A New Adaptation Lever in 360° Video Streaming

Lucile Sassatelli, Marco Winckler, Thomas Fisichella, Ramon Aparicio-Pardo, Anne-Marie Déry-Pinna

A New Adaptation Lever in 360° Video Streaming

Lucile Sassatelli, Marco Winckler, Thomas Fisichella, Ramon Aparicio, Anne-Marie Pinna-Déry*
Université Côte d'Azur, CNRS, I3S
Sophia Antipolis, France

ABSTRACT

Despite exciting prospects, the development of 360° videos is persistently hindered by the difficulty to stream them. To reduce the data rate, existing streaming strategies adapt the video rate to the user's Field of View (FoV), but the difficulty of predicting the FoV and persistent lack of bandwidth are important obstacles to achieve best experience. In this article we exploit the recent findings on human attention in VR to introduce a new additional degree of freedom for the streaming algorithm to leverage: Virtuall Walls (VWs) are designed to translate bandwidth limitation into a new type of impairment allowing to preserve the visual quality by subtly limiting the user's freedom in well-chosen periods. We carry out experiments with 18 users and confirm that, if the VW is positioned after the exploration phase in scenes with concentrated saliency, a substantial fraction of users seldom perceive it. With a double-stimulus approach, we show that, compared with a reference with no VW consuming the same amount of data, VW can improve the quality of experience. Simulation of different FoV-based streaming adaptations with and without VW show that VW enables reduction in stalls and increases quality in FoV.

CCS CONCEPTS

• **Human-centered computing** → **User studies**; **Virtual reality**; • **Networks** → *Network simulations*.

KEYWORDS

360 video, streaming, user attention, limited bandwidth, freedom

ACM Reference Format:
Lucile Sassatelli, Marco Winckler, Thomas Fisichella, Ramon Aparicio, Anne-Marie Pinna-Déry. 2019. A New Adaptation Lever in 360° Video Streaming. In *29th ACM SIGMM Workshop on Network and Operating Systems Support for Digital Audio and Video (NOSSDAV '19), June 21, 2019, Amherst, MA, USA.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3304112.3325610

1 INTRODUCTION

Virtual Reality (VR) is on the rise, with Cisco predicting a 20-fold increase in Internet traffic generated by immersive applications by 2021. 360° videos are an important modality of VR enabling applications in story-telling, journalism or remote education. Despite these exciting prospects, the development of immersive applications is persistently hindered by the difficulty to access them through Internet streaming. Their bandwidth requirements are indeed two orders of magnitude that of a regular video for equivalent perceived quality [2]. To reduce the required data rate, the non-visible part of the sphere can be sent with lower quality. Doing so requires however to predict, at the time of transmission, that is possibly several seconds ahead of playback, where the user is going to look at. Such prediction is only partly possible over very short time horizons owing to the complex dependency on previous motion and content, and inherent randomness [7]. The difficulty of correct prediction and the often substantial discrepancy between the video rate of the highest quality and the available bandwidth (that will worsen with future higher resolution headsets) make the quality displayed in the user's Field of View (FoV) often too low to be satisfactory, entailing a low contentment and possible sickness.

This article contributes to improve the user's Quality of Experience (QoE) in VR by taking a new stance: given the richness of the VR experience, we posit that the visual quality is not the only and not always the best dimension in which the content can be degraded to fit the available bandwidth. A lower visual quality can sharply degrade the feeling of immersion as well as negatively impact the vestibular system. In this article we exploit the recent findings on the human attentional process in VR to introduce a new additional degree of freedom for the streaming algorithm to leverage to improve QoE. **Contributions**:

• We introduce a new effect named Virtual Wall (VW) which translates bandwidth limitation into a new type of impairment allowing to preserve the visual quality: it limits the user's freedom in well-chosen yet frequent periods. We make the hypothesis that this effect can improve the user's QoE compared with legacy FoV-based adaptation using compression only, for a given bandwidth budget.

• To verify the hypothesis, we implement the effect in a DASH-SRD 360° video player and carry out experiments based on a double-stimulus approach with 18 users to collect subjective ratings and head motion traces. The results confirm that, when placed appropriately, the users seldom sense the VW, and that a majority of users prefer the version with VW. We identify that users are mostly not affected by the freedom restriction for ca. 10s. We analyze the importance of quality and responsiveness in the user's preference.

• We simulate different streaming adaptation logics fed with the collected head motion traces, and show that, even in the presence of substantial playback buffers and with FoV-based adaptation, VW enables reduction in stalls and increases quality in FoV, thereby validating it as an additional adaptation lever.

