

Service : DIH/EQS		External diffusion
Reference : FT.BD.FTR&D/DIH/EQS/462/02/JLB		Date : 07/10/2002

**SEOVQ software tool for
quality, preference and acceptability evaluation
of multimedia images**

*Present document contains information proprietary to France Télécom R&D.
Its content cannot be revealed, duplicated by any way without authorization.*

	Name	Service	Date	Visa
Author(s)	J-L BLIN	DIH/EQS	07/10/02	

DIFFUSION LIST					
Function	Service	Name	Nb of copy	For	
				Action	Info
Laboratory manager	DIH/EQS	T. ALPERT	1		
URD manager	DIH/EQS	A. CROUZET	1		

TABLE OF CONTENTS

- 1. INTRODUCTION..... 4**
- 2. USEFUL FUNCTIONS OF THE SEOVQ SOFTWARE TOOL..... 4**
- 3. PRESENTATION OF THE GENERAL INTERFACE OF SEOVQ..... 5**
- 4. PREPARATION INTERFACE OF A TEST SESSION..... 6**
- 5. QUALITY EVALUATION INTERFACE..... 10**
- 6. EVALUATION INTERFACE OF PREFERENCE BY COMPARISON..... 12**
- 7. EVALUATION INTERFACE OF ACCEPTABILITY..... 14**
- 8. ANALYSE INTERFACE OF TEST RESULTS..... 16**
 - 8.1. THE ASSEMBLING OF THE OBSERVER VOTES 16
 - 8.2. THE PRESENTATION OF RESULTS IN QUALITY EVALUATION MODE..... 16
 - 8.3. THE PRESENTATION OF RESULTS IN THE COMPARISON OR ACCEPTABILITY MODES 19

1. INTRODUCTION

Due to the multiplicity of audio-visual transmission projects via IP, everyone wonders about the present real perceived quality according to chosen transmission bitrate. All these multimedia images have near "MPEG-4" functionalities. Unlike the MPEG-2 standard for television images, updating frequency of image data, format (number of lines, number of columns) and size are not fixed.

It becomes necessary to have a software tool at one's disposal that is versatile to different audiovisual contexts of services, in order to be able to evaluate the quality, find preferences, indeed the determination of a level of acceptability.

2. USEFUL FUNCTIONS OF THE SEOVQ SOFTWARE TOOL

The SEOVQ software (Subjective Evaluation and Optimization of Video Quality) operates on a Windows 2000 compatible PC and can play audiovisual sequences from files using proprietary formats or not. This tool is able to indifferently manage and associate the following functionalities:

- ☞ Type(s) of decoders(s)
- ☞ Image size
- ☞ Image format
- ☞ Extraction of images from the proprietary skin
- ☞ Contextual parameters
- ☞ Delay simulation.
- ☞ Programming sequence display duration.
- ☞ 3 Image zoom in presetting.

It is imperative to *associate different MPEG-4 decoders* or proprietary compatible. It deals with being able to compare the performances of the codecs within the same test. The codecs are recognized by their suffix attached to the files in order to play the current sequences by the right decoder-player. This suffix acts as a switch to the correct decoder

All the *image formats* can be displayed within the limits of the available screen format.

The *size of the images* is adjustable by changing either the size of the screen or the screen display format.

The images are *separated from the proprietary skin* to get an anonymous display in order to evaluate the reproduced quality performances of the players without being influenced by the knowledge of an environment.

It is possible to add *contextual parameters* such as the cost, the way you see the images in a mobile or nomadic use.

It is possible to add a 1 second step by step *delay* to simulate downloading times of scenes in order to evaluate the acceptability. The value by default of the delay is set to 0.

The *duration of the scenes* can be adjusted from 3 seconds to 3 minutes. A 3 second duration is considered as the limit under which it is not possible to appreciate quality with reliability. 3 minutes is the maximum duration for which the global note of quality remains significant.

The software, controls the *zoom* values of the decoders. Default value is 1, but it is possible to preset the zoom value to 2 or 4. This allows to indifferently evaluate the perceived quality of the images, using several zoom in values during a test.

3. PRESENTATION OF THE GENERAL INTERFACE OF SEOVQ

General interface (cf. figure 1) includes 3 main functionalities:

- ☞ 1 Prepare and organise a test session
- ☞ 2 Presentation of the images being evaluated
- ☞ 3 Analysis of test results

Functionalities 1 and 3 are only accessible by the operator who runs the test via a required password, which password is typed under the specific interface "Change operator password". The observers may directly start a session presentation (field 2). They are allowed to quit the session only after having watched each sequence at least once and scored them.

Field 3 brings together, scene by scene and the overall scenes, all votes provided by all the sessions. A test can be made of several sessions.

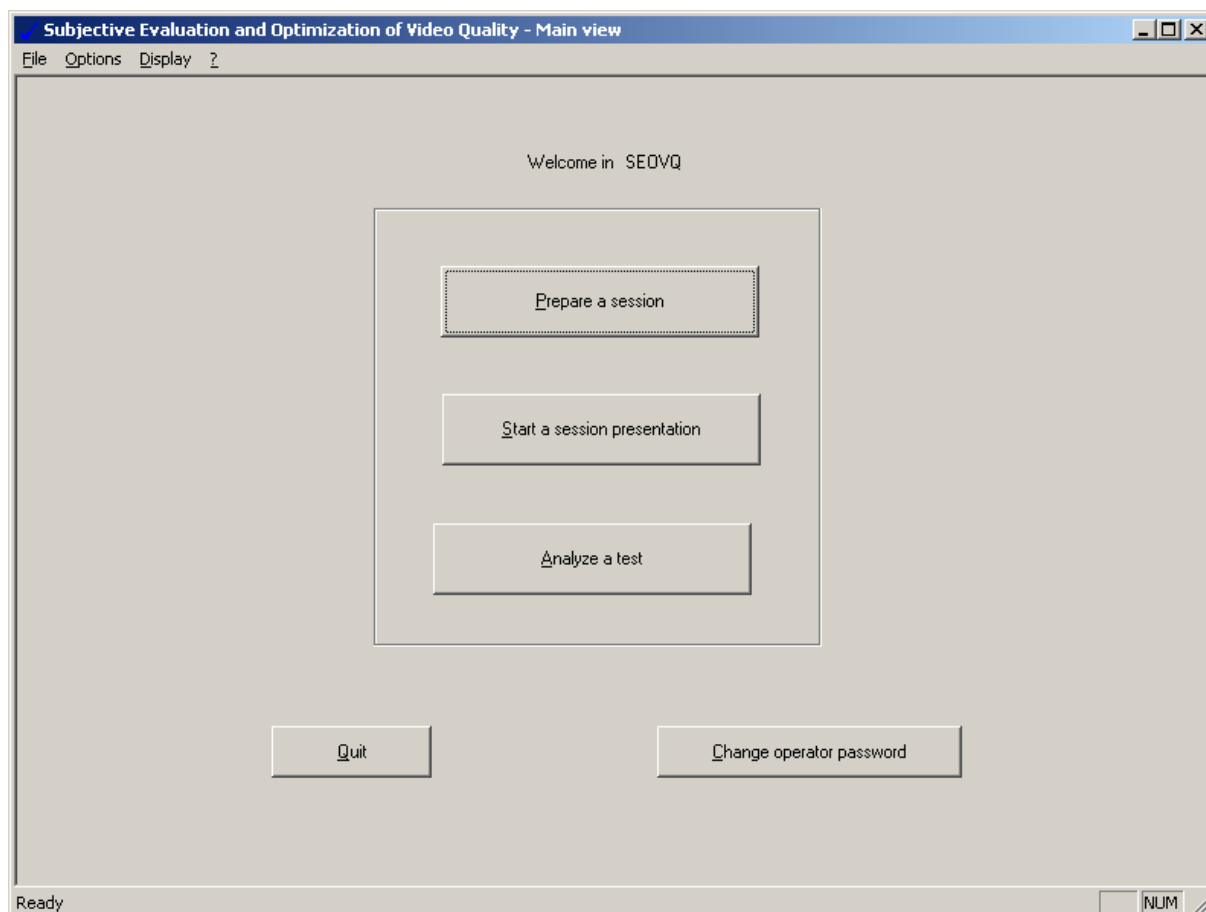


Figure 1 : Main interface of SEOVQ

The window below (figure 2) is a software control to fit the current format of the working screen. The default format is 1024*768.

