DOI: 10.2352/ISSN.2470-1173.2016.16HVEI-109

Does visual quality depend on semantics? A study on the relationship between impairment annoyance and image semantics at early attentive stages

Ernestasia Siahaan, Alan Hanjalic, Judith A. Redi; Delft University of Technology; Delft, The Netherlands

Abstract

We hypothesize that the semantics of image content affects how humans judge the perceptual quality of images. The recognition of image content has been shown to be processed within the first 500 ms of observation (and mostly in a pre-attentive stage). We look at whether or not participants are also able to detect impairments and judge their annoyance at early attentive stages. As the presence of impairments may slow down the early semantic recognition process, we investigate whether or not different semantic content impacts people's judgment of image quality. Our results show that participants do recognize image content despite the presence of impairments even at very early stages of vision (within the first fixation). In addition, we show that semantic categories have an influence on people's detection of image impairments at early attentive stages. People seem to be able to correctly detect very obvious impairments within one fixation, but more subtle impairments are not perceived. Finally, we show that people are more tolerant toward impairments on images portraying outdoor scenes than images portraying indoor scenes; additionally, users seem to be more critical toward images containing animate objects (humans or animals) in the region of interest compared with those with inanimate objects.

Introduction

Designing and developing multimedia systems which deliver a high Quality of Experience [1] requires mechanisms for controlling the visual quality of images and videos: e.g. to adaptively optimize visual quality during (live) video streaming, or to benchmark coding or processing algorithms off-line [2]. These mechanisms, commonly referred to as objective quality metrics, aim at automatically estimating the perceived quality of an image or video when affected by impairments due to e.g. image compression and/or transmission errors, mimicking perceptual mechanisms. For this reason, the design of objective metrics often resorts to the more or less explicit modeling of the human visual system (HVS) and its response to impairments in the image (or video) under consideration.

Within the past 25 years, research on objective quality metrics has yielded remarkable results [3, 4]. Nevertheless, the room for improvement is still large [2]. One main limitation of objective metrics is their tendency to focus on predicting the visibility of impairments, which is assumed to map directly to the annoyance and the overall quality impression of the image. Recently, though, research has pointed out that this approach may be too simplistic. In fact, there are other elements besides impairment visibility that contribute to the overall annoyance experienced by the viewer [1, 5], such as the perceived naturalness and usefulness of the image [6], the image's aesthetic appeal [7], and the social context in which the image or videos are viewed [8]. In this paper, we argue that another of such elements is the semantic content of the image.

Semantics refers to the meaning of words, phrases, or systems[1]. In studies related to vision, semantics refers to the content of the image, and thus relates with meaningful entities that people recognize to appear in the image. One widely accepted theory is that there is a hierarchy of structures that one could observe in an image's content, which enlists edges, surfaces, objects, and scenes [9]. Image semantics are usually categorized based on the higher-level structures, *i.e.* objects and scenes, observed in the image, and have mainly been studied in the context of understanding what information people recognize when observing images [10]. Although the recognition of semantic categories related to depicted objects (e.g. chair, table, person, face) and scenes (e.g. landscape, cityscape, indoor, outdoor) is a process deeply embedded in the functioning of the HVS, the latter has hardly been studied in relation to visual quality perception.

Furthermore, studies have shown that impairments located in visually important areas of images are perceived as more annoying [11, 12]. This phenomenon has been explained from a visibility point of view, i.e. artifacts located in those areas are more likely to attract visual attention, thus are more visible and therefore more annoying. Pre-attentive processes such as semantic categorization [10, 13] have not been considered in this explanation. As (top-down) visual attention is known to be (also) driven by the semantic content of the scene [14], the intrinsic interest for specific semantic categories (e.g. faces) might influence visual annoyance as well.

This paper investigates the relationship between image semantics, presence and strength of visual impairments and people's judgment of visual quality. We perform an extensive study involving images representing different semantic content and impaired with artifacts at different strengths, and we investigate the ability of participants to recognize the semantic content and the presence of impairments at very early attentive stages. In addition, we study how the judgment of perceived image quality evolves at early attentive stages and depending on the image semantics.

The remainder of this paper is structured as follows. The subsequent section will describe our research questions. The setup of the experiment performed to answer these questions will be described in the section after. Afterwards, we will present the results of the experiment we conducted, as well as the discussions on what the results imply. We will then draw conclusions in the last section.

[1] Oxford dictionary, http://www.oxforddictionaries.com/

DOI: 10.2352/ISSN.2470-1173.2016.16HVEI-109

Figure 1. From Torralba et al. [18]. Resolving object categories in impaired images is challenging. Thus, our brain relies on context (other elements in the scene) to estimate object categories. The blob in the red circles is the same in every picture, yet every time is associated with a different semantic category.

Semantic Content and Visual Annoyance

To better understand whether semantic content may influence the way we judge the visual quality of images, we first look into the reason why impairments may provoke visual displeasure or annoyance.

