigemura, Yoshino, Kobayashi, Takahashi and Nomura (2004) also found that differences occur in the fusiform gyrus and that these differences occurred between 350 and 389ms. Growing evidence suggests that early perceptual processing of possible and impossible objects are similar, with differentiation between categories occurring later (Freud et al., 2013a; 2013b; Freud et al, 2015b; Shigemura et al., 2004). Evidence suggests that local and global properties of objects are processed early (e.g., Beaucousin, Simon, Cassotti et al., 2013; Han, He & Woods, 2000; Leek, Roberts, Oliver, Cristino & Pegna, 2016; Oliver, Cristino, Roberts, Pegna & Leek, in revision; Proverbio, Minniti & Zani, 1998; Yamaguchi, Yamagata & Kobayashi, 2000), we suggest, therefore, that the point at which processing of possible and impossible objects differs is the point of integration of these two types of information to form a 3D representation.

In Chapter 5, we identified a marker for integration of local and global information using low-level visual stimuli. We were interested to find out if this marker would generalise to a completely different paradigm. We believe that the detection of object impossibility should reflect the same process as local/global integration. When there are differences in the ERPs between possible and impossible objects, this reflects the inability to integrate local and global levels of the stimulus.

Our goal is to examine whether the N2/P3 temporal ERP signature for the perceptual integration of shape information across global and local scales, found in Chapter IV, generalises to a task involving image classification of geometrically possible and impossible objects. Using event-related potentials (ERPs), we will compare the processing of possible and impossible objects in a simple classification task. As the local and global elements, alone, are possible, determining whether an object is possible depends on the integration of local and global information. Therefore, the point where the ERPs differ between possible and impossible objects should reflect the integration of local and global components for a 3D representation to be made.

5.3 Methods

5.3.1 Participants

28 Bangor University students (24 female, mean age 21.46, *SD*=3.16, 3 left-handed) participated for course credit. The sample was recruited through an online participation portal. All participants had normal or corrected-to-normal visual acuity. Ethics approval was granted by Bangor University. Informed consent was obtained and participants were free to withdraw from the study at any time without penalty. The data from 3 participants were removed due to bad EEG recordings.

5.3.2 Stimuli

40 possible and 40 impossible objects (some adapted) from Williams and Tarr (1997) presented in 3 views with 120 degree intervals in viewing angle were used (see Figure 41.). Some stimuli were adapted to ensure that possible and impossible objects were matched for complexity in terms of contours and vertices, t-tests showed that object complexity did not

differ significantly: possible vertices (*M*=29.15, *SD*=5.69), impossible vertices (*M*=29.35, *SD*=5.65), *t*(39) = 1.275, *n.s*; possible contours (*M*=38.98, *SD*=7.9), impossible contours (*M*=38.63, *SD*=7.78), *t*(39) = 1.617, *n.s*. Stimuli were scaled to260x260 pixels, Stimuli subtended 7.7 x 7.7 degrees of visual angle. Stimuli were shown on a 60Hz, 27" AOC 3D monitor (D2769VH), at a resolution of 1920x1080.

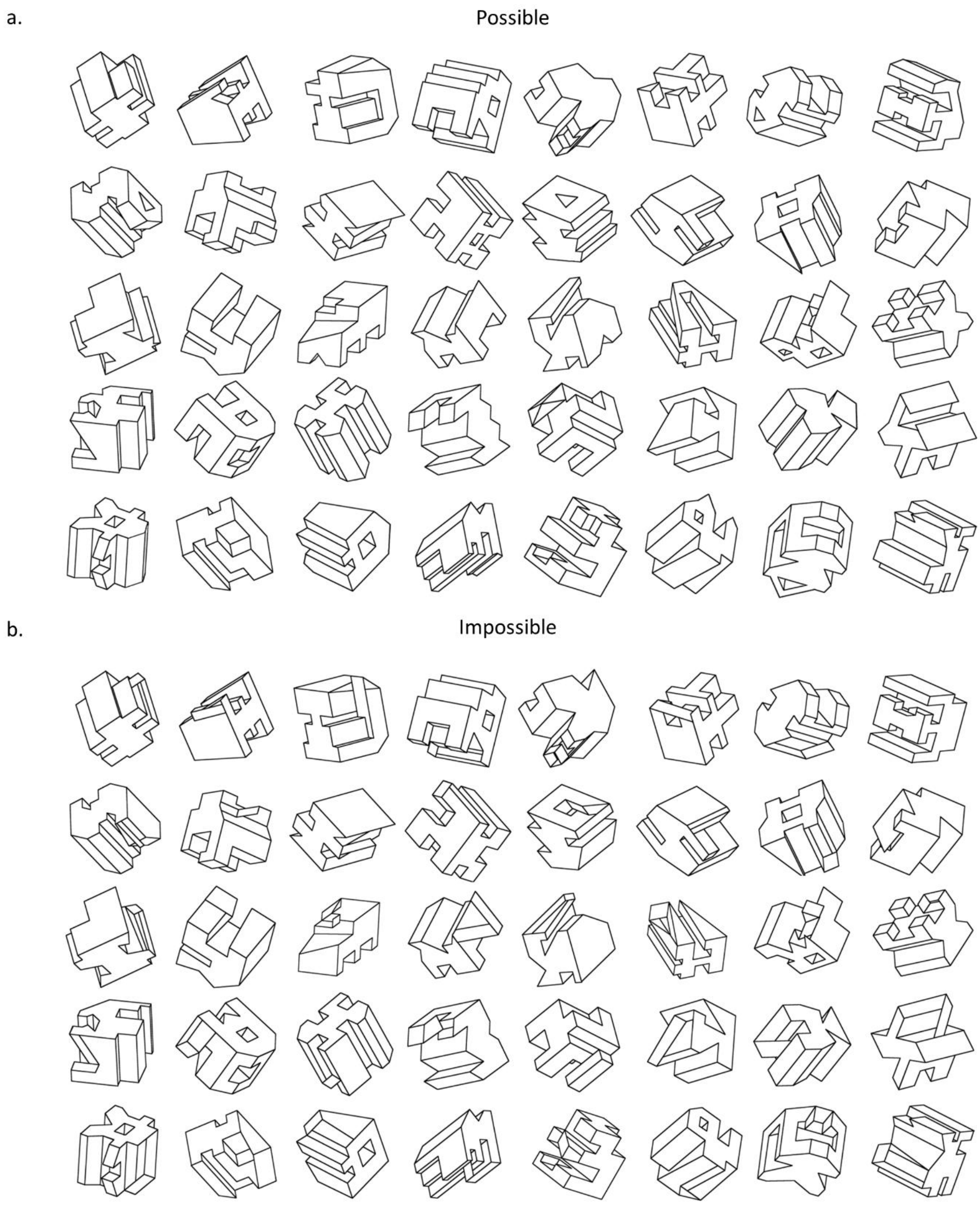


Figure 41. The stimulus set used in Experiment 3 (adapted from Williams and Tarr, 1997) comprising 40 possible and 40 impossible objects. The original set was modified to equate low-level image statistics (N contours and vertices).

5.3.3 Design

A within subjects design was used, with object type (possible/impossible) as a factor. There were 120 trials for each object type comprising each of the 40 objects presented once at three orientations (0, 120 and 240 degrees). Total number of trials was 240. Stimulus presentation order was randomised.

5.3.4 Procedure

ERPs were recorded while participants performed an object decision task, in which an image randomly selected from one of the 80 objects (40 possible and 40 impossible) was presented and participants decided if the image was geometrically possible or impossible. Images were shown at 3 different viewpoints to increase the number of trials.

The trial sequence was as follows: first, a small central fixation cross was shown on the screen. Fixation duration was jittered, lasting between 500 and 800ms. Second, following this, the test image was shown for 1000ms. The stimulus was replaced by a response prompt (question mark). Participants were asked to press a button on a keyboard indicating whether the object was possible or impossible, “1” for possible, “2” for impossible, the response screen remained on the screen until a response was made (see Figure 42). Button order was not counterbalanced, however, as RTs were not recorded, this should not be an issue. RTs were not recorded to reduce any motor contamination of the ERP recording. Recording started 100ms before stimulus onset.

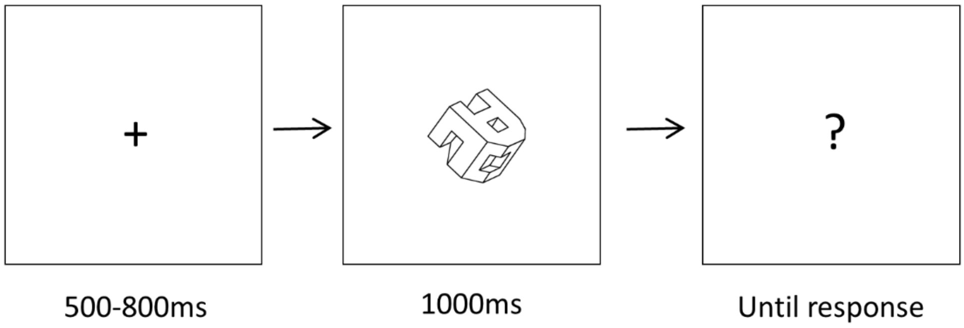


Figure 42. Example of the trial layout for a possible object, jittered fixation from 500-800ms, then stimulus presentation for 1000ms, then a response screen with question mark until a response is made.

5.3.5 Electrophysiological recording and processing

The electroencephalograph (EEG) was recorded continuously through 128 electrodes placed on an ECI cap (Electro-Cap International, Ohio, USA) using the Active-Two Biosemi EEG system (Biosemi V.O.F Amsterdam, Netherlands). Eye movements and blinks were corrected using the ICA protocol in Analyser 2 software and segmented data was then visually inspected with trials containing artefacts rejected. Epochs that contained muscle or skin potential artefacts were rejected. Only trials on which participants gave a correct response were included and the mean number of correct trials after artefact rejection was: 106 (possible), 84.5 (impossible). Activity from all electrodes was sampled at a rate of 1024Hz. Offline 30 Hz (48 db/oct slope) lowpass and 0.1 Hz (48 db/oct slope) highpass filters were applied to the data. All data was re-referenced to an average reference which was then used to generate the grand averages. We used a 100ms pre-stimulus interval for the baseline correction. Continuous recording took place during the test phase of the experiment and trials were epoched/segmented from 100ms pre-stimulus to 1000ms post-stimulus onset.

5.3.6 ERP analyses

Five early ERP waves (P1, N1, P2, N2 and P3) were identified based on the topography, global field power (GFP), deflection and latency characteristics of the respective grand average ERPs time-locked to stimulus presentation. Preliminary epochs of interest for each wave were defined on the basis of deflection extrema in mean local field power (e.g., Brunet, Murray & Michel, 2011; Lehmann & Skrandies, 1980; Murray, Brunet & Michel, 2008). Peak detection was time-locked to the electrode of maximal amplitude for each wave. The latency of peak amplitude was used to define time epochs for analyses of the waves: P1 (75-125ms; Peak latency (B6) = 100ms); N1 (135–185ms; Peak latency (B7) = 160ms); P2 (230-260ms; Peak latency (D28) = 245ms); N2 (250-290ms; Peak latency (A24) = 270ms); P3 (300-390ms; Peak latency (A4) = 345ms).

Two symmetrical clusters over the left (LH) and right (RH) hemispheres were extracted each consisting of 8 spatially adjacent posterior electrodes: RH: A32, B3, B5, B6, B7, B8, B18, B19 and LH: A5, A6, A8, A9, A10, A11, D17, D28, and central electrodes: A4, A19, A23, A24 and A25, which correspond/overlap with electrode locations P4, P6, PO8, P2, CP4 and P3, P5, PO7, P1, CP3, and central Pz and Oz, of the extended 10–20 system respectively. These electrode clusters formed the region-of-interest for the subsequent analysis of local and global integration, namely possible vs. impossible conditions. Mean amplitudes were analysed using the General Linear Model by way of repeated measures ANOVA. Greenhouse-Geisser corrections were applied to all analyses of ERP data.

5.3.7 Mass Univariate Analyses

Mass Univariate analyses (e.g., Groppe, Urbach & Kutas, 2011; Guthrie & Buchwald, 1991) were used to elucidate the time course of the integration of local and global shape properties. This involved using pair wise, time-frame by time-frame, permutation tests based on repeated measures t-tests across all 128 electrodes from 0-800ms. An a priori criterion for significance testing was adopted in which a threshold of $p<.01$ (two-tailed) must be attained for at least 10 consecutive time frames in at least 5 neighbouring electrodes (Guthrie & Buchwald, 1991; Murray, Brunet & Michel, 2008).