The working directory can be modified and the default directory is "C:\SEOVQ".

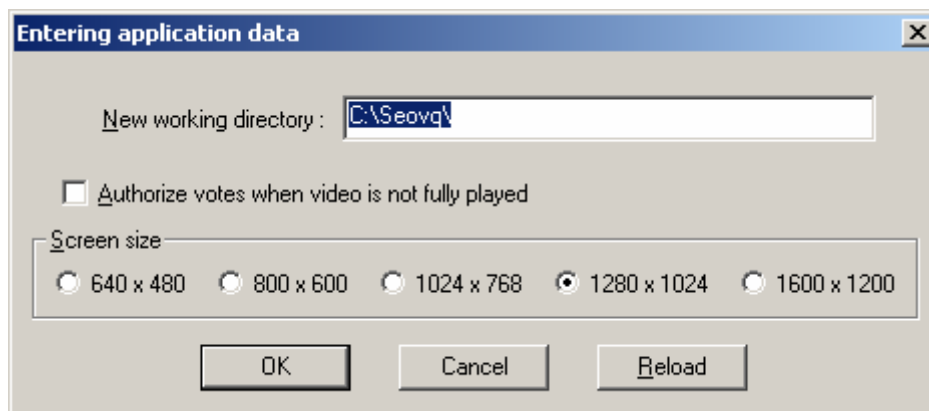


Figure 2: Field Option, "Entering application data"

The interrogation point gives you access to the "About" window that show the version number of SEOVQ. (cf figure 3).



Figure 3 : SEOVQ version

4. PREPARATION INTERFACE OF A TEST SESSION

The first preparation interface is shown in figure 4. The quality evaluation, the comparison (preference) and the acceptability are the 3 proposed test methods. Once a type of test has been chosen, a test name is given, then the session number. A test can be made up of several sessions. Each scene is named because a test is carried out scene by scene. In the example below, there are 4 scenes and the duration of a sequence display is limited to 10 seconds. The number of algorithms corresponds to the number of combined parameters for each scene, which are equally a type of codec, a compressed bit-rate, a zoom, an image format, a programmable delay and up to 2 contextual parameters.

Each file describing all the parameters to play a test is recognized with its suffix :

- ☞ Quality evaluation : « file_name_test » + « .sdte »
- ☞ Comparison, preference : « file_name_test » + « .sdtc »
- ☞ Acceptability : « file_name_test » + « .sdta »

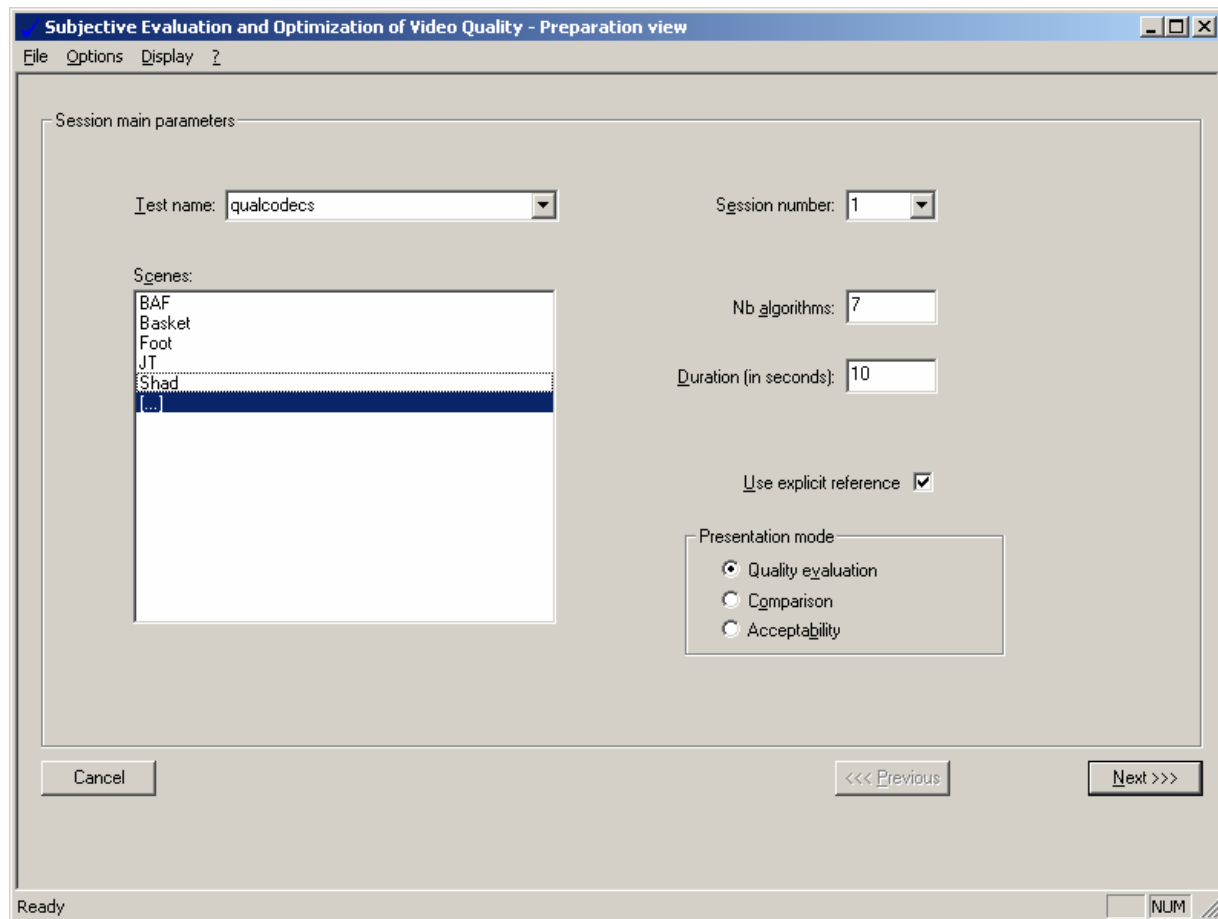


Figure 4 : First window of a test session preparation

Figures 5 and 5bis represent the second part of the preparation interface. These two figures are just a unique window, which is not big enough to show all the programming fields. The below example describes the main features of a quality evaluation test as follows:

- ☞ Session: 1
- ☞ Number of scenes: 5
- ☞ Number and rank of the algorithms: 7 + reference

The 5 first fields are automatically processed from the previous information described in the first part of the preparation interface. Field number 3 indicates the number of algorithms that systematically start out from 1 to n (here from 1 to 5), whereas number 0, when it is required, is exclusively reserved for the explicit reference. The name of each scene is written in field number 4. In field number 5 the rank (A, B, C, ...) is registered from the button which will serve as an access to the corresponding sequence. This rank is randomised from one scene to another in order to prevent the observers from detecting a ranking of algorithms. Then, the observers don't get the ranking from one scene to another. The rank R is reserved for the explicit reference and is not affected by the randomisation of the presentation. The field number 5 provides the right way to find the current sequence. Zoom value of 2 or 4 can be inserted in field number 6. The default value of the zoom is 1. The delay in seconds between the play button action and visualization of the sequence can be inserted in field number 7. The default value of the delay is 0. Finally, the last 2 added fields are contextual parameters likely to bring about a change of the observer quality scores. These 2 fields are empty by default.

