Psychophysical Study of Human Visual Perception of Flicker Artifacts in Automotive Digital Mirror Replacement Systems

Nicolai Behmann[1,▲], Sousa Weddige[1], Holger Blume[1]

[1]*Institute of Microelectronic Systems, Leibniz University Hannover, Hannover, Germany*
E-mail: behmann@ims.uni-hannover.de

Abstract. *Aliasing effects due to time-discrete capturing of amplitude-modulated light with a digital image sensor are perceived as flicker by humans. Especially when observing these artifacts in digital mirror replacement systems, they are annoying and can pose a risk. Therefore, ISO 16505 requires flicker-free reproduction for 90 % of people in these systems. Various psychophysical studies investigate the influence of large-area flickering of displays, environmental light, or flickering in television applications on perception and concentration. However, no detailed knowledge of subjective annoyance/irritation due to flicker from camera-monitor systems as a mirror replacement in vehicles exist so far, but the number of these systems is constantly increasing. This psychophysical study used a novel data set from real-world driving scenes and synthetic simulation with synthetic flicker. More than 25 test persons were asked to quantify the subjective annoyance level of different flicker frequencies, amplitudes, mean values, sizes, and positions. The results show that for digital mirror replacement systems, human subjective annoyance due to flicker is greatest in the 15 Hz range with increasing amplitude and magnitude. Additionally, the sensitivity to flicker artifacts increases with the duration of observation.* © *2021 Society for Imaging Science and Technology.*

1. INTRODUCTION

Digital side- and rear-view mirror replacement systems in vehicles increase the safety for the vehicle occupants and other road users by reducing the vehicle blind spots. Conventional mirrors are replaced by cameras on the outside, and displays reproduce the environment inside the vehicle. In addition to a larger field of view of the cameras compared to mechanical mirrors, air resistance and fuel consumption are reduced, and additional information can be visualized.

Alongside the benefits brought about by these developments, new risks arise from the time-discrete perception and reproduction of the vehicle environment. More and more incandescent lighting is replaced by LED light. It is, among other applications, used in vehicle light (like day running light and rear light), street lighting, and fuel station price tags. When the image sensor of a camera is discretely sampling amplitude-modulated light, the captured image intensity can appear non-constant in contrast to direct human perception.

During the exposure time of the image sensor, the brightness of the surrounding in every pixel is accumulated in a phototransistor, quantized, and digitally outputted. Temporal mismatches between the exposure time of the sensor and the irradiation period of the modulated light result in temporal flickering as well as spatial aliasing. This so-called LED flicker is depicted in Figure 1. The irradiation time of an exemplary pulse width modulated LED light as well as the exposure time of the image sensor and the resulting intensity captured by the image sensor are shown as a function over time.

Unintended, disturbing artifacts can occur when reproducing the picture on the screen. While machine vision algorithms are affected in terms of non-captured traffic lights or variable speed limit signs, the human vision is more sensitive to illumination modulation. Especially in peripheral view, the amount of rods leads to high sensitivity to movements and brightness differences. Risks in terms of traffic (accidents, wrong information transfer) and health (headache, epileptic seizures [18]) are possible consequences. Consequently, LED flicker has to be eliminated in digital rear- and side-view mirrors.

In order to quantify the effect of flicker on visual perception from humans, key performance indicators (KPIs) need to be standardized and correlated to psychophysics studies, which is ongoing work in the IEEE P2020 working group [10]. Those are necessary for the government to set margins for legal admission of mirror replacement systems and for original equipment manufacturers to compare different flicker mitigation systems. Moreover, these KPIs are needed to evaluate the effect of flicker mitigation algorithms. Example presented by the authors in [1]. These KPIs must also take into account the subjective human visual perception.

In this article, we present a novel detailed psychophysics study on the human visual perception of flicker in side mirror applications. Therefore, real-world and synthetic driving sequences have been captured and flicker has been added manually. In the form of a laboratory and online study, those synthetic sequences were presented to the test persons, overlaid with synthetic flicker artifacts (varying flicker frequency, amplitude, and mean value). The perception of flicker and subjective impression on the disturbance of different flicker settings were evaluated.
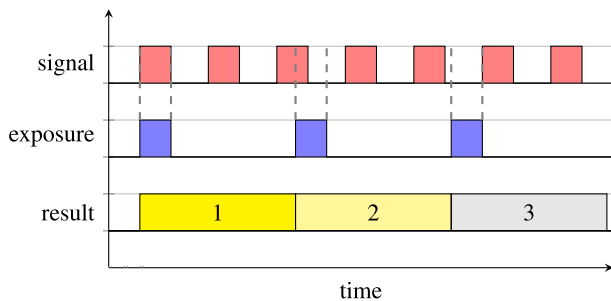
▲ IS&T Member.

J. Percept. Imaging
IS&T International Symposium on Electronic Imaging 2021

010401-1

Jan.-June 2021
Human Vision and Electronic Imaging 2021

Figure 1. Origin of LED flicker artifacts. The pulse width modulated LED signal, exposure time of the image sensor, and the resulting image are shown as a function over time. The first exposure time includes a full LED impulse and the LED is perceived as switched on by the image sensor. The second exposure time and an LED impulse are partly overlapping and the LED is perceived as dimmed by the image sensor. The third irradiation time is not captured by the image sensor and appears switched off on the display. The light is perceived by humans in direct vision as constantly switched on.

