Deliverable D2.4.2

ICoSOLE identifier: ICoSOLE-D2.4.2-BBC-Evaluations2_v12_FINAL.docx
Deliverable number: D2.4.2
Main author of Deliverable: Dave Marston (BBC), Rik Bauwens (VRT), Nikki Peeters (VRT), Mike Matton (VRT)
Internal reviewer: Georg Thallinger (JRS)

Work package / task: WP2 / T 2.3
Document status: Final
Confidentiality: Public

<table>
<thead>
<tr>
<th>Version</th>
<th>Date</th>
<th>Reason of change</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2016-09-16</td>
<td>Document created</td>
</tr>
<tr>
<td>2</td>
<td>2016-09-23</td>
<td>BBC content added</td>
</tr>
<tr>
<td>3</td>
<td>2016-09-28</td>
<td>iMinds content added</td>
</tr>
<tr>
<td>4</td>
<td>2016-10-06</td>
<td>JRS content added with some BBC bits</td>
</tr>
<tr>
<td>6</td>
<td>2016-10-12</td>
<td>VRT input</td>
</tr>
<tr>
<td>7</td>
<td>2016-10-19</td>
<td>BBC input</td>
</tr>
<tr>
<td>8</td>
<td>2016-10-23</td>
<td>iMinds input</td>
</tr>
<tr>
<td>9</td>
<td>2016-10-25</td>
<td>Version for internal review</td>
</tr>
<tr>
<td>11</td>
<td>2016-11-07</td>
<td>Addressing review comments</td>
</tr>
<tr>
<td>12</td>
<td>2016-11-10</td>
<td>Final version</td>
</tr>
</tbody>
</table>
The work presented in this document was partially supported by the European Community under the 7th framework programme for R&D.

This document does not represent the opinion of the European Community, and the European Community is not responsible for any use that might be made of its content.

This document contains material, which is the copyright of certain ICoSOLE consortium parties, and may not be reproduced or copied without permission. All ICoSOLE consortium parties have agreed to full publication of this document. The commercial use of any information contained in this document may require a license from the proprietor of that information.

Neither the ICoSOLE consortium as a whole, nor a certain party of the ICoSOLE consortium warrant that the information contained in this document is capable of use, nor that use of the information is free from risk, and does not accept any liability for loss or damage suffered by any person using this information.
Table of Contents

List of Figures ........................................................................................................................................... v

List of Tables ............................................................................................................................................... vi

1 Executive Summary ................................................................................................................................. 1

2 Introduction ............................................................................................................................................... 2
  2.1 Purpose of this Document .................................................................................................................. 2
  2.2 Scope of this Document ...................................................................................................................... 2
  2.3 Status of this Document ..................................................................................................................... 2
  2.4 Related Documents ........................................................................................................................... 2

3 Terminology ............................................................................................................................................. 3

4 Field Trial Overviews ............................................................................................................................... 4

5 User evaluations ....................................................................................................................................... 5
  5.1 Committee of the Regions ................................................................................................................ 5
  5.2 Leffingeleuren Music Festival .......................................................................................................... 8
  5.3 Immersive Experiences of Music Festivals from the Living Room ................................................. 13
  5.4 Multi-depth Layered Video Streaming .............................................................................................. 17
  5.5 BBC Taster Trials – Proms in Binaural ............................................................................................ 25
  5.6 General Feedback from Trials ......................................................................................................... 28

6 Technical evaluations ............................................................................................................................... 30
  6.1 Professional Audio and Video Capture and Live Streaming ............................................................. 30
  6.2 Immersive Audio Production .......................................................................................................... 30
  6.3 Content Synchronisation .................................................................................................................. 30
  6.4 Content Selection ............................................................................................................................... 34
  6.5 Visual Matching ................................................................................................................................ 37
  6.6 Multi-depth Layered Video Streaming ............................................................................................. 40

7 Conclusions ............................................................................................................................................. 44

8 References ............................................................................................................................................... 45

9 Glossary ................................................................................................................................................... 46

10 Appendix .............................................................................................................................................. 47
  10.1 WanderCouch Interview Script ....................................................................................................... 47
  10.2 Multi-depth Layered Video Streaming User Study Survey of Demographics ............................... 54
  10.3 Multi-depth Layered Video Streaming User Study Written Instructions ..................................... 55
  10.4 Multi-depth Layered Video Streaming Interview Script ................................................................. 56
  10.5 Leffingeleuren Questionnaire (Dutch) ............................................................................................. 57
List of Figures

Figure 1: Number of Moments uploaded by participants 5
Figure 2: Motivation to create Moments 5
Figure 3: General appreciation of the Moments app 6
Figure 4: Appreciation of quality feedback 6
Figure 5: Edit clips or not? 7
Figure 6: Watched Wall of Moments 7
Figure 7: Effect of the Wall of Moments 8
Figure 8 Representative frames taken from respectively the captain, concert and poolhall videos. 17
Figure 9 Quality comparison: traditional (top) versus Chroma Keying (middle) versus Alpha Mask (bottom). 20
Figure 10 Absolute preference numbers for each of the streaming technique comparisons and tested video clips. 22
Figure 11 Examples of visual gaze analysis results that illustrate a user’s artefact scanning behaviour roughly along the boundaries of the foreground object. For the sake of comprehensibility, the boundary of the foreground object has been outlined in the topmost image. 24
Figure 12: BBC Taster Proms - visitor stats 26
Figure 13: BBC Taster Proms - binaural listener views 27
Figure 14: BBC Taster Proms - more binaural listener views 27
Figure 15: Example frame generated for time measurement from the image. 31
Figure 16: Measurement of synchronisation results for three different network conditions. 32
Figure 17: Frame from capturing the PTP display at the production machine in Edinburgh. 33
Figure 18: Frame captured on UGC IP Studio machine 34
Figure 19: Prediction results for the LNF event, using a model trained on the COR data. 35
Figure 20: Prediction results for the Leffingeleyuren event, using a model trained on the COR data. 36
Figure 21: Fraction of false positives if automatic selection is applied at a specific cut-off score, measured on the Leffingeleyuren data. 37
Figure 22: True positive matching rate at fixed false positive rate for different configurations and lossy compression rates (for different values of $\theta$) of local descriptors. 38
Figure 23: Jaccard index of temporal localisation for different configurations and lossy compression rates (for different values of $\theta$) of local descriptors. 39
Figure 24: Resulting global and local descriptor sizes for different configurations and lossy compression rates (for different values of $\theta$) of local descriptors 39
Figure 25 The experimental setup of the subjective evaluation of the MLV streaming methodology incorporated an optical gaze tracker. 41
Figure 26 Foreground gaze hit ratio for a specific test participant. 42
List of Tables

Table 1 Average time spent per constituent WanderCouch screen. ................................................. 14
Table 2 Usage of WanderCouch’s UGC consumption screen across test subject populations. .......... 15
Table 3 Average results of test users’ self-reported WanderCouch experience. .............................. 15
Table 4 Back- versus foreground bitrate, in kbps. ........................................................................ 18
Table 5 Objective video quality assessment results for the 27 stimuli featured in the subjective evaluation of the MLV streaming methodology; traditional (TR) versus Chroma Keying (CK) versus Alpha Mask (AM) implementation. .................................................. 40
1 Executive Summary

This deliverable contains the results and analysis of user studies and technical evaluations that have taken place during the project period after August 2015. The main evaluations have been organised around the field trials at the Edinburgh Festival in August 2016, and the Leffingeleuren music festival in September 2016.
2 Introduction

2.1 Purpose of this Document
Evaluations v2: Report on technical evaluations and field trials.

2.2 Scope of this Document
This deliverable covers results of technical and user tests that have been performed up to month 36 (September 2016) of the project.

2.3 Status of this Document
Final version.

2.4 Related Documents
The first evaluations document is deliverable [D2.4.1] which covers the field trials in the first period of the project.
3 Terminology

Due to different areas of study often using the same terminology for different purposes, it can lead to confusion when these areas of study are brought together. In this report, we have this situation where the word ‘immersion’ or ‘immersive’ has two distinctly different meanings depending upon the context in which it is used:

- **Immersive audio** - Often the phrase ‘3D audio’ is used, and it means audio that is perceived to come from any direction around the listener. Immersive audio is a step up from surround audio, which only covers the horizontal plane.

- **Immersive experience** – This is where a performance in a remote event is being presented to a user and how involved or immersed they feel in that event. Another word could be ‘presence’, as this is how they feel detached from their current physical situation (such as sat in a living room), and being part of the media they are watching and listening to.

Clearly the purpose of immersive audio is to give the listener an immersive experience, however the immersiveness of an experience is dependent upon many factors including the user’s desire to be immersed.
4 Field Trial Overviews

Three major field trials took place in the last 12-month period of the project. These were at the Edinburgh Festival in Scotland (http://www.eif.co.uk/), the Leffingeleuren music festival in Belgium (http://leffingeleurenfestival.be), and the Committee of Regions in Brussels. Further details of the field trials can be found in D6.4 "Second Demonstration and Field Trial". There have also been other trials carried out which have been useful in gathering relevant results for the project.
5 User evaluations

5.1 Committee of the Regions

Shortly after the closing of the Committee of the Regions event, we sent out a survey to our 29 test users. Most of them were students or journalists, ranging from 20 to 40 years old.

The response to our call for sending in videos was higher than expected, which is confirmed by the pie chart below (Figure 1).

![Figure 1: Number of Moments uploaded by participants](Image)

People made clips for various reasons, but most of the time, clips were made with the purpose of testing the app.

![Figure 2: Motivation to create Moments](Image)

In general, the appreciation for the new version of the app was positive, with only 11.1% giving it a score below 5/10 (Figure 3).
Figure 3: General appreciation of the Moments app

Also the feedback to how well someone filmed was well received (Figure 4).

Figure 4: Appreciation of quality feedback

Most interestingly, most of the time, people didn’t edit their clips at all before uploading them to the Wall of Moments (Figure 5). This was in contrast with the observation after Dranouter, where people indicated this as a much-wanted feature missing from that prototype. This could also be due to the nature of the event: a lot of youngsters, present at a festival, tend to have a higher motivation to experiment with filters and video effects.
A lot of people also showed interest in the screens displaying the Walls on several strategic locations (Figure 6), especially when their own clip was displayed.
5.2 Leffingeleuren Music Festival

All test users at the Leffingeleuren field trial have been asked to complete a questionnaire. We distributed the link to the online questionnaire (see questionnaire in Section 10.5) via social media and e-mail. We received 17 responses from the nearly 100 test users.

5.2.1 General analysis and main conclusions

The concept of our second field trial was to offer multiple sharing opportunities, combine a livestream with footage from festivalgoers and be able to make full use of the feedback tool, to engage with our test users. Via the Moments app, test users could upload the clips and pictures they make with different apps. They would be updated about what was happening to their clips (uploaded, being processed, awaiting approval…). Our research team was able to send requests and respond to inquiries via direct messaging integrated into the Moments and Trademark app. The uploading of footage went really well. We were even able to push pictures posted on social media (Twitter and Instagram) to the Wall.

Although test users are able to alter their footage via other apps, they would have also valued the possibility to add filters and make adjustments directly in the Moments app. They wanted the app to be more than a medium to send content to the Wall, they would like it to generate and alter content as well. They also valued the instant feedback on their content but wanted even more information, such as where their clips are used (i.e. when they were shown on which Wall). This is a feature that was available in a previous iteration of the app, but was not integrated in the app used for Leffingeleuren, due to timing constraints. They were able to see what was happening to their clip, e.g. whether it was uploading or under review, but were also annoyed when they saw that the clips were not uploading without knowing the specific reason.

Figure 7: Effect of the Wall of Moments

<table>
<thead>
<tr>
<th>Option</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>A I was pleased to see my own video</td>
<td>7 (87.5%)</td>
</tr>
<tr>
<td>B It was interesting to see what was happening on the different venues of the EU Open Days</td>
<td>3 (37.5%)</td>
</tr>
<tr>
<td>C No effect at all</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>D I didn’t watch the screens</td>
<td>1 (12.5%)</td>
</tr>
<tr>
<td>E Other</td>
<td>0 (0%)</td>
</tr>
</tbody>
</table>
Reporters also appreciated the direct messaging capabilities and took their reporting to a next level, filming specific content by receiving guidance via the app. Furthermore, test users suggested that dividing tasks and instructions among reporters would make the app more interactive. Interviewing a performer or filming from a specific angle to cover the entire performance would be examples of the options. At first, we did not give tasks because we wanted to let the test users use the app freely. Starting at the end of prep day, we started to post tasks and polls via Trademark to boost interaction. On the last day, we experimented by posting challenges as part of a contest. For example, we asked them to post clips of them playing air guitar. Feedback later proved test users enjoyed these extra assignments.

