

VQone MATLAB toolbox: A graphical experiment builder for image and video quality evaluations

Mikko Nuutinen

Institute of Behavioural Science, PO Box 9
(Siltavuorenpenger 1 A), University of Helsinki, e-mail:
mikko.nuutinen@helsinki.fi

Toni Virtanen

Institute of Behavioural Sciences, University of Helsinki

Olli Rummukainen

Department of Signal Processing and Acoustics, Aalto
University

Jukka Häkkinen

Institute of Behavioural Sciences, University of Helsinki

This article presents VQone, a graphical experiment builder, written as a MATLAB toolbox, developed for image and video quality ratings. VQone contains the main elements needed for the subjective image and video quality rating process. This includes building and conducting experiments and data analysis. All functions can be controlled through graphical user interfaces. The experiment builder includes many standardized image and video quality rating methods. Moreover, it enables the creation of new methods or modified versions from standard methods. VQone is distributed free of charge under the terms of the GNU general public license and allows code modifications to be made so that the program's functions can be adjusted according to a user's requirements. VQone is available for download from the project page (<http://www.helsinki.fi/psychology/groups/visualcognition/>).

Introduction

Image and video quality assessment plays an important role in development and optimization of image acquisition, encoding and transmission schemes (Bovik, 2013; Chandler, 2013). In a typical experiment an observer is seeing images in a sequence and the task is to evaluate some property from the images, such as overall quality, sharpness, graininess or saturation or to give a magnitude of difference between the images. Many image rating methods, such as Forced Choice Paired Comparison (PC), Triplet, Absolute Category Rating (ACR), Double Stimulus Impairment Scale (DSIS), Double Stimulus Categorical Rating (DSCR) and Single Stimulus Continuous Quality Evaluation (SSCQE) have been standardized (ISO 20462-1, 2005; ISO 20462-2, 2005; ITU-T P.910, 2008; ITU-R BT.500, 2012). The standards of image rating describe how to display test images and video and possible reference stimuli to observers as well as how to collect rating scores. The reference stimuli can be images with known properties or quality, helping the observers anchor their ratings to something concrete. Moreover, images are often used in human behavioral research (e.g., Leisti, Radun, Virtanen, Nyman, & Häkkinen, 2014; Coco & Keller, 2014; To, Gilchrist, Troscianko, & Tolhurst, 2011).

Different applications and experimental settings require different methods for rating images and videos. For example, there is a compromise between the amount of test images and observer fatigue due to prolonged test durations. The PC method displays two images side by side or sequentially, and the observer's task is to select the image with more of the attribute in question, for example image quality or sharpness. The method excels in finding small, near detection threshold differences between the test stimuli mostly with unidimensional differences. However, the PC method is only suited for experiments with a relatively small number of test images, because the number of image pairs increases exponentially with the number of images (Mantiuk, Tomaszewska, & Mantiuk, 2012).

The ACR method displays one image at a time and image is rated without a reference image. The ACR method is the fastest method for assessing many test stimuli. However, the ACR method can be inaccurate because, when providing rating values for test images, observers use and compare test images with their own internal references which leads to individual differences in the use of the given rating scale.

If the reference image is available and the number of test images is high, the use of rating methods such as DSIS or DSCR can be justified. The DSIS method displays the reference and test images and the observer's task is to define the preference category for the test image. The DSCR method defines the categories of both the reference and test images.

This paper introduces the VQone toolbox for MATLAB, a

This project has been supported by Academy of Finland project "Mind, Image, Picture" (MIPI), (project no. 267061).

still and video quality rating experiment builder. The VQone contains the main elements needed for building and conducting experiments and data analysis. That is, the VQone is a tool for showing prior manipulated stimuli and for recording responses. The VQone toolbox is free of charge and offers an intuitive and comprehensive graphical user interface (GUI). Currently there are many free software packages that are viable tools for creating image rating experiments (see Table 2). However, these packages are limited compared with VQone as far as graphical interface, available rating methods and flexibility are concerned.

VQone enables the creation of a new Dynamic Reference Absolute Category Rating method (DR-ACR) (Virtanen, Nuutinen, Vaahteranoksa, Oittinen, & Häkkinen, 2014) and a wide range of experiments according to image quality standards (PC, triplet, ACR, DSIS, DSCR, SSCQE). All of the setups can also be augmented with the possibility of gathering qualitative free answers from the observers. That is, observers write down in one or two words in the text input field the most important aspects that influenced their quality rating. This allows the researchers to gain descriptive data about the reasons behind the observers' judgments (Nyman et al., 2006; Radun et al., n.d.). The descriptive data complements the standard rating methods by offering a description of what was seen in the test stimuli when a quality decision was made or a preference was expressed.

VQone is not limited to standardized experiments, it can also be used to construct entirely new experimental setups. The user can create and name rating scales, radio buttons and check boxes. Furthermore, the user can modify the sizes and locations of stimulus windows and add different reference stimuli. Because all the settings and experiments are built through the GUIs, VQone provides tools for forming complex image and video rating without the need for programming.

In the first section of the present article, we provide a non-technical description of the basic functionality offered by the experiment builder unit of VQone. All settings and possibilities offered by VQone are presented in the VQone user manual located within the files needed to run VQone on MATLAB. In the second section, we present and analyze samples (distributed with the VQone package) of a typical image rating study. In the third section of this article, we describe how VQone compares with existing software.

At the time of writing, the version of VQone reviewed in the present article is 0.94. It has been used in more than 60 experiments (with our academic and industrial partners) with more than 1800 participants. The VQone software, the user manual and updated list of our academic studies in which VQone has been used can be found from the VQone project page (<http://www.helsinki.fi/psychology/groups/visualcognition/>).

Program description

Software and hardware requirements

The VQone toolbox is written in MATLAB (R2012a) and has been tested under Windows XP/7/8. The processing power required to run VQone depends on the type of experiment. For an experiment, consisting of a sequence of still images, the requirements are modest, and a standard computer system will suffice. We have run VQone on a laptop: 8-GB Ram, 2.50 GHz Intel i5-2520M CPU. When using very large stimulus files, such as 1080p video files, it is recommended that the user test that the system is capable of running the stimuli without freezing or other temporal artifacts.