Sec. 2 presents the main related works. The VW effect is motivated and introduced in Sec. 3. Sec. 4 details the design of the user experiments. The results are analyzed in Sec. 5 and streaming simulations in Sec. 6. Sec. 7 concludes the article.

*Corresponding author: sassatelli@i3s.unice.fr

NOSSDAV '19, June 21, 2019, Amherst, MA, USA

ACM ISBN 978-1-4503-6298-6/19/06…$15.00
https://doi.org/10.1145/3304112.3325610

2 RELATED WORKS

To cope with bandwidth limitations, the concept of adaptive video streaming has been extended to 360° videos, where the encoding rate can be adapted in time and space, deciding transmissions based on the FoV. Such adaptation is enabled by the SRD extension to the MPEG-DASH standard [8] and by the recent MPEG-OMAF standard including these capabilities.Several FoV-based adaptation logics have been proposed, for example [12] which leverages the responsiveness brought by HTTP/2 to make replacements striving to maintain a high quality in the FoV when possible. The dependence of the streaming performance on the user's behavior requires understanding the human attention in VR to design better streaming algorithms. In [10], Sitzmann et al. show (with 169 users) that the average exploration time, that is the time a user takes to scan the whole 360°-wide longitude span, is 19 seconds. They also show that this duration tends to decrease when the scene is made of a lower number of well-isolated Regions of Interest (RoIs). In [4], David et al. present a dataset of 20s-long 360° videos and head motion traces from 57 users. They show that the exploration phase in their videos last between 5 and 10s. In [1], Almquist et al. propose a taxonomy of 360° content by analyzing the distribution of the head positions obtained from 32 users on 30 videos of average duration 3 minutes. From these findings, the authors identify video classes on which the head position prediction task is made easier, and how the streaming algorithms can consequently be adapted. In [3], Dambra et al. show how film editing can be designed to better predict the head position, thereby easing streaming and making the proof of the concept that user's attention driving tools can be designed jointly with the transmission algorithm in order to improve streaming. We build on all these key tools and findings to propose a new lever to improve the QoE of VR streaming.

3 A NEW ADAPTATION LEVER: VIRTUAL WALLS

Motivation: It has been recently shown in [10] and [1] that, when presented with a new VR scene[1], a human first goes through an exploratory phase that lasts about 10 to 15s ([1, Fig. 18], [10, Fig. 2]), before settling down on RoIs, that are salient areas of the content. Almquist et al. have identified the following main video categories for which they could discriminate significantly different users' behaviors: *exploration, static focus, moving focus* and *rides*. In *exploration* videos, the spatial distribution of the users' head positions tend to be more widespread; for that reason the somewhat homogeneous content (i.e., with high-entropy) in *exploration* videos hardly allows to predict where the users will watch and possibly focus on. *Static focus videos* are made of a single salient object (e.g., a standing-still person), making the task of predicting where the user will watch easy: an angular sector can be identified (as there is a single or few RoIs), and will remain the same over time. In *Moving focus*, the RoIs move over the sphere and hence the angular sector where the FoV will be likely positioned changes over time. *Rides* videos are characterized by the attracting angular sector being the direction of the camera motion which is substantial.

[1]Hereafter, we use the term "scene" as defined in [6] as a period of the video between two edits with space discontinuity.

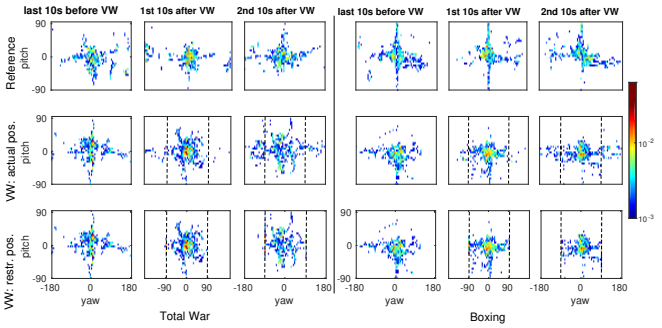

Figure 1: Heat maps of most utilized yaw and pitch angles in 3 periods (columns). The visible sector of the VW is the 180° span between the vertical dotted lines.

The rule-of-thumb so far (see, e.g., [5, Sec. 3] or the Oculus Rift developer guidelines [9, p. 5]) has been to always send something for the user to watch in any part of the sphere. In this article, we hypothesize that it is possible to restrict temporarily the angular sector the user can access, in order to save sending some part of the sphere and being able to increase the quality in the accessible sector. We therefore investigate if and how, after the exploration phase, in *static focus* and *ride* scenes, placing a so-called Virtual Wall can improve QoE compared with a higher compression factor.

