

Template optimization and transfer in perceptual learning

Ilmari Kurki

Institute of Behavioural Sciences, University of Helsinki,
Helsinki, Finland



Aapo Hyvärinen

Department of Computer Science, University of Helsinki,
Helsinki, Finland

Jussi Saarinen

Institute of Behavioural Sciences, University of Helsinki,
Helsinki, Finland

We studied how learning changes the processing of a low-level Gabor stimulus, using a classification-image method (psychophysical reverse correlation) and a task where observers discriminated between slight differences in the phase (relative alignment) of a target Gabor in visual noise. The method estimates the internal “template” that describes how the visual system weights the input information for decisions. One popular idea has been that learning makes the template more like an ideal Bayesian weighting; however, the evidence has been indirect. We used a new regression technique to directly estimate the template weight change and to test whether the direction of reweighting is significantly different from an optimal learning strategy. The subjects trained the task for six daily sessions, and we tested the transfer of training to a target in an orthogonal orientation. Strong learning and partial transfer were observed. We tested whether task precision (difficulty) had an effect on template change and transfer: Observers trained in either a high-precision (small, 60° phase difference) or a low-precision task (180°). Task precision did not have an effect on the amount of template change or transfer, suggesting that task precision per se does not determine whether learning generalizes. Classification images show that training made observers use more task-relevant features and unlearn some irrelevant features. The transfer templates resembled partially optimized versions of templates in training sessions. The template change direction resembles ideal learning significantly but not completely. The amount of template change was highly correlated with the amount of learning.

1980), vernier acuity (Fahle & Morgan, 1996; Saarinen & Levi, 1995), and texture segmentation (Ahissar & Hochstein, 1993) is known to substantially improve perceptual performance. Evidence from psychophysical studies with external noise shows that the most important mechanism by which perceptual learning operates is improved sampling of the stimulus information in the visual system (Doshier & Lu, 1998; Gold, Sekuler, & Bennett, 2004; Li, Klein, & Levi, 2008; Li, Levi, & Klein, 2004). The set of internal weights that the visual system associates with input-stimulus features is referred to as the perceptual template. According to this idea, practice changes the template so that more weight will be placed on input-stimulus features that allow efficient discrimination of the practiced stimuli, and less weight on nonefficient features. A theoretical upper limit to this optimization can be analyzed using statistical decision theory, yielding a Bayesian ideal-observer template with maximal performance to discriminate the stimulus in noise (Geisler, 2011; Green & Swets, 1966).

The classification-image method (Beard & Ahumada, 1998; Eckstein & Ahumada, 2002; Murray, 2011) is a psychophysical technique that uses a statistical approach to directly estimate the perceptual template. In a classification-image experiment, low-contrast, near-threshold stimuli are masked with pseudorandom noise. The contrast is adjusted so that the observer makes both correct and incorrect perceptual decisions (e.g., false alarms, reporting that the target was present when only the noise was given). Classification-image estimation is based on trial-to-trial correlation between the noise-stimulus values and the corresponding observer’s responses. The basic idea is first to factor out the effect of the target-stimulus presence. The unknown template weights are then estimated, by regressing the known stimulus-noise values with the observed re-

Introduction

Classification images and perceptual learning

Extensive practice in low-level perceptual tasks such as orientation discrimination (Fiorentini & Nicoletta,

Citation: Kurki, I., Hyvärinen, A., & Saarinen, J. (2016). Template optimization and transfer in perceptual learning. *Journal of Vision*, 16(10):16, 1–18, doi:10.1167/16.10.16.



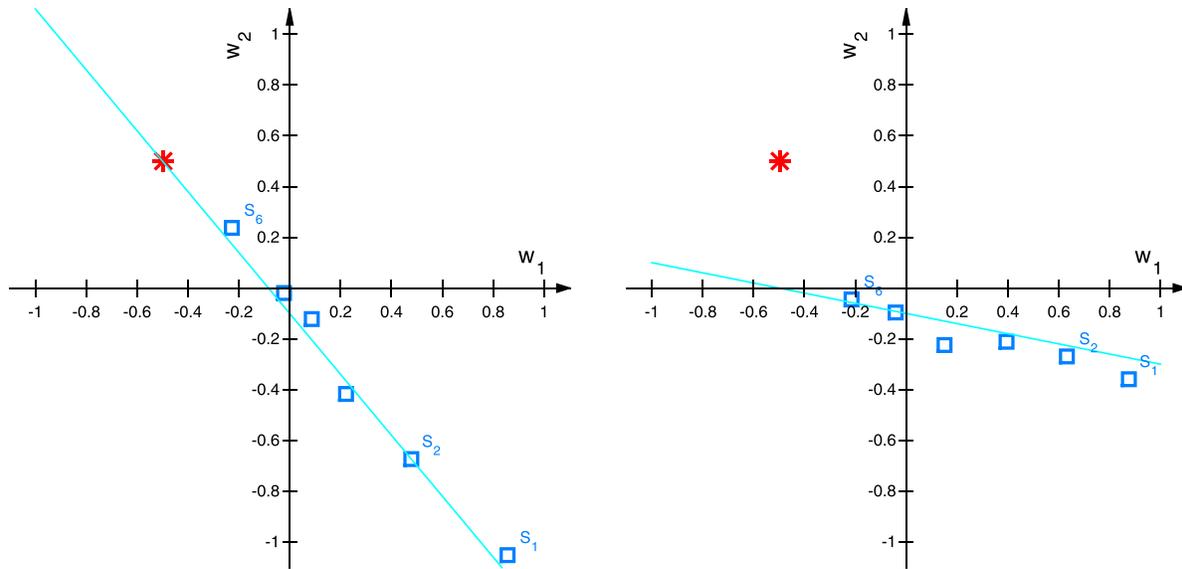


Figure 1. Illustration of perceptual template change in learning. The simplified example stimulus space consists of just two perceptual-template weights (w_1 and w_2). The ideal template (red asterisk) has identical absolute weights in w_1 and w_2 but an opposite sign. Templates for the training sessions (s_1, s_2, \dots, s_6) are marked by cyan squares. To infer the direction of the template change (cyan lines) and extrapolate whether the template change leads to an ideal template (ideal direction), we fitted a linear regression to the templates. In the left panel, learning changes the template weights to an ideal direction. In the right panel, learning changes the template to a direction that is not ideal but correlates with it. The rate of template change is constant in the right panel and saturating in the left panel.

sponses. The resulting template estimate reveals how information at each stimulus location is weighted in decisions.

The classification-image method was long considered to be impractical for studying learning. A standard two-dimensional classification-image experiment with two-dimensional pixel noise requires the estimation of thousands of weights (for every pixel in the stimulus), requiring thousands of trials and several hours of experiment. However, by reducing the dimensionality of the stimulus, it is possible to make variants that have enough statistical power to estimate the template reliably in a single session and to track template changes from one session to the next. For example, in a “position noise” paradigm, the stimulus configurations (for example, lines) are composed from a number of local elements such as the difference of Gaussian blobs whose relative position is varied (Li et al., 2004). These studies have shown that for some stimuli, perceptual learning can expand the perceptual-template area (Dobres & Seitz, 2010; Kurki & Eckstein, 2014; Li et al., 2004). Initially, the template weights only the elements that have the highest signal-to-noise ratio, either because of their location in the visual field (fovea) or because of the stimulus design. Learning later expands the template to weight also elements with a lower signal-to-noise ratio (in the periphery; Kurki & Eckstein, 2014; Li et al., 2004).

Template change in perceptual learning

Here we introduce a method to estimate and visualize the template change in learning, yielding a “template change vector” that describes how perceptual weights—i.e., different parts of the classification image—are updated in the course of learning. We then directly test the idea that optimization by learning changes the template toward an ideal-observer template. Previous studies have shown that the match between the estimated template and an ideal template increases in the course of learning, and predicts the amount of perceptual performance increase (Kurki & Eckstein, 2014; Li et al., 2004). However, this measure is not very informative about how the template changes in learning. Especially because sampling efficiency is generally very low, it is possible that learning changes the template in a direction that merely correlates weakly with the ideal template (see Figure 1). The template change vector will give more information on how learning changes the processing of stimulus features. For example, we can test whether the processing of certain features benefits more from learning than others; this can reveal differences in the plasticity of neural mechanisms. Comparison with an ideal observer is particularly beneficial here, as it shows the optimal direction of change on the basis of stimulus information. Such analysis of template change will also allow for more stringent testing of computational

