Methodology to Assess Quality, Presence, Empathy, Attitude, and Attention in 360-degree Videos for Immersive Communications

Marta Orduna, Pablo Pérez, Jesús Gutiérrez, and Narciso García

Abstract—This paper proposes a methodology to assess video quality, spatial and social presence, empathy, attitude, and attention in 360-degree videos for immersive communications. The methodology is validated in an experiment which simulates an immersive communication environment where participants attend three conversations of different genre (everyday conversation, educational, and discussion) and from actor and observer acquisition perspectives. We consider three experimental conditions: (A) visualizing and rating the perceptual quality of contents in a Head-Mounted Display (HMD), (B) visualizing the contents in an HMD, and (C) visualizing the contents in an HMD where participants can see their hands and take notes. In all conditions participants visualize the same 360-degree videos, designed and acquired in the context of international experiences. Fifty-four participants were evenly distributed among A, B, and C conditions taking into account their international experience backgrounds (working or studying in a foreign country), obtaining a balanced and diverse sample of participants. In this paper, video quality is evaluated with Single-Stimulus Discrete Quality Evaluation (SSDQE) methodology. Spatial and social presence are evaluated with questionnaires adapted from the literature. Initial empathy is assessed with Interpersonal Reactivity Index (IRI) and a questionnaire is designed to evaluate the attitude after the visualization of each video. Attention is evaluated with three questions about the conversations of the contents that had pass/fail answers. The results from the subjective test validate the proposed methodology in immersive communications, showing that video quality experiments can be adapted to conditions imposed by experiments focused on the evaluation of socioemotional features in terms of contents of long-duration, different acquisition perspectives, and genre. In addition, the positive results related to the sense of social and spatial presence imply that technology can be relevant in the analyzed use case. Other main result is that the acquisition perspective greatly influences social presence. Finally, the annotated dataset, Student Experiences Around the World dataset (SEAW-dataset), obtained from the experiment is made publicly available for the research community.

Index Terms—Quality of Experience, Video Quality, Subjective Assessment, Presence, Social Presence, Spatial Presence, Empathy, Attitude, Attention, Virtual Reality, Immersive Communications, 360º video, 360-degree video

✦

1 INTRODUCTION

Virtual Reality (VR) is an emerging field that is achieving great interest in applications for social purposes, such as assistive, entertainment or educational applications, and also in teleconferencing scenarios [1], [2].

The main reason is that the use of real-time 360-degree video provides additional value as a communication platform that goes one step beyond traditional audio and video transmission [3]. Typically, this type of platform is based on the architecture shown in Figure 1, where 360-degree video and audio are transmitted from a particular location to a remote user, and are displayed by a Head-Mounted Display (HMD). It allows the transmission of additional non-verbal signals such as facial expressions or body postures that are exchanged during a conversation, which greatly improve the effectiveness of face-to-face communications [4]. On the other side, the remote user application displays some

- M. Orduna, J. Gutiérrez and N. García are with the Grupo de Tratamiento de Imágenes, Information Processing and Telecommunications Center and Escuela Técnica Superior de Ingenieros de Telecomunicación, Universidad Politécnica de Madrid, Madrid 28040, Spain. E-mail: {moc, jgs, narciso}@gti.ssr.upm.es.

- P. Pérez is with Application Platforms and Software Systems Labs, Nokia Bell Labs, Madrid 28050, Spain. E-mail: pablo.perez@nokia-bell-labs.com
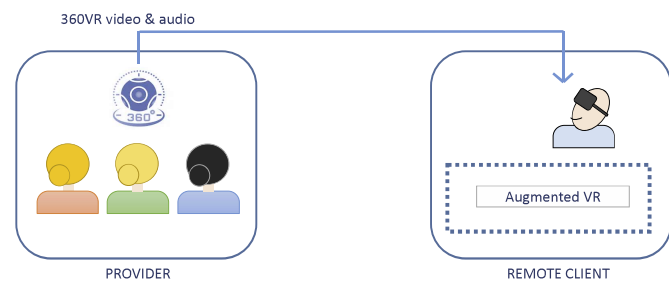
Fig. 1. Simulated immersive communication environment of the experiment

augmented information and provides interactivity using VR controllers [5] or the user's hands [6], [7]. Besides, a return channel (not shown in the figure) normally exists, but its implementation varies significantly between different works, mainly depending on the use case.

As a conclusion, VR provides more immersive environments and interactive experiences than today's communications technology to the user who attends the conversation with an HMD [8]. Thanks to these engaging environments, users can evoke psychological effects such as a sense of presence or other affective skills that we refer to as socioemotional features [9].

In order to satisfy users' demands and expectations, it is essential to study and guarantee a high Quality of Experience (QoE), which is affected by technical parameters but also by socioemotional features. In reference to the technical features, omnidirectional content is much more demanding than traditional content [10]. Specifically, a higher resolution than 2D videos is required to provide similar video quality due to the fact that the pixels are distributed in a 360-degree sphere. Also, higher framerates, ideally equal to the refresh rate of the HMDs used for the visualization, are required to avoid annoyance for participants. These requirements dramatically increase bandwidth. In this sense, several works in the literature have explored techniques to save bandwidth while offering acceptable video quality. Most approaches are based on non-uniform schemes, encoding and transmitting with higher quality only the field of view that the user is visualizing [11]. Other works are based on the fact that users tend to look at certain parts of the scene that are more attractive. In these cases, saliency or attention maps are computed to efficiently distribute the bitrate [12].

The evaluation of the video quality achieved in these solutions or, in general, in the transmission of 360-degree video, is typically performed using subjective assessment tests based on methodologies highly proven with traditional contents [13], [14], [15], and recently adapted to immersive video [16]. Commonly, the stimuli used with these methodologies are short-duration videos without narrative, which are randomly displayed in different qualities and consumer devices to be rated. The problem is that this kind of methodologies has not been designed taking into account the requirements of socioemotional aspects. The analysis of empathy, spatial and social presence, attention or other affective skills that are experienced in a VR scenario requires long-duration videos with a narrative and genre adapted to the purposes of the experiment [17], [18], [19], [20]. The whole purpose of using VR for teleconferencing is benefiting from its higher immersion and sense of presence; if the quality evaluation test does not provide such features, then its results might be not valid for the desired use case.

Considering the current situation with the COVID-19 pandemic and the relevance of teleconferencing scenarios, VR technology can foster a change in communications. However, it is necessary to further investigate and provide standardized methodologies to consider all aspects that influence the QoE for the final boost of this technology [21], [22]. Once there is literature that analyzes different socioemotional and technical aspects, an important advance should be to evaluate them together in experiments closer to real scenarios and use cases, increasing ecological validity and reliability. In addition, experiments that consider aspects that have already been independently evaluated saves time and resources. So, in this paper, we not only present an experiment with a methodology designed to evaluate both technical and socioemotional aspects; we launch a renewed point of view: how the evaluation of technical aspects influences socioemotional aspects, and vice versa, accelerating immersive communications as a solution to the current situation.

1.1 Contributions

This paper contributes to the fields of affective computing and quality of experience, providing an experiment where video quality and socioemotional aspects are jointly addressed. Here, we present in detail our main contributions:

- **Methodology**. We propose and validate a methodology to jointly assess video quality and presence, empathy, attitude, and attention in immersive communications. This methodology is a solution for experiments in a controlled environment but more realistic than those presented in the literature. We propose the use of Single-Stimulus Discrete Quality Evaluation (SSDQE) method to measure the quality during the test session and the aggregate quality in a post-questionnaire using the Absolute Category Rating (ACR) on the same five-grade scale. Spatial and social presence are evaluated with an aggregate score obtained from 5 items based on the literature and adapted to our experimental environment. The initial empathy is evaluated using the Interpersonal Reactivity Index (IRI). The attitude is measured in pre-questionnaire and post-questionnaire designed using facet theory. Due to the reliability of the scale, we propose to use only the post-questionnaire. The attention is addressed with three questions about the scene that have pass/fail answers.
- **Video quality assessment in immersive communications**. We propose and verify an assessment of video quality for immersive communications using long-duration videos specifically designed and acquired for the exploration of socioemotional contents.
- **Dataset**. We make publicly available a Student Experiences Around the World dataset (SEAW-dataset) of 3 video sources (stereoscopic raw format) designed and acquired specifically for the purposes of the experiment. During the recording we considered three genres and both actor and observer acquisition perspective in the same context, international experiences, working or studying in a foreign country. Additionally, the questionnaires and the associated rates obtained from a diverse and balanced sample of 54 participants are provided.

The rest of the paper is structured as follows. Firstly, an overview of related works is presented in Section 2. Then, Section 3 presents the research questions. Section 4 explains the main features of the experiment design: experimental conditions, test material, methodology, scenario, test session, and observers. The experimental results are presented in Section 5 and finally, Section 6 includes general conclusions.

2 RELATED WORK

The studies conducted in the literature present limitations that influence the results and conclusions and should be considered for the design of the experiment. In this section, we present an overview of the works mainly related to the quality and socioemotional features assessment.