In general, annoyance has been hypothesized to be related to the potential of disturbance to endanger Darwinian fitness [15, 16]. For example, literature on environmental noise has pointed at annoyance being a reaction to the disturbance that noise causes to activities such as communication [17]. To the best of our knowledge, no similar study has been conducted towards explaining visual annoyance. Our hypothesis is that the presence of visible impairments causes visual annoyance by creating hindrance in recognizing content in an image, and may further impact activities such as task performance or decision making (as exemplified in figure 1). Visual annoyance may be a reaction to this hindrance, and may depend on the entity of the hindrance as well as on the semantic category of the content to be recognized. Indeed, some categories may be more urgent to be recognized, e.g. because of evolutionary reasons. It is known, for example, that human faces and outdoor scenes are consistently recognized correctly even within the first fixation. This is not true, for example, for inanimate objects and indoor scenes [10]. Impairments visible in images representing faces or outdoor scenes may be tolerated differently than those appearing in images belonging to other categories, because impairments hinder the fast recognition process.

It is important to note here that the recognition of semantic categories in vision happens very fast, mostly within the first fixation (~500 ms [19]) and with a consistent part of the recognition already achieved within the first 100 ms of observation. Therefore, if a relationship between semantics and visual annoyance exists, it is interesting to investigate it at these early stages of vision. If the recognition of certain semantic categories is slowed down by the presence of impairments, it is interesting to verify whether the delay in recognition is proportional to the annoyance brought about by impairments. In addition, it is also interesting to investigate the extent to which users are aware of the presence of impairments at such early vision stages (visibility), and whether the annoyance for the artifact visibility can be already quantified.

This work contributes to better understanding the mechanisms by which people appreciate and evaluate visual quality of images in relation to their semantic content. Such knowledge would allow the development of "smarter" visual quality metrics, able to adapt to the context in which the visual media is being used. For example, people's tolerance to video impairments may be different when they are looking at video surveillance footage, watching videos of activities performed outdoors, or enjoying a series of wildlife footage.

Hence, in this paper, we investigate the relationship between (1) impairment visibility, (2) their annoyance and (3) the semantics of the image elements on which impairments appear. We specifically tackle the following research questions:

RQ 1. To what extent do impairments hinder the recognition

Table 1. Image and participant groups in the experiment setup

Impairment levels	Image Group (IG)	Participant Group (PG)											
		PG1	PG2	PG3	PG4	PG5	PG6	PG7	PG8	PG9	PG10	PG11	PG12
Original image	IG1	v							v			v	
	IG2		v					v					v
	IG3			v			v			v			
	IG4				v	v					v		
Medium impairment level	IG5				v			v					v
	IG6	v				v				v			
	IG7		v						v		v		
	IG8			v			v					v	
Strong impairment level	IG9			v			v			v			
	IG10				v				v		v		
	IG11	v				v						v	
	IG12		v					v					v

DOI: 10.2352/ISSN.2470-1173.2016.16HVEI-109

Figure 2. Examples of stimuli used in our experiment, representing (a) indoor scenes, (b) outdoor natural scenes, and (c) outdoor manmade scenes

of semantic elements in images?

RQ 2. How early in the vision stages do people detect the presence of impairments in images with different semantic content, and how early are they are able to discriminate between different levels of strength of impairments?

RQ 3. Is annoyance or tolerance toward the presence of impairments in images influenced by semantic categories?

Experiment Setup

We designed an experiment to investigate quality perception of images with different semantic content at early stages of vision.

We started from a set of images representing different object and scene categories (further described in the following subsection). To vary their quality, we compressed them at two different levels, obtaining three versions per image, with clearly distinguishable impairment strengths (as revealed by a pilot study). To investigate image quality perception and semantics recognition at early attentive stages (within which most of the semantic content of an image is resolved), we set participants to observe and evaluate images after three different presentation times, lasting 40, 107 and 500 ms. These values were chosen to investigate impairment detection, annoyance and content recognition within the first fixation (40 ms and 107 ms) and after the first fixation (500 ms) which, according to previous studies, lasts on average about 400 ms [19]. These presentation times have also been shown to lead to different levels of image semantic content understanding [10].

After each short presentation of an image, we then asked participants:

(1) whether or not they perceived impairments
(2) to quantify the visual quality of the image
(3) to describe the semantic content of the image.

Our goal was then to verify whether their answers to these questions would depend on the semantics of the image, the presentation time (i.e. the attentive stage) and the level of impairments of the images.

Stimuli

We selected 79 images, which were 1024x768 in size, from the MIT-CSAIL database of objects and scenes [20]. Examples of the images we selected can be seen in figure 2. As previous studies in vision science have shown that image semantics can be expressed both in terms of scenes (e.g. landscape, forest, bar, gym, slum) or objects (e.g. apple, car, plane, dog, person), and that the two are resolved roughly at the same time [10], we diversified our stimuli in terms of both. Specifically we selected:

(1) three scene categories, i.e. indoor, outdoor natural, and outdoor manmade, as studies have shown that humans recognize these scene categories differently at pre-attentional stages. In particular, at very early attention stages, indoor scenes tend to be recognized as outdoor images instead [10].
(2) two object categories, i.e. animate (breathing) objects, and inanimate objects, as previous studies have shown that humans can recognize animate objects in detail more consistently even within one fixation, as opposed to inanimate objects.

Although we limit ourselves to these general categories for this study, more detailed sub-categories (for example, offices, shops, bars as sub-categories of indoor scenes, and human, animal as sub-categories of animate objects) should also be considered in future work.