Another remark raised by test users was that the Wall of Moments did not capture sound. During the festival, this seems obsolete. For visitors of the Wall at home, however, sound was available to get a full grasp of the festival.

Test users are happy to see their own clips on screens throughout the festival. Therefore, they wanted to know when their clips were being shown on the screens, be able to see the entire Wall of Moments via the same app (for this test, we used a separate app) and have more screens on the festival. The screens displaying the Wall of Moments definitely proved to be an added value. Test users wanted even more visibility than the screens we were able to install on-site. Due to high costs of large LED walls and the extra overhead of providing internet at more locations on the festival site, we were limited in how many screens we placed, as well as how bright and large they could be. The quality of the videos on the screens was sufficient for most of our test users.

Aside from the Wall of Moments onsite, we also shared the Wall on the web and mobile (iOS/Android) via social media for friends and family at home to watch an interactive version of the Wall. When they signed in, the current Wall was shown. Test users could click or tap the current Moment to go to the next one. They could tap and hold to rewind to the most recent Moment, swipe to the left to fast-forward and swipe to the right to reveal a menu. This menu allowed access to a list of our reporters, along with an individual playout of the approved Moments they shared.

The Wall displayed an (intelligent) loop of sent-in Moments. The question was raised in which order the videos were being shown (research team selection, duration, popularity…). Some found it interesting to see videos from the past days, other found this to be “in contrast with” the live stream.

5.2.2 Survey results

Once again, our test users were very active. On average, we received about 10 clips per user. A breakdown can be found below.
We also asked what users would have done differently in the app, such as enhancements, tips and ideas. We got some interesting suggestions, by most valued first:

- Ability to see the clips of others in the app
- Ability to add filters and effects via the app
- Ability to alter clips before uploading
- Adding sound to the Wall of Moments
- The speed of the application
- Making it accessible to other types of smartphones
- Making the length of the clips longer
- Allowing shorter clips (for example Snapchat)
- A button with which you can switch to the Wall of Moments

While there was a possibility to see clips from others, this functionality was not integrated into the Moments app. Technically; it was easier to package the Wall in another app. It would have been an improvement, however, to add a cross-link in both apps to make transitioning between apps easier.

Most of the test users made clips solely for the purpose of the Wall of Moments, and to share their Moments with friends (who were able to follow Leffingeleuren via the Wall app).
Most of our test users also visited the ICoSOLE – Wall of Moments booth at Leffingeleuren. They came in for questions, made suggestions and to follow what was happening on the festival terrain on the two screens we installed in the booth.

We also offered the Wall as a standalone app as a catchup service for our test users or for friends and family at home or underway. The Wall was available on the web, iOS and Android. According to our test users, this was more a nice-to-have feature.
Lastly, we asked our test users how they would rate the Moments app with a score from 0 (lowest) to 10 (highest). It turned out that most of them had a positive to very positive experience.

Also the comments and suggestions about the app yielded overall positive results. An extract can be found below.

**What do you like about the Wall of Moments?**

- The experience of contributing to such a project.
- The entire feeling.
• Nice concept because people are able follow it.
• The general concept of portraying the atmosphere of a festival. I therefore found the challenges of day 3 to be an added value.
• That the public is engaged as a medium for sharing with a broader audience, outside of the concert.
• Capturing the random moments apart from the performances. The flags surrounding the church, the first beer etc... It is also nice to see your own contributions on the televisions. I did expect them to be distributed on a larger scale than only the booth and bar.
• Sharing content in a fast way.
• Everything comes together on the screens across the festival. Now and then you see your own clip returning or an image of a band you were not able to see.
• Nice to browse through the clips.
• Nice concept.
• The live part.
• That everyone can share his or her moments with others. If you would use this app on bigger festivals, then it is handy to gather the atmosphere on other locations of the terrain. That way, you do not have to walk the distance and finding out that you liked it better some place else, or the other way around.
• The atmosphere was there sometimes, and the reviews of the performances.
• It is nice that you can co-create – (can) be part of the visual reproduction of the festival experience.
• People who are not able to make it to the festival can also enjoy the music, atmosphere...
• It is nice to gather the festival from the perspective of non-professional photographers, more personal.

5.3 Immersive Experiences of Music Festivals from the Living Room

Last project year’s deliverable [D2.4.1] has reported on iMinds’ initial efforts in subjectively evaluating the WanderCouch prototype, a Smart TV application for the immersive experience of music festivals from the living room. In the final year of the ICoSOLE project's lifecycle, this evaluation has been broadened by adding 5 prospective users test the prototype. Other than enlarging the participant sample (which grew from 6 to 11 test subjects), no changes were applied to either the experiment setup, the scientific objectives, the featured content, or the employed test methodology. Therefore, the interested reader is kindly referred to deliverable [D2.4.1] for a thorough discussion of either of these aspects of the experimental design. Here, we will focus on describing the demographic characteristics of the 5 extra test participants, on disambiguating the intent of the WanderCouch evaluation, and on offsetting the findings from the extended test population against the original results reported in deliverable [D2.4.1].

5.3.1 Demographics of the extended test population

The 5 extra participants (1 female) who were recruited for extending the subjective WanderCouch evaluation were either colleagues from our research institute or acquaintances of such colleagues. Each of the additional test users was a music festival enthusiast and was between 22 and 45 years old. Furthermore, they all had prior experience with music festivals, both through physical attendance and via broadcast coverage of such real-world events on (Belgian) TV. In line with the recruitment criteria employed during the initial qualitative WanderCouch test, the new participants felt some form of affinity for at least 2 of the 3 musical artists appearing in the study. Finally, as was the case with the original test population, none of the additional test subjects had ever visited the Dranouter music festival in

1 Please see deliverable [D5.4] for a description of the WanderCouch prototype.
This is important, as part of the subjective evaluation revolved around investigating WanderCouch's ability to grant users the opportunity to form a “mental map” of the festival environment and to grasp its geospatial traits.

5.3.2 Intent and objectives

During the Y2 review of the ICoSOLE project, some of the assessors raised concerns about the presented WanderCouch evaluation. We believe these concerns primarily stem from a lack of proper contextualization of the intent and objectives of the study.

The WanderCouch user study was never meant to be a formal one. For example, we deliberately did not resort to specialized methods like the ITC-SOPI scale to implement immersion and presence assessment [Lessiter, 2001]; instead, easily understandable, high-level statements were formulated to gauge informants’ opinions concerning these measures. This decision is grounded on the fact that the overall intent of the user study was to evaluate, at a rather high level of abstraction, the potential held by the WanderCouch application towards convincingly experiencing distant music festivals (and, while doing so, to collect intuitive feedback from representatives of the application's target audience). Please also note that the WanderCouch approach does not aim to accurately imitate a real-world event visit but instead is designed to produce a novel type of substitute event experience, which further convinced us not to factorize complex measures like immersion and engagement into their contributing components in the post-usage evaluation stage.

5.3.3 Results

Please recall from deliverable [D2.4.1] that the performed evaluation yielded two types of results: quantitative findings that emanated from the observational part of the test methodology, and subjective findings that were collected during the post-experience self-reporting and interviewing phase. For each result category individually, we will now investigate the impact of extending the test population by comparing the findings yielded by respectively the initial and extended participant sample.

It is explicitly stressed here that the WanderCouch user study was not intended to be statistically significant. Instead of targeting statistically relevant insights, the subjective evaluation aimed to collect intuitive, high-level feedback from test users. In line with this intent, the result comparisons that will be discussed in the remainder of this subsection will also be presented in a rather informal manner (i.e., without resorting to any form of statistical analysis).

**Observational study**

Table 1 summarizes the main differences between the observational study results pertaining to respectively the original and the extended test population. In particular, this table lists the average amount of time spent on each composing screen of the WanderCouch prototype by respectively the original and extended participant sample set.

<table>
<thead>
<tr>
<th>Table 1 Average time spent per constituent WanderCouch screen.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>WanderCouch screen</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Concert playback screen</td>
</tr>
<tr>
<td>- Triggerfinger</td>
</tr>
<tr>
<td>- Intergalactic Lovers</td>
</tr>
<tr>
<td>- Bart Peeters</td>
</tr>
<tr>
<td>360 degree video</td>
</tr>
</tbody>
</table>

Footage from the ICoSOLE Y2 Dranouter field trial was utilized to content-wise populate the WanderCouch prototype for the purpose of conducting the user study.

This time is the sum of the viewing time of the concert playback video and the 360° video. In the extended sample set, the sum of 03:38 is calculated analogously.
Table 2 communicates either absolute or average figures for the amount of UGC items consumed, the number of UGC ratings issued, and the number of UGC content filters installed by the initial versus the enlarged test participant sample. All interactions enumerated in Table 2 were performed in the context of the UGC consumption screen of the WanderCouch prototype.

**Table 2 Usage of WanderCouch’s UGC consumption screen across test subject populations.**

<table>
<thead>
<tr>
<th></th>
<th>Original Sample Set (6 participants)</th>
<th>Extended Sample Set (11 participants)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Average number of UGC items consumed</strong></td>
<td>27</td>
<td>33</td>
</tr>
<tr>
<td>(on the UGC screen)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Pictures</td>
<td>10</td>
<td>15</td>
</tr>
<tr>
<td>- Videos</td>
<td>17</td>
<td>18</td>
</tr>
<tr>
<td><strong>Absolute number of UGC ratings issued</strong></td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td><strong>Absolute number of content filters applied</strong></td>
<td>9</td>
<td>19</td>
</tr>
</tbody>
</table>

**Self-reported experience**

Recall from the evaluation methodology description in deliverable [D2.4.1] that, after users had tested the WanderCouch prototype, they were asked to reflect on their experience by rating a number of statements using a 5-point Likert scale (with the middle response option indicating a neutral position). For processing and analysis purposes, the issued ratings were encoded by mapping the 5 response categories to the [1,5] numerical interval (i.e., values 1 and 5 respectively denote a “strongly disagree” and “strongly agree” opinion).

Table 3 compares the results with respect to self-reported WanderCouch experience, both for the original and the extended test population. All numbers in Table 3 represent average values of the numerically encoded responses to the 5-point Likert scale questions; higher values denote better scores, with 5 being the highest achievable score.

**Table 3 Average results of test users’ self-reported WanderCouch experience.**

<table>
<thead>
<tr>
<th></th>
<th>Original Sample Set (6 participants)</th>
<th>Extended Sample Set (11 participants)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Immersion</td>
<td>4.16</td>
<td>4.09</td>
</tr>
<tr>
<td></td>
<td>Original Sample Set (6 participants)</td>
<td>Extended Sample Set (11 participants)</td>
</tr>
<tr>
<td>------------------------------------</td>
<td>-------------------------------------</td>
<td>--------------------------------------</td>
</tr>
<tr>
<td>Enjoyment of festival</td>
<td>4.16</td>
<td>3.91</td>
</tr>
<tr>
<td>Felt like visited festival in person</td>
<td>3</td>
<td>3.45</td>
</tr>
<tr>
<td>Ability to discover interesting content</td>
<td>3.33</td>
<td>4.18</td>
</tr>
<tr>
<td>Conveys a good festival impression</td>
<td>4.66</td>
<td>4.63</td>
</tr>
<tr>
<td>See more from festival</td>
<td>4</td>
<td>4.18</td>
</tr>
<tr>
<td>Choose performances to watch</td>
<td>3.33</td>
<td>4</td>
</tr>
<tr>
<td>Enjoyment of music concerts</td>
<td>4.66</td>
<td>4.45</td>
</tr>
<tr>
<td>Transcends classic TV coverage</td>
<td>4.66</td>
<td>4.81</td>
</tr>
<tr>
<td>Increased content consumption freedom</td>
<td>4.66</td>
<td>4.54</td>
</tr>
<tr>
<td>More variety of content compared to traditional TV coverage</td>
<td>4.33</td>
<td>4.36</td>
</tr>
<tr>
<td>See more non-concert content compared to traditional TV coverage</td>
<td>4.5</td>
<td>4.72</td>
</tr>
<tr>
<td>Favourability of concert viewpoint personalization</td>
<td>3.83</td>
<td>3.91</td>
</tr>
<tr>
<td>Favourability of inclusion of 360 degree video</td>
<td>3.83</td>
<td>3.91</td>
</tr>
<tr>
<td>Panoramic map helps to convey festival impression</td>
<td>4.33</td>
<td>4.27</td>
</tr>
<tr>
<td>Favourability of watching content captured at locations other than stages</td>
<td>3.66</td>
<td>4</td>
</tr>
<tr>
<td>UGC improves experience</td>
<td>2.67</td>
<td>3.45</td>
</tr>
<tr>
<td>Necessity of UGC filtering scheme</td>
<td>3.33</td>
<td>3.18</td>
</tr>
<tr>
<td>Favourability of Social Media integration</td>
<td>3.00</td>
<td>3.27</td>
</tr>
<tr>
<td>Favourability of surround sound</td>
<td>4.66</td>
<td>4.45</td>
</tr>
</tbody>
</table>

The experience-related feedback solicitation, the results of which have been presented in Table 3, was implemented via a (paper-based) survey. This survey did not undergo any changes when enlarging the participant sample. Stated differently, users of both the initial and the extended test population were requested to fill in an identical questionnaire. For the sake of comprehensiveness, a digital copy of the employed questionnaire (and the script the assessor adhered to in order to acquaint the participant with the study in general and the *WanderCouch* prototype in particular) has been included in Appendix 10.1 of this deliverable.
Finally, the following are two example quotes from the post-usage interviews with the additional test participants:

- P9: “People sometimes capture completely different things [compared to professional coverage]”
- P10: “[the application] gives you the opportunity to switch between concerts in an easy way. When you are at the festival you have to really walk over to the stage, which takes a lot of time”

Summary

Although Table 1, Table 2 and Table 3 reveal that the extension of the test subject sample absolutely did introduce differences in terms of observed user behaviour and self-reported user experience, we consider these differences to be marginal and non-substantial. In effect, the general trend of the results pertaining to both the observational study and the self-reported experience seems to have largely persevered when growing the test population from 6 to 11 representative users. The qualitative feedback collected during the post-usage interviews did not uncover any appreciable disruptions between the original and extended test participant sample either.