TABLE 1
Summary of datasets of 360-degree videos and evaluations

Authors	Number of 360-degree videos	Participants	Socioemotional aspect	Additional data
Li et al. [23]	73	95 (56 female, 39 male)	Valence and arousal	Participants' rotational head movements
Jun et al. [24]	80	551 (247 female, 262 male, 2 other)	Presence, arousal, simulator sickness	Future use intention of the video. Tracking data of the VR headset
Corbillon et al. [25]	70	59		Tracking data of the VR headset
Lo et al. [26]	10	50		Tracking data of the VR headset. Image saliency map from videos
David et al. [12]	19	57 (25 female)		Tracking data of the head and eye movement. Statistics related to exploration behaviors
Yang et al. [27]	13	30		Mean opinion scores

2.1 Quality Evaluation

One of the main features to take into account during the design of subjective experiments is the test content. Despite the increase in consumption and therefore the creation of 360-degree content, high technical requirements are necessary for this kind of experiments. For example, problems caused during video acquisition or post-processing (e.g., stitching errors) or audio artifacts can influence the quality evaluations and affect the understanding of the content narrative. Another aspect to take into account when selecting 360-degree content is its characterization in terms of exploration properties [28]. Generally, contents can be classified as directed or exploratory. Directed videos can help the observer to guide attention in the scene. Although participants move freely around the scene in exploratory contents, most of them fully explore the whole scene (360°) in 20 seconds [24]. Nevertheless, contents of long-duration can improve the engagement and enhance the emotions of the participants [20], [29]. In addition to the duration, the genre and context of the video influence the success of the research. Specific genres of content should be considered based on the socioemotional features addressed in the experiment [24], e.g: horror stimulus to test fear [30]. Taking this into account, we examined some 360-degree datasets in the literature with different characteristics, summarized in Table 1. For example, Li et al. [23] released a public database of 360-degree videos covering a wide range of arousal and valence. Also, Jun et al. [24] published a dataset containing 80 videos that were used to investigate a set of socioemotional features with a sample of 551 participants. They provided video sources with the corresponding report ratings and head movements. In addition, there are several datasets created to analyze exploration behaviors of the users when watching the content, such as the ones from Corbillon et al. [25] and Lo et al. [26] providing also head-movement data, or the one from David et al. [12] that includes both head and eye tracking data. Regarding quality evaluation, some annotated datasets have been published, mainly containing short-duration videos, such as the one from Yang et al. [27].

The use of short-duration videos is a common approach on audiovisual quality evaluation, which is supported by several international recommendations related to subjective quality assessment, such as ITU-T P.910 and P.913 [14], [15], [16]. In these recommendations, standard assessment methodologies are proposed to subjectively evaluate the impact of typical video artifacts (e.g., coding degradations) and also the guidelines for data processing. For each video artifact, the mean of the evaluations of the observers with the associated Confidence Intervals (CI) are computed to analyze the distribution of the means and their cumulative frequency of appearance [14]. In the case of video quality, means are called Mean Opinion Score (MOS) and typically, are presented with the associated 95% CIs. As these methodologies have been highly tested in the literature with 2D video, the distribution obtained with the representation of the MOS with the associated CIs of video quality evaluations helps the researcher to validate her/his experiment. Generally, it is more difficult for observers to appreciate the differences between very high quality content. However, in the videos encoded with intermediate qualities, the observers are able to find differences, but when the video quality is very low and annoying artifacts appear, the ratings saturate. These methodologies, which were originally designed for 2D video, have been used in 360-degree video experiments and, somehow, adapted to address the new perceptual factors involved in VR [31], [32], such as simulator sickness or exploration behavior. However, there is a research line supporting that quality assessment should be done under the most realistic conditions when services and applications are addressed to end users [33], [34].

Based on the constraints presented, mainly related with content, methodologies, and context, we decided the design and acquisition of the 360-degree contents taking into account the purposes and the final devices used in the experiment [30]. With this, we propose a quality evaluation on long-duration videos with a context that interests or affects the participant and with a genre selected according to the purpose of the research. Also, we choose an environment where the participant is isolated, facilitating the real world disassociation.

2.2 Socioemotional aspects evaluation

Due to the limitations of the traditional QoE assessments, a great effort has been made in the analysis of socioemotional features in VR. Riva et al. [35] demonstrated the effectiveness of VR as an *affective medium*, a medium able to elicit different emotions through the interaction with its contents. Furthermore, the study demonstrated that the perceived sense of presence, related to a sense of being in a place [36], influences the emotional state. Following this research line, many studies have already confirmed the ability of VR to create more immersive environments, improving the socioemotional features. For instance, Fonseca et al. [20] demonstrated the highest emotional involvement of the participants viewing two types of narrative 360-degree contents with an HMD. MacQuarrie et al. [30] obtained a significant improvement of enjoyment of users using the HMD. In addition, VR emphasizes the phenomenon called Fear of Missing Out (FoMO) [37], defined, in the context of VR, as the apprehension that others might be having rewarding experiences from which the user with the HMD is absent. FoMO and the fact that users can freely move around omnidirectional scene can influence on the attention.

TABLE 2
Technical specifications of the test material used in the pilot study [41]

Source content	Resolution	Framerate (fps)
Alento	3840x1920	25
AngelFalls	3840x2160	30
Flamenco	3840x2160	30
LionKing	3840x2048	30
Lions	3840x1920	30
SwissJet	3840x1920	50

Some works in the literature analyze methods for assessing attention in this kind of environment [38], [18].

Several works go one step further analyzing the use of this technology for empathy purposes and even for behaviour change purposes. Empathy is defined as the ability to view the world from another person's perspective combined with an emotional reaction to that perspective, including feelings of concern for others [39]. These studies are based on the fact that involvement created by VR environments facilitates empathy for users and can be used for specific purposes [17]. Aitamurto et al. [40] evaluated the responsibility for resolving gender inequality visualizing a 360-degree content in which participants could choose to watch the narrative from the male or female character's perspective. Likewise, Tussyadiah et al. [19] confirm the effectiveness of VR technology in shaping consumers' attitude and behavior for tourism purposes.

Most of the literature can be divided into two main areas. One area focuses on the analysis of specific socioemotional features or a small subset of these independently. Many of those works do not address technical features such as resolution, framerate, or encoding parameters of the video [40]. The other area focuses mainly on technical parameters and only some socioemotional aspects are evaluated [42], [43]. We propose a methodology, following the experience of the literature, to assess video quality and several socioemotional features in the same experiment, reporting technical features, questionnaires, and sample diversity.

2.3 Pilot study

To further examine the findings from the literature, we conducted a pilot study where the influence of the HMD, usability, and fatigue in 360-degree video quality assessments were examined [32]. The equipment used in the experiment consisted of two of the most popular HMDs with different evaluation methods, Samsung Galaxy S8 with Samsung Gear VR which includes a touchpad on its right side, and Lenovo Mirage Solo with a handheld controller. Regarding the stimuli, Table 2 presents the technical specifications of the six representative sources with audio selected for the experiment. As recommended in ITU-T P.910 [14], they cover a wide range of characteristics in terms of spatial and temporal information. Additional information is provided in the repository [41]. Clips of 25 seconds from the sources were encoded with ITU-T H.265/High Efficiency Video Coding (HEVC) using fixed Quantization Parameters (QPs): 22, 27, 32, 37, and 42 [44]. Video quality was evaluated using the ACR-HR (Absolute Category Rating with Hidden Reference) with a five-level rating scale, as recommended in ITU-T P.910 [14]. Presence was assessed with two of the

highly tested questionnaires: the Temple Presence Inventory (TPI) [45] and the Presence Questionnaire (PQ) [46]. As a result of this work, we provided a repository that contains[1]:

- Dataset of video sources with the associated objective metrics results (PSNR, WS-PSNR, CPP-PSNR, VMAF, SSIM, MSSSIM) and details (Spatial and Temporal Indicators [14], resolution, framerates, and brief descriptions).
- Head tracking data and video quality rates obtained from 48 participants during free-viewing experiments with two HMDs: Samsung GearVR and Lenovo Mirage Solo.
- Presence questionnaire scores, specifically TPI (Lombard et al.) and PQ (Witmer & Singer), obtained from 48 participants.
- Statistical analysis notebook.

Following the literature, we corroborate that it is difficult to evaluate socioemotional aspects in short-duration clips where there is neither narrative nor context. In addition, the fact of repeatedly visualizing the same clips in different qualities and with two devices, made the participants initially evaluate the aspects related to presence in a positive way but nevertheless, as the experiment session progressed, the sense of presence decreased notably, what we call as fatigue effect. Additionally, some participants after the session told the researcher responsible for the experiment that the presence was highly depending on the content. As we had not collected this information in a structured way, we considered in the experiment that we present in this paper higher-level aspects such as acquisition perspective, camera location, and interactive elements that could influence on socioemotional aspects. The fact that the fatigue and higher-level aspects may affect the evaluation of socioemotional features is a huge motivation for the methodology for video quality evaluation proposed in this experiment. Also, the results of the pilot study help us to select the handheld controller as evaluation method to increase the comfort of the observers. Additionally, we present a comparison of the scores obtained during the video quality evaluation of the pilot study and the experiment presented in this paper.

3 RESEARCH QUESTIONS

Based on the previous analysis, we pose the following Research Questions (RQs):

- RQ1: Is it possible to evaluate video quality in videos of long-duration designed for the evaluation of socioemotional features?
- RQ2: Which technical aspects, such as the position of the camera, the type of conversation, the video quality or the acquisition perspective influence socioemotional features?
- RQ3: Which interactive elements can be provided to the remote client to improve some socioemotional aspects such as presence or attention?

