Haptic Saliency Model for Rigid Textured Surfaces

Anna Metzger, Matteo Toscani, Matteo Valsecchi and Knut Drewing

Justus-Liebig University Giessen
anna.metzger@psychol.uni-giessen.de

Abstract. When touching an object, we focus more on some of its parts rather than touching the whole object's surface, i.e. some parts are more salient than others. Here we investigated how different physical properties of rigid, plastic, relieved textures determine haptic exploratory behavior. We produced haptic stimuli whose textures were locally defined by random distributions of four independent features: amplitude, spatial frequency, orientation and isotropy. Participants explored two stimuli one after the other and in order to promote exploration we asked them to judge their similarity. We used a linear regression model to relate the features and their gradients to the exploratory behavior (spatial distribution of touch duration). The model predicts human behavior significantly better than chance, suggesting that exploratory movements are to some extent driven by the low level features we investigated. Remarkably, the contribution of each predictor changed as a function of the spatial scale in which it was defined, showing that haptic exploration preferences are spatially tuned, i.e. specific features are most salient at different spatial scales.

Keywords: haptic salience, attention, haptic exploration

1 Introduction

When humans haptically explore an object, they do not intensively touch all its areas; rather they select some of its parts. If this selection is not completely random, it follows that some parts are more salient than others. Saliency can be defined as the perceptual property of a physical stimulus which makes it stand out from competing stimulation. This concept is tightly related to attention. In fact, attention is commonly described with the metaphor of a searchlight that intensifies the incoming sensation from a selected part of sensorial input (e.g.[1, 2]). Obviously, the selection process that determines which of the information that is entering our perceptual systems plays a central role in sensation. This process reflects bottom-up aspects (i.e. properties of the sensory signals), as well as influences from the internal state of the organism (top-down aspects) [3]. Here we focus on the bottom-up aspects.

Bottom-up saliency had been extensively investigated in the visual domain. Probably the highest impact implementation of a bottom-up saliency model has been formalized by Itti and Koch [4–6], based on the results from visual search experiments (e.g. [2]), and inspired by cortical visual processing. According to their model, the visual input is first linearly filtered at several spatial scales, then color, intensity and orientation maps

are computed separately, resembling the computations carried out by neurons in the early stages of visual processing. These maps are then combined with different weights in order to achieve a single saliency map. Fixations are predicted by a winner-takes-all system applied to the combined map. Despite its limitations (see [7]), Itti and Koch's original work paved the way to the development of a large body of computational models [8–15] proposed to predict gaze allocation.

In analogy to the investigation of visual salience (e.g., [2, 16]), haptic saliency was often measured using search tasks. According to this paradigm, participants are presented with haptic stimuli comprising one target among a different number of distractors. They are asked to detect the target as soon as possible: The quicker the answer, the higher the saliency of the target. Rough stimuli pop out among smooth ones, movable targets pop out among stationary distractors, hard stimuli pop out among soft ones (e.g.,[17–19]). When stimulus parts differ only slightly in their properties (e.g. roughness), the search for a target among distractors takes longer. However, in search studies target and distractors differ usually only in one property and the difference was varied only in two levels (e.g. rough vs smooth). Thus, from these studies it can be indeed concluded which features are salient relative to a fixed class of distractors, but not which features and feature contrasts are salient relative to other features and how this affects exploratory behavior. Here we aim to model haptic exploratory behavior of complex stimuli.

We produced texture stimuli the surfaces of which were defined by haptic grating elements that randomly varied in certain features such as amplitude, isotropy, frequency or orientation. We assumed that the duration of touching certain locations at a stimulus ("touch duration") is directly linked to its saliency. In fact, fixation behavior is related to visual saliency (e.g. [8]). Fixations are defined as maintained gaze at a certain position, as opposed to transitional eye movements such as saccades or smooth pursuit. Alternation between moving and static phases has been reported also during haptic exploration [20]. Although the static phases could be functionally similar to fixations, their role is not yet established. Hence, we decided to keep duration information as dependent variable, instead of segmenting recorded movements into moving and static phases and focusing on the latter ones. Touched position was recorded while participants explored couples of stimuli with the whole hand in order to judge their similarity. We recorded only the position of the index finger because we assumed its position to be highly correlated to the one of the other fingers. In fact, for five-finger haptic search on a planar surface it was shown that finger positions were highly correlated, indicating that they were moving as a single unit [21]. Finally, we used a linear regression model to predict the spatial distribution of touch duration as a function of the spatial distribution of selected features and their spatial gradients (as a measure of local contrast). Thus, we could describe the link between features and haptic saliency by feature and feature-contrast specific weights.