5.4 Multi-depth Layered Video Streaming

The pragmatically designed adaptive and Web-compliant Multi-depth Layered Video (MLV) streaming methodology, developed by iMinds, has been subjected to a qualitative evaluation to investigate whether it holds the capacity to improve on traditional, frame-based\(^4\) streaming in terms of user-perceived video quality. The interested reader is kindly referred to deliverables [D5.1] and [D5.4] for a discussion of the methodology’s rationale and technical implementation; here, we will focus exclusively on the outcome of its subjective analysis. Please note that the evaluation of the proposed MLV streaming approach also encompassed an objective component, which will be elaborated on in Section 6.6 of this deliverable.

The concrete goal of the subjective evaluation consisted of comparing the perceptual performance of respectively the Chroma Keying (CK) and Alpha Mask (AM) implementations of the proposed MLV streaming methodology, not only mutually but also with respect to traditional, frame-based video streaming.

5.4.1 Content sample

The evaluation featured three distinct video fragments. Two of these fragments (i.e., clips “captain” and “concert”) were taken from the IRCCyN IVC 1080i video quality database [Péchard, 2008], while the third video (called “NTIA poolhall”) was fetched from the Consumer Digital Video Library (http://www.cdvl.org/). All involved videos have a Full HD spatial resolution, have a playback duration between 8 and 10 seconds, are free of shot transitions (e.g., scene cuts), and were collected in raw YUV422 format. If present, the video fragments’ audio track was dropped.

![Figure 8 Representative frames taken from respectively the captain, concert and poolhall videos.](image)

Content-wise, the three clips are quite similar, in that they all depict scenes featuring a human actor in front of a rather trivial background. In particular, the captain clip shows a man looking through a spyglass in front of an artificial fountain, the concert clip portrays musical performer Jean-Michel Jarre walking around on a stage and talking in his microphone, while the poolhall scene depicts a talking

---

\(^4\) The term “frame-based” in this context refers to the traditional video coding method where the encoder takes integral video frames as input and, via specific Rate-Distortion (RD) optimizations, divides the target encoding bitrate over the constituting frames and then over the constituting pixels of each frame.
person seated in a restaurant. The first two video fragments did not undergo any content-related editing, whereas the last clip was converted into a single shot video fragment by removing the trailing scene of the pool table. The captain and poolhall clips were recorded with a statically positioned and oriented camera, while in the concert fragment the movement of the performer was tracked by rotating a fixed camera. Representative frame excerpts from the content sample are shown in Figure 8. We hypothesized that the three considered videos could match well with the proposed MLV streaming approach, as we expected prospective viewers to focus their attention primarily on the foreground actor and much less on the (somewhat irrelevant) background.

The spatial perceptual information (SI) and temporal perceptual information (TI) figures (as defined by Recommendation ITU-T P.910 [ITU-T, 2008]) of the content sample are as follows: the captain, concert and poolhall videos have a SI value of 33.15, 45.86 and 57.96, respectively, while their TI values respectively equal 20.35, 24.94 and 32.26. Intuitively speaking, these SI and TI measures quantify the intra-frame visual complexity and the amount of inter-frame motion, respectively. Whereas the items in the employed content set are conceptually and content-wise largely analogous, they nonetheless span a rather large area in the SI and TI continua (which range from 0 to infinity). We would have preferred to include videos with more comparable SI and TI values in the evaluation, yet this was found to be irreconcilable with the other requirements that were put forward for the content sample (i.e., the videos had to be available in uncompressed form and in addition had to depict conceptually largely similar content).

5.4.2 Content preparation

For each item in the content sample, 3 discrete bitrate levels were determined (which will be abstractly denoted with the terms low, medium and high) so that each bitrate transition resulted in a clearly discernible quality difference as ascertained by an independent video quality expert. Then, via a manual and offline object segmentation step, the prominent actor in each of the three video clips was converted into a video object, with the remainder of the video scene being classified as belonging to the background. The three established bitrate amounts were finally distributed over respectively the scene background and the segmented foreground object for each of the processed videos (again in close consultation with the video quality expert). As can be read from Table 4, this was always done in such a way that the foreground object was allotted considerably more encoding bitrate compared to the background. The bitrate budget divisions as listed in Table 4 were enforced for both the CK and AM implementations.

<table>
<thead>
<tr>
<th></th>
<th>Captain</th>
<th>Concert</th>
<th>Poolhall</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Background</td>
<td>Foreground</td>
<td>Background</td>
</tr>
<tr>
<td>Low</td>
<td>200</td>
<td>400</td>
<td>200</td>
</tr>
<tr>
<td>Medium</td>
<td>300</td>
<td>700</td>
<td>400</td>
</tr>
<tr>
<td>High</td>
<td>800</td>
<td>1200</td>
<td>800</td>
</tr>
</tbody>
</table>

Certain practical quality adaptation considerations had tangible implications on the content preparation. First of all, it has been empirically established that temporal resolution variation among video fragments contributing to a single scene might lead to the introduction of spatial patches in the recomposed frames for which no colour data is available (this way effectively creating transparent “holes”). It was consequently decided not to apply any temporal resolution scaling to the MLV footage that featured in the evaluation. This concretely implies that, for each of the three videos in the considered content sample, the frame-rate of respectively the scene background and its associated foreground object always equalled the temporal resolution of the unprocessed video. Secondly, although compositing footage with varying spatial resolution might introduce object alignment issues, the consulted video expert nonetheless recommended factoring in spatial resolution manipulation in the evaluation. Therefore, whereas the foreground objects were always streamed in the original (i.e., Full HD) resolution, the scene backgrounds were spatially downgraded to a 1280x720 resolution (except for the concert background in the high bitrate setting, which the video expert preferred to stream in the original resolution, partly because the foreground object in this clip is spatially considerably smaller compared to the two other clips). To intrinsically mask any resulting perceptual errors, the scene background was in
the evaluation always streamed integrally, without the foreground being “cut out” from it. This decision undoubtedly impaired the overall visual quality of the background (i.e., there is a bitrate cost associated with encoding the presence of the foreground object as opposed to a uniformly coloured area), yet this penalty was deemed to be outweighed by the fact that it enabled the exploitation of spatial resolution as an additional degree of streaming freedom in the evaluation.

The content preparation actions described thus far yielded a total of 6 video clips: the unprocessed material (which immediately also served as background in the MLV footage) as well as the segmented foreground of three distinct videos. These were all VBR (Variable BitRate) encoded using the H.264 Main profile according to the target bitrates listed in Table 4. Please note that the bitrate sum of each back- and foreground pair in Table 4 served as target encoding bitrate for the corresponding non-MLV (i.e., frame-based) video. The resulting videos were next temporally divided into 2 second long MPEG-DASH Media Segments. Then, the MLV material was streamed via the proposed methodology and processed at client side using the CK and AM MLV streaming implementations. The resulting output was screen-grabbed (by periodically reading the raw pixel data of recomposited frames from the WebGL frame buffer) and subsequently losslessly encoded (Quantization Parameter value of 0) using the H.264 High profile. Although the non-MLV video footage could in theory have been evaluated directly, it nonetheless underwent the same processing (by treating it as an MLV scene consisting of only a single video object), just to exclude the impact of any perceptual effect introduced by the applied screen grabbing technique from the quality comparison results.

In summary, the content preparation finally amounted to 27 video clips (3 video fragments streamed in 3 different bitrates and in either a traditional or MLV fashion, with two implementations being available for the latter streaming technique). Figure 9 displays a corresponding frame from the captain sequence when streamed using the three investigated techniques in the medium bitrate setting. Although complicated by the limited size of the images in print, one should be able to discern that the detail and general quality of the background in the middle and bottom frame is lower compared to that in the traditional streaming snapshot. In contrast, the traditional approach is seen to perceptually suffer from a less sharp representation of the foreground actor, most notably his face. Colour contamination issues, caused by the lossy (de)compression of the video footage (a problem that has been described in depth in deliverable [D5.1]), are noticeable in the middle image in Figure 9, especially around the person’s hat in the right-hand side of the frame; such issues do not appear in the MLV streaming implementation that applies the Alpha Mask optimization (see bottom image in Figure 9 and deliverable [D5.4] for more information).
5.4.3 Subjective Assessment Methodology

The subjective evaluation of the proposed MLV streaming approach was implemented as a user study in which the Pair Comparison (PC) method as defined in Recommendation ITU-T P.910 [ITU-T, 2008] was applied. In the PC method, users are requested to mutually compare two video stimuli presented as a pair. The primary motivation for resorting to this evaluation technique is its high discriminatory power. As the user-perceptible visual differences that exist between the three tested streaming approaches were anticipated to be potentially small, Pair Comparison was deemed an excellent assessment method to quantify these differences. As a subordinate motivation, the PC mechanism involves a simple cognitive task that does not require specific knowledge or expertise from the assessor.

The subjective video quality assessment study logically encompassed three consecutive phases. In the first phase, some demographic information was collected by requesting that the participant fill in a short survey (primarily) about his or her video consumption habits. Then, the participant was handed written instructions about the test procedure. After verbally addressing any potential questions the participant had about the test procedure, he or she was then asked to complete a training session consisting of 5

---

5 A digital copy of the involved questionnaire can be found in Appendix 10.2 of this deliverable.
6 A digital copy of the written instructions can be found in Appendix 10.3 of this deliverable.
representative test conditions (i.e., 5 video pairs exhibiting quality differences comparable to those appearing in the actual test later on). The objective of this trial run was three-fold:

- familiarize the assessor with the test procedure in general and the employed rating scale in particular,
- counter potential learning effects, and
- stabilize the observer’s opinion with respect to the range of quality differences that items in a pair might exhibit.

The video content that featured in the practice session was not re-used in the actual test and the outcome of the session was discarded.

Once the training session was completed, the researcher left the room and the second phase of the user study commenced, which involved the participant conducting the actual experiment. Here, the participant was asked to assess a total of 57 pairs. For each of the three videos and predefined bitrate levels, the three investigated streaming techniques were paired in both the possible orders, this way giving rise to 54 pairs. Three control conditions, each consisting of two (arbitrarily chosen) identical video fragments, were randomly interspersed among these 54 meaningful pairs. The constituting items of a pair were always presented sequentially on the screen, with a fixed 2 second time interval being enforced between the two (during which the screen turned grey). After the presentation of the second item in each pair, the assessor was asked to grade the statement “I prefer the second video over the first” on a 5-point Likert scale ranging from “strongly disagree” to “strongly agree”, with the middlemost value representing a neutral opinion. After the assessment of the 19th as well as the 38th pair, the participant was given the opportunity to relax (for as long as he or she saw fit) before continuing with the experiment. The pair presentation order was counterbalanced across test subjects (generalized Latin square experimental design) to minimize the impact of fatigue and other confounding factors on the aggregated set of observations.

While the participant was performing the actual PC test, the researcher observed his or her progress in real-time via an IP camera feed. As soon as the final pair had been rated, the researcher rejoined the participant in the study room to implement the third and final phase in the study consisting of a structured post-experiment interview’. During the interview, the three tested streaming techniques were explained to the participant and then discussed. In its totality, a test session on average lasted approximately 1 hour, distributed nearly linearly over the three consecutive phases.