Definition: We define a Virtual Wall (VW) as a restriction of the visible angular sector. There are many strategies for restricting the visible part of the sphere to prevent the transmission of parts of the content. The most obvious is to replace not transmitted sphere sectors with black patches. However, entering black areas may be both (i) utterly unsettling and risking the user to lose their footing (as when one closes their eyes while standing up), and (ii) consciously perceived as an unacceptable (cheap) solution. That is why we design the VW effect as a subtle degradation of the user interaction with the content: when the longitude of the user's position reaches the limit of the visible sector, the FoV only refreshes in latitude until the user comes back in the visible sector. As only the longitude is affected, the user does not risk to lose her footing. Fig. 1 (e.g., Col. 2-3) shows the impact of the VW on the accessible FoV positions.

Implementation: An XML file describes the periods and angular sectors where to position VWs. The Android player is modified to monitor the head position and feed the video renderer with the desired FoV, corresponding to either the actual head position if in the visible sector, or the wall position if the head is outside.

4 EXPERIMENTAL DESIGN

The user experiments aim at testing the following hypotheses:
[H1] If the VW is positioned after the exploration phase in scenes with concentrated saliency (*static focus* and *rides*), a substantial fraction of users will seldom perceive it.
[H2] Compared with a reference with no VW consuming the same amount of data, VW can improve the user's QoE (by increasing the visual quality without impairing the experience), if placed in contents and periods where the users are much likely to focus on a known region.

The users are hence presented with two versions of each video, translating the same bandwidth budget into different impairments:

Class	Scene, duration	VW period	C	D	E
Ride	F1, 31s	18s-31s	10	5	10
Ride	Trike, 51s	25s-51s	10	5	10
Ride	Assassin, 51s	20s-46s	10	5	10
Ride	Total War, 42s	22s-42s	10	5	10
Static focus	Boxing, 85s	25s-85s	12	3	6

Table 1: Description of videos (scenes from [1], classes, encoding rates). Column C: Video encoding rate (Mbps) outside the VW period in both reference and effect versions. Column D: Video encoding rate (Mbps) inside the VW period in the reference version. Column E: Video encoding rate (Mbps) inside the VW period in the effect version.

the *reference version* has the video coding rate reduced (legacy adaptive streaming approach) while the *effect version* displays high quality at the expense of reducing the freedom of the user with a VW. We use a double-stimulus approach following the guidelines in [11] to identify which type of impairment yields the higher user's QoE. The details of the encoding and VW are provided next.

4.1 Videos to assess

The video scenes and their features are detailed in Table 1. All the scenes are freely available on Youtube with the IDs listed in [1, Table 1]. The VW effect is tested on 5 scenes corresponding to two videos: *Comb. Rides* compiles four scenes classified as Rides, and *Boxing* is made of one scene classified as Static focus in [1, Table 1]. All VWs are positioned after the exploration phase, i.e., after about 20s of the start of the scene, and last for a few tens of seconds until the end of the scene. Their angular sector is 180° in longitude (i.e., in yaw angle). No restriction is made on latitude (i.e., in pitch angle) to maintain balance. The video encoding rates are chosen to be representative of a network scenario where VW could be triggered as an alternative lever to quality degradation: if a bandwidth drop happens in a focus phase. For the user experiments, we consider two qualities and a simple case where the bandwidth allows streaming the highest quality for the entire sphere then drops to a level allowing to fetch the low quality for the entire sphere, or the higher quality in some part only. We hence position the VW at this time and accordingly choose the encoding ratio between high and low quality with the same factor as the reduction in the visible sector with the VW, that is 2 in our experiments (360° to 180° and see col. D-E in Table 1). The reference version therefore displays low quality over the whole sphere during the VW period, while the effect version displays high quality in the restricted visible sector.

4.2 Experimental procedure

The videos are described with MPEG-DASG SRD [8], tiled in 3x3 and played in the Samsung Gear VR headset with Samsung S7 Edge phones. The 360° video streaming player is the Android app made available in [3] and adapted to implement VW. The users are repositioned at the center of the visible sector at the beginning of each VW with the technique introduced in [3]. We control the displayed qualities over time by restricting to a single one the representation available for each segment in the manifest file.