2   Experiment

In every trial of the experiment participants explored two stimuli, which were relieved textures defining a square in the horizontal plane (Fig. 1C). In order to engage participants in a perceptual task , thus promoting the exploration of the stimuli, we asked participants to explore the stimuli sequentially and report how similar they were. The choice of a similarity judgment as a task demand was arbitrary and results were not analyzed. From the movement data we computed touch duration for all the different locations. We used a linear regression model to relate touch duration with the physical features which defined the experimental stimuli.

2.1   Methods

Participants. 10 female students (average age 22.6 years) −naïve to the purpose of the experiment–volunteered to participate and were reimbursed for their participation (8€/h). All participants were right handed with no reported sensory or motor impairments of the right hand. Informed consent was obtained from each participant. The study was approved by the local ethics committee LEK FB06 at Giessen University and is in accordance with the declaration of Helsinki (2008).

Apparatus. Participants sat in front of a visuo-haptic workbench, consisting of a PHANToM 1.5A haptic force feedback device (spatial resolution 0.03 mm, used temporal resolution 333 Hz), a 22"-computer screen (120 Hz, 1280x1024 pixel), stereo glasses, a mirror and an exploration table. We used the PHANToM to measure the position of the right index finger. The finger was attached to the PHANToM by a custom made adapter, consisting of a metallic pin with a ball at its end and a plastic fingernail. This fingernail was connected to the participant's fingernail via an adhesive substance and via the magnet to the metallic pin. The adapter left the finger pad uncovered and allowed for free finger movements including six degrees of freedom. Attached to the PHANToM participants were able to move freely in a 38x27x20 cm workspace. The exploration table consisted of two square slots one aside the other. The test stimuli were placed in the left slot and the comparison stimuli in the right slot. In the left slot at the four corners of the stimulus we placed wooden toothpick tips, for the sake of calibration. These points were used to define a projective transformation which mapped the touched position on the horizontal stimulus plane onto arbitrary stimulus coordinates ranging from 0 to 1, assuming the stimulus to be squared. At the end of each trial, observers were asked to touch one randomly chosen calibration point. The average Euclidian distance between the touched position of each calibration at the beginning of the experiment and the touched positions of that point at the end of every trial corresponded to 3.5 +/- 2.3 mm, 2.5% of the stimulus length.

To guide the participants through the experiment and control the available visual information a virtual, schematic 3D-representation of the finger and the stimuli was occasionally displayed. To indicate the position of the stimuli the calibration points (red spheres of 8mm diameter) were displayed. The finger was visualized throughout the experiment as a green sphere of 8 mm diameter. Participants looked at the virtual

representation from 40 cm viewing distance (fixated by a chin rest) via stereo glasses and via a mirror. The mirror, prevented participants from seeing their hand and the stimuli. Similarity adjustments were done using a virtual slider (displayed above the stimuli), rendered as a horizontal yellow bar with "same" and "different" marking its ends. When participants moved the hand towards the right end the bar turned green, indicating the level of similarity. White noise presented via headphones masked sounds.
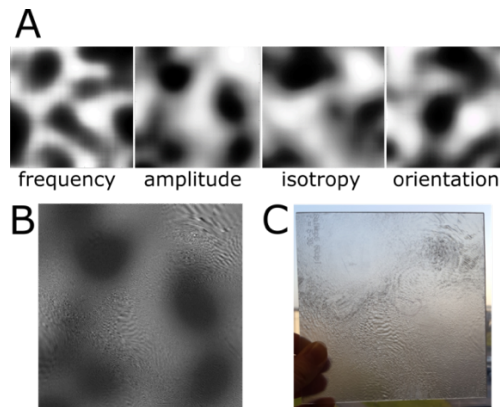

Fig. 1. Experimental stimuli. **A)** Pink noise two-dimensional distributions corresponding to the features maps. **B)** Combined stimulus texture. **C)** 3D Printed version of B).