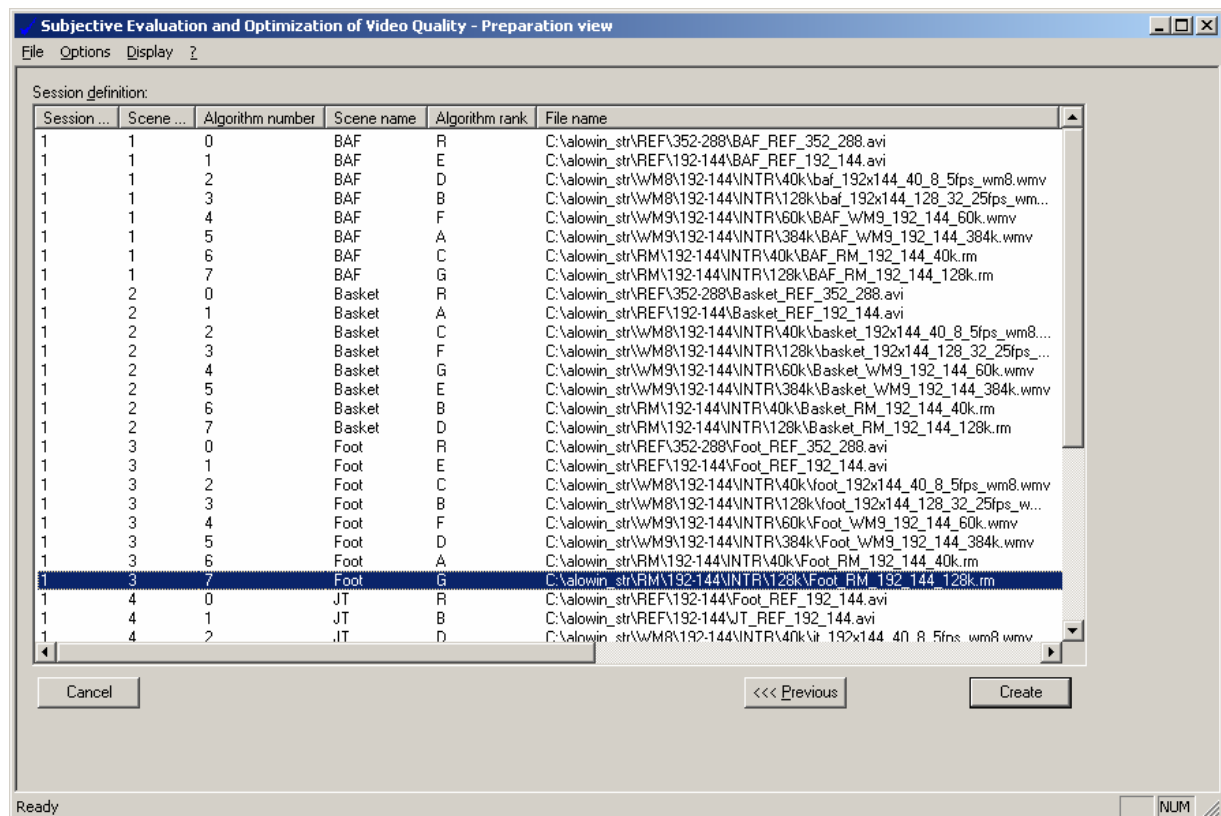


Figure 5 : Test session organisation

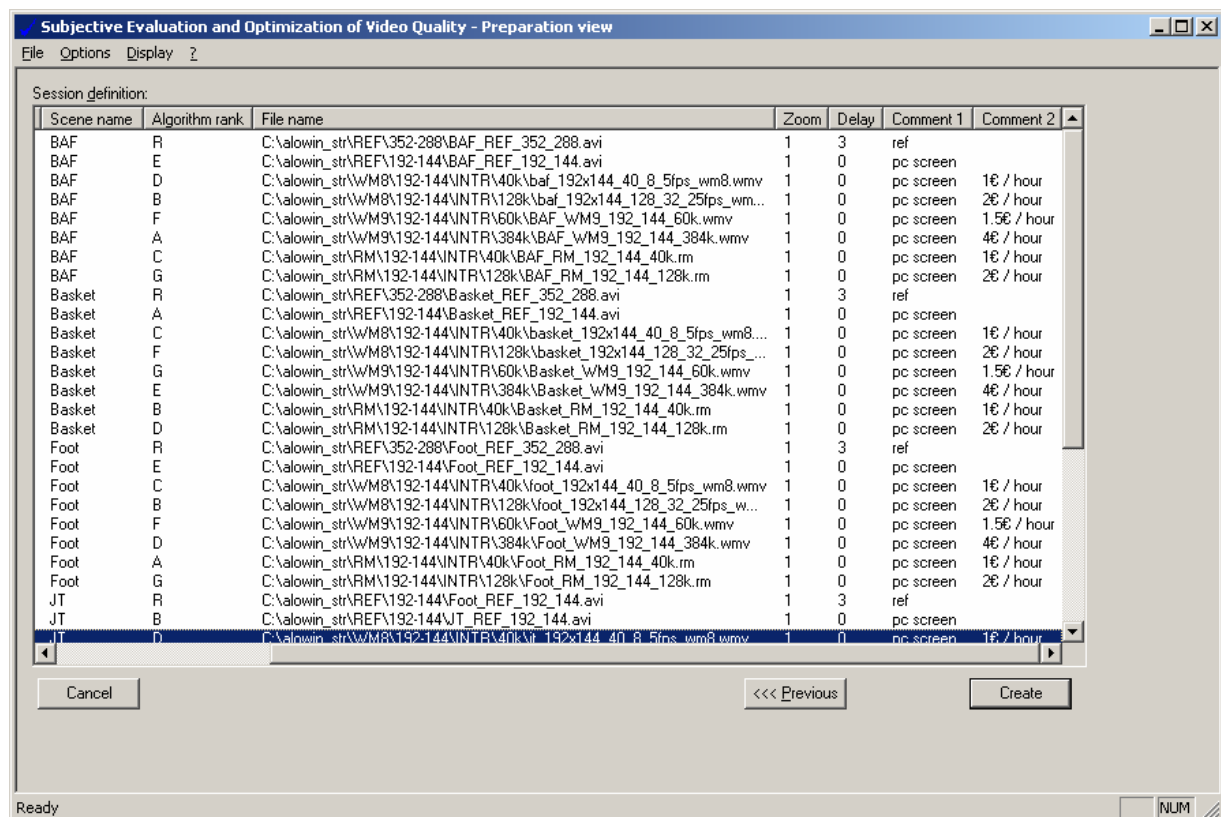


Figure 5 bis : Test session organisation

The button "Create" (cf. figure 5) saves a file in text format with all the parameters of the test session, as in the following example:

```
[DEFINITION]
DURATION=10
NB_ALGO=7
USE_REF=Y
DATE=23/07/2002
SCENES=BAF; Basket; Foot; JT;Shad;
[SEQUENCES]
1_0=C:\ALOWIN_STR\REF\352-288\BAF_REF_352_288.avi;1;R;0;;;
1_1=C:\ALOWIN_STR\REF\192-144\BAF_REF_192_144.avi;1;G;0;;;
1_2=C:\ALOWIN_STR\WM8\192-144\INTR\40k\BAF_192x144_40_8_5fps_wm8.wmv;1;C;0;;;
1_3=C:\ALOWIN_STR\WM8\192-144\INTR\128k\BAF_192x144_128_32_25fps_wm8.wmv;1;E;0;;;
1_4=C:\ALOWIN_STR\WM9\192-144\INTR\60k\BAF_WM9_192_144_60k.wmv;1;D;0;;;
1_5=C:\ALOWIN_STR\WM9\192-144\INTR\384k\BAF_WM9_192_144_384k.wmv;1;F;0;;;
1_6=C:\ALOWIN_STR\RM\192-144\INTR\40k\BAF_RM_192_144_40k.rm;1;B;0;;;
1_7=C:\ALOWIN_STR\RM\192-144\INTR\128k\BAF_RM_192_144_128k.rm;1;A;0;;;
2_0=C:\ALOWIN_STR\REF\352-288\Basket_REF_352_288.avi;1;R;0;;;
2_1=C:\ALOWIN_STR\REF\192-144\Basket_REF_192_144.avi;1;E;0;;;
2_2=C:\ALOWIN_STR\WM8\192-144\INTR\40k\Basket_192x144_40_8_5fps_wm8.wmv;1;F;0;;;
2_3=C:\ALOWIN_STR\WM8\192-144\INTR\128k\Basket_192x144_128_32_25fps_wm8.wmv;1;B;0;;;
2_4=C:\ALOWIN_STR\WM9\192-144\INTR\60k\Basket_WM9_192_144_60k.wmv;1;A;0;;;
2_5=C:\ALOWIN_STR\WM9\192-144\INTR\384k\Basket_WM9_192_144_384k.wmv;1;C;0;;;
2_6=C:\ALOWIN_STR\RM\192-144\INTR\40k\Basket_RM_192_144_40k.rm;1;G;0;;;
2_7=C:\ALOWIN_STR\RM\192-144\INTR\128k\Basket_RM_192_144_128k.rm;1;D;0;;;
3_0=C:\ALOWIN_STR\REF\352-288\Foot_REF_352_288.avi;1;R;0;;;
