

Processing of haptic texture information over sequential exploration movements

Alexandra Lezkan¹ · Knut Drewing¹

Published online: 3 October 2017
© The Psychonomic Society, Inc. 2017

Abstract Where textures are defined by repetitive small spatial structures, exploration covering a greater extent will lead to signal repetition. We investigated how sensory estimates derived from these signals are integrated. In Experiment 1, participants stroked with the index finger one to eight times across two virtual gratings. Half of the participants discriminated according to ridge amplitude, the other half according to ridge spatial period. In both tasks, just noticeable differences (JNDs) decreased with an increasing number of strokes. Those gains from additional exploration were more than three times smaller than predicted for optimal observers who have access to equally reliable, and therefore equally weighted, estimates for the entire exploration. We assume that the sequential nature of the exploration leads to memory decay of sensory estimates. Thus, participants compare an overall estimate of the first stimulus, which is affected by memory decay, to stroke-specific estimates during the exploration of the second stimulus. This was tested in Experiments 2 and 3. The spatial period of one stroke across either the first or second of two sequentially presented gratings was slightly discrepant from periods in all other strokes. This allowed calculating weights of stroke-specific estimates in the overall percept. As predicted, weights were approximately equal for all strokes in the first stimulus, while weights decreased during the exploration of the second stimulus. A quantitative Kalman filter model of our assumptions was consistent with the data. Hence, our

results support an optimal integration model for sequential information given that memory decay affects comparison processes.

Keywords Haptics · Temporal Processing · Texture

Textures are preferably judged by touch. Heller (1982, 1989) reported a greater contribution from touch compared with vision to texture perception. Given that textures are defined by repetitive small spatial structures on an object's surface, exploration covering a greater extent results in repetitive, redundant, intake of the same stimulus signals. Texture perception can benefit from integrating sensory information over time. Current models of information integration mostly refer to simultaneously presented redundant signals (Ernst & Banks, 2002; Drewing et al., 2008), e.g., holding a pen in the hand simultaneously results in both tactile and kinesthetic information about its diameter. In the present study, we investigated information integration for sequentially gathered signals in texture perception. In three experiments, we challenge predictions from models on simultaneous information and develop and test a more general Kalman filter model that allows accounting for specific observations in the integration of sequential information (Knill & Pouget 2004) by memory-decay affected comparison processes.

To describe the integration of simultaneous redundant information, the Maximum Likelihood Estimation (MLE) model is well-established (overview in Ernst & Bühlhoff, 2004). Jacobs (2002) suggested that integration uses all signals available for a property. First, signal-specific estimates s_i for the property are derived from each signal i . Second, all estimates are combined into a coherent percept P by weighted averaging:

✉ Alexandra Lezkan
alexandra.lezkan@psychol.uni-giessen.de

¹ Justus-Liebig University Giessen, Giessen, Germany

$$P = \sum_i w_i s_i \text{ where } \sum_i w_i = 1 \text{ and } w_i \in [0, 1]. \quad (1)$$

Estimates derived from each signal are prone to noise σ_i^2 . Averaging different estimates can decrease the perceptual variance (σ_s^2) of the combined percept (Landy, Maloney, Johnston, & Young, 1995). According to the maximum likelihood estimation (MLE) model, the variance (σ_s^2) of a percept is lowest and the weights (w_i) are optimal if the weights are proportional to the inverse variances of the signal-specific estimates ($1/\sigma_i^2$):

$$w_j = \frac{1/\sigma_j^2}{\sum_{i=1, \dots, j, \dots, N} 1/\sigma_i^2} \quad \text{with} \quad \sigma_s^2 = \frac{1}{\sum_i 1/\sigma_i^2}. \quad (2)$$

Weighted averaging (Eq. 1) describes the percept of a property, when stimuli with signals slightly conflicting in their information on this property are created (Ernst and Banks, 2002). Experimental data also quantitatively confirm the predicted reduction of perceptual variance (measured via discrimination thresholds) in multi-estimate compared with single-estimate situations (Eq. 2), and even the predicted optimal weights, e.g., for the case of visuo-haptic and visuo-auditory integration of size and location (Alais & Burr, 2004; Ernst & Banks, 2002). Recent studies found neurophysiological correlates of optimal multisensory integration (Fetsch, DeAngelis, & Angelaki, 2013; Helbig et al., 2012).

Within haptic perception, observers use multiple redundant signals that are simultaneously available and integrate them in agreement with MLE predictions (Drewing & Ernst, 2006; Drewing, Wiecki & Ernst, 2008). However, in haptic perception, the integration of information over time is at least as important as integration over different sensory sources (Henriques & Soechting, 2005). Typical haptic exploratory procedures extend over time and space and can be decomposed into several exploration segments. For specific object dimensions, such as surface orientation or texture, exploratory behavior comes along with a systematic repetition of the same stimulus information. In texture exploration, individual exploration segments refer to scans of the finger over the same spatial region. Thereby, extending the exploration by repeating exploration segments increases the amount of redundant information. To formulate a model for such sequential and not simultaneous information, a Kalman filter (Kalman, 1960) may be better suited than the MLE model. The Kalman filter takes a more general approach to optimal information integration. It is able to describe how a series of sequential estimates are used for estimating a property in a way that the variance of the final estimate is minimized. The Kalman filter uses Bayesian interference, combining prior with present information, and can account for changes in the estimates over time. For example, a Kalman filter approach can model if memorized information from sequentially

gathered signals gets noisier over time. First empirical studies observed correlates of fundamental Kalman filter characteristics, prediction and updating, in the brain activity of mice (Funamizu, Kuhn, & Doya, 2016). The MLE model and its predictions are captured within the Kalman filter framework as a (simple) special case with noninformative prior information and estimates that are stable over time (Battaglia, Jacobs, & Aslin, 2003; Ernst & Bühlhoff, 2004).

The present study was designed to challenge predictions from the MLE model and to develop a better-suited Kalman filter model for the sequential integration of texture information. The exploratory procedure for textures includes several lateral strokes in different directions (Lederman & Klatzky, 1987). We define an exploration segment as a single unidirectional stroke across the texture. Then, a segment-specific estimate for a property is derived from the information gathered during a single stroke. We assume that each exploration segment i yields an estimate with equal variance ($\sigma_i^2 = \sigma_0^2$, with σ_0^2 being a constant value σ_i^2). The assumptions underlying the MLE model predict that all estimates are weighted equally in the percept (Eq. 2, left) and the final variance of the percept (σ_s^2) can be computed by $\sigma_s^2 = \sigma_0^2/N$ (Eq. 2, right) with N being the number of redundant estimates. Given that the discrimination threshold (t_s^2) assesses the percept's variance (σ_s^2) with $t_s^2 = 2\sigma_s^2$ (Jovanovic & Drewing, 2014; Lezkan et al., 2016), it follows for discrimination thresholds:

$$t_s = \sqrt{2\sigma_0^2/N} \quad \text{and} \quad \log(t_s) = -\frac{1}{2}\log(N) + \text{const.} \quad (3)$$

That is, discrimination thresholds should depend on the number of exploration segments in a well defined fashion and a linear fit on log-log scales should have a slope of $-1/2$. Previous research on sequential integration of extended haptic stimulation seems not to support these predictions. Quick (1974) had already suggested in his model that visual thresholds linearly decrease with increasing stimulation on a log-log scale, but with diverse slopes. For haptic detection thresholds, the observed slope in Quick's model was close to -1 (Gescheider, Berryhill, Verrillo, & Bolanowski, 1999; Gescheider, Bolanowski, Pope, & Verrillo, 2002; Gescheider, Güçlü, Sexton, Karalunas, & Fontana, 2005; Louw, Kappers, & Koenderink, 2005) and thus clearly below the slope of $-1/2$ predicted from the assumptions underlying the MLE model. However, performance in detection tasks might not be relevant, because detection does not require perceiving the magnitude of a stimulus property (Louw et al., 2005). In a discrimination task on felt surface orientation, thresholds decreased with increasing length of exploration, and the decrements were smaller the longer the explored surface was (Giachritsis, Wing & Lovell, 2009). This is qualitatively in line with the threshold predictions but was not quantitatively analyzed and thus is not conclusive. Importantly,

results from Metzger, Lezkan, and Drewing (2017) are at odds with the prediction of equal weights in the integration of sequential haptic information. The authors investigated softness discrimination, where people typically indent a soft stimulus repeatedly, and determined the weights of indentation-specific softness estimates for the first and the second stimulus in a trial. While a rather equal weighting was visible for the indentations of the first stimulus, during the exploration of the second stimulus weights decreased for later indentations.

Thus, Metzger et al.'s (2017) results casts the assumptions of the MLE model into doubt and call for a more complex model of the processes of sequential integration during discrimination tasks. These results seem to agree with a model of the comparison process between first and second stimulus that can be derived from single cell measurements on monkeys. In a vibrotactile discrimination task, Romo and colleagues (Romo, Hernández, Zainos, Lemus, & Brody, 2002; Romo & Salinas, 2003) found that neuronal responses in area SII are different for the first and the second stimulus in a trial. Whereas the response to the first stimulus was only associated with the first stimulus' characteristics, the response to the second stimulus also included information about the first remembered stimulus. This is to say, neural responses during the second stimulus reflected the comparison between the two stimuli, which was the task of the monkey. Hernández et al. (2010) measured the monkey's cortical activity during vibrotactile discrimination. The activity of frontal lobe circuits was associated with the result of the sensory decision which of the two stimuli had higher frequency as well as with the past information about the stimuli. Most importantly, cortical areas that receive inputs from area SI were reported to combine present sensory information from SI with sensory representations stored in working memory. Overall, the results suggested that comparison processes take place during the presentation of the second stimulus after the first stimulus has been captured and memorized as a reference.

