The influence of spatial frequencies on navigation by visual guidance

Masterarbeit

der Mathematisch-Naturwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

vorgelegt von

Diễm My Lilian Levinh

Tübingen, März 2019

Erklärung:

Hiermit erkläre ich,

- dass ich diese Arbeit selbst verfasst habe.
- dass ich keine anderen als die angegebenen Quellen benutzt und dass ich alle wörtlich oder sinngemäß aus anderen Werken übernommenen Aussagen als solche gekennzeichnet habe.
- dass die eingereichte Arbeit weder vollständig noch in wesentlichen Teilen Gegenstand eines anderen Prüfungsverfahrens gewesen ist.

Tübingen, den

Primary Examiner: Prof. Dr. Hanspeter Mallot
Cognitive Neuroscience
Dept. of Biology, Faculty of Science
Eberhard Karls University Tübingen

Secondary Examiner: Prof. Dr. Jan Benda
Neuroethology
Dept. of Biology, Faculty of Science
Eberhard Karls University Tübingen

Contents

Zusammenfassung

Menschen ist es möglich, in kürzester Zeit ihre Umgebung zu erfassen und dementsprechend zu reagieren. Um dies zu erklären, wurden bereits mehrere Filtermechanismen vorgeschlagen, eine davon ist die Filterung nach Ortsfrequenzen. Es ist bereits bekannt, dass hohe und niedrige Ortsfrequenzen jeweils andere Arten an Information vermitteln, und es wurden bereits mehrere Studien zu dem Einfluss von Ortsfrequenzen auf Ort und Objekterkennung durchgeführt. Weiterhin legen Ergebnisse nahe, dass die Verwendung von Ortsfrequenzen aufgabenabhängig ist. Daher untersuchen wir in dieser Studie die bis jetzt wenig erforschte Bedeutung von hohen und niedrigen Ortsfrequenzen in der Orientierung anhand von Landmarken. In unserem Versuch hatten Versuchspersonen die Aufgabe Landmarken zu verfolgen, die entweder in hohe oder niedrige Ortsfrequenzen gefiltert wurden. Weiterhin wurde der Kontrast der Landmarken verändert, um die Auswirkung von Kontrast auf die landmarkenbasierte Orientierung zu untersuchen. Zudem wurden zwei verschiedene Umgebungen verwendet, die sich in ihrem Bildinhalt und den dadurch möglichen Landmarken unterschieden. Während die erste Szenerie leicht erkennbare freistehende Landmarken aufwies, waren die Landmarken der anderen Szenerie schwerer unterscheidbar und bildeten weiterhin eine unübersichtliche Landschaft. Auch wurde ein lineares Modell an das Verhalten der Versuchspersonen gefittet. Das Verhalten der Versuchspersonen konnte mithilfe eines linearen Modells beschrieben werden, welches auf Landmarkenbewegung beruht. Die Leistung der Versuchspersonen, wenn sowohl hohe wie auch niedrige Ortsfrequenzen erkennbar waren, war vergleichbar oder höher als in den Konditionen, in denen die Ziel Ortsfrequenz fast ausschliesslich benutzt wurde. Dies könnte darauf hinweisen, dass die Versuchspersonen Informationen von hohen und niedrigen Ortsfrequenzen benutzten. Unsere Ergebnisse legen zudem nahe, dass die Versuchspersonen verschiedene Orientierungsstrategien in den Szenerien benutzten. Versuchspersonen in der Szenerie mit schwer erkennbaren Landmarken, orientierten sich vornehmlich an Objekten, wenn sie hohen Ortsfrequenzen folgen sollten. Sollten jedoch niedrigen Ortsfrequenzen gefolgt werden, wurden die Relationen ganzer Bildelemente zueinander bevorzugt. Währenddessen schien sich, unabhängig von den Ortsfrequenzen, die Orientierung in der übersichtlichen Szenerie auf Objektmerkmale zu stützen. Weiterhin unterschied sich der präferierte Anteil von hohen und niedrigen Landmarken zwischen den Bildern. Im Unterschied zu der allgemeinen Bevorzugung von niedrigen Ortsfrequenzen in der Früherkennung, waren in unserem Experiment Reaktionszeiten auf niedrige Ortsfrequenzen signifikant länger als zu hohen Ortsfrequenzen. Die längeren Reaktionszeiten könnten auf eine Präferenz von hohen Ortsfrequenzen in der Bewegungsdetektion und in der Objektserkennung und einer Bevorzugung von Kanten bei Sakkaden zurückzuführen sein. Es wurde jedoch nur in einer der Szenerien ein Effekt auf Reaktionszeiten der Versuchspersonen gefunden. Dies legt nahe, dass ähnlich wie bei der Gesichtsverarbeitung eine Präferenz von Ortsfrequenzen flexibel ist, und mit Landmarken und Umgebung wechseln kann.

Abstract

Humans are able to recognize and act accordingly to their surroundings in less than a second. To explain this capability several early vision filtering mechanisms have been proposed, one of them being the filtering of spatial frequencies. It is known, that high and low spatial frequencies carry different kinds of information, and there have been several studies on the role of spatial frequencies regarding scene and object recognition. Findings suggest, that spatial frequency usage is task dependent. In this study we therefore investigate the, as of yet little explored, importance of high and low spatial frequencies on landmark based guidance. Subjects were to follow high or low spatial frequency filtered landmarks in a virtual reality environment. Landmark contrast was varied to investigate the effect of contrast on landmark based guiding. Further we used two different sceneries, which varied in potential landmarks and their surroundings. While one scenery featured distinct freestanding landmarks, possible landmarks contained in the other scenery were harder to distinguish and formed a cluttered environment. Subject data was fitted to a linear response model. Subject behavior could be described using a linear response model based on landmark movement changes. Subject performance was as good or better at ratios with both high and low spatial frequencies distinguishable compared to when the target spatial frequency was used almost exclusively. This may indicate, that subjects used information of both high and low spatial frequency components. Our results further suggest, that subjects'

orientation strategies were dependent on the scenery. It is possible, that subjects used different strategies when following either high or low spatial frequencies in the cluttered scenery. While subjects may have focused on object features when following high spatial frequencies, subjects may have oriented themselves using changes on a configural scale, when following low spatial frequencies. Meanwhile subjects orienting themselves in the clearly arranged scenery may have used object features regardless of target spatial frequency. It is also notable, that the preferred spatial frequency ratio varied with scenery. Contrary to the general preference of low spatial frequencies in early recognition tasks, subject reaction times were longer to low spatial frequency filtered landmarks than to high spatial frequency filtered ones. The longer reaction times may stem from a preference of high spatial frequencies in change detection and object analysis and a preference of saccades towards edges. However, an effect of spatial frequency on reaction time was only observed in one of the used sceneries. The results therefore suggest that, like in face processing tasks, spatial frequency preference in landmark based guidance is flexible and may change with the nature of scenery and landmarks.

1 Introduction

Orientation and wayfinding have been important skills throughout the ages, be it for finding food sites or the way home. One strategy, found both in humans (Learmonth, Newcombe, & Huttenlocher, 2001; Epstein & Vass, 2014) and many animals (Collett, 1995; Cartwright & Collett, 1983; Nouhuys & Kaartinen, 2007), is the usage of landmarks. Even though there is no universal consensus on what constitutes a landmark exactly, landmarks are generally defined as easily identified local position information, that can be used in orientation. Landmarks do not have to be distinct single objects, however, instead humans can also use geometrical features (e.g. room shape) (Hermer & Spelke, 1994) or even a color gradient (Gillner, Weiß, & Mallot, 2008) to orient themselves. Further, landmarks can be used in a flexible manner. Humans can use both local landmarks and global landmarks for orientation and route finding tasks, and a formerly global landmark may be used as a local landmark under a different situation (Steck & Mallot, 2000). In guidance the course is kept either by keeping the bearing to a landmark steady or by changing the bearing in a specific way. Usage of landmarks for guidance or direction can mitigate error accumulation contrary to using path integration only, and visual feedback can be crucial for even seemingly simple tasks; for instance, most any driver will fail to execute a mundane lane change when deprived of visual feedback (Wallis, Chatziastros, & Bülthoff, 2002). It has also been found, that, if landmarks are present, humans will forego learning cognitive maps, which were also more inaccurate, in favor of landmarks (Foo, Warren, Duchon, & Tarr, 2005). Moreover, not only do humans need less time for learning route descriptions containing landmarks over descriptions containing only street names, they were also able to recall them more accurately (Tom & Denis, 2004). A requisite for successful landmark use and visual orientation in general is a sound scene and object recognition. It takes considerably less than even a half second for humans to infer the basic gist of a scenery (Davenport & Potter, 2004; Potter, 2012) and several filtering mechanisms and models have been proposed to explain and emulate rapid scene recognition (Oliva & Torralba, 2001, 2006; Siagian & Itti, 2007). It is assumed that spatial frequencies are important in early vision tasks, as the result of spatial filtering forms the basis for higher order processes. Spatial frequencies can be filtered into high and low spatial frequencies (HSF and LSF), which carry different kinds of information. While LSFs carry coarse information such as brightness and color, HSFs carry information on finer aspects of an image such as edges.

1.1 Temporal differences in the perception of spatial frequencies

It is commonly assumed that a coarse to fine bias exists for most early vision tasks meaning that LSFs are processed faster and that LSF information is preferred in early vision tasks (Breitmeyer, Levi, & Harwerth, 1981; Lupp, Hauske, & Wolf, 1976; Vassilev & Mitov, 1976). This is supported by LSF information being transferred by the magnocellular pathway, while HSF information is conveyed through the slower parvocellular pathway (Bassi, Lehmkuhle, et al., 1990). Further LSF information and the magnocellular pathway, which responds to high temporal frequencies, are better suited for motion perception (Derrington & Lennie, 1984). It is assumed that LSFs are used to generate a first broad perception which is only afterwards modulated by HSFs enabling a more accurate object recognition (Bar, 2004). The coarse to fine bias is further supported by various experiments using a variety of stimuli from simple gratings (Breitmeyer et al., 1981) to complex stimuli such as faces (Pourtois, Dan, Grandjean, Sander, & Vuilleumier, 2005) and sceneries (Schyns & Oliva, 1994). However, there is evidence, that the coarse to fine bias may not be fixed. Using a composite image that consisted of two different scenes, one high and one low filtered, it was possible to prime subjects for both HSF and LSF scenes, although exposure was only 30ms (Oliva & Schyns, 1997). Further, it has also been found, that spatial frequency preference depends on the task and saliency of the spatial frequency information (Goffaux, Hault, Michel, Vuong, & Rossion, 2005). This would indicate, that the main reason of the coarse to fine bias is not due to faster processing of LSFs but that using LSF first is advantageous for most early vision tasks.

1.2 Usage of HSF and LSF in orientation

Lower spatial frequencies are more prevalent in natural environments (Tolhurst, Tadmor, & Chao, 1992), and experiments have shown, that humans can make good assessment of sceneries by using LSF information alone (Schyns & Oliva, 1994). However, important information such as an object's shape is better discerned using HSF information and object recognition relying solely on LSF information may therefore be more prone to misinterpretation. Although subjects can identify scenes without recognizing the objects present in that scene (Schyns & Oliva, 1994; Oliva & Schyns, 2000), landmark usage is only sensible when landmarks are reliably identified. Further, eye movements in scene and object analyzation are focused on areas with high spatial variance such as edges and corners while avoiding areas with only low spatial variance (Krieger, Rentschler, Hauske, Schill, & Zetzsche, 2000; Baddeley & Tatler, 2006). It would therefore be reasonable to expect HSF information to be used alongside of LSF information during orientation using landmarks.

1.3 Purpose of this study

Our study aims to compare how well HSF and LSF information can be used in landmark based guidance. Subjects were tasked to follow either HSFs or LSFs in a virtual environment. The scenery was filtered in HSF and LSF components, which could move independently. We varied the proportion of HSF and LSF contrast, in order to find a point of contrast at which HSF and LSF information would be used equally well. Also, we tested whether there would be a significant difference in reaction times to either HSFs or LSFs. Due to the coarse to fine bias and the higher suitability of LSFs in motion detection, we expect subjects to have a shorter reaction time when following LSFs. Moreover, the salience of spatial frequencies may vary depending on the attributes of the landmark and surroundings. Therefore, we tested two different sceneries, and examined, whether scenery would have an influence on the point of contrast in which both HSF and LSF could be followed equally, and the reaction time. Further we tested whether it would be possible to use a linear response model similar to the pursuit model described in (Fajen & Warren, 2007) to predict subjects' movements dependent on the landmark movements.

2 Material and Methods

2.1 Subjects

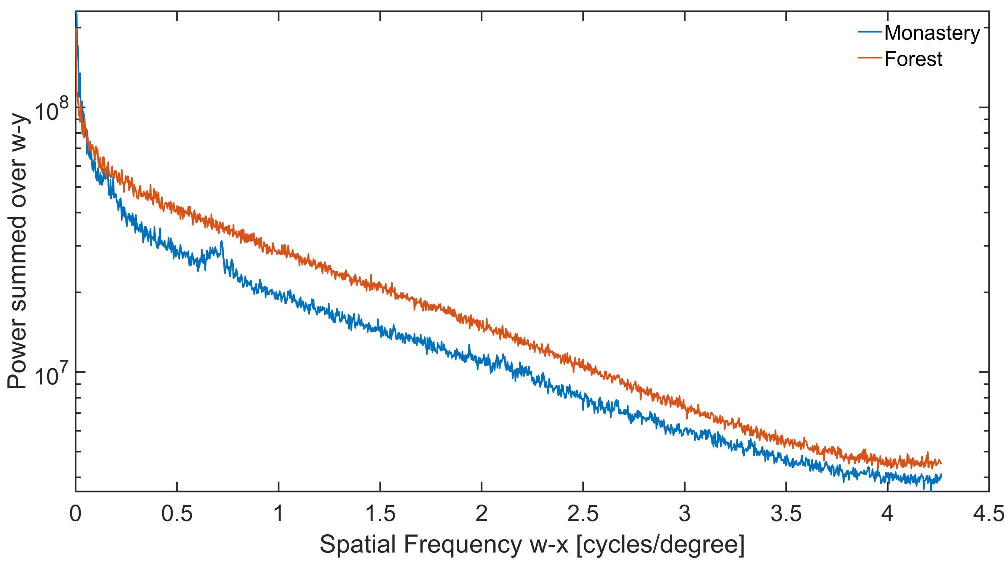
Twelve paid volunteers participated in this study (8 females, age range 20-26), one of the subjects being the author. The subjects had normal or corrected to normal vision. All subjects were given information on the general experiment and gave their informed written consent. Subjects could choose between monetary payment or course credit for their participation in the experiment.

2.2 Video sequence generation

Video sequences were generated using Matlab R2018a (MathWorks, Natick, MA 01760, USA) at 60 frames/second. As video sources two greyscale 360° panorama images with a size of 490x3072 pixel were used. Both used images held distinctly different features : The image of a village monastery (Monastery) in Bebenhausen (48°33'43.5"N 9°03'36.5"E, 9°075106), Germany and a forest (Forest) near a bridge crossing the Kirnbachriver, Germany (48°32'59.3"N 9°04'30.5"E) (see Fig.1).
The images were first separated into HSF and LSF components through filtering via Fourier transformation. The Fourier transformed images were respectively multiplied with a filter (see Eq.1 and see Fig. 2) to acquire the HSF or LSF filtered components. We chose differing σ_h and σ_l, in order to have little overlap between HSFs and LSFs. Further, before HSF and LSF image components were again merged into a composite image the respective image components were weighted to obtain different contrasts conditions. There were five contrast conditions for each scenery (Contrasts HSF/LSF : 0.10/0.90, 0.25/0.75, 0.50/0.50,0.75/0.25, 0.90/0.10).

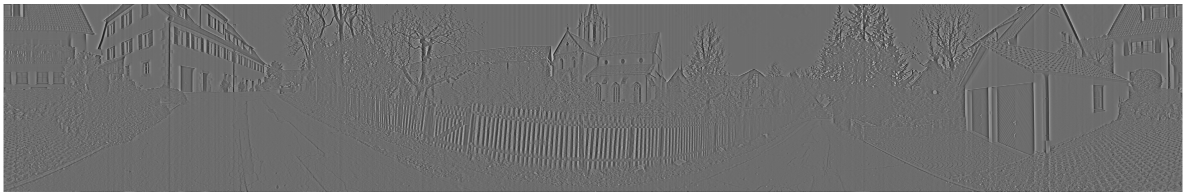
(a)

(b)

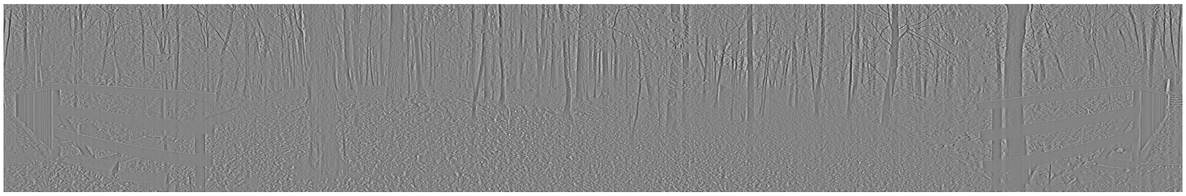


(c)

Figure 1: **(a) Monastery scenery; (b) Forest scenery; (c) Powerspectrum of Monastery and Forest sceneries** Though subjects could look around freely, they were asked to limit their head direction within roughly -90° to +90° centered around the center of the scenery, in order for all subjects to respond to roughly the same landmarks. Though the Forest scenery is energetically higher at most frequencies, most of it's features (e.g. trees, leaves) are hard to discern from the background and each other, especially as the bridge was not seen during the experiments, as it was situated outside the subjects' line of view at all times. In contrast the spire in the Monastery scenery provides a readily visible landmark (also see 3 Results).