5.4 Results

The behavioural analysis and the standard waveform analysis was done on N=19 participants, having removed 6 participants with high error rates (based on *d'* scores).

5.4.1 Behavioural Analyses

A Wilcoxon signed-rank test revealed that accuracy (% correct) for possible objects (*Mdn*=87%, *SD*=7.16) was greater than that for impossible (*Mdn*=77%, *SD*=12.49) objects, Z=2.96, p=.003.

5.4.2 Analyses of waveforms

In the following analyses we wanted to investigate the time course of integration of local and global object features. Based on our hypothesis, differential processing of possible and impossible objects will indicate the time point where local and global shape features are integrated for recognition. We conducted both standard waveform analyses and mass univariate contrasts, only correct response trials were analysed.

P1. Using peak amplitude measures, with a 50ms time window (75-125ms). A 2(Condition: possible, impossible) x 2(Laterality) repeated measures ANOVA revealed that there was a main effect of hemisphere, $F(1,18) = 13.76$, $p = 0.002$, with greater amplitudes over the right hemisphere (*M*=3.26, *SD*=1.99) than the left hemisphere (*M*=1.53, *SD*=1.29), see Figure 43. There were no other main effects of interactions. There were no significant main effects or interactions for the latency data.

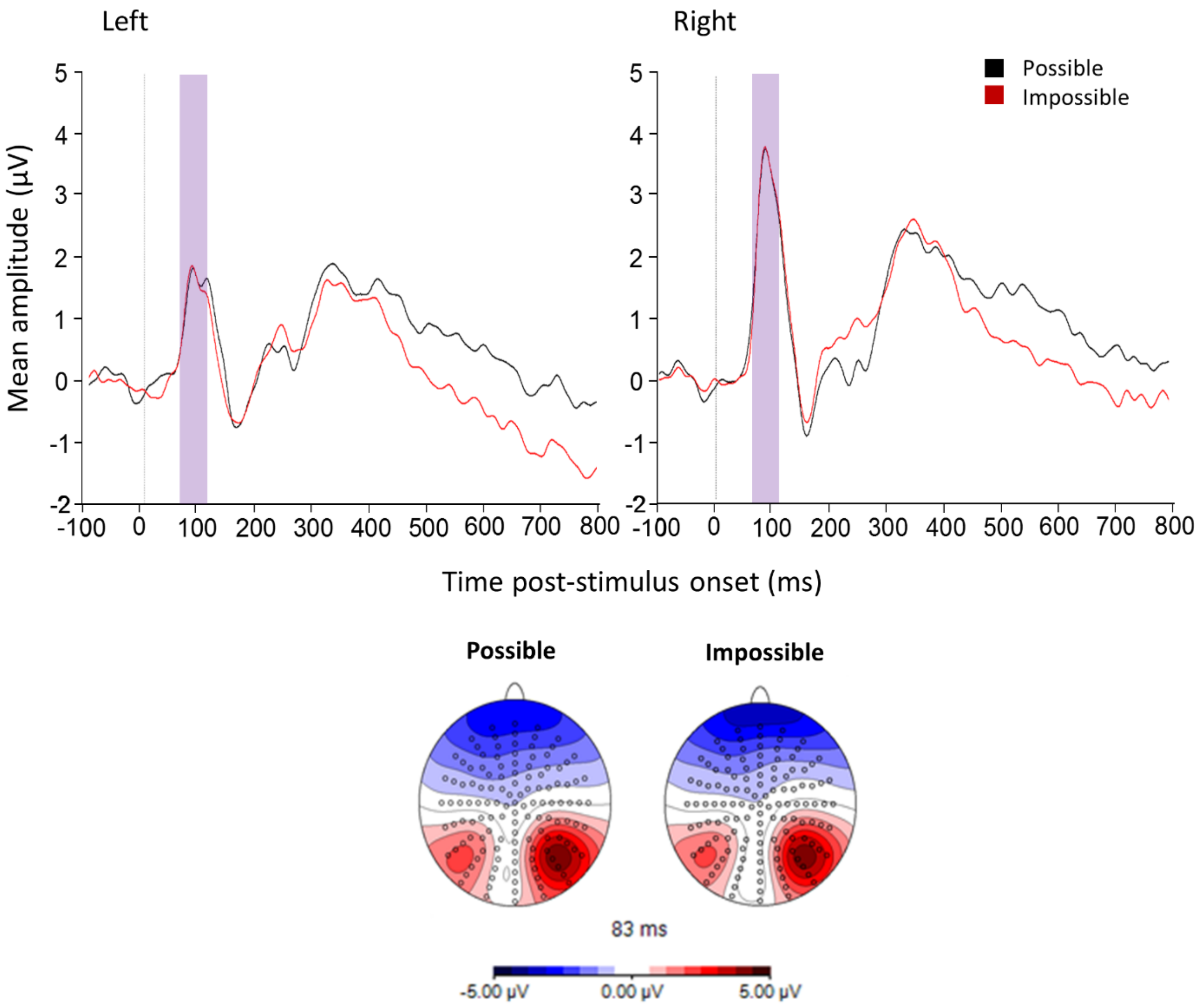


Figure 43. Waveforms showing the P1 (highlighted) average 81ms post-stimulus onset for the left and right hemispheres on occipital electrodes for possible and impossible stimuli.

N1. Using peak amplitude measures, with a 50ms time window (135-185ms). A 2(Condition: possible, impossible) x 2(Laterality) repeated measures ANOVA revealed that there were no significant main effects or interactions. There were no significant main effects or interactions for the latency data.

P2. Using peak amplitude measures with a time window of 30ms (230-260ms), a 2(Condition: possible, impossible) x 2(Laterality) repeated measures ANOVA revealed that there were no significant main effects or interactions. There were no significant main effects or interactions for the latency data.

N2. Using peak amplitude measures with a time window of 40ms (250-290ms), a 2(Condition: possible, impossible) x 2(Laterality) repeated measures ANOVA revealed that there were no significant differences in the amplitude data, but in the latency data, there was a main effect of condition, $F(1, 18)=6.71$, $p=.018$. The N2 for impossible stimuli was earlier ($M=264.65$, $SD=10.91$) than that of possible stimuli ($M=270.91$, $SD=14.27$). There were no other main effects of interactions.

P3. Using peak amplitude measures over a 90ms time window (300-390ms), a 2(Condition: possible, impossible) x 2(Laterality) repeated measures ANOVA revealed that there was a main effect of stimulus type, $F(1,18)=11.94$, $p=.003$, with greater peak amplitudes for impossible ($M=3.68$, $SD=2.18$) than possible ($M=3.19$, $SD=2.26$) stimuli (see Figure 44). There were no other main effects of interactions.

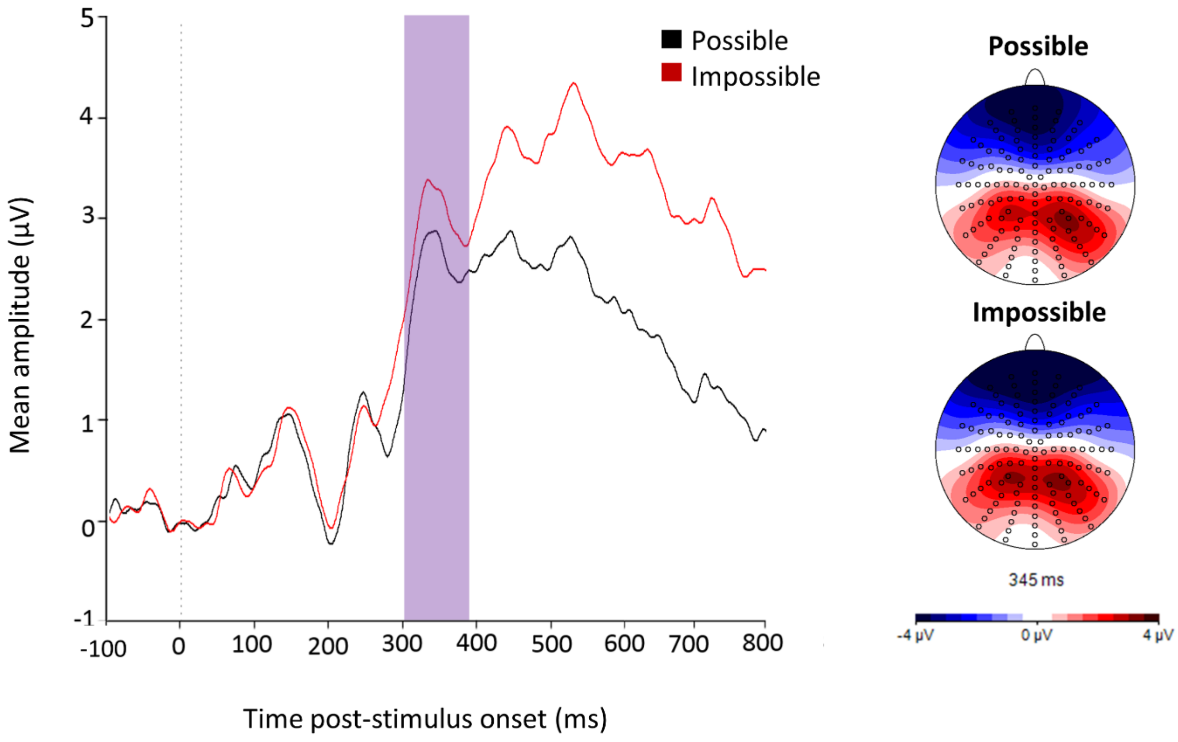


Figure 44. Waveforms from electrode cluster in the central-parietal area corresponding to Pz in 10-20 system. Showing the P3, between 300 and 390ms post-stimulus onset for possible and impossible stimuli.

5.4.3 Mass Univariate Analyses

We wanted to further investigate differential processing of possible and impossible objects. To do so, mass univariate analyses (MUAs) were used to identify a temporal marker defining the earliest time point of differential ERP sensitivity to possible versus impossible objects and, therefore, integration of local and global shape features. MUA can be used to provide an additional 'bias free' measure of statistical contrasts across all electrodes, as opposed to using selected clusters as in standard waveform analysis (see Luck & Gaspelin, 2017). Here, we use MUA to verify the statistical robustness of our earlier analyses. A point-wise mass univariate analysis performed on the possible versus impossible stimuli showed that the earliest differences began during the N2/P3. The difference affected a large group of central-parietal and right anterior leads beginning at about 290ms until around 360ms encompassing the N2 and P3 components, see Figure 45. This confirms that integration of local and global information encompasses the N2/P3 component starting as a latency shift during the N2.

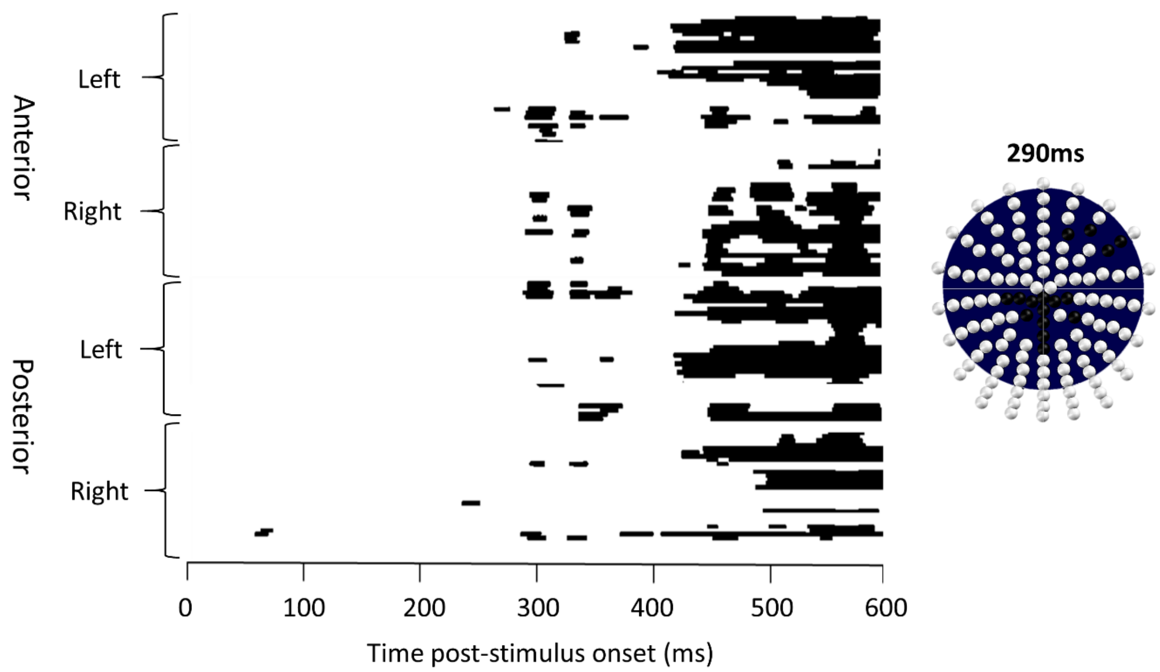


Figure 45. Mass univariate contrasts showing time (x axis) and electrodes (y axis) for the possible/impossible stimuli contrast. All 128 electrodes are shown, dark areas indicate periods where electrodes are significant at p<.01. The electrode montages show the electrodes significant at p<.01 at 290ms post-stimulus onset in black for each contrast.