To answer these RQs, we designed a subjective experiment where an immersive communication between a

1. https://www.gti.ssr.upm.es/data/360VR

TABLE 3
Overview of the three experimental conditions with the associated interactive element and features assessed in the experiment

Condition	Assessment		Interactive element
	Quality	Socioemotional	Hands
A	X	X	
B		X	
C		X	X

Fig. 2. Participant of condition C at the environment of the experiment

provider and a *remote client* was simulated, presented in Figure 1. At the provider side, a conversation among several people took place, and the remote client attended virtually wearing an HMD. In the subjective test, the observer took the role of the remote client and visualized pre-recorded 360-degree videos with fluctuations of quality, simulating a VR streaming communication.

The contents used in the experiment showed simulated conversations around a common topic: international experiences, i.e. working or studying abroad. The main idea behind choosing this specific context was our ability to gather a balanced sample of people who have had international experiences and with people who have not. We acquired 360-degree videos with different acquisition perspectives (actor and observer) and genre (everyday conversation, educational, and discussion). For that, student volunteers were recruited for the recordings, both exchange and national students from the university, making the conversations more realistic and fluent. Conversations were in English, making the experiment accessible to different nationalities and mother tongues and increasing the diversity of the sample.

4 EXPERIMENT DESIGN

The experiment was designed to jointly assess socioemotional features such as the sense of presence, empathy, attitude, and attention with video quality in a specific use case: a 360-degree communication.

4.1 Experimental Conditions

The experiment considered three test conditions, summarized in Table 3, and each participant was assigned a condition. However, in all conditions, participants visualized the same video, with the same fluctuations of the quality. After each video, they were requested to rate its visual quality, as well as to evaluate the socioemotional features of interest: empathy and attitude, spatial and social presence, and attention.

Participants assigned to condition A had the additional task of periodically rating the visual quality of the video during its playback, whenever its quality changed. This is a conventional design to evaluate the subjective quality of the video sequence under different intensities of impairment. However, this focused task might have impact on the evaluation of socioemotional features compared to the baseline scenario without the task (condition B).

Finally, participants in condition C were provided with an additional interactivity element: the possibility to see their own hands and take handwritten notes about the conversation, as shown in Figure 2. We hypothesize that this could enhance socioemotional features such as presence and attention with respect to the other conditions.

4.2 Test Dataset

The set of source videos, Student Experiences Around the World dataset (SEAW-dataset) consists of three stereoscopic contents in 4K resolution at 30 fps and a duration of approximately five minutes each were acquired and prepared specifically for the experiment. Figure 3 shows a screenshot of the source videos and the original ones can be found in the supplementary material[2].

As it can be observed in all sequences, student volunteers were sitting around a table far enough from the camera to avoid stitching problems affecting the user's QoE and video quality scores. In addition, the camera was placed at the position and average height of the head of a person sitting at the same table, facilitating the engaging experience [47]. Table 4 summarizes the genre, perspective-taking, and a brief description of the contents used in the experiment. In contents with the actor acquisition perspective, student volunteers during the recording looked at the camera, and even waved their hands to increase the immersion of the participant of the experiment visualizing the 360-degree content with the HMD.

The contents were encoded with HEVC switching to a different fixed QP each 25 seconds to create one Processed Video Sequence (PVS) per source content [13]. The QPs selected for the experiment were: 15, 22, 27, 32, 37, and 42 [44]. These QPs were randomized along the video encoding, following Rec. ITU-R BT.500-13 [13]. Based on the assumption that each video source maintains the features in terms of color, texture, composition, and light, participants rated the quality of each one of the 25-second units along the whole sequence, avoiding the repetition of the same clip [29]. Due to the duration of the contents, each QP was rated at least two times in each PVS. Finally, the original audio quality was maintained through the experiment, improving the immersion and the QoE of the observers [48].

4.3 Methodology

Here, we explain in detail the methodology considered in the experiment. Table 5 summarizes the items evaluated in the three experimental conditions.

2. The dataset will be made public in www.gti.ssr.upm.es/data/seaw-dataset upon acceptance of the paper.

(a) *Coffee shop* (b) *International office* (c) *Study in Spain*

Fig. 3. Video sources screenshots

TABLE 4
Characteristics of the 360-degree videos considered in the experiment

Name	Genre	Perspective-taking	Description
Coffee shop	Everyday conversation	Observer	*A coffee conversation between foreign and local students about cultural differences*
International office	Educational	Actor	*A presentation given by a professor to students about the foreign application process*
Study in Spain	Discussion	Actor	*A conversation about the differences between transport and rental prices in different countries*

TABLE 5
Structure of the test session questionnaires

Condition	Pre-questionnaire (once)			During each content		Post-questionnaire (for each content)			
	Personal information	Empathy (IRI)	Attitude (EM1-EM4)	Quality (SSDQE)	Quality (ACR)	Attention survey	Attitude (EM5-EM8)	Spatial Presence (PP1-PP5) & Social Presence (SP1-SP5)	Notes
A	X	X	X	X	X	X	X	X	
B	X	X	X		X	X	X	X	
C	X	X	X		X	X	X	X	X

Personal information: For each participant, we collected age, gender, vision (corrected or normal), nationality, experience living in a foreign country and which one, and English level. This was used to characterize our observers and guarantee diversity.

Empathy. The initial empathy of each of the observers was evaluated using the Interpersonal Reactivity Index (IRI) [39]. This questionnaire is a psychometrically invariant empathy measure based on 28 statements related to the Perspective-Taking scale (PT), Fantasy Scale (FS), Empathic Concern scale (EC), and Personal Distress scale (PD). For each statement, the observer was required to indicate how well it described her/him on a five-level scale (where 1 = "Does not describe me well", to 5 = "Describes me very well").

Attitude. A survey was designed to measure the attitude towards the context of the videos, international experiences. As there was no validated questionnaire to measure the attitude of the participants towards other cultures and foreigner experiences, we decided to apply the Facet theory [49]. Facet theory consists of distinguishing the facets in which the designers of the experiment are interested. From the identified facets, the items of the questionnaire are defined and associated. In our case, we identified four characteristics that a person with a positive attitude towards foreigners and other cultures must have: interest, tolerance, respect, and social sensitivity. We established four statements related to the **interest**, **respect**, **tolerance,** and **social sensitivity**

towards other cultures and traditions. These four items were evaluated before starting the session and after the visualization of each of the three videos analyzed in the experiment. In this way, we could compare the empathy and attitude evolution throughout the session. To do this, four questions (EM1-EM4) were designed for the first evaluation at the beginning of the test, and another four questions (EM5-EM8) for the evaluation after each video. The idea behind this design was to compare the ratings before the visualization of the 360-degree content and after it. Observers provided ratings on a seven-point Likert scale (where 1 = "Strongly disagree", to 7 = "Strongly agree") based on most works in the literature [40], [32]. Specifically, it was measured with the following questions: *I just need to know the traditions of my country of origin (EM1), I think that some traditions of other cultures should not be allowed in my country (EM2), I feel comfortable with traditions different from mine (EM3), and I am worried about the experiences of foreign people in my country (EM4), I like participating in this kind of conversation (EM5), I think that an intercultural society has a positive impact for the people (EM6), I would feel comfortable sharing traditions of other cultures (EM7), I would like to participate in a buddy program or in a project to know more about foreign experiences in my country (EM8).*

Quality. A Single-Stimulus Discrete Quality Evaluation (SSDQE) method [50] was applied to measure the quality in observers assigned to condition A. SSDQE uses long-duration contents to evaluate quality guaranteeing the con-

TABLE 6
Attention survey: True/False statement, short answer, and multiple choice question for each of the 360 videos

Content	True/False statement	Short answer question	Multiple-choice question
Coffee shop	Students think that the second year is easier	How long ago has the university education system in Spain changed?	What city are the exchange students from?
International office	All students can apply for an internship both in research groups and in companies	The deadline to apply for double degree students	The deadline to apply for an internship or exchange program for the whole year
Study in Spain	Norwegians spend on average less money on public transport	The price of the public transport card in France	The rent per month in Norway

tinuity of the narrative. For this, the content is divided into segments, trying to mimic realistic situations of video consumption. As represented in Fig. 4, impairments are inserted throughout the content used as stimuli (PVS) in alternate segments ("processed segments") and participants rate the perceived quality during the following ones ("evaluation segments"). Note that during the evaluation segment, video playback continues and is encoded with the same quality as the previous processed segment. Specifically, they evaluated the quality on a five-grade quality scale [13], where the categories: "Bad", "Poor", "Fair", "Good", and "Excellent" were displayed on the screen. Additionally, the aggregate quality was asked, following the literature, in the post-questionnaire using the Absolute Category Rating (ACR) on the same five-grade scale [1], [51].

Attention. Observer attention was assessed with three questions about the conversations taking place in the videos that had pass/fail answers [18], [30]. For each content, we designed a multiple-choice question, a short answer question, and a True/False statement, presented in Table 6. In this way, participants scored zero or one point for each correct answer, resulting in a maximum score of three points for the total attention score for each video.