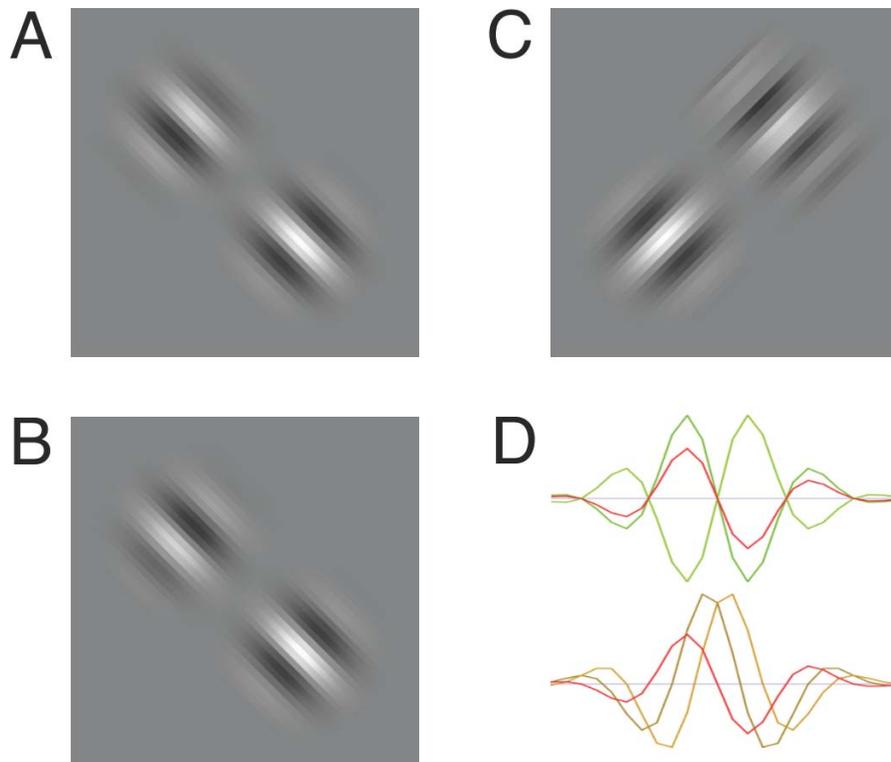


Figure 2. Stimuli. Stimuli were Gabor patches; the phase of the upper (target) Gabor was varied, whereas the lower (reference) Gabor always had the same 0° phase. Contrast threshold for discrimination was measured by varying the contrast of the upper Gabor, using a one-interval task. (A) Low-precision condition with a $+90^\circ$ phase angle ϕ . (B) Low-precision condition with -90° phase angle. (C) High-precision condition with $+30^\circ$ phase angle; one-dimensional noise (lines with random contrast) was added to the stimulus. (D) Cross-section of two possible stimuli in low-precision (top, green profiles) and high-precision (bottom, brown profiles) conditions. Both conditions had the same ideal observer (ideal sampling strategy), shown by the red profile. The global orientation of the stimuli was held constant in the training sessions at either -45° (A, B) or $+45^\circ$ (C), then rotated by 90° in the transfer session.

models of perceptual-learning dynamics. Lastly, previous studies have studied template changes by analyzing the classification difference between the first and the last session. However, classification images from single sessions are often very noisy, and this approach does not use the data in a very effective manner to find the systematic changes in template weighting, as it discards most of the sessions.

In this study we used a standard and much-studied Gabor stimulus (a sinusoid grating in a Gaussian spatial window). This stimulus is ideal for investigating low-level psychophysical processing, as its spatial profile resembles receptive fields of neural mechanisms in early visual areas, and thus this kind of stimulus activates only a local population of neurons (De Valois & De Valois, 1988). This allows for complementing and testing the generalizability of results obtained from position-noise studies. A position-noise stimulus effectively tests how the visual system integrates responses from a large number of local units that are activated by different position-noise elements. In contrast, the task here was to discriminate between two possible Gabor targets that had opposite phase; it thus tested

processing within the local units, as the only difference between the targets is the phase of the sinusoid. The Gabors and the noise were implemented by changing the contrast of 24 lines (see Figure 2), making the stimulus space of such low dimension that it is possible to reliably estimate the template changes from session to session.

By using this well-understood stimulus, we sought to get a more principled account of how the perceptual template changes with learning. Our approach is to estimate the perceptual template in every session of learning and then use regression analysis to find the direction of the systematic template change. The template change vector can then be used to assess how closely the change resembles a change toward an ideal template. The perceptual template for each session can be represented as points in the stimulus space (the dimensionality of the stimulus space equals the number of stimulus parts, here 24) that determine how different stimulus parts are weighted for decisions. The ideal template is the difference between two possible target stimuli with opposite phase. This is illustrated using a toy stimulus space with two dimensions in Figure 1.

The hypothesis that the template becomes more ideal provides a simple prediction of how the templates should change: Initially the template weights may be quite nonideal. If the weights change toward an ideal direction, the template change vector should point toward the ideal template, and with enough training, the template will eventually converge with the ideal template.

We used classification images as template estimates in the change analysis. To simplify the template analysis, we assume that template changes are linear—i.e., the template changes along a straight line in the stimulus space. The main interest in classification images has been in relative weights of different stimulus parts; less attention has been paid to the magnitude or length (distance from the origin) of the classification image. The length is determined by the internal-to-external noise level (Abbey & Eckstein, 2002; Ahumada, 2002; Knoblauch & Maloney, 2008; Murray, Bennett, & Sekuler, 2002). If the internal noise changes systematically in the course of learning, this may bias the estimated direction somewhat. We did not attempt to normalize the template lengths (magnitudes) in any way, due to worries that it would increase the estimation noise, as classification images are quite noisy. Instead, based on the internal noise-level analysis, we assumed that the internal noise level is roughly stable after the first two sessions and does not decrease more with further training. In order to compare the classification images with the ideal template, we scaled the ideal template length by the estimate of the internal noise level in the last four training sessions of the experiment—i.e., we assumed that even with optimal weighting the observer will have this amount of internal noise.

In addition to changes in the perceptual template, learning can reduce the amount of internal noise in the system. We estimated the internal noise level for each subject and session, using the double-pass noise technique (Burgess & Colborne, 1988). In this method, two identical noise samples are shown twice to the observer within an experiment run. Response consistency between these two passes can then be used to estimate the variance of internal noise in the system compared with the (known) variance of external noise.

It is well known that there are large individual differences in the amount of perceptual learning (see, e.g., Dobres & Seitz, 2010). However, traditional threshold measurements have not been informative on what constrains learning. It could be caused by either a lack of template change—i.e., observers stick to the initial, nonoptimized template—or nonoptimal template change—i.e., weak learners change their templates as much as strong learners, but in a nonoptimal manner that does not increase efficiency. Here we measured the amount of systematic template change.

We tested the hypotheses by computing the correlation of template change magnitude and the amount of performance change: If weak learning is caused by nonoptimal template change direction rather than amount of change, these two should not be correlated.

To analyze the template change direction and magnitude, we used two complementary linear regression techniques. The first is multivariate linear regression using template weight as a dependent variable and session index (time) as the predictor. This method is simple and powerful but makes the assumption that changes in weights are approximately constant over time. However, many studies have shown that at least performance improvements in learning saturate over time (see, e.g., Karni & Sagi, 1993; Poggio, Fahle, & Edelman, 1992). Moreover, even from a theoretical point of view, the change cannot be truly constant, as it should converge to an ideal template. We therefore also use linear orthogonal regression (also known as total least squares), which is a regression technique that does not require a separate predictor variable but uses singular value decomposition to infer the direction of change (Golub & Van Loan, 1980); in this analysis the rate of weight change does not have to be constant.

Transfer of template optimization

Our second main objective was to investigate the transfer of perceptual learning. Several studies have suggested that learning in low-level tasks can be highly specific: Performance gains with learning do not generalize to a new stimulus if its low-level properties such as orientation and spatial frequency (Fiorentini & Nicoletta, 1980) or location (Schoups, Vogels, & Orban, 1995) are changed. However, in some studies, substantial transfer has been reported (Ahissar & Hochstein, 1997; Liu & Weinshall, 2000). Often there is large individual variability in the amount of transfer (see, e.g., Zhang, Cong, Song, & Yu, 2013). Since generalization of learning has a large practical value, and may also provide information about the neural structures where learning takes place (reverse hierarchy theory; see Ahissar & Hochstein, 1997), the factors that affect the transfer of learning have been intensively studied.

It has been suggested that transfer can be dependent on task difficulty (Ahissar & Hochstein, 1997; Liu & Weinshall, 2000) or precision (Jeter, Doshier, & Lu, 2009). Ahissar and Hochstein (1997) showed that there was almost full transfer of improvement in an easy or low-precision task, where the orientation difference between the visual-search target and distractors was 30°, but almost no transfer in the difficult or high-precision task, where the difference was 16°. Using the orientation-discrimination task with the Gabors, Jeter

et al. (2009) suggest that the relevant factor in transfer is the orientation difference between stimuli—i.e., task precision. They report almost complete transfer in a low-precision condition, where the orientation difference between the Gabors was large, and virtually no transfer in a high-precision condition, with a small orientation difference.

In the current study, observers trained in the phase-discrimination task with the same stimuli for six sessions (days); the seventh session measured the transfer of training. In this seventh session the stimuli and task were the same except that the global orientation was rotated by 90° (Figure 2). We ask if there is a systematic relationship between the features in the original template, the template optimized by learning, and the template in the transfer condition. In other words, do some reweightings and optimizations systematically transfer to the template for transfer stimuli and some not? We designed both a low-precision and a high-precision version of our stimuli to evaluate the potential role of task difficulty in template optimization—i.e., whether learning in low- and high-precision tasks would be different. The low-precision task had a 180° phase difference between two possible Gabor phases; the high-precision task had a 60° difference. This stimulus design makes the same ideal observer (an odd-symmetric Gabor with the same frequency as the targets) for both high- and low-precision stimuli. This makes comparing the two conditions straightforward, as both conditions have the same stimulus difference, and thus the potential differences in templates cannot reflect differences in the available stimulus information. Moreover, we can directly compare the template change across conditions, as the ideal template is the same in both conditions.

Methods

Subjects

Twelve subjects (eight women; age range = 22–33 years) participated in the experiments. The experimental procedure was in accordance with the Declaration of Helsinki and was approved by the Ethics Committee of the Institute of Behavioural Sciences, University of Helsinki. All subjects were volunteers and gave their written consent to experiments.

Apparatus and stimuli

Experiments were conducted in a dimly lit laboratory. Stimuli were created using MATLAB 2013b

(MathWorks, Natick, MA) using custom software and PsychToolbox 3 extensions (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007; Pelli, 1997). Stimuli were displayed on a Mitsubishi Diamond Pro 2070 SB monitor using a Cambridge Research Systems (Cambridge, UK) ViSaGe MK II stimulus generator with 15-bit luminance resolution.

The stimuli were two Gabor stimuli placed diagonally on top of each other (see Figure 2). The orientation of both Gabors was the same, either 45° or 135°. The lower Gabor served as a reference and always had a constant 0° phase, whereas the phase of the upper Gabor was either +90° or –90° (low-precision condition) or +30° or –30° (high-precision condition).

