MPEG DASH - Some QoE-based insights into the tradeoff between audio and video for live music concert streaming under congested network conditions

Rafael Rodrigues*§, Peter Pocta†, Hugh Melvin‡, Manuela Pereira§ and Antonio M. G. Pinheiro*§
*IVT Lab, Optics Center, Universidade da Beira Interior, Covilhã, Portugal
Email: rafael.rodrigues@ubi.pt
†Department of Telecommunications and Multimedia, Faculty of Electrical Engineering, University of Zilina, Slovakia
Email: pocta@fel.uniza.sk
‡Discipline of Information Technology, National University of Ireland, Galway
Email: hugh.melvin@nuigalway.ie
§Instituto de Telecomunicações, Universidade da Beira Interior, Covilhã, Portugal
Email: mpereira@di.ubi.pt, pinheiro@ubi.pt

***Abstract*—The rapid adoption of MPEG-DASH is testament to its core design principles that enable the client to make the informed decision relating to media encoding representations, based on network conditions, device type and preferences. Typically, the focus has mostly been on the different video quality representations rather than audio. However, for device types with small screens, the relative bandwidth budget difference allocated to the two streams may not be that large. This is especially the case if high quality audio is used, and in this scenario, we argue that increased focus should be given to the bit rate representations for audio. Arising from this, we have designed and implemented a subjective experiment to evaluate and analyse the possible effect of using different audio quality levels. In particular, we investigate the possibility of providing reduced audio quality so as to free up bandwidth for video under certain conditions. Thus, the experiment was implemented for live music concert scenarios transmitted over mobile networks, and we suggest that the results will be of significant interest to DASH content creators when considering bandwidth tradeoff between audio and video.**

***Keywords*—MPEG DASH; Audio-visual quality; Mean Opinion Score; Adaptive media streaming**

I. Introduction

Although a relatively recent development, the evolution and penetration of HTTP Adaptive Streaming (HAS) has been rapid over the last 10 years. This has been driven by the very strong commercial case, as evidenced by proprietary solutions that were initially developed by Apple, Adobe and Microsoft. The common objective across these solutions was to provide a media consumption platform that piggy-backed on existing web infrastructure, and that was client driven. This allowed the client to make informed decisions based on realtime network characteristic estimates, user device type/capabilities, and client preferences, which required a backend server that provides the media for consumption, divided into short chunks of a few seconds and rendered multiple times. The server provides metadata, both at a semantic (eg. genre) level and physical level (media structure/formats/bit rates/video frame rates etc) on its stored media in the form of a Media Presentation Description (MPD) file. The client firstly pulls this file and makes decisions based on this and the other variables, listed above. Such a model fits very well with best-effort Internet infrastructure and maps well to user demands to consume media on a wide variety of devices under differing scenarios. The proliferation of these proprietary solutions to meet user needs, and the resulting interoperability challenges, necessitated work on standardisation, and culminated in the release of MPEG DASH (Dynamic Adaptive Streaming over HTTP) standard in 2011 [1], [2]. With YouTube and Netflix as key adopters, it has received huge support and adoption rates. Consequently, DASH has been the subject of very significant research, that has examined the many variables that make up the full system, and their interaction. A key objective of much of this research is driven by the need to maximise the end user Quality-of-Experience (QoE).

978-1-5090-0354-9/16/$31.00 ©2016 IEEE

II. Background & Related Work

With the huge growth over the last 15 years in multimedia traffic, significant research has been undertaken in both subjective and objective assessment of multimedia quality as perceived by the end-user. However, most studies to date have focused on individual modalities , i.e. audio and video separately. This has resulted in relatively mature and well researched subjective approaches and objective metrics. Regarding objective metrics for audio, these include PEAQ (Perceived Audio Quality) [3] and POLQA (Perceptual Objective Listening Quality Assessment) Music [4] with a comparison provided in [5]. For video, a whole range of metrics have emerged, such as the basic PSNR (Peak Signal Noise Ratio), SSIM (Structural similarity), and PEVQ (Perceptual Evaluation of Video Quality) [6]. However, subjective tests have shown [7] that there is a strong inter-relationship between audio and video, and thus research has more recently focused on developing a combined audio-visual model. In [8], a review of

audio and video metrics is presented as well as an investigation into the key issues in developing joint audio-visual quality metrics. In particular, it outlines the common approach to deriving audio-visual quality (AV_Q) from the audio quality (A_Q) and visual quality (V_Q) as follows:

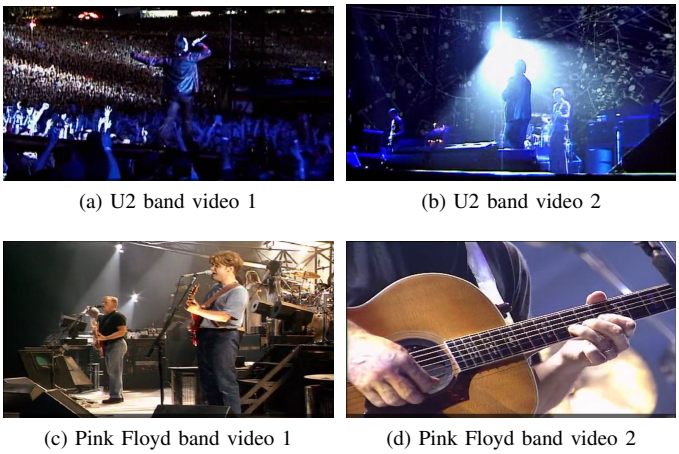
$$AV_Q = a_0 + a_1 A_Q + a_2 V_Q + a_3 A_Q V_Q \quad (1)$$

where parameters (a_0, a_1, a_2) denote the different weights of audio and video and the multiplication factor, with a_0 as a residual term. Undoubtedly, this is a significant challenge with many variables and contextual factors. Our research aims to add to the knowledge base in designing such a joint model.

Many studies up to now, according to [9], have studied different aspects of HAS from a video quality perspective, e.g. the impact of quality switches [10], stalling vs. switches [11] and initial delay vs. stalling and starting bit rate [12]. On the other hand, to the best of our knowledge, only one study up to now has dealt with the impact of audio content on quality experienced by the end user in the context of HAS. In [13], Tavakoli et al. investigated an influence of audio presence on an evaluation of video related impairments. This study shows that audio has a minor impact (Pearson correlation coefficient of 0.93 between Audio and No audio test reported) on video quality perceived by the end user, assessed according to the methodology defined in ITU-T Rec. P.910 [14]. Moreover, when it comes to quality adaptation strategies, a correlation between MOS (Mean Opinion Scores) obtained for a whole sequence and MOS for processed sequences was always lower when an audio part was involved in the test.

To the best of our knowledge, no study exists that explicitly deals with the impact of audio quality, and more generally the trade-off in relative bandwidth utilization on audio-visual quality experienced by the end user in the context of HAS. We believe that such insights may be very useful for TV broadcasters and video content delivery providers, such as Netflix, YouTube, Amazon, and Hulu, that are interested in optimizing their client-side quality adaptation strategies. Such insights can inform decisions about the range of both audio and video content quality rendered, so as to provide the end user with the best quality possible considering the mix of corresponding network conditions, user device capabilities, and user preferences. It is worth noting here that with very few exceptions, the quality adaptation strategies up to now have uniquely focused only on adapting the quality of the video content. In this paper we thus investigate the effect of reducing quality level of audio content on audio-visual quality experienced by the end user in the context of HAS. To do so, we have run a subjective test according to ITU-T Rec. P.911 [15] simulating a concert broadcast over a mobile network. It is worth noting here that in terms of content, we have deployed live music performances as this content fully reflects a typical content of the broadcast service of our interest. Moreover, this scenario represents a good example of the situation whereby the quality of audio plays a crucial role. Insights arising from this study will allow DASH content providers to optimise the use of limited bandwidth in terms of the tradeoff between video and audio.

As further evidence of the extent to which the challenge of evaluating HAS remains very current, the ITU-T are currently working on P.NATS - Parametric non-intrusive assessment of TCP-based multimedia streaming quality, considering adaptive streaming. The aim is to develop a collection of objective parametric quality assessment modules that predict the impact of observed IP network impairments on quality experienced by the end-user in multi-media mobile streaming and fixed network applications using progressive download, including adaptive streaming methods [16].



(a) U2 band video 1 (b) U2 band video 2

(c) Pink Floyd band video 1 (d) Pink Floyd band video 2

Fig. 1: Examples of frames from the videos used in the subjective tests.

The remainder of the paper is organized as follows. Section 3 describes the subjective test carried out within this study and its results. In section 4, subjective test scores are presented and analyzed in detail using One-way and Two-way ANOVA tests. Section 5 provides the final conclusions.