For running all settings of VQone, it is required that Microsoft Excel and Windows Media Player (WMP) be installed. Moreover, MATLAB distribution should contain the Image Processing Toolbox. All experimental data are reported in the form of Excel spreadsheets. VQone uses the MATLAB `actxserver` function to create, read and write to Excel spreadsheets. The MATLAB `actxcontrol` function is used to open WMP (which is the default video player) on a stimulus window when an experiment, in which video is acting as stimuli, is launched. Moreover, the user will need to ensure that the required codec packs for video and audio compression and the file formats are installed. If still images are only used as stimuli, WMP need not be installed. VQone uses still image codecs (.jpg, .tif, .png, etc) offered by the Image Processing Toolbox for reading image files.

The program structure

The program structure is described in Fig. 1. The handling of VQone is controlled by GUIs, where the user can create a new setup, load already created setups or analyze data. All parameter settings during experiment creation are written in the experiment-specific MATLAB structure array that we call the setup structure. When a new experiment is created, an experiment-specific folder is created and the setup structure is saved there. When a new trial for an already created experiment is conducted or already collected raw data is analyzed, the user should select the setup structure file from the experiment-specific folder. VQone reads from the setup structure file how to display stimuli and questions and how to analyze and visualize raw data.

When a new experiment is built the Excel spreadsheets (.xls) for stimulus filenames, practise data and experiment data are created. The filenames spreadsheet contains the filenames of stimuli that will be used in the experiment. Image and video files (stimuli) are stored in the folders, named 'images' and 'videos', from where they are loaded. Every time that an observer rates stimulus, data is written to the experiment data spreadsheet. A practise data spreadsheet is used when a new experiment setup is tested or piloted.

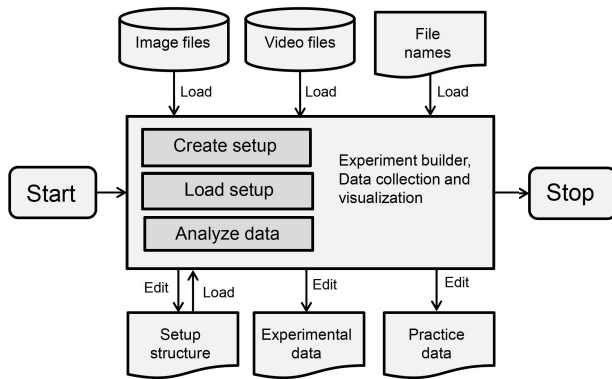


Figure 1. Structogram of the VQone software architecture

The graphical user interface

After typing "VQone" into the MATLAB command window, the *main panel* appears, which contains three subpanels (see Fig. 2). On the left there is a panel for building a new experiment (Create setup). In the center there is a panel for modifying or conducting a previously built experiment (Load setup). On the right there is a panel for analyzing raw data from the conducted experiments (Analyze data).

The name for the experiment is written into the "Setup name" field and according to it, VQone creates and names a new experiment-specific folder. The experiment-specific folder is the place where VQone saves all the experiment-specific files, such as the setup structure file and data spreadsheets. The rating method for a new experiment is selected from *Test Method* pop-up menu (see Fig. 3). The quality rating methods of the ACR, still PC, still triplet, video ACR, video PC, video SSCQE or questions only are available as pre-selection for the experiment building process. It is possible to build modifications of the ACR method, such as the DR-ACR, DSIS and DSCR, via the selection of the ACR.

When the desired quality rating method is selected from the main panel the button *Create* opens the experiment building panel (see Fig. 4 for the PC setup and Fig. 7 for the ACR setup). The content of this panel varies depending on the chosen rating method. In the experiment building panel factors such as the number of stimuli and the contents and randomization of the stimuli as well as what is displayed on different windows (reference, stimuli or questions) can be determined.

The experiment building panel: the PC method. Figure 4 shows the experiment building panel when the still PC quality rating method is selected from the main panel. The number of stimuli is selected from the *CVs* (Content Variable) subpanel. The text fields of "CV1 name" and "CV2 name" determine the names for the content variables. For the example presented in Fig. 4 content variables are named as scenes and images. For this example, the numbers of scenes and images are set to be 3 and 5. That is, in the experiment

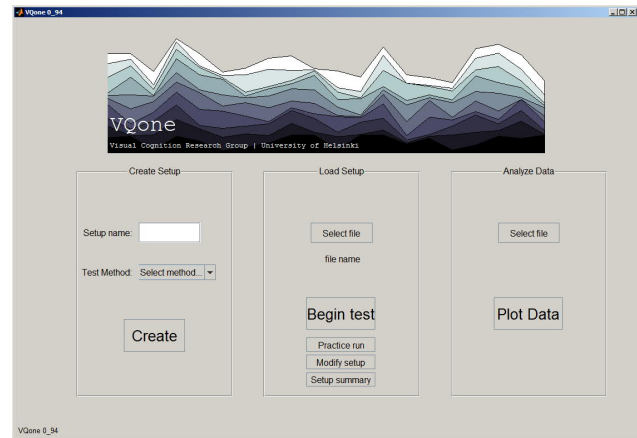


Figure 2. The VQone main panel on start-up. Subpanels: Experimental builder (left), Experimental loader (central), Data analysis and visualization (right)

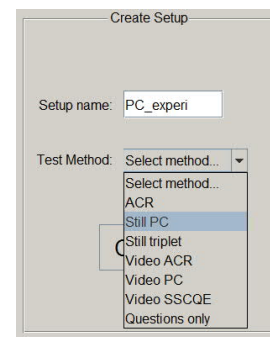


Figure 3. The quality rating method is selected from the pop-up menu of the experiment builder panel

the observers will evaluate 15 images (3 scenes x 5 images). Figure 5 shows the thumbnails for the five images from three scenes used for clarifying the settings.