We first selected an equal number of indoor and outdoor images, and then divided the outdoor images into the same number of natural and manmade ones. As most of the images did not have complete information on their main object categories, we had four people (including the main author of this paper) annotating them. The four people looked at the 79 images with no impairments and no restriction in viewing time, and were then asked to categorize the images in terms of scene (either indoor, outdoor natural or outdoor manmade) and object category (either animate or inanimate).

Images with 100% agreement on their scene and object categories among the annotators were then used in the analysis of our data, i.e. 73 images for recognition of scene categories, and 73

Table 2. Distribution of Presentation Times (PT) for all images and image groups

Impairment Level	Image ID	PT (in ms)	Image Group
None	1	40	IG1
	2	107	
	3	500	
	4	40	
	5	107	
	…		
	20	107	
	21	500	IG2
	…		
	40	40	
	...		
	79	40	IG4
Medium	1	500	IG5
	...		
	79	500	IG8
Strong	1	107	IG9
	...		

DOI: 10.2352/ISSN.2470-1173.2016.16HVEI-109

images for recognition of object categories. Out of the 73 images considered for scene category recognition, 38 were indoor images, 15 were outdoor natural, and 20 were outdoor manmade. Out of the 73 images considered for object category recognition, 23 had animate objects as their point of interest, and 50 had inanimate objects.

All 79 images (including those with less than 100%) were anyway employed in our experiment. Prior to that, they impaired through JPEG compression at two levels (strong and medium, corresponding to quality parameters Q = 15, 30, respectively when using the Matlab JPEG compression implementation). All original images, in addition to those with medium and strong impairments, were then involved in the evaluations.

Procedure

Our dataset included 237 images (79 contents x 3 impairment levels, each to be evaluated at 3 different presentation times). As we were willing to conduct a full factorial design, with presentation time and impairment level to be investigated strictly between subjects to avoid memory effects (i.e., we wanted subjects to never see the same content twice), we resorted to an incomplete balanced block design. Each participant was asked to rate 60 images in one experimental session, with a break after the first 30 evaluations. To have at least 10 participants rating each of the 237 images at each presentation time, we recruited 120 participants in total for this experiment. We divided our whole image set into groups of 20, and assigned participants into different participant groups that would evaluate 60 different images, in random order, in two sessions of rating 30 images each, having a short break in between to prevent fatigue. Table 1 illustrates this setup. The table shows that four image groups (IG) were created for each impairment level. These IGs consist of 19-20 different contents, either impaired or not. Each participant group (PG) evaluated three IGs, including different contents (i.e., making sure that no content would be seen twice by the same person). The presentation time distribution is shown in Table 2.

When entering the experimental room, participants were firstly briefed about their task. They performed a short training to familiarize with the experimental interface and task and with the range of impairments that they would evaluate during the experiment. In addition, they were given explanation and examples of indoor, outdoor natural, and outdoor manmade scenes, as well as animate and inanimate objects.

Once the training was complete, the actual evaluations started. To visualize the first (next) image to be evaluated, participants had to click on a button on the screen. This was necessary to make sure that participants would pay attention to the screen, given the short image presentation times involved. An image would then appear for a certain presentation time. At the end of the presentation time, the image was masked [21] to ascertain the restricted presentation time of the image. The image evaluation screens then appeared, asking the participant, in the following order:

1. Whether or not they detected impairments in the image (Yes/No question)
2. How they would rate the visual quality of the image (on a 5 point ACR scale [22])
3. How they would rate the aesthetic appeal of the image (on a 5-point ACR scale)
4. Which scene category would best describe the content of the image (multiple choice question) : (a) indoor, (b) outdoor natural, and (c) outdoor manmade

Table 3. Confusion matrix for scene category recognition across all participants (RQ1)

		Annotators		
		Indoor	Outdoor Natural	Outdoor Manmade
Participants	Indoor	3345	1	35
	O. Natural	12	1421	62
	O. Manmade	90	208	1717

Table 4. Confusion matrix for object category recognition across all participants (RQ1)

		Annotators	
		Animate	Inanimate
Participants	Animate	1982	108
	Inanimate	195	4607

5. Which object category would best describe the main subject(s) of the image (multiple choice question) : (a) animate, and (b) inanimate
6. What particular object(s) did they recall from the image (open question).

All experiments were performed in a controlled lab environment compliant to the ITU-R BT.500 recommendation [23]. Images were visualized on a 23" Samsung LED monitor with resolution 1920x1080. The images were displayed at native resolution on a neutral (gray) background. Participants viewed the images from 100 cm distance, with constant illumination at approximately 70 lux. In general, one experiment session took 1-1.5 hours.

Results and Discussions

124 subjects divided into 12 user groups (UG) (refer to Table 1) participated in the experiment. 2 out of the 124 participants could not finish their whole experiment session, and so only rated a fraction of the 60 images that they were supposed to do in one session. We did not include the 2 participants in our data analysis. On average, each of the 79 images was rated 90 times, with 9 different conditions of impairment level and presentation times combined. Every image with one impairment level and one presentation time was rated by at least 10 people.

