Immersive Gastronomic Experience with Distributed Reality

Pablo Perez Ester Gonzalez-Sosa Redouane Kachach Jaime Ruiz Ignacio Benito
Francisco Pereira Alvaro Villegas

Nokia Bell Labs*

Figure 1: *Left)* first-person view of the experience; *center)* external view of the experience; *right)* graphical model of the elements of the Distributed Reality.

ABSTRACT

We have developed an immersive gastronomic experience as a proof of concept of Distributed Reality, a type of Augmented Virtuality which combines a reality transmitted from a remote place, using 360° video, with a local reality, using video see-through. In order to reach fully immersive experience, local objects of interest such as hands and local food are segmented using red chrominance keying. Only those segmented objects are merged with the remote reality, enabling this way to increase self-presence and to allow user interaction. More concretely, the gastronomic experience consists of tasting small pieces of food, while being immersed in a remote place designed to pair with the food, thus creating an innovative concept with potential impact in hospitality and tourism industries. An evaluation performed with 66 users shows that it provides good levels of immersion, local interactivity, and general user satisfaction. The application achieves real time performance and has been developed for a smartphone mounted on a consumer headset, thus being easy to deploy and to reuse in other use cases.

Index Terms: Human-centered computing— Human computer interaction (HCI)—Interaction paradigms;

1 INTRODUCTION

Mixed Reality (MR) can be described as a continuum between a totally virtual reality (VR) and a local reality (the physical space surrounded the user wearing the VR headset). In this continuum lies Augmented Reality (AR), where the local reality is augmented with virtual elements, or Augmented Virtuality (AV) [16] where a virtual reality is augmented with elements from the local reality. In both cases, there is a *main* environment (either virtual or real), which is *augmented* with elements from the other one.

We are starting to explore the concept of Distributed Reality (DR), a similar concept to AV where the virtual reality is replaced by a remote reality. Unlike virtual reality, which is computer-generated, the remote reality is a 360° immersive video of a real place in real time, streamed to the user device. This remote reality is then merged with objects of interest (e.g. hands and food) from the local reality using a image-based segmentation algorithm based on red-chrominance key. Likewise AV, DR provides therefore a good sense of self-perception and it allows for interaction with local objects without removing the headset [15]. All in all, we believe the combination of merging realities with self-perception and interaction that provides the DR concept may become the new paradigm for human communications and also for human to machine interaction in our daily life in the future.

As a first approach to this DR concept, we have developed an immersive experience for tasting *tapas*[1]. Wearing a Head Mounted Display (HMD), the user can eat or drink a small piece of food or beverage (a *tapa*) while being completely immersed in a remote reality. As can be seen from Fig. 1 *left*, the user is able to see his own hands, a table and some food next to him, while feeling immersed in a scene of the beach. Only the table and the food belongs to his local reality (Fig. 1 *center*), while the beach scene is a 360° video which is played by the VR system of his HMD. With a menu designed by a professional cook, there is a pairing between each of the *tapas* and its surrounding remote reality, thus creating a very powerful gastronomic experience, which may have a relevant impact in hospitality and tourism industries.

The idea of eating in a virtual environment was already addressed by Korsgaard *et al.* [12]. They conducted an AV experience in which the user had to interact with real food placed in front of him. There are two main key differences between this prior work and our novel proposal. While the work by Korsgaard *et al.* [12] used a computed generated environment, our novel proposal replaces this with a remote reality captured as a 360° video (notice that in both cases local reality is captured with video see-through). Besides, in [12], there is no real merge between the two environments but rather a transition to the food area when the head orientation angle is lower than 25° or to the computer-generated environment when the head orientation angle is greater than 30°, preventing therefore to achieve full immersion. In our proposal full immersion is achieved

*Authors' e-mails:{pablo.perez, ester.gonzalez, redouane.kachach, jaime.ruiz_alonso, ignacio.benito_fronterlo, francisco.pereira_vega, alvaro.villegas}@nokia-bell-labs.com

[1] A *tapa* is an appetizer or snack typical from Spanish gastronomy.

by merging at frame level the remote and local reality, allowing the user to see herself immersed in the remote place at all times.

The task of eating in an immerse world has been attracting attention not only from the VR community. A psychological study conducted by Andersen *et al.* [1] explored whether a contextual exposure by immersive VR could induce effects on desires and liking aligning with the situation. For instance, one of their conclusions was that immersive experiences at the beach increased desires for cold beverages. Based on that, it could be possible that VR technology will be useful for food consumer research in the mid-term future. Our work can be a first milestone on this road.