Stimuli. We printed haptic stimuli (13.97x13.97x0.3 cm; Fig. 1C) using a 3D printer (Object30Pro, Stratasys, material VeroClear, nominal resolution 600 to 1600 dpi). The stimuli were generated as 2D-images (Fig. 1B), and then translated into printable 3D models using the OpenSCAD *surface()* function. The dimensions of the stimuli were chosen in a way that they could be covered by the palm of the average hand with slightly spread fingers. The upper surface of the stimuli was defined to spatially vary in the following features: amplitude (vertical depth), spatial frequency, orientation and isotropy. The spatial distribution of each feature was defined by a greyscale feature map in which corresponding feature values were color coded (Fig. 1A). For instance in the amplitude map, white color referred to high amplitude and in the frequency map white color indicated high spatial frequency. The feature maps were generated as two-dimensional white noise distributions with the same size as the stimulus. This distribution was then low pass filtered so that only the spatial frequencies lower than the average size of a fingertip (1.27cm) remained. Then we scaled them to a fixed range [0 1]. By this, we created feature distributions in which the feature values appear in blobs with a minimal size of average size of a fingertip, so that changes can be detected by moving the index finger. To combine the features in one stimulus (Fig. 1B), we filtered a new white noise [0 1] base texture with a filter defined by the feature values in each position of the feature maps, and used the values of the filtered texture at the given position to define the combined stimulus in that same position. Thus, the filter was oriented at a given spatial frequency and with a given level of isotropy. Filters were defined in the frequency domain by the following equation:

$$F_i\left(\varphi_x, \varphi_y\right) = e^{\left(-\frac{\left(\sqrt{\varphi_x{}^2+\varphi_y{}^2}-\varphi_{xyi}\right)^2}{\sigma_f^2}\right)} e^{\left(-\sigma_{ai}\cos(2(\alpha_{xy}-\alpha_i))\right)} \qquad (1)$$

F_i being the filter correspondent to the pixel i, defined as a function of the two dimensional Fourier frequency space (φ_x, φ_y); φ_{xyi} the frequency corresponding to the frequency band prescribed by the frequency map (ranging from 5 to 30 cycles per image) at the pixel i; σ_f^2 the width of the band pass filter (fixed to 0.01 cycles per image); σ_{ai} the width of a circular Gaussian function, corresponding to the degree of isotropy from the isotropy map (ranging from 0° to 103°) at pixel i (with $\sigma_{ai} = 0°$ corresponding to a completely isotropic filter); α_{xy}, the angle in the polar representation of φ_x, φ_y, and α_i the orientation of the filter, as prescribed by the orientation map (ranging from 0° to 90°) at pixel i. Intuitively, to produce the i pixel of the combined map, the first Gaussian term in equation (1) defines a Gaussian ring around the frequency given by the i pixel of the frequency map; the second circular Gaussian term defines a range of orientations: The bigger the range, the more isotropic the filter. The multiplication of the two Gaussians gives an oriented band pass filter. Amplitude was imposed afterwards, by multiplying the combined map with the amplitude feature map (ranging from 0.1 to 0.3 cm). Fig. 2 shows examples of individual feature differences, applied uniformly in space on a random noise texture. Ten different maps were produced by using ten different seeds (1-10) of pseudorandom number generator of MATLAB (The MathWorks, Inc., Natick, MA).
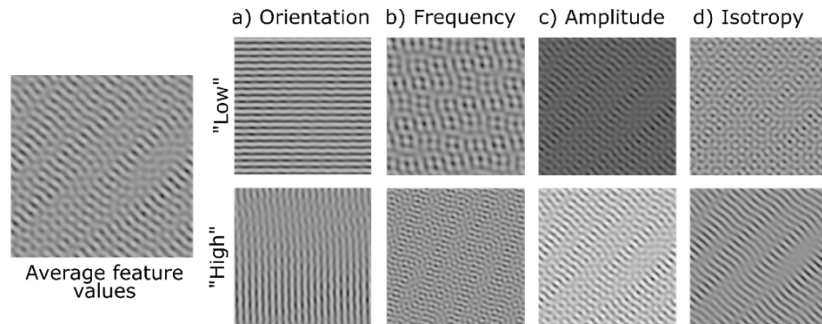

Fig. 2. Examples of individual features differences. Top: "low" feature values, or close to horizontal for orientation; bottom: "high" feature values (or close to vertical for orientation). From left to right: Change of orientation ("high" or "low" α_i), spatial frequency ("high" or "low" φ_{xyi}), amplitude, as imposed by multiplicative scaling after filtering, isotropy ("high" or "low" σ_{ai}). For the example of each feature (a-d), "High" and "Low" values were set close to the extremes of the range for that feature, whereas the values for the other features were set at average (as shown in the "Average feature values" example).