This can explain the data from Metzger et al. (2017) on decreasing weight of sequential estimates during the exploration of the second stimulus in softness discrimination, as follows. During the exploration of the second stimulus, present sensory signals are continuously compared with the remembered estimate. Within this comparison process, the variance of the estimate of the remembered first stimulus increases due to memory decay. Hence, information gathered sooner after the first stimulus may lead to a more precise judgment on the difference between the two stimuli than later information and therefore is weighted higher. Such a process will not be captured by the rather simple assumptions underlying the MLE model but requires a Kalman filter model that can additionally account for changes in the estimates' variance.

In the first experiment of the present study, we investigated for texture discrimination how the (spatio-temporal) extension

of exploratory movements, i.e., the number of strokes across the texture, affects discrimination thresholds. The assumptions underlying the MLE model predict that the reduction of thresholds follows a power function of the number of strokes with exponent $-1/2$, whereas the outlined model on the comparison process with memory decay predicts less reduction (i.e., a larger exponent). In the second experiment, we tested whether stroke-specific estimate weights are unequal and follow the pattern predicted from the outlined model on the comparison process. Finally, in Experiment 3 we tested quantitative predictions for the estimate weights that stem from a Kalman filter model of optimal integration given memory decay affected the comparison process.

Experiment 1

We created haptic texture stimuli by using a PHANToM force-feedback device. The device is attached to a finger via a thimble. It simulates objects by monitoring 3D-finger position and by applying an appropriate reaction force. We used virtual gratings that consisted of sinusoidal ridges on an otherwise planar surface. Different grating stimuli differed in ridge height or the distance between adjacent ridges (= period). On each trial, participants explored one of the two possible standard gratings and one comparison grating. Afterwards, half of the participants decided which grating had felt higher (amplitude judgment), the other half decided about grating period (period judgment). Participants were instructed to explore with back and forth movements having a defined finger velocity and force to avoid confounds. As a consequence, participants had to simultaneously focus on the discrimination task and their exploratory movement. To reduce the attention needed for movement control, the movement was guided by intuitive visual feedback and participants initially practiced the instructed force and velocity.

The experiment started with this "practice phase." Afterwards, in the "exploration phase," we varied the number of strokes (1...8) that participants used to explore each stimulus. We measured just-noticeable differences (JNDs; assessing discrimination thresholds) for either task by using the adaptive staircase procedure called BestPEST (Lieberman & Pentland, 1982). We expected that JNDs would decrease with the number of strokes conducted following a power function. Furthermore, we tested the exponent of the power function against $-1/2$, which is the value predicted by the assumptions underlying the MLE model.

Participants

A total of 16 healthy participants, students from Giessen University, were tested (mean age: 22 years, range: 19-26 years; 9 females, 7 males). All participants had normal or

corrected-to normal visual acuity, were right-handed, and none of them reported cutaneous or motor impairments. Participants were naïve to the purpose of the study. They participated for course credit. Methods and procedures of both experiments were approved by the local ethics committee LEK FB06 at Giessen University, and they were in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki. Participants gave written, informed consent.

Apparatus and Stimuli

The apparatus can be seen in Fig. 1a. Participants sat in front of a custom-made visuo-haptic workbench, which comprised a PHANToM 1.5A haptic force feedback device and a 22"-computer screen (120 Hz, 1024 x 1280 pixel). The right index finger was connected to the PHANToM via a thimble-like holder, which allows for free finger movements having all six degrees of freedom in a 38 x 27 x 20 cm³ workspace. Simultaneously, the participants looked through stereoglasses (CrystallEyes™) and via a mirror onto the screen (40-cm viewing distance). The mirror prevents participants from seeing their hand and enables spatial alignment of the 3D-visual with the haptic display. The participants' heads were stabilized by a chinrest. The devices were connected to a PC. A custom-made software controlled the experiment, collected responses, and recorded finger positions and reaction forces (from PHANToM, every 2 ms). Noise presented via headphones and ear plugs masked sounds generated by the PHANToM.

Both stimuli were presented after each other in front of the participants. The stimuli were virtual gratings covering an area

of approximately 30-mm width (x -axis) x 15-mm depth (z -axis). Gratings consisted of ridges (width 1 mm; extending over the entire depth) on an otherwise planar surface. Ridge height was a sine-function (within 0 to π) of x -position. Programmed peak amplitudes of the ridges varied between 0.16 and 0.74 mm; the peak-to-peak period between ridges varied between 2 and 9 mm. In each single stimulus, ridge amplitudes and periods were constant. Strokes started left or right from the grating. Haptic grating stimuli were created using the PHANToM force feedback device. The device simulates objects by applying reaction forces \vec{F}_p as a function of the 3D-finger position P . Force magnitude linearly increases with the indentation depth of the finger into a virtual object (i_p) and force direction is normal to the object's surface (\vec{n}_p : normal vector, D : spring constant):

$$\vec{F}_p = \vec{n}_p * |\vec{F}_p| \text{ and } |\vec{F}_p| = D * i_p \quad (4)$$

The spring constant D was replaced by the variable K to keep object indentation constant under differing finger forces. The variable K was defined such that for the target indentation I (set to 1 mm) the magnitudes of finger force and reaction force were (approx.) equal. Vertical finger force was estimated from the device's reaction forces in y -direction $F_y(j)$ (y -axis = height) in the previous device cycles $j = 1 \dots n$ (~previous 300 ms):

$$K = \frac{1}{n} \sum_{j=1 \dots n} F_y(j) / I \quad (5)$$

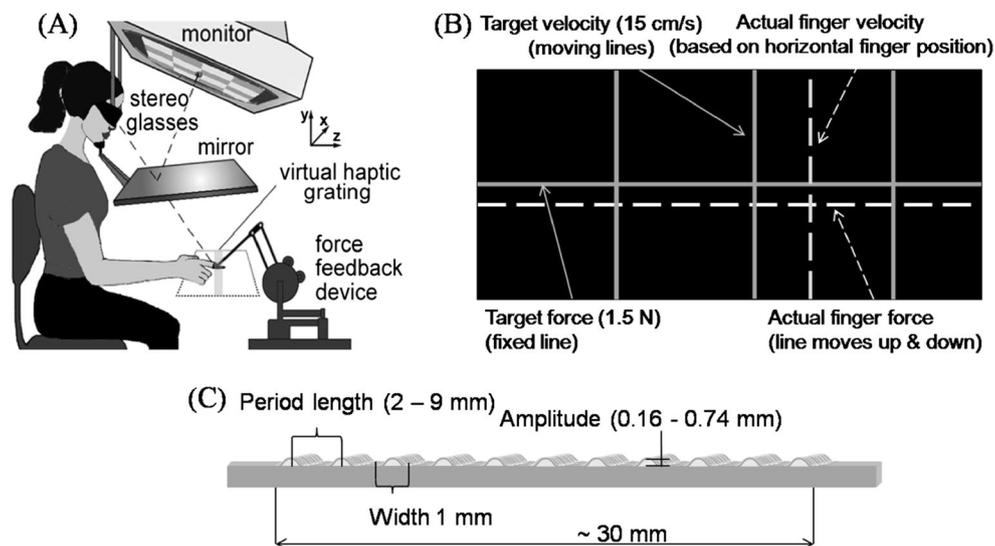


Fig. 1 Sketch of the visuo-haptic setup (a), the visualization presented during exploration (b), and a stimulus (c). (a) Participants were sitting in front of the workbench, wearing earplugs and headphones. A head and chin rest limited head movements. (b) Visual feedback on the two movement parameters velocity and force. Feedback lines were only displayed

while the finger was outside the grating area. Please notice: what is depicted as solid lines were actually blue lines and what is depicted in dashed lines were red lines. (c) Stimuli were virtual gratings, which varied in the period length for half of the participants and in the amplitude for the other half

Design and Procedure

Participants successively explored two gratings. Between participants we varied the Judged Dimension (Amplitude, Period). Half of the participants judged which of the two gratings had felt higher (Amplitude); the other half judged which grating had higher spatial period (Period). We further varied the *Number of strokes* (1, 2, 3, 4, 5, 6, 7, 8) that participants used to explore each of the two stimuli (within-participant variable). A single stroke was defined by a single unidirectional exploratory movement across the grating. We measured 75% discrimination thresholds (JNDs) for two standard stimuli. The standard stimuli in the Amplitude group had amplitudes of 0.4 or 0.5 mm and periods of 5 mm. In the Period group, the standard stimuli had periods of 5 or 6 mm and amplitudes of 0.4 mm.

JNDs were determined using the BestPEST adaptive staircase procedure combined with the two-interval forced-choice task. In the BestPEST method (Lieberman & Pentland, 1982) before each stimulus presentation, the likelihood distribution of possible thresholds is calculated by using the sigmoid-shaped psychometric function with a slope of one, on the basis of all previous responses of the participant. The value with the maximum likelihood of being the threshold value is then chosen as the comparison stimulus. This method is an optimum strategy for fast threshold determination. In effect, the procedure raises the difference between the values of comparison and standard after a wrong response and lowers it after a correct response. We terminated the procedure after 26 trials per staircase, estimating the 75% threshold (JND) by the final maximum-likelihood estimate. For each *Number of strokes* and each standard stimulus, two up and two down staircases measured the upper and lower JNDs, respectively. In the Amplitude condition, initial amplitudes of the comparison stimuli were given by the standard's amplitude ± 0.35 mm; the comparisons' period was always 5 mm. In the Period condition, initial periods of the comparisons were the standard's period plus/minus 3 mm; the comparisons' amplitude was always 0.4 mm. Trials from all staircases were randomly interleaved in the measurement phase. Overall, the measurement phase consisted of 2 [standards] * 2 [staircases] * 26 [staircase length] * 2 [repetitions] * 8 [*Number of strokes*] = 1,664 trials. The entire experiment consisted of five sessions lasting approximately 2 hours each. Before to the experiment, participants were trained for approximately 30 min to execute exploratory movements with constant instructed finger velocity (15 cm/s) and force (1.5 N). The training consisted of two parts. In the first part, participants trained on a virtual plane without ridges. In the second part of the training, movements were performed on virtual gratings. Each part ended after participants had performed 20 trials in a sequence with maximally 3 movement errors. We defined movement errors as a deviation of actual velocity or force values by more than 60% from the target velocity and force.