The structure of the article is as follows. Section 2 describes the implementation of the Distributed Reality pilot system built for the gastronomic experience whereas Section 3 depicts the questionnaire to evaluate it. Then, Section 4 presents the gastronomic experience and the results of its evaluation. Finally Section 5 reports some general conclusions and future work.

2 Distributed Reality implementation

In this section we describe in more detail the two elements from the DR, the remote and the local realities, and the procedure followed to blend them in real time. The DR application is built using a Samsung Galaxy S8 attached to a GearVR HMD.

2.1 Remote Reality

Remote reality videos have been prepared so that their content pairs with the *tapa*. Some of them have been recorded by ourselves from scenarios with touristic relevance, using a professional Nokia OZO camera. Others have been obtained from publicly available sources. In all cases, videos are rendered in monoscopic equirectangular projection with a resolution of 3840×2160 pixels and 30 fps, encoded in HEVC at 20 Mbps, and packaged in MPEG-DASH, using the proprietary professional software packager provided by Nokia OZO.

Contents are stored in an http server, and they are delivered using MPEG-DASH adaptive streaming protocol. For local laboratory tests, IEEE 802.11ac wireless connection is used to guarantee that there is no bandwidth bottleneck in any part of the system. For field tests, the http server has been installed in a Multi-access Edge Computing (MEC) node providing service to a LTE radio cell.

The remote reality texture is then obtained by using a modified ExoPlayer[2], which plays in real-time a 360° video stream and projects it onto a sphere of radius R_R surrounding the user (Fig. 1 *right*).

2.2 Local Reality

Local Reality is captured by the smartphone rear camera following a video see-through approach. To create the Distributed Reality blend, it is first required to segment which elements of the local reality are merged into the final scene and which ones are not. In the particular case of the gastronomic experience, we are interested on segmenting local food and hands.

Several techniques have been studied to address the local reality segmentation: green chroma keying [15], skin detection [2], usage of depth information on its own [17,23] or combined with color [5], or more complex computer vision algorithms [4]. We have selected a red chrominance keying approach as the segmentation technique, as it provides good performance for segmenting both hands and food with low computational cost, and it can be directly executed on the smartphone with the images captured from its camera.

2.2.1 Segmentation: Red Chrominance Keying

Instead of using the traditional approach of a green background, which imposes the experience to take place inside a green room, we have based our solution in YCbCr skin detection. In particular, we have added a transparency alpha layer to the local reality texture, based on red chrominance:

$$\alpha(Cr) = \begin{cases} 1 & \text{if } Cr \geq C_1 \\ (Cr - C_0)/(C_1 - C_0) & \text{if } C_0 \leq Cr < C_1 \\ 0 & \text{if } Cr \leq C_0 \end{cases} \quad (1)$$

where,

$$C_r = 0.5R - 0.418688G - 0.081312B \quad (2)$$

For $R, G, B \in [0,1]$ we have used threshold values $C_0 = 0.06$ and $C_1 = 0.12$. Those thresholds are less strict than what it is normally used for skin detection, but this way the solution provides also a simple methodology to introduce elements in the local reality: red or yellow objects with high saturation will be displayed as local, while blue and neutral tones will be filtered out as background. With this method, it is possible to deploy the system in any room whose walls and furniture have black, white, grey and blue tones, instead of needing a full green room. The main drawback of using chroma keying is that this algorithm is quite sensitive to illumination conditions. To prevent this limitation from impairing the experience, light sources must be controlled when building the physical setup.

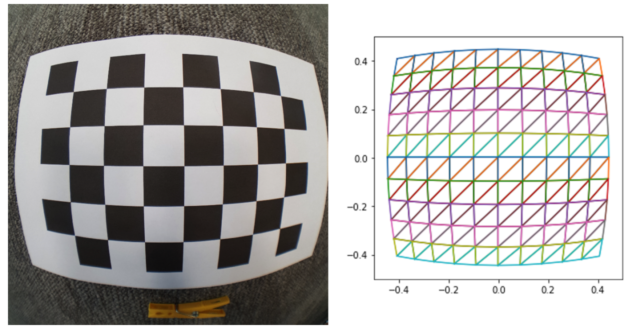

Figure 2: *Left)* checkerboard image used for calibration; *Right)* geometry of the local plane after the calibration process.