The remainder of the article is structured as follows: First, previous work from literature related to psychophysics studies on flicker are presented. Subsequently, the psychophysics study's structure, the previous coarse investigation and our capture and testing setup is described in detail. The results from the studies are presented and discussed in the following chapter, before concluding this article.

2. RELATED WORK

First studies on flicker perception were performed by Brown [3] and Kelly [11] 50 years ago regarding flicker in cinematic applications. Using a sinusoid light source, the critical flicker perception frequency was evaluated for large-area illumination. Later, these studies were extended from the temporal-frequency domain to the spatial-frequency domain.

With a psychophysical study [12], the perception of white flicker in front of a white and a black background was investigated. For several frequencies between 50 and 70 Hz and duty cycles from 20 to 90 %, it appeared that flicker in front of a white, respectively, black background is barely perceived at 60 Hz in conjunction with 70 % duty cycle and 65 Hz in conjunction with 90 % duty cycle. For a greater reference stimulus (white background), accordingly a higher flicker amplitude was necessary for the human observer's perception, which confirms the law of Weber [8].

An empirical technique for measuring the perceived flicker on refresh displays has been developed by [17]. It introduced a flicker matching technique, where the perceived flicker from a refresh display was compared to a lamp with constant luminance and adjustable temporal frequency for different display contrasts. It provided a basis for a predictive model of flicker perception on displays.

By extending the flicker prediction model by phosphor persistence, refresh frequency, luminance, and display size, Farrell [9] contributed a model to predict flicker appearance on video display terminals.

Another flicker prediction model from Denes [6] focused on temporally changing images. The participants watched 18 stimuli consisting of partly flickering picture pairs and were requested to mark those parts that appeared to flicker with a computer mouse. The results show that the flicker perception degrades with increasing refresh rates and increases with the blur.

An extension of the flicker prediction model with motion is found in [5]. Participants rated moving circles consisting of color-changing dots whether they appear to flicker or not. In several rounds, the circles consisted of a different number of dots (24, 18, 12, 10), the circles moved in different velocities, and the dots flickered in various frequencies ($\frac{1}{12}, \frac{1}{6}, \frac{1}{4}, \frac{1}{3}, \frac{1}{2}$ Hz). From this, a strongly correlating prediction model was implemented. The results show that flicker frequency, velocity, and object spacing impact motion silencing.

The spatial flicker effect in video scaling has been investigated [16] focusing on noise and blur flicker. These appear mostly in video streaming systems because of adaptive video sizes and compression. Participants rated videos containing two alternating layers on mobile devices. The results show that low frequencies can relieve the annoyance of the flicker effect, but at some point, a further decreasing frequency does not bring significant effect. The amplitude has a dominant effect and should be kept as low as possible.

The effect of a flickering stimulus appearing brighter than a steady stimulus of equal mean luminance was investigated in [19]. Participants adjusted the amplitude of matching stimulus to match it in brightness to the flickering stimulus. The results show that the brightness enhancement increased with increasing modulation frequencies, peaking at about 16 Hz at full modulation.

The IEEE P2020 working group on automotive image quality [10] implements a subgroup working on KPIs for LED flicker artifacts in both visual and machine vision use cases. First, KPIs for flicker detection and modulation amplitude in video sequences will be released soon. A first psychophysics study on area flicker was conducted, in which testers were presented three different flickering videos with varying flicker edge sharpness, frequency, and contrast. Testers were then requested to rate the video to be assessed in regards to the two other fixed reference flickering videos. An increasing flicker sensitivity was observed for sharper edges, higher temporal frequencies, and contrast.

A study on the perception of flickering red rear lights with 28 test persons [15] showed that the critical flicker frequency peaks at a viewing angle of 20 to 30 to the optical axis of the eye with 54 to 56 Hz. During the study, the participants focused on 50 different measuring points around a pulse width modulated red rear light, and their critical flicker frequency is determined through decreasing and increasing of the pulse width modulation frequency.

Human perception of flicker is a widely studied topic, as can be seen in the related work presented above. All studies have in common, the flicker sensitivity peaks around

15 Hz temporal frequency with increasing contrast, motion, amplitude, and mean.

However, so far no real-world driving situations for mirror replacement systems were examined. Today, the majority of flicker artifacts arise out of local, direct pointing LED lights (daytime head- and tail lights, fuel station price tags, marketing banners) and require a novel application-specific psychophysics study which is contributed in the following.

3. PSYCHOPHYSICS STUDY

Many parameters such as flicker frequency, amplitude, mean value, position in the field of view, and vehicle speed influence the perception of flickering light. The evaluation of all combinations of these parameters would have taken too long to be assessed in one study setup. Therefore, the study was divided into two parts using the coarse-to-fine method.

The first assessment was executed as a laboratory study with real-world video sequences and a coarse selection of parameters and has been presented in parts by the authors in [2]. Based on the evaluation results, being revisited in the first part of this psychophysics study, an in-depth analysis on higher flicker frequencies and finer resolution of flicker amplitudes follows in the second, fine-granular psychophysics study. Latter was executed in an interactive online study.

Both studies followed closely the recommendations of the International Telecommunication Union (ITU-R) BT.500-14 [4] for procedures and environmental conditions for the subjective assessment of the quality of television images. In that case, the ability of the human visual perception to retain quality under non-optimum conditions that relate to transmission (impairment assessments) were conducted. For rating, the same following rating scale for subjective perception was adopted.