Each trial started with a visual representation of the upcoming stimulus and start point (left or right of the grating, balanced). Participants initiated the trial with a button press at the start point location. Then, participants stroked across a first grating back and forth. The computer program stopped the stimulus presentation, when the required number of strokes had been conducted. Afterwards, a second grating was explored using the same number of strokes as for the first grating. Finally, participants had to decide by a button press (done with the PHANToM), which grating had felt higher in amplitude/had higher spatial period. During the strokes, a vertical line that moved forth or back along the exploratory axis indicated the prescribed finger velocity (15 cm/s) and stroke direction. A stationary horizontal line indicated prescribed force (1.5 N). Participants monitored their current velocity and force by further feedback lines, which were displayed while the finger was outside the grating area. A vertical line displayed the current 1D-finger position on the x-axis; a horizontal line moved up and down with exerted force. Trials were repeated later in the session when a movement error was detected.

Data Analysis

We calculated individual JNDs per *Number of strokes* condition by averaging across the two upper and the two lower JNDs for each standard stimulus (8 JND values). These values were log-transformed (base 10) before analyses. According to the predictions it is the log JNDs that should linearly decrease with the log *Number of strokes*. In addition, the log-transformation allows comparing gain ratios in the amplitude and the period conditions. It transforms the ratios between JNDs for different *Number of strokes* into differences, which can be directly analyzed by an ANOVA.

Results

Individual log JND values entered an ANOVA with the within-participant variable *Number of strokes* (1..8) and the between-participant variable Judged Dimension (Amplitude, Period). For the variable *Number of strokes*, we calculated linear contrasts, which provide a targeted test of our hypotheses. The linear contrast of *Number of strokes* was significant, $F(1,14) = 15.326$, $p < 0.001$ (one-tailed), confirming the predicted decrease of JNDs with an increasing *Number of strokes*. The interaction *Number of strokes* (linear contrast) X *Judged dimension* failed to reach significance, $F(1,14) = 0.350$, $p = 0.563$, which may suggest that both amplitude and period JNDs depend in similar manner on the *Number of strokes*. To be more precise, the lack of effects on log values suggests that the ratios between the JNDs of different *Number of strokes* conditions are similar. Finally, the main effect of *Judged Dimension* was significant, $F(1,14) = 584.050$, $p <$

0.001, which is, however, essentially not interesting, because it only reflects the fact that (log) amplitude and period JNDs differ in scale. Figure 2 shows a log-log plot of the JNDs.

We fit a power function separately to the amplitude JNDs and to the period JNDs. To achieve this, we linearly regressed log transformed JNDs on log transformed stroke numbers. As a consequence, the slope of the fitted line corresponds to the exponent of a power function fitted to the non-logarithmized data. In both cases, the fitted line described the data well. For the Amplitude group, the regression line explained $r^2 = 88\%$ of the variance. For the Period group, the explained variance was $r^2 = 80\%$. According to the MLE predictions, the slope is expected to be -0.5 . In contrast, the slopes of the fitted lines reached values of -0.148 for the Amplitude group and -0.112 for the Period group. By fitting regression lines to the individual log-log data, we were able to calculate a t test against the predicted slope of -0.5 . In the Amplitude group ($M = -0.148$, $SD = 0.151$) and the Frequency group ($M = -0.112$, $SD = 0.066$), the slopes differed significantly from the MLE prediction, $t(7) = 6.580$, $p < 0.001$ and $t(7) = 16.673$, $p < 0.001$.

Discussion Experiment 1

In Experiment 1, we found that participants discriminate grating stimuli the more precisely the longer they explore them. Such redundancy gains were smaller than predicted by the assumptions underlying the MLE model. According to these assumptions each single estimate is weighted according to its inverse variance. In case of repeated strokes across the same stimulus, estimates from each single stroke should have equal variance and, hence, each estimate should obtain equal

weight. The present results disprove the MLE predictions, and thus extend the previous evidence (Metzger et al.'s, 2017), suggesting that the assumptions underlying the MLE model do not apply to sequential integration.

As outlined in the introduction, an alternative model, which may explain the present and previous observations on sequential integration, links to memory decay during the comparison process of the discrimination task. There is evidence that discrimination performance is based on a continuously ongoing comparison process between a remembered estimate from the first stimulus and present sensory signals from the second stimulus (Romo et al., 2002; Romo & Salinas, 2003; Hernández et al., 2010). During the comparison process, i.e., during exploration of the second stimulus, the memory trace of the first stimulus might diminish from stroke to stroke, and thus the variance of the remembered estimate increases. Memory decay and increasing variance, as observed, will lead to lower redundancy gains than predicted from the MLE assumption of equal variance and higher overall estimate variance. An optimality model, including these factors, in sequential presentation, would further predict that strokes within the second stimulus are not weighted equally but decrease for later strokes. We designed further experiments to test whether information from different strokes during the exploration is unequally weighted in the grating percept.

Experiment 2

In Experiment 2, participants discriminated a standard and a comparison stimulus according to grating period using a two-

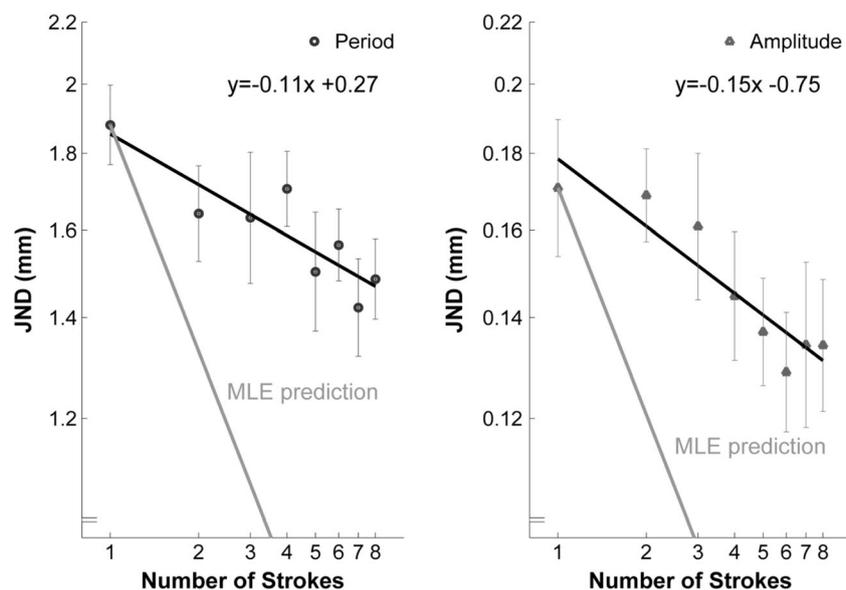


Fig. 2 Exp. 1: Log-Log plot. Average JNDs for frequency discrimination (left; expressed as period) and amplitude discrimination (right) and standard errors as a function of *Number of strokes* and *Judged dimension*. The gray line represents the MLE model prediction of an optimal integration

interval forced choice task combined with the method of constant stimuli. They stroked three times across each stimulus. While participants explored the standard stimulus, we presented slightly discrepant period information in one of the strokes. That is, the grating period of each stroke in the standard stimulus could take one of two values. The stroke with the deviant period in the standard stimulus is the discrepant stroke. We defined several standard stimuli by varying the *Position* [1, 2, 3] of the discrepant stroke within the presentation of the standard. Additionally, the standard was either presented as the first or second stimulus of the trial, which is represented in the variable *Stimulus order* [first vs. second]. Each standard stimulus was combined with 14 comparisons. The comparisons differed in their periods, but for the strokes across each single comparison stimulus the period was kept constant. For each of the standards, we determined the point of subjective equivalence (PSE) with the comparison. Based on this, we calculated the weight of the discrepant information in the standard stimulus for each combination of *Position* and *Stimulus order*. We predicted an interaction between those two variables. Weights were expected to decrease with higher *Position* in the second but not in the first stimulus.

Participants

The final sample included 11 students (8 females, 3 males). Four additional participants had been excluded, because they had problems with the task, either with the movement (>30% trials with movement error) or with the discrimination ($JND \geq 6$ mm in experimental conditions, exceeding the effective range of measurement of the present design). Participants in the final sample were naïve to the purpose of the study, right handed, had an age range of 19–26 years, no sensory or motor impairments, and participated for course credit.

Apparatus and Stimuli

The apparatus and the virtual gratings were the same as in Experiment 1. The ridges of all grating stimuli had peak amplitudes of 0.5 mm. Typically, a standard stimulus was explored by three strokes. For strokes over standard stimuli, we used periods of 6 and 4.5 mm. In the experimental conditions, the period presented in two of three strokes is called the dominant period. In the remaining stroke, the participant was presented a discrepant period. Thus, if the dominant period was 4.5 mm, we presented in one stroke the discrepant period of 6 mm. The discrepant period of 6 mm could be presented in the first, middle, or last stroke, whereas in the other two strokes the dominant information of 4.5 mm would be presented. Additionally, in control conditions with one or three strokes we used standard stimuli, in which no discrepant information was presented. Furthermore, we presented 14 comparison stimuli that

varied in period (2–8.5 mm in steps of 0.5 mm) and in which no discrepant information was presented.