2.2.2 Merging Segmented Objects with Remote Reality

As the DR scene is rendered for each eye with an approximately square aspect ratio, the smartphone camera is also set up to capture at 1:1 aspect ratio. To allow the video processing to work smoothly on real time in the device, the resolution of the captured video has been set to 720×720 pixels, and 30 fps. This relatively low resolution and frame rate can potentially lead into cybersickness. To prevent that, a careful design of the experimental setup has been done to naturally restrict the user's potential motion without impacting the overall experience, which will be described later.

The square aspect ratio forces the application to discard about half of the horizontal resolution that the camera is able to provide and, as a result, the visual angle of the camera is much narrower than the $\sim 100°$ of field of view (FOV) provided by the HMD. To correct this effect, a fish-eye lens has been attached in front of the camera, providing up to 210° visual angle in the horizontal plane, and $\sim 100°$ in the vertical plane, which is enough to cover the whole FOV of the virtual scene.

However, this also introduces a heavy radial distortion component in the image, which needs to be rectified. The radial distortion parameters of the whole camera system (fish-eye + phone) have been estimated using a 9×6 chessboard, and the camera parameter estimation performed by OpenCV[3], which is largely based on [26].

[2] https://github.com/google/ExoPlayer

[3] https://docs.opencv.org/3.4/dc/dbb/tutorial_py_calibration.html

Table 1: Distributed Reality Experience Questionnaire (DREQ). The ten *main* questions use 5-point scales and address [1]presence (Likert scale), [2]video quality (Absolute Category Rating, ACR [8]), [3]cybersicknes (CS, *Vertigo* scale [19]), and [4]Quality of Experience (ACR). The extra question (*WDRC*), used to measure the Net Promoter Score, uses a 0-10 probability scale [20]. *RINT* question was not asked in the particular experience described in this article.

Factor	Question
Spatial Presence (SPRE)[1]	I felt like I was actually there in the remote place.
Remote Interaction (RINT)[1]	*I was able to interact with the objects of the remote place.*
Local Perception (LPER)[1]	I was aware of the events occurring around me.
Local Interaction (LINT)[1]	I was able to interact with the objects of the real world.
Task Completion (TASK)[1]	I was able to taste the food as in a normal restaurant.
Remote Quality (REMQ)[2]	Rate the perceived quality of the remote place.
Local Quality (LOCQ)[2]	Rate the perceived quality of your local reality.
In-Experience CS (IECS)[3]	Did you feel any sickness or discomfort during the experience?
Post-experience CS (PECS)[3]	Are you feeling any sickness or discomfort after the experience?
Global QoE (GQOE)[4]	How would you rate the quality of the experience globally?
Would Recommend (WDRC)	How likely is that you would recommend this experience to a friend or colleague?

We have used a 3-parameter radial distortion model:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = R \begin{bmatrix} x \\ y \end{bmatrix} = (1 + k_1 r^2 + k_2 r^4 + k_3 r^6) \begin{bmatrix} x \\ y \end{bmatrix} \tag{3}$$

Where (x, y) are the coordinates of the pixel in the distorted image, (x', y') are the coordinates in the rectified image, r is the distance of each pixel to the center of the image, and (k_1, k_2, k_3) are the radial distortion parameters. Once those distortion parameters of the lens have been obtained, the distortion is rectified by projecting the texture provided by the 1:1 camera into a triangle mesh (see Fig. 2), whose vertices are defined by:

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = R \begin{bmatrix} x \\ y \end{bmatrix} \quad \text{for} x, y \in \{-0.5, -0.4, \dots 0.5\} \tag{4}$$

In order to blend the local with the remote reality, the local one is projected onto a square of side W which is centered in front of the user at position $(0, 0, -z_L)$, with $z_L < R_R$ (notice that the remote reality has been already projected it onto a sphere of radius R_R surrounding the user, see Fig. 1 *right*). As there is only one camera in the phone that has been used (as in most smartphones), local reality capture is monoscopic. To ensure that the same image is perceived by both eyes, the position of the local reality plane in the virtual scene z_L is set much higher than the interpupillary distance: $z_L = 100$m. The side length of the square where the camera is projected (W) has been determined by manual fine-tuning, so that the subjective perception of the reality was considered natural: $W = 200$m. The radius of the remote reality sphere has been computed accordingly so that it does not interfere with the camera plane: $R_R = 200$m.