The randomization type of stimuli presentation can be set in the *Stimulus options* panel. Figure 4 shows the settings in which "random presentation" is selected from the pop-up menu. The text "CV1 is locked after the first pick and the CV2s are gone through randomly for each CV1" meaning the experiment is one in which all image pairs from a random scene will be displayed in a random order before moving on to the next random scene. The other option of the pop-up menu is "serial presentation", in which all image pairs are gone through serially for each scene.

The text field of "Set viewing time for the pair" determines the length of time a stimulus pair is displayed. If the selection is set to be e.g. 3, the stimulus pair will be shown on the display for 3 seconds. If the selection is set to be 0, the stimulus pair will only change when the observer clicks the "Next Image Pair" button.

A mask stimulus can also be presented between stimuli using the check box *Masking image between trials*, and if each

CVs

CV1 name: Number:

CV2 name: Number:

☐ Ask for feedback after the test
Please feel free to give us any feedback from the ex...

Stimulus Options

Random presentation

CV1 is locked after the first pick and CV2s are gone through randomly for each CV1.

Set viewing time for the pair. Seconds or 0 until button press:

☐ Masking image between trials
☐ Pair repetition

Set maximum difference for product indices:

Monitor options

		Stimulus Size		Stimulus position	
Stimulus 1	Left	600	376	1	560
Stimulus 2	Center	600	376	560	560
Disabled	Right	600	376	1120	560
Questions	Bottom	700	500	200	37

☐ Randomized mo...

Figure 4. In the experiment builder panel of the PC method, the stimulus options such as randomization, viewing time and masking can be set. In addition factors such as the number of images and scenes and the stimuli window sizes and positions can be selected

stimulus pair will be repeated in the experiment using the check box *Pair repetition*. If pair repetition is enabled all the image pairs will be presented twice in each experiment. The default masking image is a white noise stimulus with a center fixation point. The masking image, which aims to clear the iconic memory buffer in the human visual system, prevents the observers from focusing only on differences between two concurrent stimuli (Sperling, 1960; Rensink, 2014). The default masking stimulus file (masking_image.jpg) is located in the image folder and users can change or modify it according to their research requirements.

The field of "Set maximum difference for product indices" determines whether one stimulus will be compared to all the other stimuli, or for example only to the few most similar stimuli. If each stimulus is to be compared to all the other stimuli, the default choice of Inf should remain; if the stimuli should only be compared to a few others, the number of stimuli pairs for each stimulus should be defined. In that case the closest stimuli will be determined according to the order of files in the stimulus filename spreadsheet file, which must be edited for the desired combinations. For instance, if the maximum difference for product indices is set to be 3, the first image in the stimulus filename spreadsheet list will be compared to the second and third image listed, the second one will be compared to the third and fourth, and so forth. The option to compare a stimulus only to the most similar

stimuli can be used for reducing the number of image pairs and the duration of an experiment.

The positions and sizes of the experiment windows are set from the *Monitor options* panel (Fig. 4). The VQone can show three images or videos at the same time on three different experiment windows. These windows are named Left, Center and Right on the *Monitor options* panel. In addition, the fourth window (Bottom) is reserved to show the response buttons or sliders. The fields of "Stimulus size" and "Stimulus position" determine the sizes and positions of the experiment windows. It is worth to mention that experiments could also be designed to rate the differences between different types of monitors by positioning the experiment windows on the different displays and presenting identical stimuli on all experiment windows. Figure 4 shows the settings for the experiment, in which the "Left" experiment window is selected for Stimulus 1, the "Right" for Stimulus 2 and the "Bottom" for questions. By pressing the *Preview displays* button the positions and sizes of stimulus windows are shown and the experiment window options are easy to fine tune (Fig. 6).

The experiment building panel: the ACR method.

Figure 7 shows the ACR experiment building panel. Compared to the experiment building panel of the PC method, the ACR method contains the *Referencing* and *Randomizing options* subpanels. The *Randomizing options* subpanel

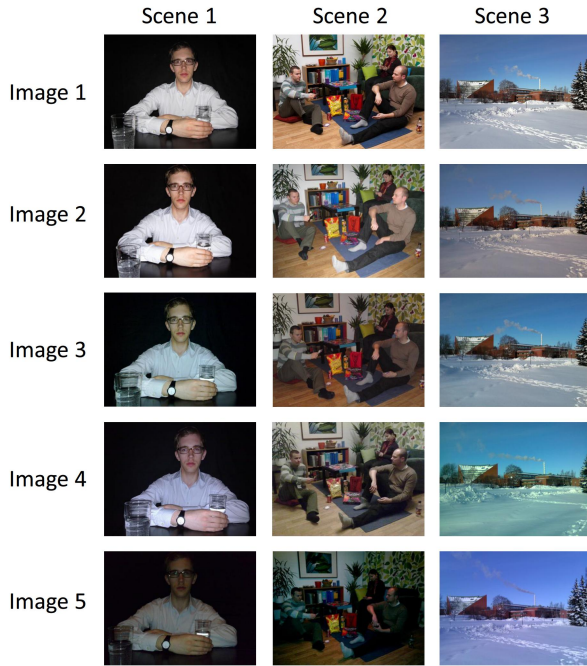


Figure 5. An example matrix of five images from three scenes

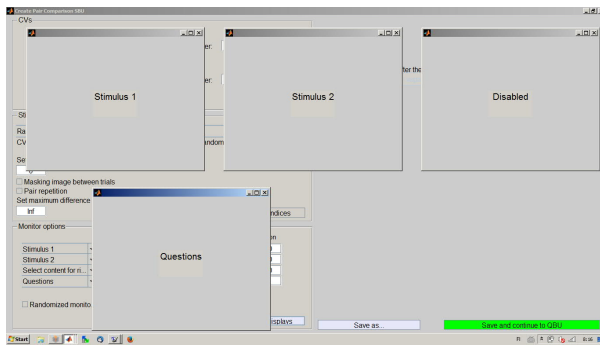


Figure 6. The positions and sizes of the stimulus windows can be previewed while building new experiment setups

is used for setting stimuli randomization options, as its name indicates. For example in the example showed in Fig. 7 the randomization of both variables (CV1 and CV2) are chosen and CV1 (named scenes in this example) is set as the dominant variable. That is, in the experiment the scene (CV1) will be first randomly chosen, and then the images (CV2) will be presented in random order. By pressing the *Preview indices* button an order of content variables is showed on the MATLAB command window, from which one can check that the variable settings are correct.

