Diving Into Perceptual Space: Quality Relevant Dimensions for Video Telephony

Falk Schiffner & Sebastian Möller
Quality and Usability Lab, Technische Universität Berlin
Email: [falk.schiffner, moeller] @ tu-berlin.de

Abstract—**This paper reports a study to unveil the quality relevant perceptual space of video degradations in the domain of video telephony. The perceptual space was explored using a *Semantic Differential* (SD) test paradigm with a subsequent *Principal Component Analysis* (PCA). This paper provides a view on the test itself as well as on the analysis of the results.**

Keywords—*quality of experience; video quality; video telephony*

I. Introduction

Quality of Experience (QoE) of telecommunication services, such as video telephony, represents a crucial subjective evaluation from the user's perspective. Since audio-visual communication services are used broadly [1], service providers are constandly interested in improving and monitoring their services. For the further understanding of the user's experience and rating of video quality, more information is needed. It is known that the formation processes of quality rating consists of multiple factors [2] that lead to a multidimensional feature space. These features can be seen as the quality relevant perceptual dimensions. This is also true in the case of video telephony. The presented study is a further step towards the implementation of a perceptual-based predictor for video quality. This is supposed to lead to a better prediction of the perceived video quality and a deeper understanding of the subjective video quality judgment process.

II. Related Work

While the quality relevant perceptional space for speech telephony has already been mapped (e. g. [3]), only little work was done for video (e. g. in the domain of IPTV, Tucker [4]). Tucker obtained three preceptual dimensions (*Fragmentation, Movement Disturbance and Frequency Content*). In the domain of IPTV, the user consumes streamed video content passively, whereas in the domain of video telephony the user is much more involved in interaction. Besides that, the video material is less diverse than in the IPTV context. In most cases the video shows a classical *head-and-shoulder* scene (see Figure 1). The requirements of video telephony in comparison to IPTV are therefore different. However, the work conducted by Tucker was chosen as a starting point to investigate the quality relevant space.

III.Experiment

Test Material: A set of video files were prepared with the aim to cover a large range of potential video degradations of video telephony. The degradations were selected on the basis of an expert survey and only focused on potential transmission degradations. To include degradations into the video, the *Reference Impairment System for Video* (RISV [5]) was employed. The RISV is an adjustable system that can be used to create reference conditons and produces video degradations that can occur in digital video systems. In addition, effects of coding and packet loss were also included. 2-pass coding was applied, to ensure the same level of degradation over the whole sample. For more details on the test material see Table I and II. The degradations were processed via *MATLAB, ffmpeg, NetEm*, and *Traffic Control* (TC). Examples of the test material are shown in Figure 1.

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

TABLE I: Overview – Test Material

Material	Video Telephony / Head-and-Shoulder Scene
Number of Files	30 Files – 4 Persons (2 male / 2 female)
Resolution	$640\,x\,480$ Pixel
Frame Rate	$25\,fps$
Screen Size (diagonal)	$18,5\,cm$ ($7,3\,inch$)
Viewing Distance	ca. $60\,cm$

TABLE II: Description of the Impairments in the Test Material.

Name	Description
Reference	Unimpaired Material
RISV Artifical Blockiness 5x5 / 8x8	All Frames (2 Block Sizes Settings)
RISV Artifical Blurring ITU(F1) / Filter7	All Frames (2 Filter Settings)
RISV Artifical Jerkiness 6 / 11 Frames	Jerkiness (6 resp. 11 Frames holded)
RISV Artifical NoiseQ 3% / 15%	Salt & Pepper Noise (x% Pixel/Frame)
H264 Bitrate 28 / $56kbps$	H.264-Codec Bitrate (2-pass Coding)
Packet Loss 0.5% / 1.5%	H.264-Codec, TC, NetEm
Luminance Impairment I (darker)	Luminance Value reduced
Luminance Impairment II (lighter)	Luminance Value raised

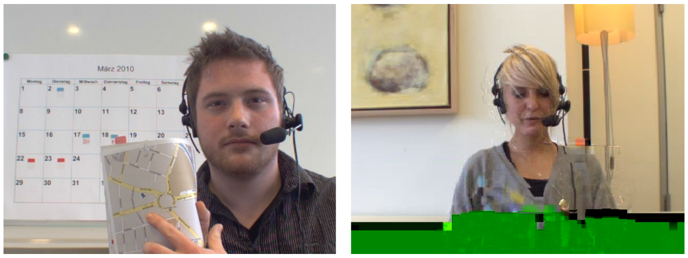

Fig. 1: Examples of the Video Material – left: unimpaired Reference; right: impaired Sample (Packet Loss $1.5\,\%$).

Test Participants and Procedure: For this study, 23 participants were recruited. The group consists of 11 female and 12 male participants. The average age was 31.4 years ($\sigma = 6.7$). Every participant was subjected to a vision test (*Ishihara-Test, Snellen Table*) before the experminent to check for normal eyesight.

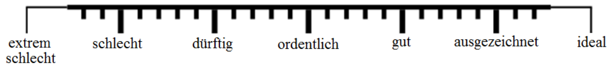

Fig. 2: 7-point continuous scale with German labels left-to-right (corresponding values in brackets): extremely bad (1), bad (2), poor (3), fair (4), good (5), excellent (6) and ideal (7).