3 Evaluation Questionnaire

We have designed a questionnaire entitled Distributed Reality Experience Questionnaire (DREQ) as a tool to evaluate different Distributed Reality applications (Table 1). It is aimed at covering as many presence and quality aspects with as few questions as possible, thus making it possible to recover feedback from users outside a laboratory environment (e.g. in trade shows).

The first five items considered in DREQ are related to presence and, among all its perspectives [25], we focus on the concept of spatial presence (or place presence), which is the sensation of the user to be actively present in the virtual environment [14]. As in DR both realities (remote and local) are equally important, and their perception is mediated through the HMD in either case, it is necessary to measure the presence in both of them (being present in the remote space without fully abandoning the local environment). Based on recent studies of the reduction of dimensionality of spatial presence questionnaires, we have decided to address the main two components of spatial presence *at the remote reality*: [7]: self-location (*SPRE* in our questionnaire) and possible actions (*RINT*). The same components are also assessed for the spatial presence *at the local reality*: *LPER* and *LINT* respectively. As DR is intended to create experiences with some functionality or task (in this case, tasting the *tapa*), the effectiveness of such task is also measured (*TASK*). Whenever possible, DREQ takes the wording of the items from Witmer and Singer's Presence Questionnaire (PQ) [24].

The next five items cover video quality (two items), global Quality of Experience (QoE, one item) and cybersickness (two items). Those factors are typically measured in the evaluation of immersive 360° video [21, 22], or interactive experiences including immersive video [3], and measuring them allows comparing DR experiences with other video-based studies. As the standardization of subjective assessment of immersive video is still ongoing [9], audiovisual quality is normally measured by reusing the tools from conventional bidimensional video, e.g. Recommendation ITU-T P.910 [8]. Our questionnaire uses the Absolute Category Rating (ACR) scale from that recommendation to measure audiovisual quality and global QoE (*REMQ*, *LOCQ*, *GQOE*).

The most common tool to measure cybersickness (also called "simulator sickness") is the Simulator Sickness Questionnaire (SSQ) [11], which provides four different sickness measurements (oculomotor, nausea, disorientation, and total) from the self-report of 16 symptoms. As adding 16 items to the questionnaire is excessive for our purposes, we have opted for asking about a global measurement of sickness at two different moments: during the experience (*IECS*), as it is where presence and quality measurements refer to, and after the experience (*PECS*), as it is where SSQ refers to. A recently developed scale, specifically designed for 360° video, has been used for it [19].

Finally, users are asked whether they would recommend the experience to a friend or colleague, in a scale of 0 to 10 (*WDRC*). This is used to compute the Net Promoter Score (NPS), a *de facto* standard metric used for customer satisfaction studies [20]. NPS is computed by segmenting users, according to their vote, into detractors (6 or less), neutral (7 or 8) and promoters (9 or 10), so that:

$$NPS = 100\% \frac{\text{Promoters} - \text{Detractors}}{\text{Promoters} + \text{Detractors} + \text{Neutral}} \tag{5}$$

Even though its reported validity as single predictor of customer loyalty and firm growth is arguable [10], the NPS is still widely used due to its simplicity, and has been adopted to assess satisfaction in

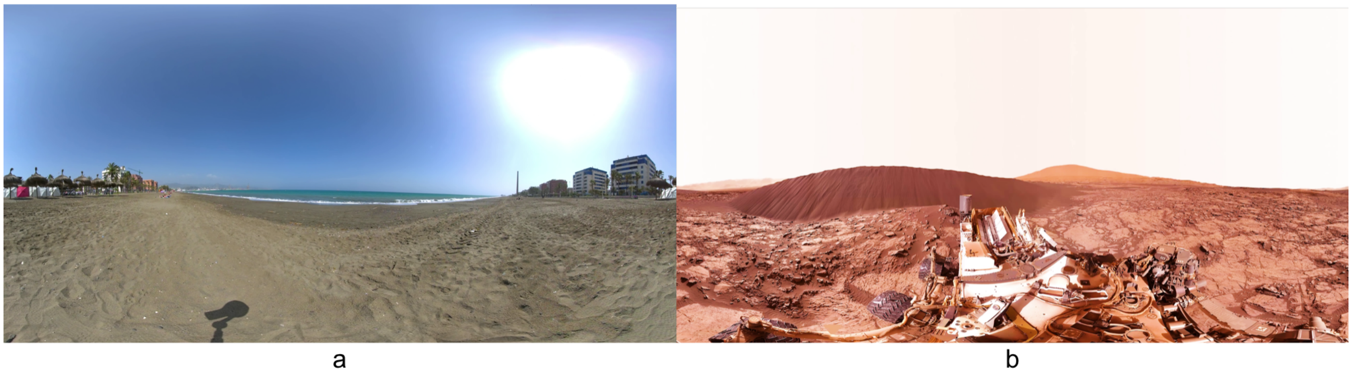

Figure 3: Remote reality videos used for the experience: *a)* Beach shore, *b)* Mars surface seen from NASA Curiosity rover.