The stimulus was made by modulating the contrast of 24 thin, oriented lines (width = 0.04°, length = 1°) defined by a Gabor function. More specifically, the contrast of the line at position x in the Gabor \mathbf{g}_ϕ in phase ϕ was defined as

$$\mathbf{g}_\phi(x) = a \exp(-x^2/\sigma_x^2) \cos(2\pi x/\lambda - \phi), \quad (1)$$

where a is the contrast of the Gabor, σ_x is the width of the Gaussian envelope (here 0.17°), and λ is the cycle length of the sinusoid (0.25°). The vector of white random noise values \mathbf{n} , and the mean luminance I_m were then added to this pattern, which defined the contrast of oriented lines. After that, the lines were windowed by a one-dimensional Gaussian function at the orientation orthogonal to the Gabor function. The resulting stimulus (without noise) is the standard two-dimensional Gabor (a two-dimensional Gaussian multiplied by a one-dimensional sinusoid; see Figure 2, for example). The noise had the same contrast energy in each line. The Gabor stimuli (and lines) were oriented globally at either 45° or 135°.

Procedure

A one-interval phase-discrimination task with a four-point confidence-rating response was used. Confidence rating was used instead of a simple yes/no because it provides more information about the outcome of the perceptual processing and can thus yield classification images with less estimation error (Murray et al., 2002).

Each trial started with the presentation of a small central fixation crosshair for 250 ms followed by a blank screen for 250 ms. After that, the stimulus (Gabor pair) appeared for 300 ms in a randomized location, drawn from a Gaussian distribution with the mean at the center of the screen and a horizontal and vertical standard deviation of 0.25°. The observer's task was to indicate the phase of the top Gabor—i.e., whether the target Gabor appeared to have shifted to the left (–30°, –90°) or the right (+30°, +90°), compared

to the bottom (reference) Gabor. The response was given using a keyboard.

The contrast of the upper (target) Gabor was controlled using an adaptive QUEST procedure (Watson & Pelli, 1983) so that the average performance was 75% correct responses. The contrast threshold estimate of an experimental run was used as an initial contrast value for the next run, except for the first run, when 50% contrast was used. The contrast threshold for the last training session was used as the initial value in the first transfer run.

In order to use double-pass internal noise analysis with an adaptive method, experiments were divided into 12 trial miniblocks, where each stimulus (noise mask and Gabor) was shown twice in random order. Within a miniblock, the target contrast was kept constant—i.e., the contrast was updated only once every 12 trials. However, in the first experimental and the first transfer run, when the contrast threshold fluctuated more, the contrast was updated on every trial, and these runs were removed from the consistency analysis.

Experiments were done in 120 trial runs, lasting about 5 min. In each session (day), observers did 10 experiment runs. The whole experiment had seven sessions: six training sessions and one transfer session. In the training sessions, the global orientation of the stimulus was held constant (45° or 135°), being chosen randomly for each subject. In the transfer session, an orthogonal orientation was used. Experiments were carried out on separate days and over a period of about 2 weeks. Before starting the experiment, and in the beginning of the transfer session, observers practiced the task using high-contrast noise-free versions of the stimulus for one run, in order to make sure that they were familiar with the stimuli and understood the task.

Data analysis

Performance measures

Contrast threshold at the end of every run was estimated by using the mean of the QUEST posterior probability distribution. The amount of learning transfer in the contrast threshold was measured by using the transfer index, defined as the ratio of threshold change between the initial session c_1 and the transfer session c_t to change between the initial session and the last training session c_6 : $\tau = (c_1 - c_t)/(c_1 - c_6)$. A value of 1 means that any decrease in the contrast threshold is fully transferred; 0 means no transfer.

Classification-image analysis

The perceptual template (classification image) was estimated using the generalized linear model (Knoblauch & Maloney, 2008, 2012). In a given trial, the

stimulus \mathbf{s}_t was randomly either a Gabor with a positive phase shift \mathbf{g}_ϕ or a Gabor with a negative phase shift $\mathbf{g}_{-\phi}$, added to a Gaussian-distributed external noise \mathbf{n}_t . We assume that the internal response r_t on experimental trial t is dependent on the cross-correlation between the noisy stimulus \mathbf{s}_t and the internal template \mathbf{w} . In addition, we modeled the effect of internal noise by adding a random Gaussian-distributed noise e_t :

$$r_t = \mathbf{w}^T \mathbf{s}_t + e_t. \quad (2)$$

By construction, we can define the positive values of the internal-response mean stimulus with phase shift to the right, and the negative values with phase shift to the left when the absolute value represents confidence. The observer is assumed to give a confidence-rating response by comparing the response with a set of internal criteria, so that rating response i is given when the response falls between criteria c_i and c_{i+1} . Since there were only two possible target stimuli (\mathbf{g}_ϕ and $\mathbf{g}_{-\phi}$), the cross-correlation between the target and the template has only two values. We used a dummy variable o_t to represent the match in trial t and used only external random noise in the regression analysis. This ensures that target shapes cannot bias template estimates.

Considering the model's behavior with respect to a single criterion c_i , the expectation for a positive, “positive phase shift” response is

$$E(r_t > c_i) = \Phi(\mathbf{w}^T \mathbf{n}_t + o_t - c_i), \quad (3)$$

where Φ is the cumulative normal distribution function. A generalized linear ordinal probit model (linear regression with a nonlinear Gaussian link function) can be used to solve this kind of regression with multiple criteria—i.e., estimate the unknown template weights \mathbf{w} from the known stimulus values and responses (for a review, see Knoblauch & Maloney, 2012). We used 24 regressors to estimate the internal weights (classification image), one regressor to estimate the constant target-stimulus match, and three regressors for the internal criteria that correspond to the four response alternatives.

Sampling-efficiency analysis

We estimated the sampling efficiency ρ^2 of the estimated templates from the square of the cross-correlation of the estimated template $\hat{\mathbf{w}}$, normalized to unit length, and an ideal template \mathbf{w}^* :

$$\rho^2 = (\hat{\mathbf{w}}^T \mathbf{w}^*)^2, \quad (4)$$

Internal noise analysis

We used a double-pass method (Burgess & Colborne, 1988) to estimate the internal-to-external noise

ratio. The method is based on calculating the probabilities of obtaining the same rating response on two passes of the same noise sample and target stimulus. This is then compared with the expected consistency, given the amount of internal noise in the system. We used a generalization of the double-pass method for the rating-scale response, explained in detail elsewhere (Kurki & Eckstein, 2014).

Template change analysis

The aim of the template change analysis is to reveal the direction of template weight change from session to session. We assume that the change trajectory is linear and can be represented as a vector that goes through measured templates and shows the direction of the change. We then compare this with an ideal learning vector: a vector that points from the measured templates toward the ideal template. To get enough samples to analyze the template change, we divided the data from each training session into two 600-trial parts and then computed the classification images separately, yielding 12 template estimates $\hat{\mathbf{w}}_k$ for each session k .

The first method to estimate the direction of template change is to use a multivariate linear regression and regress the template weights from each session on the index $k = 1, 2, \dots, 12$ of the session, representing the time dimension. This will then give the direction of the change in the course of learning. The data were made zero mean by subtracting the template average. Then we regressed these templates with a linear, zero-mean session index, using the regress function of MATLAB.

The second method uses linear orthogonal regression, also known as linear total least squares. This technique does not require a predictor variable (time), and thus we do not need to assume that the rate is constant. Linear orthogonal regression was computed by using principal-component analysis of 12 templates obtained in the learning phase. The regression weights are given by the first principal component, which is the direction in the stimulus space where the template projections have the largest variance and thus the largest template change. Principal-component analysis does not maintain the order of templates—i.e., the vector either points from the last session to the first or vice versa. We forced the template change vector to point toward the last-session template by flipping it if the template projections in it were in descending order (based on the sign of the correlation coefficient). In addition to the template change vector, we computed the template projections on this vector, showing the rate of the template change.

We then compared the template change vectors with an ideal learning direction. If the templates were centered at the origin of the space, it would be the

vector from the origin to the ideal template. However, since the templates are generally not centered at the origin, we must center the ideal template \mathbf{w}^* with the measured templates by subtracting the mean of all classification images $\hat{\mathbf{w}}_k$. The magnitude of the classification images estimated with the generalized linear model is dependent on the internal noise level (Knoblauch & Maloney, 2008; Kurki, Saarinen, & Hyvärinen, 2014). We scaled the ideal template with an internal noise estimate \hat{e} in order to make the template magnitudes comparable. The internal noise estimate was the mean of the four last sessions, based on the observation that the internal noise level saturated after about two sessions.

Thus the centered ideal learning direction \mathbf{d}^* was estimated using

$$\mathbf{d}^* = \left(\frac{1}{\hat{e}} \mathbf{w}^* - \sum_k \hat{\mathbf{w}}_k / k \right) \quad (5)$$

where k is the session index. The learning direction was then compared with the ideal learning direction by computing the dot product between the ideal and the observed $\hat{\mathbf{d}}$ template change vectors, both normalized to a unit length, divided by the vector length $|\hat{\mathbf{d}}|$. We refer to this measure as the change optimality index δ :

$$\delta = \hat{\mathbf{d}}^T \mathbf{d}^* / (|\hat{\mathbf{d}}| |\mathbf{d}^*|) \quad (6)$$

Lastly, we estimated the confidence intervals for the template change direction and the change optimality index by using bootstrap resampling (Efron & Tibshirani, 1993). For each subject and session, we resampled (with replacement) the noise mask and response data so that within each session, the number of the stimulus (left/right phase shift) and response classes were the same, generating 1,024 bootstrap replicas of classification-image data. After that we calculated classification images for each replica and, using these, made 128 replica template change vectors. The confidence interval for the template change vector was calculated from the standard deviation of these replicas; the confidence interval for the optimality measure was similarly estimated from the standard deviation of the optimality measure of the replicas.

