Journal of
Electronic Imaging

JElectronicImaging.org

Why is quality estimation judgment fast? Comparison of gaze control strategies in quality and difference estimation tasks

Jenni Radun
Tuomas Leisti
Toni Virtanen
Göte Nyman
Jukka Häkkinen

SPIE. IS&T imaging.org

Why is quality estimation judgment fast? Comparison of gaze control strategies in quality and difference estimation tasks

Jenni Radun,* Tuomas Leisti, Toni Virtanen, Göte Nyman, and Jukka Häkkinen
University of Helsinki, Institute of Behavioral Sciences, P.O. Box 9, FI-00014 University of Helsinki, Finland

Abstract. To understand the viewing strategies employed in a quality estimation task, we compared two visual tasks—quality estimation and difference estimation. The estimation was done for a pair of natural images having small global changes in quality. Two groups of observers estimated the same set of images, but with different instructions. One group estimated the difference in quality and the other the difference between image pairs. The results demonstrated the use of different visual strategies in the tasks. The quality estimation was found to include more visual planning during the first fixation than the difference estimation, but afterward needed only a few long fixations on the semantically important areas of the image. The difference estimation used many short fixations. Salient image areas were mainly attended to when these areas were also semantically important. The results support the hypothesis that these tasks' general characteristics (evaluation time, number of fixations, area fixated on) show differences in processing, but also suggest that examining only single fixations when comparing tasks is too narrow a view. When planning a subjective experiment, one must remember that a small change in the instructions might lead to a noticeable change in viewing strategy. [DOI: 10.1117/1.JEI.23.6.061103]

Keywords: image quality; subjective estimation; eye movements; task.

Paper 14125SSP received Mar. 17, 2014; revised manuscript received Jun. 24, 2014; accepted for publication Jul. 10, 2014; published online Aug. 7, 2014.

1 Introduction

In the area of quality estimation, the aim is often to objectively measure the quality of an image or a video without the help of actual human viewers. The subjective estimations from observers are, however, the ground truth against which the models are tested. The objective metrics perform adequately when the differences in quality are clear.[1–3] However, in the cases where image quality changes are small and subjective estimations are based more on preferences than on differentiating artifacts, things become more complex. The models try to use the knowledge of how the human visual system functions to make more accurate measures, and visual attention is one feature that is sometimes taken into account.[4] To understand what influences visual attention in the subjective tasks that serve as ground truth for objective measures, eye movement research on viewing strategies in different tasks is needed.

Both top-down and bottom-up influences control our gaze and where we allocate our attention, since without this interaction we would not be able to act in our environment. Top-down influences are internal factors coming from the observer and bottom-up influences are factors coming from the environment. Top-down influences are the observer's aims and states, such as knowledge of what is needed to accomplish a task, and bottom-up influences are, for example, the physical characteristics of an object. However, many studies have found that top-down influences direct gaze more than low-level bottom-up influences.[5,6] The evidence in favor of a strong top-down guidance of gaze control has led to the need for a better understanding of visual tasks.[7]

Image quality estimation is one example of a visual value estimation task. Image quality can be defined as the degree to which something fulfills the requirements imposed on it;[8] in other words, how valuable an image is under certain circumstances. One definition of image quality stresses that the image must have an adequate combination of discriminability and identifiability in the items depicted.[8] These requirements are good for distinguishing low-quality images from high-quality images, but they reveal little about differences in a high image quality. In the case of high image quality, the requirements of discriminability and identifiability are no longer sufficient, since the images are always both recognizable and distinguishable. Research examining the subjective dimensions of multivariate image quality has found that high-quality images were separated from the others based on naturalness, and lower-quality images from the others according to sharpness and darkness.[9] At high levels of image quality, estimation is a highly subjective preference task where the value of an image is determined by subjective impressions that are not directly convertible to low-level image changes but instead show a subjective interpretation of changes in certain contexts.[10] High-level image quality estimation can be considered a preference task, and preference tasks can be linked to gaze control studies where the effect of the tasks on eye movement patterns has been examined.

Common tasks in gaze control studies include free viewing,[11,12] memory,[13,14] search,[13–15] and detection tasks.[16,17] However, after Antes[18] examined eye movements in preference estimation related to line drawings, preference estimation has received little attention in studies on gaze control in

*Address all correspondence to: Jenni Radun, E-mail: jenni.radun@helsinki.fi natural scenes. Compared to the common tasks in the eye movement studies listed above, the preference task is special, since it is a subjective evaluation task where the subjective value of an object lies in the interpretations the observer gives to the object under certain circumstances, or is otherwise stated in the interaction between the object and observer's cognitive and affective processes.[19] Other tasks have objectively correct answers or do not require an answer at all (free viewing).

Even with the limited attention paid to preference tasks in gaze control research, some studies on the subject have been done. For example, Antes[18] concluded that when people viewed an image and evaluated their preference, the first part of the act consisted of viewing the informative aspects of the image and the latter part the examination of less informative details. Later, in scene viewing studies, saccade lengths have been linked to different ways of watching an image: long saccades (>8 deg) have been identified as transitions to new image areas and short saccades have been associated with a detailed examination of certain areas.[20,21] In addition, fixation durations have been connected with saccade lengths: long fixations connected with short saccades define local processing and short fixations connected with long saccades are relocations to new regions of the scene.[21,22] These different types of processing have been examined in relation to different viewing tasks. A study that compared gaze control in four visual tasks found that when estimating pleasantness, viewing was described as having global processing because of its long saccade lengths (with the average being above a visual angle of 7 deg).[23] In addition, it found that fixation durations were shorter in the pleasantness and search tasks than in the memory and free-viewing tasks.

Earlier eye movement research in quality estimation studies has shown that when estimating quality, fewer fixations were made on regions of interest (ROIs) than in a free-viewing task.[24] The ROIs were defined using areas that the observers were watching the most in the free-viewing task. Furthermore, it has been shown that the fixations' spatial distribution in the free-viewing task differed from their distribution in the quality estimation task[25] and that the information from the free-viewing task was more useful for objective quality measures.[26] However, the free-viewing task is different from the quality estimation task since it does not require an estimation or answer to any question related to images. Compared with the free-viewing tasks, the degradations in images in the quality estimation task were indeed noticed to influence the spatial viewing, even though there was no difference in fixation durations.[25] One might ask whether the differences between the tasks are still visible if one compares two estimation tasks.

1.1 *Present Study*

In this study, we examine how the task changes viewing strategies when comparing quality estimation and difference estimation tasks that are performed on the same material with comparisons between a pair of images and only the instructions of what is being estimated are changed. For both quality and difference estimation tasks, it is possible to apply the magnitude estimation of differences or change in quality;[17,27] however, only one involves value evaluation and can, therefore, be considered a preference task.[28] Magnitude estimation requires the subject to assign numbers to a series of stimuli under the instruction to make the numbers proportional to the apparent magnitudes of the sensations produced.[29] Therefore, both tasks require forward and backward comparisons, since both tasks involve a magnitude estimation of an image pair. In one task, the observers estimate the perceived difference between two images with a numerical scale defined by a reference image pair, and in another task, the observers estimate the difference in quality between two images with a numerical scale defined by a reference image. The answers were given on a scale with 0 indicating no difference or no difference in quality in an image pair and number 20 indicating the same difference or difference in quality as in the reference image pair.

Furthermore, both of these tasks are important in the area of image quality estimation, since sometimes the objective is to determine whether some artifacts are detectible, and sometimes whether the observers' preferences have changed due to these artifacts. Therefore, in image quality studies, it is important to understand how the instructions change viewing strategies. In our study, we used natural image material having small global changes in image quality. The difference estimation in this case is not a simple search task with one correct answer, but requires scanning and evaluating all changes in the different parts of an image. Image quality estimation with high-quality material is a preference task with no correct answer. The main emphasis of the study is not the subjective ratings the observers give in different tasks, but whether the tasks of quality and difference estimation show differences in strategies.

We examine viewing strategies from the perspective of both the temporal information from eye movements, such as fixation durations, and the spatial allocation of attention. To clarify the allocation of attention in different tasks, we calculated two measures concerning image areas: the semantic ROIs and the low-level saliency of the images. These measures are used to define different bottom-up influences: one concentrating on semantical meanings of the image and the other on low-level image features, such as color, intensity, and orientation. The phenomenon of change blindness has underlined the importance of semantically significant areas by showing that observers often fail to detect quite large changes if they are outside the ROIs, defined as semantically meaningful image areas.[30] On the other hand, some models suggest that it is possible to estimate visual attention allocation on the basis of available low-level image features.[31,32] We formed five hypotheses about the differences between the tasks.

H1: Fixation durations are shorter in a quality estimation task than in a difference task

Shorter fixation durations have been found in a preference task when compared with memory and free-viewing tasks.[23] Shorter fixation durations mean less information is needed per fixation.

H2: Quality estimation tasks show a global viewing strategy demonstrated by long saccade lengths

A global viewing strategy would mean short fixations and many relocations, many relocations, similar to a preference task.[23] Therefore, a vast area would be scanned. When estimating quality, fewer fixations have been made on ROIs than in a free-viewing task,[24] which could be in line with the global viewing strategy since there is a less detailed examination of small areas; however, this would have to be confirmed.

H3: There is a difference between spatial allocation of attention in an image quality task and difference estimation task

Earlier differences in spatial attention allocation have been found in the tasks of free viewing and quality estimation.[24,25] In this study, we want to confirm this difference in strategies due to tasks that are more similar than free viewing and quality estimation. We will use two magnitude estimation tasks, where the goal of one is to estimate the magnitude of the quality difference and that of the other is to estimate the magnitude of the perceived difference.