Procedure. At the beginning of the experiment, there was a calibration procedure. Each participant was required to touch four calibration points (at the corners of the

stimulus) with the right index finger. At the calibrated corner a red circle (16 mm diameter) was displayed around the wooden toothpick tip. Participants were instructed with a schematic drawing to position their index finger in a way that the toothpick tip was below the middle of the fingertip. After the calibration participants received two stimuli in every trial. They were free to use the whole right hand for the exploration and were instructed to explore the stimuli with a continuous sweep (without lifting the hand). The test stimulus was always placed left and explored first, to prevent serial effects. At the end of each trial, participants were asked to move the finger to a randomly assigned corner of the left stimulus slot and position it on the toothpick tip like in the calibration, for the purpose of measuring the position error with respect to the initial calibration. Afterwards they indicated how similar they perceived the stimuli. Between the trials participants moved the finger to the waiting position (at the left corner closer to the participant), while the experimenter exchanged the stimuli. The experiment consisted of a single session of four blocks. Each block consisted of ten trials in which participants explored the ten different stimuli presented in random order as a test stimulus. For each trial, the comparison stimulus was randomly chosen among the nine remaining stimuli. Thus, each stimulus was explored as a test stimulus 4 times, resulting in 40 trials in total, which were completed on average within 1.13h.

Analysis. Exploratory behavior was only analyzed for the test stimulus, by computing touch durations for each position of the stimulus. We used a multiple linear regression approach to analyze touch duration as a function of stimulus features. Touch duration for each position of the stimulus was computed as number of samples in which the finger was in that position multiplied by the temporal resolution of the PHANToM (3ms). In order to consider the dimension of the finger, we low-pass filtered the sampled traces in space, so that one assessed position corresponded to an area of the stimulus of approximately the size of the finger pad. The sigma of the Gaussian low-pass filter was chosen to be half of the finger size, estimated as 1.27cm. Afterwards, the touch durations of each individual trial were z-score transformed, so that each trial weighted the same in the regression analyses. The stimulus features were also z-score transformed, so that the regression coefficients were expressed in standard deviation units (β-weights) and thus comparable between each other. We considered the four features: Amplitude, isotropy, frequency and orientation and their gradients, at 30 different spatial scales. Gradients were computed with a Sobel operator [22] at each spatial scale. The spatial scales were obtained by low pass filtering the feature maps and their gradients with Gaussian spatial filters defined by increasing sigma (width). Sigmas were linearly spaced from zero (no filtering) to ~80% of the image size, in order to largely remove the spatial content of the feature maps. These extreme filtering levels are useful to show spatial tuning for the touch duration as a function of each predictor (i.e. feature map); in fact predictors at these spatial scales should be minimally related to the exploratory behavior. We filtered the feature maps and their gradients with the *imfilter()* MATLAB function, set to replicate the edges of the image when the filter kernel area was exceeding the border of an image. We believe that the relationship between orientation and touch duration, and spatial frequency and touch duration is non monotonic, therefore not suited for linear regression. In fact, orientation is a periodic

magnitude, and sensorial responses to both orientation and frequency are described by tuning curves [23, 24]. For instance the high and low spatial frequencies of the perceivable range are unlikely to be salient. Hence, we included only their gradients in the linear regression analyses. Additionally, the use of gradients allowed us to include potentially perceptually relevant information in our analyses, since the perceptual systems can detect second order information which is not present in the Fourier spectrum [25, 26]. Fig. 3A shows the predictors used in the regression analyses, for one of the ten stimuli. Linear regression assumes independent predictors. The four feature maps are generated to be uncorrelated, but they relate to their gradients and are correlated across different spatial scales. Therefore, as a first step we performed a bivariate linear regression for each predictor separately, and selected the spatial scale at which the single predictor was most predictive. Also, we selected the best predictor between amplitude and its gradient, and isotropy and its gradient. We thus could use four independent predictors for a multiple regression analysis. We performed a regression analysis on each participant separately. After fitting a regression model with the main effects of each predictor and all the interaction terms, we tested its β-weights across participants. Finally, we evaluated consistency between participants by predicting touch duration of each participant based on the regression analyses on the other participants' data.