RQ1: Recognizing content with impairments in images and in pre-attentive stages

In this subsection, we report the results related with our first research question, i.e. whether or not the recognition of semantic content is hindered by the presence of impairments in images, and whether this hindrance would affect the amount of time that an observer needs to resolve semantic categories. To check this, we first looked into cases where participants incorrectly recognized the scene or object categories of the images presented to them. The results are shown in Tables 3 and 4 for scene and object recognition, respectively. The tables report the number of times each image was assigned by the experiment participants to each semantic category, against the categorization given (unanimously) by the annotators. We are going to assume that annotators, who could see the image for unlimited time, gave a correct categorization for it, as also proven by the high agreement in

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.16HVEI-109

categorization at that stage; hence, we refer here as "incorrect" to the cases in which the category assigned by the participant differed from the one indicated by the annotators. The categorizations of participants mismatched those of annotators in less than 10% of the cases, except in the case of outdoor natural images, which were often recognized as outdoor manmade images.

We first verified whether incorrect responses were given mostly for the same images, perhaps with ambiguous content (i.e., belonging to more than one category). For both scene and object categories, less than 10 out of the 79 images were categorized incorrectly more than 10% of the times (8 for scene categories, and 6 for object categories), and none of them was categorized incorrectly more than 50% of the times. This observation suggests that the mismatches are unlikely to be caused by the ambiguity of specific images.

We then verified whether the mismatches in categorization were occurring more often for images with certain impairment levels or presentation times. Figures 3 and 4 show the frequency of incorrect recognition across levels of impairment and presentation times. Both figures show that the number of incorrect responses for scene or object category fluctuates with the change of presentation time or impairment level. We used a generalized linear mixed model (GLMM) to model this relationship. The correctness of image categorization was set as the dependent variable, whereas impairment level (level 1 = no impairment, level 2 = medium impairment, level 3 = strong impairment) and presentation time (40 ms, 107 ms, and 500 ms) were set as fixed factors. Participants were treated as random effects. The GLMM used a binomial distribution with a logit link function. Because our data were unbalanced (the amount of incorrect responses was much lower than that of correct responses) we performed over-sampling of our data points to fit the model. We randomly selected and created duplicates of data points which represented incorrect recognition of scenes or object categories, until we had a balanced number of data points representing correct and incorrect recognition of the semantic categories.

When considering participant's responses for recognizing scene categories as independent variable, our model indicated that there is a significant effect of *presentation time* ($p=0.000$), and its interaction with impairment level (*PT*impairment level,* $p=0.000$) on participants' ability to correctly recognize the scene category of an image. When controlling for all other variables, viewing the image for 40 ms seems to have a positive influence on participants' recognition of scene categories in an image (*PT=40ms,* coefficient β=0.334 in the model), whereas the 107 ms presentation time influences participants' recognition negatively (*PT=107ms,* β=-0.408). The interaction between impairment level and presentation time also has a significant effect on the recognition of scenes in images. When there is no impairment, and the presentation time is 40 ms, the recognition of scene categories in images is influenced positively (*Impairment level=1*PT=40ms,* β=0.665). The influence becomes more positive when the presentation time is at 107 ms, while the impairment level stays at no impairment (*Impairment level=1*PT=107ms,* β=0.752). When the impairment level is raised to medium level impairment, and the presentation time is at 107 ms, the influence is still positive despite it being lower than the previous two interactions (*Impairment level=2*PT=107ms,* β=0.459).

For object category recognition, the model shows that there is a significant effect of *impairment level* (p=0.000), *presentation time* (p=0.000), and their interaction (*Impairment level*PT,*

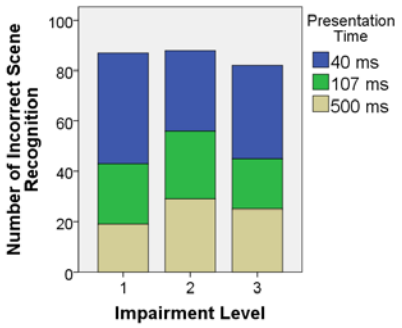

Figure 3. Frequency of incorrect recognition of scene categories across presentation times and impairment levels (level 1=no impairment, level 2=medium, and level 3=strong impairment) (RQ1)

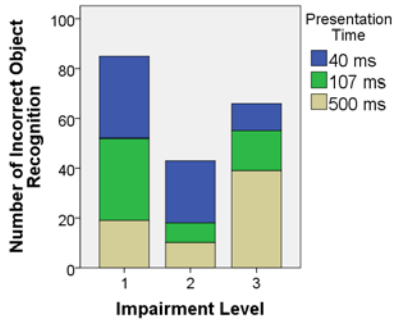

Figure 4. Frequency of incorrect recognition of object categories across presentation times and impairment levels (level 1=no impairment, level 2=medium, and level 3=strong impairment) (RQ1)

p=0.000). When controlling for other variables, impairment levels give a negative influence on the recognition of object categories, where the influence becomes more negative as the impairment level becomes stronger (*Impairment level=1,* β=-0.956, and *Impairment level=2,* β=-1.681). Presentation time also has a negative influence on the correctness of the recognition of object categories. When controlling for other variables, the lower the presentation time, the more negative is its influence on participants' ability to recognize object categories correctly (*PT=40ms,* β=-1.523; *PT=107 ms*, β=-1.111). The interaction between impairment level and presentation has a positive influence on the recognition of object categories in images (*Impairment level=1*PT=40ms*, β=2.281; *Impairment level=1*PT=107ms*, β=1.998; *Impairment level=2*PT=40ms*, β=2.926; *Impairment level=2*PT=107ms*, β=0.849).