Presence. Spatial and social presence experienced by the observers were evaluated with five questions obtained from the state of the art [35], [40]. The questions of social presence questionnaire were mainly related to factors that influence the involvement in the meeting, such as the feeling that people in the meeting are looking at us, talking to as or where is the group attention focused on [36]. Observers provided ratings on a seven-point Likert scale (where 1 = "Strongly disagree", to 7 = "Strongly agree"). Specifically, spatial presence was measured with the following questions: *I felt I was present in the places shown in the video (PP1), I felt surrounded by the actions in the video (PP2), I felt I was sitting by the table at the place of the video (PP3), I felt I could have reached out and touched the items on the table of the video (PP4), and I felt that all my senses were stimulated at the same time (PP5)*. Likewise, social presence was measured with the following ones: *I felt that people were talking to me (SP1), I felt that I was listening to the others in the video (SP2), I felt I was present with the other people in the video (SP3), I felt like the people in the video could see me (SP4), and I felt I was actually interacting with other people (SP5)*.

Notes: Participants assigned to condition C assessed the usefulness of the notes taken during the test session. Specifically, the question *Have your annotations helped you*
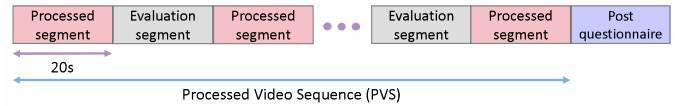

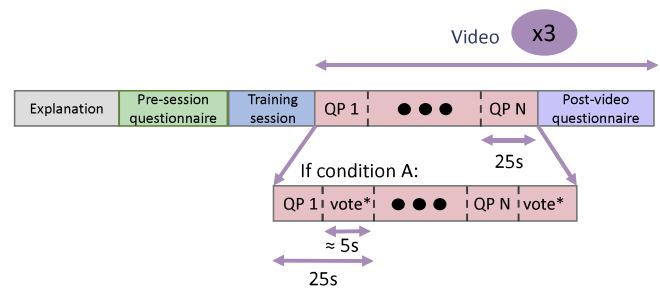Fig. 4. Structure of the test sequences used with SSDQE methodology

Fig. 5. Test session structure

to solve the questions? was used to consider whether the correct answers were correct from the annotations or from the memory of the participants.

4.4 Environment and equipment

All participants visualized the contents with a Samsung Galaxy S8 and the last model of Samsung Gear VR headset endowed with head tracking. The maximum resolution that viewers could perceive with this HMD (assuming a field of view of 85°x100° and a smartphone native resolution of 1440x2960 pixels), is about 680x822 pixels [52]. Monophonic audio was heard through headphones.

In all conditions, A, B, and C, the questionnaires were presented and answered using a web application. Observers who were assigned to condition A evaluated the quality of the video during the session. For this purpose, a VR application that allows users to visualize contents and answer customized questionnaires without having to take off their goggles was used [53]. They used a handheld controller as the evaluation method because it is more natural than the touchpad, avoiding any sign of discomfort [32]. Observers assigned to condition C were able to see their own hands, as well as a small whiteboard to take some notes, using an Augmented Virtuality (AV) approach, as shown in [54]. The local environment was captured by the smartphone camera and displayed in front of the 360-degree video. Background

(a) Condition A   (b) Condition B   (c) Condition C
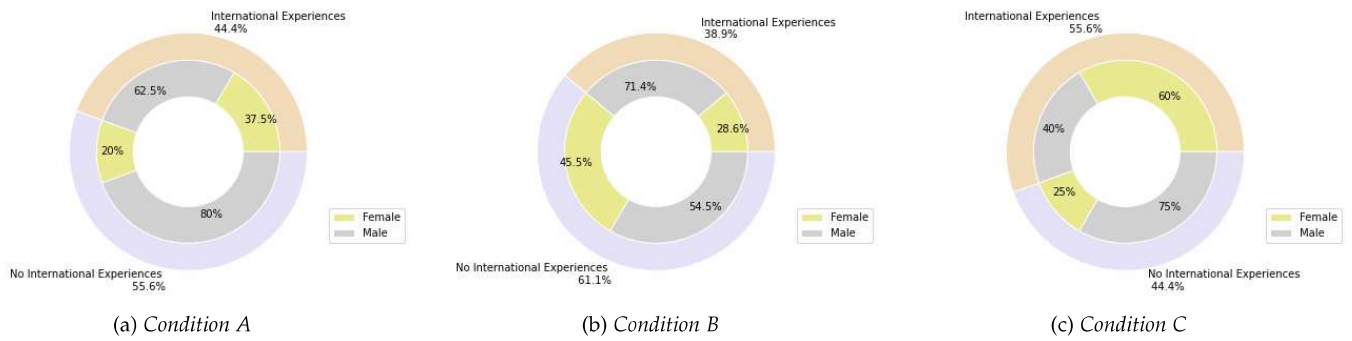
Fig. 6. Observers distribution in conditions A, B, and C taking into account the gender and international experience
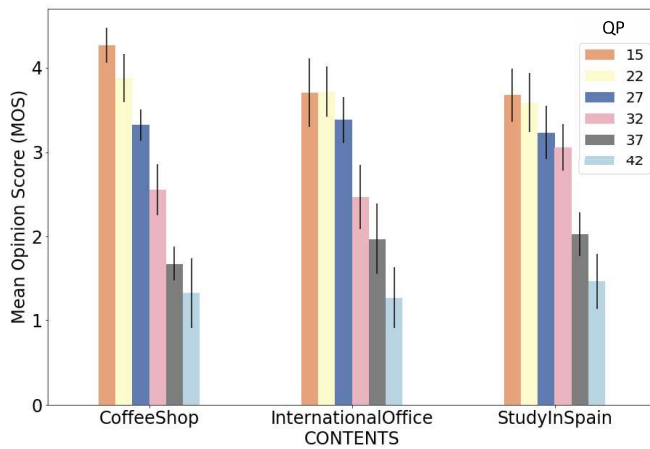

Fig. 7. The mean opinion scores (y-axis) on a five-level scale obtained from 17 participants assigned to condition A who evaluated the perceived video quality in the processed segment, encoded with specific QP, every 20 seconds while visualizing each of the three contents (x-axis), following the SSDQE methodology. Error bars represent 95% CI.
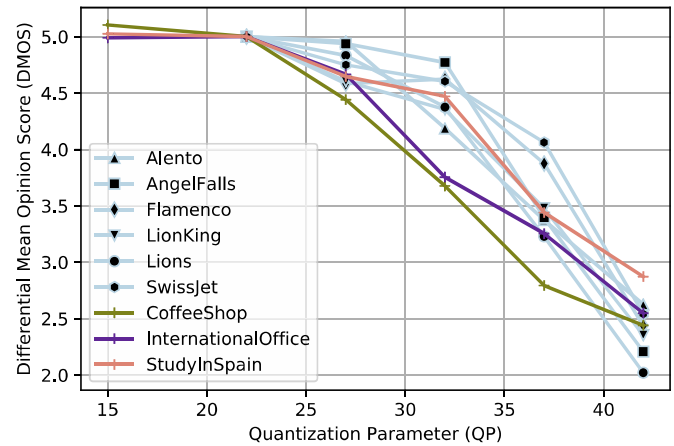
Fig. 8. Comparison of DMOS (y-axis) on a five-level scale obtained from 17 participants assgined to condition A, and from participants from the pilot study. Both participants evaluated clips of short duration encoded with fixed quantization parameters (x-axis). However, participants assigned to Condition A evaluated perceived quality following SSDQE methodology and participants from the pilot study evaluated it following ACR methodology.

was removed from the camera image using chroma-keying based on red chrominance.

Regarding the local environment, the observers were seated in a swivel chair in front of a table. This chair allowed them to spin around without more limitations than the three degrees of freedom, imposed by the HMD. The table in front of them was a requirement imposed by the videos, since, as presented in Figure 3, the three contents simulate a meeting around a table. In this way, observers could identify the table of the videos with the real one. Additionally, participants were located in totally isolated cubicles, facilitating the immersion in the content and avoiding any external distraction that increases the sense of FoMO.

4.5 Test session

The test session structure is presented in Figure 5. At the beginning, participants received a brief explanation of the experiment. Also, they were informed and signed a consent form that allowed us to process the information in accordance with the General Data Protection Regulation (GDPR) of the European Union. The experiment started with the pre-questionnaires: a personal information survey, the empathy questionnaire (IRI), and the initial attitude

survey. The training session consisted of a visualization of a discussion around the table with a duration of two minutes approximately. The PVS used for the training sessions was encoded with the best and worst qualities offered in the experiment (QP values of 15 and 42) every 25 seconds. Observers assigned to condition A tested the evaluation method with the handheld controller. After the training session, the assessment session started. All participants visualized the same three PVS in a randomized order following Recommendation ITU-R BT.500-13 [13]. For observers assigned to condition A, every 20 seconds the SSDQE question appeared without a time limit. After each video, all participants, regardless of the assigned condition, answered a post-questionnaire with questions about the quality and the socioemotional features. Here, participants assigned to condition C also answered the notes question.