A time series plot of the frequency distribution of significant differences between possible and impossible stimuli, sub-sampled into 10ms bins is shown in Figure 46.

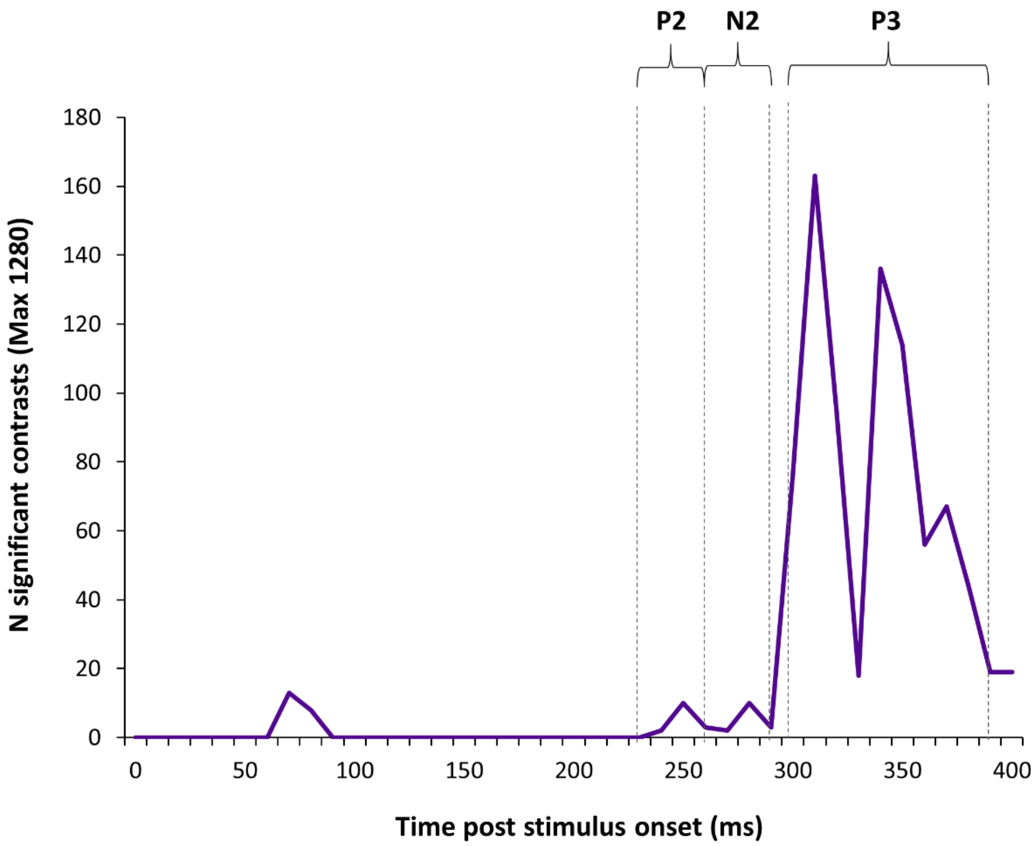


Figure 46. Time series distribution showing the frequency of significant difference contrasts from the mass univariate analysis between 0 and 400ms. Contrasts shown are between possible and impossible stimuli subsampled to 10ms bins.

5.4.4 Further Mass Univariate analyses - accuracy

We wanted to verify that the N2/P3 effects seen in both the standard waveform analyses and the MUA reflect a perceptual sensitivity to integration. To do this, the sample was split based on accuracy, using all 24 participants. Mass univariate analyses were used to identify a temporal marker defining the earliest time point of differential ERP sensitivity to possible versus impossible objects in the 50% of lowest accuracy participants and 50% highest accuracy participants. If the effects do reflect a perceptual sensitivity to integration, we might expect it to correlate with accuracy. Therefore, if the modulation is not present we might expect this to result in an inaccurate response.

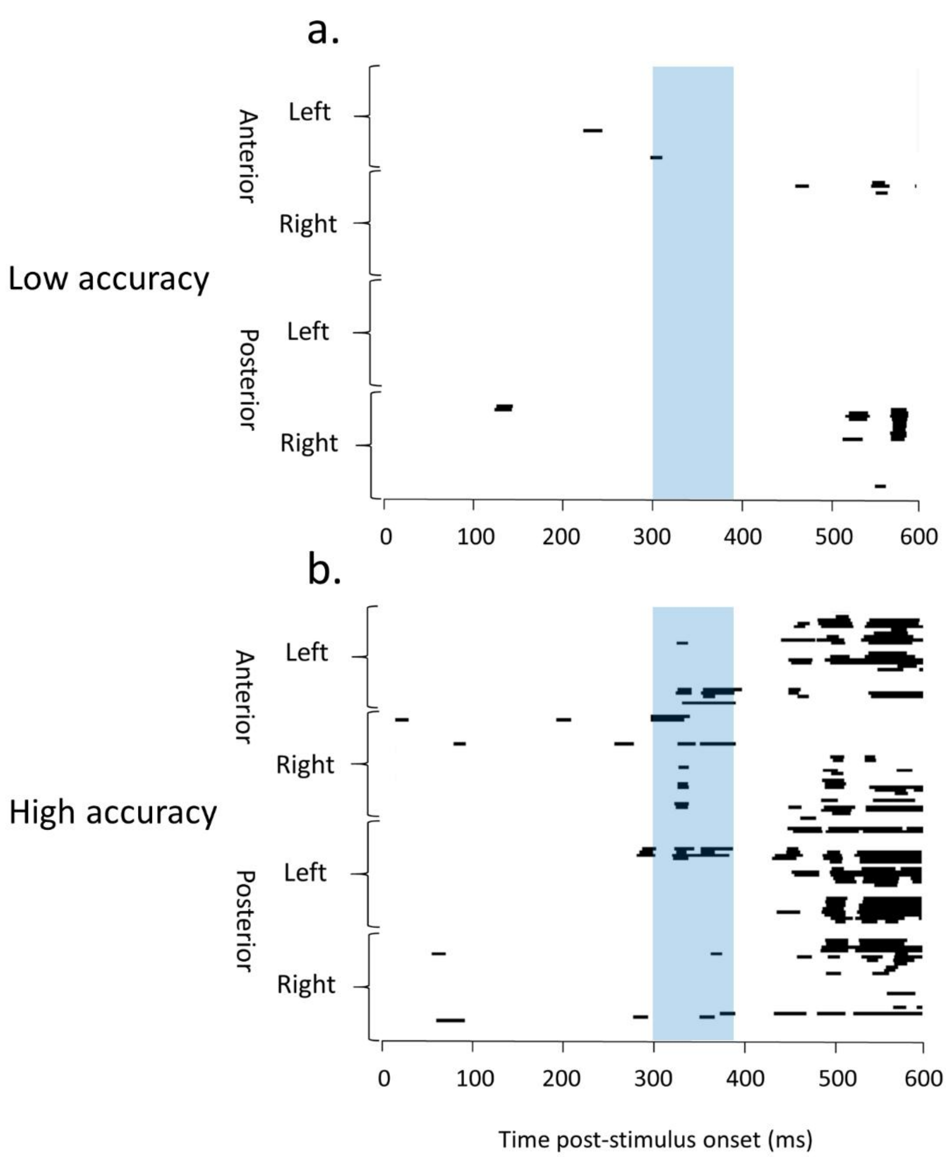


Figure 47. Mass univariate contrasts showing time (x axis) and electrodes (y axis) for the possible/impossible stimuli contrast in (a) lowest 50% accuracy and (b) highest 50% accuracy. All 128 electrodes are shown, dark areas indicate periods where electrodes are significant at *p*<.01. The highlighted section shows the P3 (300-390ms).

Figure 48 shows a time series plot of the frequency distribution of significant differences between possible and impossible stimuli, sub-sampled into 10ms bins. These data were analysed as a non-parametric time-series using the Friedman test, which showed that at the P3, frequency distributions were significantly different between low and high accuracy groups, $\chi^2(1)=9$, $p=.003$.

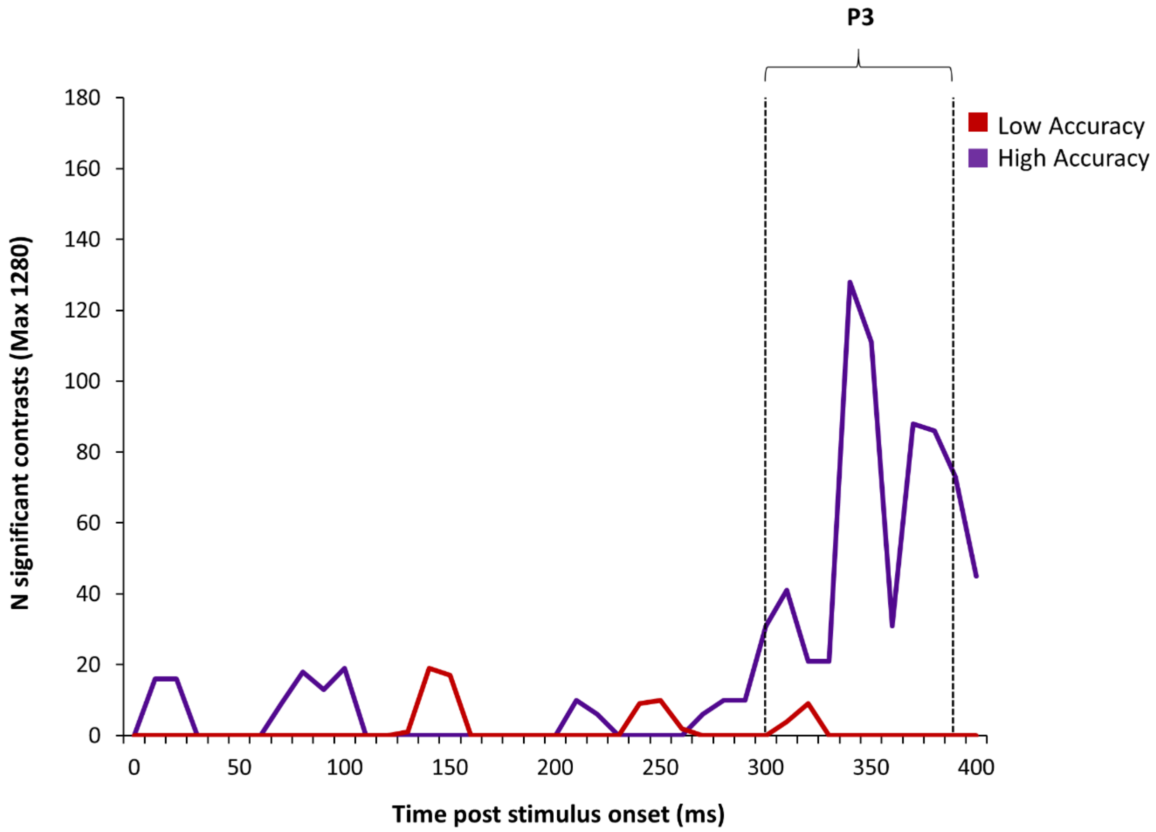


Figure 48. Time series distribution showing the frequency of significant difference contrasts from the mass univariate analysis between 0 and 400ms. Contrasts shown are between possible and impossible stimuli for participants with the lowest accuracy (red) and highest accuracy (purple).

5.5 Discussion

The aim was to find identify an ERP signature associated with the perceptual integration of shape information across local and global spatial scales. Also, we wanted to establish whether the results from a Navon-type experiment using hierarchical stimuli (Chapter IV) generalised to an image classification task. To summarise, we found that first, the behavioural data showed an advantage for possible over impossible objects, with greater accuracy for possible objects.

Second, early ERP components showed no differences between possible and impossible objects; the first differences arose at the N2, around 270ms post-stimulus onset, with earlier N2 for impossible objects. Third, we found that differential sensitivity to object impossibility in amplitude occurred at the P3 (from 300ms post-stimulus onset). Fourth, we verified that the N2/P3 effect reflects a perceptual sensitivity to integration. The effect was present for highly accurate participants, but not those with lower accuracy.

