DOI: 10.2352/ISSN.2470-1173.2016.16HVEI-124

# Image Quality of Experience: A Subjective Test Targeting the Consumer's Experience

***Michele A. Saad, Intel Corp., Austin, TX, USA; Margaret H. Pinson, NTIA's Institute for Telecommunication Sciences (NTIA/ITS), Boulder, CO, USA; David G. Nicholas, Intel Corp. Hillsboro, OR, USA; Niels Van Kets, Ghent University–iMinds, Ghent, Belgium; Glenn Van Wallendael, Ghent University–iMinds, Ghent, Belgium; Ramesh V. Jaladi, Intel Corp., Santa Clara, CA, USA; and Philip J. Corriveau, Intel Corp. Hillsboro, OR, USA.***

## Abstract

*This paper presents a subjective image quality experiment design that specifically targets understanding consumer perception of visual quality. With this consumer-oriented goal in mind, images were captured from a large number of representative consumer devices. This design choice is fundamentally different from previous work in the field, in that it is not based on simulated software impairments, but rather on naturally occurring artifacts due to the device and conditions. New subjective questions were posed, which allowed exploration of differences in the quality of experience (QoE) provided by various devices. This paper presents the experimental design implemented to capture ratings and feedback on these consumer devices. The study is analyzed to highlight and demonstrate the unique capabilities provided by this method.*

## Introduction

Subjective image quality tests are an important tool for understanding human perception. Historically, image quality researchers looked to broadcasters for solutions, re-applying the subjective methods and experiment designs developed for broadcast video quality (e.g., ITU-T Rec. BT.500). For broadcast video applications, it is reasonable to exclude camera impairments from the experiment, as broadcasters begin with studio level contribution quality video. Subjective video tests typically focus on quality drops within the distribution chain.

However, with the pervasive proliferation of video onto all types of devices, the interesting topics for image quality shift to center on consumer created content. Understanding consumer demands for better quality of experience from cameras embedded into smartphones becomes critical. For such applications, the camera is the primary source of impairment. Image files are small and so often left in the camera's native format—or post-processed to improve the image quality. Image quality subjective test methods and experimental designs should reflect this difference.

Most image quality tests follow the traditional experiment design used by many subjective video testing facilities and experts. A set of low resolution images (e.g., 0.2 to 0.4 MP) is impaired with defined levels of simulated distortions. This experiment design is reflected in the databases publicly available for researchers, such as the LIVE Image Quality Assessment Database [1], the Tampere Image Database (TID) [2], the Categorical Subjective Image Quality Database (CSIQ) [3], the IRCCyN/IVC Image Quality Database [4], and the MICT Toyama Image Quality Evaluation Database [5].

A major limitation of [1]–[5] is that they contain image distortions and distortion levels that are not commonly encountered in consumer scenarios. These experiments are designed to analyze simulated distortions, such as JPEG compression, blur, or additive noise. Such distortions are not representative of the types of artifacts encountered in photographs taken by consumer devices. Cameras inherently introduce a complex mixture of subtle artifacts such as poor white balance, focus problems, lens distortion, and chromatic aberration. Another limitation of [1]–[5] is that the images have much lower resolutions than the resolutions of photos coming from today's state-of-the-art devices: mobile devices, point-and-shoot cameras (compacts), and digital single-lens reflex cameras (DSLRs).

Table 1 compares the characteristics of images used in traditional image quality tests with those captured by consumer devices. Traditional image quality subjective tests cannot analyze or model the complex and subtle interaction of artifacts in consumer content.

**Table 1: Comparison between Traditional Image Quality Subjective Tests and Consumer Devices**

| | Traditional Experiment | Consumer Devices |
|---|---|---|
| Resolution | < 1MP | 5 to 40 MP |
| Impairments | Software | Hardware plus software |
| Artifacts | Single distortion | Confounding distortions |
| Content | Low diversity | High diversity |
| Camera | Excluded | Included |

The mismatch between traditional experiments and consumer content can be demonstrated by examining no-reference image quality assessment (NR-IQA) models that were trained on publicly available databases. NR-IQA model design relies on subjective ground truth data from subjective tests for accurate modeling, score prediction, and algorithm testing. Saad, Corriveau, and Jaladi [10] show that the prediction accuracy typically drops when NR-IQA models are tested on consumer content instead of singly occurring simulated distortions. This is attributed to a lack of representative consumer content and associated subjective scores to train and develop models on.