3_1=C:\ALOWIN_STR\REF\192-144\Foot_REF_192_144.avi;1;F;0;;;
3_2=C:\ALOWIN_STR\WM8\192-144\INTR\40k\Foot_192x144_40_8_5fps_wm8.wmv;1;G;0;;;
3_3=C:\ALOWIN_STR\WM8\192-144\INTR\128k\Foot_192x144_128_32_25fps_wm8.wmv;1;A;0;;;
3_4=C:\ALOWIN_STR\WM9\192-144\INTR\60k\Foot_WM9_192_144_60k.wmv;1;B;0;;;
3_5=C:\ALOWIN_STR\WM9\192-144\INTR\384k\Foot_WM9_192_144_384k.wmv;1;C;0;;;
3_6=C:\ALOWIN_STR\RM\192-144\INTR\40k\Foot_RM_192_144_40k.rm;1;D;0;;;
3_7=C:\ALOWIN_STR\RM\192-144\INTR\128k\Foot_RM_192_144_128k.rm;1;E;0;;;
4_0=C:\ALOWIN_STR\REF\352-288\JT_REF_352_288.avi;1;R;0;;;
4_1=C:\ALOWIN_STR\REF\192-144\JT_REF_192_144.avi;1;A;0;;;
4_2=C:\ALOWIN_STR\WM8\192-144\INTR\40k\JT_192x144_40_8_5fps_wm8.wmv;1;G;0;;;
4_3=C:\ALOWIN_STR\WM8\192-144\INTR\128k\JT_192x144_128_32_25fps_wm8.wmv;1;F;0;;;
4_4=C:\ALOWIN_STR\WM9\192-144\INTR\60k\JT_WM9_192_144_60k.wmv;1;E;0;;;
4_5=C:\ALOWIN_STR\WM9\192-144\INTR\384k\JT_WM9_192_144_384k.wmv;1;C;0;;;
4_6=C:\ALOWIN_STR\RM\192-144\INTR\40k\JT_RM_192_144_40k.rm;1;D;0;;;
4_7=C:\ALOWIN_STR\RM\192-144\INTR\128k\JT_RM_192_144_128k.rm;1;B;0;;;
5_0=C:\ALOWIN_STR\REF\352-288\Shad_REF_352_288.avi;1;R;0;;;
5_1=C:\ALOWIN_STR\REF\192-144\Shadocks_REF_192_144.avi;1;C;0;;;
5_2=C:\ALOWIN_STR\WM8\192-144\INTR\40k\Shadocks_192x144_40_8_5fps_wm8.wmv;1;B;0;;;
5_3=C:\ALOWIN_STR\WM8\192-144\INTR\128k\Shadocks_192x144_128_32_25fps_wm8.wmv;1;F;0;;;
5_4=C:\ALOWIN_STR\WM9\192-144\INTR\60k\Shad_WM9_192_144_60k.wmv;1;D;0;;;
5_5=C:\ALOWIN_STR\WM9\192-144\INTR\384k\Shad_WM9_192_144_384k.wmv;1;A;0;;;
5_6=C:\ALOWIN_STR\RM\192-144\INTR\40k\Shad_RM_192_144_40k.rm;1;G;0;;;
5_7=C:\ALOWIN_STR\RM\192-144\INTR\128k\Shad_RM_192_144_128k.rm;1;E;0;;;
```

