universität wien

MASTERARBEIT

Titel der Masterarbeit

“Visuo-haptic integration during binocular rivalry: Studying the roles of intermodal congruence and object familiarity“

verfasst von

Dalibor Andrijević, BSc

angestrebter akademischer Grad

Master of Science (MSc)

Wien, 2014

Studienkennzahl lt. Studienblatt:	A 066 013
Studienrichtung lt. Studienblatt:	Masterstudium Joint Degree Programme Mei:CogSci Cognitive Science
Betreuerin / Betreuer:	Univ.-Prof. Dr. Ulrich Ansorge

A person perceiving a familiar object is not aware that what is perceived

is as much an expression of memory as it is of perception.

Endel Tulving, 1990

All parts of the nervous system are connected together and no part is even probably

ever capable of reaction without affecting and being affected by various other parts

and it is a system certainly never absolutely at rest.

Charles Sherrington, 1906

My body is the fabric into which all objects are woven,

and it is, at least in relation to the perceived world,

the general instrument of my 'comprehension'.

Maurice Merlau-Ponty, 1945

Acknowledgments

Openness and trust, patience and personal support are along with curiosity and reason the fertile soil for the flourishing of a scientific culture. This holds true on the macro scale, like in a big community of people as it holds true on a smaller scale, like with family, friends and the wizards or reason. I want to thank all those who showed those great virtues, because without them the world would be void of endeavours like the present.

I want to thank Prof. Dr. Ulrich Ansorge who showed the openness to supervise a project which seemed not to fall in his primary research field. Yet, his profound methodological knowledge made it possible for me to overcome many obstacles. I enjoyed our amicable discussions.

I address also a big thank to Prof. Dr. Hermann Prossinger, whose lectures opened my eyes to the deep world of statistics and its possibilities as a powerful tool, but also its limits in epistemology. It was truly delightful to gain a taste of destilled reason during his lectures.

I also appreciate the short, but inspiring remarks by Prof. Dr. Peter König who made kind and helpful mail responses without even knowing my face.

Some very special people showed the willingness and support to participate in the study as test persons and to proofread the work. Much of their patience was needed during the experimental sessions, for which I am really grateful, as they gave it away for free. So, thanks to Sebastian Günnel, Andre Hoffmann, Max Sänger, Berndt Mahr, Joe Hörleinsberger, Francisca Tan, Clemens Dürr, Alex Hentschel and some others. For his English proficiency and proofreading, a very special ‘thank you’ to Vladan Šutanovac.

And lastly, and foremost importantly, I thank Barbara, her family and my relatives for the continuous support, the patience and trust. No words describe how much I appreciate your love and good will. Thank you for being the eye of the storm. Thank you for everything.

List of contents

Abstract

Does the nervous system take advantage of conceptual information during visuo-haptic integration? The present study investigated whether touch as a way to prime action-related and conceptual knowledge can facilitate multisensory integration. Specifically, it sought to test whether haptically primed, object-intrinsic information of affordances (in addition to the object shape information, its structural descriptions) would modulate stronger visuo-haptic processing during binocular rivalry than object shape information alone. During binocular rivalry, each eye receives a different visual input, resulting in alternating dynamics of perceptual switches. This phenomenon has been used to study the psychophysics and the neurobiology of visual perception. Only recently, it has also been used to study multisensory interactions. It has been demonstrated that auditory, olfactorial or haptic signals can stabilize the visual image which corresponds to the signal of the other sensory modality, presumably in early multisensory convergence zones. While previous studies used 'minimal', abstract cross-modal stimuli pairs (like Gabor patches), in the current study multisensory effects were investigated with functional familiar objects which engage a dense network of fronto-parietal areas involved also during conceptual and involuntary motor processing. The network's signals converge on posterior visuo-haptic convergence areas, like the lateral occipital complex (LOC). Crucially, some of those areas are also involved in perceptual stabilization during binocular rivalry. In contrast, unfamiliar nonsense objects were used which engage a smaller somatosensory network which also converges on LOC. It was hypothesized that visuo-tactile integration during binocular rivalry would be stronger stabilized when observers interacted with familiar objects, engaging the denser fronto-parietal network with greater top-down signals, than with unfamiliar objects which engage the less dense bottom-up network. Perceptual stabilization was measured by the frequency of the perceptual switches and the duration of the perceptual dominance of the haptically congruent visual stimulus across the two object classes. In addition, effects of visuo-haptic priming were measured at the trial onset.

Although initial cross-modal priming was present and binocular rivalry was consistently modulated by haptic input, familiar objects did not show an advantage in visuo-tactile binding over unfamiliar objects. This implies that the signals from the fronto-parietal top-down network of the familiar objects stabilize visuo-haptic processing comparably strong to the signals of the bottom-up somatosensory network of the unfamiliar objects. This suggests that the nervous system relies on sparse representations in multisensory processing during object interaction without necessarily taking advantage of conceptual information. Since haptic space is invariant and its interpretative frame is very limited, the resolution of ambiguities in the visual system can rely on structural descriptions alone, rendering conceptual information unnecessary during visuo-haptic integration.

Zusammenfassung

In der vorliegenden Arbeit wurde untersucht, ob haptischer Input als Mittel, handlungsrelevantes und konzeptuelles Wissen zu bahnen, multisensorische Integration erleichtert. Im Besonderen wurde der Frage nachgegangen, ob haptisch gebahnte funktionale und handlungsbezogene Informationen, die (zusätzlich zur Objektform bzw. strukturellen Beschreibungen) Alltagsobjekten innenwohnen, stärker visuo-haptische Integration während binokulärer Rivalität fördern als Objektform allein. Bei binokulärer Rivalität wird jedem Auge ein anderes Bild dargeboten, was dazu führt, dass die beiden Eindrücke nicht verschmelzen, sondern sich abwechseln. In vergangenen Jahren wurde binokuläre Rivalität eingesetzt, um multisensorische Wechselwirkungen zu untersuchen, bei welchen z. B. gezeigt wurde, dass auditive, olfaktorische und haptische Signale dasjenige visuelle Bild stabilisieren, welches mit dem nicht-retinalen Input kongruent ist. Während man in vorangegangenen Studien abstrakte Stimuli einsetzte (z. B. Gabor-Gitter), wurden in der vorliegenden Arbeit Alltagsobjekte verwendet, die über handlungsrelevante Informationen verfügen. Von diesen ist bekannt, dass sie bei Beobachtern ein dichtes top-down-Netzwerk von fronto-parietalen Gehirnarealen aktivieren, das auch bei semantisch-konzeptueller und unwillkürlicher Handlungsplanung beteiligt ist. Die neuronalen Signale konvergieren auf posteriore, visuo-haptische Verarbeitungsareale, wie den lateralen okzipitalen Komplex (LOC). Einige dieser Areale sind auch bei perzeptueller Stabilisierung während binokulärer Rivalität beteiligt. Zur Kontrolle wurden nicht-funktionale Objekte verwendet, die ein kleineres, somatosensorisches Netzwerk aktivieren, dessen Signale ebenso auf LOC kovergieren. Da die Interaktion mit bekannten und funktionalen Objekten das größere top-down-Netzwerk aktiviert, wurde angenommen, dass hierbei visuo-taktile Wahrnehmung während binokulärer Rivalität stärker stabilisiert wird. Während der Interaktion mit unbekannten, nicht-funktionalen Objekten, die das kleinere bottom-up-Netzwerk aktiviert, wird die visuo-taktile Wahrnehmung hingegen weniger stark stabilisiert. Wahrnehmungsstabilisierung wurde gemessen als die perzeptuelle Wechselfrequenz und die zeitliche Dominanzdauer der haptisch kongruenten visuellen Objekte. Zusätzlich wurden Effekte der visuo-haptischen Bahnung zu Beginn der Versuchsdurchgänge gemessen.

Zwar konnte sowohl die visuelle Verarbeitung haptisch gebahnt, als auch die binokuläre Rivalität haptisch moduliert werden; bekannte, funktionale Objekte haben jedoch visuo-haptische Integration weder verstärkt noch die visuo-haptische Wahrnehmung stärker stabilisiert im Vergleich zu unbekannten Objekten. Dies impliziert, dass die top-down-Signale aus dem fronto-parietalen Arealen für die bekannten Objekte die visuo-haptische Verarbeitung gleichermaßen stabilisieren wie die Signale des bottom-up Netzwerks aus den somatosensorischen Zentren für unbekannte Objekte. Das suggeriert, dass das Nervensystem spärliche Repräsentationen während multisensorischer Verarbeitung bei Objekt-Interaktion verwendet, ohne notwendigerweise konzeptuelle Informationen heranzuziehen. Da der haptisch-perzeptuelle Raum invariant und sein Interpretationsrahmen stark begrenzt ist, kann das visuelle System sensorische Zweideutigkeiten allein nach strukturellen Objektbeschreibungen auflösen. Somit sind konzeptuelle Informationen während visuo-haptischer Integration nicht notwendig.

List of abbreviations

2AFC	Two-alternative forced choice
FAM	Object familiarity
FAM+	Experimental session with familiar objects
FAM-	Experimental session with unfamiliar objects
PSF	Peceptual switches frequency
SSC	Sensory stimulation condition
VO	Visual-only trial/block
VTC+	Visuo-tactile congruency trial/block
VTC-	Visuo-tactile incongruency trial/block
IPS	Intraparietal sulcus
iTC	Inferior temporal cortex
LOC	Lateral occipital complex
lpPC	Left posterior parietal cortex
lpFC	Left prefrontal cortex
vPMC	Ventral premotor cortex
mFG	Medial fusiform gyri
pmTG	Posterior middle temporal gyri
pvTC	Posterior ventral temporal cortex
SII	Secondary somatosensory area

1. Introduction

When thinking of sensory perception, we demarcate with rather strong lines one sensory quality from another. A series of seminal findings from psychophysics and neurophysiology from the past two decades, however, indicates that this conception of sensory perception may be all too simplified and incomplete. Recent evidence shows that sensory systems are closely intertwined, overlap in the nervous system (Ghazanfar & Schroeder, 2006) and that the integration of incoming sensory signals at that level is a necessary condition for generating rich perceptual experience (Murray & Wallace, 2012; Naumer & Kaiser, 2010; Stein & Meredith, 1993).

In this context, current study investigated whether touch as a means of accessing motor and conceptual knowledge of real objects can influence multisensory processing. Specifically, it sought to test whether object-intrinsic information of affordances – in addition to the 'structural descriptions' (object shape information) – would modulate stronger the visuo-haptic processing than object shape information alone. With binocular rivalry as a tool, this question has not yet been addressed.

1.1 Unimodal, bimodal, multimodal: Multisensory processing and integration

This section provides the systemic approach and the broad conceptual framework of the current study. It also outlines the importance of sensorimotor activity and observer-object interaction for the construction of multisensory object representations which may underlie visuo-haptic processing. For unambiguous multisensory experience to emerge, it takes, on the one hand, activity by means of movement and exploration of the environment; on the other hand, the nervous system needs to resolve perceptual ambiguities and to integrate various incoming sensory events. Basic principles of multisensory integration and interaction are reviewed as well as the consequences of the tightly connected operations of the sensory modalities.

1.1.1 The constructive nature of multimodal perception

Observers are exposed to a vast number of different physical stimuli which reveal the state of the environment. The nervous system has to organize the information stream which constantly perturbs different sensory channels. Such stimulation is referred to as 'cross-modal' (Stein & Stanford, 2008). In cross-modal stimulation, receptors belonging to distinct sensory modalities simultaneously receive information of different physical formats. This information builds the basic interpretable components of an event or enacted object as it gets

passed on to the nervous system which constructs the consistent content of experience. This is also referred to as multisensory or multimodal integration. Unifying all such cues to a unitary perceptual event is referred to as ‘binding’. Electrophysiological studies of the past 30 years point to the temporal binding hypothesis (‘binding by synchrony’), which states that multiple cues encoded in different cortical areas are merged by synchronized neural spike activity with fine temporal resolution giving rise to conscious perceptual awareness (Freeman, 1975; Varela et al., 2001). But “synchronized oscillatory activity is not only stimulus driven but does occur also across widely distributed networks of interconnected cortical areas in anticipation of an attention demanding discrimination task“ (Singer, 2007). In this sense, the nervous system with its sensory interfaces is not a passive information receiver, but rather an active generator of internal models about the world and sensory predictions about upcoming sensory stimulations. During the adaptive sensory signal processing, internal models and predictions are evaluated on the basis of sensory evidence (Engel, Fries & Singer, 2001; Friston, 2009). This idea, initially formulated by Helmholtz during the 1860ies (Helmholtz, 1867), lies at the core of visual input prediction on the basis of haptically derived object information.

1.1.2 Movement and action generate coinciding sensory signals

We process unimodal information when we rely on one sensory channel only. For instance, when we explore the environment only visually, we derive information from retinal images. Sometimes sensory systems can sometimes receive information independently of each other. This is the mode of operation when we extract information from the extrapersonal space – the distance outside the reaching area of our hands. We process multisensory information, in contrast, in peripersonal space – the sphere within the reach of the limbs – where we exploit proprioceptive, mechanoceptive and visual signals, for example during the interaction with objects. Here we are not only exposed to the sensory stimulation passively, but we inevitably engage and combine simultaneously more of our sensory systems (Gibson, 1966). In general, multisensory processing takes place when we encounter unitary events – the moments where the involved senses, such as vision, audition and touch, are stimulated *simultaneously*. This is the case when we touch a particular object while simultaneously perceive it visually. Hence, perceptual experience is embodied activity whereby the sensory organs actively sample information and interact with the environment when the observer is in action – be it during grasping for an object, or during saccades to fixate on a location (Holst & Mittelstedt, 1950; Noe, 2004; Saig et al., 2012; Sperry, 1950; Varela et al. 1992). A series of experiments conducted by Held et al. during the 1960s point to the idea that the very foundation for the sampling of spatially and temporally simultaneously occuring cues are sensory-motor contingencies. They suggest that, when the observer is in action, sensory-

motor feedback loops provide the nervous system with structured information to guide behavior, navigate properly and give rise to a perceptually consistent environment. They showed with young kittens that if self-induced movement and *active* exploration are absent during exposure to visual stimuli, severe impairment in visuo-motor task performance occurs.

Evidently, coincidentally incoming multimodal cues of unitary events can be biologically plausibly encoded and associated via the Hebbian synapse: „Cells that fire together, wire together“ (Hebb, 1949; Lewis, 2010; Wallace et al., 2004). This kind of correlated neural activity – signals about self-induced movement and their sensory consequences in the external world – can be considered the basic mechanism for detecting and integrating multimodal signals in a constant stream of sampled information. This includes signals which represent features of enacted objects. Departing from this idea, it is legitimate to regard sensory-motor coupling, coincidence detection and associative learning as generative forces behind multisensory integration and the generation of multisensory representations.

1.1.3 Neural consequences of coinciding stimulation of multiple sensory systems

The existence of a somatotopic map (Penfield & Boldrey, 1937), retinotopic map (Engel, Glover & Wanderbell, 1997) or tonotopic map (Romani, Williamson & Kaufman, 1982) points to a need for a neural architecture, which enables modality-specific processing of signals which have physically distinct origins (mechanic pressure, electromagnetic and acoustic waves). Nevertheless, the presence of different co-occurring signals has implications for the representation of events and objects in the nervous system, few of them being highlighted hereinafter, since they are relevant for the present study.

The most peculiar case where the intersensory associations and specific connectivity of certain brain areas lead to anomalous sensory experience is sensory synaesthesia, where for instance subjects hearing a tone see a specific colour (Bargary & Mitchell, 2008; Ramachandran & Hubbard, 2001; Rich & Mattingley, 2002). Apart from that very special case, there are multisensory neural cell assemblies which either respond similarly to stimuli from different sensory systems or they respond stronger when the stimulation is cross-modal compared to unimodal (‘multisensory enhancement‘, Stein & Stanford, 2008). This means that they have receptive fields for the combined stimulation of more sensory modalities. For instance, neurons in the superior temporal sulcus (STS) (Wright et al., 2003) and in the superior colliculus (Wallace, Wilkinson & Stein, 1996) preferably fire when visual and auditory cues are combined, in contrast to their response to either single modality stimulation. Likewise, there are visuo-tactile neurons in the lateral occipital complex (LOC) (Beauchamp, 2005; Macaluso, 2006) and in the intraparietal sulcus (Lacey et al., 2001). Both of these cortical structures appear to be highly responsive during visuo-haptic shape

exploration and recognition, as revealed by fMRI. Also, many putative unisensory areas can be considered multimodal, since it is evident that primary sensory areas also exhibit multisensory properties to a certain degree (Ghazanfar & Schroeder, 2006). To see somebody moving their lips so as to speak without vocalization engages the auditory cortex (Calvert et al., 1997). It was demonstrated consistently that haptic object exploration and recognition, especially in respect to recovering an object's geometric shape, involved middle and lateral occipital areas – brain regions traditionally regarded as unisensory extrastriate visual areas (Amedi, et al., 2002; James et al., 2002; Reed, Shoham & Halgren, 2004). Therefore, the visual and haptic modality – as far as we consider them separate – share neural representations of object features.

1.2 Behavioral and psychophysical consequences of coinciding stimulation of multiple sensory systems

In the research area of visuo-tactile multisensory processing, behavioral and psychophysical studies can be categorized according to at least three threads of investigation. That is, the conclusions in the studies aim to contribute to one of the three threads. The first deals with sensory cooperation to enhance task performance, object recognition and to maximize sensory reliability and perceptual optimality. The second focuses on the question of whether the internal representations of objects used in experiments are multimodal or unimodal. The third is concerned with the extent to which one modality can influence another; here, the pominent notion can be coined as sensory 'capturing', whereby it is investigated whether one sensory modality is more dominant than any other sensory modality in respect to a task. The work here deals mainly with the first notion, albeit it considers also the nature of internal representations. Sensory cooperation is crucial because one principal aspect of multisensory integration is to reduce perceptual ambiguity by taking advantage of cross-modal congruency.