H4: Attention is allocated to areas of semantic interest more in a quality estimation task than in difference estimation task

We hypothesized that because preferences are context-dependent[33] and interpretations of quality changes differ depending on the content,[10] these would be seen in the concentration of attention on semantically important image areas.

H5: Attention is allocated to areas of high saliency estimated from the low-level image features more in a difference task than in a quality estimation task

We hypothesized that a difference estimation task would direct the subjects' gaze more to areas of high saliency, since there is no reason for context dependency and areas with artifacts are important.

2 Methods

The observers completed two visual tasks during which their eye movements were recorded. The first was a memory task, which was used to define the semantic ROIs. Following this, all observers were shown pairs of images with a nonprocessed (A) and processed (A') image in an adjusted flicker paradigm setting (A, A', A). Half of the observers estimated the perceived difference (hereafter the difference task) and the other half the perceived quality difference (hereafter the quality estimation task). The processing of the images was selected to show the variety of artifacts common in optical systems. The artifacts were small global changes.

2.1 *Participants*

Twenty observers participated in the experiment. They reported to be naïve in image quality estimation. All had normal or corrected to normal vision, which was tested for near visual acuity, near contrast vision (near F.A.C.T.), and color vision (Farnsworth D-15). Four observers were excluded from further analysis due to problems in eye tracking data (calibration or missing data). The final number of observers was 16 (8 for both tasks). One observer did not perform the memory task, and thus, the number of observers in this task was 15. Six observers were male and 10 female. The mean age of the subjects was 24 years (the youngest 20 and the oldest 28). They were recruited from Helsinki University's students' e-mail lists and each received two movie tickets for their participation.

2.2 *Stimuli*

In this study, we used the term content to refer to one scene and the term image to refer to different versions of one content scene. Altogether, there were seven different contents (Figs. 1 and 2). The contents presented close-ups of people [Figs. 1(a) and 2(a)], people further away with many surrounding details [Figs. 1(b) and 2(b)], a town scene with people [Fig. 2(c)], a town scene without people [Fig. 1(c)], and a nature scene [Fig. 2(d)]. The image content woman [Fig. 2(b)] was a test image specifically designed and developed for the evaluation of color still image processes.[34] The contents boy [Fig. 2(a)] and scenery [Fig. 2(c)] were also captured and utilized for the purposes of image quality evaluation.[35] The Belgian café [Fig. 1(c)] is the content that the ISO recommends for evaluating the results of image processing.[36] The remaining three images [children, town, and party, Figs. 1(a), 1(b), and 1(c)] were chosen to present different photographic contents, such as people close-up and at a distance, inside and outside images, and small details and scenery. These images formed image pairs, where image A was the original image and image A' the processed version of the same content. Image processing was performed differently for two different content groups, group A (Fig. 1) and group B (Fig. 2), to obtain more variability in the contents

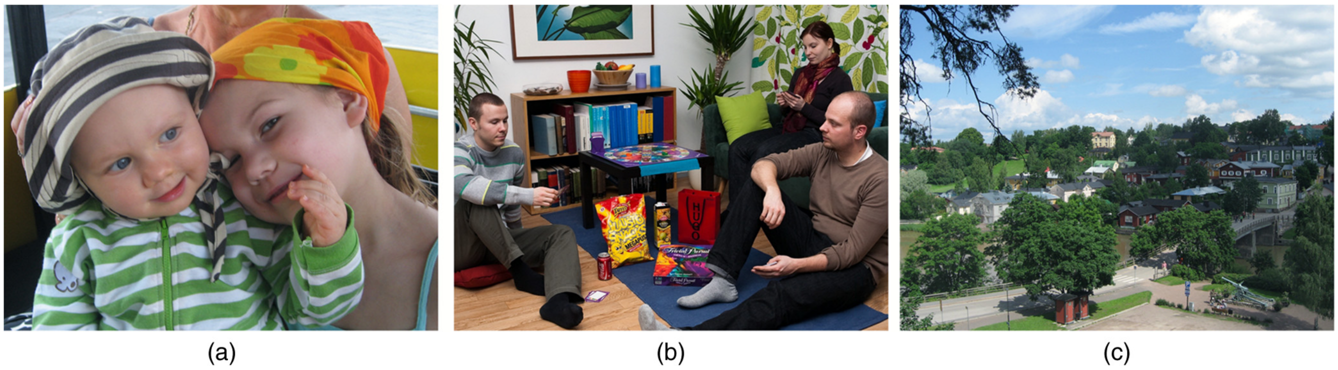


Fig. 1 The contents in group A were (a) children, (b) party, and (c) town. These contents were manipulated by using blur, noise, white point, and luminance (plus or minus).

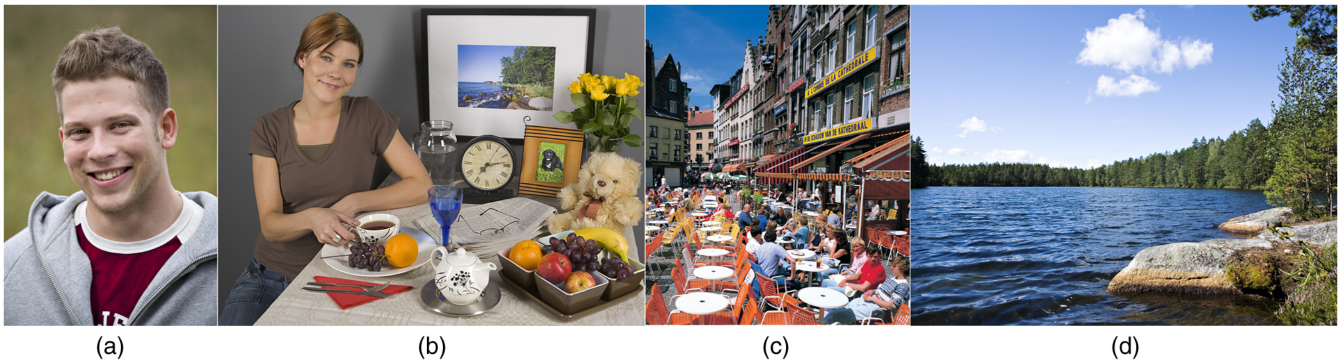


Fig. 2 The contents in group B were (a) boy, (b) woman, (c) Belgian café, and (d) scenery. The contents in group B were manipulated by using different JPEG bitrates (0.1068, 0.21173, and 1.708 bpp).

and manipulations. The groups were defined so that different content types were present in both groups (close-ups of people, people further away, and scenery).

We selected both structural (blur, noise, and jpeg-compression) and nonstructural (white point, increased, and decreased luminance) manipulations. The nonstructural manipulations are especially important for understanding the estimations of high-quality images since these do not change the discriminability of the image. Three contents (group A, Fig. 1) were processed with five different types of processing: blur, noise, the white point, and increased and decreased luminance. The images were processed in a MATLAB® software (MathWorks Inc., Massachusetts) environment using the image processing toolbox. Noise manipulation was performed using the imnoise function with Gaussian additive noise with a mean of zero and a standard deviation of 0.005. Image blurring was performed using the imfilter function with a Gaussian low-pass filter with a standard deviation of 0.75. Luminance manipulation was performed by first transforming the image from the sRGB color space to the La*b* and adding a value of 10 or subtracting a value of 12 from the L channel and then transforming the image back to the sRGB color space. White point manipulation was performed by adjusting the color temperature of the images to simulate the incorrect setting of the white point of the camera: the white point of the camera was either at 5100 K when the color temperature of the illumination was 6500 K (town, children) or at 5500 K when the color temperature of the illumination was 2700 K (party). For four image contents (group B, Fig. 2), JPEG2000 compressions were made with a publicly available codec, Kakadu 6.0,[37] using three different bitrates: 0.1068, 0.21173, and 1.708 bpp.

These images from both content groups A (3×5) and B (4×3) formed 27 image pairs, with an original and a processed image in each pair. The images were presented using the full screen (1600×1200 pixels), which means they subtended a visual angle of 30.8 deg × 23.3 deg at a viewing distance of 80 cm. Two contents had a small horizontal width (Belgian café: 991 pixels, visual angle of 19.1 deg; boy: 800 pixels, visual angle of 15.4 deg). They were presented at the center of the screen with middle gray areas to the sides.

2.3 *Apparatus*

The stimuli were displayed on an EizoColorEdge CG210 (EIZO Corporation, Ishikawa, Japan) 21.3 in. monitor using an NVIDIA GeForce 8800 GT graphic card with a refresh rate of 60 Hz. The display was viewed in a darkened midgray tent with dim background lighting from behind at 30 lux and 5300 K on the surface of the display. The calibration was set for gamma 2.2 and was measured with an Eye-One Monitor/pro, (X-rite, Michigan), which gave a white point of 80.1 $\mathrm{cd/m^2}$, CIE [x,y], 0.31, 0.33.

Eye movements were monitored using a Tobii ×120 standalone eye tracking device (Tobii Technology, Stockholm, Sweden), with an accuracy of 0.5 deg and a data rate of 120 Hz. The distance of the observer was checked at the beginning of the experiment when calibration was performed for the eye tracker, so that the observer's distance from the eye tracker was ~65 cm and that from the monitor was ~80 cm. The observer's distance and head position were checked at every stage of the experiment. The observers were instructed to remain in approximately the same position as they were when calibrating the eye tracker. For calibration, the observer fixated on a dot appearing at five points on the display.