5. QUALITY EVALUATION INTERFACE

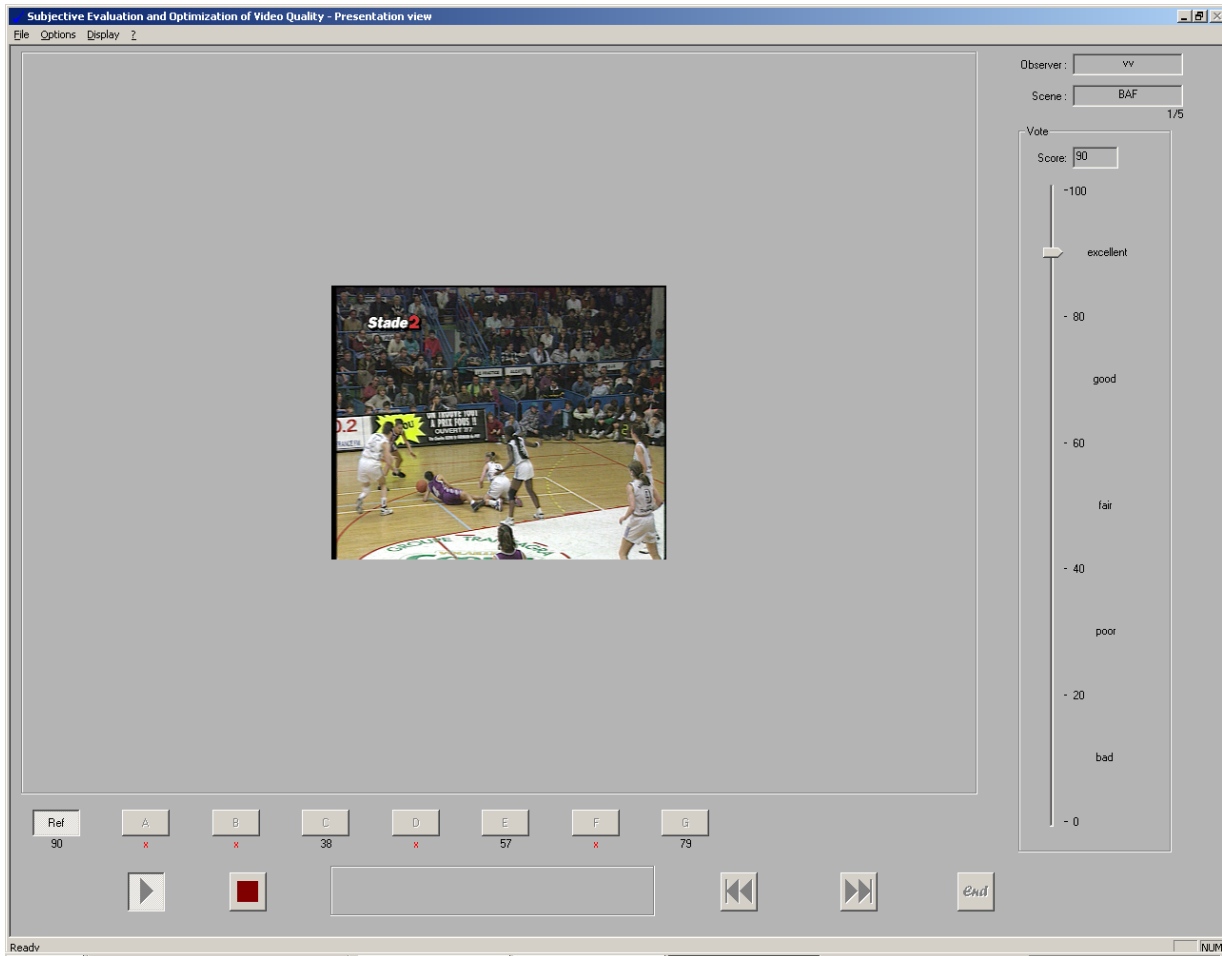


Figure 6 : Quality evaluation interface

On the screen above, the observer selects one of the sequences to assess, with the following buttons :



Starts playing the sequence.

At the end of the sequence, the observer evaluates the quality, by positioning the cursor of the slider, located on the vertical right-hand area of the interface. The score is then written under the button of the selected sequence.



Stops the sequence at any time.



Skips to the following scene. (active only if all the sequences from a given scene have been scored).



Returns to the previous scene. (active once the first scene has been scored).

From one scene to another a randomised presentation prevents the observers from attempting to vote in an identical way according to an established order.



Closes the session and saves all the observers scores in order to subsequently analyse them. This button is only active when all the sequences in the test have been scored.

6. EVALUATION INTERFACE OF PREFERENCE BY COMPARISON.

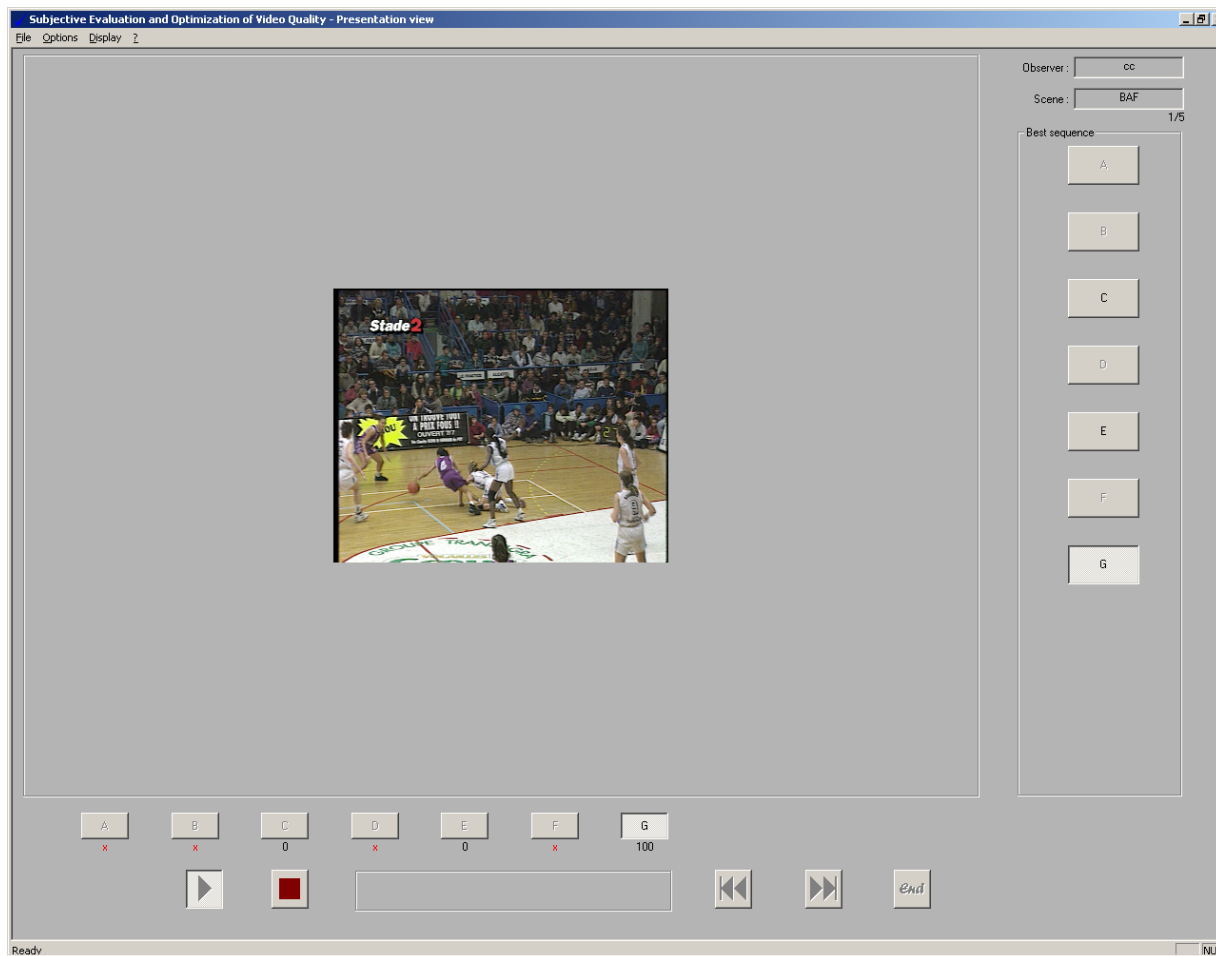


Figure 7 : Evaluation interface of preference by comparison

On the screen above, the observer selects one of the sequences to assess, with the following buttons :







-  Starts playing the sequence.
-  Stops the sequence at any time.
-  Skips to the following scene. (active only if all the sequences have been visualized and a preference has been established with one of the seven corresponding buttons : see fig. 8).
-  Return to the previous scene. (active once a preference has been established for the first scene : see fig. 8).