We recruited 18 users using a convenience sample. Exact gender-balance was met. After a video to get familiar with the gear and the virtual environment, the users were presented back-to-back with the effect and reference versions of each video. The order of the videos was that of Table 1, while the order of the versions was picked randomly, established prior the experiments for all the user indexes by drawing a Bernouilli random variable. Using a scale from 1 (the worst) to 5 (the best), they were asked to rate each version of the video w.r.t.: the *visual quality* of the video, the *responsiveness* of the system to their head motion and their *comfort*. After seeing the second version of the video, they indicated which version they did prefer. As also considered in [10], the videos were watched standing up in order not to restrict motion, with the back of a chair in reach to keep balance if needed.

We also embedded logging threads into the player to collect objective measurements of the user's motion. In particular, we recorded the head position (yaw and pitch angles). When VWs were active, we also recorded the positions of the displayed FoV, as well as the actual head position, enabling us to compute the depth of each hit, the hit duration, and the number of hits.

5 EXPERIMENTAL RESULTS

We first present how the users interact with the VW by analyzing log data. We confirm that, when placed appropriately, the users seldom sense the VW, even more so in high camera motion rides (88% sense the VW at most twice). We then analyze the subjective ratings showing that a majority of users prefer the VW version, and we identify how the importance of quality and responsiveness in the preference depends on the scene category.

5.1 Impact of VW onto users' behavior

The logs are analyzed for 16 of the 18 users (due to a technical issue with the first 2 users, but below subjective ratings are for 18). First, Fig. 2, which depicts how many times did users hit each VW (one in each scene), allows to confirm our first hypothesis [H1]: in all scenes, at least 50% (9 in 16) of the users did not hit a wall more than twice, and 88% (14 in 16) in the ride scenes characterized by a substantial camera motion (F1, Assassin's Creed and Total War).

Fig. 3 depicts the time of hit (counted from the start of the VW) depending on its order. Interestingly, we observe that much fewer hits happen in the first 10 seconds after the onset of the VW than in the subsequent 10 seconds. In all scenes but F1, there is often at most one hit in the first 10 seconds (and 75% of users in F1 experience at most a hit). This leads to hypothesize that the user's attention is more focused in the first 10 seconds, where a VW therefore go mostly unnoticed. This is confirmed by Fig. 1 depicting the heat maps of two scenes (Total War and Boxing, the others show similar features). The top row represents the reference version, for which the head position and FoV position are the same (no VW). For the effect version, the middle and bottom rows represent the head position and the FoV position, respectively. The first column allows to verify the similarity of the head motion distribution before the wall starts, in the reference and effect versions. The wall limit is depicted by vertical dotted lines, and we indeed observe that in the first 10 seconds after the wall (middle column), only a few instances of the head position are beyond the wall. In the next 10 seconds however

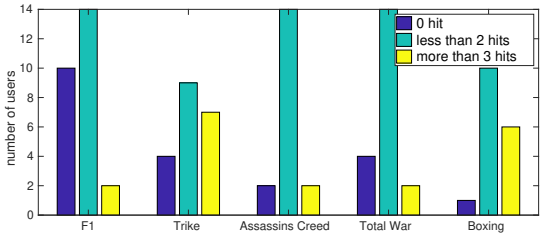

Figure 2: Histogram of the number of users (total of 16) never hitting the wall, at most twice or more for each scene.

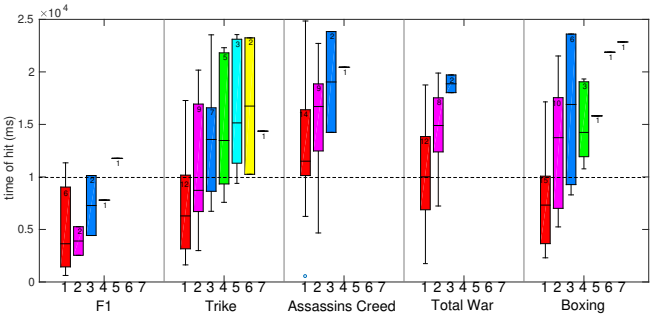

Figure 3: Boxplots of hit times, in order of occurrence. The number of samples is indicated in each bar.

(last column), the head position is more often beyond the wall, the FoV position blocking at the wall limit. From this we can extract a guideline to position a VW if needed: wait about 20s for the exploration phase of the current scene to be over, do not make the wall last more than 10 seconds if possible. This visible sector reduction over 10 seconds shall already give helpful slack to the streaming algorithm when the bandwidth is too low to stream HQ or allow replacements, as shown in Sec. 6.