4.6 Observers

A total of 54 observers (20 females, 34 males) took part in this experiment. There were participants in the age range between 17 and 26 years, with a Mean age (M) of 22.18 and a Standard Deviation (SD) of 1.95. All observers were checked

TABLE 7
Difference in aggregate quality and socioemotional features between the three conditions

Questionnaire items	Condition A	Condition B	Condition C	Significance
Aggregate quality 5-level scale	$M = 3.537$ $(SD = .719)$	$M = 3.111$ $(SD = 1.022)$	$M = 3.167$ $(SD = .885)$	$F_{2,153} = 3.687, p < .05, \eta_p^2 = .045, \gamma = .687$
Spatial Presence (7-level scale)	$M = 5.463$ $(SD = 1.019)$	$M = 5.185$ $(SD = 1.318)$	$M = 5.411$ $(SD = .942)$	$F_{2,153} = .900, p > .05, \eta_p^2 = .011, \gamma = .203$
Social Presence (7-level scale)	$M = 5.059$ $(SD = 1.398)$	$M = 5.133$ $(SD = 1.287)$	$M = 5.144$ $(SD = 1.271)$	$F_{2,153} = .005, p > .05, \eta_p^2 = .000, \gamma = .05$
Attitude post-questionnaire (7-scale level)	$M = 5.875$ $(SD = .250)$	$M = 6.042$ $(SD = .108)$	$M = 6.236$ $(SD = .191)$	$F_{2,153} = 4.660, p < .05, \eta_p^2 = .055, \gamma = .782$
Attention (3-level scale)	$M = 1.981$ $(SD = .765)$	$M = 1.833$ $(SD = .885)$	$M = 1.685$ $(SD = .748)$	$F_{2,153} = 1.839, p > .05, \eta_p^2 = .023, \gamma = .391$

TABLE 8
Difference in aggregate quality and socioemotional features between the three contents

Questionnaire items	Coffee shop	International office	Study in Spain	Significance
Aggregate quality (5-level scale)	$M = 3.111$ $(SD = .904)$	$M = 3.222$ $(SD = .883)$	$M = 3.481$ $(SD = .885)$	$F_{2,153} = 2.485, p > .05, \eta_p^2 = .03, \gamma = .496$
Spatial Presence (7-level scale)	$M = 5.326$ $(SD = 1.173)$	$M = 5.200$ $(SD = 1.137)$	$M = 5.533$ $(SD = .991)$	$F_{2,153} = 1.394, p > .05, \eta_p^2 = .017, \gamma = .297$
Social Presence (7-level scale)	$M = 4.748$ $(SD = 1.364)$	$M = 4.752$ $(SD = 1.280)$	$M = 5.837$ $(SD = .964)$	$F_{2,153} = 15.710, p < .01, \eta_p^2 = .169, \gamma = 1$
Attitude post-questionnaire (7-scale level)	$M = 6.111$ $(SD = .253)$	$M = 5.866$ $(SD = .234)$	$M = 6.176$ $(SD = .098)$	$F_{2,153} = 3.271, p > .05, \eta_p^2 = .038, \gamma = .605$
Attention (3-level scale)	$M = 2$ $(SD = .777)$	$M = 1.704$ $(SD = .743)$	$M = 1.796$ $(SD = .877)$	$F_{2,153} = 1.925, p > .05, \eta_p^2 = .024, \gamma = .407$

TABLE 9
Cronbach's α obtained for the questionnaires used in the experiment about spatial presence, social presence, and attitude

Questionnaire	Cronbach's α
Spatial Presence	0.857
Social Presence	0.865
Attitude (pre-questionnaire)	-0.094
Attitude (post-questionnaire)	0.710

for normal or corrected-to-normal vision. All participants were required at least an intermediate level of English to understand the conversations of the videos. They received a small financial reward for participating. In this way, we obtained a sample of participants with international experiences or nationalities from 15 countries in Europe, America, and Asia. The representation of user diversity was an additional value of the experiment, since it increased its reliability [55], [56]. Furthermore, as it can be observed in Figure 6, participants with international experiences and taking into account gender were distributed almost uniformly under conditions A, B, and C, guaranteeing a balanced sample.

5 EXPERIMENTAL RESULTS

For each one of the RQs, one or more hypotheses have been laid out and investigated, to look for relevant conclusions. Besides, the methodology to analyze the results was performed according to the nature of the data. The quality evaluation in condition A was examined with the MOS and the associated 95% CIs obtained from the scores, presented in Figure 7. In regard to the quality and socioemotional features, the Pearson & D'Agostino normality test was computed to validate the normal distribution of the collected data. For cases where the distribution was normal, the 2-way Analysis of Variance (ANOVA) was performed to examine the differences among the evaluated videos

and conditions. For social and spatial presence, due to the condition of non-normality, the following transformation was implemented: $\arcsin(\sqrt{P/7})$, where P is the presence rating and it is divided by seven because social and spatial presence were evaluated in a seven-level scale. Once the data was transformed, it was analyzed under the normality condition. Post-hoc analyses using Bonferroni correction for multiple comparisons were applied to examine the differences among the evaluated videos and conditions. The considered level of significance was 0.05. Table 7 and Table 8 present a summary of the scores of the items evaluated in the experiment and the significance (F, p, partial eta-squared η_p^2, and observed power values γ) between conditions and contents, respectively.

To investigate the factor structure of the questionnaires of presence and attitude, scale reliability has been addressed using Cronbach's α, presented in Table 9. Considering a reliability scale with $\alpha > 0.7$, the attitude responses from the pre-questionnaire have not been considered for the analysis.

5.1 Video quality assessment

Regarding RQ1, we investigated the first hypothesis (**H1**): video quality evaluation can be adapted to long-duration videos designed for socioemotional features assessment purposes. In this sense, we were interested in analyzing the effect of evaluating the quality of the video during the visualization of continuous sequences in which the scene features remain similar. Figure 7 was obtained from the scores of the 17 participants assigned to condition A. Note that the ratings of one of the observers were not collected correctly, so we remove an observer from condition A. As the evaluation of video quality on a 5-level score can be modeled by a Gaussian random process [57], we use parametric analysis for the evaluation of the scores, following the common practices for video quality data evaluation [58].

We have performed a ANOVA to assess the dependency of the scores on each source video and each QP value. Results show that the QP is significant ($F_{5,596} = 186.4$, $p < .001$, $\eta^2 = .598$), while the source content is not ($F_{2,596} = 0.85$, $p > .05$, $\eta^2 = .018$). Bonferroni-corrected pairwise t-tests show that all pairs of QPs are significantly different between them, except the two higher qualities QP values of 15 and 22, which are not. Note that due to the different duration of the videos and the randomization of the QPs, each QP was not evaluated the same number of times.

These quality scores were compared with the MOS values obtained from the pilot study presented previously [32], which was executed using a conventional ACR methodology with randomized 10-second video sequences. As the source contents were different in both experiments, we computed the Differential Mean Opinion Scores (DMOS), according to ITU-T Recommendation P.910 [14]. We used QP 22 as the hidden reference, as it was the highest quality available in the pilot study, and it was shown not to be significantly different from QP 15 in our new experiment. Figure 8 shows that both methodologies offer comparable results: good distribution of the ratings and a consistent decrease of the perceived quality when augmenting the QP, as expected in this type of tests [59].

These results show that subjects are able to effectively assess the video quality of individual QPs, and the content does not distract them from the task. This is in line with the results already reported in the literature for conventional 2D video and similar evaluation methodologies [50], [60], [61]. Furthermore, having the subjects engaged in the content increases the ecological validity of the quality evaluation compared to traditional methods [34], [62].

The aggregate quality scores rated at the end of each video in a five-level scale were analyzed statistically to find differences between videos and conditions. Due to the normality condition, 2-way ANOVA was applied. Table 8 shows that there is no significant difference between videos. MOS lay somewhere in the middle between the lowest and highest scores obtained for individual QPs, which is also expected [63]. It is known that several factors, such as the amplitude, frequency, and time location of the quality switches have an effect on the formation of the overall quality opinion [33], but addressing them is outside the scope of our experiment.

However, there is a significant difference among conditions, as seen in Table 7. Student's t-test with Bonferroni correction shows that this difference is significant between conditions A and B ($p = .0307$). Participants assigned to condition A scored the aggregate quality higher than participants assigned to condition B and C. It means that participants that are focused on the quality evaluation throughout the sequence, change their perspective about the perceived global quality.

To the authors' knowledge, this result is new in the literature. Some authors have used similar methods to evaluate the video quality continuously during the content playback, and then a single *endpoint quality* score at the end to assess the overall quality of the sequence [61], [64]. However, none of them has also had the same sequences evaluated just at the end, as it is proposed, for instance, by ITU-T [65]. Our results show that the evaluation of quality during the

sequence has a significant influence on the endpoint quality.

5.2 Spatial and social presence

In reference to RQ2, we investigated the second hypothesis (**H2**): acquisition perspective, type of the conversation, and experimental condition have influence on: spatial and social presence. As said before, the items of the sense of spatial presence and social presence were measured on a seven-point Likert scale independently. As presented in Table 7 and Table 8, the analysis was twofold. Once the non-normality condition of the social and spatial presence ratings was corrected, ANOVA was applied to analyze differences between experimental conditions. The aggregate measure of the five spatial and social presence items respectively show that there is not a significant difference. Nevertheless, a notable result is that the perceived social and spatial presence were very high in all conditions. In this sense, we want to point out that during the design of the experiment we presumed significant differences for condition C. We consider that the absence of differences is due to the fact that there were no specific tasks that required hands-on interaction with the VR environment.

Due to the non-normality condition and relative samples, Friedman test

Likewise, ANOVA was applied to examine the differences between videos. The aggregate measure of the five items of spatial presence shows that there is no significant difference, but the aggregate measure of the five items of social presence does show a difference. Student's t-test with Bonferroni correction shows that there is a significant difference between "Study in Spain" and "Coffee shop" contents ($Z = -4.887$, $p < .01$) and "Study in Spain" and "International office" content ($Z = -5.023$, $p < .01$).