2.2   Results

Spatial tuning. Fig. 3B shows the bivariate regression results for each of the six chosen predictors (feature maps) at each spatial scale. Since regression coefficients are represented by β-weights, they are comparable across predictors. Therefore, higher absolute value at a spatial scale than another means that the predictor at that spatial scale has a higher linear effect than at the other spatial scale. Thus, we can look at the β-weights as a function of the spatial scale as a tuning function. It is clear from the figure that β-weights change for each predictor systematically with the spatial scale. One-way repeated measures ANOVAs confirm this impression for all the predictors but isotropy gradient ($F_{29,261}$ = 9.69, 8.30, 3.18,1.22, 15.00, 26.4240; for orientation gradient, frequency gradient, amplitude gradient, isotropy gradient, amplitude and isotropy, respectively. All p-values < 0.005, except for isotropy gradient, p = 0.207). Features at the most predictive spatial scales were chosen for further analyses (red circles in Fig. 3B). Because of the higher β-weights, amplitude and isotropy were preferred to their gradients.

Multiple regression. Amplitude, isotropy, orientation gradient and frequency gradient were chosen as independent predictors for a multiple linear regression analysis, at the spatial scale which best related to touch duration (red circles in Fig. 3B). A multiple regression with all the interaction terms was performed for each participant separately. Fig. 4A depicts the average β-weights for each regression term. The statistical significance of these β-weights was tested by comparing them with an empirically determined baseline by means of multiple Bonferroni-corrected t-tests. To

compute the baseline we determined a new set of *β*-weights under the null hypothesis that the touch duration is randomly distributed on our stimuli. To do so, we randomized the correspondence between touch duration and the positon on the stimulus, and repeated the multiple regression for every participant. This procedure was repeated 50 times and the resulting *β*-weights were averaged across repetitions. Multiple comparisons revealed a significant effect of amplitude, isotropy and their bivariate interaction (t_9 = 5.99, 6.14, 5.50; all *p*-values < 0.0033, which is the Bonferroni-corrected significance level for 15 comparisons). The effect of the other terms did not reach statistical significance. Results suggest that participants spent more time exploring elevated elongated texture modulations. Presumably, at low amplitude isotropy becomes less relevant.
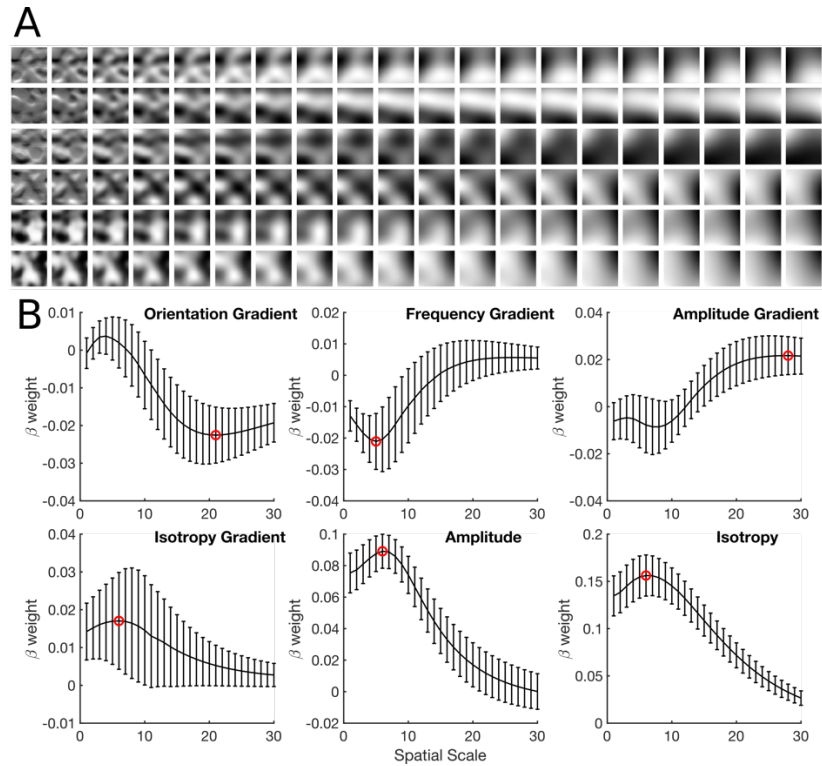

Fig. 3. Predictors and spatial tuning. A) Feature maps at different spatial scales: Gradients of orientation, frequency, amplitude, isotropy, and amplitude and isotropy from top to bottom; spatial scales from left to right. First column represents the non-filtered stimuli. The sigma of the Gaussian low-pass filter increases and the high spatial content decreases from left to right. **B)** Spatial tuning of the features. Average *β*-weights for each of the feature maps at each spatial scale, on y-axis. On x-axis the index of the spatial scale corresponding to the order of the filtered feature maps as represented in A). Error bars represent standard errors of the mean, computed across participants. Red circles indicate the most predictive spatial scales, chosen for further analyses.