5.2 Analysis of subjective ratings

Fig. 4.a represents the fraction of users having declared preferring the effect version. The data is fitted to a Bernoulli distribution, whose 90% confidence interval on the probability parameter is represented on each bar. Fig. 4.a shows that a majority of users tend to prefer the version with increased quality and VW: about 58% in Comb. Rides and 68% in Static focus (Boxing). Despite the large confidence margin (including the 0.5 level) not allowing to formally conclude on hypothesis [H2], analyzing how does the preference depend on visual quality and responsiveness enables to explain the difference between the scene categories and extract potential improvements. Fig. 4.b (resp. 4.c) depicts the fraction of users preferring the effect version with respect to the score they have given the visual quality of the reference version (resp. w.r.t. the responsiveness score given to the effect version). On the one hand, the user's preference is negatively correlated with the visual quality score of the reference decreases for both Comb. Rides and Boxing. Also, the marginal probability that a user rates the reference version with poor scores is substantially higher: Fig. 5.a reveals that as much as 40% of users rate the visual quality of the reference with a score lower than 3 for Comb. Rides, vs. 10% for the effect version (and ca. 30% vs. 0%

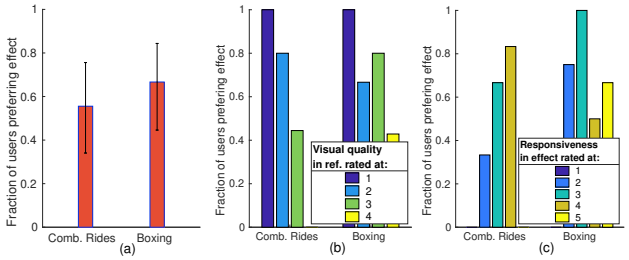

Figure 4: (a): Fraction of users preferring the VW effect over the reference for each video clip. (b): Fraction of users preferring VW over ref. conditionally to their visual quality score of the ref. version. (c): The same, but conditionally to the responsiveness score given to the effect version.

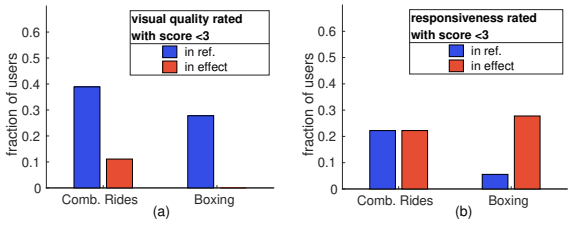

Figure 5: Fraction of users rating each version of each video with a score lower than 3 for (a) quality, (b) responsiveness.

for Boxing). Therefore, the preference is strongly correlated with the visual quality *and* the visual quality of the reference is rated low by many users. On the other hand, Fig. 4.c shows that the preference is clearly correlated with the responsiveness score (of the effect version) only for Comb. Rides, in which case Fig. 5.b shows that the marginal probability of a low responsiveness score is equivalent between both version. This explains why the responsiveness score (representing the sense of impairment brought by the VW) has an overall importance in the preference lower than the visual quality. This analysis is therefore another element supporting hypothesis [H2]: users prefer the effect version because they perceive more the lower visual quality than the presence of a VW. The open comments made appear that the implementation of the VW can be enhanced in Rides, as the camera motion worsens the feeling of the VW. A possible solution is implementing a slow-down of the playback based on the FoV position, and is part of future work.

Finally Fig. 6.a shows the quality score distributions obtained by both versions as boxplots, with the advantage of the effect version. Fig. 6.b shows as expected that the effect version obtains lower responsiveness scores (the stronger difference for Static focus corroborating the higher number of hits than in Comb. Rides). However, it is interesting to see in Fig. 6.c that, despite the lower responsiveness, the users did not rate their overall comfort lower in the effect version than in the reference: this demonstrates that the VW effect is acceptable, and is hence a valid solution to help maintain visual quality in a context of degraded bandwidth.

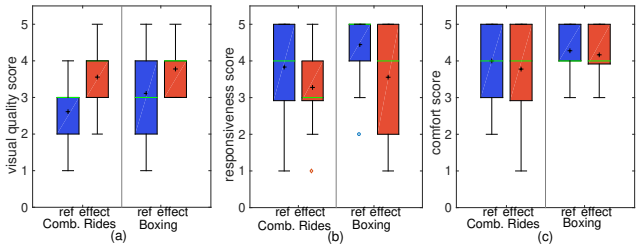

Figure 6: Boxplots of scores in (a) quality, (b) responsiveness, (c) comfort. The greater, the better. The black cross marks the average, the green line the median.

