

# A QoE study of different stream and layout configurations in video conferencing under limited network conditions

Simon N.B. Gunkel, Marwin Schmitt, Pablo Cesar  
 CWI: Centrum Wiskunde & Informatica  
 Science Park 123, 1098 XG Amsterdam, Netherlands  
 gunkel@cwi.nl, schmitt@cwi.nl, p.s.cesar@cwi.nl

**Abstract**— One particular problem of QoE research in video conferencing is, that most research in the past concentrated on one-to-one video conferencing or simply video consumption. However, video conferencing with two people (one-to-one) and within a group (multi-party) is different. Particularly, limitations of one participant might have an effect on the QoE of the whole group. This possible effect however is not well studied. Therefore, this paper aims to better understand the impact of individual limitations towards the groups QoE. To do so, we show a study about different video stream configurations and layouts for multi-party conferencing in respect to individual network limitations. For this, we conduct a user study with 20 participants in 5 groups, in a semi-controlled setup. Such a setup, combines supervising participants locally while still using our software infrastructure deployed in the internet. Furthermore, we use an asymmetric experiment design, by putting every participant under a different condition, as this proposes a more realistic scenario. Within our study, we look at three different factors: layout, video quality and network limitations. To foster conversation between participants, the group engaged in a discussion about different survival questions. Our findings show that packet loss and the resulting distortions have a greater impact on the QoE as reducing the video quality by its resolution. Furthermore, our findings indicate that participants are more satisfied in a visually equal layout (showing participants in a similar size) and a more balanced stream configuration.

**Keywords**—*Quality of Experience; Video Quality; User Perception; Computer-mediated Communication; Multi-party Video Conferencing; Error; Loss*

## I. INTRODUCTION

The wide popularity of video conferencing systems does not necessarily translate on the provision of fully satisfying user experiences. Technical difficulties, delays, and blurred videos are still commonplace in commercial services such as Skype and Google+ Hangouts. Even though some optimization mechanisms do exist [1][2][3], the use of conversation dynamics of participants in such systems is rather limited (e.g. whether the participants are in front of the computer or not). This is also large due to limited research work in the aspect of group conversation analysis in video conferencing. One particular problem in group video conferencing is that even if only one computer is on a low internet connection, or suffering

of problems, the experience for all participants will suffer. Thus, the study presented in this paper explores the relationship between QoE and three different factors: layout, video quality (resolution) and network limitations (packet loss). In detail our study analyses four different layout/stream configurations, with different bandwidth profiles and thus different network problems (packet loss). This study included 20 participants in 5 groups (4 people in each group) that were invited to our institute. We structured our study in a semi-controlled environment which allowed us to control the environment of participants (similar PC, room, lighting) while utilizing our server infrastructure in the internet. This is a practical approach that still results into realistic network conditions. Furthermore, the four layout/stream configurations were conducted in an asymmetric experiment design, giving every participant a different condition (in random order). This is also a practical and realistic approach and currently under the investigation of the ITU [4].

## II. RELATED WORK

Research regarding the QoE in multi-party video-conferencing has gained track in the last years. At the moment, the challenge of providing users with an optimal video-conferencing experience is tackled from two sides: understanding the social context of video-conferencing [5] and providing reactive systems based on this [6] and on the other hand understanding the influence of system, mainly network characteristics, on the QoE [7]. This is done by extensive user trials while emulating different conditions of the network [1][3]. However, the relationship between network performance values and the perception of users is still not fully clear [8]. Furthermore, those approaches are limited, as they do not incorporate information about the environment and the user context (e.g. the user task) [1][2][3]. Recently there have been improvements in understanding the influence of network factors based on the speech-patterns of the ongoing conversation [9][10]. In a similar manner, those audio cues are at the moment what orchestration relies on most. Since the appreciation of dynamic orchestration as well as the influence of network properties is shown to be connected to conversation dynamic, this study investigates how (or whether there is) an interaction between these factors.



