1Mbps is enough:
Video Quality and Individual Idiosyncrasies in Multiparty HD Video-Conferencing

Marwin Schmitt[1], Judith Redi[2], Pablo Cesar[1,2], Dick Bulterman[1]
[1]CWI: Centrum Wiskunde & Informatica, Amsterdam, The Netherlands
[2]Delft University of Technology, Delft, The Netherlands
m.r.schmitt@cwi.nl, J.A.Redi@tudelft.nl, p.s.cesar@cwi.nl, dick.bulterman@cwi.nl

***Abstract*— Most video platforms deliver HD video in high bitrate encoding. Modern video-conferencing systems are capable of handling HD streams, but using multiparty conferencing, average internet connections in the home are on their bandwidth limit. For properly managing the encoding bitrate in video-conferencing, we must know what is the minimum bitrate requirement to provide users an acceptable experience, and what is the bitrate level after which QoE saturates?. Most available subjective studies in this area used rather dated technologies. We report on a multiparty study on video quality with HD resolution. We tested different encoding bitrates (256kbs, 1024kbs and 4096kbs) and packet loss rates (0, 0.5%) in groups of 4 participants with a scenario based on the ITU building blocks task. We discuss the influence of group interaction and individual idiosyncrasies based on different mixed models, and look at covariates engagement and enjoyment as further explanatory factors. We found that 256kbs is still sufficient to provide a fair overall experience, but video quality is noticed to be poor. On the higher bitrate end, most people will not perceive the difference between 1024kbs and 4096kbs, considering in both cases the quality to be close to excellent. Independent on bitrate, packet loss has a small but significant impact, quantifiable in, on average, less than half a point difference on a 5-point ITU scale.**

Keywords—qoe, subjective study, multiparty video-conferencing

I. INTRODUCTION

Desktop video-conferencing is one of the fastest growing technologies in real-time communication: CISCO reported a growth of 30% between 2013 and 2014[1], and Skype reported a 10x growth of group video-calling in the last two years[2]. Desktop video-conferencing systems have advanced in the recent years with improved imaging technologies (HD is available for displays and cameras, codecs have improved) and being capable of multiparty sessions. The user choice for one system over the other is, for a major part, dictated by the Quality of Experience (QoE) these systems can deliver. Hence, it is of major importance to understand how system factors and network conditions typical of regular households influence user satisfaction with multiparty video conferencing systems.

To fully understand QoE in multiparty video-conferencing, it is essential to perform interactive tests, besides passive ones. In passive tests, participants assess the quality of video clips [1] or recordings from video-conferencing sessions [2][3]. While easy to setup, these tests suffer from limited ecological validity: the focus in a video-conferencing session is on effectively interacting with the other participants via audiovisual communication; the level to which system factors impair this ability is not accounted for in passive tests. Interactive tests tackle this limitation by involving multiple test participants performing a (joint) task over a video-conferencing system.

Interactive studies, especially for multiparty scenarios, have been mostly looking at the effect of delay on QoE [4][5][2], or at different stream encodings in combination with different dynamic layouts [6]. The impact of encoding and loss rate on quality has been studied only in two-party scenarios [7][8][9]. The maximum quality studied in these experiments was VGA encoded with 2Mbps in H.264, but no significant improvement in QoE was recorded for streams encoded at bitrates higher than 0.77Mbps [8]. This is a relevant result, as it poses an upper bound to the bitrate to which streams need to be encoded to ensure a satisfactory experience, which does not exceed the bandwidth availability of most households. Nevertheless, it is unclear whether this bound still holds for higher resolution video, such as HD. Popular video delivery services encode their 720p videos within 2Mbps and 4Mbps[3]. A multiparty conference, with such bitrates, easily maxes out the typical broadband connections available at households[4]. It is therefore important to establish (1) what is the minimal encoding quality for 720p video to still deliver a satisfactory experience; (2) what is the maximum bitrate above which no significant QoE improvement can be achieved and (3) the extent to which packet loss, typically occurring in situations like a home WLAN, impacts on QoE, also depending on the encoding bitrate.