1.2.1 Visuo-tactile priming

When one sensory modality encodes a piece of information successfully, another sensory system can show a facilitated response to the same input, suggesting that it is possible to transfer representations across the sensory domains. This is referred to as cross-modal priming or intersensory bias (Welch & Warren, 1980). A series of experiments demonstrated that changing the modality from the study procedure of 2-D and 3-D objects to the subsequent recognition test on them did not attenuate cross-modal priming and memory

transfer. The studies showed evidence for multisensory nature of object representations for implicit and explicit memory systems (Easton, Greene & Srinivas, 1997b). Also, in visuo-tactile priming, cross-modal facilitation does not involve modality specific, but rather shared multisensory representations or structural descriptions of objects – their volumetric "global form and structure" (Lacey et al., 2009; Reales & Ballesteros, 1999). In the context of this study is assumed that functional objects – beside 'structural descriptions' – bear action-related, semantic information (affordances) which might facilitate visuo-haptic integration.

1.2.2 Multisensory interactions in multistability and binocular rivalry

Is it possible that implicit knowledge about object-intrinsic, functional properties evoked by touch (of functional 3D objects) can modulate visual perception and facilitate multisensory integration? Such a question has not yet been addressed. Since multistable perception, including binocular rivalry, is sensitive to non-visual sensory perturbations and ecologically valid stimuli (Baker & Graf, 2009) – to which real objects are counted among –, it is chosen here as a tool to answer this question. It is therefore elaborated in the following.

Multistability is a perceptual phenomenon in which a single sensory stimulus has several perceptual interpretations. The stimuli can for example be ambiguous figures. Binocular rivalry, which is also a multistable phenomenon (Logothetis, Leopold & Sheinberg, 1996), is an extensively studied peculiarity of the visual system, serving as a psychophysical tool to investigate the dynamics of visual cognition and neural mechanisms of visual perception (Blake & Logothetis, 2002; Tong, Meng & Blake, 2006). During binocular rivalry each eye is presented with a different visual stimulus, resulting in a perceptual conflict situation where a perceptual alternation takes place (like with ambiguous figures). Numerous factors can resolve the perceptual ambiguity – that is prolong the perceptual stabilization phases – like the stimulus strength (Levelt, 1966) intermittent disappearance of the stimulus itself (Leopold et al., 2002), interocular grouping (Kovacs, 1996) or exogenous and endogenous attention (Chong; Chong & Blake, 2006; Mitchell, Stoner & Reynolds, 2004; Tadin & Blake, 2005).

More recently, the number of studies increased rapidly which show that binocular rivalry can be resolved by introducing signals from other modalities. They utilized cross-modal tasks in which an unambiguous stimulus in the non-visual modality is combined with the presentation of congruent and incongruent binocular stimuli. Evidence was shown for a 'olfactorial capture' in binocular rivalry (Zhou et al., 2010), while others found significant interactions of auditory signals with the rivalry dynamics (Conrad et al., 2010; Conrad et al., 2013; van Ee et al., 2009). It was reported that even semantic context arising from auditory signals can modulate visual perceptual selection (Chen, Yeh & Spence, 2011). The authors showed that the semantic content of a soundtrack constrained the perceptual duration of a

target image which was not congruent with the soundtrack. This suggests that the non-target image which matched the auditive stimulus was preferred indirectly although it was not attended to and although it had a lower stimulus strength in terms of luminance contrast. It was concluded that the semantically loaded auditory stimulus exerted a perceptual rather than a conceptual effect on binocular rivalry.

In the visuo-haptic domain, Blake, Sobel & James (2004) demonstrated that the visual rotation direction of a random-dot globe, exhibiting a bistable motion pattern, changed upon touching a haptic globe according to its felt haptic rotation direction. They suggested the involvement of middle temporal visual area to resolve visual ambiguity – but only when the two sensory channels are stimulated concurrently. Subsequently, other authors found that multistability can be resolved in favour of the stimulus which is under voluntary motor control (Hu & Knill, 2010; Maruya, Yang & Blake, 2007). Likewise, rivalry between Gabor patches can be biased by haptic skimming of a corresponding grating patch. The chance of switching an established visual percept increased if the haptic grating orientation did not match it; conversely, if the visual percept and the haptic signal matched, it was above chance likely that it would persist (Lunghi & Alais, 2013; Lunghi, Binda & Morrone, 2010). The effect was selective for the spatial frequency match between the two modalities, allowing the conjecture that somatosensory signals interact with neurons of early visual areas as early as in V1, which have the narrowest spatial frequency sensitivity (Lunghi & Morrone, 2012). Again, the question whether affordances and functional object properties can modulate the dynamics of binocular rivalry has not yet been adressed. Yet, given the abovementioned, semantic context might play a role during visuo-haptic processing.

The multisensory effects in binocular rivalry discussed so far are in accordance with previous propositions claiming that during binocular rivalry lower and higher neural levels of processing are actively involved in perceptual pattern formation. Lower neural stages inhibit and excite each other laterally for low-level pattern grouping, while higher level signals with full-fledged perceptual representations project to the lower-level neural populations via feedback connections (Leopold & Logothetis, 1999; Tong, Meng & Blake, 2006). At this point, it seems possible that multisensory signals can perturb the pattern formation mechanisms during binocular rivalry as there are many entry hubs for non-visual and representational signal inputs.

1.3 The world sketched within – semantic object representations as multisensory constructs

In this section, the nature and the role of object representations are discussed with respect to multisensory perception, since they seem to be involved in audio-visual and visuo-

olfactorial processing. Observers enact objects by means of sensorimotor activity engaging vision and haptics. The nervous system extracts their salient properties – their *structural descriptions* (Schacter, 1992), like shape and texture, and functional features, their *affordances* (Chemero, 2003; Gibson, 1979; v. Uexküll, 1920). They are the elemental building blocks of conceptual representations – internal models of the object's properties which have significance for the observer. These models are based on the physical features of the objects and the associated motor programs contingent to how they are interacted with, suggesting 'embodied abstraction' – the grounding of the object representations in perception and action (Binder & Desai, 2011; Goldberg, Perfetti & Schneider, 2005; Martin, 2007; Martin & Chao, 2001). For instance, recognizing an object becomes aggrieved when objects are explored haptically with gloves, suggesting that explicit memory is comprised of "distinctive, low-level, sensory-based information about objects" (Reales & Ballesteros, 1999). Analogous to the ideas in connectionism and parallel distributed processing (Rumelhart & McClelland, 1986), the current paradigm is that meaning is embodied by the spreading activation of a distributed associative network of sensory and motor brain areas. It enables on-line encoding and retrieval of task-relevant information. Taking this into account, do these representations play a role during real-time visuo-haptic integration or is visuo-haptic integration mandatory – taking place without necessarily relying on them?

Because of the grounding in sensorimotor activity and the fact that haptics is akin to the visual system „in recovering the topology of the physical space" (Gaißert, Bülthoff & Wallraven, 2011), the representations are considered multisensory (Norman et al. 2004; Pietrini et al., 2004), spatial and "flexibly accessible via both bottom-up and top-down inputs" (Lacey, Campbell & Sathian, 2007; Lacey & Sathian, 2012). It was argued that visuo-tactile object representations – though accessible by both of the modalities and subject to mutual priming to a high degree – are viewpoint-dependent, the reason being the hand's preference for sampling the "back" information of an object and the eye's preference for the "front" (Newell et al., 2001). In contrast, when controlling for equal accessability of the object information for both modalities, the constructed representations become viewpoint-independent (Lacey, Peters & Sathian, 2007). Though the question of viewpoint-dependency remains not conclusively resolved, the tendency is that the representations are viewpoint-independent, as indicated by cross-modal perceptual learning and the transfer of object category knowledge across the modalities (Lacey & Sathian, 2012; Yildirim & Jacobs, 2012). This suggests that they might inform the visual system during visuo-tactile processing. For familiar objects, the representations are built up of a more robust network of different structural descriptions, compared to those of unfamiliar objects which rely mostly on visual information (Lacey & Campbell, 2006). Encoding of both object types involves covert verbal descriptions, as suggested by interference effects, and hence rendering them possibly

linguistic as well. From the records of patients with semantic dementia it was concluded that the visuo-haptic cross-modal perceptual representations can be dissociated from the functional and conceptual knowledge of the same objects (Capitani et al., 2003; Forti & Humphreys, 2005; Tulving & Schacter, 1990). Nevertheless, it must be stated that both systems, though being dissociable in test situations and in semantic disorders, interact tightly via task-sensitive, reciprocal feedback loops in normal observers (Binder & Desai, 2011), so that the representations still could modulate visual perception.

The substrate of the categorial conceptual and perceptual knowledge and the affordances of manipulable objects, tools for example, comprises of a distributed network of brain areas. A number of studies allow the hypothesis that affordances are automatically detected and processed in regions which are also involved in motion related to how the object is used and regions involved in naming the object (Martin, 2007). Visual presentation and recognition of tools elicits activation in primary visual areas of the occipital cortex, medial fusiform gyri (mFG), posterior middle temporal gyri (pmTG) of the posterior ventral temporal cortex (pvTC), inferior temporal cortex (iTC), the LOC, left posterior parietal cortex (lpPC) near intraparietal sulcus (IPS) and left ventral premotor cortex (vPMC) (Chao, Haxby & Martin, 1999; Chao & Martin, 2000; Kellerbach, Brett & Patterson, 2003; Martin et al., 1996; Pietrini et al., 2004). The network is interpreted as a motor-based feature space and as a web of action knowledge about manipulable objects which ultimately form the semantic representations of objects. It is automatically activated upon the visual presentation of the objects and many areas exhibit stronger neural activation upon the presentation of manipulable in comparison to non-manipulable objects regardless of the intention to act (Chao & Martin, 2000; Grèzes et al., 2003; Kellerbach, Brett & Patterson, 2003; Markis, Hadar & Yarrow, 2011). This strongly suggests a potential contribution of object-intrinsic features to multisensory integration. Haptic exploration of real objects, in contrast, engages secondary somatosensory area (SII), IPS, left prefrontal cortex (lpFC), insular cortex and left premotor cortex (FEF), lateral and medial secondary motor cortices, LOC (neighboring retinotopic areas), inferior temporal cortices (iTC) (including fusiform gyri) and medial temporal areas (Amedi et al, 2002; Reed, Shoham & Halgren, 2004). Finally, due to their shared capacity to uncover spatial properties, both modalities have overlapping processing brain areas, on which bisensory information converges. These areas are involved in multisensory binding and conceptual processing. They include the LOC, IPS, superior parietal gyrus (sPG), ventral temporal cortex (along with the fusiform gyri), pmTG and the anterior SII (Amedi et al., 2002; Binder & Desai, 2011; Hoffman et al., 2012; Lacey et al., 2007; Zhang et al., 2004). It is plausible to assume, that, at the onset of appropriate object-related stimulation, the spreading activation in this web reflects the affordances, the motor

based knowledge how to use the object and probably some episodic and verbal associations related to it.

1.4 Predictive coding underlying multisensory processing and binocular rivalry

An essential feature of multisensory processing is the capacity to resolve perceptual ambiguities which occur from the equivocal interpretability of the same physical stimulus. Confronted with multiple perceptual interpretations, for observers to survive it is essential to operate on mutually exclusive and reliable perceptual representations (Leopold & Logothetis, 2000). The interpretation of experimental data of numerous recent studies favours an inferential theory of perception, which claims that “sensory input stands in the same relationship to the percept as evidence does to a conclusion” (Richeimer, 2006). It traces back to Herrman von Helmholtz (Helmholtz, 1867) and is currently the favoured epistemology of perception in enactive cognitive science, computational neuroscience or robotics. Herein, the nervous system acts not as a passive information receiver, but more like an active generator of hypotheses or internal models about the causes of the sensory stimulation. The models in turn construct predictions about upcoming sensory events. Analogous to the Kalman filter (Kalman, 1960), the state-updating and prediction is a cyclical, recursive process which builds the core of the predictive coding framework (Friston & Kiebel, 2009). The generation of predictions is context-sensitive and both the lower-level sensory signals as well as higher-level cognitive factors, like semantic cues, can trigger specific expectations. In predictive coding, bottom-up multisensory signals are matched to top-down prediction signals along hierarchical cortico-cortical pathways. If the expectations are not met, a prediction error signal drives the re-adjustment of the models which in turn either change the predictions or change the interpretation of the current physical stimulus. Thus, for the generation of coherent, veridical multisensory perceptual states, minimization of prediction error is vital. This means that discrepant sensory signals which are not consonant with a unitary event or context will be suppressed or discarded while consonant signals will be preferably selected for (Bruner, 1957; Firston & Kiebel, 2009; Hohwy, Roepstorff & Friston, 2008; Welch & Warren, 1980).

For example, at Easter the Jastrow duck-rabbit bistable image was interpreted as a rabbit, while in autumn as a bird (Brugger & Brugger, 1993), suggesting that the seasons play a role in setting up a higher-level context which biases the generation of perceptual hypotheses. Similarly, the “my wife or my mother-in-law” bistable figure is longer interpreted as the “mother-in-law” when the accompanied soundtrack is the voice of an old lady, and conversely as “my wife” when the voice of a young woman is heard (Hsiao et al., 2012). During visuo-haptic interaction, less reliable, noisy visual cues lose weight in light of more

reliable haptic cues when estimating an object's shape (Ernst & Banks, 2002; Helbig & Ernst, 2007). As pointed out earlier, similar effects can be observed in situations with binocular rivalry in which the haptic or auditory sensory channel constructs a perceptual model. The associated preference for the congruent visual stimulus fits well the idea that the prediction error is minimized (Conrad et al., 2010; Conrad et al., 2013; Chen, Yeh & Spence, 2011; Lunghi & Alais, 2013; Lunghi, Binda & Morrone, 2010; van Ee et al., 2009). Also, the visual signal from one eye persists longer and is suppressed shorter if it is congruent to the haptic signal compared to the incongruent visual signal from the other eye.

In binocular rivalry multiple stages of neural processing are involved: early stages with monocular channels and late stages with pattern representations (Tong & Blake, 2006). Passing on predictions from frontal areas to earlier processing stages and feed-forwarding error signals to later stages may well determine what perceptual content will be dominant during binocular rivalry (Hohwy, Roepstorff & Friston). In fact, there is evidence that the neural activity correlated with a specific percept is located in fronto-parietal areas, suggesting their role in perceptual stabilization (Leopold et al., 2002; Logothetis, Leopold & Sheinberg, 1996; Sterzer, Kleinschmidt & Rees, 2009; Sterzer & Rees, 2008).

Reed, Shoham and Halgren (2004) point out that during nonsense object palpation a different, smaller and weaker activated web of brain areas is involved compared to sensing functional objects. This suggests that the interaction with functional objects is associated with more neural signals which might contribute to more stable perception. The approximate ratio of the counts of the actived voxels for the two object types is 1:3. They conclude that perceptual tactile processes can be told apart from higher-level haptic cognition. This observation squares with effective connectivity data as revealed by Granger causality: while a more dense and more clearly segregated (mainly frontal) network feeds into LOC and IPS representing a 'top-down' prediction network comprised of frontal areas during interaction with familiar objects, a less segregated and more sparse 'bottom-up' network feeds the LOC consisting of mainly somatosensory inputs when interacting with unfamiliar objects. This significant difference is reflected by a ratio of 1:1.7 for the path weights of the connections between all of the involved areas for the two object classes (Deshpande et al., 2010). This suggests that there are more neural signals generated during interaction with known objects compared to non-sense objects – a fact which might bring about more internal associations for familiar objects. In the context of this study, LOC would be activated by haptic exploration and match the incoming signals with the top-down predictions. Given that one of the two visual images corresponding to the haptic object is selected and held in consciousness, there would be less error signals propagated to higher areas. In that case, the visuo-haptic integration would be rendered more stable, since error signals drive perceptual alternations in higher, fronto-parietal areas. In addition, if the observer explores familiar objects, the

bistable perception might be even stronger stabilized due to the more dense reciprocal network between LOC and the frontal areas. This idea is supported by the fact that LOC shows higher activity when the bistable stimulus is resolved in favour of one specifically grouped shape representation (Fang, Kersten & Murray, 2008).

Hence, the predictive coding framework would suggest that upon the presentation of meaningful stimuli the nervous system not only infers the causes of the stimulation but – in addition to basic feature processing – it performs conscious and unconscious classification and extraction of meaningful associated sensory patterns which are projected upon bottom-up signals (Grèzes et al., 2003; Sanguinetti, Allen & Peterson, 2013), and, by doing so, stabilizes perceptual states. We like to refer to this conjecture as 'volleys of conceptual processing' during perceptual experience.

1.5 Aims, the scope and the hypotheses of the present study

How does haptic modality along with other factors contribute to the resolution of perceptual conflicts and to the stabilization of visual perception? In the current study, we investigated whether motor-cognitive factors, affordances and object-intrinsic features, exert stronger crossmodal binding effects during interaction with familiar objects compared to unfamiliar objects. We chose binocular rivalry, since it is sensitive to non-visual sensory perturbations and ecologically valid stimuli – to which real objects are counted among. Moreover, up to now all of the multisensory binocular rivalry studies addressed the question using fairly simple stimuli, void of contextual salience, naturalistic features and non-perceptual associations. Yet, as pointed out earlier, approaches with "naturalistic pairs" and functional objects are also necessary, instead of "arbitrary or minimalist pairs" to investigate crossmodal binding during unitary events (Baker & Graf, 2009; de Gelder & Bertelson, 2003; Reales & Ballesteros, 1999). In this sense, this study seeks also to investigate whether its findings would tie in with the findings of former studies on multisensory integration during binocular rivalry in which also semantic features and naturalistic pairs were used as bimodal stimuli (Chen, Yeh & Spence, 2011; Zhou et al., 2010).