Consistency between participants. In order to evaluate the generality of our results, we used a regression model based on the regression analyses performed on all but one participant, to predict that participant's touch duration. Prediction performance for each participant is a measure of consistency across participants, i.e. generality of our regression results. Specifically, for each participant, we computed a linear model by averaging the β-weights previously obtained by the individual multiple regressions after excluding her results (n-1 participants model). To compute the predicted touch duration, the regression model was applied to each of the experimental stimuli (example in Fig. 4B). Measured touch durations were expressed as a function of predicted touch durations. R^2 was computed for each participant and used as a measure of prediction performance (Fig. 4C).
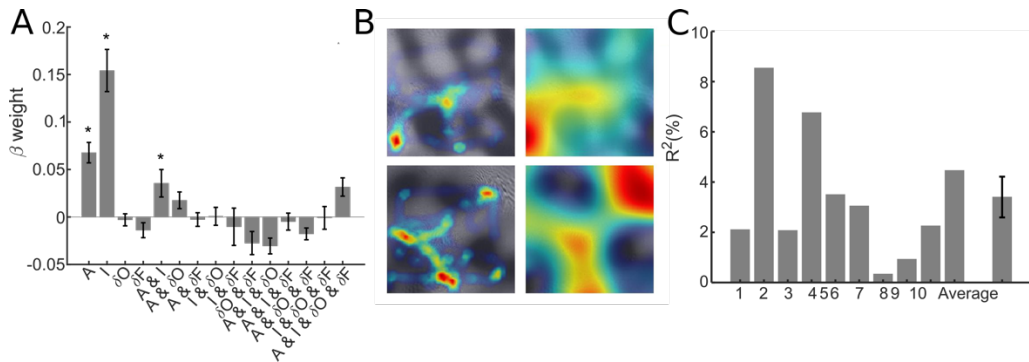

Fig. 4. Multiple regression results. **A)** Average β-weights on the y-axis. Regression terms on the x-axis: A (amplitude), I (isotropy), δO (Orientation gradient), δF (Frequency gradient) and their interactions. Stars indicate which β-weights were on average different from zero, after Bonferroni correction. **B)** Observed (left) and predicted by the n-1 participants model (right) touch duration for 2 different stimuli (bottom and top). Touch duration is color coded (blue refers to short and red to long times). Observed touch duration is the average over 4 trials from participant 2 ($R^2 = 8.54\%$). **C)** Prediction performance of the n-1 participants model expressed in terms of R^2 percent, on the y-axis. Participant index and average on the x-axis. The error bar is the standard error of the mean computed between participants.

Although small for some participants, R^2 showed that on average 3.4% of the variance of the touch duration was explained by the n-1 participants model, indicating that, predictions can be generalized to different participants. Pearson's correlation coefficient were on average (mean Pearson's r = 0.172) different from zero, $t_9 = 7.73$, $p < 0.05$, indicating that the n-1 participants model could on average significantly predict–to some extent–that one participant's exploratory behavior.

2.3 Discussion

We recorded touched position when participants were exploring relieved textures to judge their similarity. We used a set of physical properties of the textures (amplitude,

isotropy, amplitude gradient, isotropy gradient, spatial frequency gradient, and orientation gradient) to predict touch duration through a linear regression model. We first observed for each individual predictor, despite isotropy gradient, spatial tuning, i.e. touch duration was best predicted at an intermediate spatial scale. As a second step, we selected the best individual predictors across spatial scales, and between original features maps and their gradients. Finally, we modelled touch duration as a linear function of amplitude, isotropy, spatial frequency gradient, and orientation gradient. Regression results demonstrated that it is possible to predict exploratory behavior based on local texture information. Significant β-weights (amplitude, isotropy and their interaction) suggest that participants tended to preferentially explore high amplitude modulations and anisotropic areas (aligned rather than directionally uniform) of the texture. The positive interaction between amplitude and isotropy presumably means that when amplitude is low, isotropy is irrelevant. Otherwise, anisotropic areas are indeed carrying edge information, thus potentially informative about shape of objects which would be relevant to recognize in everyday life. Additionally, we found that one person's touching behavior is predictable on the bases of other participants' exploratory behavior at an above-chance level, given the local properties of the texture stimuli. This indicates that, the relationship between exploratory behavior and local texture information is to some extent consistent across participants.