In this study, we tackle these questions by reporting about an interactive subjective test on visual quality in a HD multiparty video conferencing system. The study is designed around a task which requires audiovisual interaction, i.e., the ITU recommended building blocks task [10], extended for multiparty interaction and HD video. We use a desktop-based 4-way video-conferencing scenario with WQHD screens and 720p video-streams encoded in H.264. We test the effect of encoding the

[1] http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/VNI_Hyperconnectivity_WP.html

[2] http://blogs.skype.com/2016/01/12/ten-years-of-skype-video-yesterday-tod

[3] http://www.lighterra.com/papers/videoencodingh264/

[4] https://www.stateoftheinternet.com/resources-connectivity-2015-q3-state-of-the-internet-report.html

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

video streams to bitrate levels fitting the bandwidth delivered by different internet access technologies typical for domestic households: broadband, DSL and mobile. We inject packet-loss in the videostreams, simulating typical slightly impaired wireless networks. We then investigate the impact of bitrate and packet-loss on overall, audio and video quality as perceived by participants.

As QoE is not only a result of system factors [11], but largely depends on user and context factors [12], we also look at individual and group idiosyncrasies. Ratings from users reflect their personal preferences, expectations and previous experiences. In interactive scenarios, every test run has its unique conversation dynamic, adding further divergence in the experience. Group dynamics and group cohesion cast further confounding factors in assessing the impact of the system factors on QoE [13]. Thus, in our analysis, we also investigate which role group interaction and individual preferences play in the appreciation of QoE, by means of variance analysis of different mixed linear effect models [14] [15].

Our results are clear: 256Kbps encoding still gives participants a fair experience, even though they find the video quality poor. On the other hand, 1Mbps seems to be enough to convey high QoE. Distortions due to packet loss are noticed, but no dramatic decrease in quality is to be expected. The group interaction plays in most cases a smaller role than individual idiosyncrasies, which instead explain a large extent of the variance in the collected QoE ratings.

II. Study Design

The goal of this study was to measure the QoE of current state-of-the-art desktop video-conferencing systems in typical home situations. Of special interest are for us the factors which can vary dynamically: the network conditions. We thus designed the study to investigate the effect on QoE of two independent variables, bitrate and packet loss rate, set at levels typical of domestic environments.

With respect to bitrate, we envisioned three conditions. The "low encoding" one (256kbs up and 768kbs down), reflected mobile broadband or slow xDSL connections; the "medium encoding" condition mimicked a typical xDSL connections (1Mbps up / 3Mbps down) and the "high encoding" one (4Mbps up and 12Mbps down) reflected broadband like TV cable connections. Each bitrate level was further combined with either of two levels

Figure 1 Screenshot from Experiment

Table 1 System Setup

Hardware	Model Nuc 5i5ryh: Core i5u, 8GB Ram, SSD		
	Displays	Dell 27" 2560 x 1440 (WQHD)	
	Headsets	Creative Soundblaster Xtreme	
	Webcams	Logitech C920	
Fixed System Parameters	Resolution	1280x720 – per participant	
	Framerate	24 fps	
	Encoding	H264 (x264) with Tune zero-latency, ultrafast speed-preset, GOP size 24, no b-frames, sliced threads encoding	
	Audio	AMR encoded	
	Delay	One-way ca. 120 ms	
Conditions	Encoding Bitrate		
	LowEnc: 256kbps	MediumEnc: 1024kbps	HighEnc: 4096kbps
	Loss		
	None (0%)	Random (0.5%)	

of packet loss, i.e. (1) no packet loss, as would occur on a wired connection and (2) 0.5% packet loss, likely over an impaired wireless network [3]. This resulted in a full factorial design with 6 conditions.

As six conditions would be too many to be assessed by each participant (risking fatigue), we opted for a mixed blocked design. 28 participants (18 female, average age: 31.9, sd: 10), performed the experimental task in groups of 4, to mimic real multiparty video-conferencing activities. Each group was exposed to 4 of the 6 conditions, counterbalanced in order. Eventually, each condition was assessed by at least four groups, i.e. 16 participants. The QoE assessment was performed, after exposure to each condition, via three ITU questions with 5-point ACR scales, targeting overall, audio and video quality [10]. The questionnaire contained further 12 items regarding experience, engagement and conversation dynamics. At the end of the whole experimental session participants also filled in a 9-item questionnaire regarding demographical information and enjoyment of the task.