Table 8 shows that "Study in Spain" scored higher in social presence. To better explore the difference between contents, Wilcoxon Signed-Rank test with Bonferroni correction were applied to the items of the social presence questionnaire (SP1-SP5) [40]. The analysis shows significant differences between "Study in Spain" and the other videos, "Coffee Shop" and "International office", in questions related to the perception that people in the conversation speak, look at, and interact with the participant: SP1 ($Z = 47.5$, $p < .01$ and $Z = 108$, $p < .01$), SP3 ($Z = 113.5$, $p = .0007$ and $Z = 136$, $p = .015$), SP4 ($Z = 115.5$, $p < .01$ and $Z = 108.5$, $p = .0001$), and SP5 ($Z = 107$, $p = .0005$ and $Z = 86$, $p < .01$). The reason is that in "Study in Spain" content the actors appeal to the camera more frequently, emphasizing the non-verbal side of the conversation.

5.3 Empathy and attitude towards international experiences

Following with RQ2, we investigated the third hypothesis (**H3**): acquisition perspective, type of the conversation, and experimental condition have influence on: empathy and attitude.

Firstly, IRI ratings were examined to obtain an adequate measure of the initial empathy of the participants, avoiding any deviation that may affect the subsequent analysis of the attitude. Given the condition of normality, ANOVA test was conducted to examine the IRI scores depending on gender

TABLE 10
The mean and standard deviations on a seven-level scale of the items of the attitude survey: interest, respect, tolerance, and social sensitivity in the three experimental conditions

Condition	Post-questionnaire			
	Interest	Respect	Tolerance	Social Sensitivity
A	$M = 5.167$	$M = 6.685$	$M = 6.407$	$M = 5.241$
	$(S = 1.411)$	$(S = .571)$	$(S = .806)$	$(S = 1.17)$
B	$M = 5.389$	$M = 6.574$	$M = 6.333$	$M = 5.870$
	$(S = 1.177)$	$(S = .71)$	$(S = .861)$	$(S = 1.171)$
C	$M = 5.537$	$M = 6.704$	$M = 6.407$	$M = 6.296$
	$(S = 1.343$	$(S = .565)$	$(S = .913)$	$(S = .936)$

and international experiences. It shows that there are not significant differences ($F_{1,50} = .76$, $p > .05$ and $F_{1,50} = 3.838$, $p > .05$, respectively). Based on the literature [66], [67], we expected significantly higher scores for females than for males. In our case, there are not significant differences but on average females scored higher empathy than males both for participants with international experiences ($M = 3.373$; $SD = .228$ and $M = 3.316$; $SD = .237$) and for participants without international experiences ($M = 3.246$; $SD = .302$ and $M = 3.182$; $SD = .199$).

Secondly, the attitude was evaluated with the questionnaire asked after the visualization of each of the three PVS. Table 10 summarizes the obtained results for each facet in the post-questionnaires ("Post"). Note that the data presented in the table is calculated in the original seven-level scale. The attitude was measured with the aggregation of four items of the designed survey: interest, respect, tolerance, and social sensitivity. Table 8 shows that there is not a statistically significant difference between contents but Table 7 presents a significant influence of the condition in which the content was visualized. Participants assigned to condition C achieved the highest attitude index, followed by participants from condition B and A. After finding that the condition greatly influences on the attitude, Student's t-test with Bonferroni correction was applied to find differences between conditions. They show that the significant difference is only between A and C conditions ($Z = -3.146$, $p = .002$). It makes sense because participants assigned to condition A had the video quality assessment task, distracting them from the conversations taking place in the video. From this analysis, another main result is that there is an important positive impact in the three videos, and as presented, in the three conditions.

5.4   Interactive element

Finally, to answer RQ3 we investigate the fourth hypothesis (H4): Observers who can take notes get higher total attention scores. Participants scored one point for each correct answer, resulting on a scale from 0 to 3. Due to condition of normality, ANOVA test was applied to find differences between conditions. As presented in Table 7, the scores show that there is no a significant difference between participants assigned to condition A, B, and C. Among the 18 participants assigned to condition C, 10 of them reported that the notes they had taken helped them to answer the questions. However, their scores are not significantly different from the ones obtained by the other 8 participants.

5.5   A method to simultaneously assess video quality and socioemotional features

Our experiment shows that the methodology used for condition A is suitable for the simultaneous evaluation of video quality and socioemotional features. As shown in section 5.1, SSDQE is valid to evaluate individual quality variations. Additionally, SSDQE does not affect the evaluation of presence or attention, which has two implications: on the one hand, it confirms that socioemotional features can be assessed despite having the extra task of continuous video quality evaluation; on the other, it shows that SSDQE does not reduce the observer immersion, making it a real content-immersive method.

There are, however, at least three caveats. First, using SSDQE does affect the evaluation of the overall quality of the sequence. Results obtained using this method will not be exactly the same as assessing the quality just with an endpoint evaluation. Second, as described in Section 5.3, using SSDQE during the video has some impact on the attitude of the observers. This means that, although the simultaneously evaluation of quality and socioemotional features is possible, it is not completely neutral, and some interaction between evaluation tasks may exist. Finally, it is worth noting that the experiment has been done with a specific type of content and visualization (360-degree videos simulating conversations on international experiences). Other types of videos or visualization setups might have different behavior. Further research is needed to address these items.

6   CONCLUSIONS

We have proposed a methodology where video quality, presence, empathy, attitude, and attention are jointly assessed in VR communications. We have simulated that user attend meetings remotely with the HMD and all meetings are focused on the international experiences context. Additionally, we have evaluated three conditions for the attendants. As a result, we have provided a dataset of three source videos designed and acquired for the purposes of the experiment. In addition, we have made them publicly available with the associated scores of the questionnaires and head-tracking of the participants.

We can conclude that video quality assessment can be adapted to conditions imposed by socioemotional feature methodologies, such as contents of longer duration where the scene background is mainly static. This is an important contribution to the state of the art, since it shows that methodologies can be designed to simultaneously evaluate technical features and socioemotional features that go one step further. Thus, it allows this type of experiment in more realistic environments with final VR applications.

The prototype evaluated for VR communications provides high scores in terms of social and spatial presence. Significant differences in the sense of social presence have been obtained between sequences. Then, we can assure that social presence is highly influenced by the acquisition perspective, narrative, and non-verbal behaviour of the participants on the provider side, enriching the effectiveness of the conversation.

We have designed a questionnaire to evaluate attitude among participants. We have found significant differences

between the experimental conditions and we can confirm that a positive impact has been achieved in all participants.

Finally, we cannot assure that the interactive element, the proper hands of the participants and a whiteboard with a whiteboard marker to take notes, significantly influences attention and spatial and social presence.

ACKNOWLEDGMENTS

REFERENCES

[1] S. N. B. Gunkel, H. M. Stokking, M. J. Prins, N. van der Stap, F. B. t. Haar, and O. A. Niamut, "Virtual Reality Conferencing: Multi-user Immersive VR Experiences on the Web," in *ACM Multimedia Systems Conference*, Jun. 2018, pp. 498–501.

[2] J. Li, V. Vinayagamoorthy, R. Schwartz, W. IJsselsteijn, D. A. Shamma, and P. Cesar, "Social VR: A New Medium for Remote Communication and Collaboration," in *CHI Conference on Human Factors in Computing Systems*, Apr. 2020, pp. 1–8.

[3] M. Orduna, "Quality, Presence, and Emotions in Virtual Reality Communications," in *IEEE Conf. on Virtual Reality and 3D User Interfaces - Doctoral Consortium*, Mar. 2020, pp. 561–562.

[4] F. Grondin, A. M. Lomanowska, and P. L. Jackson, "Empathy in Computer-Mediated Interactions: A Conceptual Framework for Research and Clinical Practice," *Clinical Psychology: Science and Practice*, vol. 26, no. 4, p. e12298, Jul. 2019.

[5] T. Rhee, S. Thompson, D. Medeiros, R. Dos Anjos, and A. Chalmers, "Augmented virtual teleportation for high-fidelity telecollaboration," *IEEE transactions on visualization and computer graphics*, vol. 26, no. 5, pp. 1923–1933, 2020.

[6] G. A. Lee, T. Teo, S. Kim, and M. Billinghurst, "A User Study on mr Remote Collaboration using Live 360 Video," in *IEEE Int. Symposium on Mixed and Augmented Reality*, 2018, pp. 153–164.

[7] R. Kachach, M. Orduna, J. Rodríguez, P. Pérez, Á. Villegas, J. Cabrera, and N. García, "Immersive telepresence in remote education," in *International Workshop on Immersive Mixed and Virtual Environment Systems*, 2021, pp. 21–24.

[8] J. Thies, M. Zollhöfer, M. Stamminger, C. Theobalt, and M. Nießner, "FaceVR: Real-time Gaze-Aware Facial Reenactment in Virtual Reality," *ACM Transactions on Graphics*, vol. 37, no. 2, pp. 1–15, Jun. 2018.

[9] M. Salminen, S. Järvelá, A. Ruonala, V. Harjunen, G. Jacucci, J. Hamari, and N. Ravaja, "Evoking Physiological Synchrony and Empathy Using Social VR with Biofeedback," *IEEE Transactions on Affective Computing*, Dec. 2019, early Access.

[10] Recommendation ITU-T G.1035, "Influencing Factors on Quality of Experience for Virtual Reality Services," May 2020.