1 very annoying
2 annoying
3 slightly annoying
4 perceptible, but not annoying
5 imperceptible

4. COARSE INVESTIGATION OF FLICKER PARAMETERS

The coarse part of the psychophysics study was executed with manually manipulated real-world driving sequences in a vehicle simulator in the laboratory. These real-world driving sequences were captured by a modified car with three cameras—one mounted to the windshield capturing the scene in front of the car, and one at each side window recording backwards. All cameras were synchronized and captured at a fixed frame rate of 25 Hz and a resolution of [1920×1080]. Long exposure times were used to suppress flickering lights during the capture. Flicker was intentionally added by framewise masking and multiplying the masked regions by a factor. If the factor was one, the masked region

Table I. Flicker sequences of the masked regions.

ID	Sequence	Frequency	Amplitude
0	1.0 (original)	25.0 Hz	0.00
1	1.0, 0.5	12.5 Hz	0.25
2	1.0, 0.0	12.5 Hz	0.50
3	1.0, 0.9	12.5 Hz	0.05
4	0.5, 0.1	12.5 Hz	0.20
5	1.0, 0.9, . . ., 0.0. . ., 0.9	1.25 Hz	0.50
6	1.0, 0.9, . . ., 0.5. . ., 0.9	2.5 Hz	0.25
7	random	div	0.5

looked the same as in the original picture. If the factor was zero, the masked region was black. Consequently, a smaller factor led to a darker light in the masked region. The factor for each frame was chosen to fit the frequency, amplitude, and mean value needed.

During the coarse part of the study, different scenes and environments were assessed. They varied in weather, daytime, and environment (city/highway). The size of the flickering area varied as well. Small, medium, and large flicker areas were assessed. The position in the viewing field varied, too.

Twenty participants aged between 20 and 50 with normal or corrected-to-normal vision took part in the experiment. All participants had at least some driving experience and five of them were female.

4.1 Study Setup and Execution
The participants took seat on a real driver seat in the laboratory. In front of them a dashboard and a large 55" screen (SONY KDL-55X4500, 420 cd/m^2) were mounted. On the left and right side of the dashboard, a 21" screen (EIZO S2100, 300 cd/m^2) was positioned, simulating the digital side mirrors (see Figure 2). The front screen was approximately 120 cm, the left screen 60 cm, and the right screen 130 cm apart from the participant to recreate a situation comparable to driving a real car. The room light was stabilized at approximately 750 Lux.

First, the participants coped with some introducing tests, whereat their critical flicker frequency was measured and flickering light was introduced through the rating of a modified, flickering front light. Afterward, the participants rated 40 flickering video sequences with different flicker sequences. The sequences were shown until the participants rated, but at most three times. Five different scenes and eight different flicker sequences including the original video sequence were used and are listed in Table I. The flicker frequency varied between 1.25 Hz and 12.5 Hz. The amplitude varied between 0.05 and 0.5. The duration of the sequences varied between 5 and 10 seconds each. The mean value was chosen to always reach the maximum intensity at the peak of the amplitude modulation.
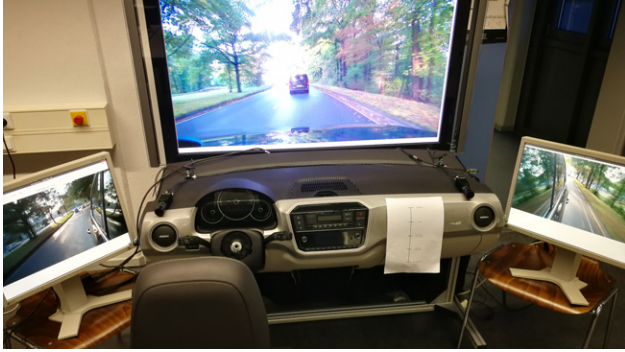
J. Percept. Imaging
IS&T International Symposium on Electronic Imaging 2021

010401-3

Jan.-June 2021
Human Vision and Electronic Imaging 2021

Figure 2. Psychophysical study setup. A central 55″ LCD Screen replayed the central camera, while the left and right 21″ monitor acted as the digital side mirror replacement system in the open car environment. The evaluation scale was printed on a paper.

Table II. Subjective rating of flicker annoyance by 20 participants depending on the flickering area size. Rating scale from 1 (very annoying) to 5 (not perceptible).

	Small (S)	Medium (M)	Large (L)
Mean rating	2.93	2.54	2.05
95 % Conf.	0.10	0.13	0.01

4.2 *Results and Evaluation*

Video sequences with overlaid flicker sequences were evaluated in a randomized order for each test candidate. It was ensured that the same scene is not evaluated several times in a row with different flicker settings.

In Table II, the mean subjective grade and standard deviation of all sequences structured for different light source sizes are depicted. With increasing size of the flickering region, the mean grade drops from 2.93 for small (S) over 2.54 for middle (M) to 2.05 for large (L) flickering regions. For the original sequences, a mean grade of 4.5 is reached. However, already a small flickering area is perceptible and slightly annoying to the test persons.

The study's results are evaluated for comparable sizes and scenes in the following. Data sets for different amplitudes are shown in Figure 3 as a function of the frequency. For the existing values of the amplitudes 0.25 and 0.5, a rating's decrease for an increasing frequency can be observed. If the values for 12.5 Hz are compared, a better rating for smaller amplitudes is observed. During the second, detailed part of the study frequencies between 2.5 and 12.5 Hz and higher should be investigated. Additionally, there is a huge difference between the rating for an amplitude of 0.05 and 0.2. The range between those amplitudes has to be investigated in more detail as well.

