

# Visual assessment of HDR video

Vittorio Baroncini\*, Federica Mangiardi\*, Emiliano Pallotti\*, Massimiliano Agostinelli\*\*, \* Fondazione Ugo Bodoni \*\* Tretag s.r.l.

## Abstract

*The future of Multimedia can be depicted in a scenario where almost all the fruition of audiovisual content will be delivered to the final user by means of streaming over any kind of transmission media. This migration towards streaming will change drastically most of today's scenarios, ranging from TV receivers down to any kind of mobile terminals, including 5G devices. On the other side the content itself will be of better quality; in fact it will be in HD and UHD format, it will be scalable and it will also use a Wider Color Gamut (WCG) and High Dynamic Range (HDR), to improve the User Experience. Furthermore the users will be able to access the Broadcasting services (i.e. TV) also through the new generation 4G/5G networks. This paper deals with the formal subjective assessment evaluation of WCG and HDR video content, and it is located in the framework of the standardization activities of MPEG (ISO-JEC/SC29/WG11) that, together with ITU-T (through the joint committee JCTVC) is active in the definition of new technologies for WCG and HDR video coding. Here the results of the comparison between a new test method designed to evaluate HDR video content, i.e. the HDR-Side By Side (HDR-SBS), and a classic subjective video test method, the DCR (Degradation Category Rating), are described. The new HDR-SBS test method, was designed to evaluate HDR and WCG video material produces by the Submissions made in response to the Call for Evidence (CfE) recently issued by MPEG [1]; the HDR-SBS method is basically a category rating method based on a side-by-side approach. The paper demonstrates that, despite the effort made by MPEG to design a new test method devoted to the assessment of HDR content, and despite what suggested in other papers describing similar tests [2] [3] [6], the DCR method shows better performances. Furthermore the DCR test method results to be less stressing for the human subjects participating to a formal subjective assessment experiment.*

## Introduction

The International Standardization Body MPEG (Moving Picture Expert Viewing) [ISO-JEC/SC29/WG11] issued in the early 2015 a "Call for Evidence on High Dynamic Range and Wide Color Gamut", to assess evidence of new technologies able to initiate the roadmap towards a new standard in this area [1]. Several "Submissions" were received in response to the Call for Evidence (CfE) and were visually assessed, by means of formal subjective visual quality assessment, by two test laboratories: EPFL (École Polytechnique Fédérale de Lausanne) and FUB (Fondazione Ugo Bordonni). In both cases a side by side approach was applied, in the aim of being able to assess also very small differences in quality, mainly for what concerns the color variations. The evaluation was made to verify the "evidence" of improvement in visual quality provided by the "Submissions" against the existing technology, developed in MPEG i.e. HEVC Main 10 Profile (here defined as the "Anchor"). The evaluation done at EPFL compared directly the "Submissions" against the "Anchor".

The evaluation made at FUB considered the "Anchor" as one of the "Submissions".

In the EPFL experiment one half of the screen was showing the "Anchor" and other half of the screen was showing the PVS produced by the "Submissions", at the same bit rate.

Here the viewing subjects were asked to express their opinion performing a "pair comparison" between the "Anchor" and the "Submissions".

In the FUB experiment, here called HDR-SBS (HDR- Side By Side), one half of the screen was showing the uncompressed video sequences (SRC) and other half of the screen was showing the PVS produced by the "Submissions" (considering the "Anchor" as one of the "Submission").

In the FUB experiment the viewing subjects were asked to rate any degradation they could see comparing the SRC with the PVS.

In both the EPFL and FUB experiments the screen was showing only one half of the original video signal, and for each video source (SRC) a different portion of the screen was selected equal to one half of the horizontal dimension. For each SRC the portion shown on the screen was carefully selected by the EPFL and FUB test managers trying to include the more interesting features.

During the activity done to prepare the CfE, the experts of MPEG, decided to adopt the "side by side" approach because it was expected to be more performing for a better evaluation of very small differences in color. In fact it was widely agreed that color differences (mainly when very small) may be better assessed when the reference and the coded pictures are located on the same screen in a side by side fashion. This arrangement was demonstrated in some previous subjective evaluations tests [2] [3] and [4] where the side by side approach was adopted in a successful way. It has to be noted that the above cases were related to both fixed [2] [4] and moving [3] images. At the same time for the subjective evaluation of HDR fixed images described in [5], the ITU-T Recommendation P.910 Degradation Category Rating (DCR) test method was used. Finally in [6] both sequential and side by side presentation strategies were compared assessing visual quality of moving images, demonstrating better performances of the side by side approach.

The main argument against temporal presentation of stimuli (typical of the DCR test) was related to the supposed limitations of the short time memory of the Human Visual System (HVS), mainly when assessing small color variations; furthermore the results of described in [6] pushed the final decision in favor of a side by side approach.

The side by side test method designed for the CfE Submissions evaluation, here called HDR-SBS, was therefore designed splitting the screen in two equal sections and showing SRC and PVS side by side. The presentation was always made two times, taking care of inverting the position the SRC and PVS video during the second presentation; this solution allowed to minimize the impact of any eventual non uniform behavior of the screen of the SIM2 HDR display. At the end of the two presentations a message on the screen asked the viewers to express their vote. The viewing subjects always knew which of the two portion of the screen was presenting the SRC and were asked to assess any degradation