6 ADAPTIVE STREAMING STRATEGIES WITH VIRTUAL WALLS

The goal of this section is to assess how much gain can a VW bring to and compared with reference FoV-based adaptations. We express the streaming decision as an optimization problem, design heuristics using pyramidal pre-fetching based on user's FoV, and propose two competitors to assess the gains brought by VW in terms of quality in the FoV and stalls. We simulate the streaming performance using the head motion traces collected in the experiments without and with VW. We show that, even in the presence of substantial playback buffers and with FoV-based adaptation, VW enables reduction in stalls and maximizes quality in FoV, confirming the interest of the method both from a QoE and system point of view.

6.1 Problem description

The download decisions are taken at most each Δ_{Dl} seconds and aim at maximizing the expected quality in the FoV by adapting online to the FoV while ensuring some level of playback buffer to avoid stalls as much as possible. The optimization problem is formulated in Problem 1. Notation is provided in Table 2.

$$\max_{\{x_{ijl}\}} \sum_{i=1}^{M} \sum_{j=j_t}^{j_t+K} q_l p_i(j) x_{ijl}, \quad \text{s.t.} \tag{1a}$$

$$x_{ijl} = 0, \quad \forall i,j,l : buf_i(t) \geq B_{max} \text{ or } j < j_{t_i} \tag{1b}$$

$$buf_i(t) - \Delta_{Dl} + \sum_j \sum_l x_{ijl} \Delta_{Dl} \geq B_{min}, \quad \forall i \in \mathcal{M} \tag{1c}$$

$$\sum_{i,j,l} x_{ijl} s_{ijl} \leq C_t \Delta_{Dl}, \quad \sum_l x_{ijl} \leq 1, \quad \forall i \in \mathcal{M}, \forall j \in [j_t, j_t + K] \tag{1d}$$

$$\sum_l x_{ijl} \leq \sum_l x_{i(j-1)l}, \quad \forall i \in \mathcal{M}, \forall j \in [j_t + 1, j_t + K] \tag{1e}$$

$$x_{ijl} = 0, \quad \forall i,j,l : j \in \mathcal{W}_{per} \text{ AND } i \notin \mathcal{W}_{angle} \tag{1f}$$

The parameters $j_{t_i}, \forall i \in M$, denote the first segment index not in tile i's buffer, and $j_t = \min_i j_{t_i}$. Constraint (1b) ensures that only tiles with non-full buffers are scheduled for download, that a segment already there is not downloaded again and that no segment beyond the lookahead window is fetched; (1c) aims at maintaining a minimum buffer level for each tile; (1d) enforces the bandwidth limitation and ensures that at most one quality is chosen for each piece to download, (1e) ensures that segments closer in time are given priority.

Parameter	Definition
M ($\mathcal{M}$), L ($\mathcal{L}$), J	number (set) of tiles, quality levels, segments
Δ_{Dl}	minimum period between 2 download decisions
K	look-ahead window in number of segments
q_l	quality rating of level l
$buf_i(t)$	num. of sec. stored in buffer of tile i at time t
$p_i(j)$	proba. that tile $i \in$ FoV at segment j
s_{ijl}	size (in B) of tile i of seg. j at quality level l
C_t	estimated bw. for download from t for dur. Δ_{Dl}
B_{min}, B_{max}	min and max buffer size
$\mathcal{W}_{per}$ ($\mathcal{W}_{angle}$)	set of seg. (tiles) in a VW period (visible sector)
Decision var.	indicates whether tile i of seg. j is scheduled
$x_{ijl} \in \{0,1\}$	for download, $\forall i \in \mathcal{M}, j \in \mathcal{J}, l \in \mathcal{L}$

Table 2: Parameters and variable of the optimization problem

Finally, constraint (1f) is active when one or more VWs are positioned along the video: the segments in the corresponding period and outside the pre-determined visible sector are not required.

6.2 Algorithms

Targeting an implementation at the client (on the phone), we design a heuristic algorithm to solve Problem 1. This heuristic is named *Hwall* and described in Algo. Let $dist(FoV(t), i)$ denote the distance between the current FoV and the center of tile i. At time t, $p_i(j)$ is estimated and updated with

$$p_i(j) = \frac{\left(\max_{i \in \mathcal{M}} dist(FoV(t), i)\right) - dist(FoV(t), i)}{\sum_i \left(\left(\max_{i \in \mathcal{M}} dist(FoV(t), i)\right) - dist(FoV(t), i)\right)}.$$

We define $j_{i,min}$ so that $[j_{t_i}, j_{i,min}]$ is the minimum set of segments that must be downloaded from t to ensure that constraint (1c) is satisfied (i.e., $buf_i(t + \Delta_{Dl}) \geq B_{min}$). We hence typically (but not always) have $j_{t_i} \leq j_{i,min} \leq j_t + K$.