Based on these results, we observe that:

1) The high percentage of correct responses on recognizing scene and object categories (tables 3 and 4) is in line with previous studies that show humans' ability to grasp the content of an image even within one fixation [10]. However, we should mention that we only analyzed, so far, the success in recognizing high-level categories (indoor, outdoor natural, outdoor manmade), and have not looked into participants' recognition of more specific categories. Effects of impairments may still show when asking participants to identify specific object or scene categories (e.g. animals or trees or humans, bars or forests).

2) In cases where incorrect recognition of scene and object categories happen, the presence of impairments in images, along with the limited presentation time to view them do have an effect on participants' ability to correctly recognize scene and object

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.16HVEI-109

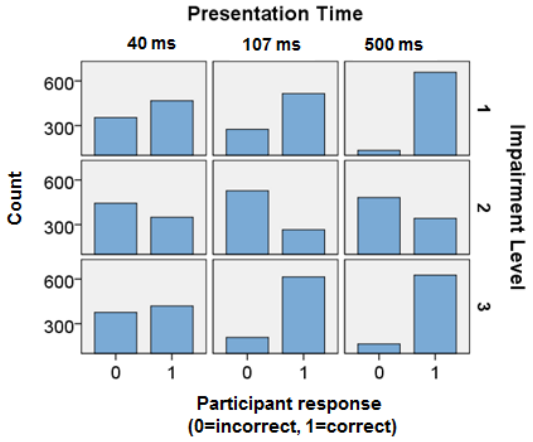

Figure 5. Comparison of incorrect and correct responses of recognizing impairments in images across presentation times and impairment levels (level 1=no impairment, level 2=medium, and level 3=strong impairment) (RQ3)

categories of images. This seems to be more pronounced for objects than for scenes.

RQ2: Detecting impairments in early attentive stages

Literature has shown that humans can recognize and categorize the content of an image already within one fixation [19, 10]. As we are interested in looking at the relationship of semantic categories with human judgment of visual quality, we also take interest in the vision stages in which impairments are noticed by viewers. More specifically, we look into how early participants detect the presence of impairments in images, whether this detection depends on the impairment strength, and whether it differs depending on the semantic content of the image.

One of the questions that we asked our participants in the experiment was whether or not they saw impairments in the image that was presented to them. We first looked into participants' responses across impairment levels. We considered a response as incorrect when participants replied 'No' to the question above while viewing an impaired image (impairment levels 2 and 3), or when they answered 'Yes' in presence on an unimpaired image (impairment level 1).

Figure 5 shows the distribution of incorrect and correct responses given by participants across impairment levels and presentation times. The percentage of correct responses grows, as expected, with presentation time: at 40 ms presentation time it is 51.28%, at 107 ms it becomes 57.98%, and 67.58% at 500 ms presentation time. From the figure, we can also observe that participants could discriminate images that have no impairments and strong impairments from early attention stages, i.e. within only 40 ms of image viewing. However, for images with medium-level impairments, participants tended to conclude that the images they saw were not impaired. The percentage of correct responses given for images with medium-level impairment is 39.71%. This is in contrast with 68.24% of correct responses for images with no impairment, and 68.92% correct responses for images with strong impairments, regardless the presentation time. It would seem, therefore, that whereas people are able to resolve semantic categories within as little as 40 ms, this is less the case for the detection of impairments. In fact, in total, the percentage of correct responses on impairment detection across presentation times and impairment level is 58.94%, indicating that within the first 500 ms of vision the perception of impairments (and thereby, their judgment) is volatile.

To verify the findings above, we modeled the impact of impairment levels, presentation times, and semantic categories on the participants' ability to recognize the presence of impairments in images using a GLMM. Our dependent variable was the (binary) correctness of the detection of impairments in an image. The independent variables for our model were impairment levels, presentation time, and semantic categories (scene or object category). Participants were treated as random effects. The model used binomial probability distribution, with logit link function.

Using scene category as one of the independent variables, the model reports significant effects of *impairment level* (p=0.000), *presentation time* (p=0.000), and the interactions *impairment level*presentation time* (p=0.000), *presentation time*scene category* (p=0.042), as well as *impairment level*presentation time*scene category* (p=0.038). As our main interest is the attention stage in which participants' recognize impairments in images, we look into the coefficient values of presentation time, and the coefficient values of its interaction with other variables in the model.

Controlling for all other variables, 107 ms presentation time (*PT=107ms,* β=0.867) gives a less positive influence on participants' ability to correctly recognize the presence of absence of impairments compared with 40 ms presentation time (*PT=40ms,* β=1.754). This is also true when interacting with scene category. When observing images with outdoor natural scene category, 107 ms presentation time (*scene category=outdoor natural*PT=107ms,* β=-1.252) gives a more negative influence than 40 ms presentation time (*scene category=outdoor natural*PT=40ms,* β=-0.889). When comparing the influence of 107 ms presentation time interacting with different scene categories, it seems that the people recognize impairments more clearly in indoor images (*scene category=indoor*PT= 107ms,* β=-0.754) than outdoor natural images. This hints at the influence of scene category on people's annoyance/tolerance toward impairments in images, which will be further discussed in the next section.

