# Sensory enhancement of peripheral vision

**Dissertation**

kumulativ

zur Erlangung des Grades eines

Doktor der Naturwissenschaften

(Dr. rer.nat.)

des Fachbereichs Psychologie der Philipps-Universität Marburg

Vorgelegt von

**Mustafa Nedim Göktepe**

Aus Izmir, Türkei

Marburg, 2023

Vom Fachbereich Psychologie der Philipps-Universität Marburg (Hochschulkennziffer 1180)
als Dissertation angenommen am 31.10.2023

Erstgutachter(in): Prof. Dr. Alexander C. Schütz

Zweitgutachter(in): Prof. Dr. Karl R. Gegenfurtner

Tag der Disputation: 29.11.2023

# Table of Contents

Summary ..... 1
Introduction ..... 4
*Transsaccadic learning and peripheral-foveal associations* ..... 6
*Foveal retinotopic cortex supplements peripheral object recognition* ..... 10
*Combining peripheral information with other senses* ..... 14
Experimental Approaches ..... 17
Study I: The *role of familiarity in peripheral-foveal associations* ..... 17
Study II: *Foveal-feedback follows spatial frequency processing* ..... 22
Study III: *Combining peripheral information with other senses* ..... 28
General Discussion ..... 34
*Peripheral-foveal predictions with and without saccades* ..... 34
*Potential limitations of foveal-feedback in daily life* ..... 36
*Combining sensory information from within and across senses* ..... 37
References ..... 39
Appendix ..... 47
*Study I* ..... 47
*Study II* ..... 60
*Study III* ..... 84
Eigenständigkeitserklärung ..... 105
Authorship Statement ..... 106
Acknowledgments ..... 107

## Summary

More than 99% of the visual information is sampled by peripheral vision. Despite covering the majority of the visual field, the peripheral vision offers lower visual resolution than the fovea, that is responsible from gathering high resolution information from the central visual field. Although visual sensitivity changes drastically across the retina as a function of eccentricity, our visual experiences appear to be homogeneous. This apparent visual homogeneity is achieved by our visual system striving to optimize information gathering and minimize biological costs. To this end, the visual system uses various heuristics stemming from priors and expectations while dividing the labor of gathering information between available sensory systems. The aim of this dissertation is to provide an account of how various sensory mechanisms support peripheral vision. Particularly, in three studies, it investigated how peripheral vision and the execution of peripheral tasks are supported by transsaccadic learning and prediction, neural feedbacks providing additional processing resources, and supplementary information from other senses.

Study I investigated how transsaccadic learning and object predictions of familiar objects supports peripheral vision. Through transsaccadic learning the visual system associates how the appearance of an object or a feature change from periphery to fovea. Using these object specific associations visual system generates predictions about these objects and how they would look like at different eccentricities. In addition, through lifelong experience on how object appearance changes as a function of eccentricity, the visual system could generate predictions even for novel objects. However, it was unclear whether object specific predictions reserved for familiar objects provide an advantage over general predictions that are also available for novel objects in visual tasks. Study I addressed this question in two experiments where observers unknowingly familiarized with a subset of the objects by performing a sham task that required them to make saccades to these objects. On the following day, they either performed a peripheral-foveal matching or transsaccadic change detection task with familiarized and novel objects. We found that the presence of familiar objects improved the performance in both tasks by providing more precise object specific predictions from previous peripheral-foveal associations that generalize across the visual hemifields. Thus, Study I shows that object specific predictions unique to familiar objects provide additional support to the peripheral vision and execution of peripheral tasks.

Study II investigated a neural feedback mechanism that allows peripheral information to be processed in the fovea retinotopic cortex and supports peripheral discrimination. The support of

the foveal-feedback mechanism in peripheral discrimination can be impaired when a foveal input is presented asynchronously with peripheral targets. However, it was not clear whether the peripheral object information has to compete with the foveal input for the same neural resources, or if it is masked by it. Study II tested both explanations with a peripheral letter discrimination using both novel and familiar characters. Crucially, we manipulated the spatial frequency compositions of the foveal noise. Thus, if the foveal noise is masking the foveal-feedback, we would expect the efficiency of the foveal noises to vary depending on the amount of shared spatial frequency with the peripheral characters. Alternatively, if foveal noise is competing with the foveal-feedback, we would expect a more general effect of foveal noise independent from how they are similar to the peripheral characters. We found that low spatial frequency foveal noise was more effective at impairing the peripheral discrimination of both familiar and novel characters, indicating a frequency specific masking of foveal-feedback. We follow-up this result with a control experiment where the low and medium spatial frequency noises were presented overlappingly with the peripheral and foveal characters. As anticipated, we found that low frequencies were more effective at masking peripheral characters than medium frequencies while the opposite pattern was true for the foveal characters. Additionally, behavioral oscillation analyses suggested that the masking of foveal-feedback is periodic at around 5 Hz. Thus, Study II shows that the peripheral discrimination of both novel and familiar objects is supported by a foveal-feedback mechanism that periodically processes peripheral information and subjects to masking.

Study III investigated how imprecise peripheral information can be combined with sensory information from other modalities. More specifically, virtual and augmented reality applications are promising for augmenting user performance and experience by providing supplementary information across senses. However, one major bottleneck for these applications is to supplement information within a tight spatiotemporal window across different sensory modalities. Therefore, if and how spatiotemporally incongruent information from different sources is an important theoretical question with direct implications. Study III addressed this question by testing how imprecise peripheral information can be combined with supplementary tactile information when they are spatially and temporally incongruent. Using a custom-built setup, observers performed visual displacement judgments with or without spatially or temporally incongruent or congruent tactile displacement cues. Using their performance in the visual only condition, we modeled how observers combine visual and tactile information in the visuotactile conditions. We found that the combination weights systematically shifted towards tactile cues under temporal incongruency

compared to congruency condition. In contrast, spatial incongruency altered how visual and tactile information are combined and hinted possible individual differences in cue combination strategies. Thus, the weighting of visual and tactile information is modulated and altered by spatial and temporal incongruency which might have important consequences for multisensory applications. Nevertheless, Study III suggests that despite large temporal and spatial incongruencies tactile cues can supplement peripheral visual information.

In three studies, this dissertation seeks to understand how peripheral vision is supported by diverse neural and sensory mechanisms. In particular, peripheral vision is supported by precise object associations for familiar objects that are acquired through transsaccadic learning. These familiar object associations benefit peripheral matching and transsaccadic change detection by providing more precise peripheral to foveal and foveal to peripheral predictions than the general predictions available also for novel objects. Regardless of their familiarity, the peripheral discrimination of objects is also supported by a foveal-feedback mechanism that periodically processes peripheral object information in the foveal retinotopic areas. However, the processing of peripheral information is prone to masking by delayed foveal inputs with matching spatial frequency composition as the peripheral object. On the other hand, supplementary tactile information can be combined with imprecise peripheral information despite spatiotemporal incongruencies. However, while temporal incongruencies shift the weighting of visual and tactile information, spatial incongruencies can alter combination strategies differently across different individuals. In conclusion, these sensory interactions between peripheral vision and other sensory mechanisms support peripheral vision and offer better peripheral estimates for performing various tasks.

## Introduction

When we open our eyes, we immediately gain an abundant amount of information about our environment. Vision gives us an instant access about objects in the environment: how they are shaped, what color they are, what they are made of, whether they are stationary or not. While we experience the flow of visual information like watching a video, we rarely notice or have access to intricate mechanisms that afford us to have a stable and homogenous visual experience. Behind this video like flow, lies an effort to estimate and approximate the surrounding world through complex neural computations. Hence, the visual system provides us with an impression of a world that is not always consistent with the world we live in. During most of our daily lives, we rarely notice the difference between our visual experience and the veridical world and how fallible is the visual system. When we do, we call these occurrences illusions and study them to better understand how our visual experiences are fabricated. These impressions are affected by our beliefs and assumptions but most importantly by our perceptual system's goal of effectively gathering information with minimum cost.

The dual mandate of the visual system striving to optimize information gain and the biological cost is one of the key factors that sets a part visual perception from a video recording. For this, the visual system relies on various anatomical structures and neural mechanisms to process an incoming light (for a review see Stewart, Valsecchi, & Schütz, 2019). An incoming photon first received by the cornea, a concave structure that focuses the incoming light towards the pupil and the lens. Then the pupil, analogous to camera aperture, adjusts how much light will be let to enter the eye by contracting and dilating. Next, the lens further focuses light by stretching or shrinking depending on the focusing distance. Finally, the light arrives at the next major structure, the retina. The retina can be divided into three functional subsections: optic disc, fovea, and the periphery. The human retina is inverted in a sense that the light has to first travel through blood vessel and neural structures before reaching to the photoreceptors located far back. This inverted organization necessitates an opening for the ganglion cell axons to exit the eyeball. The optic disc provides this exit point but in return this part of the retina is vacant from photoreceptors. This creates a blind spot in the visual field (Mariotte 1668 cited in Mariotte, 1740; Ferree & Rand, 1912; Watanabe, 1958). The fovea is located centrally in macula and responsible for less than 1% of the visual field. It has the highest cone photoreceptor density in the retina but in return it does not contain any rod photoreceptors (Curcio et al., 1990; Oesterberg, 1935). Crucially, the high ratio of ganglion cells

per photoreceptor affords the fovea to have high-resolution (Virsu & Rovamo, 1979). In contrast, the rest of the retina, periphery, covers more than 99% of the visual field but it is mainly populated by rods. More importantly, the ratio of ganglion cells per photoreceptor gradually reduces with increasing eccentricity. The varying rate of ganglion connectivity leads to the overrepresentation of central visual field throughout the visual processing. The lateral geniculate nucleus (LGN), an important hub for parsing and relaying visual information from retinal ganglion cells to early visual cortex (Weyand, 2016), shows this characteristic by having exponentially more cellular structure allocated for smaller eccentricities (Malpeli & Baker, 1975). Another major structure, V1, continues to show retinotopic organization and cortical magnification (Benson et al., 2021). Thus, overrepresentation of central visual field continues along the further in visual processes by cortical structures that either continue to show cortical magnification or receiving input from such areas.

While the visual perception is heavily biased by the information sampled with the central vision including the fovea, the visual system benefits from having two modes of visual information provided by periphery and fovea. For instance, only periphery contains rod receptors that are more sensitive to light and primary supplier of visual information under scotopic conditions (Zele & Cao, 2015). Another distinction between visual information sampled by fovea and periphery is the degree of visual detail. The peripheral vision is characterized by reduced overall contrast sensitivity compared to fovea with abrupt cut-offs for higher frequencies (Pointer & Hess, 1989; Robson & Graham, 1981). One reason for the lower contrast sensitivity in the periphery compared to fovea arises from the larger receptive fields of the retinal ganglion cells in the periphery and linearly related to degree of cortical magnification (Virsu & Rovamo, 1979). The reduced cortical representation of peripheral information does not afford high acuity information from the peripheral retina and contributes to crowding (Kwon & Liu, 2019). This leads to a pooling of peripheral information that provides summary statistics about the visual environment (Rosenholtz et al., 2012). Thus, peripheral vision delivers coarse information about the environment at a low biological cost (lesser energy consumption due to lower ganglion cell connectivity). Therefore, there is a division of labor between fovea and periphery that affords the visual system a capacity to sample fine visual information via fovea while having a coarse information about the overall visual environment.

This division of labor extends to other modalities as peripheral vision provides information about the environment at a glance while slower sensory information from other modalities provides

fine tune adjustments for action planning and final perceptual estimates (Vater et al., 2022). For instance, the visual system guides manually tracking or intercepting moving targets with eye movements (Fooken et al., 2021). By covering a large portion of the visual field, peripheral vision allows monitoring of multiple objects, when the visual information required for the task can be resolvable in the periphery (Cámara et al., 2020).

Another major advantage of dividing labor across senses is that combining information across senses can provide better overall estimate of the environment. The reliability of the sensory information is constrained by both sensory and environmental factors. The sensory constraints include the inherently noisy measurements of physical properties and the physiological resolution limits such as cone density or rate of ganglion cell connectivity. Combined with external factors like low lightning conditions, sensory estimates from individual modalities might be unreliable. In order to obtain more reliable measurements, these individually less reliable estimates can be combined or integrated (Ernst & Bülthoff, 2004). Thus, there exists a larger division of labor between all perceptual systems for reliably estimating the physical world by collecting complimentary as well as redundant information from the environment.

A major contribution of peripheral vision for generating reliable estimates in visual and multisensory perception is to economically sample information from a large visual field. In return, when the peripheral resolution is not adequate for the task at hand, other more reliable sources of information such as foveal vision and other senses can support the peripheral tasks by supplementing information or processing power. The aim of this dissertation is to investigate how these divisions of labor between peripheral vision and other sensory mechanisms support peripheral vision. More specifically, the subsequent sections will describe how peripheral and foveal information are associated through saccades for providing prediction about peripheral objects, how foveal feedback processes supports the discrimination of peripheral objects, and how tactile cues influence peripheral visual displacement judgments.

### *Transsaccadic learning and peripheral-foveal associations*

Humans execute ballistic eye movements called saccades three to four times per second (Rayner, 1998). As a result, fovea is shifted to various locations of interest in visual scenes. This allows us to sample high resolution information from a portion of the field larger than the actual size of the fovea. More importantly, by executing saccades to the different regions of the visual

scene, the visual system obtains both low resolution information from periphery and higher-resolution information from fovea for the same object of interest. Thus, the visual processing of saccade targets begins in the periphery. However, each saccade destabilizes visual perception by shifting the retinotopic coordinates of the object without any real change in spatial (world) coordinates. To keep the perceptual stable, visual system remaps the visual scene when it is preparing for the next saccade (Duhamel et al., 1992; Gottlieb et al., 1998; Kusunoki & Goldberg, 2003; Umeno & Goldberg, 2001; Walker et al., 1995, see also Cavanagh et al., 2010). When a saccade is planned, receptive fields are shifted to the saccade target to preemptively process the object if the object falls to those receptive fields after the saccade. While this stabilizes object position in retinotopic coordinates, object features are also transferred with remapping (Melcher, 2005, 2007, see also Melcher, 2011).

The remapping of object features with each impending saccade is one of the gateways for facilitating the division of labor between peripheral and foveal vision. Through remapping, when a saccade is planned and executed towards a peripheral object, the presaccadic object information sampled by the periphery is linked to the postsaccadic object information sampled by the fovea (Zirnsak & Moore, 2014). Within the ideomotor framework, this establishes a crucial link between saccades and their perceptual consequences. According to the ideomotor framework, coupling actions and their sensory consequences can allow learning the sensory outcomes of one's action (Shin et al., 2010 ; see also Moeller & Pfister, 2022; Pfister, 2019). In return, the associative learning of actions and their consequences creates a wealth of experience that can be used to generate predictions about the outcomes of the planned action (Herwig & Schneider, 2014). Thus, allowing planning of actions to yield the desired sensory outcome based on the predictions generated from previous experiences. This has a fundamental role in everyday actions such as knowing how far to reach or how much force grip one should apply to a wine glass for picking it up without breaking it.

For vision, one benefit of associating actions with their sensory outcomes is that it can be used to optimize saccade planning. For instance, saccade end point can be planned and adapted by the expected informational value of the presaccadic target (Schütz et al., 2012; Wolf et al., 2019). For novel scenes, objects, or tasks, however, this can be achieved by relying on the predictions generated based on the previous action outcomes. This is a highly desirable strategy for the visual system to follow as it can economize information sampling by collecting fast but low-resolution

information from periphery while analyzing foveal information with scrutiny (Ludwig et al., 2014). The general predictions stemming from the overall visual experience can serve both foveal and peripheral vision. If the task requires scanning of most of the visual scene as in visual search, or if the object of interest can be resolvable in the periphery, the visual system can economize saccades by collecting information primarily with the peripheral vision. In contrast, if the task requires decisions based on fine-grain details like in radiological evaluations (e.g., discriminating benign from malignant tissues), the visual system can adapt its saccade planning for minimizing saccade errors (Lisi et al., 2019) and sample foveal information guided by action-consequence predictions (Herwig & Schneider, 2014).

Besides optimizing visual scanning and processing of novel objects, action-consequence pairs provide object or feature specific predictions about familiar items in visual scenes. The visual system can transsaccadically learn how the appearance of specific objects or features change with each saccade and use this information for generating object or feature specific predictions. For instance, Valsecchi & Gegenfurtner (2016) showed that perceptual estimations are shaped by the peripheral-foveal predictions and can be altered upon establishing a new action-consequence association. More specifically, they showed that size estimations can be re-calibrated when the size of the objects were systematically changed across saccades. First, observers acquired new action-consequence coupling by comparing the size of a central and peripheral stimulus. Critically, after the size comparison, they had to execute a saccade to the peripheral object which changed it size during saccade unbeknownst to the observers. Thus, observers transsaccadically either learned that object size is over or underestimated in periphery relative to fovea. This re-calibration of object size led to a bias in subsequent size estimations consistent with the newly acquired peripheral-foveal associations.

In their seminal work Herwig & Schneider (2014), propose a transsaccadic learning mechanism able to generate object-specific predictions by associating peripheral and foveal object information belonging to the same object. The transsaccadic association of objects then can be used generate predictions about what is like to see an object in different viewing conditions. They showed that these predictions support peripheral object recognition and visual search. More specifically, the transsaccadic feature prediction mechanism can help recognizing peripheral objects by making predictions about what it would be like to see the same object more foveally. Similarly, it can facilitate visual search by generating predictions about what it would be like to

see the search target peripherally. Hence, the transsaccadic feature prediction mechanism could generate predictions from the peripheral visual input about the foveal object information as well as from foveal visual input to generate predictions about peripheral object information. Subsequent studies have shown that these predictions contribute to discrimination of simple shapes (Herwig, 2015), spatial frequency gratings (Herwig et al., 2018; Herwig & Schneider, 2014), as well as complex daily life objects (Osterbrink & Herwig, 2019, 2021) and faces (Osterbrink & Herwig, 2021).

An important question is whether these predictions are limited to the location of acquisition or can generalize to other locations. Considering the retinotopic organization found in visual areas, it is reasonable to assume that the peripheral and foveal information are associated and provide predictions in a retinotopic manner. On one hand, if such predictions can only be used for specific retinotopic coordinates, then the visual system would need to store multiple instances of the same association for different retinotopic coordinates. On the other hand, if the peripheral and foveal associations can be extrapolated to provide predictions generalizable to other retinotopic coordinates, this would provide a greater flexibility and less redundancy but may require using more computational resources for calculating complex extrapolations (Osterbrink & Herwig, 2023). Both arguments found support from behavioral studies showing either generalization or location specificity. The participants in the study of Herwig, Weiss, & Schneider (2018) acquired arbitrary object associations where spatial frequency of the objects differ between peripheral and foveal percepts. In the subsequent test phase, they performed a spatial discrimination task with objects that are located in the same spatial and/or retinotopic locations as the objects in the acquisition phase or in different locations. The authors found that only targets located in the same retinotopic coordinates bias the perceptual judgments consistent with the location specificity and redundancy argument. In contrast, Valsecchi & Gegenfurtner (2016) found that the arbitrary object associations suggesting a difference in peripheral and foveal size estimation, bias subsequent size estimations for objects located at the same location as in the acquisition as well as at the opposite visual hemifield. Although these results seem conflicting, their divergence might be explained by the differences in designs. Perhaps the most important difference between studies is the stimuli they employed. In one hand, Valsecchi & Gegenfurtner (2016) manipulated the object size that is expected to be perceived similarly at opposite hemifields with the same eccentricity. On the other hand, Herwig, Weiss, & Schneider (2018) manipulated the spatial frequency objects, which can be

perceived drastically different in periphery and fovea. Therefore, the visual system already expects some discrepancy between the peripheral and foveal objects.

The location specificity of transsaccadic learning and feature predictions are still an open debate. However, in either case, these mechanisms portray how the visual system benefits from having two different sources of visual information. The division of labor between the peripheral and foveal vision is clearly advantageous for quickly collecting coarse information from visual scenes and using this information for identifying regions of interest that require further processing by fovea.

### *Foveal retinotopic cortex supplements peripheral object recognition*

While the eye movements provide a unique tool for the visual system for exploiting the merits of both peripheral and foveal vision, the organizational differences in how peripheral and foveal information is processed in the cortex continues to support the division of labor even when the eyes are relatively still.

Despite covering the majority of the visual field, the peripheral vision is underrepresented in the visual hierarchy starting from as early as LGN (Purves et al., 2001). Instead, the representation of central vision takes up a substantial proportion of the cortical structure. As a result, more resources are allocated in the processing of foveal information. Thus, it was conventionally assumed that the peripheral and foveal information were processed by distinct retinotopically organized structures in the early visual cortex. Recently, this idea has been challenged by a new body of research suggesting a striking interplay between the peripheral vision and the foveal retinotopic cortex. In a functional magnetic resonance imaging (fMRI) study, Williams et al. (2008) showed that peripheral object information is also processed in the foveal retinotopic regions of the visual cortex. They measured blood oxygenation level dependent (BOLD) activity of participants performing an object discrimination task in the periphery. These objects were from a collection of algorithm-generated complex objects that form distinct categories based on their features (Op De Beeck et al., 2006). The task of the participants was to make within-category discrimination judgments for objects presented 7° away from fixation. Astonishingly, they found activations in the foveal retinotopic regions correlated with the objects presented in periphery. This correlated activity was specific to objects from the same category even when they were presented at the opposite hemifield, suggesting that the foveal retinotopic areas processes

position invariant peripheral object information. They postulated that this is a feedback process that feeds the peripheral object information from higher cortical areas to the foveal retinotopic cortex to further process the peripheral information. Subsequently, a follow-up study from the same group (Chambers et al., 2013) provided the first causal evidence for the so called foveal-feedback mechanism. Using the same peripheral discrimination task and objects, they showed that inhibiting the activity in the foveal retinotopic cortex with transcranial magnetic stimulation (TMS) impaired discrimination of the peripheral objects. This impairment provided further evidence for foveal-feedback by showing that the foveal retinotopic cortex receives and processes peripheral object information. The inhibition of the foveal retinotopic regions modulated the peripheral discrimination performance only when the magnetic stimulation registered 350 ms after the onset of peripheral objects. This indicates that the foveal-feedback has a specific time course locked to peripheral stimulus onset. The 350 ms stimulus onset asynchrony (SOA) suggest that the peripheral object information becomes available to foveal-retinotopic cortex at a later stage of the visual processing at a time frame associated more with feedback than feedforward processes (Lamme & Roelfsema, 2000).

The impairment of peripheral discrimination with TMS pulses provided the first direct evidence for the role of foveal-feedback in peripheral object processing. More importantly, it provided the critical behavioral measure for investigating the foveal-feedback. Yu & Shim (2016) was the first to study foveal-feedback with a purely behavioral paradigm. They used a peripheral matching task where participants had to report whether the peripheral object was identical to or different from a target object. Crucially, another object at fixation was briefly presented with a 150 ms SOA. This object was either identical to target, from a different category or a scrambled object. When the central object was identical to the peripheral object, the peripheral matching performance was improved compared to when it was from a different category or scrambled. Critically, this improvement was specific to identical targets that were presented at fixation with 150 ms SOA and not when they were presented simultaneously (0 SOA) with the peripheral objects or in other locations. Thus, asynchronous foveal visual stimulation interacts with the foveal-feedback process comparable to a TMS stimulation to the foveal retinotopic cortex. Moreover, results of Yu & Shim (2016) suggest that the foveal-feedback benefits from incoming foveal input from the same peripheral object. As in Chambers et al. (2013), this modulation was limited to foveal processes and had a specific timeframe. However, it is important to note that the foveal-feedback effect in TMS was at 350 ms SOA while it was 150 ms SOA. However, this difference can be attributed to

differences in the manipulation. While the foveal visual in Yu & Shim (2016) needs to be processed by the earlier mechanisms before it is fed back to foveal retinotopic cortex, TMS pulse can directly target the foveal-feedback. Thus, 150 SOA (150-183 SOA including the presentation duration of the foveal input) provides a behavioral benchmark for studying foveal-feedback with purely behavioral designs.

The ability to study foveal-feedback with purely behavioral design motivated a number of studies investigating the key questions for better understanding the foveal-feedback mechanism. One important question is what is the informational content of the foveal-feedback? In other words, what is the nature of the information processed by the foveal-retinotopic cortex and by extension what is the relationship between the delayed foveal objects with the peripheral targets. Williams et al (2008) already provided the first evidence that the foveal-feedback processes peripheral object information at least at the category level. In line with this, coinciding foveal information of objects from different categories impairs peripheral discrimination performance while foveal input of the same object supplements it (Yu & Shim, 2016). However, another study (Weldon et al., 2016) reports that the peripheral discrimination of the objects from the same stimulus set was equally impaired by foveal distractors from the same as well as different categories. This is in contradiction with the previous results suggesting that the foveal-feedback processes peripheral object information at least at object category. Considering that the object in these object sets have maximal similarity among category members and minimal similarity with the objects from the different categories (Op de Beeck, Baker, Di Carlo, & Kanwisher, 2006), contradicting results from these studies is puzzling. Thus, the exact relationship between the delayed foveal inputs and the informational content of the foveal-feedback is still an open question.

Another question directly related to the information processed by the foveal-feedback is: What kind of objects does foveal-feedback process to supplement peripheral discrimination? Multiple studies provided evidence for the involvement of foveal-feedback in the processing of complex peripheral objects (Chambers et al., 2013; Contemori et al., 2022; Fan et al., 2016; Ramezani et al., 2019; Weldon et al., 2016, 2020; Williams et al., 2008; Yu & Shim, 2016, for a review Oletto et al., 2022). However, the low-pass filtered versions of the same objects were unaffected by the same foveal distractors (Fan et al., 2016). The evidence for the involvement of foveal-feedback in processing of simple objects and their features is even less conclusive. In one hand, peripheral color and shape discrimination is still effective by an asynchronous foveal input

(Weldon et al., 2020), suggesting that the foveal-feedback processes shape and color information. On the other hand, neither peripheral orientation judgments (Yu & Shim, 2016) or nor peripheral motion speed judgments of moving dots (Fan et al., 2016) were affected by foveal distractors. Therefore, the current evidence seems to suggest that the foveal-feedback is involved in the processing of objects that may require far greater processing resources than those allocated to the peripheral vision. Thus, the visual system might be employing the foveal retinotopic regions to further process lower-resolution peripheral information with additional resources when the peripheral tasks require detailed processing of the peripheral information. In line with this argument, foveal-feedback supports peripheral discrimination when the task requires subordinate category level (e.g., duck/non-duck, race car/non-race car), or basic category level (e.g., bird/non-bird, car/non-car) but not when the task requires a superordinate category level (e.g., animal/non-animal, vehicle/non-vehichle) discrimination (Ramezani et al., 2019).

Another important characteristic of foveal-feedback is the specific time window where the peripheral information is processed by the foveal retinotopic cortex. While the temporal discrepancy between TMS and behavioral manipulation can be explained by the difference in the methodologies, the exact time window of the foveal feedback is still a point of debate among behavioral studies (see Table 1 in Oletto et al., 2022 for a summary). For instance, while Weldon et al. (2016, 2020) reports an impairment on peripheral discrimination performance by a foveal input as early as 117 ms SOA, Fan et al. (2016) reports a similar impairment at 150 ms SOA, and Ramezani et al. (2019) finds an impairment at 300 ms SOA. The large variances in the reported time windows might be reflecting the differences among studies. For instance, in Ramezani et al. (2019), the peripheral objects were pictures of animals and vehicles in everyday environment, while the other studies have used the same artificial and isolated objects. Thus, these objects might be available at different timepoints to the foveal-feedback. Similarly, the complexity of foveal distractors used in these studies quite variably ranging from 1/f noise (Ramezani et al., 2019) to complex objects (Weldon et al., 2016) that might be intercepting foveal-feedback at different time points. Hence, the time window of foveal-feedback might be contingent on the relationship between the peripheral objects and the delayed foveal inputs. Another potential explanation could be the difference in task demands between discriminating natural subordinate objects and the more systematic artificial objects. In line with this argument, foveal-feedback has been found to be modulated by the task difficulty. When a mental rotation has to be carried out before making a comparison between two peripheral objects, the time window of foveal-feedback shifts to a later

time point with increasing mental rotation demand (Fan et al., 2016). Nevertheless, the reported time window of foveal-feedback still varies among studies using the same objects and tasks (117 ms for Weldon et al., 2016; Weldon et al., 2020 vs. 150 ms for Fan et al., 2016; Contemori et al., 2023). Finally, previous studies sparsely sampled SOAs for determining the time window of foveal-feedback. Therefore, their temporal resolution was limited by the long foveal presentation durations between 83-117 ms. This can partially explain the seemingly conflicting reports since when the total duration of the foveal presentation is considered, the reported time windows overlap considerably. Finally, recent evidence suggests that the foveal-feedback may not be a one-off but a periodic process. Contemori et al (2023) densely sampled the SOAs to improve their resolution in estimating the time window of foveal-feedback. They report a major impairment in peripheral discrimination performance at around 150 ms SOA. More importantly, densely sampling SOA allowed them to identify an oscillatory pattern in the behavioral performance suggesting that the foveal-feedback occurs periodically at a rate of around 5 Hz.

The evidence so far describes foveal-feedback as a mechanism periodically supporting the peripheral discrimination of certain objects and features. It appears that when there is a need to peripherally discriminate objects without executing a saccade, the visual system employs processing power normally allocated to central retina to further process the incoming peripheral information (Fan et al., 2016). Hence, foveal-feedback might be another mechanism showing how the visual system exploits having two channels of information while compensating the shortcomings peripheral vision.

### *Combining peripheral information with other senses*

While foveal vision can support the execution of peripheral tasks through object predictions, or by foveal-feedback, peripheral vision also interacts with modalities other than vision. By covering the majority of the visual field peripheral vision provides visual information about objects and their features that collectively contribute to our perceptual experience. However, our experiences are rarely unimodal. When we take a walk around the Lahn, we see the river flows slowly, ducks and other birds fly by, and the trees form a green curtain between us and the river. At the same time, we hear the river when we pass by the water gates or hear birds chirping while smelling grass after rain. A pleasant scene as it is, it is a bombardment of information through different senses. Therefore, it is a constant challenge for the perceptual systems to filter and combine information from multiple senses into one holistic percept, known as the binding problem.

However, one great advantage of sampling object information with different senses is to have a more reliable estimation of what and where the objects are. Consider an owl sitting fully camouflaged on a tree right ahead. It would be very challenging to spot it maybe even for a trained eye. However, hearing a hoot once would almost immediately give away the owl's location that now became harder to miss.

Being able to effortlessly collect multisensory information affords us to combine or integrate information that reduces ambiguity and lead to have more reliable interpretation of the world (Ernst & Bülthoff, 2004). While peripheral retina collects visual information from the large portion of the visual field, the reliability of the collected information may not always be sufficient to make decisions (Rosenholtz, 2016). Thus, a secondary source of information either in the form of prediction from transsaccadic associations (i.e., Herwig & Schneider, 2014; (Valsecchi & Gegenfurtner, 2016) or from other senses can greatly serve to peripheral tasks.

These secondary sources of information could supplement or substitute peripheral vision and improve visual (in)abilities in various visual tasks. For instance, spatial navigation in visually impaired groups improved by technologies that substitutes visual information with auditory information (Jicol et al., 2020; Neugebauer et al., 2020; Scalvini et al., 2022). Similarly, visual information can be supplemented by other senses when the available visual information is not sufficient to perform a task (Brown et al., 2019). Finally, virtual reality applications could improve user performance by providing an additional source of information even when the normally available information is sufficient by itself to perform the task. Thus, peripheral information can benefit from information collected by other senses when the visual information is not reliable or completely missing.

Unlike secondary sources of visual information that can be directly mapped to the appropriate peripheral information through remapping (Duhamel et al., 1992; Gottlieb et al., 1998; Kusoniki & Goldberg, 2002; Umeno & Goldberg, 1997; Walker et al., 1995, see also Cavanagh et al., 2010), mapping of information collected by other senses might have additional constraints. For instance, multisensory neurons encode information from different modalities (Alais et al., 2010; Stein & Meredith, 1993). These multisensory neurons encode information from single modalities as well as from multiple modalities if the signals from those modalities spatially (Meredith & Stein, 1996) and temporally proximal to each other (Meredith et al., 1987). In multisensory integration research, the so called "spatial rule" dictates that signals that the degree of multisensory integration

of two signals depends on their spatial proximity to each other (e.g., Gepshtein et al., 2005; Slutsky & Recanzone, 2001). Thus, multisensory integration occurs not only for spatially overlapping signals but also signals that are sufficiently close to each other like in a ventriloquist show. The size of this spatial window depends on the individuals' spatial discrimination threshold (Slutsky & Recanzone, 2001). However, the spatial rule is not a prerequisite for multisensory integration and predominantly relevant for spatial tasks (e.g., localization, spatial discrimination, Spence, 2013). In contrast, the temporal congruency of signals seems critical for a wide range of temporal (e.g., temporal order judgment) and non-temporal tasks and across modalities (Diederich & Colonius, 2015; Jagini, 2021). For instance, while audiovisual cues can be integrated within a 200 ms (-100 to 100 ms discrepancy) window (Slutsky & Recanzone, 2001), visuoproprioceptive cues can be integrated within a wider window of 400 ms (Kannape et al., 2019). Moreover, multisensory cues with larger temporal discrepancy can be still integrated if the lag between cues is informative (van Polanen et al., 2019). Similarly, spatial and temporal congruency might be also crucial for determining if and how peripheral information supplemented by additional sensory information.

Supplementing peripheral vision with additional sensory information from different modalities can provide a more reliable perceptual estimate than visual estimates (Brown et al., 2019, RealSense, Intel). Therefore, it can be a powerful tool for overcoming the limitations of peripheral vision (e.g., crowding) when performing peripheral tasks. Thus, like the division of labor between peripheral and foveal vision, there is a division of labor between peripheral vision and other senses for generating a more reliable estimate of the world.

## Experimental Approaches

### Study I: The *role of familiarity in peripheral-foveal associations*

Goktepe, N., & Schütz, A. C. (2023). Familiar objects benefit more from transsaccadic feature predictions. *Attention, Perception, & Psychophysics*. doi:10.3758/s13414-022-02651-8

With frequently executed saccades the visual system samples object information from different parts of the retina. As a result, the visual system obtains information from the same object with different resolutions. In contrast to the abrupt changes to object appearances, our perceptual experience is fairly stable. One key mechanism contributing to visual stability across the visual field is the ability to predict "what would it be like to see an object in other parts of the visual field" from previous transsaccadic associations of object-specific information. For instance, in the face of new transsaccadic associations the visual system can re-calibrate its predictions for objects (Osterbrink & Herwig, 2019; Osterbrink & Herwig, 2021) or visual features such as object size (Valsecchi & Gegenfurtner, 2016).