Apparatus. Four rooms (one per participant in the conversation) with similar lighting and background conditions were equipped with identical desktop computers, displays, webcams and headsets (see Table 1 for detail). The video-conferencing software used was QoE-TB [12], which was specifically designed for conducting QoE experiments. The software uses GStreamer for media processing and UDP as the transport protocol. The packet loss was introduced by dropping RTP packets on the senders buffer, so every participant would see the same distortions. The experiment conductor monitored the session, but was only visible and audible for the introduction. The screenshot in Figure 1 shows the timecode embedded in the video which was cropped in the video of participants (8px).

Experimental task and protocol. To elicit a conversation that would stress the use of video, we employed the building blocks task from ITU-T P.920 [10][4][5], and we adapted it for a multiparty scenario. Each participant in each group had the same unassembled Lego® model but only part of the instructions. The task of the group was to communicate the respective instructions over the system to complete the model construction. To stress the HD setting, we chose a model with fine pieces (smallest 0.5x0.5 cm).

Before the actual experiment began, participants were briefed about the research scope, gave written consent and were explained the procedure of the study. They were then escorted each to a different experimental room and seated 68cm from the monitor, as recommended by ITU-T P.913 [16]. The video-conferencing software was then started, and the experimenter joined the conferencing in the beginning to make sure everything was working properly (e.g. sound volume adjustments). Participants were then asked to begin the experimental task, and they were exposed to the first of four experimental conditions. Each condition lasted for 7 minutes, after which the system would automatically display the ACR scales for QoE rating. Before each new condition, the experimenter shortly spoke over the video-conferencing, asking participants whether they needed a pause. After all rounds, the final questionnaire was administered and participants debriefed.

III. Analysis

Our basic assumptions were that a higher bitrate leads to higher or equal quality ratings and higher packet loss to a lower or equal rating. Figure 2 shows average scores with 95% confidence intervals for the three dependent variables (overall, audio and video quality) in the six experimental conditions, ordered according to the expected perceived quality.

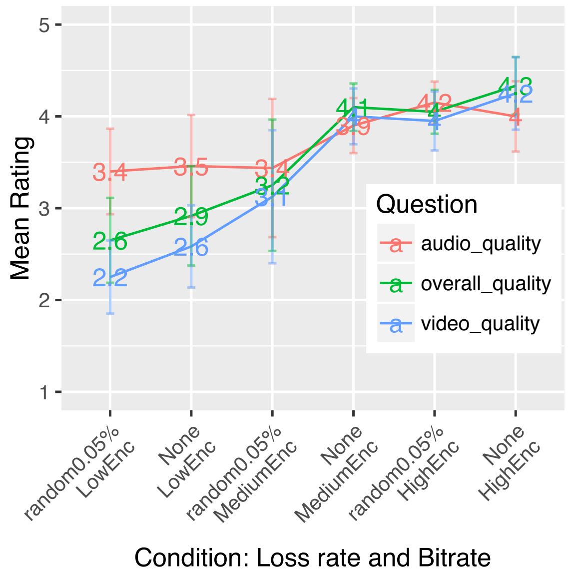

Figure 2 Mean ratings with 95% confidence intervals per condition

Methodology. To fully understand the impact that bitrate and loss have on QoE ratings,we need to untangle it from that of individual idiosyncrasies of participants and group dynamics of a specific sessions. To this purpose, we resort to linear mixed effect models, which extend linear models (such as ANOVAs) by introducing the concept of random factors. Linear models assume that dependent variables (in our case the QoE scores) can be modeled by a linear combination of the levels of the indpendent variables of interest, the so-called fixed effects (here: *bitrate* and *loss*).

In a linear mixed model, the concept of 'random effect' is introduced to explain the systematic impact that unobserved variables, uncorrelated with the fixed effects, may have on the dependent variable. This is especially useful to model data obtained from within-subjects designs (such as ours), where observations (ratings) cannot be considered to be indipendent (as they would be in linear models), since they are expressed by the same user or within the same session. For example, ratings from the same user may be correlated due to unobserved factors (e.g. mood, prior experience). This correlation can be modeled as a random effect, and in mixed models it is taken into account by estimating an individual intercept and slope for the fixed effect(s), depending on the level of the random effect. A visualization of this is proposed in *Figure 3*, where we model group as random factor per bitrate. In analyzing our data, we consider the influene of two random factors:

- *User*: Having repeated measures from (a random sample of) participants, we may expect them to use the rating scale in different ways, or to have different quality preferences (or idiosyncrasies).
- *Group*: The specific group interaction and conversation (again randomly sampled from all possible interactions and conversations) may influence the experience of all members of this group, and thereby their ratings.