When we consider object category as one of the independent variables in our model, there are significant effects of *impairment level* (p=0.000), *presentation time* (p=0.000), and the interactions *impairment level*presentation time* (p=0.000)), *impairment level*object category* (p=0.037), *presentation time*object category* (p=0.000), as well as *impairment level*presentation time*object category* (p=0.002), on participants' ability to recognize the presence or absence of impairments in images.

Both models confirm that there is an influence of presentation time on whether or not people notice impairments in images, as well as its interaction with the other independent variables, impairment levels and semantic category. In addition, people can detect strong levels of impairments as well as pristine images quite well since early attention stages. However, when shown images with visible yet less perceptually strong impairments, people seem to be less able to detect them, even when the presentation time is longer than one fixation.

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.16HVEI-109

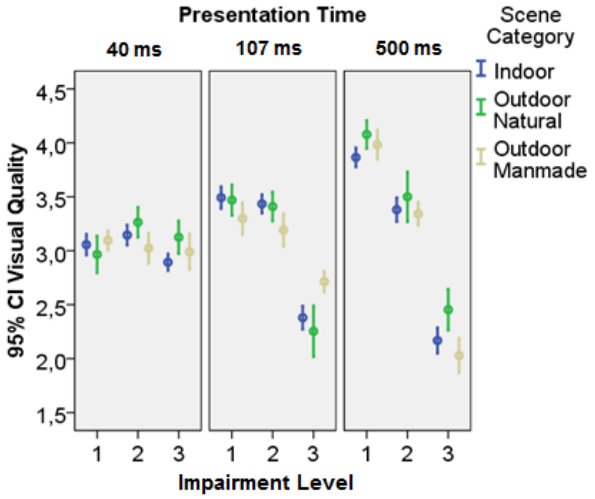

Figure 6. Changes in visual quality over different impairment levels and presentation times for the three scene categories (RQ2); (impairment level 1=no impairment, level 2=medium, and level 3=strong impairment)

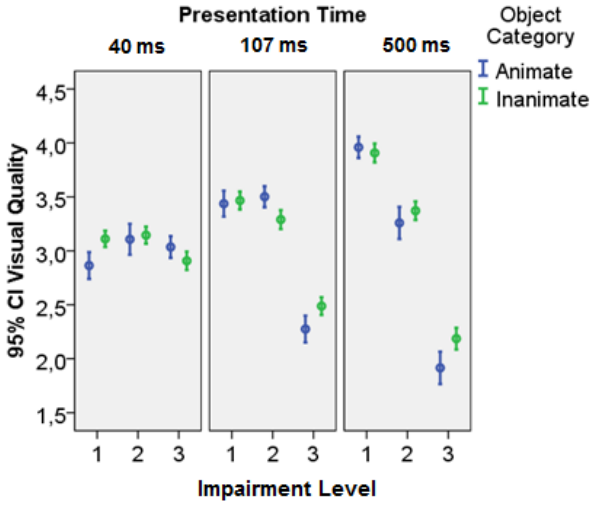

Figure 7. Changes in visual quality over different impairment levels and presentation times for the two object categories (RQ2); (impairment level 1=no impairment, level 2=medium, and level 3=strong impairment)

RQ3: Tolerance towards impairments in images with different semantic categories

Our last and core research question concerned people's tolerance or annoyance towards impairments in images representing different semantic categories. To look into this, we again use a GLMM to model the effect of impairment level, presentation time, and semantic categories and all their interactions of the first and second order on the visual quality ratings. Again, we built two different models, one to investigate the influence of scene category, and the other taking into account object categories. The models used a multinomial probability distribution, with logit link function. In both models, participants were treated as random factors.

Influence of Scene Category

Our first model used scene categories as the independent variable representing image semantics. The model suggests that all independent variables have a significant effect on participants' judgement of image visual quality: *impairment levels* (p=0.000), *presentation times* (p=0.000), *scene category* (p=0.01). So did all the first and second order interactions (*impairment level*PT*, p=0.000; *impairment level*scene category*, p=0.006; *PT*scene category*, p=0.000; *impairment level*PT*scene category*, p=0.000).

As expected, controlling for all other variables but impairment levels, visual quality rating becomes higher as the impairment level becomes lower (*Impairment level=1*, β=4.061; *Impairment level=2*, β=2.634). When controlling for other variables, the shorter the presentation time, the higher the visual quality is rated (*PT=40ms*, β=1.769; *PT=107ms*, β=1.340). When it comes to semantic categories, controlling for all other variables, visual quality tends to be more positively rated for outdoor natural images (*scene category=outdoor natural,* β=0.948).

The coefficient values for the interactions terms show that in general, people are more critical towards indoor category images compared with outdoor images, particularly outdoor natural image. At no impairment level, and 107 ms presentation time, indoor scene images have a lower influence on visual quality (*Impairment level = 1 * PT = 107ms * scene category = indoor,* β=1.286) compared to outdoor natural images (*Impairment level=1*PT=107ms*scene category=outdoor natural,* β=1.842). The same tendency can be observed at medium-level impairment, and 107 ms presentation time, indoor scene images having β=1.230, while outdoor natural images β=1.935. This implies that, given the same presentation time or impairment level, participants tend to rate indoor images lower than outdoor images.