Firstly, we found that there was greater accuracy for possible than impossible objects. Object impossibility may be more difficult to detect due to the inability to combine the local and global elements of the image and therefore create a 3D representation of the object. Freud et al., (2015a, experiment 2b), however, found no difference in accuracy between possible and impossible stimuli. They used same object decision task as the current study, but with different stimuli. It may be that exposure duration was the cause of the difference between this study and ours. Exposure duration is posited to play a role in accuracy for a shape discrimination task using possible and impossible objects: with a short exposure duration (85ms), no difference between possible and impossible objects was found, however when using a long exposure duration (986ms) there was greater accuracy when using possible than impossible stimuli (Freud, Hadad, Avidan & Ganel, 2015). They found that exposure duration is an important consideration for finding differences in processing possible and impossible objects; long exposure duration leads to inevitable processing of finer spatial scales and avoids participant using the gist of the object; coarse analysis. Therefore, it might be that in the current study, as we displayed stimuli for 1 second, there was enough time for object impossibility to be identified, which may not have been the case in other studies.

Early ERP components showed no differences between possible and impossible objects, supporting the idea that early processing of possible and impossible objects does not differ (Freud et al., 2013a; 2013b), and that object impossibility is not distinguished this early in processing. We suggest that there are no early processing differences between possible and impossible objects due to the processing of local parts and global configuration, separately. Both local and global elements of the impossible shapes are individually possible. Therefore, the earliest point where the processing differs, we suggest, is the point at which the local and global elements of the object are integrated. In Freud's work (Freud et al., 2013b) they suggest that the global 3D structure is incoherent in impossible objects, whereas though we agree that it is the 3D structure that is incoherent, our 'global' element is spatial configuration. We have different definitions of local and global. In our impossible objects, the local parts are, themselves, possible, as is the global configuration, it is when these are combined to create a 3D representation that the impossibility becomes apparent. These 'impossible' areas of the objects may be high frequency in nature, but this is not the same as our 'local parts'. Therefore, using objects whose global configuration and local parts, alone, can exist in the 3D world but whose 3D structure is impossible should allow us to identify the point of integration of local and global features for a coherent 3D object.

In the ERP data, we found an earlier N2 response for impossible objects. We posit that this may be the onset of a temporal marker for integration of local and global information; the point at which the output from the global processing, and the output from the local processing are combined, and the system cannot complete processing as usual. If we assume Bar's (2003) interactive processing model, we might suppose that the difference in N2 latency reflects the early inconsistencies revealed by top-down feedback from LSF information during the more fine-grained analysis. The N2 for possible objects was later due to further updating based on the feedback from LSF information – it takes longer to integrate the local and global for possible objects, whereas it is quicker for impossible objects as the updating from frontal areas (LSF information) requires fewer feedback loops as inconsistencies become apparent. Shigemura et al. (2004) also investigated spatiotemporal differences in processing possible and impossible objects using diamond shapes. They used a passive viewing task (counting the number of different coloured stimuli). Our task was different, asking participants to actively discriminate between possible and impossible objects, therefore looking for the inconsistencies, and our key manipulation was the use of objects that included local parts that were possible. However, their findings do corroborate ours, as they found that differences between possible and impossible objects occur between 350 and 390ms post-stimulus onset. Furthermore, our findings support those from Chapter IV. Using low-level visual stimuli and a Navon paradigm, we previously found that the integration of information at local and global spatial scales occurs at the N2/P3. We have shown that this ERP marker for local and global integration generalises to an image classification task.

In order to verify that the N2/P3 effects we found reflect a perceptual sensitivity to integration, the data was analysed split by accuracy. The comparison of high and low accuracy trials allowed us to determine that the differential activity we observed at the N2 and P3 did reflect perceptual sensitivity to integration as the N2/P3 integration effect was present for high accuracy participants, but not for those with low accuracy. As there are differences between possible and impossible objects, it must be the case that the perceptual system (at the N2/P3) is responding to the 3D geometric possibility of the stimulus; the processing requires more information than 2D/image based properties of the stimuli as the impossibility only arises at the level of 3D object geometry. One possible conclusion is that the results show that the perceptual system is computing a reconstruction of the 3D object from its 2D sensory input – this is another piece of evidence supporting theories of shape processing that include the importance of 3D structure – and suggest that 3D structure matters. Some models of shape processing (such as HMAX) are based solely on the 2D 'image-based' projection.

The findings presented here could help elucidate the mechanisms involved in object representation and recognition. Our findings could be taken as support for a coarse-to-fine processing mechanism, or simple to complex processing – as only later in the perceptual processing there are differences between objects, this suggests that the intact local and global processing is occurring as normal, whereas the integration of the local and global information is occurring later, from around 250ms post stimulus onset. Freud, Hadad, Avidan and Ganel

(2015) found that exposure duration plays a role in accuracy for a shape discrimination task using possible and impossible objects. With a short exposure duration (85ms), no difference between possible and impossible objects was found, however when using a long exposure duration (986ms) there was greater accuracy when using possible than impossible stimuli. This provides evidence that early perceptual processes involved in identifying object impossibility do not differ, whereas the impossibility can be identified with longer exposure duration. Freud et al. (2015a) did find that high SF information was more important for processing impossible objects, they concluded that holistic representations of objects relied on mainly LSF information, whereas identifying object impossibility (via spatial incoherence of impossible objects) was mainly based on processing of HSF information.

In Freud's work, the differences between possible and impossible objects are suggested to reflect post-recognition processes, however, we suggest that the differences in the ERPs reflect integration of the earlier perceptual processing of local and global elements for the formation of a 3D structural representation, which presumably precedes recognition. We found that the early perceptual processing of the objects was the same and the later processes differed, we can explain this by assuming Bar's (2003) model, whereby low SF (more global) information is processed in a feed-forward manner, then updates possible interpretations of an object whilst the more fine- grained analyses (more local) take place. We suggest that the differences between processing of geometrically impossible and possible objects reflects the point where the top-down modulation and updating of information based on more fine-grained analyses reveals inconsistencies, which occurs before recognition. Our behavioural results may also be consistent with Bar's (2003) interactive theory as the more fine-grained analysis stage may reveal the inconsistencies in the impossible objects, therefore requiring longer time to process them and leading to more errors.

To summarise, we found that early processing of local and global information in geometrically possible and impossible objects does not differ, however, later differences at the

N2 and P3 reflect the integration of local and global information. This supports the findings from Chapter IV, and provides evidence that the ERP marker for local/global integration generalises to an image classification task. Taken together with Freud's work on the perceptual processing of possible and impossible objects, these findings provide support for Bar's interactive model of object recognition.

5.6 Summary

- This study examined the integration of information from local and global spatial scales.
- We used a simple image classification task whilst recording ERPs. The task was to decide whether an object was geometrically possible or impossible.
- The main results were that early processing did not differ between geometrically possible and impossible objects and integration of local/global information occurs at the N2/P3.
- We provide additional evidence that the N2/P3 reflects a perceptual mechanism involved in the integration of shape information across spatial scales using a non-Navon paradigm.

6 Chapter VI

General Discussion

The empirical chapters were aimed at examining (1) how different kinds of shape information across local and global spatial scales are computed, and integrated, during the perception of 3D object shape, (2) the role of stereo information in 3D shape processing and (3) the temporal dynamics of shape information processing. The experiment presented in Chapter III investigated the differences between local and global processing in 3D objects under stereo and non-stereo viewing conditions. We found evidence for the distinct perceptual processing of shape information at local and global spatial scales during 3D object recognition. This raises the question of when information is integrated during the online perceptual processing of object shape. The experiment presented in Chapter IV was designed to look at the potential integration of the local and global levels of stimuli. We used very low-level visual stimuli and a Navon-type paradigm to avoid the problems associated with using Navon letters. We found evidence for the integration of information at local and global spatial scales at the N2/P3 using basic visual stimuli with a Navon task. The experiment presented in Chapter V was designed to find out if the temporal marker identified in Chapter IV generalised to a completely different paradigm; using more complex stimuli and an image classification task. The findings from Chapter V provided evidence that the integration of local/global information occurs at the N2/P3. I will outline the main findings from each empirical chapter, then discuss the broader implications of the findings for object recognition literature and global and local processing literature. I will then discuss methodological implications and finally, suggestions for future research.

6.1 Summary of findings

The experiment reported in Chapter III examined the time-course of local and global information processing for object recognition and how this differed when viewing in mono or stereo. ERPs were recorded whilst participants completed a recognition memory task where they decided whether objects presented were from a set learned in prior training sessions. Participants learned objects and completed the recognition task in either mono or stereo viewing conditions, using polarising stereo glasses. The main findings from the study were that during mono viewing perceptual sensitivity was greatest for distracters with different local parts to targets, and this difference occurred during the N1 component, and was restricted to trained views only. For stereo viewing perceptual sensitivity was greatest for distracters with different 3D spatial configuration, and this difference occurred during the N2/P3 component, and generalised across trained and untrained views. The results show that object recognition is modulated by stereo information about 3D object structure. The findings challenge theoretical models of object recognition that do not attribute functional significance to both 2D and 3D shape information.

In Chapter IV, we investigated the integration of information at local and global spatial scales. ERPs were recorded whilst participants made orientation decisions about hierarchical stimuli made up of Gabor patches oriented to either the left or right. A Navon-type paradigm was used, whereby participants' attention was directed to either the local or global level of the stimuli and the local and global levels could be either congruent or incongruent. The main findings from the study were that congruency effects were evident at the P1 and N1 for both local and global report. Also, global interference occurred at the N2/P3, which reflects the integration of local and global information. The results show that the integration of local and global levels of the stimuli occurs at the N2/P3 time range, around 300ms post-stimulus onset.

Finally, Chapter V used more complex stimuli and an image classification task to examine the robustness of our integration effect from Chapter IV. We utilised impossible objects, whereby the local parts were geometrically possible, as was the global configuration. The objects' impossibility, and geometric incoherence becomes apparent when integrating the local and global levels of information. The rationale, therefore, was that the first point that the ERPs differed between possible and impossible object conditions would reflect the integration of information at local and global spatial scales. ERPs were recorded whilst participants made decisions as to whether objects presented were geometrically coherent (possible or impossible). The main finding was that the integration of local and global information occurs at the N2/P3. Also, we were able to verify that the N2/P3 effect did reflect the perceptual integration of local and global shape information, as we found that the effect was present for high accuracy trials, but not low accuracy. We would expect that accuracy is correlated with the perceptual integration of local and global information. If this is not occurring, participants will make more incorrect decisions. The results provided evidence of the generalisability of the integration effect at the N2/P3, from around 250ms post-stimulus onset, with a more complex stimulus set.

In sum, the results from all three empirical chapters suggest that local and global processing occur at least partly in parallel, first processing of local and global information occurs at the N1. Information from local and global levels are then integrated at the N2/P3, evidenced in low-level visual stimuli and more complex 3D objects. Also, stereo information has a role in object recognition and ought to be included in models of object recognition.

6.2 Temporal dynamics of shape processing at local and global spatial scales

The time course of local and global processing is unclear, as discussed in Chapter 1: the task and stimuli are important determinants of global and local processing differences in ERP studies. To briefly recap, several different components have been suggested to relate to local and global processing including the P1 (Han, Fan, Chen & Zhuo, 1997; Han, He & Woods, 2000; Jiang & Han, 2005), N1 (Beaucousin et al., 2013) and the later N2 and P3 (Heinze & Münte, 1993; Malinowksi et al., 2002; Volberg & Hübner, 2004; Yamaguchi et al., 2000). Several ERP studies, find evidence of global and local processing occurring at an early (from 150ms), then a later N2 or P3 component (Boeschoten, Kemner, Kenemans & Engeland, 2005; Han, Yund & Woods, 2003; Heinze & Münte, 1993). The ERP results from Chapter III showed differential perceptual sensitivity to local parts and global spatial configuration at the N1 (see Figure 17). This is in line with other ERP studies (e.g., Beaucousin et al., 2013; Leek et al., 2016). In Chapter III, the differential sensitivity to local parts and global spatial configuration continued to an N2/P3 component from 260-380ms (see Figures 18, 19 and 20), this finding lends support to those studies that find an early and later component for local and global processing differences.