Fig. 1. Focus+Context Client Layout



Fig. 2. Fixed Client Layout

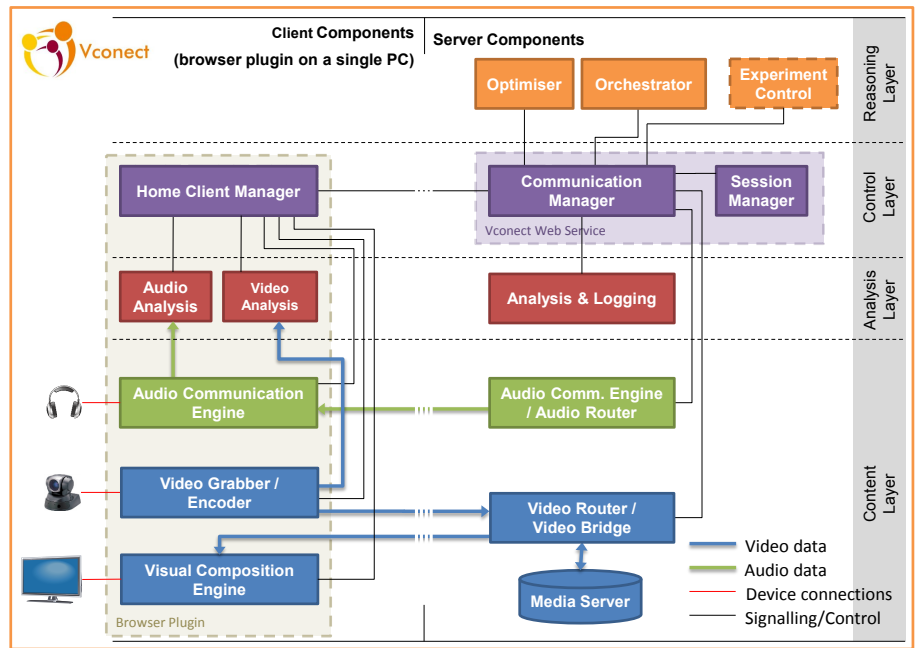


Fig. 3. System Architecture Diagram

In this paper we deal with spatial presentation of the participants (layout and orchestration) and network characteristics (bandwidth and packet loss) which are normally studied as separate entities. In the one-to-one video-conferencing systems, it was not necessary to provide a special layout for the participants. The two major alternatives which are used in current video-systems are to distribute all participants in equal size (Figure 1, e.g. used in Skype) or a focus+context approach (Figure 2, e.g. used in Google+ Hangouts) where one participant is displayed larger in focus and the others with small previews only. The focus+context approach requires the system to take decisions on when to put which participant into focus, often referred to as orchestration [11]. For tele-immersive video-conferencing different setups were investigated which tried to replicate the face-to-face setup as well as possible (e.g. employing life-size displays, preserving gazing and gesturing) [12][13]. For desktop video-conferencing, the aspect of gaze awareness and window layouts has been studied [14].

Packet loss is a common problem in real-time communication and has extensively studied, e.g. for video-streaming [15] [16], voip calls (e.g. [17]) and extensive analyses based on encoding and streaming characteristics (e.g. for H264 video [18]). In the context of video-conferencing, it has been investigated for one-to-one videoconferencing with subjective tests [19] and simulations [20] but not yet in a multiparty environment.

### III. SYSTEM AND ARCHITECTURE

Our video communication system used was developed in the European Framework 7 project Vconnect. It uses a server-centric architecture, due to the advantages in transmission delay against a P2P approach [21][22]. Figure 3 shows the architecture of our communication system. It is divided into client components (endpoints) on the left hand side, and server-

side components on the right. Functionality is grouped into 4 layers. Whereas other systems mainly implement layers for communication management, we added layers for measuring and reasoning. This allows global management, like optimisations and reconfiguration of components across the whole system.

#### A. Client Components

At the Content Layer, the client side of the platform includes components for capturing, encoding and transmitting video streams from a high-definition camera and audio streams from a microphone. These components are largely equivalent to those found in conventional video chat or conferencing systems. This layer also incorporates components which receive, decode and compose multiple audio and video streams for presentation to the user. Moving up to the Analysis Layer, the client includes components for the automated analysis of the captured audio and video streams.