2.4 *Procedure*

Figure 3 presents a flow chart for the procedure. The observers first read general instructions, which explained that the

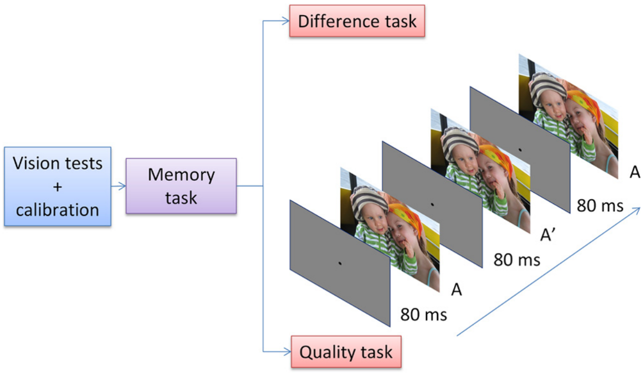


Fig. 3 A schema of the procedure and the adjusted flicker paradigm showing that the first stages were the same for all observers, after which they were divided into two groups with different tasks that, nevertheless, used the same image material and the same flicker paradigm, where a nonprocessed image (A) was shown first, after which a processed version of the same content (A') and the nonprocessed image (A) was shown again. Before each image, a fixation point on a mid-gray background appeared for ~80 ms. The observers controlled the viewing time of the test images themselves by pressing a button when they were ready to move on.

research was investigating visual perception in photographs using eye movement tracking. Then their vision was tested. The observers faced a monitor where the images were presented; the eye tracking device was below it. Below the eye tracking device was a laptop computer on which they entered their answers. Responses were made on the laptop only when no test image was present on the display screen. After the vision tests, the experiment leader explained the memory task, checked the observers' head position, and calibrated the eye tracker.

First, the observers completed a memory task, which was used to define the semantic ROIs. Only the nonprocessed images (A) were presented in this task (altogether seven images, one of each content). The instructions were to look at the image and later write down what was in it as if describing the objects in the pictures to someone who had not seen it. When the observers felt that they had seen the image long enough, they clicked the button on a mouse and a gray screen with information on the image number and the instruction to write down the answer in a free text field appeared. The experimenter went through the first image together with the observer to make sure that the observer had understood the instructions. Then the observer continued the task alone. Five different randomizations were used in the experiment. This meant that the same randomization was used for four observers, two for each task. Before each image was shown, a black fixation point appeared in the middle of the screen on a gray background for ~80 ms.

After the memory task, the participants were divided into two groups: a difference task group and a quality estimation task group. The experimenter then gave new instructions. For the difference task group, the instructions were to estimate how large the change in an image pair was; for the quality estimation group, the instructions were to estimate how large the change in quality in an image pair was. The setting was modified from Ref. 17. The scales for the estimations were determined by a reference image pair (Fig. 4), which was shown at the beginning of the task, after four practice contents, and throughout the test as every tenth image pair. The participants were informed that the numerical amount of change or the amount of change in quality in the reference image pair was 20 and that the value 0 described no visible difference between the two images presented. The reference image content was a picture of a parking lot. The reference images were chosen to show a moderate change in quality in multiple image artifacts. For this purpose, two images processed with different imaging pipelines creating optical artifacts were chosen, since these images showed simultaneously moderate changes in colors as well as sharpness and graininess (Fig. 4). Value 0 was defined as no visible difference or no visible difference in quality and the value 20 was defined only by the image pair, since the graphical scales with quality terms associated with different steps cannot be divided into intervals of equal size.[38] The test image pairs were presented in five different random orders, so that four observers always did the test with the same randomizations, two observers from each task.

The observers viewed image pairs in a setting that resembled the flicker paradigm[30] or the three-interval paradigm[17] commonly used in change detection studies. Each observer first saw the original image (image A) and then a processed version of the same image (image A') followed by the original image again (image A) (Fig. 3). Contrary to the typical flicker paradigm, the observers themselves decided on the length of time they would look at each image by pressing the button on the mouse. We considered it important that the viewing time would not be fixed, since different viewing tasks have been shown to require different amounts of time to finish.[39] They saw each test image pair only once. Before each image was shown, a fixation point appeared in the middle of the screen on a mid-gray background for ~80 ms. After watching all three images, the number of the image pair and the instructions to answer the questions appeared. For practice, the experimenter and observer together went through four practice contents that were different from the test image contents. The material was presented in the same way in both tasks.

2.5 *Eye Movement Data Analysis*

In the change and quality estimation tasks, only the data from fixations on the processed images (images A') that were inside the image area were included in the analysis. Two consecutive data points were calculated to be in the same fixation if they were within a 35 pixel (visual angle of 0.67 deg) radius of each other. The first fixation was defined as the first fixation that started after the test image had appeared. In addition, fixations lasting <90 ms or >2000 ms were removed from the data as outliers.[15] This meant that 2.9% of the fixations were removed (2.5% <90 ms and 0.4% >2000 ms). When examining the strategies used in different tasks (when analyzing fixation durations and saccade lengths), we removed the last fixations from the analysis, since these were the fixations when the mouse button was pressed to move to the next image and the function of the

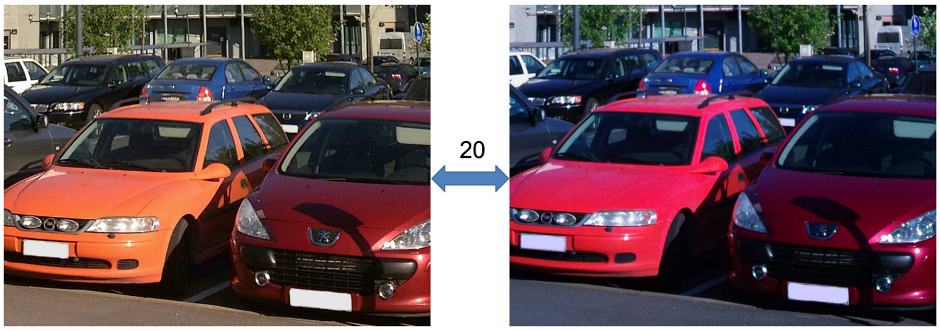


Fig. 4 The reference image pair gave a reference value of 20 for the change or quality difference.

fixation probably was not the same as with the other fixations. It has been suggested that the final fixation before the execution of task-related movement is connected to the tasks' cognitive demands.[40] We, however, kept this fixation in the analysis when examining the spatial distribution of fixations, since we cannot state that the processing of image ends before this final fixation. The saccade amplitudes were calculated in visual angles using Euclidean distance.

2.6 *Semantic Regions of Interest*

The semantically meaningful places in the contents were calculated from the memory task to show the areas on which the observers fixated the most when examining semantically meaningful image areas. The instructions were to watch the images and then write down the most important things about them as if they were describing the images to someone who had not seen them. For defining ROIs in images, similar separate, brief verbal descriptions were used in a seminal study on change blindness.[30] However, they used the verbal descriptions to define the ROIs and we used the fixations recorded while performing the task. Defining the ROIs allows for a comparison of the other two tasks from the perspective of how much the semantically important areas in the images are attended to.

To define the regions of the semantically important areas, a fixation distribution map of each image was convolved with a Gaussian kernel. The full width at half maximum (FWHM) of the Gaussian kernel that defined the size of the patch was set to a visual angle of 2 deg (104 pixels).

$$\mathrm{FWHM} = 104/2\sqrt{2\ln 2}.$$

Each fixation was weighted according to its duration and the Gaussian filter approximated the area of accurate vision. In other words, the Gaussian filter was calculated with the standard MATLAB® (MathWorks Inc.) function fspecial, where the FWHM was the standard deviation and the size of fixation was its duration. From the resulting fixation density map (FDM), we defined the regions where the concentration of fixations was high. From the FDM, the value of 0.25 on the z axis was defined as the cut-off point for ROI since our qualitative examination of the distributions showed this to be the single value that best suited most of the images. When comparing the tasks, we calculated how many fixations were present inside this area defined as the ROI.

2.7 *Low-Level Salient Areas*

We used Saliency Toolbox 2.2 (Ref. 41), which was downloaded in July 2011. The toolbox is based on the modeling work of Itti and Koch[31] and was modified by Walther and Koch.[32] It models attention with a biologically plausible model that calculates the salient areas of images using contrast, color, and orientation information. The saliency toolbox compiles a 1/16 cell saliency map for the images, which shows the salient areas of a given image using the information on contrast, color, and orientation information weighted with the winner-take-all maps. We used the Saliency Toolbox 2.2's default settings (color, intensity, and four orientations with a weight of 1) and added a skin color feature with a weight of 1, since our images contained human beings. The parameters of the pyramid levels were dyadic, the normalization type was iterative and the number of iterations was 3, and the shape mode was a feature map. The saliency maps were calculated only from the nonprocessed images (A), and since the toolbox provides a 1/16 cell saliency map, this map was converted into the size of the original images using nearest-neighbor interpolation. The areas that had positive saliency values were defined as salient image areas. When comparing the tasks, we calculated the fixations within the salient areas.

2.8 *Fixated Area*

The fixations from all observers in one experimental group (eight observers in both groups) were combined and two fixation distribution maps were compiled in the same manner as when defining the semantic ROIs. The cut-off point selected from the FDM was 0.02 on the z axis, since our qualitative examination of the distributions showed this to be the single value that best suited most of the images.