Figure 8 : Preference buttons

The observer must choose one and only one proposition between n (sequence) based on a preference criterion. He has the possibility to modify his choice at any point. Clicking on the button corresponding to the current sequence, selects this sequence as the preferred one that is scored to 100. The sole sequences previously visualized are scored to 0, otherwise a red cross remains under the corresponding buttons.

From one scene to another a randomised presentation prevents the observers from attempting to vote in an identical way according to an established order.



Closes the session and saves all the observers choices in order to subsequently analyse them. This button is only active when all the sequences in the test have been tested.

7. EVALUATION INTERFACE OF ACCEPTABILITY

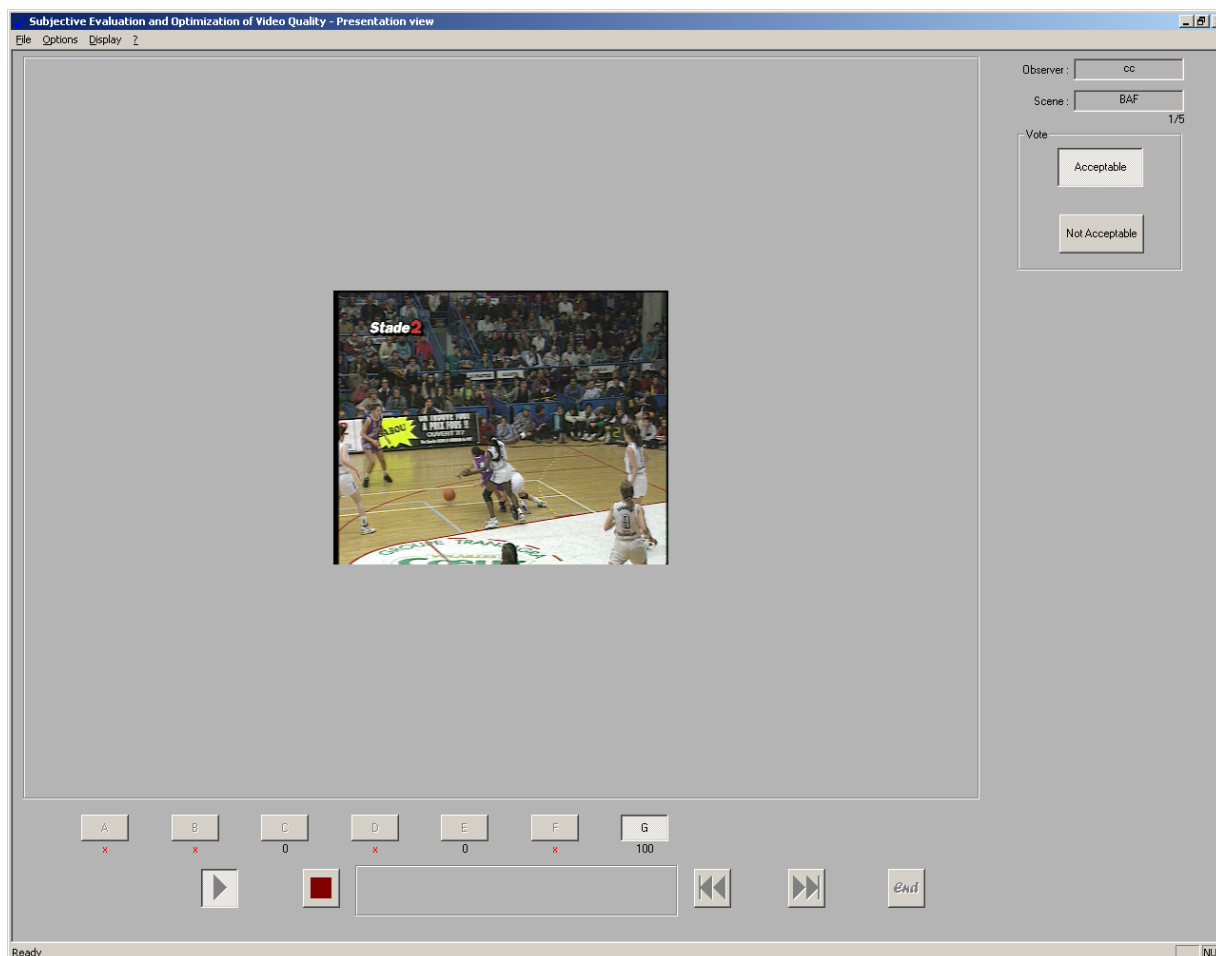


Figure 9 : Evaluation interface of acceptability

On the screen above, the observer selects one of the sequences to assess, with the following buttons :



Starts playing the sequence.



Stops the sequence at any time.



Allows to decide if the current sequence is "acceptable" or "non acceptable". The decision (acceptable =100, non acceptable = 0) is then written underneath the button of the selected sequence.



Skips to the following scene. (active only if all the sequences have been visualized and a "acceptable" or "non acceptable" choice has been established).



Return to the previous scene. (active once a "acceptable" or "non acceptable" choice has been established for the first scene).

From one scene to another a randomised presentation prevents the observer from attempting to vote in an identical manner according to an established order.



Closes the session and saves all the observers choices in order to subsequently analyse them. This button is only active when all the sequences in the test have been tested.

PS : Here, 2 contextual parameters may be added, this is likely to change the observer behaviour. For example, a cost, the way you see the images in a mobile or nomadic use are possible parameters.

8. ANALYSE INTERFACE OF TEST RESULTS

8.1. The assembling of the observer votes

After having chosen to analyse a test by selecting a test name, all the observer names are displayed as indicated in figure 10. It is possible to see the observer notes by double-clicking on his name. It is also possible to select only one part of the observers and to obtain partial results by clicking on the "next" button that produces the display of results scene by scene and on the average of all the scenes (cf. paragraphs 8.2 and 8.3). The display remains identical whichever type of test it is.