**Apparatus and Setup**

The user study was carried out in a dedicated room in the Expertise centre for Digital Media (EDM) research institute in Diepenbeek, Belgium. The apparatus of the study consisted of a desktop PC running Windows 8.1, a 22” Full HD Samsung SyncMaster S22B300 monitor, and the Tobii EyeX optical gaze tracker ([http://www.tobii.com/xperience/](http://www.tobii.com/xperience/)). The Pair Comparison test was implemented as a Web application. The desktop PC ran an Apache HTTP server to locally host this Web application as well as the 27 prepared video stimuli (see Section 5.4.2). The test application was executed in a commodity Web browser (i.e., Google Chrome version 48.0.2564.116 m) that was set to full screen mode. The brightness and contrast settings of the employed monitor were calibrated using the Windows “Display Color Calibration” tool.

**Participants**

A total of 18 users (4 female) participated in the subjective evaluation. All but two participants were between 20 and 30 years old, with the two outliers being older. All participants were either colleagues or university students. The subjects were screened for (corrected-to-)normal visual acuity and for absence of colour vision deficiencies. The former test was implemented with a Snellen eye chart, the latter by resorting to Ishihara coloured plates. On average, participants indicated to watch about 11 hours of video per week, with a PC being the most frequently used video consumption device, followed at considerable distance by smartphones and TV sets; tablets were found not to be a popular video consumption context in our study population.

---

[A digital copy of the script to which the researcher adhered to structure the post-experiment interview with individual test participants can be found in Appendix 10.4 of this deliverable.](#)
5.4.4 Subjective results

A statistical analysis of the subjective results was conducted to ascertain whether the factors content (i.e., video fragment), bitrate and presentation order (i.e., when comparing streaming techniques A and B, did the participant first see them in either the AB or BA order) had a significant effect on the variable under investigation, in case participants’ preference with regard to the three compared streaming techniques. In the remainder of this subsection, we will denote the involved streaming schemes using the terms traditional (TR), Chroma Keying (CK) and Alpha Mask (AM).

Please recall that the evaluation adhered to a within-subjects experimental design (every participant assessed each of the 57 pairs) in which the presentation order of the factor combinations was randomized. After removing the results pertaining to the three control conditions (see Section 5.4.3) from the data set, the remainder of the recorded preference ratings was divided into three disjoint collections depending on the two video streaming techniques they applied to. The content of the resulting groups was then halved by dropping the ratings that users expressed when comparing the streaming techniques in inverse presentation order (for any given video content and bitrate). In effect, the data resulting from reversing the presentation order was solely used as a reliability metric in the statistical analysis (see Section 5.4.5). Finally, for the three result groups separately, either the repeated measures ANOVA method or the Kruskal-Wallis non-parametric test was applied depending on whether Bartlett’s test revealed a violation of the homogeneity of variances. If the ANOVA revealed a factor to have a statistically significant effect on the evaluation variable, a pairwise t-Test with Bonferroni corrections was applied as a post hoc method.

Concerning the Alpha Mask versus traditional streaming comparison results, the ANOVA did not show a significant main effect of any of the factors on users’ preference for either of the two involved streaming techniques. For the comparison of the Chroma Keying versus the traditional approach, the ANOVA revealed a significant main effect of the bitrate factor on streaming technique preference (F(2,153)=3.75, p=2.56e⁻²). However, the post hoc analysis did not show any significant differences among the three different bitrate levels. In the remaining result cluster (i.e., Alpha Mask versus Chroma Keying), the ANOVA detected a significant main effect of the content factor on streaming technique preference (F(2,153)=21.78, p=6.15e⁻³). The post hoc analysis did show statistically significant differences between respectively the captain and poolhall videos (p=3.2e⁻³) and the concert and poolhall videos (p=9.9e⁻³). No statistically significant effect of the technique presentation order was discovered for any of the technique comparisons, nor were significant interactions among the three considered factors.

Figure 10 summarizes the results of participants’ streaming technique preferences, accumulated across the three considered bitrate levels. In these plots, again only those ratings are retained that users expressed in the original (instead of in the inverse) streaming technique presentation order. The bar charts cardinaly plot the number of times participants preferred one or the other paired streaming technique, without taking the amplitude of their preference into account. For each streaming technique combination and tested video clip, the number of neutral responses issued by assessors can be calculated by subtracting the number of votes jointly received by the two techniques from the aggregated number of comparisons, being 54. For example, for the Alpha Mask versus Chroma Keying comparison with the captain clip, a total of 28 neutral responses were registered. The numbers that are printed on top of the individual bars express the mean and standard deviation of users’ comparison ratings when preference amplitude is taken into account (using numerical values 1 and 2 to respectively denote “(dis)agree” and “strongly (dis)agree” responses). Although we believe these results to be valuable, the reader is reminded that their statistical relevance was found to be limited.

Figure 10 Absolute preference numbers for each of the streaming technique comparisons and tested video clips.
5.4.5 Reliability of the subjective Results

As a data reliability measure, the PC test included three hidden control conditions (in which identical videos were compared), while assessors also needed to grade all of the streaming technique comparison pairs twice, in alternative presentation orders. Analysis of the control condition data revealed that only once there was a strong preference for one of the compared videos was expressed by a test subject, whereas more than half of the grades (correctly) corresponded with a neutral opinion (i.e., 29 out of the total 54). We numerically encoded users’ preference ratings for the control conditions as follows: 0 for neutral responses, 1 for (dis)agree preferences, and 2 for the strongly (dis)agree options. Using this coding scheme, the average preference value turned out to be 0.48 with a 0.53 standard deviation. Please remark that in an ideal scenario, both values would be 0 (denoting neutrality).

Concerning the results pertaining to the presentation order permuting, the analysis revealed little to no intra-subject variation. In absolute figures, in 204 out of the total 486 cases, the participants expressed perfectly consistent preference ratings when comparing the same two streaming techniques in the two alternative presentation orders. In 121 of the residual 282 cases, participants maintained their preference yet with a different magnitude, whereas the remaining 161 cases yielded inconsistent ratings. By discretely mapping the numerical interval [-2,2] to the 5-point preference scale, the average rating difference for the 486 cases was found to equal -0.16 with a 1.19 standard deviation. Again, the closer these values are to 0, the better.

5.4.6 Discussion

The statistical analysis of the subjective assessment findings did not turn out to be especially favourable to the proposed MLV streaming technique. However, we believe that this result needs to be put into proper perspective by also considering the outcome of the post-experiment interviews and participants’ recorded gaze information (the latter of which will be objectively analysed in Section 6.6.2 later on in this deliverable).

As part of the post-experiment interview phase, we informed about the criteria participants applied when comparing the investigated video streaming techniques. It turned out that, even though we explicitly asked assessors to express their preference per paired streaming techniques (without explicitly mentioning the term video quality in this context), participants nearly unanimously interpreted their role in the experiment to primarily be that of a video quality assessor. In particular, many participants indicated that they actively searched for visual artefacts in the compared videos and, if present, were typically inclined to favour the video showing the least amount of artefacts. Recognized types of artefacts included general video blockiness or pixelation, the presence of contours surrounding segmented objects, and (overly noticeable) quality differences between background and foreground. It is apparent that the application of such evaluation criteria hurt the appreciation of the tested MLV sequences. Some participants explicitly mentioned that their appreciation of the MLV approach would likely be different if they would simply be watching videos at home at their leisure.

Participants’ active scanning behaviour (in search of visual artefacts), is objectively confirmed by the gaze tracking data. In particular, whereas some users initially tended to focus on the foreground objects, their attention was found to diverge to increasingly also include the scene backgrounds as the experiment progressed. This behaviour is comprehensively illustrated in Figure 11 and of course unarguably defies the tested MLV streaming premise (i.e., shift bitrate allocation from background to foreground, as this is the part of the video the viewer will most likely concentrate on). Besides artefact scanning, the observed gaze evolution can partly also be attributed to the large number of repetitions of the same 3 video clips (albeit in different bitrates or using different streaming techniques).
Figure 11 Examples of visual gaze analysis results that illustrate a user’s artefact scanning behaviour roughly along the boundaries of the foreground object. For the sake of comprehensibility, the boundary of the foreground object has been outlined in the topmost image.

Many test users also indicated that, during the test, they had often tended to prefer the traditional streaming approach because it yielded the familiar scenario in which visual quality is rather uniformly distributed among back- and foreground. However, when we explained the rationale behind the investigated MLV approach, all 18 participants unanimously were found to be receptive to the idea. This is evidenced by the fact that nearly every test subject was able to independently devise at least one use case in which the proposed methodology could prosper. Some notable examples of imagined use cases were the streaming of a fashion show (with the models wearing the showcased clothes acting as foreground objects), video conferencing, task-centric or educational video applications (e.g., surgery training), and discourse scenarios (e.g., a human news presenter).

Finally, some participants remarked that the MLV streaming approach could have benefited from the inclusion of audio in the experiment. Especially for the concert video in the content corpus, they thought
that the presence of an accompanying audio track would have caused their attention to intrinsically be pulled more towards the foreground object.

5.4.7 Broader implications

The subjective evaluation has revealed two broader video research implications. First of all, the contour artefact turned out to be either very annoying or very distracting (or both) for nearly all of the 18 test participants. The Chroma Keying implementation considerably suffered from the presence of such artefacts, which we consider to be a determining factor for its slightly lower subjective appreciation compared to the alternative Alpha Mask implementation (although the difference was not found to be statistically relevant). We argue that the negative impact of the contour artefact on the viewing experience is likely to be extrapolatable to video streaming applications in general. Secondly, the quality differences, as applied between respectively the back- and foreground in the evaluated MLV content sample, were also classified by many test participants as being an undesirable video artefact. Quantifying the acceptable amount of quality variation among back- and foreground therefore represents an essential avenue for follow-up research.

5.4.8 Summary and conclusions

Statistically speaking, no decisive preference difference has been found to exist between respectively the proposed MLV methodology and classical video streaming for the tested content sample. However, we believe that the reported subjective results correspond with the worst-case scenario (from the proposed methodology's perspective), given the high amount of content repetition in the evaluation (causing assessors' attention to diverge to the background of the tested scenes), the lack of sound output, and the finding that participants overreacted to the presence of “artefacts” in the compared videos (although they were really asked to intuitively express their preference). This observation, combined with structurally elicited test participant feedback, causes us to conclude that the proposed methodology nonetheless holds the potential to outperform traditional streaming with respect to perceptual appreciation, at least in specialized video use cases involving salient foreground objects and plain backgrounds.

5.5 BBC Taster Trials – Proms in Binaural

BBC Taster (http://www.bbc.co.uk/taster/) is the BBC’s Connected Studio’s project that exposes innovative ideas to the general public with on-line prototypes. The BBC R&D’s audio team has used BBC Taster to try out some binaural productions, to let the public listen to them and gather feedback from them. The BBC audio team were involved in the annual Proms concerts from the Royal Albert Hall in September 2016. To achieve a binaural sound an additional microphone array was installed in the venue, and this was used to generate an immersive mix, which was rendered binaurally. The BBC Taster site allows gathering of statistics and a short questionnaire was produced to gather information about the listeners' experience. Figure 12 shows the statistics for visitors to the site, showing over 7000 page visits amongst other information.
Figure 12: BBC Taster Proms - visitor stats

The results of the questionnaire are shown in Figure 13 and Figure 14.
Figure 13: BBC Taster Proms - binaural listener views

Figure 14: BBC Taster Proms - more binaural listener views
The overall view is that the majority of listeners found the binaural experience enjoyable and would listen to it again. The demographics of the group was largely male and older (55+), so for any future trials it would be useful to choose material that would attract a different demographic.

5.6 General Feedback from Trials

iMinds WP3&4 multi-panoramic video production system was used during year 2 at the Dranouter festival, and year 3 at the Leffingeleuren festival, among several other non-integrated field trials, such as the concert of the Belgian band dEUS in December 2015 in the bourla theatre in Antwerp, and the gay pride concerts in Amsterdam, in front of the royal palace on the Dam, on July 24, 2016. Some of this content has been presented to directors to assess usability and likability of the concept. Post-production trials have been organised on Thursday November 28, 2015 and Monday February 22, 2016 with Leonid Adamopoulos at partner VRT and Monday August 8, 2016 with Francesco Indaco at United4All in Hilversum (Netherlands).

The sessions started by a brief introduction of the concept of capturing events using high resolution wide angle static cameras during the event, and deferring not only cutting but also framing to the post production stage. The keyboard and mouse interface was explained, literally in no more than 10 minutes, after which the directors started exploring and experimenting. Both of them had not been previously exposed to the concept or our system before – it was totally new for them. In all cases, they managed to deliver final edits of 3 to 5 minutes in 2 hours or less.