2.9 *Statistical Tests*

The data were normalized for the subjective evaluation of difference and quality difference to make the distributions of the evaluations comparable between observers. The evaluations of perceived difference or the difference in quality were divided (scaled) by the observer's median evaluation. The median was chosen because it was not possible to assume that the estimations would be normally distributed.[17]

Since the data were not normally distributed and there were dependencies due to repeated estimations of different images, we used generalized estimating equations (GEEs) for the analysis. GEEs can be used for non-normal correlated data and with data having missing values, since they use within-cluster similarity of the residuals to estimate the correlation in order to reestimate the regression parameters and to calculate standard errors.[42] With GEEs, one can select the distribution that fits the data: for image duration, fixation duration, first fixations duration, and saccade amplitudes; the distribution was defined as gamma with a log-link, and for the number of fixations, it was defined as a Poisson distribution with a loglinear link. For subjective magnitude estimations, the distribution was defined as both a tweedie and link function identity. For fixation durations and saccade amplitudes, the means from one person's eye movements on one image were used. In the analysis, the subject was defined as a subject, the within-subject variable was the image, and the factors were defined to be image contents and processing types as well as the task.

To compare the two groups in terms of the areas fixated on, and the proportion of fixations on ROIs as well as on salient areas, we used generalized linear models (GLMs) for the averages across all participants performing the same task on each image, since by using a link function that defines the relationship between the systematic component of the data and the outcome variable, they can deal with the data that do not fulfill the requirements of normality.[43] This was due to the distribution of values: when examining the proportions of single images, the distributions were nonnormal because they had many values of 1 (in ROIs) or 0 (in salient areas). By taking the means per image, the distributions began to approach normal as would happen according to the central limit theorem. GLMs do not require normality and can deal with categorical predictors. The probability

Table 1 The medians of variables describing strategy in the tasks of quality estimation and difference estimation are presented in the table as well as the significance of the comparison between tasks.

	Quality estimation	Difference estimation	*p* value
Viewing time (ms)	2465	3914	**
Fixation durations (ms)	333	319	ns
Saccade amplitude (deg)	5.4	6.1	**
First fixation duration (ms)	350	300	**
Fixation count	6	10	**
Area fixated on (%)	15.2	26.1	***
Proportion in salient areas (%)	14.0	15.0	*
Proportion in regions of interest (%)	75.4	64.8	***

Note: ns, nonsignificant, *$p < 0.5$, **$p < 0.01$, ***$p < 0.001$.

distribution for the proportion of fixations on ROIs and salient areas was normal and, for the area fixated on, Gaussian.

3 Results

Table 1 summarizes the results of the different tasks by giving the median for each variable as well as the significance of differences between the groups. Below is a more detailed explanation of the results.

3.1 *Estimation Task*

The starting point for comparing the tasks was to test whether the task affected the estimated magnitude of differences and quality estimations. The results indicate that the observers' evaluations of the magnitude of difference or quality did not differ between the tasks [Wald $\chi^2(1) = 1.0$, $p > 0.05$]. This means that if the observers had a different strategy, it would not influence the estimations.

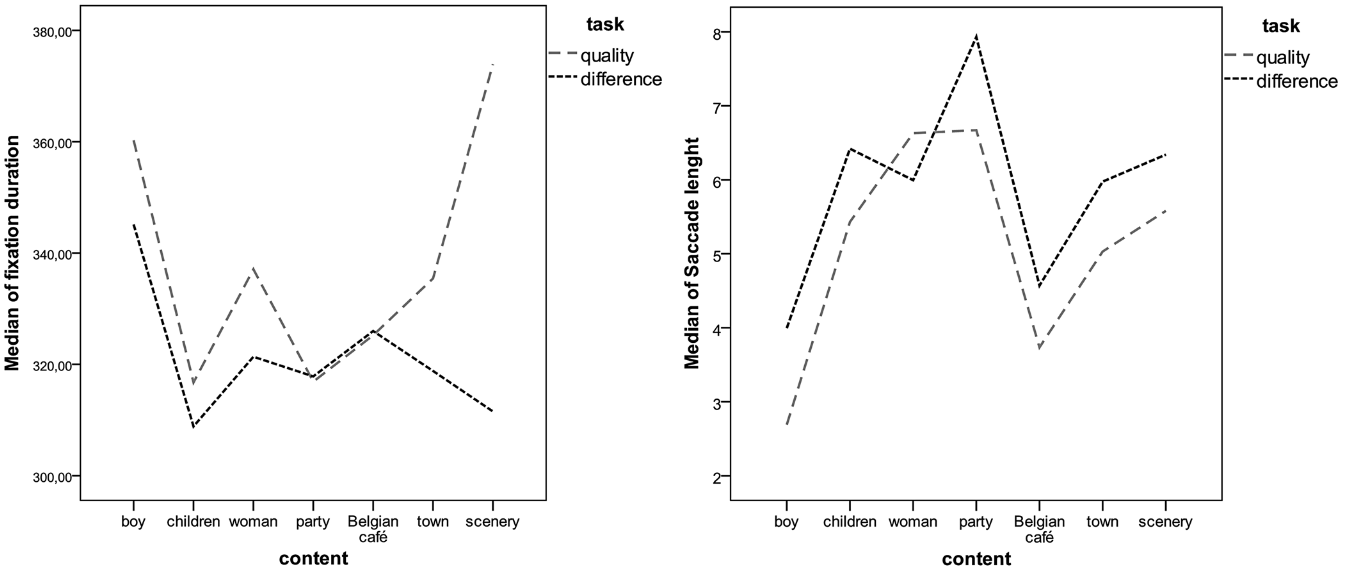
3.2 *Description of Differences Between Viewing Strategies*

Our observers spent more time watching the processed image (image A') in the difference task (median 3.91 s) than in the quality estimation task (median 2.47) [Wald $\chi^2(1) = 9.2$, $p < 0.01$] and needed more fixations in a difference task [Wald $\chi^2(1) = 7.2$, $p < 0.01$]. In other words, the difference task needed more time than the quality estimation task.

However, for fixation duration averages per image per observer, the main influence of the task was not significant [Wald $\chi^2(1) = 1.1$, $p > 0.05$), but the interactions between the task and the contents [Wald $\chi^2(5) = 19.8$, $p < 0.01$] as well as the task and the manipulation [Wald $\chi^2(6) = 13.6$, $p < 0.01$] were significantly different [Figs. 5(a) and 6(a)]. Therefore, the differences due to the task in fixation durations are visible only if the type of the test material is taken into account.

We further examined the durations of the first fixations, since when viewing a scene the initial fixation provides an abstract scene representation that is used to plan subsequent eye movements through the scene.[44] The first fixation duration reflects the immediate information processing, in this case, the processing of the first actively chosen fixation point.[45] The first fixations were longer in the quality estimation task (median 350 ms) than in the difference task (median 300 ms) [Wald $\chi^2(1) = 7.5$, $p < 0.01$]. Planning where to look next consequently took longer in the quality estimation task than in the difference task.

The saccades' average length per image was longer in the difference task (median 6.1 deg of visual angle) than in the quality task (median 5.4 deg of visual angle) [Wald



(a)

(b)

Fig. 5 The medians of fixation durations (a) and saccade lengths (b) are presented as a function of task and image content.

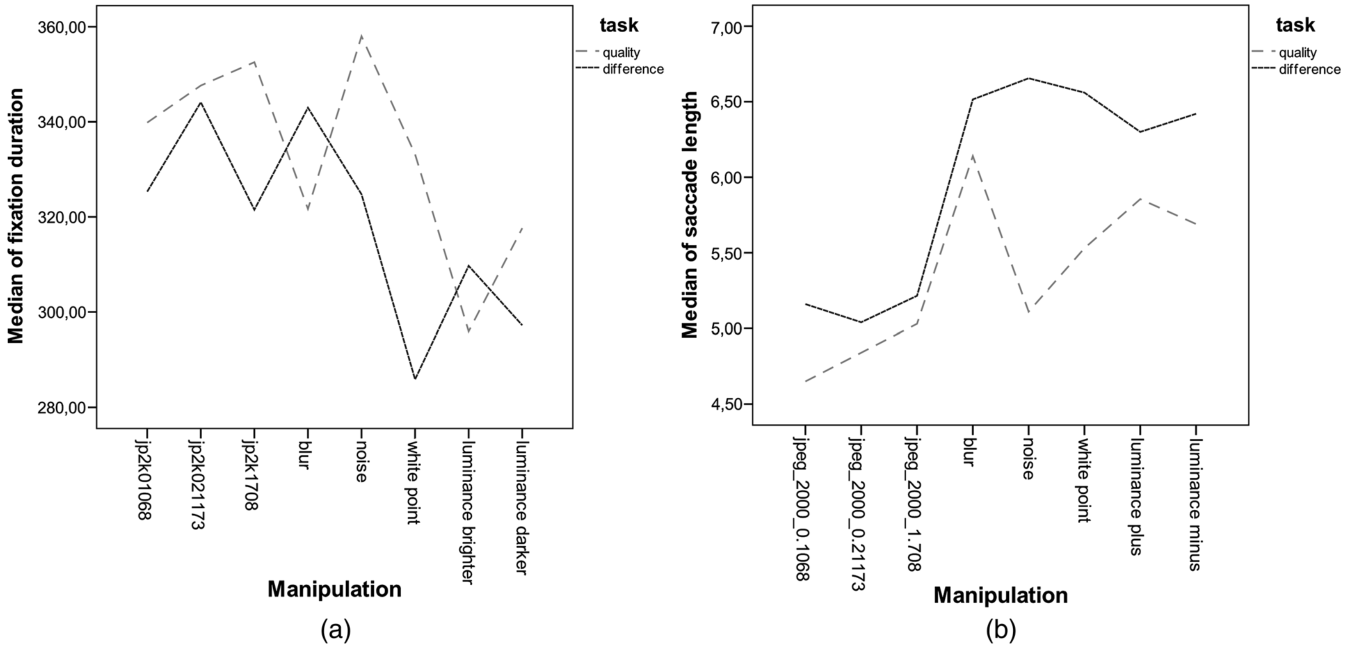


Fig. 6 The medians of fixation durations (a) and saccade lengths (b) are presented as a function of task and image manipulation.