#### B. Server-side Components

The role of the server-side components is to facilitate social communication between a large number of endpoints. At the Content Layer, two main components provide scalable transmission of audio-visual streams: the Audio-Router and the Video Router. Both components function as an efficient packet switch and replicator which connects multiple source audio/video streams to multiple clients. The Control Layer includes the Communication Manager and the Session Manager. Together, they provide the hub of the platform's communication framework, enabling messages to be transmitted between components, and enabling users to find each other and join a session.

Finally, the Reasoning Layer contains the Optimiser and Orchestrator component, supplemented by an Experiment Control component. The Optimiser ensures optimal use of the

network at lowest cost, while maximizing the Quality of Experience for users. However, its functionality is limited in this study to follow our specific experimental design, instead of executing dynamic optimisations. Driven by conversational properties the Orchestrator takes decisions on the visual layout and whom to show. Finally the Experiment Control adjusts the algorithm to the different study conditions, and helps to synchronize log data from the different components.

Our optimisation process takes three types of information into consideration: the network (e.g. network component workload, traffic limits, and congestion), the client (bandwidth, delay, video analysis, and audio analysis), and the user (context, task, social network analysis, and real-time feedback). To utilise this information the Optimiser includes different models that map the information to optimisation rules. These rules trigger specific settings for the network and client components. In this study, we follow a fixed rather than dynamic optimization approach, based on our 4 study conditions. The goal here is to better understand the possible QoE impact of different strategies and client layouts.

#### IV. STUDY SETUP AND METHODOLOGY

In this study, we focus on the relationship between QoE and three different factors: layout, video quality (resolution) and network limitations (packet loss). Our study included 20 participants in 5 groups, with 4 people being in each group. The participants are from 16 different nationalities, with an average age of 27 (SD 5), and 6 different professions (with 70% being students or PhD candidates from different fields). Each group engaged in a wilderness survival task that included 12 questions, in 4 rounds of 3 questions. Each question had to be answered in a group discussion. To give the discussion some structure we selected in each group a moderator to keep the conversation flow. Furthermore, we introduced four different layout/stream configurations to our participants, each resulting in a different visual representation, a different bandwidth profile and a finally a differently network packet loss rate. This resulted into the following research questions:

**RQ1. How is the relationship between individual influencing factors and the QoE?**

**RQ2. What is the impact of network restrictions to the choice of layouts?**

**RQ3. Can we improve the QoE by choosing a particular layout and stream configuration based on the network limitations?**

TABLE I. OVERVIEW OF CONDITIONS

Condition	Layout	Streams	Available bandwidth
C1	Focus+Context	1 HQ, 2 LQ (2 HQ and 3 LQ during switch)	High
C2	Fixed	2 HQ, 1 LQ	High
C3	Fixed & Focus+Context	C1/C3 (3 HQ and 3 LQ during switch)	High
C4	Fixed	3 LQ	Low

##### 1) Study Conditions

Table 1, Figure 1 and Figure 2 show the different configurations of conditions. We are using an asymmetric setup in which each client is a different condition. Three of the clients are configured for a connection with a high bandwidth (C1-C3) while one client is configured for a low bandwidth connection (C4). The low bandwidth client uses a fixed layout (all participants in equal size in a 2x2 grid layout, see Figure 2) while we tested a different layout in each of the other setups. The configuration C1 employs Focus+Context style layout (see Figure 1). The currently speaking participant will be shown in high-quality (HQ) while the other participants are shown in low-quality (LQ). Thus this client receives normally 1 HQ stream and 2 LQ streams. But in the moment when streams are switched (i.e. a different participant is shown in the main view) the new streams (HQ for the future main view participant, LQ for the current main view participant) are requested in parallel to the old still active streams. This is necessary to make a smooth transition, when switching the participant in the main view [23]. Furthermore the low bandwidth client is shown as a LQ stream. The configuration C2 is a fixed layout (see Figure 2), in the same way as C4, but with HQ streams from the participants on a high bandwidth configuration. The configuration C3 combines the approaches of C1 and C2. Normally C3 uses the Focus+Context approach of C1, but switches to the fixed view if the user on the low bandwidth client is currently speaking. To make a not too fast switch between both layouts, the fixed layout is maintained for at least 15 seconds before switching again to the Focus+Context layout. To encode the streams we use the Fraunhofer Video Encoder with the following settings:

- LQ – low quality stream: resolution = 320x180, framerate = 15fps, bitrate = 128kb
- HQ – high quality stream: resolution= 1280x720, framerate = 15fps, bitrate = 800kb

##### 2) Procedure

In the structure of the experiment, we first meet as a group in person to give an introduction and to let everyone meet each other. Afterwards we divide people into different (but similar) rooms, with a similar PC setup:

- Desktop PCs (Core i7, 16GB Ram, SSD, windows 7)
- Webcam (Logitech HD C920)
- Headset (creative sound blaster tactic3d alpha)
- Screens (Dell UltraSharp U2713HM, set on 2560x1440 WQHD resolution)
- Private Room (similar in furniture setting and lighting)

While setting up the system each participant was encouraged to look at all survival questions in order to form an opinion. In every conversation round each client was in a different condition (in a random order) so that each client was in each condition. Thus, we had a fully randomized within-subject design. Each conversation round was closed by filling out a questionnaire. Each questionnaire included 12 questions, assessed on a 9-Point likert-like scale (only endpoints marked). We originally had three questions directly concerned with



TABLE I. QUESTIONNAIRE RATING AND LOSS-RATE AS AVERAGE VALUE

Condition	Overall Experience	Video Quality	Audio Quality	Distortion	Total Loss-rate
Overall	7,66 (SD 1,124)	7,11 (SD 1,232)	8,01 (SD 1,049)	4,66 (SD 2,283)	0,115% (SD 0,162%)
C1	7,75 (SD 1,118)	7,15 (SD 1,349)	8,25 (SD 0,911)	4,38 (SD 2,419)	0,108% (SD 0,145%)
C2	7,50 (SD 1,192)	6,85 (SD 1,309)	8,05 (SD 1,146)	4,75 (SD 2,082)	0,153% (SD 0,107%)
C3	7,55 (SD 1,050)	7,20 (SD 0,834)	7,75 (SD 1,118)	5,69 (SD 1,815)	0,182% (SD 0,245%)
C4	7,85 (SD 1,182)	7,25 (SD 1,410)	8,00 (SD 1,026)	3,81 (SD 2,536)	0,018% (SD 0,018%)

QoE, we asked participants for their overall experience and the quality of the audio and video. We further asked for details of the visual quality (how good they could see facial expressions and how life-like participants appeared), different aspects of the conversation dynamics (how easy it was to keep track of the conversation, how lively the discussion was, if they and others started to speak at the same time (false start) and whether they noticed awkward silences in the conversation) and group dynamics (how “close” they felt to other participants and how satisfied they were with the group discussion). In the debriefing of the first group, we noticed that, there were differences whether participants would reflect visual distortions in the video quality and overall experience ratings. Thus we decided to add a question, asking whether they saw any distortions in the video (thus for this question we have only responses from 16 participants). Finally, each participant had to answer a final questionnaire and we finished the study with a semi-structured face-to-face group interview. In this way we collect information from the System (network data and audio communication data) and questionnaire data, related to each round and the overall study.

## V. RESULTS

In this section we show the different results of our study, this includes 3 different data: (A) the questionnaire rating and network loss according to the condition, (B) differences in the correlation of properties in the different conditions and, (C) the ranking of the different conditions.