Taking into account it was their first time using the system, the time spent in post production was estimated to be about 50% more than in conventional editing systems. This slight increase in post production cost outweighs possible enormous savings in the cost of capturing: deferring framing from production to post-production highly reduces production cost at a moderate increase in post-production cost. The latter is only spent on material needing post-production, and not on everything that gets captured.

In all three sessions, likability and usability have been assessed to be very high: the system worked stable and was very responsive and easy to understand and get working with, but most importantly, the additional creative freedom of framing in post-production was highly appreciated and recognized as a major innovation. In conventional post production, framing can only be corrected in minor ways, and an editor feels “slave” to the cameramen. The feeling of editing with the multi-panoramic system is much more pleasant, as creative control of how an event is brought into picture now largely becomes a post-production issue, with the editor being in power.

On the down-side, the editor still has to live with limitations and mistakes in how static cameras are set up during capturing. In part, this is because the capturing itself for the trials was done by iMinds staff, who has no education in this area. In part, our system is a digital zoom system and thus limited by resolution: only about a factor 3 zooming is feasible for HD production. This is enough for most shots, especially wide and medium, but regularly not enough for close-ups. For close-ups, a mixed approach, where most cameras are static and fixed, and some are operated on pan-tilt units and equipped with a zoom motor, would do, while still allowing single operator capturing in far most cases.

In all three sessions, numerous, mostly small, enhancements to the keyboard and mouse interface have been suggested, many of which have been implemented in the mean while, and others are still planned.

One other recurring suggestion was to develop a custom interface console for the purpose, especially with a joy-stick for virtual pan-tilt-zoom. We have no plans to develop a custom interface console, but rather opted to make it easy, for ourselves or for third parties, to develop such consoles using our low level API, and the RESTful web API that has been developed on top of that in the startup company AZiiPix.
The final edits resulting from these tests can be downloaded from
https://www.dropbox.com/sh/8971feb73wisloi/AADzCNyomm5cGdAOVREImW4za?dl=0
https://www.dropbox.com/sh/w50sz0ii84nngzd/AAC7u5Ni7rJEv9DqTFjKCRX1a?dl=0
https://www.dropbox.com/sh/wqw5zlnmuzcnd4h/AABf0G2ojOciyhF7dj2NQrf-a?dl=0

Left: Leonid Adamopoulos at VRT; right: Francesco Indaco at United4All (Hilversum Netherlands), in action.
6 Technical evaluations

6.1 Professional Audio and Video Capture and Live Streaming

The BBC’s Edinburgh Festival site presents a perfect setting to trial outside broadcast innovations because of the range of radio productions made in one venue, over a short timescale, and with a dedicated infrastructure and team of venue and broadcast audio engineers.

Audio sources from the stage in the BBC’s Blue Tent and play-ins from the outside broadcast truck were shared with the FOH (front of house) mixing desk and with the BBC R&D audio team via MADI. This interfaced seamlessly with our IP Studio system, as the IP Studio audio capture workstation contained a MADI capture card. We relied on the software, which is part of the MADI driver to monitor the incoming MADI levels, but this did not allow any analysis on the reliability of the MADI signal. So further work might improve the monitoring of incoming MADI audio signals.

The video was captured by the Primer project team, who were also using IP Studio. They had four fixed 4k cameras situated around the seating area pointing towards the stage. Each of these cameras was captured by IP Studio machines, and from one of those video feeds, they generated a downscaled (to standard HD) H.264 coded video stream, which they streamed via RTP to our IP Studio video capture machine. This proved to be a reliable solution, and as their high-performance workstation handled the downscaling and H.264 encoding, it meant our lower-performance workstation could easily record the video frames into the sequence store.

6.2 Immersive Audio Production

We received speech microphone, audience microphone, automated effects (e.g. contestant buzzer), and pre-mixed play-in signals with a view to creating a 3D object-based mix with simultaneous automated stereo mix. Each source was flagged in our Editor software depending on which category it belonged to (e.g. “speech”, “audience” or “play-in”) and streamed as an individual audio object to the receiving Player app. This allowed the viewer to adjust the relative levels of the main elements of the programme while maintaining our internal balance of the individual objects in each group. The user could also choose to listen to a stereo or binaural render.

Ideally for an immersive audio production we should use un-mixed audio material in order to have full control over the balance of multiple sources in three dimensions. However there were several factors that compromised this approach: The Editor’s live mixing facilities are limited, with only mouse and keyboard control (this could be improved in future with the use of MIDI-controlled hardware or an appropriate IP audio mixer); there were few/short sound-checks (rehearsals) for each production so opportunities to optimise incoming levels and produce satisfactory balances were limited; we had inadequate knowledge of each production/script and no producer present to advise. So to overcome the need for traditional live mixing, the sources were sent post-fader from the outside broadcast truck so that any live mixing/adjustments to each source were completed by the broadcast engineer allowing us to concentrate on 3D panning and stereo compatibility.

There are particular aesthetic judgements required when mixing 3D audio for 2D video, especially with regard to positioning sources relative to their visual position in the picture. This is easier with a static shot, as in this trial, but with a multi-camera production, audio source levels and positions often change with each shot. Synchronisation between audio and video was satisfactory throughout the trial.

Our control room was compromised by noisy plant and IP Studio computers in the same room so noise-cancelling headphones were essential to achieve satisfactory monitoring but, in future, a quieter listening environment is preferable so that loudspeakers can be used for the stereo mix.

6.3 Content Synchronisation

For synchronising user generated content streams with the backend system (and thus professional content), we have implemented the synchronisation protocol between the capture app and the backend system described in D4.5.
We have performed two experiments to evaluate the synchronisation performance: one in a controlled setting to evaluate the synchronisation protocol, and one as part of the Edinburgh field trial to assess synchronisation with a production system under real production conditions.

### 6.3.1 Evaluation of synchronisation protocol

In order to evaluate the precision of the synchronisation method, we use the following setup. The machine running the synchronisation service displays a frame showing the system time, and the time is also rendered as VITC [VITC] (scaling it over multiple lines to simplify extraction). Figure 15 shows an example of a generated frame. The mobile phone records the screen using the capture app, which also performs synchronisation and sends the time stamps together with the video to the backend system. The backend system records the list of time stamps received and the video frames. The VITC from the video frames is then extracted using the VITC detector from VidiCert\(^8\). As some frames of the mobile video may capture the time range when the time code switches, we perform median filtering of the measured times to eliminate such frames. We then compare the list of time stamps and calculate the statistics of the offsets.

We perform the test under three network conditions: using a dedicated Wifi network (Wifi1), using a shared office Wifi network (Wifi2) and using a 3G network connection (3G). The 3G connection has been made via a VPN client in order to access a local machine displaying the time code. The offsets between the time stamps associated with the frames and the visual content of the frames have been measured. The tests have been made using a Samsung Galaxy S6 running Android 6.0.1.

Figure 15: Example frame generated for time measurement from the image.

Figure 16 shows the measured results from a series of recordings. In the dedicated Wifi, the synchronisation error typically stays within 30ms, i.e., is about one frame. There are some outliers reaching an error of up to 350ms. In a shared Wifi, the typical error increases slightly to about 50ms. However, the number of outliers and the corresponding error increases. Interestingly, the results for 3G (tested in a city area) are quite similar to those of the shared Wifi network. However, there are also some extreme outliers (outside the ±1s range shown in the plot), which is not the case for the Wifi networks.

---

\(^8\) http://www.vidicert.com/en
As outliers occur in most cases only for short times, and the order of frames is usually preserved, adding an overall buffering delay in the range of seconds and correct the time stamps based on assuming temporal consistency in the sequence could help to reduce the problem, and get most errors within the range of about 2-3 frames. Of course, the situation may be different if the network being used is under high load.

6.3.2 Evaluation at Edinburgh Field Trials

The Edinburgh Festival field trials involved two parts for synchronisation; the live UGC captured video and the professional audio and video. The professional audio and video capture, based in the BBC Blue Tent venue, was done using the IP Studio system. This is governed by a PTP master clock, which ensures all the machines (and the packets of data they handle) on the network are synchronised with each other and accurately time-stamped. However, the PTP grandmaster used for the field trial was not GPS-locked (getting a GPS signal was a major extra expense), so was free-running.

With the UGC capture, once the live video was streamed from the mobile phone to the cloud service it was then time-stamped there. As PTP cannot be used over WANs, we used a commercial tool called TimeKeeper⁹ that combines multiple NTP sources to obtain a GPS synchronised time source in the

⁹ http://www.fsmlabs.com/timekeeper/
cloud. As the IP Studio PTP grandmaster clock was not GPS locked, so that there is no absolute time reference available, or any common clock between the two systems. We can thus perform the following measurements:

- Measuring the time offset between the local clock in Edinburgh (from visual capture) and the time stamp assigned on the capture machine in the cloud.
- Measure the offset between the timestamps assigned locally in Edinburgh and the timestamps assigned after capturing the signal again in Salford. This includes the offset between the local clock in Edinburgh and the GPS-locked grandmaster in Salford, as well as the signal runtime. Measure the offset between the timestamps assigned in Salford to the timestamps assigned on the machine in the cloud. This includes the signal runtime, as well as the possible imprecision of the clock in the cloud.

![Figure 17: Frame from capturing the PTP display at the production machine in Edinburgh.](image)

From the time displayed in Edinburgh, and the timestamp assigned by the cloud, an offset of -0.86s was determined.

The UGC capture IP Studio in Salford, which received an SDI stream from the cloud player, and pushed these video frames into the sequence store for DASH playout. Each DASH video segment was given a timestamp in the store. The segments were 5 seconds long, and the timestamp for the first grain in each segment was available.

As above, the smartphone used for the UGC capture was pointed at the screen of the professional audio capture IP Studio in Edinburgh, which was displaying the PTP clock used on that VLAN. By comparing the time shown on this screen-grab and the timestamp on the UGC capture machine in Salford, the offset between the two could be determined.

The screenshot in Figure 18 shows the time for a particular frame. The table below shows how the times of the UGC capture machine in Salford compare with the Edinburgh audio capture machine and the capture machine in the cloud. If the system timestamp is taken as the reference, then there is less than 4 seconds difference. Comparing with the PTP clock it is about 35 seconds difference.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>UGC capture (Salford)</strong></td>
<td><strong>DASH segment number</strong></td>
</tr>
<tr>
<td></td>
<td>294186755</td>
</tr>
<tr>
<td><strong>UGC capture (Salford)</strong></td>
<td><strong>DASH segment timestamp (first grain)</strong></td>
</tr>
<tr>
<td></td>
<td>16:45:50.00</td>
</tr>
<tr>
<td><strong>Edinburgh system time (screen-grab)</strong></td>
<td><strong>16:45:46.03</strong></td>
</tr>
<tr>
<td><strong>Edinburgh PTP time (screen-grab)</strong></td>
<td><strong>16:46:25.33</strong></td>
</tr>
</tbody>
</table>

10 This time is shown as HH:MM:SS.ss to match the DASH timestamp format. The screen-grab shows the time in the HH:MM:SS:FF where FF is the frame number, the frame number is converted to seconds here.
The differences are small enough to allow different scenes to be switched between without too many problems, though they would be too large for capturing the same scene and maintaining a plausible flow. If both the Edinburgh and cloud-based clocks were GPS-locked (or at least locked with each other), then it is likely these delays would be significantly reduced.

### 6.4 Content Selection

#### 6.4.1 Approach

As described in D4.2, we predict the likelihood of an incoming Moment to be accepted by the operator of the Trademark content selection tool. The likelihood is used to rank the Moments, which is then used to present to the operator using Trademark the incoming content in a way that puts clips likely to be accepted at the top of the list. The following metadata are considered as inputs:

- Media type (image, video)
- User ID
- Device (manufacturer, model)
- Location (coordinates, place identifier), determine from one or more of: GPS, user input, visual matching
- Overall quality score
- Visual uniqueness (determined from matching scores with top $k$ other content items within a time window)
- Tag uniqueness (determined from tag TF/IDF of a content item within a time window)

The impact of some of these properties on the relevance of the content, are event independent (e.g., device properties, quality), while others are very specific (e.g., location) and may even change during the event. The event-independent group of properties can be used to build an initial model for the prediction of the likelihood of selection of the content, and thus the position in the ranking. Such a model can be trained offline on data from previous events. However, it must be possible to quickly adapt the
model to a new event or the current situation during an event. We thus use online random forests, as they can constantly update the model based on selection made by the operator. Every time the operator accepts or rejects an item in Trademark, the model is updated. Thus the model will also adapt to changing selection criteria of the operator during the event.