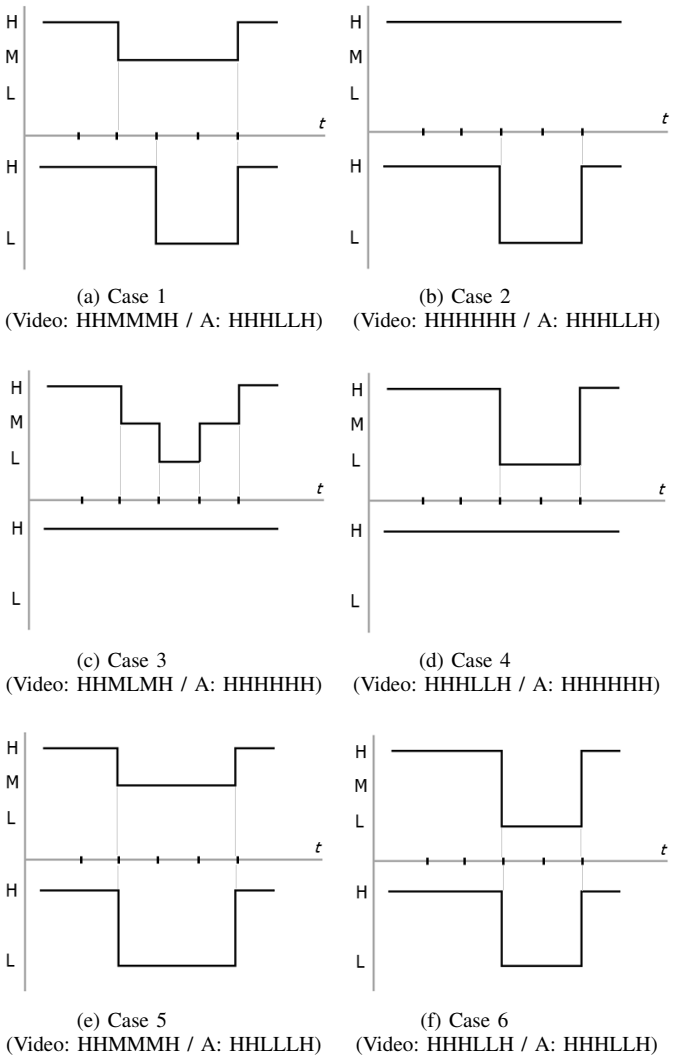
III. Subjective Test Design

A. Source videos and impairment design

Given the focus of this paper, a key requirement was to select content where audio quality might play an important role in quality perception. Moreover, test sequences should include diverse scenarios, both in terms of audio and video content. Thus, we selected scenes from live music performances to use as source videos for the subjective experiment. Two concerts, from two different bands (U2 and Pink Floyd), were ripped from DVD to provide the source content. We believe that the selected videos include all the typical scenes in terms of spatial and temporal information, which may occur in real-life situations in this context.

Source videos were then resized to 480p resolution (854x480), which is the standard definition for mobile streaming [17]. Four 1-minute long scenes were cut from the available content (Fig. 1) and chopped into 10 second chunks, following the results of the study published in [13], with video and audio streams demuxed. FFmpeg software [18] was used to encode video chunks at three different compression rates - 512 (H), 256 (M) and 128 Kbps (L) - using the H.264/AVC video coding standard [19].

FFmpeg was also used to encode the corresponding audio chunks with the High Efficiency Advanced Audio Coding v2 (HE-AAC v2) scheme. HE-AAC v2 extends the AAC range



(a) Case 1 (Video: HHMMMH / A: HHHLLH)

(b) Case 2 (Video: HHHHHH / A: HHHLLH)

(c) Case 3 (Video: HHMLMH / A: HHHHHH)

(d) Case 4 (Video: HHHLLH / A: HHHHHH)

(e) Case 5 (Video: HHMMMH / A: HHLLLH)

(f) Case 6 (Video: HHHLLH / A: HHHLLH)

Fig. 2: Temporal dynamics of impairments test cases. Video profiles are in the upper half (H. 512 Kbps, M. 256 Kbps, L. 128 Kbps) and audio profiles in the lower half (H. 128 Kbps, L. 24 Kbps).

of operation to as low as 24 Kbps [20]. Considering the main objective of this study, audio was encoded at 128 (H) and 24 Kbps (L).