We investigated the integration of local and global shape information in two studies (Chapters V and VI) using different stimuli and tasks. The results revealed that the integration of information at local and global spatial scales occurs at the N2/P3 time frame. In Chapter IV we found congruency effects at the N2/P3 (from around 350ms), for only local report, which is evidence of global interference (see Figures 37 and 38). There were considerably more differences in activity for incongruent trials in the local report condition than the global report condition, indicating that the local information in the incongruent global report trials did not interfere significantly with global processing. However, due to the differences in local report trials, it seems that global information was interfering with the local processing. In Chapter V we replicated this finding: we observed an N2/P3 difference which we attribute to the integration of local and global shape information (see Figures 44, 45 and 46). The effects replicate findings from previous studies where a component around the N2 and P3 has frequently been reported to reflect differences in local and global processing with hierarchical stimuli (e.g. Han, He & Woods, 2000; Han, Yund & Woods, 2003; Heinze et al., 1998; Heinze & Münte, 1993; Malinowksi et al., 2002; Volberg & Hübner, 2004; Yamagucchi, Yamagara & Kobayashi, 2000).

In Chapter IV we found evidence of integration of local and global shape information at the N2/P3 (see Figures 44, 45 and 46). It is difficult to compare findings regarding integration of local and global information between the experiments in Chapters IV and V as in Chapter IV, we used low-level Gabor stimuli and a Navon-task, whereas the experiment presented in Chapter V used more complex object stimuli and an image classification task. We did not find evidence of differential processing of local and global information in Chapters IV or V (as in Chapter III). Presumably, the early processing of local and global elements of our stimuli were the same in both of our conditions in Chapter V and in Chapter IV our stimuli design did not allow us to separately identify local and global processing in the ERPs.

One potential issue, still, is whether we can equate our N2/P3 findings from Chapters IV and V with the same process – integration of local and global information. The stimuli and task were considerably different, from very basic Gabors to more complex 3D objects. We think that with the definitions of global and local we use allow comparisons between Chapters IV and V to be made. To briefly recap, we define local as information from a narrow sampling window and global information from a broader sampling window, but acknowledge that both sampling windows contain low and high spatial frequency information. The stimuli presented in both experiments fulfil the definitions we use; therefore, we suggest the findings are comparable.

Our findings go some way to elucidating the parallel/serial nature of local and global processing. A question in global/local processing concerns whether object shape processing is based on analyses of information from different spatial scales that is serial (in a top-down or bottom-up fashion) or parallel. In Chapter III, we showed that local and global processing both occur at the N1 (see Figure 17), indicating that the processes are at least partly parallel, and as our findings from Chapters IV and V show that the integration occurs at the N2/P3, we presume that the parallel processing continues to this point of integration. This is consistent with a conception of object shape processing based on analyses across multiple spatial scales (e.g., Bar, 2003; Bar et al., 2006; Hedge, 2008; Heinz et al., 1994; Heinz et al., 1998; Lamb & Robertson, 1988, Lamb, Robertson & Knight, 1989; Leek et al., 2016; Peyrin et al., 2003; Robertson, Lamb & Knight, 1988). However, it is not clear from our findings whether global or local processing begins first. Though, our findings from Chapter IV lend partial support to the theory of global precedence: we found that whilst there were only congruency effects in the behavioural data, we observed global-to-local interference in the ERP data (see Figures 37 and 38), indicating that global information seems to be processed automatically, therefore may be processed preferentially and earlier than local information.

In Chapter III, we found differential sensitivity to local shape information and global spatial configuration that was modulated by mono and stereo viewing. To recap, under mono viewing conditions, local shape information was weighted more strongly during early perceptual analysis, whereas under stereo viewing, global spatial configuration was also computed (see Figure 21). This appears to show that computation of local shape properties happens earlier than that of global 3D spatial configuration. This hypothesis appears to be at odds with theories of coarse-to-fine analyses of visual input (e.g., Bar, 2003; Bar et al., 2006; Hedge, 2008; Heinz, Johannes, Münte & Mangun, 1994; Heinz, Hinrichs, Scholz, Burchert & Mangun, 1998; Navon, 1977; Peyrin, Chauvin, Chokron & Marendaz, 2003; Peyrin, Baciu, Segebarth & Marendaz, 2004; Peyrin et al., 2010). However, this is only the case if we suppose that coarse analyses are exclusively supporting the derivation of 3D spatial configuration and fine analyses exclusively support the perception of local part structure. We suggest that the derivation of global spatial configuration and local part structure is more likely to be supported by both coarse and fine analyses, containing both high and low spatial frequency information. Similarly, in Chapter IV we found global-to-local interference (see Figures 37 and 38), which indicates that global information is processed preferentially to local, automatically; this also fits with coarse-to-fine processing models.

An alternative possibility is that visual processing order is flexible, and does not operate so rigidly as supposed by coarse-to-fine theories (Morrison & Schyns, 2001 (for review); Gosselyn & Schyns, 2001; Oliva & Schyns, 1997; Schyns & Oliva, 1999. The different findings from mono and stereo viewing groups in Chapter III (see Figure 21) lend support to the notion that different spatial scales may be processed in different orders, dependent on the task and stimuli used and the information available. On the basis that processing order of local and global information might depend on the complexity of the stimuli, familiarity with the stimuli and the task, we do not claim that there is an absolute timing for integration, or order for processing local and global spatial scales. However, we are able to show that different types of information are most relevant for a particular task. For example, in Chapter III, it seems that local parts information is more important for recognition in the absence of stereo disparity, whereas with stereo disparity, global spatial configuration becomes more salient. Whereas, in Chapter IV, in a Navon-style task, global information is processed automatically, and presumably first, when attending to the local level of a stimulus. As the experiments presented in Chapters IV and V used different tasks and stimuli, we do not claim that integration of information at local and global spatial scales occurs at the specific time points we found, but that the components that we identified are important for the integration of information at different spatial scales.

A key finding in the local/global literature is the global precedence effect (GPE). The GPE, originally described by Navon (1977) has been reported in many studies using hierarchical stimuli (e.g., Beaucousin, Simon, Cassotti, Pineau, Houde & Poirel, 2013; Han, He & Woods, 2000; Poirel, Pineau & Mellet, 2008; Proverbio, Minniti & Zani, 1998; Yamaguchi, Yamagata & Kobayashi, 2000). The results from Chapter V demonstrate partial support for the GPE without some of the problems associated with using letter stimuli as in traditional Navon experiments, such as the possibility of biasing hemispheric effects due to the left-hemisphere dominance for lexical processing. In Chapter IV we found that there were higher error rates for incongruent trials and faster RTs for congruent stimuli (see Table 2). However, we observed this effect for both local report and global report trials. We suggest that the reason for the discrepancy in results (the local interference) may be due to the nature of the stimuli and task we used as the task required attention to both local and global levels for all trials. We did, however, observe global-to-local interference effects in the ERPs, this demonstrates the generalisability of this element of the GPE and of the automaticity of global processing. A limitation in Chapter IV was design of the stimuli. The Gabor stimuli were designed to maximally elicit a GPE, however due to the necessity to attend to both the local and global levels in both local and global report trials, we were unable to see a clear global precedence effect (in the behavioural data particularly).

The global interference findings from Chapter IV, and the GPE in general, could be described to be consistent with a coarse-to-fine model of visual recognition, with the global information conveyed by rapid magnocellular visual channels, allowing for rapid initial perceptual analysis of visual inputs. This early analysis allows for guidance of the subsequent analysis of local information conveyed by slow parvocellular visual channels, this could also be explained by early global information sent to the orbitofrontal cortex (OFC) and the high SF information takes a slower route along the ventral stream (Bar, 2003; Bar et al., 2006).

The results in Chapter V also provide some support for Bar's (2003) interactive processing model. Using an image classification task with objects that were geometrically possible or impossible, we found that the early perceptual processing of objects was the same, whereas later processes differ, presumably at the point where geometric incoherence is identified. It has been reported that high SF information is more important than low SF information for processing impossible objects; identifying object impossibility (via spatial incoherence of impossible objects) is mainly based on processing of HSF information (Freud et al., 2015a). Therefore, the lack of early differences in the ERPs could be due to the fast, feed-forward processing of mainly low spatial frequency information from the images and the differences between processing of geometrically impossible and possible objects should reflect the point where the recurrent processing (top-down modulation) and updating of information based on more fine-grained analyses reveals inconsistencies.

6.3 Implications for models of object recognition

The findings from Chapters III, IV and V are discussed in relation to theories of object recognition. Object recognition models differ in terms of the low-level features that are used to make up shape representations. The low-level features could be as basic as pixels (Liu, Knill & Kersten, 1995); more complex such as edges and vertices (Lowe, 1987; Poggio & Edelman, 1990); or collections of edges and vertices (Fukushima & Miyake, 1982; Riesenhuber & Poggio, 1999; 2002) or volumetric parts (Marr & Nishihara, 1978); categorical properties of object parts (Biederman, 1987; Hummel & Biederman, 1992; Hummel, 2001) and surfaces (Leek, Reppa & Arguin, 2005).

In Chapter III, we found differential sensitivity in the ERPs to shape differences between targets and non-targets defined by either shared local parts or 3D shape configuration. These results provide new evidence consistent with theoretical models that propose that the representation of complex 3D object shape involves the specification of higher-order part structure and global 3D part configuration (e.g., Behrmann, et al., 2006; Behrmann & Kimchi, 2003; Biederman, 1987; Hummel & Stankiewicz, 1996; Marr & Nishihara, 1978). The results from Chapter III, therefore, challenge theoretical models which do not attribute functional significance to these properties of object shape representations - including the hierarchical, feed-forward HMAX deep (i.e., multi-layer) network architecture (e.g., Riesenhuber & Poggio, 1999; Serre et al., 2007), other recent approaches to image classification based on hierarchical deep networks (e.g., Cichy, Khosla, Pantazis, Torralba & Oliva, 2016; Khaligh-Razavi & Kriegeskorte, 2014; Krizhevsky, Sutskever & Hinon, 2012) and others (e.g., Bulthoff & Edelman, 1992; Chan et al., 2006; Li & Pizlo, 2011; Li et al., 2009; Pizlo, 2008). Results from Chapters IV and V suggest that theories of object perceptual processing should also include the integration of information from local and global spatial scales (as in Bar, 2003).

In Chapter III, we also found that recognition of 3D object shape was modulated by stereo visual input, shown in both the behavioural and ERP data. This finding challenges theoretical models of object recognition that do not attribute functional significance to both 2D and 3D shape information. We showed that stereo visual input can modulate perceptual sensitivity to different attributes of 3D shape - contrary to the predictions of theoretical models that attribute little, if any, functional significance to stereo information in the derivation of 3D object representations (e.g., Bulthoff & Edelman, 1992; Chan et al., 2006; Cichy, Khosla, Pantazis, Torralba & Oliva, 2016; Khaligh-Razavi & Kriegeskorte, 2014; Krizhevsky, Sutskever & Hinon, 2012; Li & Pizlo, 2011; Li et al., 2009; Pizlo, 2008; Reisenhuber & Poggio, 1999; Serre et al., 2007).

Findings from Chapter V also provide some support for object recognition theories that highlight the importance of 3D information. As we found evidence of the integration of local and global information for the processing of geometric coherence, it must be the case that the perceptual system is responding to the 3D geometric possibility of the stimulus. The processing requires more information than 2D/image based properties of the stimuli as the impossibility only arises at the level of 3D object geometry. One possible conclusion is that the results show that the perceptual system is computing a reconstruction of the 3D object from its 2D sensory input – this is another piece of evidence supporting theories of shape processing that include the importance of 3D structure – and suggest that 3D structure matters. Some models of shape processing (such as HMAX) are based solely on the 2D 'image-based' projection.

6.4 Methodological considerations

There are problems inherent in Navon-style experiments: one limitation is that Navon displays typically combine object elements that would not occur in the natural world such as large letters made of smaller ones. It may be that Navon letters, or similar hierarchical stimuli are problematic in that they are too artificial, and therefore the effects may be paradigmatic. This is evidenced by the widely varying findings from studies using slightly different stimuli and tasks. Though, our stimuli in Chapter IV were also very unnatural, we designed them in such a way to avoid the possibility of hemispheric effects being biased by lexical processing (predominantly a left hemisphere process). Using such low-level visual stimuli allowed us to investigate local and global processing without representations being an issue and without a 'semantic' interpretation. The experiment presented in Chapter IV used simple edge elements (e.g., that might make up textures) and we found that global-to-local interference occurs in even very low-level stimuli. Our stimuli in Chapter IV, however, still had the problem of global being large and local, small. It could be that the effects were due to the size of the global elements, relative to local. However, Krakowski, seems to have countered this argument, in an experiment with an intermediate stimulus level as well as the traditional local and global. Krakowski (2015) found that size of the level was not the issue, as global and intermediate levels were processed in the same way. They controlled for the size of elements, which seems to rule out the possibility that global elements are processed first, simply because they are larger than local, or intermediate.