Two sequences, both with middle sized flicker artifacts, one captured in daytime and the other one at night, were compared to evaluate the influence of environmental lightning on the perception of flicker annoyance. The night sequence was slightly worse rated with 2.46 against 2.62. However, the 95% confidence intervals are overlapping a lot wherefore no significant difference can be determined.
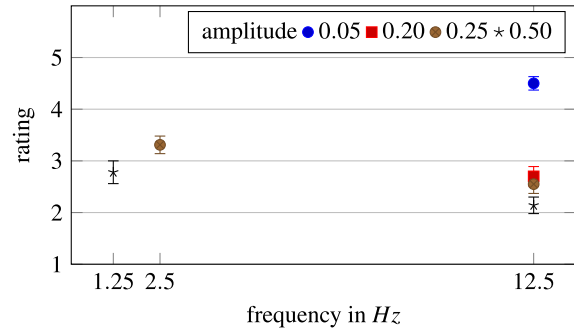

Figure 3. Evaluation of the coarse psychophysical study. 20 participants rated the annoyance of different flicker sequences. The rating of different amplitudes' data sets as a function of the frequency in Hz is shown for comparable flicker sizes and scenes. Hereby the rating's mean values and 95% confidence intervals are used.

Further environmental conditions between both video sequences (partial rain at night, sunlight during day) overlap the environmental lightning difference, which makes a clear conclusion difficult.

In conclusion, the perception and annoyance of flicker increased through larger flickering regions, higher amplitudes, and higher frequencies. The flickering frequency in this study was limited by the original videos' frame rate (25 Hz) to 12.5 Hz (Nyquist–Shannon sampling theorem). Consequently, the second part of the study should cover higher frequencies. The amplitudes less or equal to 0.2 should be assessed in detail, because there is a large difference between the assessment of an amplitude of 0.01 (imperceptible) and 0.2 (slightly annoying). Additionally, the flicker mean value should be investigated according to the Weber–Fechner law [8] in the second part of the study.

5. DETAILED INVESTIGATION OF SELECTED FLICKER PARAMETERS

For the investigation of human visual perception and annoyance of high frequency flicker, in particular, an investigation of higher frequencies was necessary. The available recording system was limited to a frame rate of 25 Hz because of high data rates and limited possibilities to eliminate flicker in the recording. Consequently, a synthetic data set with a higher frame rate was generated and used for the detailed study.

5.1 *Synthetic Data Set Acquisition*

The synthetic video sequence was generated using the open-source simulator for autonomous driving research "CARLA" [7]. The video sequence was generated at a frame rate of 60 Hz and a resolution of [1980 × 1200] from a synthetic camera mounted at the position of the left side mirror. For streaming, the flickering videos were encoded with an AVC video codec with a visually lossless quality setting (CRF equals 17). Because of the higher frame rate, flicker frequencies of 30 Hz could be reached according to the Nyquist sampling theorem. The captured video sequence had a length of 11.1 seconds and showed the left side mirrors' view of a car driving through a city. The car went around a

J. Percept. Imaging
IS&T International Symposium on Electronic Imaging 2021

010401-4

Jan.-June 2021
Human Vision and Electronic Imaging 2021

Figure 4. Synthetic capture of the left side-view mirror in chronological order of the test sequence. Captures with the following car's day running light switched on are shown at the top, and switched off at the bottom. The scene starts on a ramp (left), afterward the car turns around a corner (middle) and finally go straight through a roundabout (right). The rightmost column shows a closeup of the car's day running light, depicted from the frame shown in the third column.

corner and subsequently continued straightforward through the roundabout. During all this, a second car was following the camera car. To achieve controlled flickering light, the same sequence was captured once with the following car's daytime headlights on or off. Exemplary images are depicted in Figure 4.

Afterward, the video captures were combined framewise to one video sequence with selected frequency, amplitude, mean value, and phase shift. Therefore, the appropriate frame i of both video captures was proportionately added up. The video sequence with the daytime running light switched on was weighted by α and the other video sequence by $\beta = 1 - \alpha$. The weighting factor α at frame i was calculated as follows:

$$\alpha(i) = m + A \cdot \cos\left(\frac{2 \cdot \pi \cdot f \cdot i}{F} + n \cdot \pi\right). \qquad (1)$$

The mean value is represented by m, the amplitude by A, the flicker frequency by f, the frame rate by F, and the phase shift by n. A frame multiplied by 1 results to look like the original capture, a frame multiplied by 0 results in a black frame.

All frames connected in chronological order resulted in one video sequence, where every pixel that is different between both captures was flickering with the described parameters. The sky and its reflections in the vehicle needed to be masked using the simultaneously generated semantic mask, as those were rendered randomly by the game engine and were not identical in the driving sequence with daytime headlights on or off. The segmentation mask was differentiating between regions (e.g., vehicles, plants, street, road marking and sky) through unique IDs, which are shown in Figure 5.