A recent study by Virtanen et al. [6] begins to address some of the limitations of the experiments in [1]–[5]. The authors of [6] collected photos from consumer devices. Their subjective test included non-standard changes to the subjective method that are not relevant to this paper's focus (i.e., regarding image presentation order and the use of anchors). The test also included a non-standard rating method, which will be discussed later.

While [6] is an important step towards understanding complex and subtle consumer content artifacts, the experiment design has a number of limitations. First, the content diversity is low. The images

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.16HVEI-124

in the dataset are derived from eight scenes, six of which are photographs of people and none of which are night scenes. Second, the resolution of the images is fixed. The images were all scaled to one resolution before subjective testing and distribution; the original images are unavailable. Third, all details of the image capture protocol are missing (e.g., camera settings, use of tripod).

This paper proposes an experiment design that is targeted at understanding consumers' visual quality of experience. This method includes camera impairments and is intended for comparisons between multiple cameras. This subjective image quality test design is analyzed with respect to the new types of questions that can be asked. Future work is needed on how to evaluate quality of experience.

## Consumer Content Experiment Design

In this section, an experiment design is proposed to compare the image quality of consumer cameras. This design is called the *consumer content experiment design*. This experiment design reflects the real world situation where a consumer compares several different cameras by taking pictures of the same scene with multiple cameras.

### Taking Pictures to Compare Cameras

The first design element in the experiment is inclusion of images from multiple cameras. To allow for comparisons, the experimenter must closely reproduce the same photo using a variety of different cameras.

A popular method to compare cameras is to take the same photo with different cameras under extremely controlled circumstances. For example, a test chart (a chart of artificial patterns and textures) is set up in a lab and a photograph is taken with each camera. The goal is precision (e.g., using controlled lighting, tripod, precise framing). In this scenario, the quality evaluation consists of a comparison between the photograph of the chart and the actual high quality laboratory chart. Examples include the Digital Photography Review (DPReview.com) studio shot comparison tool, the Imatest Master software, and the Imaging Resource Comparometer. Figure 1 is an example of a test chart from Imatest. This technique omits much of the natural variation found in consumer images.

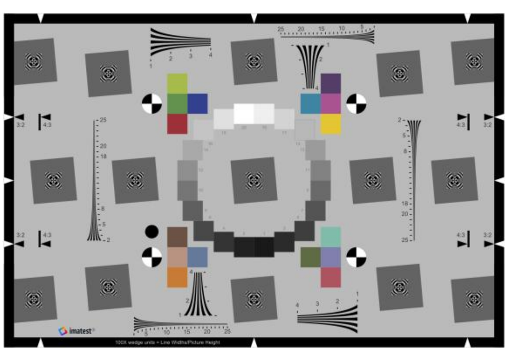

*Figure 1. This eSFR ISO chart from Imatest implements the ISO 12233:2014 standard, with some added features. See http://www.imatest.com/docs/esfriso_instructions/*

The consumer content experiment design instead uses *equivalent images*: photographs that depict approximately the same scene. Equivalent images are created by picking up two or more cameras in succession and taking photos of a scene, without the use of a tripod or measuring tool. Equivalent images could also be created using different camera settings (e.g., automatic focus vs. manual focus). The goal is to emulate the level of repeatability that an average person could obtain.

All pictures associated with a single scene will be referred to as an *equivalent image set*. There will be inherent differences within an equivalent image set due to camera differences (such as aspect ratio, focal length, variation in exact camera positioning) and slight changes in the scene (such as moving clouds, fluctuations in natural light). Unless post-processing options are to be considered, the range of image quality and array of image artifacts comes exclusively from the camera systems used to capture the photos for the study.