Design and Procedure

Similar to Experiment 1, in each trial, participants explored a standard and a comparison grating in random order. A trial was constructed as in Experiment 1. Participants always judged which grating had higher spatial frequency. In the experimental conditions, each stimulus was explored with three strokes. For the majority of the standards, one stroke (discrepant period stroke) differed in its spatial period from the two others (dominant period strokes). We varied the *Position* of the discrepant stroke within the standard stimulus (1st, 2nd, or 3rd stroke) and the *Stimulus order* (standard as 1st or 2nd stimulus) as within-participant variables. Additionally, we included control conditions, in which we presented standard stimuli with dominant period information from each stroke, either 4.5 mm or 6 mm. Participants explored these stimuli with three or one stroke. In contrast to Experiment 1, the point of subjective equivalence (PSE) and just noticeable differences (JND) of the standard periods were assessed using the method of constant stimuli: for each stimulus order each standard was compared 8 times to each of the 14 comparisons. Overall, the experiment comprised 10 [standards] * 14 [comparisons] * 2 [stimulus order] * 8 [repetitions] = 2,240 trials. The entire experiment consisted of 4 sessions lasting about 2–2.5 hours each. In one session, each standard-comparison pairing was repeated four times. The first sessions started with a phase for training instructed finger force and velocity as in Exp. 1.

Data Analysis

We determined individual psychometric functions for each standard stimulus and each *Stimulus order* (standard is 1st vs. 2nd stimulus). The percentage of trials in which the participant perceived the standard to be higher in spatial frequency than the comparison was calculated as a function of the comparison stimulus. We fitted cumulative Gaussian functions to the psychometric functions, using the `psignifit` toolbox that implements maximum-likelihood estimation procedures (Wichmann & Hill, 2001) and estimated PSEs by the Gaussian parameter μ and JNDs by σ (84% discrimination thresholds). We calculated individual weights of the discrepant stroke (w_d) from the PSEs in the experimental conditions (P_e), and from the two average PSEs in the control conditions (P_d : PSEs for standard with the same period as the discrepant stroke, P_o : PSEs relating to period of dominant strokes):

$$w_d = (P_e - P_o) / (P_d - P_o) \quad (6)$$

We averaged over the two weights for the two standard stimuli in each condition. Additionally, all weights were restricted to

have values within 2 standard deviations from the condition average (5 outliers in 66 cases). The individual average weights of the discrepant stroke were analyzed by ANOVAs.

Results

PSEs

In the control condition, participants explored either with one or three strokes two sequential gratings without any discrepant information within the standard. The PSEs represent the perceived period of the stimuli and are plotted in Fig. 3. We analyzed the PSEs by an ANOVA with the three factors *Period of the standard stimulus* (4.5 mm vs. 6 mm), *Number of strokes* (1 vs. 3) and *Stimulus order* (1st vs. 2nd). The PSEs in the control conditions differed significantly regarding the spatial *Period of the standard stimulus*, $F(1,10) = 166.57$, $p < 0.001$, which ensures our manipulation. There was no significant effect of the *Number of strokes*, $F(1,10) = 0.22$, $p = 0.651$, the *Stimulus order*, $F(1,10) = 3.90$, $p = 0.077$, *Number of strokes* \times *Stimulus order*, $F(1,10) = 1.02$, $p = 0.336$, or the *Number of strokes* \times *Period* \times *Stimulus order*, $F(1,10) = 0.001$, $p = 0.981$. However, the interaction between *Stimulus order* and *Period* was significant, $F(1,10) = 18.29$, $p = 0.002$. As shown in Fig. 3, the difference between the percepts of the 4.5-mm stimulus and the 6-mm stimulus was higher in the second in contrast to the first stimulus. It is important to note that these effects will not affect our predictions about the weights, as average PSEs measured in the control condition are accounted for in the computation of weights.

To check whether the discrepant information influenced perception, we compared the PSEs from experimental conditions, i.e., from standards including discrepant information, to the PSEs from the control conditions. This analysis was done

separately for the first and the second stimulus and each dominant period. As expected, discrepant stimuli with the dominant period of 4.5 mm were perceived to have higher period than the corresponding control stimuli ($t(11) = 2.242$, $p = 0.024$ and $t(11) = 4.986$, $p < 0.001$, one-tailed, for first and second stimulus, respectively), and discrepant stimuli with the dominant period of 6 mm were perceived as having lower period ($t(11) = -3.050$, $p = 0.006$ and $t(11) = -3.332$, $p = 0.004$).

Weights of Discrepant Information

The position of a stroke in a stimulus differently affected this stroke's weight depending on whether the first or the second stimulus was considered (Fig. 4). Individual weights were entered into an ANOVA with the within-participant variables *Stimulus order* (1st vs. 2nd in trial) and *Position* within stimulus (1st vs. 2nd vs. 3rd stroke). The *Position* of the discrepant stroke within the stimulus did not show a significant main effect on the weight, $F(2,20) = 0.166$, $p = 0.849$. The main effect of *Stimulus order* also was not significant, $F(1,10) = 0.019$, $p = 0.894$. More importantly and as expected, the interaction of *Stimulus order* and *Position* was significant, $F(2,20) = 4.666$, $p = 0.022$. We tested further whether, as also predicted, weights in the first stimulus do not depend on the *Position* within the stimulus, while weights in the second stimulus decrease the further their position is from the first stimulus. We calculated linear contrast analyses separately for the first and the second stimulus. In the first stroke, these analyses did not reveal a significant linear effect of position, $F(1,10) = 4.065$, $p = 0.071$. Also as predicted, in the second stroke weights systematically decreased with increasing stroke position, $F(1,10) = 6.233$, $p = 0.016$ (one-tailed).

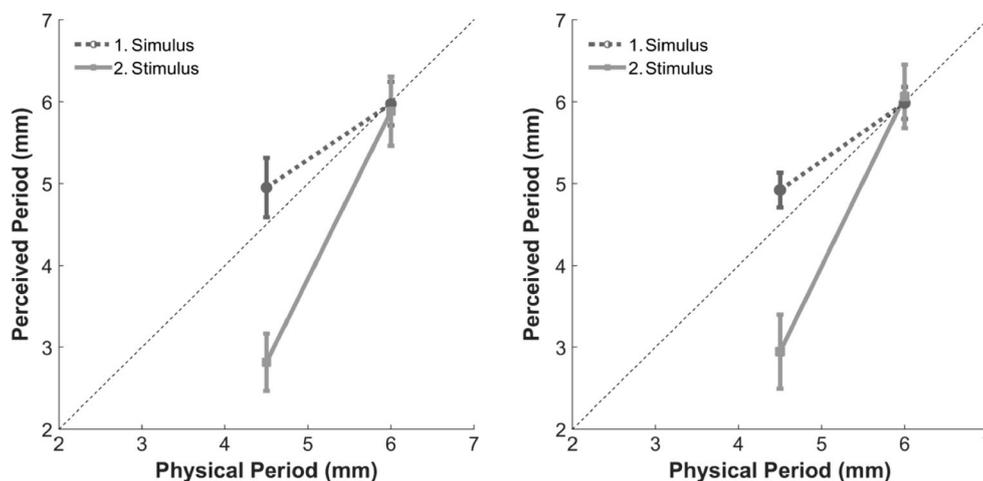


Fig. 3 Exp. 2, control condition. Average PSEs and standard errors (11 participants) as a function of the spatial period of the standard and of the Stimulus order. Left is the control condition with 1 stroke, right the control condition with 3 strokes.

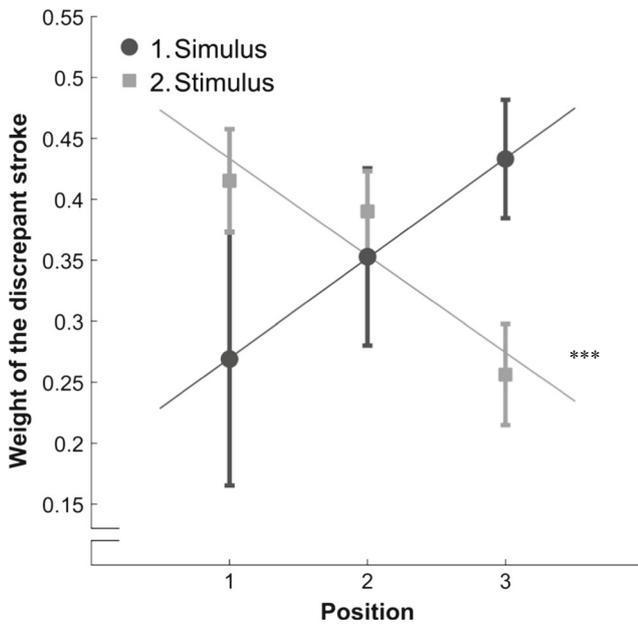


Fig. 4 Exp. 2, Average estimated weights and standard errors of the discrepant stroke as a function of Position within the stimulus and Stimulus order within the trial.

JNDs

In Experiment 1, the participants showed better discrimination thresholds for increasing numbers of strokes. In the present Experiment, we can test with the two control conditions whether this effect can be replicated. We expect better discrimination thresholds in the 3-stroke condition than in the 1-stroke condition. A paired one-tailed t test of the log-transformed (base 10) JNDs showed a significant difference between the two control conditions, $t(10) = 3.347, p = 0.004$ (JNDs 1-stroke condition: $M = 4.32, SEM = 0.70$; JNDs 3-stroke condition: $M = 3.24, SEM = 0.43$).

Discussion Experiment 2

We introduced slight discrepancies in spatial period information in a one of several strokes across a grating stimulus. We varied the position of the discrepant spatial period information within the standard stimulus presentation. None of the participants reported to have noticed the discrepant periods when being asked after the experiment. Discrepant information contributed to the grating percept, as shown in the significant PSE shifts in the expected directions. From PSEs, we calculated individual weights of the discrepant stroke for each condition. Our results confirmed our predictions: weights depended differently on stroke position for the first and the second stimulus. Weights did not significantly change within the first stimulus. However, in the second stimulus, a stroke's weight was higher the closer the discrepant stroke was to the first stimulus. Our data are consistent with the assumption that

the comparison process during the exploration of the second stimulus becomes—due to decay of the memory trace of the first stimulus—increasingly more variable over time and later strokes are weighted less.