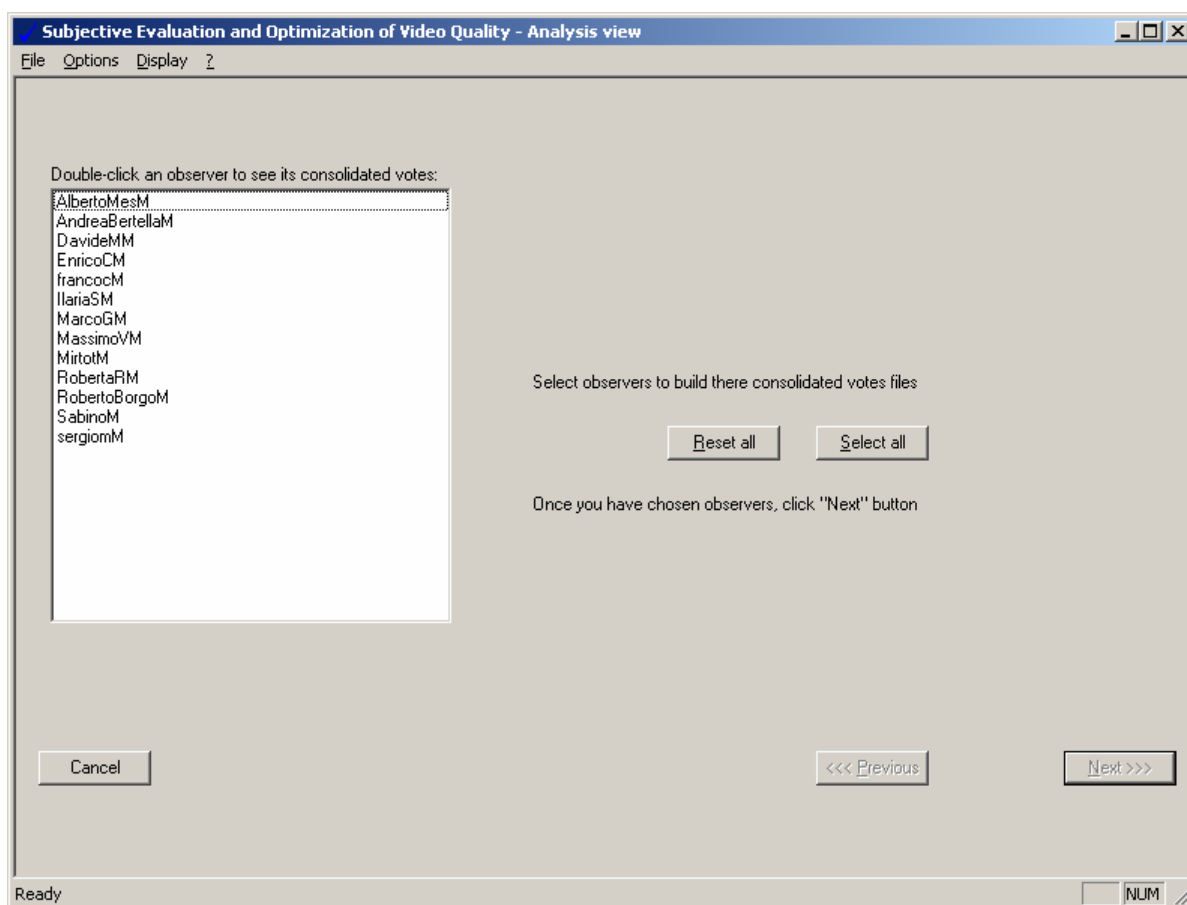


Figure 10 : Observers selection window for a given test.

8.2. The presentation of results in quality evaluation mode.

The results scene by scene and the scenes averages are presented in figures 10 and 10bis :

- ☞ Average
- ☞ Standard deviation
- ☞ 95% confidence interval
- ☞ Asymmetry factor or skewness
- ☞ Beta 2 or Kurtosis: gaussian distribution = 3