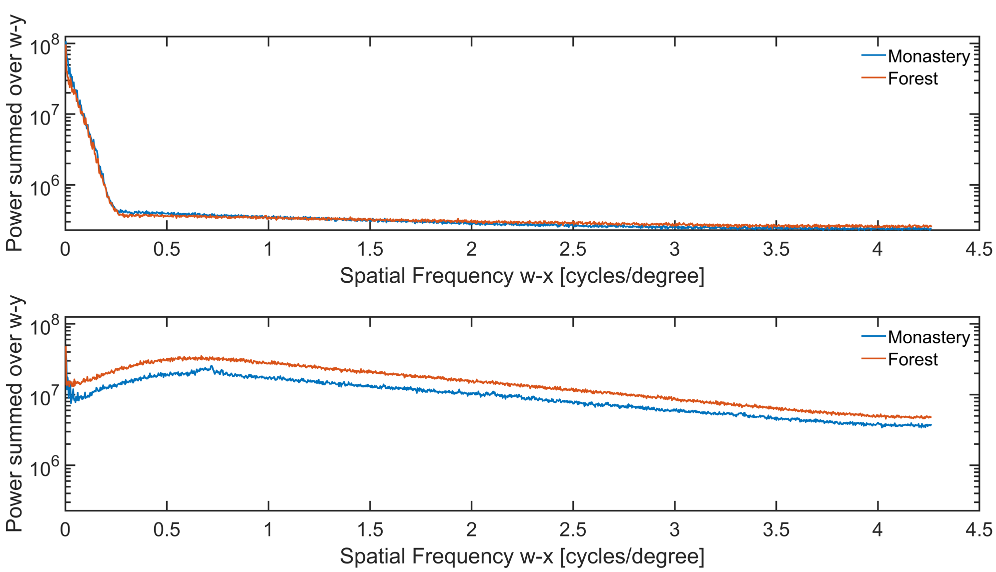
(a) Monastery scenery HSF filtered



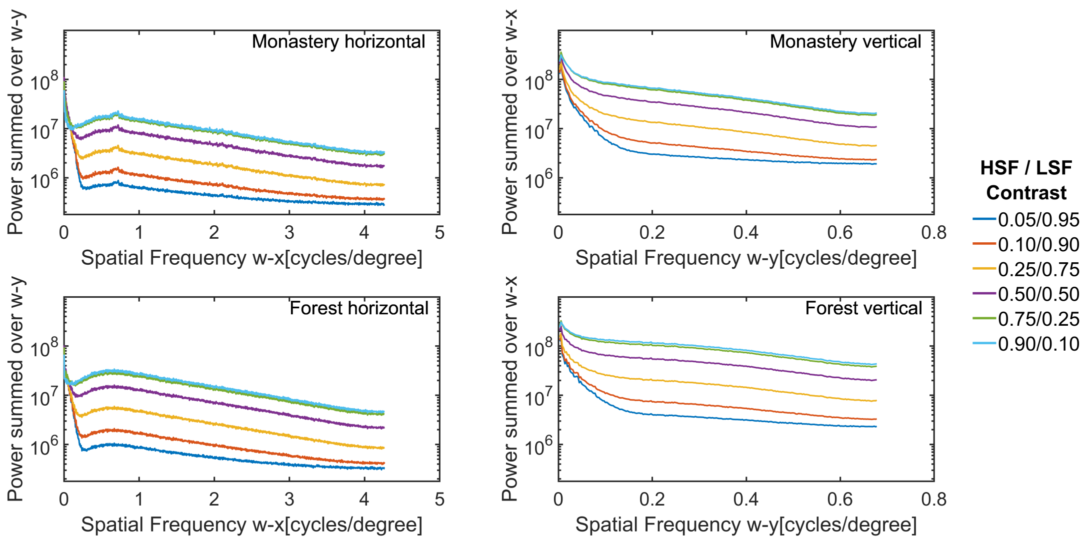
(b) Forest scenery HSF filtered

(c) Monastery scenery LSF filtered

(d) Forest scenery LSF filtered



(e)



(f)

Figure 2: **(a-d) Spatial frequency filtered sceneries. Powerspectrum of spatial frequency filtered Monastery and Forest scenery; (f) Powerspectrum of HSF/LSF composites of the Monastery and Forest scenery**
(a-d) As mentioned in 2.4 Experimental Setup and Procedure, the HSFs are harder to identify in this figure while the LSFs are easier to perceive compared to in the actual experiment due to scaling. Thus a sixth contrast condition with a rate of 0.05/0.95 (HSF/LSF) was added. The HSF and LSF components were merged to a composite. The rate of the respective components and their relative position towards the center were differed to attain different contrast conditions and landmark orientations. **(e)** The powerspectra of the HSF components at a contrast of 1.00/0.00 (HSF/LSF) and the powerspectra of the LSF components at a contrast of 0.00/1.00 (HSF/LSF). **(f)** The powespectra of the Monastery and Forest scenery composites plotted against a semi logarithmic axis. While the overall power increases with HSF content, the power at low spatial frequencies decreases.

$$f_H(\omega) = 1 - \exp\left(-\frac{\omega^2}{2\sigma_H^2}\right) \qquad \sigma_H = 191\,[pixel] \tag{1}$$

$$f_L(\omega) = \exp\left(-\frac{\omega^2}{2\sigma_L^2}\right) \qquad \sigma_L = 48\,[pixel]$$

Source: http://www.cs.uregina.ca/Links/class-info/425-nova/Lab5/index.html

ω = frequency

σ = filter width $\rightarrow$ cutoff frequency

The cutoff frequencies were chosen, in order for HSF-LSF components to be perceptually equally strong in a composite image containing 50% of HSF and LSF image components each. Additionally, we generated the powerspectra of the filtered HSF and LSF images and all resulting composite images (see Fig. 2).

During video creation, the HSF and LSF image components were respectively shifted at a predefined angle relative to the image center, according to a semi-random pattern. The subject faced the same image portion, the scenery's center, at the start of each trial, and both HSF and LSF image components started at 0° relative to the scenery's center. In each 63s long trial both HSF and LSF components´ angular directions were shifted in a stepwise fashion 10 times each. The shifts occurred at a mean interval of 3s ± 0.5s. Further it was pseudo randomized whether the shift would affect either HSF or LSF image components. The mean shift strength was 6° ± 3° in a random direction. There were four trials of each contrast condition. In the first and third trial the subjects were instructed to follow the HSF image components, in the second and fourth the LSF image components (see Fig.4). The landmark movements of all trials were generated independently, and therefore all trials differ in their landmark movements.

2.3 Virtual Environment

The virtual environment was generated using Unity (vers. 5.511f1, Unity Technologies, San Francisco, CA 94103, United States). The ground had a randomly generated black and white pattern with no viable points of orientation, the sky plane held the same pattern with the black substituted for blue (see Fig. 3). Both sky and ground plane provided optic flow information, to emulate a forward movement, but were uniformly patterned to provide no landmark information. Throughout the experiments the subjects were situated in the middle of a cylinder on which the videos were projected (see. Fig 3). The cylinder moved along with the subject, in order to maintain a fixed distance to the subject. However, the cylinder maintained a fixed angle during the entire trials and was therefore not influenced by the subjects' rotations. The cylinder's video texture, provided the only viable landmark information in the virtual environment.

2.4 Experimental Setup and Procedure

The subjects were seated during the whole experiment, wearing a HTC-Vive headset (HTC Corporation, Taoyuan City 330, Taiwan, (R.O.C)) which could display 1080 × 1200 pixels per eye at a maximum of 90 Hz, the field of view was 75° and goggles allowed up to 16 pixels/degree $\widehat{=}$ 8cpd. Before the actual experiment subjects were shown a short demonstration featuring a definition of HSFs and LSFs and their task, which also allowed them to accustom to the VR environment. Also, in order for the subjects to get better accustomed to the virtual environment, at the start of each block the subjects started at a plane patterned to resemble a highway, which indicated the image center. Also,

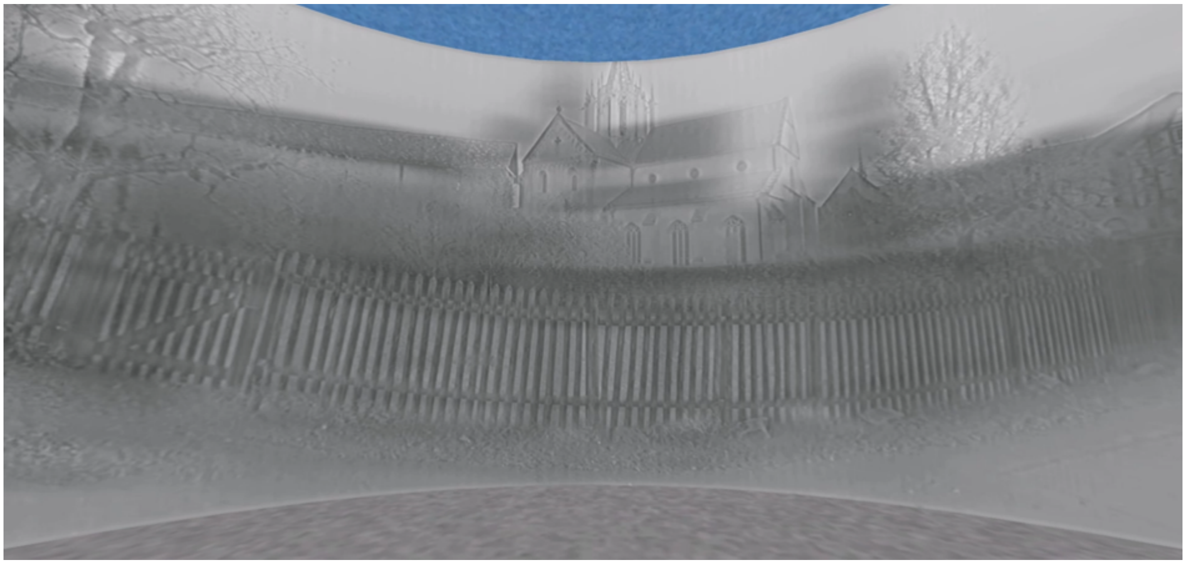


Figure 3: A part of the subjects' field of view during the experiment, depicting the section around the scenery's center. The HSF to LSF ratio in this trial was 0.50/0.50 (HSF/LSF). HSF and LSF components move pseudo-randomly, with no overlap between two movements. Ceiling and floor are uniformly patterned and provide no viable landmarks.

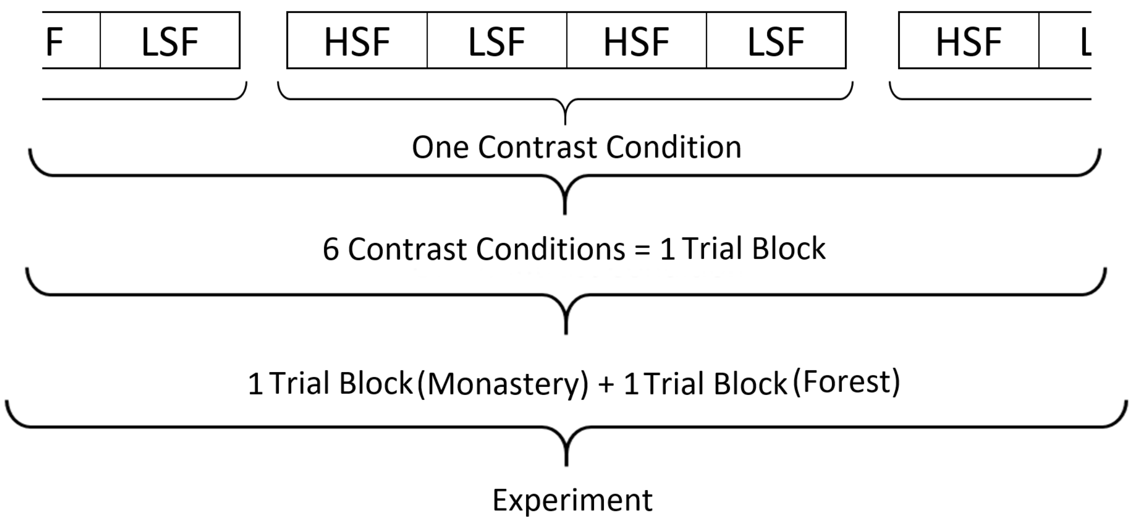


Figure 4: The Structure of the Experiment. All trials for each contrast condition were shown successively. Each trial lasted 63s and there was a 1s gap between two trials, in which subjects were shown which spatial frequency they were to follow in the coming trial. Each trial block spanned all contrast conditions (Contrasts HSF/LSF : 0.10/0.90, 0.25/0.75, 0.50/0.50,0.75/0.25, 0.90/0.10), and all trials within a trialblock featured the same scenery. After the Monastery-trial block subjects could take a break, if they wanted, before starting the Forest-trial block. After the experiment a short debriefing was held to question subjects on which landmarks they had used.

there was no landmark movement during these initial 10s. At the beginning of each trial the subjects were shown, whether they had to follow the HSF or LSF image components. The subjects' orientation was centered on the middle of each scenery, at the start of all trials. The subjects task was to maintain their original bearing towards the selected spatial frequency component, while the subjects moved continuously in the virtual environment throughout the experiment. While the subjects could not change their movement speed, they could change their heading angle freely using an Xbox One controller (Microsoft Corporation, Redmond, WA 98052-6399, USA). Both the subjects position, the subjects head direction and the subjects heading angle in lab coordinates were recorded at a rate of 60Hz. The cutoff frequencies of HSF and LSF were chosen, in order for both spatial frequencies to be equally perceivable at a HSF/LSF ratio of 0.50/0.50. The discernment however was made on a computer screen and the image therefore spanned a much smaller viewing angle than the 360° panorama in the VR. As the image was greatly expanded in the VR, HSF components became more easily perceived, due to HSFs being easier to perceive at larger scales (Morrison & Schyns, 2001), while LSFs decreased in cycles/° viewing angle even further, thus losing information. Consequently, the HSFs were generally more easily perceived during the experiments, than as intended. We rebalanced our experiment by adding another contrast condition (HSF/LSF ratio: 0.05/0.95) starting with the sixth subject.

Each trial lasted for 63s. Each contrast condition contained four trials, two in which subjects were to follow HSF landmarks and two in which LSF landmarks were to be followed (see Fig. 4). All six different contrast conditions of one scenery formed one trial block, which therefore consisted of 24 trials. The whole experiment consisted of two trial-blocks. Subjects could take a pause between the two trial-blocks, if they desired to. At a short debriefing at the end of the experiment subjects were asked which landmarks they had used during the experiments.

2.5 Analysis

Subject heading angles in lab coordinates as provided by Unity were recorded on a frame by frame basis i.e. with 60 samples per second. To acquire the subjects' reaction time for all trials, we calculated the cross correlation between the landmark changes of both HSF and LSF components and the subject heading angle change. The time difference with the maximum absolute correlation was considered the average reaction time, if the time difference was within 2s. Further, the maximum correlation had to be both at least two standard deviations higher than the second largest correlation. Additionally, as we had calculated the correlation of the subject heading angle change with both HSF and LSF landmark changes, we had two possible reaction times per trial. The reaction was only then deemed valid, if the maximum correlation of subject heading angle change and target landmark angle was higher than the maximum correlation to the non-target landmark. Subject reaction time data, which did not meet these requirements were excluded from further analysis.

To determine the subject performance i.e. how well subjects were able to follow either HSF or LSF components, the sum of squared residuals between the subjects' heading angle (see Fig.5) and the respective landmark angle was calculated. It was determined using a general linear F-Test (see Eq.2), whether subjects could perform significantly better than in a null-hypothesis of making no turn at all. For each trial the F*-value was compared to a threshold to determine whether the subject was able to perform better than chance. It was deemed that subjects could follow the landmarks in a condition, if subjects could perform significantly better than in the null hypothesis in at least 50% of all trials in that condition. Further, all degrees of Freedom were the same for all trials in all conditions, as each trial had the same length, thus if a trial would have a bigger than average F*-value, the cause would be a bigger difference between the subject's performance and the null hypothesis, in favour of the subject (see Eq. 2). Therefore we compared the F*-values against each other, to determine whether subjects' performance was better for a particular condition.The subject performance F*-Values were compared using a 3-Way ANOVA which factors were contrast (HSF/LSF ratio), followed spatial frequency (HSF and LSF) and scenery (Monastery/Forest) for the whole dataset. Further a 2-Way ANOVA for the datasets separated by the two sceneries was used, which had as factors contrast (HSF/LSF ratio) and followed spatial frequency (HSF and LSF).
As there were two trials per condition featuring different landmark movements, we effectively had two datasets. We used each dataset to establish a multiple linear regression model (linear response model)

in order to describe the subjects' movements.