While previously acquired object-specific associations can help explain how visual stability can be achieved for these familiar objects, it cannot be the only mechanism as the perception of novel object lacking these associations is also stable. Therefore, it is unclear if previous experience with specific objects is needed for making transsaccadic predictions or object specific associations provides more precise predictions. It is reasonable to assume that there could be a more general mechanism that can generate predictions for novel objects based on the cumulative experience of how object appearance changes with eccentricity. In contrast, the predictions for familiar objects might be more precise since they can be generated from the previously acquired object associations. In two experiments, Study I tests if object-specific associations for familiar objects provide an extra benefit to familiar objects in peripheral matching (Experiment 1) and transsaccadic change detection (Experiment 2). As individual experiences can vary between observers, we controlled object familiarity by using an object set completely novel to our observers. Therefore, in both experiments observers first completed a training task to be familiarized with a subset of the stimulus set. In the training task, they performed a transsaccadic orientation change detection task, thereby

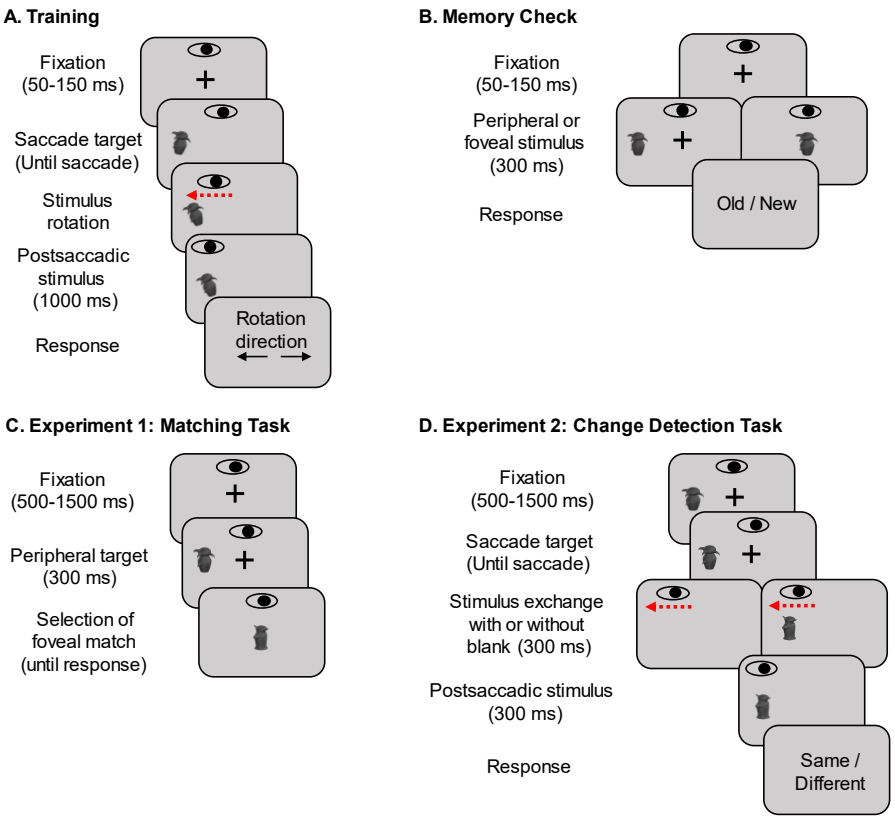



*Figure* 1. General design of experiments in Study 1. Example trials for the sham training task (A), recognition task in Memory Check (B), peripheral-foveal matching task in Exp 1 (C), and transsacadic change detection task in Exp 2 (D). Trials are not drawn to scale. The eye icon represents gaze position. Red arrow indicates saccade direction.

unknowingly acquired transsaccadic associations of peripheral and foveal object information without any explicit instruction (Figure 1). On the next day, they performed an object recognition task to check their familiarity with the objects from the previous day. Then the observers in Experiment 1, proceeded to a peripheral matching task where they were asked to match one of the three foveal options to the briefly presented peripheral targets. We manipulated the familiarity of both peripheral targets and foveal options to compare peripheral and foveal predictions for familiar and novel objects as follows. The foveal options for novel peripheral targets were either all novel, two familiar, or one familiar object. Similarly, foveal options for familiar targets were either all familiar, two familiar, or one familiar object (Figure 2A). We first tested if novel objects benefit from peripheral-foveal predictions by analyzing the matching performance in trials with novel peripheral targets (gray solid line in Figure 2A) and all-novel foveal options (0 Familiar Options in Figure 2A). We found that observers were able to match novel peripheral targets with only novel foveal options suggesting that previously acquired object-specific associations are not necessary for making peripheral-foveal predictions. Next, by analyzing trials with novel peripheral targets

with different number of familiar options, we found that object-specific predictions from non-matching familiar objects improves peripheral matching performance for novel objects. Thus, object-specific predictions from foveal familiar objects were beneficial for discounting non-matching options. Conversely, we found that presence of non-matching familiar options did not improve the peripheral matching of familiar peripheral targets. However, the absence of this effect for familiar peripheral targets can be attributed to ceiling effect, since the overall peripheral matching performance for familiar peripheral targets were already high (~80%).

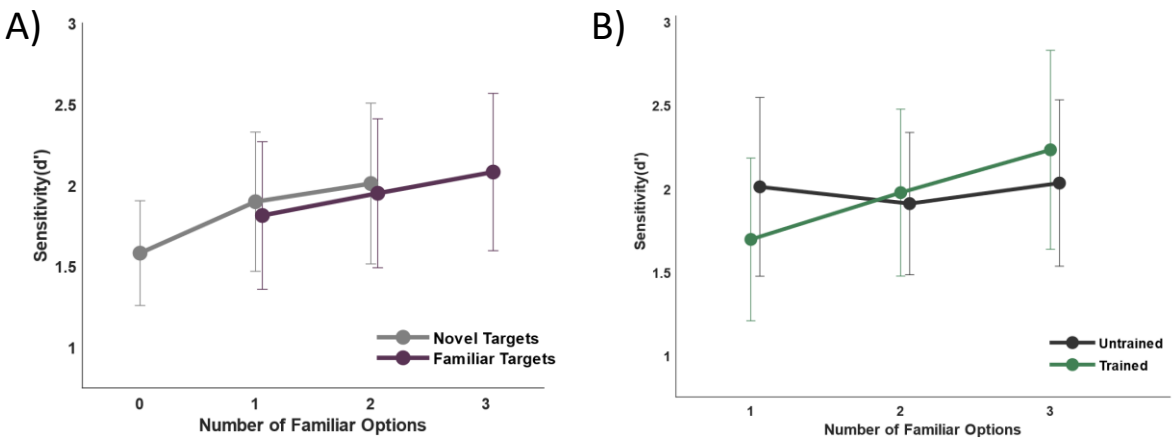



*Figure* 2. Results of Experiment 1. A) Peripheral discrimination performance, with familiar/novel foveal options (x-axis) and familiar/novel peripheral targets (separate lines). B) Peripheral discrimination performance for familiar targets presented on trained (green line) or untrained (gray line) locations for different number of foveal options (x-axis). Error bars show 95% confidence intervals.

In Experiment 1 we also tested whether the object-specific predictions are bounded to where they have been acquired or they could be generalized to a mirrored location in the other hemifield. During the training, observers acquired object-specific associations from objects that were always presented on the left side of the display. In contrast, familiar peripheral targets in Experiment 1 were presented either at the same location as the training or on right side at a mirrored eccentricity. Hence, we tested the location specificity of the object-specific associations by comparing the trials where the familiar peripheral target presented on the trained location with the trials where they were on the opposite hemifield (Figure 2B). We found that the peripheral matching performance did not depend on the training location and generalized to the opposite visual hemifield. Therefore, as Valsecchi & Gegenfurtner (2016) and unlike Herwig et al. (2018) we suggest that peripheral-foveal predictions can be generated from previously acquired object associations regardless of where they have been acquired. Overall, Experiment 1 showed that peripheral-foveal predictions can be made for novel objects lacking the previously acquired object-specific associations. On the

other hand, object-specific predictions reserved for familiar objects generalize to opposite hemifield and improve the peripheral matching performance.

While Experiment 1 allowed us to compare different combinations of novel and familiar object predictions, it was missing a critical component of daily life, saccades. Stitching pre- and post-saccadic information together is crucial for achieving perceptual stability and marrying peripheral and foveal visual information (Herwig, 2015). With each saccade the visual system has to decide if the pre- and post-saccadic information are similar enough for integration or not (Atsma et al., 2016; Demeyer et al., 2010). While Experiment 1 shows that peripheral-foveal predictions are not unique to familiar objects, object-specific associations of familiar objects were advantageous for detecting mismatches between pre- and post-saccadic objects to be more readily detected when object-specific associations are available.

A robust way to improve the detection of small transsaccadic changes that otherwise go undetected is to introduce a “blank” period is introduced following the saccade onset and prior the onset of the postsaccadic object (e.g., Deubel et al., 1996; Weiß et al., 2015). The blanking is argued to separate pre- and post-saccadic information that otherwise integrated (Atsma et al., 2016; Vaziri et al., 2006) or suppressed (Takano et al., 2020). Therefore, we reasoned that separation of pre- and post-saccadic information with blanking could further improve the benefit of having precise object-specific predictions by better separating the predictions for both objects.

We conducted Experiment 2 to test if object-specific predictions for familiar objects improves the detection of transsacadic changes with or without blanking. After completing the training and memory check, observers completed a transsaccadic change detection task where the pre-saccadic objects were replaced by an identical or a different object either immediately with saccades or after a 300 ms blanking period (Figure 1D). We manipulated the availability of object-specific predictions by varying the familiarity of pre- and post-saccadic objects. Since the change detection performance depends on the degree of dissimilarity between pre- and post-saccadic objects, we controlled task difficulty by varying the magnitude of change (Easy: change in 2 features, Hard: change in 1 feature).

The results of Experiment 2 suggest that presence of a familiar object increased the overall change detection performance (Figure 3). Hence, in line with our first hypothesis, we suggest that object-specific associations reserved for familiar objects improve change detection performance by providing richer object predictions compared to novel objects. In addition, we found an

interaction between blanking and familiarity indicating that the familiar objects benefited more from blanking than novel objects. Therefore, we suggest that blanking improves detection of changes to familiar objects by segregating object-specific predictions more effectively.

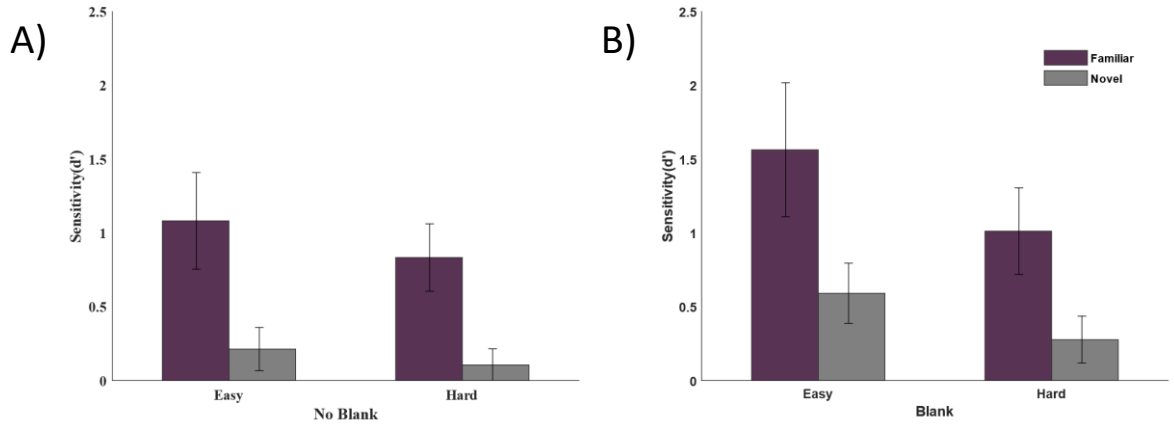



*Figure* 3. Results of Experiment 2. Transsaccadic change performance without (A) or with (B) 300 ms transsaccadic blanking. Detection performance is separately shown for familiar object pairs (purple bars) and novel pairs (gray bars) across task difficulty (x-axes). Error bars show 95% confidence intervals.

The Experiment 2 also allowed us to test whether object-specific associations provide an advantage to familiar peripheral objects by providing more precise foveal predictions over novel objects lacking specific transsaccadic associations. This was not possible in Experiment 1 since familiar peripheral targets always had to be paired with at least one familiar foveal option that could provide foveal to peripheral predictions. In order to test whether object-specific associations provide better peripheral to foveal predictions, we compared trials in Experiment 2 where the pre-saccadic targets were familiar objects to those with novel objects. We found that observers were better at detecting changes to peripheral familiar objects than novel objects, suggesting that object-specific predictions were beneficial.

In conclusion, we showed that the ability of making peripheral-foveal predictions was not limited to familiar objects. Hence, these predictions available for both novel and familiar objects can support visual stability (Valsecchi & Gegenfurtner, 2016) and various visual tasks (Herwig & Schneider, 2014). While peripheral-foveal predictions are not unique to familiar objects, object specific associations acquired through transsaccadic learning provide more precise predictions that are generalizable across the visual field. Thus, we conclude that familiar objects benefit more from transsaccadic feature predictions.

## Study II: *Foveal-feedback follows spatial frequency processing*

Goktepe, N., & Schütz, A. C. (in review). Frequency-specific and periodic masking of peripheral characters by delayed foveal input. *Scientific Reports*.

Previous studies have shown that peripheral discrimination is modulated by delayed foveal input depending on the relationship between the peripheral target(s) and foveal input (Fan et al., 2016; Weldon et al., 2016; Yu & Shim, 2016). However, the exact nature of this relationship is unclear. While discrimination performance benefits by an asynchronous foveal input that is identical to peripheral target (Yu & Shim, 2016), it is impaired by a foveal object regardless of its similarity to the peripheral target (Weldon et al., 2016). Thus, whether the foveal input is competing with or masking the information processed by foveal-feedback is an open question.

Another open question is whether foveal-feedback process is limited to trained complex objects. The evidence so far suggests that foveal-feedback is involved in processing of trained complex objects (e.g., Fan et al., 2016, Weldon et al., 2016; Weldon et al., 2020) or familiar objects (Ramezani et al., 2019) but not involved in processing of motion (Fan et al., 2016) or orientation (Yu & Shim, 2016). Finally, the exact time window of foveal-feedback, defined as SOA, is still debated as the current estimations vary from 117 ms (Weldon et al., 2016, Weldon et al., 2020) to 300 ms (Ramezani et al., 2019). This variation can be partially attributed to sparse sampling of SOAs with long non-overlapping presentation durations. More importantly, a recent study using dense sampling with overlapping SOAs has shown that foveal input impairs foveal-feedback periodically at a rate of 5 Hz (Contemori et al., 2023).

The aim of Study II was to address these open questions and particularly investigate the relationship between the foveal input and peripheral targets. In order to disentangle the masking from competition explanations, we used a peripheral letter discrimination task with asynchronous low, medium, high spatial frequency foveal noise (LSF, MSF, and HSF respectively, see Figure 4). We hypothesized that if the foveal input is masking the information processed by foveal-feedback, then the peripheral discrimination performance should be impaired by the foveal noise depending on the degree of overlap between foveal and peripheral stimuli in spatial frequencies.

On the other hand, if the foveal noise competes with the foveal-feedback information, the peripheral discrimination performance should be similarly impaired regardless of the spatial composition of the foveal noise. In order to investigate if training or familiarity is a precondition for foveal-feedback, we asked one group of observers to discriminate Latin characters that are over-trained by our observers in the course of their lives and asked the remaining observers to discriminate Korean characters, which were novel to them. Hence, if the foveal feedback only supports the processing of familiar stimuli, then the foveal noise should not systematically impair the discrimination of Korean characters.

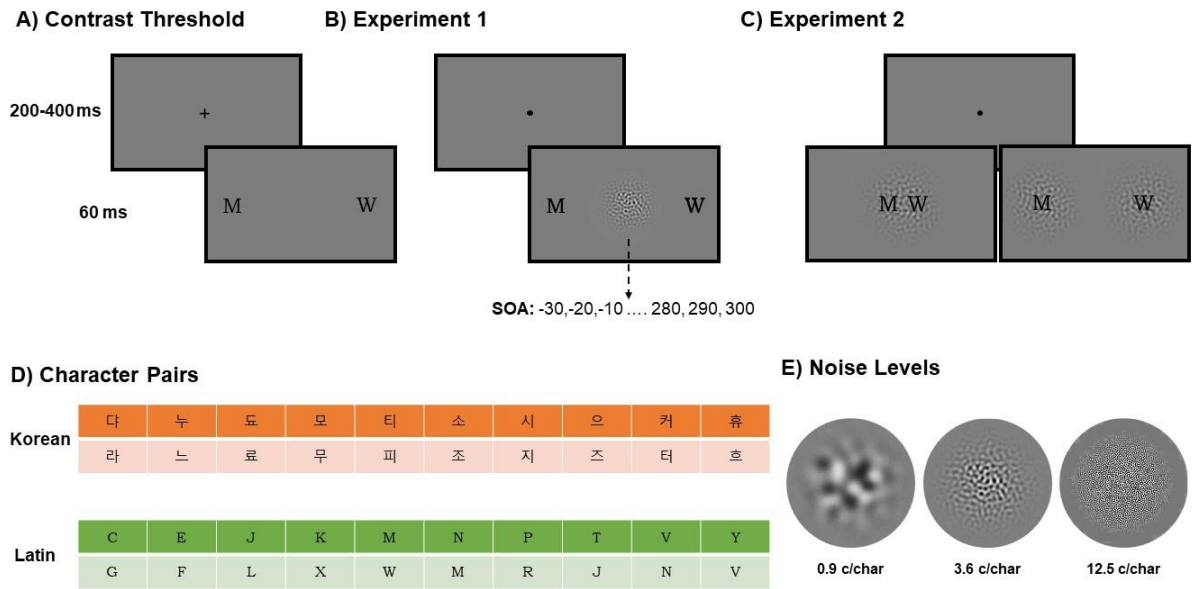



*Figure* 4. General design of Study 2. A) Example trials from the adaptive character contrast threshold measures without foveal noise in Experiment 1. B) Peripheral same-different task with various SOA foveal noise in Experiment 1, C) Adaptive character contrast measures with overlaid spatial frequency. D) Latin and Korean characters. Each column shows pairs for "different" trials. E) The three spatial frequency bands (LSF, MSF, HSF) used in Experiment 1 and 2. For illustration purposes stimuli are not drawn to scale and have higher contrast.

Finally, in order to achieve a high temporal resolution for identifying the time window of foveal-feedback, we densely sampled SOAs between -30 to 290 ms (negative values indicate noise onsets preceding stimulus onset) for 60 ms. As we were interested in the modulation of peripheral discrimination performance, we used an adaptive procedure (QUEST, Watson, 2017) to determine the character contrast yielding 82% discrimination performance for each observer at the beginning of each session. This also allowed us to control for training effects that can improve discrimination performance over the multiple sessions, especially for Korean characters.

In order to assess the relationship between foveal noises and both familiar and novel peripheral characters, we first analyzed the change in discrimination performance across SOAs for noise frequency and character conditions. To this end, we first calculated the moving average performance with an averaging window of 60 ms using the overlapping SOAs for each condition. Thus, we obtained the average discrimination performance as a function of SOA, more specifically time intervals where there was a foveal noise on the screen (Figure 4). Finally, we calculated the change in discrimination performance across SOAs by subtracting the mean discrimination performance from each SOA (Solid lines in Figure 5).

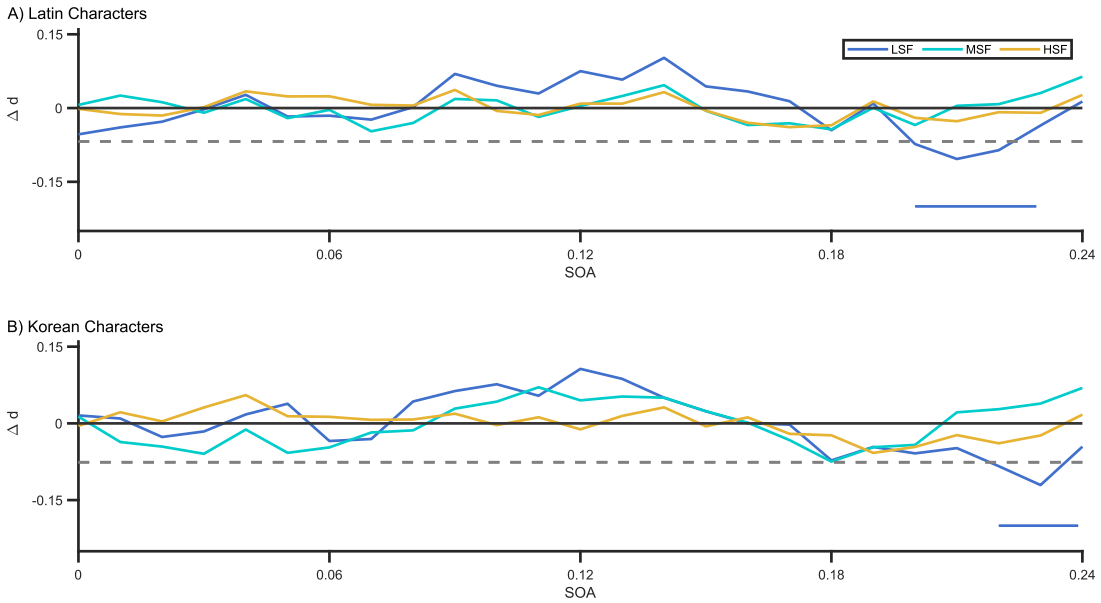



*Figure* 5. Change in the peripheral discrimination performance for Latin (A) and Korean (B) characters for LSF, MSF, and HSF noise (separate lines) across SOA (x-axes). The gray dashed line shows the most conservative critical value across noise conditions. In statistical analyses exact permuted critical values were used. The sperate horizontal lines show SOAs that lead to significant impairment in the respective noise condition.

We used a permutation-based analysis to detect time periods where the foveal noise led to significant impairment of performance. To this end, we created 10000 permuted observers for each real observer by shuffling their responses in each trial. This procedure allowed us to severe any statistical relationship between foveal noise and discrimination performance. Then, we followed the same averaging procedure to obtain 10000 permuted group level discrimination performance. These group level permuted performances form the null distribution for the given noise and character condition. Hence, the bottom 95$^{th}$ percentiles of these distributions are critical values for omnibus tests equivalent to one-sided statistical tests with $\alpha = 0.05$ (i.e., dashed lines in Figure 5).

Therefore, any observed impairment more extreme than these critical values are deemed to be statistically significant impairments of performance.

We observed that the discrimination performances changed almost in unison in foveal noise and character conditions. These changes can be commonly characterized by three distinct periods starting with a period of impairment from 0 to 90 ms SOA. This is to be expected as during this period the presentation of peripheral characters largely overlapped with the foveal noise, thereby foveal noise could have competed with peripheral targets for attentional resources (Fan et al., 2016). Next, the peripheral discrimination performance recovers between 90-170 ms SOAs. Given the oscillatory nature of foveal-feedback, the relative recovery of performance is expected and might be indicating a period where the foveal noise fails to intercept the foveal-feedback processes. Finally, the period between 180-240 is a second period of impairment. Critically, during this period we see a local dip at 180 ms SOA which overlaps with the previous estimations (Fan et al, 2016; Weldon et al., 2016; Weldon et al., 2020). More importantly, LSF noise became more separated from MSF and HSF noises and led to a stronger impairment. We found that LSF noise led to a significant impairment for both Latin (200-220 ms SOA) and Korean characters (220-230 ms SOA). This specific impairment caused by LSF noise was also significantly stronger than MSF noise for Latin characters and HSF noise for both character conditions.

The sinusoidal change in the discrimination performance is in line with the periodic occurrence of foveal-feedback signals as reported previously (Contemori et al, 2023). We analyzed the Fourier spectrum of the observed and permuted data to identify possible behavioral oscillations. We again used the permuted data to generate null distributions but this time for identifying significant peaks in amplitudes across foveal noise and character conditions. For LSF foveal noise condition, we found significant amplitudes peaking around 5 Hz both for Latin and Korean characters (Figure 6). Additionally, in Korean character condition amplitudes for MSF foveal noise significantly peaked around 8 Hz. In line with the above results, the peak amplitudes for LSF foveal noise were statistically stronger than other conditions for Latin characters but not for Korean characters.

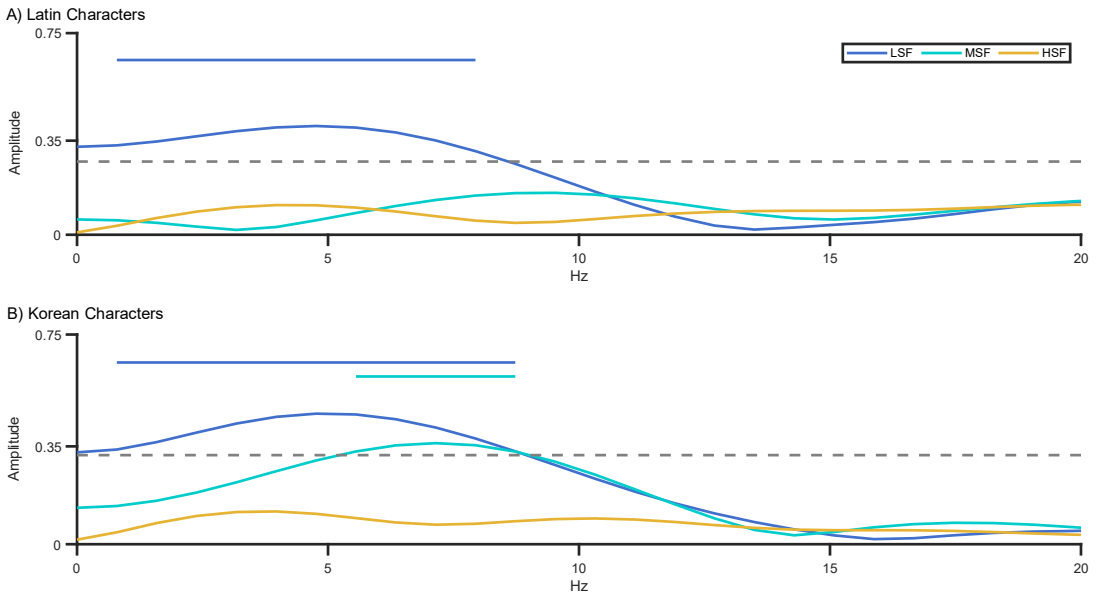



*Figure 6.* Behavioral oscillations analysis for Latin (A) and Korean (B) characters for LSF, MSF, and HSF noise (separate lines). The gray dashed lines show the most conservative critical value is used for demonstration purposes. The exact permuted critical values were used in statistical analyses. The sperate horizontal lines show frequencies with significantly high amplitudes.

We found that the peripheral discrimination of Latin and Korean characters is impaired by delayed LSF foveal noise. This indicates that foveal-feedback is involved in the processing of both familiar and novel peripheral characters. More importantly, the frequency-specific nature of this impairment is more compatible with the masking explanation, suggesting that the foveal noise is masking the information processed by foveal-feedback.

We directly tested this prediction in a follow-up experiment, where observers discriminated peripheral or central Korean and Latin characters embedded in either LSF or MSF noise (Figure 4). We hypothesized that if the LSF noise in Experiment 1 is impairing peripheral discrimination performance by masking the information processed by foveal-feedback, it should be more efficient at masking peripheral characters than MSF noise. In other words, the masking explanation would predict that the discrimination of peripheral characters embedded in LSF noise would be harder than those embedded in MSF noise. To test this hypothesis, we measured the contrast threshold of observers for discriminating noise embedded characters presented at periphery (10°) and fixation (0°). We found that MSF noise was more efficient at masking central Korean and Latin characters, replicating previous studies (e.g., Majaj et al., 2002; Solomon & Pelli, 1994, Figure 7). More importantly, this pattern was reversed for the characters presented in periphery. In line with our

hypothesis, the peripheral characters were better masked by LSF than MSF noise, in line with two previous study (Anderson & Thibos, 1999; Chung et al., 2002). This suggests that the peripheral discrimination is impaired by delayed foveal inputs that can effectively mask the peripheral information processed by foveal-feedback.

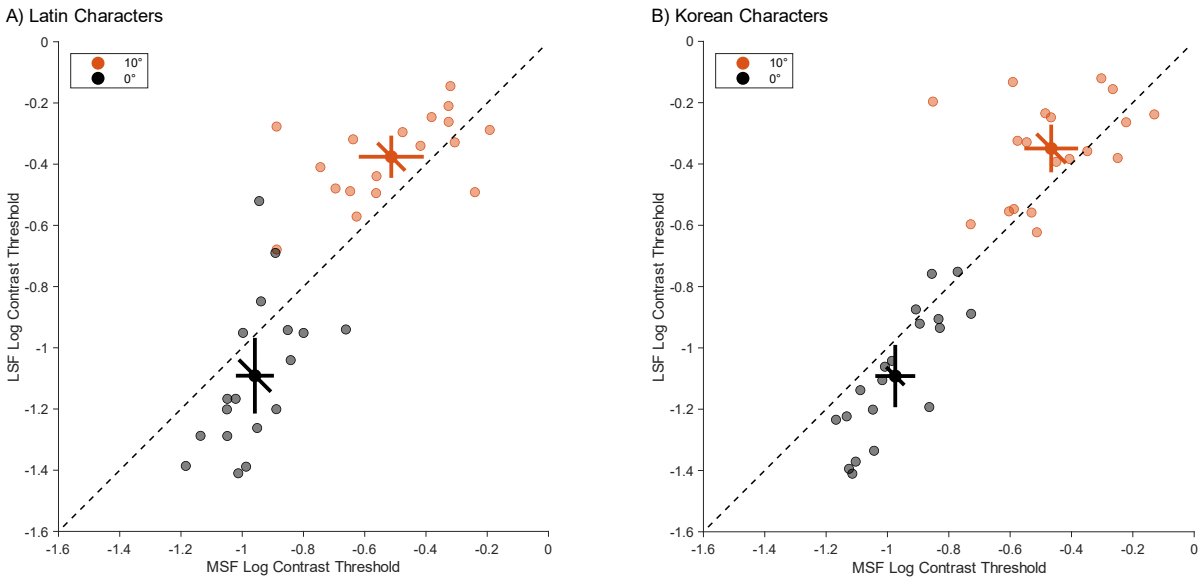



*Figure* 7. Character contrast thresholds yielding 70.7% discrimination performance at 0° (fixation) and 10° for Latin (A) and Korean (B) characters. Mean contrast thresholds for Latin and Korean characters for LSF and MSF are shown with error bars. Error bars show 95% confidence intervals for MSF noise (horizontal), LSF noise (vertical), and their pairwise comparisons (diagonal).

In sum, we suggest that the delayed foveal inputs periodically mask the peripheral object information processed by foveal-feedback. Crucially, we found that the efficacy of foveal masking depended on the spatial frequency overlap with the peripheral information and not to the frequency preference of fovea. We show that foveal-feedback is involved in processing of peripheral characters. Moreover, we found that previous experience is not a prerequisite for foveal-feedback processing as the delayed LSF noise impaired the peripheral discrimination of both familiar and novel objects alike.

## Study III: *Combining peripheral information with other senses*

Goktepe, N., Drewing, K., & Schütz, A. C. (in review). Spatiotemporal congruency modulates weighting of visuotactile information in displacement judgments. *IEEE Transactions on Haptics*.

The need for understanding how vision and other senses increased exponentially with the wide adoption of virtual (VR) and augmented reality (AR) devices capable of rendering multisensory objects and experiences. These multisensory environments are prone to have spatially and temporally discrepant signals due to the lack of temporal and spatial synchronization across devices responsible for rendering underlying unimodal cues (Cornelio et al., 2021). Rendering discrepant unimodal cues may have adverse outcomes such as cybersickness (Viaud-Delmon et al., 2006) or alter virtual experiences, such as by extending peripersonal space (Kilteni et al., 2012). As the spatiotemporal window for integrating sensory information depends on the task and modality (Diederich & Colonius, 2015), determining the spatiotemporal tolerances of virtual systems requires experimenting with various tasks and in different modalities.

Study III was motivated by this need and investigated how peripheral visual displacement judgments were influenced by spatially or temporally incongruent tactile displacement cues. To this end, we built a custom setup that can independently generate visual and tactile cues (Figure 8). Our setup consisted of a visual display attached to a commercially available braille cell (Metec Ag.) mounted on a fully adjustable platform. Using this platform, we made sure that from observers' perspective, the visual stimuli on the display were spatially coinciding with the tactile stimuli located on the back of the display. The tactile stimulation was delivered to the right index finger of observers via individual pins of the braille cells.

The task of observers in all conditions was to report the visual displacement direction of peripheral objects while keeping fixation. Critically, the peripheral stimuli were briefly masked by a full screen random noise before being displaced to their new location. The masking of peripheral stimuli mimics the visual masking naturally occurring during saccades and leads to discrimination

performance similar to well-known saccadic suppression of displacement (Born, 2019; Zimmermann et al., 2013, 2014). In three conditions, the visual displacements were paired by uninformative tactile displacements (i.e., 50% of time towards visual displacement, Figure 8). In temporal incongruency condition, the tactile displacements preceded the visual displacements by 200 ms. In spatial incongruency conditions, visuotactile displacements were temporally congruent but the visual object presented on the opposite side of the screen 50 visual degrees away from the braille cell. In the third visuotactile condition, the visual and tactile displacements were spatially and temporally congruent (i.e., congruency condition).

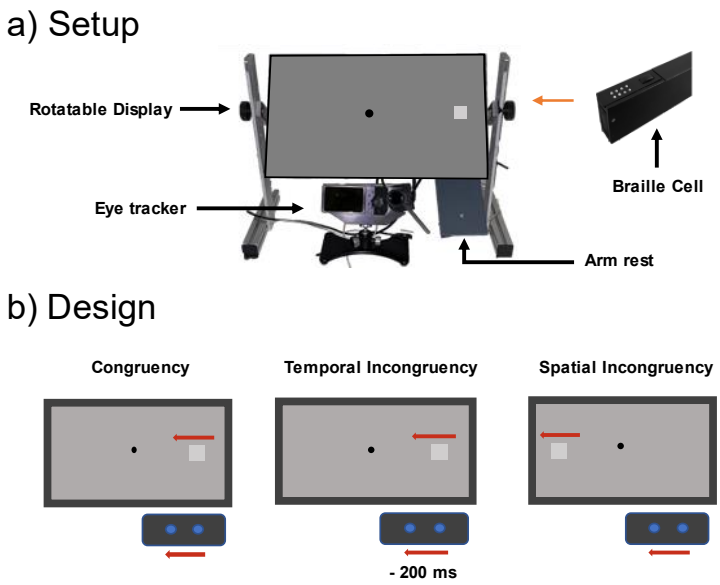



*Figure* 8. Illustration of equipment and the visuotactile conditions in Study III. a) The braille cell (on the right) was directly located under the display and corresponded to the visual stimulus (light square). Observers kept their index finger on the braille device in all conditions. b) Example trials from visuotactile congruency and incongruency conditions. Black dots indicate fixation and gaze. The black box with 2 dots represents the braille cells. Red arrows show the direction of visual and tactile displacement. For illustration purposes stimuli are not drawn to scale and have higher contrast.

There were also two visual only conditions where the observers still reached the braille cell but did not receive any tactile stimulation. The only difference between the two visual only conditions was the location of the visual stimuli relative to the index finger. In one condition the visual stimuli were located directly above the index finger of the participant, whereas in the other condition it was located on the other side of the display as in the spatially incongruent visuotactile condition. This allowed us to control for any bias in visual displacement judgments related to the

location of the visual stimulus (Irwin & Robinson, 2018). Similarly, when the index finger is underneath the visual stimulus, it can act as a landmark and facilitate discrimination like visual landmarks (Deubel et al., 1998). All observers completed all conditions in separate 5 randomly ordered blocks with eye tracking.

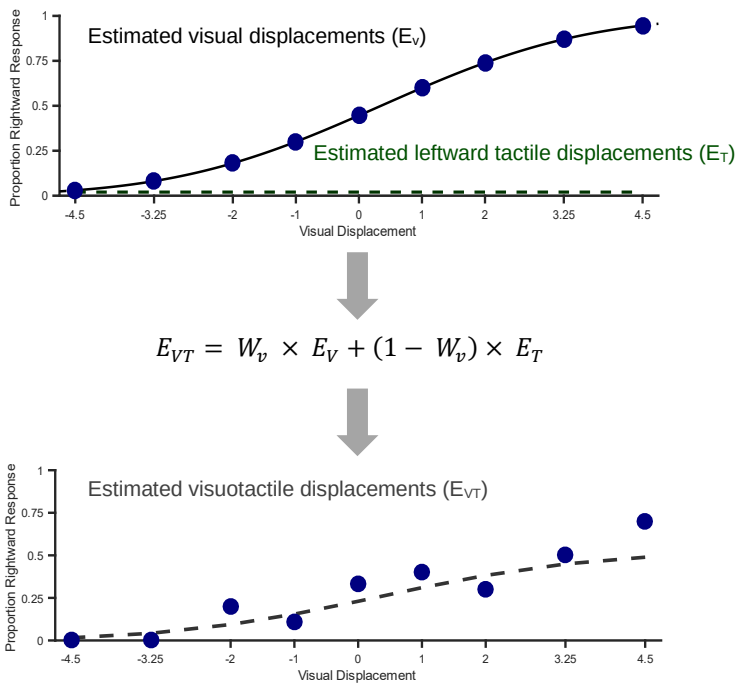



*Figure 9***.** Data and model for one participant. First, the visual estimates were derived by fitting a psychometric function to the visual only condition (solid line). Then the tactile estimations for leftward tactile displacements (green line) and combination weights ($W_v$) were fitted to obtain a model for visuotactile displacements (gray line) with least-squared errors. Blue dots show the observed data in visual only and visuotactile conditions with leftward tactile displacements.