Figure 4: A user testing the full gastronomic experience. The system has been set up in a white tent, with a grey metallic structure. The three *tapas* can be seen from right to left: ham toast, *gazpacho*, and mousse. Both the food and the cocktail cutlery have high values of red chrominance, therefore appearing in the Distributed Reality blend.

health care [6] or education [18], though it normally needs to be complemented with other questions or metrics [13].

For each experience and when suitable, questions are reworded to simplify its understanding (e.g. the *TASK* question always refers to the specific task that the user performs in the experience). Items regarding properties that are not present in the experience are removed from the questionnaire (in this particular case the *RINT* item is not included).

In addition, some demographic information is retrieved with the questionnaire: sex, age, occupation, and experience with VR.

4 Gastronomic Experience

The Distributed Reality application described in the previous section has been used to implement a gastronomic experience where users can eat some *tapas* while immersed in a remote place.

4.1 Physical setup

The experience takes place on a white room or tent, with grey or black floor. The space does not need to be completely empty: white, grey or black elements or objects are allowed as well, as they will not be seen in the blended reality. As can be seen from Fig. 4, the user is sitting on a rotating bar chair in front of a bar table, both white. The food is selected to have red, yellow or brown tones, so that it is visible through the segmentation algorithm, and it is presented on small plates of the the same colors on top of the table. White uniform diffuse light is used to ensure that the segmentation works correctly. With this setup, the user can only see her own hands (and her arms, if she is wearing a short sleeve shirt), plus the food and the plate. She can also sit comfortably and look around with the rotating chair, but the height of the chair and the position of the table forces the rotation to be slow in a natural way, as it would happen in a conventional bar environment. Limiting the number and size of visible local reality elements in the experience, as well as limiting the rotation freedom of the user, allows minimizing the cybersickness in the experience despite the relatively reduced resolution and frame rate of the local reality video capture.

4.2 Performance testing

The Distributed Reality application was tested in a Samsung Galaxy S8 with GearVR Framework by Oculus. A 10 mm fish-eye lens (Apexel APL 10MM) was attached in front of the smartphone lens to increase the field of view. The VR framework worked consistently at a refresh rate of 60 fps. Both local and remote videos were correctly displayed at their respective native resolutions and frame rates (30 fps). Segmentation algorithm was consistently executed in less than 15 ms, thus achieving the target of less than one frame of delay in the segmentation. The setup was completed with high-quality headphones (Bose QuietComfort 25) to achieve good audio quality and aural isolation from the environment.

4.3 Experience design

With the setup that has already been described, a complete gastronomic experience was designed: the tasting of different *tapas* paired with their respective immersive environments. The menu was designed by a professional chef, and it consisted of: *i)* ham toast, paired with a video of a quiet environment in a forest; *ii)* raspberry *gazpacho*, paired with a video of the beach shore; and *iii)* orange and berries mousse, paired with a video from *Curiosity* mission on Mars surface. All the videos included ambient sound and voice-over description of the *tapa*. The full experience duration was 8 minutes. Additionally, an introductory 5 minute 360° video was shown to the users just before the experience, including a recording of the chef preparing and explaining the *tapas* in his kitchen. This full version of the experience was successfully demonstrated in a public event in the city of Segovia (Fig. 4).

Due to the difficulty to replicate the full environment of the experience in a laboratory, a reduced version was also prepared for technical evaluation (a formal evaluation of the full environment is left for future work). In this reduced version, only the last two videos were shown (beach and Mars, Fig. 3), with a total duration of 5 minutes. The professional *tapas* were also replaced by simpler versions, and the voice-over was removed. This way, all the technical and presence aspects are preserved, while removing the dependency