Using the diverse encoded streams, 6 different impairment cases described in Fig. 2 were designed by concatenation of audio and video chunks into 1-min long *mp4* files. Impairment cases were designed to simulate different situations of network congestion, with variable tradeoff between audio and video bitrates. Case 2 includes audio degradation only, while cases 3 and 4 include video degradation only. The remaining cases simulate degradations of both audio and video quality simultaneously, with different combinations of the chosen representations. the total bandwidth required for the transmission of both audio and video stream in the cases 1 and 5 is roughly the same at each quality level as that required by cases 3 and 4. On the other hand, the total bandwidth required for each quality level is even lower, when it comes to case 6. We believe that the impairment profiles chosen for this test

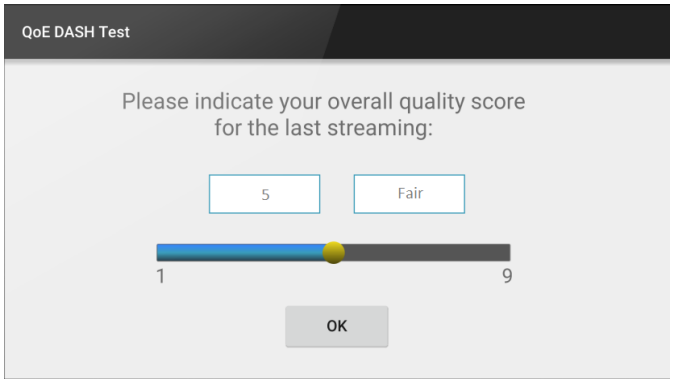


Fig. 3: Rating screen in Android App used for subjective tests.

cover the most important cases occurring in real-life situations.

B. Test methodology

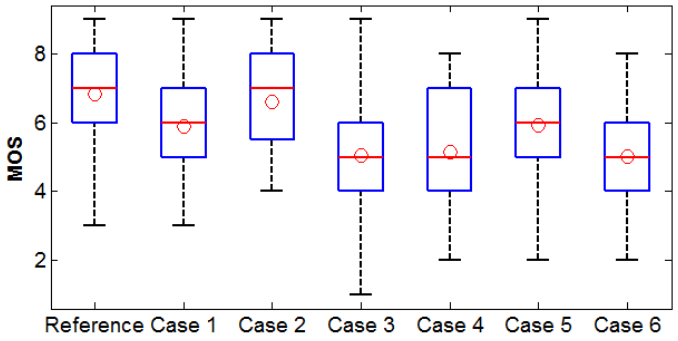
A single-stimulus study (Absolute Quality Rating) with hidden reference was conducted at the Image and Video Technology laboratory of the Optics Center - Universidade da Beira Interior (UBI). The study followed the ITU-T Rec. P.911 methodology for subjective audiovisual quality assessment in multimedia applications [15]. 32 naive subjects, mostly students at UBI, participated in the study, from which 21 were male, with ages ranging between 18 and 35 (mean 24 years), and 11 were female, with ages ranging between 18 and 22 (mean 20 years). The selected subjects represented the target end user group of live music concert streaming services.

Experiment sessions were carried out in a controlled ambiance with subjects using stereo headphones (Philips SL3060). An Android application was developed specifically to run the experiment on LG Nexus 5 smartphones (quad-core, 4.95" screen with resolution of 1920x1080), which provided full screen visualization of the clips. After each presentation, a calibrated rating bar was presented to the participant (Fig. 3) to provide overall audiovisual quality score, considering the nine-level quality scale depicted in Table I.

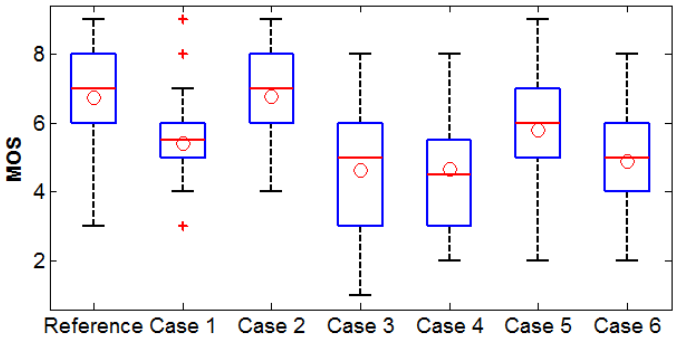
Average session duration was 20 minutes. Considering 4 different scenes and 6 cases per scene, there were 24 different clips involved in the test set. Every subject attended one single session and the test was designed in such a way that each clip was viewed the same number of times, i.e. 16. This approach was used to prevent over-visualization of the presented contents and consequent biasing of the results, given the relatively long duration of the test sequences. Ref-