One may wonder whether correlated errors between stroke-specific estimates can alternatively explain the results, as had been the case for other failures of MLE predictions (Oruç, Maloney, & Landy, 2003; Rosas, Wichmann, & Wagemans, 2007): Positively correlated errors reduce the effect of an additional estimate on the percept's overall variance as compared to the MLE predictions (Eq. 2). That is, the higher the correlation between the additional estimate and previous estimates, the higher the variance of the final percept. In the case of a sequential, step-by-step integration of correlated estimates, estimates gathered later would correlate more with the previously collected information than earlier estimates, and hence, effectively decrease variance less and obtain less weight in the percept (Oruç et al. 2003). That is, correlated errors between stroke-specific estimates predict lower weights for later strokes. This prediction applies to both strokes across the first stimulus and strokes across the second stimulus in a trial. However, for the first stimulus we did not observe such a downward trend, rejecting the alternative explanation by correlated errors.

Sensory adaption could be another possible explanation of the data. It was reported that after repeated stimulation sensory adaptation leads to aftereffects by reducing sensitivity (Thompson & Burr, 2009). Such aftereffects were shown in different aspects of the sense of touch (Kappers & Bergmann Tiest, 2015), including the perception of vibration (Lederman, Loomis, & Williams, 1982; Hollins, Bensmaïa & Washburn, 2001). Thereby, the sensitivity should be the more reduced the more stimulations were presented. Sensory adaptation, thus, may predict that information from later strokes is noisier and hence weighted less. However, sensory adaption would predict the same pattern as correlated errors do, namely a general position effect, which applies to the first and the second stimulus. Thus, sensory adaptation can be rejected as an alternative explanation for the observed pattern of weights. A possible reducing role of adaptation for the overall variance in longer explorations might deserve further investigation in the future.

We observed, as expected, no position effect for the first stimulus. But the results on a lack of position effect are not as convincing, as we hoped. Numerically, the line of regression of weight on stroke position for the first stimulus shows an upward trend with high standard errors, which may or may not explain the lack of significance. We conducted another experiment to replicate the findings and extend them for different numbers of strokes. Importantly, Experiment 3 tests quantitative predictions from a Kalman filter model of optimal integration under conditions of memory decay during the comparison process, i.e., during exploration of the second stimulus.

Experiment 3

Experiment 3 is meant to generalize the investigations from Experiment 2 to explorations of varying lengths and to compare the results to predictions from a formal Kalman filter model. We manipulated the number of strokes used to explore standard and comparison stimulus. Participants explored each stimulus either with 2, 3, 4, or 5 strokes. Additionally, as in Experiment 2, we varied the position of the discrepant information within the standard stimulus (1st...Nth position with N being the number of strokes) and the stimulus order (standard presented first vs. second). We measured the PSEs and JNDs for each condition and calculated the weight of the discrepant stroke.

Additionally, we tested in Experiment 3 if our model of a comparison process with memory decay can quantitatively predict the data. In a nutshell, the model assumes that estimates from the individual strokes of the first stimulus are integrated to an overall percept, and that during the exploration of the second stimulus, estimates from each stroke are compared stroke-by-stroke with the integrated estimate from the first stimulus. The initial integration of the first stimulus estimate is modelled in line with the assumptions of the MLE model. However, importantly, the first stimulus' estimate is affected by memory decay. To account for the comparison process during the exploration of the second stimulus, we used a more complex Kalman filter model of optimal integration.

Model

We assume that for each stroke (i) of the second stimulus the stroke-specific estimate is compared to the overall estimate from the first stimulus, resulting in a sensory difference value $D^{(i)}$. The posterior estimate of the difference between first and second stimulus after this stroke $\hat{D}^{(i+1)}$ is based on the present sensory difference value $D^{(i)}$ and a prior that is given by the difference estimate from the previous stroke $\hat{D}^{(i)}$ (= posterior estimate after *previous* stroke) (Shadmehr & Mussa-Ivaldi, 2012):

$$\hat{D}^{(i+1)} = \hat{D}^{(i)} + k^{(i)}(D^{(i)} - \hat{D}^{(i)}) \tag{7}$$

That is, the present difference estimate $\hat{D}^{(i+1)}$ is the previous estimate $\hat{D}^{(i)}$ plus the prediction error of the previous estimate $(D^{(i)} - \hat{D}^{(i)})$ weighted by the Kalman gain $k^{(i)}$. The Kalman gain describes the ratio between the prior variance ($p^{(i-1)}$) and the sensory variance ($\sigma_{D^{(i)}}^2$). For determining the Kalman gain, it is important to consider that our model is based on multiple comparisons with the first stimulus estimate and thus the first stimulus estimate is included in the

computation of each difference estimate. The resultant covariance between prior and sensory estimate has to be taken into account (cf. Oruç et al., 2003, Eqs. 5 & 7):

$$k^{(i)} = \frac{p^{(i-1)} - cov(D^{(i)}, \hat{D}^{(i)})}{p^{(i-1)} + \sigma_{D^{(i)}}^2 - 2cov(D^{(i)}, \hat{D}^{(i)})} \tag{8}$$

$$\text{with } p^{(i-1)} = \frac{p^{(i-1|i-2)} * \sigma_{D^{(i-1)}}^2 - cov(D^{(i-1)}, \hat{D}^{(i-1)})^2}{p^{(i-1|i-2)} + \sigma_{D^{(i-1)}}^2 - 2cov(D^{(i-1)}, \hat{D}^{(i-1)})} \tag{9}$$

In our model, sensory variance of the difference value ($\sigma_{D^{(i)}}^2$) is the sum of the variance of a one-stroke based estimate ($\sigma_{N=1}^2$) and the variance of the first stimulus estimate ($\sigma_{S1N=j}^2$) modified by memory decay. The variance of the one-stroke based estimate ($\sigma_{N=1}^2$) was estimated from the corresponding JND in Experiment 1 (considering the transformation from 75% to 84% discrimination thresholds). The variance of the first stimulus overall estimate was estimated by $\sigma_{S1N=j}^2 = \sigma_0^2/j$ (Eq. 2, right), i.e., from the MLE prediction on overall variance as a function of number of strokes ($N=j$) and one-stroke based variance; it was therefore lower the more strokes over the first stimulus were performed (e.g. $\sigma_{S1N=4}^2 > \sigma_{S1N=5}^2$).

Additionally, an effect of memory decay was modelled for the variance of the first stimulus estimate ($\sigma_{S1N=j}^2$). The rate of the decrease due to memory decay is usually described by a power function of the time t with a negative exponent (Wixted & Ebbesen, 1991, 1997). Murray, Ward, and Hockley (1975) reported such a power function for an experiment that resembles the present one: Two-point thresholds T at the thumb increased with the prolongation of the time interval t (in sec) between the first and the second touch by $T=2.303t^{0.221}$. The change in thresholds can be directly linked to change in the variance of the individual measurements ($\sigma^2 = \frac{1}{2}T^2$; assuming uncorrelated errors). We modelled memory decay for the variance of the first stimulus estimate as a function of number of strokes over the second stimulus (i) with the exponent taken from Murray et al. (1975): $\sigma_{S1N=j(i)}^2 = \sigma_{S1N=j}^2 * t^{0.442}$. Assuming that the prior for the first stroke on the second stimulus is non-informative (variance set to infinite), we predicted weights of each stroke-specific estimate in the second stimulus.

Participants

Fifteen right-handers, naïve to the purpose of the experiment, were in the final sample (mean age: 25.4 years, range: 20-36

years; 10 females, 5 males). Four subjects had to be excluded from analyses because of problems with the task according to criteria described for Experiment 2.

Apparatus, Stimuli, and Procedure

The Apparatus, the configuration of the grating stimuli and the procedure in single trials were identical to those in Experiment 2.

Design and Data Analyses

In addition to *Stimulus order* and *Position*, we varied the *Number of strokes*. Participants applied 2, 3, 4, or 5 strokes per stimulus. As in Experiment 2, we measured PSEs and JNDs using the method of constant stimuli. Each standard was compared 10 times to each of 14 comparison gratings. In the present experiment, we also analyzed the movement force and velocities used in each condition.

Table 1 gives an overview of the 28 possible combinations of *Number of strokes*, the *Stimulus order*, and the *Position*. Two types of standards were used: a period of 4.5 mm could be the discrepant or the dominant information, and a period of 6 mm assumed the other role. Twenty-four standard stimuli corresponded to the conditions with more than two strokes, each of which was either presented as first or second stimulus. However, for the two-stroke condition one standard operationalized two different conditions, depending on which information is defined as being dominant. One of the two-stroke standards can be interpreted both as a 4.5-mm dominant stimulus with discrepant information in the second stroke and as a 6-mm dominant stimulus with discrepant info in the first stroke; for the other two-stroke standard it is vice versa. That is, the two-stroke conditions are operationalized by only two standard stimuli, either presented as first or second stimulus. Overall, the experiment consisted of 3,640 trials = (24 + 2) [standards] x 2 [stimulus order] x 14 [comparisons] x 5 [repetitions] divided into 5 sessions, each lasting 2.5–3 hours. The first sessions started with a training of finger force and velocity similar to Exp. 2.