We modelled the visuotactile displacement judgments using unisensory estimates to investigate how spatial and temporal incongruency modulates the weighting of visual and tactile estimates. For visual estimates ($E_V$), we fitted psychometric functions to spatially congruent condition as there were no significant difference between JNDs and PSEs in the non-blanking trials of spatially congruent and incongruent visual only conditions. The tactile estimates ($E_T$) were approximated by a cumulative gaussian with mean of 0 (i.e., no displacement bias) and standard deviation (i.e., JND) as free parameter. Then for each visuotactile conditions, visual and tactile estimates were linearly combined by separate free weighting parameters ($W_V$). $W_V$ is ranging from 0 to 1 where the visuotactile estimates are equal to visual estimates when $W_V = 0$. The modelled

visuotactile estimates ($E_{VT}$) shared the same visual and tactile estimates but had different combination weights (see Figure 9 for an example). From the visuotactile models we extracted the combination weights for congruency, spatial and temporal incongruency conditions. We found that the combination weights in temporal incongruency was positively correlated with the congruency condition (Figure 10). More importantly, we found a significant shift in combination weights favoring tactile information over visual information in general. Thus, when the tactile displacement cues preceded the visual displacement cues, the use of tactile information was overall systematically increased compared to when they were simultaneously presented.

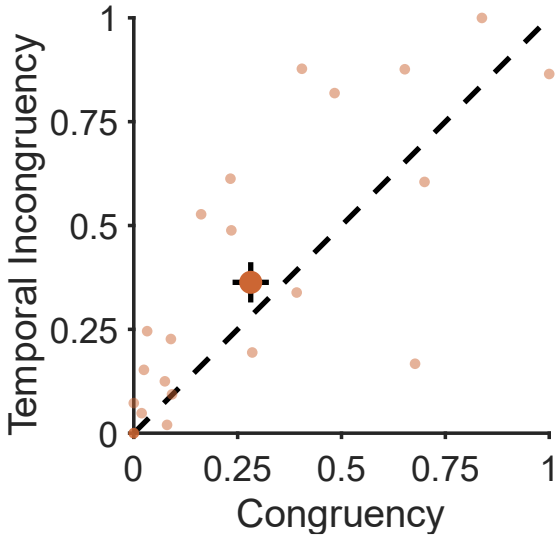



*Figure 10.* Modelled tactile weights for spatiotemporal congruency and temporal incongruency conditions. Mean haptic weights for the temporal incongruency and congruency conditions are presented with error bars showing 95% confidence interval for congruency (horizontal), temporal incongruency (vertical) conditions.

In contrast to temporal incongruency condition, we observed changes in combination weights independent from the congruency condition (Figure 11). We found that the change in combination weights in the spatial incongruency was not correlated with the weights in the congruency condition. In addition, spatial incongruency did not significantly shift the combination weights towards visual or tactile information. Thus, when the tactile and visual displacement were spatially misaligned, the use of tactile information changed differently for different participants hinting inter-individual differences.

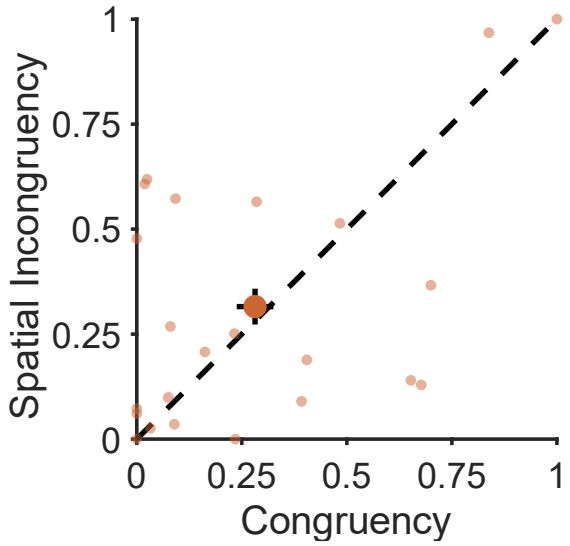



*Figure 11*. Modelled tactile weights for spatiotemporal congruency and spatial incongruency conditions. Mean haptic weights for the spatial incongruency and congruency conditions are presented with error bars showing 95% confidence interval for congruency (horizontal), spatial incongruency (vertical) conditions.

Upon closer inspection we noticed that while combination weights for some participants shifted towards tactile estimates like in the temporal incongruency condition, for the others they drastically changed from favoring one estimate to another. We further analyzed the individual differences hinted for the spatial discrepancy condition by comparing the change in combination weights in temporal and spatial incongruency conditions relative to the congruency condition (Figure 12). To this end, we subtracted the combination weights in the congruency condition from the incongruency conditions. We found that the change in combination weights in the temporal incongruency condition was orthogonal to the those for spatial incongruency condition ($r$ (22) = -0.028, $p$ = .9). In line with this, we observed that almost half of the relative combination weights were shifted to opposite directions in incongruency conditions (Shaded areas in Figure 12). However, an exact McNemar test suggests that this trend was not sufficient to form statistically two distinct groups.

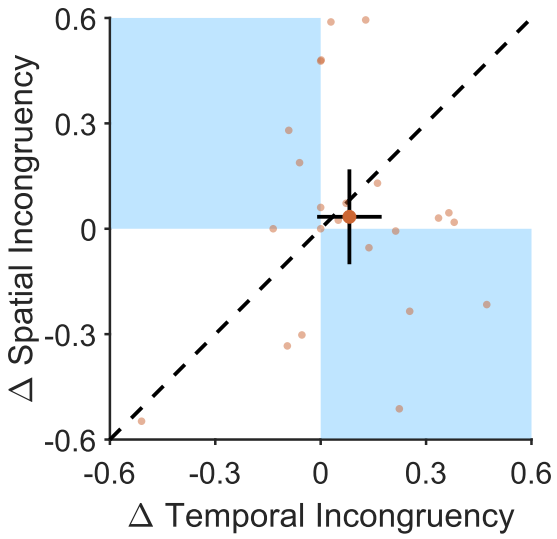



*Figure 12*. Tactile weights in the incongruency conditions relative to the spatiotemporal congruency condition. Error bars show 95% confidence interval for temporally incongruent (horizontal) and spatially incongruent (vertical) visual conditions.

Overall, we found that spatiotemporal incongruency modulates how visual and tactile information are combined for supplementing peripheral vision for detecting small visual displacements. More specifically, while temporal incongruency systematically shifted the combination weights towards unisensory estimates, spatial incongruencies led to different strategies. The changes in combination weights in the presence of a spatial congruency is reminiscent of cue switching where multisensory estimates are driven by either of the unisensory estimates. Here, we observed that combination weights switched between visual and tactile estimates in spatial incongruency condition compared to congruency condition. Thus, if visual estimates had more weight in congruency condition, they gain more weight in temporal incongruency condition whereas tactile estimates became dominant in the spatial incongruency condition. This is a particularly interesting finding as it shows that spatiotemporal congruency is an important determinant of how the sensory information from different sources will be combined. This has direct consequences for virtual reality applications, where rendering multiple sensory cues congruently is a challenge.

## General Discussion

The peripheral retina provides visual information from more than 99% of the visual field at a lower resolution than its counterpart, fovea. While the lower resolution allows sampling information at a lower biological cost, it may not be feasible for performing perceptual tasks especially at higher eccentricities. Thus, the perceptual system resorts to a division of labor where the coarse peripheral information informs future actions and the shortcoming of peripheral vision at performing peripheral tasks are compensated by various mechanisms. The current dissertation project investigated three of these mechanisms that support perceptual judgments at periphery. In three studies, it shows how peripheral perception is supported by transsaccadic learning for generating object predictions, foveal retinotopic cortex for processing peripheral objects with a distinct neural feedback mechanism, and tactile perception for combining supplementary information for informing peripheral judgments. More specifically, Study I found that object specific predictions for familiar objects to be particularly beneficial for peripheral matching or detecting transsaccadic changes compared to general predictions for novel objects that lack object specific predictions. In Study II, we found that foveal-feedback supports the peripheral discrimination of both novel and familiar object. Crucially, we showed that the foveal-feedback is periodically processing peripheral object information which can be masked by incoming foveal information following spatial frequency principles. Finally, in Study III we found that visual displacement judgments are influenced by tactile displacement cues. Critically, we observed that the combination of visual and tactile information modulated depending on the spatiotemporal alignment of visual and tactile information. While these studies provided evidence on how these mechanisms contribute to peripheral perception, some of the remaining questions on these mechanisms are discussed in the following subsections.

### *Peripheral-foveal predictions with and without saccades*

We found that peripheral and foveal object predictions can be generalized to locations other than where the peripheral and foveal associations are acquired. However, it is not clear how the visual system chooses which peripheral and foveal information needs to be associated. As we demonstrated in the training phase of Study I, it is not necessary to explicitly attempt to acquire these associations or perform the same task as the test trial (see also (Valsecchi & Gegenfurtner, 2016). The coupling of peripheral and foveal information for association could be realized by the spatiotemporal correspondence of the to be associated information, which we found to be critical for sensory combination as well as multisensory integration (Ernst & Banks, 2002; Ernst &

Bülthoff, 2004). In daily life associating peripheral and foveal information is argued to be primarily realized by transsacadic learning (Weiss, Schneider, & Herwig, 2014). Therefore, peripheral and foveal object information directly follow each other. More importantly, the remapping of receptive field could establish spatial correspondence between peripheral and foveal information from the same object. Considering that the visual and visuomotor neurons displaying remapping properties are typically found in retinotopically organized structures (for a review see Hall & Colby, 2011), it is reasonable to assume that the spatial correspondence of peripheral and foveal information depends on retinotopic coordinates. Thus, the visual system could use remapping for transsaccadically associating peripheral and foveal object information. However, peripheral and foveal object information can be associated without saccades when an object appears at periphery then reappears at fixation after a small delay (Herwig & Schneider, 2014). Object associations acquired this way are weaker, which underlies the importance of saccades and potentially remapping. On the other hand, they might be hinting at a flexible temporal window for forming object associations. Temporal flexibility might be mandatory for the visual system given that the time between the arrival of pre- and post- saccadic information depends on the eccentricity of the object and saccade amplitude.

Another important question pertaining to the eccentricity of the peripheral object is whether the weighting of peripheral to foveal predictions depends on the eccentricity of the peripheral target. Decreasing visual sensitivity with increasing eccentricity can potentially modulate object predictions at two different levels. First, when the information of an object located at far periphery is associated with the foveal information of the same object for instance with a saccade, the peripheral information might not be sufficient from discriminating that object from another. This would either mean that the fuzzy peripheral object information are not associated or these associations may not be as reliable as the ones acquired from smaller eccentricities. Another consequence of eccentricity could be that the peripheral object information might not be diagnostic enough to trigger a specific foveal prediction, especially if there is competition between predictions from visually similar objects. One way of resolving both kinds of ambiguities could be not associating or predicting object specific information if the peripheral information is not diagnostic. The evidence from transsaccadic integration studies support this notion. The current body of evidence suggests that the visual system holds on to pre- and post-saccadic information and integrates them only when they are sufficiently similar to each other (Atsma et al., 2016; for a review see Stewart et al., 2020). Thus, the visual system might also associate peripheral information

with foveal information when they are sufficiently similar. On the other hand, associating peripheral and foveal object information might be the default mode of the visual system as it is possible to acquire arbitrary object associations even when the peripheral and foveal information are belonging to different complex objects (Osterbrink & Herwig, 2019). Finally, the perceptual estimates can be optimized by weighting available information based on their reliability (e.g., Maximum Likelihood Estimation, Alais & Burr, 2004, Ernst & Banks, 2002) Therefore, a moderate solution can be obtained if the visual system takes into account the reliability of object information when associating or predicting object information. Hence, the weighting of object prediction for making decision would be scaled by the reliability of information that was used to generate that prediction. Thus, it is unclear if and how the visual system associates and predicts object information with unreliable information.

### *Potential limitations of foveal-feedback in daily life*

In Study II, we used delayed foveal noise patches to investigate the informational content and the temporal dynamics of the foveal-feedback. While gaussian noise patches are hardly typical in daily life, the frequency specific masking of foveal-feedback has direct consequences for the utility of foveal-feedback in everyday life. More importantly, with the frequently executed eye movements, peripheral objects are brought to fovea. The interaction of presaccadic and postsaccadic information has been repeatedly shown to alter how we perceive objects (for a review see Stewart, Valsecchi, and Schütz, 2020). The postsaccadic information usually updates presaccadic information by means of transsaccadic integration or overwriting it (Grzeczkowski et al., 2020; Parker & Tas, 2020; Tas et al., 2012, 2021). Does postsaccadic information mask foveal-feedback of presaccadic object information? There is some evidence that foveal-feedback is not involved when a saccade is planned (Fan et al., 2016). The postsaccadic masking might be the reason for why foveal-feedback is not engaged when saccades are involved. Then, another important question is whether the visual system is not engaging foveal-feedback by default, or it has to learn to discard or not to engage the foveal-feedback when it is not reliable.

Apart from postsaccadic masking, the benefit of foveal-feedback might be limited for some tasks or visual scenes. For instance, when searching for an object in clutter, in addition to peripheral-foveal predictions, foveal-feedback could support the visual search by processing peripheral information in foveal retinotopic cortex. However, the processing of peripheral information can be disrupted by the other foveally sampled objects in the scene. Based on our results, the efficiency of such objects at masking foveal-feedback would depend on the degree of

overlap between the spatial frequency composition of the peripheral and foveal stimuli. Thus, the role of foveal-feedback in daily life requires further investigation.

### *Combining sensory information from within and across senses*

While peripheral to foveal and foveal to peripheral predictions can be generated for a variety of simple features or complex objects, the association of peripheral and foveal information is not mandatory after each saccade. In daily life, it is not typical to encounter objects that change their appearance during saccade. Hence, it might have been a reasonable strategy for the visual system to directly associate presaccadic and postsaccadic information. However, the visual system is selectively sensitive to large change transsaccadic changes and associates presaccadic and postsaccadic information based on the discrepancy of the two information (Köller et al., 2020). Presumably, this provides greater flexibility to the visual system to re-calibrate when it is necessary and discard less persistent and discrepant miss percepts. This is also in line with the evidence from transsaccadic integration studies suggesting that the presaccadic and postsaccadic information is not integrated if they are discrepant (Atsma et al., 2016; Demeyer et al., 2010; Laurin et al, 2021).

Similarly, visual information can be combined or integrated with sensory information from other senses. However, unlike combining visual information within the same modality, combining information across senses requires a crossmodal mapping. When there are multiple information present from the same modality, combining or integrating information corresponding to the same object should mainly be limited by the reliability of the sensory information. On the other hand, when there is multiple information from different senses, combining or integrating information across senses is not only limited by the reliability of individual sensory information but also how well this information can be mapped across senses. The crossmodal mapping of sensory information can be realized by multisensory neurons (Alais, Newel, & Mamassian, 2010; Stein & Meredith, 1993), that limits combinations to the sensory information within the same spatiotemporal window. The results of Study III also underline how multisensory combination patterns modulated by the spatiotemporal alignment of the visual and tactile information. In addition to the spatiotemporal alignment, sensory combination and integration could be potentially guided by existing crossmodal correspondences (Spence, 2011). Crossmodal correspondences describe nonarbitrary relationship between perceptual dimensions that are acquired via statistical learning, structural similarity, or semantic mediation (Spence, 2011). These relationships such as between size and pitch (i.e., smaller objects are associated with higher pitches than larger objects.) modulates binding information from different senses (Bien et al., 2012; Orchard-Mills et al., 2016).

Therefore, if and how sensory information will be combined or integrated seem to depend on their correspondence and spatiotemporal alignment in general. However, it is not clear how these factors weighed for determining if two sensory information needs to be combined. For instance, corresponding audiovisual stimuli can still be integrated despite being spatially misaligned (Bien et al. 2012). Similarly, temporal alignment can be a more important determinant factor for multisensory integration for performing non-spatial tasks (Spence, 2013).

Finally, there seems to be differences in how sensory information from different senses are combined with peripheral or foveal information. The most striking difference between peripheral and foveal information could be their reliability as the fovea offers higher resolution. Hence, according to MLE, peripheral information should have a lower weighting in integration than foveal information in general. However, this may not be the whole picture as the foveal information can be preferred over peripheral information even it is less reliable (e.g., (Gloriani & Schütz, 2019). Astonishingly, audiovisual and visuotactile integration with peripheral and foveal information can be drastically different. For instance, Chen et al. (Chen et al., 2017) found that while fission illusion (i.e., a visual flash perceived to be double when paired with double tap or beeps) is stronger in periphery than fovea, the fusion illusion (i.e., two flashes perceived as one when paired with a single tap or beep) is stronger at fovea than periphery. Considering the high similarity of illusions in terms of visual reliability, these results might indicate a difference in how foveal and peripheral information are integrated beyond differences in visual sensitivity.

In conclusion, the results of Study I, II, and III clearly demonstrate the intricate interactions between peripheral vision and other perceptual mechanisms. These interactions portray a division of labor within and across sensory modalities for optimizing perceptual estimates and decisions with the available sensory information. While periphery has a limited visual resolution compared to fovea, it provides coarse information from a large visual field at a relatively low biological cost. Thus, it provides essential summary statistics for planning future actions and perception in general. On the other hand, limited resolution especially at the far periphery may not always be sufficient for discriminating visual targets or making decisions. Thus, the perceptual system uses transsaccadic predictions from experience, additional processing resources normally reserved for other perceptual systems, or available supplementary information from other senses to improve the reliability of perceptual estimates for performing peripheral tasks.

## References

Alais, D., & Burr, D. (2004). The Ventriloquist Effect Results from Near-Optimal Bimodal Integration. *Current Biology*, *14*(3), 257–262. https://doi.org/10.1016/j.cub.2004.01.029

Alais, D., Newell, F., & Mamassian, P. (2010). Multisensory Processing in Review: From Physiology to Behaviour. *Seeing and Perceiving*, *23*(1), 3–38. https://doi.org/10.1163/187847510X488603

Anderson, R. S., & Thibos, L. N. (1999). Relationship between acuity for gratings and for tumbling-E letters in peripheral vision. *Journal of the Optical Society of America A*, *16*(10), 2321. https://doi.org/10.1364/JOSAA.16.002321

Atsma, J., Maij, F., Koppen, M., Irwin, D. E., & Medendorp, W. P. (2016). Causal Inference for Spatial Constancy across Saccades. *PLOS Computational Biology*, *12*(3), e1004766. https://doi.org/10.1371/journal.pcbi.1004766

Benson, N. C., Kupers, E. R., Barbot, A., Carrasco, M., & Winawer, J. (2021). Cortical magnification in human visual cortex parallels task performance around the visual field. *ELife*, *10*, e67685. https://doi.org/10.7554/eLife.67685

Bien, N., Ten Oever, S., Goebel, R., & Sack, A. T. (2012). The sound of size. *NeuroImage*, *59*(1), 663–672. https://doi.org/10.1016/j.neuroimage.2011.06.095

Born, S. (2019). Saccadic Suppression of Displacement Does Not Reflect a Saccade-Specific Bias to Assume Stability. *Vision*, *3*(4), 49. https://doi.org/10.3390/vision3040049

Brown, F. E., Sutton, J., Yuen, H. M., Green, D., Van Dorn, S., Braun, T., Cree, A. J., Russell, S. R., & Lotery, A. J. (2019). A novel, wearable, electronic visual aid to assist those with reduced peripheral vision. *PLOS ONE*, *14*(10), e0223755. https://doi.org/10.1371/journal.pone.0223755

Cámara, C., López-Moliner, J., Brenner, E., & De La Malla, C. (2020). Looking away from a moving target does not disrupt the way in which the movement toward the target is guided. *Journal of Vision*, *20*(5), 5. https://doi.org/10.1167/jov.20.5.5

Cavanagh, P., Hunt, A. R., Afraz, A., & Rolfs, M. (2010). Visual stability based on remapping of attention pointers. *Trends in Cognitive Sciences*, *14*(4), 147–153. https://doi.org/10.1016/j.tics.2010.01.007

Chambers, C. D., Allen, C. P. G., Maizey, L., & Williams, M. A. (2013). Is delayed foveal feedback critical for extra-foveal perception? *Cortex*, *49*(1), 327–335. https://doi.org/10.1016/j.cortex.2012.03.007

Chen, Y.-C., Maurer, D., Lewis, T. L., Spence, C., & Shore, D. I. (2017). Central–peripheral differences in audiovisual and visuotactile event perception. *Attention, Perception, & Psychophysics*, *79*(8), 2552–2563. https://doi.org/10.3758/s13414-017-1396-4

Chung, S. T. L., Legge, G. E., & Tjan, B. S. (2002). Spatial-frequency characteristics of letter identification in central and peripheral vision. *Vision Research*, *42*(18), 2137–2152. https://doi.org/10.1016/S0042-6989(02)00092-5

Contemori, G., Oletto, C. M., Cessa, R., Marini, E., Ronconi, L., Battaglini, L., & Bertamini, M. (2022). Investigating the role of the foveal cortex in peripheral object discrimination. *Scientific Reports*, *12*(1), 19952. https://doi.org/10.1038/s41598-022-23720-w

Cornelio, P., Velasco, C., & Obrist, M. (2021). Multisensory Integration as per Technological Advances: A Review. *Frontiers in Neuroscience*, *15*, 652611. https://doi.org/10.3389/fnins.2021.652611

Curcio, C. A., Sloan, K. R., Kalina, R. E., & Hendrickson, A. E. (1990). Human photoreceptor topography. *The Journal of Comparative Neurology*, *292*(4), 497–523. https://doi.org/10.1002/cne.902920402

Diederich, A., & Colonius, H. (2015). The time window of multisensory integration: Relating reaction times and judgments of temporal order. *Psychological Review*, *122*(2), 232–241. https://doi.org/10.1037/a0038696

Demeyer, M., De Graef, P., Wagemans, J., & Verfaillie, K. (2010). Object form discontinuity facilitates displacement discrimination across saccades. *Journal of Vision*, *10*(6), 17–17. https://doi.org/10.1167/10.6.17

Deubel, H., Bridgeman, B., & Schneider, W. X. (1998). Immediate post-saccadic information mediates space constancy. *Vision Research*, *38*(20), 3147–3159. https://doi.org/10.1016/S0042-6989(98)00048-0

Deubel, H., Schneider, W. X., & Bridgeman, B. (1996). Postsaccadic target blanking prevents saccadic suppression of image displacement. *Vision Research*, *36*(7), 985–996. https://doi.org/10.1016/0042-6989(95)00203-0

Duhamel, J.-R., Colby, C. L., & Goldberg, M. E. (1992). The Updating of the Representation of Visual Space in Parietal Cortex by Intended Eye Movements. *Science*, *255*(5040), 90–92. https://doi.org/10.1126/science.1553535

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429–433. https://doi.org/10.1038/415429a

Ernst, M. O., & Bülthoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences*, *8*(4), 162–169. https://doi.org/10.1016/j.tics.2004.02.002

Fan, X., Wang, L., Shao, H., Kersten, D., & He, S. (2016). Temporally flexible feedback signal to foveal cortex for peripheral object recognition. *Proceedings of the National Academy of Sciences*, *113*(41), 11627–11632. https://doi.org/10.1073/pnas.1606137113

Ferree, C. E., & Rand, G. (1912). The spatial values of the visual field immediately surrounding the blind spot and the question of the associative filling in of the blind spot. *American*

*Journal of Physiology-Legacy Content*, *29*(4), 398–417. https://doi.org/10.1152/ajplegacy.1912.29.4.398

Fooken, J., Kreyenmeier, P., & Spering, M. (2021). The role of eye movements in manual interception: A mini-review. *Vision Research*, *183*, 81–90. https://doi.org/10.1016/j.visres.2021.02.007

Gepshtein, S., Burge, J., Ernst, M. O., & Banks, M. S. (2005). The combination of vision and touch depends on spatial proximity. *Journal of Vision*, *5*(11), 7. https://doi.org/10.1167/5.11.7

Gloriani, A. H., & Schütz, A. C. (2019). Humans Trust Central Vision More Than Peripheral Vision Even in the Dark. *Current Biology*, *29*(7), 1206-1210.e4. https://doi.org/10.1016/j.cub.2019.02.023

Gottlieb, J. P., Kusunoki, M., & Goldberg, M. E. (1998). The representation of visual salience in monkey parietal cortex. *Nature*, *391*(6666), 481–484. https://doi.org/10.1038/35135

Grzeczkowski, L., Van Leeuwen, J., Belopolsky, A. V., & Deubel, H. (2020). Spatiotopic and saccade-specific transsaccadic memory for object detail. *Journal of Vision*, *20*(7), 2. https://doi.org/10.1167/jov.20.7.2

Herwig, A. (2015). Transsaccadic integration and perceptual continuity. *Journal of Vision*, *15*(16), 7. https://doi.org/10.1167/15.16.7

Herwig, A., & Schneider, W. X. (2014). Predicting object features across saccades: Evidence from object recognition and visual search. *Journal of Experimental Psychology: General*, *143*(5), 1903–1922. https://doi.org/10.1037/a0036781

Herwig, A., Weiß, K., & Schneider, W. X. (2018). Feature prediction across eye movements is location specific and based on retinotopic coordinates. *Journal of Vision*, *18*(8), 13. https://doi.org/10.1167/18.8.13

Irwin, D. E., & Robinson, M. M. (2018). How post-saccadic target blanking affects the detection of stimulus displacements across saccades. *Vision Research*, *142*, 11–19. https://doi.org/10.1016/j.visres.2017.09.004

Jagini, K. K. (2021). Temporal binding in multisensory and motor-sensory contexts: Toward a unified model. *Frontiers in Human Neuroscience*, *15*. https://doi.org/10.3389/fnhum.2021.629437

Jicol, C., Lloyd-Esenkaya, T., Proulx, M. J., Lange-Smith, S., Scheller, M., O'Neill, E., & Petrini, K. (2020). Efficiency of Sensory Substitution Devices Alone and in Combination With Self-Motion for Spatial Navigation in Sighted and Visually Impaired. *Frontiers in Psychology*, *11*, 1443. https://doi.org/10.3389/fpsyg.2020.01443

Kannape, O. A., Smith, E. J. T., Moseley, P., Roy, M. P., & Lenggenhager, B. (2019). Experimentally induced limb-disownership in mixed reality. *Neuropsychologia*, *124*, 161–170. https://doi.org/10.1016/j.neuropsychologia.2018.12.014

Kilteni, K., Normand, J.-M., Sanchez-Vives, M. V., & Slater, M. (2012). Extending Body Space in Immersive Virtual Reality: A Very Long Arm Illusion. *PLoS ONE*, *7*(7), e40867. https://doi.org/10.1371/journal.pone.0040867

Köller, C. P., Poth, C. H., & Herwig, A. (2020). Object discrepancy modulates feature prediction across eye movements. *Psychological Research*, *84*(1), 231–244. https://doi.org/10.1007/s00426-018-0988-5

Kusunoki, M., & Goldberg, M. E. (2003). The Time Course of Perisaccadic Receptive Field Shifts in the Lateral Intraparietal Area of the Monkey. *Journal of Neurophysiology*, *89*(3), 1519–1527. https://doi.org/10.1152/jn.00519.2002

Kwon, M., & Liu, R. (2019). Linkage between retinal ganglion cell density and the nonuniform spatial integration across the visual field. *Proceedings of the National Academy of Sciences*, *116*(9), 3827–3836. https://doi.org/10.1073/pnas.1817076116

Lamme, V. A. F., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences*, *23*(11), 571–579. https://doi.org/10.1016/S0166-2236(00)01657-X

Lisi, M., Solomon, J. A., & Morgan, M. J. (2019). Gain control of saccadic eye movements is probabilistic. *Proceedings of the National Academy of Sciences*, *116*(32), 16137–16142. https://doi.org/10.1073/pnas.1901963116

Ludwig, C. J. H., Davies, J. R., & Eckstein, M. P. (2014). Foveal analysis and peripheral selection during active visual sampling. *Proceedings of the National Academy of Sciences*, *111*(2). https://doi.org/10.1073/pnas.1313553111

Majaj, N. J., Pelli, D. G., Kurshan, P., & Palomares, M. (2002). The role of spatial frequency channels in letter identification. *Vision Research*, *42*(9), 1165–1184. https://doi.org/10.1016/S0042-6989(02)00045-7

Malpeli, J. G., & Baker, F. H. (1975). The representation of the visual field in the lateral geniculate nucleus of Macaca mulatta. *The Journal of Comparative Neurology*, *161*(4), 569–594. https://doi.org/10.1002/cne.901610407

Melcher, D. (2005). Spatiotopic Transfer of Visual-Form Adaptation across Saccadic Eye Movements. *Current Biology*, *15*(19), 1745–1748. https://doi.org/10.1016/j.cub.2005.08.044

Melcher, D. (2007). Predictive remapping of visual features precedes saccadic eye movements. *Nature Neuroscience*, *10*(7), 903–907. https://doi.org/10.1038/nn1917

Melcher, D. (2011). Visual stability. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1564), 468–475. https://doi.org/10.1098/rstb.2010.0277

Meredith, M. A., & Stein, B. E. (1996). Spatial determinants of multisensory integration in cat superior colliculus neurons. *Journal of Neurophysiology*, *75*(5), 1843–1857. https://doi.org/10.1152/jn.1996.75.5.1843

Meredith, M., Nemitz, J., & Stein, B. (1987). Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. *The Journal of Neuroscience*, *7*(10), 3215–3229. https://doi.org/10.1523/JNEUROSCI.07-10-03215.1987

Moeller, B., & Pfister, R. (2022). Ideomotor learning: Time to generalize a longstanding principle. *Neuroscience & Biobehavioral Reviews*, *140*, 104782. https://doi.org/10.1016/j.neubiorev.2022.104782

Neugebauer, A., Rifai, K., Getzlaff, M., & Wahl, S. (2020). Navigation aid for blind persons by visual-to-auditory sensory substitution: A pilot study. *PLOS ONE*, *15*(8), e0237344. https://doi.org/10.1371/journal.pone.0237344

Oletto, C. M., Contemori, G., Bertamini, M., & Battaglini, L. (2022). The Role of Foveal Cortex in Discriminating Peripheral Stimuli: The Sketchpad Hypothesis. *NeuroSci*, *4*(1), 9–17. https://doi.org/10.3390/neurosci4010002

Op De Beeck, H. P., Baker, C. I., DiCarlo, J. J., & Kanwisher, N. G. (2006). Discrimination Training Alters Object Representations in Human Extrastriate Cortex. *The Journal of Neuroscience*, *26*(50), 13025–13036. https://doi.org/10.1523/JNEUROSCI.2481-06.2006

Orchard-Mills, E., Van Der Burg, E., & Alais, D. (2016). Crossmodal Correspondence Between Auditory Pitch and Visual Elevation Affects Temporal Ventriloquism. *Perception*, *45*(4), 409–424. https://doi.org/10.1177/0301006615622320

Osterbrink, C., & Herwig, A. (2019). Transsaccadic prediction of real-world objects. *Journal of Vision*, *19*(10), 237c. https://doi.org/10.1167/19.10.237c

Osterbrink, C., & Herwig, A. (2021). Prediction of complex stimuli across saccades. *Journal of Vision*, *21*(2), 10. https://doi.org/10.1167/jov.21.2.10

Osterbrink, C., & Herwig, A. (2023). What determines location specificity or generalization of transsaccadic learning? *Journal of Vision*, *23*(1), 8. https://doi.org/10.1167/jov.23.1.8

Österberg GA. Topography of the layer of rods and cones in the human retina. *Acta Ophthalmol (Copenh)*. 1935; 6: 1–102.

Parker, J., & Tas, A. C. (2020). Transsaccadic Updating: Evidence for Overwriting of Color Information. *Journal of Vision*, *20*(11), 1327. https://doi.org/10.1167/jov.20.11.1327

Pfister, R. (2019). Effect-based action control with body-related effects: Implications for empirical approaches to ideomotor action control. *Psychological Review*, *126*(1), 153–161. https://doi.org/10.1037/rev0000140

Pointer, J. S., & Hess, R. F. (1989). The contrast sensitivity gradient across the human visual field: With emphasis on the low spatial frequency range. *Vision Research*, *29*(9), 1133–1151. https://doi.org/10.1016/0042-6989(89)90061-8

Purves, D., Augustine, G. J., Fitzpatrick, D., Katz, L. C., LaMantia, A.-S., McNamara, J. O., & Williams, S. M. (2004). The Retinotopic Representation of the Visual Field. In *Neuroscience: Third Edition* (pp. 263–267). essay, Sunderland, MA: Sinauer Associates.