In order to interpret spatial tuning, we need to consider that there is a portion of high spatial frequencies which is not resolvable for the perceptual system. Thus, filtering that portion out, removed noise from the analysis and improved predictions, explaining why the non-filtered version of a feature map is not the best predicting spatial scale. After a level of low-pass filtering at which most of the texture information is removed, if the exploration information is related to the local texture properties, predictions are again impaired. This is in fact what we found for the lowest spatial scales.

We only investigated a subset of the possible local haptic properties of an object. Previous research on haptic search has shown that a series of features related to 3D-shape (i.e. vertexes and edges, see [17, 27, 28]), temperature [17, 29] and compliance [17, 30] have the tendency to pop out. However, our research has focused on local texture properties, and did not include comparable features, so that, it is not known whether in a search task high amplitude modulations and anisotropic patterns tend to pop out. Another limitation is given by the simplicity of the linear models used. In fact, the relationship between local texture features and exploratory behavior does not have to be neither linear nor monotonic. We explicitly excluded spatial frequency and orientation as predictors because of their likely non monotonic relationships with exploratory behavior, with the effect of potentially loosing predictive power. Contemporary work on visual saliency is largely making use of machine learning techniques like support vector machines and deep neural networks (e.g.[31–33]). It is remarkable that with a simple linear model we predicted exploratory behavior better than chance and we aim to extend our analyses to more powerful methods in the future. In our experiments, we assumed that the attentional haptic focus was given by the index finger position, which we tracked. A different assumption could potentially change the interpretation of the results. Hence, the interpretation of the regression coefficients should be taken with caution. For visual perception several studies showed that

exploratory behavior strongly depends on the given task (e.g.,[34, 35]). Our participants were involved in a single task. Although the choice of a similarity judgment task was meant to avoid introducing a task-specific focus on particular features, we cannot assess to which extent haptic exploratory behavior is stimulus-driven or task-driven. Future research could investigate the relationship between touch duration and local texture features in a different task, in order to assess whether and how much of the exploratory behavior is purely bottom up driven and task-independent. Finally, in our analyses we neglected the dynamic aspects of explorations, i.e. we did not consider previous touch behavior. As it is for vision [11], future research can exploit dynamic aspects of touching behavior to improve predictions.

We have shown that haptic exploratory behavior is predictable based on local texture information, even with such a simple model as linear regression. Our results significantly extend research on haptic saliency from haptic search, which only allowed investigating which features are salient relative to a fixed class of distractors, towards modelling exploratory behavior on the base of local texture properties.

References

1. Crick, F.: Function of the thalamic reticular complex: the searchlight hypothesis. Proc. Natl. Acad. Sci. **81**, 4586–4590 (1984)
2. Treisman, A.M., Gelade, G.: A feature-integration theory of attention. Cognit. Psychol. **12**, 97–136 (1980)
3. Treue, S.: Visual attention: the where, what, how and why of saliency. Curr. Opin. Neurobiol. **13**, 428–432 (2003)
4. Itti, L., Koch, C., Niebur, E.: A model of saliency-based visual attention for rapid scene analysis. IEEE Trans. Pattern Anal. Mach. Intell. **20**, 1254–1259 (1998)
5. Itti, L., Koch, C.: A saliency-based search mechanism for overt and covert shifts of visual attention. Vision Res. **40**, 1489–1506 (2000)
6. Itti, L., Koch, C.: Computational modelling of visual attention. Nat. Rev. Neurosci. **2**, 194–203 (2001)
7. Schütz, A.C., Braun, D.I., Gegenfurtner, K.R.: Eye movements and perception: A selective review. J. Vis. **11**, 1–30 (2011)
8. Parkhurst, D., Law, K., Niebur, E.: Modeling the role of salience in the allocation of overt visual attention. Vision Res. **42**, 107–123 (2002)
9. Oliva, A., Torralba, A., Castelhano, M.S., Henderson, J.M.: Top-down control of visual attention in object detection. Presented at the Image processing, 2003. icip 2003. proceedings. 2003 international conference on (2003)
10. Walther, D.B., Serre, T., Poggio, T., Koch, C.: Modeling feature sharing between object detection and top-down attention. J. Vis. **5**, Art-No (2005)
11. Foulsham, T., Underwood, G.: What can saliency models predict about eye movements? Spatial and sequential aspects of fixations during encoding and recognition. J. Vis. **8**, 1–17 (2008)
12. Einhäuser, W., Spain, M., Perona, P.: Objects predict fixations better than early saliency. J. Vis. **8**, 1–26 (2008)
13. Masciocchi, C.M., Mihalas, S., Parkhurst, D., Niebur, E.: Everyone knows what is interesting: Salient locations which should be fixated. J. Vis. **9**, 1–22 (2009)