5.2 *Study Setup*

Due to the Corona pandemic, the study was executed as an online study. The participants rated videos on a website on
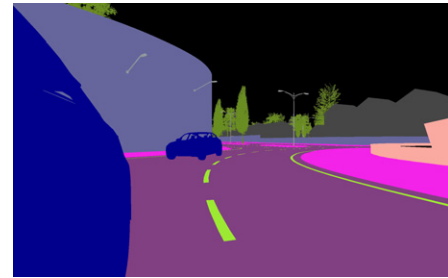

Figure 5. Automatically generated segmentation mask using "CARLA". Different regions are tagged with unique IDs and shown in different colors (sky, street, vehicles, ...).

their computer at home. In total, 32 persons participated, of which 15 additionally rated video sequences for an additional mean value investigation. Every participant rated the flicker video sequence with varying flicker settings in random order. Different frequencies, amplitudes, and mean values were investigated. An exact guideline on how to perform the test was provided to the participants at the beginning.

The single-stimulus method was used for the assessment. Only one sequence was shown at a time and was rated by the same scale used in the first assessment. The participants received information on the studies' execution on the homepage, to create most similar conditions. The participant worked at a desk and kept an arm's length distance from the computer screen. The study was executed in daylight without direct sunlight shining on the display. The screens' middle height was at eye level and the videos width on the screen was 30 cm. The rating scale was introduced on the homepage including colors, numbers, and words. "imperceptible (5)" was green and "very annoying (1)" was red. At the following pages the rating scale was shown under the video sequences. A progress bar was shown above the videos to motivate the participants to finish the assessment (see Figure 6). When the participant clicked on a rating

J. Percept. Imaging
IS&T International Symposium on Electronic Imaging 2021

010401-5

Jan.-June 2021
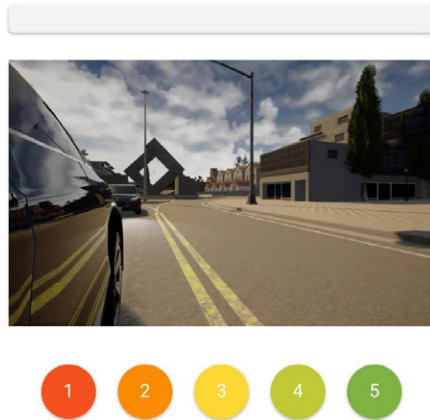Human Vision and Electronic Imaging 2021

Figure 6. Website layout for the synthetic video sequence's rating. Below the video there are five clickable colored buttons representing the rating scale. Above the video there is a progress bar which fills up during the study.

button, the next video sequence was loaded and displayed in an endless loop. The participants' rating was stored with an unique ID. Additionally, the frequency, amplitude, mean value, time and date, the videos' position in the participants viewing order, and the number of corrupted, viewed, and changed frames were saved.

The flicker artifacts resulted from the combination of seven different frequencies ($f \in 0.5, 5, 10$ 12, 15, 20, 30 Hz) and ten amplitudes ($A \in 0.02, 0.04, 0.06, 0.08, 0.10, 0.12, 0.14, 0.16, 0.18, 0.20$) to 70 different flicker sequences in total. The mean values were chosen to get a maximum value of 1, which corresponded to the maximum intensity ($m = 1 - A$).

The Weber–Fechner law states that a higher difference stimulus is needed when there is a higher reference stimulus, to observe a difference between the stimuli. This was investigated for a constant frequency of 12 Hz and a mean value of 0.25 and 0.75. For each mean value, three different amplitudes of 0.05, 0.15, and 0.25 were evaluated. This investigations' videos were merged into the other videos for 15 participants.

5.3 Results and Evaluation

The 11.1 second long video sequences consisted of 660 frames each. If the participant watched more than 3500 frames, the assessment for this video sequence was taken out of the evaluation. This long viewing indicated that the participant was absent. In total, 2020 video sequences were taken into the evaluation. On average, six frames per video sequence (1.31 % of all frames) were corrupted. In one viewing maximal 145 frames were not transmitted. No frame was transmitted with changes in it. Consequently, the user study was valid and the results could be used for further evaluation. For the evaluation, the assessments' mean values and 95 % confidence intervals were calculated [4].

In the following, the results will be evaluated by frequency and amplitude and the rating and number of watched frames in relation to the position of the videos
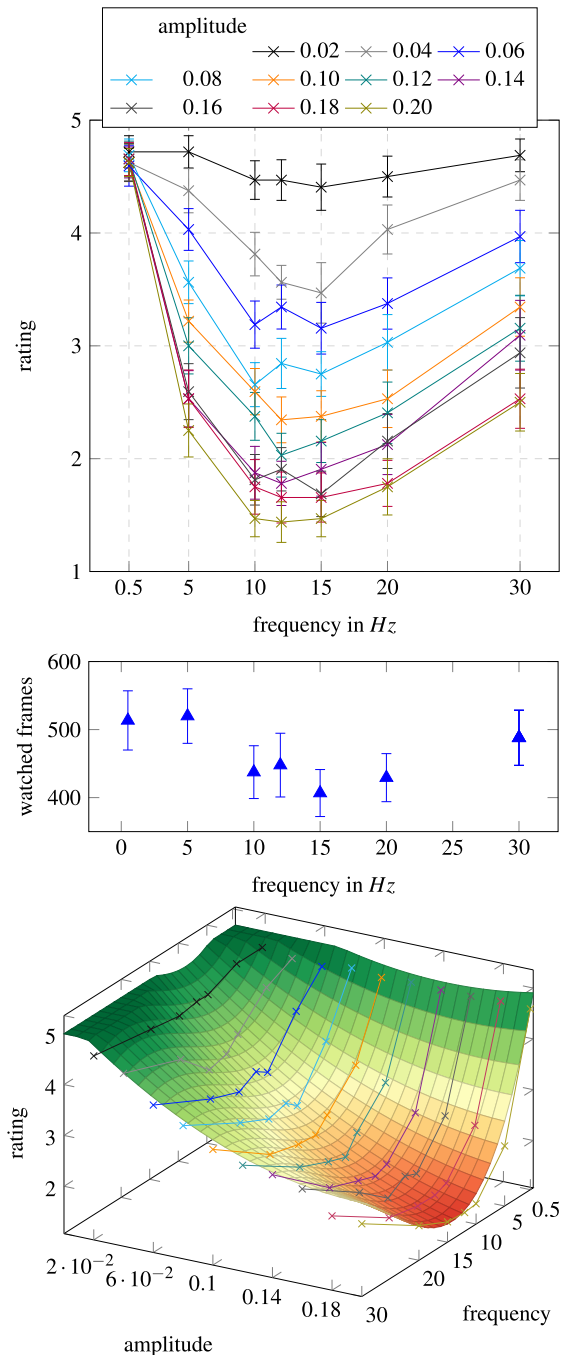