Table 1 Experiment 3: Overview of experimental conditions. Each condition was defined by the *Number of Strokes*, the *Stimulus order*, and the *Position*

Number of strokes	Stimulus order	Position of the discrepant stroke					
2	1	2	1	2			
3	1	2	1	2	3		
4	1	2	1	2	3	4	
5	1	2	1	2	3	4	5

We determined individual psychometric functions for each standard and in each experimental condition. As in Experiment 2, we calculated weights of the discrepant stroke by taking into account the average PSEs measured in the control conditions of Experiment 2 and restricting individual weights to be within 2 standard deviations from the mean (22 outliers in 420 cases). The individual weights were analyzed by linear contrast analyses over positions separated by *Number of strokes* and *Stimulus order* conditions. We expected that weights for the discrepant stroke in the second stimulus, but not in the first stimulus, systematically decrease with *Position*.

Results

Movement Parameters Velocity and Force

To check for potential confounds in weight assessment, we tested whether participants systematically varied exploratory force or velocity during the exploration of a stimulus. On average, 95% of the movement data of each participant could be used for this analysis. For each *Number of strokes* condition, we calculated a separate ANOVA with the within-participant variables stroke *Position* and *Stimulus order*. We did not find any significant effect of *Stimulus order* on movement force (2 strokes: $F(1,14) = 0.636, p = 0.438$; 3 strokes: $F(1,14) = 0.811, p = 0.383$; 4 strokes: $F(1,14) = 0.014, p = 0.907$; 5 strokes: $F(1,14) = 2.732, p = 0.121$; if necessary, p value corrected according to Greenhouse and Geisser (1959)) nor on movement velocity (2 strokes: $F(1,14) = 1.150, p = 0.241$; 3 strokes: $F(1,14) = 1.161, p = 0.694$; 4 strokes: $F(1,14) = 1.029, p = 0.328$; 5 strokes: $F(1,14) = 1.698, p = 0.214$). Also, there was no significant interaction: *Position* x *Stimulus order* (force: 2 str.: $F(1,14) = 1.902, p = 0.190$; 3 str.: $F(2,28) = 2.839, p = 0.113$; 4 str.: $F(3,42) = 0.653, p = 0.472$; 5 str.: $F(4,56) = 1.211, p = 0.312$; velocity: 2 str.: $F(1,14) = 0.000, p = 0.992$; 3 str.: $F(2,28) = 1.591, p = 0.228$; 4 str.: $F(3,42) = 1.724, p = 0.177$; 5 str.: $F(4,56) = 1.086, p = 0.372$), indicating that differences between the first and the second stimulus in the pattern of stroke-specific weights cannot be due to movement variation.

However, a main effect of *Position* can be found for each *Number of strokes* for velocity (2 str.: $F(1,14) = 7.143, p = 0.018$; 3 str.: $F(2,28) = 5.827, p = 0.024$; 4 str.: $F(3,42) = 13.924, p < 0.001$; 5 str.: $F(4,56) = 10.633, p = 0.001$) and force (2 str.: $F(1,14) = 27.987, p < 0.001$; 3 str.: $F(2,28) = 13.234, p = 0.001$; 4 str.: $F(3,42) = 10.487, p = 0.001$; 5 str.: $F(4,56) = 7.584, p = 0.003$).

Weights of Discrepant Information

The detailed results of the linear contrast analyses of the individual weights are shown in Table 2 and Fig. 5. Analyses were

Table 2 Experiment 3: Linear trend analysis of the *Position* effect separately for each combination of *Number of strokes* and *Stimulus order* conditions

Number of strokes	Stimulus order	Standard ...	<i>F</i>	df	<i>p</i>
2	First		6.213	1,14	0.026
	Second		0.065	1,14	0.401
3	First		0.022	1,14	0.885
	Second		1.892	1,14	0.096
4	First		0.281	1,14	0.604
	Second		4.642	1,14	0.025
5	First		0.014	1,14	0.909
	Second		5.461	1,14	0.018

two-tailed for the first stimulus and one-tailed for the second, because we expected a position effect only for the second stimulus. As expected for the first stimulus, we did not observe significant linear effects of position on the weights in most conditions, except for an increase in the two-stroke condition. For the second stimulus, we observed the expected significant linear decrease of weights in the 4- and 5-stroke conditions, and for the 3-stroke condition we observed a corresponding trend. Together, these data replicate and extend the findings of

Experiment 2. Both experiments offer support for the idea of a different processing for the first and the second stimulus.

Model Data vs. Empirical Data

In Fig. 6, we compare model predictions on the weights with empirical data for the second stimulus. For each combination of *Number of strokes* and *Position* conditions we calculate *t* tests between the empirical weights and the predicted value. As is the case for the predicted weights, empirical weights were normalized for each *Number of strokes* separately so that averages across all positions sum up to a value of 1. For 13 of 14 conditions, the predicted and the measured weights did not differ significantly, $t(14) \leq |1.374|$, $p \geq 0.191$. Only in the second stroke of the 3-stroke conditions, $t(14) = 3.722$, $p = 0.002$, the empirical weight was higher than expected. Predicted values explained $r^2 = 0.83$ of the empirical variance between conditions, $p < 0.001$. Overall, empirical data followed model predictions.

JNDs

We averaged the log-transformed (base 10) JNDs across all *Position* and *Stimulus order* conditions with the same

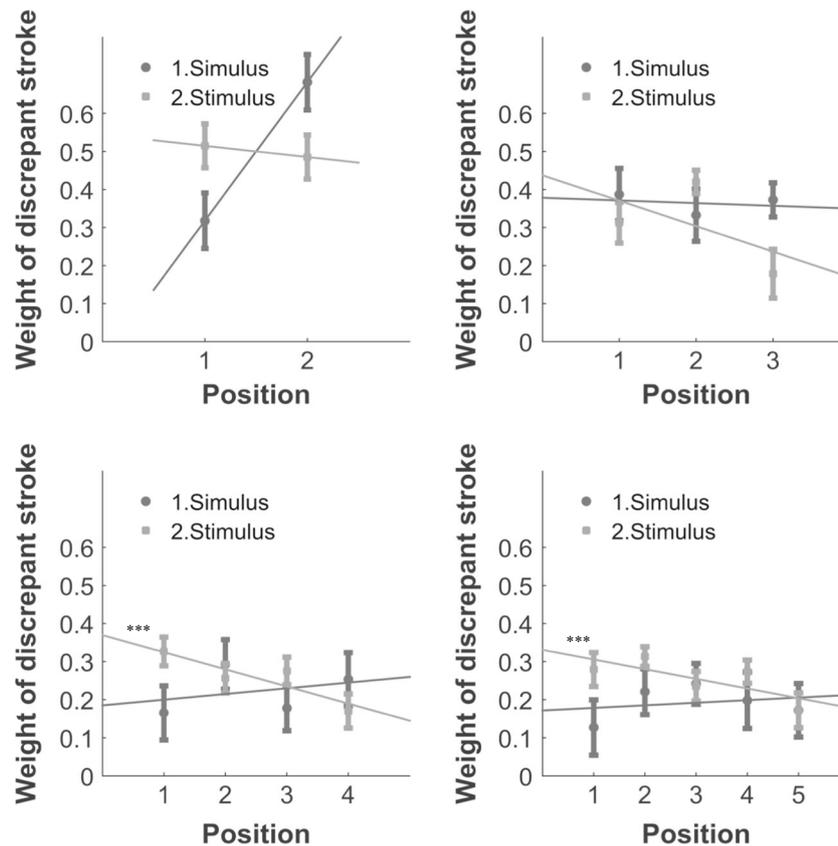


Fig. 5 Exp. 3: Average estimated weights of the discrepant stroke and standard error as a function of *Stimulus order* (first vs. second stimulus) and *Position* within the standard, plotted separately for all *Number of strokes* conditions

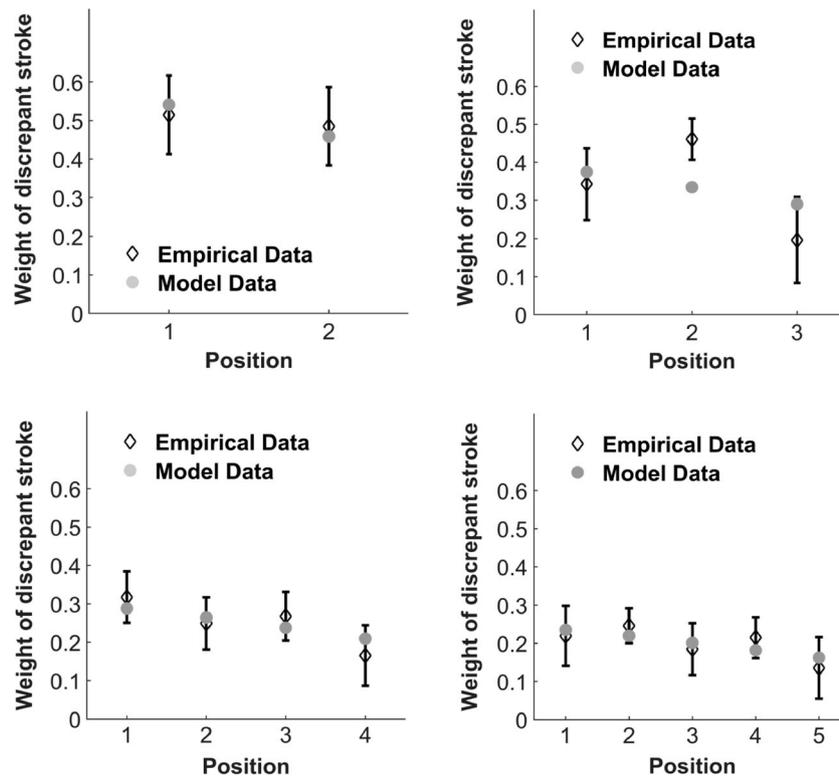


Fig. 6 Exp. 3: Average empirical (plotted with 95% confidence intervals) vs. predicted weights for the discrepant stroke in the second stimulus depending on its *Position* within the standard, plotted separately for all *Number of strokes* conditions

number of strokes (Fig. 7). The log JND values decreased with an increasing number of strokes in a linear contrast analysis, $F(1,14) = 4.161, p = 0.031$ (one-tailed). The regression of log JND on log *Number of strokes* explained $r^2 = 0.72$ of the data. The slope of the regression line is -0.146 . As in Experiment 1, this slope not in line with MLE predictions, in that it is significantly different from $-0.5, t(14) = 4.501, p < 0.001$.