While indoor and outdoor natural scene categories have an influence on people's level of annoyance or tolerance toward image impairments, our model shows that outdoor manmade scene category does not have such significant influence (β=0.000). We try to show this phenomenon through figure 6. The figure shows how the mean opinion scores (MOS) for visual quality change across impairment levels and presentation times for the three scene categories. From the figure, we can clearly see how people become more critical towards indoor category images than outdoor natural images as the impairment level and presentation time increase. It is not as obvious, however, to see how visual quality changes for outdoor manmade images compared with the other two categories.

Influence of Object Category

Our next model investigated the influence of the category (animate or inanimate) of the main object in the image on image quality. The model fitting results indicate that impairment levels (p=0.000), presentation time (p=0.005), and object categories (p=0.003) have a significant effect on the perceived quality of images. We also find a significant effect of the first order interaction term impairment level*presentation time (p=0.000), presentation time*object category (p=0.004), and of the second order interaction term impairment level*presentation time*object category (p=0.000).

Similar with the model obtained for scene categories in the previous subsection, the coefficient values for impairment level show that, when controlling for other variables, the lower the impairment level, the less critical people become in rating visual quality (no impairment, β=3.684; medium-level impairment,

DOI: 10.2352/ISSN.2470-1173.2016.16HVEI-109

β=2.441). It also shows, as with the previous model, that there is more positive influence on visual quality when the presentation time is shorter (40 ms presentation time, β=1.414; 107 ms presentation time, β=0.603). With regards to semantics, controlling for other variables, the model shows a negative influence on visual quality when the object is of animate category (β=-0.768).

The coefficient values for the interaction terms show that, when there is no impairment level, and the object category of the image is animate, the higher the presentation time, the less negative is the judgment on visual quality (at 40 ms presentation time, β=-1.655; 107 ms presentation time, β=-0.715). Figure 7 further shows the interactions among the dependent variables, and how the visual quality changes. From the figure, we can also observe the conditions in which images with animate objects are judged more critically than images with inanimate objects.

Based on the above results, we conclude that the judgment of visual quality may to be influenced by image semantics to a certain extent. When it comes to scene categories, people tend to be more critical toward images representing indoor scenes than for images representing outdoor scenes. Object categories also seem to have an influence on the judgment of visual quality. People tend to be more critical towards images with images having animate rather than inanimate objects in the visually important regions. The different behavior in rating these different scene categories and object categories may be linked to the bias that have been shown in previous studies. It has been shown that people have less difficulty picking up more details related with images having animate objects in their content, particularly images of people, compared with inanimate objects [10]. In fact, this has been linked to human having advantage for visual processing of faces and humans [23]. We might think then that as humans are more sensitive towards animate objects in an image, they would also have more sensitivity towards impairments on those objects. However, it has also been shown that people have less difficulty recognizing outdoor images than indoor images in early attention stages, although this has not been linked to any difference in the ability to perceive sensory information in the different scene categories [10]. When it comes to judging visual quality, people are less critical towards impairments on outdoor scene images. It would seem then, that whereas for objects, impairments are deemed more annoying when present on objects that are recognized faster, for scenes the opposite happens. This relationship needs further investigation also in relation to the urgency of recognizing categories.

Conclusions

In this study, we investigated how the perceived visual quality of images is influenced by their semantic content, and conversely, how the presence of impairments in images may hamper the recognition of the image semantic content in the early stages of vision.

Our results show that, in general, people can recognize the semantic content of images within a very short time span (~500 ms), despite the presence of impairments in the image. In the cases where people fail to recognize the content of the image correctly, this depends mostly on the strength of the impairments in the picture, and on the time for which the picture has been observed. These results pertain to high-level semantic categories (super categories), for both objects (inanimate or animate) and scenes (indoor or outdoor). It may be the case that the correctness of content recognition decreases when taking into account finer categories (e.g. forest or beach for outdoor scenes, man or animal for animated objects), or whether the effect of impairments becomes more evident.

We also show that the semantic content of the image does have an influence of people's tolerance toward visual impairments. People seem to be more critical toward indoor scenes compared with outdoor scenes in images, and toward images that have animate objects (people or animals) as point of interest, than images with inanimate objects in their visually important region. Finally, our results show that the judgment of image quality becomes more precise the longer people are exposed to the image. In fact, at very low presentation times the presence of impairments is incorrectly detected 50% of the times, and more prominently when impairments are not strong.

In future studies, it would be interesting to look into finer semantic categories, for objects as well as scenes, and how they affect people's annoyance or tolerance toward impairments in visual media. In this sense, a systematic analysis of the annoyance of equally visible artifacts on images presenting different semantic content is to be preferred. Additionally, the generalizability of the results presented in this papers to artifacts different from blockiness is to be verified. Finally, the reliability of the quality judgment at early stages of vision (which we have shown to be low), is interesting to further investigate, as it may have methodological implications for the subjective testing of image quality.