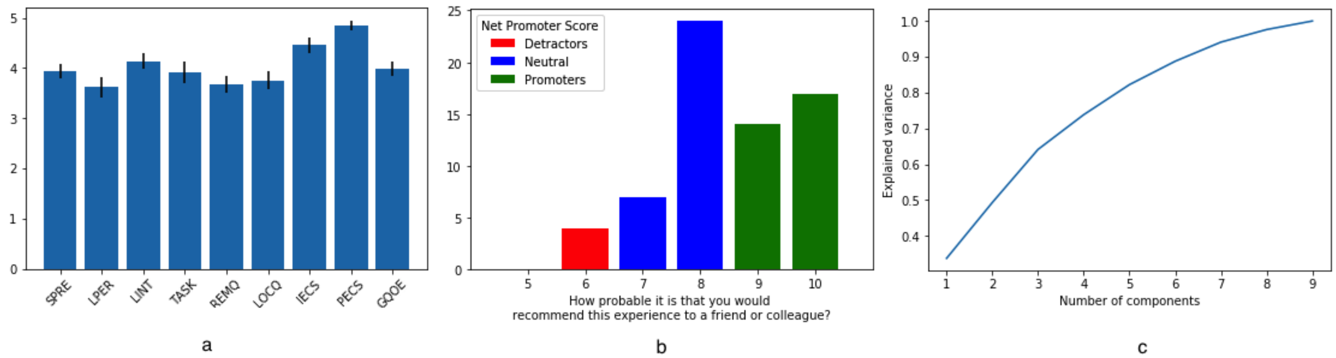

Figure 5: *a)* Mean and .95 confidence intervals for each of the main nine DREQ items. *b)* Histogram of votes for "Would Recommend" question and segmentation in detractors, neutral and promoters. *c)* Principal Component Analysis of the main nine DREQ items.

of the availability of a professional chef to prepare the food, at the cost of some reduction of the whole experience appeal.

4.4 Evaluation

The reduced version of the experience was evaluated in a local technology event. Event attendees were invited to test the experience and fill the DREQ. 66 people participated in the evaluation, 80% male and 20% female, between 22 and 55 years old, all of them with some degree of technological background, but with no direct relationship with the researchers.

Most of the questions received positive feedback from most of the users. Fig. 5.a shows the mean values and confidence intervals for the nine main items from the Distributed Reality Experience Questionnaire. Mean values for presence, video quality and Quality of Experience range from 3.6 (*LPER*) to 4.1 (*LINT*), indicating that the experience is *Good* on average. There are no statistically significant differences between the different items of these categories. Cybersickness, on the other hand, scores slightly better: 4.5 on average during the experience, and 4.9 after that. Half of the users (33) reported some level of discomfort of during the experience, though most of them (31) reported just light effects (value 4 of the scale). 85% of the users reported not having any symptom at all after the experience. Fig. 5.b shows the histogram of answers to *WDRC* question. The resulting Net Promoter Score value (40%) shows that the experience was really well appreciated by the users [20].

Those results suggest that the experience has good chance of being accepted by potential users. The technical limitations that were described in Section 2 with respect to the limited frame size and frame rate, especially of the local video, have only moderate impact in the perceived quality. Additionally, the design of the experience has proved to be successful in limiting cybersickness effects to the minimum, as average discomfort values are quite high.

As already mentioned, Fig. 5.a shows that mean values of the different items in the DREQ vary within a small range. This could suggest that most of the values are tightly related and therefore show a single global opinion of the experience strongly influencing the response to all the questions. To verify this hypothesis, we have computed the Principal Component Analysis (PCA) of the nine main items from DREQ, as well as their pairwise coefficient correlations. The graphic of accumulated explained variance of the PCA components is shown in Fig. 5.c. It is observed that the first component of the PCA only explains 40% of the variance, which means that there is some correlation between the factors, but it is not not extremely high. 7 out of 9 PCA components are needed to account for 90% of the variance. Additionally, the pairwise Pearson correlation coefficient between the different items is small, ranging from 0.04 up to 0.47. It can be concluded that the different

Table 2: ANOVA table of the influence of different factors in the NPS question. Factors labeled with an asterisk are significant with a *p*-value below 0.05. The effect is moderate ($\eta^2 \approx 0.15$).

Factor	sum_sq	df	F	PR(>F)	η^2
SPRE	0.386	3.0	0.138	0.936	0.004
LPER	4.389	3.0	1.568	0.212	0.048
LINT*	12.747	3.0	4.555	**0.008**	**0.138**
TASK	1.750	4.0	0.469	0.758	0.019
REMQ*	14.892	3.0	5.322	**0.004**	**0.161**
LOCQ	4.306	3.0	1.539	0.220	0.047
IECS	4.592	3.0	1.641	0.196	0.050
PECS*	12.156	2.0	6.517	**0.004**	**0.132**
GQOE	0.522	2.0	0.280	0.757	0.006

items measured in DREQ are relatively independent among them, thus verifying that they represent different quality factors, as it was intended with the questionnaire design. As a side effect, the low inter-factor consistency does not allow to build a composite quality index by aggregating all the factors: Cronbach's *alpha* for a composite index of the nine main items is just .72. Therefore we will use the global QoE and, more significantly, the NPS, as global indexes of quality and user satisfaction.