Another issue that merits discussion is the use of definitions for global and local in Chapters III and V. The terms local and global are restricted to local volumetric parts and global spatial configuration. Therefore, the core idea that we tested derives from structural description approaches, with independent coding of parts and the spatial configuration of parts. However, in Chapter III, we compared target/non-target image similarity models including pixel-overlap, HMAX and Gabor filterbank to see if other models could account for differences between the stimuli. The HMAX model, in particular, represents an alternative theoretical proposal to our structural descriptions account, as it is an image-based model, which does not include independent coding of parts their spatial configuration. We found that there were differences between stimulus types for the HMAX model: there was a difference between locally-similar (SD) and globally-similar distracters (DS), but only for trained viewpoint stimuli and no other differences were significant. If this accounted for our results, we would expect to see greater differences between target-DS for trained views that untrained views, and fewer differences between target-SD targets, both trained and untrained. This, however, does not appear to be the case, the ERP results in Chapter III show differences between targets and DS stimuli, but also large differences between targets and SD stimuli. Also, our Chapter V results replicated the integration findings in Chapter IV which used very basic, non-object stimuli.

In Chapter III, stereo and mono viewing conditions were used. To do this, we used polarising stereo glasses and a 3D monitor. The mono viewing group had information presented to both eyes, but this was the same image, whereas the stereo group had a slightly different image presented to each eye, creating stereo disparity. It could be suggested that the stereo effects found in Chapter III may not represent facilitation from stereo *per se*, but only the information provided in our stereo viewing condition. This information may also be available from monocular viewing in some situations, but happened to be available only in our stereo viewing condition. Nonetheless, our findings suggest that 3D information, even if not from stereo disparity, plays an important role in object recognition.

Further investigation into the perceptual processing of information at different spatial scales is required. Future research might investigate the neural correlates of the integration of local and global information in 3D objects. Also, the future research might further investigate the role of stereo information in object recognition. We found that there was some facilitation from stereo information, but it is unclear if the stereo information needs to be included in a LTM representation. A study using groups that learn novel objects in stereo and are tested in mono viewing conditions and vice versa could verify that the stereo information is being used to form a LTM representation.

6.5 Conclusions

In summary, this thesis provides novel insight into the time-course of the processes involved in shape perception and object recognition. Overall, our results provide evidence for the early processing of local and global information during object perception and the later integration of information from different spatial scales. Our results also provide support for the parallel processing of information at local and global spatial scales. Furthermore, our results challenge models of object recognition that do not include independent coding of object parts and their spatial relations. Lastly, our findings highlight the importance of stereo information in object recognition models.

References

Ahissar, M., & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends in Cognitive Science, 8(10),* 457-464.

Amirkhiabami, G., & Lovegrove, W. J. (1999). Do the global advantage and interference effects covary? *Perception and Psychophysics, 61(7),* 1308-1319.

Arguin, M. & Leek, E.C. (2003). Orientation invariance in visual object priming depends on prime-target asynchrony. *Perception & Psychophysics, 65*, 469-477.

Ban, H., & Welchman, A. E. (2015). fMRI analysis-by-synthesis reveals a dorsal hierarchy that extracts surface slant. *The Journal of Neuroscience, 35(27),* 3823-9835.

Bar, M. (2003). A cortical mechanism for triggering top-down facilitation in visual object identification. *Journal of Cognitive Neuroscience, 15*, 600-609.

Bar, M. (2004). Visual objects in context. *Nat. Rev. Neurosci, 5,* 617-629.

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., Hamalainen, M. S., Marinkovic, K., Schacter, D. L., Rosen, B. R., & Halgren, E. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences, 103*, 449-454.

Beaucousin, V., Simon, G., Cassotti, M., Pineau, A., Houdé, O. & Poirel, N (2013). Global interference during early visual processing: ERP evidence from a rapid global/local selection task. *Frontier in Psychology, 4*, 1-6.

Bennett, D. J., & Vuong, Q. C. (2006). A stereo advantage in generalizing over changes in viewpoint on object recognition tasks. *Perception & Psychophysics*, *68(7),* 1082–1093.

Behrmann, M., Peterson, M.A., Moscovitch, M., & Satoru, S. (2006). Independent representation of parts and the relations between them: Evidence from integrative agnosia. *Journal of Experimental Psychology: Human Perception and Performance, 32*, 1169-1184.

Behrmann, M., & Kimchi, R. (2003). What does visual agnosia tell us about perceptual organisation and its relationship to object perception? *Journal of Experimental Psychology: Human Perception and Performance, 29*, 19-42.

Berger, H. (1929). Uber das Elektrenkephalogramm des Menschen (On the human electroencephalogram). *Archiv f. Psychiatrie u. Nervenkrankheiten, 87,* 527–570.

Biederman, I. (1987). Recognition-by-components: a theory of human image understanding. *Psychological Review, 94(2),* 115–117.

Boeschoten, M. A., Kemner, C., Kenemans, J. L., & van Engeland, H. (2005). The relationship between local and global processing and the processing of high and low spatial frequencies studied by event-related potentials and source modeling. *Cognitive Brian Research,* 24, 228-236

Broadbent, D. E. (1977). The hidden preattentive processes. *American Psychologist, 32(2),* 109-118.

Brunet, D., Murray, M. M., & Michel, C. M. (2011). Spatiotemporal analysis of multichannel EEG: Cartool. *Computational Intelligence and Neuroscience*, 1-15.

Bullier, J. (2001). Integrated model of visual processing. *Brain Research Reviews, 36,* 96-107.

Bülthoff, H. H., & Edelman, S. (1992). Psychophysical support for a two-dimensional view interpolation theory of object recognition. *Proceedings of the National Academy of Sciences, 89(1)*, 60–64.

Bultitude, J. H., & Woods, J. M. (2010). Adaptation to leftward-shifting prisms reduces the global processing bias of healthy individuals. *Neuropsychologia, 48,* 1750-1756.

Burke, D. (2005). Combining disparate views of objects: Viewpoint costs are reduced by stereopsis. *Visual Cognition, 12(5)*, 705–719.

Burke, D., Taubert, J., & Higman, T. (2007). Are face representations viewpoint dependent? A stereo advantage for generalising across different views of faces. *Vision Research, 47(16),* 2164–2169.

Chainey, H., & Humphreys, G. W. (2001). The real object advantage in agnosia: Evidence for a role of surface and depth information in object recognition. *Cognitive Neuropsychology, 18,* 175-191.

Chan, M. W., Stevenson, A. K., Li, Y., & Pizlo, Z. (2006). Binocular shape constancy from novel views: the role of a priori constraints*. Psychophysics, 68(7),* 1124–1139.

Christie, J., Ginsberg, J. P., Steedman, J., Fridriksson, J., Bonilha, L., & Rorden, C. (2012). Global versus local processing: seeing the left side of the forest and the right side of the trees. *Frontiers in Human Neuroscience,* 6(28), 1-8.

Christman, S. D. (1993). Local-global processing in the upper versus lower visual fields. *Bulletin of the Psychonomic Society, 31(4),* 275-278.

Cichy, R.M., Khosla, A., Pantazis, D., Torralba, A., & Oliva, A. (2016). Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence. Scientific Reports, 6:27755.

Cichy, R. M., Pantazis, D. & Oliva, A. (2014). Resolving human object recognition in space and time. *Nature Neuroscience, 17*, 455-462.

Craddock, M., Martinovic, J., & Müller, M. M. (2013). Task and spatial frequency modulations of object processing: An EEG study. *PLoS One, 8(7),* e70293.

Cristino, F., Davitt, L., Hayward, W. G. & Leek, E. C. (2015). Stereo disparity facilitates view generalisation during shape recognition for solid multipart objects. *Quarterly Journal of Experimental Psychology, 68*, 2419-2436.

Dale, G., & Arnell, K. M. (2014). Lost in the forest, stuck in the trees: dispositional global/local bias is resistant to exposure to high and low spatial frequencies. *PLoS One, 9(7),* e98625.

Davidoff, J., & Wilson, B. (1985). A case of visual agnosia showing a disorder of presemantic visual classification. *Cortex, 21,* 121-134.

Delis, D. C., Robertson, L. C., & Efron, R. (1986). Hemispheric specialization of memory for visual hierarchical stimuli. *Neuropsychologia, 24,* 205-214.

Dickson, S. J., Pentland, A. P., Rosenfeld, A. (1992). 3-D shape recovery using distributed aspect matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 14(2),* 174-198.

Edelman, S., & Intrator, N. (2003). Towards structural systematicity in distributed, statically bound visual representations. *Cognitive Science, 27,* 73-109.

Edelman, S., & Bülthoff, H. H. (1990). *Viewpoint specific representation in three-dimensional object recognition (*A.I. Memo No. 1239, C.B.I.P Memo No, 53). Retrieved from http://hdl.handle.net/1721.1/6556

Fabre-Thorpe, M. (2011). The characteristics and limits of rapid visual categorization. *Frontiers in Psychology, 2*, 1-12.

Farah, M. J. (1990). *Visual agnosia: Disorders of object recognition and what they tell us about normal vision,* Cambridge, MA: MIT Press.

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex, 1,* 1-47.

Fink, G. R., Halligan, P. W., Marshall, J. C., Frith, C. D., Frackowiak, R. S. J., & Dolan, R. J. (1997). Neural mechanisms involved in the processing of global and local aspects of hierarchical organised visual stimuli. *Brain, 120,* 1779-1797.

Fink, G. R., Marshall, J. C., Halligan, P. W., & Dolan, R. J. (1999). Hemisphere asymmetries in global/local processing are modulated by perceptual salience. *Neuropsychologia,* 37, 31-40.

Flevaris, A. V., Bentin, S., & Robertson, L. C. (2010). Local or global?: Attentional selection of spatial frequencies binds shapes to hierarchical levels. *Psychological Science,* 21(3), 424-431.

Flevaris, A. V., Bentin, S., & Robertson, L. C. (2011). Attentional selection of relative SF mediated global versus local processing: Evidence from EEG. *Journal of Vision,* 11(7), 1-12.

Flevaris, A. V., Martinez, A., & Hillyard, S. A. (2014). Attending to global versus local stimulus features modulates neural processing of low versus high spatial frequencies: an analysis with event-related potentials. *Frontiers in Psychology, 5,* 277.

Foster, D. H., & Gilson, S. J. (2002). Recognizing novel three-dimensional objects by summing signals from parts and views. *Proceedings. Biological Sciences / The Royal Society*, *269*(1503), 1939–1947.

Freud, E., Avidan, G., & Ganel, T. (2015). The highs and lows of object impossibility: effects of spatial frequency on holistic processing of impossible objects. Psychonomic Bulletin Reviews, 22, 297-306.

Freud, E., Avidan, G., & Ganel, T. (2013). Holistic processing of impossible objects: Evidence from Garner's speeded-classification task. Vision Research, 93, 10-18.

Freud, E., Ganel, T., & Avidan, G. (2013). Representation of possible and impossible objects in the human visual cortex: Evidence from fMRI adaptation. NeuroImage, 64, 685-692.

Freud, E., Hadad, B-A., Avidan, G., & Ganel, T. (2015). Evidence for similar early but not late representation of possible and impossible objects. Frontiers in Psychology, 6, 1-8.

Freud, E., Rosenthal, G., Ganel, T., & Avidan, G. (2015). Sensitivity to object impossibility in the human visual cortex: Evidence from functional connectivity. Journal of Cognitive Neuroscience, 27(5), 1029-1043.

Fukushima, K., & Miyaki, S. (1982). Neocognitron: A new algorithm for pattern recognition tolerant of deformations and shifts in position. *Pattern Recognition, 15,* 455-469.

Gawryszweski, L. G., Riggio, L., Rizzolatti, G., & Umilta, C. (1987). Movements of attention in the three spatial dimensions and the meaning of "neutral" cues. *Neuropsychologia, 25,* 19-29.

Gosselin, F., & Schyns, P. (2001). Bubbles: A technique to reveal the use of information in recognition tasks. *Vision Research, 41,* 2261-2271

Grice, G. R., Canham, L., & Boroughs, J. M. (1983). Forest before trees? It depends where you look. *Perception and Psychophysics, 33,* 121-128.

Groppe, D. M., Urbach, T. P., & Kutas, M. (2011). Mass univariate analysis of event-related brain potentials/fields: A critical tutorial review. *Psychophysiology, 48,* 1711-1725.

Guthrie, D., & Buchwald, J. S. (1991). Significance testing of difference potentials. *Psychophysiology, 28,* 240-244.