### Choosing Scenes to Compare Cameras

The second design element is the scene selection criteria. From video quality subjective test research, it is known that the scenes used in an experiment are nearly as influential on the analysis and conclusions as the impairments themselves. Pinson, Barkowsky and Le Callet [10] recommend criteria for choosing a robust set of video sequences. Based on that prior work, it is recommended the experiment contain various scenes with a wide distribution of the following factors:

- Focal length (close to far)
- Lighting levels (low to high)
- Light type (natural and artificial)
- Light direction (relative to lens)

These distributions are not intended to be uniform. For example, there is more interest in various low lighting conditions (which bring out interesting camera behavior) than in high lighting conditions.

Additionally, the scenes should include interesting characteristics:

- Areas of the image containing solid color (e.g., white paper, grey backdrop), particularly in the presence of low lighting
- Reflective surfaces (e.g., glass, metal, snow)
- Fine detail (e.g., high spatial information)
- Spatial details at different frequencies
- Bright colors
- Muted colors (e.g., composition of black, white and grey)
- Smoothly graduated color or luma changes (e.g., blue sky, sunset)
- High contrast edges
- Low contrast picture
- Fine textures (e.g., grass, sand, fabric weave)
- Natural edge angle distributions (e.g., as seen in nature)
- Artificial edge angle distributions (e.g., as seen in building architecture)

Finally, the scenes should include a variety of subject matter (e.g., landscapes, portraits, night scenes). In addition to making the test more interesting for subjects, this variety increases the likelihood that the scenes will include a robust variety of interesting characteristics.

### Choosing the Display Method

The third design element is the display method. The consumer content experiment design will inevitably include pictures with differing pixel counts. Pixel count means the total number of pixels in an image, calculated as the height in pixels times the width in pixels. This is typically reported in megapixels (MP).

During the subjective test, all images are scaled to contain the same number of pixels as the monitor used for display. This allows

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.16HVEI-124

subjects to rate images on a single monitor and avoids the biasing factor that would occur if images were displayed at native pixel counts (e.g., an 8 MP image will appear larger than a 5 MP image on a 4K monitor). This solution is recommended as it matches the most popular application today (i.e., digital images viewed on a digital display). The rescaling is intended to be a transparent process.

### *Open Issues - How to Add Quality of Experience*

As presented to this point, the consumer content experiment design can be defended on the basis of prior work. The design is more realistic (i.e., better matched to the consumer content application) yet less accurate (i.e., contains confounding factors that complicate data analysis). The impact upon repeatability must be determined; this topic will be discussed later.

The fourth design element is Quality of Experience (QoE), a missing element that is needed to better understand consumer demands. QoE is a holistic approach to understanding a consumer's reaction to the images produced and displayed. Understanding the entire experience ventures away from traditional solely quantitative ratings obtained from a Likert scale to exploring qualitatively the feelings and reactions users have to the content produced. Let us first examine prior research on QoE.

Virtanen et al. [6] ask subjects to rate the overall image quality plus four image characteristics: sharpness, graininess, lightness, and saturation. The task of rating these image characteristics likely focuses attention on these pre-defined image characteristics. The overall quality scores are likely impacted by subjects' opinion of these characteristics [7]. Whether this is desirable or undesirable is not obvious. ITU-T Rec. P.835 has subjects rate the same speech stimuli on multiple scales for improved analysis of noise suppression algorithms. However, the situation for image quality is more complex. Cameras introduce impairments that are difficult to isolate. Also, preliminary work in [9] suggests that asking subjects to rate multiple image characteristics at one time can cause their responses in certain outlier cases to be artificially interdependent.