To compare *Hwall* with FoV-based streaming strategies not involving VW, we design *Href1* and *Href2* from the same logic. *Href1* is the same pyramidal strategy based on the current FoV but without any consideration of VW. It is hence described with Algo. 1 without reference to constraint (1f) in line 1 nor the *if* statement in line 6. *Href2* is meant to be less conservative by considering the knowledge of the VW position, i.e., the highest saliency region, and forcing to download high quality in this region. It is described with Algo. 1 without reference to constraint (1f) in line 1.

6.3 Simulations

Parameter K therefore tunes how conservative (trying to fill the buffer quickly) or aggressive (being more responsive to the head motion by fetching high quality at the expense of a lower buffer) the adaptation is. To assess how much gain can a VW bring to and compared with responsive reference FoV-based adaptations, we set $\Delta_{Dl} = 2$, $K = 2$, $B_{min} = 3$ and $B_{max} = 20$ for *Hwall*, *Href1* and *Href2*. A segment is one-second long, and a startup buffer of 10s is allowed to build up with low quality tiles before the playback starts. We set $L = 2$, $q_l = l$ for $l = 1, 2$ (). We present the results obtained as time series of metrics of interest for user 5 and the Boxing video ($J = 53$, results are qualitatively equivalent for the other cases).

Fig. 7 represents the typical network scenario for which the VW tool has been designed: upon sensing a bandwidth drop, the adaptation logic decides to trigger a VW alternatively to dropping the

Algorithm 1: Streaming decisions with heuristic *Hwall*

Data: Buffer states $buf_i(t), \forall i \in \mathcal{M}$
Result: $\{x_{ijl}\}, \forall i \in \mathcal{M}, l \in \mathcal{L}, j = j_t, \ldots, j_t + K$
For all i, j verifying constraints (1b) and (1f), allocate highest quality: $x_{ijL} = 1$;
Compute requested data: $data = \sum_{ijl} x_{ijl} s_{ijl}$;
$j = \min(j_t + K - 1, J)$;
while $data > C_t \Delta_{Dl}$ *AND* $j \geq j_t$ **do**
  for *i in descending order of distance to FoV(t)* **do**
    if $j \notin \mathcal{W}_{per}$ *OR* $i \notin \mathcal{W}_{angle}$ **then**
      if $j > j_{i,min}$ **then**
        decrease quality or cancel download if quality already minimum;
        update $data$;
        if $data \leq C_t \Delta_{Dl}$ **then**
          break;
      else if $j_{t_i} \leq j \leq j_{i,min}$ **then**
        decrease quality if not yet minimum;
        update $data$;
        if $data \leq C_t \Delta_{Dl}$ **then**
          break;
  $j = j - 1$;
if $j < j_t$ *AND* $data > C_t \Delta_{Dl}$ **then**
  break constraint (1c) and defer the download of as many segments as needed verifying $j > j_t$ (at least the next is kept scheduled), in descending order of playback position and distance to FoV(t)

quality (as *Href1* does) or undergoing stalls (as *Href2* does). The buffers built during the startup period allow to download high quality tiles, at the cost of decreasing the buffer. *Href1* is designed to be more conservative and we indeed observe (middle row) that it is able to maintain a relatively high buffer compared with *Href2* and *Hwall*. *Href2* is designed to systematically fetch high quality in the wall's sector. Contrary to *Hwall*, it also needs to fetch at least low quality outside the wall's sector. This amounts to excessive data with respect to what the bandwidth allows, thereby yielding stalls. Despite the download of high quality in the wall's angle, the quality in FoV of *Href2* in the wall's period is not perfectly 2 because the user's position is not constrained. Finally, *Hwall*, by not having to download the non-visible wall's sector, is able to achieve quality 2 in the FoV without any stall.