Given that upon the presentation of known objects affordances and other object-intrinsic factors automatically and implicitly drive the activity in a sensorimotor network associated with how familiar objects are used (Grèzes et al., 2003; Kellenbach, Brett & Patterson, 2003), it is anticipated here that this network generates a stronger predictive code which stabilizes stronger the visuo-haptic integration than data-driven somatosensory signals during presentation of unfamiliar objects. Similarly, given the stronger connectivity between the frontal areas and the visuo-tactile convergence sites like LOC and mFG for familiar objects ('top-down'-network) compared to the somatosensory connection to the LOC

(‘bottom-up’-network) for unfamiliar objects, a stronger visuo-haptic stabilization for familiar objects is predicted compared to unfamiliar objects. This may especially be expected because perceptual stabilization during binocular rivalry is correlated with fronto-parietal activity – a neural web which has overlaps with the ‘top-down‘ network of familiar objects (Leopold & Logothetis, 1999; Sterzer, Kleinschmidt & Rees, 2009). If this is the case, then one could conclude that multisensory processing is not only driven by the integration of structural descriptions and spatial information of familiar objects but also by conceptual information – ‘volleys of conceptual processing‘.

In contrast, it is possible that visuo-haptic integration is not facilitated or improved by object familiarity and that the perceptual fusion of visual and haptic signals occurs mandatory, automatically and implicitly – not relying necessarily on internal associations and motor-cognitive factors related to a known object. Since it was previously found that perceptual tactile processes can be told apart from higher-level haptic cognition (Reed, Shoham & Halgren (2004), that would be the alternative hypothesis. It would also reflect the possibility that familiar objects are comprised of ‘sparse‘, action-related representations which have a similar predictive code strength like unfamiliar objects, indicating that perceptual stabilization is not necessarily correlated with higher neural activity. Additionally, it is also possible that the automatic activation of motor codes by affordances of familiar objects is generated very rapidly and that the influence is too short-lived to have a lasting and penetrating effect on multisensory integration (Markis, Hadar & Yarrow, 2011).

It was anticipated here that peceptual switches frequency (*PSF*) would be the highest and perceptual dominance duration (*PDD*) for the visual target would the shortest when there is no haptic stimulation (for both familiar and unfamiliar objects). Conversely, lowest *PSF* and longest target dominance duration are expected when visuo-haptic stimuli congruency is given with familiar objects, followed by congruency with unfamiliar objects. After the initial haptic priming, with the initial visual presentation of the dichoptic stimuli (‘onset rivalry’), the observers would select the haptically congruent image more frequently in the trials with familiar objects compared to unfamiliar objects in the two-alternative forced choice task (2AFC). Overall, the reaction time was anticipated to be shorter in trials during visuo-tactile congruency with familiar objects, since the ‘multisensory redundant target effect’ was expected to hold here – the fact that bisensory stimulation speeds up the reaction time (Diederich & Colonius, 2004; Forster et al., 2002; Laurienti et al., 2003).

2. Methods

2.1 Observers

The undergraduate volunteer students (11 males, 13 females) of the University of Vienna participated in the experiment (mean age, 23.7 years; range 19–29 years). All reported normal colour vision and normal or corrected-to-normal vision and none oft hem had a record of pathological somatosensation or of any other neurological abnormalities. All signed a written consent and received credit needed to fulfil class requirements as compensation. None of them knew the purpose of the experiment.

2.2 Haptic and visual stimuli

The subject sat on a chair in front of a table with the chin positioned on a chin rest. Parallel to their viewing direction two Samsung 943BM TFT-displays were placed, one on the left and one on the right side (1280x1024 px, 300 cd/m^2, 170°x160°, 75 Hz vertical refresh rate). Over a Wheatstone mirror stereoscope two images were projected to the eyes, creating a dichoptic, competitive stimulation (setup similar to Blake, Sobel & James, 2004). A round plate-like, rotatable platform was positioned behind a covering in front of the stereoscope on which the objects could be placed in 12 random radial positions for the haptic exploration. The covering made sure the subject's view on the platform was obstructed and periphery vision was minimized. The subjects' line of gaze to the platform simulate a 'canonic' perspective angle which observers typically encounter in everyday situations, for instance when standing in front of a table and looking down on an object in order to manipulate it. The participants did not see the haptic artifacts, as the view was obstructed by the mirrors of the stereoscope. Yet the visual stimuli coincided with the objects in location in front of the stereoscope at a distance of about one armlength (fig. 1). The room was darkened during the experimental sessions. This facilitates visuo-tactile binding, a factor which – in analogy to the 'spatial rule' which claims that spatially coincident multisensory cues in combination generate enhanced neural responses – facilitates coherent sensory experience (Stein & Meredith, 1993). The room was darkened during the experimental sessions.

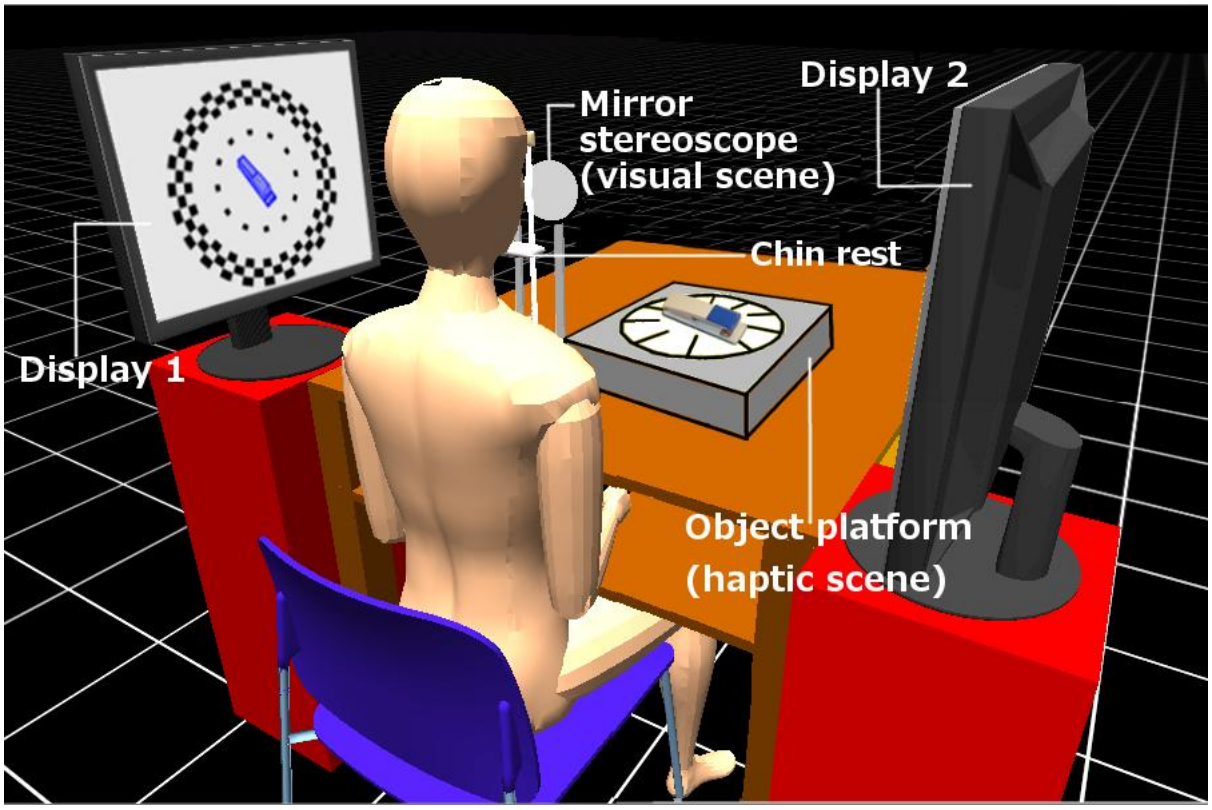

Figure 1. Experimental apparatus. Observers looked 'down' through the stereoscope on the objects which were latched on the rotatable platform. The dichoptic object images were calibrated on the displays so as to appear to the observer as having the same dimension like the haptic object in front of them. A chin rest limited head movements and a covering (not shown here) was installed between the stereoscope and the object platform to obstruct the view on the surrounding.

2.3 Haptic and visual stimuli

Because in the present work it is assumed that functional and non-perceptual features have an effect on multisensory processing, 20 well-known hand-sized everyday haptic objects were used for haptic stimulation (fig. 12 b, Appendix). To contrast the expected effect, 20 geometrical shapes with no known function or inherent meaning were presented as control (fig. 12 a, Appendix). All objects used were hand-sized, oblong and had one principal axis and one or none symmetry plane, allowing for a random placement on the platform in varying angular positions (fig. 12 c, Appendix). Since the question whether the multisensory object representations is not conclusively resolved, the variation of the angular position across the trials was expected to minimize habituation to just one viewing angle and to avoid latent spatial priming. The overall size, length and width of the objects of both categories were comparable (Mann-Whitney-U-test: Z_L = 0.230, p = 0.818; Z_B = 0.081, p = 0.935), but the height was different (Z_H = 3.45, p = 0.0006), leading to different volumentric data (Z_{Vol} = 2.178, p = 0.029). Yet, both object categories allowed for highly similar precision

and power grips for the exploration, rendering them comparable in terms handling. The 20 objects of each class were divided in 10 fixed pairs which shared similar size and complexity and were pre-arranged before each of the two experimental sessions.

The visual stimuli were images of the same objects of which in every trial a different pair was shown. The objects were photographed in 12 angular positions from the perspective in which the observer would virtually see them behind the mirrors. The images were processed with *ImageMagick*, outlining the most distinctive edges with Sobel filters. The object's visual angle was θ_L = 9.7° maximum along its principle axis and θ_B = 5.7° maximum perpendicularly to the principle axis. Shrinking the retinal image by 1° improved the exclusive appearance of the percept further (8.7° x 4.7°). This remained unnoticed by the subjects during the visuo-tacile interaction, because this was a change below visuo-tactile detection threshold (Rock & Victor, 1964). Thus, the object's angular size was sufficiently small to lower piecemal effects while retaining the size which corresponded to the real haptic objects (Blake, 2001). Moreover, suppression of one of the two visual stimuli is deeper and more coherent if both stimuli share similar global complexity (Alais & Melcher, 2006), which was realized here by pairing objects of similar size and overall appearance.

In each trial, the image pair was randomly tinted with two out of three different colours (red, green, blue). The objects' edges were darker than the objects' surface areas. To ensure equal stimulus salience, the edge colours and the surface colours were calibrated so as to have a nearly identical luminance across displays and relative to each other. This was measured by a colorimeter in the Lab colour space (table 1, appendix). A light grayscale background around the object images was chosen to enhance the brightness contrast between the depicted object and the background. Since dichoptic stimuli are fused better on darker backgrounds, a brighter background renders them more disparate and distinguishable for the perceptual switches (de Weert and Levelt, 1976).

2.4 Experimental design and independent and dependent variables

Object class (synonymous to 'object familiarity': *FAM*) and sensory stimulation condition (*SSC*) were orthogonal factors. The two object classes (familiar, *FAM+*, and unfamiliar, *FAM-*) were presented in three sensory stimulation conditions, in a total of six independent variable blocks (2 x 3 repeated measures design; fig. 2). Two experimental sessions were conducted (á ~100 min), one for each object class (*FAM+* and *FAM-*), with no more than 4 days apart. In each experimental session, one block was unimodal visual-only (*VO*) whereas the other two were bimodal, containing visuo-tactile congruent (*VTC+*; one of the two visually presented images corresponds to the haptic object by identity and spatial alignment) and incongruent trials (*VTC-*; neither of the visually presented images

corresponds to the haptic object by identity and spatial alignment; see also figure 3 b). Both experimental sessions started with the *VO* block, followed by either the *VTC+* or the *VTC-* block, for both of which the presentation order was randomly assigned and counterbalanced across subjects. Each of the three blocks comprised of 40 trials (with total of 120 trials for each familiarity condition). Every object pair was shown four times per block, whereby each object of that pair was presented two times to each eye. For the trials of the first block (*VO*), the colour and the object orientation were randomly assigned to each object of the pair; those configurations (colour and orientation) were then the same for the other two blocks (*VTC+* and *VTC-*). During the bimodal block trials, the initial haptic priming was followed by binocular presentation of either an incongruent or congruent stimulus pair in a two force choice task (2AFC; in each of the *FAM+* or *FAM-* experimental sessions).

The dependent variables were the same for all blocks (*VO, VTC+* and *VTC-*) of both experimental sessions (*FAM+* and *FAM-*). The initial visual percept and the subsequent perceptual switches were recorded by pressing the corresponding colour key. The reaction time (*RT*) and the initial target selection rate were a performance index as were the target dominance durations (at the trial onset and throughout the trial) and the number of subsequent perceptual switches (*PSF*). Since the *SSC* blocks were identical with respect to the trial sequence, object colour, object orientation and the eye-of-origin for the object in a trial, but differing only with respect to the multisensory conditions, *a target was defined as the colour in which the object was presented when it was congruent with the haptically explored object during the VTC+ block*. Thus, that colour – and to the same measure the correlated object – was also the target in the same trials of the *VO* and the *VTC-* block. Conversely, the non-target was the object image which was not congruent with the touched object.

		SSC		
FAM	Experimental session with unfamiliar objects (*FAM-)*	**Unimodal (*VO*)**	**Bimodal (*VTC+* or *VTC-*)**	**Bimodal (*VTC+* or *VTC-*)**
	Experimental session with familiar objects (*FAM+*)	**Unimodal (*VO*)**	**Bimodal (*VTC+* or *VTC-*)**	**Bimodal (*VTC+* or *VTC-*)**
		40 trials	**40 trials**	**40 trials**

Figure 2: Schematic summary of the experimental session design. It was randomly assigned and counterbalanced among the observers with which experimental session each of them started the data aquisition (with either *FAM+* first or *FAM-* first). The two sessions took place on two different days. It was also randomly assigned and counterbalanced which block (*VTC+* or *VTC-*) after *VO* was the next. It was expected that the dependent variables would differer between the congruent conditions (*VTC+*) of the two familiarity sessions, with *FAM+* exerting more facilitating responses, lower *PSF*, higher target selection rate and longer target dominance duration.

2.5 Experimental procedure

The subjects were not aware of the original purpose of the study, instead they were told that the experiments are probing colour vision in relation to haptic perception. Moreover, to avoid the bias that the participants might merely report the visual percept that happens to be congruent with the haptic object – rather than reporting the percept that is more dominant and salient in first place – the observers were asked to report the colour in which the dominant visual object was tinted. Whether a subject began the data acquisition with *FAM+* or *FAM-* condition was randomly assigned and counterbalanced across the subjects (12 versus 12). The day before the *FAM+* experimental session, the subject received a list of names to learn by which the familiar objects are normally labelled (table 2, appendix). In their first experimental session, subjects‘ handedness, colour vision (Ishihara plates 1, 11, 19, 23) and ocular dominance (Dolman method) were tested. The experimental sessions were conducted with the *Matlab®PsychToolbox* and its built-in image processing routines (Brainard, 1997). In order to learn the accurate colour-percept key-mapping, 15-20 training trials were conducted prior to the experimental procedure. Once the subjects performed at least 10 error-free trials (they do not hit the wrong colour key), the experimental session began during which grey noise was deployed over headphones to avoid distractive auditory stimuli from the lab environment.

The experiment began in in both experimental sessions (*FAM+* and *FAM-)* with the *VO* block. Herein, by the onset of the trial, the subjects were briefly presented a dichoptic pair of objects for 700 ms in a 2AFC. For the reaction time record, they were instructed to respond as quickly and as accurately as possible via the according colour-key, and by doing so to indicate which object was perceived as predominantly. Subsequently, the same image pair was presented for 20 s, during which the subjects were asked to report every perceptual switch by hitting the corresponding colour-key (fig. 3 a). A switch was defined as the perceptual state in which one object colour is perceived as being overridden for its greater part by the other object colour. Each orientation along the principal axes of the two visual objects was independently and randomly assigned to one of the 12 orientations. The trial ended by the initiation of an inter-trial-interval of 4 s. Then the next object pair was presented until the *VO* block ended after 40 trials. If they hit the wrong colour, the trial was repeated later within the same block.

The following 80 trials belonged to the two bimodal blocks (fig. 3 b). The haptic object either corresponded in one half of the cases to one of the two binocular object images (*VTC+*) and in the other they did not (*VTC-*). Humans have remarkable recognition capacities in the haptic domain when faced with familiar objects belonging to everyday sensorimotor experience. Correct naming occurs in 94 % of cases within 5 seconds (Klatzky, Lederman &

Metzger, 1985). Every bimodal trial began with a high-pitched (880 Hz) 0.1 s sine tone over headphones which afforded the participants to explore the object haptically for 5 s as well.