A linear mixed model follows the general structure:

$$y = X\beta + Z\gamma + \epsilon \qquad (1)$$

Where y is the vector of our responses (different quality ratings) of length n = 112 (7 groups, 4 participants per group, 4 ratings per participant); X is the design matrix for fixed effects, so with a maximum size of n x 11 (3 bitrate levels + 2 loss levels + 6 interactions); β are the coefficients of the fixed effects; Z is the design matrix for the random effects, with a maximum size of n x 224 (28 users + interactions of 28 users with 7 groups); γ are the coefficients of the random effects and ϵ is the vector of

Table 2 P-values from the likelihood ratio test between full model and reduced model (unfiltered dataset/filtered dataset), testing the impact of removing a fixed or random effect on goodness of fit. A p-value < 0.05 indicates that the full model is a significant better fit than the reduced model, thereby indicating that the influence of tested effect on QoE is significant. We did not find obvious violations of normal distribution and homoscedasticity of the residuals of the presented models.

Effect tested		**interaction between bitrate and loss**	**loss**	**bitrate**	**User**	**Session**
Full model		**M1)** ~ bitrate + loss + loss:bitrate + (bitrate|Group/User)	**M2)** ~ bitrate + loss + (bitrate |Group/User)	**M2)** ~ bitrate + loss + (bitrate | Group/User)	**M2)** ~ bitrate + loss + (bitrate | Group/User)	**M2)** ~ bitrate + loss + (bitrate | Group/User)
Reduced model		**M2)** ~ bitrate + loss + (bitrate | Group/User)	**M2B)** ~ bitrate + (bitrate|Group/User)	**M2L)** ~ loss + (bitrate|Group/User)	**M3)** ~ bitrate + loss + (bitrate|Group)	**M4)** ~ bitrate + loss + (bitrate|User)
Dep. Variable (y)	**overall_quality**	0.689/0.656	0.006/0.009	0/0.001	0.004/0.012	0.038/0.999
	video_quality	0.206/0.274	0.011/0.005	0/0.001	0.148/0.325	0.006/0.301
	audio_quality	0.165/0.091	0.722/0.976	0.03/0.052	0.001/0.005	0.003/1

the residuals of length n. It is assumed that the random effects are independent and distributed as $N(0, \tau^2)$, the errors are independent and distributed as $N(0, \sigma^2)$, and the random effects and errors are independent. We specify then the model eq. (1) in different ways to test the impact of fixed and random effects on QoE. For brevity, we denote the models using the *R* notation:

y ~ bitrate + loss + bitrate:loss + (bitrate | Group/User)

with the specific notation

- *bitrate:loss* denoting the interaction between bitrate and loss
- (*bitrate|Group/User*) being short for (*bitrate|Group*) + (*bitrate|Group:User*), where in turn *(bitrate|Group)* denotes a random effect of the group per bitrate and (*bitrate|Group:User*) denotes a random effect of the user per group per bitrate

To evaluate the goodness of the fit of the different models and relation between our independent variables, effects of group and user, we compute the marginal R^2 (variance explained by fixed effects, higher is better), conditional R^2 (variance explained by fixed effects and random effects, higher is better) and AIC (measurement of goodness of fit in relation too number of parameters used; where smaller is better [17]) in Figure 4. The values were obtained as described in [18].

Results. As our data set is not big enough to build random slopes per both bitrate and loss, we started our analysis by comparing a model with random slope per bitrate (M1 in Table 2) with a model with random slopes per loss (y ~ bitrate*loss + (loss|Group/User)). The model with random slopes per bitrate performed better on all three dependent variables (conditional R^2 = 0.785/AIC=287, for overall quality in M1, R^2 = 0.591/AIC=309 for the same dependent variable, with the other model; similar results were obtained for video and audio quality), thus we choose to model the random slopes per bitrate.