The subjects heading angle change was modeled using the HSF and LSF landmarks´ movements. We determined whether the linear response model was able to describe the subjects successfully by comparing the linear response model's estimate of the subject to the subjects trajectory (General Linear F-Test, see Eq.2). We compared the linear response model's estimate to the subjects' trajectory of the used model coefficients source dataset, to test whether the subjects behavior was indeed linear. Further we compared the linear response models estimate to the subjects' trajectory of the dataset not originating the used model coefficient set, to test whether the linear response model would be able to describe subject behavior outside it's source dataset. For each trial the F*-value was compared to a threshold to determine whether the model was able to describe the subject's behavior. The null hypothesis was a model without any coefficients, which therefore makes no turns. It was deemed that subjects could follow the landmarks in a condition, if subjects could perform significantly better than in the nullhypothesis in at least 50% of all trials in that condition. The F*-Values were compared using a 5-Way ANOVA with following factors to test whether the linear response model performs equally well on a foreign dataset as well as it's own: Contrast, followed spatial frequency, scenery, model coefficient set, crosstest/test on source dataset. Moreover, we tested whether the linear response model's performance differed with the used model coefficient set using a 4-Way ANOVA with the factors: Contrast, followed spatial frequency, scenery, model coefficient set. We further used a 3-Way ANOVA with contrast, followed spatial frequency and scenery as factors to determine whether the linear response model's performance differed between sceneries. A 2-Way ANOVA was used to test the effect of contrast and target spatial frequency on the linear response models performance.

Additionally we calculated the sum of the model coefficients for each trial, to acquire the response strength. The response strength describes the power of the subjects reaction to a landmark angle change. We further tested, whether there would be a correlation of response strength and subject performance, as it is possible that a decrease of subjects performance would be mainly to the subjects not responding or responding too little.

Finally, we also tested for a correlation of response strength and the linear response model's performance, as a positive correlation may indicate, that a decrease in the linear response model's performance was caused by the subjects not responding to all landmark angle changes.

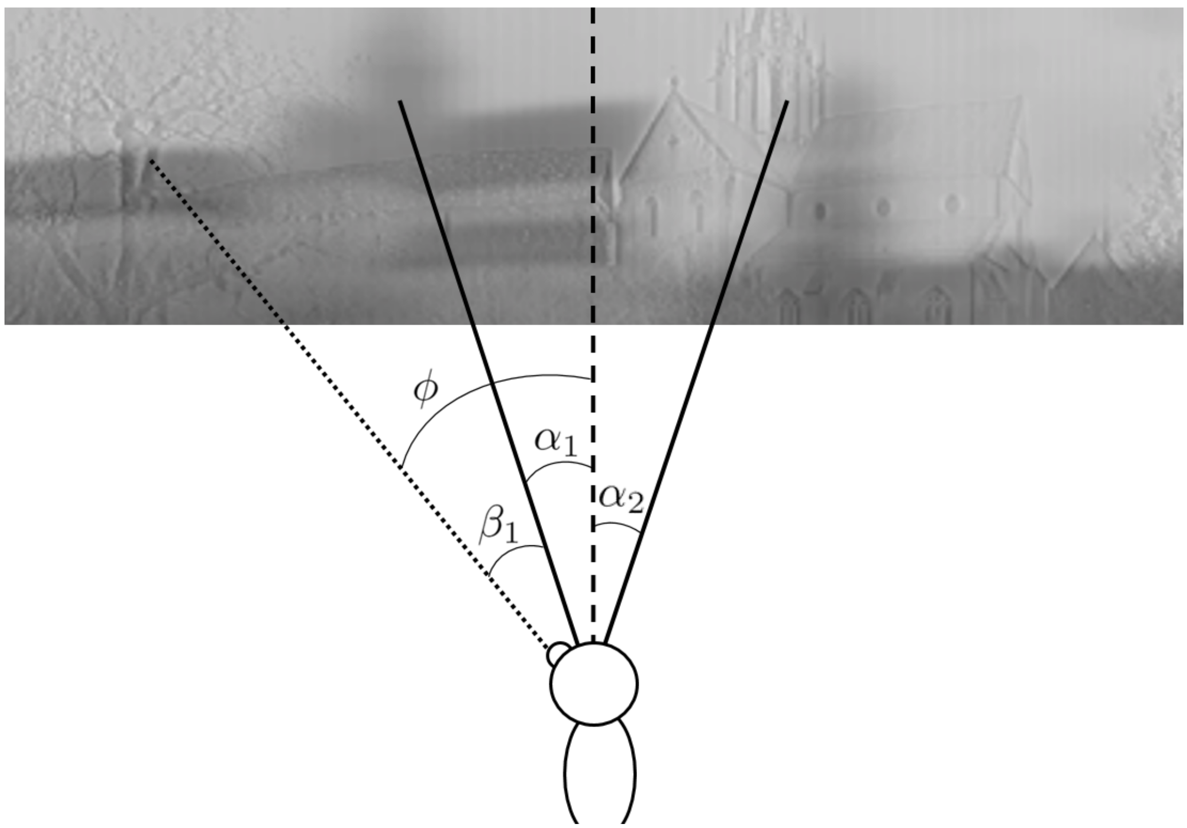


Figure 5: Explanation of used variables. During the experiment the landmarks will change their orientation. The landmark orientation ($\alpha_{1/2}$), is the angle between an absolute orientation axis (dashed line) and the respective spatial frequencies scenery's center. If for example instructed to follow LSFs (moving with α_1) the subjects will try to keep their bearing (β_1) low by changing their heading (ϕ)

$$F^* = \frac{SSR(R) - SSR(F)}{df_R - df_F} / \frac{SSR(F)}{df_F} \tag{2}$$

Source: 6.2 - The General Linear F-Test — STAT 501 (17.04.2019),

https://newonlinecourses.science.psu.edu/stat501/node/295/

n = number of trials

SSR(R) = sum of squared residuals of the reduced model (Null-hypothesis)

SSR(F) = sum of squared residuals of the full model

df_R = Degrees of Freedom of the reduced model → n - 1

df_F = Degrees of Freedom of the full model = Degrees of Freedom of the reduced model - number of added variables

γ_0 = overall mean angle

ϵ = error

$\text{Nullhypothesis}_{subject\ performance} \rightarrow \quad \dot{\phi}_i = \gamma_0 + \epsilon_i; \qquad \gamma_0 = 0$

$\text{Fullmodel}_{subject\ performance} \rightarrow \quad \dot{\phi}_i = \dot{\alpha}_i + \epsilon_i$ also see Eq.3

$\text{Nullhypothesis}_{model\ performance} \rightarrow \quad \dot{\phi}_{model,i} = \gamma_0 + \epsilon_i; \qquad \gamma_0 = 0$

$\text{Fullmodel}_{model\ performance} \rightarrow$ see Eq.4 and 5

df_F(subject performance)→ n - 1 - 1 (subject input)

df_F(linear response model performance)→ n - 1 - 18 (number of model coefficients) - 2 (HSF and LSF landmark angle changes)

We expected subjects to primarily react to movement during the experiments, rather than absolute position. Therefore, subjects will react to landmark angle changes ($\dot{\alpha}_{1/2}$), with a change of their own heading direction ($\dot{\phi}$) to keep their relative angle towards the landmarks, thus keeping a globally low bearing change (β).

$$\dot{\beta} = \dot{\alpha} - \dot{\phi} \rightsquigarrow 0 \tag{3}$$

We expected the subjects to react after a landmark angle change within a maximum response time (K).We modeled the subjects heading change at a time (t) using the landmark angle changes ($\dot{\alpha}$) in a time frame of $t - K$ to $t - 1$, weighted by coefficients (c) which describe the subjects behavior within a time frame of $1 - J$, with J being the length of the whole trial.

$$\text{for} \quad i = K_1, K_2 ... J$$

$$\dot{\phi}_i = \sum_{i=1}^{J} = \dot{\alpha}_{1,i} * c_{1,i-k} + \dot{\alpha}_{2,i} * c_{2,i-k} + \epsilon_i \tag{4}$$

This leads to the regression equation

$$\begin{pmatrix} \dot{\phi}_{K+1} \\ \dot{\phi}_{K+2} \\ . \\ . \\ \dot{\phi}_J \end{pmatrix} = \begin{pmatrix} \dot{\alpha}_{1,1} & \dot{\alpha}_{1,2} & .. & \dot{\alpha}_{1,K} & \dot{\alpha}_{2,1} & \dot{\alpha}_{2,2} & .. & \dot{\alpha}_{2,K} \\ \dot{\alpha}_{1,2} & \dot{\alpha}_{1,3} & .. & \dot{\alpha}_{1,K+1} & \dot{\alpha}_{2,2} & \dot{\alpha}_{2,3} & .. & \dot{\alpha}_{2,K+1} \\ . & . & .. & . & . & . & .. & . \\ . & . & .. & . & . & . & .. & . \\ \dot{\alpha}_{1,J-K} & \dot{\alpha}_{1,J-K+1} & .. & \dot{\alpha}_{1,J-1} & \dot{\alpha}_{2,J-K} & \dot{\alpha}_{2,J-K+1} & .. & \dot{\alpha}_{2,J-1} \end{pmatrix} * \begin{pmatrix} c_{1,K} \\ c_{1,K-1} \\ . \\ . \\ c_{1,1} \\ c_{2,K} \\ c_{1,K-1} \\ . \\ . \\ c_{2,1} \end{pmatrix} \quad (5)$$

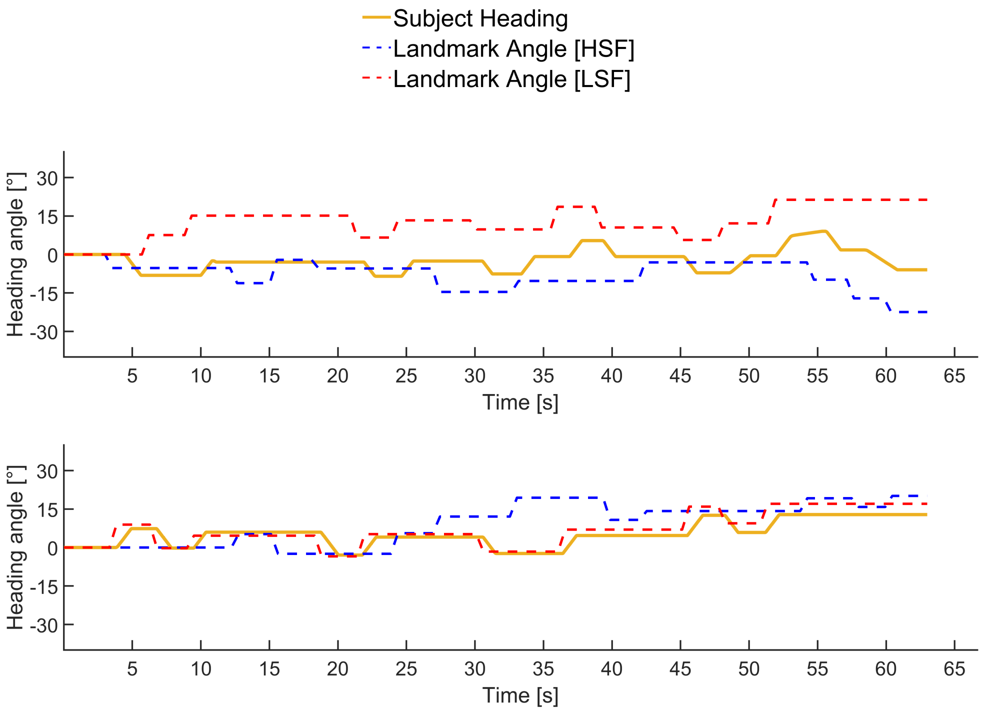
3 Results

We recorded the subjects' heading and head orientation, while they tried to follow either HSF or LSF landmarks moving in a VR environment. We varied contrasts between HSF and LSF components between trials.

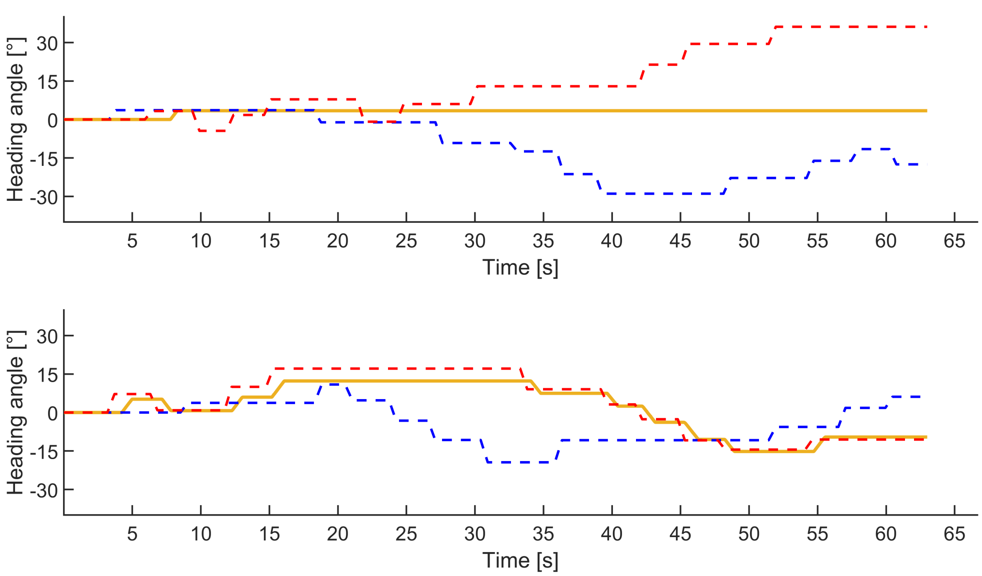
3.1 Subject Performance

Subjects showed no substantial bias towards either left or right during the experiments, as the only significant bias was merely $4.80 * 10^{-4}$ degrees to the right (One Sample T-test, $t_{Monastery}(62)$=-3.047, p=0.006; One Sample T-test, $t_{Forest}(62)$= 0.223, p=0.826; One Sample T-test, $t_{both\ sceneries}(123)$= -0.493, p=0.627).
We measured subject performance by comparing their error to the target landmarks' trajectory variability (General Linear F-Test). A significant majority of subjects (>64.28%) performed above chance in all trials (General Linear F-Test, $F_{both\ sceneries}$ (3688,3689)$\geq$ 1.360, p< 0.05) except in the Forest scenery for the 0.05/0.95 (HSF/LSF) contrast condition when following HSFs (35.71%) (see Tab. 1).

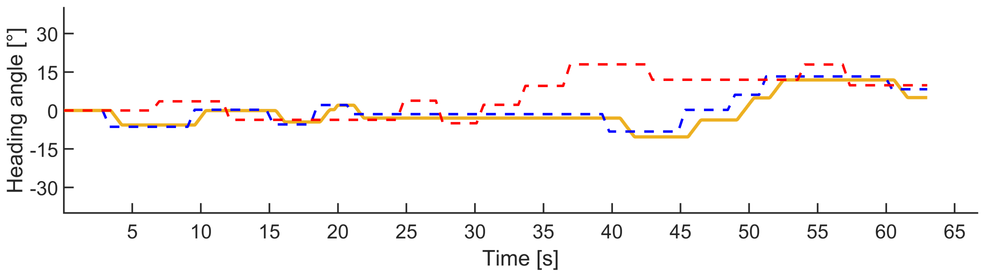
There was no difference in subject performance between the first and second trial within one condition (4-Way, factors Contrast, Followed Spatial Frequency, Scenery, Trial order, ANOVA,$F_{both\ sceneries}$ (1, 509) =2.61, p=0.106). Contrast had a significant effect on subject performance both in the Monastery scenery (2-Way ANOVA, factors: Contrast, Followed Spatial Frequency,$F_{Monastery}$ (5,256) =8.38, p<0.001), and in the Forest scenery (2-Way ANOVA, factors: Contrast, Followed Spatial Frequency,F_{Forest} (5, 256) =13.99, p<0.001)(see Figs. 6 and 7). Further subjects performed equally in the Monastery scenery and Forest scenery (3-Way ANOVA, factors: Contrast, Followed Spatial Frequency, Scenery,$F_{both\ sceneries}$ (1,509) =0.59, p=0.443). Also, subjects performed better when following HSFs in the Monastery scenery (2-Way ANOVA, factors: Contrast, Followed Spatial Frequency, $F_{Monastery}$ (1, 256) =4.01, p=0.046) while there was no difference in performance between HSFs and LSFs in the Forest scenery (2-Way ANOVA, factors: Contrast, Followed Spatial Frequency, Scenery,F_{Forest} (1,256) =0.2, p=0.658).