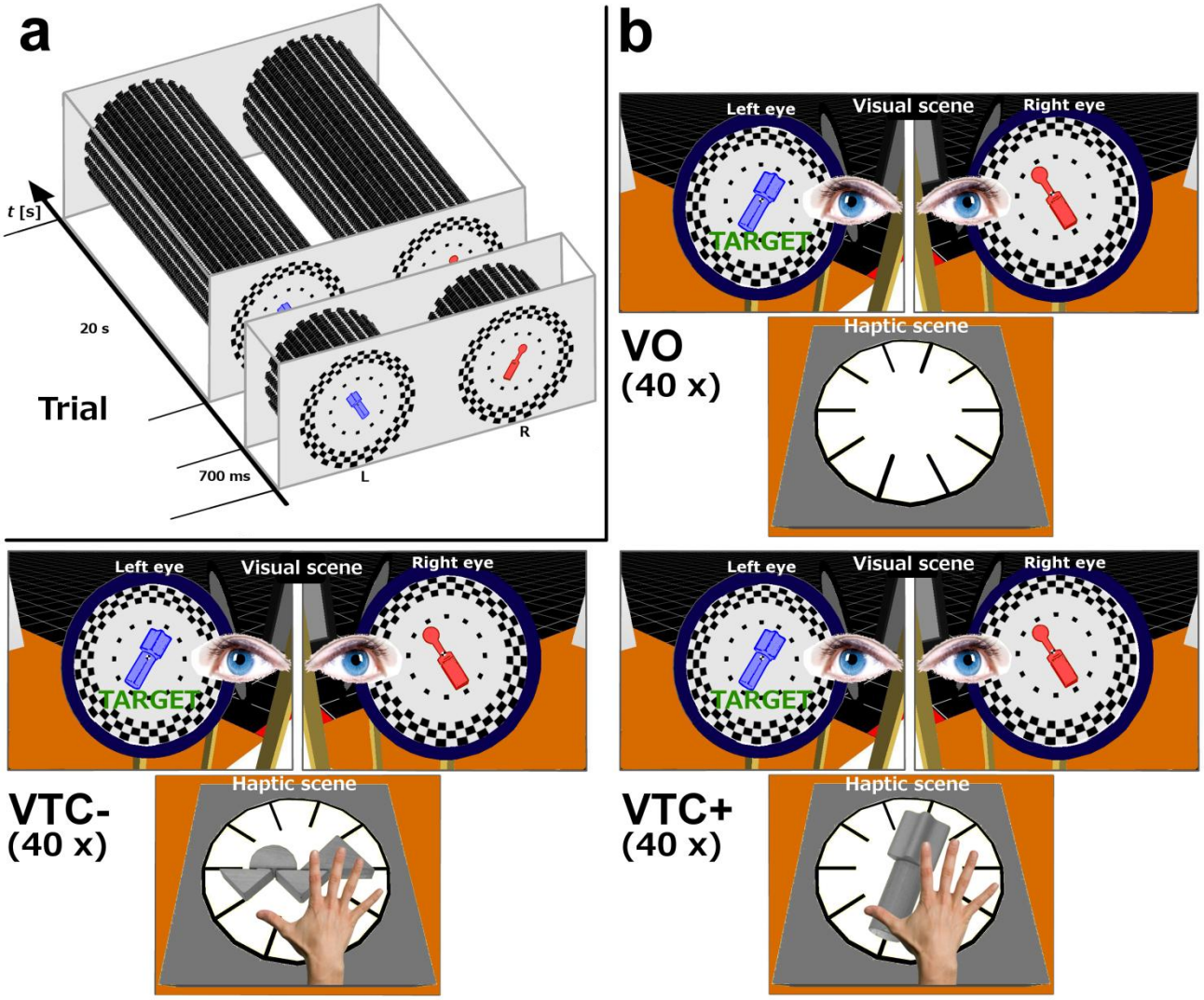

Figure 3. Experimental design (exemplified with a pair of unfamiliar objects) **(a)** in all three blocks the observers were briefly presented (700 ms) with a dichoptic image pair at the trial onset. Subsequently, the screen was cleared and the reaction time and perceptual selection were recorded. Then, the trial continued for 20 more seconds in which the same object image pair was shown and the perceptual switches were tracked. In the multimodal blocks haptic stimulation occurred concurrent to binocular rivalry. **(b)** Scheme of the sensory stimulation conditions (*SSC*) of all three blocks. Upper right: visual-only (*VO*), lower left: visuo-tactile incongruent (*VTC-*), lower right: visuo-tactile congruent (*VTC+*) condition. The first block was always *VO*. It was randomized which of the two bimodal blocks was the next. The blue object is the target in *VO* and *VTC-* since it is congruent to the haptic object during an identical trial of the *VTC+* block. Each of the blocks comprised of 40 trials, totalling up in 120 trials per experimental session.

The object was latched beforehand by the experimenter on the rotatable plate in a randomly assigned orientation. The subjects were instructed to palpate after the edges, the texture and the stiffness of the object. This standardized their hand movement towards typical ‘exploratory procedures‘ (Lederman & Klatzky, 1987). The end of the exploration was signalled by a 100 ms low pitched sine tone (440 Hz).

If the experimental session was conducted with familiar objects, then the subjects had to name the object after the exploration according to the list, and call the orientation according to the numbers 1 to 6 (due to rotational symmetry of the plate, the orientations with the numbers 7 to 12 were just the same as the orientations 1 to 6). With unfamiliar objects, they were asked to call the orientation only. Subsequently, a 2AFC followed again with the binocular object images for 700 ms after which the subject had to report the dominant colour via a key stroke as quickly as possible. If they called the wrong name, did not know the object or if they hit the wrong colour, the trial was repeated later at a random point of time. After the brief binocular stimulation, they had to report the perceptual switches for 20 s. The next trial was initiated after an inter-trial-interval of 5 s.

Upon the completion of 40 trials of the bimodal condition (40 min), the subject took a 5 min pause. After the experimental session, an informal inquiry was conducted regarding the overall impression of the binocular stimuli, general remarks and whether they could tell what the experiment was about.

3. Data analysis and results

The data of three subjects was discarded from analysis. There were either too many trials where reaction time was repeatedly above 2 seconds or in too many trials no response was shown at all. Also, in one case the subject made too many errors during the training trials and kept the high error-rate during the experimental session. Data values below 1.00 (probabilities) undergone arcsine and data values over 1.00 (target dominance durations, reaction time and *PSF*) undergone square-root transformation to homogenize the variances before they were submitted to the ANOVAs. For the ANOVAs, Mauchly's tests were applied to check the sphericity of the data. Violations from sphericity were compensated for by adjusting the degrees of freedom with Greenhouse-Geisser's ε and the adjusted *p*-values are reported. Once the normality assumption of the data was rejected, non-parametric procedures were applied.

3.1 Perceptual selection at the trial onset

As haptic priming is known to affect visual perception (Easton, Greene & Srinivas, 1997; Reales & Ballesteros, 1999), the target selection rate at the rivalry onset was assessed. Once again, the target was defined in the *VO* and *VTC*- blocks as the same colour/visual object which was congruent to the haptic object during the *VTC*+ block. To test for an advantage of object familiarity for the probability of selecting the target, the selection rate for the same object which was congruent during the *VTC*+ block was compared again to its selection rate during the *VO* and *VTC*- blocks. The target selection rates for the latter conditions were expected to lie at chance level, while during the *VTC*+ block they would increase, especially during the *FAM*+ session.

For each subject, the selection ratio was computed by adding up all the target hits within a block and dividing it by the sum of the hits and misses which yielded the probability for selecting the target in a trial of the respective block:

$$p_{SSC} = \frac{N_{Hit}}{N_{Hit} + N_{Miss}}$$

Next, the data sets' variances were homogenized by the arcsine-transformation and submitted to the within-subject 2 x 3 ANOVA, again with *FAM* and *SSC* as factors.

There was a highly significant main effect of multisensory stimulation condition on the target selection rate ($F(2, 40) = 10.922$, $p = 0.004$, $\eta_p^2 = 0.351$). The object familiarity,

however, did not exert a main effect ($F(1, 20) = 0.169$, $p = 0.799$). There was also no interaction between the two predictors ($F(2, 40) = 1.644$, $p = 0.206$). As the data violates the normality assumption, four post-hoc Wilcoxon signed rank tests with Bonferroni-Holm correction showed a significantly higher target selection rate when there was visuo-tactile congruency for both familiar ($P_{VTC+} = 0.607$ vs. $P_{VO} = 0.479$ and $P_{VTC.} = 0.505$) and unfamiliar object classes ($P_{VTC+} = 0.594$ vs. $P_{VO} = 0.517$ and $P_{VTC.} = 0.500$). A binomial test confirms this finding; for instance, given the mean probability of $p_{VO} = 0.479$ for the target selection in the VO_{FAM-} *block*, one would expect only with a probability of $p = 0.045$ that the mean number of target selections ($i = 24$) out of $n = 40$ trials in the $VTC+_{FAM-}$ block was obtained by that probability ($B(24|0.479, 40) = 0.045$). The familiarity conditions did not differ significantly ($W = 96.5$, $p = 0.531$). See also table 1 and figure 4 for the summary of the multisensory effects.

Table 1. Pairwise comparisons via Wilcoxon's signed rank tests (including Bonferroni-Holm corrected *p*-values and Cohen's *d*) demonstrate significant differences between the *SSCs* regarding the mean probability *P* for the target selection at the onset of a trial (standard deviations in brackets).

FAM	$\bar{t}_{VO}$ ***(σ)***	$\bar{t}_{VTC+}$ ***(σ)***	$\bar{t}_{VTC-}$ ***(σ)***	***W***	***p***	***Cohen's d***
FAM+	0.479 (0.052)	0.607 (0.128)		21.5	0.006	1.103
		0.607 (0.128)	0.505 (0.082)	34	0.024	0.867
FAM-	0.517 (0.078)	0.594 (0.117)		57	0.091	0.731
		0.594 (0.117)	0.500 (0.064)	37	0.027	0.899

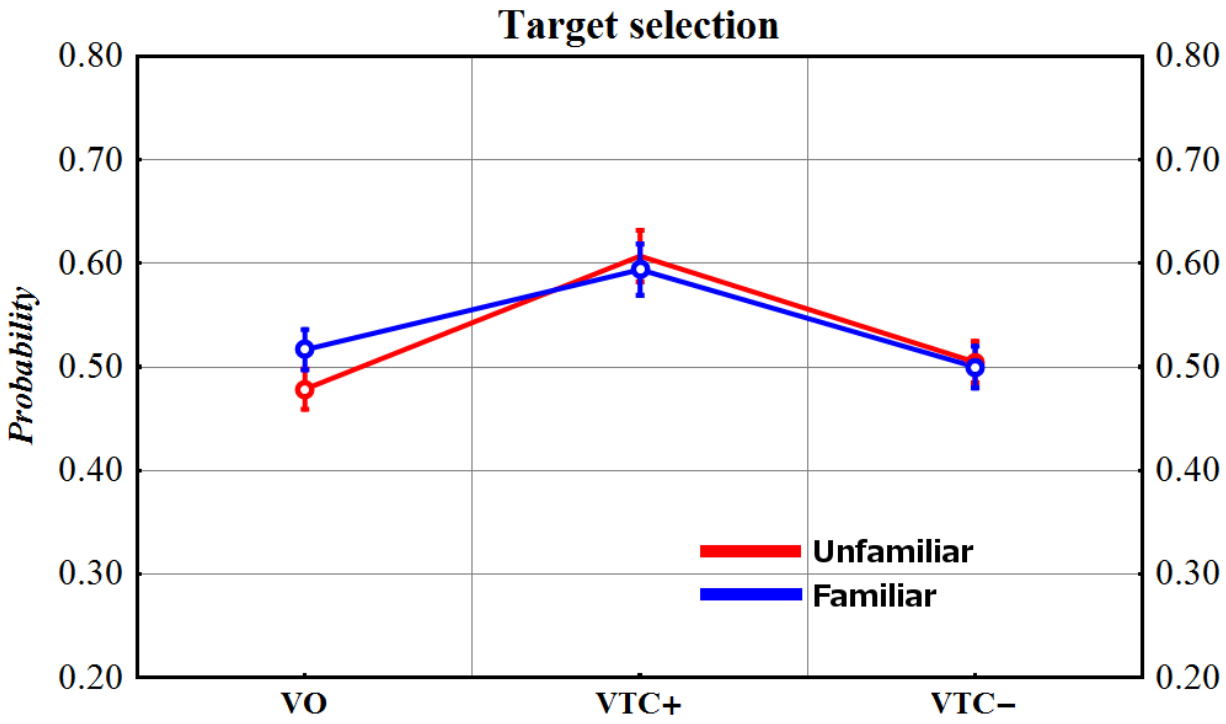

Figure 4. Target selection probability for familiar and unfamiliar object classes under different multisensory conditions; for both classes, visuo-haptically congruence yielded statistically significant effect compared to the visual-only and visuo-haptically incongruent conditions; the within-subject error bars denote the 95-% confidence interval.

3.2 Perceptual dominance durations

3.2.1 Target dominance duration during the trial

In the next step, change of the dominance phase durations was analysed across the blocks – the most important indicator for the modulation of binocular rivalry. Every button hit indicating a peceptual switch ended a perceptual duration recording and set the stop watch back to "0" to record the duration of the subsequent percept. This generated a vector of perceptual durations for each block. For each subject, dominance phases beyond 3 SD were discarded from analysis. When the temporal data across all subjects and all blocks is pooled together, the distribution of the perceptual durations can usually be described by the gamma function $\Gamma(\alpha, \beta)$, with α as the shape and β as the scale parameter (Levelt, 1967). Here, however, the distribution is better fitted by the log-normal density function (Lehky, 1995; figure 5):

$$f_X(x;\ \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}}\, e^{\frac{-(\ln(x)-\mu)}{2\sigma^2}}$$

The parameter μ denotes the distribution's mean while σ denotes the standard deviation. With the μ = 0.239 and σ = 0.544, the maximum log-likelihood is higher for the log-normal than for the best approximation with the gamma density function ($log(L(\mu, \sigma))$ = -61282.4 vs. $log(L(\alpha, \beta))$ = -64191.6).

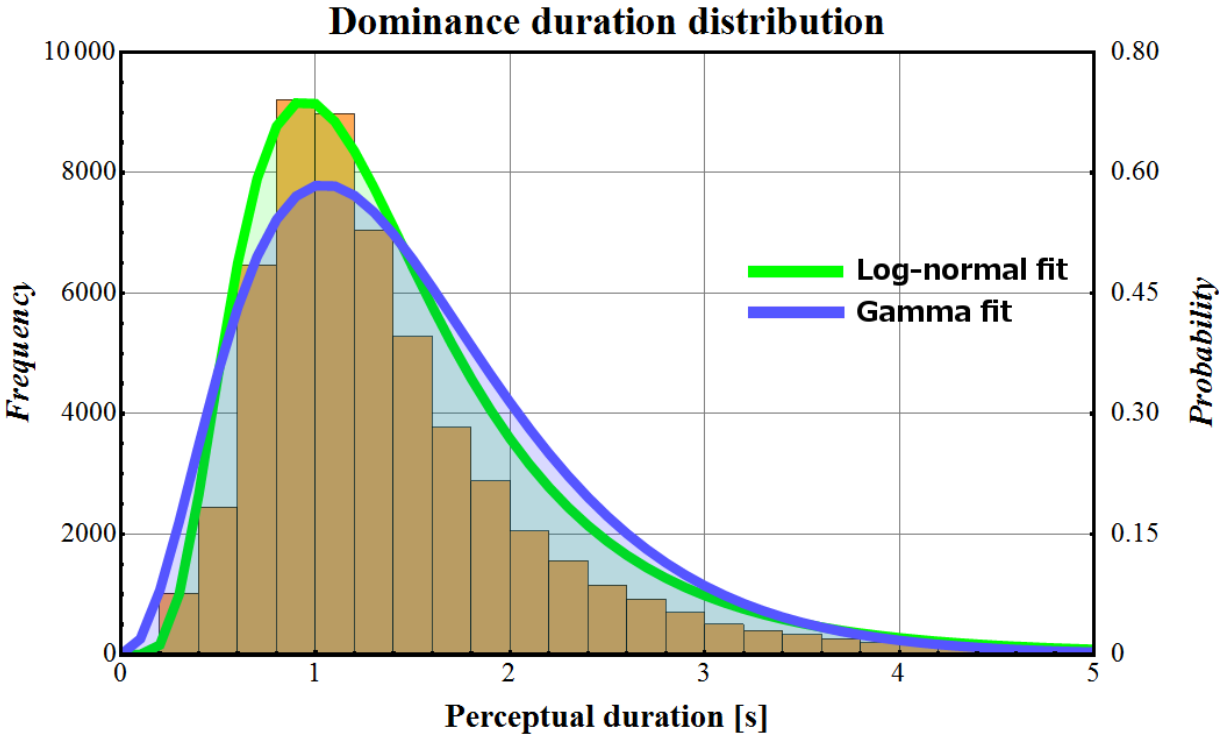

Figure 5. Frequency distribution of the pooled perceptual duration (left y-axis) fitted with the theoretical log-normal (green) and gamma (blue) probability densities (right y-axis). Values above 6 *SD* were considered outliers ($\bar{t}$ = 1.49 s, *SD* = 0.967). The maximum log-likelihood for the log-normal distribution fit was higher and thus the log-normal probability density function was selected for further analysis of the perceptual phases data.

Subsequently, the data of each block was split into subsets containing the dominance phases of the target and the non-target. Again, the target was defined in the *VO* and *VTC-* blocks as the same visual object which was congruent to the haptic object during the *VTC+* block. Conversely, the non-target was the object image which was not congruent to the touched object. It was expected that for each block the distributions of the perceptual durations for the target would overlap but that the mean durations for the target across the three blocks would differ (figure 6). To test for any significant differences, the duration values were submitted as response variable to a repeated measures 2 x 3 ANOVA again with *FAM* and *SSC* as factors. Beforehand, the duration values were square-root-transformed to homogenize the variances.