The *Referencing* subpanel contains the method options in which one or two static reference stimuli or dynamic reference stimuli are shown. By selecting one static reference the standardized methods of DSIS or DSCR can be built. The

DSIS and DSCR methods show the reference and test stimuli and it is the observer's task to evaluate the test stimulus or to evaluate both reference and test stimuli. With the option of two static references, e.g. a setup with low and high quality reference stimuli can be built. In that case the observers see e.g. in the left window the low quality reference stimulus and in the right window the high quality one. The test stimuli are displayed in the center window (see Fig. 8). As the observers are seeing low- and high-quality reference stimuli, they form anchor points for the overall quality variation within a set of test stimuli.

In addition to building standardized experiment settings, modified experiment settings are also possible. Figure 7 shows settings, in which Dynamic reference (DR) is selected from the *referencing* subpanel. The DR-ACR is a hybrid of the ACR and PC method. The DR-ACR shows a slideshow of the stimuli with the corresponding scene prior to its rating (see Fig. 9). As the observers view the other stimuli in the slide show (as a reference), they form a general idea of the overall quality variation within the set of stimuli.

In *dynamic referencing options* subpanel there are check boxes to select masking image between references, randomize references and omit current image. In addition there is the text field to set times between references. The "Randomized references" check box randomizes the order of the reference stimuli. The "Leave current image out" check box shows the other stimuli from the scene as references and omits the evaluated stimulus from the slide show. Time between references defines the reference stimuli display time for the DR-ACR method. If the reference stimulus display time is too short observers may probably not notice differences in quality and types of distortions between the test stimuli. The default display time is set to be 1 second. However, according to a study by Nuutinen et al. (2014), a display time of 0.5 second can be a good compromise between test duration and accuracy. In the example showed in Fig. 9 the component t_{DR} is the display time of one reference stimulus in a set of reference stimuli.

The experiment building panel: QBU. When the desired options are selected from the experiment building panel the button *Save and Continue* opens the Question Building Unit (QBU). The QBU is the final stage of the setup wizard and it is a tool used for generating the response method in the experiment. The interface of the QBU is the same regardless of the selected rating method (Fig. 10, left).

The *options* panel of the QBU is used for choosing different forms of response methods, such as sliders, check boxes or open question forms. For example, the standard ACR rating method can be constructed using sliders. For the standard PC method the check boxes can be a viable selection. All elements of the question window can be edited. The sizes, locations, questions and labels of check boxes and sliders can be changed by using the property editor (*Edit properties*

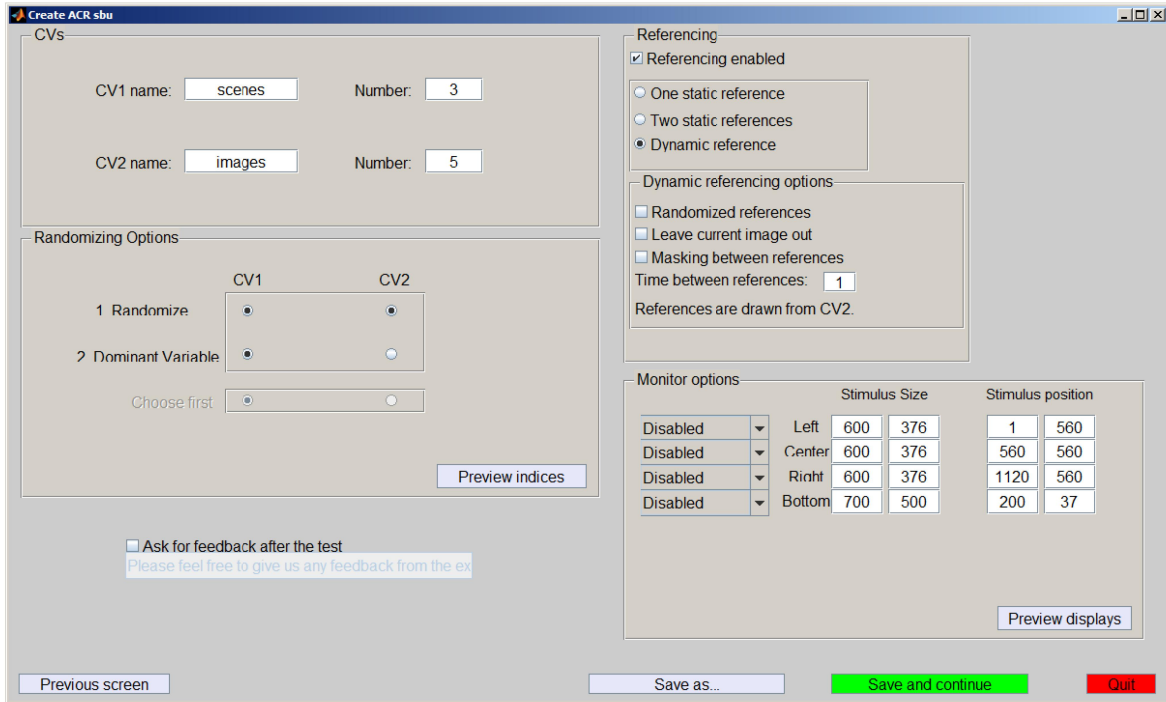


Figure 7. In the experiment builder panel of the ACR method referencing options such as one or two static reference stimuli or dynamic reference can be selected. In addition factors such as the number of images and contents, randomization of the stimuli and the stimuli window sizes and positions can be selected

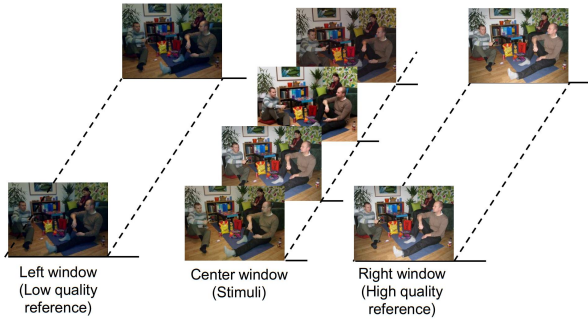


Figure 8. In the case of two static references the observers see e.g. in the left window the low quality reference stimulus and on the right window the high quality one. The test stimuli are displayed in the center window.