Ramezani, F., Kheradpisheh, S. R., Thorpe, S. J., & Ghodrati, M. (2019). Object categorization in visual periphery is modulated by delayed foveal noise. *Journal of Vision*, *19*(9), 1. https://doi.org/10.1167/19.9.1

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372–422. https://doi.org/10.1037/0033-2909.124.3.372

Robson, J. G., & Graham, N. (1981). Probability summation and regional variation in contrast sensitivity across the visual field. *Vision Research*, *21*(3), 409–418. https://doi.org/10.1016/0042-6989(81)90169-3

Rosenholtz, R. (2016). Capabilities and Limitations of Peripheral Vision. *Annual Review of Vision Science*, *2*(1), 437–457. https://doi.org/10.1146/annurev-vision-082114-035733

Rosenholtz, R., Huang, J., Raj, A., Balas, B. J., & Ilie, L. (2012). A summary statistic representation in peripheral vision explains visual search. *Journal of Vision*, *12*(4), 14–14. https://doi.org/10.1167/12.4.14

Scalvini, F., Bordeau, C., Ambard, M., Migniot, C., Argon, S., & Dubois, J. (2022). Visual-auditory substitution device for indoor navigation based on fast visual marker detection. *2022 16th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*, 259–266. https://doi.org/10.1109/SITIS57111.2022.00029

Schütz, A. C., Trommershäuser, J., & Gegenfurtner, K. R. (2012). Dynamic integration of information about salience and value for saccadic eye movements. *Proceedings of the National Academy of Sciences*, *109*(19), 7547–7552. https://doi.org/10.1073/pnas.1115638109

Shin, Y. K., Proctor, R. W., & Capaldi, E. J. (2010). A review of contemporary ideomotor theory. *Psychological Bulletin*, *136*(6), 943–974. https://doi.org/10.1037/a0020541

Slutsky, D. A., & Recanzone, G. H. (2001). Temporal and spatial dependency of the ventriloquism effect: *Neuroreport*, *12*(1), 7–10. https://doi.org/10.1097/00001756-200101220-00009

Solomon, J. A., & Pelli, D. G. (1994). The visual filter mediating letter identification. *Nature*, *369*(6479), 395–397. https://doi.org/10.1038/369395a0

Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics*, *73*(4), 971–995. https://doi.org/10.3758/s13414-010-0073-7

Spence, C. (2013). Just how important is spatial coincidence to multisensory integration? Evaluating the spatial rule: Spatial coincidence and multisensory integration. *Annals of the New York Academy of Sciences*, *1296*(1), 31–49. https://doi.org/10.1111/nyas.12121

Stewart, E. E. M., Valsecchi, M., & Schütz, A. C. (2020). A review of interactions between peripheral and foveal vision. *Journal of Vision*, *20*(12), 2. https://doi.org/10.1167/jov.20.12.2

Takano, S., Matsumiya, K., Tseng, C., Kuriki, I., Deubel, H., & Shioiri, S. (2020). Displacement detection is suppressed by the post-saccadic stimulus. *Scientific Reports*, *10*(1), 9273. https://doi.org/10.1038/s41598-020-66216-1

Tas, A. C., Moore, C. M., & Hollingworth, A. (2012). An object-mediated updating account of insensitivity to transsaccadic change. *Journal of Vision*, *12*(11), 18–18. https://doi.org/10.1167/12.11.18

Tas, A. C., Mordkoff, J. T., & Hollingworth, A. (2021). Object-mediated overwriting across saccades. *Journal of Vision*, *21*(2), 3. https://doi.org/10.1167/jov.21.2.3

Umeno, M. M., & Goldberg, M. E. (2001). Spatial Processing in the Monkey Frontal Eye Field. II. Memory Responses. *Journal of Neurophysiology*, *86*(5), 2344–2352. https://doi.org/10.1152/jn.2001.86.5.2344

Valsecchi, M., & Gegenfurtner, K. R. (2016). Dynamic Re-calibration of Perceived Size in Fovea and Periphery through Predictable Size Changes. *Current Biology*, *26*(1), 59–63. https://doi.org/10.1016/j.cub.2015.10.067

Van Polanen, V., Tibold, R., Nuruki, A., & Davare, M. (2019). Visual delay affects force scaling and weight perception during object lifting in virtual reality. *Journal of Neurophysiology*, *121*(4), 1398–1409. https://doi.org/10.1152/jn.00396.2018

Vater, C., Wolfe, B., & Rosenholtz, R. (2022). Peripheral vision in real-world tasks: A systematic review. *Psychonomic Bulletin & Review*, *29*(5), 1531–1557. https://doi.org/10.3758/s13423-022-02117-w

Vaziri, S., Diedrichsen, J., & Shadmehr, R. (2006). Why Does the Brain Predict Sensory Consequences of Oculomotor Commands? Optimal Integration of the Predicted and the Actual Sensory Feedback. *The Journal of Neuroscience*, *26*(16), 4188–4197. https://doi.org/10.1523/JNEUROSCI.4747-05.2006

Viaud-Delmon, I., Warusfel, O., Seguelas, A., Rio, E., & Jouvent, R. (2006). High sensitivity to multisensory conflicts in agoraphobia exhibited by virtual reality. *European Psychiatry*, *21*(7), 501–508. https://doi.org/10.1016/j.eurpsy.2004.10.004

Virsu, V., & Rovamo, J. (1979). Visual resolution, contrast sensitivity, and the cortical magnification factor. *Experimental Brain Research*, *37*(3). https://doi.org/10.1007/BF00236818

Walker, M. F., Fitzgibbon, E. J., & Goldberg, M. E. (1995). Neurons in the monkey superior colliculus predict the visual result of impending saccadic eye movements. *Journal of Neurophysiology*, *73*(5), 1988–2003. https://doi.org/10.1152/jn.1995.73.5.1988

Watanabe, K. (1958). On the Achromatic Zone around the Spot of Mariotte. *The Tohoku Journal f Experimental Medicine*, *67*, 235–243.

Watson, A. B. (2017). QUEST+: A general multidimensional Bayesian adaptive psychometric method. *Journal of Vision*, *17*(3), 10. https://doi.org/10.1167/17.3.10

Weiß, K., Schneider, W. X., & Herwig, A. (2015). A "blanking effect" for surface features: Transsaccadic spatial-frequency discrimination is improved by postsaccadic blanking. *Attention, Perception, & Psychophysics*, *77*(5), 1500–1506. https://doi.org/10.3758/s13414-015-0926-1

Weldon, K. B., Rich, A. N., Woolgar, A., & Williams, M. A. (2016). Disruption of Foveal Space Impairs Discrimination of Peripheral Objects. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00699

Weldon, K. B., Woolgar, A., Rich, A. N., & Williams, M. A. (2020). Late disruption of central visual field disrupts peripheral perception of form and color. *PLOS ONE*, *15*(1), e0219725. https://doi.org/10.1371/journal.pone.0219725

Weyand, T. G. (2016). The multifunctional lateral geniculate nucleus. *Reviews in the Neurosciences*, *27*(2), 135–157. https://doi.org/10.1515/revneuro-2015-0018

Williams, M. A., Baker, C. I., Op De Beeck, H. P., Mok Shim, W., Dang, S., Triantafyllou, C., & Kanwisher, N. (2008). Feedback of visual object information to foveal retinotopic cortex. *Nature Neuroscience*, *11*(12), 1439–1445. https://doi.org/10.1038/nn.2218

Wolf, C., Wagner, I., & Schütz, A. C. (2019). Competition between salience and informational value for saccade adaptation. *Journal of Vision*, *19*(14), 26. https://doi.org/10.1167/19.14.26

Yu, Q., & Shim, W. M. (2016). Modulating foveal representation can influence visual discrimination in the periphery. *Journal of Vision*, *16*(3), 15. https://doi.org/10.1167/16.3.15

Zele, A. J., & Cao, D. (2015). Vision under mesopic and scotopic illumination. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.01594

Zimmermann, E., Born, S., Fink, G. R., & Cavanagh, P. (2014). Masking produces compression of space and time in the absence of eye movements. *Journal of Neurophysiology*, *112*(12), 3066–3076. https://doi.org/10.1152/jn.00156.2014

Zimmermann, E., Morrone, M. C., & Burr, D. C. (2013). Spatial Position Information Accumulates Steadily over Time. *The Journal of Neuroscience*, *33*(47), 18396–18401. https://doi.org/10.1523/JNEUROSCI.1864-13.2013

Zirnsak, M., & Moore, T. (2014). Saccades and shifting receptive fields: Anticipating consequences or selecting targets? *Trends in Cognitive Sciences*, *18*(12), 621–628. https://doi.org/10.1016/j.tics.2014.10.002

# Appendix

## *Study I*

# Familiar objects benefit more from transsaccadic feature predictions

Nedim Goktepe[1] &Alexander C. Schütz[1,2]

Accepted: 20 December 2022
# The Author(s) 2023

Abstract
The transsaccadic feature prediction mechanism associates peripheral and foveal information belonging to the same object to make predictions about how an object seen in the periphery would appear in the fovea or vice versa. It is unclear if such transsaccadic predictions require experience with the object such that only familiar objects benefit from this mechanism by virtue of having peripheral-foveal associations. In two experiments, we tested whether familiar objects have an advantage over novel objects in peripheral-foveal matching and transsaccadic change detection tasks. In both experiments, observers were unknowingly familiarized with a small set of stimuli by completing a sham orientation change detection task. In the first experiment, observers subsequently performed a peripheral-foveal matching task, where they needed to pick the foveal test object that matched a briefly presented peripheral target. In the second experiment, observers subsequently performed a transsaccadic object change detection task where a peripheral target was exchanged or not exchanged with another target after the saccade, either immediately or after a 300-ms blank period. We found an advantage of familiar objects over novel objects in both experiments. While foveal-peripheral associations explained the familiarity effect in the matching task of the first experiment, the second experiment provided evidence for the advantage of peripheral-foveal associations in transsaccadic object change detection. Introducing a postsaccadic blank improved change detection performance in general but more for familiar than for novel objects. We conclude that familiar objects benefit from additional object-specific predictions.

Keywords Saccades · Transsaccadic prediction · Perceptual learning · Object recognition · Visual stability

## Introduction

The visual system on average executes three to four saccades per second to reposition the fovea to sample high-resolution information about objects or locations of interest in the visual scene (Rayner, 1998). Thus, by bringing an object of interest to the fovea, it samples information at two different resolutions about that object: Prior to a saccade, the visual system samples information in low resolution with peripheral vision. When gaze is shifted to the target object, it samples high-resolution information with foveal vision. This leads to the fundamental question about how the visual system achieves perceptual stability despite the large differences in object information from the peripheral and foveal visual field (for reviews, see Mathôt & Theeuwes, 2011; Melcher, 2011; Rolfs, 2015; Stewart et al., 2020; Wurtz, 2008; Wurtz, 2018). Different mechanisms might contribute to the seamless perception across saccades.

Recent evidence suggests that the visual system can optimally integrate peripheral and foveal information across saccades (e.g., Ganmor et al., 2015; Hübner & Schütz, 2017; Stewart & Schütz, 2019a; Wolf & Schütz, 2015; for a review, see Stewart et al., 2020). However, it is not always adequate or possible to integrate peripheral and foveal information. When the sampled information from the periphery and the fovea is discrepant (e.g., when the saccade target is displaced or exchanged with another object), the visual system can segregate pre- and postsaccadic information (e.g., Atsma et al., 2016; Demeyer et al., 2010; Laurin et al., 2021; Tas et al., 2012; Tas et al., 2021). Furthermore, in some cases, the visual system samples object information with only peripheral vision (Treisman, 1986), and visual search is surprisingly unaffected by blocking foveal vision (David et al., 2021; Nuthmann, 2014; Nuthmann & Canas-Bajo, 2022). According to the transsaccadic feature prediction mechanism (Herwig & Schneider, 2014), object recognition and visual search is supported by predictions based on previous associations of

* Nedim Goktepe
goektepe@staff.uni-marburg.de

1 AG Allgemeine und Biologische Psychologie, Philipps-Universität Marburg, Marburg, Germany

2 Center for Mind, Brain and Behavior, Philipps-Universität Marburg, Marburg, Germany

Published online: 31 January 2023

peripheral and foveal information of objects. The associated information serves as a base of prediction of object features and identities. Herwig and Schneider (2014) argued that prediction from the associated information works bi-directionally: While peripheral object information is used to predict the foveal information of peripheral objects (peripheral object recognition), in return, foveal object information is used to predict the peripheral information of the same object to facilitate visual search. Thus, when peripheral and foveal information of an object is associated, recognizing the said object in the periphery is facilitated through sampled peripheral information and the associated foveal prediction (i.e., peripheral-foveal prediction). Similarly, when the same object becomes the visual search target, the available foveal information and the associated peripheral prediction facilitates visual search (i.e., foveal-peripheral prediction). They provided direct evidence for their account in two experiments in which object features were manipulated across eye movements. Manipulation of peripheral-foveal associations by means of swapping pre- and postsaccadic objects in training shifted subsequent peripheral judgments of these objects towards to the swapped object, quantifying the role of transsaccadic prediction for trained objects. Studies using similar designs provided evidence that the transsaccadic feature prediction mechanism operates with simple visual features such as shapes (Herwig et al., 2015) as well as complex stimuli such as faces and everyday objects (Osterbrink & Herwig, 2021). Furthermore, Köller et al. (2020) showed that the contribution of transsaccadic feature predictions is modulated by the discrepancy between the associated peripheral and foveal information. Hence, transsaccadic predictions could be another factor to consider for understanding the transsaccadic integration puzzle (Herwig, 2015).

Previous studies corroborated evidence that the visual system uses previously associated peripheral and foveal information belonging to the same object to make predictions and to facilitate visual tasks. However, in daily life we are surrounded by a myriad of objects, some of which are familiar (i.e., previously encountered objects), while others are completely novel. With its more limited peripheral vision (for reviews, see Rosenholtz, 2016; Strasburger, 2020; Strasburger et al., 2011; Whitney & Levi, 2011), the visual system is tasked with identifying saccade targets to quickly and efficiently sample information (e.g., Najemnik & Geisler, 2005; for a review, see Eckstein, 2011). Therefore, an ecologically important question is whether transsaccadic predictions are limited to familiar objects or whether they also apply to novel objects. While object-specific predictions for familiar objects can be made based on previously acquired transsaccadic associations as demonstrated by Herwig and Schneider (2014), novel objects that lack these object-specific predictions could generate predictions based on experience in transsaccadic changes (i.e., how visual appearance typically changes at different eccentricities). Hence, object predictions could be conceptualized as a ladder where at the lower end predictions are made for novel objects solely based on coarse expectations of transsaccadic changes. However, at the upper end of the ladder, predictions could be based on more precise object-specific predictions available for familiar objects. Hence, one would expect the precision and accuracy of object predictions to increase with increased familiarity. Previous studies compared visual search and peripheral recognition performance of objects that are swapped or not swapped with another object during training to measure transsaccadic prediction effects (Cox et al., 2005; Herwig et al., 2018; Herwig & Schneider, 2014; Köller et al., 2020; Osterbrink & Herwig, 2021; Paeye et al., 2018; Valsecchi et al., 2020; Valsecchi & Gegenfurtner, 2016). Therefore, observers were equally familiar with both classes of objects during the test phase where the effect of transsaccadic prediction was measured. Thus, whether object-specific predictions from previous transsaccadic associations provide an advantage to familiar objects compared to novel objects with coarse predictions remains unanswered.

Another important question for the everyday relevance of the transsaccadic prediction mechanism is the specificity of the peripheral-foveal predictions. Do these associations reflect a generative mechanism translating between peripheral and foveal representations independent of location and object or do they reflect rather narrow object- and location-specific associations? Two studies (Valsecchi et al., 2020; Valsecchi & Gegenfurtner, 2016) provided evidence that the visual system uses peripheral-foveal associations of object size for predicting the size of objects at the same as well as at the opposite hemifield. In contrast, Herwig et al. (2018) reported that the use of peripheral-foveal associations in prediction is restricted to the location of learning. However, it is important to note the differences between the two groups of studies. First, while Valsecchi and Gegenfurtner (2016) and Valsecchi et al. (2020) manipulated object size, Herwig et al. (2018) manipulated spatial frequency of the transsaccadic objects. Other than being different visual features, perceived size is expected to be uniformly changing across symmetrical eccentricities at opposite hemifields. On the other hand, even if perceived spatial frequency changes symmetrically across visual hemifields, Herwig et al. (2018) did not use symmetric locations to test location specificity. Thus, the location specificity of object predictions is an open question.

To address these questions, we conducted two experiments. Since familiarity with an object varies with individual experience, we devised a training phase where observers unknowingly familiarized with a set of novel objects by performing a sham task. In the first experiment, we manipulated available foveal and peripheral information to test if peripheral-foveal or foveal-peripheral predictions provide an advantage to familiar compared to novel objects. Due to the design of the

first experiment, observers did not execute saccades. However, in daily life we constantly execute saccades to sample information. Therefore, saccades are also important for understanding the contribution of object-specific predictions for familiar objects. We addressed this in the second experiment, by testing whether having object-specific predictions for familiar objects improves transsaccadic object change detection compared to novel objects that lack object-specific predictions.

## Experiment 1: Peripheral-foveal object matches

Based on the transsaccadic feature mechanism, object-specific predictions stemming from previously associated peripheral and foveal information should provide additional advantage compared to novel objects that lack such predictions. One premise of the transsaccadic feature prediction mechanism is that object-specific transsaccadic associations improve the matching of peripheral and foveal information. To test this hypothesis, we asked observers to match peripheral and foveal objects in a 3-AFC (three-alternative forced choice) matching task (Fig. 1).

An important aspect of the peripheral-foveal association is the assumption that those associations could be used to make bi-directional predictions (Herwig & Schneider, 2014). That means peripheral sensory information could predict foveal information (peripheral-foveal prediction) and foveal sensory information could predict peripheral information (foveal-peripheral prediction). Therefore, the familiarity of both peripheral targets as well as foveal options could affect the matching performance. On one hand, familiar peripheral targets could provide object-specific peripheral-foveal predictions that provide a direct comparison for foveal information for the different response options. On the other hand, familiar foveal options could benefit the matching of both familiar and novel peripheral targets by providing specific foveal-peripheral predictions. Therefore, we systematically manipulated the number of familiar options to test how the bi-directional predictions contribute to peripheral-foveal matching. We would expect better performance in trials with familiar than novel peripheral targets if observers make use of peripheral-foveal predictions. Similarly, we would expect to see performance to improve with increasing number of familiar foveal response options, if foveal-peripheral predictions are used by observers.

### Methods

#### Observers

In total, 15 observers (11 females, mean age = 24.33 ± 4.29 years) with normal or corrected-to-normal vision participated in the experiment. Observers were university students; they signed an informed consent form and were compensated with 8 euros per hour. All experiments were approved by the local ethics commission of the Department of Psychology of Marburg University (proposal number 2015-35k) and were conducted in accordance with the Declaration of Helsinki (1964).

#### Stimuli and apparatus

Stimuli were taken from greebles stimulus set (tarrlab.com). Greebles are systematically varying humanlike objects that have been used extensively in psychological research (Gauthier & Tarr, 1997; Gauthier et al., 1999). We chose to use greebles since they systematically capture similarities (e.g., shape of an orange and a tennis ball) as well as differences (e.g., color of an orange and a tennis ball) between everyday objects. Each greeble has a gender (male or female), a family (one of four family body features), and a family trait (one of four body parts). All greebles were rendered achromatic using MATLAB's rgb2gray function. We chose four greebles to be familiarized during the training phase. These greebles only shared one of their features with any other greeble in the same set. Therefore, each familiar greeble had two different features compared to the rest. Similarly, the novel greebles were selected to share only one feature with all other greebles including familiar greebles. Greebles subtended approximately an area of 3 × 5 visual degrees on a neutral gray background.

Throughout the experiments, the stimuli were presented at a 106-cm viewing distance on a back-projection setup, including a PROPixx projector (VPixx Technologies, Saint Bruno, QC, Canada), with a resolution of 1,920 × 1,080 pixels at 120 Hz and a 91 × 51-cm screen from Stewart Filmscreen (Torrance, CA, USA). The display was gamma corrected for linear luminance output and calibrated to minimize the central hot spot. After calibration, it had a luminance of 2.07, 71, and 140 cd/m$^2$ for black, gray, and white pixels, respectively. We used MATLAB (MathWorks Inc., Natick, MA, USA) along with Psychtoolbox (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007) and EyelinkToolbox (Cornelissen et al., 2002) for programing and running the experiments. Eye movements were recorded with an Eyelink 1000+ eye tracker (SR Research Ltd., Kanata, ON, Canada) at 1,000-Hz sampling rate. Eye-tracker calibration was done using the right eye of the observers with a 9-point grid. Fixation points on the grid were placed around the center of the display at 10° horizontal and 5° vertical eccentricities. A fixation check was carried out at the beginning of each trial. The eye tracker was re-calibrated after half of the trials were completed. The heads of the observers were stabilized using a chin and forehead rest.

#### Procedure

Training phase The training consisted of a transsaccadic orientation change detection task using a set of four Greebles

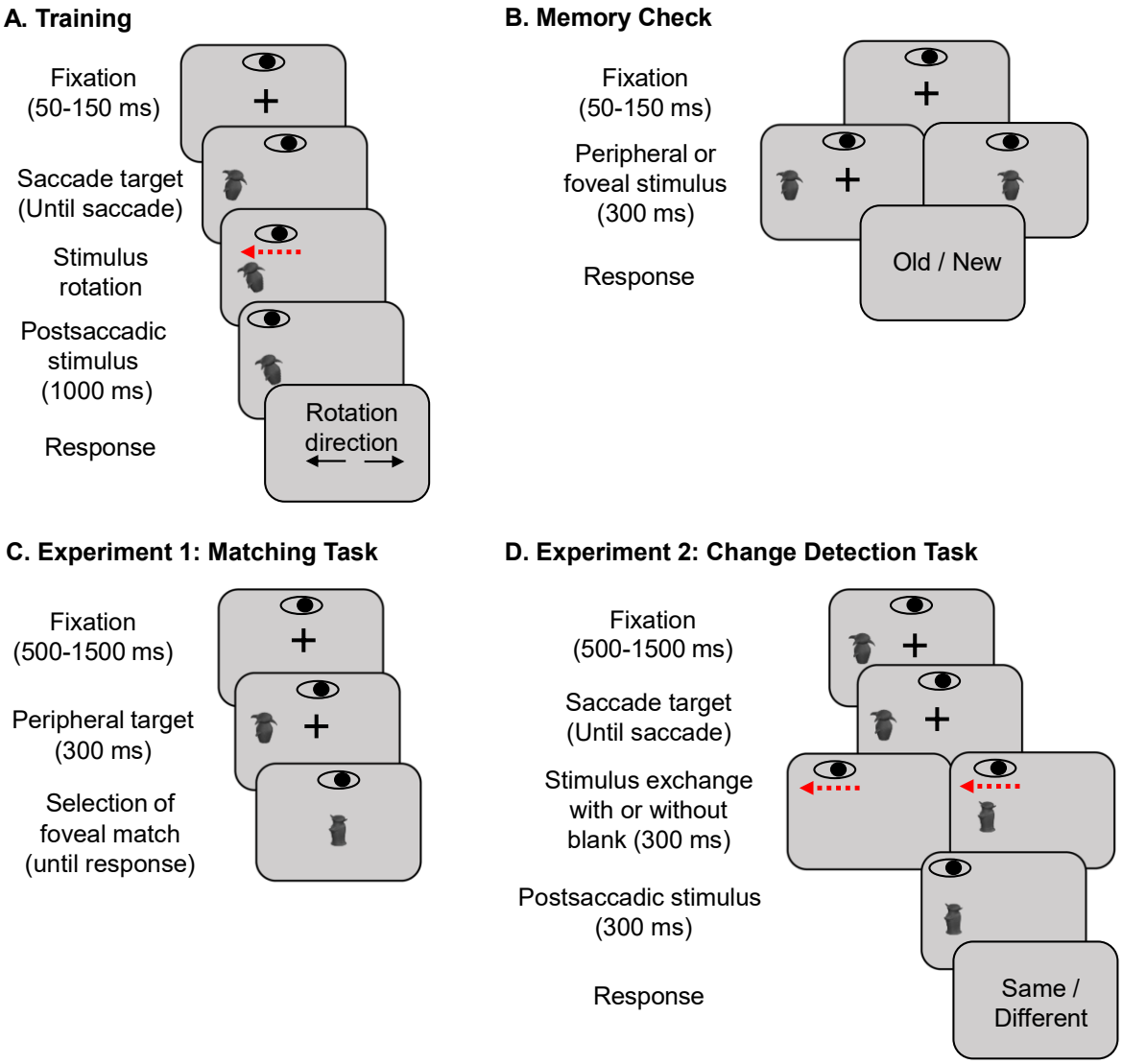



Fig. 1 Schematic illustration of trials in training (A), memory check (B), and the experimental task in Experiments 1 and 2 (C, D). Observers performed the same training and memory check on day 1, then completed Experiment 1 or Experiment 2 on the following day. For illustration purposes, stimulus size and rotation are exaggerated. The response screen was blank in the experiments. The eye icon represents horizontal gaze position and is vertically offset for illustration purposes

chosen for familiarization. The observer's task was to make a saccade to the peripheral object and report the direction of the orientation change. In each trial, after a random fixation period of 50–150 ms, one of the four greebles was randomly selected and presented at 10° left of the fixation target. The peripheral object had a random orientation between -45 and 45°. The peripheral object was tilted by ± 5° when the measured eye position deviated more than 2° away from the fixation target. The now foveated object stayed on the screen for 1,000 ms before it was extinguished. Observers reported the direction of the rotation by using the left and right arrow keys. The experiment consisted of 640 trials in total (160 trials per greeble) and took about 1.5 h.

Memory check On the following day, observers first completed a recognition task to check their memory of the greebles. Four greebles from training and four novel greebles differing in one dimension were used. Observers were asked to indicate whether they recognized the presented greeble from the previous day. In the first half of the trials, greebles were presented peripherally at 10° left or right of the center of the display. In

the second half, greebles were presented foveally at the center of the display. The greeble was extinguished if the gaze position was outside of an imaginary 4°× 7° square central to display. The greeble reappeared when gaze returned to the imaginary square. Greebles were presented for a total duration of 300 ms in five blocks for each viewing condition.

Test phase Following the memory check, observers proceeded with the 3-AFC peripheral-foveal matching task. Each trial started with a fixation check followed by a central fixation cross. After a random period of 500–1,500 ms, a novel or familiar greeble was displayed for a total of 300 ms at 10° either left or right of the fixation with equal probability. To ensure peripheral presentation, the greeble (peripheral target onwards) was extinguished when the measured gaze position was more than 2° away from the fixation. At the end of the peripheral target presentation, one of three options (one matching the peripheral target and two different greebles) was randomly presented in the center of the screen at fixation. The non-matching options always shared two out of three features with the peripheral target and were systematically varied to be both familiar, both novel, or one of each. Therefore, for novel peripheral targets, the overall number of familiar response options varied between zero and two. For familiar peripheral targets, the overall number of familiar response options varied between one and three. Observers were instructed to scroll through the options by using up and down arrows until they identified the target. To make sure that only foveal information was presented to the observers during the response period, the current option was replaced by a noise mask when the gaze deviated more than 2° from the center. In addition, to avoid potential aftereffects from quickly scrolling through options, a central noise mask was presented for 300 ms each time the observers scrolled to another option. The experiment consisted of 240 trials (2 (peripheral target familiarity: familiar, novel) × 2 (target location: left, right) × 3 (number of familiar options: 0, 1, 2 for novel peripheral targets; 1, 2, 3 for familiar peripheral targets) × 20 repetitions).

## Data analysis

We intended to use the memory check as a control measure for familiarization training. To assess if the memory check was an adequate measure of observers' familiarity, we checked the relationship between the performance in the memory check and in the matching task of Experiment 1. We correlated the memory performance with the difference between all novel and all familiar conditions, which reflects the strongest familiarity contrast across conditions. There was no significant relationship between memory check performance and familiarity-dependent performance in Experiment 1 ($r(14) = 0.235$, $p = 0.4$, $BF_{10} = 0.441$). Thus, we used the data of all observers in further analyses.

Data acquired from both days were converted to d′ scores according to Hacker and Ratcliff (1979). Unless otherwise stated, all analyses were carried out in MATLAB and JASP (JASP Team, 2022, JASP (0.15) for Windows). Mean performances for each condition and marginal means for each factor were reported as d′ scores with the associated standard errors. We chose the sample size for both experiments following Herwig and Schneider (2014), which yielded an observed statistical power of 0.849 for a one-way repeated-measures ANOVA. Bayesian factors (BFs) for main effects and posterior odds (POs) for post hoc comparisons were also provided for Experiment 1.

## Results

On the first day, observers were familiarized with the designated greebles by completing a sham task without any instruction about the true purpose of the task. Then, they performed the memory check and the peripheral-foveal matching task. Figure 2 shows peripheral-foveal matching performances for familiar and novel peripheral targets across different number of familiar options. It can be seen that observers were able to perform the matching task even when the peripheral target and all matching options were novel. We first compared the performance in this condition to chance (d' = 0) to confirm that peripheral-foveal predictions are not unique to familiar objects. A one-sample t-test yielded a significantly better than chance level performance in this condition ($d' = 1.58 \pm 0.3$, $t(14) = 10.513$, $p < 0.001$, $d = .583$, $BF_{10} = 2.65 \times 10^7$). Thus, observers were able to match targets that are only briefly seen on the periphery without needing the object-specific associations that familiar objects enjoy.

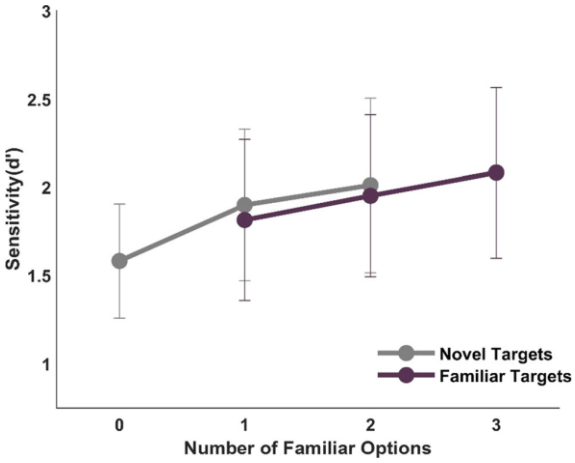



Fig. 2 Mean performance of peripheral-foveal matching for familiar and novel peripheral targets across a different number of familiar foveal options. Error bars show 95% confidence intervals

Figure 2 also shows that matching performance increased with increasing number of familiar options, hinting at the additional advantage of having familiar objects. Object-specific associations are suggested to be working bi-directionally to generate object-specific predictions. In other words, when peripheral or foveal information is present, a prediction is made for the counterpart information about how it would be like to see that item in the other viewing condition. We therefore analyzed trials with novel and familiar peripheral targets and with different number of familiar options separately to test the two different directions of predictions.

We first analyzed trials with novel peripheral targets and with none, one, or two familiar options (gray line in Fig. 2) to see how foveal-peripheral predictions from the familiar foveal options affect matching performance. A one-way repeated-measures ANOVA showed a significant effect of number of familiar foveal options (No familiar option: d′ = 1.58 ± 0.3, one familiar option: d′ = 1.9 ± 0.39 and two familiar options: d′ = 2.01 ± 0.45) on matching performance ($F(2,28) = 5.686$, $p = 0.008$, $\eta^2 = 0.289$, $BF_{10} = 5.536$). Holm-Bonferroni-corrected post hoc comparisons indicated a significant performance improvement from no familiar option to one ($p = 0.046$, PO = 1.617) and two familiar options ($p = 0.009$, PO = 3.694).

Our design does not allow us to test peripheral-foveal prediction in isolation, since trials with familiar peripheral targets have at least one familiar option, which allow foveal-peripheral prediction. Therefore, we tested if the peripheral-foveal prediction from the familiar peripheral targets provides additional improvement to the matching performance by using trials with one or two familiar foveal options for familiar and novel peripheral targets (overlapping data points in Fig. 2). Thus, the additional advantage of peripheral-foveal predictions was tested by comparing performance in trials with novel peripheral targets with familiar peripheral targets with matched number of familiar options. A 2 (familiarity: Familiar: d′ = 1.88 ± 0.2, Novel: d′ = 1.95 ± 0.2) × 2 (Number of familiar choices: 1 familiar option: d′ = 1.86 ± 0.2, 2 familiar option: d′ = 1.98 ± 0.2) ANOVA yielded no significant difference for familiarity ($F(1,14) = 0.523$, $p = 0.482$, $BF_{10} = 0.318$), suggesting that there was no additional advantage of peripheral-foveal prediction. For completeness, we report the main effect of number of familiar options, which partially repeats the post hoc test comparing one familiar option to two familiar options, this time for the combination of familiar and novel peripheral target trials ($F(1,14) = 1.473$, $p = 0.245$, $BF_{10} = 0.496$). There was also no interaction between the two factors ($F(1,14) = 0.013$, $p = 0.912$, $\eta^2 = 0.001$, $BF_{inc} = 0.39$).

We also analyzed separately the trials with familiar peripheral targets, for the sake of completeness. A one-way ANOVA suggested no significant effect of number of familiar options (1 option: d′ = 1.81 ± 0.42, 2 options: d′ = 1.95 ± 0.42, 3 options: d′ = 2.08 ± 0.44) on matching performance for familiar peripheral targets ($F(2,28) = 1.59$, $p = 0.222$, , $BF_{10} = 0.476$).

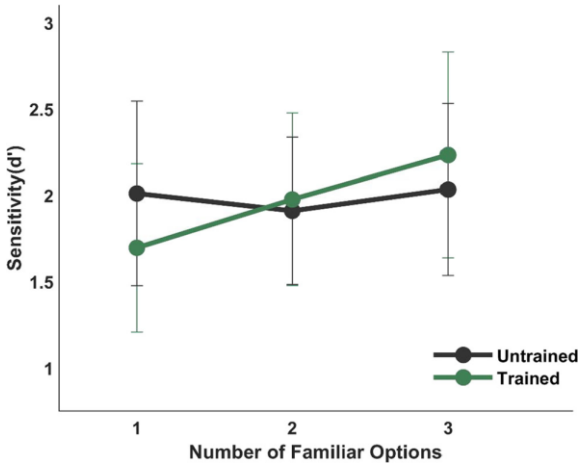



Fig. 3 Mean peripheral-foveal matching performance for familiar targets at the trained (light green) or untrained (dark green) location across a different number of familiar options. Error bars show 95% confidence intervals

Location specificity of transsaccadic learning is an important question for understanding the role of transsaccadic predictions in everyday life. Previous studies have reported conflicting results on the location specificity of the transsaccadic learning (Herwig et al., 2018; Valsecchi et al., 2020; Valsecchi & Gegenfurtner, 2016). During training, visual targets were always presented on the left side of the screen. However, peripheral targets in Experiment 1 were randomly presented either on the left or on the right side of the screen with equal probability. Therefore, if object-specific predictions are dependent on learning location, we would expect the matching performance for familiar peripheral targets to be better when presented on the training location compared to the opposite location. Thus, we tested location specificity by comparing trials from Experiment 1, where the familiar peripheral target was presented either on the same or opposite side of the training stimuli (Fig. 3). A 2 (Location: Trained: d′ = 1.97 ± 0.21, Untrained: d′ = 1.99 ± 0.21) × 3 (Number of Familiar options: one option: d′ = 1.67 ± 0.22, two options: d′ = 1.48 ± 0.22, three options: d′ = 1.39 ± 0.22) Bayesian repeated-measure ANOVA comparing the location factor to the null model provided moderate evidence ($BF_{10} = 0.217$, $F(2,28) = 0.019$, $p = 0.892$) indicating that matching performance was independent from the training location. This is consistent with the findings by Valsecchi and Gegenfurtner (2016) and Valsecchi et al. (2020). As a sanity check, we also performed the same analysis on novel peripheral targets where we do not expect any effect of location. As expected, matching performance did not depend on the novel peripheral location ($F(2,28) = 0.02$, $p = 0.892$, $BF_{01} = 3.54$).

## Discussion

On the first day, observers familiarized with novel targets by executing saccades that are known to lead to stronger learning

effects compared to passive viewing (Herwig & Schneider, 2014). The task of Experiment 1 was to match a briefly presented peripheral target with one of the three centrally presented options. We systematically varied the familiarity of the target and options. Our results show that peripheral-foveal matching is possible in the absence of object-specific predictions. After having confirmed this we returned to the main question of Experiment 1: Do object-specific predictions provide an additional advantage by means of bi-directional object-specific predictions (Herwig & Schneider, 2014)? That is, through association, peripheral sensory information could predict foveal information (peripheral-foveal prediction) as foveal sensory information could predict peripheral sensory information (foveal-peripheral prediction). Therefore, if observers benefit from peripheral-foveal predictions, we would expect better performance in trials with familiar peripheral targets than in trials with novel peripheral targets. Similarly, if observers benefit from foveal-peripheral predictions, we would expect performance to improve with increasing number of familiar options (in particular for novel peripheral targets). We observed an increase in peripheral matching performance with increasing number of familiar options for both familiar and novel peripheral targets. However, only the differences between no familiar option to one and two options for novel peripheral targets reached significance (Table 1). Given the high overall matching performance (~80%) in Experiment 1, the absence of a difference between one and two familiar options most likely represents a ceiling effect. Overall, having object-specific associations of familiar objects was advantageous when observers could compare the sensory information from the peripheral targets with the peripheral predictions from familiar foveal objects.

One explanation for foveal-peripheral predictions being more advantageous than peripheral-foveal predictions could be the higher reliability of the foveal information due to foveal vision having higher acuity and being less prone to crowding (for reviews, see Rosenholtz, 2016; Strasburger, 2020; Strasburger et al., 2011; Whitney & Levi, 2011). The difference in reliabilities could have been further amplified by our design since peripheral objects were briefly presented while foveal objects were presented until observers selected a response. Even if peripheral-foveal predictions provide an advantage for the peripheral-foveal matching task, our design does not allow testing for the isolated effect of peripheral-foveal prediction because trials with familiar targets had at least one familiar option. Thus, if there is any advantage of peripheral-foveal associations in peripheral-foveal matching, it is undermined by the effect of foveal-peripheral associations that are based on more reliable sensory information in our task.