The duration of the experiment was between $40\,min - 60\,min$. The participants were allowed to take a $5\,min$ break, if needed. In the beginning, a small training was placed to allow the participants to get familiar with the rating task and the degradations. The first task was to rate the overall quality of the video samples via a 7-point continuous scale (Figure 2) [6]. The second task was to describe the video via a *SD*. Through this method one can unveil which attribute is more pronounced in the sample. A set of 17 antonym pairs used to get a polarity profile for each degradation. These pairs were obtained through severval pre-tests and based on Tucker [4]. Between one antonym pair was a discrete 7 step scale, where the participants have to weight which one of the two words discribes the sample best. The columns 2 and 3 in Table III show the list of antonym pairs.

IV. Data Analysis

The analysis of the overall quality rating (in terms of *Mean Opinion Score* (MOS)) revealed that the bigger the degradation, the smaller the perceived quality (see Figure 3). These MOS compared to the MOS of a previous study [7] show a very strong correlation ($r = .97$). Therefore, we consider the overall quality rating to be stable. The scores of the antonym pairs obtaind in the SD were analyzed and a PCA was conducted. The rotation method was VARIMAX with Kaiser normalisation. This reveals 4 components with Eigenvalues above 1 (see column $4 - 7$ in Table III). To interpret which antonym pairs loads on which component, the authors only take factor loadings above 0.7 into account. Component 1 is loaded by the 7 antonym pairs (1, 2, 4, 12, 13, 14, 15) and explains 56.7% of the variance. We label the component *Unclearness* since the describing antonym pairs are related to an unclear video image. This dimension has temporal and spatial aspects. Component 2 explains 17.7% of the variance and is loaded by pairs 5, 6, 8, 10, 16. The antonym pairs describing impairments are related to a broken and incomplete video. We label the component *Incompleteness*. This dimension seems to have more temporal aspects. The component 3 is loaded by pairs 9, 17 and explains 8.9% of the variance. The label for that is *Noisiness*. It seems only to be related to the inserted noise. The component 4 is loaded only by pair 7 and is labeled *Luminosity*. It can be linked to both luminance impairments and explains 6.0% of the variance. It is not possible to clearly distinguish between *spatial* and *temporal* dimensions from the data.

TABLE III: Column $2, 3$ show the Antonym Pairs – Used in the SD-Test (only English Translation); Column $4 - 7$ show the Factor Loading of the Antonym Pairs on the Components (CP) with Eigenvalues above 1 (last row).

	Adjective	Antonym	CP 1	CP 2	CP 3	CP 4
1	pixely	uniform	.70			
2	daubed	not daubed	.84			
3	shredded	not shredded				
4	high contrast	low contrast	−.86			
5	dismembered	not dismembered		.97		
6	jerking	constant		.96		
7	overexposed	underexposed				.99
8	blocky	not blocky		.76		
9	flickery	not flickery			.88	
10	blurred movement	sharp movement		.70		
11	overlapped	not overlapped				
12	color distorted	color correct	.80			
13	stripy	not stripy	.70			
14	blurred	sharp	.87			
15	artifical	natural	.72			
16	waggly	stable		.90		
17	noisy	noiseless			.96	
		Eigenvalues:	9.64	3.01	1.52	1.02

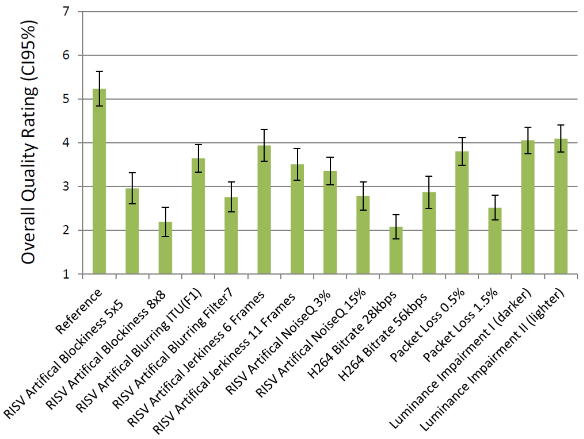

Fig. 3: Overal Quality Rating with Confidence Interval (CI95 %).

V. Conclusion and future work

This study unveils that there are 4 relevant quality dimension for video in the context of video telephony services. It is planed to further explore the perceptual space to verify and deeper explain the findings. In future, a perceptual-based video quality predictor will be developed. It will be combined with a speech quality estimator, to predict audio-visual quality for video telephony.

References

[1] B. Belmudez and S. Möller, *Audiovisual quality integration for interactive communications*. EURASIP Journal on Audio, Speech, and Music Processing, Nov,2013:24.

[2] S. Möller and A. Raake, Eds., *Quality of Experience: Advanced Concepts, Applications and Methods*, 1st ed. Heidelberg, Germany: Springer, 2014, chapter 5.

[3] M. Wältermann, *Dimension-based Quality Modeling of Transmitted Speech*. Springer-Verlag, D-Berlin, 2013.

[4] I. Tucker, "*Perceptual Video Quality Dimensions*," Master Thesis, 2011, Technische Universität, D-Berlin.

[5] ITU-T Rec. P.930, *Principles of a reference impairment system for video*, International Telecommunication Union, CH-Geneva, 04/1996.

[6] S. Möller, *Quality Engineering - Qualität kommunikationstechnischer Systeme*. Springer-Verlag, Heidelberg, 2010.

[7] F. Schiffner and S. Möller, *Audio-Visuelle Qualität: Zum Einfluss des Audiokanals auf die Videoqualitäts- und Gesamtqualitätsbewertung*. DAGA, D-Aachen, 2016.