A Perceptual Study of the Decoding Process of the SoftCast Wireless Video Broadcast Scheme

Anthony Trioux, Giuseppe Valenzise, Marco Cagnazzo, Michel Kieffer, François-Xavier Coudoux, Patrick Corlay, M Gharbi

A Perceptual Study of the Decoding Process of the SoftCast Wireless Video Broadcast Scheme

Anthony Trioux, Giuseppe Valenzise, Marco Cagnazzo, Michel Kieffer, François-Xavier Coudoux, Patrick Corlay, Mohamed Gharbi

A Perceptual Study of the Decoding Process of the SoftCast Wireless Video Broadcast Scheme

Anthony Trioux*, Giuseppe Valenzise†, Marco Cagnazzo‡, Michel Kieffer†, François-Xavier Coudoux*, Patrick Corlay*, Mohamed Gharbi*

*UMR 8520 - IEMN, DOAE, Univ. Polytechnique Hauts-de-France, CNRS, Univ. Lille, YNCREA, Centrale Lille, F-59313 Valenciennes, France

†LTCI, Télécom ParisTech, Institut Polytechnique de Paris, France

‡Univ. Paris-Saclay, CNRS, CentraleSupélec, L2S, 91192 Gif-sur-Yvette, France

{anthony.trioux, francois-xavier.coudoux, patrick.corlay, mohamed.gharbi}@uphf.fr,

{giuseppe.valenzise, michel.kieffer}@l2s.centralesupelec.fr, cagnazzo@telecom-paris.fr

***Abstract*—The SoftCast scheme has been proposed as a promising alternative to traditional video broadcasting systems in wireless environments. In its current form, SoftCast performs image decoding at the receiver side by using a Linear Least Square Error (LLSE) estimator. Such approach maximizes the reconstructed quality in terms of Peak Signal-to-Noise Ratio (PSNR). However, we show that the LLSE induces an annoying blur effect at low Channel Signal-to-Noise Ratio (CSNR) quality. To cancel this artifact, we propose to replace the LLSE estimator by the Zero-Forcing (ZF) one. In order to better understand the perceived quality offered by these two estimators, a mathematical characterization as well as an objective and subjective studies are performed. Results show that the gains brought by the LLSE estimator, in terms of PSNR and Structural SIMiliraty (SSIM), are limited and quickly tend to null value as the CSNR increases. However, higher gains are obtained by the ZF estimator when considering the recent Video Multi-method Assessment Fusion (VMAF) metric proposed by Netflix, w hich evaluates the perceptual video quality. This result is confirmed b y the subjective assessment.**

***Index Terms*—SoftCast, Linear Video Coding, Quality assessment, Visual artifacts, Joint Source-Channel Coding**

I. Introduction

SoftCast and recent extensions referred as Linear Video Coding and Transmission (LVCT) systems [1]–[6] have been recently proposed as a promising alternative to H.264/AVC or HEVC-based [7] wireless video transmission schemes. Compared to these schemes which experience the so-called cliff-effect [8] referring to a sudden and abrupt loss of quality, the received video quality obtained with LVCT schemes such as SoftCast [1], WaveCast [2], etc. scales linearly with the Channel Signal-to-Noise Ratio (CSNR) [3]. Particularly useful in broadcast and mobile applications, they provide quality of service even in the presence of suddenly degraded channel quality as shown in Fig. 1. This property comes from the linear processing applied to the pixels, avoiding quantization or entropy coding, and the transmission carried out without channel coding.

While LVCT avoid annoying freezes and glitches of the video, they reconstruct videos presenting distortions referred as snow effect. This artifact is illustrated with the SoftCast scheme in Fig. 2. It is strongly visible for low CSNR values ($\leq$ 10dB) but becomes almost invisible for higher values.

(a) H.264/AVC (b) SoftCast

Fig. 1: Example from [1] of visual quality comparison between SoftCast and H.264/AVC video coding and transmission scheme at CSNR=9dB. The complete videos are available at http://people.csail.mit.edu/szym/softcast/videos.html.

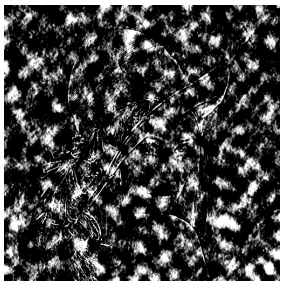
In order to reduce this distortion, traditional LCVT approaches use a LLSE estimator at the receiver, which maximizes the reconstructed PSNR. However, we show in this paper that although the PSNR obtained is higher, it modifies at low CSNR, the edges' sharpness of the transmitted video. This modification introduces a blur effect that can be annoying for the user as shown in Fig. 2a. In contrast, the edges remain sharp when using the ZF estimator (Fig. 2b), however the price to be paid is a stronger snow effect over the video.