We investigated the impact of the effects (fixed: bitrate and loss, random: group and user) on QoE by comparing, via a likelihood ratio test (lrt) [14], different versions of the same model, herewith called full and reduced models. The reduced

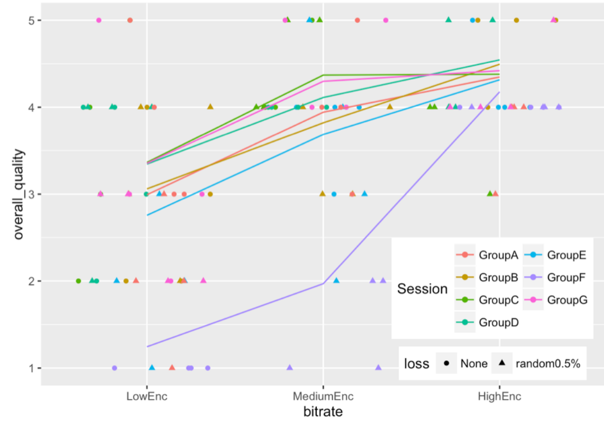

Figure 3 Ratings and fitted Model 2 per Session

models include only a subset of the effects considered by the full models; if the lrt returns a p-value smaller than 0.05, this indicates that the reduced model has a significantly worse performance than the full one, and hence the omitted effect has a significant influence on the dependent variable. For example, to test whether an interaction of *bitrate* and *loss* has a significant impact on QoE, we perform the lrt between the full model with interaction (M1) against the model without interaction (M2 in *Table 2*). The lrt shows that for here is no significant interaction between *bitrate* and *loss* for all our assessed quality ratings ($p > 0.05$ for overall, video and audio quality). Based on this finding, we proceed to test the different effects from M2.

The p-values resulting from the lrt between M2 and its reduced versions investigating the effect of loss, bitrate, user and group on QoE are shown in *Table 2*. *Loss* is a significant factor for all dependent variables except audio quality (see p-values for M2B in *Table 2*). *Bitrate,* on the other hand, is a main effect for all dependent variables (see M2L). The *User* effect impacts the ratings of overall quality and audio quality (M3). *Group,* on the other hand, impacts all dependent variables (M4).

Upon closer inspection of the data, we observed that there was one group showing a markedly different behavior. Figure 3