$\chi^2(1) = 8.7$, $p < 0.01$]. The average saccade lengths' interaction with task and content was significant [Wald $\chi^2(5) = 15.4$, $p < 0.01$], but interaction between task and manipulation was not [Wald $\chi^2(6) = 2.3$, $p > 0.05$] [Figs. 5(b) and 6(b)]. This means that in the difference tasks, the viewing consisted of fixations that were further apart and there was less detailed examination with repeated fixation in one area than in the quality task.

Furthermore, as a group the observers' fixations were distributed over a larger area in the difference task than in the quality estimation task (medians: 26.1% of the image area in the difference task and 15.2% in the quality task) [Wald $\chi^2(1) = 232.0$, $p < 0.001$].When comparing the areas covered by the different tasks at a group level, we noticed that the area attended to in the quality estimation task was on average 16.7% of the whole image area, and only an average of 4.1% of this area was not covered by the fixations in the difference task. This means that in the difference task, the observers fixated on the areas considered important in the quality estimation task. However, they also fixated on a large area outside of this region.

3.3 *Semantic Regions of Interest and Low-Level Saliency*

To examine how important the semantically informative areas of the images were, we calculated the semantic ROIs for the image content. Here we defined the semantic ROIs as the meaningful places that convey the message of the image measured by the eye movements from the memory task. Figure 7 shows the semantic ROIs with rounded yellow contours. In the portrait images [Figs. 7(e) and 7(f)], the semantic ROIs were mostly focused on the faces. Figure 8 shows the proportion of the image area that belonged to the semantic ROIs, salient areas, or both. It also shows that the area of the semantic ROIs varied from 6.1% (boy) to 41% (Belgian café) of the image area (Figs. 7 and 8).

To compare the semantic ROIs with the prediction based on low-level saliency, we used a saliency model to calculate the salient areas in each image.[31,32] These areas are shown with angular green contours in Fig. 7. The salient areas were widely distributed across the entire image areas. Figure 8 shows that the salient areas per image content changed from 10.4% for the content woman to 6.4% for the image scenery. When comparing semantic ROIs and salient areas, it must be kept in mind that the starting point for these measures is different, since the semantic ROIs are based on a memory task and the saliency models are developed using visual attention based on a search task as the starting point.[31] Therefore, these describe different areas of images and are suitable for examining whether the tasks of difference estimation and quality estimation use high-level or low-level image features as bases for their evaluations.

Figures 7 and 8 also show the amount of overlap between semantic ROIs and salient areas. The area of overlap was the largest for the contents woman [Fig. 7(a); 4.8% of the area] and party [Fig. 7(b); 4.3%] and smallest for the content children [Fig. 7(e); 0.7%] (Fig. 8). Semantic ROIs and low-level saliency displayed the least amount of overlap in images where a large part of the image area consisted of one or two faces [Fig. 7(e), children and Fig. 7(f), boy], which is consistent with earlier results suggesting the dominance of social cues over saliency cues.[46] Here social cues mean information about humans and their relations.

We calculated the proportion of fixation durations on the salient areas and semantic ROIs compared with all fixation durations in the different tasks. A significantly smaller proportion of time was spent in the salient image areas than in the ROIs (Figs. 9 and 10). The task had an effect on fixation locations, as the proportion of time inside the semantic ROIs was higher in the quality estimation task than in the difference task [Wald $\chi^2(1) = 251.0$, $p < 0.001$] and vice versa in the salient areas [Wald $\chi^2(1) = 4.3$, $p < 0.05$] (Figs. 9 and 10). Further, the task and the content had a

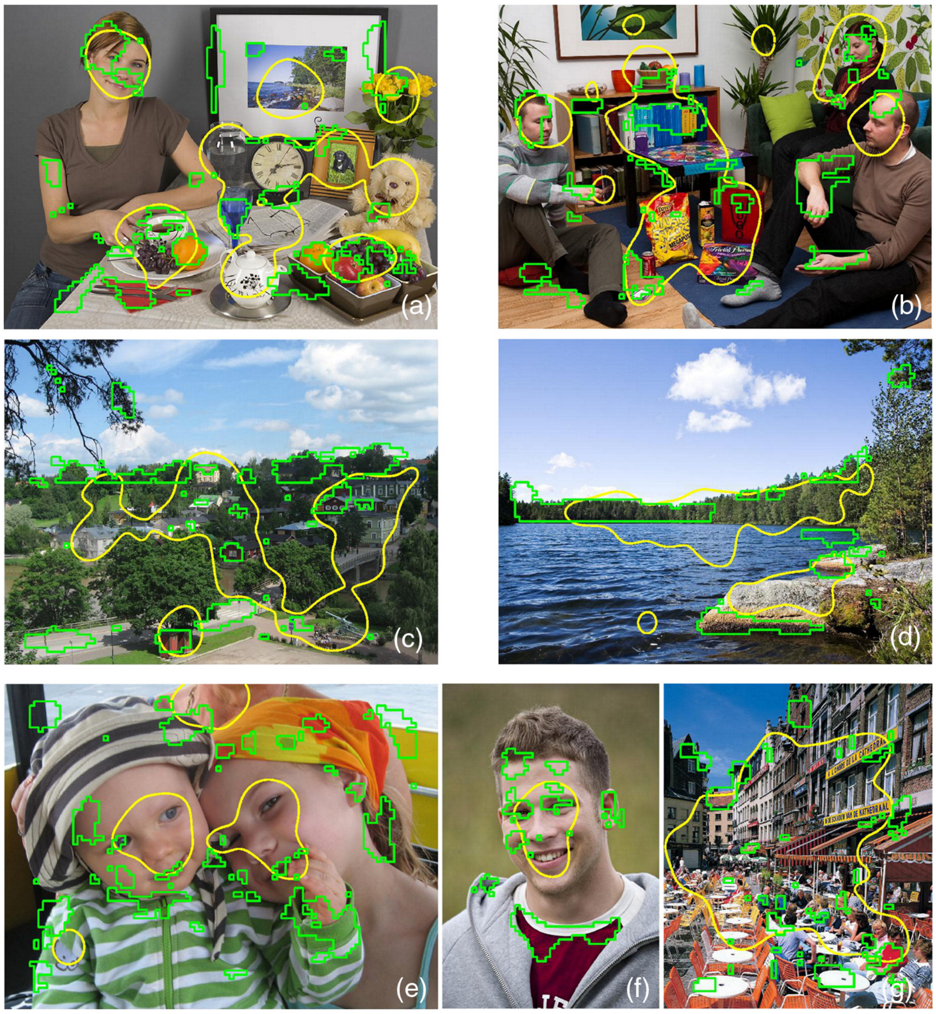


Fig. 7 The semantic regions of interest (ROIs) (rounded yellow lines) and low-level salient areas (angular green lines) are shown for all test image content: (a) woman, (b) party, (c) town, (d) scenery, (e) children, (f) boy, and (g) Belgian café. Only the nonprocessed images (A) were used for calculating the semantic ROIs and saliency.

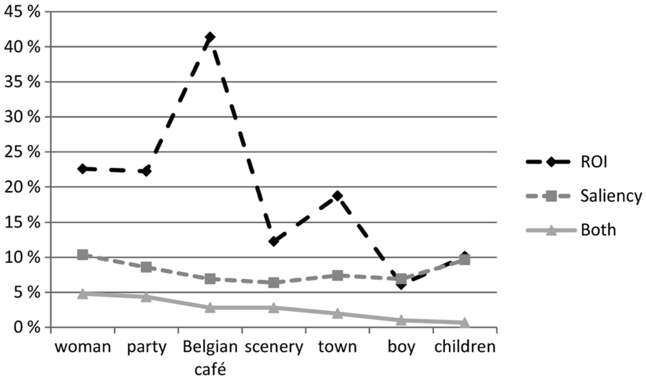
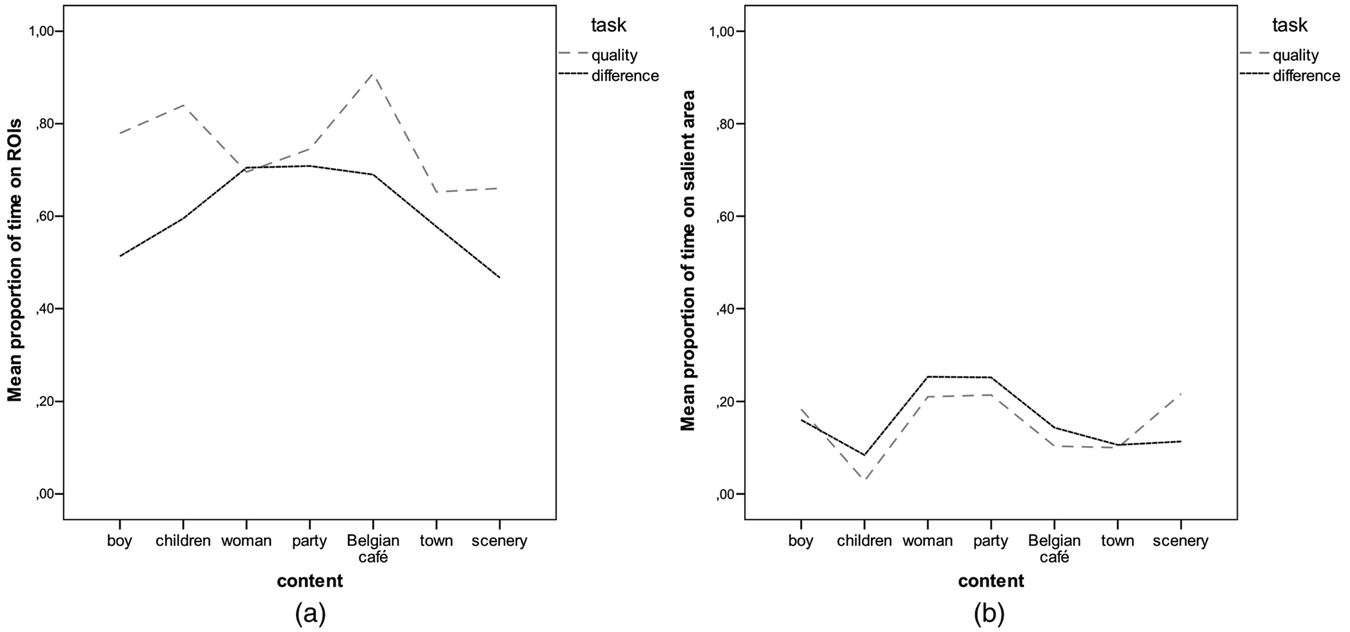