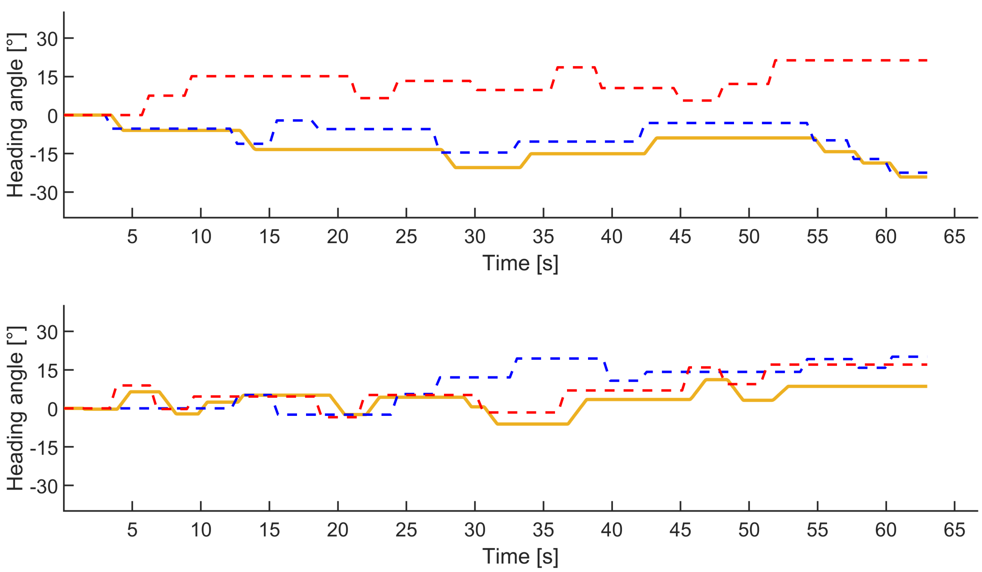
(a) Monastery, 0.05/0.95 (HSF/LSF), HSF as target (upper), LSF as target as target (lower)



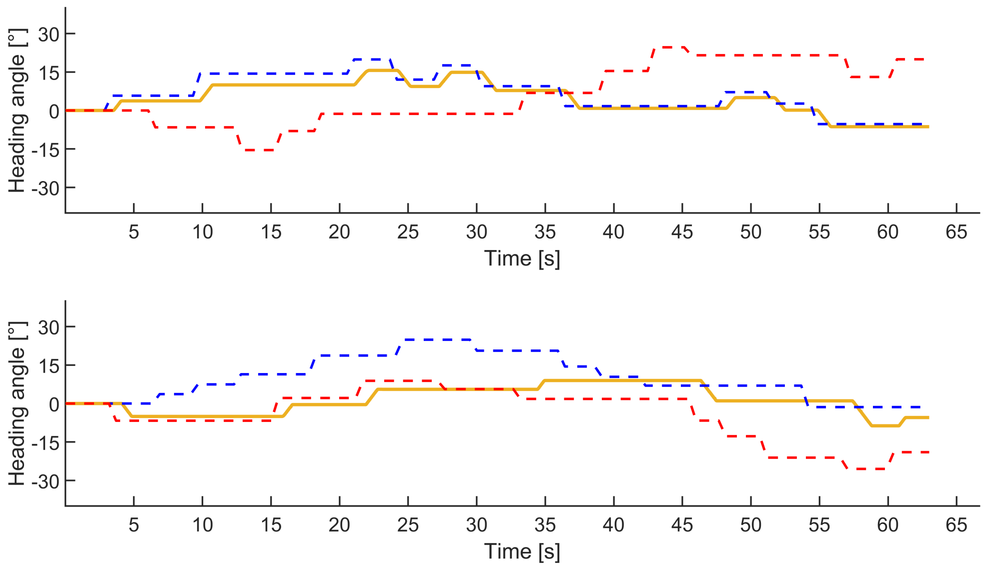
(b) Forest, 0.05/0.95 (HSF/LSF), HSF as target (upper), LSF as target as target (lower)



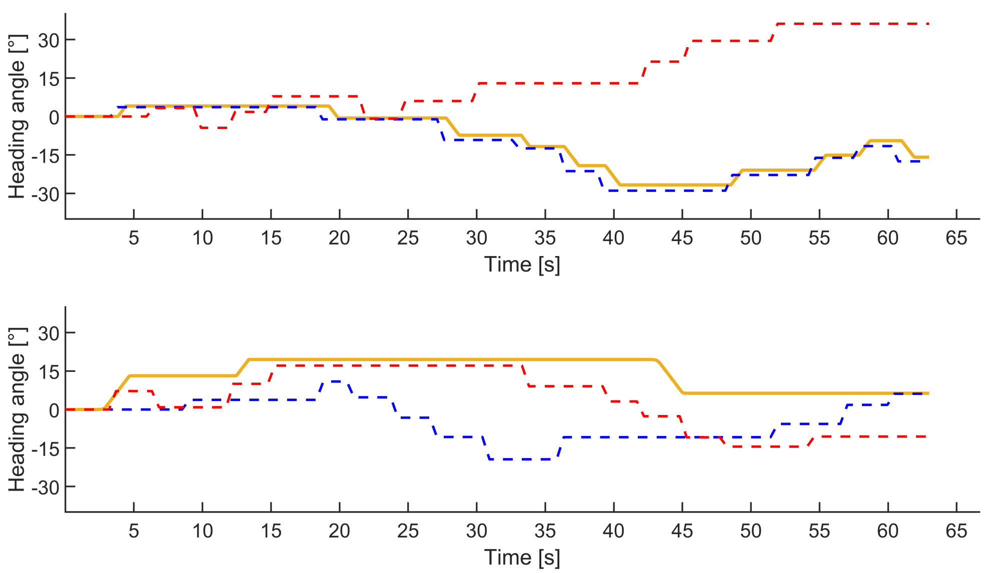
(c) Monastery, 0.25/0.75 (HSF/LSF), HSF as target (upper), LSF as target as target (lower)



(d) Forest, 0.25/0.75 (HSF/LSF), HSF as target (upper), LSF as target as target (lower)



(e) Monastery, 0.90/0.10 (HSF/LSF), HSF as target (upper), LSF as target as target (lower)

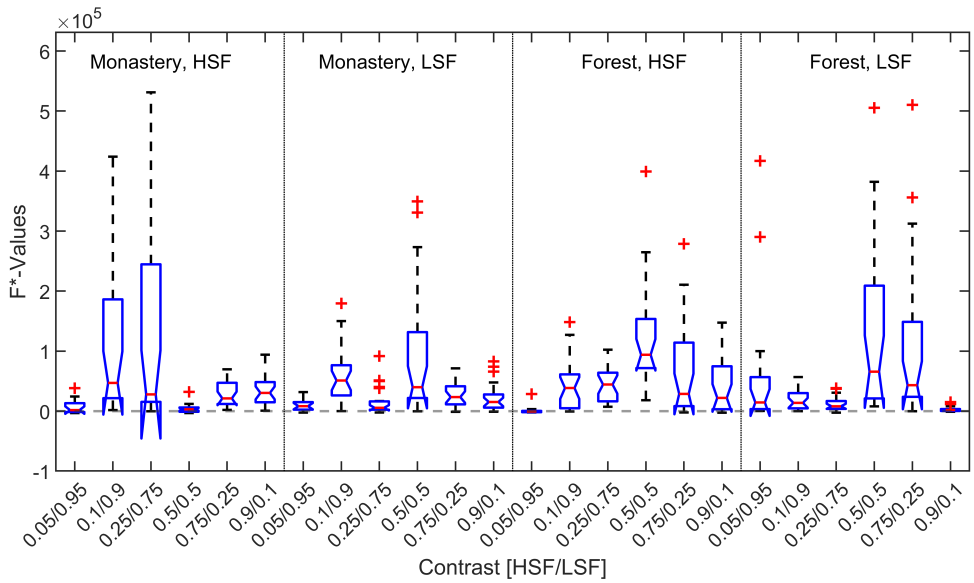


(f) Forest, 0.90/0.10 (HSF/LSF), HSF as target (upper), LSF as target as target (lower)

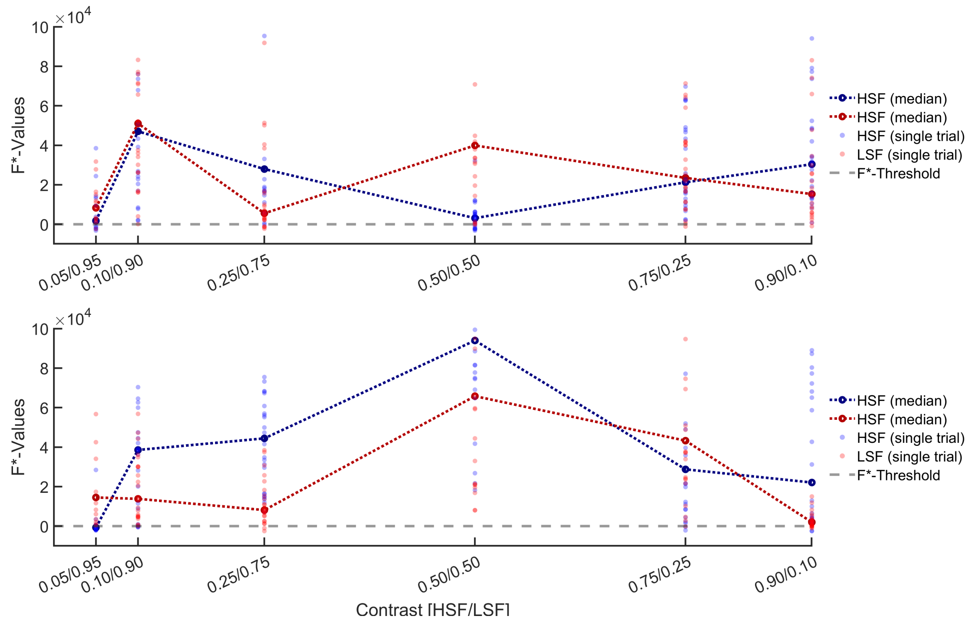
Figure 6: **Data of subject TM (male) Subject heading and landmark orientation.** At a contrast of 0.25/0.75 (HSF/LSF) both HSFs and LSFs were roughly equally perceptually strong in the VR environment. It is notable, that our calculated subject performance and subject response strength can already be estimated from the raw data.

Percentage of all trials of one condition (%)				
	M_HSF	M_LSF	F_HSF	F_LSF
0.05/0.95 (HSF/LSF)	64.28	92.85	35.71	100.00
0.10/0.90 (HSF/LSF)	100.00	100.00	83.33	95.83
0.25/0.75 (HSF/LSF)	95.83	83.33	100.00	91.67
0.50/0.50 (HSF/LSF)	70.83	95.83	100.00	100.00
0.75/0.25 (HSF/LSF)	100.00	95.83	91.67	95.83
0.10/0.90 (HSF/LSF)	100.00	95.83	91.67	83.33

Table 1: **Percentage of trials, in which the subjects were able to follow the target landmark.** The total number of trials in the 0.05/0.95 (HSF/LSF) contrast conditions were 12 for HSF and LSF conditions each. The total number of trials for all other conditions were 24 for each condition. The first column states the contrast. The first row states the used scenery (M/F) and the target spatial frequency (HSF/LSF). It was assumed, that the subjects were able to follow the target spatial frequency in a condition, if subjects were better than would be expected in the the null-hypothesis (subjects make no turns) in at least 50% of all trials of the condition.



(a)



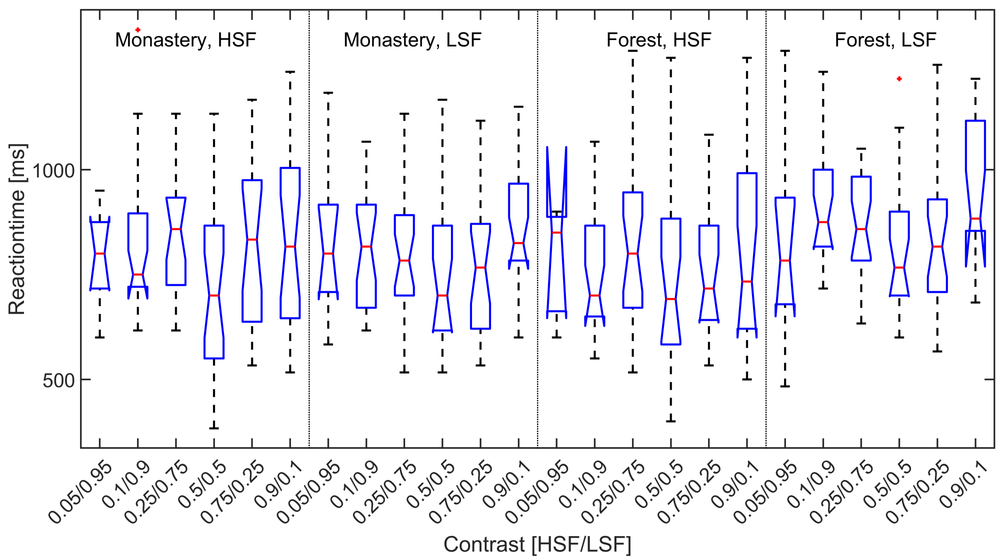
(b) Monastery (upper) and Forest (lower)

Figure 7: **Subjects' performance.** Note that **(a)** and **(b)** have different y-axes and that far outliers are not shown in favour of clarity. The grey horizontal lines denote the significance threshold, F*-Values larger than 3.89 were considered significant. **(a)** The error of all subjects sorted by conditions. Red bands indicate the median error; notches signify the median's 95% confidence interval. The 25 and 75 percentile are represented as the boxes' edges. Whiskers indicate the data points farthest from the median which are not considered outliers, outliers are shown as red crosses. **(b)** The median of subjects' error and error of single trials in the Monastery (upper) and Forest (lower) scenery. The error of single trials is indicated by transparent dots. Confidence Intervals were omitted for clarity, but can be seen in **(a)**. Generally, HSFs and LSFs were followed equally well. Further, subjects performed equally well in both sceneries.

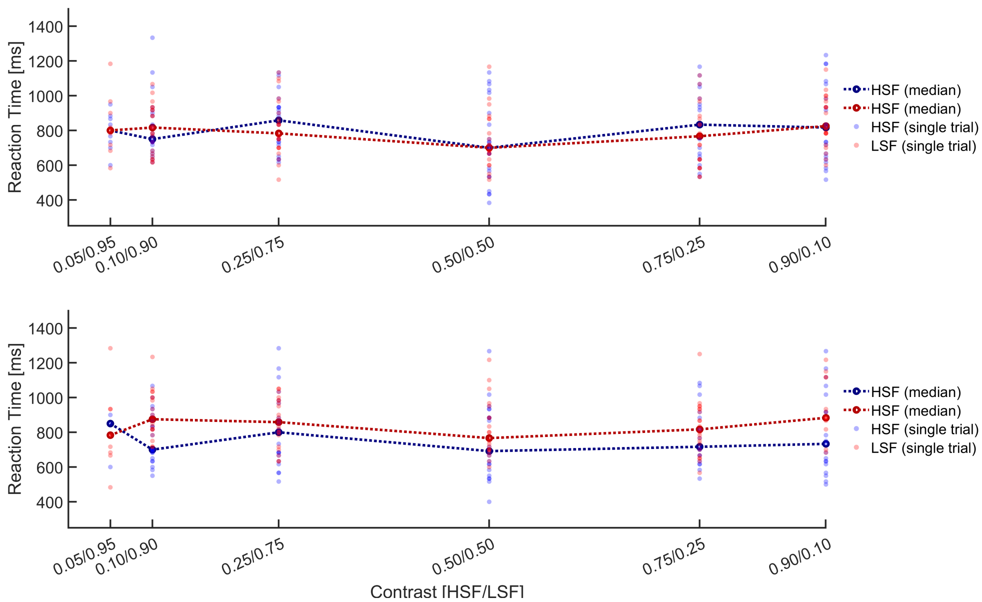
3.2 Reaction Time

The mean reaction time in the Monastery condition was 800.215ms ($TR_{Monastery}$= 800.215 ms ± 176.848 ms) and 815.818 ms (RT_{Forest}= 815.818 ms ± 178.836 ms) in the Forest scenery.

The subjects' reaction time did not differ between sceneries (3-Way ANOVA, factors: Contrast, Followed Spatial Frequency, Scenery,$F_{both\ sceneries}$ (1,430) =0.46, p=0.499) (see Fig. 8). Further there was only a significant effect of contrast on the reaction times, if the reaction times of both sceneries were considered jointly (3-Way ANOVA, factors: Contrast, Followed Spatial Frequency, Scenery,$F_{both\ sceneries}$ (1,430)=3.11, p=0.009), whereas contrast had no significant effect on subject reaction time both in the Monastery scenery (2-Way ANOVA, factors: Contrast, Followed Spatial Frequency,$F_{Monastery}$ (5,221)=2.09, p=0.068) and Forest scenery (2-Way ANOVA, factors: Contrast, Followed Spatial Frequency,F_{Forest} (5,204)=1.20, p=0.313) if considered alone. Subjects reaction times were shorter for HSFs in the Forest scenery (2-Way ANOVA, factors: Contrast, Followed Spatial Frequency,F_{Forest} (1,204) =7.54, p=0.007), while spatial frequency had no effect in the Monastery scenery (2-Way ANOVA, factors: Contrast, Followed Spatial Frequency,$F_{Monastery}$ (1,221) =1.80, p=0.899)



(a)

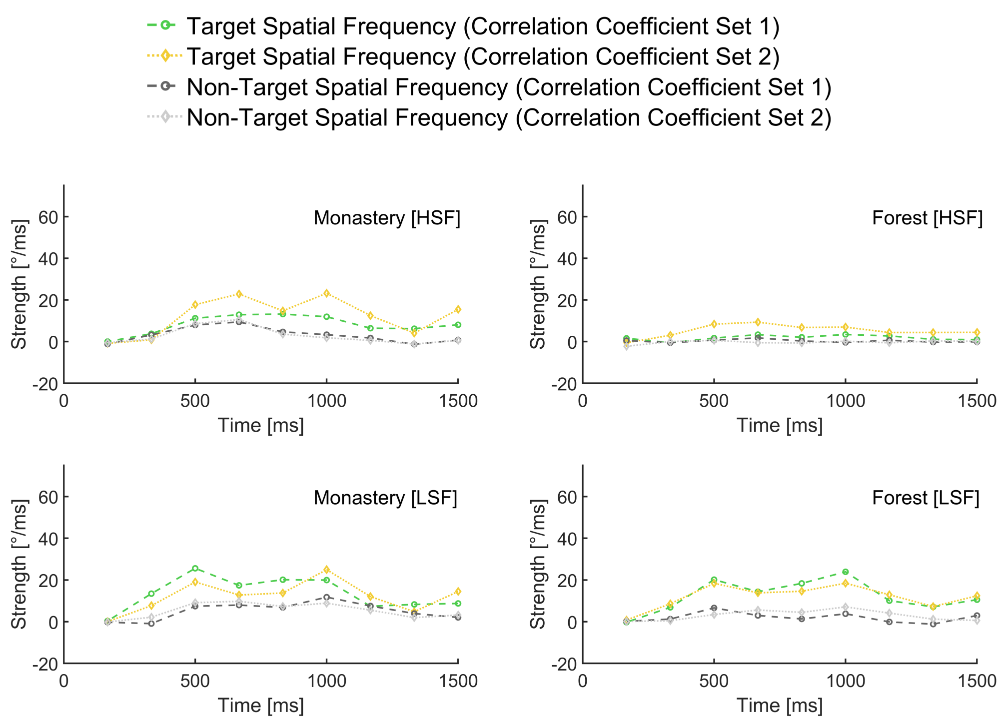


(b) Monastery (upper) and Forest (lower)