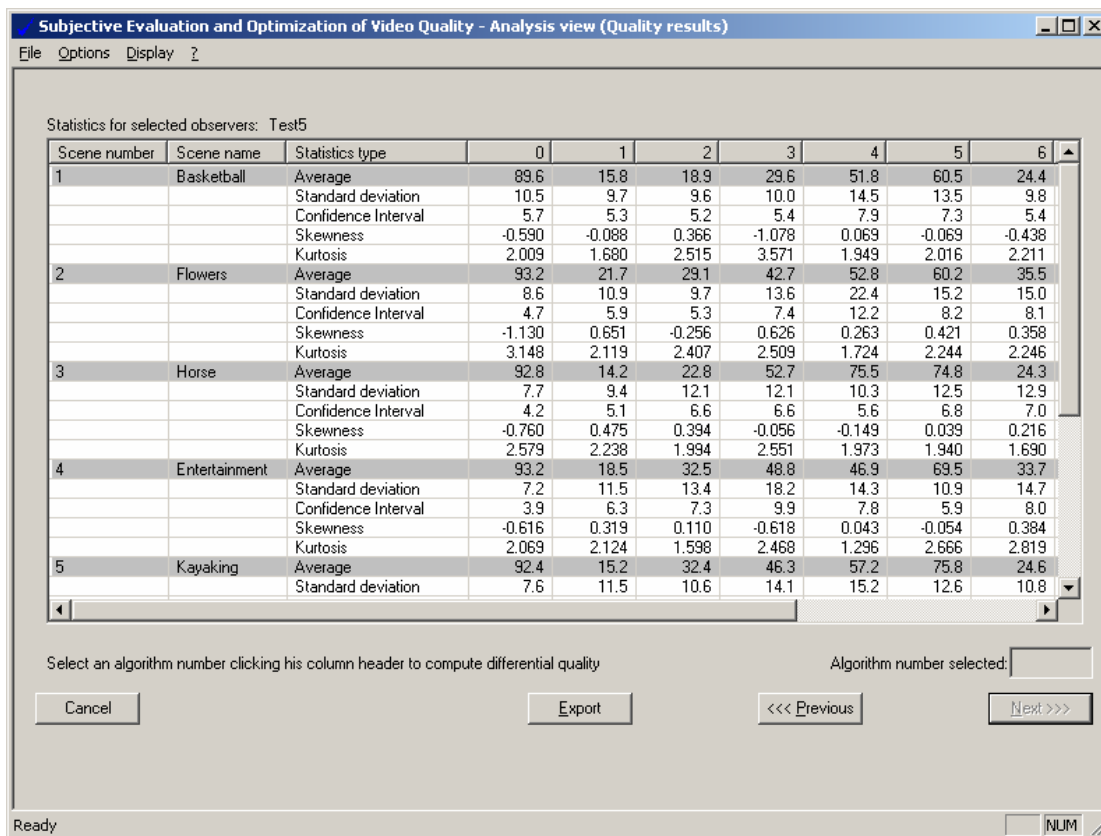


Figure 11 : Presentation window of quality evaluation results

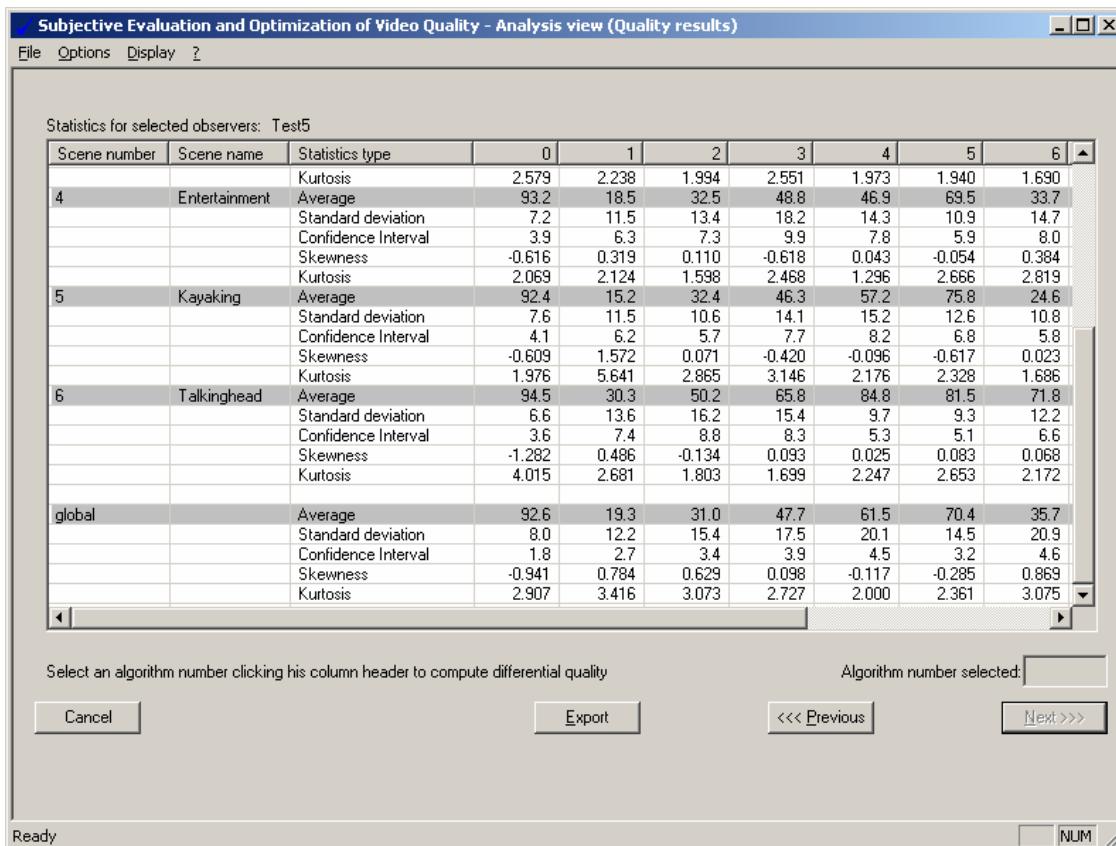


Figure 11 bis : Presentation window of quality evaluation results

These results are exportable under EXCEL allowing a modification of a presentation. Then it may be copied into a WORD document.

The distribution against the scores may be symmetric but not exhibiting a gaussian profile. That is the skewness score that tends towards zero that gives us this information.

In the case where the beta 2 score would rather be far from 3 (non gaussian) and the skewness score far from 0 (non symmetric), it becomes necessary to look at the distribution. First of all, the asymmetry of votes can be linked to a non-linearity between degradation stimulus and the scale of votes. Secondly, a multimode distribution indicates that there are different decision criteria according to observer population. With a simple click on the current average, it is possible to open a window giving the distribution of votes as shown in figures 11. This distribution is exportable directly in a WORD document as shown in figures 12 and 13.

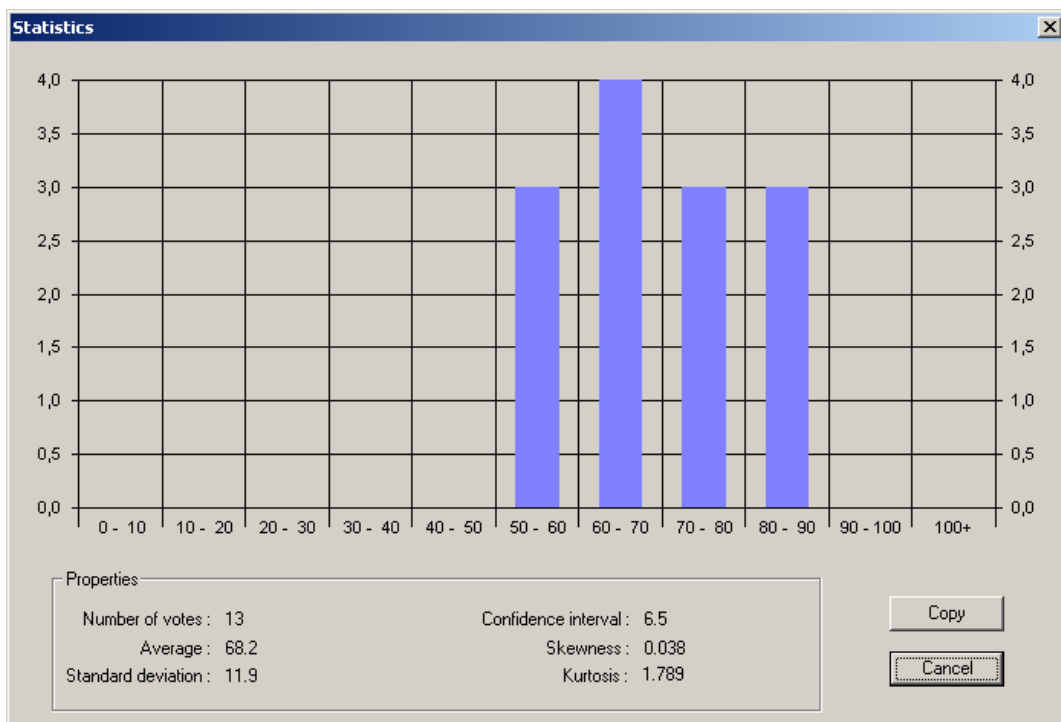


Figure 12 : window of distribution of scores

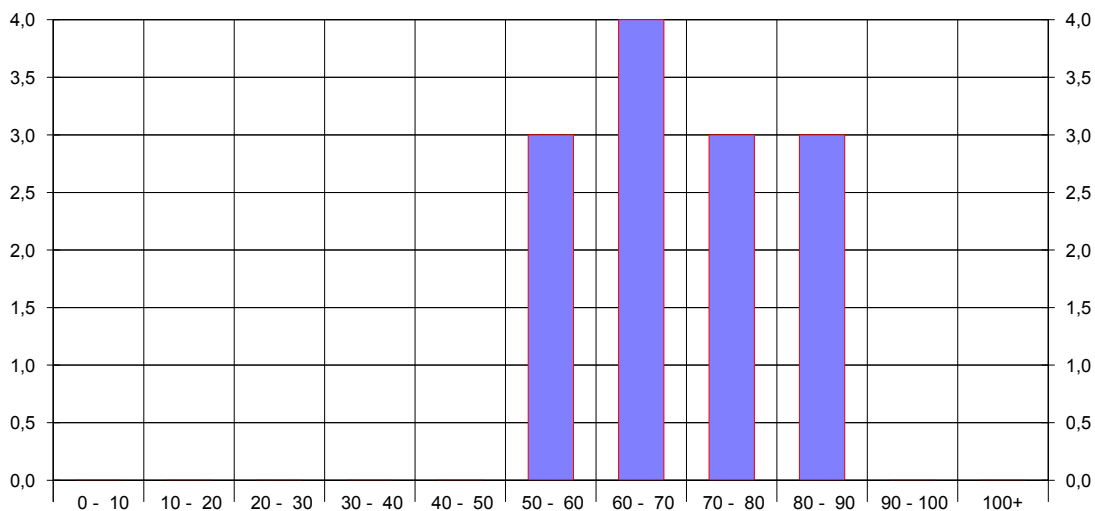


Figure 13 : exported distribution of scores under a WORD document

8.3. The presentation of results in the comparison or acceptability modes

Here a simple sum of votes is carried out. A result in percentage is added that results in either a preference or a level of acceptability scene by scene or by all the scenes. These results are exportable under EXCEL allowing a modification of a presentation. Then it may be copied into a WORD document.

Statistics for selected observers: acceptcinewm8

Scene number	Scene name	Statistics type	1	2	3	4	5
1	basket	Average	94.4	94.4	38.9	5.6	0.0
		Number of votes	17	17	7	1	0
2	flower	Average	100.0	94.4	83.3	22.2	0.0
		Number of votes	18	17	15	4	0
3	horse	Average	100.0	94.4	83.3	33.3	0.0
		Number of votes	18	17	15	6	0
4	mobile	Average	100.0	88.9	44.4	38.9	0.0
		Number of votes	18	16	8	7	0
5	tennis	Average	100.0	88.9	72.2	27.8	0.0
		Number of votes	18	16	13	5	0
global		Average	98.9	92.2	64.4	25.6	0.0

Buttons: Cancel, Export, <<< Previous

Status: Ready NUM