[11] C. Ozcinar, J. Cabrera, and A. Smolic, "Visual Attention-aware Omnidirectional Video Streaming Using Optimal Tiles for Virtual Reality," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 9, no. 1, pp. 217–230, Jan. 2019.

[12] E. David, J. Gutiérrez, A. Coutrot, M. P. Da S., and P. Le Callet, "A Dataset of Head and Eye Movements for 360 Videos," in *ACM Multimedia Systems Conference*, Jun. 2018, pp. 432–437.

[13] Recommendation ITU-R BT.500-14, "Methodology for the Subjective Assessment of the Quality of Television Pictures," Oct. 2019.

[14] Recommendation ITU-T P.910, "Subjective Video Quality Assessment Methods for Multimedia Applications," Apr. 2008.

[15] Recommendation ITU-T P.913, "Methods for the Subjective Assessment of Video Quality, Audio Quality and Audiovisual Quality of Internet Video and Distribution Quality Television in any Environment," Mar. 2016.

[16] J. Gutierrez, P. Perez, M. Orduna, A. Singla, C. Cortes, P. Mazumdar, I. Viola, K. Brunnstrom, F. Battisti, N. Cieplinska, D. Juszka, L. Janowski, M. I. Leszczuk, A. Adeyemi-Ejeye, Y. Hu, Z. Chen, G. Van Wallendael, P. Lambert, C. Diaz, J. Hedlund, O. Hamsis, S. Fremerey, F. Hofmeyer, A. Raake, P. Cesar, M. Carli, and N. Garcia, "Subjective Evaluation of Visual Quality and Simulator Sickness of Short 360 Videos: ITU-T Rec. P.919," *IEEE Transactions on Multimedia*, pp. 1–14, 2021, Early Access.

[17] N. S. Schutte and E. J. Stilinović, "Facilitating Empathy Through Virtual Reality," *Motivation and Emotion*, vol. 41, no. 6, pp. 708–712, Oct. 2017.

[18] A. Voinescu, L. Fodor, D. S. Fraser, M. Mejías, and D. David, "Exploring the Usability of Nesplora Aquarium, a Virtual Reality System for Neuropsychological Assessment of Attention and Executive Functioning," in *IEEE Conf. on Virtual Reality and 3D User Interfaces*, Mar. 2019, pp. 1207–1208.

[19] I. Tussyadiah, D. Wang, T. Jung, and M. tom Dieck, "Virtual Reality, Presence, and Attitude Change: Empirical Evidence from Tourism," *Tourism Management*, vol. 66, pp. 140–154, Jun. 2018.

[20] D. Fonseca and M. Kraus, "A Comparison of Head-Mounted and Hand-Held Displays for 360 Videos with Focus on Attitude and Behavior Change," in *International Academic Mindtrek Conference*, Oct. 2016, pp. 287–296.

[21] P. Pürcher and M. Höfler, "Technology Meets Psychology: Psychological Background in Virtual Realities," in *Int. Convention on Information and Communication Technology, Electronics and Microelectronics*, May 2018, pp. 633–637.

[22] A. Perkis et al., "QUALINET White Paper on Definitions of Immersive Media Experience (IMEx)," 2020, arXiv:2007.07032.

[23] B. J. Li, J. N. Bailenson, A. Pines, W. J. Greenleaf, and L. M. Williams, "A Public Database of Immersive VR Videos with Corresponding Ratings of Arousal, Valence, and Correlations between Head Movements and Self Report Measures," *Frontiers in Psychology*, vol. 8, Dec. 2017.

[24] H. Jun, M. R. Miller, F. Herrera, B. Reeves, and J. N. Bailenson, "Stimulus Sampling with 360-Videos: Examining Head Movements, Arousal, Presence, Simulator Sickness, and Preference on a Large Sample of Participants and Videos," *IEEE Transactions on Affective Computing*, Jun. 2020, early Access.

[25] X. Corbillon, F. De Simone, and G. Simon, "360-degree Video Head Movement Dataset," in *ACM on Multimedia Systems Conference*, Jun. 2017, pp. 199–204.

[26] W. Lo, C. Fan, J. Lee, C. Huang, K. Chen, and C. Hsu, "360 Video Viewing Dataset in Head-Mounted Virtual Reality," in *ACM on Multimedia Systems Conference*, Jun. 2017, pp. 211–216.

[27] J. Yang, T. Liu, B. Jiang, H. Song, and W. Lu, "3D Panoramic Virtual Reality Video Quality Assessment based on 3D Convolutional Neural Networks," *IEEE Access*, vol. 6, pp. 38 669–38 682, Jul. 2018.

[28] F. De Simone, J. Gutiérrez, and P. Le Callet, "Complexity measurement and characterization of 360-degree content," in *Human Vision and Electronic Imaging*, vol. 2019, no. 12, Jan. 2019, pp. 216–1–216–7.

[29] L. Janowski, L. Malfait, and M. H. Pinson, "Evaluating Experiment Design with Unrepeated Scenes for Video Quality Subjective Assessment," *Quality and User Experience*, vol. 4, no. 2, Jun. 2019.

[30] A. MacQuarrie and A. Steed, "Cinematic Virtual Reality: Evaluating the Effect of Display Type on the Viewing Experience for Panoramic Video," in *IEEE Virtual Reality*, Mar. 2017, pp. 45–54.

[31] A. Singla, S. Fremerey, F. Hofmeyer, W. Robitza, and A. Raake, "Quality assessment protocols for omnidirectional video quality evaluation," *Electronic Imaging*, vol. 2020, no. 11, pp. 69–1–69–7, Jan. 2020.

[32] M. Orduna, P. Pérez, C. Díaz, and N. García, "Evaluating the Influence of the HMD, Usability, and Fatigue in 360VR Video Quality Assessments," in *IEEE Conf. on Virtual Reality and 3D User Interfaces*, Mar. 2020, pp. 683–684.

[33] M.-N. García, F. De Simone, S. Tavakoli, N. Staelens, S. Egger, K. Brunnström, and A. Raake, "Quality of Experience and HTTP Adaptive Streaming: A Review of Subjective Studies," in *International Workshop on Quality of Multimedia Experience*, Sep. 2014, pp. 141–146.

[34] M. Pinson, M. Sullivan, and A. Catellier, "A New Method for Immersive Audiovisual Subjective Testing," in *Int. Workshop on Video Process. and Quality Metrics for Consumer Electronics*, Jan. 2014.

[35] G. Riva et al., "Affective Interactions using Virtual Reality: the Link Between Presence and Emotions," *CyberPsychology & Behavior*, vol. 10, no. 1, pp. 45–56, Feb. 2007.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TAFFC.2022.3149162, IEEE Transactions on Affective Computing

IEEE TRANSACTIONS ON AFFECTIVE COMPUTING VOL. XX, NO. X, XXX XXXX
13

[36] M. Slater and S. Wilbur, "A Framework for Immersive Virtual Environments (FIVE): Speculations on the Role of Presence in Virtual Environments," *Presence: Teleoperators and Virtual Envirnonments*, vol. 6, no. 6, pp. 603–616, Dec. 1997.

[37] A. K. Przybylski, K. Murayama, C. R. DeHaan, and V. Gladwell, "Motivational, Emotional, and Behavioral Correlates of Fear of Missing Out," *Computers in Human Behavior*, vol. 29, no. 4, pp. 1841–1848, Jul. 2013.

[38] L. Tong, S. Jung, and R. W. Lindeman, "Action Units: Directing User Attention in 360-degree Video based VR," in *ACM Symposium on Virtual Reality Software and Technology*, Nov. 2019, pp. 1–2.

[39] M. H. Davis, "Measuring Individual Differences in Empathy: Evidence for a Multidimensional Approach," *Journal of Personality and Social Psychology*, vol. 44, no. 1, pp. 113–126, Jul. 1983.

[40] T. Aitamurto, S. Zhou, S. Sakshuwong, J. Saldivar, Y. Sadeghi, and A. Tran, "Sense of Presence, Attitude Change, Perspective-Taking and Usability in First-Person Split-Sphere 360 Video," in *CHI Conf. on Human Factors in Computing Systems*, Apr. 2018, pp. 1–12.

[41] "Evaluating the Influence of the HMD, Usability, and Fatigue in 360VR Video Quality Assessments - Supplemental material," https://www.gti.ssr.upm.es/data/360VR.

[42] S. Schmidt, S. Zadtootaghaj, S. Wang, and S. Möller, "Towards the Influence of Audio Quality on Gaming Quality of Experience," in *Int. Conf. on Quality of Multimedia Experience*, 2021, pp. 169–174.

[43] S. Vlahovic, M. Suznjevic, N. Pavlin-Bernardic, and L. Skorin-Kapov, "The Effect of VR Gaming on Discomfort, Cybersickness, and Reaction Time," in *Int. Conference on Quality of Multimedia Experience*, 2021, pp. 163–168.

[44] JCT-VC and ISO/IEC JTC1/SC29/WG11, "Common HM Test Conditions and Software Reference Configurations," Output doc. M27343, 102th MPEG Meeting, Shanghai, China, Oct. 2012.

[45] M. Lombard, T. B. Ditton, and L. Weinstein, "Measuring Presence: the Temple Presence Inventory," in *Annual International Workshop on Presence*, Jan. 2009, pp. 1–15.