Pinson, Sullivan and Catellier [14] propose the use of distractor questions. Subjects answer multiple questions for each stimulus, yet (unlike [6]) only one question rates the stimulus on a perceptual scale. Distractor questions are used in addition to the traditional mean opinion score (MOS) scale. Distractor questions serve to:

- Focus the subject on the clip as a whole (instead of only the clip's quality)
- Emphasize the intended application
- Provide added information

The same distractor questions are asked for all stimuli, thus the distractor questions must be generally applicable. Multiple-choice questions are recommended to keep the overall cognitive task low. Narwaria et al. [8] demonstrated this technique.

Radun et al. [7] used Interpretation-Based Quality (IBQ) to solicit descriptive user feedback. In IBQ, subjects are asked to describe their own subjective quality impressions of images in their own words. This task is in addition to quantitatively rating the stimuli. In [7], the authors designed a subjective study that requires participants to rank printed images of simulated image signal processor (ISP) pipeline outcomes with the aim of understanding the effects of ISP tuning parameters on image quality.

Distractor questions and IBQ are worth further investigation. In addition to focusing the subject on the intended application, distractor questions can be QoE questions that are unrelated to MOS. IBQ can yield better understanding of the consumer perspective. The fourth design element is the optional inclusion of QoE, via one or both of these techniques.

## Implemented Test: CCRIQ

Saad et al. [12] describe the Consumer-Content Resolution and Image Quality (CCRIQ) test: a subjective test conducted according to the *consumer content experiment design*. The goal of CCRIQ was to better understand the relationship among camera type, image pixel count, monitor resolution (HD vs 4K), camera characteristics (optics and post processing), and the overall perceived quality. See [12] for their conclusions. The CCRIQ dataset is made available to the research community in the Consumer Digital Video Library (CDVL, www.cdvl.org)[1]. This section provides an overview of CCRIQ in the context of analyzing the consumer content experiment design. For more details, we refer the reader to [12].

### *Overview*

The CCRIQ test spans 23 cameras and 18 equivalent image sets. The cameras were 11 phones, 2 tablets, 6 compacts, and 4 digital single reflex lens (DSLR) cameras. The cameras were chosen to have sensor resolutions ranging from 1 megapixel (MP) to 20 MP. The 18 equivalent image sets were photographed and selected according to the criteria described above. Scene content was limited to five content categories: landscape scenes, outdoor night shots, portraits, indoor still life arrangements, and flat surfaces. The scene depths were also chosen to vary widely. CCRIQ contains all natural images and none with artificially simulated distortions. The experiment design was a full matrix of equivalent image sets and cameras.

CCRIQ consists of two studies. The primary study (described in [12]) analyzes the combined effects of image pixel count and monitor resolution on the overall perceived quality. This study was conducted using the absolute category rating (ACR) method from ITU-T Rec. P.913. The secondary study analyzes a smaller set of images using the QoE techniques described above (IBQ and distractor questions) and is omitted from [12].

The secondary study applies the IBQ approach described in [7] to ask subjects to describe in their own words their subjective quality impressions of the stimuli presented to them in addition to quantitatively rating the stimuli. This secondary study also includes a multiple choice question that asks subjects to guess what type of device was used to take the photo (whether phone, tablet, point-and-shoot camera (compact), or professional camera (DSLR)). We note that in the remainder of the text we use the terms *point-and-shoot* and *compact* interchangeably and the terms *professional camera* and *DSLR* interchangeably. Understanding quality from a consumer perspective requires one to shift the focus from the traditional experimental design to one that addresses subjective quality assessment holistically. The coupling of quantitative data in the form of mean opinion scores (MOS) with qualitative feedback from consumers enables a different perspective.

The authors note that [12] only describes a portion of the study with the primary focus being on understanding specific camera

[1] The images in the dataset were left in their original file resolution as saved by the camera.

DOI: 10.2352/ISSN.2470-1173.2016.16HVEI-124

characteristics. Paper [12] does not elaborate on the complete study design and there is no mention of the secondary study. In this paper the focus is on describing a subjective test experiment design that is non-traditional and consumer oriented. Data will be presented that begins to explore the link between image quality ratings and how consumers describe those ratings, along with why they describe the ratings as they do. This new approach brings a level of understanding and investigation around consumers' expectations and why they like or dislike an image that can help further define, develop, and deploy objective models; it also provides a richer dataset to the research community.