### A. Questionnaire Rating and Network Loss

Table II shows the average rating of the questionnaire answers and the measured total packet loss-rate. The Overall Experience, Video Quality and Audio Quality is rated similar across the whole study. The responses were not normal distributed according to the Kolmogorov-Smirnov test and the Friedman rank sum test showed no significant differences between conditions. This is not the case for the distortion which show, particularly in C3, significant differences to the other conditions (responses normal distributed according to Kolmogorov-Smirnov test, sphericity was not violated according to Mauchly Tests for Sphericity, a one-way repeated measure ANOVA showed a significant effect of the Conditions on the distortion rating ( $F(3,45) = 3.018$ ,  $p < 0.05$ )). This suggests, that it is difficult for our participants to differentiate between the different factors that influence the QoE, as one participant describes it [P8] “sometimes I found it hard to disentangle what is purely a graphical problem or graphical advantage and what due to the blur in front of it”. This also lead to some clashing of opinions about the system for example like [P10] “to be honest I have used skype before with this multiple video thing in the past, but it was very annoying. This

one (our system) was much better and much easier.” and [P13] “it (our system) is like skype”. This suggests that users were not always engaged or focused on the video but relied more on the audio in some cases. This is also visible when looking at the Conditions and Correlations (B).

### RQ2. What is the impact of network restrictions to the choice of layouts?

What is not visible in Table II, is that the loss mainly occurs in the high resolution streams. The reason for this is likely to be due to higher data transfer rates and peaks while switching streams, as well as transmitting I-frames. While this needs further investigation it is clear that participants suffered from distortions even with a very low error rate. This however behaves different with the lower resolution streams, showing less errors and distortions. Overall, therefore that the distortion rating and the loss rate follow the same pattern we can say that the error has more impact on the QoE than the lower quality due to lower resolution and bitrate.

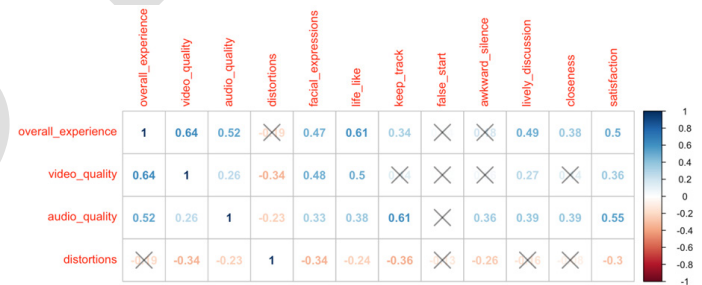


Fig. 4. Overall Questionnaire Correlation

### B. Conditions and Correlations

Figure 4 and Figure 5 show correlations between the different ratings of the questionnaire throughout the conditions. The correlations focus on 4 properties: the overall rating of the experience (overall\_experience), the video quality, the audio quality, and the distortions. All values shown in the different figures are Pearson correlation with significance of  $p < 0.1$  (whereas values with significance of  $p > 0.1$  are crossed out). Figure 4 shows the overall questionnaire correlation throughout all conditions. It is visible that the most values correlate (at least weak) with a negative correlation of distortion. Particularly the rating of the overall experience correlates with the most values, as to be expected.

Figure 5 shows the correlation between questionnaire answers according to a specific condition. It is visible that the correlation between the different ratings changes within the different conditions, particularly how audio and video quality correlates to other ratings. In condition 2 (Figure 5, top right), the video quality both correlates to facial expressions and