### 6.4.2 Results

We have trained a model on the data collected at the field trial at the Open Days of the Committee of the Regions (COR) in October 2015. We have split the features into event independent ones (e.g., content quality) and event specific ones (e.g., location). When using 70% of the COR data for training and 30% for testing, we obtain a true classification rate (TCR) of 0.8537 when using all features and of 0.8415 when using only the event independent ones.

The next step is to transfer the model to another event. At the Austrian Research Night (LNF) in April 2016, we have collected a smaller set of content items. We use COR for training and predict on the LNF data. In addition to the TCR, we also measure the average precision (AP). The TCR implies that we would use the random forest to actually make a binary classification, while the AP measures how good the resulting ranking is. As we have an operator in the loop, we do not make the decision automatically, but we can best support the operator by putting the items to be likely acceptable at the top of the list. The AP measures this by assessing how many of the items down to each rank are relevant.

![Figure 19: Prediction results for the LNF event, using a model trained on the COR data.](image)

We repeated the same experiment at the Leffingeleuren festival in September 2016. This event span over three days (in comparison to the single evening of LNF), and thus a much higher amount of videos has been received. Figure 20 shows the results. Again, it takes some time until the model initially converges. As the content and the selection preferences change more over the three days event, there are intermediate variations of the performance until the model adapts again. The stable model achieves a TCR of about 0.75 and an AP of 0.86.

As the performance at the top of the ranked list is quite good, we analysed how far one could go with automating content selection (which should be activated only after the initial adaptation of the model).
As we have an operator in the loop, we mostly care about false positives (i.e., Moments that the automatic system accepted, but the operator would have rejected). False negatives are not an issue, because they can remain in the list to be processed, and can still be accepted by the operator if needed. Figure 21 plots the expected false positive rate at a cut-off score. The score is the selection likelihood returned by the random forest, with 1.0 meaning accepted and 0.0 not accepted. We see that we can set the score to about 0.8 with for staying within a 1% false positive rate. If the workload is high, Moments with a score above this value could be automatically accepted, while the other remain in the list to be processed by the operator.

![Graph](image-url)

**Figure 20:** Prediction results for the Leffingeleuren event, using a model trained on the COR data.
6.5 Visual Matching

We report evaluation results on the compact video descriptor presented in D4.2. While the proposed descriptor could be implemented using different local descriptors and aggregation methods, we base the compact image sequence descriptor on the MPEG CDVS descriptor, making use of the global and local parts of the descriptor. A CDVS descriptor contains a set of local SIFT descriptors [Lowe, 2004] sampled around ALP interest points [CDVS, 2015], which are quantised to a ternary representation. In addition, it contains an aggregated global descriptor, represented as Scalable Compressed Fisher Vector (SCFV) [Lin, 2014] as a binary vector. As we are only interested in the similarity of entire key frames, we only match the aggregated global descriptors.

MPEG has collected a data set for an activity called Compact Descriptors for Visual Analysis (CDVA) for evaluating technologies for this purpose [CDVA CfP]. We use this data set for our experiments. The dataset contains in total around 23,000 video clips with durations ranging from about one minute to more than an hour. The material contains broadcast and user generated content in different resolutions and frame rates, and with diverse contents. It is divided into a set of reference and query clips, which contain different views of one object or scene, embedded into noise clips. In addition, a part of query clips have been modified with transformations (e.g., resolution and frame rate changes, overlays, screen capture). The rest of the set contains distractor material for retrieval experiments.

We perform pairwise matching of the 9,715 queries against the 5,128 reference clips, and report the true positive rate at 1% false positive rate and the temporal localisation performance measured as Jaccard index. Further details on the data and the evaluation metrics can be found in [CDVA Eval]. In addition to matching performance metrics, we measure the reduction in descriptor sizes due to the proposed compression.

As a first step, shot boundary detection using matching of colour histograms is performed, and subsequent frames with high similarity discarded. This creates an irregularly sampled set of key frames for each shot. We then extract CDVS descriptors for each of the remaining frames. We compare an

Figure 21: Fraction of false positives if automatic selection is applied at a specific cut-off score, measured on the Leffingeleuren data.
uncompressed version of the descriptor (i.e., a set of single frame descriptors) with compressed versions that apply lossless compression to the global descriptors and lossy compression to the local descriptors.

We have evaluated the algorithm with three configurations, differing in the resulting descriptor sizes. Figure 24 shows the obtained mean descriptor sizes for a segment descriptor. For the compressed descriptor, lossy compression is applied with different thresholds for discarding similar descriptors: $\theta_l = \{0, 3, 5, 7, 10, 15, 30, 40, 50\}$. Both the sizes after differential encoding of the global and local descriptors and after additionally applying adaptive binary arithmetic coding are reported. For the local descriptor, the lossy compression provides the main size reduction, while binary arithmetic coding only contributes a small additional benefit. It is also interesting that the smaller value for $\theta_l$ already allows gaining most of the size reduction, resulting in 48% of the original descriptor size, while setting $\theta_l$ to 5 only gains one additional percent. This shows that there is a significant number of very similar descriptors, but the distance to further descriptors is then typically larger, so that increasing the threshold by a small margin does not have much impact. Only for higher values of $\theta_l$ (40+) there is again an additional size gain.

The reported numbers are per segment, and translate to about 1.7-3.0kB per second of video for the uncompressed descriptor, and about 600B-1kB for the compressed descriptor (at $\theta_l=30$).

Figure 22 and Figure 23 show the resulting performance metrics. The lossless compression of global descriptors does of course not have any impact on the performance, but the results show that at the chosen moderate compression rates, both the true positive rate and the temporal localisation performance remain unchanged. Actual results for individual videos do vary in terms of matching scores, with both changes in true and false positives between the descriptors with different compression. However, these results are balanced over the large data set, resulting in nearly constant performance. Only if the threshold for local compression is set to values of 30 and up, there is a clear impact on both the matching and temporal localisation performance.

![Figure 22: True positive matching rate at fixed false positive rate for different configurations and lossy compression rates (for different values of $\theta_l$) of local descriptors.](image)
Figure 23: Jaccard index of temporal localisation for different configurations and lossy compression rates (for different values of $\theta_l$) of local descriptors.

Figure 24: Resulting global and local descriptor sizes for different configurations and lossy compression rates (for different values of $\theta_l$) of local descriptors.
6.6 Multi-depth Layered Video Streaming

6.6.1 Objective video quality assessment

The perceptual fidelity of the 27 content configurations that were prepared to conduct the subjective evaluation of iMinds’ proposed MLV streaming methodology (see Section 5.4.2) was also assessed objectively. This was implemented by comparing the quality of the prepared video stimuli against that of their respective source signals (i.e., the unprocessed video on which the object segmentation was applied) using the Peak Signal-to-Noise Ratio (PSNR), Structural Similarity (SSIM) and Visual Information Fidelity (in the pixel domain, VIFp) metrics [Sheikh, 2006]. The results are listed in Table 5. The PSNR, SSIM and VIFp value spaces respectively equal \([0, +\infty]\), \([-1,1]\) and \([0,1]\), with higher values denoting better video quality. Please note that the reported objective figures were generated using a specialized Web implementation that overcomes the inter-video sync inaccuracies witnessed in contemporary Web browsers (please see deliverable [D5.4] for more details) by operating on decomposed frame sequences instead of video input. Contrary to video input, image-based input does allow tightly synchronized playback of the constituting entities of an MLV scene in the Web browser. Such frame-accurate entity playback sync is a prerequisite to enable objective frame-by-frame comparison with the source signal.

| Table 5 Objective video quality assessment results for the 27 stimuli featured in the subjective evaluation of the MLV streaming methodology; traditional (TR) versus Chroma Keying (CK) versus Alpha Mask (AM) implementation. |
|---------------------------------|-----------------|-----------------|-----------------|-----------------|
|                                | Captain        | Concert         | Poolhall        |
|                                | TR             | CK              | AM              | TR             | CK              | AM              |
| PSNR                           | 31.22399       | 29.76373        | 29.65177        | 30.82720       | 27.22029        | 27.14064        |
| SSIM                           | 0.856273       | 0.855087        | 0.851589        | 0.810569       | 0.780561        | 0.777704        |
| VIFp                           | 0.191511       | 0.204516        | 0.184914        | 0.118853       | 0.083744        | 0.075360        |
|                                |                 |                 |                 |                |                 |                 |
| PSNR                           | 32.98046       | 31.22409        | 31.20216        | 33.43121       | 31.08041        | 31.05073        |
| SSIM                           | 0.873295       | 0.865927        | 0.864808        | 0.837656       | 0.818014        | 0.816874        |
| VIFp                           | 0.242913       | 0.243927        | 0.234792        | 0.184861       | 0.138733        | 0.133284        |
|                                |                 |                 |                 |                |                 |                 |
| PSNR                           | 35.11651       | 33.81326        | 33.84330        | 34.83295       | 32.24666        | 32.26169        |
| SSIM                           | 0.896118       | 0.885625        | 0.885589        | 0.851363       | 0.829584        | 0.829355        |
| VIFp                           | 0.323718       | 0.308757        | 0.306523        | 0.228273       | 0.174132        | 0.171264        |

The following high-level findings can be derived from Table 5:

- The objective quality scores of all 3 streaming methods monotonically increase as the bitrate rises (as was to be expected).
- The two MLV streaming implementations always achieve slightly lower objective scores compared to the traditional approach (e.g., a 2 point PSNR difference). We argue that this finding can be attributed to the fact that the beneficial effect of enhanced foreground quality on the objective grade is outweighed by the detrimental contribution of the (severely) degraded background quality.
- The objective grades of the 2 MLV streaming implementations are near identical. This outcome applies across all tested content and bitrates settings.

In all, Table 5 reveals that the employed objective metrics found the tested video streaming techniques to be roughly comparable in terms of produced visual quality for the considered content sample. Therefore, it was decided to discontinue the objective evaluation track for the MLV streaming research in favour of subjective evaluation techniques (see Section 5.4).
6.6.2 Gaze tracking analysis

Please recall from Section 5.4.3 that the apparatus for the subjective evaluation of the proposed MLV streaming approach included an optical gaze tracker that recorded viewers’ region of focus during the experiment (see Figure 25). The resulting gaze metadata has been analysed to estimate the amount of time test participants spent on watching respectively the back- and foreground content of the evaluated video stimuli. To this end, each recorded gaze point was classified as either a back- or foreground hit by mapping its screen coordinates to the coordinate space of the segmented foreground object of the involved video fragment. If the resulting video coordinates indexed a pixel carrying the chroma value denoting transparency (i.e., a pixel with a pure green colour), the user was (likely) watching the background of the video scene at the time his or her gaze was sampled; if not, the user was (likely) focussing on the foreground object.

The motivation for conducting the back- versus foreground classification of gaze records is the following: as the two investigated MLV streaming implementations favour the foreground object over the background in terms of assigned bandwidth budget, the subjective video quality as perceived by test users will likely have differed depending on the amount of time they spent on watching the back- versus foreground. The conducted gaze analysis allowed us to quantify the temporal distribution of a test participant’s region of focus over respectively the back- and foreground during the experiment.

When analysing the sampled gaze metadata, a distinction was made between respectively the first and the second item in each assessed pair. Remember that each pair depicted the same content streamed at an identical bitrate yet using a different technology (i.e., TR versus CK versus AM). As such, it became possible to investigate potential differences in gaze behaviour between the constituting items of each pair. This was deemed important, since certain test participants had claimed (as part of their post-usage interview, see Section 5.4.6) the tendency to shift their focus from fore- to background when transitioning from the first to the second video stimulus in a pair. In particular, some participants stated that, since the first video stimulus in the pair had already familiarized them with the foreground material of the scene, they shifted their attention to increasingly include the background as the playback of the second item in the pair took place.

An example outcome of the objective gaze analysis is shown in Figure 26. Here, the percentage of foreground gaze hits per constituting item of each of the 57 pairs assessed by a particular test participant is plotted as respectively a box and a line chart.
Figure 26 reveals that, on an individual participant basis, there sometimes definitely existed some variability with respect to the amount of time spent on watching the foreground object. This participant-specific variability was exhibited both on an intra- and inter-pair basis. However, when averaged out over the 18 participants who were involved in the subjective evaluation of the MLV streaming methodology, a nearly identical foreground gaze hit percentage was witnessed for the constituting items of video stimuli pairs. In particular, the average foreground gaze hit percentage for the first and second paired stimuli equalled 64.92% and 63.23%, respectively.

The objective findings from the conducted gaze analysis have been related to the qualitative evaluation results presented in Section 5.4.4. The following is a non-exhaustive list of (high-level) correlations that could be identified from this mapping effort:
Participant P2’s foreground gaze hit ratio was relatively low (around 45% on average).
Throughout the entire experiment, this particular user literally never expressed a preferential score for the tested MLV streaming techniques when comparing them against traditional streaming.