### *The Subjective Test*

CCRIQ was conducted by three laboratories: 1) NTIA/ITS in Boulder, CO, USA; 2) Intel Corp. in Santa Clara, CA, USA; and 3) Ghent University-iMinds in Ghent, Belgium. ITS and Ghent University provided 18 participants each, and Intel provided 17. None of the participants were imaging experts. Visual acuity and color deficiency tests as well as lab illumination and viewing distance conditions were in compliance with ITU-T Rec. P.913 (summarized in [13]). The test consisted of two sub-studies: the primary study and the secondary study. Each was preceded by a short training phase to familiarize the subjects with the tasks.

The primary study used the standard single stimulus absolute category rating (ACR) method and augmented it to include the presentation of the stimuli on two identical 28" monitors: one configured for HD resolution and the other for 4K. The order of stimulus presentation on one monitor or the other was fully randomized. When a stimulus was presented on one monitor, the other monitor displayed the rating page that instructed the subjects to provide a quality rating. A discrete 5 point Likert scale was used (where 5 corresponded to Excellent, 4 to Good, 3 to Fair, 2 to Poor, and 1 to Bad). A total of 392 images were rated in the primary study. To avoid subject fatigue, the set of 392 images was split into two overlapping pools. Each subject rated either 218 or 221 images, and 45 images were rated by all subjects. Each stimulus was rated by at least 26 subjects. All data associated with two images in the primary study were discarded due to naming errors (i.e., the images were taken by the wrong camera).

The secondary study extended the consumer oriented focus of the test to include user qualitative feedback in addition to the ACR ratings. In this sub-study, the stimuli were presented on the 4K configured monitor only. The second monitor was used to display the study questions. Twenty image stimuli were used in this portion of the study: 5 from DSLR cameras, 6 from point-and-shoot-cameras, 4 from tablets, and 5 from phones. The secondary study was small, because it was considered to be an exploratory test. The goal was to evaluate IBQ and distractor questions.

Participants were asked three questions for each stimulus presented to them. Similar to the primary study, the first question was to rate the quality of the stimulus on a 5 point Likert scale. The second question was a multiple choice question that required subjects to guess the type of device that was used to capture the photo. The multiple choice options were 1) phone, 2) tablet, 3) point-and-shoot camera, or 4) professional camera. This question ties into the study consumers' expectations of the various device types. Finally, the third question used the IBQ approach to ask subjects to describe in their own words their subjective experience of the stimulus by explaining the rating they awarded it. The subjects were allowed to type their open-ended responses. There was no limit on the number of words allowed.

No subject data was discarded in post analyses. The IBQ free-responses were modified as follows: translated into English (where necessary), spelling corrected, and common English words discarded (e.g., "a", "the", "and").

## Analysis

A summary of the conclusions reached by the primary study will be presented first. See [12] for other analyses of the primary study data, including camera characteristics and their relationship to perceived image quality. Next we present an analysis of the stability of the experiment design and the secondary study. Along the way, the relationship between our experiment design and the data will be examined.

### *Primary Study Analysis*

The HD and 4K MOS ratings were highly linearly correlated, with a Pearson linear correlation coefficient of 0.979. Analyses were conducted using a two-sided Student's *t*-test at the 95% significance level. Within the [3.0..5.0] range, the 4K monitor MOSs were 0.2 higher than the HD monitor MOSs. Within the [1.0..3.0] range, the 4K and HD monitor MOSs were statistically equivalent. Due to the close similarity between these two sets of MOS values, our other analyses combine the 4K and HD ratings into a single MOS value.

The 23 cameras in CCRIQ were chosen to represent the variety of popular consumer devices on the market in 2014. These cameras span a wide range of characteristics such as sensor size, resolution, lens maximum aperture, price, and form factor. Popular consumer devices on the market tend to generally produce good quality images.