To this end, this paper proposes to investigate the trade-off between blur and snow effects and therefore to better understand the quality of experience [9] offered by these two estimators in a SoftCast context. First, the estimators are characterized mathematically. Then, an objective evaluation of the received quality offered by the two estimators is performed by considering the metrics suggested by [10] as they offer the best correlation with human judgment. Finally, a subjective evaluation based on a forced-choice PairWise Comparison (PWC) and statistical analysis is proposed.

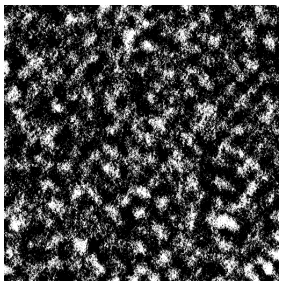
The paper is organized as follows: Section II introduces and reviews the SoftCast scheme. In Section III, the estimators are characterized mathematically. The performance of the two linear estimators is then assessed through objective assessment

(a) LLSE: PSNR=24.63dB

(b) ZF: PSNR=23.24dB



(c) LLSE: PSNR=24.63dB



(d) ZF: PSNR=23.24dB

Fig. 2: Illustration of the artifacts generated by SoftCast on the *Lena* image transmitted without compression in a channel with CSNR=0dB. (a) Reconstructed images with SoftCast(LLSE) depicting snow and blur effects. (b) Reconstructed image with SoftCast(ZF) depicting snow effect. (c),(d) Resulting LLSE and ZF error images, respectively.

in Section IV. The subjective quality assessment is finally presented in Section V. Conclusions and discussions are given in Section VI.

In what follows and for ease of reading, we denote the SoftCast with LLSE estimator and the SoftCast with ZF estimator: SoftCast(LLSE) and SoftCast(ZF), respectively.

II. SoftCast scheme Review

The basic scheme of SoftCast [1] is introduced in Fig. 3. SoftCast first takes a Group of Pictures (GoP) and uses a 3D full-frame DCT as a decorrelation transform. The DCT frames are divided into N small rectangular blocks of transformed coefficients called *chunks*. The data compression can be done in SoftCast after the decorrelation transform. Specifically, when the available channel bandwidth for the transmission is smaller than the signal bandwidth, SoftCast discards the chunks with less energy *i.e.,* only $M < N$ chunks may be transmitted. This is generally the case especially for the transmission of High Definition (HD) content as mentioned in [6]. At the receiver side, these discarded chunks are replaced by null values [1]. To represent the bandwidth limitation, one may use the Compression Ratio (CR) [6] defined as:

$$\text{CR} = M/N. \tag{1}$$

The third block at the transmitter consists of a chunk scaling operation to match the transmission power constraints. The scaling coefficients denoted $g_i, i = 1, 2, \cdots, M$ are chosen so as to distribute the available power P to all the chunks in a

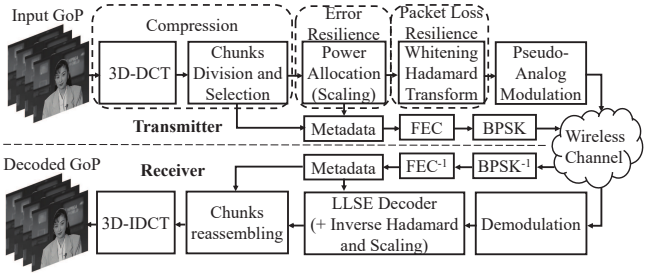


Fig. 3: Block diagram of the SoftCast scheme.

way that minimizes the Mean Square Error (MSE) between transmitted and decoded chunks. This is a typical Lagrangian problem and the solution is given by:

$$g_i = c \cdot \lambda_i^{-1/4}, \tag{2}$$

where $c = \sqrt{\frac{P}{\sum_j \sqrt{\lambda_j}}}$ and λ_i is the energy of the i^{th} chunk.

A Hadamard transform is then applied to the scaled chunks to provide packet loss resilience. This process transforms the chunks into slices. Each slice is a linear combination of all scaled-chunks. Finally, the slices are transmitted in a pseudo-analog manner using Raw-OFDM [1]. Classical channel coding and modulation are skipped.