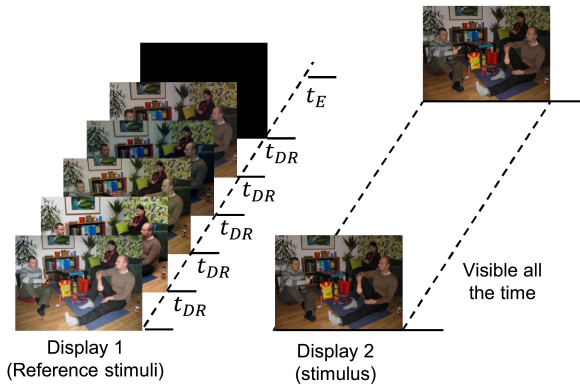


Figure 9. DR-ACR image rating method presents a reference image set in one display and a test image in the other.

button). One important property is that random starting positions for the pointers in the sliders can be enabled. Moreover, fine tuning can be done by choosing "property editor" from the top "View" menu. In that way everything from the fonts, colors and text fields can be edited. One can also add onscreen instructions for the observers by adding a text box from the top "Insert" menu.

The QBU enables the selection of up to 11 sliders, which can be individually edited. For the example presented in Fig.

10 (right) we selected two sliders, which we named "Image quality" and "Sharpness". The endpoints of the sliders are set to "min" and "max". In this example setting a subject response plot is also selected. It is a graph showing a record of rating scores, which helps observers to remember their previous answers and encourages them to use the whole scale. This reduces the variations due to individual tendencies to use the scale.

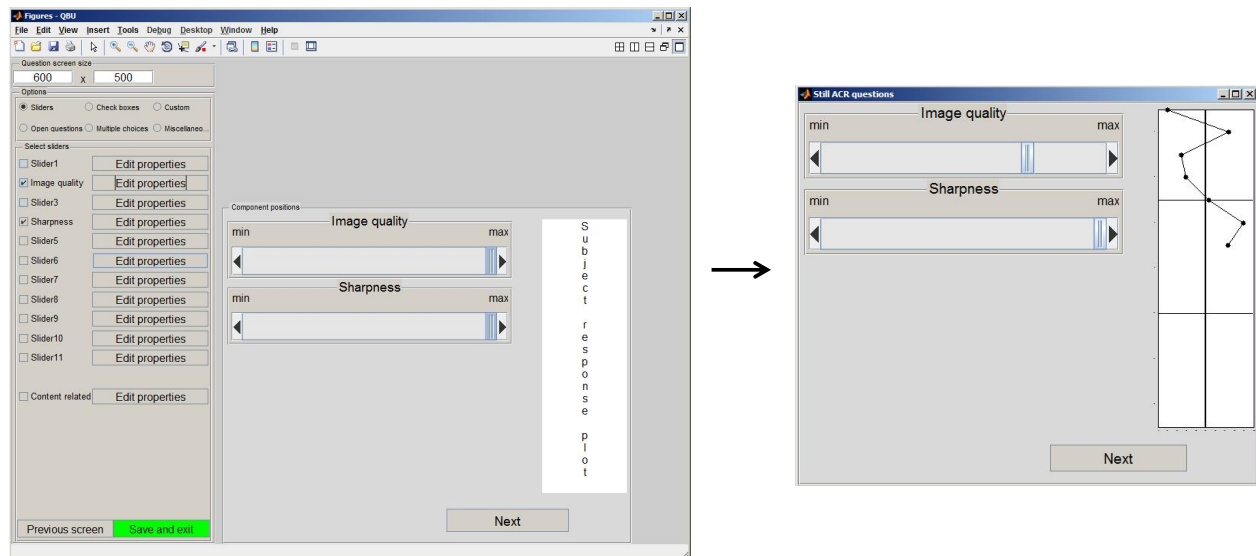


Figure 10. The question building unit (left) is used to generate the questions for the experiment (right)

The experiment building panel: Stimulus filename spreadsheet. Before a new experiment can be launched, the filename spreadsheet file should be edited. The stimulus filename spreadsheet file contains the file names of the experiment stimuli. Figure 11 shows an example structure of the filename spreadsheet for an experiment in which five images (CV2) from two scenes (CV1) are being evaluated. The filenames of the stimuli should be ordered first by CV1 number and then by CV2 number.

If the experiment uses fixed reference stimuli, their filenames should be written in the "Reference 1 (optional)" and "Reference 2 (optional)" columns. The "CV1 (optional)" column should contain the index of CV1 (scene number in this example) for which each reference stimulus is intended. In the example presented in Fig. 11, the image files "im1D.png" and "im1B.png" are used as too dark and too bright reference stimuli for the CV1 related question for scene 1. The image files "im2L.png" and "im2F.png" are used as too loud and too faded colors reference stimuli for scene 2. CV1 specific questions can be written into the "CV1 related questions (optional)" column. The CV1 related question should be selected and edited in the QBU (see the check box of "Content related" and *Edit properties* button in Fig. 10). In this example the sliders are named "Lightness" for scene 1 and "Color saturation" for scene 2.

Samples of typical quality rating test and comparison of two rating methods

VQone is the standard experiment building tool in our laboratory for subjective image and video quality ratings. The VQone software distribution also contains some sample data. These sample data include stimulus files, setup structure files

and result spreadsheet files for a still image quality rating experiment in which the performance of the standard PC and a new DR-ACR image rating methods was compared.