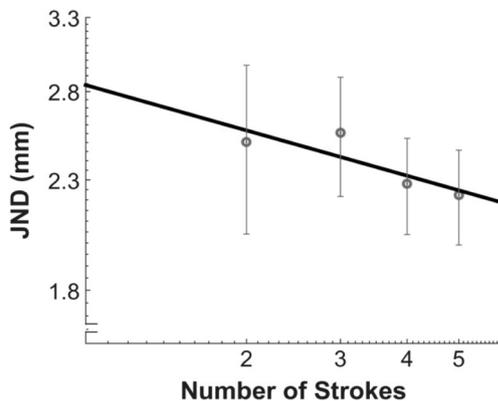


Fig. 7 Exp. 3: Log-log plot. Log average JNDs and standard errors as a function of log *Number of strokes* collapsed across all *Position* and *Stimulus order* conditions

Discussion Experiment 3

Experiment 3 replicated and extended the results of Experiment 2 by including different exploration lengths (*Number of strokes*) and comparing the results to model predictions. As in Experiment 2, we found evidence for a different processing of information from the first and the second stimulus. While stroke-specific estimates were rather equally weighted for all strokes across the first stimulus, weights decreased with the position of the stroke in the second stimulus. Predictions from a Kalman filter model of a comparison process that is affected by memory decay fit the weight data well. The model has no free parameter; the rate of memory decay was estimated from a previous study (Murray et al., 1975). We conclude that in discrimination tasks on sequentially gathered information memory decay affects the comparison process.

General Discussion

The present study addressed the integration of redundant texture signals from sequentially sampled strokes. The integration of simultaneously presented, redundant signals had been successfully described by the MLE model of optimal

integration. As expected, the present results show that the assumptions underlying this simple model do not describe the integration of sequentially presented texture information. The MLE assumptions predict a specific rate with which discrimination thresholds decrease with a prolonged exploration over the textures, and it predicts that equally reliable estimates should contribute equally to the percept. We found lower rates of threshold decrease as predicted by the MLE assumptions (Exp. 1 & 3) and unequal weights of estimates from different strokes (Exp. 2 & 3). However, the data can be explained by an extended model of an optimal observer that we had derived from the literature (Romo et al., 2002; Romo & Salinas, 2003; Hernández et al., 2010; Metzger et al., 2017; Kalman, 1960). We state that the two stimuli in a trial, when presented sequentially, are not processed in the same way. Information from the first stimulus is integrated in a MLE fashion (with equal weights) into a final estimate. We speculate that the final estimate is transferred to a different structure where it is stored in memory. This memorized estimate from the first stimulus gets noisier over time. Due to this circumstance, information from the second stimulus is processed differently. For each exploration segment of the second stimulus, a comparison process between the overall first stimulus estimate and the segment-specific second stimulus estimate is performed. The model predicts that the information coming from different strokes of the first stimulus are integrated with equal weights, whereas segment-specific weights should systematically decrease over time for the second stimulus. Because the comparison process is affected by memory decay and not the integration process, the empirical weights assessed for various exploration lengths are in line with this prediction (Exp. 2 & 3). A Bayesian-type Kalman filter model of the process, which uses a literature-based rough estimation of the memory decay and has no free parameter, can quantitatively predict the weights assessed in Experiment 3. Our experiments help to better understand the haptic integration of signals over time. Optimality, in the sense of seeking the lowest variance of the final percept is still the goal of our system. However, more complicated system properties, such as memory, need to be taken into account to describe sequential compared with simultaneous integration processes.

Our result might be surprising given the fact that recent studies on visual perception did find hints for MLE integration of sequential information. For instance, Wolf and Schütz (2015) reported close to MLE-optimal trans-saccadic integration of information. The authors compared weights of presaccadic, peripheral and postsaccadic, foveal signals with predictions of the MLE model. One reason why MLE predicted integration might occur in this case, but not in our study, is the task itself. In the study by Wolf and Schütz (2015) participants had to indicate whether the vertical component of a plaid stimulus was tilted clockwise or counterclockwise. Thus, in contrast to comparing a memorized first stimulus to a sequentially

experienced second stimulus, participants compared one sequentially experienced stimulus to a fixed reference. This task, consequently, did not include memory transfer and storage of a representation of the reference's perceptual estimate, which possibly decays over time. These results may be similar to the integration of information in the first stimulus in a trial within our experiments. That is, the tasks in Wolf and Schütz' (2015) study required a single overall estimate of the stimulus, rather than comparing sequentially gathered information from a (second) stimulus to a memorized and therefore decaying reference.

Other studies do provide examples for perceptual optimization under conditions of memory decay. A recent study on the comparison between a memorized reference stimulus and a comparison stimulus showed that a Bayesian model that includes memory decay can explain the so-called contraction bias (Ashourian & Loewenstein, 2011). In a delayed comparison task, participants compared the visual length of two bars, the first of which was memorized. Participants tended to report the size of the memorized bar to be closer to the overall mean of the used sample of bars than the size of the second bar (= contraction bias). The authors suggest from their data a Bayesian model of optimal processing in that the sample of overall used bars provides a prior for the judgment on the memorized size of the first stimulus, and in that this prior gets weighted higher the more the memorized stimulus representation is affected by memory decay. Their conclusions are in good agreement with our model, in which memory decay is as well assumed to add variance to the memorized representation of the first stimulus. Similarly, in the field of color vision, Olkkonen, McCarthy, and Allred (2014) reported a central tendency bias in a delayed color estimation task, which also was modelled by a Bayesian model including memory decay. In a similar manner, Fassihi, Akrami, Esmaeili, and Diamond (2014) were able to explain performance of humans and rats in a tactile working memory task.

The task determines which factors need to be considered to achieve optimal perceptual estimates. For some tasks, the assumptions underlying the MLE model are sufficient; however, in sequential comparison tasks memory decay needs to be taken into account. Additionally, other factors may play a role; Fischer and Whitney (2014) recently suggested that visual perception is "serially dependent" in the sense that it uses both prior information and the present sensory input to inform perception at the present moment. Interestingly, the authors showed in their data that attention is able to modulate the impact of the prior information. Future research focusing on haptic sequential integration also may include attention as a potentially modifying factor.

Our proposed model of a comparison process that is affected by memory decay has interesting implications on how

participants should ideally explore texture stimuli in a discrimination task when they are less constrained in their exploratory behaviour. Yet, Wismeijer, Erkelens, van Ee, and Wexler (2010) described that sensory estimates as predicted by an optimal observer model can predict subsequent visual exploration movements. It has been argued that movements performed in free exploration are aimed to optimize the gathering of sensory information and to enhance task performance (Kaim & Drewing 2011). Given the proposed model, certain exploration strategies should lead to more precise discrimination than other strategies and therefore should be more preferentially performed by the observers. For instance, when participants are free to choose how often they stroke across each of two successively explored texture stimuli, the model would predict that more strokes are conducted across the first than across the second stimulus. The reason is that memory decay is assumed to take place only for the first stimulus estimate during the exploration of the second stimulus. As a consequence, benefit from additional strokes across the second stimulus is counteracted by memory decay, but not benefit from additional strokes across the first stimulus. To give another example, in completely free exploration, participants might prefer to go back to the first stimulus, when its memory traces gets too noisy and frequent changes between the two stimuli can be expected. It would be interesting to address those points in further research.

It is further noteworthy that in Experiment 2 the control conditions revealed considerable biases in the perception of the 4.5-mm period stimulus, when it was presented as the second one. As argued previously, we used the average measured values in the control conditions to calculate the weights of discrepant information and therefore the biases did not affect our predictions on the weights. However, we ask why this bias might have occurred. Karim, Harris, Morley, and Breakspear (2012) described that when participants discriminate two vibrotactile stimuli they perform better when the first stimulus lies between the global mean of all stimuli and the second stimulus. This is known as the “time-order effect” (Karim et al., 2012; Preuschhof, Schubert, Villringer, & Heekeren, 2010). It was suggested that the reason for this observation is a “drift” of neural responses for the first stimulus towards the global mean. We speculate that this effect in combination with a stimulus range effect causes the biases we observe in the control conditions. While we chose an equal spacing of periods between 2 mm and 8 mm, this could be a perceptually not completely symmetric space. If you assume that the standard with the 6-mm period is closer to the perceived global mean than the standard with 4.5-mm, “time-order” effects might explain biases in the perception of the 4.5-mm standard. Interestingly, even this perceptual bias hints to the same conclusion that we draw from our main results. That is, in a task of comparing two sequentially presented stimuli the first and the second stimulus are not processed in the same way.

Conclusions

This study asked the fundamental question of how information is integrated over time in the mainly sequentially working haptic sense. Our results show that spatiotemporal integration does take place within haptic perception. However, gains from this integration were lower than we predicted by an optimal integrator model (MLE), which is usually applied to the integration of simultaneously presented information. A closer investigation of the integration process revealed that the processing in our sequential task is likely to be more complex. The perceptual system we describe takes the loss of information due to memory decay into account and counterbalances such decay with weighting this information less over time. We suggest a Bayesian model to describe the perceptual process, which focuses on comparing the two stimuli online to produce the least noisy estimate of the difference between the two stimuli.