Participant P3’s gaze metadata and subjective preference scores show exactly the opposite correlation compared to P2’s results. More formally, this user tended to focus extensively on the foreground content throughout the experiment (80% foreground gaze hits on average); in turn, P3 issued a lot of neutral and preferential scores for the MLV streaming techniques when comparing them against traditional streaming.

Participant P4’s gaze metadata revealed a steady decline in foreground gaze hit percentage as the experiment progressed. In particular, whereas this user seems to have invested considerable amounts of attention to the foreground objects prior to the first experiment intermission (see Section 5.4.3), his or her focus was then observed to shift to increasingly include the scene background. This focus shift appears to be aligned with P4’s subjective quality scores. More formally, P4 by far issued the highest amount of preferential scores for the CK and/or AM techniques while he or she was assessing the first 19 pairs of the experiment.

Each of the just described correlations between observed gaze behaviour and subjective quality scoring seems to corroborate our hypothesis that people need to chiefly focus on the foreground content in the course of the streaming session for the proposed MLV streaming technology to hold any benefits in terms of perceived video quality. However, our corpus of gaze observations also holds data points which violate this hypothesis or for which the expected gaze/preference correlations are not obvious.

For example:

- Participant P9’s gaze data uncovered an average 7% intra-pair difference in foreground hit-ratio (with the first item in pairs exhibiting more foreground hits). However, this observed behaviour did not seem to have had a salient impact on the preference ratings issued by the user. In particular, no clear preference for MLV streaming techniques that appeared as the first item in a pair was found in this user’s subjective scores.

- Similar to the gaze behaviour of P4, participant P10’s focus appears to have shifted from foreground to background as the experiment progressed. However, contrary to the P4 case, this focus shift apparently did not influence P10’s streaming technique preference ratings.

A more low-level (e.g., statistical) analysis of potential gaze/preference correlations is yet to be implemented. However, given the high-level observations described above, it is likely that the participant sample size will prove to be insufficient to be able to uncover statistically meaningful differences with respect to the relationship between gaze behaviour and reported streaming technique preference.

Please note that some reservations regarding the accuracy of the collected gaze metadata seems advocated. The Tobii EyeX hardware was calibrated per individual test user and in addition is advertised to be tolerant to lateral head movements. Unfortunately, the hardware is also known to operate less robustly in situations where the distance between the user’s head and the tracker changes. Although test participants were explicitly requested to keep their distance to the screen (and hence also to the tracker) as constant as possible throughout the experiment, numerous situations were observed (via the IP camera installed in the room where the experiment took place) where a test participant temporarily leaned either forward or backward while assessing video stimuli. Chances are that gaze samples, taken while the user’s head was located at a distance to the screen other than the one for which the tracker was calibrated, will be of a questionable accuracy.
7 Conclusions

The evaluations were categorised into two areas: user evaluations to assess how users of the technology respond to it, and technical evaluations where the technology is assessed for its performance under realistic scenarios. The field trials provided the test bed for many of these evaluations as they provided both the access to many users, realistic content and actual physical limitations of the venues.

The user evaluations covered subjective testing of multi-layered video coding, and the public’s views on immersive audio using binaural playback as well as the Wall of Moments system. The results of the subjective tests for the multi-layered video coding showed it was no worse than traditional video coding, and given that the test material could be considered to be the worst-case scenario for the MLV approach, it could potential show improvements over traditional coding. For the binaural audio research using the BBC Proms playout, there was an overwhelming view than binaural reproduction was an improvement over traditional stereo and it provides a favourable experience. This result helps justify the use of binaural playout for immersive audio in the project.
References

[D2.4.1] Evaluations v1
[D5.1] First version of content preparation and distribution
[D5.4] Playout clients and content preparation & distribution


# Glossary

## Partner Acronyms

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BBC</td>
<td>British Broadcasting Corporation, UK</td>
</tr>
<tr>
<td>BIT</td>
<td>Bitmovin GmbH, AT</td>
</tr>
<tr>
<td>DTO</td>
<td>Technicolor, DE</td>
</tr>
<tr>
<td>iMinds</td>
<td>iMinds Vzw, BE</td>
</tr>
<tr>
<td>JRS</td>
<td>JOANNEUM RESEARCH Forschungsgesellschaft mbH, AT</td>
</tr>
<tr>
<td>TaW</td>
<td>Tools at Work Hard+Soft Vertriebsgmbh, AT</td>
</tr>
<tr>
<td>VRT</td>
<td>De Vlaamse Radio en Televisieomroeporganisatie NV, BE</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AM</td>
<td>Alpha Mask</td>
</tr>
<tr>
<td>CK</td>
<td>Chroma Keying</td>
</tr>
<tr>
<td>MLV</td>
<td>Multi-depth Layered Video</td>
</tr>
<tr>
<td>PC</td>
<td>Pair Comparison</td>
</tr>
<tr>
<td>PSNR</td>
<td>Peak Signal-to-Noise Ratio</td>
</tr>
<tr>
<td>SSIM</td>
<td>Structural Similarity</td>
</tr>
<tr>
<td>TR</td>
<td>traditional</td>
</tr>
<tr>
<td>VIFp</td>
<td>Visual Information Fidelity</td>
</tr>
</tbody>
</table>

Acknowledgement: The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 610370.
10 Appendix

10.1 WanderCouch Interview Script

Interview Script for ICoSOLE Smart TV Application for Music Festivals – IMMERSION STUDY

1. Introduction
a) Welcome the participants and thank them for their time.
b) Explain the context of the study for the ICoSOLE project
   - ICoSOLE aims at developing a platform that enables users to experience live events which are spatially spread out, such as festivals, parades, marathons or bike races, in an immersive way by combining high-quality spatial video and audio and user generated content
   - The objective of this session is you to try out the ICoSOLE Smart TV application for music festivals. We offer you content from the Dranouter 2015 music festival. First I will explain you the different features that are available in the application and how to navigate through it using the remote control. Then, I will give you one hour so you can enjoy the concerts that are available using the application. Finally, I will ask you a set of questions about your experience.
c) Ask the participants to read and sign the consent form
   - Ask if we can video record the session
   - Remind the participants that all data will be kept private and will be used only for the purpose of the experiment
   - Remind the participants that they can pause the session at any moment if they need a break
   - Remind the participants that they can/have to be honest with respect to their answers
2. Guide the participants through the Application

Explain Main Screen
Volume control and mute
Back buttons returns to the previous page
Home buttons on any screen takes you to this screen
Help Key show some hints

LIVE view section:
playing stage - Enter to full screen
List of stages – Enter to activate list, up/down arrows to navigate, Enter select an object and play it

UGC Section:
Enter to access the list and left/right arrows to navigate
Enter on an element opens UGC screen

Tweets are read only no interaction is available

FILTERS section:
Enter to Access the list
Show filters button open a list of categories with filters
Choosing a filter adds a button that can be navigated and press Enter to remove
Use to filtrate the content of the LIVE and UGC sections

MAP Selection
Topographic map is a 2D view of the festival terrain
Panoramic map is a 360 video map

LIVE STREAM SCREEN
Volume control is the same
Colour buttons open and hide menus

UGC Menu
Use arrows to navigate
Enter shows the element in a PiP mode
Enter again opens the element in the UGC screen with the concert as background
Red button to close List or PiP

CAMERA Menu
Use arrows to choose a camera (point of view)
The pointer in the mini map shows the location and orientation of the camera
Enter to choose a point of view

MAP Selection
Use arrows to choose a map
Enter to confirm selection
SOCIAL MEDIA
This list has no interaction at all

ODV STREAM SCREEN
Volume control is the same
Colour buttons open and hide menus
Use arrows to change the view of the camera (pan and tilt)
Use the channel up and down keys for changing the zoom level

UGC Menu
Use arrows to navigate
Enter shows the element in a PiP mode
Enter again opens the element in the UGC screen with the concert as background
Red button to close List or PiP

CAMERA Menu
Use arrows to choose a camera (point of view)
The pointer in the mini map shows the location and orientation of the camera
Enter to choose a point of view

MAP Selection
Use arrows to choose a map
Enter to confirm selection

SOCIAL MEDIA
This list has no interaction at all

UGC SCREEN
Volume control is the same

PREVIEW section
Shows the UGC element that you chose and displays some useful data on the right side

RATINGS
Use horizontal arrows to set a rating

FILTERS
Same as in main screen

VIDEOS and PICTURES lists
Use horizontal arrows to navigate
Enter to select

TOPOGRAPHIC MAP
Use horizontal arrows to navigate
Enter to select
Only stages can be chosen
ODV MAP
Use arrows to change the view of the camera
Channel up and down to change the zoom level
Press numbers to navigate to the other location
Yellow key hides/show mini map
Numbers on stages open the concert
Experience Questionnaire

To what extent do you agree with the following statements regarding your experience?

<table>
<thead>
<tr>
<th>Statement</th>
<th>Strongly disagree</th>
<th>Neutral</th>
<th>Strongly Agree</th>
<th>Don't know</th>
</tr>
</thead>
<tbody>
<tr>
<td>I feel like I visited the festival in person.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>With respect to real life experience</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I enjoyed the concert(s) I watched</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I had to choose what to watch from the alternative concurrent performances (I was interested in more than one artist)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I could see more things from the festival</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I could discover interesting content easier</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>With respect to a traditional TV broadcasting</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I had more freedom to choose what to watch</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I could see more things from the festival</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>There was more variety of content</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>With respect to the concert(s) you watched</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The tent was crowded</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The crowd was actively involved with the performer</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I had enough freedom to choose a point of view when watching a concert</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Answer the following questions.

1. Can you describe the atmosphere of festival as you were able to perceive it via the application?

2. What size do you think this festival had?
   a. Comparable to Rock Werchter
   b. Comparable to Pukkelpop
   c. Comparable to Genk on stage
   d. Comparable to Gentse feesten
   e. Smaller than any of the festivals above

3. Name 3 to 5 things that you remember from the festival (beyond the concerts)
4. Did you see a girl in a fairy costume dancing with children? (In the footage from other people at the festival)

5. Did you see a baby crawling? (In the footage from other people at the festival)

6. What do you think about the combination of professional content and user generated content?

7. Did the “Live” nature of the application give you the feeling that you could miss out things? If it did how do you feel about this characteristic?

8. Would you like the application to include “on-demand” functionality for the professional content? E.g. for catching-up at the end of the day.

To what extent do you agree with the following statements regarding your experience?

<table>
<thead>
<tr>
<th>Statement</th>
<th>Strongly disagree</th>
<th>Neutral</th>
<th>Strongly agree</th>
<th>Don’t know</th>
</tr>
</thead>
<tbody>
<tr>
<td>The application…</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>let me enjoy the festival</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>helped me to discover interesting content of the festival</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>immersed me in the festival</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>offered me a good general impression of the festival</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>It is important to have multiples points of view (vantage points) when watching a concert</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The panoramic map helped me to get an impression of the festival terrain</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The panoramic map improved my experience</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The content made by other visitors of the festival improved my experience</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>It is valuable to be able to watch content captured at locations other than the stages (e.g. the camping in the topographic map)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The experience offered by the application transcends the kind of experiences that are attainable with classic TV broadcasts</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Being able to watch a musical performance in 360 degrees is a valuable feature</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The filtering mechanism is necessary to help me find the content I’m looking for</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The social media integration (Twitter feeds in this case) is a valuable feature

Including a surround sound setup would improve my experience

On a scale from 1 to 5, 1 being the lowest score, how much do you like the bands that are featured in the Smart TV application?

1. Triggerfinger
2. Intergalactic Lovers
3. Bart Peeters

Do you have any further comments?
10.2 Multi-depth Layered Video Streaming User Study Survey of Demographics

Demographics

Participant ID: ____________

1. Gender: Male ☐ Female ☐

2. Age range:
   15 – 24 ☐ 25 – 34 ☐ 35 – 44 ☐ 45 – 54 ☐ 55+ ☐

3. Regarding your weekly video consumption, how many hours do you on average watch ...

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1 – 5</th>
<th>6 – 10</th>
<th>11 – 15</th>
<th>16 – 20</th>
<th>21 +</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short online videos</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(e.g., music videos</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>on YouTube, Facebook</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>clips, etc.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full movies on</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Internet (e.g.,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>YouTube, Netflix,</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>etc.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full movies on TV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>channels</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full movies from a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DVD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4. Regarding your weekly video consumption, how many hours do you use on average each one of the following types of devices ...

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1 – 5</th>
<th>6 – 10</th>
<th>11 – 15</th>
<th>16 – 20</th>
<th>21 +</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smartphone</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tablet</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Personal computer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Television</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5. What type of programs/content do you usually watch? Mark as many as apply.

