Reviewer #2 comments

The paper presents a software tool to demonstrate that VMAF can be used to evaluate the video quality of WebRTC communications.

Although the development of the tool, as well as the experiments performed, are explained in detail and appropriately, I think the details provided are not relevant enough to justify a journal publication. This detail is ok for a technical report, but not for a research dissemination. Indeed, the conclusion obtained is that VMAF and VIFp correlate with subjective measure better than SSIM and PSNR. But this result is not new, as it had already been proved in previous works.

- Author Action Point (AAP): We have removed some low-level details of the objective evaluation (section 2.2.1). Moreover, the process for the extraction of the padding video has been significantly improved, and the manual part described in the former version of the paper is now completely automatic. All in all, the paragraph:

>“To do that, we use also FFmpeg to extract each frame of the presenter and viewer videos as JPG images. At this point, a manual step needs to be carried out. By inspecting the generated images, we need to identify the exact frame in which the actual video starts and ends. In other words, we need to exclude the padding video sent at the beginning and the end. All in all, we need to include the frame numbers in which the actual content starts in the automated bash script which is used to carry out this step (called calculate_qoe_metrics.sh) and execute the same script again.”

... has been removed, substituted by the following enhanced and simplified paragraph:

>“To do that, the Linux command convert is used to check when the color sequence displayed in padding video bars (white, cyan, purple, blue, yellow, green, red, and black) is present in each frame. This way, the script is able to determine automatically the exact frames in which the actual content starts and finishes.”

Although it has not been explicitly explained in the paper, this enhancement is also available in the GitHub repository which supports the experimentation. The major changes were committed in this patch.

Regarding the VMAF/VIFp correlation with subjective results, indeed this work is aligned with previous results showing that the same correlation also happens in WebRTC, which is our specific use case and it has not been reported before in the literature, to the best of our knowledge.

Moreover, the authors justify the tests because WebRTC is a different use case than the one for which VMAF was developed. However, they do not test some of the possible problems, as audiovisual desynchronization.

- AAP: Audiovisual desynchronization is out of the scope of the paper due to the fact that VMAF (which is a major topic in our work) is a QoE assessment method just for video, not for video and audio. Nevertheless, we agree that audiovisual desynchronization is a potential problem to be studied. All in all, we have included this issue as possible future work in the discussion section.

Finally, in order to validate the results, I believe more than one sequence should be used, considering different types of scenarios.
AAP: Different scenarios has been studied in the paper regarding the packet loss rate between presenter and viewer. Using a different sequence has been proposed as future work in the discussion section, together with different scenarios, or topologies:

“... For instance, our work can be extended by evaluating audiovisual desynchronization or audio quality, for instance using QoE assessment methods such as PESQ (Perceptual Evaluation of Speech Quality) or POLQA (Perceptual Objective Listening Quality Assessment) to name a few. Moreover, other WebRTC use cases can be exercised, for example using additional input video sequences, or different topologies such as the broadcasting scenario (one to many) or video chat (many to many), and using heterogeneous QoS scenarios (e.g. simulating network congestion or different traffic constraints).”