In parallel, the SoftCast transmitter sends an amount of data referred as metadata. These data consist of the mean and the variance of each transmitted chunk as well as a bitmap, indicating the positions of the discarded chunks in the GoP. Metadata are strongly protected and transmitted in a robust way [1] to ensure correct delivery and decoding.

At the receiver side, a Linear Least Square Error (LLSE) decoder is used to estimate the values of the chunks based on channel noise estimation. Using the metadata, the decoded values are then reassembled to form DCT-frames, which are then passed through an inverse 3D-DCT process.

III. Characterization of the two estimators

In this section, we analyze the behaviors of the two estimators. For ease of understanding and without loss of generality, we consider the transmission of images *i.e.,* only the spatial DCT is considered for this analysis. Furthermore, the Hadamard transform is not considered in the following analysis as it does not change either the transmission power or channel noise characteristics [3].

At the transmitter side, we recall that SoftCast first scales the magnitude of the chunks to offer a better protection against transmission noise:

$$y_i = g_i \cdot x_i. \tag{3}$$

where x_i and y_i represent the i^{th} chunk and scaled chunk, respectively.

The transmitted signal is then corrupted by Additive White Gaussian Noise (AWGN):

$$\begin{aligned} \hat{y}_i &= y_i + n_i, \\ &= g_i \cdot x_i + n_i. \end{aligned} \tag{4}$$

where n_i is the channel noise defined as $N(0, \sigma^2)$.

When the ZF estimator is used, the received chunks are simply estimated by performing the inverse scaling operation:

$$\begin{aligned} \hat{x}_i(\text{ZF}) &= \frac{\hat{y}_i}{g_i}, \\ &= x_i + \frac{n_i}{g_i}. \end{aligned} \tag{5}$$

After undergoing an inverse DCT process, the reconstructed image I_{rec} using the ZF estimator is the sum of the original image I_{ori} and B_i the resulting noise after inverse DCT process:

$$\begin{aligned} \mathrm{I}_{rec} &= \text{DCT}^{-1}\{\hat{x}_i(\text{ZF})\}, \\ &= \text{DCT}^{-1}\{x_i\} + \text{DCT}^{-1}\{\frac{n_i}{g_i}\}, \\ &= \mathrm{I}_{ori} + B_i. \end{aligned} \tag{6}$$

In the case of the LLSE estimator, the chunks are estimated by leveraging channel noise estimation [1], [11]:

$$\begin{aligned} \hat{x}_i(\text{LLSE}) &= \frac{g_i\lambda_i}{g_i^2\lambda_i + \sigma^2} \cdot \hat{y}_i, \\ &= \frac{1}{1 + \frac{\sigma^2}{g_i^2\lambda_i}} \cdot \frac{\hat{y}_i}{g_i}. \end{aligned} \tag{7}$$

By substituting the values of g_i according to (2), we get:

$$\hat{x}_i(\text{LLSE}) = \frac{1}{1 + \frac{\sigma^2}{c^2\sqrt{\lambda_i}}} \cdot \hat{x}_i(\text{ZF}). \tag{8}$$

As observed in (8), the estimated DCT coefficients obtained by the LLSE estimator are attenuated version of the ones obtained by the ZF estimator:

- When considering high CSNR values, *i.e.,* low σ^2 value, the LLSE and ZF estimator perform similar.
- When considering low CSNR values, we observe that the lower the energy λ_i of the chunk is, the higher the attenuation is. It is well known that high frequencies represent the edges of an image, and that for still images, these high frequencies carry low energy after DCT process. Therefore, although the LLSE estimator allows to reduce the MSE *i.e.,* increase the reconstructed PSNR, it acts as a filter that attenuates more strongly the edges of the images. The decoded images in Fig. 2 clearly illustrates this phenomenon, where using the LLSE estimator allows to improve PSNR value of about 1.4dB. However, the perceived quality may not be as good as with the ZF estimator since the edges of the image obtained with the LLSE estimator are clearly modified as shown in the error image.

Finally, we note that to be able to perform the decoding process with the SoftCast(LLSE) scheme, the receiver needs an additional information, which is an estimate of the channel's noise (σ^2) as shown in (7). This data may not be available for some applications. In that case, only the ZF estimator can be used since it does not require any channel quality information to decode the video. Therefore, a study needs to be conducted to compare the two estimators and to better understand the quality of experience offered by themselves. In the following, we first evaluate their performance based on well-known objective metrics.