Materials and methods

The sample data consists of the data of 32 observers. Prior to experiments, the observers confirmed that they did not personally or professionally work with image processing or image-quality evaluation. All the observers had normal or corrected to normal vision. The observers were divided into two groups of 16 observers each. Different groups evaluated the test images using the DR-ACR or PC setup. The experiments were performed in a dark room with controlled lighting directed towards a wall behind the displays, which produced an ambient illumination of 20 lux to avoid flare. The setup included two colorimetrically calibrated 24" 1920 x 1200 displays (Eizo Color Edge CG210) for displaying one test image at a time and its reference images (DR-ACR method) or test image pairs (PC method). The third smaller display underneath the primary displays was used for presenting questions.

In the experiments, the observers evaluated the 45 test images which had 3 scenes (the numbers of CV1 and CV2 were 3 and 15). VQone presented test images (DR-ACR method) or test image pairs (PC method) in random order, one scene at a time, to each observer. For the DR-ACR method, the setup was built to show a slider measuring general quality of stimuli. The scale of the slider was analogue without visible discrete steps or numbers. Observers did not know how many steps were in the slider in order to avoid any tendency to favor certain numbers (e.g., even tens and quartiles). The resolution of the scale was 100 steps. That is, the ratings

	A	B	C	D	E	F	G	H	I
1	CV1	CV2	Filename	CV1 (optional)	Reference 1 (optional)	Reference 2 (optional)	CV1 related questions (optional)	Min (optional)	Max (optional)
2	1	1	im10.png	1	im1D.png	im1B.png	Lightness	Too dark	Too Bright
3	1	2	im11.png	2	im2L.png	im2F.png	Color saturation	Too loud colors	Too faded colors
4	1	3	im12.png						
5	1	4	im13.png						
6	1	5	im14.png						
7	2	1	im20.png						
8	2	2	im21.png						
9	2	3	im22.png						
10	2	4	im23.png						
11	2	5	im24.png						
12									
13									

Figure 11. VQone software uses a stimulus filenames spreadsheet to retrieve the filenames for the test stimuli and reference images. In addition optional content specific questions can be retrieved from the spreadsheet

were recorded from 0 to 100 using step sizes of 1. For the PC method, VQone showed two images side by side, and the observers' task was to select their preferred image. The images were selected using radio buttons. Figure 12 shows the screen shots from the stimulus and question windows of both experiments.

VQone offers simple data analyzing tools which can be used to check e.g. that the experiment setup functions as it should or when the number of observers is high enough. The options for data histograms, mean values and 95% confidence intervals and standard deviation values as a function of the number of observers or the mean rating scores are available. VQone calculates these values based on the information retrieved from the Setup structure file. It is recommended that more demanding statistics analyses for raw data should be done using statistics software packages. The GUI for separate data analyzing panel is presented in Fig. 13 in which standard deviation values as a function of the number of observers (n) are plotted for the data of the DR-ACR experiment. From the plot it can be seen that the standard deviation value saturates, when the number of observers is more than ten ($n > 10$), and the number of observers ($n = 16$) was high enough in that experiment.

Results

The main goal of the experiment distributed as sample data was to compare the performance of the DR-ACR and PC methods. The performance of the methods was measured in terms of speed and discriminability. The term discriminability refers to the ability of the method to identify statistically significant differences between the test images. The term speed refers to the time effort (as a function of the number of test images) needed to conduct the rating test.

The number of significant differences between image pairs was based on the linear mixed models (DR-ACR) and Chi-square tests (PC method). These metrics are presented in more detail elsewhere (Nuutinen et al., 2014). The maximum number of statistically significant differences for the 15 test images is calculated as $\frac{15 \cdot (15-1)}{2} = 105$. Table 1 shows the

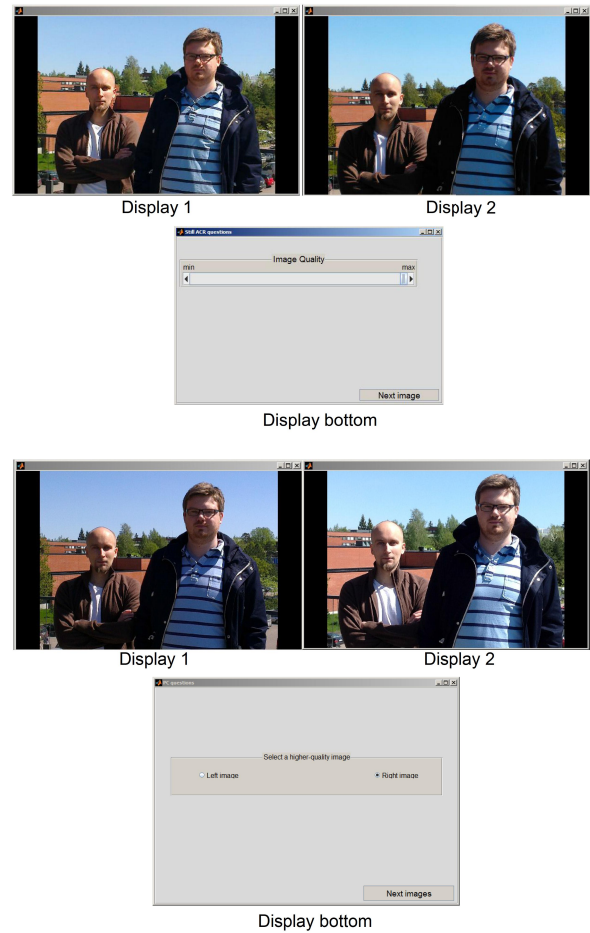


Figure 12. The DR-ACR experiment user interface (up) and the PC experiment user interface (bottom)

number of statistically significant image pairs for the DR-ACR and PC method. According to the results the accuracy of the DR-ACR was 65.1% (68.3/105 image pairs). On the other hand, the DR-ACR was faster, the test duration was only 54% (16.3/30.1 minutes, $3 \cdot 15$ test images) of the duration of the PC experiment.