Results

Performance change

We observed robust learning in contrast thresholds in both the high- and low-precision conditions (Figure 3). On average, the thresholds in the final (sixth) session were about 58% percent lower in the low-

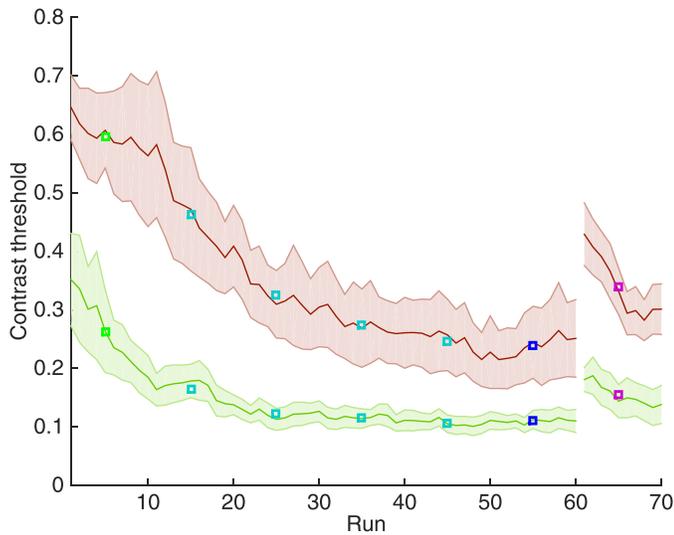


Figure 3. Performance. Contrast thresholds for 75% correct responses plotted against the run number (10 runs per session/day). Runs 1–60 were training runs; 61–70 measured the transfer. The green square shows the mean contrast threshold of the initial session, the cyan squares are the mean thresholds in the training sessions, the blue dot is the final session’s mean threshold, and the purple square is the seventh, transfer session’s mean threshold. The brown graph line shows the average threshold in the high-precision condition run by run, and the green graph is the low-precision condition. Shaded areas in the graphs represent ± 1 SEM (across six observers per condition).

precision condition and 60% percent lower in the high-precision condition than the initial threshold. The contrast-threshold decrease was statistically significant between the first and sixth sessions—low-precision: $t(5) = 3.18$, $p = 0.025$; high-precision: $t(5) = 4.66$, $p = 0.0056$. Every individual observer had a lower threshold in the final session (for individual data, see Table 1).

Most learning was seen in the early sessions. Especially in the low-precision condition, performance improvement after the second session was saturating and slow. Interestingly, there was almost no threshold improvement during the first high-precision session (Figure 3), and most learning seems to happen at a later time than in the low-precision condition. This was not caused by a ceiling effect, as the thresholds for all subjects were well below 100% contrast even in the first session.

We observed partial transfer of learning in both the high- and low-precision conditions. The average transfer index across all subjects and both conditions was 0.61; the average low-precision transfer index was 0.52, and the high-precision transfer index was 0.71. We observed large individual variation in the amount of transfer in thresholds, ranging from a complete transfer to no transfer (see Table 1), and the difference between the low- and high-precision groups was not statistically significant, $t(5) = 0.626$, $p = 0.56$. In the low-precision condition, the contrast threshold in the transfer session (seventh) was about 41% higher than in the final (sixth) learning session, but also 41% lower than the initial threshold. Neither difference was statistically significant: $t(5) = -1.59$, $p = 0.17$; $t(5) = 1.55$, $p = 0.18$. In the high-precision condition the transfer threshold was about 43% higher than the final threshold and about 43% lower than the initial threshold. Both of these differences were significant: $t(5) = -3.44$, $p = 0.018$; $t(5) = 3.76$, $p = 0.013$.

Classification images

The classification image is an estimate of perceptual-template weights in space—i.e., how the different parts of the stimulus are weighted for perceptual decisions. Figure 4 shows the average classification images across the subjects in every session (1–6: training sessions; 7:

Condition	Subject	c_1	c_6	c_t	f_1	f_6	f_t
Low-precision	S1	0.23	0.14	0.16	0.018	0.052	0.036
	S2	0.25	0.13	0.29	0.018	0.046	0.01
	S3	0.22	0.08	0.14	0.015	0.201	0.075
	S4	0.14	0.07	0.13	0.072	0.236	0.103
	S5	0.18	0.08	0.11	0.048	0.216	0.153
	S6	0.54	0.16	0.11	0.005	0.032	0.076
High-precision	S7	0.31	0.15	0.32	0.076	0.174	0.053
	S8	0.56	0.13	0.21	0.016	0.244	0.102
	S9	0.64	0.53	0.49	0.006	0.021	0.019
	S10	0.82	0.20	0.34	0.004	0.116	0.037
	S11	0.59	0.18	0.29	0.012	0.152	0.046
	S12	0.67	0.25	0.38	0.013	0.036	0.032

Table 1. Performance measures. *Notes:* Contrast thresholds c_1 in the first session, c_6 in the final (sixth) session, and c_t in the transfer session; absolute efficiency (squared ratio of human and ideal-observer efficiency) f_1 in the first session, f_6 in the final session, and f_t in the transfer session.

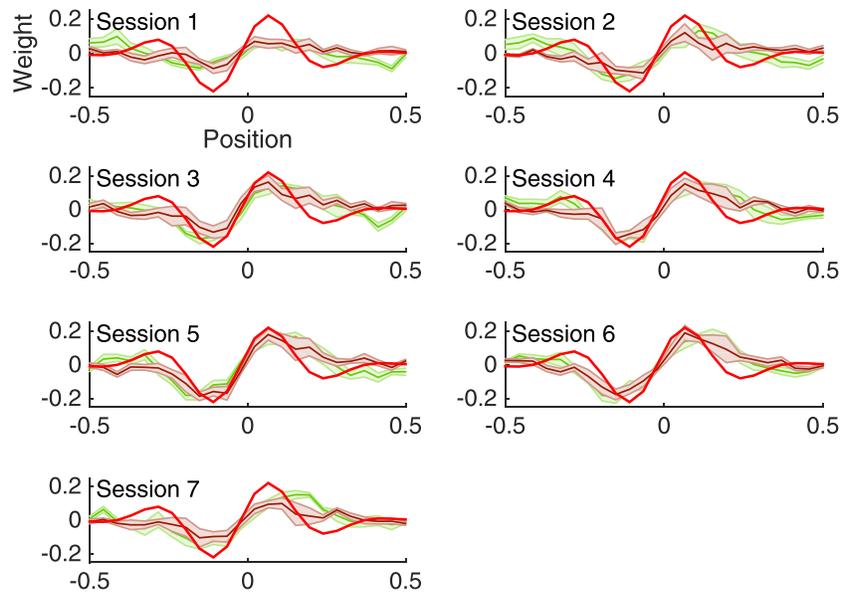


Figure 4. Average classification images for each session. The classification image shows the estimated template weight that the visual system assigns to stimulus information at each position (x-axis scale in the diagonal from the top to the bottom of the screen). The green graph lines show the average classification image in the low-precision condition, and the brown graph lines are the average high-precision classification images (six subjects each). Sessions 1–6 were the training sessions and Session 7 was the transfer session, where the stimulus was flipped to an orthogonal orientation. The red curve is the ideal-observer template. Shaded areas represent ± 1 SEM.

transfer session). Figures 5 and 6 show the classification images for the first and the last training sessions as well as the transfer session for each individual observer. The red curve is an ideal observer, which is the same in both tasks.

All classification images resemble odd-symmetric Gabors. However, in the first sessions the profile of the templates does not generally match the ideal-observer strategy well. Templates may contain some irrelevant and nonoptimal features; for example, observers S3, S5, and S8 initially weight information at the ends of the Gabor where there is little information (see Figure 5). These irrelevant features are also visible in the average classification images, especially in the low-precision condition (Figure 4). In addition, many templates initially miss at least some stimulus parts—for example, observers S3, S4, S6, and S10 are not initially able to use the information at the top half of the Gabor (almost flat classification image above the horizontal meridian).

With learning, the templates become better matched with an ideal observer. More useful features are included in templates and some irrelevant features are unlearned (see S3, S5, and S8). For example, the final templates of S4 and S8 match the ideal closely when the initial template misses the top part. However, even the final templates are generally not very ideal, having on average wider tuning than the ideal (Figure 5; Figure 6: S1, S2, S5, S6, S9, and S12).

We investigated how much more ideal the templates become with learning by estimating the template

sampling efficiency in every session (Figure 7). We found a steady increase in average sampling efficiency, rising steeply at first but saturated in the last sessions. In the low-precision condition, practice increased sampling efficiency from 13% to 36%, or by 173%, $t(5) = -3.95$, $p = 0.010$. In the high-precision condition the increase was 119%, from 21% to 46%, $t(5) = -3.22$, $p = 0.023$. There was not a statistically significant difference between conditions in sampling efficiency, $F(1, 10) = 0.55$, $p = 0.66$ (repeated-measures ANOVA, with session as the within-subject variable). We also estimated absolute efficiency change by comparing the squared ratio of observed d' and ideal-observer's d' . The data are shown in Table 1. On average (across the sessions and subjects), absolute efficiency was 8.9% in the low-precision condition and 7.7% in the high-precision condition.