☐ News reports
☐ Sports (Which ones: ________________________________________________________________)
☐ Movies (Which types: _____________________________________________________________)
☐ TV Series (Which types: __________________________________________________________)
☐ Music videos
☐ Other _________________________________________________________________________

6. To what extent would you rate yourself as being a video quality expert?

☐ No expert at all
☐ Rather not an expert
☐ Neutral
☐ Somewhat an expert
☐ An expert
Hello and thank you for participating in our user study!

In this experiment you will watch short video sequences (i.e., with a duration of approximately 10 seconds each) on the screen that is in front of you. The content set of the experiment will consist of three distinct video sequences. Multiple versions of each of these video sequences have been created using different streaming techniques. The video clips will be presented in pairs, meaning that two different versions of the same video clip will always be shown consecutively. A two second intermission interval will be applied to temporally space out the constituting items in a pair.

You will compare the second video clip against the first video clip in each pair and then will be asked to express your preference.

You should carefully watch each of the presented video clips in a pair. Then, at the end of the playback of the second item in the pair, please record your opinion using the following scale:

I prefer the second video over the first video:
Strongly Disagree – Disagree – Neutral – Agree – Strongly Agree

Please note that if you prefer the first video over the second video in the pair, you should select a rating situated in the left-hand side of the scale. Conversely, the right-hand side of the scale applies if you prefer the second video over the first video.

It is very important to try to maintain a consistent interpretation of the rating scale throughout the entire study (i.e., while grading the different pairs featured in the experiment).

No time constraint applies to the pair rating task. To enter your rating, simply click the correct radio button and then click the “Submit” button.

Please try to express your preference with respect to the presented videos as honestly and truthfully as possible. Remember, there are no good or bad answers, we are only interested in your (subjective) opinion!

Once your rating for the current video pair has been recorded, there will be a short pause before the presentation of the next pair of video clips commences. In total, you will be asked to grade 57 pairs of videos. At two predefined points in time during the experiment, you will be given the opportunity to take a (short) break.

IMPORTANT NOTE: Due to potential computational overload of the PC hosting the experiment, it might accidentally happen that the playback of videos stalls or stutters for a short period of time. You must NOT take such stalls or stutters into account when assessing the corresponding video!

During the experiment, your gaze will be (optically) tracked. The tracker will be calibrated for your eyes and sitting position before the start of the experiment. We therefore kindly ask you to maintain your initial position as much as possible throughout the experiment, as this will ensure maximal reliability of the gaze tracking records. The tracker is robust to eye blinking, so please feel free to blink your eyes during the experiment.

We will begin with a short practice session to familiarize you with the test procedure. This practice session is purely intended to get yourself acquainted with the video quality assessment method; no information will be collected from this practice session. Once you feel comfortable with the video assessment method, the researcher will leave the room and the actual experiment will begin. The experiment will consume about 30 minutes of your time.

Once the experiment has ended, the researcher will re-join you to conduct a short post-comparison interview to collect subjective feedback about the experiment.
10.4 Multi-depth Layered Video Streaming Interview Script

Post Experiment Interview.  

Participant ID: __________

1. Did you notice any difference in the different videos you watched? Yes / No  
   a. Can you explain what difference you noticed?

2. Explain the two different techniques that were compared.  
   a. Do you think these streaming techniques have potential benefits? If so, elaborate your opinion.

3. Detailed subjective evaluation of the techniques.  
   a. Low bitrate evaluation  
      i. Show unprocessed video as baseline - Show Chroma key version - Show alpha mask version  
      ii. What do you think about the streaming techniques for this bitrate?
   b. Medium bitrate evaluation  
      i. Show unprocessed video - Show Chroma key version - Show alpha mask version  
      ii. What do you think about the streaming techniques for this bitrate?
   c. High bitrate evaluation  
      i. Show unprocessed video - Show Chroma key version - Show alpha mask version  
      ii. What do you think about the streaming techniques for this bitrate?

4. The tested streaming optimizations have a number of (potentially detrimental) visual implications. How annoyed were you with the following introduced artifacts? [Textual description + Please rate this on a scale from 1 to 5, 1 being totally not annoyed and 5 being very annoyed]  
   a. Non-perfect synchronization between background and foreground  
   b. Greenish outline surrounding the foreground objects  
   c. Difference in quality between background and foreground

5. What were the criteria you applied when expressing your preference while comparing the video pairs?

6. What was your principal region of focus during the experiment?  
   Foreground / Background / Combination of both

7. Perceived applicability of the techniques in “everyday life” video streaming.  
   a. In which scenarios do you think these techniques can be useful (type of scenes in the videos)?  
   b. What limitations do you think the techniques have?
   c. Which video genres seem most appropriate to you to apply the techniques?

8. How tedious or boring did you think the experiment was?
10.5 Leffingeleuren Questionnaire

The original version of the questionnaire is Dutch, an English translation is attached below.

Feedbackvragen

1. Wat is jouw geslacht?
   a. Man
   b. Vrouw

2. Wat is jouw leeftijd?
   a. 10-15
   b. 15-25
   c. 25-35
   d. 35-45
   e. 45-55
   f. 55-65
   g. 65 - 75

3. In welke provincie woon je?
   a. Limburg
   b. Oost-Vlaanderen
   c. West-Vlaanderen
   d. Vlaams-Brabant
   e. Antwerpen
   f. Brussel Wat is jouw leeftijd?

4. Wat doe je van beroep?
   a. Arbeider
   b. Bediende
   c. Kaderlid/management
   d. Zelfstandige/vrij beroep
   e. Loopbaanonderbreking/tijdskrediet
   f. Gepensioneerd
   g. Student
   h. Huisman/-vrouw
   i. Invalide
   j. Werkzoekend
   k. Ander

5. Wat was jouw eerste indruk van de Moments-app?

6. Hoeveel filmpjes heb je gemaakt?
   a. Geen
   b. 1-5
   c. 6-20
   d. 21-50
   e. Meer dan 50
   f. Geen idee

7. Hoeveel zou je de app scoren op 10?

8. Met wat voor smartphone heb je de app gebruikt?

9. Wat vind je leuk aan de Wall of Moments?

10. Wat zou je veranderen aan de app? (meerdere antwoorden zijn mogelijk)
    a. De lengte van de filmpjes langer maken.
    b. De filmpjes kunnen aanpassen voor je ze uploadt
    c. Toevoegen van filters of effecten via de Moments-app
    d. Andere filmpjes ook zien via jouw app
e. De snelheid van de applicatie  

f. De snelheid van het opladen van de filmpjes  
g. Geluid toevoegen aan de Wall of Moments  
h. Grootte van het geheugen  
i. Wifi-verbinding  
j. Ook toegankelijk maken voor andere types smartphones  
k. Ander:  

11. Waarom heb je voornamelijk filmpjes gemaakt?  
   a. Ik wou de sfeer van het festival en mijn ervaringen delen met anderen  
   b. Ik wou mijn filmpjes graag terugzien op de schermen  
   c. Het optreden was fantastisch en ik wou dit vastleggen  
   d. Ik wou de applicatie testen  
   e. Ik wou een herinnering aan Leffingeleuren  
   f. Ander:  

12. Waarom heb je sommige filmpjes niet geüpload?  
   a. Ik had geen verbinding met het internet  
   b. Ik heb de applicatie niet kunnen opstarten  
   c. De applicatie liep vast  
   d. Ik heb de filmpjes niet kunnen doorsturen  
   e. De applicatie sloot automatisch af  
   f. Ik heb niet kunnen filmen via de applicatie  
   g. Ik kon de applicatie niet starten  
   h. De applicatie was te moeilijk in gebruik  
   i. Ik heb er niet aan gedacht  
   j. Ik wou vooral genieten van de optredens en het maken van filmpjes verstoorde dit  
   k. Iets anders, namelijk:  

13. Heeft de Wall of Moments jouw beleving van het festival versterkt?  
   a. Ja  
   b. Nee  

14. Heeft de redactietool jouw ervaring als testreporter versterkt?  
   a. Ja  
   b. Nee  

15. Ben je tijdens het festival langs de VRT booth geweest?  
   a. Ja  
   b. Nee  

16. Heb je de Wall of Moments via de thuislink bekeken?  
   a. Ja  
   b. Nee  

17. Wat was voor jou de belangrijkste reden om via de thuislink te stream te volgen?  
   a. Ik was nieuwsgierig  
   b. Ik wou vooral de optredens bekijken  
   c. Ik wou de filmpjes van de testreporters zien  
   d. Ik wou de sfeer van het festival ervaren  
   e. Ik ken een van de testreporters  
   f. Ik wilde mijn Moments herbeleven  
   g. Ander:  

18. Heeft het achteraf bekijken van de filmpjes jouw festivalbeleving versterkt?  
   a. Ja  
   b. Nee  

19. Welke tips heb je voor de Wall of Moments te verbeteren?  

20. Was de uitleg en communicatie voldoende voor het gebruiken van de app?  
   a. Ja
21. Heb je de dome bezocht waar live concerten van De Zwerver gestreamd worden?
   a. Ja
   b. Nee

22. Welke score op 10 zou je geven aan de ervaring van de dome?
23. Wat vond je het leukst aan de dome?
24. Wat vond je niet zo leuk aan de dome?
25. Welke dagen ben je geweest naar Leffingeleuren?
   a. Vrijdag
   b. Zaterdag
   c. Zondag
   d. Alle

The English translated version.

Leffingeleuren questionnaire

1. What is your gender?
   a. Man
   b. Woman

2. What is your age?
   a. 10-15
   b. 15-25
   c. 25-35
   d. 35-45
   e. 45-55
   f. 55-65
   g. 65-75

3. In which province do you live?
   a. Limburg
   b. Oost-Vlaanderen
   c. West-Vlaanderen
   d. Vlaams-Brabant
   e. Antwerpen
   f. Brussel?

4. Which profession do you occupy?
   a. Labourer
   b. Clerk
   c. Officer/management
   d. Independent/vrij beroep
   e. Professional leave
   f. Retired
   g. Student
   h. Houseman/-wife
   i. Disabled
   j. Looking for work
   k. Other

5. What was your first impression on the Moments app?
6. How many clips did you make?
   a. None
   b. 1-5
   c. 6-20
   d. 21-50
   e. More than 50
   f. No idea

7. Out of 10, how many point would you give to the app?

8. Which smartphone did you use the app on?

9. What do you like about Wall of Moments?

10. What would you improve in the app? (several options possible)
    a. Increasing the length of the clips.
    b. Being able to adapt the clips before sending.
    c. Adding filters and effects through the Moments app
    d. Watching other clips through your app
    e. The speed of the application
    f. The speed of uploading clips
    g. Adding sounds to the Wall of Moments
    h. The size of the memory
    i. The Wi-fi connection
    j. Making the app available on more types of smartphones
    k. Other:

11. Why did you create clips?
    a. I wanted to share the atmosphere of the festival and my experiences with others
    b. I wanted to see my clips on the large event screens
    c. The performance was fantastic and I wanted to capture this
    d. I wanted to test the application
    e. I wanted a memory on Leffingeleuren
    f. Other:

12. Why did you refrain from uploading some clips?
    a. I did not have an internet connection
    b. I was unable to start the app
    c. The app crashed
    d. I was unable to send the clips
    e. The application closed automatically
    f. I was unable to record through the app
    g. The application was too difficult
    h. I forgot about it
    i. I wanted to enjoy the performances, and recording clips disturbs this
    j. Other:

13. Did the Wall of Moments improve your experience of the festival?
    a. Yes
    b. No

14. Did the editing tool increase your experience as test reporter?
    a. Yes
    b. No

15. Did you visit the VRT booth during the festival?
a. Yes  
b. No  
16. Did you watch the Wall of Moments using the "use at home" link?  
a. Yes  
b. No  
17. What was your main motivation for following the stream via "use at home"?  
a. I was curious  
b. I mainly wanted to see the  
c. I wanted to see the clips of the test reporters  
d. I wanted to experience the atmosphere of the festival  
e. I know a test  
f. I wanted to re-experience my Moments  
g. Other:  
18. Did watching your clips after the festival increase your festival experience  
a. Yes  
b. No  
19. Which tips do you have to improve the Wall of Moments concept?  
20. Was the explanation and communication on the usage of the app satisfactory.  
a. Yes  
b. No  
21. Did you visit the dome where gigs in De Zwerver were live streamed?  
a. Yes  
b. No  
22. Out of 10, which score would you give to the experience in the dome?  
23. What did you like the most about the dome?  
24. What did you like not so much about the?  
25. During which days have you been at Leffingealeuren?  
a. Friday  
b. Saturday  
c. Sunday  
d. All days