Consumer images, however, span a wide quality range. One factor is aesthetics, so CCRIQ emphasizes scenes with good aesthetics. Another factor is lighting, so CCRIQ scenes span a wide range of light conditions:

- Natural light, full sun (3)
- Natural light, shade (2)
- Bright artificial light (3)
- Dim artificial light (5)
- High contrast ratio, bright to dark (2)
- Night scene (3)

This combination of scenes and cameras produced scores that are fairly well distributed between 1 and ≈4.8, with a slight peak around 3.8 (see Figure 5 of [12]). This distribution was achieved through the use of a preliminary test using seven subjects, run to check and balance the experiment design. The scenes were adjusted and the cameras selection did not change. Note that this reverses the traditional experiment design, where the scenes are retained and the impairment levels are adjusted—likely by adjusting simulating distortion levels to more uniformly sweep the quality range.

CCRIQ contains two compositions that were each photographed both with autoflash and with flash disabled (if autoflash triggered). This yielded 29 image pairs that compare camera performance with and without flash. Each scene composition depicted a portrait (head and shoulders) of a model with professional makeup and hair. Figure 2 shows the difference in MOS between the flash and no flash images, and Figure 3 shows sample image pairs. The quality of DSLR and compact cameras was sometimes dramatically improved by the auto-flash, while the quality of phone cameras typically was not. Neither tablet in the study had a flash.

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.16HVEI-124

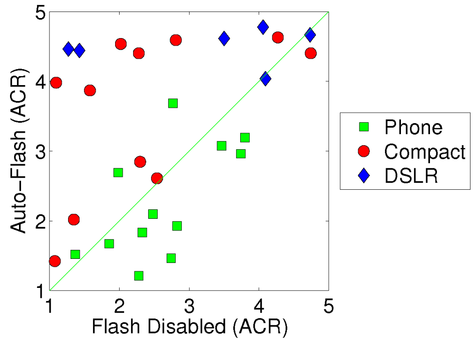


*Figure 2. Comparison between MOSs for autoflash and flash-disabled.*

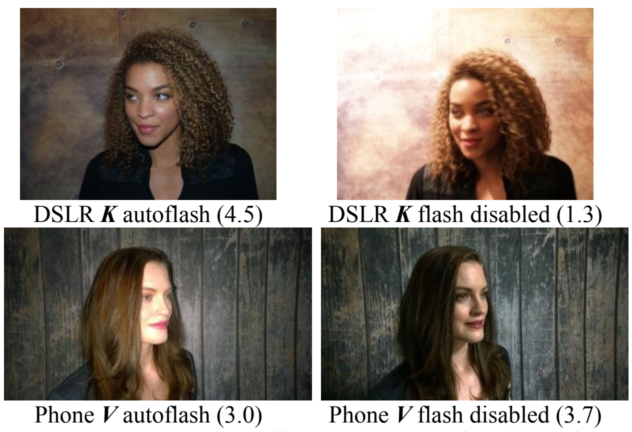


*Figure 3. Photos demonstrate the different quality impact of autoflash (left) and flash disabled (right) based on camera type. Shown are two sample cameras: DSLR **K** and phone **V**. MOS is given in parentheses. Camera **K** with autoflash disabled yields surprisingly poor perceptual quality yet closely matches the photographer's memory of the scene.*

## *Lab-to-Lab Comparisons*

The inter-lab MOS scores from both the primary and secondary tests were highly linearly correlated. Table 2 shows the Pearson correlation coefficients between MOS scores from each of the test labs. The high correlation scores between the labs are a strong indication of the stability of the study design and the repeatability of the obtained scores.