Han, S., Fan, S., Chen, L., & Zhuo, Y. (1997). On the different processing of wholes and parts: A psychophysiological analysis. *Journal of Cognitive Neuroscience, 9(5),* 687-698.

Han, S., He, X. & Woods, D. L. (2000). Hierarchical processing and level-repetition effect as indexed by early brain potentials. *Psychophysiology, 37*, 817-830.

Han, S., Weaver, J. A., Murray, S. O., Kang, X., Yund, E. W., & Woods, D. L. (2002). Hemispheric asymmetry in global/local processing: Effects of stimulus position and spatial frequency. *Neuroimage, 17,* 1290-1299.

Han, S., Yund, E. W., & Woods, D. L. (2003). An ERP study of the global precedence effect: The role of spatial frequency. *Clinical Neurophysiology, 114,* 1850-1865.

Harris, I., Dux, P. E., Benito, C. T. & Leek, E. C. (2008). Orientation sensitivity at different stages of object processing: Evidence from repetition priming and naming. *PLoS ONE, 3 (5),* e2256*.*

Hedge, J. (2008). Time course of visual perception: Coarse-to-fine processing and beyond. *Progress in Neurobiology, 84,* 405-439.

Heinz, H. J., Johannes, S., Münte, T. F., & Mangun, G. R. (1994). The order of global- and local-level information processing: Electrophysiological evidence for parallel perception processes. In *Cognitive Electrophysiology.* H. Heinz, T. Münte & G. R. Mangun (Eds.). pp 1-25. Birkhauser, Boston.

Heinz, H. J., Hinrichs, M., Scholz, M., Burchert, W., & Mangun, G. R. (1998). Neural mechanisms of global and local processing: A combined PET and ERP study. *Journal of Cognitive Neuroscience, 10,* 485-498.

Heinze, H-J., and Münte, T. K. (1993). Electrophysiological correlates of hierarchical stimulus processing: dissociation between onset and later stages of global and local target processing. *Neuropsychologia,* 31(8), 841-852.

Hochstein, S., & Ahissar, M. (2002). View from the top: hierarchies and reverse hierarchies in the visual system. *Neuron, 36,* 791-804.

Hong Liu, C., Ward, J., & Young, A. W. (2006). Transfer between two- and three-dimensional representations of faces. *Visual Cognition*, 13, 51–64.

Hübner, R., & Volberg, G. (2005). The Integration of Object Levels and Their Content: A Theory of Global/Local Processing and Related Hemispheric Differences. *Journal of Experimental Psychology: Human Perception and Performance,* 31(3), 520-541.

Hughes, H. C., Nozawa, G., & Kitterle, F. (1996). Global precedence, spatial frequency channels, and the statistics of natural images. *Journal of Cognitive Neuroscience, 8(3),* 197-230.

Hummel, J. E. (2001). Complementary solutions to the binding problem in vision: Implications for shape perception and object recognition. *Visual Cognition, 8,* 489-517.

Hummel, J.E. (2013). Object recognition. In D. Reisburg (Ed.). *Oxford Handbook of Cognitive Psychology*. pp 32-46. Oxford. Oxford University Press.

Hummel, J. E., & Biederman, I. (1992). Dynamic binding in a neural network for shape recognition. *Psychological Review, 99(3),* 480-517.

Hummel, J. E., & Stankiewicz, B. J. (1996). An architecture for rapid, hierarchical structural description. In T. Inui & J. McCelland (Eds.), *Attention and Performance XVI: On information integration in perception and communication* (pp.93-121). Cambridge, MA: MIT Press.

Humphrey, G. K., Goodale, M. A., Jakobson, L. S., & Servos, P. (1994). The role of surface information in object recognition: Studies of a visual form agnosic and normal subjects. *Perception, 23,* 1457-1481.

Humphrey, G. K., & Khan, S. C. (1992). Recognizing novel views of three-dimensional objects. *Canadian Journal of Psychology/Revue Canadienne de Psychologie, 46(2),* 170–190.

Ivry, R., & Robertson, L. C. (1988). *The two sides of perception.* Cambridge, MA: MIT Press.

Jiang, Y., & Han, S. (2005). Neural mechanisms of global/local processing of bilateral visual inputs: an ERP study. *Clinical Neurophysiology,* 116, 1444-1454.

Johnson, J. S., & Olshausen, B. A. (2003). Timecourse of neural signatures of object recognition. *Journal of Vision,* 3, 499-512.

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, 299-312.

Kay, K. N., Naselaris, T., Prenger, R. J., & Gallant, J. L. (2008). Identifying natural images from human brain activity. *Nature, 452,* 352-355.

Khaligh-Razavi, S-M & Kriegeskorte, N. (2014). Deep supervised, but not unsupervised, models may explain IT cortical representation. *PLOS Computational Biology*, 10 (11), e1003915.

Kimchi, R., & Palmer, S. E. (1985). Separability and integrality of global and local levels of hierarchical patterns. *Journal of Experimental Psychology: Human Perception and Performance, 11(6),* 673-688.

Kinchla, R. A., & Wolfe, J. M. (1979). The order of visual processing: “Top-down”, “bottom-up”, or “middle-out”. *Perception and Psychophysics, 25,* 225-231.

Kirchner, H., & Thorpe, S. J. (2006). Ultra-rapid object detection with saccadic eye movements: visual processing speed revisited. *Vision Research, 46 (11),* 1762-1776.

Koenderink, J. J., van Doorn, A. J., & Kappers, A. M. L. (1992). Surface perception in pictures. Perception & Psychophysics, 52(2), 487–496.

Krakowski, C-S., Borst, G., Pineau, A., Houde, O., & Poirel, N. (2015). You can detect the trees as well as the forest when adding the leaves: Evidence from visual search tasks containing three-level hierarchical stimuli, *Acta Psychologica, 157,* 131-143.

Krizhevsky, A., Sutskever, I. & Hinton, G.E. (2012). ImageNet classification with deep convolutional neural networks. In: Pereira, F., Burges, C.J.C., Bottou, L & Weinberger, K.Q. (Eds). *Advances in Neural Information Processing Systems 25*. Curran Associates, Inc. (pp 1097-1105).

Kverega, K., Boshyan, J., & Bar, M. (2007). Magnocellular projections as the trigger of top-down facilitation in recognition. *The Journal of Neuroscience, 27(48),* 13232-13240.

Lamb, M. R., & Robertson, L. C. (1988). The processing of hierarchical stimuli: Effects of retinal locus, locational uncertainty and stimulus identity. *Perception and Psychophysics, 44,* 172-181.

Lamb, M.R. & Robertson, L.C. (1990). The effect of visual angle on global and local reaction times depends on the set of visual angles presented. *Perception & Psychophysics, 47,* 489-496.

Lamb, M. R., Robertson, L. C., & Knight, R. T. (1989). Effects of right and left temporal parietal lesions on the processing of global and local patterns in a selective attention task, *Neuropsychologia, 27,* 471-483.

Lamb, M. R., Robertson, L. C., & Knight, R. T. (1990). Component mechanisms underlying the processing of hierarchically organised patterns: Inferences from patients with unilateral cortical lesions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 471-483.

Lamb, M. R., & Yund, E. W. (1993). The role of spatial frequency in the processing of hierarchically organized stimuli. *Perception and Psychophysics, 54,* 773-784.

Lee, T. S. (2003). Computations in the early visual cortex. *Journal of Physiology, 97,* 121-139

Lee, Y. L., & Saunders, J. A. (2011). Stereo improves 3D shape discrimination even when rich monocular shape cues are available. *Journal of Vision, 11(9),* article no. 6.

Leek, E. C. (1998a). The analysis of orientation-dependent time costs in visual recognition*.* *Perception, 27,* 803–816.

Leek, E. C. (1998b). Effects of stimulus orientation on the identification of common polyoriented objects. *Psychonomic Bulletin & Review, 5,* 650–658.

Leek, E. C., Atherton, C. J. & Thierry, G. (2007). Computational mechanisms of object constancy for visual recognition revealed by event-related potentials. *Vision Research, 5*, 706-713.

Leek, E. C., Davitt, L. & Cristino, F. (2015). Implicit encoding of extrinsic object properties in stored representations mediating recognition: Evidence from shadow specific repetition priming. *Vision Research, 108*, 49-55.

Leek, E. C. & Johnston, S. J. (2006). A polarity effect in misoriented object recognition: The role of polar features in the computation of orientation-invariant shape representations. *Visual Cognition, 13,* 573-600.

Leek, E. C., Reppa, I., & Arguin, M. (2005). The structure of three-dimensional object representations in human vision: Evidence from whole-part matching. *Journal of Experimental Psychology: Human Perception and Performance, 31(4),* 668–684.

Leek, E. C., Roberts, M. V., Oliver, Z. J., Cristino, F., & Pegna, A. (2016). Early differential sensitivity of evoked-potentials local and global shape during the perception of three-dimensional objects. *Neuropsychologia*, *89,* 495-509.

Lehmann, D., & Skrandies, W. (1980). Reference-free identification of components of checkerboard-evoked multi-channel field potentials. *Electroencephalography and Clinical Neurophysiology*, 48, 609-621.

Levine, M. W., & McAnany, J. J. (2005). The relative capabilities of the upper and lower visual hemifields. *Vision Research, 45(21),* 2820-2830.

Li, Y., & Pizlo, Z. (2011). Depth cues vs. simplicity principle in 3D shape perception. *Topics in Cognitive Science, 3,* 667-685.

Li, Y., Pizlo, Z., & Steinman, R. M. (2009). A computational model that recovers the 3D shape of an object from a single 2D retinal representation. *Vision Research, 49(9)*, 979–991.

Loftus, G. R., & Harley, E. M. (2004). How different spatial-frequency components contribute to visual information acquisition. *Journal of Experimental Psychology: Human Perception and Performance, 30(1),* 104-118

Liu, Z., Knill, D. C., & Kersten, D. (1995). Object classification for human & ideal observers. *Vision Research, 35,* 549-568.

Lowe, D. G. (1987). Three-dimensional object recognition from single two-dimensional images. *Artificial Intelligence, 31(3),* 355-395.

Luck, S. J., & Gaspelin, N. (2017). How to get statistically significant effects in any ERP experiment (and why you shouldn't). *Psychophysiology, 54,* 146-157.

Luna, D., Merino, J., y Marcos-Ruiz, R. (1990). Processing dominance of global and local information in visual patterns. Acta Psychologica, 73(2), 131-143.

Lundh, B. L., Lennerstrand, G., & Derefeldt, G. (1983). Central and peripheral normal contrast sensitivity for static and dynamic sinusoidal gratings. *Acta Ophthalmologica, 61,* 171-182.

Malinowski P, Hübner R, Keil A, Gruber T (2002) The influence of response competition on cerebral asymmetries for processing hierarchical stimuli revealed by ERP recordings. *Experimental Brain Research, 144*, 136–139

Mangun, G. R., Heinz, H. J., Scholz, M., & Hinrichs, H. (2000). Neural activity in early visual areas during global and local processing: a reply to Fink, Marshall, Halligan and Dolan. *Journal of Cognitive Neuroscience, 12,* 357-359.

Marr, D (1982). *Vision*. San Francisco: W. H. Freeman.

Marr, D., & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three dimensional shapes. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 200(1140)*, 269–294.

Martens, U., & Hübner, R. (2013). Functional hemispheric asymmetries of global/local processing mirrored by the steady-state visual evoked potential. *Brain and Cognition,* 81, 161-166.

Martin, M. (1979). Local and global processing: The role of sparsity. *Memory and Cognition, 7(6),* 476-484.

Morrison, D. J., & Schyns, P. G. (2001). Usage of spatial scales for the categorisation of faces, objects, and scenes. *Psychonomic Bulletin and Review, 8,* 434-469.

Murray, M. M., Brunet, D., & Michel, C. (2008). Topographic ERP analyses: a step-by-step tutorial review. *Brain Topography, 20(4),* 249-264.

Murray, I., MacCana, F. & Kulikowski, J. J. (1983). Contribution of two movement detecting mechanisms to central and peripheral vision. *Vision Research, 23,* 151-159.

Navon, D. (1977). Forest before trees: The precedence of global feature in visual perception. *Cognitive Psychology, 9*, 353-383.

Navon, D., & Norman, J. (1983). Does global precedence really depend on visual angle? *Journal of Experimental Psychology: Human Perception and Performance, 9(6),* 955-965.

Norman, J. F., Swindle, J. M., Jennings, L. R., Mullins, E. M., & Beers, A. M. (2009). Stereoscopic shape discrimination is well preserved across changes in object size. *Acta Psychologica, 131(2),* 129–135.