Experiment 1 also provided evidence for spatial transfer of predictions from transsaccadic associations. The question of location specificity of transsaccadic learning is an important aspect for the ecological usefulness of familarity-based predictions. If transsaccadic associations of familiar objects were to be spatially bound, then the advantage of object-specific predictions would be highly specific and limited, as in some forms of perceptual learning (e.g., Karni & Sagi, 1991; for reviews, see Fahle, 2005; Fine & Jacobs, 2002; Goldstone, 1998; Watanabe & Sasaki, 2015). On the other hand, if the transsaccadic associations generalize then the visual system could use them for generating object-specific predictions to quickly identify and pre-process familiar peripheral objects in visual scenes. Previously, Valsecchi and Gegenfurtner (2016) and Valsecchi et al. (2020) provided evidence supporting the independence of transsaccadic learning for size, which is opposed by Herwig et al. (2018) showing specificity for spatial frequency. However, those studies used different simple visual features (size vs. spatial frequency) and eccentricities (opposite hemifield vs. various other locations) that can explain the contradictory results. In the current study, we used complex objects and compared matching performance in the trained and the mirrored location in the opposite hemifield to that in Valsecchi and Gegenfurtner (2016) and Valsecchi et al. (2020). Our results favor the spatial transfer of transsaccadic learning at least to the opposite hemifield. Together, the results of Experiment 1 suggest that the matching of peripheral and foveal information benefits

Table 1 Summary of hypotheses and associated results from Experiment 1. Each row shows which variables were used to test the given effect. Non-contributing variables are indicated with "-"

| | Peripheral target | Number of familiar options | Location | p | $BF_{10}$ |
|---|---|---|---|---|---|
| Novel Object Prediction | Novel | 0 | - | < 0.001 | $2.65 \times 10^7$ |
| Foveal-Peripheral Prediction (Novel) | Novel | 0, 1, 2 | - | 0.008 | 5.536 |
| Peripheral-Foveal Prediction (Additional) | Novel, Familiar | - | - | 0.482 | 0.318 |
| Foveal-Peripheral Prediction (Familiar – Additional) | Familiar | 1, 2, 3 | - | 0.222 | 0.476 |
| Location Specificity | Familiar | - | Trained, Untrained | 0.892 | 0.217 |

from the spatially independent transsaccadic associations of familiar objects at least when foveal to peripheral predictions can be made.

## Experiment 2: Transsaccadic change detection

The first experiment showed that the presence of familiar foveal distractors improved the matching with novel peripheral targets. This implies that observers were able to detect the mismatch between the peripheral sensory information of the novel objects and the peripheral predictions of familiar distractors based on foveal sensory information. In the first experiment, observers were required to match peripheral and foveal objects when they were not allowed to execute saccades to alter the retinal location of the objects. However, in daily life we constantly make saccades to peripheral objects to obtain foveal information that offers better resolution. By executing a saccade, the visual system integrates pre- and post-saccadic information related to saccade targets as well as to other attended locations (Schut et al., 2018; Stewart & Schütz, 2019a; Stewart & Schütz, 2019b). However, if there is a high discrepancy between pre- and post-saccadic objects, then the visual system segregates this information (e.g., Atsma et al., 2016; Demeyer et al., 2010; Tas et al., 2012; Tas et al., 2021). Peripheral-foveal associations should provide richer object-specific predictions compared to generative predictions of novel objects that can further facilitate the detection of discrepancies between pre- and post-saccadic information. Therefore, in the second study we hypothesized that transsaccadic changes would be more detectable in familiar objects than in novel objects.

Transsaccadic change detection of location (Deubel et al., 1996; Poth et al., 2015) and features (Hübner & Schütz, 2021; Weiß et al., 2015) is known to be improved by a blank period introduced after saccades. Atsma et al. (2016) suggested that the visual system retains both peripheral and foveal information to arbitrate between integration or segregation of transsaccadic information. If there is a high discrepancy between the two types of information or a blank, which suggests a violation of the stable world assumption (Deubel et al., 1996), the visual system segregates the foveal and peripheral information. As a result, the transsaccadic change becomes more apparent to the observer. Within the framework of the transsaccadic feature prediction mechanism, we consider two possibilities for how the presence of a blank could differentially improve transsaccadic change detection of familiar and novel objects. On the one hand, blanking could be more effective at improving the change detection in novel stimuli since detecting changes in familiar stimuli should already be facilitated by the peripheral-foveal associations. On the other hand, blanking could improve the change detection performance for familiar stimuli since the separation of information by blanking would help to segregate more precise object-specific predictions from the sensory information of the postsaccadic stimulus.

To test whether peripheral-foveal prediction of familiar objects improves transsaccadic change detection and how blanking contributes to the detection of changes in familiar and novel stimuli, we conducted a second experiment. Since we used complex objects that varied in high level features, we varied the magnitude of change (change difficulty here after) as a control for large discrepancies between transsaccadic objects that could lead to segregation of transsaccadic information.

### Methods

#### Observers

In total, 18 new observers (13 females, mean age = 23.83 ± 3.07 years) with normal or corrected-to-normal vision participated to the experiment. Observers were university students; they signed an informed consent form and were compensated with 8 euros per hour.

#### Stimuli and apparatus

The stimuli and apparatus were identical to the training except for the following changes. The same four familiarized greebles from the training phase and eight novel greebles were used. For each of the 12 greebles, a hard change greeble and an easy change greeble that vary in one or two dimensions, respectively, were chosen to serve in change trials.

#### Procedure

The observers underwent the same training and memory check procedure as the observers in the first experiment. After the memory check, observers proceeded with the transsaccadic change detection task. Each trial started with a fixation period of 500–1,500 ms. While fixation was maintained, a presaccadic peripheral target was presented either 10° left or right of the fixation target. Saccades were detected online when the horizontal gaze position traveled more than 4° from the target. The detection of the saccade triggered the exchange of the presaccadic target with a different (novel) or an identical postsaccadic target. For the blank trials, the presaccadic target was extinguished and exchanged with the postsaccadic target only after a 300-ms blank period. In all conditions, the postsaccadic target was presented for 300 ms. The task of the observer was to report whether the pre- and post-saccadic target were identical or not. Each observer completed blank trials and no-blank trials in separate blocks. The order of the blank and no-blank blocks was counterbalanced by reversing the order for every other observer. Typically, blank and no-blank trials are

interleaved in this paradigm, but we decided to block these conditions to avoid temporal variation in the onset of the postsaccadic target and to maximize the consistency of trials within a block. The viewing condition (familiar target viewed peripherally or foveally) and the difficulty (hard: one different feature; easy: two different features) of the target was manipulated across observers in the following way: Half of the observers completed the task where familiar stimuli were presented as a postsaccadic target for hard change trials and as a presaccadic target for easy change trials. For the other half of the observers the position of the familiar stimuli was switched for easy and hard change trials.

### Data analysis

Similar to Experiment 1, we intended to use the memory check as a control measure for familiarization training. To asses if the memory check was an adequate measure of observers' familiarity for Experiment 2, we checked the relationship between performance in the memory check and the change detection task. We correlated the memory performance with the difference between novel and familiar objects in no-blank conditions, separately for easy and hard change trials. There was no significant relationship between memory check performance and easy change trials ($r(17) = 0.171$, $p = 0.49$, $BF_{10} = 0.361$) or hard change trials ($r(17) = -0.158$, $p = 0.53$, $BF_{10} = 0.35$) . Thus, we did not exclude any observers on the basis of their memory check performance.

As in the first experiment, the data acquired during the training phase were converted to d′ scores. The change difficulty manipulation for the transsaccadic change detection task does not allow us to make a direct d′ calculation since the signal-absent (same) trials cannot be assigned to signal-present (different) trials with two difficulties. Therefore, the data acquired from the experiment phase were converted to d′ using the revised d′ table from Hacker and Ratcliff (1979). We compared main effects to the null model for obtaining BFs. For interactions, we added main effects to the null model to calculate BFs. Posterior odds were reported for post hoc tests.

Observers were separately tested in two viewing groups where the order of the exchanged stimulus for the familiar stimulus trials was switched across the difficulty conditions. By means of a one-way mixed ANOVA, we first compared the two viewing groups using familiar trials for the blanking and no-blanking conditions. There was no significant difference between the two viewing groups ($F(1,16) = 0.691$, $p = 0.418$, $BF_{10} = 0.378$). Therefore, we merged the data from the groups for the subsequent analyses. In the merged data, trials with at least one familiar stimulus were assigned to the familiar trial group regardless of change difficulty or being pre- or post-saccadic. Observed statistical power for the all main effects and the interaction from a 2 × 2 × 2 ANOVA was larger than 0.8.

## Results

Similar to Experiment 1, observers in Experiment 2 were familiarized with the training stimuli on day 1 then proceeded to the test phase. We conducted a 2 (Familiarity: Familiar, Novel) × 2 (Blanking: Blank, No-blank) × 2 (Change Difficulty: Easy, Hard) repeated-measures ANOVA (Fig. 4). There was a main effect of familiarity ($F(1,17) = 27.101$, $p < 0.001$, $\eta^2 = 0.061$, , $BF_{10} = 7.81$) indicating a better change detection performance for familiar ($d' = 0.86 \pm 0.08$) than for novel ($d' = 0.56 \pm 0.08$) stimuli. Similarly, change detection performance improved with the presence of a blanking period ($F(1,17) = 10.084$, $p = 0.006$, $\eta^2 = 0.06$, , $BF_{10} = 7.08$, blank: $0.86 \pm 0.09$ vs. no blank: $0.56 \pm 0.09$). As expected, the change difficulty also modulated the performance ($F(1,17) = 49.156$, $p < 0.001$, $\eta^2 = 0.447$, , $BF_{10} = 1.13 \times 10^{15}$, easy: $1.22 \pm 0.01$ vs. hard: $0.3 \pm 0.01$). Change difficulty did not interact with blanking ($F(1,17) = 0.144$, $p = 0.709$, $BF_{inc} = 0.286$), familiarity ($F(1,17) = 3.676$, $p = 0.072$, $BF_{inc} = 0.977$), or a combination of the two ($F(1,17) = 0.365$, $p = 0.554$, $BF_{inc} = 0.171$), which suggests that the difficulty manipulation was balanced for familiar and novel stimuli.

However, we did find a significant interaction between blanking and familiarity ($F(1,17) = 10.444$, $p = 0.005$, $\eta^2 = 0.011$, $BF_{inc} = 2.152$). To check if familiar and novel objects benefitted from blanking to a different degree, we compared the difference between blank and no-blank conditions for familiar and novel conditions. Since the change difficulty could modulate the blanking effect, we analyzed easy and hard change trials separately. We calculated blanking effects for familiar and novel conditions by subtracting the performance

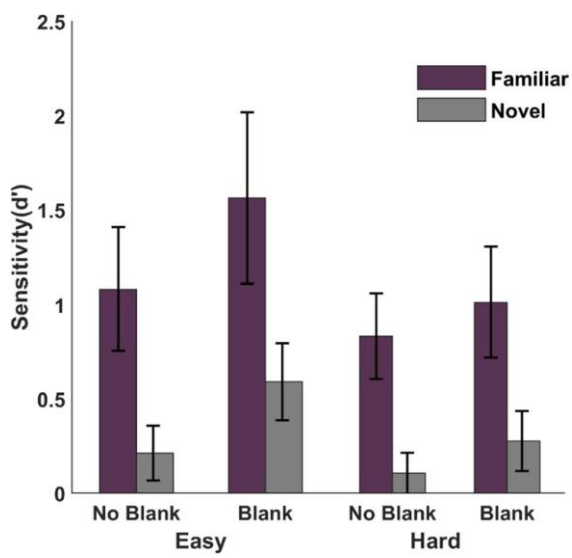



Fig. 4 Mean change detection performance with or without blank for familiar (purple) and novel (gray) stimuli across change difficulty (left panel: easy change; right panel: hard change). Error bars show 95% confidence intervals

in blank conditions from no-blank conditions for easy and hard change trials. We found that familiar objects benefitted from blanking more than novel objects in both easy ($t(17) = 2.361$, $p = 0.03$, $d = 0.556$, $BF_{10} = 2.13$) and hard change conditions ($t(17) = 2.747$, $p = 0.014$, $d = 0.647$, $BF_{10} = 4.06$).

Finally, unlike Experiment 1, Experiment 2 allowed us to test the isolated effect of peripheral-foveal predictions. The design of Experiment 1 only allowed us to test the additional contribution of peripheral-foveal predictions since for familiar peripheral targets there was always a familiar foveal option that could provide foveal-peripheral predictions. However, in Experiment 2, this can be directly tested by comparing the familiar presaccadic stimulus trials to the novel stimulus trials. Since change difficulty of these trials was manipulated across groups, for one group these trials were hard changes and for the other group they were easy change trials. We controlled change difficulty by comparing familiar presaccadic trials with novel trials with the same change difficulty. Therefore, if the familiar presaccadic stimulus trial was also a hard change trial, the performance of the observer was paired with their performance in hard change novel trials. By matching the change difficulty of familiar and novel trials, we tested whether peripheral-foveal predictions provide extra benefit in change detection. A 2 (peripheral object: familiar, novel) × 2 (blanking: blank, no blank) repeated-measures ANOVA comparing trials with familiar and novel peripheral targets (with matched change difficulty) yielded a significant main effect of peripheral object familiarity ($F(17,1) = 12.614$, $p = 0.002$, $\eta^2 = 0.135$, $BF_{10} = 7.023$, $d' = 1.02 \pm 0.14$ vs $d' = 0.66 \pm 0.14$) on change detection performance. As expected, there was also an effect of blanking ($F(17,1) = 5.204$, $p = 0.036$, $\eta^2 = 0.129$, $BF_{10} = 6.054$, $d' = 1.01 \pm 0.15$ vs $d' = 0.68 \pm 0.15$) and interaction ($F(17,1) = 5.914$, $p = 0.026$, $\eta^2 = 0.034$ $BF_{inc} = 0.812$).

## Discussion

In Experiment 2, we asked observers to report changes to presaccadic objects during saccades. In line with our hypothesis, the results showed that transsaccadic changes to familiar objects were more readily detectable compared to novel objects especially when they were blanked (Table 2.).

The familiarity effect that we observed can be well explained by the framework suggested by Atsma et al. (2016). They argued that the visual system holds on to the peripheral and foveal information to arbitrate integration or segregation of information based on the degree of discrepancy between the two types of information. The more precise object-specific predictions for familiar objects provide an advantage in detecting transsaccadic changes. Therefore, when the visual system evaluates the discrepancy between a familiar object with another object, it can compare the object-specific transsaccadic prediction with the sensory information of the other object. However, any transsaccadic exchange between two novel objects provides less precise predictions and thereby makes the change detection harder.

Another way of improving the detection of transsaccadic changes is to introduce a blank period before the transsaccadic change (e.g., Deubel et al., 1996; Hübner & Schütz, 2021; Weiß et al., 2015). We found that familiar objects benefitted more than novel objects regardless of change difficulty. In other words, the separation of pre- and post-saccadic information by blanking was more beneficial for detecting changes in familiar objects compared to novel objects. One reason could be that the object-specific predictions are informative for the visual system but were masked by the postsaccadic input, while general predictions generated for novel stimuli were also masked but not as informative as object-specific predictions to improve the change detection. Therefore, removing the postsaccadic masking or overwriting by means of blanking (Grzeczkowski et al., 2020; Tas et al., 2021) might have amplified the difference between the two types of prediction.

Finally, we tested if peripheral-foveal predictions from transsaccadic associations can be effective at signaling transsaccadic changes. Experiment 1 showed that the foveal-peripheral predictions were largely responsible for the familiarity effect that we observed in the peripheral-foveal matching task. However, the design of Experiment 1 did not allow comparison of the isolated effects of peripheral-foveal since familiar peripheral target trials always contained familiar options. This comparison can be made by analyzing the trials with peripheral familiar targets in Experiment 2. Our results showed that, at least for the transsaccadic object change detection, having peripheral-foveal predictions was advantageous. We argue that the different pattern of results in the two experiments could be explained by the differences in the relative reliability and usefulness of information in the two tasks. Although peripheral and foveal predictions were useful for both tasks, only in the change detection were both predictions equally informative. On the one hand, in Experiment 1, observers had one source of peripheral-foveal prediction with time-constrained sensory information and three sources of foveal-peripheral prediction without any time constraints on the sensory information. Therefore, the visual system had multiple sources of information to perform the task. Thus, the foveal-peripheral predictions could have been weighted more by having more reliable sensory information. In contrast, the availability of peripheral and foveal information was more balanced in Experiment 2, making the peripheral-foveal prediction more informative compared to Experiment 1. Thus, we argue that object-specific predictions work bi-directionally but the prediction in each direction is used based on their reliability and relevance for the task.

Table 2. Summary of the hypotheses and associated results from Experiment 2. Each row shows which variables were used to test the given effect. Non-contributing variables are indicated with "-"

| | Familiarity | Blanking | p | $BF_{10}$ |
|---|---|---|---|---|
| Familiarity | Novel, Familiar | - | < 0.001 | 7.81 |
| Blanking | - | Blank, No blank | 0.006 | 7.08 |
| Familiarity x Blanking | Familiar, Novel | Blank, No blank | easy: 0.03<br>hard: 0.014 | 2.13<br>4.06 |
| Peripheral-Foveal (Familiar) | Novel, Familiar (saccade targets) | - | 0.002 | 7.023 |

## General discussion

Through the course of life, the visual system samples information with peripheral and foveal vision. Each saccade directed to an object provides peripheral information and foveal information about the targeted object. The visual system uses the available sensory information to make transsaccadic predictions about how an object would look like in different viewing conditions (e.g., Herwig & Schneider, 2014; for reviews, see Huber-Huber et al., 2021; Stewart et al., 2020).

We argue that object predictions can be made at two levels. First, object predictions can be made based on a generative model about how object appearance changes at different eccentricities without requiring object-specific learning. Second, a more precise prediction can be made based on the previously associated peripheral and foveal information belonging to the same object, as suggested by Herwig and Schneider (2014). We argue that object predictions can be conceptualized as a ladder where at the lower end coarse predictions can be made for both novel and familiar objects based on a generative model about their appearance at different eccentricities and at the upper end more precise object-specific predictions can be made for familiar objects from previously associated peripheral and foveal object information. In this study, we investigated if object-specific predictions for familiar objects provide an advantage in peripheral-foveal matching and transsaccadic change detection tasks compared to novel objects, which lack these predictions. Our results suggest that while more general predictions are sufficient to perform peripheral-foveal matching and change detection tasks with novel objects, familiar objects with more precise predictions had a significant advantage over novel objects on both tasks. Previous studies have shown that these predictions work bi-directionally to support various visual tasks such as peripheral recognition or visual search (Herwig & Schneider, 2014; Osterbrink & Herwig, 2021). The design of Experiment 1 allowed us to systematically manipulate the familiar object information available to peripheral and central vision. We found that observers benefitted more from foveal-peripheral predictions. In addition, the availability of peripheral-foveal predictions was advantageous in Experiment 2. Therefore, our results not only support bi-directionality of object-specific predictions, but also show how task demands and available information could influence the direction of prediction.

Object-specific predictions that can be made for familiar objects could be decisive when performing demanding visual tasks such as medical image interpretation (e.g., Bertram et al., 2013; Lago et al., 2021; Williams et al., 2021). The association of transsaccadic information for familiar objects could be an important factor in visual expertise and everyday life, even for very overtrained stimuli, such as faces. Familiar faces are processed more rapidly and with less attention compared to novel faces (Gobbini et al., 2013). Furthermore, face processing benefits from transsaccadic effects such as the preview benefit (Buonocore et al., 2020; Huber-Huber et al., 2019). Moreover, Osterbrink and Herwig (2021) showed that transsaccadic predictions can be made for faces and that they bias judgments towards the acquired peripheral-foveal associations. Following our ladder metaphor, the visual system can generate predictions about how faces look like in different viewing conditions when we encounter a stranger. When we encounter the same face again or start to become more acquainted with it, in addition to our first general predictions, the visual system can make predictions specific to that face. By having these more precise predictions, we can now more easily recognize the face or detect changes to it.

The transsaccadic predictions for the pre- and post-saccadic information can also help to arbitrate between integration and segregation of pre- and post-saccadic information (Atsma et al., 2016). Previous studies have shown that the visual system is sensitive to transsaccadic changes to objects and could adapt itself within a brief period (Cox et al., 2005; Valsecchi et al., 2020; Valsecchi & Gegenfurtner, 2016). As a result, the visual system builds transsaccadic associations of peripheral and foveal information reflecting the change caused by adaptation (Cox et al., 2005; Herwig et al., 2018; Herwig & Schneider, 2014; Köller et al., 2020; Osterbrink & Herwig, 2021; Paeye et al., 2018; Valsecchi et al., 2020; Valsecchi & Gegenfurtner, 2016) or learning as in the current experiments. Previous studies showed that pre- and post-saccadic information is integrated based on their reliabilities (e.g., Ganmor

et al., 2015; Hübner & Schütz, 2017; Stewart & Schütz, 2019a; Wolf & Schütz, 2015). However, when the saccade target is drastically changed beyond adaptation, for instance when it is exchanged with another object (Demeyer et al., 2010; Tas et al., 2012; Tas et al., 2021), the presaccadic information is not integrated or overwritten by the postsaccadic information. The object-specific predictions could further help the visual system to arbitrate between perceptual continuity and discontinuity. Observers in Experiment 2 were better at detecting changes to familiar objects than novel objects. Thus, object-specific transsaccadic predictions are more informative than their more general counterparts at detecting changes.

In conclusion, we argue that general transsaccadic predictions can be based on a generative model of typical differences in appearance at different eccentricities and that more precise object-specific predictions can be acquired by transsaccadic learning. While the visual system can perform object matching and change detection using general predictions, it specializes in familiar objects by directly associating peripheral and foveal information for better precision. The higher precision provided by object-specific predictions of familiar objects could be supporting visual expertise as well as help the visual system to detect changes in the environment more readily.

Acknowledgments We thank Lena Weinert and Melanie Zylenas for helping with the data collection.

Funding Open Access funding enabled and organized by Projekt DEAL. NG and ACS were supported by the German-Canadian International Research Training Group (IRTG) 1901 "Brain in action" by the German Research Foundation (DFG). ACS was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 676786) and by "The Adaptive Mind", funded by the Excellence Program of the Hessian Ministry of Higher Education, Science, Research and Art.

Data availability The original stimulus images are courtesy of Michael J. Tarr, Carnegie Mellon University, http://www.tarrlab.org/ Data generated during experiments can be found at https://doi.org/10.5281/zenodo.7520182. None of the experiments was preregistered.

## Declarations

Competing interests The authors have no competing interest to declare that are relevant to content of this article

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

## References

Atsma, J., Maij, F., Koppen, M., Irwin, D., & Medendorp, W. (2016). Causal inference for spatial constancy across Saccades. PLoS Computational Biology, 12(3), e1004766. https://doi.org/10.1371/journal.pcbi.1004766

Bertram, R., Helle, L., Kaakinen, J. K., & Svedström, E. (2013). The effect of expertise on eye movement behaviour in medical image perception. PLoS One, 8(6), e66169.

Brainard, D. (1997). The psychophysics toolbox. Spatial Vision, 10(4), 433–436. https://doi.org/10.1163/156856897x00357

Buonocore, A., Dimigen, O., & Melcher, D. (2020). Post-Saccadic face processing is modulated by pre-saccadic preview: Evidence from fixation-related potentials. The Journal of Neuroscience, 40(11), 2305–2313. https://doi.org/10.1523/jneurosci.0861-19.2020

Cornelissen, F., Peters, E., & Palmer, J. (2002). The Eyelink toolbox: Eye tracking with MATLAB and the psychophysics toolbox. Behavior Research Methods, Instruments, & Computers, 34(4), 613–617. https://doi.org/10.3758/bf03195489

Cox, D. D., Meier, P., Oertelt, N., & DiCarlo, J. J. (2005). 'Breaking'position-invariant object recognition. Nature Neuroscience, 8(9), 1145–1147.

David, E., Beitner, J., & Võ, M. (2021). The importance of peripheral vision when searching 3D real-world scenes: A gaze-contingent study in virtual reality. Journal of Vision, 21(7), 3. https://doi.org/10.1167/jov.21.7.3

Demeyer, M., De Graef, P., Wagemans, J., & Verfaillie, K. (2010). Object form discontinuity facilitates displacement discrimination across saccades. Journal of Vision, 10(6), 17–17.

Deubel, H., Schneider, W., & Bridgeman, B. (1996). Postsaccadic target blanking prevents saccadic suppression of image displacement. Vision Research, 36(7), 985–996. https://doi.org/10.1016/0042-6989(95)00203-0

Eckstein, M. P. (2011). Visual search: A retrospective. Journal of Vision, 11(5), 14–14.

Fahle, M. (2005). Perceptual learning: Specificity versus generalization. Current Opinion in Neurobiology, 15(2), 154–160.

Fine, I., & Jacobs, R. A. (2002). Comparing perceptual learning across tasks: A review. Journal of Vision, 2(2), 5–5.

Ganmor, E., Landy, M., & Simoncelli, E. (2015). Near-optimal integration of orientation information across saccades. Journal of Vision, 15(16), 8. https://doi.org/10.1167/15.16.8

Gauthier, I., & Tarr, M. J. (1997). Becoming a "Greeble" expert: exploring mechanisms for face recognition. Vision Research, 37(12), 1673-1682. https://doi.org/10.1016/S0042-6989(96)00286-6

Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P., & Gore, J. C (1999) Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects. Nature Neuroscience, 2(6), 568-573. https://doi.org/10.1038/9224

Gobbini, M., Gors, J., Halchenko, Y., Rogers, C., Guntupalli, J., Hughes, H., & Cipolli, C. (2013). Prioritized detection of personally familiar faces. PLoS One, 8(6), e66620. https://doi.org/10.1371/journal.pone.0066620

Goldstone, R. L. (1998). Perceptual learning. Annual Review of Psychology, 49(1), 585–612.

Grzeczkowski, L., Deubel, H., & Szinte, M. (2020). Stimulus blanking reveals contrast-dependent transsaccadic feature transfer. Scientific Reports, 10(1), 1–9. https://doi.org/10.1038/s41598-020-75717-y

Hacker, M., & Ratcliff, R. (1979). A revised table of d' for M-alternative forced choice. Perception & Psychophysics, 26(2), 168–170. https://doi.org/10.3758/bf03208311

Herwig, A. (2015). Transsaccadic integration and perceptual continuity. Journal of Vision, 15(16), 7. https://doi.org/10.1167/15.16.7

Herwig, A., & Schneider, W. (2014). Predicting object features across saccades: Evidence from object recognition and visual search.

Journal of Experimental Psychology: General, 143(5), 1903–1922. https://doi.org/10.1037/a0036781
Herwig, A., Weiß, K., & Schneider, W. (2015). When circles become triangular: how transsaccadic predictions shape the perception of shape. Annals of the New York Academy of Sciences, 1339(1), 97–105. https://doi.org/10.1111/nyas.12672
Herwig, A., Weiß, K., & Schneider, W. (2018). Feature prediction across eye movements is location specific and based on retinotopic coordinates. Journal of Vision, 18(8), 13. https://doi.org/10.1167/18.8.13
Hübner, C., & Schütz, A. C. (2017). Numerosity estimation benefits from transsaccadic information integration. Journal of Vision, 17(13), 12. https://doi.org/10.1167/17.13.12
Huber-Huber, C., Buonocore, A., Dimigen, O., Hickey, C., & Melcher, D. (2019). The peripheral preview effect with faces: Combined EEG and eye-tracking suggests multiple stages of trans-saccadic predictive and non-predictive processing. NeuroImage, 200, 344–362.
Huber-Huber, C., Buonocore, A., & Melcher, D. (2021). The extrafoveal preview paradigm as a measure of predictive, active sampling in visual perception. Journal of Vision, 21(7), 12. https://doi.org/10.1167/jov.21.7.12
Hübner, C., & Schütz, A. (2021). A bias in saccadic suppression of shape change. Vision Research, 186, 112–123. https://doi.org/10.1016/j.visres.2021.05.005
JASP Team. (2022). JASP (Version 0.16.3) [Computer software]. https://jasp-stats.org/faq/how-do-i-cite-jasp/
Karni, A., & Sagi, D. (1991). Where practice makes perfect in texture discrimination: evidence for primary visual cortex plasticity. Proceedings of the National Academy of Sciences, 88(11), 4966–4970.
Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3? Perception, 36, 14, ECVP Abstract Supplement.
Köller, C., Poth, C., & Herwig, A. (2020). Object discrepancy modulates feature prediction across eye movements. Psychological Research, 84(1), 231–244. https://doi.org/10.1007/s00426-018-0988-5
Lago, M. A., Jonnalagadda, A., Abbey, C. K., Barufaldi, B. B., Bakic, P. R., Maidment, A. D., & Eckstein, M. P. (2021). Under-exploration of three-dimensional images leads to search errors for small salient targets. Current Biology, 31(5), 1099–1106.
Laurin, A. S., Bleau, M., Gedjakouchian, J., Fournet, R., Pisella, L., & Khan, A. Z. (2021). Post-saccadic changes disrupt attended pre-saccadic object memory. Journal of Vision, 21(8), 8–8.
Mathôt, S., & Theeuwes, J. (2011). Visual attention and stability. Philosophical Transactions of the Royal Society, B: Biological Sciences, 366(1564), 516–527.
Melcher, D. (2011). Visual stability. Philosophical Transactions of the Royal Society, B: Biological Sciences, 366(1564), 468–475.
Najemnik, J., & Geisler, W. S. (2005). Optimal eye movement strategies in visual search. Nature, 434(7031), 387–391.
Nuthmann, A. (2014). How do the regions of the visual field contribute to object search in real-world scenes? Evidence from eye movements. Journal of Experimental Psychology: Human Perception and Performance, 40(1), 342–360. https://doi.org/10.1037/a0033854
Nuthmann, A., & Canas-Bajo, T. (2022). Visual search in naturalistic scenes from foveal to peripheral vision: A comparison between dynamic and static displays. Journal of Vision, 22(1), 10. https://doi.org/10.1167/jov.22.1.10
Osterbrink, C., & Herwig, A. (2021). Prediction of complex stimuli across saccades. Journal of Vision, 21(2), 10. https://doi.org/10.1167/jov.21.2.10
Paeye, C., Collins, T., Cavanagh, P., & Herwig, A. (2018). Calibration of peripheral perception of shape with and without saccadic eye movements. Attention, Perception, & Psychophysics, 80(3), 723–737. https://doi.org/10.3758/s13414-017-1478-3
Poth, C. H., Herwig, A., & Schneider, W. X. (2015). Breaking object correspondence across saccadic eye movements deteriorates object recognition. Frontiers in Systems Neuroscience, 9, 176.
Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. Psychological Bulletin, 124(3), 372–422. https://doi.org/10.1037/0033-2909.124.3.372
Rolfs, M. (2015). Attention in active vision: A perspective on perceptual continuity across saccades. Perception, 44(8–9), 900–919.
Rosenholtz, R. (2016). Capabilities and limitations of peripheral vision. Annual Review of Vision Science, 2, 437–457.
Schut, M. J., Van der Stoep, N., Fabius, J. H., & Van der Stigchel, S. (2018). Feature integration is unaffected by saccade landing point, even when saccades land outside of the range of regular oculomotor variance. Journal of Vision, 18(7), 6–6.
Stewart, E., & Schütz, A. (2019a). Transsaccadic integration is dominated by early, independent noise. Journal of Vision, 19(6), 17. https://doi.org/10.1167/19.6.17
Stewart, E., & Schütz, A. (2019b). Transsaccadic integration benefits are not limited to the saccade target. Journal of Neurophysiology, 122(4), 1491–1501. https://doi.org/10.1152/jn.00420.2019
Stewart, E., Valsecchi, M., & Schütz, A. (2020). A review of interactions between peripheral and foveal vision. Journal of Vision, 20(12), 2. https://doi.org/10.1167/jov.20.12.2
Strasburger, H. (2020). Seven myths on crowding and peripheral vision. i-Perception, 11(3), 2041669520913052.
Strasburger, H., Rentschler, I., & Jüttner, M. (2011). Peripheral vision and pattern recognition: A review. Journal of Vision, 11(5), 13–13.
Tas, A., Moore, C., & Hollingworth, A. (2012). An object-mediated updating account of insensitivity to transsaccadic change. Journal of Vision, 12(11), 18–18. https://doi.org/10.1167/12.11.18
Tas, A., Mordkoff, J., & Hollingworth, A. (2021). Object-mediated overwriting across saccades. Journal of Vision, 21(2), 3. https://doi.org/10.1167/jov.21.2.3
Treisman, A. (1986). Features and objects in visual processing. Scientific American, 255(5), 114B–1125B.
Valsecchi, M., & Gegenfurtner, K. (2016). Dynamic re-calibration of perceived size in fovea and periphery through predictable size changes. Current Biology, 26(1), 59–63. https://doi.org/10.1016/j.cub.2015.10.067
Valsecchi, M., Cassanello, C., Herwig, A., Rolfs, M., & Gegenfurtner, K. R. (2020). A comparison of the temporal and spatial properties of trans-saccadic perceptual recalibration and saccadic adaptation. Journal of Vision, 20(4), 2–2.
Watanabe, T., & Sasaki, Y. (2015). Perceptual learning: Toward a comprehensive theory. Annual Review of Psychology, 66, 197–221.
Weiß, K., Schneider, W., & Herwig, A. (2015). A "blanking effect" for surface features: Transsaccadic spatial-frequency discrimination is improved by post-saccadic blanking. Attention, Perception, & Psychophysics, 77(5), 1500–1506. https://doi.org/10.3758/s13414-015-0926-1
Whitney, D., & Levi, D. M. (2011). Visual crowding: A fundamental limit on conscious perception and object recognition. Trends in Cognitive Sciences, 15(4), 160–168. https://doi.org/10.1016/j.tics.2011.02.005
Williams, L., Carrigan, A., Mills, M., Auffermann, W., Rich, A., & Drew, T. (2021). Characteristics of expert search behavior in volumetric medical image interpretation. Journal of Medical Imaging, 8(04), 041208. https://doi.org/10.1117/1.jmi.8.4.041208
Wolf, C., & Schütz, A. (2015). Trans-saccadic integration of peripheral and foveal feature information is close to optimal. Journal of Vision, 15(16), 1. https://doi.org/10.1167/15.16.1
Wurtz, R. H. (2008). Neuronal mechanisms of visual stability. Vision Research, 48(20), 2070–2089.
Wurtz, R. H. (2018). Corollary discharge contributions to perceptual continuity across saccades. Annual Review of Vision Science, 4, 215–237.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

*Study II*

# Frequency-specific and periodic masking of peripheral characters by delayed foveal input

Nedim Göktepe[1]*, Alexander C. Schütz[1,2]

[1]*AG Allgemeine und Biologische Psychologie, Philipps-Universität Marburg, Marburg, Germany*

[2]*Center for Mind, Brain and Behavior, Philipps-Universität Marburg, Marburg, Germany*

Abstract

The foveal-feedback mechanism supports peripheral object recognition by processing information about peripheral objects in foveal retinotopic visual cortex. When a foveal object is asynchronously presented with a peripheral target, peripheral discrimination performance is affected differently depending on the relationship between the foveal and peripheral objects. However, it is not clear whether the delayed foveal input competes for foveal resources with the information processed by foveal-feedback or masks it. In the current study, we tested these explanations by measuring the effect of foveal noise at different spatial frequencies on peripheral discrimination of familiar and novel characters. Our results showed that the impairment of foveal-feedback was strongest for low-spatial frequency noise. A control experiment revealed that for spatially overlapping noise, low-spatial frequencies were more effective than medium-spatial frequencies in the periphery, but vice versa in the fovea. This suggests that the delayed foveal input selectively masks foveal-feedback when it is sufficiently similar to the peripheral information. Additionally, this foveal masking was periodic as evidenced by behavioral oscillations at around 5 Hz. Thus, we conclude that foveal-feedback supports peripheral discrimination of familiar and novel objects by periodically processing peripheral object information.