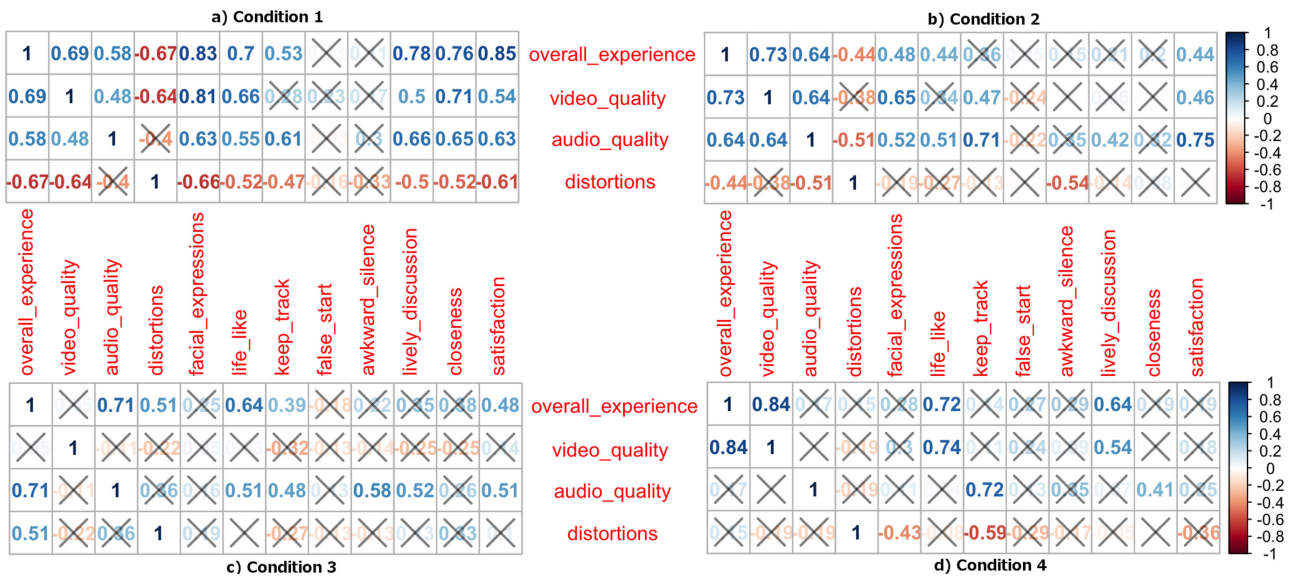


Fig. 5. Condition 1 Questionnaire Correlation. Condition 1 (a, top left), Condition 2 (b, top right), Condition 3 (c, bottom left), Condition 4 (d, bottom right)

keep\_track, but not to closeness or life like. This drastically changes in condition 3 with no significant correlation of video quality to any other rating. However, the audio quality shows many correlations in condition 3. Further, distortions show a positive correlation to the overall experience. This is a strong indication that audio became more important than video under condition 3. This was also indicated by some participants during the interview. Participant 18 said about the condition 3: “I did not have problems with anyone ... I mean in the sense I did not notice it, nobody disappeared ... I had a bad image, but I did not even look at people a lot, there is too many things changing on the screen, that you can really look at anything. But I had no audio visual problems. I could follow the conversation, no problem. I guess the quality of an image was changing.” This suggests that participants were more focused on the audio rather than the video under condition 3.

**RQ1. How is the relationship between individual influencing factors and the QoE?**

With this study we cannot fully answer that question, particularly because of the problem described earlier that participants have trouble to distinguish between different influencing factors towards the overall experience. Over all conditions the QoE seems to be rated very similar. However, what we can say is that the conditions have very different correlations. In this way, the relevance of different factors varies between conditions, in particular with regard to audio and video quality ratings.

**C. Overall ranking**

The overall ranking of conditions is shown in Table III. Our participants ranked condition 1 and 2 worst and condition 3 and 4 best. This however does not match the general feedback the participants gave in the interviews, as there were many complaints regarding distortions and confusing switches in condition 3 and 1. [Participant 8 about round 3] “This is annoying, I changed between modes, but in that round it was easy, we finished very fast.”

TABLE II. QUESTIONNAIRE OVERALL RANKING OR CONDITIONS

Condition	Number of Best	Number of Worst	Expert vote Best	Expert vote Worst
C1	3	7	3	5
C2	3	7	2	3
C3	7	4	3	4
C4	7	2	6	2

To get a better overview of the different answers between users we clustered our participants according to their expertise in using video conferencing solution. This resulted in 6 non-experts and 14 experts. The ranking of the “experts” (shown in Table III) is much more balanced, with a clear preference on condition 4. This was also reflected in the interviews, where a majority of people answered to prefer the fixed view (C2 + C4, Figure 2) over the Focus+Context view (C1, Figure 1) as “it’s just too disturbing and I cannot follow the conversation ... it’s too much” [Participant 18].