We have computed a one-way ANOVA for the response to *WDRC* against each of the main items. The results are displayed in Table 2. According to our experiment, local interaction, remote quality, and post-experience cybersickness are the ones whose variation affects significantly the answer to *WDRC*, and therefore to NPS. This suggests that those dimensions might be differential when determining the acceptability of a Distributed Reality application, and they should be taken into consideration for future designs.

5 Conclusions

We have implemented a simple system, yet very effective, to test the concept of Distributed Reality. Unlike other MR approaches, DR merges two realities: local reality captured with video see-through (for instance eating some food) while visiting a remote reality (e.g. the beach) captured as a 360° video. With its limitations, our system provides an effective framework to test DR applications, as it includes a very flexible method to introduce local elements into the remote reality, using an Augmented Virtuality approach based on red chrominance segmentation.

To be able to evaluate the user opinion of the experience, the Distributed Reality Experience Questionnaire has been designed, covering the most relevant factors influencing spatial presence, audiovisual quality, cybersickness, and QoE. Results show that the selected questions show a good level of independence considered pairwise, which proves that distinct presence and quality factors are actually being accounted for with the different items of the questionnaire.

We have tested the system with a reduced version of the experience. It was evaluated by 66 users, which provided positive feedback of the experience, reaching a 40% value of Net Promoter Score (NPS). Mean quality of experience was 4.0 (*Good*) and cybersickness measurement results were even better (light to no effect during the experience, no remaining effect after it). Therefore we can conclude that the system that we have presented in this article can be used *as-is* to deploy distributed reality experiences.

We have developed a fully functional DR experience: eating a variety of *tapas* in different remote places, which are selected to be paired with the food by a professional chef. Having a complete experience which is close to a functional use case is beneficial for two reasons: on the one hand, it allows system tests to have ecological validity (the system is tested under realistic conditions); on the others, it demonstrates the potential of the distributed reality approach for the industry (in this case, for hospitality and tourism).

In our future work we will also use this framework to explore the DR concept more in depth. On the one hand, it will be the baseline platform to test new DR use cases, such as bidirectional communication or remote car driving, where there is more interactivity between both realities. The most relevant factors influencing the NPS are the ability to interact with the local elements (the food), the quality of the remote video, and the long-term effect of cybersickness, which suggest us future research lines. Addressing them should be the target of following versions of this technology, which can focus on: deep-learning based object segmentation, stereoscopic local camera, better resolution and frame rate, high-end VR HMD, etc.

Acknowledgments

The authors wish to thank chef Diego Isabel for his collaboration in the preparation of the experience, as well as Telefónica Investigación y Desarrollo and Hotel San Antonio del Real in Segovia (Spain) for their support in the field test. A special thanks for Ignacio Benito, David Tenorio, and Miguel Garrido for their help in the preparation of the test sequences and infrastructure. This work has been partially supported by Spain's Center for the Development of Industrial Technology (CDTI) under project CDTI IDI-20180015 (VINEDO).

References

[1] I. N. S. K. Andersen, A. A. Kraus, C. Ritz, and W. L. Bredie. Desires for beverages and liking of skin care product odors in imaginative and immersive virtual reality beach contexts. *Food Research International*, 2018.

[2] G. Bruder, F. Steinicke, K. Rothaus, and K. Hinrichs. Enhancing presence in head-mounted display environments by visual body feedback using head-mounted cameras. In *International Conference on CyberWorlds*, pp. 43–50. IEEE, 2009.

[3] K. Brunnström, M. Sjöström, M. Imran, M. Pettersson, and M. Johanson. Quality of experience for a virtual reality simulator. In *Human Vision and Electronic Imaging (HVEI), Burlingame, California USA, 28 January-2 February, 2018*, 2018.

[4] A. P. Desai, L. Pena-Castillo, and O. Meruvia-Pastor. A window to your smartphone: exploring interaction and communication in immersive vr with augmented virtuality. In *Computer and Robot Vision (CRV), 2017 14th Conference on*, pp. 217–224. IEEE, 2017.