## Introduction

The visual system is characterized by a division of labor between the foveal and peripheral visual field. Foveal vision allows for high-acuity and color vision in a small part of the visual field and the peripheral visual field allows for a large field of view at lower resolution (Malpeli & Baker, 1975; Curcio, Sloan, Kalina, & Hendrickson, 1990; Levi 2008; Strasburger, Rentschler, & Juttner, 2011; Rosenholtz, 2016; Simpson, 2017). Interestingly, peripheral and foveal processing are not independent of each other. Peripheral object information is fed back to foveal retinotopic cortex to support peripheral discrimination (Williams et al., 2008; Chambers, Allen, Maizey, & Williams, 2013; Fan et al. 2016, for a review see Stewart et al. 2020; Oletto et al. 2022). An impairment of peripheral object discrimination occurs when foveal-feedback is interrupted by TMS of foveal retinotopic areas (Chambers, Allen, Maizey, & Williams, 2013) or by asynchronously presenting a foveal stimulus (Yu & Shim, 2016). However, the time window and the precise informational content of this foveal-feedback mechanism is still unclear.

Previous results are somewhat inconclusive about the informational content and the temporal dynamics of foveal-feedback. For instance, while asynchronously presented foveal noise impairs peripheral discrimination, presenting a foveal object identical to the peripheral target improves peripheral discrimination. This suggests that the foveal input interferes with the foveal-feedback only when it does not match the peripheral object information processed by the foveal-feedback. This extends previous neuroimaging reports suggesting that the foveal-feedback processing is selective for the object's category (Williams et al., 2008; Fan et al. 2016). In stark contrast to neuroimaging results, a foveal object from the same category as the peripheral object does not lead to a stronger impairment than an object from another category (Weldon, Woolgar, Rich, & Williams, 2016). Thus, it is not clear to what extent the information processed by foveal-feedback is similar to the peripheral object. A plausible explanation accommodating these results would be that the foveal stimulus competes with the content of the foveal feedback(Weldon, Woolgar, Rich, & Williams, 2016). This implies that while an identical foveal stimulation supports peripheral discrimination, other foveal objects are detrimental to it. Therefore, when there is a competition between the foveal stimulus and the content of foveal-feedback, peripheral discrimination suffers. Another explanation is that the foveal object is not competing for foveal resources but acting like a mask (Stromeyer, & Julesz, 1972; White & Lorber, 1976; Legge, Cohen, & Stromeyer, 1978; Solomon & Pelli, 1994 Hutchinson, & Ledgeway, 2004; Bachmann, & Francis,

2014). Thus, the similarity between the peripheral and the foveal information determines the strength of impairment. However, when the mask is identical to the peripheral object, the foveal object no longer acts as a mask and aids foveal-feedback as a secondary source of object information. Current results do not distinguish if there is a competitive or a masking relationship between the foveal stimulus and the foveal-feedback.

Characters of written language present a unique opportunity to test these explanations and identify the informational content of foveal-feedback. Characters are one of the best studied visual objects and are highly relevant for daily life (Parish, & Sperling, 1991; Solomon & Pelli, 1994; Legge, Mansfield, & Chung, 2001; Majaj, Pelli, Kurshan, & Palomares, 2002; Grainger, & Dufau, 2008). On the one hand, they are more complex than simple visual features, which are immune to foveal-feedback interruption (Fan et al. 2016; Yu & Shim, 2016). On the other hand, they are still simple objects in the sense that they can be identified by a narrow spatial frequency band (Solomon & Pelli, 1994; Majaj et al., 2002). This makes them particularly useful to understand if the foveal object is competing or masking the foveal-feedback information. If the information processed by foveal-feedback is masked by the foveal object, then a foveal object that contains similar spatial frequencies as the peripheral object would impair the peripheral discrimination more than a foveal object that overlaps less in spatial frequencies. In contrast, if the foveal object is competing with the foveal-feedback information, then foveal objects would impair peripheral discrimination in a similar manner regardless of the overlap in spatial frequency bands. Thus, we tested the predictions of the “masking” and “competition” explanations by using foveal noises defined by different frequency bands.

As previous studies widely vary in their estimation of the time window of foveal-feedback (for a comparison see Table 1 in Oletto et al. 2022), we densely sampled the stimulus onset asynchrony (SOA) of the foveal noise. Recently, it has been shown that this variation might be due to foveal-feedback not having one but multiple critical time windows of reduced peripheral discrimination oscillating at 5 Hz (Contemori et al., 2022). Another reason to densely sample SOAs is the fact that the speed of visual processing depends on the spatial frequencies of the signal. This is captured by coarse-to-fine visual processing where low spatial frequency information is processed before high spatial frequency information (for a review see Kaufmann, Ramanoel, & Peyrin, 2014). Hence, dense sampling allows us to identify the time window of foveal-feedback, which might shift between different spatial frequencies of foveal noise.

Finally, foveal-feedback has been studied almost exclusively with trained (e.g., Weldon, Woolgar, Rich, & Williams, 2016; Weldon, Woolgar, Rich, & Williams, 2020). or familiar stimuli (Ramezani, Kheradpisheh, Thorpe, & Ghodrati, 2019). Thus, it is still unclear if foveal-feedback is restricted to familiar stimuli or if it also supports the peripheral discrimination of novel objects. Transsaccadic feature prediction is another mechanism that supports peripheral object discrimination by using both peripheral and foveal information. It supports discrimination by generating object predictions using previously associated peripheral and foveal object information (Herwig, & Schneider, 2014; Osterbrink, & Herwig, 2021; Poth, Herwig, & Schneider, 2015). We recently showed that peripheral discrimination of novel objects similarly benefits from these predictions albeit in a more general way and to a weaker extent (Goktepe & Schütz, 2023). To check if foveal-feedback applies to novel objects at all, we compared Korean characters that were novel to our observers and Latin characters that were over-trained through lifespan. If foveal-feedback depends on previous experience then foveal noise should impair only the peripheral discrimination of Latin characters.

## Results

### The temporal dynamics and the informational content of foveal-feedback

The main purpose of the current experiment was to investigate the information that is processed by foveal-feedback, how it interacts with foveal input (e.g., masking vs. competition) and to check if foveal-feedback generalizes to novel peripheral targets. To this end, we asked observers to perform a peripheral same-different judgment task using characters from their native script (Latin alphabet) and a foreign script (Korean). A bandpass-filtered foveal noise centered on one of three frequency bands was presented at different SOAs (see Fig. 1). Thus, any difference across foveal noise types provides evidence for masking of foveal feedback. Similarly, a reduction in peripheral discrimination of Korean characters would provide evidence for the generalizability of foveal-feedback to novel objects.

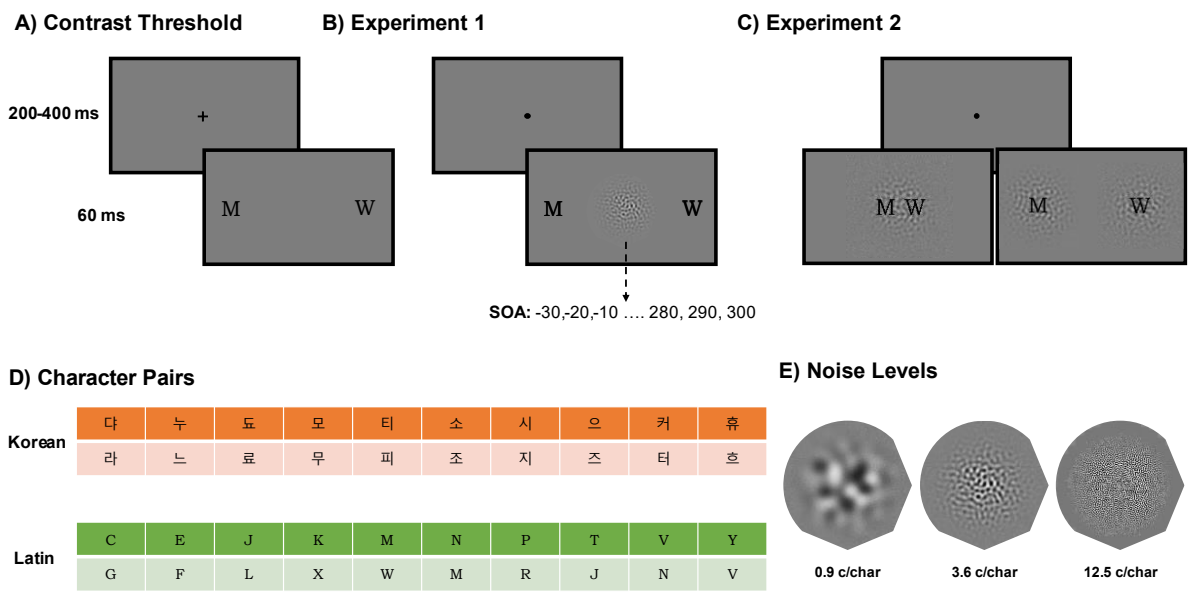



*Figure 1*. Stimuli and Methods. Example of trials from contrast threshold (A), Experiment 1 (B), and Experiment 2 (C). A, Experiment 1 used two QUEST procedures to determine the character contrast threshold of each observer. B, in Experiment 1 observers fixated to a central dot for a random duration of 200-400 ms. Following the offset of the fixation dot, either two Korean or Latin characters were presented 10° left and right of the fixation location for 60 ms. Low, Medium, or High level noise was presented at fixation for 60 ms at a stimulus onset asynchrony (SOA) of -30 to 290 ms. A-C, Stimuli are not drawn to scale for illustration purposes. D, List of Latin and Korean characters used in the experiments. E, Three frequencies of noise that were used in the experiments. As the width of the characters was approximately one degree of visual angle, one c/char$_{width}$ corresponds to one c/deg.

For each character and noise condition, we densely sampled a relatively wide time window of SOAs to determine the onset and the duration of foveal-feedback processing with a resolution much higher than most of the previous studies. As a result, we measured foveal-feedback in overlapping SOAs as Contemori et al. (2022). In contrast to Contemori et al. (2022), we calculated discrimination performance (d') for each SOA as the average of trials where the foveal noise was present at that time window (for details see Data Analysis). In other words, our analysis shows at which interval foveal noise impairs peripheral discrimination of characters.

Following previous studies (Weldon, Woolgar, Rich, & Williams, 2016; Weldon, Woolgar, Rich, & Williams, 2020), we identified the time window of foveal-feedback by checking the change in the relative discrimination performance across noise intervals within each noise and character conditions. To this end, we created 10000 permuted-datasets by shuffling the SOA labels of each trial separately for each observer and conditions. For each empirical and permuted-dataset in each condition, we set the mean discrimination performance to zero by subtracting the overall

mean across SOAs. Hence, negative values in Fig. 2 indicate a reduction in performance compared to the mean. The top 95th percentile performance of the permuted-datasets in each condition (dashed lines in fig. 2) served as the critical value of the underlying omnibus statistical test that is equivalent to a one-sided test at α= 0.05.

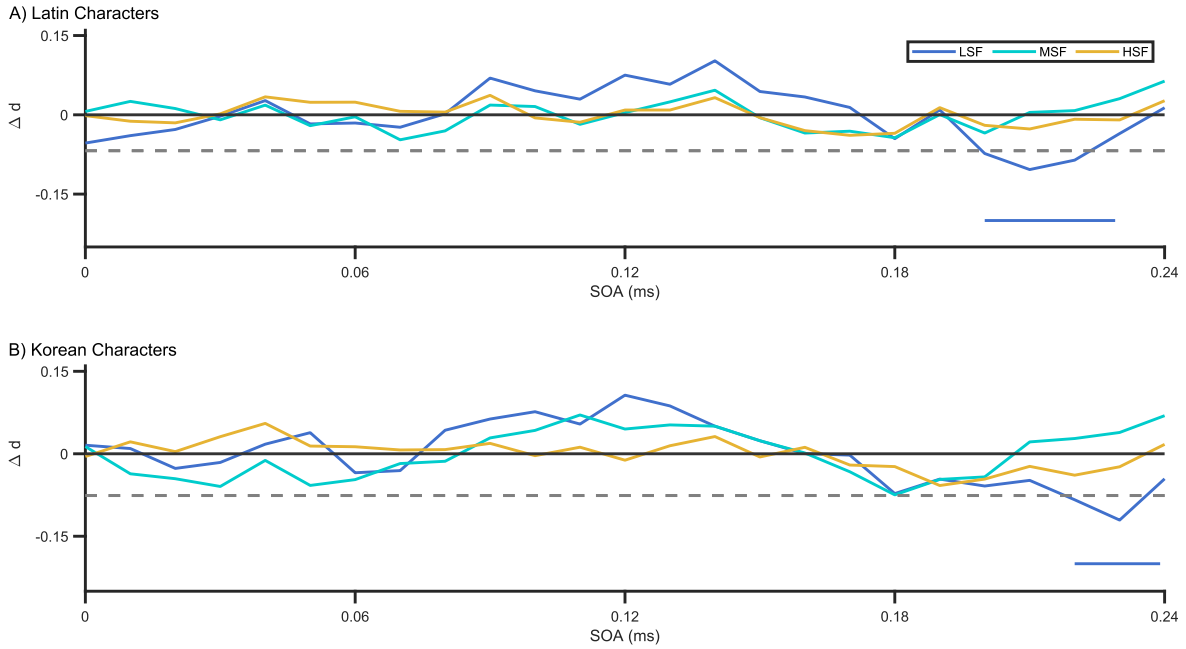



Figure 2. Change in peripheral discrimination performance (*d'*) for Latin (A) and Korean (B) characters with low-, medium-, and high-spatial frequency foveal noise. Negative values indicate discrimination performance below the mean. The dashed line shows the 95th percentile of the permuted performance corresponding to one-sided statistical test (for illustration purposes only the smallest critical value is shown). The lower horizontal bars indicate SOAs with significant reductions in performance.

The temporal profile of peripheral discrimination was remarkably similar for each character and noise condition. In both conditions, discrimination performance in all noise conditions was lower than the no-noise baseline condition (see Appendix 1). The peripheral discrimination performance appeared particularly vulnerable to foveal noise for the period of 90 ms after the stimulus onset (0 SOA). During the first 60 ms of this period, the foveal stimulus and peripheral targets were both present at the same time. Therefore, this period of vulnerability could be also attributed to divided attention (Fan et al. 2016). The following period between 90-170 was characterized by a relatively stronger peripheral discrimination performance and a recovery of performance. Following the period of recovery, the discrimination performance declined again at 180 ms after stimulus onset. The interval between 180-190 ms was a period of major reduction in

all condition, which overlaps with the effective time windows reported by multiple studies (Yu & Shim, 2016; 150-200 ms; Fan et al. 2016, 150-233 ms; Weldon et al. 2016; 117-234 ms; Weldon et al., 2020 117-234; Contemori et al. 2022, 2023 94-177 ms). However, the period of significant reduction we observed was slightly later around 220 ms and was present only for the LSF noise condition. Peripheral discrimination of Latin characters was significantly impaired by LSF noise at 200 ms ($p$ = 0.038), 210 ms ($p$ = 0.006), and 220 ms ($p$ = 0.019). Similarly, discrimination of Korean characters was significantly impaired at 220 ms ($p$ = 0.035) and 230 ms ($p$ = 0.005) by LSF noise. At a first glance, this significant dip after the common 180-190 interval looks at odds with previous results. However, most of the foveal-feedback periods reported by studies employing fewer and wider noise time windows. Thus, our noise sampling allowed us to detect critical periods within the reported periods with higher precision. Hence, like the major dip at 180 ms, the significant reduction period around 220 ms is also consistent with previous reports and maybe the epicenter of the foveal-feedback period. Critically, while the reduction of 180 ms was common in all noise conditions, the significant reduction around 220 ms was specific to LSF noise.

We then compared the strength of foveal noise modulation across noise conditions separately for each character condition. To this end, we used the SOAs that yielded the strongest impairment in each noise condition. This allowed us to avoid any temporal bias in the analysis as different spatial frequencies could mask foveal-feedback at different time intervals. By means of shuffling noise condition labels for each participant we obtained 10000 permuted-datasets (for details see Data Analysis). Next, we checked if the differences between the observed conditions are more extreme than the central 95$^{th}$ percentile of differences in the respective differences in the permuted-datasets. Thus, for each noise comparison we tested significance with a two-tailed statistical test at $\alpha = 0.05$. We found that for Latin characters, LSF noise impaired performance significantly stronger than MSF ($p$ = 0.042) and HSF noise ($p$ = 0.042). There was no significant difference between MSF and HSF noise ($p$ = 0.828). The discrimination performance for the Korean characters followed the same trend but LSF was only significantly stronger than HSF ($p$ = 0.034). The differences between LSF and MSF ($p$ = 0.057) and MSF and HSF ($p$ = 0.668) did not reach significance.

So far, our analysis compared differences in discrimination performance between SOAs. Comparing discrimination performance in trials with different SOAs essentially checks if there are temporally limited periods of major reductions. This is highly desirable as the foveal noise could

impact discrimination performance in other ways, for instance as a distractor (Fan et al. 2016). However, it is a conservative method that would completely miss an effect of periodic reductions in multiple time intervals. To capture potential periodic reductions, we conducted a complementary frequency analysis reported in the next section.

Periodicity of foveal-feedback

It has been recently suggested that foveal-feedback is a recurrent process instead of a one-off occurrence (Contemori et al., 2022). The periodicity of a mechanism is typically measured by analyzing the oscillatory signals from EEG or from perceptual performance (Drewes & VanRullen, 2011; Landau & Fries, 2012; Fiebelkorn, Saalmann, & Kastner 2013; Tomassini et al., 2015; Benedetto & Morrone, 2017 ; Zhang, Morrone, & Alais, 2019). Like neural oscillations, behavioral oscillations are analyzed using Fourier transformations to estimate the contribution of each frequency band. Contemori et al. (2022) showed that the influence of a central noise on the peripheral discrimination performance oscillates at around 5 Hz. In the current study, we found that LSF noise is more effective than MSF and HSF noise. Therefore, if the behavioral oscillations reported earlier underly foveal-feedback processes, then we should find similar oscillations to be stronger in the LSF condition. To link foveal-feedback to the periodic changes in the perceptual performance, we analyzed the Fourier spectrum of each character and noise condition. In order to identify significant peaks in amplitudes, we generated permuted amplitudes from the shuffled-datasets and perform a one-tailed omnibus statistical test with $\alpha = 0.05$ for each noise condition (see Data Analysis for details). We found frequency amplitudes peaking at around 5 Hz for LSF and MSF noise for both character conditions. As expected, in the LSF condition, amplitudes for these frequencies were significantly higher than the permuted amplitudes (dashed lines in Fig. 3). For the Latin characters these frequencies ranged from 0.79 to 7.94 Hz ($p_{max}$= 0.035, $p_{min}$ = 0.01). For Korean characters we found that similar frequencies from 0.79 to 8.73 Hz ($p_{max}$= 0.041, $p_{min}$ = 0.011) had significantly more amplitude. In contrast, in the MSF noise condition the peak was located around 8 Hz. More importantly, frequencies ranging from 5.56 to 8.73 Hz had significantly more power only in the Korean character condition ($p_{max}$= 0.045, $p_{min}$ = 0.034). This is probably reflecting the common dip at 180 ms, which appeared to be more pronounced for Korean than Latin characters.

Next, we compared the amplitudes across noise conditions by analyzing the Fourier spectrum of the permuted data generated for the noise comparisons (for more details see Data

Analysis). Then, we calculated the amplitude differences between the maximum amplitude in each real and permuted noise condition. Next, we used the central 95th percentile of the permuted differences to assess statistical significance at $\alpha = 0.05$ with a two-tailed test. We found that for the discrimination of Latin characters, the LSF noise condition had stronger oscillations than MSF ($p = 0.009$) and HSF ($p < 0.001$). The amplitude difference between MSF and HSF noise conditions did not reach significance ($p = 0.85$). For Korean characters, LSF noise had the highest amplitude but none of the comparisons reached significance (LSF vs MSF: $p = 0.423$; LSF vs HSF: $p = 0.057$; MSF vs HSF: $p = 0.276$).

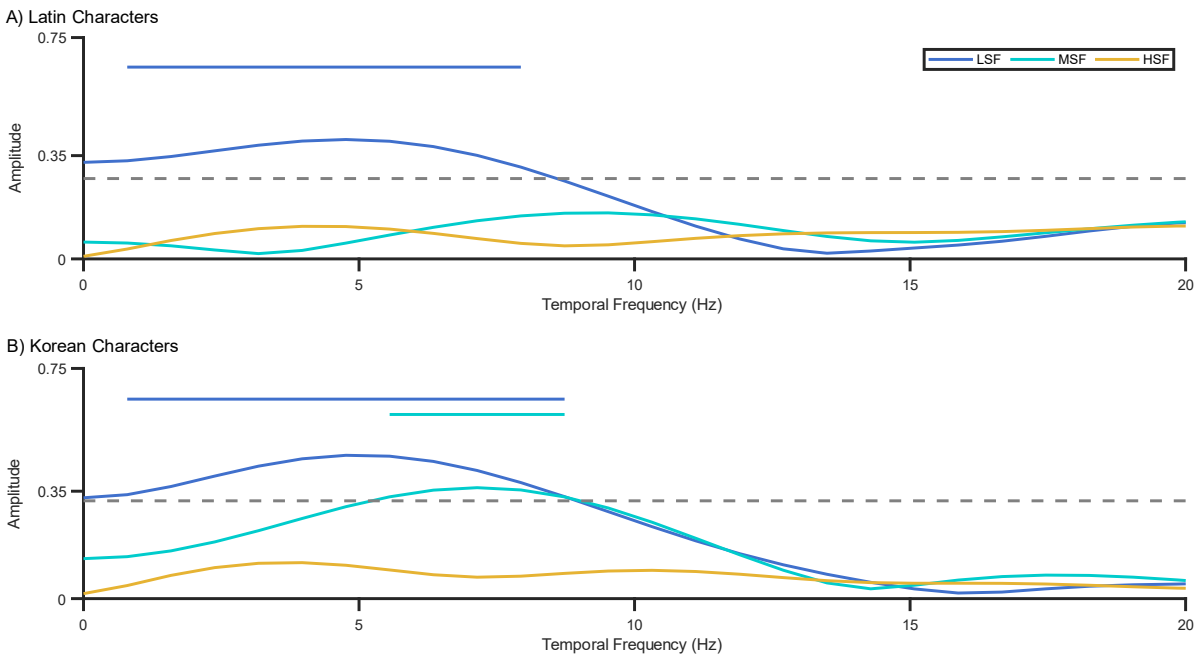



Figure 3. Perceptual performance oscillations for Latin (A) and Korean (B) characters across LSF, MSF, and HSF conditions. Dashed line shows 95th percentile amplitude calculated from permuted data corresponding to a one-sided statistical test (for illustration purposes only the largest critical value is shown). Horizontal bars indicate frequencies with a significantly larger amplitude.

Overall, we found a significant reduction of discrimination performance around the SOA of 220 ms for the LSF noise condition for both character conditions. Therefore, we suggest that foveal-feedback is also supporting the peripheral discrimination of novel objects. LSF noise was particularly detrimental to discrimination performance and led to stronger impairments. This frequency-specific effect is more compatible with masking rather than competition or attention (Fan et al. 2016).

Foveal-feedback follows spatial frequency processing

Previous studies provided some evidence that the informational content of foveal-feedback is specific at least at the level of object category (Williams et al., 2008; Fan et al. 2016; Yu & Shim, 2016, but see Weldon et al. 2016). So far, our results suggest that peripheral discrimination performance of characters was particularly impaired by LSF noise acting as a mask. The specific effect also implies that foveal-feedback processes information that is more vulnerable to a specific type of foveal input. This systematic impairment due to masking of LSF noise provides us with a unique opportunity to better describe what is exactly processed by foveal-feedback. The superiority of LSF noise over MSF and HSF can be explained by fundamental spatial frequency (Parish, & Sperling, 1991; Solomon & Pelli, 1994; Legge, Mansfield, & Chung, 2001; Majaj, Pelli, Kurshan, & Palomares, 2002; Grainger, & Dufau, 2008). Within this framework, the most effective foveal noise should be the most effective mask for the information processed by foveal-feedback. This means that foveal-feedback is processing object information that has a spatial frequency distribution very similar to LSF noise. Considering previous evidence suggesting that the foveal-feedback is processing peripheral object information, LSF noise should be masking Korean and Latin characters more efficiently than the MSF or HSF noise. If LSF noise is better at masking foveal-feedback information, then our peripheral characters should be processed by the same frequency channels that are responsible for processing LSF noise. Hence, if it is the precise object information encoded by peripheral retinae processed by foveal-feedback, then LSF noise should be better at masking peripheral characters than other frequency bands.

It is well established that character processing in the fovea occurs at a higher spatial frequency band, equivalent to our MSF noise (Majaj et al., 2002). Therefore, if the peripheral characters are masked by the foveal noise according to foveal processing of letters, the strongest performance reduction would be expected for MSF instead of LSF noise. However, peripheral processing of letters is supported by frequency channels closer to LSF than MSF noise (Anderson & Thibos, 1999; Chung, Legge, & Tjan, 2002). Hence, if the foveal noise is masking the peripheral information processed by foveal-feedback, then LSF should be a more effective mask because the spatial frequency composition of the peripheral information is more similar to LSF than MSF. Thus, we conducted a follow-up experiment to test whether peripheral LSF or MSF noise is a better mask for peripheral character discrimination. We measured the discrimination performance at 0° (fixation) and 10° eccentricity for Latin and Korean characters (Fig. 1). Both characters were

embedded in LSF or MSF noise with the same RMS contrast. Then, using a staircase method, we varied the character contrast to determine the discrimination threshold of observers.

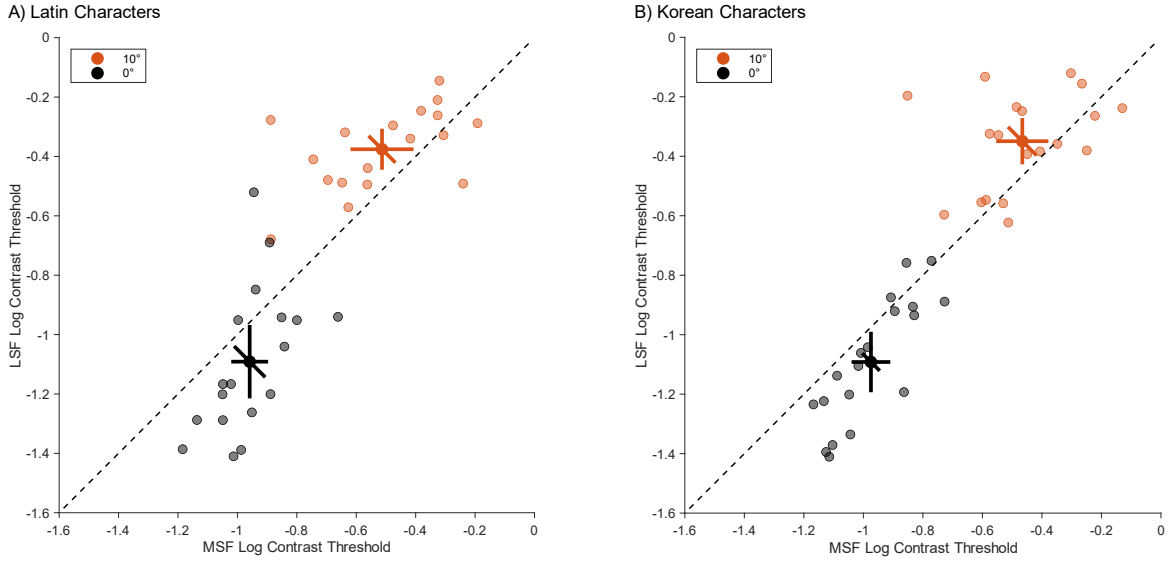



Figure 4. Character contrast thresholds for Latin (A) and Korean (B) characters at 0° (fixation) and 10° eccentricity. Light colors indicate individual participants and dark color the average across participants. Error bars represent 95% confidence intervals for MSF noise (horizontal), LSF noise (vertical), and their pairwise comparisons (diagonal).

Figure 4 shows character contrast thresholds of each observer for peripheral and foveal discrimination judgments of characters embedded in LSF or MSF noise. For both character conditions, character contrast thresholds were higher for LSF noise than MSF noise for targets presented at 10° eccentricity. To compare contrast thresholds against both noise and character conditions we used a Linear Mixed Model (LMM), where character, noise frequency, and stimulus eccentricity were fixed factors and participants was a random factor. The model yielded significant main effects of stimulus eccentricity ($t(119) = -13.6$, $p < 0.001$) and noise frequency ($t(119) = -2.136$, $p = 0.035$) but no effect of character ($t(199) = -0.394$, $p = 0.694$). Crucially, there was a significant interaction between noise frequency and target eccentricity suggesting that the effectiveness of noise masking depended on stimulus eccentricity ($t(119) = 3.031$, $p = 0.003$, see Methods for complete statistical results). Post-hoc Holm-Bonferroni corrected comparisons on eccentricity and noise interaction showed that contrast thresholds were higher for MSF than LSF noise at fixation ($p = 0.011$), as expected. Crucially, in line with previous studies (Anderson & Thibos, 1999; Chung, Legge, & Tjan, 2002) the pattern was reversed for characters presented at 10° eccentricity and contrast thresholds were higher for LSF than for MSF noise ($p = 0.009$). Hence, we suggest that the effectiveness of foveal noise at impairing peripheral discrimination

depends on the efficacy of the noise for peripheral processing and not for foveal processing. Moreover, our results provide the first direct evidence that foveal-feedback is processing peripheral information that is vulnerable to masking.

Discussion

The foveal-feedback mechanism supports peripheral object discrimination by processing information related to peripheral objects in foveal retinotopic visual cortex (Williams et al., 2008; Fan et al. 2016). Previous studies provided mixed evidence about the informational content of foveal-feedback and its relation to peripheral objects: It was not clear whether the input from foveal object competes or masks the information processing by the foveal-feedback. In the current study, we tested these explanations by varying the relationship between the peripheral target and the foveal noise. According to the competition explanation, all foveal noises should have comparable effects on peripheral discrimination as they all compete with foveal-feedback. In contrast, according to the masking explanation, the similarity of the foveal and peripheral objects should dictate how foveal-feedback will be affected. Our results favor the latter argument as the foveal-feedback impairment was stronger for noise with low- than with medium- or high-spatial frequency. In Experiment 2, we showed that the foveal-feedback follows spatial frequency processing of peripheral vision as low-spatial frequency noise was more efficient at masking peripheral information than medium-spatial frequency noise.

Previous studies described foveal-feedback as a mechanism supporting the discrimination of complex objects (Fan et al. 2016; Yu & Shim, 2016), but not simple features. While higher-level objects such as rendered complex objects (Weldon, Woolgar, Rich, & Williams, 2016; Weldon, Woolgar, Rich, & Williams, 2020) or complex objects from daily life (Ramezani, Kheradpisheh, Thorpe, & Ghodrati, 2019). are clearly affected by interrupting foveal-feedback, the discrimination of simple visual features like motion (Fan et al. 2016) and orientation(Yu & Shim, 2016) does not seem be affected by interrupting foveal-feedback in the same way. Our results suggest that the interruption of foveal-feedback processing not only depends on the properties of the peripheral information but also on the relationship between peripheral and foveal information. Thus, we argue that foveal-feedback may not be selective in which type of object it supports, instead it might be processing all kinds of objects but with varying degree. The degree of support might depend on factors such as how resolvable the peripheral object is by peripheral vision alone and the spatial

frequency composition. The absence of foveal-feedback for simple features could also be caused by using noise that was not effectively masking the foveal-feedback processing for those features.

Another topic of debate is the time window of foveal-feedback. The majority of previous studies presented foveal noise for long presentation durations at few, temporally spread-out SOAs, allowing only for a coarse temporal estimation of foveal-feedback. In Experiment 1 we used overlapping and dense sampling of SOAs to obtain a more precise estimation. Unlike Contemori et al. (2022) who followed a similar approach, we separately analyzed noise periods of 10 ms. With this finer sampling method, we showed that previous seemingly inconsistent reports of foveal-feedback onsets might be in fact consistent with one and another.

More fundamental questions for future research are what foveal retinotopic cortex brings to the table by reprocessing peripheral information that is reduced to the resolution of the peripheral visual field and how the visual system bridges the differences in processing between the peripheral and the foveal visual field, such that information can be exchanged efficiently. Previous studies showed that the visual system can learn object-specific associations that connect peripheral and foveal object information (Herwig & Schneider, 2014; Osterbrink & Herwig, 2021; Poth, Herwig, & Schneider, 2015; Goktepe & Schütz, 2023). This mechanism supposedly helps to align the peripheral and foveal representation of objects and to compensate for the differences in processing. This object-specific mechanism is supplemented by a more generative transformation between peripheral and foveal information because peripheral-to-foveal matching is also possible for completely unfamiliar objects[31]. Our finding of foveal-feedback effect for both, highly-overtrained (Latin characters) and novel (Korean characters) objects suggests that these peripheral-foveal associations and foveal feedback processing are somewhat independent of each other. On the one hand, foveal-feedback processing supports peripheral discrimination even when there is extensive experience and as a consequence, strong peripheral-foveal associations for a certain stimulus like the Latin characters in our study. This means that the presence of strong peripheral-foveal associations does not render foveal-feedback processing superfluous. On the other hand, foveal-feedback processing can support peripheral discrimination even when the objects are novel and have not been observed in dynamic, transsaccadic conditions like the Korean characters in our study. This means that peripheral-foveal associations are not necessary to accomplish foveal-feedback processing.

In conclusion, we found that foveal noise is acting like a spatial frequency mask for the foveal-feedback processing of peripheral objects. This means that the processing of foveal-feedback depends on the spatial-frequency composition of the peripheral information rather than on the spatial-frequency preference of foveal processing. The disruption of foveal-feedback was detrimental for novel and familiar objects alike, which means on the one hand that foveal-feedback processing does not require specific peripheral-foveal associations and on the other hand that experience with an object does not render foveal-feedback processing superfluous. In sum, we describe foveal-feedback as a mechanism that supports peripheral discrimination of both familiar and novel objects by periodically processing peripheral object information.

## Methods

### *Participants*

A total of 36 observers including the author N.G. completed Experiment 1. 26 observers (19 females, mean age = 24.38 ± 6.02) performed the *Korean characters* condition and 31 observers (22 females, mean age = 25.26 ± 5.65) performed the *Latin characters* condition.

In Experiment 2, 26 observers (17 females, mean Age = 27.38 ± 5.28) performed the discrimination task with both Korean and Latin characters. After exclusions, the analysis included the data of 19 observers (13 females, mean Age = 26.74 ± 5.62) for Korean and 18 observers (11 females, mean Age = 27.22 ± 5.7) for Latin characters.

All observers had normal or corrected to normal vision without any known visual condition. They all signed informed consents prior to testing and received monetary compensation. The experiment was approved by the local ethics commission of the Department of Psychology of Marburg University (proposal number 2015-35k) and were conducted in accordance with the Declaration of Helsinki (1964).

### *Stimuli*

A subset of modern Latin characters (A, C, D, E, F, G, H, J, K, L, M, N, P, R, T, U, V, W, X, Y) and a subset of Korean characters (댜, 누, 됴, 모, 티, 소, 시, 으, 커, 휴, 버, 쿄, 라, 느, 료, 무, 피, 조, 지, 즈, 터, 흐, 비, 크) rendered in the Bookman Old Style font served as stimuli. Each character subtended an area of approximately 1 × 1° on the display. For “same” trials in one block, 10 characters were randomly drawn from the respective character pool. For “different” trials, a set

of 10 character pairs was used for all observers (see Fig. 1 for the pairs). Low (LSF), medium (MSF), and high (HSF) spatial frequency noises subtended an area of approximately 6.5°× 6.5° and were bandpass (Gaussian) filtered random pink noises. Filtered noises had central frequencies of 0.9, 3.6, and 12.5 c/character and a band width of 1.6 octaves. Therefore, there was no frequency overlap between the three types of noise. They were rendered at fixation within a Gaussian envelope with unit standard deviation. They were displayed at a 20% RMS contrast which is more suitable for comparing stimuli with different spatial frequencies than Michelson contrast (Kukkonen, Rovamo, Tiippana, & Näsänen, 1993). In Experiment 2, the same characters and noises were used with the following differences: The characters were displayed at various contrasts to measure contrast thresholds. Noises were used as a background for foveally or peripherally presented characters at 5% RMS contrast.