**RQ3. Can we improve the QoE by choosing a particular layout and stream configuration based on the network limitations?**

As there is a clear difference in the correlation of user ratings in the different conditions (hence the importance of different properties to the overall experience and satisfaction is different) and we have different patterns in the distortions and error rate, it is clear that the perception and dynamics of how properties influence each other is different under the different conditions. Overall due to the users ranking and interviews we can confirm a stronger tendency of users towards the tiled layout (C2 + C4, Figure 2).

VI. CONCLUSION

The rating of the different conditions in the experiment is very similar. This however does not mean that the conversation behaviour and perception is actually the same in all conditions. The reason for the similar rating (even though a difference in perception) is that it is difficult for participants to distinguish

between the overall conversation dynamics and the different other factors (like video and audio quality) that influence the perception and QoE, as stated by participants in the interviews. Furthermore, our data shows that the correlations between different influencing factors are different in the different conditions. This means the overall experience is similar (after all we have the same group of people and task) but the way different properties influenced the final QoE is different. Particularly, which role audio and video plays towards the conversation. For example it is clear that in c3 the audio is taking over (strong correlation to overall quality and correlation towards the satisfaction) where as video does not play a role anymore (no significant correlation at all). This is also the reason why condition 3 was ranked as best condition equally as condition 4, as our conversation task does not strongly depend on visual cues. The “expert” ranking also confirms this with a clearer preference on condition 4. Condition 4 offers more visual cues that are beneficial for the conversation (as opposed to condition 3). This is indicated by the correlation in condition 4 between video quality, lifelike and the liveliness of the discussion (Figure 5, bottom right).

Overall, our findings indicate that when it comes to configuring streams and layouts the stability of the system and a balanced stream quality, as well as, a balanced visual layout seems most satisfying to the users. This is given our conversation task and network limitations. Furthermore, this can only be done by combining the knowledge of the group conversation and network conditions. Ultimately, this finding put in practice can lead to improvements in the design of video conferencing systems and optimization strategies.

#### ACKNOWLEDGMENT

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. ICT-2011-287760 (Vconnect project). Thanks goes to all partners of the Vconnect project for their valuable input. Special thanks go to Yaroslav Kryvyi for his technical support.