The results of this study contribute as initial insights into unveiling the role of semantic content in influencing people's judgment of visual quality. A more in-depth knowledge on the subject would allow creating semantic-aware video quality metrics, able to detect areas of the images where image quality would be perceived as lowest due to semantic content. Such metrics would in turn allow application-specific or task-specific optimization of visual quality for delivering images or videos to users, allowing the optimization of visual quality optimization according to the content of the media and context in which it is consumed.

References

[1] P. L. Callet, S. Möller, and A. Perkis, "Qualinet White Paper on Definitions of Quality of Experience," in European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003), 2012.

[2] A.K. Moorthy and A.C. Bovik, "Visual Quality Assessment Algorithms: What Does the Future Hold?" International Journal of Multimedia Tools and Applications, Special Issue on Survey Papers in Multimedia by World Experts, Vol: 51 No: 2, pp. 675-696, 2011

[3] N. Ponomarenko, F. Battisti, K. Egiazarian, J. Astola and V. Lukin, "Metrics performance comparison for color image database", in Proc. Fourth international workshop on Video Processing and Quality Metrics for consumer electronics, Vol: 27, 2009

[4] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," IEEE Trans. Image Processing, Vol:15 No: 11, 2006

[5] J.A. Redi, Y. Zhu, H. de Ridder, and I. Heynderickx, "How passive image viewers became active multimedia participants", in Visual Signal Quality Assessment, pp. 31-72, Springer International Publishing, 2015.

[6] H. de Ridder and S. Endrikhovski. "Invited paper: image quality is fun: reflections on fidelity, usefulness and naturalness", in SID

DOI: 10.2352/ISSN.2470-1173.2016.16HVEI-109

Symposium Digest of Technical Papers, Vol: 33, pp. 986–989, Wiley Online Library, 2002.

[7] J.A. Redi. “Visual quality beyond artifact visibility,” in Proc. SPIE 8651, Human Vision and Electronic Imaging XVIII, 2013.

[8] Y. Zhu, I. Heynderickx, and J. A. Redi. “ Alone or together: measuring users' viewing experience in different social contexts", in Proc. SPIE 9014, Human Vision and Electronic Imaging XIX, 2014.

[9] D. Marr, “Vision”. W.H. Freeman, San Fransisco, CA., 1982.

[10] L. Fei-Fei, A. Iyer, C. Koch, and P. Perona,” What do we perceive in a glance of a real-world scene?”, Journal of Vision, vol. 7, no. 1, pp. 1–29, 2007.

[11] H. Alers, H. Liu, J.A. Redi and I. Heynderickx, “Studying the risks of optimizing the image quality in saliency regions at the expense of background content”, in Proc. SPIE 7529, Image Quality and System Performance VII, 2010.

[12] U. Engelke, R. Pepion, P. Le Callet, and H.-J. Zepernick, “Linking impairment perception and visual saliency in H.264/AVC coded video containing packet loss,” in Proc. SPIE/IEEE Int. Conf. Visual Communications and Image Processing, July 2010.

[13] E. Rosch, “Principles of categorization”. In E. Rosch, B. Lloyd, Cognition and categorization (pp. 27–48). Hillsdale, NJ: Lawrence Erlbaum, 1978

[14] A. D. Hwang, H.-C. Wang and M. Pomplun, “Semantic guidance of eye movements in real-world scenes,” Vision Research, vol. 51, no. 10, pp. 1192-1205, 2011.

[15] F. L. Hall, S. M. Taylor and S. E. Birnie, “Activity interference and noise annoyance”, Journal of Sound and Vibration, vol. 103, pp. 237-252, 1985.

[16] K. T. Kalveram, “Zur Evolution des Belästigungserlebnisses. Ökopsychologische und verhaltensbiologische Betrachtungen über die Wirkung von Lärm.”, Psychologische Beiträge vol. 38, pp. 215-230, 1996.

[17] R. Guski, U. Felscher-Suhr and R. Schuemer, “The concept of noise annoyance: how international experts see it,” Journal of Sound and Vibration, vol. 223, pp. 513–527, 1999.

[18] A. Torralba, K.P. Murphy and W.T. Freeman, “Using the forest to see the trees: exploiting context for visual object detection and localization”, Communications of ACM, vol. 53, no. 3, pp. 107–114, 2010.

[19] I. Biederman, R.C. Teitelbaum, and R.J. Mezzanotte, “Scene perception: A failure to find a benefit from prior expectancy or familiarity”, Journal of Experimental Psychology: Learning, Memory, and Cognition, vol. 9, pp. 411–429, 1983.

[20] J. Xiao, J. Hays, K. Ehinger, A. Oliva, and A. Torralba, “SUN Database: Large-scale Scene Recognition from Abbey to Zoo,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2010.

[21] Li, F. F., VanRullen, R., Koch, C., & Perona, P, “Rapid natural scene categorization in the near absence of attention.” Proceedings of the National Academy of Sciences of the United States of America, 99, 9596–9601, 2002.

[22] I. T. U. (ITU), "ITU-R BT.500-13. Methodology for the subjective assessment of the quality of television pictures," International Telecommunication Union (ITU), 2012.

[23] T. Ro, C. Russel, & N. Lavie, “Changing faces: A detection advantage in the flicker paradigm,” Psychological Science, 12, pp. 94–99, 2001.