*Equipment*

The stimuli were displayed at a 106 cm distance on a back-projection system comprising of a 120 Hz PROPixx projector (VPixx Technologies, Saint Bruno, QC, Canada) with 1920 × 1080 resolution and a 91 × 51 cm Stewart Filmscreen (Torrance, CA). Prior to the experiment, gamma and hot spot corrections were carried out to ensure linear luminance output. After calibration, the display had a luminance of 2.07, 71, and 140 cd/m$^2$ for black, grey, and white pixels, respectively. An Eyelink 1000+ (SR Research Ltd., ON, Canada) was used to record eye movements. Eye movements were recorded at 1000 Hz sampling rate after a 9-point calibration. Calibration points were positioned at the screen center and 5 degrees around it. The experiment was programmed in MATLAB (MatWorks. Inc.) using Psychtoolbox (Brainard, 1997; Kleiner, Brainard, Pelli, Ingling, Murray, & Broussard, 2007) and EyelinkToolbox (Cornelissen, Peters, & Palmer, 2002).

*Procedure*

Experiment 1

In Experiment 1, both Korean and Latin characters conditions followed the same procedure and were each completed in three separate sessions. Each session consisted of a thresholding block and an experimental block.

The thresholding procedure contained two QUESTs (Watson, 2017) to determine individual character contrast values necessary for achieving 75% accuracy without a foveal noise. At the beginning of each trial, a fixation cross was presented for a random duration of 200-400 ms on a

neutral background. Following the fixation offset, two characters were presented at 10° on the horizontal meridian. Observers were asked to indicate whether these characters were the same or different. Observers were instructed to keep fixation at the center for the full duration of the trial. The mean of the two contrast values suggested by QUESTs was used in the experimental block. Following the eye tracker calibration, observers proceeded to the experimental trials.

Experimental trials consisted of no-noise (baseline), LSF, MSF, and HSF conditions. The no-noise condition was similar to the thresholding trials, but with a fixed stimulus contrast. In the noise trials, a foveal noise was presented at a random stimulus onset asynchrony (SOA). At the beginning of each session SOAs between -30 and 290 ms were randomly selected from a uniform distribution and repeated equally in each noise condition. An SOA of zero reflects simultaneous onset of the stimulus and noise. Hence, for negative SOAs, the noise preceded the stimulus. Each trial started with a fixation check followed by a fixation cross presented for a random duration between 200-400 ms. 60 ms after the offset of the fixation cross, the characters were presented for 60 ms. For noise trials, the foveal noise was presented at the respective SOA for 60 ms. If the gaze of observers left a central 6.5 × 6.5° square at any time during the trial, they received an auditory warning at the end of that trial. Each session consisted of 80 threshold trials and 620 test trials (i.e., 20 no-noise trials, and 200 trials per noise condition) in randomized order, which took about two hours.

Experiment 2

Experiment 2 used a 2-up-1-down staircase procedure to determine the discrimination threshold for characters directly embedded in either LSF or MSF noise presented at 0° (fixation) or 10°. The embedded peripheral characters were at the center of a 6.5° × 6.5° LSF or MSF noises, whereas centrally presented characters were approximately 0.7° degrees apart and embedded within LSF or MSF noise (6.5° × 6.5°). After giving their consent, Experiment 2 started with eye tracking calibration. Characters from each condition were presented in blocks. Each block was further divided into two subblocks for the foveal and 10° eccentricity conditions. Each block contained one ascending and one descending staircase. The staircases were interleaved and the character contrast was reduced after two consecutive correct responses or increased after an incorrect response. The order of blocks and subblocks was randomized. Each trial started with a fixation check, followed by the 60 ms presentation of a pair of characters within one of the two

noise backgrounds. Observers again performed a same-different judgment. The experiment had 640 trials and took about 1.5 hours.

Data Analysis

Experiment 1

Trials with erroneous fixations were discarded (< 1% of trials per observer). In Experiment 1, for each character and foveal noise condition, noise SOAs were binned with 10 ms increments based on the vertical retrace time stamps of the display. The refresh rate of the display (120 Hz) did not allow to sample a comparable number of trials for the -10 SOA. Therefore, these trials were also discarded (< 2 % of all trials). Then, for calculating the discrimination performance for each SOA, all trials with a central noise present within the 10 ms interval were used. Therefore, we ran a moving average with 60 ms window to pool together temporally overlapping SOAs. Thus, for calculating the bin at 150 ms SOA, trials with noise onset asynchronies from 100 ms to 150 ms were used.

For SOA comparison, we used an omnibus one-tailed permutation test. For this, we created 10000 permuted datasets by reshuffling the perceptual responses across trials of the same condition and observer. Permuted datasets were then analyzed in the same way as the actual dataset. For each SOA in each noise and character condition, the top $95^{th}$ percentile of the permuted datasets was used as the critical value. Any SOAs with an observed value less than the critical value were deemed significant at $\alpha = 0.05$ for a one-tailed test.

For comparing the strength of impairment between noise conditions, we generated another 10000 permuted datasets for each noise comparison by shuffling the noise labels across trials of the respective conditions. In other words, for comparing LSF and MSF noise, we shuffled the noise labels of trials only with LSF or MSF noise. Then, as previously, we calculated the group level performance of each permuted dataset separately for the two noise conditions in question. We took the minimum value for each permuted dataset in each noise condition. The middle $95^{th}$ percentile of the differences between the two noise conditions for the permuted datasets served as the critical value. Any observed difference more extreme than the critical value was deemed significant at $\alpha = 0.05$ for a two-tailed test.

For analyzing the behavioral oscillations, i.e. the periodicity of changes in peripheral discrimination performance, we analyzed the Fourier power of the SOA bins tapered by Hanning

window. As Contemori et al. (2023), we obtained amplitudes for frequencies ranging from 0-20 Hz. We used the previously generated permuted datasets for statistical testing. The permuted datasets followed the same steps as the empirical data. The top 95$^{th}$ percentile power for each condition from the permuted datasets served as a critical value for a one-tailed significance test at $\alpha = 0.05$. Amplitudes were compared across noise conditions were calculated similarly to the noise comparison analysis by Fourier transforming the permuted datasets created for the noise comparison. The difference between the maximum amplitude in the permuted datasets was used to determine the critical value for testing the real difference for the respective noise conditions. Any observed difference more extreme than the critical value was deemed significant at $\alpha = 0.05$ for a two-tailed test.

Experiment 2

In Experiment 2, we fitted psychometric functions for each observer, eccentricity, background and character condition using the psignifit toolbox for MATLAB (Wichmann & Hill, 2001). We defined the contrast threshold of each participant as the contrast yielding 70.7 % accuracy according to their psychometric function. Observers with psychometric functions that did not reach the target accuracy were excluded from the analysis of the respective character condition. One additional observer was excluded due to poor fixation throughout the experiment. After exclusions, seven observers had valid thresholds only in one of the two character conditions. Because of this we conducted a linear mixed model (LMM) with observer as random effects and eccentricity (0° vs 10°), noise frequency (LSF vs MSF), and character (Latin vs Korean) as fixed effects. In addition to the main effects and interaction reported in *Results*, the model yielded no significant interaction between character and eccentricity ($t(119) = 0.346$, $p = 0.73$), or between character and noise frequency ($t(119) = -0.267$, $p = 0.79$) or between character, eccentricity, and noise frequency ($t(119) = 0.326$, $p = 0.745$).

References

Anderson, R. S., & Thibos, L. N. (1999). Relationship between acuity for gratings and for tumbling-e letters in Peripheral Vision. *Journal of the Optical Society of America A*, *16*(10), 2321. doi:10.1364/josaa.16.002321

Bachmann, T., & Francis, G. (2014). Visual masking. *Visual Masking*, 1–108. https://doi.org/10.1016/b978-0-12-800250-6.00001-7

Benedetto, A., & Morrone, M. C. (2017). Saccadic suppression is embedded within extended oscillatory modulation of sensitivity. *The Journal of Neuroscience*, *37*(13), 3661–3670. https://doi.org/10.1523/jneurosci.2390-16.2016

Brainard, D. H. (1997) The Psychophysics Toolbox, Spatial Vision 10:433-436.

Chambers, C. D., Allen, C. P. G., Maizey, L., & Williams, M. A. (2013). Is delayed foveal feedback critical for extra-foveal perception? *Cortex*, *49*(1), 327–335. https://doi.org/10.1016/j.cortex.2012.03.007

Chung, S. T. L., Legge, G. E., & Tjan, B. S. (2002). Spatial-frequency characteristics of letter identification in central and Peripheral Vision. *Vision Research*, *42*(18), 2137–2152. doi:10.1016/s0042-6989(02)00092-5

Contemori, G., Oletto, C. M., Cessa, R., Marini, E., Ronconi, L., Battaglini, L., & Bertamini, M. (2022). Investigating the role of the foveal cortex in peripheral object discrimination. *Scientific Reports*, *12*(1). https://doi.org/10.1038/s41598-022-23720-w

Cornelissen, F. W., Peters, E. M., & Palmer, J. (2002). The Eyelink Toolbox: Eye tracking with MATLAB and the psychophysics toolbox. *Behavior Research Methods, Instruments, & Computers*, *34*(4), 613–617. https://doi.org/10.3758/bf03195489

Curcio, C. A., Sloan, K. R., Kalina, R. E., & Hendrickson, A. E. (1990). Human photoreceptor topography. *The Journal of Comparative Neurology*, *292*(4), 497–523. https://doi.org/10.1002/cne.902920402

Drewes, J., & VanRullen, R. (2011). This is the rhythm of Your eyes: The phase of ongoing electroencephalogram oscillations modulates saccadic reaction time. *The Journal of Neuroscience*, *31*(12), 4698–4708. https://doi.org/10.1523/jneurosci.4795-10.2011

Fan, X., Wang, L., Shao, H., Kersten, D., & He, S. (2016). Temporally flexible feedback signal to foveal cortex for peripheral object recognition. *Proceedings of the National Academy of Sciences*, *113*(41), 11627–11632. https://doi.org/10.1073/pnas.1606137113

Fiebelkorn, I. C., Saalmann, Y. B., & Kastner, S. (2013). Rhythmic sampling within and between objects despite sustained attention at a cued location. *Current Biology*, *23*(24), 2553–2558. https://doi.org/10.1016/j.cub.2013.10.063

Goktepe, N., & Schütz, A. C. (2023). Familiar objects benefit more from transsaccadic feature predictions. *Attention, Perception, & Psychophysics*. https://doi.org/10.3758/s13414-022-02651-8

Grainger, J., Rey, A., & Dufau, S. (2008). Letter perception: From pixels to pandemonium. *Trends in Cognitive Sciences*, *12*(10), 381–387. https://doi.org/10.1016/j.tics.2008.06.006

Herwig, A., & Schneider, W. (2014). Predicting object features across saccades: Evidence from object recognition and visual search. *Journal Of Experimental Psychology: General*, *143*(5), 1903-1922. https://doi.org/10.1037/a0036781

Hutchinson, C. V., & Ledgeway, T. (2004). Spatial frequency selective masking of first-order and second-order motion in the absence of off-frequency `looking'. *Vision Research*, *44*(13), 1499–1510. https://doi.org/10.1016/j.visres.2004.01.014

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

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in psychtoolbox-3. Perception, 36(14), 1-16.

Kukkonen, H., Rovamo, J., Tiippana, K., & Näsänen, R. (1993). Michelson contrast, RMS contrast and energy of various spatial stimuli at threshold. *Vision research*, *33*(10), 1431-1436.

Landau, A. N., & Fries, P. (2012). Attention samples stimuli rhythmically. *Current Biology*, *22*(11), 1000–1004. https://doi.org/10.1016/j.cub.2012.03.054

Legge, G. E., Cohen, M. A., & Stromeyer, C. F. (1978). Spatial-frequency masking with briefly pulsed patterns. *Perception*, *7*(2), 161–166. https://doi.org/10.1068/p070161

Legge, G. E., Mansfield, J. S., & Chung, S. T. L. (2001). Psychophysics of reading. *Vision Research*, *41*(6), 725–743. https://doi.org/10.1016/s0042-6989(00)00295-9

Levi, D. M. (2008). Crowding—an essential bottleneck for object recognition: A mini-review. *Vision Research*, *48*(5), 635–654. https://doi.org/10.1016/j.visres.2007.12.009

Majaj, N. J., Pelli, D. G., Kurshan, P., & Palomares, M. (2002). The role of spatial frequency channels in letter identification. *Vision Research*, *42*(9), 1165–1184. https://doi.org/10.1016/s0042-6989(02)00045-7

Malpeli, J. G., & Baker, F. H. (1975). The representation of the visual field in the lateral geniculate nucleus of Macaca Mulatta. *The Journal of Comparative Neurology*, *161*(4), 569–594. https://doi.org/10.1002/cne.901610407

Oletto, C. M., Contemori, G., Bertamini, M., & Battaglini, L. (2022). The role of foveal cortex in discriminating peripheral stimuli: The sketchpad hypothesis. *NeuroSci*, *4*(1), 9–17. https://doi.org/10.3390/neurosci4010002

Osterbrink, C., & Herwig, A. (2021). Prediction of complex stimuli across saccades. *Journal of Vision*, *21*(2), 10. https://doi.org/10.1167/jov.21.2.10

Parish, D. H., & Sperling, G. (1991). Object spatial frequencies, retinal spatial frequencies, noise, and the efficiency of letter discrimination. *Vision Research*, *31*(7–8), 1399–1415. https://doi.org/10.1016/0042-6989(91)90060-i

Poth, C. H., Herwig, A., & Schneider, W. X. (2015). Breaking object correspondence across saccadic eye movements deteriorates object recognition. *Frontiers in Systems Neuroscience*, *9*, 176.

Ramezani, F., Kheradpisheh, S. R., Thorpe, S. J., & Ghodrati, M. (2019). Object categorization in visual periphery is modulated by delayed foveal noise. *Journal of Vision*, *19*(9), 1. https://doi.org/10.1167/19.9.1

Rosenholtz, R. (2016). Capabilities and limitations of Peripheral Vision. *Annual Review of Vision Science*, *2*(1), 437–457. https://doi.org/10.1146/annurev-vision-082114-035733

Simpson, M. J. (2017). Mini-review: Far Peripheral Vision. *Vision Research*, *140*, 96–105. https://doi.org/10.1016/j.visres.2017.08.001

Solomon, J. A., & Pelli, D. G. (1994). The visual filter mediating letter identification. *Nature*, *369*(6479), 395–397. https://doi.org/10.1038/369395a0

Stewart, E. E., Valsecchi, M., & Schütz, A. C. (2020). A review of interactions between peripheral and foveal vision. *Journal of Vision*, *20*(12), 2. https://doi.org/10.1167/jov.20.12.2

Strasburger, H., Rentschler, I., & Juttner, M. (2011). Peripheral vision and pattern recognition: A Review. *Journal of Vision*, *11*(5), 13–13. https://doi.org/10.1167/11.5.13

Stromeyer, C. F., & Julesz, B. (1972). Spatial-frequency masking in vision: Critical bands and spread of masking. *Journal of the Optical Society of America*, *62*(10), 1221. https://doi.org/10.1364/josa.62.001221

Tomassini, A., Spinelli, D., Jacono, M., Sandini, G., & Morrone, M. C. (2015). Rhythmic oscillations of visual contrast sensitivity synchronized with action. *The Journal of Neuroscience*, *35*(18), 7019–7029. https://doi.org/10.1523/jneurosci.4568-14.2015

Watson AB. QUEST+: A general multidimensional Bayesian adaptive psychometric method. Journal of Vision. 2017;17(3):1–27. doi: 10.1167/17.3.10.

Weldon, K. B., Rich, A. N., Woolgar, A., & Williams, M. A. (2016). Disruption of foveal space impairs discrimination of peripheral objects. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00699

Weldon, K. B., Woolgar, A., Rich, A. N., & Williams, M. A. (2020). Late disruption of central visual field disrupts peripheral perception of form and color. *PLOS ONE*, *15*(1). https://doi.org/10.1371/journal.pone.0219725

White, C. W., & Lorber, C. M. (1976). Spatial-frequency specificity in visual masking. *Perception & Psychophysics*, *19*(4), 281–284. https://doi.org/10.3758/bf03204231

Wichmann, F. A., & Hill, N. J. (2001). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics*, *63*(8), 1293–1313. https://doi.org/10.3758/bf03194544

Williams, M. A., Baker, C. I., Op de Beeck, H. P., Mok Shim, W., Dang, S., Triantafyllou, C., & Kanwisher, N. (2008). Feedback of visual object information to foveal retinotopic cortex. *Nature Neuroscience*, *11*(12), 1439–1445. https://doi.org/10.1038/nn.2218

Yu, Q., & Shim, W. M. (2016). Modulating foveal representation can influence visual discrimination in the periphery. *Journal of Vision*, *16*(3), 15. https://doi.org/10.1167/16.3.15

Zhang, H., Morrone, M. C., & Alais, D. (2019). Behavioural oscillations in visual orientation discrimination reveal distinct modulation rates for both sensitivity and response bias. *Scientific Reports*, *9*(1). https://doi.org/10.1038/s41598-018-37918-4

**Acknowledgments** We thank Melanie Zylenas, Christoph Schäfer, Joni Blume, and Eva Berlips for helping with the data collection.

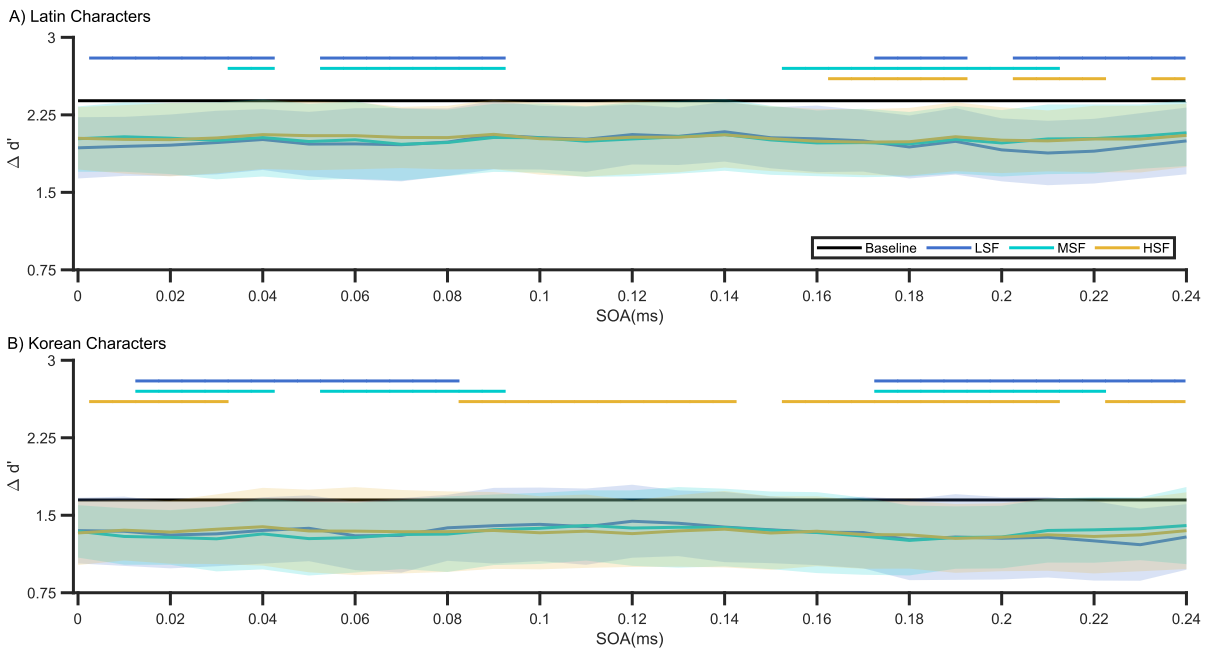
**Funding** This project received funding from the German-Canadian International Research Training Group (IRTG) 1901 “Brain in action” by the German Research Foundation (DFG) and from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No. 676786).

**Availability of data and materials** Data generated during this study will be published upon acceptance at zenodo.org. None of the experiments was preregistered. Correspondence and material requests should be addressed to N.G.

**Conflicts of interest/Competing** interest Authors have no competing interests to declare that are relevant to the content of this article.

## Supplementary materials

### *Experiment 1: Comparison to baseline no-noise condition*




Supplementary Figure 1. Character discrimination performance for Latin (A) and Korean (B) characters with foveal LSF, MSF, and HSF noise relative to the no-noise performance. The black horizontal lines show the no-noise baseline performance. Shaded areas show 99,95% confidence interval corresponding to the alpha level used for statistical analyses ($\alpha = 0.0007$, for $3 \times 25$ comparisons in each character condition). The upper horizontal bars indicate SOAs with significant reductions in performance compared to baseline.

Previous studies have used two methods for identifying the time window of foveal-feedback. In the results section, we used the method of comparing the different SOAs with each other. The other method entails comparing SOAs to a no-noise baseline condition (e.g., Fan et al., 2016). Here, we re-analyzed the data from Experiment 1 using the other method and provide a short discussion.

We compared each SOA for each noise and character condition to observers' no noise baseline performance (Supplementary Fig 1). Bonferroni corrected paired-sample t-tests revealed multiple effective SOAs that vary across noise and character conditions. Both conditions LSF noise produced the strongest reduction for similar SOAs. For both characters, we observed a performance reduction around an SOA of 30 within a window of 10-60 ms and around an SOA of 240 within a window of 10-70 ms. The timeframe of the first episode of foveal-feedback is consistent with Weldon et al., 2016, 020)while the second episode is consistent with Fan et al. (2016) and

Ramezani et al. (2019). Therefore, our results capture the wide range of reports from previous studies. Whereas MSF noise were effective within smaller time windows around 80 and 170 SOAs around 10-30 ms. Lastly, HSF noise produced inconsistent episodes of performance reductions. However, an episode of reduction starting around an SOA of 170 was common for the two characters. Taken together, in contrast to the SOA comparison method in Results section, baseline comparison suggests inconsistent time windows of foveal-feedback across different noise and character conditions. However, the difference between the two methods is not surprising. When the noise conditions are compared to the no-noise condition, there is an inherited assumption that any change in performance is due to foveal-feedback, which may not be necessarily the case. In fact, using this method Fan et al. (2016) found a reduction of performance at 50 ms SOA as we did. They attributed this reduction to attention as the noise presentation coincided with the peripheral target presentation. This is a more plausible explanation as TMS results show that foveal-feedback takes place around 300 ms after target onset[8]. While this method is prone to false positives, comparing SOAs with each other can be over conservative as it would yield a null result if there are multiple episodes of foveal-feedback causing similar amounts of reduction. Therefore, our analysis in the main results is more likely to show the time window of foveal-feedback when it is strongest.

*Study III*[*]

Spatiotemporal congruency modulates weighting of visuotactile information in displacement judgments

Nedim Goktepe[1], Knut Drewing[2], and Alexander C. Schütz[3]

Abstract

With the recent technological developments, devices capable of rendering multisensory experiences such as virtual reality (VR) and augmented reality (AR) became widely available. One challenge of these devices is to deliver multisensory signals with high spatiotemporal fidelity. A tight spatiotemporal representation of multisensory signals can be an important determinant of whether the two signals would be combined and offer an augmented sensory experience. In the current study, we investigated if and how spatially and temporally discrepant tactile displacement cues can supplement imprecise visual displacement cues. Participants performed a visual displacement task with visual and tactile displacement cues under spatial and temporal incongruency conditions. We modelled how participants combined visual and tactile information in visuotactile condition using their performance in visual only condition. We found that temporal incongruency lead to an increase in tactile weights although they were correlated with the congruency condition. In contrast, the spatial incongruency led to individual differences altering cue combination strategies. Our results illustrate the importance of spatiotemporal congruency for combining tactile and visual cues when making visual displacement judgments. Given the altered cue combination strategies and individual differences, we recommend developers to adopt individual spatiotemporal calibration procedures to improve the efficiency of the sensory augmentation.

*Index Terms*—Multisensory integration, Sensory substitution, Sensory augmentation, Vision, Haptics, Spatial Congruency, Temporal Congruency, Individual differences

## Introduction

New technologies are chasing the next best thing for creating realistic experiences in our virtual lives. These technologies often try to emulate real life by rendering virtual worlds that are similar to what we perceive or to augment sensory experiences by substituting or supplementing sensory information. An important step for augmenting user experience through virtual applications is to create environments that feature intuitive additional sensory information that substitutes or supplements an existing feature in the environment.

Virtual technologies for extending and improving the perceptual (in)abilities of its users have a wide range of applications for both clinical and non-clinical populations. For instance, visual information necessary for navigation can be substituted by auditory information to improve navigation in visually impaired populations (Jicol et al. 2020; Neugebauer et al. 2020; Scalvini et al. 2022).Virtual technologies could also replace conventional therapies by offering better therapeutic outcomes at lower costs (Cerritelli et al. 2021). Besides complete substitution, the available sensory information can also be supplemented by providing additional sensory information from a different modality. For instance, the surgical precision in teleoperated robot-assisted operations can be improved by supplementing force feedback information from the robot hand with a graphical representation (Akinbiyi et al. 2006).

A potential bottleneck for virtual reality applications is to generate multiple sensory signals through multiple devices in tandem (Jicol et al., 2020). There is a knowledge gap about the spatial and temporal tolerances of delivering parallel stimulations for substituting or supplementing sensory experiences. This is an important question both for developing necessary technologies and for understanding how a sensory signal is mapped onto a sensory dimension from another modality. One critical question is whether temporal or spatial congruency is necessary for cross dimensional mapping.

The role of spatiotemporal congruency has been primarily studied in the multisensory integration literature (Elshout et al. 2020; Lewald, Ehrenstein, and Guski 2001; Miller 1972; Serwe, Drewing, and Trommershauser 2009). The general consensus is that spatial and temporal proximity of sensory cues is a crucial factor for if and how cues should be integrated (Harrar, Harris, and Spence 2017; Jovanovic and Drewing 2014; Körding et al. 2007; Senkowski et al. 2007; Shams and Beierholm 2010) for reviews see (Stein and Meredith 1993) and (Calvert, Spence, and Stein

2004). The brain seems to assign sensory information from multiple senses to the same object or event if it falls within a close spatiotemporal window (Welch and Warren 1980) ("assumption of sensory unity", for a review see Chen and Spence 2017). In other words, according to the unity assumption, the brain expects that sensory cues belonging to the same object should be proximal to each other in terms of space and time. However, recent evidence suggests that the spatial and temporal congruency is more of a rule of thumb than a must have (Bermant and Welch 1976), for a review see (Spence 2013). As per the unity assumption, the ideal spatiotemporal distance between the sensory cues for achieving the strongest binding is zero. However, simply watching a ventriloquist reveals that this spatiotemporal window is flexible. Under controlled conditions, (Slutsky and Recanzone 2001) suggests that for achieving ventriloquism, the timing difference between the auditory and visual information should be less than 150 ms and spatial distance should be within the audience's auditory discrimination threshold. Moreover, this flexible window seems not to be constant across modalities and tasks. Instead, the spatiotemporal window for sensory integration is variable. For instance, spatial congruency seems to be mostly important for spatial tasks (Spence 2013). Likewise, the spatiotemporal window for rendering multisensory experiences for virtual applications might be also contingent on the sensory modalities of signals and the intended use of application rather than fixed.

In contrast to full multisensory integration, substitute or complimentary sensory information can be combined in a common map without integrating individual sensory information. Hence, combining sensory information for supplementing information from another sense does not require to assume that all sensory information have the same cause. Thus, sensory information could be combined in a more flexible spatiotemporal window than multisensory integration. For instance, visual search can be facilitated by an auditory cue that is spatially not aligned with the visual target (pip and pop effect, (Van Der Burg et al. 2008). Nevertheless, it is plausible to assume a spatiotemporal window for arbitrating which sensory information should be combined. Like multisensory integration, this spatiotemporal window for combining multiple senses might depend on their modality and task requirements. Thus, for developing virtual devices that supplement sensory information, it is imperative to deliver sensory signals within a spatiotemporal window that is suitable for the modality and the task at hand.

Here, we investigated the flexibility of the spatiotemporal window for combining sensory information and how spatiotemporal discrepancies impact on sensory combination in virtual reality

applications. Given the spatiotemporal flexibility for cue combination we opted for extreme spatial and temporal discrepancies in order to characterize how cue combination is affected by spatial or temporal discrepancies. Considering the potential task and modality dependency of this window, we decided to use a task that exploits natural limitations of the visual system. Humans shift their gaze up to four times per second by executing saccadic eye movements. During the execution of these eye movements, visual sensitivity drops to a degree where it may not be possible to visually detect small changes in object positions. While saccades suppress the detection of displacements, other senses could supplement this imprecise visual information and improve detection. The saccadic suppression of displacement is a robust and a well-studied paradigm where a visual target is displaced to a near location during saccades (Bridgeman, Hendry, and Stark 1975; Deubel, Schneider, and Bridgeman 1996; Mack 1970; Wallach and Lewis 1966). Here, we used a variant of this paradigm where participants were required to fixate instead of executing saccades to targets (Born 2019; Zimmermann et al. 2014; Zimmermann, Morrone, and Burr 2013). In this implementation, the visual masking that naturally occurs during saccades is simulated by a full screen noise that briefly masks the peripheral target before it is displaced. Like for saccadic suppression of displacement, the impairment caused by masking can be relieved when the masking period is followed by a blank screen before the visual target reappears at a new location (Born 2019; Deubel et al. 1996; Irwin and Robinson 2018; Zimmermann et al. 2014). Here, we supplement visual displacement information with uncorrelated tactile displacement information to test the spatiotemporal flexibility in combining these cues.

In order to render visual and tactile displacement cues, we built a custom setup where visual and tactile displacement cues are generated by two independent devices. Thus, our setup allowed us to manipulate visual and tactile cues independently from each other. Therefore, we were able to test whether imprecise visual displacement cues are substituted by uncorrelated but precise tactile displacement cues. More importantly, we checked if and how the spatial and temporal synchronicity of these cues is crucial for incorporating tactile cues into visual displacement judgments.

## Methods

### *Participants*

29 participants from Marburg University participated in the experiments. Six participants were excluded from statistical analyses as four participants failed to follow experiment instructions and two participants had points of subjective equalities (PSEs) outside of the measured range (see Data Analysis). Hence, the final sample included 23 participants (16 females, mean age = 24.74 ± 4.3). All participants had normal or corrected-to-normal (e.g., contact lenses) vision and did not report any tactile insensitivity on their right index finger where the tactile cues were administered. Participants signed written informed consent before the experiment and were compensated with 8 €/hour. The procedure was approved by the local ethics commission of the Department of Psychology (proposal number 2015-35k) and was in line with the Declaration of Helsinki (2013) except that the study was not preregistered.

*Equipment and Setup*

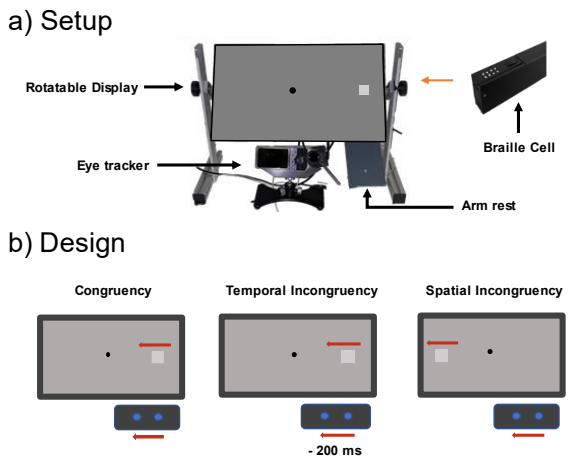



**Fig. 1.** Setup (a) and design (b) of the experiment. a) Participants reached to the braille cell that was located under the screen and correspond to the visual target in the spatial congruency conditions. b) Example trials from visuotactile conditions. None of the stimuli is drawn to scale.

The experiment was conducted with custom hardware consisting of a visual display and a braille device. The visual display was a DELL UltraSharp monitor (resolution: 2560 × 1440) that was fitted on a custom platform. With this platform, the screen height and angle were adjusted to allow participants to reach the braille device which was located on the back of the screen (Fig. 1a). The height and the angle of the display were fixed for all participants. With a chinrest and adjustable chair, we ensured that all participants viewed the screen from a similar vantage point. The visual display was calibrated to provide linear luminance output (e.g., gamma correction). Following the calibration, the luminance for black, grey, and white pixels were 1.17, 126.53, and 259.74 cd/m$^2$ respectively.

Tactile displacement cues were generated by a commercial braille cell (b11: B11.cdr (metec-ag.de), Metec Ag.) designed for both private and research purposes. This device had successfully used in a similar design (Sun and Okada 2019). The braille cell has an array of 2 × 3 pins that can be individually manipulated on command. The diameter of each pin was 0.7 mm (Min tactile force = 17 cN, Dot stroke = 0.7 mm) and the distance between neighboring pins was 2.45 mm. The braille cell was placed behind the screen at an overlapping position with the visual stimulus except in the spatial incongruency condition. The tactile cue was generated by raising a single pin and switching it with a neighboring pin.

In all conditions, participant gaze position was recorded with an Eyelink 1000+ eye tracker (SR Research Ltd., ON, Canada) at 1000 Hz. At the beginning of all conditions, the eye tracker was calibrated using a 9-point calibration with 25° horizontal and 5° vertical steps from the screen center.

We used a custom MATLAB script to control both visual and tactile devices, record gaze position and participant responses. The visual stimulus was controlled using Psychtoolbox (Brainard 1997; Kleiner, Brainard, and Pelli 2007). Gaze position was recorded through custom functions and the Eyelink toolbox extension of Psychtoolbox. The braille cell was controlled through MVBD which is a commercial software developed by the producer company.

*Stimuli*

In all trials there was a peripheral visual target located at 25° horizontal distance from the fixation on the same vertical axis. The visual stimulus was a dark gray square subtending an area of 0.53° × 0.53° on a lighter gray background. To allow participants to distinguish the different conditions, the visual target was presented as a square in the visuotactile conditions and as a diamond in the visual only control condition. There was a brief full screen visual noise increasing the difficulty of judging visual displacements. The noise consisted of an equal number of dark and bright pixels that were randomly arranged for each trial. Thus, the mean luminance of the noise was equal to the background luminance.

In all conditions except in the visual only condition, there was also a tactile stimulus accompanying the visual target. The tactile stimulus was generated by raising one of the pins of the braille device that was located just under the visual stimulus from the participant's perspective. Thus, the size of the tactile stimulus was 0.7 mm or 0.1°. The tactile displacement was generated

by retracting the raised pin and raising its neighboring pin at the same time. Thus, the size of tactile displacement was equal to the distance between two neighboring pins, which was 4.5 mm or 0.33°.

*Procedure*

Upon signing informed consents, participants were told that they will be helping to test a prototype of a new device capable of rendering objects with visual and tactile features. They were instructed to fixate on a fixation target and report the direction of the visual displacement (i.e., left or right) of the peripheral target. However, they received strict instruction that they should not totally ignore or rely on visual or tactile information alone. The experimenter did not mention or comment on the congruency of the visual and tactile cues in terms of direction or timing. However, at the beginning of the visual only block, the experimenter informed participants about the absence of tactile cues. Following the instructions, participants were seated comfortably in front of the visual display and placed their chin into the chinrest. Under the guidance of the experimenter, with their right arm, they reached to the back of the screen and placed their index finger on the braille cell.

All participants completed five separate blocks in one session. There were two control visual and three visuotactile conditions, completed in separate blocks. Hence, each block followed the same general trial structure in all conditions. At the beginning of each block, the experimenter ensured that participants were properly in contact with the braille cell by completing three test trials with the tactile cues without the visual target. Then, participants proceeded to the eye tracker calibration. Common to all conditions, each trial started with a fixation check, in which the participant had to press the space bar while fixating a fixation target located at the center of the display but randomly jittered 1° horizontally in each trial. Following a successful fixation check, the fixation target blinked to signal the start of the trial. After a random duration of 500-750 ms, the visual target appeared at a 25° horizontal distance from the fixation target. 300 ms after the visual target, a brief full screen noise occluded the visual target for 50 ms. Following the occlusion period, the visual target reappeared at a horizontally shifted location (4.5°, 3.25°, 2°, 1° either to the right or left) or at the same location. The visual target was displayed another 300 ms following the occlusion. Participants indicated the direction of the visual displacement with their left hand by pressing “F” or “J” on the keyboard for leftward and rightward displacement, respectively. At the end of each block, participants were allowed to take a short break if needed.