Internal noise

The internal-to-external noise ratio was estimated using a double-pass method (Figure 8). The average internal-to-external noise ratio was significantly lower (1.0) in the low-precision than the high-precision condition (1.5), $F(1, 10) = 5.34$, $p = 0.043$ (repeated-measures ANOVA). The average internal noise level in the high-precision condition dropped during training sessions by 54%, $t(5) = 5.44$, $p = 0.003$, whereas the change in the low-precision condition was small (11%)

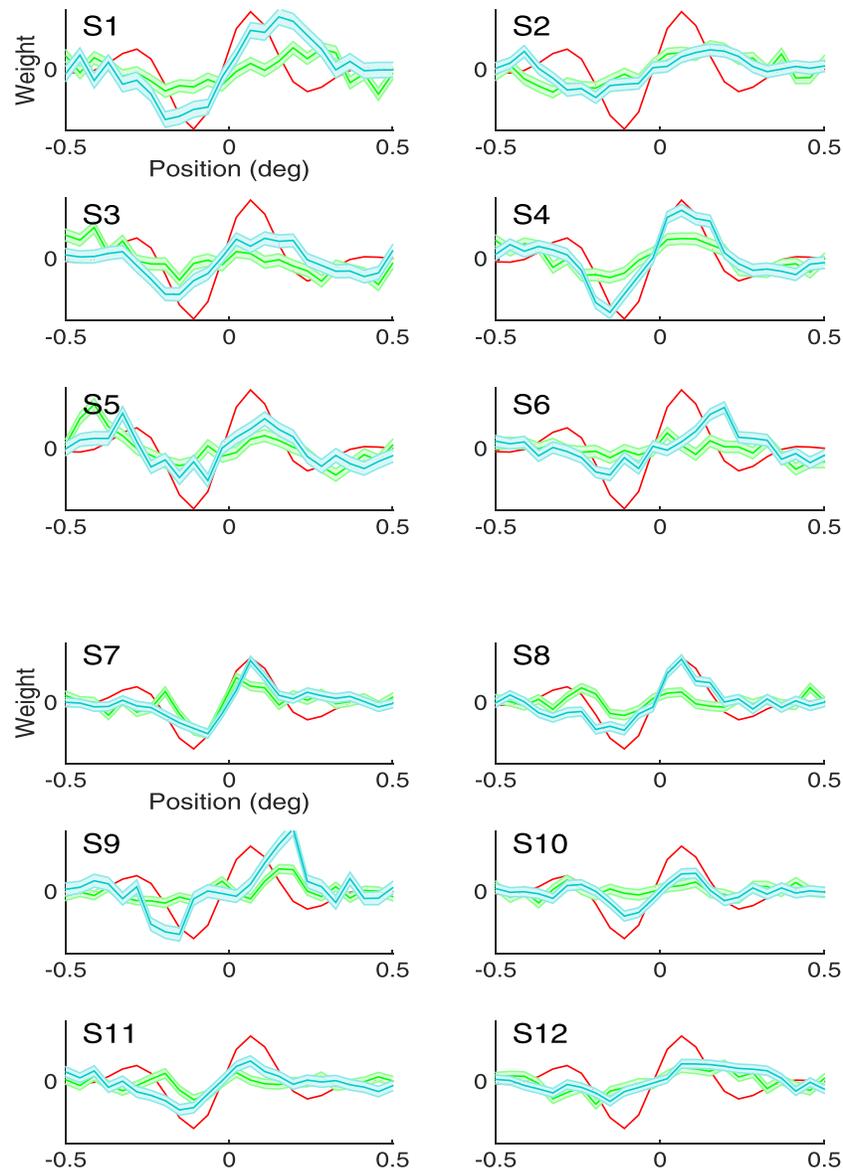


Figure 5. Classification images for each subject (S1–S12) during learning. The green profile is the classification image for the initial session; the cyan profile shows the classification image at the sixth and final session. Shaded areas show the estimated confidence intervals (± 1 SEM). The red curve shows the ideal observer. Subjects 1–6 received the low-precision stimulus, whereas Subjects 7–12 received the high-precision stimulus.

and not statistically significant, $t(5) = 1.22$, $p = 0.28$. The level of internal noise in the transfer condition remained almost constant in the low-precision condition; in the high-precision condition it was slightly higher than in the last training session, but the increase was not statistically significant, $t(5) = -0.967$, $p = 0.378$.

Template change analysis

Figure 9 shows the linear regression estimates for the template change for every subject—i.e., the rate of internal weight change in the course of learning. The red

plot shows an optimal template change vector that would eventually lead to an ideal template. As it is dependent on the weights of the initial templates, it is different for every subject. The estimated change direction is quite similar to the optimal direction; estimated correlations for individual subjects (δ_1 and δ_0) are shown in Table 2. The mean correlation across the subjects is 0.47 in the low-precision condition and 0.53 in the high-precision condition. We also calculated the session-wise projections of the templates on the template change vector (Figure 9, insets). This shows how much the template changes from one session to the next along the direction of learning. The inset also shows the projections of the ideal template and the transfer

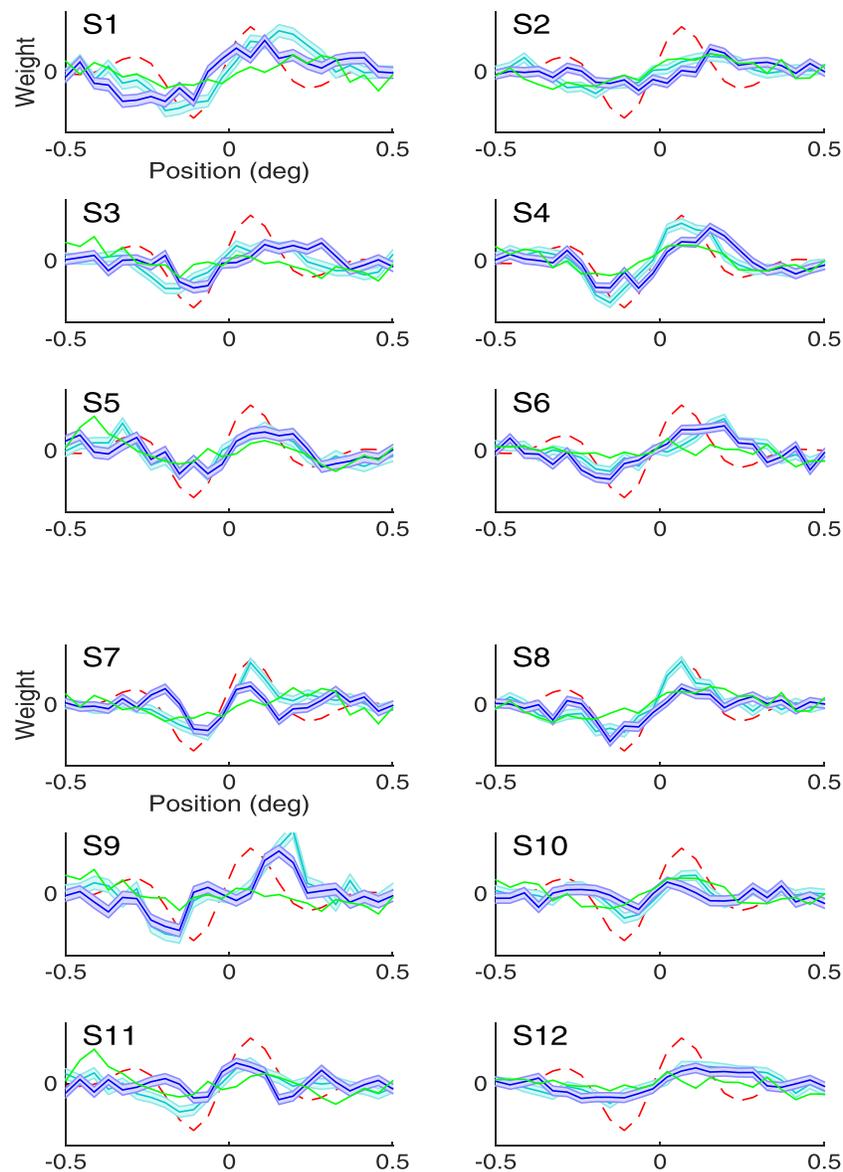


Figure 6. Classification images for each subject in the transfer session. The cyan profile is the classification image for the final (sixth) training session; the magenta profile is the transfer session. The red curve shows the ideal observer, and the green the first session. Shaded areas show the confidence intervals (± 1 SEM). Subjects 1–6 had the low-precision stimulus, whereas Subjects 7–12 had the high-precision stimulus.

template. In most cases the template projection on the learning vector increases systematically with the session number (time), suggesting that the regression captures how the weights change with learning. The projection of the ideal template on the estimated template change vector is typically quite high in absolute terms, implying that the direction is quite close to ideal, and typically much higher than the last template. This implies that template weights are still quite far from the ideal template. Projection of the transfer template is most often about halfway between the last and the first template, but there is considerable variation in the amount of transfer between the subjects.

The alternative and complementary method to estimate the template change used orthogonal regression. The results (Figure 10) are a bit noisier but again closely resemble the optimal change direction. The mean correlation with the measured and the optimal direction is 0.50 in the low-precision condition and 0.54 in the high-precision condition. The projection of templates is generally monotonically rising. The change in projection is quite constant and no obvious saturation can be seen, but the estimates are rather noisy. On average, orthogonal and linear regression estimates are quite close to each other, and the predictions are significantly correlated, $\rho = 0.87$, $p =$

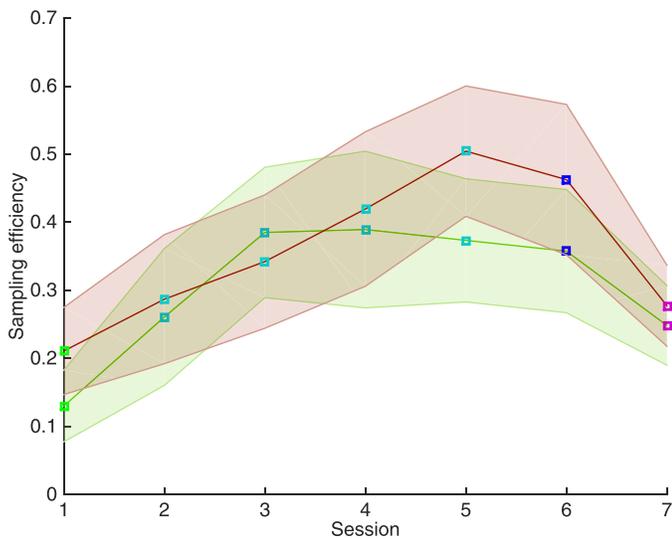


Figure 7. Average template sampling efficiency in learning. Sessions (days) 1–6 were the training sessions and 7 was the transfer session. The brown graph line shows the high-precision condition, and the green graph line the low-precision condition. Shaded areas show the confidence intervals (± 1 SEM).