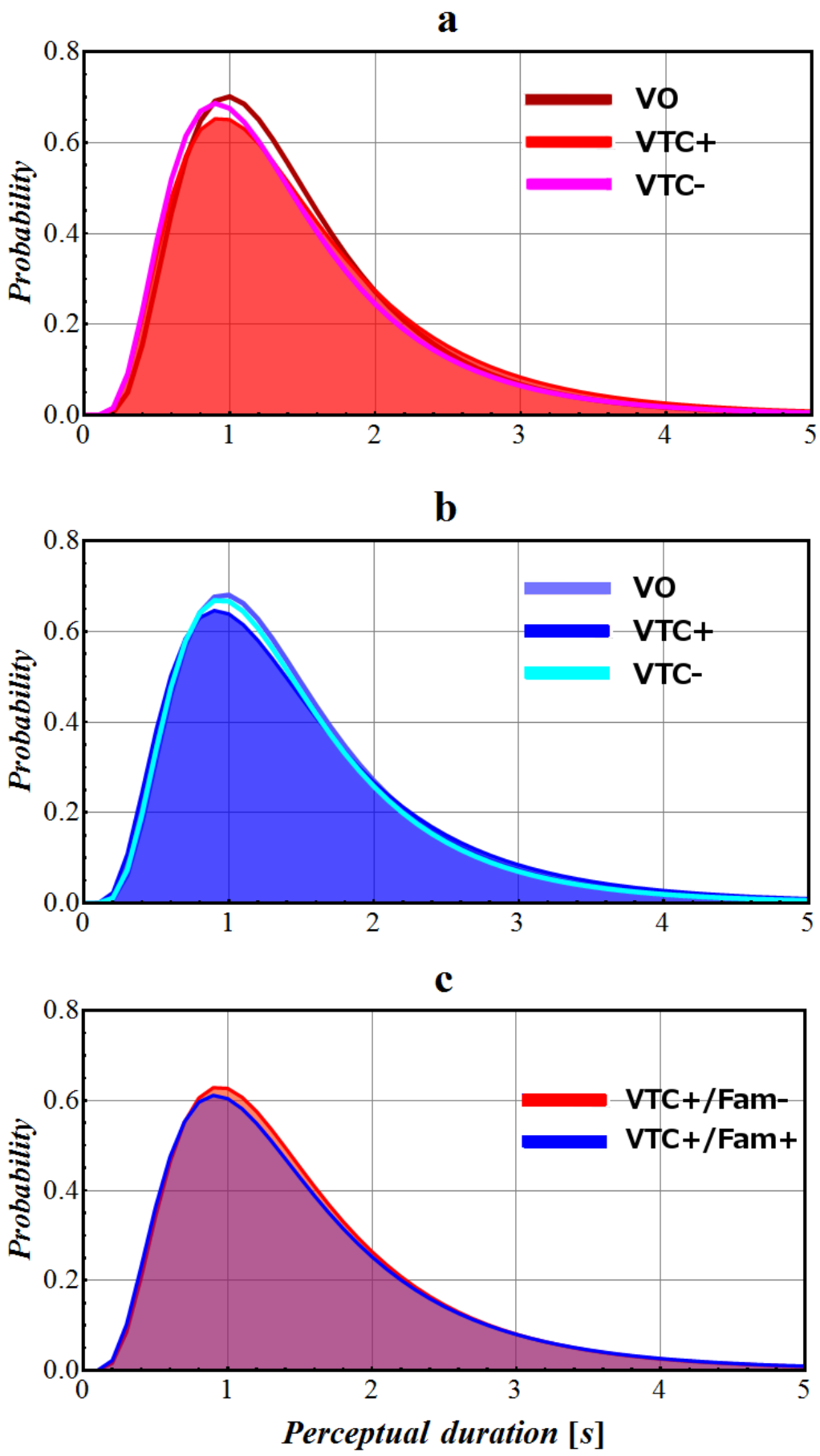

Figure 6. Log-normal fits for the distribution of target durations across the three sensory stimulation conditions for **(a)** unfamiliar objects (reddish) and **(b)** familiar objects (bluish); **(c)** direct comparison between the two visuo-tactile congruent conditions (of *FAM+* and *FAM-*).

Neither object familiarity did not show a main effect ($F(1, 20) = 0.063$, $p = 0.901$) on the target dominance duration, nor did the multisensory stimulation ($F(2, 40) = 1.996$, $p = 0.173$). The two factors did not interact ($F(2, 40) = 1.397$, $p = 0.259$). Still, four paired sample *t*-tests with the group's means yielded multisensory effects close to statistical significance: with unfamiliar objects and during visuo-haptic congruence, the target is perceived longer ($\bar{t}_{VTC+}$ = 1.793 s, SD_{VTC+} = 0.726) than during the visual-only ($\bar{t}_{VO}$ = 1.651 s, SD_{VO} = 0.693, p = 0.059) and visuo-haptically incongruent conditions ($\bar{t}_{VTC-}$ = 1.694 s, SD_{VTC-} = 0.712, p = 0.053). With familiar objects, during visuo-tactile congruence ($\bar{t}_{VTC+}$ = 1.779 s, SD_{VTC-} = 0.807) the percept was also longer stable compared to visuo-tactile incongruence ($\bar{t}_{VTC-}$ = 1.673 s, SD_{VTC-} = 0.665, p = 0.089) and to only visual stimulation ($\bar{t}_{VO}$ = 1.632 s, SD_{VO} = 0.076). See also fig. 7.

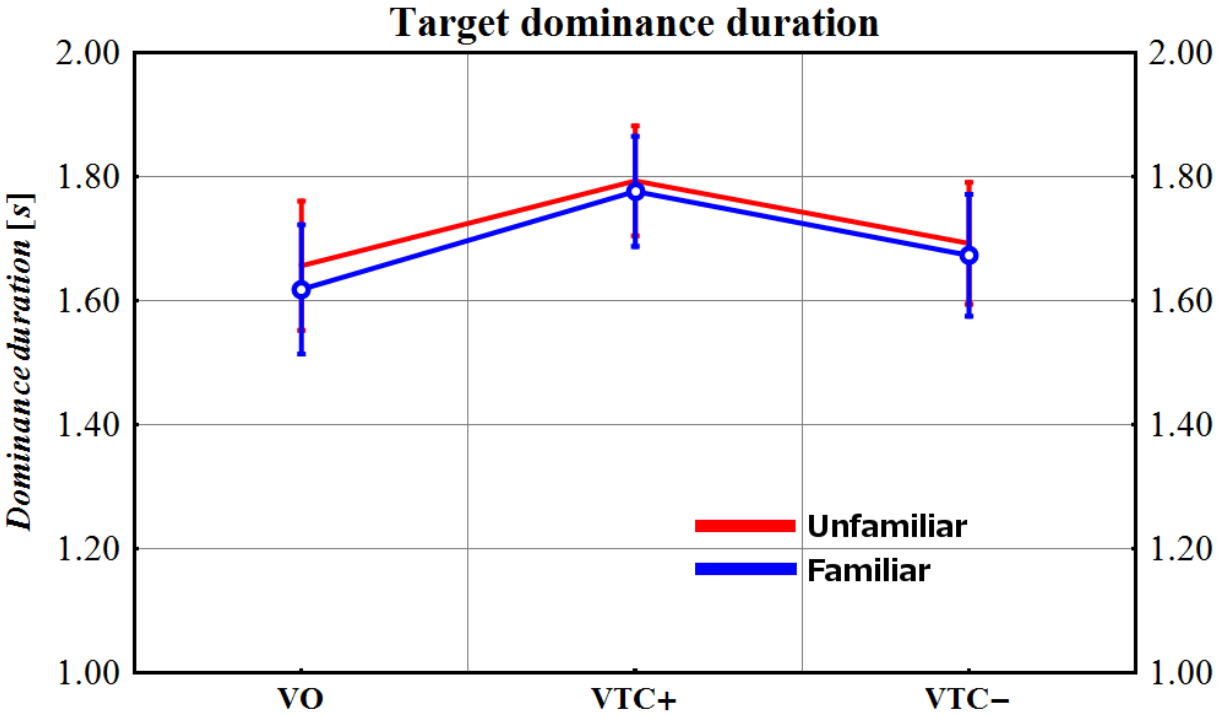

Figure 7. Dominance duration across the whole trial of the familiar or unfamiliar target objects under different multisensory conditions; for both object classes, visuo-haptically congruence showed a close to significance prolonged duration when compared to the visual-only and visuo-haptically incongruent condition; object familiarity did not influence the target dominance; the within-subject error bars denote the 95-% confidence interval.

3.2.2 Target dominance duration and suppression at the trial onset

Dominance duration and the suppression duration of the initial percept serve as additional measures for perceptual stabilization. In case the target was initially selected and maintained for a certain period, then it was *dominant*, otherwise it was *suppressed*. Thus, the dominance duration of the non-target equals the suppression duration of the target, and vice

versa. If the initial percept was the target, then one would expect it to be maintained longer during *VTC+* as compared to *VO* and *VTC-* – and especially during the associated *FAM+* session. Conversely, if the initial percept was the non-target, then one would expect the target to be suppressed shorter during *VTC+* as compared to *VO* and *VTC-* – and especially during the associated *FAM+* session. During *VO* and *VTC-* the target would show dominance duration equal to the non-target. To test this, the phase duration values were square-root-transformed to homogenize the variances and submitted to the repeated-measures 2 x 3 ANOVA (with familiarity and *SSC* as factors). *SSC* yielded a significant main effect ($F(2, 40) = 4.472$, $p = 0.047$, $\eta_p^2 = 0.183$), whereas object familiarity did not ($F(1, 20) = 0.209$, $p = 0.784$). No interaction was detected ($p = 0.761$). Four post-hoc Wilcoxon signed rank tests with *p*-value corrections via the Bonferroni-Holmes procedure revealed multisensory effects (see table 2; full lines in fig. 8). With the same ANOVA procedure the target suppression durations were compared across the two factors. With around 1.70 seconds the target is suppressed for a period approximately equal in length across all multisensory blocks with both familiar and unfamiliar objects (*SSC:* $F(2, 40) = 0.003$, $p = 0.959$; *FAM:* $F(1, 20) = 0.036$, $p = 0.932$; dashed lines in fig. 8).

Table 2. Pairwise comparison with Wilcoxon's signed rank tests (including Bonferroni-Holm corrected *p*-values and Cohen's *d*) demonstrating significant differences between the *SSCs* with respect to the initial target duration $\bar{t}$ (standard deviations in brackets)

FAM	$\bar{t}_{VO}$ ***(σ)***	$\bar{t}_{VTC+}$ ***(σ)***	$\bar{t}_{VTC-}$ ***(σ)***	***W***	***p***	***Cohen's d***
FAM+	1.63 (0.85)	2.07 (1.16)		45	0.045	1.029
		2.07 (1.16)	1.83 (1.20)	57	0.044	1.174
FAM-	1.64 (0.89)	1.92 (0.98)		34	0.019	0.922
		1.92 (0.98)	1.68 (0.78)	50	0.027	0.871

When the initial suppression time of the target is directly compared to the initial dominance time (2 x 3 repeated measures ANOVA with target duration/target suppression and *SSC* as factors), it turns out that both are equal during *VO* and *VTC*-, but different during *VTC*+: there was a significant interaction between the factors with unfamiliar objects ($F(2, 40) = 4.583$, $p = 0.016$, $\eta_p^2 = 0.186$), as well as with unfamiliar objects ($F(2, 40) = 4.592$, $p = 0.016$, $\eta_p^2 = 0.187$). At the trial onset, the target is perceived longer ($\bar{t}_{VTC+/Fam-} = 2.07$ s; $W = 187$, $p = 0.014$ and $\bar{t}_{VTC+/Fam+} = 1.92$ s; $W = 196$, $p = 0.005$) than it is suppressed (duration with both object classes $\bar{t} \approx 1.70$ s). See fig. 8.

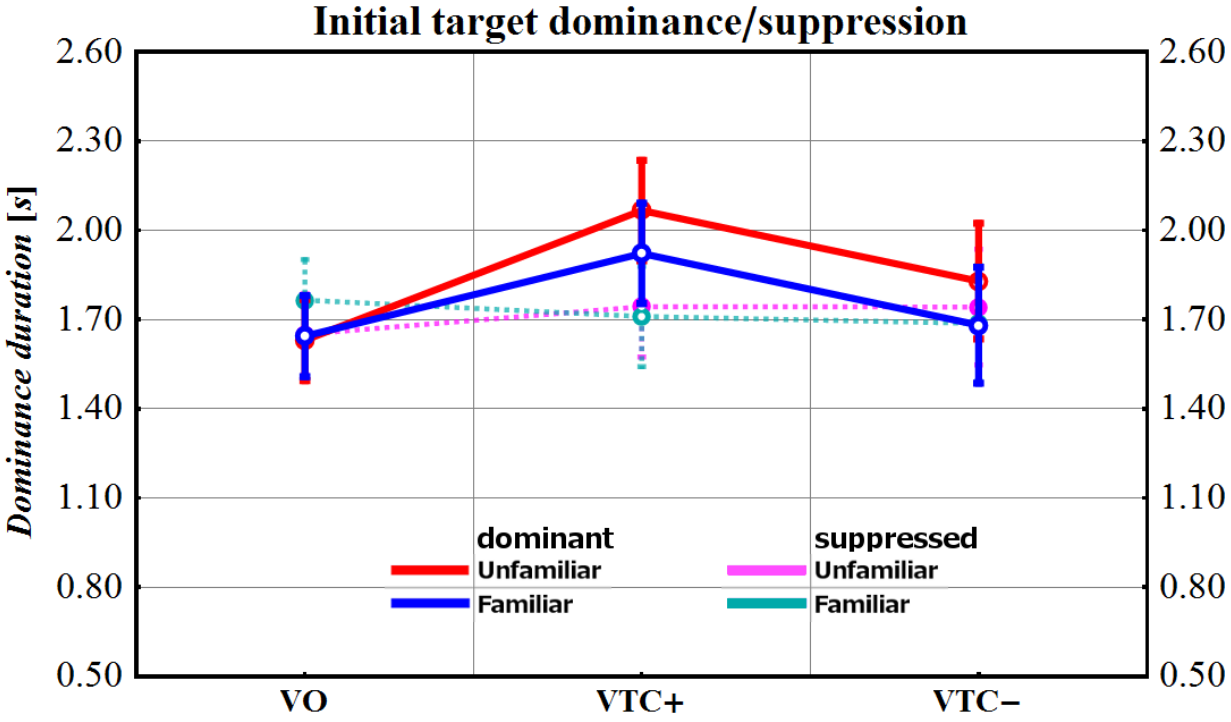

Figure 8. Mean dominance duration of the target as the first percept (solid lines); during visuo-tactile congruency the target was maintained the longest; it was, however, suppressed for a shorter period of time when it was not the first percept (dashed lines); thus, the first percept was maintained longer, if it was the target rather than the non-target; the within-subject error bars denote the 95-% confidence interval.

3.3 Reaction time analysis

As with the advantage in perceptual selection at the beginning of a trial, stimulation in one sensory modality, for example by priming, exerts facilitated response in another sensory modality (‘multisensory redundant target effect‘). Departing from this notion, what was expected here was that reaction times for the target would be shorter in the *VTC*+ block as the visual target was primed by its haptic exploration. In addition, with familiar objects there

should be even a greater *RT* advantage. Only the reaction times recorded during a target hit were pooled for further analysis. For each subject, the reaction time values beyond 2 standard deviations from the mean were treated as outliers and were bootstrapped, in order to compensate for the lost data, as the data sample was limited (~ 20 *RTs* per block per subject).

To test if there was a facilitation of response for familiar objects, a 2 x 3 repeated measures ANOVA was conducted with the square-root-transformed mean *RT* values and with *SSC* and familiarity as factors. Both, the *SSC* ($F(2, 40) = 3.136$, $p = 0.092$) and the familiarity ($F(1, 20) = 2.947$, $p = 0.071$) had no main effect on the reaction time, yet were close to significance. No interaction was detected between the factors ($p = 0.153$). There is however an overall advantage for unfamiliar objects of ca. 100 ms ($W = 699$, $p = 0.035$; see also fig. 9). This advantage was also present, albeit not being significant, in observers‘ mean reaction times which – in addition to target his – included also target misses ($W = 743$, $p = 0.071$).

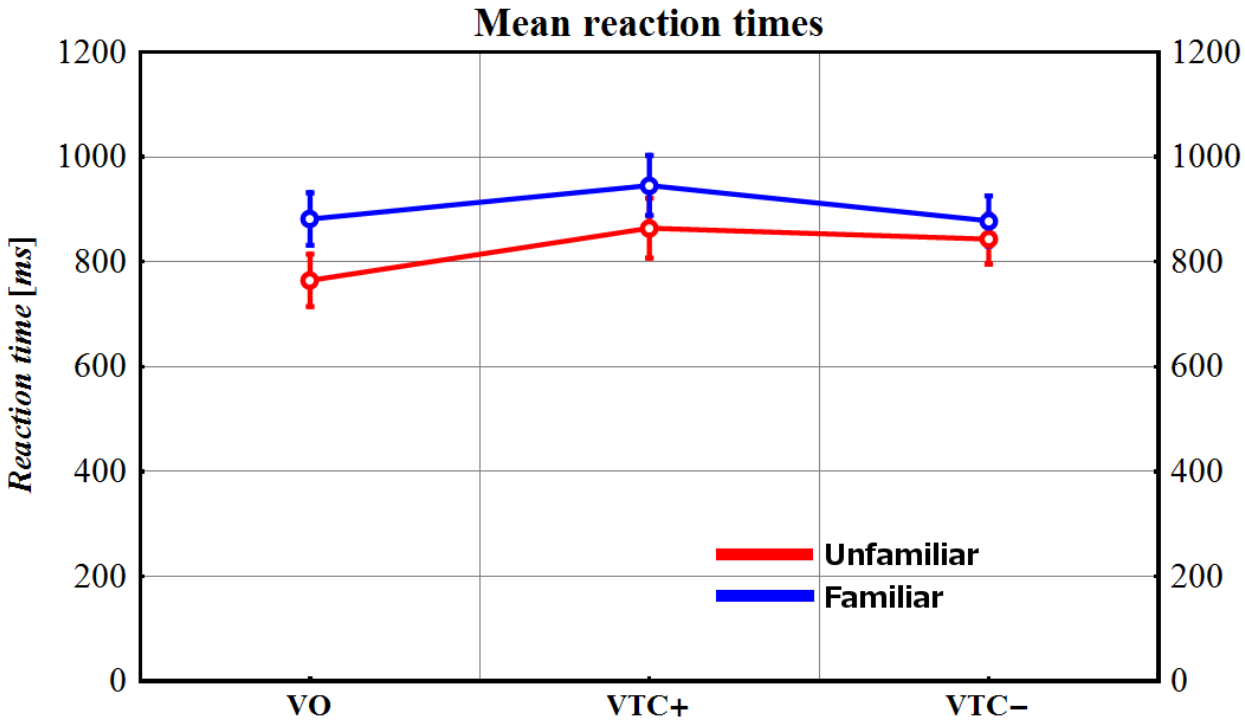

Figure 9. Results for the mean reaction time across the multisensory conditions (*SSC*) and object familiarity; there was no significant difference between the multisensory conditions, yet a small net advantage was detected for unfamiliar objects; the within-subject error bars denote the 95-% confidence interval.