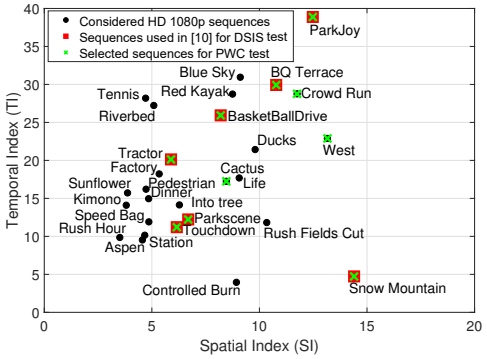


Fig. 4: Resulting Spatial and Temporal Information (SI, TI) indexes for the considered video sequences.

IV. Objective evaluation

To assess the performance of the two estimators, we select and retain the four following objective metrics: PSNR, SSIM, MS-SSIM and the recent VMAF [12] metric proposed by Netflix, which is known to evaluate the perceptual video quality. The metrics are chosen due to their high correlation with the subjective scores obtained in our recent study [10].

Configuration: Three GoP-sizes of 8, 16 and 32 frames and four CRs ranging from 0.25 (75% of discarded chunks) to 1 (no compression applied) by 0.25 step are considered as in [6]. Transmissions through AWGN channels are simulated and represented by a CSNR value varying from 0 to 27dB by 3dB step. Each frame is split into 192 chunks as in [10]. As classically done in the literature [1]–[6] and as it does not influence the perception of the blur, only the luminance is considered in this paper.

Material: The objective evaluation is performed considering different video samples, each with a duration of 5 seconds. Their spatiotemporal complexity is computed and displayed in Fig. 4 using a modified version of the Spatial and Temporal Information (SI, TI) indexes proposed by the ITU-R [13]:

$$\text{SI} = mean_{time}\{std_{space}[Sobel(F_k(i,j))]\}, \tag{9}$$

$$\text{TI} = mean_{time}\{std_{space}[F_k(i,j) - F_{k-1}(i,j)]\}. \tag{10}$$

where $F_k(i,j)$ represents the k^{th} frame, (i,j) the corresponding spatial coordinates and $Sobel()$ the Sobel filtering operation, respectively.

Note that, in the original definition of the SI,TI indexes, the highest value along the time axis is selected instead of computing the mean value. We choose to average the results over the sequence in order to better characterize the video complexity. Indeed, when using the original definition of the TI index for a video with relative slow motions that contains cut(s), the final TI value will be high due to the scene change, even if the video contains relative slow motions.