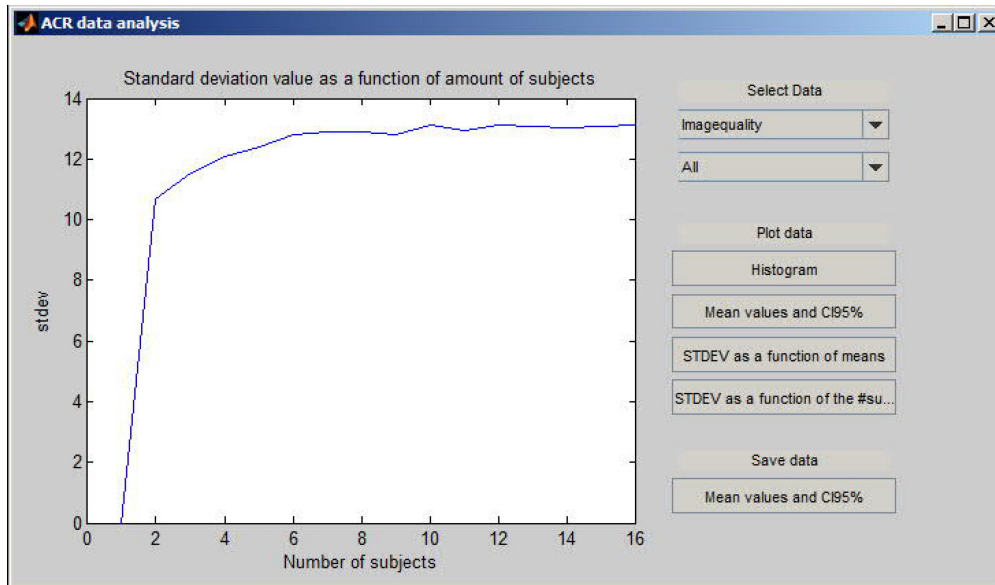


Figure 13. The data analyze panel can present data histograms, mean values with 95% confidence intervals and standard deviation values as a function of the number of observers or mean values. Data can be visualized over all contents or in a content specific manner

Table 1

The number of statistically significant image pairs for the DR-ACR and PC method

Content	1	2	3	Average
DR-ACR	75	59	71	68.3
PC	91	97	97	95

Comparison with existing software

Table 2 provides a list of the software that can be used for creating image and video quality rating experiments. There are already many software packages that can be used for creating experiments in which images are displayed and observer responses are recorded, yet the functionality offered by VQone is unique. In this section we will focus on software packages with which VQone has the most in common: Image and Video Quality Evaluation Software (IVQUEST) and MSU Perceptual Video Quality Tool. These software packages, such as VQone, are developed primarily for image quality evaluations.

The selection and number of rating methods is two of the most important performance criteria for image rating experiment builders. Table 3 lists the rating methods offered by IVQUEST, MSU and VQone. It should be noted that rating methods built by VQone can be modified forward and new rating methods can be constructed. That is, after the base method (ACR, PC, triplet etc.) is selected from the main

panel (see Fig. 3), the selected method can be modified to fulfill the needs of the application or research question.

The stimuli of VQone and IVQUEST can be still images or video. MSU can only be used with video stimuli. IVQUEST offers Single Stimulus Impairment (SS) and Single Stimulus Continuous Quality Evaluation (SSCQE) methods. The SS method is equivalent to the ACR method. The SS method is standardized in (ITU-R BT.500, 2012) and the ACR in (ITU-T P.910, 2008). MSU offers five rating methods of which four are standardized: DSIS (Double Stimulus Impairment Scale), DSCQS (Double Stimulus Continuous Quality Scale), SCACJ (Stimulus Comparison Adjectival Categorical Judgement) and SAMVIQ (Subjective Assessment Method for Video Quality rating). The fifth rating method is named MSUCQE (MSU Continuous Quality Evaluation). The MSUCQE method shows two video sequences simultaneously. The task of the observer is to point the position where a sequence starts to get worse if one sequence is worse than the other.

As Table 3 shows the VQone provides both methods (SS and SSCQE) that are offered by IVQUEST. The MSU offers two methods that VQone does not offer, namely the MSUCQE and the standardized SAMVIQ (ITU-R BT.1788, 2007). The SAMVIQ method gives access to several samples of a video sequence. The observer selects different video samples randomly. The observer can modify the score of each sample of a video sequence as desired.

A superior feature of VQone compared with the other packages is its flexibility to create and modify windows in which stimuli are presented. MSU and IVQUEST display

Table 3

Rating methods offered by IVQUEST, MSU and VQone image rating software

Method	IVQUEST	MSU*	VQone
ACR / SS	x	-	x
SSCQE	x	-	x
DSCQS	-	x	x
DSIS	-	x	x
SCACJ	-	x	x
SAMVIQ	-	x	-
MSUCQE	-	x	-
ACR-DR	-	-	x
PC	-	-	x
Triplet	-	-	x
Questions only	-	-	x

*Only video stimuli

stimuli always in the fixed size window positioned at the center of the primary display. VQone enables stimuli to be freely displayed on 1 - 3 windows, e.g., so that one stimulus window is displayed on one display and other is displayed on the other display and so on. The fourth display or window is reserved for the question panel. The positions and sizes of the windows can freely be set via the experiment building process.

VQone enables different free-named rating scales and free-ended questions (e.g. text input fields) to be formed according to the needs of different applications and research questions. Moreover, it is also possible to create content specific questions, which may be change when the next content is evaluated. IVQUEST and MSU always display the same rating questions to observers.

Conclusion

VQone is currently in use in research laboratories at the University of Helsinki. It has proven to be useful in experimental psychological experiments of image quality, image understanding and decision making. The easy handling of all experiment building, launching and analyzing settings through the GUIs provide the researcher, who is most comfortable in a graphical environment, with many useful tools for forming complex image and video rating experiments.