3.4 Frequency of perceptual alternations

Finally, *PSF* was also considered as indicative to a change in the perceptual behavior during binocular rivalry. Within a subject, dropping of the *PSF* from condition to condition suggests that the nervous system decreases perceptual volatility, rendering the perceptual content more stable. During the experiment, each time the observers perceived the colours as altering, they reported it by a key stroke.

Within each trial, the average perceptual switching frequency is defined as the number of switches *N* divided by the total time during which the switches occured:

$$\bar{f} = \frac{N_{Trial}}{t_{Trial}} \text{ [Hz]}$$

For each trial in every *SSC* block in the two sessions, the average frequency was computed, square-root-transformed and analysed with a 2 x 3 repeated measures ANOVA with object familiarity (*FAM+, FAM-*) and sensory stimulus condition (*VO, VTC+, VTC-*) as factors.

Neither the factor familiarity ($F(1, 20) = 0.013$, $p = 0.349$) nor the sensory stimulation condition ($F(2, 40) = 0.916$, $p = 0.963$) had a significant main effect on the frequency of the perceptual alternations. The mean frequency $\bar{f}$ of the *VTC+* block with unfamiliar objects was 0.639 *Hz*, in the case of familiar objects it was 0.644 *Hz* for the (fig. 10).

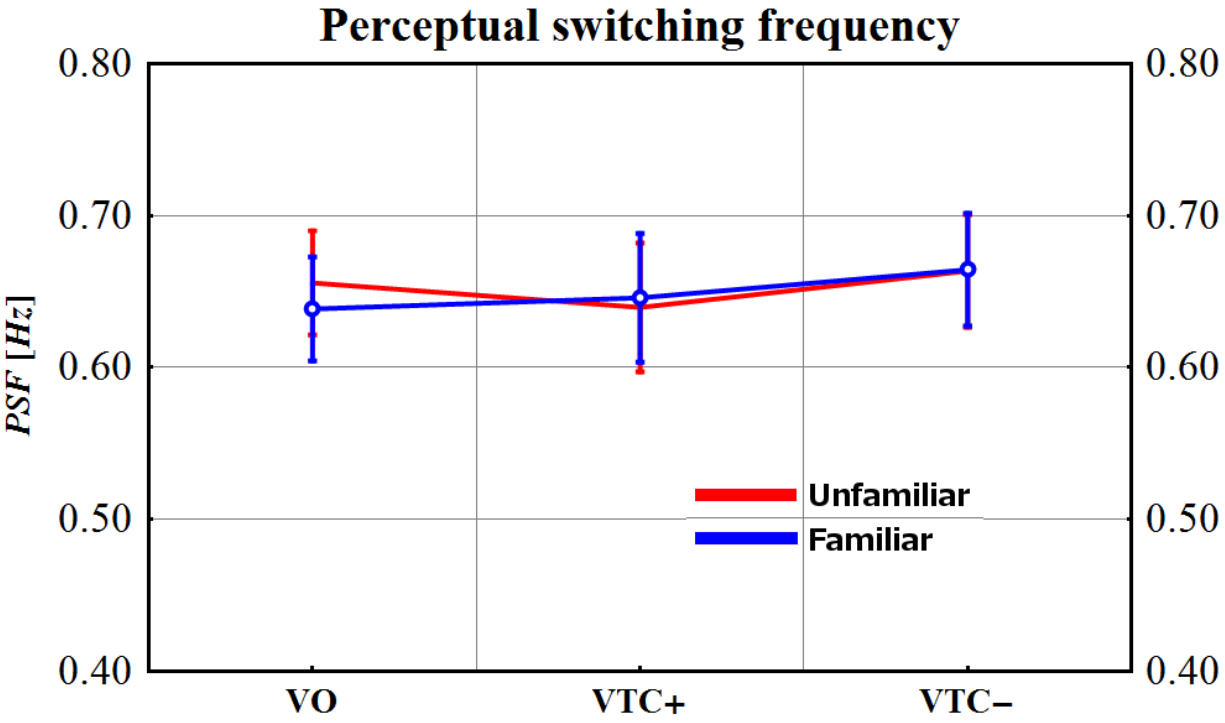

Figure 10. Average perceptual switching frequency as a function of block (*VO, VTC+, VTC-*) and session (red: *FAM-*, blue: *FAM+*). The within-subject error bars denote the 95-% confidence interval (Cousineau, 2005).

To check whether there is a difference in perceptual switching behavior in the course of each trial, data was analysed over the time span of all trials. The trial duration of 20 seconds was binned in 80 time windows (with a bin width of 250 ms). For each bin, the number of perceptual alternations of all trials of a particular block was summed up across all subjects and divided by the total number of switches of all bins belonging to that block. This computation resulted in a data set representing the temporal evolution of the mean probability for a perceptual switch at each time point t_i of a trial (figure 11 a). The temporal course of the mean probability for a perceptual switch in a trial can be modeled by a modified logistic function with four parameters (*a, b, c, d*):

$$p_{switc\,h}(t) = -a * (\frac{b - e^{ct}}{b + e^{dt}})$$

Parameter *a* scales the overall perceptual stability: the greater its value, the less stable is visual perception and consequently the more likely is a perceptual alternation at each bin; parameter *b* describes the response inertia towards the maximum switching frequency: greater values indicate later response onset (high values entail later response onset and maximum alternation frequency in the trial); parameters *c* and *d* jointly, yet antagonistically, determine the temporal decay of the switching probability: while high values for *c* gradually raise the switching probability during the trial, high values for *d* counteract this effect. For the *VTC+* block with unfamiliar objects, the non-linear curve fitting procedures in *Mathematica*® describe the temporal development of the switching probability with the parameters a = 0.0025, b = 1.8987, c = 0.2781 and d = 0.2806, resulting in the least-square fitted function (depicted in fig. 11 b, dark red):

$$p_{VTC+;Fam-}(t) = -0.0025 * (\frac{1.8987 - e^{0.2781\ t}}{1.8987 + e^{0.2806\ t}})$$

After fitting the probabilities of the remaining five conditions, it became evident that, quantitatively, the probabilities have very similar fitting parameters (see table 5, appendix). This was also reflected by the approximately same size of the area under the curve (AUC), which was 0.1616 for the *VTC+* unfamiliar condition, compared to 0.1643 for the same block of the unfamiliar session (table 5, appendix).

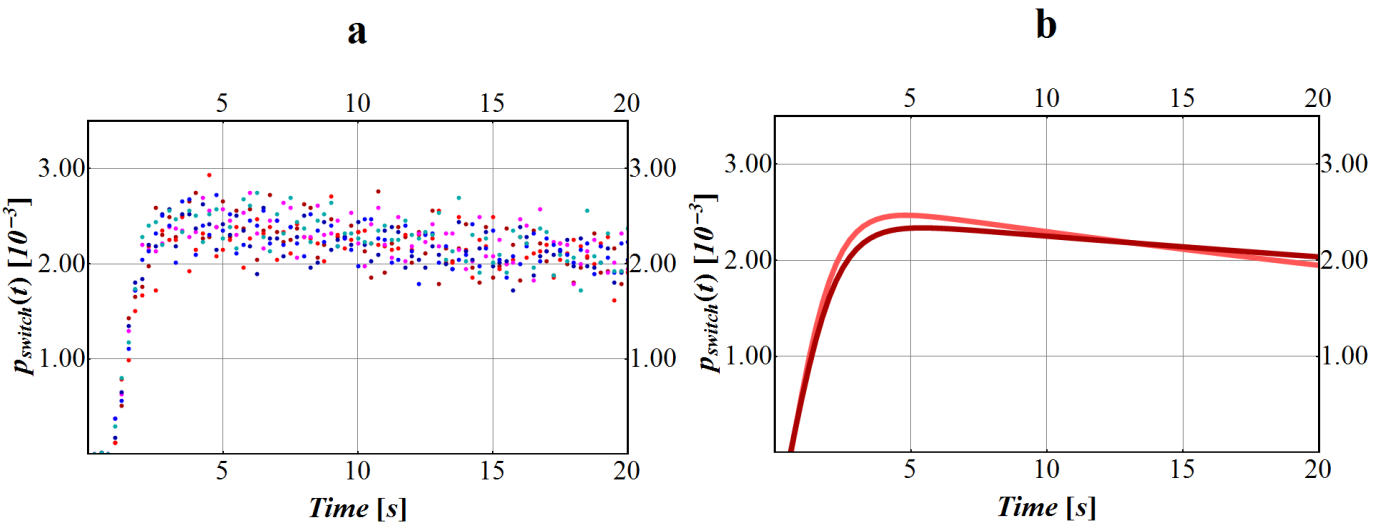

Figure 11. (a) Mean probabilities for a perceptual switch during the course of an average trial of each block, averaged across all subjects (reddish colours depict the *FAM-* session; dark red: *VO*, light red: *VTC+,* magenta: *VTC-*; bluish colours depict the *FAM+* session; dark blue: *VO*, light blue: *VTC+*, cyan: *VTC-*); **(b)** Modified logistic fit functions of the mean probability data for the *VO/FAM-* (light red) and *VTC+/FAM-* (dark red).

To test whether this small differences between the conditions were statistically significant, a residual analysis was performed (Andrews, 1971). The function $p_{VTC+/FAM-}(t)$ was selected as the reference model (RM). The squared residuals from the RM and its original data set were compared against the squared residuals of two other blocks which would indicate effects of the two independent variables (*VO/FAM-* block for multisensory effects and *VTC+/FAM+* block for familiarity effects). In the next step, two Wilcoxon's signed rank tests for paired samples were conducted: since the data of each block was ordered in 80 bins, the 80 residuals of the RM were compared against the 80 residuals of either of the two blocks which would indicate some effects. The first test compared the residuals between the RM and its original data (*VTC+/FAM-*) against the residuals from the RM and the original data set of *VO/FAM-.* Here, the test statistic $W = 1660$ with $p = 0.85$ indicated no significant difference between the the visual-only and the visuo-haptically congruent conditions. In the second test, the difference between the RM and its original data was tested against the residuals from the RM and the original data set of *VTC+/FAM+.* Also here was excluded that the data sets were different; no difference was found between the familiarity conditions ($W = 1379$, $p = 0.249$).

4. Discussion

This study aimed to answer the question whether multistable visual perception can be modulated stronger by a tactile input which invokes conceptual and action-related associatons than by a tactile input which does not. Specifically, it sought to test whether during binocular rivalry additional internal signals connected to previous knowledge and semantic information about objects (in addition to the shape information alone) would interact with visuo-haptic processing. Such prior knowledge and semantic information was introduced by familiar objects, which were contrasted with novel non-sense objects to which the observers were not exposed to before. During the course of the visual competition, observers reported the colour of the dominant percept to avoid the bias of merely reporting the visual percept that happens to be congruent with the haptic object. In the visuo-haptic blocks, the colour was associated with the visual objects during binocular competition which, in turn, were either congruent or incongruent with the haptic objects.

4.1 Perceptual selection at the trial onset

In this study, one question also addressed was to what extent would the haptic priming affect the initial perceptual selection. Previous investigations indicate that attending to cued visual stimuli before the onset of binocular rivalry can bias which percept will be dominant at the beginning of a trial (Chong & Blake, 2006; Mitchell et al., 2004; Stanley et al., 2011). In the context of multistability, however, such first-percept biases were not yet established (Blake, Sobel & James, 2004; Lunghi & Morrone, 2012). In contrast, a clear bias in the onset rivalry was demonstrated here: the touched object was more likely to be selected and was perceived for a longer period of time. A sort of implicit perceptual memory is one plausible explanation for the inconsistency of results (Leopold at al., 2002): it is known that intermittent removal of the competitive stimuli during a trial leads to the stabilization of one particular percept. In the previous multisensory studies only two rivalling alternatives were used in all the trials, so that the inter-trial-interval lead to the stabilization of one particular percept – and thus rendering the non-visual stimulus not capable of biasing the onset rivalry towards the other percept. But here, 10 different object pairs were utilized which never were shown in two successive trials, ruling out the intermittend presentation of the same stimuli, from which just one particular percept becomes stabilized. Furthermore, while previously only two colours were used to report the dominant percept, two out of three colours were randomly combined here in every trial, making it less likely that one particular colour would

be perceived as more stable. Both factors show an advantage of the current paradigm for the investigation of visuo-haptic priming effects at the onset of binocular rivalry.

Despite the higher likelihood for the initial bias, no difference was detected between the cross-modal priming with familiar and unfamiliar objects, suggesting that non-perceptual object features and motor-cognitive factors did not influence initial perceptual selection. object features.

4.2 Perceptual dominance durations

The most reliable indicator of perceptual stabilization and change during binocular rivalry is the perceived duration of a particular percept. The hypothesized object-intrinsic information and affordances did not affect this measure variable – neither at the trial onset nor in the course of the trials. Still, it was demonstrated that the visual image was consistently perceived longer when it was congruent to the haptically explored object – regardless of the object familiarity. This observation dovetails with several preceding findings (Conrad et al., 2010; Conrad et al., 2013; van Ee et al., 2009; Chen, Yeh & Spence, 2011; Lunghi & Alais, 2013; Lunghi, Binda & Morrone, 2010; Lunghi & Morrone, 2012; Maruya, Yang & Blake, 2007; Zhou et al., 2010). It was evident to a smaller (non-significant) degree for the durations recorded throughout the trials, and to a stronger (significant) degree at the trial onset. This discrepancy might arise from the possibility that the motor response behavior (indicating a perceptual switch) could converge to a constant phase oscillation during the course of the trial instead of remaining a non-constant, stochastic oscillation, as is typical for binocular rivalry (Fox & Herrmann, 1967; Lehky, 1995). This would gradually equalize the target's dominance phase with its suppression time throughout a trial (or conversely, with the dominance phase of the non-target). Actually, this notion is supported by the fact that the phase duration of the target was longer when the target was the initial percept ($\bar{t}_{VTC+/FAM-;\ in}$ = 2.07 s and $\bar{t}_{VTC+/FAM+;\ in}$ = 1.92), compared to the overall average across the whole trial ($\bar{t}_{VTC+/FAM-;\ tr}$ = 1.79 s and $\bar{t}_{VTC+/FAM+;\ tr}$ = 1.78 s). Still, it cannot be completely excluded that the dual task setting could compromise the accurate perceptual reports which might lead to stereotypical responses during the trial. But a varying response behavior would reflect different perceptual states more reliably, because it is improbable during binocular rivalry that several successive perceptual durations converge to similar values, especially when there is bimodal stimulus congruency.

Surprisingly and contrary to previous studies, target suppression phases did not become shorter when the observers faced visuo-tactile congruent stimuli. Usually, the target's perceptual dominance is prolonged and its suppression is shortened (see the

abovementioned studies in this section). Here, however, the target did not become boosted into visual awareness despite congruent haptic signals.

Visuo-haptic congruency nevertheless prolonged the target phase duration throughout the trial, in spite of not reaching statistical significance. Therefore, a convincing and statistically significant visuo-haptic priming and more stable visuo-haptic percept formation in this paradigm seems to be at the very beginning of the trial. Notwithstanding the above said, objects with functional properties did not show an advantage over novel, non-sense objects with respect to the dominance duration of the target at the trial onset.

4.3 Reaction time

Contrary to the predicton, intersensory facilitation and ‘redundant target effect‘ were not observed here. Neither did the sensory stimulation conditions nor object familiarity facilitate responses. Quite the opposite was observed: during both bisensory conditions (*VTC+* and *VTC-*), reaction time was non-significantly longer, than during the unimodal condition. This is at odds with numerous studies from the past which showed that observers responded faster to simultaneous haptic and visual stimulation than to unimodal stimulation (for example Diederich & Colonius, 2004; Forster et al., 2002; Laurienti et al., 2003). Also, during semantic and visuo-tactile congruency reaction time was even longer than during non-semantic visuo-tactile congruency. This does also not square with previous work, which demonstrated that semantically congruent multisensory stimuli lowered reaction time (Laurienti et al., 2003). There was even an overall significant disadvantage with familiar objects – to a greater extent in the unisensory condition, and to a lesser extent in the bisensory conditons. Putting the redundant target effect aside, this observation suggests that non-perceptual, object-intrinsic features influence reaction time performance. It is possible that concepts associated with the dichoptically presented visual objects compete for selection slowing down the response in the 2AFC. As conceptual knowledge was not associated to the novel objects, the reaction time during the unfamiliar session was shorter. The fact that no intersensory facilitation was found raises the question whether reaction time is a useful variable for measuring multisensory effects during binocular competition. Thus, a final interpretation of the observed reaction times remains inconclusive. object features.

4.4 Frequency of perceptual alternations

The mentioned was expected that the perceptual switching rate would decrease during the multisensory blocks (*VTC+* and *VTC-*) compared to the unisensory block (*VO*).

Two possible reasons would explain this frequency decrement. On the one side, there is a dual task during *VTC+* and *VTC-* because the observers were primarily required to report the perceptual switches while, simultaneously, exploring an object haptically. One might assume that both tasks share the attentional load in such a way that increasing the focus on the one task comes at the expense of the focus on the other task. This possibility was ruled out by the inclusion of the *VTC-* block in the design as a manipulation check so that the hypothesized effect of visuo-tactile congruency could be detected during *VTC+*. On the other side, the expected effect could – as hypothesized – cause the slowdown of perceptual alternation due to perceptual congruency and stabilization during *VTC+*.