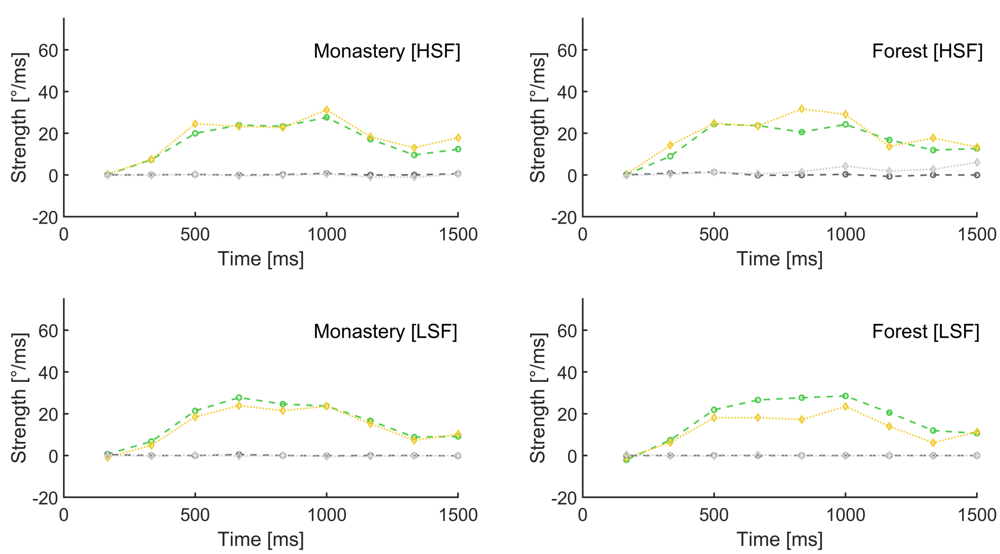
Figure 8: **Subjects' reaction times. (a)** The subjects' reaction time sorted by conditions. The first character indicates the scenery (Monastery or Forest), the second characters indicate whether HSF or LSF were followed during the trial. The last characters represent the contrasts of HSF and LSF elements. Red bands indicate the median; notches indicate the median's 95% confidence interval. The 25 and 75 percentile are represented as the boxes' edges. Whiskers indicate the data points farthest from the median which are not considered outliers, outliers are shown as red crosses. **(b)** The median reaction time and reaction times of single trials in the Monastery (upper) and Forest (lower) scenery. Subjects reaction times were slower when following LSFs in the Forest scenery. There was no difference in reaction times between HSFs and LSFs in the Monastery scenery.

3.3 Model Coefficients

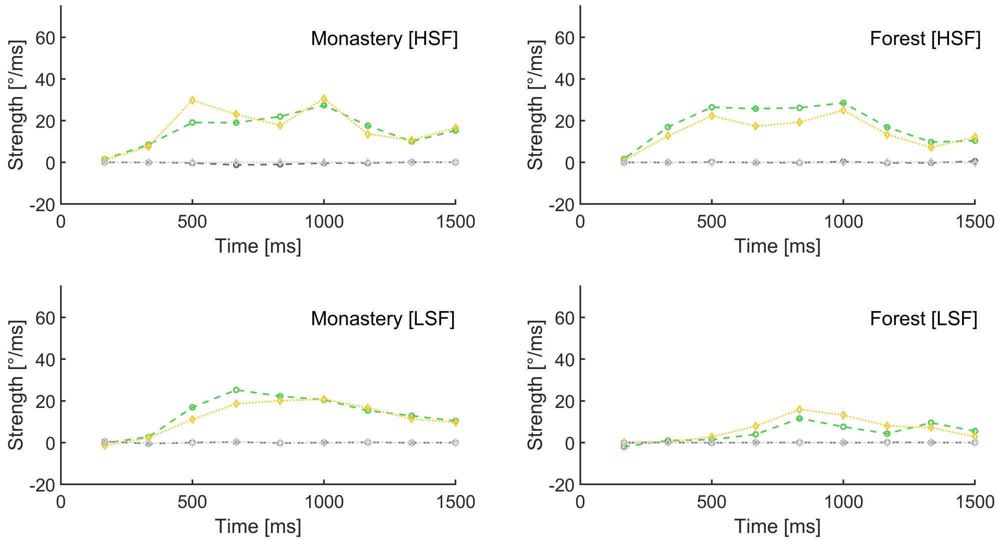
Our experiment was such structured, that for each contrast and scenery condition there were two trials were subjects had to follow HSF or LSF respectively. Thus the resulting dataset could be split into two separate datasets each containing all conditions. We acquired two different sets of model coefficients for our linear response model, as we tested the linear response model on each of the two datasets. It is noteworthy that subjects' reaction time and response strength can also be seen in the coefficients (see Fig. 9).



(a) Contrast at 0.05/0.95 (HSF/LSF)



(b) Contrast at 0.25/0.75 (HSF/LSF)

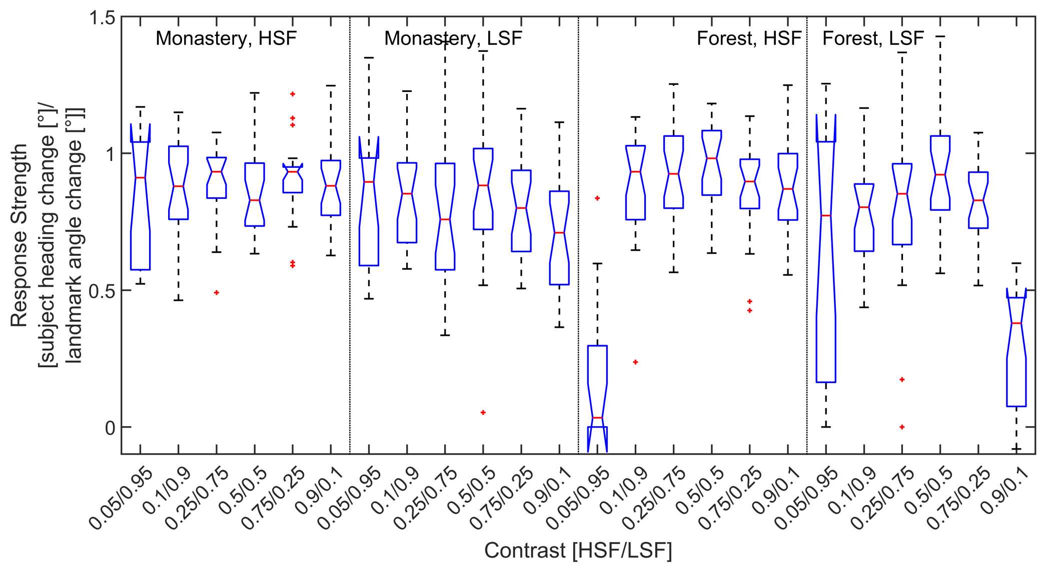


(c) Contrast at 0.10/0.90 (HSF/LSF)

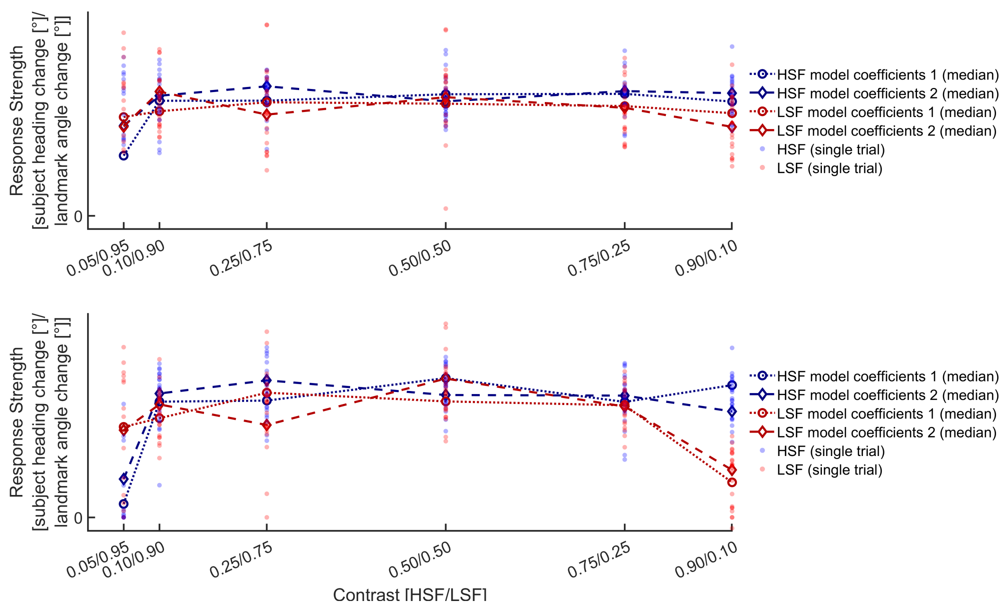
Figure 9: **Model coefficients.** It is noteworthy, that the model coefficients reflects both response strength as well as reaction time. It can further be seen, that a low contrast for the target landmark affected subject behavior in Forest scene trials more than it did in Monastery scene trials.

3.4 Subject response strength

The subjects' response strength, which describes the amplitude of the response of a subject to a landmark movement, was significantly larger in the Monastery scenery compared to the Forest scenery (3-Way ANOVA, factors: Contrast, Followed Spatial Frequency, Scenery, $F_{both\ sceneries}(1{,}517)=20.65$, $p<0.001$) (see Fig.10). Contrast had no effect on response strength in the Monastery scenery (2-Way ANOVA, factors: Contrast, Followed Spatial Frequency, $F_{Monastery}$t(5,256)=0.98, p=0.432). Further, subjects responded significantly weaker when target landmarks were presented at their lowest contrast in the Forest scenery (2-Way ANOVA, factors: Contrast, Followed Spatial Frequency, $F_{Forest}(5{,}256)=30.26$, $p<0.001$). Subject's response strength was higher when following HSF landmarks in both sceneries (2-Way ANOVA, factors: Contrast, Followed Spatial Frequency, $F_{Monastery}(1{,}256)=8.02$, p=0.005; factors: Contrast, Followed Spatial Frequency, ANOVA, $F_{Forest}(1{,}256)=6.57$, p=0.011). Also subject performance increased with subject response strength (r(534)=0.125, p= 0.004).



(a)



(b) Monastery (upper) and Forest (lower)

Figure 10: **Response strength.** **(a)** Subjects' response strength sorted by conditions. The first character indicates the scenery (Monastery or Forest), the second characters indicate whether HSF or LSF were followed during the trial. The last characters represent the contrasts of HSF and LSF elements. Red bands indicate the median; notches indicate the median's 95% confidence interval. The 25 and 75 percentile are represented as the boxes' edges. Whiskers indicate the data points farthest from the median which are not considered outliers, outliers are shown as red crosses. **(b)** The response strength derived using all data of either dataset and the response strength derived from single trials in the Monastery (upper) and Forest (lower) scenery. Contrast had an effect on response strength in both sceneries. Subjects reacted stronger to HSFs in the Forest scenery while there was no effect in the Monastery scenery.

3.5 Model Performance

We tested our linear response model's model coefficient sets by cross-testing them, using the respective other dataset's subject heading as a baseline (General Linear F-Test). Further, the linear response models were tested against their own dataset to determine whether there is a linear response towards landmark changes and whether the subject response would be homogenous enough to be described using a single model. The linear response model was not able to describe a subjects' behavior in a sufficient number of trials ($\leq 42.86\%$) when following LSFs at the 0.05/0.95 condition in the Monastery scenery regardless of used model coefficient set (see. Tab. 3). Further the linear response model based on the first model coefficient data set was also not able to describe the behavior of enough subjects following LSFs in the 0.90/0.10 contrast condition in either scenery (Monastery and Forest, 41.67%). Otherwise, the linear response model was able to describe the behavior of a significant percentage of all subjects' (>57.14%). The linear response model based on the second model coefficient set was not able to describe the behavior of subjects for all conditions (>57.14%), except for the 0.05/0.95 (HSF/LSF) contrast condition when subjects were to follow LSFs in the Monastery scenery. When tested against the used model coefficients' source data set, the linear response model was able to describe most subjects' behavior in any condition ($\geq$ 50%), using the first model coefficient set. Using the second model coefficient set on it's source data set, the linear model was able to describe all conditions ($\geq$50%), except for the 0.05/0.95 (HSF/LSF) contrast condition when subjects were to follow LSFs in either scenery ($\leq$42.86%) and when subjects were to follow HSFs in at a contrast of 0.25/0.75 (HSF/LSF) in the Forest scenery.

The linear response model could describe the subject's behavior equally well regardless of whether the used model coefficients and the subject data were of the same dataset (5-Way ANOVA, factors: Contrast, Followed Spatial Frequency, Scenery, Model Coefficient Set, Crosstest/Test on source dataset, $F_{both\,sceneries}$ (1,1036) =0.63, p=0.428). Further, the linear response model's performance did not differ using either model coefficient set (4-Way ANOVA, factors: Contrast, Followed Spatial Frequency, Scenery, model coefficient Set, $F_{both\,sceneries}$ (1,509) =0.76, p=0.388) (see Fig.11 and 12). Also, there was no significant difference in model performance for either scene (3-Way ANOVA, factors: Contrast, Followed Spatial Frequency, Scenery, $F_{both\,sceneries}$ (1,517) =0.73, p=0.394), except in interaction with contrast (3-Way ANOVA, factors: Contrast, Followed Spatial Frequency, Scenery, $F_{both\,sceneries}$ (5,517) =6.20, p<0. 001)(see Fig. 12). Moreover, the 0.10/0.90 (HSF/LSF),0.50/0.50 (HSF/LSF) and 0.75/0.25 (HSF/LSF) contrast conditions were described significantly better than the 0.05/0.95 (HSF/LSF) contrast condition (2-Way ANOVA, factors: Contrast, Followed Spatial Frequency, $F_{both\,sceneries}$ (5,524) =5.74, p<0. 001). The linear response model described subject behavior significantly better when HSFs were followed (2-Way ANOVA, factors: Contrast, Followed Spatial Frequency, $F_{both\,sceneries}$ (1,524) =6.33, p=0. 012).

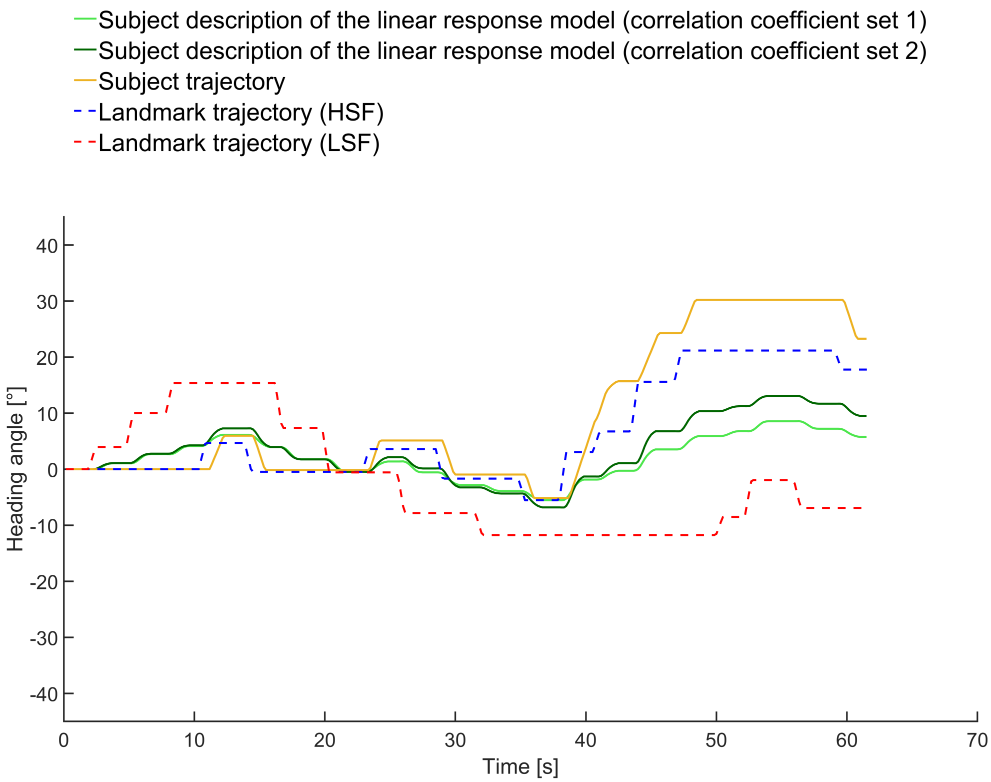
Moreover, the linear response model's performance increased with the subjects' response strength (r(534)=0.0973, p= 0.024).