[5] E. J. Fernandez-Sanchez, J. Diaz, and E. Ros. Background subtraction based on color and depth using active sensors. *Sensors*, 13(7):8895–8915, 2013.

[6] D. Hamilton, J. V. Lane, P. Gaston, J. Patton, D. Macdonald, A. Simpson, and C. Howie. Assessing treatment outcomes using a single question: the net promoter score. *The bone & joint journal*, 96(5):622–628, 2014.

[7] T. Hartmann, W. Wirth, H. Schramm, C. Klimmt, P. Vorderer, A. Gysbers, S. Böcking, N. Ravaja, J. Laarni, T. Saari, et al. The spatial presence experience scale (SPES). *Journal of Media Psychology*, 2015.

[8] ITU-T. P.910: Subjective video quality assessment methods for multimedia applications, 04 2008.

[9] ITU-T. P.360-VR: Subjective test methodologies for 360 degree video on HMD (Under Study), 2018.

[10] T. L. Keiningham, B. Cooil, T. W. Andreassen, and L. Aksoy. A longitudinal examination of net promoter and firm revenue growth. *Journal of Marketing*, 71(3):39–51, 2007.

[11] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology*, 3(3):203–220, 1993.

[12] D. Korsgaard, N. C. Nilsson, and T. Bjørner. Immersive eating: evaluating the use of head-mounted displays for mixed reality meal sessions. In *IEEE 3rd Workshop on Everyday Virtual Reality (WEVR)*, pp. 1–4. IEEE, 2017.

[13] M. W. Krol, D. de Boer, D. M. Delnoij, and J. J. Rademakers. The net promoter score–an asset to patient experience surveys? *Health Expectations*, 18(6):3099–3109, 2015.

[14] M. Lombard and T. Ditton. At the heart of it all: The concept of presence. *Journal of Computer-Mediated Communication*, 3(2), 1997.

[15] M. McGill, D. Boland, R. Murray-Smith, and S. Brewster. A dose of reality: Overcoming usability challenges in VR head-mounted displays. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pp. 2143–2152. ACM, 2015. doi: 10.1145/2702123.2702382

[16] P. Milgram. A taxonomy of (real and virtual world) display and control interactions. In *Proceedings of the 16th ACM Symposium on Virtual Reality Software and Technology*, VRST '09, pp. 10–10. ACM, New York, NY, USA, 2009.

[17] D. Nahon, G. Subileau, and B. Capel. "never blind vr" enhancing the virtual reality headset experience with augmented virtuality. In *2015 IEEE Virtual Reality (VR)*, pp. 347–348. IEEE, 2015.

[18] K. Palmer and C. Devers. An evaluation of mooc success: Net promoter scores. In *EdMedia+ Innovate Learning*, pp. 1648–1653. Association for the Advancement of Computing in Education (AACE), 2018.

[19] P. Perez, N. Oyaga, J. J. Ruiz, and A. Villegas. Towards systematic analysis of cybersickness in high motion omnidirectional video. In *QoMEX*, 2018.

[20] F. F. Reichheld. The one number you need to grow. *Harvard business review*, 81(12):46–55, 2003.

[21] R. Schatz, A. Sackl, C. Timmerer, and B. Gardlo. Towards subjective quality of experience assessment for omnidirectional video streaming. In *Proc. 9th Int. Conf. Qual. Multimedia Exp.(QoMEX)*, pp. 1–6, 2017.

[22] A. Singla, S. Fremerey, W. Robitza, P. Lebreton, and A. Raake. Comparison of subjective quality evaluation for hevc encoded omnidirectional videos at different bit-rates for uhd and fhd resolution. In *Proceedings of the on Thematic Workshops of ACM Multimedia 2017*, pp. 511–519. ACM, 2017.

[23] F. Tecchia, G. Avveduto, R. Brondi, M. Carrozzino, M. Bergamasco, and L. Alem. I'm in vr!: using your own hands in a fully immersive mr system. In *Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology*, pp. 73–76. ACM, 2014.

[24] B. G. Witmer and M. J. Singer. Measuring presence in virtual environments: A presence questionnaire. *Presence*, 7(3):225–240, 1998.

[25] C. Youngblut. Experience of presence in virtual environments. Technical report, Instutute for Defense Analyses, Alexandria, VA, 2003.

[26] Z. Zhang. A flexible new technique for camera calibration. *IEEE Transactions on pattern analysis and machine intelligence*, 22, 2000.