Acknowledgments The authors thank Manolo Hinkel and Tarek Huzail for part of the data collection. Supported by Deutsche Forschungsgemeinschaft (DR 730/1-2, FOR 560, and SFB/TRR135/1, A05). Part of the data of Experiment I and III have been pre-published in conference papers (Drewing, Lezkan & Ludwig, 2011; Lezkan & Drewing, 2014). Response data of individual observers from all experiments presented is available at [10.5281/zenodo.826653].

References

- Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology*, *14*(3), 257–262.
- Ashourian, P., & Loewenstein, Y. (2011). Bayesian inference underlies the contraction bias in delayed comparison tasks. *PLoS One*, *6*(5), e19551.
- Battaglia, P. W., Jacobs, R. A., & Aslin, R. N. (2003). Bayesian integration of visual and auditory signals for spatial localization. *JOSA A*, *20*(7), 1391–1397.
- Drewing, K., & Ernst, M. O. (2006). Integration of force and position cues for shape perception through active touch. *Brain Research*, *1078*(1), 92–100.
- Drewing, K., Lezkan, A., & Ludwig, S. (2011). Texture discrimination in active touch: Effects of the extension of the exploration and their exploitation. In: *World Haptics Conference (WHC)*, 2011 I.E. (pp. 215–220). IEEE.
- Drewing, K., Wiecki, T.V., & Ernst, M.O. (2008). Material properties determine how force and position signals combine in haptic shape perception. *Acta Psychologica*, *128*(2), 264–273.
- Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429–433.
- Ernst, M. O., & Bühlhoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences*, *8*(4), 162–169.
- Fassihi, A., Akrami, A., Esmaeili, V., & Diamond, M. E. (2014). Tactile perception and working memory in rats and humans. *Proceedings of the National Academy of Sciences*, *111*(6), 2331–2336.
- Fetsch, C. R., DeAngelis, G. C., & Angelaki, D. E. (2013). Bridging the gap between theories of sensory cue integration and the physiology of multisensory neurons. *Nature Reviews Neuroscience*, *14*(6), 429–442.
- Fischer, J., & Whitney, D. (2014). Serial dependence in visual perception. *Nature Neuroscience*, *17*(5), 738–743.

- Funamizu, A., Kuhn, B., & Doya, K. (2016). Neural substrate of dynamic Bayesian inference in the cerebral cortex. *Nature Neuroscience*, 19, 1682–1689.
- Gescheider, G.A., Berryhill, M.E., Verrillo, R. T., & Bolanowski, S.J. (1999). Vibrotactile temporal summation: Probability summation or neural integration?. *Somatosensory & Motor Research*, 16(3), 229–242.
- Gescheider, G. A., Bolanowski, S. J., Pope, J. V., & Verrillo, R. T. (2002). A four-channel analysis of the tactile sensitivity of the fingertip: Frequency selectivity, spatial summation, and temporal summation. *Somatosensory & Motor Research*, 19(2), 114–124.
- Gescheider, G. A., Güçlü, B., Sexton, J. L., Karalunas, S., & Fontana, A. (2005). Spatial summation in the tactile sensory system: Probability summation and neural integration. *Somatosensory & Motor Research*, 22(4), 255–268.
- Giachritsis, C. D., Wing, A. M., & Lovell, P. G. (2009). The role of spatial integration in the perception of surface orientation with active touch. *Attention, Perception, & Psychophysics*, 71(7), 1628–1640.
- Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, 24, 95–112.
- Helbig, H. B., Ernst, M. O., Ricciardi, E., Pietrini, P., Thielscher, A., Mayer, K. M., & Noppeney, U. (2012). The neural mechanisms of reliability weighted integration of shape information from vision and touch. *Neuroimage*, 60(2), 1063–1072.
- Heller, M. A. (1982). Visual and tactual texture perception: Intersensory cooperation. *Perception & Psychophysics*, 31(4), 339–344.
- Heller, M. A. (1989). Texture perception in sighted and blind observers. *Perception & Psychophysics*, 45(1), 49–54.
- Henriques, D. Y., & Soechting, J. F. (2005). Approaches to the study of haptic sensing. *Journal of Neurophysiology*, 93(6), 3036–3043.
- Hernández, A., Nácher, V., Luna, R., Zainos, A., Lemus, L., Alvarez, M., & Romo, R. (2010). Decoding a perceptual decision process across cortex. *Neuron*, 66(2), 300–314.
- Hollins, M., Bensmaïa, S. J., & Washburn, S., (2001). Vibrotactile adaptation impairs discrimination of fine, but not coarse, textures. *Somatosensory & Motor Research*, 18(4), 253–262.
- Jacobs, R. A. (2002). What determines visual cue reliability? *Trends in Cognitive Sciences*, 6(8), 345–350.
- Jovanovic, B., & Drewing, K. (2014). The influence of intersensory discrepancy on visuo-haptic integration is similar in 6-year-old children and adults,” *Frontiers in Psychology*, 5(57), 1–11.
- Kaim, L. & Drewing, K. (2011). Exploratory strategies in haptic softness discrimination are tuned to achieve high levels of task performance. *IEEE Transactions on Haptics* 4. 242–252.
- Kalman, R. E. (1960). A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering*, 82(1), 35–45.
- Kappers, A.M.L., & Bergmann Tiest, W.M., (2015). Aftereffects in touch. *Scholarpedia*, 10(3),32730.
- Karim, M., Harris, J. A., Morley, J. W., & Breakspear, M. (2012). Prior and present evidence: How prior experience interacts with present information in a perceptual decision making task. *PloS one*, 7(5), e37580.
- Knill, D. C., & Pouget, A. (2004). The Bayesian brain: The role of uncertainty in neural coding and computation. *TRENDS in Neurosciences*, 27(12), 712–719.
- Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modeling of depth cue combination: In defense of weak fusion. *Vision research*, 35(3), 389–412.
- Lederman, S.J. & Klatzky, R.L. (1987). Hand Movements: A window into haptic object recognition. *Cognitive Psychology*, 19, 342–368.
- Lederman, S. J., Loomis, J. M., & Williams, D. A. (1982). The role of vibration in the tactual perception of roughness. *Perception & Psychophysics*, 32(2), 109–116.
- Lezkan, A. & Drewing, K. (2014). Unequal - but fair? Weights in the serial integration of haptic texture information. *Haptics: Neuroscience, Devices, Modeling, and Applications* (pp. 386–392). Springer: Heidelberg.
- Lezkan, A. Manuel, S.G. Colgate, J.E., Klatzky, R.L., Peshkin, M.A. & Drewing, K. (2016). Multiple Fingers – One Gestalt. *IEEE Transactions on Haptics* 9(2), 255–266.
- Lieberman, H. R., & Pentland, A. P. (1982). Microcomputer-based estimation of psychophysical thresholds: The best PEST. *Behavior Research Methods & Instrumentation*, 14(1), 21–25.
- Louw, S., Kappers, A. M., & Koenderink, J. J. (2005). Haptic detection of sine-wave gratings. *Perception*, 34(7), 869–885.
- Metzger, A., Lezkan, A., & Drewing, K. (2017). Integration of serial sensory information in haptic perception of softness. *Submitted to Journal of Experimental Psychology: Human Perception and performance*.
- Murray, D. J., Ward, R., & Hockley, W. E. (1975). Tactile short-term memory in relation to the two-point threshold. *The Quarterly Journal of Experimental Psychology*, 27(2), 303–312.
- Olkkonen, M., McCarthy, P. F., & Allred, S. R. (2014). The central tendency bias in color perception: Effects of internal and external noise. *Journal of Vision*, 14(11):5, 1–15.
- Oruç, I., Maloney, L. T., & Landy, M. S. (2003). Weighted linear cue combination with possibly correlated error. *Vision Research*, 43(23), 2451–2468.
- Preuschhof, C., Schubert, T., Villringer, A., & Heekeren, H. R. (2010). Prior Information biases stimulus representations during vibrotactile decision making. *Journal of Cognitive Neuroscience*, 22(5), 875–887.
- Quick Jr, R. F. (1974). A vector-magnitude model of contrast detection. *Kybernetik*, 16(2), 65–67.
- Romo, R., Hernández, A., Zainos, A., Lemus, L., & Brody, C. D. (2002). Neuronal correlates of decision-making in secondary somatosensory cortex. *Nature Neuroscience*, 5(11), 1217–1225.
- Romo, R. & Salinas, E. (2003). Flutter discrimination: Neural codes, perception, memory and decision making. *Nature Rev. Neuroscience*, 4, 203–218.
- Rosas, P., Wichmann, F. A., & Wagemans, J. (2007). Texture and object motion in slant discrimination: Failure of reliability-based weighting of cues may be evidence for strong fusion. *Journal of Vision*, 7(6):3, 1–21.
- Shadmehr, R., & Mussa-Ivaldi, S. (2012). *Biological learning and control: How the brain builds representations, predicts events, and makes decisions*. MIT Press.
- Thompson, P., & Burr, D. (2009). Visual aftereffects. *Current Biology*, 19(1), 11–14.
- Wichmann, F. A., & Hill, N. J. (2001). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics*, 63, 1293–1313.
- Wismejer, D. A., Erkelens, C. J., van Ee, R., & Wexler, M. (2010). Depth cue combination in spontaneous eye movements. *Journal of Vision*, 10(6):25, 1–15.
- Wixted, J. T., & Ebbesen, E. B. (1991). On the form of forgetting. *Psychological Science*, 2(6), 409–415.
- Wixted, J. T., & Ebbesen, E. B. (1997). Genuine power curves in forgetting: A quantitative analysis of individual subject forgetting functions. *Memory & Cognition*, 25(5), 731–739.
- Wolf, C., & Schütz, A. C. (2015). Trans-saccadic integration of peripheral and foveal feature information is close to optimal. *Journal of Vision*, 15(16):1, 1–18.