As it is assumed that, with respect to the attentional load, there is no difference between *VTC+* and *VTC-* (for both object types), one would have expected a difference between the conditions due to multisensory congruency and especially due to object-intrinsic factors (object familiarity). However, no such difference has been observed: visuo-tactile congruency did not lower the perceptual alternation rate. What is more, the PSF during *VTC+* did not differ from the PSF recorded during the *VO* block either, suggesting no substantial attentional load due to haptic object exploration in the multisensory blocks. This seems realistic, as the observers explored the objects by highly stereotypical exploratory procedures (Lederman & Klatzky, 1987) which may not demand substantial attentional load because they are automatic in nature. Furthermore, as was previously suggested (Helbig & Ernst, 2008; Lunghi & Morrone, 2012), inter-sensory cue weighting for the final multisensory percept estimation takes place independently of modality-specific attention, rendering the process pre-attentive and automatic. Taking this into account, it is still possible that perceptual rivalry can be biased by haptic input – especially at the onset of the rivalry (Stanley et al., 2011) – while there is no change in the overall perceptual alternation rate across the conditions. The results suggest that frequency of perceptual alternations is only a limited indicator of perceptual behavior during binocular rivalry and that it should serve as a supplement to the findings of other indicators.

4.5 Strengths, limits and possible methodological improvements in the study

The present work sought to investigate multisensory semantic effects for which no evidence could be found here. Yet, it might be that this effect would be detectable when supplemented with more sensitive methods. The present experimental setting had some advantages, but also some limits.

The main strength of this study was the repeated-measure design with both sessions and with no between-subject factors. It takes into account the individual variability of responses and yields a high statistical power. In addition, it also shows for the first time that

binocular rivalry can be studied with naturalistic haptic stimuli, especially in combination with minimized spatial visuo-haptic offset which allows a more ecologically valid assessment of haptic modulations of binocular rivalry. Moreover, the study used a larger set of stimuli which never appeared twice successively as well as three colours for reporting the dominant percept (Leopold et al., 2002). This effectively eliminanted implicit perceptual memory and perceptual stabilization and revealed that there can actually be a first percept bias in binocular rivalry.

Still, a few improvements could increase the experimental sensitivity and answer the questions investigated here conclusively. Firstly, one could counteract piecemeal rivalry issues more reliably, if the participants report the exclusive dominance of a percept. The perceptual dominance data would be much more reliable, if the observers would report the exclusive perception of the non-congruent object (instead of the congruent) while at the same time exploring an object haptically. This is desired because by attending to just one of the two perceptual alternatives one can assume that motor responses would less likely converge to a constant phase oscillation. In this approach, colour coding would be useful as well. Secondly, heterochromatic flicker photometry can equalize the perceived dominance of the colour codes in every subject. This would minimize the remaining noise in the data arising from objectively (*Lab*-space) calibrated equiluminant colours, from which it cannot be excluded that the subjects perceive them slightly different. Alternatively, lowering the target's stimulus strength makes it less dominant and less likely to be selected; by biasing binocular rialry via haptic stimulation, one can however measure the gain of the enhanced probability for selection and phase duration (Levelt's second proposition; Levelt, 1966). And lastly, continuous flash suppression (CFS) and signal detection theory might be very suitable for revealing the 'volleys of conceptual processing', because it is known that stimuli with semantic context can be boosted into conscious awareness by CFS (Lupyan & Ward, 2013; Zhou et al., 2010).

4.6 General discussion

Contrary to the proposed hypotheses, object familiarity did not show an advantage in the multisensory task-setting. No interaction was observed between the rivalry and the types of objects, that is the object intrinsic properties with or without inherent meaning. This stands at odds with the expectations formulated here and predictions of others (Hohwy, Roepstorff & Friston, 2008). Still, binocular rivalry was modulated by different unimodal/bimodal input combinations. When the visual and the haptic stimuli were congruent, there was an advantage in respect to some important, but not all quantities which are indicative of a change of the multistable perception. At least with regard to the multisensory effects, this

finding is in line with previous behavioral studies on multisensory processing exploiting multistability and suggests that the experiment had the power to detect changes across conditions. It also showed that visual processing can strongly be modulated by haptic input, underpinning the notion that the nervous system's operations are essentially multisensory.

The present findings can be interpreted in favour of the idea that the nervous system takes no advantage of object familiarity and action-related object knowledge for perceptual fusion, but that it fuses visual and tactile signals rather mandatory, automatically and implicitly, without necessarily utilizing conceptual object representations. This suggests that the stronger predictive codes for familiar objects from the fronto-parietal top-down networks which feed into LOC and visuo-tactile integration sites do not stabilize more efficiently visuo-haptic perception. This implies that higher neural activity does not necessarily entail more efficient disambiguition of visual signals during multisensory integration. Instead, it seems that the driving force behind visuo-haptic integration for both object classes operates comparably strong. This seems plausible if one takes into account that during interaction with the environment in peripersonal space, for example with solid objects, the interpretative framework (the perceptual ambiguity) for shapes, or for haptic space in general, is very narrow. The physical shape and the structural descriptions of an object are invariant and unambiguous when compared to visual and auditory stimuli which often have several interpretations. Therefore, the resolution of bistable vision can rely on haptic information alone without involving object-intrinsic features and action-related knowledge to give rise to visuo-haptic integration. This might especially hold because vision and haptics recover the physical topology and solve categorization tasks equally well (Gaißert, Bülthoff & Wallraven, 2011).

Still, it is possible that the effect is present, but cannot be detected within a trial of 20 seconds. In fact, there is evidence that the automatic activation of motor codes by affordances of familiar objects is generated very rapidly and that the influence is short-lived (Markis, Hadar & Yarrow, 2011). This implies that the optimal time-frame for observation of the effect would be at the beginning of a trial. But also here, however, the effect was not observed. Thus, failing to reject the null hypothesis conclusively is not justified, especially since the reaction time was longer during the *FAM+* session suggesting that indeed some conceptual processing was taking place. Also, it is possible that the approach had not sufficient power to detect such an effect and that more sensitive methods could prove more beneficial in this respect.

5. Conclusio

In the present study, no evidence was found for ‘volleys of conceptual processing‘to resolve visual ambiguity during visuo-haptic binocular rivalry. A preliminary interpretation of the findings suggests that the nervous system relies on sparse representations during multisensory processing without necessarily engaging conceptual processing when facing known objects. Since haptic space is invariant and its interpretative frame is very limited, ambiguities in the visual system can rely on structural descriptions alone, rendering conceptual information unnecessary in visuo-haptic integration. It supplements however the literature on multistable perception with the fact, that functional haptic objects reliably modulate visual perception during binocular rivalry. But, in order to contribute conclusively to the literature on multisensory binocular rivalry with semantic features as a potential factor in crossmodal binding, a replication of the study is useful. It is likely that more sensitive methods, like CSF and signal detection theory can detect semantic effects in visuo-tactile processing. Still, the overall binocular rivalry data buttresses the view, that the brain is essentially a multisensory network with many hubs for measurable intersensory interaction.

6. References

Amedi A., Jacobson G., Hendler T., Malach R. & Zohary E. (2002). Convergence of visual and tactile shape processing in the human lateral occipital complex. *Cerebral Cortex, 12,* 1202-1212.

Alais D. & Melcher D. (2007). Strength and coherence of binocular rivalry depends on shared stimulus complexity. *Vision Research, 47,* 269-279.

Andrews D. F. (1971). Significance tests based on residuals. *Biometrika, 58 (1)*, 139-148.

Baker D. H. & Graf E. W. (2009). Natural images dominate in binocular rivalry. *Proceedings of the National Academy of Sciences of United States of America, 106 (13),* 5436-5441.

Bargary G. & Mitchell K. J. (2008). Synaesthesia and cortical connectivity. *Trends in neurosciences, 31(7),* 335-342.

Beauchamp M. S. (2005). See me, hear me, touch me: Multisensory integration in lateral occipital-temporal cortex. *Current Opinion in Neurobiology, 15,* 1-9.

Binder J. R. & Desai R. H. (2011). The neurobiology of semantic memory. *Trends in Cognitive Sciences, 15 (11),* 527-536.

Blake R. (2001). A primer on binocular rivalry, including current controversies. *Brain and Mind 2*,. 5-38.

Blake R. & Logothetis N. K. (2002). Visual competition. *Nature Reviews Neuroscience, 3 (1),* 13-21.

Blake R., Sobel K. V. & James T. W. (2004). Neural synergy between kinetic vision and touch. *Psychological Science, 15 (6)*, 397-402.

Botvinick M. & Cohen J. (1998). Rubber hands ‘feel‘ touch that eyes see. *Nature, 391 (6669),* 756.

Brainard D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10,* 433-436.

Bruckner R. L., Koutsaal W., Schacter D. L. & Rosen B. R. (2000). Functional MRI evidence for a role of frontal and inferior temporal cortex in amodal components of priming. *Brain, 123,* 620-640.

Brugger P. & Brugger S. (1993). The easter bunny in october: Is it disguised as a duck? *Perceptual and Motor Skills, 76,* 577-578.

Bruner J. (1957). On perceptual readiness. *Psychology Review, 64 (2),* 123-152.

Calvert G. A., Bullmore E. T., Brammer M. J., Campbell R., Williams S. C. R., McGuire P. K., Woodruff P. W. R., Iversen S. D. & David A. S. (1997). Activation of auditory cortex during silent lipreading. *Science, 276 (5312),* 593-596.

Capitani E., Laiacona M., Mahon B. & Caramazza A. (2003). What are the facts of semantic category-specific deficits? A critical review of the clinical evidence. *Cognitive Neuropsychology, 20 (3/4/5/6),* 213-261.

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

Chao L. L. & Martin A. (2000). Representation of manipulable man-made objects in the dorsal stream. *NeuroImage, 12,* 478-484.

Chemero A. (2003). An outline of a theory of affordances. *Ecological Psychology, 15 (2),* 181-195.

Chen Y.-C., Yeh S.-L. & Spence C. (2011). Crossmodal constraints on human perceptual awareness: Auditory semantic modulation of binocular rivalry. *Frontiers in Psychology, 2,* 1-13.

Chong S. C. & Blake R. (2006). Exogenous attention and endogenous attention influence initial dominance in binocular rivalry*. Vision Research, 46,* 1794-1803.

Chong S. C., Tadin D. & Blake R (2005). Endogenous attention prolongs dominance durations in binocular rivalry. *Journal of Vision, 5,* 1004-1012.

Conrad V., Bartels A., Kleiner M. & Noppeney U. (2010). Audiovisual interactions in binocular rivalry. *Journal of Vision, 10 (10),* 1-15.

Conrad V., Kleiner M., Bartels A., O'Brien Hartcher J., Bülthoff H. H. & Noppeney U. (2013). Naturalistic stimulus structure determines the integration of audiovisual looming signals in binocular rivalry. *PLOS ONE, 8 (8),* 1-8.

Cousineau D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorial in Quantitative Methods in Psychology. 1 (1),* 42-45.

De Gelder B. & Bertelson P. (2003). Multisensory integration, perception and ecological validity. *Trends in Cognitive Sciences, 7 (10),* 460-467.

Deshpande G., Hu X., Lacey S., Stilla R. & Sathian K. (2010). Object familiarity modulates effective connectivity during haptic shape perception. *NeuroImage, 49,* 1991-2000.

De Weert C. M. M. & Levelt W. J. M. (1976). Comparison of normal and dichoptic colour mixing. *Vision Research, 16,* 59-70.

Diederich A. & Colonius H. (2004). Bimodal and trimodal multisensory enhancement: Effects of stimulus onset and intensity on reaction time. *Perception & Psychophysics, 66 (8),* 1388-1404.

Easton R. D., Greene A. J. & Srinivas K. (1997). Transfer between vision and haptics: Memory for 2-D patterns and 3-D objects*. Psychonomic Bulletin & Review, 4 (3),* 403-410.

Engel S. A., Glover G. H. & Wandell B. A. (1997). Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cerebral Cortex, 7,* 181-192.

Engel A. K., Fries P. & Singer W. (2001). Dynamic predictions: Oscillations and synchrony in top-down processing. *Nature Reviews Neuroscience, 2 (10),* 704-716.

Ernst M. O. & Banks M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature, 415 (6870),* 429-433.

Ernst M. O. & Banks M. S., Bülthoff H. H. (2000). Touch can change visual slant perception. *Nature Neuroscience, 3 (1),* 69-73.

Fang F., Kersten D., Murray S. O. (2008). Perceptual grouping and inverser fMRI activity patterns in human visual cortex. *Journal of Vision, 8 (7): 2*, 2-9.

Forster B., Cavina-Pratesi C., Aglioti S. M. & Berlucchi G. (2002). Redundant target effect and intersensory facilitation from visual-tactile interactions in simple reaction time. *Experimental Brain Research, 143,* 480-487.

Forti S. & Humphreys G. W. (2005). Cross-modal visuo-tactile matching in a patient with a semantic disorder. *Neuropschologia, 43,* 1568-1579.

Fox R. & Herrmann J. (1967). Stochastic properties of binocular rivalry alternations. *Perception & Psychophysics, Vol. 2 (9),* 432-436.

Freeman W. J. (1975). *Mass action in the nervous system*. New York: Academic Press, Inc.

Friston K. & Kiebel S. (2009). Predictive coding under the free-energy principle. *Philosophical Transactions oft he Royal Society, 364,* 1211-1221.

Gaißert N. & Bülthoff H. H., Wallraven C. (2011). Similarity and categorization: From vision to touch. *Acta Psychologica, 138,* 219-230.

Ghazanfar A. A. & Schroeder C. E. (2006). Is neocortex essentially multisensory? *Trends in Cognitive Sciences, 10 (6),* 278-285.

Gibson J. J. (1966). *The senses considered as perceptual systems.* Boston: Houghton Mifflin.

Gibson J. J. (1979). *The ecological approach to visual perception.* Boston: Houghton Mifflin.

Goldberg R. F., Perfetti C. A. & Schneider W. (2006). Perceptual knowledge retrieval activates sensory brain regions. *The Journal of Cognitive Neuroscience, 26 (18),* 4917-4921.

Grèzes J., Tucker M., Armony J., Ellis R. & Passingham R. E. (2003). Objects automatically potentiate actoin: An fMRI study of implicit processing. *European Journal of Neuroscience, 17,* 2735-2740.

Hebb, D. O. (1949). *The organization of behavior: A neuropsychological theory*. New York: Wiley.

Helbig H. B. & Ernst M. O. (2007). Optimal integration of shape information from vision and touch. *Experimental Brain Research, 179,* 595-606.

Helbig H. B. & Ernst M. O. (2008). Visual-haptic cue weighting is independent of modality-specific attention. *Journal of Vision, 8 (1): 21*, 1-16.

Held R. (1965). Plasticity in sensory-motor systems. *Scientific American*, *213 (5),* 84-94.

Helmholtz, H. (1867). *Handbuch der physiologischen Optik*. Leipzig: Leopold Voss.

Hoffman P., Pobric G., Drakesmith M. & Ralph M. A. R. (2012). Posterior middle temporal gyrus is involved in verbal and non-verbal semantic cognition: Evidence from rTMS. *Aphasiology, 26 (9),* 1119-1130.

Hohwy J., Roepstorff A. & Friston K. (2008). Predictive coding explains binocular rivalry: An epistemological review. *Cognition, 108,* 687-701.

Holst E. v., Mittelstaedt H. (1950). Das Reafferenzprinzip. *Die Naturwissenschaften, 20,* 464-476.

Hsiao J.-Y., Chen Y.-C., Spence C. & Yeh S.-L. (2012). Assessing the effects of audiovisual semantic congruency on the perception of a bistable figure. *Consciousness and Cognition, 21,* 775-787.

Hu B. & Knill D. C. (2010). Kinesthetic information disambiguates visual motion signals. *Current Biology, 20 (10),* 436-437.

James T. W., Humphrey G. K., Gati, J. S., Servos P., Menon R. S. & Goodale M. A. (2002). Haptic study of three-dimensional objects activates extrastriate visual areas. *Neuropsychologia, 40,* 1706-1714.

Kalman R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering, 82 (D),* 35-45.

Kellenbach M., Brett M. & Patterson K. (2003). Actions speak louder than functions: The importance of manipulability and action in tool representation. *Journal of Cognitive Neuroscience, 15 (1),* 30-46.

Kerzel D. (2001). Visual short-term memory is influenced by haptic perception. *Journal of Experimental Psychology: Learning, Memory and Cognition, 27 (4),* 1101-1109.

Klatzky R. L., Lederman S. J. & Metzger V. A. (1985). Identifying objects by touch: An „expert system". *Perception & Psychophysics 37 (4)*, 299-302.

Kovacs I., Papathomas T. V., Yang M. & Fehér Á. (1996). When the brain changes its mind: Interocular grouping during binocular rivalry. *Proceedings of the National Academy of Sciences oft he United States of America, 93,* 15508-15511.

Kunde W. & Kiesel A. (2006). See what you've done! Active touch affects the number of perceived visual objects. *Psychonomic Bulletin & Review, 13 (2),* 304-309.

Lacey S. & Campbell C. (2007). Mental representation in visual/haptic crossmodal memory: Evidence from interference effects. *The Quarterly Journal of Experimental Psychology, 59 (2),* 361-376.

Lacey S., Campbell C. & Sathian K. (2007). Vision and touch: Multiple or multisensory representations of objects? *Perception, 36,* 1513-1521.

Lacey S., Pappas M., Kreps A., Lee K. & Sathian K. (2009). Perceptual learning of view-indepdendence in visuo-haptic object representations. *Experimental Brain Research, 198 (2-3),* 329-337.

Lacey S., Peters A. & Sathian K. (2007). Cross-modal object recognition is viewpoint-independent. *PLoS One, 2 (9),* e890.

Lacey S. & Sathian K. (2012). Representation of object for in vision and touch. In M. M. Murray, M. T. Wallace (Eds.), *The neural bases of multisensory processes* (pp. 179-187). Boca Raton: CRC Press, Taylor & Francis Group.

Lacey S., Tal N., Amedi A. & Sathian K. (2009). A putative model of multisensory object representation. *Brain Topography, 21(3-4),* 269-274.