Figure 7. 32 participants' rating as a function of the frequency in Hz with varying amplitude data sets (top). Number of watched frames cumulated about all amplitudes as a function of the frequency in Hz (middle). Mean values and 95% confidence intervals are used. Results from rating regression in 3D surface plot (bottom).

in the participants' viewing order. The assessment of the mean value investigation's video sequences will be evaluated separately.

5.3.1 Rating by Frequency and Amplitude

In Figure 7, the assessments' mean values and 95 % confidence intervals of 32 participants for 7 frequencies and 10 amplitudes are shown as a function of the frequency.

J. Percept. Imaging
IS&T International Symposium on Electronic Imaging 2021

010401-6

Jan.-June 2021
Human Vision and Electronic Imaging 2021

Table III. Mean values of the number of watched frames as a function of frequency and amplitude. The video sequence has 660 frames. Values above 3500 frames are considered as outliers and excluded for the evaluation. Green marking means high numbers and red low.

amplitude / frequency	0.02	0.04	0.06	0.08	0.10	0.12	0.14	0.16	0.18	0.20	mean
0.5 Hz	539	508	466	452	462	570	496	556	512	575	513
5 Hz	551	572	524	491	496	566	425	514	606	451	520
10 Hz	546	518	548	427	514	455	333	383	297	353	437
12 Hz	604	471	616	570	360	389	388	355	337	389	448
15 Hz	446	536	516	404	421	354	445	342	320	279	407
20 Hz	510	550	503	499	399	385	339	355	377	373	429
30 Hz	512	475	524	490	479	457	525	469	487	460	488
mean	**530**	**518**	**528**	**476**	**447**	**454**	**422**	**425**	**420**	**411**	**463**

The 95 % confidence interval is bigger in comparison to the coarse user study. This accounts to the well-defined environmental conditions in the laboratory, in comparison to uncontrollable, but detailed specified conditions in this study (e.g., environmental lightning). However, a mean 95 % confidence interval below half a grade allows the evaluation of the study.

With increasing amplitude, the rating was decreasing. While the amplitude of 0.02 was rated as "imperceptible" or "perceptible, but not annoying" for every frequency, an amplitude of 0.2 was rated between "annoying" and "very annoying" in the worst case at 12 Hz. The rating at 0.5 Hz was similar for every amplitude and had a strong propensity to "imperceptible". The ratings' decrease between 0.5 Hz and 5 Hz was greater for higher amplitudes. Depending on the amplitude, some ratings at 5 Hz had a tendency to "annoying". The minimum rating for each amplitude depending on the frequency shifts from 15 Hz at 0.05 toward 12 Hz at an amplitude of 0.20. For frequencies of 20 Hz and higher, the rating was increasing and led to similar results at 30 Hz as at 5 Hz. Flickering light was perceived as annoying at a frequency of about 12 Hz for amplitudes of 0.04 and higher. At amplitudes of 0.06, flickering light was perceived as annoying from 10 Hz to 30 Hz. At amplitudes 0.08 and higher, flickering light was perceived as annoying from 5 Hz to 30 Hz.

The amplitude threshold above which flicker is perceived ranged from 0.02 to 0.06, depending on the flicker frequency. The rating decreased further at higher amplitudes, instead of saturating as predicted by the results of the coarse investigation. The smaller the flicker artifacts' amplitudes, the less they were perceived as annoying.

Summarizing, flicker frequencies between 10 and 15 Hz must be avoided in any case because they are the most annoying. 20 Hz flicker has to be eliminated at amplitudes greater than 0.04. The results for flicker frequencies of 5 and 30 Hz are equivalent. These results on the dependence of the subjective annoyance on flicker frequency and amplitude are in agreement with the results from [13, 14] for medium ambient illuminations and flickering area patterns. Accordingly, the subjective perception of the subject's disturbance is due to lateral inhibition.

5.3.2 Rating and Number of Watched Frames Compared in Relation to Amplitude and Frequency

The mean value over all observed images was 465 of 660. This is equivalent to 7.72 seconds and 70.15 % relative to the video length. In Fig. 7, the watched frames on average as a function of the frequency are shown. Significantly less frames were watched for the frequencies between 10 and 20 Hz, in comparison to the remaining frequencies. This behavior matches the subjective rating of the previously described flicker annoyance.

Table III shows the average number of watched frames. Yellow colored cells were watched approximately as long as the overall mean number of watched frames. Cells with values greater than the mean value are colored green, and lower red. The mean value of watched frames in one video sequence was maximally 616 and minimally 279 frames. The absolute maximum was 3144 watched frames, minimum twelve frames.