VQone is being actively maintained and developed. New functionalities, such as the SAMVIQ rating method, implementations of faster paired comparison methods (e.g., Silverstein & Farrell, 2001) and possibility to show higher frame-rate videos than standard frame-rates are expected in future versions. A problem with the standard frame-rates can be that they do not render natural motion. Higher frame-rate (e.g. 120 fps) would be highly useful for many psychophys-

ical studies in which naturalness and smoothness of motion are required in stimulus presentation. An example implementation of high frame-rate playback using Psychophysics Toolbox can be found from Lidestam (2014). In addition support for other operating systems and better handling of audio stimuli will be considered and added. Furthermore, researchers are welcome to add functionality to the code themselves. We hope the research community finds this toolbox as useful a tool for their laboratories as we have.

References

- Bovik, A. (2013). Automatic prediction of perceptual image and video quality. *Proceedings of the IEEE*.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, 10(4), 433–436.
- Chandler, D. (2013). Seven challenges in image quality assessment: Past, present, and future research. *ISRN Signal Processing*, 2013, 905685.
- Coco, M., & Keller, F. (2014). Classification of visual and linguistic tasks using eye-movement features. *Journal of Vision*, 14(3), 11.
- Cohen, J. D., MacWhinney, B., Flatt, M. R., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments & Computers*, 25(2), 257–271.
- Geller, A., Schleifer, I., Sederberg, P., Jacobs, J., & Kahana, M. (2007). PyEPL: A cross-platform experiment-programming library. *Behavior Research Methods*, 39(4), 950–958.
- ISO 20462-1. *Photography - psychophysical experimental methods for estimating image quality - part 1: Overview of psychophysical elements* (Norm No. ISO 20462-1). (2005). ISO, Geneva, Switzerland (a).
- ISO 20462-2. *Photography - psychophysical experimental methods for estimating image quality - part 2: Triplet comparison method* (Norm No. ISO 20462-2). (2005). ISO, Geneva, Switzerland (b).
- ITU-R BT.500. *Methodology for the subjective assessment of the quality of television pictures* (Norm No. ITU-R Recommendation BT.500-13). (2012). ITU, Geneva, Switzerland.
- ITU-R Rec. BT.1788. *Methodology for the subjective assessment of video quality in multimedia applications* (Norm No. ITU-R Recommendation PBT.1788). (2007). ITU, Geneva, Switzerland.
- ITU-T Rec. P. 910. *Subjective video quality assessment methods for multimedia applications* (Norm No. ITU-R Recommendation P.910). (2008). ITU, Geneva, Switzerland.
- Leisti, T., Radun, J., Virtanen, T., Nyman, G., & Häkkinen, J. (2014). Concurrent explanations can enhance visual decision making. *Acta Psychologica*, 145(0), 65–74.
- Lidestam, B. (2014). Audiovisual presentation of video-recorded stimuli at a high frame rate. *Behavior Research Methods*, 46(2), 499–516.
- Mantiuk, R., Tomaszewska, A., & Mantiuk, R. (2012, Dec.). Comparison of four subjective methods for image quality assessment. *Comp. Graph. Forum*, 31(8), 2478–2491.
- Mathot, S., Schreij, D., & Theeuwes, J. (2012). Opensesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, 44(2), 314–324.

Table 2

An overview of software for creating image and video evaluation experiments

Name	GUI	Free	Scripting	Platform	Reference
OpenSesame	Yes	Yes	Python	Win, Mac, Lin	(Mathot, Schreij, & Theeuwes, 2012)
MATLAB Psychophysics Tool-box	No	Yes	MATLAB	Win, Mac, Lin	(Brainard, 1997)
PsychoPy	Yes	Yes	Python	Win, Mac, Lin	(Peirce, 2007)
PsyScope	Yes	Yes		Win, Mac, Lin	(Cohen, MacWhinney, Flatt, & Provost, 1993)
PyEPL	No	Yes	Python	Mac, Lin	(Geller, Schleifer, Sederberg, Jacobs, & Kahana, 2007)
IVQUEST	Yes	Yes	MATLAB	Win	(Murthy & Karam, 2010)
MSU Perceptual Video Quality Tool	Yes	Yes*		Win	(Vatolin & Petrov, n.d.)
VQone	Yes	Yes	MATLAB	Win	

*Source-code is not available

- Murthy, A., & Karam, L. (2010, June). A matlab-based framework for image and video quality evaluation. In *Quality of Multimedia Experience (QoMEX), 2010 Second International Workshop on* (pp. 242–247).
- Nuutinen, M., Virtanen, T., Leisti, T., Mustonen, T., Radun, J., & Häkkinen, J. (2014). A new method for evaluating the subjective image quality of photographs: Dynamic Reference. *Multimedia Tools and Applications*. (To appear)
- Nyman, G., Radun, J., Leisti, T., Oja, J., Ojanen, H., Olives, J.-L., ... Häkkinen, J. (2006, January). *What do users really perceive: probing the subjective image quality* (Vol. 6059). San Jose, CA.
- Peirce, J. (2007). PsychoPy - Psychophysics software in Python. *Journal of Neuroscience Methods*, 162(1–2), 8–13.
- Radun, J., Leisti, T., Virtanen, T., Häkkinen, J., Vuori, T., & Nyman, G. (n.d.). Evaluating the Multivariate Visual Quality Performance of Image-processing Components. *ACM Trans. Appl. Percept.*, 7(3).
- Rensink, R. (2014). Limits to the usability of iconic memory. *Frontiers in Psychology*, 5, 971.
- Silverstein, D., & Farrell, J. (2001). Efficient method for paired comparison. *Journal of Electronic Imaging*, 10(2), 394–398.
- Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied*, 74(11), 1–29.
- To, M., Gilchrist, I., Troscianko, T., & Tolhurst, D. (2011). Discrimination of natural scenes in central and peripheral vision.
- Vatolin, D., & Petrov, O. (n.d.). *Msu perceptual video quality tool*. Retrieved from http://compression.ru/video/quality_measure/perceptual_video_quality_tool_en.html
- Virtanen, T., Nuutinen, M., Vaahteranoksa, M., Oittinen, P., & Häkkinen, J. (2014). CID2013: a database for evaluating no-reference image quality assessment algorithms. *IEEE Transactions on Image Processing*. (To appear)