Fig. 8 The proportion of the image covered by semantic ROIs and salient areas, as well as the proportion of the image that was both a salient and a semantic ROI. The image contents are arranged according to the proportion of the areas defined by both measures with the contents having the largest common area on the left and the smallest on the right. The areas were calculated only from the nonprocessed images.

significant interaction both in the proportion of time spent on semantic ROIs [Wald $\chi^2(5) = 118.2$, $p < 0.001$; Fig. 9(a)] as well as on salient areas [Wald $\chi^2(5) = 52.6$, $p < 0.001$; Fig. 9(b)]. It seems that the attention allocation differs most between the tasks when the contents have strong attention attracters, such as faces. This might reflect the fact that in the difference task, the attention is actively allocated outside the regions of semantic interest to find the differences, whereas in the quality estimation, there is no need to avoid making the judgment based on semantically strong image areas. Also, the processing influenced attention allocation differently depending on the task [semantic ROIs: Wald $\chi^2(6) = 16.2$, $p < 0.05$; Fig. 10(a); salient areas: Wald $\chi^2(6) = 44.5$, $p < 0.001$; Fig. 10(b)]. Especially when luminance is processed, quality seems to be estimated from the semantically important areas more than difference. Therefore, the semantically important areas are more important than low-level saliency, and the semantically important areas appeared to be more important in the quality estimation task than in the difference task.



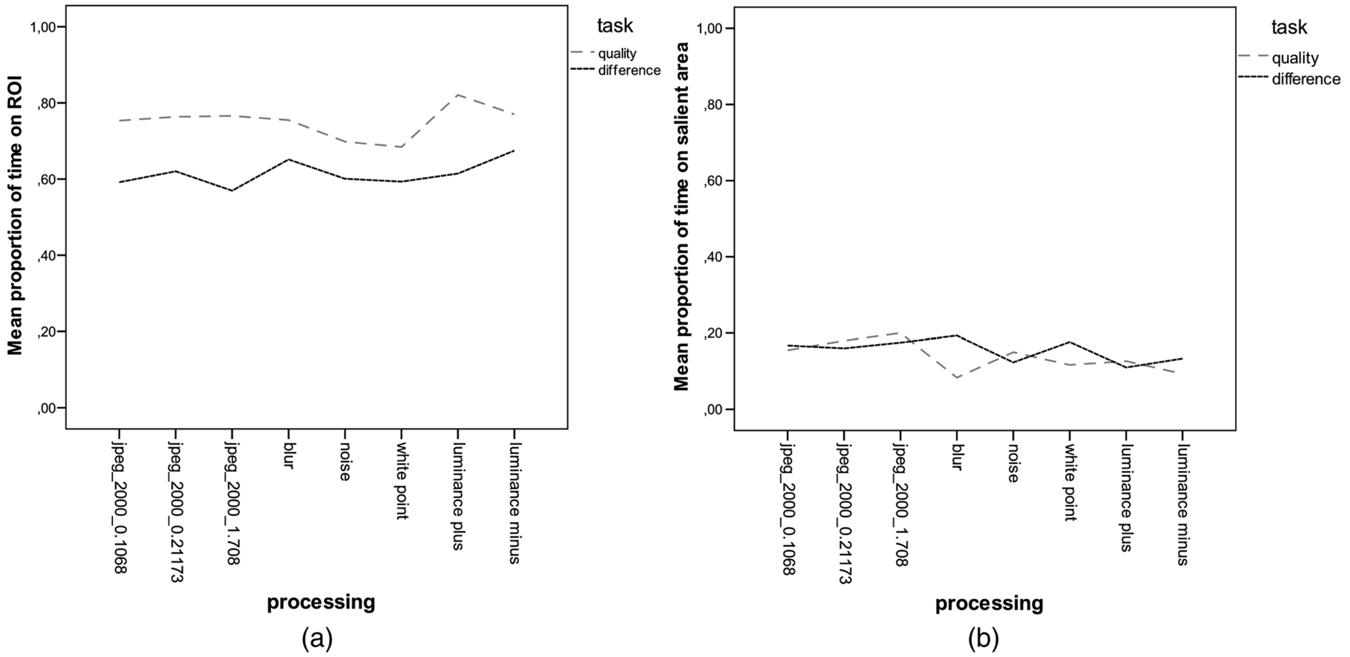
(a) (b)

Fig. 9 The average proportions of time spent on semantic ROIs (a) and salient areas (b) of all time spent watching images per image contents.

Further examination revealed that the salient areas were attended to mainly if they were also within the ROIs (Fig. 11). The proportion of fixations in the salient areas and outside the semantic ROIs was only 1.7% in the quality estimation task and 3.5% in the difference task [Fig. 11(a)]. However, although the area that was both salient and within the semantic ROIs comprised, on average, <5% of the whole image area (Fig. 8), it nevertheless accounted for 13.8% of the fixations in the quality estimation task and 12.0% in the difference task [Fig. 11(a)]. Therefore, the salient areas of the image are important only if they are also semantically important. This means that the semantics of an image overrule the low-level saliency even in the tasks where the content should not be important for the evaluation, which is the case for the difference task.

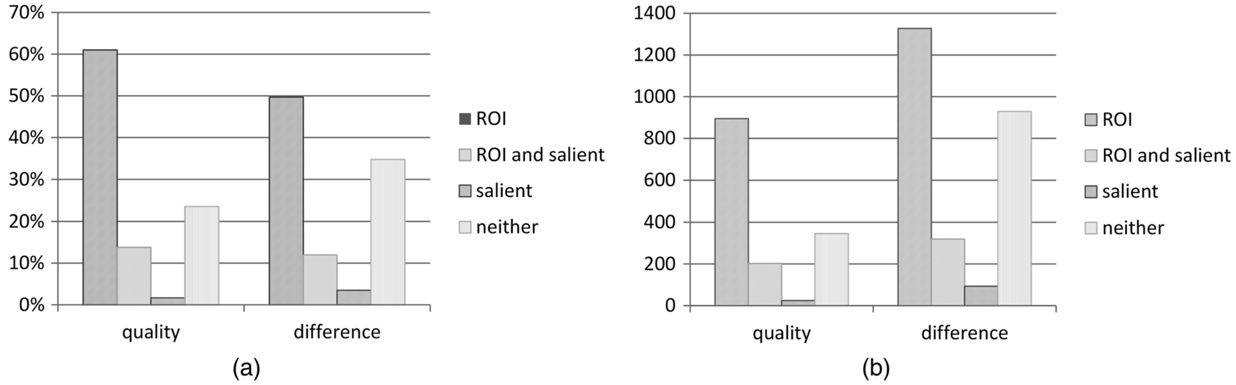
4 Discussion

The results show that the observers rated the perceived difference and quality difference equally in both tasks, but that the strategies for doing so were different. In the quality



(a) (b)

Fig. 10 The average proportions of time spend on semantic ROIs (a) and salient areas (b) of all time spent watching images per manipulation.



(a)

(b)

Fig. 11 The proportion (a) and number (b) of fixations within different image areas. The figures show that only a small proportion of fixations was present in areas that were salient but outside the ROIs.

estimation task, there were shorter viewing times, fewer fixations, smaller areas attended to, and shorter saccade amplitudes than in the difference task. Additionally, the context-dependency of the quality estimation task was visible in our results since in the quality estimation task, the fixations concentrated heavily on semantically meaningful image areas. The difference task concentrated slightly more on salient image areas than the quality estimation task. Therefore, we can say that only a small difference in the instruction influenced the strategy used to view the images.

However, even though we found differences in viewing strategies, not all accorded with our hypothesis. Considering the fixation durations and the global viewing strategy, the findings were opposite to the hypothesis. Our results show that in fixation durations, the difference between the tasks was visible only when taking the image content into account, and we hypothesized them to be shorter in the quality estimation task, as they were in a previous study concerning preference estimation[23] (H1). Furthermore, our results do not support the hypothesis that the viewing strategy would be global in a quality estimation task (H2), since the saccade lengths were shorter in the quality estimation task than in the difference estimation task, and furthermore, their median was <7 deg. We also hypothesized that there would be a difference in the spatial allocation of attention in the image quality task compared to the difference estimation task (H3), which was confirmed. As well, the fixation concentrated more on semantically important regions in the quality estimation task than in the difference task (H4), and less on salient image areas (H5) as hypothesized. However, it must be kept in mind that the proportions of fixations on salient image areas in the difference task were still low (14% median) and the salient areas also had to be semantically important to gain attention.