Percentage of all trials of one condition (%)								
↓ (HSF/LSF)	M_HSF_1	M_LSF_1	F_HSF_1	F_LSF_1	M_HSF_2	M_LSF_2	F_HSF_2	F_LSF_2
0.05/0.95	57.14	14.28	85.71	85.71	57.14	42.86	85.71	85.71
0.10/0.90	91.67	100.00	91.67	100.00	91.67	75.00	91.678	100.00
0.25/0.75	100.00	100.00	58.33	75.00	91.67	100.00	66.67	66.67
0.50/0.50	66.67	100.00	83.33	100.00	50.00	100.00	83.33	100.00
0.75/0.25	83.33	91.67	100.00	100.00	100.00	100.00	91.67	91.67
0.10/0.90	91.67	41.67	91.67	41.67	100.00	100.00	83.33	58.33

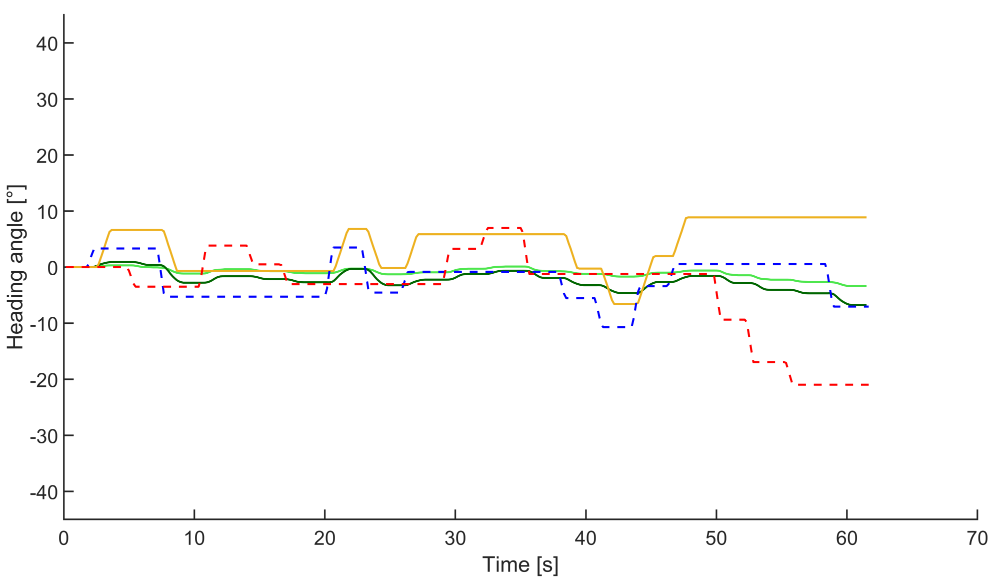
Table 3: **Percentage of trials, of a data set different from the used model coefficient's source, in which the subjects' behavior could be described using the linear response model.** The total number of trials in the 0.05/0.95 /HSF/LSF) contrast conditions were 12 for HSF and LSF conditions each. The total number of trials for all other conditions were 24 for each condition. The first column states the contrast. The first row states the used scenery (M/F), the target spatial frequency (HSF/LSF) and which model coefficient set was used (model coefficient set 1/2). It was assumed, that the linear response model could describe subject behavior in a condition, if the linear response model's estimate was better than the null-hypothesis (linear response model estimates that the subject makes no turns) in at least 50% of all trials of the condition. Except in the 0.05/0.95 (HSF/LSF) condition when subjects were to follow LSFs, the linear response model was able to describe subject behavior, originating from a data set different to the used model coefficient's source data set, in all conditions, with at least one model coefficient set

Percentage of all trials of one condition (%)								
↓ (HSF/LSF)	M_HSF_1	M_LSF_1	F_HSF_1	F_LSF_1	M_HSF_2	M_LSF_2	F_HSF_2	F_LSF_2
0.05/0.95	71.43	57.14	71.43	71.43	85.71	42.86	85.71	14.28
0.10/0.90	91.67	58.33	91.67	83.33	91.67	100.00	100.008	83.33
0.25/0.75	91.67	100.00	66.67	83.33	100.00	100.00	41.67	100.00
0.50/0.50	50.00	100.00	83.33	100.00	83.33	100.00	83.33	100.00
0.75/0.25	100.00	100.00	91.67	100.00	83.33	91.67	100.00	100.00
0.10/0.90	100.00	100.00	75.00	50.00	91.67	75.00	83.33	50.00

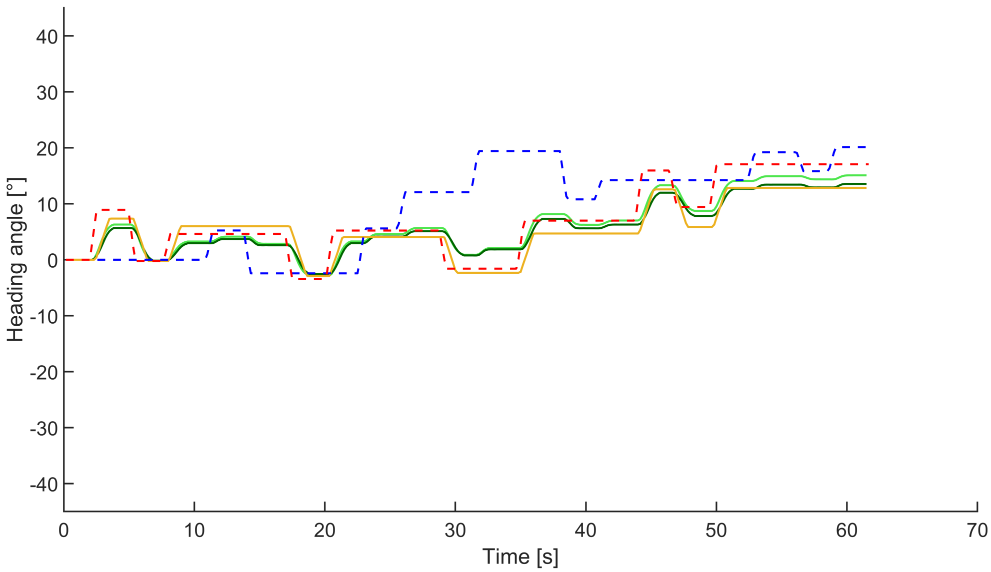
Table 4: **Percentage of trials of the source dataset, in which the subjects' behavior could be described using the linear response model.** The total number of trials in the 0.05/0.95 /HSF/LSF) contrast conditions were 12 for HSF and LSF conditions each. The total number of trials for all other conditions were 24 for each condition. The first column states the contrast. The first row states the used scenery (M/F), the target spatial frequency (HSF/LSF) and which model coefficient set was used (model coefficient set 1/2). It was assumed, that the linear response model could describe subject behavior in a condition, if the linear response model's estimate was better than the null-hypothesis (linear response model estimates that the subject makes no turns) in at least 50% of all trials of the condition. The linear response model was able to describe subject behavior, originating from the model coefficient's source data set, in all conditions with at least one data set.



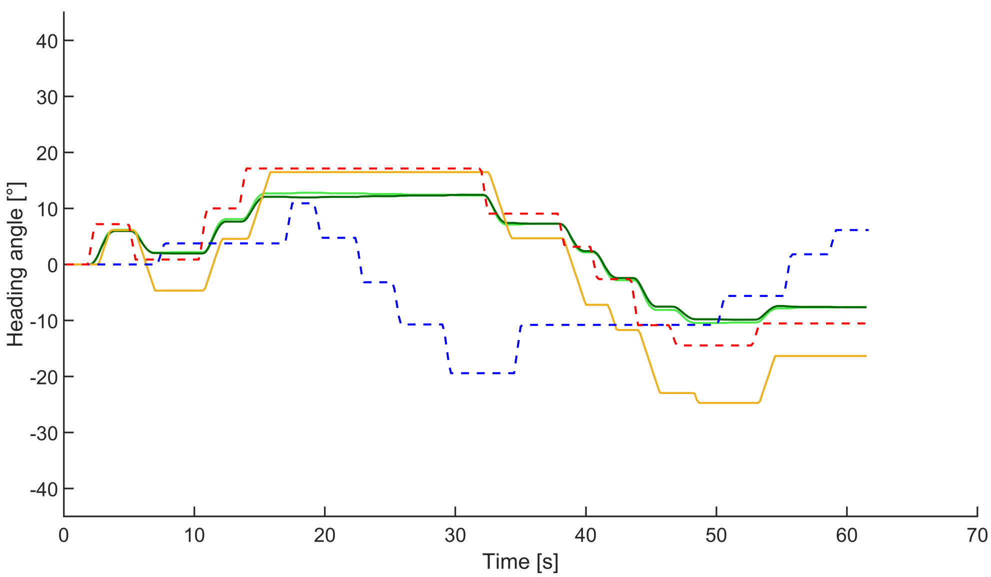
(a) Monastery, contrast 0.05/0.95 (HSF/LSF), HSFs as target spatial frequency



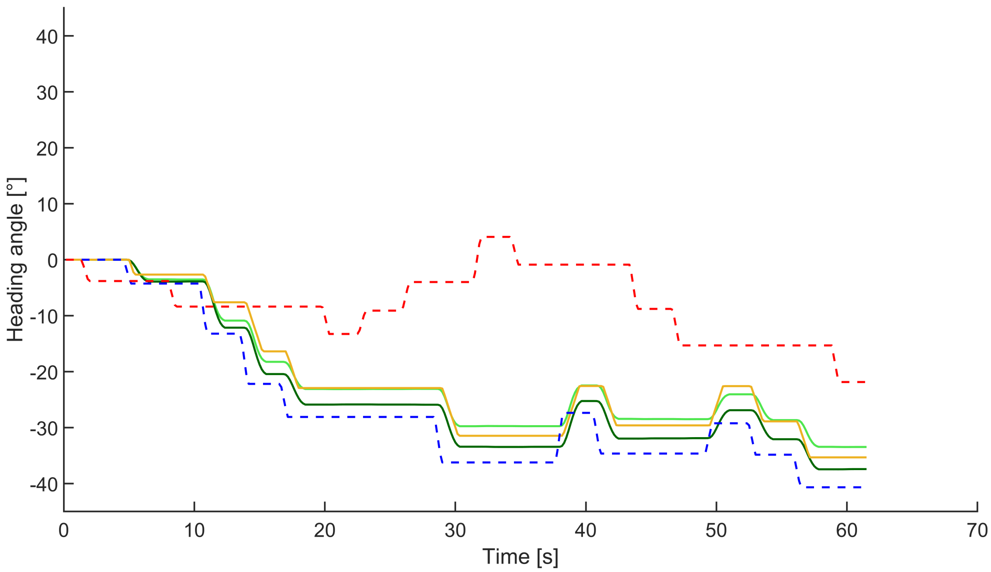
(b) Forest, contrast 0.05/0.95 (HSF/LSF), HSFs as target spatial frequency



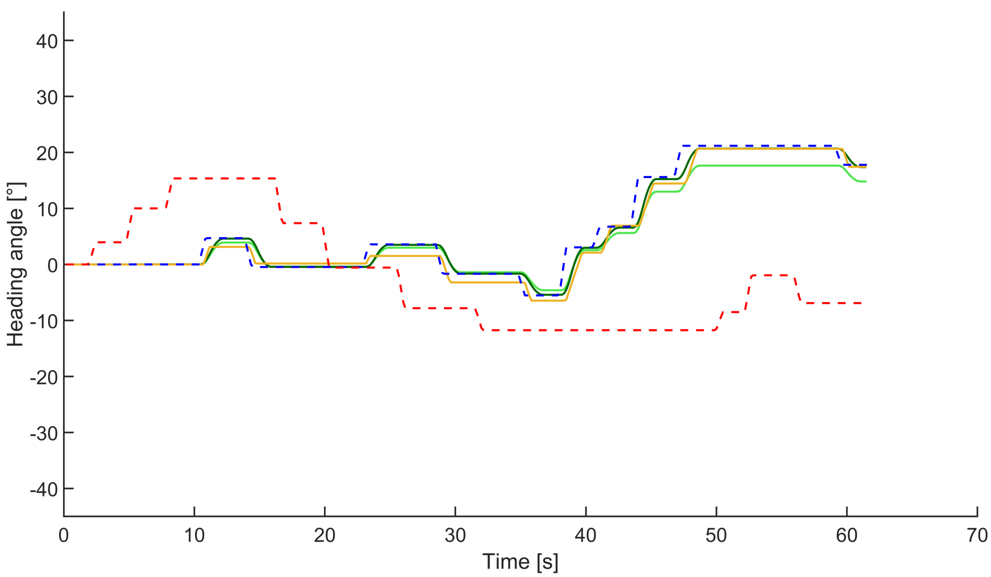
(c) Monastery, contrast 0.05/0.95 (HSF/LSF), LSFs as target spatial frequency



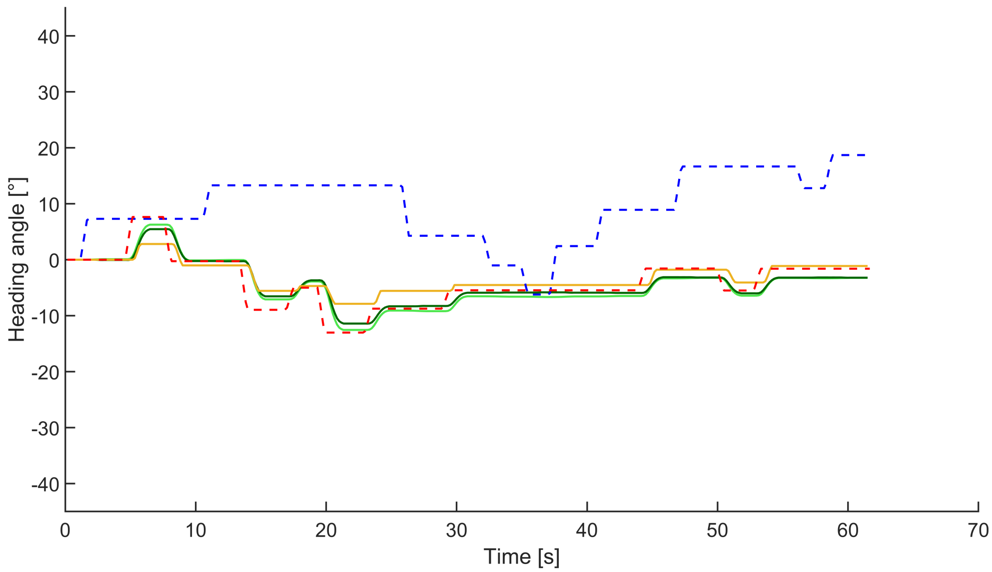
(d) Forest, contrast 0.05/0.95 (HSF/LSF), LSFs as target spatial frequency



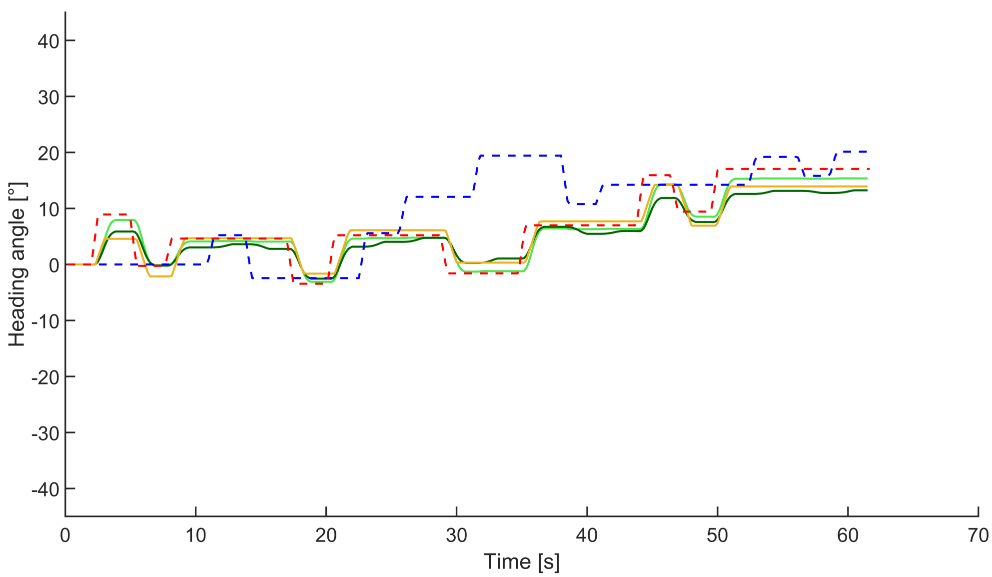
(e) Monastery, contrast 0.25/0.75 (HSF/LSF), HSFs as target spatial frequency