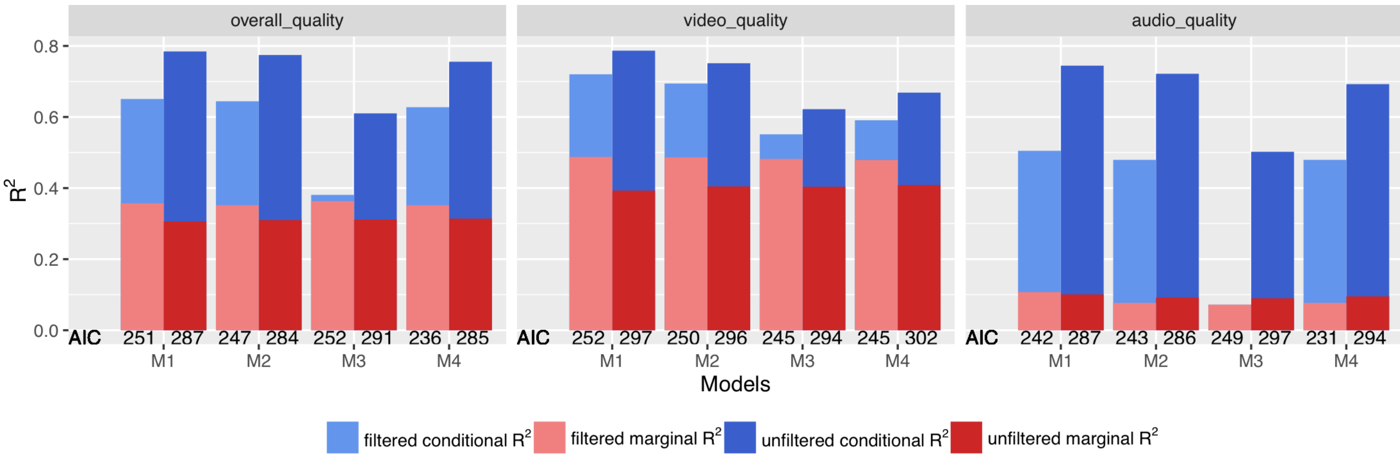

Figure 4 Marginal R2, Conditional R2 and AIC per Model

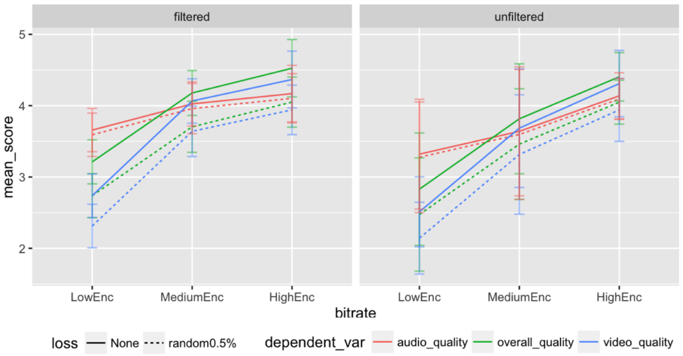

Figure 5 Fitted Model 2

plots the fit of individual groups from M1 (see Table 2) as lines, the raw ratings as points (jittered on the x-axis) against the different bitrate levels: it is clear that ratings from Group F are somewhat anomalous. As we have no indication that these are measurement errors, we performed our analysis again, on all data except those related to this anomalous group (hereon we refer to this set of data as the "filtered" set). The results for the effects *bitrate*, *loss* and *user* are unaltered, as shown in Table 2. However, leaving out the *group* factor (M4) does not result anymore in significantly worse fit

In Figure 4, we plotted a bardiagram for each dependent variable. Each bar shows the marginal R^2 (red) and the conditional R^2 (blue) of a model, for both unfiltered (darker colors) and filtered data (lighter colors). Additionally, we noted the AIC of each model above the label. Comparing the filtered with the unfiltered dataset, we observe that the proportion of variance explained by the fixed factors (red) is generally higher in the filtered data. While the total explained variance is higher for the unfiltered dataset, the AIC indicates that the models perform better on the filtered dataset in comparison to how many parameters they need for explaining the variance. The perceived audio quality is generally poorly explained by the fixed effects, even though the total explained variance is quite high. Further the difference in conditional R^2 between the model including group as a random effect (M3) and the model including user as a random effect (M4) is large, showing that the audio quality was perceived very different for participants at a base level. This is particularly obvious for the filtered dataset in which including *group* as random effect does not lead to any more explained variance. Video quality has the largest marginal R^2 indicating that *user* and *group* factors play the smallest role for these assessments. For video quality we can see that using *user* as a random effect instead of *group* leads to little improvement.

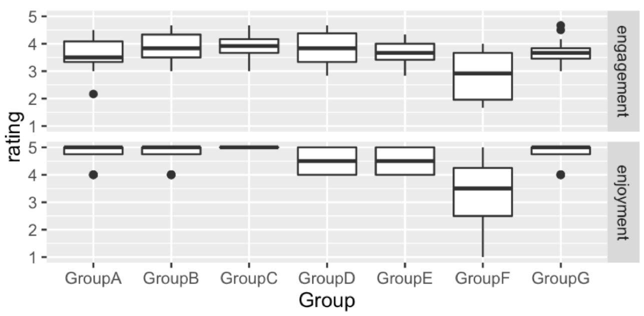

Figure 6 Boxplots of covariates per Group

Table 3 P-values by fixed factors (unfiltered dataset/filtered dataset)

question	LowEnc-HighEnc	LowEnc - MediumEnc	MediumEnc - HighEnc	None – random0.5%
overall_quality	0/0	0/0	0.22/0.25	0.01/0
video_quality	0/0	0.01/0	0.3/0.38	0.02/0.01
audio_quality	0.07/0.02	0.28/0.07	0.43/0.77	0.75/0.66

Taking into account also the interaction of *group* and *user* (M2) gives for video quality to the largest gain in R^2, indicating that group interaction as well as individual idiosyncrasies play an important role. For overall quality, in all models fixed effects explain less variance than they do for video quality; nevertheless, a similar conditional R^2 is achieved.

To precisely quantify the impact on QoE of the fixed effects, we performed a post-hoc analysis of the individual conditions using multivariate *t*-distribution adjustment for multiple comparisons. Table 3 shows the p-values of the pairwise comparisons of the QoE assessments per each pair of levels of each fixed effect. The data shows that for nearly all our variables there is a clear difference between the low bitrate condition and higher bitrates, but users cannot differentiate between medium and high bitrate encoded streams. For audio quality, only the difference between low and high quality is significant, and only in the filtered dataset. We have plotted the mean values and 95% confidence intervals in Figure 5. As we can see a lot of the variance in the unfiltered model was due to Group F.