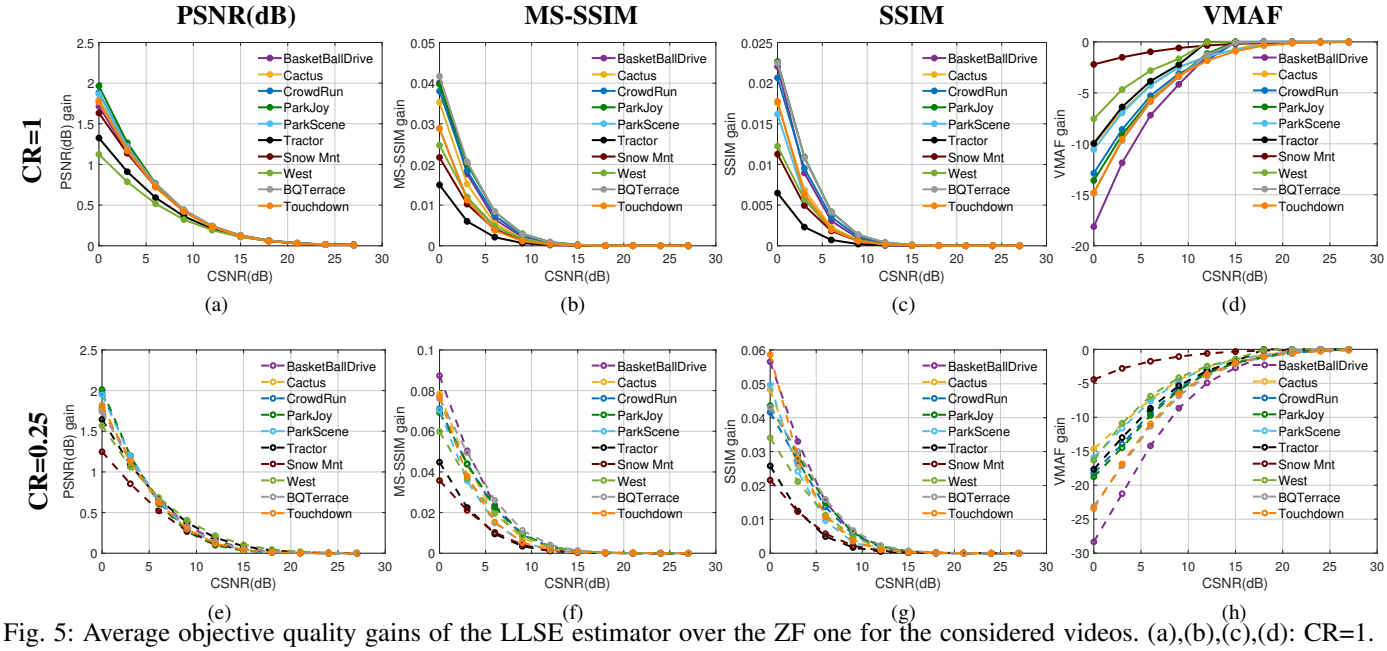


Fig. 5: Average objective quality gains of the LLSE estimator over the ZF one for the considered videos. (a),(b),(c),(d): CR=1. (e),(f),(g),(h): CR=0.25. (a),(e): PSNR(dB). (b),(f): MS-SSIM. (c),(g): SSIM. (d),(h): VMAF.

Methodology: To assess the performance of each estimator, we define the quality gain as the difference between the SoftCast(LLSE) and SoftCast(ZF) scores (the former minus the latter). For ease of reading, only the results for the green selected sequences in Fig. 4 are shown in this paper. Furthermore, due to space limitations, we only show the results for two CRs (0.25 and 1) and one GoP-size of 32 frames. Similar quality gains are obtained for the other video content displayed in Fig. 4 and when considering different GoP-sizes (8 and 16) as well as different CR (0.5 and 0.75).

Results displayed in Fig. 5 show that:

- Regardless of the considered objective metric and the configuration (GoP-size, available bandwidth, transmitted video content), the quality gain quickly decreases as the CSNR increases and becomes almost null when CSNR$\geq$10dB. This is perfectly explained in Section III, since the two estimators perform similar for high CSNR values;
- For low CSNR values ($\leq$ 10dB), as already known and regardless of the transmitted video content, the LLSE estimator outperforms the ZF one in terms of PSNR. However, this gain is limited and remains low as shown in Fig 5a and Fig 5e. Furthermore, the quality gain is less pronounced when considering the SSIM and MS-SSIM metrics with a maximum quality gain of only 0.08;
- However, as shown in Fig. 5d and Fig. 5h, regardless of the considered video, SoftCast(ZF) performs better when considering the VMAF metric with significant quality gain at low CSNR (recall that the VMAF metric ranges between [0-100]). This is due to the fact the final VMAF score is partially generated using the Detail Loss Metric (DLM) [14]. The DLM measures the loss of details affecting the content visibility. In our case and as illustrated in Fig. 2, the reconstructed videos with the SoftCast(ZF) are less affected by the loss of details than the SoftCast(LLSE) ones.

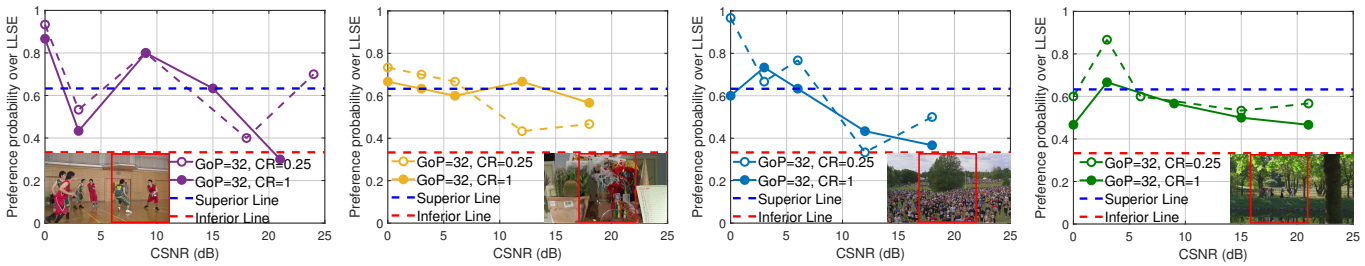
V. Subjective Evaluation

In addition to objective assessment, we also provide a subjective study described hereafter.