**Table 2: Primary and Secondary Study Inter-Lab MOS Pearson Linear Correlation**

| Primary Study | | | |
|---|---|---|---|
| | Ghent | ITS | Intel |
| Ghent | 1 | 0.952 | 0.915 |
| ITS | | 1 | 0.941 |
| Intel | | | 1 |

| Secondary Study | | | |
|---|---|---|---|
| | Ghent | ITS | Intel |
| Ghent | 1 | 0.975 | 0.964 |
| ITS | | 1 | 0.962 |
| Intel | | | 1 |

## *Secondary Study Analysis*

The secondary study asked subjects three questions:

- Please rate this photo according to its quality:
  5 (excellent) 4 (good) 3 (fair) 2 (poor) 1 (bad)
- What device do you think this photo is coming from?
  Phone, tablet, point and shoot camera, or professional camera
- Briefly describe the characteristics of this image that influenced your rating of the overall quality.

Figure 4 shows a word cloud from the responses to question #3, the IBQ free-text responses.[2] The size of the words in the cloud is proportional to their frequency of occurrence. The word cloud reveals three dominant themes: 1) scene clarity, detail, or sharpness, 2) lighting, and 3) color.


*Figure 4. Word cloud generated for user descriptive verbal feedback on stimulus quality in the secondary study. The larger the size of the words in the cloud, the more frequent their occurrence among subjects' responses.*

Each row of Table 3 contains the distribution of subject guesses from question #2, based on images associated with one device. The four devices are approximately evenly represented in the secondary study, yet not equally likely to be guessed.

**Table 3: Subject Guesses per Device Category**

| Actual Device | Subject Answers | | | |
|---|---|---|---|---|
| | Phone | Tablet | Compact | DSLR |
| Phone | 41.1% | 14.4% | 28.2% | 16.3% |
| Tablet | 39.3% | 18.5% | 30.6% | 11.6% |
| Compact | 29.3% | 16.4% | 35.2% | 19.1% |
| DSLR | 18.9% | 9.2% | 38.9% | 33% |

The top of Figure 5 shows the distribution of subject answers to questions #1 and #2 (device guess and ACR). The bottom of Figure 5 shows how the distribution changes when question #2 is

[2] Website "Wordle" was used to generate this figure.

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.16HVEI-124

replaced with the actual devices. These two figures indicate a discrepancy between consumers' preconceived notions and subjective quality. Consumers expect a large quality drop when moving from DSLR to compact to tablet or phone. In reality, the quality drop is more subtle; some phones can produce excellent quality images. Figure 6 confirms this trend, based on the distribution seen in the primary data.

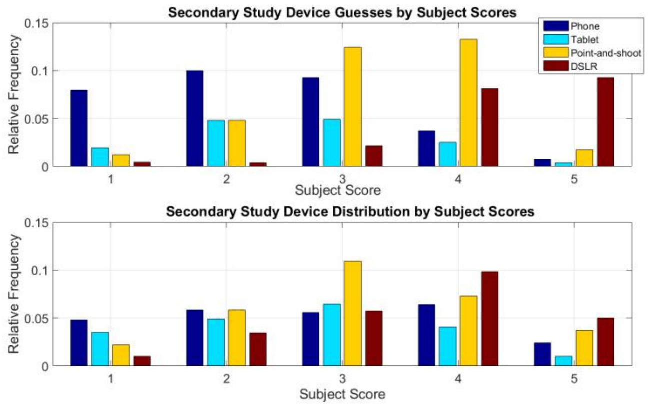


*Figure 5. Relative frequency of guessed device type (top) and actual device type distribution (bottom) for each ACR score value These charts are calculated from individual subject ratings. The x-axis is the 5-level Likert ACR scale.*

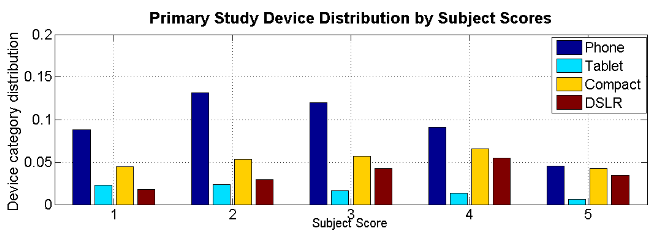


*Figure 6. Distribution of ACR scores and device types based on the primary data.*

## Conclusion

In this paper, the consumer content experiment design is described and assessed. This experiment design is intended to compare the image quality produced by several different cameras. The CCRIQ experiment is presented as an example of the novel questions that can be asked: distractor questions (multiple choice) and IBQ (free response). These added questions allow the test to delve deeper into QoE and user experience. However, the rating cycle was longer; IBQ is particularly time consuming.