[46] B. G. Witmer and M. J. Singer, "Measuring Presence in Virtual Environments: A Presence Questionnaire," *Presence*, vol. 7, no. 3, pp. 225–240, Jun. 1998.

[47] T. Keskinen, V. Mäkelä, P. Kallionierni, J. Hakulinen, J. Karhu, K. Ronkainen, J. Mäkelä, and M. Turunen, "The Effect of Camera Height, Actor Behavior, and Viewer Position on the User Experience of 360 Videos," in *IEEE Conference on Virtual Reality and 3D User Interfaces*, Mar. 2019, pp. 423–430.

[48] A. Tse, C. Jennett, J. Moore, Z. Watson, J. Rigby, and A. L. Cox, "Was I There?: Impact of Platform and Headphones on 360 Video Immersion," in *CHI Conf. Extended Abstracts on Human Factors in Comput. Syst.*, May 2017, pp. 2967–2974.

[49] I. Borg, "Facet Theory," in *Encyclopedia of Statistics in Behavioral Science*, 2005.

[50] J. Gutiérrez, P. Pérez, F. Jaureguizar, J. Cabrera, and N. García, "Validation of a Novel Approach to Subjective Quality Evaluation of Conventional and 3D Broadcasted Video Services," in *International Workshop on Quality of Multimedia Experience*, Jul. 2012, pp. 230–235.

[51] A. Covaci, R. Trestian, E. B. Saleme, I. Comsa, G. Assres, C. Santos, and G. Ghinea, "360° Mulsemedia: A Way to Improve Subjective QoE in 360° Videos," in *ACM International Conference on Multimedia*, Oct. 2019, pp. 2378–2386.

[52] L. Muñoz, C. Díaz, M. Orduna, J. I. Ronda, P. Pérez, I. Benito, and N. García, "Methodology for Fine-grained Monitoring of the Quality Perceived by Users on 360VR Contents," *Digital Signal Processing*, p. 102707, May 2020.

[53] C. Cortés, P. Pérez, and N. García, "Unity3D-based App for 360VR Subjective Quality Assessment with Customizable Questionnaires," in *IEEE International Conference on Consumer Electronics*, Sep. 2019, pp. 281–282.

[54] P. Pérez, E. Gonzalez-Sosa, R. Kachach, J. Ruiz, I. Benito, F. Pereira, and A. Villegas, "Immersive Gastronomic Experience with Distributed Reality," in *Workshop on Everyday VR*, Mar. 2019, pp. 1–6.

[55] J. Himmelsbach, S. Schwarz, C. Gerdenitsch, B. Wais-Zechmann, J. Bobeth, and M. Tscheligi, "Do We Care About Diversity in Human Computer Interaction: A Comprehensive Content Analysis on Diversity Dimensions in Research," in *CHI Conf. on Human Factors in Computing Systems*, May 2019, pp. 1–16.

[56] T. C. Peck, L. E. Sockol, and S. M. Hancock, "Mind the Gap: The Underrepresentation of Female Participants and Authors in Virtual Reality Research," *IEEE Trans. Vis. Comput. Graphic*, vol. 26, no. 5, pp. 1945–1954, Feb. 2020.

[57] L. Janowski and M. Pinson, "The accuracy of subjects in a quality experiment: A theoretical subject model," *IEEE Transactions on Multimedia*, vol. 17, no. 12, pp. 2210–2224, 2015.

[58] M. Narwaria, L. Krasula, and P. Le Callet, "Data analysis in multimedia quality assessment: Revisiting the statistical tests," *IEEE Transactions on Multimedia*, vol. 20, no. 8, pp. 2063–2072, 2018.

[59] M. Orduna, C. Díaz, L. Muńoz, P. Pérez, I. Benito, and N. García, "Video Multimethod Assessment Fusion (VMAF) on 360VR Contents," *IEEE Transactions on Consumer Electronics*, vol. 66, no. 1, pp. 22–31, Feb. 2020.

[60] N. Cranley, P. Perry, and L. Murphy, "User perception of adapting video quality," *International Journal of Human-Computer Studies*, vol. 64, no. 8, pp. 637–647, Aug. 2006.

[61] D. Ghadiyaram, J. Pan, and A. C. Bovik, "A Subjective and Objective Study of Stalling Events in Mobile Streaming Videos," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 1, pp. 183–197, Nov. 2017.

[62] P. Kortum and M. Sullivan, "The effect of content desirability on subjective video quality ratings," *Human factors*, vol. 52, no. 1, pp. 105–118, May 2010.

[63] S. Tavakoli, S. Egger, M. Seufert, R. Schatz, K. Brunnström, and N. García, "Perceptual Quality of HTTP Adaptive Streaming Strategies: Cross-experimental Analysis of Multi-laboratory and Crowdsourced Subjective Studies," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 8, pp. 2141–2153, Aug. 2016.

[64] C. G. Bampis, Z. Li, A. K. Moorthy, I. Katsavounidis, A. Aaron, and A. C. Bovik, "Study of Temporal Effects on Subjective Video Quality of Experience," *IEEE Transactions on Image Processing*, vol. 26, no. 11, pp. 5217–5231, Jul. 2017.

[65] A. Raake, M.-N. García, W. Robitza, P. List, S. Göring, and B. Feiten, "A Bitstream-based, Scalable Video-Quality Model for HTTP Adaptive Streaming: ITU-T P. 1203.1," in *International Conference on Quality of Multimedia Experience*, Jun. 2017, pp. 1–6.

[66] S. T. Hawk, L. Keijsers, S. J. Branje, J. V. Graaff, M. Wied, and W. Meeus, "Examining the Interpersonal Reactivity Index (IRI) among Early and Late Adolescents and their Mothers," *Journal of Personality Assessment*, vol. 95, no. 1, pp. 96–106, 2013.

[67] A. Gilet, N. Mella, J. Studer, D. Grühn, and G. Labouvie-Vief, "Assessing Dispositional Empathy in Adults: A French Validation of the Interpersonal Reactivity Index (IRI)." *Canadian Journal of Behavioural Science*, vol. 45, no. 1, p. 42, Jan. 2013.

Marta Orduna received the Bachelor of Engineering in Telecommunication Technologies and Services in 2016 and the Master in Telecommunication Engineering (accredited by ABET) in 2018 (Master Graduation Award), both from the Universidad Politécnica de Madrid (UPM), Madrid, Spain. She has been a member of the Grupo de Tratamiento de Imágenes (Image Processing Group) of the UPM since 2016, where she has been actively involved in several research projects. Her current research is in the area of virtual reality, video encoding and streaming, and quality of experience.

Pablo Pérez received the Telecommunication Engineering degree (integrated BSc-MS) in 2004 and the Ph.D. degree in Telecommunication Engineering in 2013 (Doctoral Graduation Award), both from Universidad Politécnica de Madrid (UPM), Madrid, Spain. From 2004 to 2006 he was a Research Engineer in the Digital Platforms Television in Telefónica I+D and, from 2006 to 2017, he has worked in the R&D department of the video business unit in Alcatel-Lucent (later acquired by Nokia), serving as technical lead of several video delivery products. Since 2017, he is Senior Researcher in the Distributed Reality Solutions department at Nokia Bell Labs. His research interests include multimedia quality of experience, video transport networks, and immersive communications.

Jesús Gutiérrez is a research fellow (Juan de la Cierva) at the Image Processing Group (GTI) of the Universidad Politécnica de Madrid (UPM), Spain. He received the Telecommunication Engineering degree (five-year engineering program) from the Universidad Politécnica de Valencia (Spain) in 2008, the master's degree in Communications Technologies and Systems (two-year M.S. program) in 2011, and the Ph.D. degree in Telecommunication in 2016, both from the UPM. From 2016 to 2019 he was a post-doctoral researcher (firstly, Marie Curie Fellow within the PROVISION ITN, and then, Marie Curie PRESTIGE Fellow) at the Image, Perception and Interaction (IPI) group of the Laboratoire des Sciences du Numérique de Nantes (LS2N) of the Université de Nantes (France). His research interests are in the area of image and video processing, evaluation of user quality of experience, immersive media technologies, and visual attention and human perception.

Narciso García received the Ingeniero de Telecomunicación degree (five years engineering program) in 1976 (Spanish National Graduation Award) and the Doctor Ingeniero de Telecomunicación degree (PhD in Communications) in 1983 (Doctoral Graduation Award), both from the Universidad Politécnica de Madrid (UPM), Madrid, Spain. Since 1977, he has been a member of the faculty of the UPM, where he is currently a Professor of Signal Theory and Communications. He leads the Grupo de Tratamiento de Imágenes (Image Processing Group), UPM. He has been actively involved in Spanish and European research projects, also serving as an evaluator, a reviewer, an auditor, and an observer of several research and development programs of the European Union. He was a Co-Writer of the EBU proposal, base of the ITU standard for digital transmission of TV at 34–45 Mb/s (ITU-T J.81). He was an Area Coordinator of the Spanish Evaluation Agency (ANEP) from 1990 to 1992 and he was the General Coordinator of the Spanish Commission for the Evaluation of the Research Activity (CNEAI) from 2011 to 2014. He has been the Vice-Rector for International Relations of the Universidad Politécnica de Madrid from 2014 to 2016. He was a recipient of the Junior and Senior Research Awards of the Universidad Politécnica de Madrid in 1987 and 1994, respectively. His current research interests include digital video compression, computer vision, and quality of experience.