4.1 *Differences in Viewing Strategies*

The length of a fixation is related to how much information needs to be extracted from an area.[47] Fixation duration can, therefore, be related to the information retrieval requirements of the task. In our study, there was no difference in fixation durations between the tasks. This is similar to a result that found no difference in fixation durations between the fixations on impaired image in the image quality estimation task and free viewing of the original image.[25] However, pleasantness and search tasks have earlier been reported to have shorter fixation durations compared with free-viewing and memory tasks, which had long fixation durations.[23] The difference between the findings in Ref. 23 and ours might indicate that the magnitude estimation of image quality differences between two images requires more processing on each fixation than when simply estimating the pleasantness of one image. In our quality estimation task, the observers had to remember the previous image to be able to answer the quality estimation question.

Even if there were no differences in fixation durations between the tasks, the interaction between image content and fixation duration was significant. The fixation durations were longer in the quality tasks with contents such as scenery and town, which did not have strong attention attracters, like human forms, faces, text, or clear objects. Also, the fixations concentrated more on semantically important areas in portraits in the quality estimation task than in the difference task. The importance of image contents for the estimation of quality had been noted earlier when the spatial distribution of fixations was found to be more useful for objective quality metrics if the images had strong or medium attention attracters.[48] Also, the fixation durations were different depending on the manipulations and the task. It seems that the structural manipulations need longer fixations than non-structural, especially in the difference task, which might be related to the fact that these present clearer artifacts. These results show that eye movement strategies are constantly built in the interaction between task requirements and attention attracters, both of which should always be taken into account in order to understand human viewing strategies.

Our difference estimation task required more time to complete than the quality estimation task. This might be because of the need to examine the whole image area for possible changes. At the level of single fixations, however, there were no differences between the tasks. In our study, more samples from a larger area and more time to integrate this information were needed for the difference estimation than quality estimation. Our results may show a phenomenon similar to that observed when comparing visual search and memorization tasks.[13] The study found no differences in single fixation durations, but found that the task influenced the number of fixations within a gaze at a given object.

Earlier research has found an average saccade amplitude of >7 deg for pleasantness estimation, suggesting the dominance of global processing.[23] Our study found the median saccade amplitude in the quality estimation task to be much shorter (5.4 deg of the visual angle), which might suggest more local than global scanning. Five degrees of the visual angle is considered to be a limit for parafoveal vision, and short and long saccades have been connected with different modes of visual attention.[49]

4.2 *Differences in Spatial Attention Allocation*

We hypothesized that low-level saliency could be important in the difference estimation task. We found slightly more fixations on salient areas in the difference task than in the quality estimation task. In our study, however, low-level saliency accounted for a small proportion of fixations, only 14% in the quality task and 15% in the difference task, when the proportion of fixations on the regions of semantic interest was 75% in the quality estimation task and 65% in the difference task. The importance of semantic ROIs was clearly emphasized in the quality estimation task. The meaning of semantically important areas is especially strong, since in this study, the observers had seen the image contents beforehand in the memorization task and as the first images in the image pairs. Furthermore, the areas of low-level saliency could attract attention only if they were also within semantically meaningful areas. This confirms the suggestion that saliency models mainly work because objects are often salient, and therefore, salient areas are also semantically meaningful.[50] Our results show that when modeling a task, semantically meaningful image areas should be taken into account, rather than areas of low-level saliency.

For the quality estimation, the places fixated on were mostly the semantically important areas of an image, which were then examined in detail with short saccades indicating local processing. The same semantic ROIs were attended to in the difference task, but there were also areas fixated on that were both outside the ROIs and outside the area of low-level saliency. The semantic ROIs were crucial for both tasks, as is suggested by change blindness studies.[51] Nevertheless, our results also show the influence of the task: in the difference task, ROIs alone were insufficient. The quality estimation may have depended solely on semantically important places, but for detecting differences, an evaluation of the whole image was needed and a larger area was scanned. Contrary to this, an earlier study comparing free viewing and quality estimation found that free viewing attracted more attention to the ROIs than quality estimation, where fixations were spread also outside this area.[24] This could reflect the differing demands of free-viewing and quality estimation tasks: in the free-viewing task, no aspect of an image is important for the task's sake, since there is no question to answer later, and in the quality estimation task, there is. However, the results could also be influenced by the fact that the fixations from the free-viewing task were used as the bases for calculating the ROIs. The differences between the tasks could have been related to differing cognitive requirements. The requirements have also been shown to change when the range of image quality changes. Poor image quality has been shown to be estimated using low-level attributes, such as sharpness and darkness, but high image quality is estimated using high-level attributes, such as naturalness.[9] In the current experiment, the observers estimated differences in high-quality images, so it might be that higher-level quality concepts were needed in the quality estimation more than in the difference estimation. This, however, needs further research.

4.3 *Limitations of the Study*

For the current study we used a between-subject setup, since we thought that the differences between the tasks would be so small that one task would influence the strategies chosen for the other if done by the same person. In a within-subject setup, however, the differences between individuals are easier to control. Furthermore, in this study, the number of subjects was modest, which we took into account in the analysis, where the dependencies between individuals were controlled for using GEEs, which takes into account the dependencies from repeated estimations by a subject. However, a next step for confirming the results would be a within-subject study with a wide variety of test images and more subjects.

In this study, we used only one saliency model,[32] but there are also many other models that might form different salient areas. However, we decided to use this model since it is one of the most used ones and a lot of research has been conducted using this model or the model it is based on[31] and since it estimates the saliency based on low-level features that are important for our visual system (color, intensity, orientations). Also, many studies on visual attention allocation have used these models.[46,50] The main focus in this study was to get a measure of low-level saliency for which purpose we find this model suitable.

In addition, the magnitude estimation with one reference image pair might not be the best way to get good image quality estimations, since even if the image pair presents a variety of changing image quality features, it will direct the attention to the ones that are the most obvious. However, we did not consider this a problem since the main interest was in comparing differences in viewing strategies between two tasks and not in getting the best possible image quality ratings. This method makes the magnitude estimation possible with both quality and difference tasks.

5 Conclusions

The visual tasks of quality estimation and difference estimation that we examined showed differences in viewing strategies. The quality estimation task was faster and the attention was allocated with fewer fixations on semantically meaningful image areas than in the difference task. At the level of single fixations, there was no significant difference in fixation durations between the tasks if the image content was not taken into account. It seems that the fixation durations and selected strategy also depend on the information demands of the task and that the need for information is calculated as a combination of information from one fixation and from repeated fixations on a certain area. We conclude that the strategies for accomplishing the tasks are different and that the quality estimation task is faster. However, the speed of processing was not visible in single fixations, but was instead due to the longer planning and efficient selection of fixated areas that were mostly semantically meaningful. The difference estimation task showed less planning during the first fixation, and more fixations with longer saccades than in the quality estimation task. The salient areas were not important unless they were also semantically interesting.

Our conclusion is that the fixation durations and saccade amplitudes themselves may not be enough when estimating the differences between tasks. This is a conclusion similar to that of an earlier study, which found the number of fixations within an object serves to be a better measure of the importance of the area than the length of single fixations.[13] One reason for the differences between the tasks might be the efficient selection of information needed. Further, we concluded that low-level saliency influences attention allocation only if these areas are also semantically important, which confirms the conclusion of Ref. 50.

These results show that changing one word in the instructions given to participants may change the way they direct their attention and the viewing strategy they use for evaluation. This is important to bear in mind when planning a subjective image quality experiment, especially if the results of the experiment will be used as a reference for the development of image quality algorithms. Contrary to Ref. 4, the use of free viewing as a task is not recommendable since there the observers themselves decide how they view the images and this might cause too much variation in the data, which means a requirement for more subjects. However, when choosing between difference and quality estimation tasks, one must understand that the implications might be different. In the difference task, a more thorough examination of all artifacts in an image might take place, whereas in the quality task, the areas that are semantically important and their change might be more pronounced. Therefore, when gathering or using subjective image quality data, it should be kept in mind that subjects that view images should always have a predefined task and even small changes in these instructions might change the way observers view the images.

Acknowledgments

This work was supported by the Finnish Doctoral Program in User-Centered Information Technology (UCIT) and by the Nokia Foundation. We wish to thank the Aalto School of Science, Department of Media Technology and, especially, Raisa Halonen and Professor Pirkko Oittinen for allowing us to use the material they produced, and Jussi Hakala for methodological advice. We also wish to thank Tero Vuori from the Nokia Camera team for letting us use its images as material.