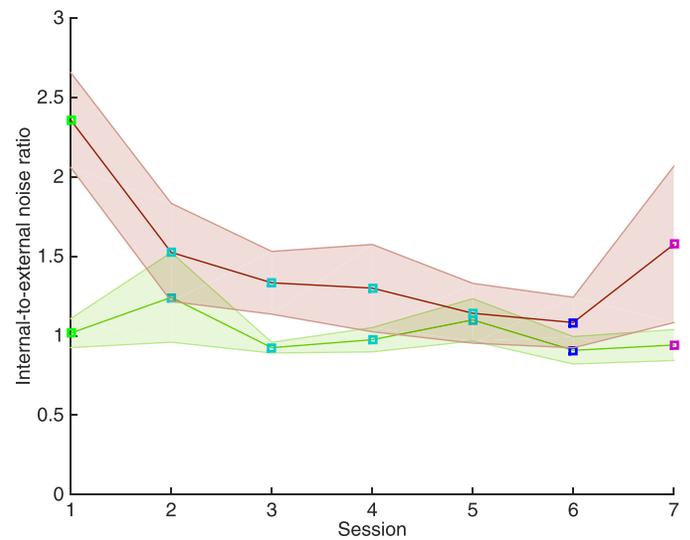


Figure 8. Estimated internal-to-external noise ratio in each session (day). Session 7 is the transfer session. The brown graph line shows the high-precision condition, and the green graph line shows the low-precision condition. Shaded areas represent the confidence intervals (± 1 SEM).

0.0001. However, linear estimates have less estimation error, as can be seen from the estimated confidence intervals.

We estimated the amount of systematic template change l for each observer, namely the distance between the projection of the first and the last training templates on the template change vector (using the linear regression estimate). The data are shown in Table 2. The amount of template change was highly correlated with the percentage of threshold decrease, $\rho = 0.62$, $p = 0.030$.

Discussion

Performance change

We observed consistent learning in phase discrimination: The average contrast threshold after six sessions of training was about 60% lower than the initial threshold. Although there were individual differences in the amount of learning, there were no nonlearners; all 12 subjects became better at the task

Condition	Subject	ρ_1^2	ρ_6^2	ρ_t^2	τ	δ_i	δ_o	l
Low-precision	S1	0.034	0.300	0.145	0.80	0.705	0.793	0.105
	S2	0.083	0.150	0.047	-0.27	0.242	0.254	0.390
	S3	0.087	0.400	0.252	0.58	0.505	0.630	0.050
	S4	0.387	0.700	0.363	0.18	0.798	0.807	0.544
	S5	0.120	0.490	0.443	0.71	0.183	0.015	0.398
	S6	0.071	0.110	0.240	1.14	0.400	0.477	0.352
High-precision	S7	0.457	0.660	0.320	-0.02	0.376	0.670	0.208
	S8	0.295	0.580	0.500	0.80	0.572	0.516	0.163
	S9	0.011	0.070	0.050	1.30	0.586	0.774	0.024
	S10	0.160	0.750	0.260	0.77	0.698	0.585	0.300
	S11	0.242	0.520	0.270	0.73	0.529	0.489	0.191
	S12	0.102	0.190	0.250	0.67	0.434	0.225	0.143

Table 2. Template characteristics. Notes: Estimated template sampling efficiencies ρ_1^2 for the first session, ρ_6^2 for the final session, and ρ_t^2 for the transfer session, for each subject. τ = threshold transfer index (1 = complete transfer, 0 = no transfer). δ_i = optimality index of the template change vector compared with change towards the ideal template, obtained by multivariate linear regression. δ_o = optimality index of the orthogonal regression template change vector. l = magnitude of template change (the distance between the first and the last template in the linear regression template change vector).

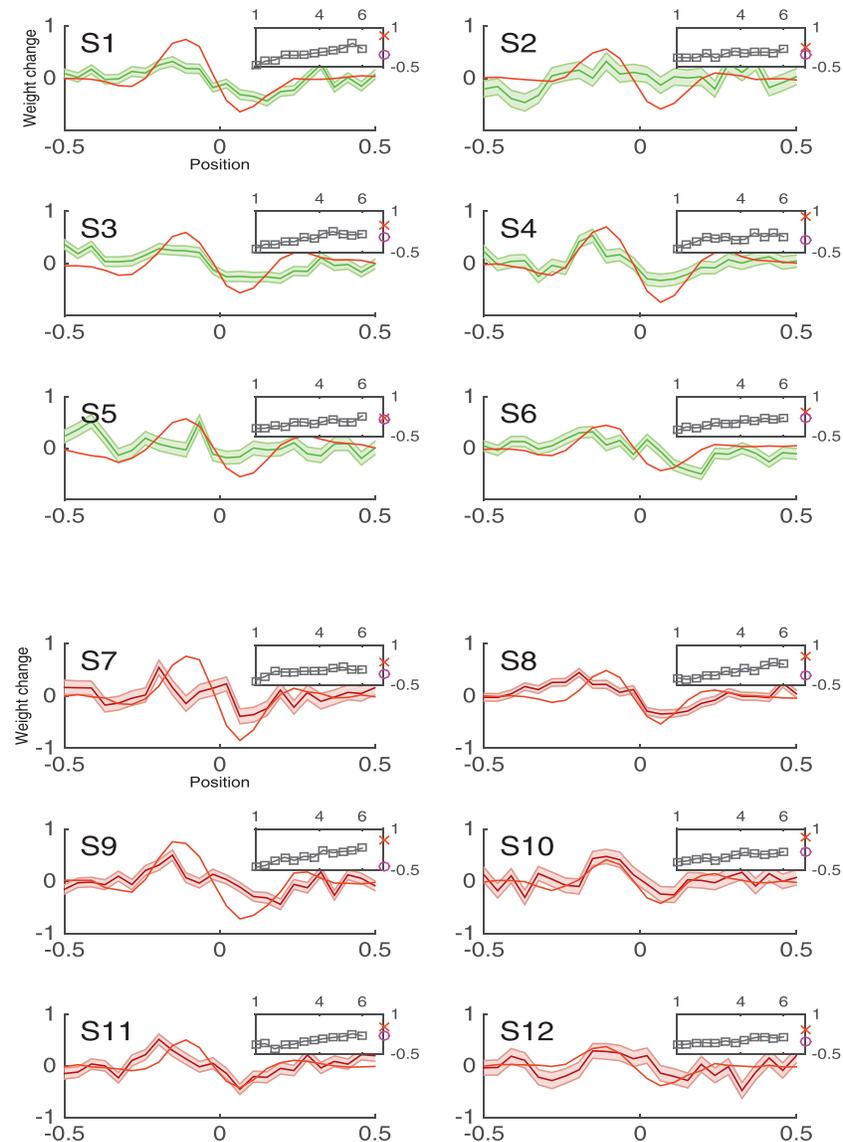


Figure 9. Template change. The template change vector (change direction) was estimated using multivariate linear regression (template weight against the session number) and then normalized to the unit norm. The results show the relative template weight changes with learning (top three rows, green: low-precision condition; bottom three rows, brown: high-precision condition). The red curve shows the optimal template change for each subject. Shaded areas represent ± 1 SEM, estimated using bootstrap resampling. The squares in the inset show the template projections in the template change vector (x-axis: session number; two template estimates per session). The red cross in the inset marks the projection of the ideal template, and the purple circle shows the transfer template projection.

with practice. In the transfer session with a Gabor target in an orthogonal orientation, we found partial transfer of learning: Thresholds were between the initial and the final threshold (62% transfer on average).

Classification-image analysis

We found that classification images in general resemble odd-symmetric Gabors, but initially they did not very closely resemble the ideal observer, often

lacking some important features. Classification images also show interesting individual differences, especially in the initial templates, where some observers relied on irrelevant features that were not present in the stimuli. Compared to an ideal observer, the sampling efficiency of the initial templates was rather low. With learning, the template started to better resemble the ideal-observer template, yielding an increase in sampling efficiency.