Laurienti P. J., Kraft R. A., Maldjian J. A., Burdette J. H. & Wallace M. T. (2003). Semantic congruence is a critical factor in multisensory behavior performance. *Experimental Brain Research, 158,* 405-414.

Lederman S. J. & Klatzky S. J. (1987). Hand movements: A window into haptic object recognition. *Cognitive Psychology, 19,* 342-368.

Lehky S. R. (1995). Binocular rivalry is not chaotic. *Proceedings of the Royal Society, 259,* 71-76.

Leopold D. A. & Logothetis N. K. (1999). Multistable phenomena: Changing views in perception. *Trends in Cognitive Sciences, 3 (7),* 254-264.

Leopold D. A., Wilke M., Maier A. & Logosthetis N. K. (2002). Stable perception of visually ambiguous patterns. *Nature Reviews Neuroscience, 5 (6),* 605-609.

Levelt W. J. M. (1965). Note on the distribution of dominance times in binocular rivalry. *British Journal of Psychology, 58 (1),* 143-145.

Levelt W. J. M. (1966). The alternation process in binocular rivalry. *British Journal of Psychology, 57 (3-4),* 225-238.

Lewis J. W. (2010). Audio-visual perception of everyday natural objects – hemodynamic studies in humans. In M. J. Naumer & J. Kaiser (Eds.), *Multisensory object perception in the primate brain* (pp. 155-190). New York: Springer Science+Business Media.

Logothetis N. K., Leopold D. A. & Sheinberg D. L. (1996). What is rivalling during binocular rivalry? *Nature, 380 (18),* 621-624.

Lunghi C. & Alais D. (2013). Touch interacts with vision during binocular rivalry with a tight orientation tuning. *PLOS ONE, 8 (3),* 1-8.

Lunghi C., Binda P. & Morrone M. C. (2010). Touch disambiguates rivalrous perception at early stages of visual analysis. *Current Biology, 20 (4),* 142-144.

Lunghi C. & Morrone M. C. (2012). Early interaction between vision and touch during binocular rivalry. *Multisensory Research (2013),* 1-16.

Macaluso E. (2006). Multisensory processing in sensory-specific cortical areas. *The Neuroscientist, 12 (4),* 327-338.

Markis S., Hadar A. A. & Yarrow K. (2011). Viewing objects and planning actions: On the potentiation of grasping behaviours by visual objects. *Brain and Cognition, 77,* 257-264.

Martin A. (2007). The representation of object concepts in the brain. *Annual Review of Psychology, 58,* 25-45.

Martin A. & Chao L. L. (2001). Semantic memory and the brain: Structure and processes. *Current Opinion in Neurobiology, 11,* 194-201.

Martin A., Wiggs C. L., Ungerleider L. G. & Haxby J. V. (1996). Neural correlates of category-specific knowledge. *Nature, 379 (6566),* 649-652.

Maruya K., Yang E. & Blake R. (2007). Voluntary action influences visual competition. *Psychological Science, 18 (12),* 1090-1098.

Mitchell J. F., Stoner G. R. & Reynolds J. H. (2004). Object-based attention determines dominance in binocular rivalry. *Nature, 429 (6990),* 410-413.

Murray M. M. & Wallace M. R. (2012). *The neural bases of multisensory processes*; CRC Press, Taylor & Francis Group.

Naumer M. J. & Kaiser J. (2010). *Multisensory Object Perception in the Primate Brain*; Springer Science+Business Media.

Newell F. N., Ernst M. O., Tjan B. S. & Bülthoff H. H. (2001). Viewpoint dependence in visual and haptic object recognition. *Psychological Science, 12 (1),* 37-42.

Noe A. (2004). *Action in Perception*. Cambridge, MA, US: MIT Press.

Norman J. F., Norman H. F., Clayton A. M., Lianekhammy J. & Zielke G. (2004). The visual and haptic perception of natural shape. *Perception & Psychophysics, 66 (2),* 342-351.

Pavani F., Spence C. & Driver J. (2000). Visual capture of touch: Out-of-body experiences with rubber gloves, *Psychological Science, 11 (5),* 353-359.

Penfield W. & Boldrey E. (1937). Somatic motor and sensory representation in the cerebral cortex of man as studied by electrical stimulation. *Brain: A journal of neurology, 60,* 389-443.

Petrini P., Furey M. L., Ricciardi E., Gobbini M. I., Wu W.-H. C, Cohen L., Guazzelli M. & Haxby J. V. (2004). Beyond sensory images: Object-based representation in the human ventral pathway. *Proceedings of the National Academy of Sciences of United States of America, 101 (15),* 5658-5663.

Ramachandran V. S. & Hubbard E. M. (2001). Synaesthesia – a window into perception, thought and language. *Journal of Consciousness Studies, 8 (12),* 3-34.

Reales J. M. & Ballesteros S. (1999). Implicit and explicit memory for visual and haptic objects: Cross-modal priming depends on structural descriptions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25 (3),* 644-663.

Reed C. L., Shoham S. & Halgren E. (2004). Neural substrates of tactile object recognition: An fMRI study. *Human Brain Mapping, 21,* 236-246.

Rich A. & Mattingley J. B. (2002). Anomalous perception in synaesthesia: A cognitive neuroscience perspective*. Nature Reviews Neuroscience, 3 (1)*, 43-52.

Richeimer J. (2006). Familiarity and the inferential theory of perception. *Theory Psychology, 16 (4),* 505-525.

Rock I. & Victor J. (1964). Vision and touch: An experimentally created conflict between the two senses, *Science, 143 (3606),* 594-596.

Romani G. L., Williamson S. J. & Kaufman L. (1982). Tonotopic organization of the human auditory cortex. *Science, 216 (4552),* 1339-1340.

D. E. Rumelhart & J. L. McClelland eds. (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. Cambridge, MA, US: MIT Press.

Saig A., Gordon G., Assa E. & Arieli A. Ahissar E. (2012). Motor-sensory confluence in tactile perception. *The Journal of Neuroscience, 32 (40),* 14022-14032.

Sanguinetti J. L., Allen J. J. B. & Peterson M. A. (2013). The ground side of an object: Perceived as shapeless yet processed for semantics. *Psychological Science, 24 (11),* 1-9.

Schacter D. L. (1992). Priming and multiple memory systems: Perceptual mechanisms of implicit memory. *Journal of Cognitive Neuroscience, 4 (3),* 244-256.

Singer W. (2007). Binding by synchrony. *Scholarpedia*, 2 (12): 1657. http://www.scholarpedia.org/article/Binding_by_synchrony.

Sperry R. W. (1950). Neural basis of the spontaneous optokinetic response produced by visual inversion. *Journal of comparative and physiological psychology, 43,* 482-489.

Stanley J., Forte J. D., Cavanagh P. & Carter O. (2011). Onset rivalry: The initial dominance phase is independent of ongoing perceptual alternations. *Frontiers in Human Neuroscience, 5*, 1-9.

Stein B. E. & Meredith M. A. (1993). *The merging of the senses.* Cambridge, MA, US: MIT Press.

Stein B. E. & Stanford T. R. (2008). Multisensory integration: Current issues from the perspective of the single neuron. *Nature Reviews Neuroscience, 9 (4)*, 255-267.

Sterzer P., Kleinschmidt A. & Rees G. (2009). The neural bases of multistable perception. *Trends in Cognitive Sciences. Vol. 13 (7),* 310-318.

Sterzer P. & Rees G. (2002). A neural basis for percept stabilization in binocular rivalry; *Journal of Cognitive Neuroscience, 20 (3),* 389-399.

Tong F., Meng M. & Blake R. (2006). Neural bases of binocular rivalry. *Trends in Cognitive Sciences, 10 (11),* 502-511.

Tulving E. & Schacter D. L. (1990). Priming and human memory systems. *Science, 247 (4940),* 301-306.

Violentyev A., Shimojo S. & Shams L. (2005). Touch-induced visual illusion. *Neuroreport, 16 (10),* 1107-1110.

van Ee R., van Boxtel J. J. A., Parker A. L. & Alais D. (2009). Multisensory congruency as a mechanism for attentional control over perceptual selection. *The Journal of Cognitive Neuroscience, 29 (37),* 11641-11649.

Varela F. J., Thompson E. & Rosch E. (1993). *Cognitive Science and Human Experience*. MIT Press.

Varela F. J., Lachaux J.-P., Rodriguez E. & Martinerie J. (2001). The brainweb: Phase synchronization and large-scale integration. *Nature Reviews Neuroscience, 2 (4)*, 229-239.

von Uexküll, J. (1980). *Kompositionslehre der Natur*. ed. by T. v. Uexküll, Frankfurt am Main: Propyläen Verlag.

Wallace M. T., Perrault Jr T. J., Hairston W. D. & Stein B. (2004). Visual experience is necessary for the development of multisensory integration. *The Journal of Neuroscience, 24 (43),* 9580-9584.

Wallace M. T., Wilkinson L. K. & Stein B. E. (1996). Representation and integration of multiple sensory inputs in primate superior colliculus. *Journal of Neurophysiology, 76 (2),* 1246-1266.

Welch R. B. & Warren D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin, 88 (3),* 638-667.

Wright T. M., Pelphrey K. A., Allison T., McKeown M. J. & McCarthy G. (2003). Polysensory interactions along lateral temporal regions evoked by audiovisual speech. *Cerebral Cortex, 13,* 1047-3211.

Zhang M., Weisser V. D., Stilla R., Prather S. C. & Sathian K. (2004). Multisensory cortical processing of object shape and ist relation to mental imagery. *Cognitive, Affective & Behavioral Neuroscience, 4 (2),* 251-259.

Zhou W., Jiang Y., He S. & Chen D. (2010). Olfaction modulates visual perception in binocular rivalry. *Current Biology, Vol. 20 (15),* 1356-1358.

Yildririm I. & Jacobs R. A. (2012). Transfer of object category knowledge across visual and haptic modalities: Experimental and computational studies. *Cognition, 126,* 135-148.

7. Appendix

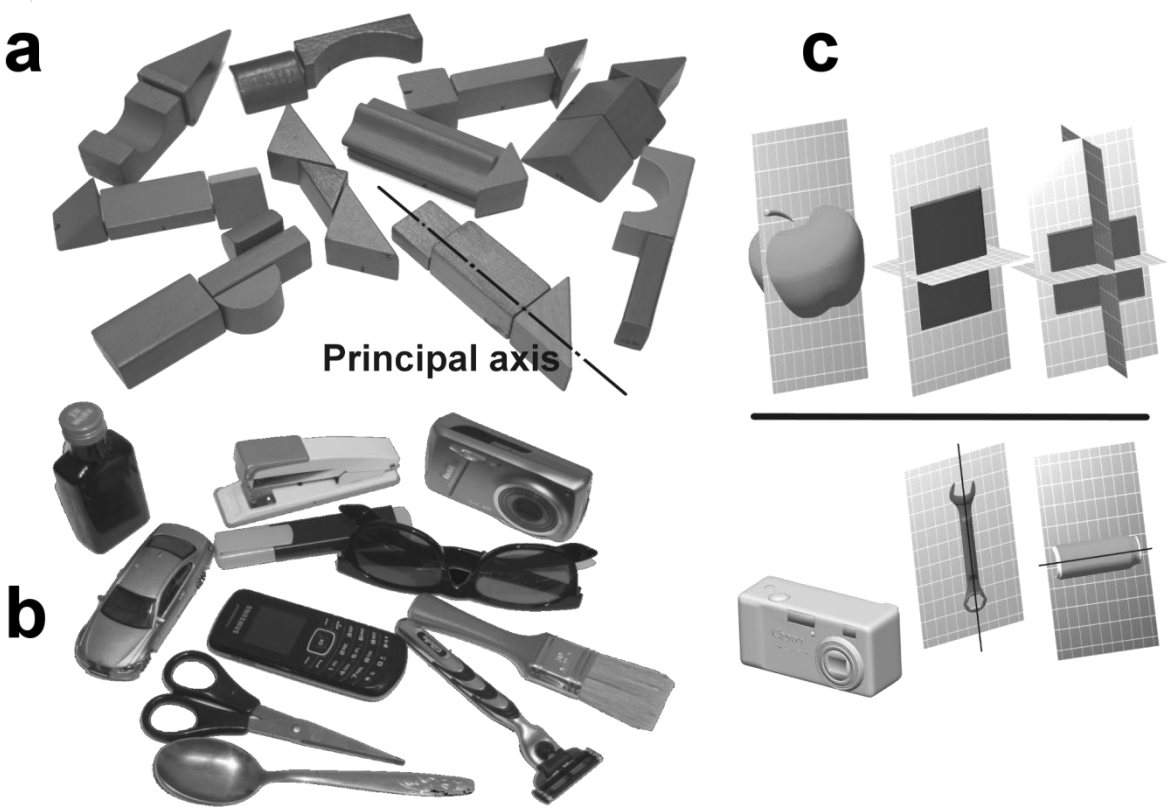

Figure 12: A sample of artifacts which were utilized in the present study. Both object classes depicted have one principle axis, along which the objects can be oriented on the object platform, allowing a positional variation. **(a)** Objects (20 in total) used exclusively in the unfamiliar condition in which the observers call the object's orientation only **(b)** Known objects (20 in total) which are used in the familiar condition, in which the observers need to call the artifact's name. **(c)** Schematic depiction of objects with different number of planes of symmetry. The left upper and lower objects have only one or none plane of symmetry and contain therefore maximum shape variation (=information). The greater the number of planes of symmetry, the less shape variation has the object. High shape variation (at least with respect to planes of symmetry) ensures that participants recognize the familiar objects and that there is no haptic ambiguity.

Table 4. The ‘Lab‘ colour space coordinates for darker (edges) and brighter (surfaces) RGB values.

		Left display			Right display		
		L	***a***	***b***	***L***	***a***	***b***
Edges	red	33.8	50.4	27.8	33.8	51.9	33.1
	green	33.5	-36.5	26.4	33.8	-37.2	29.2
	blue	34.6	74.0	-127.6	33.9	58.7	-115.8
Surfaces	red	57.3	79.0	56.0	51.8	73.4	59.9
	green	57.4	-63.0	51.1	57.0	-61.6	51.9
	blue	56.9	34.1	-93.6	57.1	21.9	-79.3

Table 5. Familiar object pairs during the familiar (*FAM*) condition.

Pliers	Glasses	Digital camera	Hole puncher
Shaving razor	Bottle opener	Brush	Cutter
Flask	Mobile phone	Glue stick	Marker pen
Toy car	Lighter	Tube	Pocket lamp
Scissors	Spoon	Stapler	Fork

Table 6. Parameters for the modified logistic function fitting for each sensory stimulation condition the time-dependent probability for a perceptual switch.

		FAM-			***FAM+***		
		VO	***VTC+***	***VTC-***	***VO***	***VTC+***	***VTC-***
Parameter	***a***	0.0027	0.0025	0.0027	0.0025	0.0026	0.0027
	b	2.011	1.8987	1.9143	2.1647	2.0724	2.0347
	c	0.2924	0.2781	0.2932	0.3279	0.3156	0.3142
	d	0.2966	0.2806	0.2964	0.3312	0.3186	0.3176
AUC	**Integral**	0.1646	0.1616	0.1683	0.1619	0.1643	0.1697

Curriculum Vitae

Persona

Name:	Dalibor Andrijević
Date of birth:	24. 07. 1982
Nationality:	Croatian

Studies

Oct 2010 – Oct 2013 **University of Vienna**

Joint Degree Programme MEi:CogSci Cognitive Science;
Master thesis: "*Visuo-haptic integration during binocular rivalry: Studying the roles of intermodal congruence and object familiarity*"

Mar 2008 – Oct 2010 **University of Vienna**

Bachelor degree studies in Physical anthropology and human behavioral ecology;
Bachelor-thesis: "*The crowing ape: The frequency of automanipulative behaviors as an indicator of stress in crowds*"; Grade: 1.9

Oct 2005 – Mar 2008 **University of Ratisbona**

Biology degree studies; completed intermediate diploma; Grade: 1,8

School

Sep 2003 – Jul 2005 **Vocational secondary school Ratisbona**

High school degree/A-levels
(Intensive subjects: physics, mathematics, English, German); Grade: 2.4

Sep 2000 – Jul 2003 **AUDI AG Ingolstadt**

Vocational training to become construction mechanic specializing in sheet construction technology; final qualification: Craft certificate, skilled worker

Scholarships

Oct 2011 – Mar 2012

Performance scholarship by the University of Vienna; Grade: 1.26

Sep 2011 – Mar 2012

Erasmus Scholarship by IEAC Austria: exchange semester in Hungary at the Budapest University of Technology (BME), Department of Cognitive Science; project collaboration in *SCANDLE* (Acustic scene analysis for detecting living entities) at the Hungarian Academy of Sciences (Magyar Tudományos Akadémia Pszichológiai Kutatóintézet)

Publications

Article:

Böhm T. M., Shestepalova L., Garreau G., Andrijević D., Georgiu J., Bendixen A., Denham S. L. & Wilkler I. (2012). Do visual cues affect the perceptual organization of sound sources moving in space? *Proceedings of the workshop on sounds and sound processing in natural and artificial systems;* Plymouth, UK

Language skills

German (like mother tongue)

Bosnian, Serbian and Croatian (mother tongue); Knowledge of cyrillic writing

English (fluent)

Spanish (basic knowledge)

Engagement, hobbies

Honorary member of the club “Kultureller Untergrund Riedenburg e. V.”: Co-organization and marketing of the open-air music festival “AgratAmAgatha” in Bavaria, Germany

Team sports (soccer, basketball), Boxing, Climbing, Hiking, Chess

Graphic design, photography and writing

References

Dr. Tamas M. Böhm
(Hungarian Academy of Sciences, Budapest; bohm@cogpsyphy.hu)

Univ. Prof. Dr. Ulrich Ansorge
(Faculty of Psychology, University of Vienna; ulrich.ansorge@univie.ac.at)