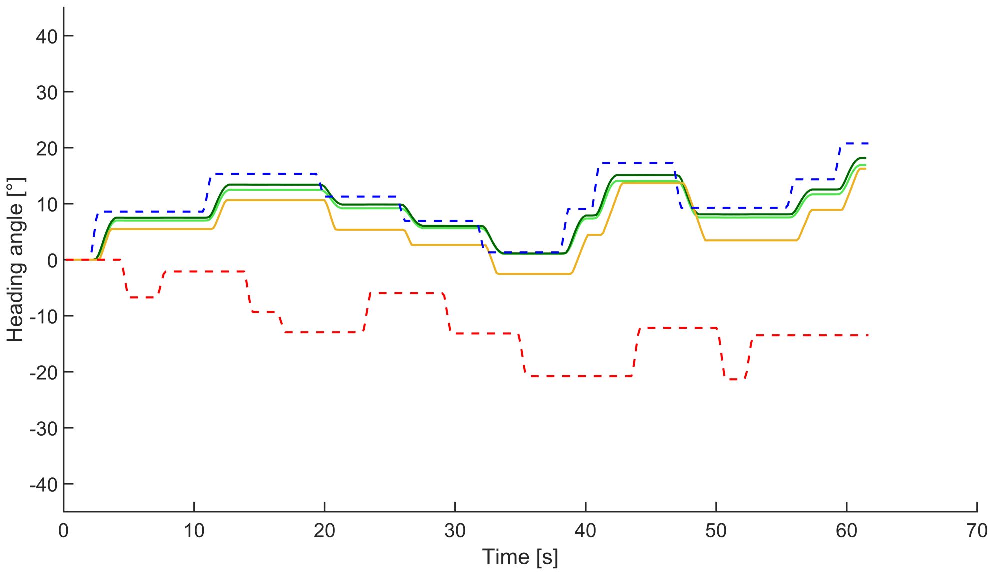
(f) Forest, contrast 0.25/0.75 (HSF/LSF), HSFs as target spatial frequency



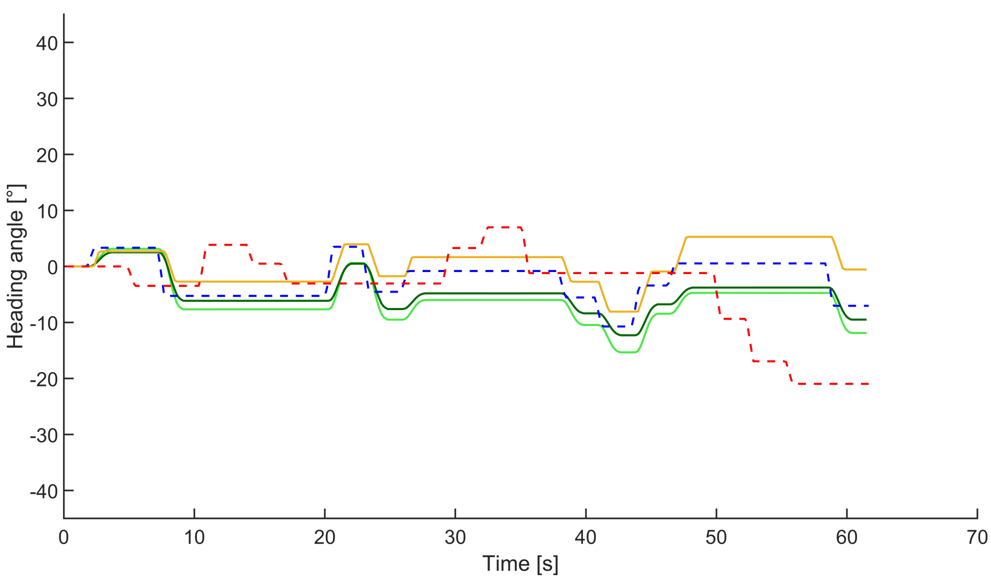
(g) Monastery, contrast 0.25/0.75 (HSF/LSF), LSFs as target spatial frequency



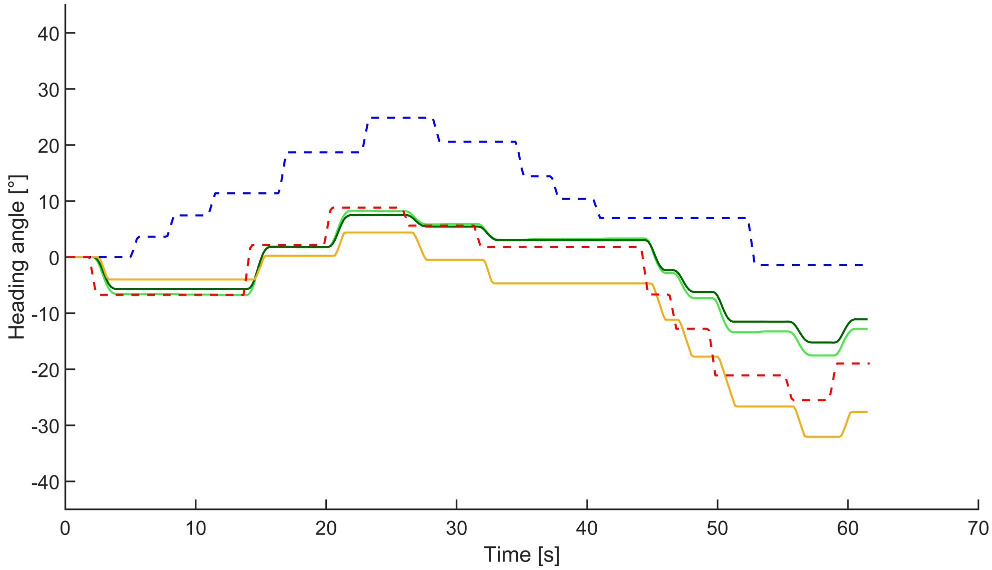
(h) Forest, contrast 0.25/0.75 (HSF/LSF), LSFs as target spatial frequency



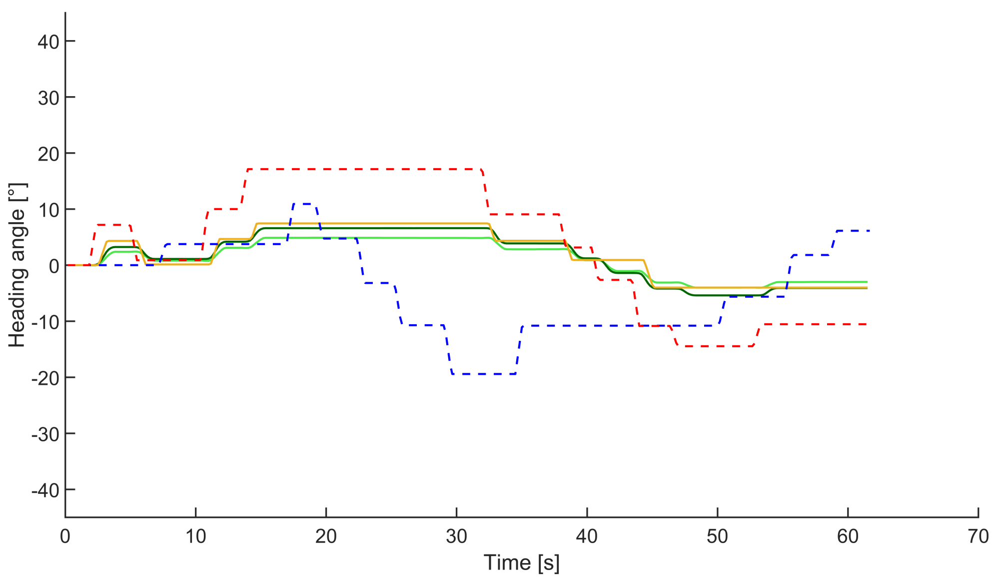
(i) Monastery, contrast 0.90/0.10 (HSF/LSF), HSFs as target spatial frequency



(j) Forest, contrast 0.90/0.10 (HSF/LSF), HSFs as target spatial frequency

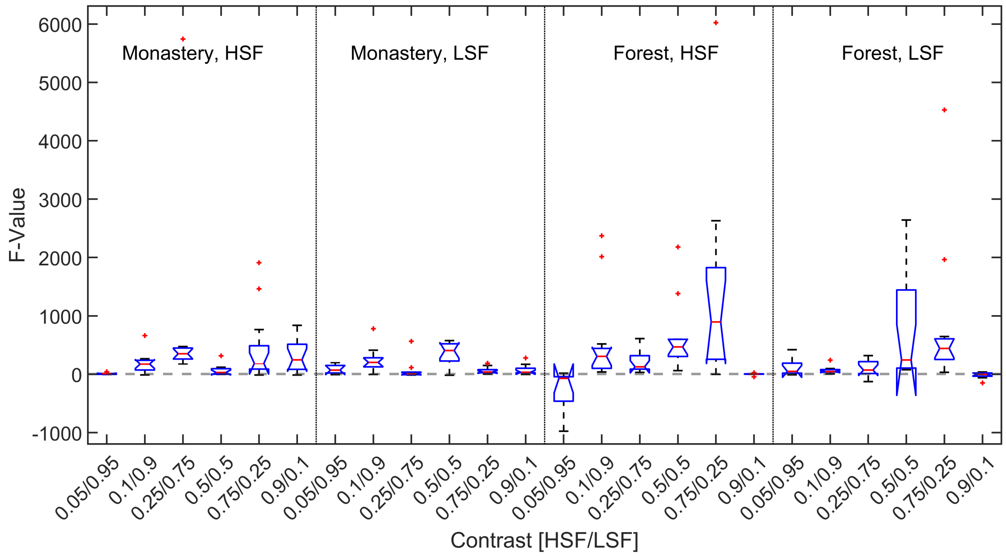


(k) Monastery, contrast 0.90/0.10 (HSF/LSF), LSFs as target spatial frequency

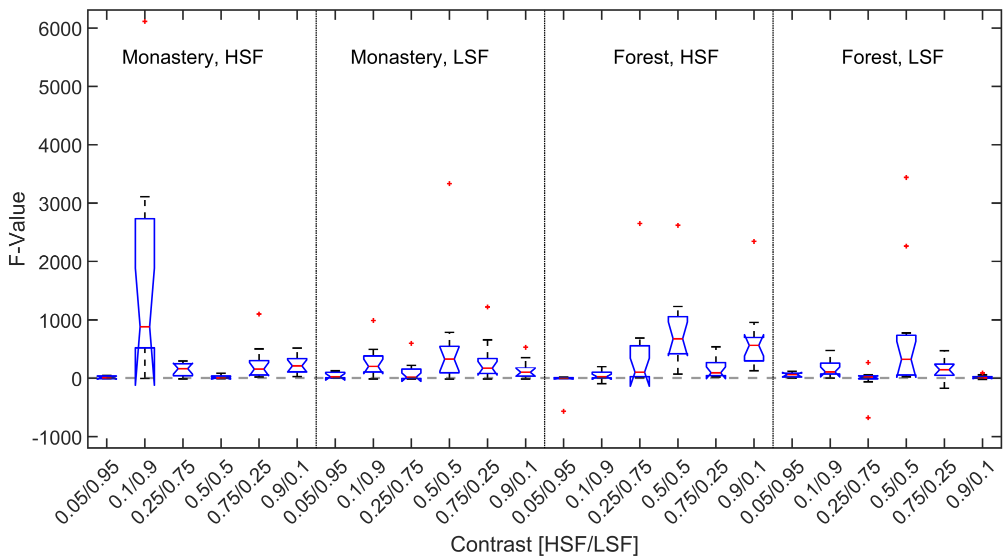


(l) Forest, contrast 0.90/0.10 (HSF/LSF), LSFs as target spatial frequency

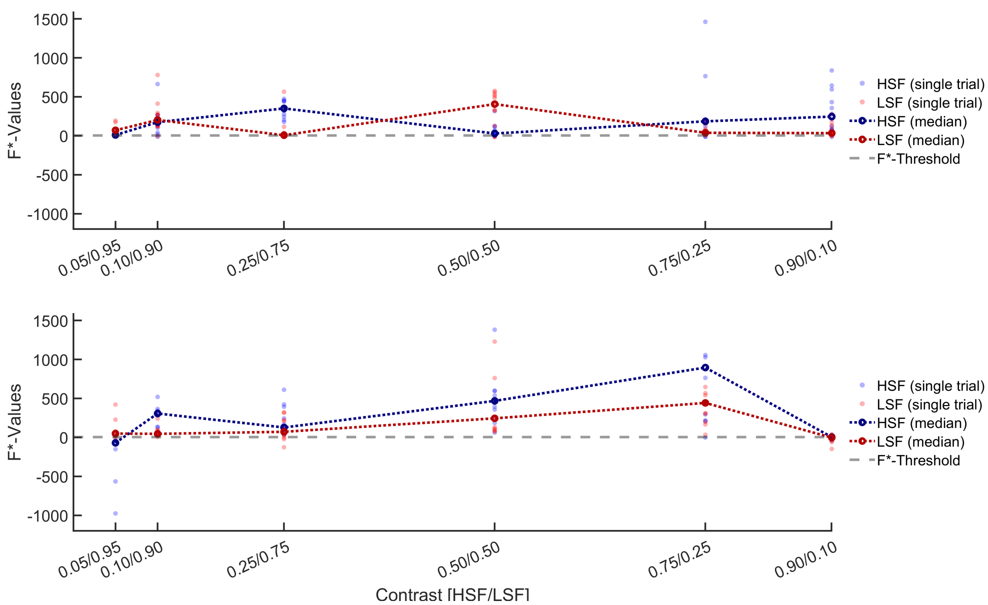
Figure 11: **Comparison of landmark angle, subject heading angle and the linear response model's estimate of the subject heading angle based on either model coefficient set.** The linear response model was not able to describe subject behavior at a contrast of 0.05/0.95 (HSF/LSF) when following HSFs in the Monastery scenery. For all other conditions at least one model coefficient set allowed the linear response model to describe the behavior of most subjects. Using either model coefficient set,the linear response model was able to describe the subjects' behavior equally well.



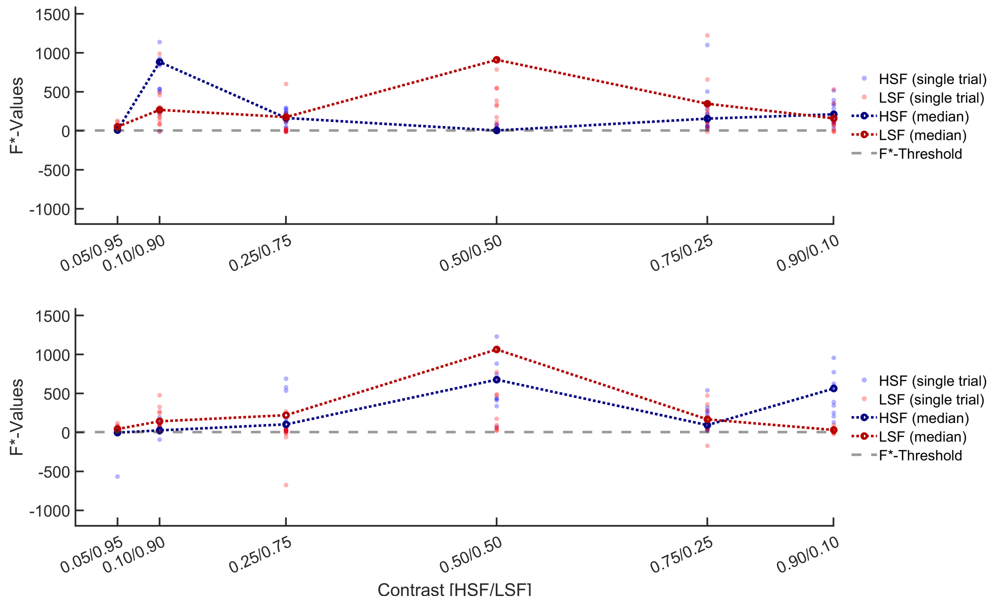
(a) model coefficient set 1



(b) model coefficient set 2



(c) model coefficient set 1, Monastery (upper) and Forest (lower)



(d) model coefficient set 2, Monastery (upper) and Forest (lower)

Figure 12: **Linear response model performance.** The grey line denotes the significance threshold, F*-Values larger than 3.85 were considered significant. Note that y-axes differ between **(a & b)** and **(c & d)**, further extreme outliers are not shown in textbf(c & d) in favour of clarity. **(a & b)** The linear response model's performance sorted by conditions. The first character indicates the scenery (Monastery or Forest), the second characters indicate whether HSF or LSF were followed during the trial. The last characters represent the contrasts of HSF and LSF elements. Red bands indicate the median; notches indicate the median's 95% confidence interval. The 25 and 75 percentile are represented as the boxes' edges. Whiskers indicate the data points farthest from the median which are not considered outliers, outliers are shown as red crosses. **(c & d)** The linear response model's performance derived using all data and the linear response model's performance derived from single trials in the Monastery (upper) and Forest (lower) scenery. The linear response model could describe subjects following HSFs better than subjects following LSFs. The linear response model performed equally well for each scenery regardless of the model coefficient set used.

3.6 Used Landmarks

At the end of the experiments subjects were asked, which landmarks they had used. The monastery's spire was reported to have been used as a landmark by 75% of all subjects when following HSFs. 83.33% of all subjects reported to have used the spire when following LSFs. When following HSFs in the Forest scenery 75% of the subjects reported they had used a tree as a landmark, though the described tree differed between subjects, 50% reported to have used trees when following LSFs and 33.33% reported using blobs such as a patch of ground.

4 Discussion

The aim of this study was to investigate the influence of spatial frequencies in landmark based guiding, depending on the landmark's contrast. We had assessed subjects performance by comparing their trajectory to the positional variance of the target landmark.