9 8	Excellent
7 6	Good
5 4	Fair
3 2	Poor
1	Bad

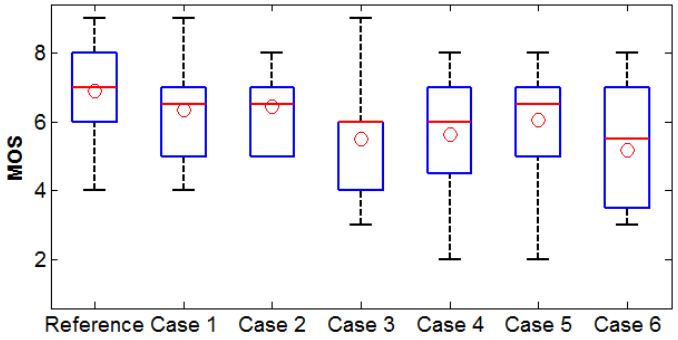
TABLE I: Scale used for subjective quality assessment.



(a) All videos



(b) U2 band videos



(c) Pink Floyd band videos

Fig. 4: Box plot of subjective test data (MOS of each impairment is indicated by a red circle).

erences were included in every session (non-explicit) and the respective quality ratings were collected. Thus, each session consisted of the randomized visualization of 12 impaired test sequences plus the 4 reference sequences, preceded by 2 training presentations. Training clips were chosen from a group of initially designed impaired clips not included in the final testing set. These clips included similar impairment profiles, but using different scenes of the available content, to promote an adaptation to the context and conditions presented in the test.

Scores obtained from the experiment were statistically analysed to observe both audio and video influence on global quality perception. In order to properly analyse if a reduction in the quality of audio content has an impact on audio-visual quality experienced by the end user, MOS from the different impairment conditions will be compared directly with MOS from the respective reference clips. Furthermore, different impairment cases, with audio or video-only impairment, will be compared against simultaneous audio and video impairment cases.

Effect	SS	DF	MS	*F*-ratio	*p*-value
Test condition	211.66	6	35.2760	**16.77**	**<0.0001**
Signal	64.15	3	21.3839	**10.16**	**<0.0001**
TC*Signal	57.13	18	3.1739	**1.51**	**0.0830**
Error	883.63	420	2.1039		
Total	1216.56	447			

TABLE II: Summary of two-way ANOVA test conducted on the MOS values.

IV. Results and Analysis

Subjective test data distribution is presented in the box plots of Fig. 4. Considering all the collected data, a two-way analysis of variance (ANOVA) test was conducted using signal (i.e. audio-visual content used in the subjective test) and test condition (i.e. audio-visual impairments designed for the test) as fixed factors (Table II). The highest F-ratio (F = 16.77, $p < 0.0001$) was achieved for the test condition factor, closely followed by the signal factor (F = 10.16, $p < 0.0001$). Moreover, the effect of both signal and test condition was found to be highly statistically significant. Regarding the interaction of all the involved factors, i.e. signal and test condition, the results show that it was not statistically significant (F = 1.51, p = 0.0830). Thus, these results reveal that subjects were more sensitive to the test conditions than to all the investigated signals, and also that there was statistically insignificant interaction between test condition and signal factor.

Fig. 5 shows box plots of subjective test results for the cases where only audio quality was varied (with video at constant maximum quality) (a), and where only video was varied (with maximum audio quality) (b), alongside with the data from corresponding reference cases. Hence, regarding audio impairments, scores from case 2 only were retrieved, whereas for video impairments, data from cases 3 and 4 were considered. Throughout the results analysis, reference data used for comparison is a selection of the paired reference scores in each test. For example, if a given subject saw impairment case 2 for U2 video 1 and Pink Floyd video 2, scores reported for U2 video 1 and Pink Floyd video 2 references in that same test are collected.