Finally, we looked into what could explain the different results for Group F. We found obvious differences in this group ratings in another two covariates: reported level of engagement and reported level of enjoyment. In the boxplot in Figure 6 it is noticeable that the group was less engaged and enjoyed the session less. An ANOVA with the different Groups as fixed factor showed that for engagement (which was assessed after each round) the ratings from Group F were lower those of the other groups ($F(6,105)=6.533$, $p<0.001$). For enjoyment (assessed at the end of the experiment session) the results are not as clear but still show a negative trend ($F(6,21)=2.1538$, $p=0.08949$, contrast for Group F being different $p=0.0135$).

IV. DISCUSSION

Low quality video, with visible encoding and loss artifacts, was rated by users as, mostly, poor quality (mean unfiltered/filtered 2.33/2.52). For overall quality users tended more towards *fair* (mean unfiltered/filtered 2.65/ 2.97). Both medium and high bitrate delivered a good quality of experience with most users rating good or excellent (overall quality 3.95/4.29). Except for our anomalous group (1.5 – 4.0), the difference between both conditions is small (0.33). The video quality is again rated slightly lower (3.86 - 4.15). The influence of packet loss, while significant, was rather small (average 0.367/ 0.425 difference between a loss and no loss condition). Surprising is that there was no interaction between packet loss and encoding bitrate. The impact of packet loss was also so small, that it did not impact the audio quality ratings negatively.

Our results showed that the perceived audio quality was clearly affected due to the impairment of the video quality. This confirms studies on cross-modal effects of audio and visual quality [19].

The initial analysis of our complete dataset showed that differences between groups play a big role (compare R^2 values between models M2 and, M3 and M4 on unfiltered dataset). On closer observation this was one group who seemed to have a very different experience than the other groups (compare R^2 model M2 between filtered and unfiltered dataset). Only for video quality leaving out the user factor (model M3) did not lead to a significant worse fit, suggesting that the impression of the video quality, was the most consistent rated within groups from our questions. The impact of bitrate and loss was the clearest on video quality (highest marginal R^2). It is interesting that while overall quality was quite consistently rated a bit higher than the video quality, the variance here was mostly due to user factors and not on a group level. This suggest that the impact of video quality on the overall experience is even in visual challenging scenarios, a personal preference.

The building blocks task has been used in previous studies to evaluate video quality in video-conferencing [4][9][8]. However the setup and assessed conditions are too different for a direct comparison

V. Conclusion

We investigated different encoding and packet loss settings that could be typical situations in the home. Participants that have to use the low bitrate encoding (up 256kbs/down 756kbs) still have an okay (poor-fair) experience. If the connection has also packet loss, similar to an impaired wireless connection, the ratings tend more towards *poor*. We would conclude that 256kbs delivers an overall satisfactory experience, but participants will not be impressed by the video and often have a poor impression of it (research question 1). The difference in quality perception between 1024kbs and 4096kbs is rather small (too small to have a significant impact in our study) and most people will give it a good or even excellent rating. Users on broadband connection thus rarely will have a better experience than users with a good DSL connection and we can conclude that 1Mbps is enough as encode bitrate for HD streams in video-conferencing (research question 2). The impact of packet loss was noticeable but rather small in our experiment (research question 3). In the high bitrate cases, it seems thus clear that it should be recommended to assign more bandwidth to forward error correction (FEC) than upping the bitrate more. In the low bitrate case both approaches seem to be viable. The bandwidth usage of FEC and quality improvements between 256kbs and 1024kbs would need to be studied in detail to give concrete advise. Packet loss will be noticed but is acceptable to most participants, so lossy protocols like UDP should still be the favorable choice for video-conferencing.

In our analysis we showed that user and group factors both are crucial in understanding the obtained participant ratings. In general the correlation between ratings of the same individual is stronger than the one of ratings within the same group. Groups seem to be more consistent when it comes to ratings of the video quality than the overall quality. Generally, we can see that our models often deliver high overall R^2, but a rather small marginal R^2. This indicates strong influences from either the user or the group interaction. In future work we plan to extend our models with more covariates from our additional questionnaire ratings and analysis of the recorded videos.