14. Chikkerur, S., Serre, T., Tan, C., Poggio, T.: What and where: A Bayesian inference theory of attention. Vision Res. **50**, 2233–2247 (2010)

15. Mahadevan, V., Vasconcelos, N.: Spatiotemporal saliency in dynamic scenes. IEEE Trans. Pattern Anal. Mach. Intell. **32**, 171–177 (2010)

16.   Wolfe, J., Horowitz, T.S.: Visual search. Scholarpedia. **3**, 3325 (2008)

17. Lederman, S.J., Klatzky, R.L.: Relative availability of surface and object properties during early haptic processing. J. Exp. Psychol. Hum. Percept. Perform. **23**, 1680 (1997)

18. Plaisier, M.A., Tiest, W.M.B., Kappers, A.M.: Haptic pop-out in a hand sweep. Acta Psychol. (Amst.). **128**, 368–377 (2008)

19. van Polanen, V., Tiest, W.M.B., Kappers, A.M.: Haptic pop-out of movable stimuli. Atten. Percept. Psychophys **74**, 204–215 (2012)

20. Grunwald, M., Muniyandi, M., Kim, H., Kim, J., Krause, F., Mueller, S., Srinivasan, M.A.: Human haptic perception is interrupted by explorative stops of milliseconds. Front. Psychol. **5**, 292 (2014)

21. Morash, V.S.: Systematic Movements in Haptic Search: Spirals, Zigzags, and Parallel Sweeps. IEEE Trans. Haptics. **9**, 100–110 (2016)

22. Sobel, I.: An isotropic 3× 3 image gradient operator. Mach. Vis. Three-Dimens. Scenes. 376–379 (1990)

23. Hsiao, S.S., Lane, J., Fitzgerald, P.: Representation of orientation in the somatosensory system. Behav. Brain Res. **135**, 93–103 (2002)

24. Johansson, R.S., Landstro, U., Lundstro, R.: Responses of mechanoreceptive afferent units in the glabrous skin of the human hand to sinusoidal skin displacements. Brain Res. **244**, 17–25 (1982)

25. Kovács, I., Fehér, Á.: Non-Fourier information in bandpass noise patterns. Vision Res. **37**, 1167–1175 (1997)

26. Seebeck, A.: Beobachtungen über einige Bedingungen der Entstehung von Tönen. Ann. Phys. **129**, 417–436 (1841)

27. Plaisier, M.A., Tiest, W.M.B., Kappers, A.M.: Salient features in 3-D haptic shape perception. Atten. Percept. Psychophys. **71**, 421–430 (2009)

28. van Polanen, V., Tiest, W.M.B., Kappers, A.M.: Integration and disruption effects of shape and texture in haptic search. PloS One. **8**, e70255 (2013)

29. Plaisier, M.A., Kappers, A.M.: Cold objects pop out! Presented at the International Conference on Human Haptic Sensing and Touch Enabled Computer Applications (2010)

30. van Polanen, V., Tiest, W.M.B., Kappers, A.M.: Haptic search for hard and soft spheres. PloS One. **7**, e45298 (2012)

31. Kienzle, W., Franz, M.O., Schölkopf, B., Wichmann, F.A.: Center-surround patterns emerge as optimal predictors for human saccade targets. J. Vis. **9**, 1–15 (2009)

32. Wang, W., Shen, J., Shao, L.: Video salient object detection via fully convolutional networks. IEEE Trans. Image Process. **27**, 38–49 (2018)

33. Kruthiventi, S.S., Ayush, K., Babu, R.V.: Deepfix: A fully convolutional neural network for predicting human eye fixations. IEEE Trans. Image Process. (2017)

34. Ballard, D.H., Hayhoe, M.M.: Modelling the role of task in the control of gaze. Vis. Cogn. **17**, 1185–1204 (2009)

35. Tatler, B.W., Hayhoe, M.M., Land, M.F., Ballard, D.H.: Eye guidance in natural vision: Reinterpreting salience. J. Vis. **11**, 1–23(2011)