From simply observing the data box plots, it is clear that the video impairments have more influence on the quality experienced by the end user than the audio impairments, when compared to reference conditions. Given that the considered subjective test data is normally distributed, which was confirmed by a Kolmogorov-Smirnov test [21], one-way ANOVA tests were performed to see if there are statistically significant differences between mean values of these two groups. The

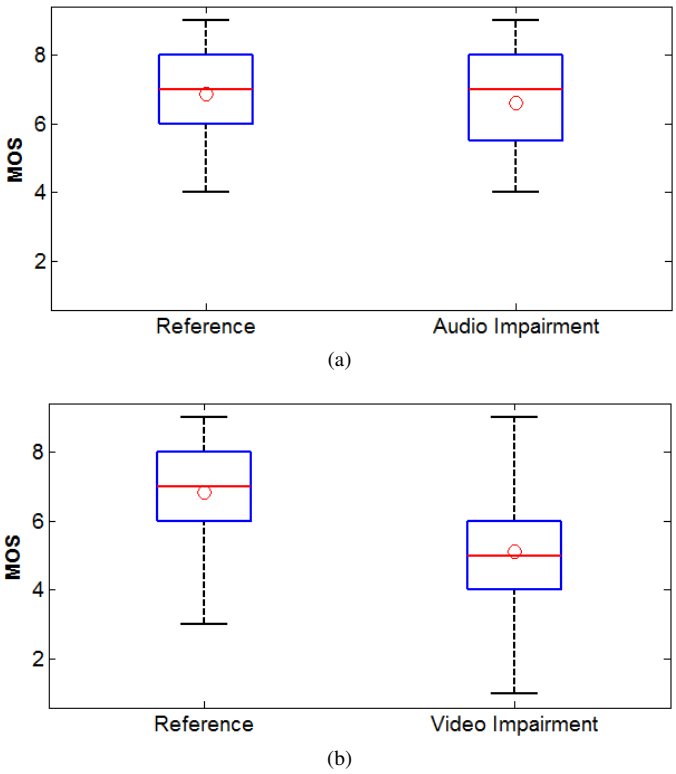


Fig. 5: Box plots of audio (a) and video (b) impairment subjective test data compared to respective references (MOS is indicated by a red circle).

p-value	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
Case 1						
Case 2	0.0039					
Case 3	0.0040	<0.0001				
Case 4	0.0120	<0.0001	**0.7995**			
Case 5	**0.8590**	0.0068	0.0022	0.0072		
Case 6	0.0017	<0.0001	**0.9130**	**0.7111**	0.0011	
Reference	<0.0001	**0.2502**	<0.0001	<0.0001	0.0226	<0.0001

TABLE III: One-way ANOVA p-values for impairment comparison with all videos (results above $p = 0.05$ are highlighted).

p-value	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
Case 1						
Case 2	0.0005					
Case 3	**0.0587**	<0.0001				
Case 4	**0.0673**	<0.0001	**0.9443**			
Case 5	**0.3101**	0.0116	0.0068	0.0079		
Case 6	**0.1461**	<0.0001	**0.5421**	**0.5911**	0.0168	
Reference	0.0003	**0.5916**	<0.0001	<0.0001	**0.1098**	<0.0001

TABLE IV: One-way ANOVA p-values for impairment comparison with U2 videos (results above $p = 0.05$ are highlighted).

p-value	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
Case 1						
Case 2	**0.7722**					
Case 3	0.0237	0.0060				
Case 4	**0.0718**	0.0280	**0.7548**			
Case 5	**0.4421**	**0.2596**	**0.1336**	**0.2763**		
Case 6	0.0037	0.0007	**0.4260**	**0.2980**	0.0284	
Reference	**0.0507**	**0.2617**	<0.0001	0.0004	**0.1071**	<0.0001

TABLE V: One-way ANOVA p-values for impairment comparison with Pink Floyd band videos (results above $p = 0.05$ are highlighted).

p-value obtained from the first one-way ANOVA test was 0.4431, representing audio-only impairments compared with the reference. In the case of video impairments, the obtained p-value was 1.71×10^{-17}. From these values, we can conclude that differences in quality perception of the audio quality variation/impairments are not statistically significant ($p > 0.05$), whereas differences in quality perception of the video variation/impairments are statistically significant ($p < 0.05$).

Further ANOVA tests were then carried out considering all the different impairment cases used in the subjective tests. Kolmogorov-Smirnov test was applied to subjective test data relative to such cases confirming the normality of these subsets.