Unlike previous learning studies using the position-noise paradigm (Kurki & Eckstein, 2014; Li et al.,

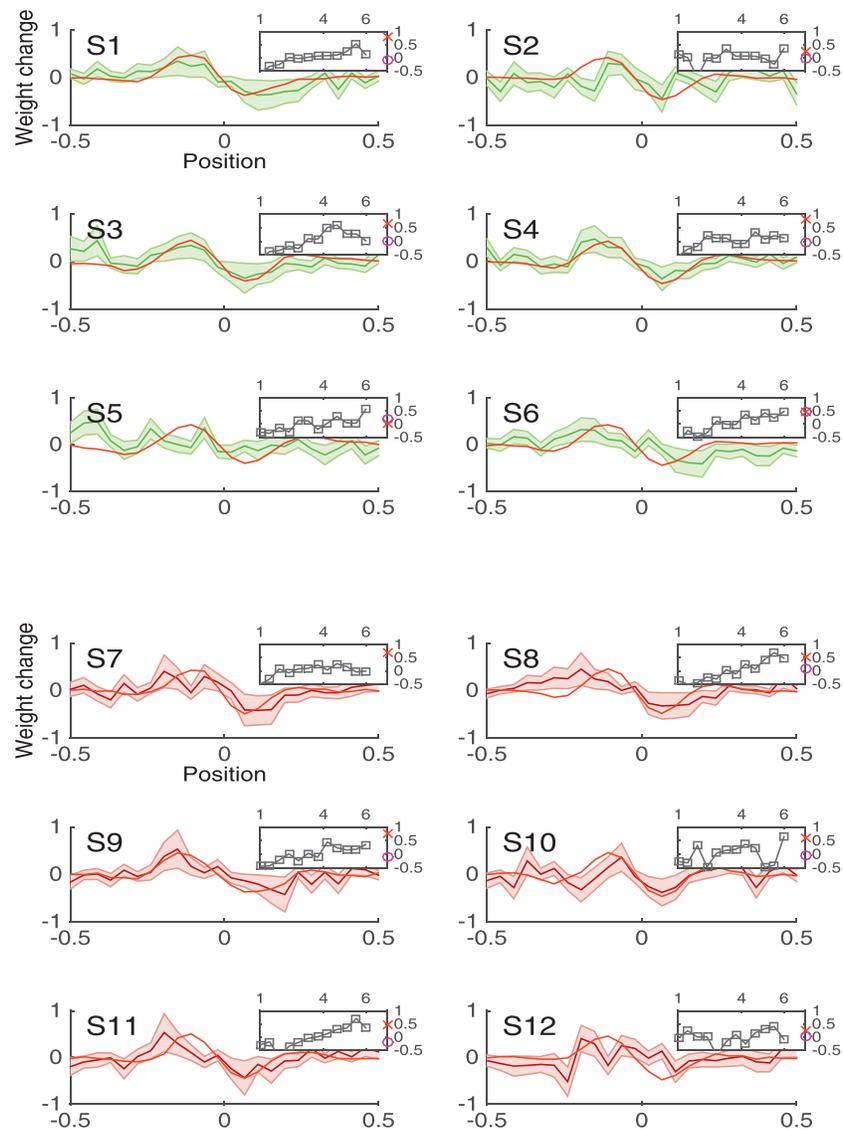


Figure 10. Template change vector (direction) estimates using orthogonal regression. The results show the rate of template weight change in learning. The red curve shows the optimal change vector. Shaded areas show ± 1 SEM, estimated using bootstrap resampling. The inset shows the template projections in the template change vector (x-axis: session number). The red cross in the inset marks the projection of the ideal template change vector, and the purple circle shows the transfer template projection.

2004), the change in templates could not be well described as an expansion of the template area. Even in initial sessions, the observers seemed to be able to reliably use features from all over the stimulus area; for example, we did not see a preference for the bottom half of the Gabor that was closer to the fovea. A possible reason is that two stimuli probe different kinds of processing limitations: The Gabor stimulus here is small and localized both in spatial frequency and in space. Thus it may be possible to process the whole stimulus by a highly local neural population even without training, as it does not require extensive integration over neurons tuned to different spatial locations. Here, the initial nonoptimality seems to be the sampling of irrelevant features outside the signal

area and omission of relevant features. This may reflect a problem in adjusting the weights of the neural mechanisms that have the best spatial-location and spatial-frequency tuning; with learning, weighting becomes more optimal.

Interestingly, in many observers the template peaks more peripherally and tuning is consistently wider than an ideal tuning, even after learning (see the averages in Figure 4, session 6; Figure 5, subjects S1, S2, S3, S5, S6, S9, and S12). This could in principle reflect off-frequency looking, where the visual system may prefer other stimulus frequencies than the target peak frequency because of channel integration (Goris, Wichmann, & Henning, 2009) or early nonlinearities (Kilpeläinen, Nurminen, & Donner, 2012). However,

the spatial frequency of the Gabor target was $4\text{ c}/^\circ$, which is thought to be near the peak of the human contrast sensitivity function (De Valois & De Valois, 1988), so it is not clear why observers would prefer lower frequencies than the target's. Another possibility is that observers have spatial uncertainty about the exact location of the target. This can cause “smearing” of the classification image (Kurki, Hyvärinen, Laurinen, & Hyvärinen, 2006; Tjan & Nandy, 2006). One should be cautious in making strong conclusions based on our data, as there seems to be quite a lot of individual variation in tuning width, with some observers having quite ideal tuning, and in one case (S7) even too tight tuning. One should also note that the technique used here shows the template only relative to the axis of Gabor modulation. It is possible that there are some changes in the orthogonal axis that are not captured by our 1-D classification images.

Internal noise

We estimated the internal noise level during learning using a double-pass technique. We found that the internal noise level was initially higher in the high-precision condition and decreased by about 50% through training. In the low-precision condition the internal noise level was lower and not affected by training.

Earlier studies have provided somewhat conflicting accounts on the role of internal noise in perceptual learning. Doshier and Lu (1998) used an equivalent noise-masking technique and reported a decrease in noise through learning. However, Gold, Bennett, and Sekuler (1999), Li, Levi, and Klein (2004), and Kurki and Eckstein (2014) used a double-pass technique and did not find internal noise-level changes during learning. Interestingly, both conditions seem to converge to an internal-to-external noise level of about 1; it is possible that this reflects some kind of hardwired limit in this task that cannot be improved upon further.

It is thus possible that whether learning operates by internal noise reduction or not depends on the task demands. One can speculate that in the high-precision condition, the neural pathway discriminating the minute difference in the Gabor phase may initially weight many neural filters that respond about equally well to both versions of the stimulus. This would add vast amounts of noise to the output, without benefiting stimulus discriminability. In the low-precision condition it may be easier to omit such weakly discriminating filters and have less problem with noise. On the other hand, this experiment design does not allow us to make strong conclusions about the issue, as the stimulus contrast was not kept constant. A number of previous studies have suggested that the amount of internal

noise may be dependent on contrast (Green & Swets, 1966), which was much lower at the end of the experiment and in the low-precision condition. It is thus possible that the drop in the internal noise level is caused by some kind of contrast-related change rather than learning.

Template transfer

We investigated the transfer of template change by using a stimulus in an orthogonal orientation in the seventh session. On average, observers showed partial transfer of learning, but there was also substantial variation, ranging from no transfer to complete transfer. On visual inspection, transfer templates typically look quite similar to partly optimized templates in the middle of learning (depending on the amount of transfer; those with high transfer typically have templates that are very similar to final templates). The sampling efficiency of the last training template is higher than the initial efficiency but lower than the sampling efficiency of the template in the final training session. Also, the transfer template projection on the template change vector is about on the halfway line between the initial and final templates. However, there are also some differences: Optimization of the irrelevant features seems to be largely transferred (this can be seen at the end of the average template in early sessions; Figure 4; see also Figure 5, S3, S5, S9, and S11).

We did not find a difference between the high- and the low-precision task in the amount of threshold transfer. Moreover, task precision did not have any effect on the amount of template change or template optimization (sampling-efficiency change). In other words, observers used a similar “optimization strategy” in both conditions.

According to the reverse hierarchy theory of perceptual learning (Ahissar & Hochstein, 1997), easy and difficult (low- and high-precision) tasks have different processing demands and different neural loci. Easy tasks would suffice with a coarse low-level representation of stimuli; in this case, learning takes place in higher brain areas and transfers over low-level stimulus differences. In difficult tasks the locus of learning will progress to lower-level mechanisms that have a higher signal-to-noise ratio, explaining the selectivity of learning for low-level stimuli. Jeter et al. (2009) redefined the task difficulty as task precision. In the current study, we did not find any evidence that task precision would necessarily lead to a different optimization strategy and a coarser template; differences between templates and optimization in high- and low-precision conditions were minute. In this experiment, the available stimulus information and the ideal

observer were matched, whereas Jeter et al. (2009) used a design in which the available stimulus information in the high-precision task was smaller than in the low-precision task. It is thus possible that available stimulus information, rather than task difficulty per se, determines whether the training transfers. For example, we speculate that if the amount of potential information is large, the visual system is faced more with a problem of integrating the different early neural filters in an appropriate manner, whereas if the amount of potential information is small, the problem is in focusing on the relevant neural filters. These two processes may have different transfer characteristics (i.e., integration having more transfer), which would explain why transfer has been found to be dependent on potential stimulus information in earlier studies. On the other hand, it is also possible that the discrepancy in the result is caused by other differences in stimuli and tasks. It is also possible that learning in easy and difficult conditions reflects different mechanisms; for example, template optimization including external noise exclusion (Doshier & Lu, 1998, 1999) might have a more prominent role in a high-noise condition. However, we did not find a clear difference between the conditions in sampling efficiency or efficiency change. More studies using different tasks will be needed to clarify the issue.

Template change direction

We used a new approach to analyze the template change direction in perceptual learning. The results show that the direction is quite close to, but not identical with, an optimal change direction that would eventually lead to an ideal-observer template. Both linear regression and orthogonal regression estimated the correlation between the observed direction and optimal direction to be about 50%, which is high but not perfect. This provides direct evidence that at least with this kind of elementary stimulus, learning operates by quite literally reweighting the template toward the ideal template. However, in most cases, reweighting was quite far from being complete, as template efficiency was only about 50% in the last session.

We further analyzed the amount of systematic template change by computing the distance between the first and the last template in the change direction vector. We found that the amount of change was highly correlated with the amount of threshold improvement. As a further illustration, four observers with the least threshold improvement had the second-, eighth-, seventh-, and third-least template change. This clearly supports the idea that learning is constrained by the lack of template change rather than nonoptimal template change. We could not find observers who changed their templates consistently but in a very

nonideal manner, and the amount of learning was highly correlated with the amount of template change.