Norman, J., Todd, J. T., & Phillips, F. (1995). The perception of surface orientation from multiple sources of optical information. *Perception & Psychophysics, 57*, 629–636.

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,* 72-107.

Olshausen, B. A., Anderson, C. H., & Van Essen, D. C. (1993). A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. *The Journal of Neuroscience, 13(11),* 4700-4719.

Paquet, L., & Merikle, P. M. (1984). Global precedence: The effect of exposure duration. *Canadian Journal of Psychology, 38,* 45-53.

Pasqualotto, A., & Hayward, W. G. (2009). A stereo disadvantage for recognizing rotated familiar objects. *Psychonomic Bulletin & Review, 16(5)*, 832–838.

Pegna, A. J., Darque, A., Roberts, M. V., & Leek, E. C. (2016). 3D viewing modulates early ERPs associated with unfamiliar object classification. *In submission.*

Peyrin, C., Michel, C. M., Schwartz, S., Thut, G., Seghier, M., Landis, T., Marendaz, C., & 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,* 2768-2780.

Peyrin, C., Baciu, M., Segebarth, C., & Marendaz, C. (2004). Cerebral regions and hemispheric specialization for processing spatial frequencies during natural scene recognition: An event-related fMRI study. *Neuroimage, 23,* 698-707.

Peyrin, C., Chauvin, A., Chokron, S., & Marendaz, S. (2003). Hemispheric specialization for spatial frequency processing in the analysis of natural scenes. *Brain and Cognition, 53,* 278-282.

Previc, F. H. (1990). Functional specialisation in the lower and upper visual fields in humans: Its ecological origins and neurophysiological implications. *Behavioural and Brain Sciences, 13,* 519-575.

Pizlo, Z. (2008). *3D Shape: Its unique place in visual perception.* MIT Press. Cambridge, MA.

Pizlo, Z. (2010). 3D Shape. Its unique place in visual perception*. Literary and Linguistic Computing.*

Pizlo, Z., Sawada, T., Li, Y., Kropatsch, W. G., & Steinman, R. M. (2010). New approach to the perception of 3D shape based on veridicality, complexity, symmetry and volume. *Vision Research, 50(1)*, 1-11.

Poggio, T., & Edelman, S. (1990). A network that learns to recognize three-dimensional objects. *Nature, 348,* 263-266.

Poirel, N., Pineau, A., & Mellet, E. (2008). What does the nature of the stimuli tell us about the Global Precedence Effect? *Acta Psychologica,* 127, 1-11.

Pomerantz, J. R. (1983). Global and local precedence: Selective attention in form and motion perception. *Journal of Experimental Psychology: General, 112,* 516-540.

Pizlo, Z., Sawada, T., Li, Y., Kropatsch, W. G., & Steinman, R. M. (2010). New approach to the perception of 3D shape based on veridicality, complexity, symmetry and volume. *Vision Research, 50(1)*, 1-11.

Proverbio, A. M., Minniti, A. & Zani, A. (1998). Electrophysiological evidence of a perceptual precedence of global vs. local visual information. *Brain Research, 6*, 321-334.

Reppa, I., Greville, W. J., & Leek, E. C. (2015). The role of surface-based representation of shape in visual object recognition. *The Quarterly Journal of Experimental Psychology, 68(12),* 2351-2369.

Reynolds, R. I. (1981). Perception of an illusory contour as a function of processing time. *Perception, 10,* 107-115.

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience, 2(11),* 1019–1025.

Riesenhuber, M., & Poggio, T. (2002). Neural mechanisms of object recognition. *Current Opinion in Neurobiology, 12,* 162-168.

Rijsdijk, J. P., Kroon, J. N., & van der Wildt, G. J. (1980). Contrast sensitivity as a function of position on the retina. *Vision Research, 20,* 235-241.

Rizzolatti, L., Riggio, L., Dascola, I., & Umilta, C. (1987). Reorienting attention across the horizontal and vertical meridians. Evidence in favour of a premotor theory of attention. *Neuropsychologia, 25,* 31-40.

Roalf, D., Lowery, N., & Turetsky, B. I. (2006). Behavioural and physiological findings of gender differences in global-local visual processing. *Brain and Cognition,* 60, 32-42.

Robertson, L. C., & Ivry, R. (2000). Hemispheric asymmetries: Attention to visual and auditory primitives. *Current Directions in Psychological Science, 9,* 59-63

Robertson, L. C. & Lamb, M. R. (1991). Neuropsychological contributions to theories of part/whole organisation. *Cognitive Psychology, 23*, 299-330.

Robertson, L. C., Lamb, M. R. & Knight, R. T. (1988). Effects of lesions of temporal-parietal junction on perceptual and attentional processing in humans. *Journal of Neuroscience, 8*, 3757-3769.

Rock, I., & DiVita, J. (1987). A case of viewer-centered object perception. *Cognitive Psychology, 19(2)*, 280–293.

Romei, V., Driver, J., Schyns, P. G., & Thut, G. (2011). Rhythmic TMS over parietal cortex links distinct brain frequencies to global versus local visual processing. *Current Biology, 21,* 334-337.

Rousselet, G. A., Thorpe, S. J., & Fabre-Thorpe, M. (2004). How parallel is visual processing in the ventral pathway? *Trends in Cognitive Science, 8,* 363-370.

Sanocki, T. (1991). Effects of early common features on form perception. *Perception and Psychophysics, 50,* 490-497

Sanocki, T. (1993). Time course of object identification: Evidence for a global o local contingency. *Journal of Experimental Psychology: Human Perception and Performance, 19,* 878-898

Sanocki, T. (2001). Interaction of scale and time during object identification. *Journal of Experimental Psychology: Human Perception and Performance, 27,* 290-302

Sasaki, Y., Hadjikhani, N., Fischl, B., Liu, A. K., Marret, S., Dale, A. M., & Tootell, R. B. H. (2001). Local and global attention are mapped retinotopically in human occipital cortex. *PNAS, 98(4),* 2077-2082.

Schacter, D. L., Cooper, L. A., & Delaney, S. M. (1990). Implicit memory for unfamiliar objects depends on access to structural descriptions. *Journal of Experimental Psychology: General, 119(1),* 5-24.

Schacter, D. L., Cooper, L. A., Delaney, S. M., Peterson, M. A., & Tharan, M. (1991). Implicit memory for possible and impossible objects: Constraints on the construction of structural descriptions. *Journal of Experimental Psychology: Learning, Memory and Cognition, 17(1)*, 3-19.

Schacter, D. L., Reiman, E., Uecker, A., Polster, M. R., Yun, L. S., & Cooper, L. A. (1995). Brain regions associated with the retrieval of structurally coherent visual information. *Nature, 376*, 587-590.

Schlösser, J., Hübner, R., & Studer, T. (2009). The effect of element spacing on hemispheric asymmetries for global/local processing. *Experimental Psychology, 56(5),* 321-238.

Schyns, P. G., & Oliva, A. (1994). From blobs to boundary edges: Evidence for time- and spatial-scale-dependent scene recognition. *Psychological Science, 5(4),* 195-200.

Serre, T., Oliva, A., & Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. *Proceedings of the National Academy of Sciences, 104,* 6424–6429.

Serre, T., Wolf, L., Bileschi, S., Riesenhuber, M., & Poggio, T. (2007). Robust object recognition with cortex-like mechanisms. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 29(3),* 411-426.

Servos, P., Goodale, M. A., & Humphrey, G. K. (1993). The drawing of objects by a visual form agnosic: Contribution of surface properties and memorial representation. *Neuropsychologia, 31,* 251-259.

Shigemura, J., Yoshino, A., Kobayashi, Y., Takahashi, Y., & Nomura, S. (2004). Spatiotemporal differences between cognitive processes of spatially possible and impossible objects: a high-density electrical mapping study. Cognitive Brain Research, 18, 301-305.

Shulman, G. L., Sullivan, M. A., Gish, K., & Sakoda, W. J. (1986). The role of spatial-frequency channels in the perception of local and global structure. *Perception, 15,* 259-273.

Shulman, G. L., & Wilson, J. (1987). Spatial frequency and selective attention to local and global information. *Perception, 16,* 89-101.

Shuwairi, S. M., Albert, M. K., & Johnson, S. P. (2007). Discrimination of possible and impossible objects in infancy. Psychological Science, 18(4), 303-307.

Simons, D. J., Wang, R. F., & Roddenberry, D. (2002). Object recognition is mediated by extraretinal information. *Perception & Psychophysics, 64 (4),* 521-530.

Smith, M.L., Gosselin, F., Schyns, P.G. (2006). Perceptual Moments of Conscious Visual Experience. *Proceedings of the National Academy of Sciences,* 103, 5626-5631.

Soldan, A., Hilton, H. J., & Stern, Y. (2009). Bias effects in the possible/impossible object decision test with matching objects. Memory and Cognition, 37(2). 235-247.

Tarr, M. J., & Bulthoff, H. H. (1998). *Object recognition in man, monkey, and machine.* MIT Press, Cambridge, MA.

Thorpe, S. J., Fize, D. & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 381*, 520-522.

Ullman, S. (2006). Object recognition and segmentation by a fragment-based hierarchy. *Trends in Cognitive Sciences, 11*, 58-64.

Ullman, S., & Basri, R. (1991). Recognition by linear combinations of models. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 13(10),* 992-1006.

Van Kleek, M. H. (1989). Hemispheric differences in global versus local processing of hierarchical visual stimuli by normal subjects: new data and a meta-analysis of previous studies. *Neuropsychologia, 27(9),* 1165-1178.

VanRullen, R., & Thorpe, S. J. (2001). The time course of visual processing: from early perception to decision-making. *Journal of Cognitive Neuroscience, 13(4),* 454-461.

Volberg, G., & Hübner, R. (2004). On the role of response conflicts and stimulus position for hemispheric differences in global/local processing: an ERP study. *Neuropsychologia, 42*, 1805-1813.

Volberg, G., & Hübner, R. (2007). Do the hemispheres differ in their preparation for global/local processing? *Experimental Brain Research,* 176, 525-531.

Volberg, G., & Hübner, R. (2008). Deconfounding the Effects of Congruency and Task Difficulty on Hemispheric Differences in Global/Local Processing. *Experimental Psychology,* 54. 83-88.

Welchman, A. E., Deubelius, A., Conrad, V., Bülthoff, H. H., & Kourtzi, Z. (2005). 3D shape perception from combined depth cues in human visual cortex. *Nature Neuroscience, 8(6),* 820–827.

Wexler M., & Ouarti, N. (2008). Depth affects where we look. *Current Biology*, *18*, 1872–1876.

Williams, P., & Tarr, M. J. (1997). Structural processing and implicit memory for possible and impossible figures. Journal of Experimental Psychology: Learning, Memory and Cognition, 23, 1344-1361.

Wismeijer, D. A., Erkelens, C. J., Ee, R. van, & Wexler, M. (2010). Depth cue combination in spontaneous eye movements. *Journal of Vision, 10(6*), 25.

Wu, X., Li, W., Zhang, M., & Qui, J. (2012). The neural basis of impossible figures: Evidence from an fMRI study of the two-pronged trident. Neuroscience Letters, 508, 17-21.

Yamaguchi, S. Yamagata, S. & Kobayashi, S. (2000). Cerebral asymmetry of the "top-down" allocation of attention to global and local features. *Journal of Neuroscience, 20*, 72.

Yovel, G., Yovel, I., & Levy, J. (2001). Hemispheric asymmetries for global and local visual perception: Effects of stimulus and task factors. *Journal of Experimental Psychology: Human Perception and Performance, 27(6),* 1369-1385.

7 Appendices

7.1 Appendix 1 – Pilot data for Chapter IV

To ensure that our stimuli and task were sufficient to elicit a global precedence effect, we first conducted the experiment without ERP recording. The accuracy results showed that there was interference when displays were incongruent, for both local and global report, $t(17)=4.65$, $p<.001$ and $t(16)=5.58$, $p<.001$, respectively. The reaction time (RT) results also showed interference from incongruent trials for both local and global report, $t(17)=6.83$, $p<.001$ and $t(17)=6.35$, $p<.001$, respectively.

Appendix 1. Table showing accuracy (% incorrect) and RTs for congruent and incongruent trials, for both local and global report.

	Local Report				Global Report			
	Congruent		Incongruent		Congruent		Incongruent	
	Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
Accuracy (%)	91.45	11.71	70.82	23.18	88.67	13.48	67.4	26.04
RT (sec)	1.24	0.11	1.51	0.12	1.14	0.15	1.32	0.08