7 CONCLUSION

Building on the recent characterization of human attention in VR, we have introduced Virtual Walls, a new degree of freedom for the streaming adaptation. VW translates bandwidth limitation into a new type of impairment allowing to preserve the visual quality. User experiments have confirmed that (i) if the VW is positioned after the exploration phase in scenes with concentrated saliency, a substantial fraction of users seldom perceive it, and that (ii) compared with a reference with no VW consuming the same amount of data,

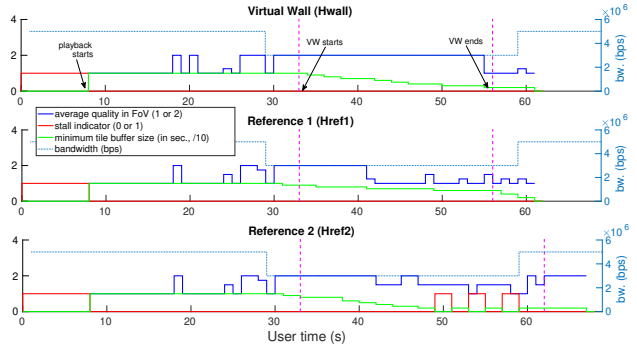

Figure 7: Time series for *Hwall*, *Href1*, *Href2*. Purple dotted lines mark the VW period (set in video time, shown in user time).

VW can improve QoE. Simulations have shown that VW can advantageously complement FoV-based adaptations, enables reduction in stalls and increases quality in FoV. Deciding when to trigger a VW instead of decreasing the fetched qualities, and with which parameters (duration and visible angular sector) should be made based on (i) the network (past bandwidh samples and playback buffers), (ii) the content (scene categories), and (iii) the user's state (whether she is in an exploratory or focusing phase). The design of such strategies is left for future work but will be key in future systems, particularly for future headsets with significantly increased resolution (such as the Varjo with 50 megapixels per eye).

ACKNOWLEDGMENTS

We are grateful to Antoine Dezarnaud and Daniela Trevisan for their valuable input. This work has been supported by the French government, through the UCA JEDI and EUR DS4H Investments in the Future projects ANR-15-IDEX-0001 and ANR-17-EURE-0004.

REFERENCES

[1] M. Almquist, V. Almquist, V. Krishnamoorthi, N. Carlsson, and D. Eager. 2018. The Prefetch Aggressiveness Tradeoff in 360 Video Streaming. In *ACM MMSys*.
[2] E. Bastug, M. Bennis, M. Medard, and M. Debbah. 2017. Toward Interconnected Virtual Reality: Opportunities, Challenges, and Enablers. *IEEE Comm. Mag.* 55, 6 (June 2017), 110–117.
[3] S. Dambra, G. Samela, L. Sassatelli, R. Pighetti, R. Aparicio, and A.-M. Pinna. 2018. Film Editing: New Levers to Improve VR Streaming. In *ACM MMSys*.
[4] E. J. David, J. Gutiérrez, A. Coutrot, M. Da Silva, and P. Le Callet. 2018. A Dataset of Head and Eye Movements for 360 Videos. In *ACM MMSys*.
[5] M. Graf, C. Timmerer, and C. Mueller. 2017. Towards Bandwidth Efficient Adaptive Streaming of Omnidirectional Video over HTTP: Design, Implementation, and Evaluation. In *ACM MMSys*. 261–271.
[6] J. Magliano and J. M. Zacks. 2011. The Impact of Continuity Editing in Narrative Film on Event Segmentation. *Cognitive Science* 35, 8 (2011), 1489–1517.
[7] A. Nguyen, Z. Yan, and K. Nahrstedt. 2018. Your Attention is Unique: Detecting 360-Degree Video Saliency in Head-Mounted Display for Head Movement Prediction. In *ACM Multimedia Conference*. 1190–1198.
[8] O. Niamut, E. Thomas, L. D'Acunto, C. Concolato, F. Denoual, and S.-Y. Lim. 2016. MPEG DASH SRD: Spatial Relationship Description. In *ACM MMSyS*.
[9] Oculus. 2017. Oculus Best Practices. Version 310-30000-02.
[10] V. Sitzmann, A. Serrano, , A. Pavel, M. Agrawala, D. Gutierrez, B. Masia, and G. Wetzstein. 2018. Saliency in VR: How Do People Explore Virtual Environments? *IEEE Trans. on Vis. and Comp. Graphics* 24, 4 (April 2018), 1633–1642.
[11] International Telecommunication Union. 2012. Methodology for the subjective assessment of the quality of television pictures. Rec. ITU-R BT.500-13.
[12] M. Xiao, C. Zhou, V. Swaminathan, Y. Liu, and S. Chen. 2018. BAS-360: Exploring Spatial and Temporal Adaptability in 360-degree Videos over HTTP/2. In *IEEE INFOCOM 2018*. 953–961.