In comparison to Table III, in Table IV the ratings' mean values by frequency and amplitude are shown. There is a clear connection between the number of watched frames and the rating. The number of watched frames as well as the rating decreased for increasing amplitude (change in color from left (green) to right (red)). Both factors are green for 0.5 Hz. The change for 5 Hz and 30 Hz is low, too. The biggest rating decrease can be recognized for frequencies between 10 and 20 Hz.

It can be seen from these inspections that there is a relation between the rating and the time needed for a decision. The lower the rating the less time was needed to decide. The more annoying flickering light was, the faster it was noticed. If flickering light was less perceptible, the participants considered their decision longer.

5.3.3 Mean Value Investigation

The mean value investigation's results are shown in Table V. An amplitude of 0.05 was rated as "slightly annoying" for both mean values. For greater amplitudes, there was a difference between the rating of both mean values. The flicker sequences with lower mean value were rated worse. At the same stimulus difference, the participants were less sensitive for sequences with a higher mean value. Consequently, a greater amplitude was needed to achieve a

J. Percept. Imaging
IS&T International Symposium on Electronic Imaging 2021

010401-7

Jan.-June 2021
Human Vision and Electronic Imaging 2021

Table IV. Mean values of the rating as a function of frequency and amplitude. Green coloring means high rating and red low rating.

amplitude frequency	0.02	0.04	0.06	0.08	0.10	0.12	0.14	0.16	0.18	0.20	mean
0.5 Hz	4.72	4.63	4.59	4.69	4.63	4.66	4.63	4.63	4.66	4.63	4.64
5 Hz	4.72	4.38	4.03	3.56	3.22	3.00	2.53	2.59	2.53	2.25	3.28
10 Hz	4.47	3.81	3.19	2.66	2.59	2.38	1.88	1.81	1.75	1.47	2.60
12 Hz	4.47	3.56	3.34	2.84	2.34	2.03	1.78	1.91	1.66	1.44	2.54
15 Hz	4.41	3.47	3.16	2.75	2.38	2.16	1.91	1.69	1.66	1.47	2.50
20 Hz	4.50	4.03	3.38	3.03	2.53	2.41	2.13	2.16	1.78	1.75	2.77
30 Hz	4.69	4.47	3.97	3.69	3.34	3.16	3.09	2.94	2.53	2.50	3.44
mean	4.57	4.05	3.67	3.32	3.00	2.83	2.56	2.53	2.37	2.21	3.11

Table V. Results of the mean value investigation. The table shows the ratings' mean values as a function of flicker mean values and amplitudes. The frequency is 12 Hz constantly.

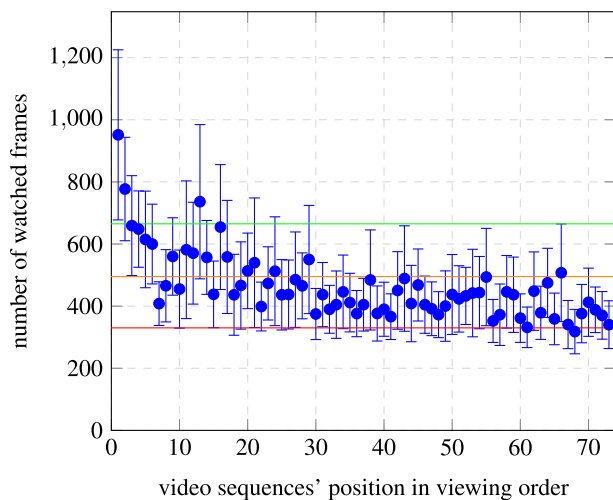| mean value amplitude	0.25	0.75
0.05	3.27	3.27
0.15	1.53	2.07
0.25	1.27	1.60

Figure 8. Number of watched frames as a function of the study's time course, respectively, the video sequence's position in the viewing order of each participant. Items under the red line were watched less than 50% of the video sequence's frames. Below the orange line 75% of the video sequence's frames were watched and above the green line more than 100%.

perception difference. According to that, the Weber–Fechner law is verified for the application of flickering digital mirrors.

5.3.4 *Behavior of the Participants During the Study*

In total, 1915 data sets were included in this evaluation. The participants rated the video sequences in changing random order. Position 1 represents the mean value for those video sequences, which the participants watched and rated first. Position 76 represents the mean value for that video sequences, which have been watched and rated lastly.

In Figure 8, the mean values and 95% confidence intervals are shown for the watched frames as a function

of the position in the participants' viewing order. The first video sequence was watched above average for about 1.5 times. A significant decrease can be seen here, as well as in the previous evaluation. On average, the video sequences were watched halfway only, lastly. This may be influenced by the participants increasing knowledge about the video sequences' order of events. The average rating per video position is closely related to the number of watched frames, as previously evaluated. Accordingly, a decrease in the subjective assessment can be observed with the progress of the study.

The slight decrease in ratings over the duration of the study suggests that participants become more sensitive to flickering light as the study progresses. As the number of watched frames also decreases over time, subjects needed fewer frames to make their judgement about the viewing quality and focused on known flickering areas. Consequently, especially long lasting flickering light has to be obviated. If flickering light did only occur briefly, it was perceived as less annoying.