4.1 Subject Behavior

4.1.1 Subject Strategy based on scenery

Subjects were able to follow the target landmark in all contrast conditions, except in the Forest scenery when following HSFs at a 0.05/0.95 (HSF/LSF) contrast.
There was no significant difference in subject performance between trials in the Monastery and Forest scenery. However, the used scenery affected the effect of contrast on subject performance. Further, subjects were able to follow HSFs in the Monastery scenery better, while there was no effect of spatial frequency on subject performance in the Forest scenery. Contrary to the Forest scenery, the Monastery scenery provided well distinguished landmarks, especially the monastery spire, which most subjects reported to have followed. Therefore, subjects may have focused on single distinct objects or features, as saccades favour high spatial frequency edges (Krieger et al., 2000; Baddeley & Tatler, 2006). Subject response and performance were higher for HSFs and both response strength and performance are correlated. It is thus possible, that the poorer subject performance when following LSFs was mainly caused by the subject's responding too little to LSFs in the Monastery scenery. One possibility why subjects in the Monastery scenery followed HSFs better and had a higher response strength is that subjects used the same strategy when following LSFs as when following HSFs, which may be relying mainly on edges and single objects, as edges are blurred at LSFs. Given the clear layout of the Monastery scenery, spatial relations or motion should not be harder to perceive than in the Forest scenery. Therefore, if most subjects had used a different strategy for HSFs and LSFs, as relying on spatial relations and motion detection, we would have expected a similar performance.
Subjects following LSFs in the Forest scenery performed as well at a contrast of 0.50/0.50 (HSF/LSF) and 0.75/0.25 (HSF/LSF) as when LSFs were almost exclusively used (contrast 0.05/0.95 (HSF/LSF)), and even outperformed the other conditions. This may indicate, that subjects employed a strategy which focuses on the motion resulting from configural changes of the whole scene using both HSF and LSF components. While subjects could easily name their used landmark in the Monastery scenery, subjects had trouble specifying which object they used in the Forest scenery and reported an overall greater difficulty in identifying objects in the Forest scenery. Also, LSFs are better suited for motion detection (Derrington & Lennie, 1984), and it has already been found that LSFs are preferred for a holistic perception (Hughes, Fendrich, & Reuter-Lorenz, 1990). Additionally, they may have assessed object identity by using the scenery's configuration, rather than by using object features, as spatial relations inferred from LSFs can be used to estimate the nature of scenes and objects (Schyns & Oliva, 1994; Lipson, Grimson, & Sinha, 1997; Oliva & Torralba, 2001; Goffaux et al., 2005). It is however notable, that while subject performance did not differ for HSFs and LSFs in the Forest scenery, subjects had shorter reaction times and greater response strengths to HSFs. If subjects had employed the same strategy when following HSFs and LSFs, a longer reaction time to HSFs would be expected, if guiding had been based on configural change detection. It is therefore possible, that subjects resorted to another strategy when following HSFs. It has been found, that a lack of LSFs can diminish a preference of global structures in favour of local features (Hughes et al., 1990). Thus it is possible, that like in the Monastery scenery, subjects focused on single objects or edges when following HSFs in the Forest scenery. It is thus possible that subjects' strategy differed in the Forest condition depending on target spatial frequency.

4.1.2 Influence of Landmark Contrast on Subject Behavior

It is of interest, that subjects performed as well or better when both HSFs and LSFs were perceivable compared to when the target spatial frequency was almost exclusively used. A reason may be that the

non-target spatial frequency components functioned as a background to use as a reference (Montagne & Laurent, 1994; Lenoir et al., 1999), as was mentioned when discussing subject strategy for following LSFs in the Forest scenery. Further findings in the middle temporal visual area of owl monkeys suggests that, speed difference in target and background may also enhance perception of a moving object (Allman, Miezin, & McGuinness, 1985). However, the effect of background on motion tasks is still controversial (van der Kamp, Savelsbergh, & Smeets, 1997).
Given that subject performance did not continuously increase with the target spatial frequencies' contrast, we conclude, that there is no such contrast turning point, at which subjects would be able to follow a given spatial frequency better. A detection threshold, with HSFs and LSFs possibly showing different thresholds, may however be investigated using a finer scaled contrast ratio. Also, there is a preferred HSF/LSF ratio for both sceneries, which differs between sceneries. It might thus be of interest, to test how scenery composition influences the preferred HSF/LSF contrast ratio.

Though there was a positive correlation of subject response strength and performance, the correlation was weak. It is therefore possible, that the correlation was mainly caused by trials in which subjects gave only little input, which would result both in a low response strength and performance.

4.1.3 Reaction Time

We had acquired subject's reaction times by determining the lag of the maximum correlation between subject's and followed landmark trajectory. Subjects' reaction times following either HSFs or LSFs were compared to test, whether a possible coarse to fine bias would result in lower reaction times when following LSF filtered landmarks. It is notable, that contrast only had a significant effect on subject reaction times when looking at data from both sceneries. There were no significant differences in reaction time lengths for different contrast within one scenery. This may indicate, that there may be an effect of contrast on reaction time, if looking at a larger dataset. However, the effect may be small, and may simply an effect of landmark visibility.

Contrary to our expectations, that reactions times would be faster for subjects following LSFs compared to subjects following HSFs, overall reaction times to LSFs were significantly longer. Though reaction times between subjects following HSFs and LSFs were not significantly different in the Monastery condition, reaction times to LSFs were even longer than to HSFs in the Forest scenery. A cause for the higher reaction times to LSFs may be, that it is easier to detect movement of the fine edged HSFs than of the coarse blurry edged LSFs. A shorter reaction time may also have originated due to a preference of saccades on high spatial frequency edges (Baddeley & Tatler, 2006).

4.2 Linear Response Model Performance

Overall, the linear response model was able to describe the subject behavior in all conditions, except when subjects were to follow LSFs in the Monastery scenery at a 0.05/0.95 (HSF/LSF) . Though the subjects did not perform better than chance, the linear response model was able to describe subject behavior at the 0.05/0.95 (HSF/LSF) contrast conditions when following HSFs in the Forest scenery. This may be explained by an overall low activity of the subject during these trials, which is reflected both in the respective raw subject heading data and the model coefficients.
In the 0.05/0.95 (HSF/LSF) contrast condition when HSFs were followed in the Monastery scenery subjects performed significantly better than chance, albeit the linear response model was not able to describe the subject movements . The data suggests that some subjects may have confused HSFs for LSFs in the 0.05/0.95 (HSFLSF) Monastery condition, as some subjects' performance was better if their trajectory was compared to the not to be followed landmark instead. As subjects in any condition always had to follow HSFs first, the lack of perceivable HSF elements may have prompted some subjects to follow the distinctly visible LSF elements instead. Thus the reduced subject performance in this condition would not only stem from a difficulty in HSF detection but also from an insufficient knowledge of spatial frequencies. However, subjects were not tested on which spatial frequency they regarded as HSFs or LSFs in the 0.05/0.95 (HSF/LSF) condition specifically, moreover the data also suggests

that subjects correctly followed the target landmarks at other contrast conditions. The apparent contradiction may also be explained by the subjects correcting their movements, either by moving right after their first motion to a landmark change or by using their bearing towards the landmark. The first possibility would be hard to model based on landmark movements as subjects would likely not correct all movements, whereas a corrective motion of the subjects by using their bearing does not have to be tied to landmark movement at all.
It is of notice, that despite overall equal performance of subjects when following HSFs and LSFs, the linear response model was able to describe subject behavior when following HSFs better. One possibility would be, that subjects were more variable in their used strategies when following LSFs, with some subjects relying on object features, while others rely on configural changes. Subjects switching between strategies is also possible. Further, in the Forest scenery subjects' response strength was weakest when they were to follow LSFs, yet their performance in following LSFs and HSFs in the Forest scenery was equal. A possibility why the linear response model could describe subject behavior less well when following LSFs in the Forest scenery may therefore be a higher heterogeneity in subject behavior. It is possible, that at least some subjects, due to their difficulty in identifying landmarks, were conservative in their input. If they had only responded to landmark angle change, when they were certain, they may still have a good performance. Yet their overall response would be low, and the linear response model would be able to describe their behavior less, as not all landmarks would be followed. The linear response model describing HSFs better does thus reinforce, that subjects may, in the Forest scenery at least, have used different strategies when following either HSFs or LSFs.

4.3 Experiment Design

Our original experiment design, was similar to the presented experiment, with one major change: Subjects were to be unaware of the HSF and LSF components moving independently. Instead subjects would not told of HSFs and LSFs at all, and HSF and LSF components would be presented as one singly moving scenery. Instead of having to follow either HSFs or LSFs, subjects would be tasked to keep their bearing towards the scenery center stable. Additionally we would superimpose a regular small shift on the subjects movements, which would be cancelled out via a baseline in the analysis. The small shifts would serve to mask that HSF and LSF components were moving independently. Our goal would have been to compare the subjects' ability to follow either HSFs or LSFs, dependent on contrast and scenery. Therefore we would have compared the subjects' deviation from either spatial filtered landmark's movement and compared the subjects' response strength to HSFs or LSFs. The advantage of concealing HSFs and LSFs components moving independently would have been, that subjects would be unbiased towards either spatial frequency, instead of concentrating on one of them. It has been found, that attention can influence the perception. As subjects were explicitly tasked to follow either HSF or LSF, they are more likely to have a narrower focus, concentrating more on fewer select image features. A narrower attention may have particularly enhanced HSF perception (Balz & Hock, 1997). Further, attention can be divided into transient (attention caused by a sudden stimulation) or sustained attention (conscious focus). It has been found, that transient attention can enhance perception of HSFs (Yeshurun & Carrasco, 1999; Gobell & Carrasco, 2005). The task of following either HSFs or LSFs may have benefited sustained attention.

However, using covert independent HSF and LSF movements carries several difficulties. Our experiment was roughly one hour long. It is therefore likely, that several subjects would have noticed, that HSFs and LSFs were moving independently. Thus, the experiment would result in data sets either differing in the subjects' behavior, or in their conditions, if the experiments were aborted. Further, excluding data of subjects, that noticed a difference in HSF and LSF, may cause a selection for orientation strategies. For instance, a subject using an object oriented approach may be less prone to noticing than a subject using a strategy based on changes on a configural level. Moreover, in order for subjects not to notice HSF and LSF image components moving independently, the landmark movements would have to be small. This would have decreased the visibility of landmark movements as a whole. In contrast, our current experimental design allowed landmark movements to be chosen freely.

Thus, especially considering that the overall experimental design is in itself new, we have chosen our current experiment design over our original idea. The concept of covert independent HSF and LSF movement may however be incorporated into future experiments.

4.4 Conclusion

In conclusion our findings suggest, that subjects use both high and low spatial frequencies in landmark based guiding. Considering subjects' better performance when following HSFs in the Monastery scenery and shorter reaction time, HSFs may play a greater role in landmark based guiding than in scene recognition tasks. Further, subjects may resort to different orientation strategies depending on the landmarks and their surroundings. Therefore spatial frequency preference may change depending on the scenery's features. It has already been found that preference of spatial frequencies differs depending on the task (Morrison & Schyns, 2001; Goffaux et al., 2005). A comparison of spatial frequency usage between landmarks most distinguishable through their spatial relations and landmarks which are mainly distinguishable by their features may therefore be of interest.

References

Allman, J., Miezin, F., & McGuinness, E. (1985). Direction-and velocity-specific responses from beyond the classical receptive field in the middle temporal visual area (mt). *Perception*, *14*(2), 105–126.

Baddeley, R. J., & Tatler, B. W. (2006). High frequency edges (but not contrast) predict where we fixate: A bayesian system identification analysis. *Vision research*, *46*(18), 2824–2833.

Balz, G. W., & Hock, H. S. (1997). The effect of attentional spread on spatial resolution. *Vision research*, *37*(11), 1499–1510.

Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience*, *5*(8), 617.

Bassi, C. J., Lehmkuhle, S., et al. (1990). Clinical implications of parallel visual pathways. *Journal of the American Optometric Association*, *61*(2), 98–110.

Breitmeyer, B., Levi, D. M., & Harwerth, R. S. (1981). Flicker masking in spatial vision. *Vision Research*, *21*(9), 1377–1385.

Cartwright, B., & Collett, T. S. (1983). Landmark learning in bees. *Journal of comparative physiology*, *151*(4), 521–543.

Collett, T. (1995). Making learning easy: the acquisition of visual information during the orientation flights of social wasps. *Journal of Comparative Physiology A*, *177*(6), 737–747.

Davenport, J. L., & Potter, M. C. (2004). Scene consistency in object and background perception. *Psychological Science*, *15*(8), 559–564.

Derrington, A., & Lennie, P. (1984). Spatial and temporal contrast sensitivities of neurones in lateral geniculate nucleus of macaque. *The Journal of physiology*, *357*(1), 219–240.

Epstein, R. A., & Vass, L. K. (2014). Neural systems for landmark-based wayfinding in humans. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1635), 20120533.

Fajen, B. R., & Warren, W. H. (2007). Behavioral dynamics of intercepting a moving target. *Experimental Brain Research*, *180*(2), 303–319.

Foo, P., Warren, W. H., Duchon, A., & Tarr, M. J. (2005). Do humans integrate routes into a cognitive map? map-versus landmark-based navigation of novel shortcuts. *Learning, Memory*, *31*(2), 195–215.

Gillner, S., Weiß, A. M., & Mallot, H. A. (2008). Visual homing in the absence of feature-based landmark information. *Cognition*, *109*(1), 105–122.

Gobell, J., & Carrasco, M. (2005). Attention alters the appearance of spatial frequency and gap size. *Psychological science*, *16*(8), 644–651.

Goffaux, V., Hault, B., Michel, C., Vuong, Q. C., & Rossion, B. (2005). The respective role of low and high spatial frequencies in supporting configural and featural processing of faces. *Perception*, *34*(1), 77–86.

Hermer, L., & Spelke, E. S. (1994). A geometric process for spatial reorientation in young children. *Nature*, *370*(6484), 57.

Hughes, H., Fendrich, R., & Reuter-Lorenz, P. (1990). Global versus local processing in the absence of low spatial frequencies. *Journal of Cognitive Neuroscience*, *2*(3), 272–282.

Krieger, G., Rentschler, I., Hauske, G., Schill, K., & Zetzsche, C. (2000). Object and scene analysis by saccadic eye-movements: an investigation with higher-order statistics. *Spatial vision*, *13*(2-3), 201–214.

Learmonth, A. E., Newcombe, N. S., & Huttenlocher, J. (2001). Toddlers' use of metric information and landmarks to reorient. *Journal of experimental child psychology*, *80*(3), 225–244.

Lenoir, M., Savelsbergh, G. J., Musch, E., Thiery, E., Uyttenhove, J., & Janssens, M. (1999). Intercepting moving objects during self-motion: Effects of environmental changes. *Research Quarterly for Exercise and Sport*, *70*(4), 349–360.

Lipson, P., Grimson, E., & Sinha, P. (1997). Configuration based scene classification and image indexing. In *Proceedings of ieee computer society conference on computer vision and pattern recognition* (pp. 1007–1013).

Lupp, U., Hauske, G., & Wolf, W. (1976). Perceptual latencies to sinusoidal gratings. *Vision research*, *16*(9), 969–972.

Montagne, G., & Laurent, M. (1994). The effects of environmental changes on one-handed catching. *Journal of motor behavior*, *26*(3), 237–246.

Morrison, D. J., & Schyns, P. G. (2001). Usage of spatial scales for the categorization of faces, objects, and scenes. *Psychonomic Bulletin & Review*, *8*(3), 454–469.

Nouhuys, S., & Kaartinen, R. (2007). A parasitoid wasp uses landmarks while monitoring potential resources. *Proceedings of the Royal Society B: Biological Sciences*, *275*(1633), 377–385.

Oliva, A., & Schyns, P. G. (1997). Coarse blobs or fine edges? evidence that information diagnosticity changes the perception of complex visual stimuli. *Cognitive psychology*, *34*(1), 72–107.

Oliva, A., & Schyns, P. G. (2000). Diagnostic colors mediate scene recognition. *Cognitive psychology*, *41*(2), 176-210.

Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International journal of computer vision*, *42*(3), 145–175.

Oliva, A., & Torralba, A. (2006). Building the gist of a scene: The role of global image features in recognition. *Progress in brain research*, *155*, 23–36.

Potter, M. C. (2012). Recognition and memory for briefly presented scenes. *Frontiers in psychology*, *3*, 32.

Pourtois, G., Dan, E. S., Grandjean, D., Sander, D., & Vuilleumier, P. (2005). Enhanced extrastriate visual response to bandpass spatial frequency filtered fearful faces: Time course and topographic evoked-potentials mapping. *Human brain mapping*, *26*(1), 65–79.

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

Siagian, C., & Itti, L. (2007). Rapid biologically-inspired scene classification using features shared with visual attention. *IEEE transactions on pattern analysis and machine intelligence*, *29*(2), 300–312.

Steck, S. D., & Mallot, H. A. (2000). The role of global and local landmarks in virtual environment navigation. *Presence: Teleoperators & Virtual Environments*, *9*(1), 69–83.

Tolhurst, D., Tadmor, Y., & Chao, T. (1992). Amplitude spectra of natural images. *Ophthalmic and Physiological Optics*, *12*(2), 229–232.

Tom, A., & Denis, M. (2004). Language and spatial cognition: Comparing the roles of landmarks and street names in route instructions. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *18*(9), 1213–1230.

van der Kamp, J., Savelsbergh, G., & Smeets, J. (1997). Multiple information sources in interceptive timing. *Human Movement Science*, *16*(6), 787–821.

Vassilev, A., & Mitov, D. (1976). Perception time and spatial frequency. *Vision research*, *16*(1), 89–92.

Wallis, G., Chatziastros, A., & Bülthoff, H. (2002). An unexpected role for visual feedback in vehicle steering control. *Current Biology*, *12*(4), 295–299.

Yeshurun, Y., & Carrasco, M. (1999). Spatial attention improves performance in spatial resolution tasks. *Vision research*, *39*(2), 293–306.