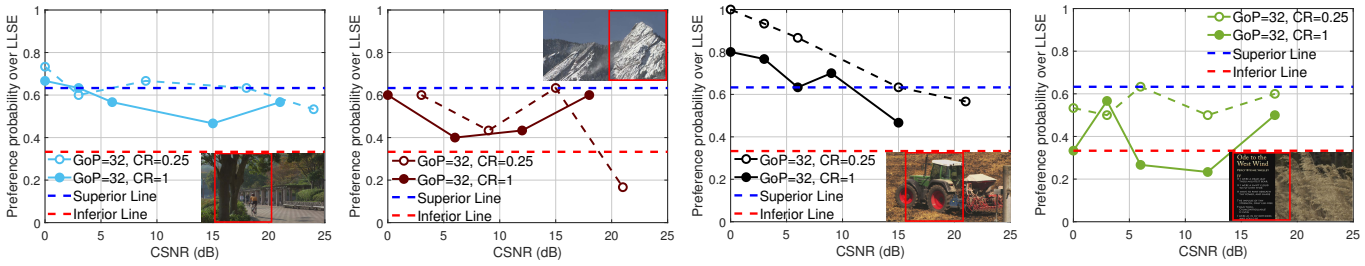
Environment: The test was performed respecting the BT.500-14 recommendation [13] provided by the ITU-R. Specifically, it took place in a dark and quiet room, with a measured ambiant luminance of 2 lux and color temperature of 6500K. The screen used for display had a 1920×1080 resolution and a height H of 40cm. Users were placed at a fixed distance from the screen which equals three times the height of the display.

Observers: Thirty observers including 21 men and 9 women took part in the experiment. The group's average age is 33 varying between 25 and 62. All of the observers have normal or corrected to normal visual acuity.

Test methodology: The pairwise comparison with forced choice was used in this study. Specifically, the reconstructed videos for each estimator were presented to the user in a side-by-side fashion. Each video was cropped using a window of 955×1980 pixels so as to separate the video with a black border of 10 pixels. Due to the limited duration of the videos (5 seconds), each pair was presented two times to the users before asking them to select the video that they preferred (left or right). Users were first familiarized with the procedure and environment through a training test. Training videos, *i.e., BQTerrace* and *Touchdown* are not considered in the results. Each participant was asked to evaluate 88 stimuli including four dummies (scores not saved) at the beginning of the test. These dummies were replayed at the end of the



(a) *BasketBall*, crop (966:1920) (b) *Cactus*, crop (465:1419) (c) *CrowdRun*, crop (485:1439) (d) *ParkJoy*, crop (305:1259)



(e) *ParkScene*, crop (185:1139) (f) *Snow Mountain*, crop (966:1920) (g) *Tractor*, crop (210:1164) (h) *West*, crop (55:1009)

Fig. 6: Evolution of the preference probability to select the ZF estimator over the LLSE one as a function of the CSNR. Sequences: a) *BasketBallDrive*, b) *Cactus*, c) *CrowdRun*, d) *ParkJoy*, e) *ParkScene*, f) *Snow Mountain* g) *Tractor*, h) *West*. The red rectangle on the starting frame of each video represents the crop used for the PWC test.

test. The stimuli were carefully randomly presented to each user by making sure that each content does not appear twice consecutively.

Test material: Ten HD 1080p sequences from [15] and [16] were selected and used in this study to represent different levels of spatiotemporal complexity. Among the ten sequences, seven sequences were used in [10]. In addition, three more content that present different features (text on the videos, rich textures, etc.) were added. The crop over the videos were made in order to keep a maximum of information over the duration of the video. An illustration of the selected videos and corresponding crops is available in Fig. 6.

In order to satisfy the suggested maximum duration of a subjective test [13] and based on previous observation [10], we carefully selected a subset of all generated video content. Specifically, we selected two CRs of 0.25 and 1 to study whether or not the compression influences the user's preference. Furthermore, among all the GoP-size, we chose the GoP-size of 32 frames as it usually gives the best reconstructed video [3]. Finally, more stimuli were considered in the low CSNR range [0-15dB] since as explained in Section III noticeable difference between SoftCast(ZF) and SoftCast(LLSE) only appears for such CSNR values.