Table III shows ANOVA results for impairment comparisons using both U2 and Pink Floyd videos. Based on the results of ANOVA tests it is possible to draw a very important conclusion for this study. There are four cases in which data subsets were reported as being statistically similar, excluding the similarity between audio impairment (case 2) and reference scores which was already discussed. Similarity between cases 1 and 5 ($p = 0.8590$) show that audio distortions, even for a longer period, do not affect global quality perception. Moreover, in cases 3 and 4 MOS values are statistically similar to the values reported for case 6, with $p = 0.9130$ and $p = 0.7111$, respectively. These results indicate that the quality perception does not change significantly with lower audio quality, considering cases where video distortions are more noticeable.

Furthermore, some marginal conclusions of our study can be derived regarding the analysis of data from undifferentiated content. Whilst cases 3 and 4 both vary video quality down to the lower level, case 3 does so gradually whereas case 4 does so abruptly. However, it is clear that in the subjective tests, they are statistically similar to each other ($p = 0.7995$). Thus, we may conclude that highly noticeable video distortions lead to a great impact on quality perception of an audiovisual sequence, regardless of bit rate changes being gradual or abrupt.

As already mentioned, the experiment included videos which are representative of two very different contexts. In the U2 videos, there is constant movement involving fast camera and light changes. On the other hand, Pink Floyd videos have less on-stage movement and both camera and light changes are, generally, slower. In terms of audio, U2 videos are louder and have a lot more interference from the audience. Taking this into consideration, ANOVA test was also run on subjective test results separately for each band, in order to analyse the influence of the type of content (Tables IV and V). Similar cases discussed in the previous paragraph were also found for both U2 and Pink Floyd bands video groups. Regarding the U2 band test data, it is possible to draw relevant conclusions from the similarities between case 5 and the reference, characterized by a p-value of 0.1098. In case 5, we simulated a gradual reduction of bandwidth requiring a combined minimum of 280 Kbps (V + A: 256 Kbps + 24 Kbps) without loss of audio-visual quality experienced by the end user. Using audio reduction, bit rate levels are dropped to a value close to the 256 Kbps simulated in cases 3 or 4, where quality perception is very much affected by low video quality. When looking at the results from Pink Floyd videos, it seems that audio bit rate influence is less negligible as opposed to

video. Nonetheless, case 5 is also similar to the reference with a p-value of 0.1071, along with case 1 (p = 0.0507), which simulates abrupt bandwidth reduction to 280 Kbps with video and audio quality levels dropping at the same time to 256 and 24 Kbps, respectively.

Some other marginal conclusions arise from separate content data analysis. For example, case 1 (video bit rate drops to 256 Kbps) for the U2 videos shows similarity with cases 3 (p = 0.0587), 4 (p = 0.0673) and 6 (p = 0.1461), where video bit rate drops to 128 Kbps. This shows that both smaller or bigger changes in video quality may cause identical losses in overall quality experienced by the end user, when visual content includes rapid movements and/or camera changes.

V. Conclusion

Developing accurate joint audio-visual models to predict perceived quality is very much a work in progress. Added to this is the complications arising from the MPEG DASH scenario whereby encoding rates vary frequently. In this paper, we analysed the joint effect of audio and video content quality on audio-visual quality experienced by the end user in the context of video streaming, using MPEG-DASH. Specifically, we aim to generate insights into possible trade-offs in relative bandwidth allocation to audio and video, when it comes to live music concert streaming.

A subjective test using mobile equipment was defined simulating a live music concert broadcast over a mobile network with varying aggregate bandwidth. We designed a number of different cases reflecting different relative bandwidth allocation. The audio content was by default encoded at a high bit rate (128 Kbps) as required for such live music concert scenarios. As typical with MPEG DASH, the videos were divided in chunks (of 10 seconds), the network overload was simulated on 2 or 3 consecutive chunks, by reducing the encoding bit rates of the video, audio or both media simultaneously.

On the basis of the results obtained from the subjective test, and relating back to our core research questions, we can conclude the following:

- Reducing the audio information bit rate during a small number of 10 seconds chunks does not affect the perceived quality of the audio visual information by the end user.
- By contrast, when the quality reduction was made to video, subjects perceive a reduction in quality. For instance, it is better to reduce the audio information bit rate from 128 to 24 Kbps in two chunks, instead of reducing the visual information from 256 to 128 Kbps for just one chunk.

In conclusion, we believe that these results will be of significant interest to DASH content providers and add significant insights into the search for an effective joint audio-visual model, which is being considered for future work.

Acknowledgment

The authors are very grateful to the Instituto de Telecomunicações - Fundação para a Ciência e Tecnologia (project UID/EEA/50008/2013) under internal project QoE-VIS, and to the Optics Center of Universidade da Beira Interior where this work has been conducted.