The use of cameras as impairments adds time and complexity to the experiment implementation. In addition to the cost of purchasing various cameras, the image collection process was prone to human error. It is easy to forget to take a picture with one camera among many, and ephemeral scenes may be impossible to recreate. Other errors included selecting the wrong camera setting, dead batteries, malfunctioning cameras, and mislabeling images. Despite these challenges, the proposed experiment design is worth the extra effort, due to the wealth of insights into camera behaviors.

## References

[1] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik, “LIVE image quality assessment database release2,” 2006, http://live.ece.utexas.edu/ research/quality/subjective.htm.

[2] N. Ponomarenko et al., “TID2008 Ð a database for evaluation of full-reference visual quality assessment metrics,” 2008, http://www. ponomarenko.info/tid2008.htm.

[3] E. C. Larson and D. M. Chandler, “Consumer subjective image quality database,” 2009, http://vision.okstate.edu/index.php?loc=csiq.

[4] P. Le Callet and F. Autrusseau, “Subjective quality assessment IRCCyN/IVC database,” 2005, http://www.irccyn.ec-nantes.fr/ivcdb/.

[5] Z. M. Parvez Sazzad, Y. Kawayoke, and Y. Horita, “MICT image quality evaluation database,” 2008, http://mict.eng.u-toyama.ac.jp/mictdb.html.

[6] Virtanen, T., Nuutinen, M., Vaahteranoksa, M., Oittinen, P. and HŠkkinen, J. “CID2013: a database for evaluating no-reference image quality assessment algorithms,” IEEE Transactions on Image Processing, vol. 24, no. 1, pp. 390-402, Jan. 2015.

[7] J. Radun, T. Leisti, T. Virtanen, J. HŠkkinen, T. Vuori, and G. Nyman, “Evaluating the multivariate visual quality performance of image-processing components,” ACM Transactions on Applied Perception, vol. 7, no. 3, June 2010.

[8] M. Narwaria, M. P. Da Silva, P. Le Callet, and R. Pepion, “Single exposure vs tone mapped High Dynamic Range images: A study based on quality of experience,” Signal Processing Conference (EUSIPCO), Sept. 2014.

[9] M.A Saad P. Corriveau, R. Jaladi, “Revealing the Dark Side of a Subjective Study: Learnings from Noise and Sharpness Ratings”, 7th International Workshop on the Quality of Multimedia Experience (QoMEX), May, 2015.

[10] M. H Pinson, M. Barkowsky, P. Le Callet, “Selecting scenes for 2D and 3D subjective video quality tests,” EURASIP Journal on Image and Video Processing, 2013.

[11] M.A Saad P. Corriveau, R. Jaladi, “Objective Consumer Device Photo Quality Evaluation,” IEEE Signal Processing Letters, vol. 22, no. 10, pp. 1516-1520, February, 2015.

[12] M.A. Saad, M.H. Pinson, D.G. Nicholas, N. Van Kets, G. Van Wallendael, R. Da Silva, R.V. Jaladi, and P.J. Corriveau, “Impact of Camera Pixel Count and Monitor Resolution on Perceptual Image Quality”, IEEE Color and Visual Computing Symposium (CVCS), August 2015.

[13] M. Pinson and L. Janowski, “A new subjective audiovisual and video quality testing recommendation,” VQEG eLetter, no. 2, vol. 1, 2014. Available: http://www.vqeg.org/.

[14] M. H. Pinson, M. Sullivan, A. A. Catellier, “A new method for immersive audiovisual subjective testing,” Eighth International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM 2014), Chandler, AZ, Jan. 30-31, 2014.