Statistical analysis: Prior to the data processing step, the outlier detection mechanism proposed by Mantiuk *et al.* [17] was used. The latter is based on the computation of the maximum likelihood of each observer with respect to the whole observers. At the end of this detection mechanism, a score is assigned to each observer and those obtaining a score close or above the threshold of 1.5 should be further investigated. In such case, a visual comparison (plot) should be made between the responses of this observer and the rest of the group. Note that the final decision on whether the latter observer should be considered as an outlier or not is left to the designer of the test. In order to improve the detection of a potential outlier, we randomly added six "obvious" comparison stimuli in the test. By "obvious", we mean for instance a pair that is composed of the reconstructed video with the ZF estimator in a bad channel (0dB) compared to the one reconstructed with the LLSE estimator in a good channel (15dB). Among the 30 observers, the mechanism of Mantiuk *et al.* only highlighted one user that required more attention. After having carefully checked the results of this observer and based on the results he provided for the "obvious" comparisons, we concluded that there was no outlier in the panel.

To analyze the obtained results, we verified if a preference for one of the estimators could be statistically observed by using the analysis proposed in [18].

Specifically, for stimulus k, we computed the preference probability of choosing the ZF estimator over the LLSE one:

$$\mathrm{P}_{\mathrm{ZF}k} = \frac{w_{\mathrm{ZF}k}}{V}, \tag{11}$$

where $w_{\mathrm{ZF}k}$ represents the number of votes in favor of SoftCast(ZF) for the k^{th} stimulus and V represents the total number of observers.

We then define two thresholds or critical regions to statistically determine whether the ZF estimator outperforms, is equal or is inferior to the LLSE estimator. We start from the hypothesis that the two estimators have an equal probability to be preferred. The preference probability therefore comes from a Bernoulli process B(V, p) where $p = 0.5$. By using the Cumulative Distribution Function (CDF) B($w_{\mathrm{ZF}k}$, V, p) and selecting 5% and 95% as the level of significance of the test, we can determine whether the choice is statistically significant or not. Since B(19,30,0.5)=0.9506, we consider that if there

are more than 19 votes for the ZF estimator it offers significant better quality. Reciprocally, since B(10,30,0.5)=0.0493, we consider that if there are less than 10 votes for the ZF estimator it offers lower quality than the LLSE one. These thresholds are represented in Fig. 6 with dashed blue ($\frac{19}{30} = 0.63$) and red lines ($\frac{10}{30} = 0.33$). Values on or above the superior line (dashed blue line) indicate statistically significant preference for the ZF estimator whereas values on or below the inferior line (dashed red line) indicate statistically significant preference for the LLSE estimator.

The resulting database with associated objective metrics as well as the preference probability for each pair are available at https://ieee-dataport.org/open-access/perceptual-study-decoding-process-softcast-wireless-video-broadcast-scheme.

Results in Fig. 6 show that:

- When considering high CSNR values ($\geq$ 10dB), for most of the cases and as expected, there is no statistical preference since the results lie in between the two dotted lines. This is in accordance with the mathematical analysis and the objective study;
- There is no statistical preference for the *Snow Mountain* sequence even at low CSNR values. Indeed, when selecting the material for the test, we noticed only very small visual differences between the two estimators;
- Although the *West* sequence contains a lot of blur due to the scrolling text, the preference is not statistically significant. Users mentioned at the end of the test that the black background where the text scrolls was particularly noisy with the ZF estimator, therefore influenced their overall judgment;
- Users also mentioned that the choice was not easy for the sequences with strong motions such as *Cactus*;
- On average, the preference for the ZF estimator is observed at low CSNR. It is especially true for the *Tractor* sequence where observers mentioned that the blur effect was clearly annoying and especially visible on the tractor's logo;

VI. Conclusion

In this paper, we study the perceived quality considering two different estimators (LLSE and ZF) for the SoftCast scheme. To this end, objective and subjective studies are performed. Based on extensive simulations on video content with different characteristics (spatiotemporal complexity, text, textures, etc.), different CR and CSNR values, we show that the ZF estimator can be a good choice when considering the perceived quality. The differences between the two estimators are mainly observed at low CSNR ($\leq$ 10dB) since for higher CSNR values, they perform similar. According to the objective assessment, the gains brought by the LLSE estimator at low CSNR values are limited when considering the PSNR, SSIM and MS-SSIM metrics. In contrast, when considering the VMAF metric higher gains are obtained with the ZF estimator due to the fact that the edges are kept sharp. At low CSNR values, the preference for the ZF estimator is also statistically observed with the subjective comparison for most of the videos. Indeed, the preference probability not only depends on the estimator used but also on the spatiotemporal complexity of the video as well as the content itself since blur may not be well perceived for videos with strong motions due to temporal masking effect. Future works may concern the design of a linear estimator that further reduces the noise while keeping the edges of the images sharp.