References

[1] I. Sodagar, "The mpeg-dash standard for multimedia streaming over the internet," *IEEE MultiMedia*, no. 4, pp. 62–67, 2011.

[2] T. C. Thang, Q.-D. Ho, J. W. Kang, and A. T. Pham, "Adaptive streaming of audiovisual content using mpeg dash," *Consumer Electronics, IEEE Transactions on*, vol. 58, no. 1, pp. 78–85, 2012.

[3] ITU-R Recommendation, BS.1387, "Method for objective measurements of perceived audio quality," 2001.

[4] ITU-T Recommendation, P.863, "Perceptual objective listening quality assessment," 2011.

[5] P. Pocta and J. G. Beerends, "Subjective and objective assessment of perceived audio quality of current digital audio broadcasting systems and web-casting applications," *Broadcasting, IEEE Transactions on*, vol. 61, no. 3, pp. 407–415, 2015.

[6] ITU-T Recommendation, P.247, "Objective perceptual multimedia video quality measurement in the presence of a full reference," 2008.

[7] J. G. Beerends and F. E. De Caluwe, "The influence of video quality on perceived audio quality and vice versa," *Journal of the Audio Engineering Society*, vol. 47, no. 5, pp. 355–362, 1999.

[8] J. You, U. Reiter, M. M. Hannuksela, M. Gabbouj, and A. Perkis, "Perceptual-based quality assessment for audio–visual services: A survey," *Signal Processing: Image Communication*, vol. 25, no. 7, pp. 482–501, 2010.

[9] M.-N. Garcia, F. De Simone, S. Tavakoli, N. Staelens, S. Egger, K. Brunnstrom, and A. Raake, "Quality of experience and HTTP adaptive streaming: A review of subjective studies," in *Quality of Multimedia Experience (QoMEX), 2014 Sixth International Workshop on*. IEEE, 2014, pp. 141–146.

[10] S. Tavakoli, K. Brunnström, K. Wang, B. Andrén, M. Shahid, and N. Garcia, "Subjective quality assessment of an adaptive video streaming model," in *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics, 2014, pp. 90 160K–90 160K.

[11] R. K. Mok, E. W. Chan, and R. K. Chang, "Measuring the quality of experience of HTTP video streaming," in *Integrated Network Management (IM), 2011 IFIP/IEEE International Symposium on*. IEEE, 2011, pp. 485–492.

[12] T. Hoßfeld, R. Schatz, E. Biersack, and L. Plissonneau, "Internet video delivery in Youtube: From traffic measurements to quality of experience," in *Data Traffic Monitoring and Analysis*. Springer, 2013, pp. 264–301.

[13] S. Tavakoli, K. Brunnström, J. Gutiérrez, and N. García, "Quality of experience of adaptive video streaming: Investigation in service parameters and subjective quality assessment methodology," *Signal Processing: Image Communication*, vol. 39, pp. 432–443, 2015.

[14] ITU-T Recommendation, P.910, "Subjective video quality assessment methods for multimedia applications," 1999.

[15] ITU-T Recommendation, P.911, "Subjective audiovisual quality assessment methods for multimedia applications," 1998.

[16] ITU-T SG-12, "Parametric non-intrusive assessment of TCP-based multimedia streaming quality, considering adaptive streaming P.NATS Terms of Reference (ToR)."

[17] F. Bossen, "Common test conditions and software reference configurations. doc. jctvc-k1100," in *11th Meeting: Joint Collaborative Team on Video Coding (JCT-VC) of ITU-T SG*, vol. 16, 2013.

[18] F. Bellard, M. Niedermayer *et al.*, "Ffmpeg," *http://ffmpeg. org*, 2012.

[19] T. Wiegand, G. J. Sullivan, G. Bjøntegaard, and A. Luthra, "Overview of the h. 264/avc video coding standard," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 13, no. 7, pp. 560–576, 2003.

[20] J. Herre and M. Dietz, "MPEG-4 high-efficiency AAC coding [standards in a nutshell]," *Signal Processing Magazine, IEEE*, vol. 25, no. 3, pp. 137–142, 2008.

[21] F. J. Massey Jr, "The kolmogorov-smirnov test for goodness of fit," *Journal of the American statistical Association*, vol. 46, no. 253, pp. 68–78, 1951.