The only difference between the two visual conditions was the spatial alignment of the visual target with respect to the braille cell. In the spatially congruent visual condition, the visual target appeared on top of the braille cell and the right index finger of the participants from their perspective. In the spatially incongruent condition, it appeared at the same horizontal distance to the screen center but on the opposite side of the braille cell. In both visual only conditions, the visual target either appeared immediately following the brief noise period or after a 300 ms blank period to relieve the effect of masking. During the blank period, only the fixation target was present on the screen.

In the visuotactile conditions, visual and tactile cues were spatially, temporally, or spatiotemporally aligned. The tactile displacement cues were generated by presenting a single pin for 300 ms then switching it with a neighboring pin that stayed raised for another 300 ms. Thus, the tactile displacement mimicked the visual displacement except, the size of tactile displacements was fixed. The direction of tactile displacements was the same as the visual displacement in half of the trials. Thus, the direction of the tactile displacement was uninformative about the visual displacement direction. In the spatial incongruency condition, the visual and tactile displacements occurred at opposite sides of the screen center. Therefore, in both visual and visuotactile spatial incongruency conditions, the visual stimuli located opposite to the participants index finger and the braille cell. In the temporal incongruency condition, the visual and tactile displacements were spatially aligned but the tactile displacement preceded the visual displacement by 200 ms.

Participants completed five blocks in random order. Each block contained 180 trials. In the visual only conditions, half of the trials had a 300 ms blank period which were not used in the current study. Therefore, the visual conditions tested with a single within-subjects factor of spatial congreuncy (Spatial Congruency: Aligned vs Misaligned) . The visuotactile conditions followed a within-subjects design with one factor determining the spatial and temporal congruency (Congruency, Spatial incongruency, Temporal incongruency). The order of trials with and without blanking was random. The direction and the size of the visual displacements were fully randomized. The direction of the tactile displacement followed a random sequence which was fixed for all participants due to equipment restraints. Participants completed 900 trials, 180 trials per condition in around 100 minutes.

*Data Analysis*

Participants were instructed to fixate for the full duration of the trial until the response phase. We tracked their gaze location during this period. We discarded trials where the gaze location was detected outside of an imaginary square of 3° × 3° around the fixation target. We identified and excluded four participants who regularly executed saccades to the visual target. For the remaining participants, we only excluded trials that violated our fixation criterion (5.42% of trials in all conditions). With the remaining data, we first fitted psychometric functions (cumulative gaussians) to the spatially congruent visual only condition which also served as a control condition for visual displacement judgments. The estimated PSEs of two participants were outside of the measured displacement range. Therefore, we excluded these participants as their psychometric functions were not viable for modelling.

We modeled the visual displacement judgments of participants in the visual only conditions by fitting individual psychometric functions separately for blanking and non-blanking conditions using Psignifit (Wichmann and Hill 2001). We estimated psychometric functions using normal cumulative gaussian functions with a beta prior of 5, 0 lapse rate and asymptote.

Visual displacement tasks with saccades reported to produce bias opposite to the saccade direction (Irwin and Robinson 2018). While our task did not involve saccades, participants were always in contact with the braille cell, which could serve as a reference for discerning visual displacement direction. This would result in participants having different visual biases or different thresholds for the visual information in spatial incongruency and the other visuotactile conditions would be different. As a control for any difference in discrimination thresholds and directional bias a in visual judgments, we first compared the spatially congruent and incongruent visual only conditions (see *Supplementary Materials*). We did not find any evidence suggesting a difference in discrimination threshold or bias in non-blanking trials between the two conditions. Hence, we used the spatially aligned visual only condition for estimating the perceptual performance in all visuotactile conditions.

For the visuotactile conditions, we observed large biases in the direction of tactile displacement. For some participants, these biases did not allow to estimate a psychometric function for the measured visual displacements. Thus, we estimated the perceptual performance in these conditions using the tactile and visual displacement cues, detailed in the *Modelling* section.

Analysis and modelling were done with custom scripts in MATLAB 2022a (Math Works Inc.).

Modeling

We modelled participant judgments in the visuotactile conditions with a model that is more parsimonious than other, more popular models in multisensory integration literature such as maximum likelihood estimation (MLE, (Alais and Burr 2004; Ernst and Banks 2002) and causal inference (Körding et al. 2007; Shams and Beierholm 2010). Our model linearly combines the

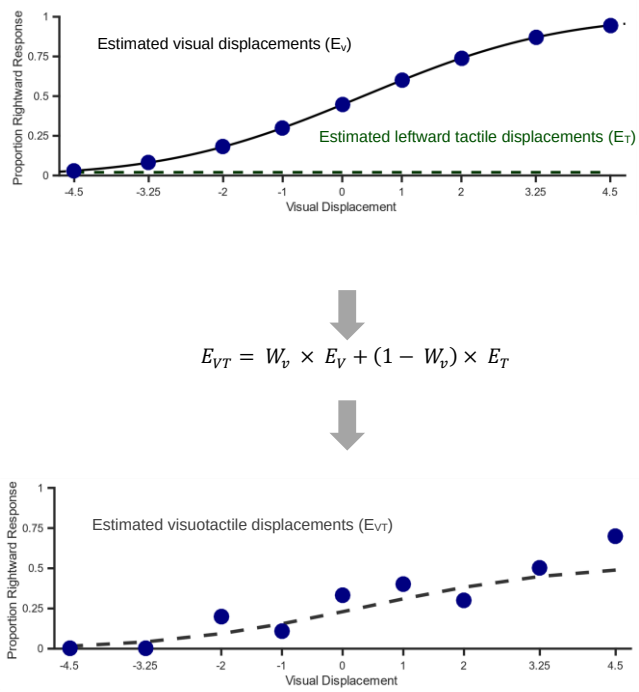



**Fig 2.** Data and model for one participant. First, the visual estimates were derived by fitting a psychometric function to the visual only condition (solid line). Then the tactile estimations for leftward tactile displacements (green line) and combination weights ($W_v$) were fitted to obtain a model for visuotactile displacements (gray line) with least-squared errors. Blue dots show the observed data in visual only and visuotactile conditions with leftward tactile displacements.

visual and tactile displacement estimates in the form of weighted summation (Fig. 2). Since we manipulated the spatial and temporal congruency of the displacement cues, participants would be expected to weight visual and tactile information differently across conditions. Thus, we modelled the perceptual judgments in each condition using independent weights for combining the visual and tactile information. We directly derived visual estimates from the psychometric functions fitted to the spatially congruent visual condition, which was otherwise identical to visuotactile conditions apart from the addition of tactile information.

We estimated the tactile estimates also with cumulative gaussian functions with a mean of zero (i.e. assuming no direction bias) and a free parameter for determining the width (i.e., just-noticeable-difference). Thus, the perceptual judgment for a visuotactile condition ($E_{VT}$) was estimated by linearly combining the visual ($E_V$) and the tactile estimate ($E_T$) with the free weighting parameter ($W_v$). $E_V$ is estimated through psychometric functions using visual displacement judgments in the spatially aligned visual only condition. $E_T$ is estimated for the leftward and rightward displacements ([-0.33 0.33]) by a cumulative gaussian function with a mean of 0 and fitted standard deviation. Each visuotactile condition was modelled by combining the same visual and tactile estimates with a different weighting parameter.

Results

From the visuotactile combination models we extracted weights of visual and tactile information in each condition, to check how the combination of visual and tactile cues is modulated by large spatial or temporal discrepancies.

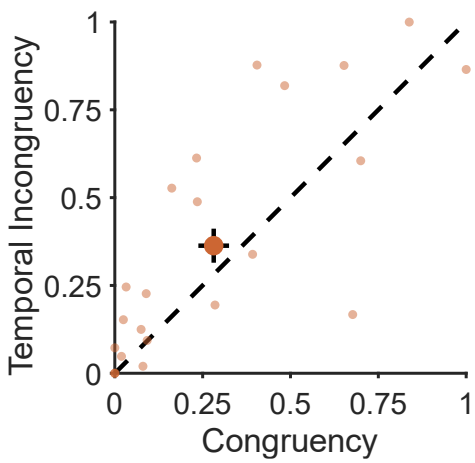



**Fig. 3.** Modelled tactile weights for spatiotemporal congruency and temporal incongruency conditions. Mean haptic weights for the temporal incongruency and congruency conditions are presented with error bars showing 95% confidence interval for congruency (horizontal), temporal incongruency (vertical) conditions.

Figure 3 shows the relationship between the influence of tactile information in the temporal incongruency and the congruency conditions. We found that the influence of tactile information was correlated in the congruency condition and the temporal incongruency condition (*rho*(22) = 0.84, $p < 0.001$). In other words, there was a linear relationship in how participants weighted the tactile information when it was spatiotemporally aligned or when it preceded the visual targets by 200 ms. In addition, they tended to use more tactile information under temporal discrepancy

compared to congruency condition. In order to test whether temporal incongruency significantly changed the weighting of tactile information, we first standardized the changes in tactile weights across visuotactile conditions by applying a Z transformation to tactile weights of each participant. A Wilcoxon signed rank on the z-transformed data yielded a significant difference in tactile weights ($z$ = 2.128, $p$ = 0.033), suggesting an increase in tactile weighting with temporal incongruency. Thus, under temporal discrepancy participants simply increased their use of tactile information without altering their overall weighting pattern.

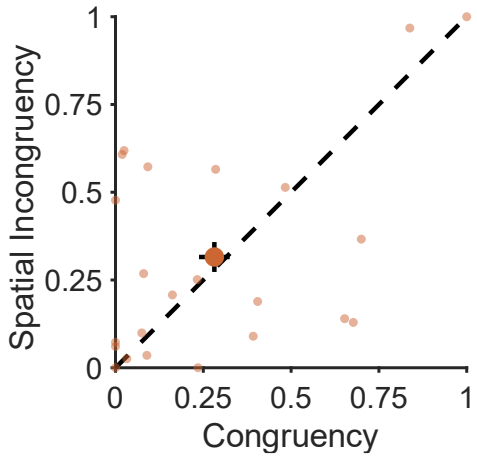



**Fig. 4**. Modelled tactile weights for spatiotemporal congruency and spatial incongruency conditions. Mean haptic weights for the spatial incongruency and congruency conditions are presented with error bars showing 95% confidence interval for congruency (horizontal), spatial incongruency (vertical) conditions.

We followed the same approach and compared the tactile weightings in the spatial incongruency condition to spatiotemporal condition (Fig. 4).

In contrast to temporal incongruency condition, participants seemed to adjust their weighting of tactile information independently from the tactile weights in the congruency condition. More specifically, while for some participants the weighting of tactile information was increased with spatial discrepancy, for others it decreased hinting at individual differences on how participants used the spatially discrepant tactile information. In line with this, there was no correlation between tactile weights in spatial incongruency and spatiotemporal congruency conditions (*rho*(22) = 0.35, $p$ = 0.103). This relationship was also significantly weaker than the relationship between the temporal incongruency and congruency conditions as indicated by the Fisher $r$ to $z$ transformed data (*Fisher* $z$= 2.706, $p$ = 0.003). In contrast to the temporal incongruency condition, we did not observe an increase in tactile weights compared to congruency condition ($z$ = 1.025, p = 0.301).

Interestingly, unlike temporal incongruency, the spatial discrepancy seemed to modulate tactile weights differently for different observers.

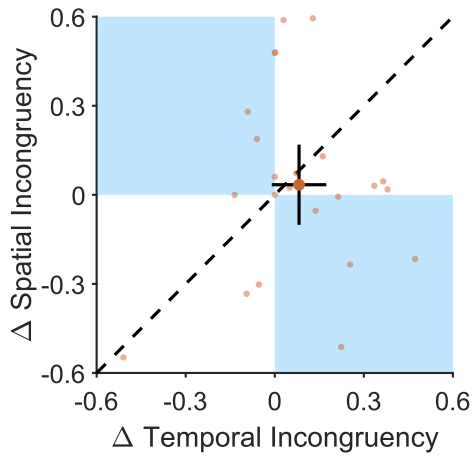



**Fig. 5.** Tactile weights in the incongruency conditions relative to the spatiotemporal congruency condition. Error bars show 95% confidence interval for temporally incongruent (horizontal) and spatially incongruent (vertical) visual conditions.

To further analyze these inter-participant differences, we calculated the change in tactile weights in incongruency conditions relative to the congruency condition (Fig. 5). There was no correlation between the relative change in tactile weights in spatial and temporal incongruency conditions ($r(22)= -0.028$, $p = 0.9$). We observed that the tactile weights in

spatial and temporal incongruency conditions were shifted towards opposite directions for 9 out of 23 participants (Shaded area in Fig. 5). We tested if these participants form a distinctly separate group by comparing the frequency of participants showing different weighting patterns (the four quadrants in Fig 5) using an exact McNemar test. However, there was no significant difference between the frequency of participants showing different patterns of combination ($\chi^2 = 1.923$, $p = 0.166$).

Discussion

We found that tactile displacement cues can be combined with visual displacement cues and influence visual displacement judgments. We modelled the influence of the tactile cues using a linear model that combines estimation for visual and tactile unisensory estimates. Our results suggest that the weighting of unisensory estimates in cue combination are prone to change in the face of large spatiotemporal discrepancy between the underlying unisensory estimates. This is in

line with previous research on multisensory integration suggesting that the strength of multisensory integration, how likely is to perceive two signals to be belonging to the same source, is dependent on the degree of spatial and temporal alignment of the multisensory signals. in general (Lewald et al. 2001). We observed that spatial and temporal discrepancies impact differently the weighting of individual cues.

The temporal discrepancy increased the weighting of tactile information but was still correlated with the congruent condition. This change in combination weights is to be expected as both adults and infants can use temporal discrepancy to segregate different cues (Oster and Werner 2018). This is also in line with previous studies emphasizing the role of temporal congruency in multisensory integration (Harrar et al. 2017; Senkowski et al. 2007; Serwe et al. 2009; Van Atteveldt et al. 2006). However, it is important to note that the majority of evidence for the importance of temporal alignment stems from studies using auditory and visual cues (Harrar et al. 2017; Slutsky and Recanzone 2001; Van Atteveldt et al. 2006). These studies suggest an integration window of 200 ms for audiovisual stimuli (Slutsky and Recanzone 2001) as opposed to visuoproprioceptive cues with a window of 400 ms (Kannape et al. 2019). Crucially, the window of integration is not contingent on the physical simultaneity but when the simultaneity of arrival of the two signals to the brain (Bresciani et al. 2005). Moreover, different neural mechanisms are responsible for binding asynchronous audiotactile cues within different time windows (Zumer, White, and Noppeney 2020). Thus, the efficiency of sensory augmentation systems might depend on the temporal sensitivity of different neural mechanisms that combine cues from various modalities.

While temporal congruency has systematically shifted the tactile weights, we found that spatial discrepancy yielded a very different outcome. We observed that in the face of spatial incongruency, tactile weights changed independently from the congruent condition. Interestingly, tactile weights for almost half of the participants drastically changed compared to the congruency and temporal incongruency conditions. The spatial incongruency altered the combination weights for these participants to a degree that the weights for tactile and visual information switched places. This is reminiscent of a phenomenon called cue switching, where individual unisensory cues are alternatingly used for estimating multisensory signals (Serwe et al. 2009). This interparticipant variation could have also hindered any change in tactile weights as a result of spatial incongruency.

We further inspected the apparent individual differences caused by spatial incongruency by comparing relative weight changes in temporal and spatial incongruency conditions. We did not find significantly distinct groups. However, we had a relatively small sample for testing individual differences. We encourage prospective studies to check these differences which can stem from various innate or developmental factors. For instance, individuals with autism spectrum disorder (ASD) are less affected from the brightness-weight illusion when judging the weight of visuotactile objects (Hadad and Schwartz 2019). Investigating individual differences in cue combination could have direct implications for applications such as augmented reality. For instance, the spatiotemporal window for cue combination might be smaller for professional or frequent video gamers who are experienced with using proximal cues. Similarly, the spatiotemporal window could be different or highly variable for patient groups, such as autism spectrum disorder (Feldman et al. 2018) or schizophrenia (Williams et al. 2010). Additionally, the spatiotemporal window might change over the course of treatment for patients recovering from physical injuries. Therefore, the development of multisensory displays in general could benefit from studying how spatiotemporal combination window varies in both healthy and patient user groups.

In the current study, we tested the impact of spatial and temporal discrepancies within controlled laboratory conditions. While the richness of the rendered environment boosts its realism and sensory integration (Bailey et al. 2018), it could lead to a tidier spatiotemporal window as multiple cues would be competing for integration and combination (Fleming, Noyce, and Shinn-Cunningham 2020). Therefore, future studies could address how scene complexity, or the number of competing cues modulate the spatiotemporal window for combining cues from different objects and senses.

The rapid advance in technology makes it possible to render realistic experiences for a wide range of applications including gaming and digital therapies. These applications augment sensory experience by supplementing proximal cues that provide information that otherwise wouldn’t be available to users. In this study, we found that the spatiotemporal congruency of visual and tactile cues to be a crucial factor for combining visual and tactile information for making visual displacement judgments. We argue that temporally discrepancy tips the combination weights towards one source of sensory estimate. In contrast, our results hint at potential individual differences in how individuals combine spatially discrepant sensory information. Thus, we

recommend that multisensory applications assess individual spatiotemporal limits for combining sensory cues for improving task performance and user satisfaction for various subpopulations.

Acknowledgement

We thank Christoph Schaefer, Eva Berlips, and Joni Blume for their help with the data collection and Hartmut Studenroth and his team for their help with the experimental setup.

References

Akinbiyi, Takintope, Carol E. Reiley, Sunipa Saha, Darius Burschka, Christopher J. Hasser, David D. Yuh, and Allison M. Okamura. 2006. 'Dynamic Augmented Reality for Sensory Substitution in Robot-Assisted Surgical Systems'. Pp. 567–70 in *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*. New York, NY: IEEE.

Alais, David, and David Burr. 2004. 'The Ventriloquist Effect Results from Near-Optimal Bimodal Integration'. *Current Biology* 14(3):257–62. doi: 10.1016/j.cub.2004.01.029.

Bailey, Hudson Diggs, Aidan B. Mullaney, Kyla D. Gibney, and Leslie Dowell Kwakye. 2018. 'Audiovisual Integration Varies With Target and Environment Richness in Immersive Virtual Reality'. *Multisensory Research* 31(7):689–713. doi: 10.1163/22134808-20181301.

Bermant, Robert I., and Robert B. Welch. 1976. 'Effect of Degree of Separation of Visual-Auditory Stimulus and Eye Position upon Spatial Interaction of Vision and Audition'. *Perceptual and Motor Skills* 43(2):487–93. doi: 10.2466/pms.1976.43.2.487.

Born, Sabine. 2019. 'Saccadic Suppression of Displacement Does Not Reflect a Saccade-Specific Bias to Assume Stability'. *Vision* 3(4):49. doi: 10.3390/vision3040049.

Brainard, David H. 1997. 'The Psychophysics Toolbox'. *Spatial Vision* 10(4):433–36. doi: 10.1163/156856897X00357.

Bresciani, Jean-Pierre, Marc O. Ernst, Knut Drewing, Guillaume Bouyer, Vincent Maury, and Abderrahmane Kheddar. 2005. 'Feeling What You Hear: Auditory Signals Can Modulate Tactile Tap Perception'. *Experimental Brain Research* 162(2):172–80. doi: 10.1007/s00221-004-2128-2.

Bridgeman, Bruce, Derek Hendry, and Lawrence Stark. 1975. 'Failure to Detect Displacement of the Visual World during Saccadic Eye Movements'. *Vision Research* 15(6):719–22. doi: 10.1016/0042-6989(75)90290-4.

Calvert, Gemma A., Charles Spence, and Barry E. Stein, eds. 2004. *The Handbook of Multisensory Processes*. The MIT Press.

Cerritelli, Francesco, Marco Chiera, Marco Abbro, Valentino Megale, Jorge Esteves, Alberto Gallace, and Andrea Manzotti. 2021. 'The Challenges and Perspectives of the Integration Between Virtual and Augmented Reality and Manual Therapies'. *Frontiers in Neurology* 12:700211. doi: 10.3389/fneur.2021.700211.

Chen, Yi-Chuan, and Charles Spence. 2017. 'Assessing the Role of the "Unity Assumption" on Multisensory Integration: A Review'. *Frontiers in Psychology* 8. doi: 10.3389/fpsyg.2017.00445.

Deubel, Heiner, Werner X. Schneider, and Bruce Bridgeman. 1996. 'Postsaccadic Target Blanking Prevents Saccadic Suppression of Image Displacement'. *Vision Research* 36(7):985–96. doi: 10.1016/0042-6989(95)00203-0.

Elshout, J. A., N. Van Der Stoep, T. C. W. Nijboer, and S. Van Der Stigchel. 2020. 'Motor Congruency and Multisensory Integration Jointly Facilitate Visual Information Processing before Movement Execution'. *Experimental Brain Research* 238(3):667–73. doi: 10.1007/s00221-019-05714-9.

Ernst, Marc O., and Martin S. Banks. 2002. 'Humans Integrate Visual and Haptic Information in a Statistically Optimal Fashion'. *Nature* 415(6870):429–33. doi: 10.1038/415429a.

Feldman, Jacob I., Kacie Dunham, Margaret Cassidy, Mark T. Wallace, Yupeng Liu, and Tiffany G. Woynaroski. 2018. 'Audiovisual Multisensory Integration in Individuals with Autism Spectrum Disorder: A Systematic Review and Meta-Analysis'. *Neuroscience & Biobehavioral Reviews* 95:220–34. doi: 10.1016/j.neubiorev.2018.09.020.

Fleming, Justin T., Abigail L. Noyce, and Barbara G. Shinn-Cunningham. 2020. 'Audio-Visual Spatial Alignment Improves Integration in the Presence of a Competing Audio-Visual Stimulus'. *Neuropsychologia* 146:107530. doi: 10.1016/j.neuropsychologia.2020.107530.

Hadad, Bat-Sheva, and Sivan Schwartz. 2019. 'Perception in Autism Does Not Adhere to Weber's Law'. *ELife* 8:e42223. doi: 10.7554/eLife.42223.

Harrar, Vanessa, Laurence R. Harris, and Charles Spence. 2017. 'Multisensory Integration Is Independent of Perceived Simultaneity'. *Experimental Brain Research* 235(3):763–75. doi: 10.1007/s00221-016-4822-2.

Irwin, David E., and Maria M. Robinson. 2018. 'How Post-Saccadic Target Blanking Affects the Detection of Stimulus Displacements across Saccades'. *Vision Research* 142:11–19. doi: 10.1016/j.visres.2017.09.004.

Jicol, Crescent, Tayfun Lloyd-Esenkaya, Michael J. Proulx, Simon Lange-Smith, Meike Scheller, Eamonn O'Neill, and Karin Petrini. 2020. 'Efficiency of Sensory Substitution Devices Alone and in Combination With Self-Motion for Spatial Navigation in Sighted and Visually Impaired'. *Frontiers in Psychology* 11:1443. doi: 10.3389/fpsyg.2020.01443.

Jovanovic, Bianca, and Knut Drewing. 2014. 'The Influence of Intersensory Discrepancy on Visuo-Haptic Integration Is Similar in 6-Year-Old Children and Adults'. *Frontiers in Psychology* 5. doi: 10.3389/fpsyg.2014.00057.

Kannape, Oliver A., Ethan J. T. Smith, Peter Moseley, Mark P. Roy, and Bigna Lenggenhager. 2019. 'Experimentally Induced Limb-Disownership in Mixed Reality'. *Neuropsychologia* 124:161–70. doi: 10.1016/j.neuropsychologia.2018.12.014.

Kleiner, M., David H. Brainard, and D. Pelli. 2007. 'What's New in Psychtoolbox-3?' Perception.

Körding, Konrad P., Ulrik Beierholm, Wei Ji Ma, Steven Quartz, Joshua B. Tenenbaum, and Ladan Shams. 2007. 'Causal Inference in Multisensory Perception' edited by O. Sporns. *PLoS ONE* 2(9):e943. doi: 10.1371/journal.pone.0000943.

Lewald, Jörg, Walter H. Ehrenstein, and Rainer Guski. 2001. 'Spatio-Temporal Constraints for Auditory–Visual Integration'. *Behavioural Brain Research* 121(1–2):69–79. doi: 10.1016/S0166-4328(00)00386-7.

Mack, Arien. 1970. 'An Investigation of the Relationship between Eye and Retinal Image Movement in the Perception of Movement'. *Perception & Psychophysics* 8(5):291–98. doi: 10.3758/BF03212596.

Miller, Epp A. 1972. 'Interaction of Vision and Touch in Conflict and Nonconflict Form Perception Tasks.' *Journal of Experimental Psychology* 96(1):114–23. doi: 10.1037/h0033468.

Neugebauer, Alexander, Katharina Rifai, Mathias Getzlaff, and Siegfried Wahl. 2020. 'Navigation Aid for Blind Persons by Visual-to-Auditory Sensory Substitution: A Pilot Study' edited by M. Ptito. *PLOS ONE* 15(8):e0237344. doi: 10.1371/journal.pone.0237344.

Oster, Monika-Maria, and Lynne A. Werner. 2018. 'Infants Use Onset Asynchrony Cues in Auditory Scene Analysis'. *The Journal of the Acoustical Society of America* 144(4):2052–60. doi: 10.1121/1.5058397.

Scalvini, Florian, Camille Bordeau, Maxime Ambard, Cyrille Migniot, Stéphane Argon, and Julien Dubois. 2022. 'Visual-Auditory Substitution Device for Indoor Navigation Based on Fast Visual Marker Detection'. Pp. 259–66 in *2022 16th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*. Dijon, France: IEEE.

Senkowski, Daniel, Durk Talsma, Maren Grigutsch, Christoph S. Herrmann, and Marty G. Woldorff. 2007. 'Good Times for Multisensory Integration: Effects of the Precision of Temporal Synchrony as Revealed by Gamma-Band Oscillations'. *Neuropsychologia* 45(3):561–71. doi: 10.1016/j.neuropsychologia.2006.01.013.

Serwe, S., K. Drewing, and J. Trommershauser. 2009. 'Combination of Noisy Directional Visual and Proprioceptive Information'. *Journal of Vision* 9(5):28–28. doi: 10.1167/9.5.28.

Shams, Ladan, and Ulrik R. Beierholm. 2010. 'Causal Inference in Perception'. *Trends in Cognitive Sciences* 14(9):425–32. doi: 10.1016/j.tics.2010.07.001.

Slutsky, Daniel A., and Gregg H. Recanzone. 2001. 'Temporal and Spatial Dependency of the Ventriloquism Effect': *Neuroreport* 12(1):7–10. doi: 10.1097/00001756-200101220-00009.

Spence, Charles. 2013. 'Just How Important Is Spatial Coincidence to Multisensory Integration? Evaluating the Spatial Rule: Spatial Coincidence and Multisensory Integration'. *Annals of the New York Academy of Sciences* 1296(1):31–49. doi: 10.1111/nyas.12121.

Stein, Barry E., and M. Alex Meredith. 1993. *The Merging of the Senses*. Cambridge, Mass London: MIT Press.

Sun, Limin, and Yoshio Okada. 2019. 'Vibrotactile Piezoelectric Stimulation System with Precise and Versatile Timing Control for Somatosensory Research'. *Journal of Neuroscience Methods* 317:29–36. doi: 10.1016/j.jneumeth.2019.02.002.

Van Atteveldt, N. M., E. Formisano, L. Blomert, and R. Goebel. 2006. 'The Effect of Temporal Asynchrony on the Multisensory Integration of Letters and Speech Sounds'. *Cerebral Cortex* 17(4):962–74. doi: 10.1093/cercor/bhl007.

Van Der Burg, Erik, Christian N. L. Olivers, Adelbert W. Bronkhorst, and Jan Theeuwes. 2008. 'Pip and Pop: Nonspatial Auditory Signals Improve Spatial Visual Search.' *Journal of*

*Experimental Psychology: Human Perception and Performance* 34(5):1053–65. doi: 10.1037/0096-1523.34.5.1053.

Wallach, Hans, and Charles Lewis. 1966. 'The Effect of Abnormal Displacement of the Retinal Image during Eye Movements'. *Perception & Psychophysics* 1(1):25–29. doi: 10.3758/BF03207816.

Welch, Robert B., and David H. Warren. 1980. 'Immediate Perceptual Response to Intersensory Discrepancy.' *Psychological Bulletin* 88(3):638–67. doi: 10.1037/0033-2909.88.3.638.

Wichmann, Felix A., and N. Jeremy Hill. 2001. 'The Psychometric Function: I. Fitting, Sampling, and Goodness of Fit'. *Perception & Psychophysics* 63(8):1293–1313. doi: 10.3758/BF03194544.

Williams, Lisa E., Gregory A. Light, David L. Braff, and Vilayanur S. Ramachandran. 2010. 'Reduced Multisensory Integration in Patients with Schizophrenia on a Target Detection Task'. *Neuropsychologia* 48(10):3128–36. doi: 10.1016/j.neuropsychologia.2010.06.028.

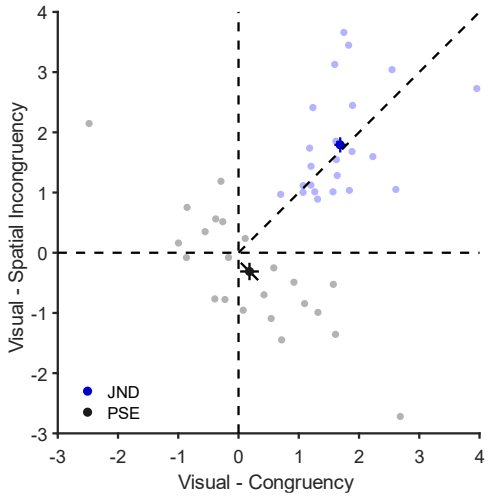
Zimmermann, Eckart, Sabine Born, Gereon R. Fink, and Patrick Cavanagh. 2014. 'Masking Produces Compression of Space and Time in the Absence of Eye Movements'. *Journal of Neurophysiology* 112(12):3066–76. doi: 10.1152/jn.00156.2014.

Zimmermann, Eckart, M. Concetta Morrone, and David C. Burr. 2013. 'Spatial Position Information Accumulates Steadily over Time'. *The Journal of Neuroscience* 33(47):18396–401. doi: 10.1523/JNEUROSCI.1864-13.2013.

Zumer, Johanna M., Thomas P. White, and Uta Noppeney. 2020. 'The Neural Mechanisms of Audiotactile Binding Depend on Asynchrony'. *European Journal of Neuroscience* 52(12):4709–31. doi: 10.1111/ejn.14928.

Supplementary Materials

Participants performed two visual conditions. In the spatially aligned conditions, the visual target appeared directly above their right index finger whereas in the other condition it was at a mirrored location. When the index finger is aligned with the visual target, it can serve as a landmark and improve visual displacement judgments even in the absence of any tactile displacement cues.




**Supplementary Fig. 1.** PSEs and JNDs in the visual only conditions without blanking. Mean JND and PSE presented with error bars showing 95% confidence interval for spatially congruent (horizontal), spatially incongruent (vertical) visual conditions and their comparison (diagonal).

To test this possibility, we compared the just noticeable differences (JNDs) in the visual only conditions without blanking when the index finger was a the same or the mirror location of the visual target (Supplementary Fig. 1). A paired sample t-test yielded no significant differences ($t(22) = 0.577$, $p = 0.567$) indicating that there was no significant improvement due to location of the visual stimulus relative to the finger.

Visual displacement tasks with saccades reported to produce bias opposite to the saccade direction (Irwin and Robinson 2018). In both conditions, there was a slight bias to report more inward displacements (towards fixation) than outward displacement. However, there was no difference in bias between the two conditions ($t(22)= 1.175$, $p = 0.253$). Since there were no differences in JNDs and PSEs between the non-blanking spatially congruent and incongruent conditions, we used the trials from the spatially congruent visual only condition to separately model the visuotactile conditions. Our model allows us to measure the weighting of visual and tactile information in our participants' visual displacement judgment separately in spatiotemporal, spatial, and temporal congruency condition.

Eigenständigkeitserklärung

Hiermit versichere ich, dass ich die vorliegende Dissertation selbstständig, ohne unerlaubte Hilfe Dritter angefertigt und andere als die in der Dissertation angegebenen Hilfsmittel nichtbenutzt habe. Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten oderunveröffentlichten Schriften entnommen sind, habe ich als solche kenntlich gemacht. Dritte waren an der inhaltlich-materiellen Erstellung der Dissertation nicht beteiligt; insbesondere habe ich hierfür nicht die Hilfe eines Promotionsberaters in Anspruch genommen. Kein Teildieser Arbeit ist in einem anderen Promotions-oder Habilitationsverfahren der Antragsstellerin oder Antragsstellersver wendet worden. Mit dem Einsatz von Software zur Erkennung von Plagiaten bin ich einverstanden.

## Authorship Statement

Study I: *Familiar objects benefit more from transsaccadic feature predictions.*

N.G. and A.C.S designed the experiments. N.G. collected and analyzed the data. N.G. prepared the first draft of the manuscript. N.G. and A.C.S revised the manuscript.

**Contributions**: N.G. 75%, A.C.S. 25%

Study II: *Frequency-specific and periodic masking of peripheral characters by delayed foveal input.*

N.G. and A.C.S designed the experiments. N.G. collected and analyzed the data. N.G. prepared the first draft of the manuscript. N.G. and A.C.S revised the manuscript.

**Contributions**: N.G. 75%, A.C.S. 25%

Study III: *Spatiotemporal congruency modulates weighting of visuotactile information in displacement judgments.*

N.G., K.D., and A.C.S designed the experiments. N.G. and A.C.S prepared the experiment. N.G. collected and analyzed the data. N.G., K.D., and A.C.S contributed to modelling of the data. N.G. prepared the first draft of the manuscript. N.G., K.D., and A.C.S revised the manuscript.

**Contributions**: N.G. 65%, K.D. 15%, A.C.S. 20%

Nedim Göktepe

Alexander Schütz

## Acknowledgments

This dissertation offers only a glimpse of the journey of a naïve pupil to a bit more seasoned researcher, which is commonly known as Ph.D. training. Although I learned a great deal from the studies mentioned here, I learned a lot more from things that are not. Those instances can be found in different experiment rooms and datasets but more importantly in the memories of people who witnessed or supported my transition.

Naturally, my supervisor Alexander Schütz witnessed my self-claimed progress probably like watching participants performing in an adaptive learning task where they gradually learn from most but not all of their wrong turns. Unlike such a participant, I had the great fortune and privilege to have him as my guide during our journey to the unknown (mostly to me). He has been a constant source of reason, and experience but also spatial frequency grating, parameters, and challenging questions. He always has been the greatest critic and supporter of my scientific curiosities.

While Alexander Schütz provided me with the stability for developing my scientific persona, it is fair to say Müge Cavdan provided the rest. She has been the bastion of motivation even when motivation meant pushing me harder or when it was as easy as pressing the send button of an email that I was reluctant to send. Although hardly imaginable, she has the ability to be happier and more excited about my achievements than me. At the same time, she got frustrated more by my setbacks, so I tried to make less mistakes. At the end of the day, no matter how beaten I was, I knew everything would be all right because she was home.

I was also fortunate enough to have people who could celebrate the good and weather the bad even when we were on different continents. Particularly, I am thankful for both of my grandmothers who believed in me even before I did. I am grateful to my mother whose support and dedication often surpassed my capabilities. Finally, I would like to thank my father whose attention to my well-being exceeded what can possibly be given on a weekly phone call.