6. CONCLUSION

In this article, we conducted a novel detailed psychophysics study on the human visual perception on flicker artifacts caused by amplitude-modulated light sources captured with time-discrete digital image sensors, with a focus on automotive mirror replacement systems. Based on the coarse laboratory study with real-world driving sequences in the first part, in-depth knowledge on the human visual perception of flicker with frequencies up to 30 Hz and 20 % amplitude is gathered. A novel synthetic data set with a frame rate of 60 Hz has been created to quantify high-frequency flicker by 32 test persons.

Flicker sensitivity increases with the amplitude of the flicker, and is perceived above a threshold of approx 5 % amplitude. The frequency dependency showed worst ratings for flicker in the range of 15 to 12 Hz, with increasing amplitude. High-frequency flicker at 30 Hz is comparable to low-frequency flicker at 5 Hz and match similar observations by Kelly [14] with a flickering pattern. Additionally, the mean value of the flicker was evaluated, showing that a smaller mean value is more annoying to the human and thereby proving Weber's law. During the study, both the viewing duration and the average rating decreased.

Using these results, flicker suppression algorithms and systems can be tuned to suppress flicker in mirror replacement systems to ensure a safe and non-distracting experience.

REFERENCES

[1] N. Behmann, G. Schewior, S. Hesselbarth, and H. Blume, "Selective LED flicker detection and mitigation algorithm for non-HDR video sequences," IEEE Intl. Conf. Consumer Electronics **8** (2018).

[2] N. Behmann and H. Blume, "Psychophysics study on LED flicker artefacts for automotive digital mirror replacement systems." *IS&T Electronic Imaging: Human Vision and Electronic Imaging Proceedings* (IS&T, Springfield, VA, 2020), pp. 234-1–234-6.

[3] J. L. Brown, "Flicker and intermittent stimulation," Vis. Vis. Perception **1**, 251–320 (1965).

[4] R. I.-R. BT "Methodology for the subjective assessment of the quality of television pictures," (Radiocommunication Sector of International Telecommunication Union, Switzerland, 2019).

[5] L. Choi, A. Bovik, and L. Cormack, "A flicker detector model of the motion silencing illusion," J. Vis. **12**, 777–777 (2012).

[6] G. Denes and R. Mantiuk, "Predicting visible flicker in temporally changing images," *IS&T Electronic Imaging: Human Vision and Electronic Imaging Proceedings* (IS&T, Springfield, VA, 2020), pp. 233-1–233-8.

[7] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "CARLA: An open urban driving simulator," *Proc. 1st Ann. Conf. on Robot Learning* (CoRR, Barcelona, Spain, 2017), pp. 1–16.

[8] W. H. Ehrenstein, *Psychophysik* (Spektrum Akademischer Verlag, 2000), Abrufdatum: 19.04.2020. [Online]. Available: www.spektrum.de/lexikon /neurowissenschaft/psychophysik/10517.

[9] J. Farrell, B. Benson, and C. Haynie, "Predicting flicker thresholds for video display terminals," *Proc. SID* (Palisades Institute for Research Services, New York, NY, 1987).

[10] IEEE P2020 Working Group, "IEEE P2020 Automotive Imaging: White Paper," *IEEE P2020 Automotive Imaging. IEEE-SA* (IEEE, Piscataway, NJ, 2018). [Online]. Available: https://www.image-engineering.de/content/li brary/white_paper/P2020_white_paper.pdf.

[11] D. H. Kelly, *Flicker* (Springer, Berlin Heidelberg, 1972), pp. 273–302, [Online]. Available: https://doi.org/10.1007/978-3-642-88658-4_11.

[12] S. Kitsinelis, "Led flicker: A drawback or an opportunity?," Opt. Photonics J. **3**, 63–66 (2013).

[13] D. Kelly, "Flicker fusion and harmonic analysis," J. Opt. Soc. Am. **51**, 917–918 (1961).

[14] D. H. Kelly, "Flickering patterns and lateral inhibition," J. Opt. Soc. Am. **59**, 1361–1370 (1969).

[15] B. Mühlstedt and J. Roßner Patrick, "Die dunkle Seite des Lichts. Diskomfort durch Flicker bei (LED-)Lichtern im Straßenverkehr in Bezug zu peripheren Flimmerverschmelzungsfrequenzen," *Berliner Werkstatt Mensch-Maschine-Systeme* (Universitätsverlag der TU Berlin, Berlin, Germany, 2013), Vol. 10, pp. 408–416.

[16] P. Ni, R. Eg, A. Eichhorn, C. Griwodz, and P. Halvorsen, "Spatial flicker effect in video scaling," *2011 Third Int'l. Workshop on Quality of Multimedia Experience* (IEEE, Piscataway, NJ, 2011), pp. 55–60.

[17] B. E. Rogowitz, "A practical guide to flicker measurement: using the flicker-matching technique," Behav. Inform. Technol. **5**, 359–373 (1986).

[18] A. Wilkins, J. Veitch, and B. Lehman, "LED lighting flicker and potential health concerns: IEEE standard PAR1789 update," *IEEE Energy Conversion Congress and Exposition* (IEEE, Piscataway, NJ, 2010), pp. 171–178.

[19] S. Wu, S. A. Burns, A. Reeves, and A. E. Elsner, "Flicker brightness enhancement and visual nonlinearity," Vis. Res. **36**, 1573–1583 (1996).

J. Percept. Imaging
IS&T International Symposium on Electronic Imaging 2021

010401-9

Jan.-June 2021
Human Vision and Electronic Imaging 2021