The template change analysis here assumes that template change trajectory is linear. How reasonable is this assumption? Previous classification-image studies (Kurki & Eckstein, 2014; Li et al., 2004) have shown that template weights typically change in gradual manner: Classification images in the middle of learning resemble a weighted average of the original and final templates, suggesting that a linear approximation might be valid. Moreover, in general the templates at different stages of learning project to the learning vector in a constantly ascending manner. For many nonlinear trajectories, this would not be the case. Finally, as this is the first study to estimate the direction of learning, it is reasonable to start with a linear model, as linear approximations are mathematically simplest and often most robust.

The results here show that the classification-image method can provide rich information about visual learning and can complement traditional threshold-performance measurements. Template change estimation adds further power to classification-image analysis in learning, giving a more detailed and principled view of how the template is optimized through learning. We show direct evidence that the template optimization process is well approximated by a model in which the weightings change gradually from an initial nonoptimal weighting toward an ideal weighting: The estimated direction of template change is highly correlated with an ideal change trajectory, and the amount of template change correlates highly with the amount of learning. The latter finding shows also that despite large individual differences generally found in perceptual-learning studies, template changes are always toward a rather ideal direction. We did not find observers who changed their templates consistently but toward a nonoptimal direction.

On the other hand, the estimated template changes are less than perfectly correlated with the ideal direction; this also implies that some part of template change is not reflected in performance increase. In a sense, the performance data underestimate the processing change in learning. Classification images also allow investigation of the individual differences in optimization strategies in a meaningful way. For example, some observers here initially had a clearly nonoptimal template that used irrelevant features at the end of the stimuli; classification images show that these parts of the template were quickly optimized and transferred to an orthogonal orientation. Here we used the method in an elementary task where overall efficiency was rather high; however, it might be even more useful in exploring learning in more complex tasks such as face recognition. The method allows the testing of specific hypotheses on what features are

optimized through learning, and is fruitful in evaluating the computational models of learning.

Keywords: perceptual learning, classification image, ideal-observer analysis

Acknowledgments

Commercial relationships: none.
Corresponding author: Ilmari Kurki.
Email: ilmari.kurki@helsinki.fi.
Address: Institute of Behavioral Sciences, University of Helsinki, Helsinki, Finland.

References

- Abbey, C. K., & Eckstein, M. P. (2002). Classification image analysis: Estimation and statistical inference for two-alternative forced-choice experiments. *Journal of Vision*, 2(1):5, 66–78, doi:10.1167/2.1.5. [PubMed] [Article]
- Ahissar, M., & Hochstein, S. (1993). Attentional control of early perceptual learning. *Proceedings of the National Academy of Sciences, USA*, 90, 5718–5722.
- Ahissar, M., & Hochstein, S. (1997). Task difficulty and the specificity of perceptual learning. *Nature*, 387(6631), 401–406.
- Ahumada, A. J. (2002). Classification image weights and internal noise level estimation. *Journal of Vision*, 2(1):8, 121–131, doi:10.1167/2.1.8. [PubMed] [Article]
- Beard, B., & Ahumada, A. J. (1998). A technique to extract the relevant features for visual tasks. *Proceedings of SPIE*, 3299, 79–85.
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10(4), 433–436.
- Burgess, A., & Colborne, B. (1988). Visual signal detection: IV. Observer inconsistency. *Journal of the Optical Society of America A*, 5(4), 617–627.
- De Valois, R. L., & De Valois, K. K. (1988). *Spatial vision*. New York: Oxford University Press.
- Dobres, J., & Seitz, A. R. (2010). Perceptual learning of oriented gratings as revealed by classification images. *Journal of Vision*, 10(13):8, 8–11, doi:10.1167/10.13.8. [PubMed] [Article]
- Dosher, B. A., & Lu, Z.-L. (1998). Perceptual learning reflects external noise filtering and internal noise reduction through channel reweighting. *Proceedings of the National Academy of Sciences, USA*, 95(23), 13988–13993.
- Dosher, B. A., & Lu, Z.-L. (1999). Mechanisms of perceptual learning. *Vision Research*, 39(19), 3197–3221.
- Eckstein, M., & Ahumada, A. (2002). Classification images: A tool to analyze visual strategies. *Journal of Vision*, 2(1): i, doi:10.1167/2.1.i. [PubMed] [Article]
- Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York: Chapman & Hall.
- Fahle, M., & Morgan, M. (1996). No transfer of perceptual learning between similar stimuli in the same retinal position. *Current Biology*, 6(3), 292–297, doi:10.1016/S0960-9822(02)00479-7.
- Fiorentini, A., & Nicoletta, B. (1980). Perceptual learning specific for orientation and spatial frequency. *Nature*, 287(4), 43–44.
- Geisler, W. S. (2011). Contributions of ideal observer theory to vision research. *Vision Research*, 51(7), 771–781, doi:10.1016/j.visres.2010.09.027.
- Gold, J., Bennett, P. J., & Sekuler, A. B. (1999). Signal but not noise changes with perceptual learning. *Nature*, 402(6758), 176–178, doi:10.1038/46027.
- Gold, J., Sekuler, A. B., & Bennett, P. J. (2004). Characterizing perceptual learning with external noise. *Cognitive Science*, 28(2), 167–207, doi:10.1016/j.cogsci.2003.10.005.
- Golub, G. H., & Van Loan, C. F. (1980). An analysis of the total least squares problem. *SIAM Journal on Numerical Analysis*, 17(6), 883–893, doi:10.1137/0717073.
- Goris, R. L. T., Wichmann, F. A., & Henning, G. B. (2009). A neurophysiologically plausible population code model for human contrast discrimination. *Journal of Vision*, 9(7):15, 1–22, doi:10.1167/9.7.15. [PubMed] [Article]
- Green, D., & Swets, J. (1966). *Signal detection theory and psychophysics*. New York: Wiley.
- Jeter, P. E., Dosher, B. A., & Lu, Z. (2009). Task precision at transfer determines specificity of perceptual learning. *Journal of Vision*, 9(3):1, 1–13, doi:10.1167/9.3.1. [PubMed] [Article]
- Karni, A., & Sagi, D. (1993). The time course of learning a visual skill. *Nature*, 365(16), 250–252.
- Kilpeläinen, M., Nurminen, L., & Donner, K. (2012). The effect of mean luminance change and grating pedestals on contrast perception: Model simulations suggest a common, retinal, origin. *Vision Research*, 58, 51–58, doi:10.1016/j.visres.2012.02.002.
- Kleiner, M., Brainard, D. H., & Pelli, D. G. (2007).

- What's new in Psychtoolbox-3? *Perception*, 36, (ECPV Abstract Supplement).
- Knoblauch, K., & Maloney, L. T. (2008). Estimating classification images with generalized linear and additive models. *Journal of Vision*, 8(16):10, 1–19, doi:10.1167/8.16.10. [PubMed] [Article]
- Knoblauch, K., & Maloney, L. T. (2012). *Modeling psychophysical data in R*. New York: Springer, doi:10.1007/978-1-4614-4475-6.
- Kurki, I., & Eckstein, M. P. (2014). Template changes with perceptual learning are driven by feature informativeness. *Journal of Vision*, 14(11):6, 1–18, doi:10.1167/14.11.6. [PubMed] [Article]
- Kurki, I., Hyvärinen, A., Laurinen, P., & Hyvärinen, A. (2006). Collinear context (and learning) change the profile of the perceptual filter. *Vision Research*, 46(13), 2009–2014, doi:10.1016/j.visres.2006.01.003.
- Kurki, I., Saarinen, J., & Hyvärinen, A. (2014). Investigating shape perception by classification images. *Journal of Vision*, 14(12):24, 1–19, doi:10.1167/14.12.24. [PubMed] [Article]
- Li, R. W., Klein, S. A., & Levi, D. M. (2008). Prolonged perceptual learning of positional acuity in adult amblyopia: Perceptual template retuning dynamics. *The Journal of Neuroscience*, 28(52), 14223–14229.
- Li, R. W., Levi, D. M., & Klein, S. A. (2004). Perceptual learning improves efficiency by re-tuning the decision “template” for position discrimination. *Nature Neuroscience*, 7(2), 178–183, doi:10.1038/nn1183.
- Liu, Z., & Weinshall, D. (2000). Mechanisms of generalization in perceptual learning. *Vision Research*, 40(1), 97–109, doi:10.1016/S0042-6989(99)00140-6.
- Murray, R. F. (2011). Classification images: A review. *Journal of Vision*, 11(5):2, 1–25, doi:10.1167/11.5.2. [PubMed] [Article]
- Murray, R. F., Bennett, P. J., & Sekuler, A. B. (2002). Optimal methods for calculating classification images: Weighted sums. *Journal of Vision*, 2(1):6, 79–104, doi:10.1167/2.1.6. [PubMed] [Article]
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10(4), 437–442.
- Poggio, T., Fahle, M., & Edelman, S. (1992, May 15). Fast perceptual learning in visual hyperacuity. *Science*, 256(5059), 1018–1021, doi:10.1126/science.1589770.
- Saarinen, J., & Levi, D. M. (1995). Perceptual learning in vernier acuity: What is learned? *Vision Research*, 35(4), 519–527.
- Schoups, A. A., Vogels, R., & Orban, G. A. (1995). Human perceptual learning in identifying the oblique orientation: Retinotopy, orientation specificity and monocularly. *The Journal of Physiology*, 483(3), 797–810.
- Tjan, B. S., & Nandy, A. S. (2006). Classification images with uncertainty. *Journal of Vision*, 6(4):8, 387–413, doi:10.1167/6.4.8. [PubMed] [Article]
- Watson, A. B., & Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception & Psychophysics*, 33, 113–120.
- Zhang, G., Cong, L.-J., Song, Y., & Yu, C. (2013). ERP P1-N1 changes associated with Vernier perceptual learning and its location specificity and transfer. *Journal of Vision*, 13(4):19, 1–13, doi:10.1167/13.4.19. [PubMed] [Article]