VI. References

[1] J. Nightingale, Q. Wang, C. Grecos, and S. Goma, "The impact of network impairment on quality of experience (QoE) in H.265/HEVC video streaming," *IEEE Trans. Consum. Electron.*, vol. 60, no. 2, pp. 242–250, May 2014.

[2] G. Berndtsson, M. Folkesson, and V. Kulyk, "Subjective quality assessment of video conferences and telemeetings," in *Packet Video Workshop (PV), 2012 19th International*, 2012, pp. 25–30.

[3] M. Ndiaye, M. C. Larabi, H. Saadane, G. L. Lay, C. Perrine, C. Quinquis, and L. Gros, "Subjective assessment of the perceived quality of video calling services over a real LTE/4G network," in *2015 Seventh International Workshop on Quality of Multimedia Experience (QoMEX)*, 2015, pp. 1–6.

[4] F. Brauer, M. S. Ehsan, and G. Kubin, "Subjective evaluation of conversational multimedia quality in IP networks," in *2008 IEEE 10th Workshop on Multimedia Signal Processing*, 2008, pp. 872–876.

[5] K. Schoenenberg, A. Raake, and P. Lebreton, "Conversational quality and visual interaction of video-telephony under synchronous and asynchronous transmission delay," in *2014 Sixth International Workshop on Quality of Multimedia Experience (QoMEX)*, 2014, pp. 31–36.

[6] S. N. Gunkel, M. Schmitt, and P. Cesar, "A QoE study of different stream and layout configurations in video conferencing under limited network conditions," in *Quality of Multimedia Experience (QoMEX), 2015 Seventh International Workshop on*, 2015, pp. 1–6.

[7] S. Egger, M. Ries, and P. Reichl, "Quality-of-experience beyond MOS: experiences with a holistic user test methodology for interactive video services," in *21st ITC Specialist Seminar on Multimedia Applications-Traffic, Performance and QoE*, 2010, pp. 13–18.

[8] B. Belmudez and S. Möller, "Audiovisual quality integration for interactive communications," *EURASIP J. Audio Speech Music Process.*, vol. 2013, no. 1, pp. 1–23, Nov. 2013.

[9] B. Belmudez, S. Moeller, B. Lewcio, A. Raake, and A. Mehmood, "Audio and video channel impact on perceived audio-visual quality in different interactive contexts," in *Multimedia Signal Processing, 2009. MMSP'09. IEEE International Workshop on*, 2009, pp. 1–5.

[10] ITU-T RECOMMENDATION, "ITU-R P.920 - Interactive test methods for audiovisual communications," 2000.

[11] Patrick Le Callet, Andrew Perkis, and Sebastian Möller, Eds., "Qualinet White Paper on Definitions of Quality of Experience (2012). European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003). Version 1.2." Mar-2013.

[12] M. Schmitt, S. Gunkel, P. Cesar, and P. Hughes, "A QoE Testbed for Socially-aware Video-mediated Group Communication," in *Proc. of the 2nd International Workshop on Socially-aware Multimedia*, New York, NY, USA, 2013, pp. 37–42.

[13] J. Skowronek and A. Raake, "Conceptual Model of Multiparty Conferencing and Telemeeting Quality," in *2015 Seventh International Workshop on Quality of Multimedia Experience (QoMEX)*, 2015.

[14] A. C. Davison, *Statistical Models*, 1 edition. Cambridge University Press, 2008.

[15] J. Pinheiro and D. Bates, *Mixed-Effects Models in S and S-PLUS*. Springer New York, 2000.

[16] P. ITU-T RECOMMENDATION, "ITU.P913: Subjective video quality assessment methods for multimedia applications," 1999.

[17] H. Akaike, "Information Theory and an Extension of the Maximum Likelihood Principle," in *Selected Papers of Hirotugu Akaike*, E. Parzen, K. Tanabe, and G. Kitagawa, Eds. Springer New York, 1998, pp. 199–213.

[18] S. Nakagawa and H. Schielzeth, "A general and simple method for obtaining R2 from generalized linear mixed-effects models," *Methods Ecol. Evol.*, vol. 4, no. 2, pp. 133–142, Feb. 2013.

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