VII. Acknowledgment

We would like to thank the French National Research Network GdR 720 ISIS for the mobility support. We also would like to thank Dr. E. Zerman and M. F. Comps for their helps in preparing the room test; Dr. Y. Yuan for her helpful comments on this paper as well as all the observers for their participation.

References

[1] S. Jakubczak and D. Katabi, "Softcast: one-size-fits-all wireless video," in *Proceedings of the ACM SIGCOMM 2010 conf.*, 2010, pp. 449–450.

[2] X. Fan, R. Xiong, F. Wu, and D. Zhao, "Wavecast: Wavelet based wireless video broadcast using lossy transmission," in *Proc. IEEE Visual Commun. and Image Process. (VCIP)*, Nov. 2012.

[3] R. Xiong, F. Wu, J. Xu, X. Fan *et al.*, "Analysis of decorrelation transform gain for uncoded wireless image and video communication," *IEEE Trans. Image Process.*, vol. 25, no. 4, pp. 1820–1833, Apr. 2016.

[4] A. Trioux, F.-X. Coudoux, P. Corlay, and M. Gharbi, "A reduced complexity/side information preprocessing method for high quality softcast-based video delivery," in *European Workshop on Visual Inform. Process. (EUVIP)*, 2019.

[5] S. Zheng, M. Cagnazzo, and M. Kieffer, "Optimal and suboptimal channel precoding and decoding matrices for linear video coding," *Signal Proc.: Image Commun.*, vol. 78, pp. 135–151, Oct. 2019.

[6] A. Trioux, F.-X. Coudoux, P. Corlay, and M. Gharbi, "Temporal information based GoP adaptation for linear video delivery schemes," *Signal Proc.: Image Commun.*, vol. 82, p. 115734, 2020.

[7] G. J. Sullivan, J. R. Ohm, W. J. Han, and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) Standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1649–1668, Dec. 2012.

[8] S. Kokalj-Filipovi and E. Soljanin, "Suppressing the cliff effect in video reproduction quality," *Bell Labs Tech. Journal*, vol. 16, no. 4, pp. 171–185, Mar. 2012.

[9] J. Nightingale, Q. Wang, C. Grecos, and S. Goma, "The impact of network impairment on quality of experience (QoE) in h. 265/HEVC video streaming," *IEEE Trans. Consum. Electron*, vol. 60, no. 2, pp. 242–250, 2014.

[10] A. Trioux, G. Valensize, M. Cagnazzo, M. Kieffer *et al.*, "Subjective and objective quality assessment of the softcast video transmission scheme," in *IEEE Visual Commun. and Image Process. (VCIP)*, 2020.

[11] T. Fujihashi, T. Koike-Akino, T. Watanabe, and P. V. Orlik, "High-Quality Soft Video Delivery With GMRF-Based Overhead Reduction," *IEEE Trans. Multimedia*, vol. 20, no. 2, pp. 473–483, Feb. 2018.

[12] C. G. Bampis *et al.*, "Spatiotemporal Feature Integration and Model Fusion for Full Reference Video Quality Assessment," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 8, pp. 2256–2270, Aug. 2019.

[13] "RECOMMENDATION ITU-R BT.500-14 - Methodologies for the subjective assessment of the quality of television images," Oct. 2019.

[14] S. Li, F. Zhang, L. Ma, and K. N. Ngan, "Image Quality Assessment by Separately Evaluating Detail Losses and Additive Impairments," *Trans. Multimedia*, vol. 13, no. 5, pp. 935–949, Oct. 2011.

[15] Xiph, "Xiph.org media."

[16] VQEG, "VQEG HDTV database. video quality experts group (VQEG)." [Online]. Available: http://www.its.bldrdoc.gov/vqeg/projects/hdtv

[17] M. Perez-Ortiz and R. K. Mantiuk, "A practical guide and software for analysing pairwise comparison experiments," Dec. 2017. [Online]. Available: https://github.com/mantiuk/pwcmp

[18] P. Hanhart, M. Řeřábek, and T. Ebrahimi, "Towards high dynamic range extensions of HEVC: subjective evaluation of potential coding technologies," in *Applications of Digital Image Processing XXXVIII*, vol. 9599, 2015, p. 95990G.