#### REFERENCES

- [1] Johanson, M., Jalminger, J., Laulajainen, J.-P. and Bür, K. Audiovisual Network Service Optimization by Quality of Experience Estimation. NEM Summit 2012
- [2] Khan, S., Duhovnikov, S., Steinbach, E. and Kellerer, W. 2007. MOS-Based Multiuser Multiapplication Cross-Layer Optimization for Mobile Multimedia Communication. *Advances in Multimedia*. 2007, 1–11.
- [3] Venkataraman, M. and Chatterjee, M. 2011. Inferring video QoE in real time. *Network*, IEEE. 25, 1 (2011), 4–13.
- [4] P. 1301. ITU-T RECOMMENDATION, “ITU-P.1301 - Subjective quality evaluation of audio and audiovisual multiparty telemeetings”. 27-feb-2013.
- [5] M. G. Ames, J. Go, J. J. Kaye, and M. Spasojevic, “Making love in the network closet: the benefits and work of family videochat”, in *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, 2010, pp. 145–154.
- [6] M. Ursu, P. Torres, V. Zsombori, M. Frantzis, and R. Kaiser, “Socialising through orchestrated video communication”, in *Proceedings of the 19th ACM international conference on Multimedia*, 2011, pp. 981–984.
- [7] 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.
- [8] Mehmood, M.A., Wundsam, A., Uhlrig, S., Levin, D., Sarrar, N. and Feldmann, A. 2012. QoE-Lab: Towards evaluating Quality of Experience for Future Internet Conditions. *Testbeds and Research Infrastructure. Development of Networks and Communities*. (2012), 286–301.
- [9] K. Schoenenberg, A. Raake, S. Egger, and R. Schatz, “On interaction behaviour in telephone conversations under transmission delay”, *Speech Commun.*, vol. 63–64, pp. 1–14, sep. 2014.
- [10] M. Schmitt, S. Gunkel, C. Pablo, and D. Bulterman, “Asymmetric Delay in Video-Mediated Group Discussions”, *Proceedings of in 6th International Workshop on Quality of Multimedia Experience (QoMEX)*, 2014, 2014.
- [11] M. Falelakis, M. Groen, M. Frantzis, R. Kaiser, and M. Ursu, “Automatic orchestration of video streams to enhance group communication”, in *Proceedings of the 2012 international workshop on Socially-aware multimedia*, New York, NY, USA, 2012, pp. 25–30.
- [12] D. T. Nguyen and J. Canny, “Multiview: Improving Trust in Group Video Conferencing Through Spatial Faithfulness”, in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, New York, NY, USA, 2007, pp. 1465–1474.
- [13] R. Vertegaal, I. Weevers, C. Sohn, and C. Cheung, “GAZE-2: conveying eye contact in group video conferencing using eye-controlled camera direction”, in *Proceedings of the SIGCHI conference on Human factors in computing systems*, 2003, pp. 521–528.
- [14] S. Junuzovic, K. Inkpen, R. Hegde, and Z. Zhang, “Towards ideal window layouts for multi-party, gaze-aware desktop videoconferencing”, in *Proceedings of Graphics Interface 2011*, School of Computer Science, University of Waterloo, Waterloo, Ontario, Canada, 2011, pp. 119–126.
- [15] M. Venkataraman and M. Chatterjee, “Effects of internet path selection on video-QoE”, in *Proceedings of the second annual ACM conference on Multimedia systems*, New York, NY, USA, 2011, pp. 45–56.
- [16] M. Fiedler, T. Hossfeld, and P. Tran-Gia, “A generic quantitative relationship between quality of experience and quality of service”, *Netw. IEEE*, vol. 24, nr. 2, pp. 36–41, 2010.
- [17] W. Jiang and H. Schulzrinne, “Comparison and optimization of packet loss repair methods on VoIP perceived quality under bursty loss”, in *Proceedings of the 12th international workshop on Network and operating systems support for digital audio and video*, 2002, pp. 73–81.
- [18] M. Mu, R. Gostner, A. Mauthe, G. Tyson, and F. Garcia, “Visibility of individual packet loss on H. 264 encoded video stream: a user study on the impact of packet loss on perceived video quality”, in *IS&T/SPIE Electronic Imaging*, 2009, pp. 725302–725302.
- [19] J. Xu and B. W. Wah, “Exploiting just-noticeable difference of delays for improving quality of experience in video conferencing”, in *Proceedings of the 4th ACM Multimedia Systems Conference*, New York, NY, USA, 2013, pp. 238–248.
- [20] P. Calyam, M. Haffner, E. Ekici, and C.-G. Lee, “Measuring Interaction QoE in Internet Videoconferencing”, in *Real-Time Mobile Multimedia Services*, D. Krishnaswamy, T. Pfeifer, and D. Raz, Red. Springer Berlin Heidelberg, 2007, pp. 14–25.
- [21] Lu, Y., Zhao, Y., Kuipers, F. and Van Mieghem, P. 2010. Measurement study of multi-party video conferencing. *NETWORKING* 2010. 96–108.
- [22] Xu, Y., Yu, C., Li, J. and Liu, Y. 2012. Video telephony for end-consumers: measurement study of Google+, iChat, and Skype. *Proceedings of the 2012 ACM conference on Internet measurement conference* (New York, NY, USA), 371–384.
- [23] S. Gunkel, J. Jansen, I. Kegel, D. Bulterman, “The Optimiser: monitoring and improving switching delays in video conferencing”, *Proceedings of Workshop on Mobile Video Delivery*, 2014

©2015 IEEE. Reprinted, with permission, from Gunkel, S.N.B.; Schmitt, M.; Cesar, P., "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*, vol., no., pp.1-6, 26-29 May 2015, doi: 10.1109/QoMEX.2015.7148085

<http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7148085&isnumber=7148077>