References

1. H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Trans. Image Process.* **15**(2), 430–444 (2006).
2. C. Li, A. C. Bovik, and X. Wu, "Blind image quality assessment using a general regression neural network," *IEEE Trans. Neural Netw.* **22**(5), 793–799 (2011).
3. L. Zhang et al., "FSIM: a feature similarity index for image quality assessment," *IEEE Trans. Image Process.* **20**(8), 2378–2386 (2011).
4. U. Engelke et al., "Visual attention in quality assessment," *IEEE Signal Process. Mag.* **28**(6), 50–59 (2011).
5. W. Einhäuser, U. Rutishauser, and C. Koch, "Task-demands can immediately reverse the effects of sensory-driven saliency in complex visual stimuli," *J. Vis.* **8**(2), 2 (2008).
6. J. M. Henderson, G. L. Malcolm, and C. Schandl, "Searching in the dark: cognitive relevance drives attention in real-world scenes," *Psychon. Bull. Rev.* **16**(5), 850–856 (2009).
7. D. H. Ballard and M. M. Hayhoe, "Modelling the role of task in the control of gaze," *Vis. Cogn.* **17**(6–7), 1185–1204 (2009).
8. T. J. W. M. Janssen and F. J. J. Blommaert, "A computational approach to image quality," *Displays* **21**(4), 129–142 (2000).
9. J. Radun et al., "Evaluating the multivariate visual quality performance of image-processing components," *ACM Trans. Appl. Percept.* **7**(3), 1–16 (2010).
10. J. Radun et al., "Content and quality: interpretation-based estimation of image quality," *ACM Trans. Appl. Percept.* **4**(4), 21.1–15 (2008).
11. T. Betz et al., "Investigating task-dependent top-down effects on overt visual attention," *J. Vis.* **10**(3), 15.1–14 (2010).
12. C. M. Masciocchi et al., "Everyone knows what is interesting: salient locations which should be fixated," *J. Vis.* **9**(11), 25.1–22 (2009).
13. M. S. Castelhano, M. L. Mack, and J. M. Henderson, "Viewing task influences eye movement control during active scene perception," *J. Vis.* **9**(3), 6.1–15 (2009).
14. G. Harding and M. Bloj, "Real and predicted influence of image manipulations on eye movements during scene recognition," *J. Vis.* **10**(2), 8.1–17 (2010).
15. M. S. Castelhano and C. Heaven, "The relative contribution of scene context and target features to visual search in scenes," *Atten. Percept. Psychophys.* **72**(5), 1283–1297 (2010).
16. A. Hollingworth, G. Schrock, and J. M. Henderson, "Change detection in the flicker paradigm: the role of fixation position within the scene," *Mem. Cognit.* **29**(2), 296–304 (2001).
17. M. P. S. To et al., "Perception of suprathreshold naturalistic changes in colored natural images," *J. Vis.* **10**(4), 12.1–22 (2010).
18. J. R. Antes, "Time course of picture viewing," *J. Exp. Psychol.* **103**(1), 62–70 (1974).
19. R. Reber, N. Schwarz, and P. Winkielman, "Processing fluency and aesthetic pleasure: is beauty in the perceiver's processing experience?," *Pers. Soc. Psychol. Rev.* **8**(4), 364–382 (2004).
20. B. W. Tatler, R. J. Baddeley, and B. T. Vincent, "The long and the short of it: spatial statistics at fixation vary with saccade amplitude and task," *Vis. Res.* **46**(12), 1857–1862 (2006).
21. B. W. Tatler and B. T. Vincent, "Systematic tendencies in scene viewing," *J. Eye Mov. Res.* **2**(2), 1 (2008).
22. P. J. A. Unema et al., "Time course of information processing during scene perception: the relationship between saccade amplitude and fixation duration," *Vis. Cogn.* **12**(3), 473–494 (2005).
23. M. Mills et al., "Examining the influence of task set on eye movements and fixations," *J. Vis.* **11**(8), 17 (2011).
24. H. Alers, L. Bos, and I. Heynderickx, "How the task of evaluating image quality influences viewing behavior," in *2011 Third Int. Workshop on Quality of Multimedia Experience*, pp. 167–172, IEEE, Mechelen, Belgium (2011).
25. A. T. A. Ninassi et al., "Task impact on the visual attention in subjective image quality assessment," in *Proc. 14th European Signal Processing Conf.*, Florence, Italy, pp. 1–5 (2006).
26. H. Liu and I. Heynderickx, "Visual attention in objective image quality assessment: based on eye-tracking data," *IEEE Trans. Circuits Syst. Video Technol.* **21**(7), 971–982 (2011).
27. J. Radun, T. Virtanen, and G. Nyman, "Explaining multivariate image quality—interpretation-based quality approach," in *Int. Congress of Imaging Science*, pp. 119–121, IS&T, Rochester (2006).
28. E. U. Weber and E. J. Johnson, "Mindful judgment and decision making," *Annu. Rev. Psychol.* **60**, 53–85 (2009).
29. S. S. Stevens, "On the psychophysical law," *Psychol. Rev.* **64**(3), 153–181 (1957).
30. R. A. Rensink, J. K. O'Regan, and J. J. Clark, "To see or not to see: the need for attention to perceive changes in scenes," *Psychol. Sci.* **8**(5), 368–373 (1997).
31. L. Itti and C. Koch, "A saliency-based search mechanism for overt and covert shifts of visual attention," *Vis. Res.* **40**(10–12), 1489–1506 (2000).
32. D. Walther and C. Koch, "Modeling attention to salient proto-objects," *Neural Netw.* **19**(9), 1395–1407 (2006).
33. P. Slovic, "The construction of preference," *Am. Psychol.* **50**(5), 364–371 (1995).
34. H. Salmi et al., "Development of a balanced test image for visual print quality evaluation," *Proc. SPIE* **7242**, 72420B (2009).
35. P. Oittinen et al., "Framework for modeling visual printed image quality from the paper perspective," *Proc. SPIE* **6808**, 68080L (2008).
36. International Organization for Standardization, ISO 12640-1, International Organization for Standardization, Genève, Switzerland (1997).
37. Kakadu 6.0, www.kakadusoftware.com.
38. K. Teunissen, "The validity of CCIR quality indicators along a graphical scale," *SMPTE Motion Imaging J.* **105**(3), 144–149 (1996).
39. M. DeAngelus and J. B. Pelz, "Top-down control of eye movements: Yarbus revisited," *Vis. Cogn.* **17**(6–7), 790–811 (2009).
40. C. P. Kaller et al., "Eye movements and visuospatial problem solving: identifying separable phases of complex cognition," *Psychophysiology* **46**(4), 818–830 (2009).
41. D. Walther and C. Koch, Saliency Toolbox 2.2, http://www.saliencytoolbox.net/ (July 2011).
42. J. A. Hanley, "Statistical analysis of correlated data using generalized estimating equations: an orientation," *Am. J. Epidemiol.* **157**(4), 364–375 (2003).
43. J. Gill, *Generalized Linear Models: A Unified Approach, Issue 134*, p. 101, SAGE Publications, Thousand Oaks (2001).
44. M. S. Castelhano and J. M. Henderson, "Initial scene representations facilitate eye movement guidance in visual search," *J. Exp. Psychol. Hum. Percept. Perform.* **33**(4), 753–763 (2007).
45. K. Holmqvist et al., *Eye Tracking: A Comprehensive Guide to Methods and Measures*, p. 560, Oxford University Press, Oxford (2011).

46. E. Birmingham, W. F. Bischof, and A. Kingstone, "Saliency does not account for fixations to eyes within social scenes," *Vis. Res.* **49**(24), 2992–3000 (2009).
47. K. Rayner, "Eye movements and attention in reading, scene perception, and visual search," *Q. J. Exp. Psychol.* **62**(8), 1457–1506 (2009).
48. H. Liu et al., "How does image content affect the added value of visual attention in objective image quality assessment?," *IEEE Signal Process. Lett.* **20**(4), 355–358 (2013).
49. S. Pannasch, J. Schulz, and B. M. Velichkovsky, "On the control of visual fixation durations in free viewing of complex images," *Atten. Percept. Psychophys.* **73**(4), 1120–1132 (2011).
50. W. Einhauser, M. Spain, and P. Perona, "Objects predict fixations better than early saliency," *J. Vis.* **8**(14), 18.1–26 (2008).
51. R. A. Rensink, "Change detection," *Annu. Rev. Psychol.* **53**, 245–277 (2002).

Jenni Radun has been working with subjective quality estimation since 2002. She is currently working at the University of Helsinki in the Visual Cognition Research Group on her PhD on subjective image quality evaluation. Previously she has worked on collaboration projects with different companies, concentrating on the visual quality of different materials, for example, cameras and high-quality print products. Her main interest is to understand the perception and evaluation of preference.

Tuomas Leisti is currently preparing his doctoral thesis on visual judgment and decision making at the University of Helsinki. He received his MA (psychology) degree from the University of Helsinki in 2005. Since graduating, he has studied different aspects of the visual quality perception process, involving both digital and printed images.

Toni Virtanen has been mainly involved in image quality research. His background comes from psychology, where he received a master's degree in 2010. His main occupation has been subjective image quality estimation in a collaboration project with Nokia at the University of Helsinki Visual Cognition Research Group. He has been with the project since 2005 and is currently working as a project manager in addition to his efforts toward a doctoral thesis on related topics.

Göte Nyman received his PhD in 1983 from the University of Helsinki, where he has worked as a professor of psychology. Currently, he works for projects at Aalto, Helsinki, and Stanford universities and in human-centered R&D in technology. His main interests include vision and image quality, future HCI technology, and collaboration networks. He is a long-time member of the Finnish Pattern Recognition Society (Hatutus).

Jukka Häkkinen received his PhD in experimental psychology from the Institute of Behavioural Sciences, University of Helsinki, Finland. He has worked as a principal scientist at Nokia Research Center and as an adjunct professor in the Department of Media Technology, Aalto University School of Science. Currently, he is the principal investigator at the Institute of Behavioral Sciences, University of Helsinki, where he leads the Visual Cognition Research Group. His research is currently focused on perceptual processes of image perception.